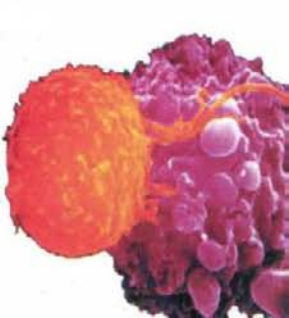
MICROBIOLOGY AND IMMUNOLOGY

PROMILA PARIHAR



Microbiology
&
Immunology

"This page is Intentionally Left Blank"

Microbiology & Immunology

Promila Parihar

SWASTIK PUBLISHERS & DISTRIBUTORS
DELHI - 110 094 (INDIA)

MICROBIOLOGY & IMMUNOLOGY

First Published 2009

ISBN 978-81-89981-30-3

Published in India by

SWASTIK PUBLISHERS & DISTRIBUTORS
31, Gali No. 1, A-Block, Pocket-5,
CRP Water Tank, Sonia Vihar
Delhi-110094 (INDIA)
email: swastik_books@yahoo.com

Printed at: Deepak Offset Press, Delhi.

PREFACE

This book is a concise review of the medically important aspects of microbiology and immunology. It covers both the basic and clinical aspects. Microorganisms have proved to be much more important to mankind than what could be thought of them earlier. During past few years application of microorganisms to human life have enhanced tremendously. Keeping in view such developments, the present project has been taken. The author of this book has taught the subject successfully for many years, and his experience and infectious enthusiasm for current ideas is clearly evident. He has chosen those aspects of the subject which are essential for the student to understand, and arranged them in such a way that he can use them as pegs on which to hang more information. In order to achieve this, most of the instructions given in larger text books have been condensed and presented concisely in simple graphs and diagrams. The student can thus acquire an easily assimilated understanding of basic of the subject which provides a firm foundation on which to build his further studies.

The entire information is gathered under fourteen main headings and sub-grouped in logical and scientific way. Packed with authentic facts and written in simple language, this has become most comprehensive.

The author expresses his thanks to all those friends, colleagues, and research scholars whose continuous inspirations have initiated him to bring this title.

The author wishes to thank the M/s. Swastik Publishers, Delhi for bringing out this book.

Constructive criticisms and suggestions for improvement of the book will be thankfully acknowledged.

Author

"This page is Intentionally Left Blank"

Contents

"This page is Intentionally Left Blank"

1

Introduction

Deterioration of soils and aquifers has become evident in the last few years as a result of inappropriate final disposal procedures for all sorts of waste materials. Oil exploitation, uncontrolled fuel spills, poor practices for final disposal of industrial wastes, overuse of pesticides, and operation of sanitary landfills are some causes of pollution of soils and groundwaters.

In the course of the last two decades a wide variety of technologies has been developed for clean-up operations of contaminated soils and aquifers. They can be classified in terms of their principle of operation: physicochemical, thermal and biological. Among the biological technologies bioremediation has evolved as the most promising one because of its economical, safety and environmental features since organic contaminants become actually transformed, and some of them are fully mineralized.

The success of bioremediation techniques is directly related to the metabolic capability of involved microorganisms and can be affected by the surrounding microenvironment. This chapter describes the mechanism of bioremediation; an overview is presented of its virtues and weaknesses.

1.1 THE SOIL AND MICRO-ORGANISMS

Soil and subsoil constitute a non-renewable natural resource that plays different roles, as described by Aguilar:

1. Filtering medium during aquifer recharge
2. Protective layer of aquifers
3. Scenario of biogeochemical, hydrologic and food chain processes
4. Natural habitat for biodiversity
5. Space for agricultural and cattle-breeding activities

6. Space for green areas to serve as sources for oxygen regeneration
7. Physical foundation for building construction
8. Sanctuary of the cultural reserve

The first four functions are the most important for the subject of this chapter, and treated as a whole they refer to what is commonly known as 'buffer capacity' of the soil. This buffering phenomenon constantly happens at recharge zones where rainfall migrates vertically downwards towards the aquifers. During seepage a large proportion of solid materials carried by the water are retained at the shallow layers and only water-borne dissolved chemicals seep downwards. Migration time depends on particle size distribution of soil; movement is faster through a fractured medium than through a granular geological material.

Seepage through the subsoil is delayed for compounds that do not migrate at the same rate as water. This frequently occurs in soils with a high organic matter content, such as clays in which organic compounds tend to be retained. There is also another important aspect related to particle size of soils. Clays are characterized by particles of small size (<2 μm); migration is therefore slower and the contact period among exogenic organic compounds and organic matter in the soil is longer, thus favouring the development of the sorption phenomenon. There the wide diversity of heterotrophic microorganisms involved in matter recycling start exerting their metabolic activity by using the existing organic compounds as carbon sources. The .longer the substrate/microorganism contact time, the higher the possibility for degradation of the organic matter. The sorption and degradation phenomena of organics in the shallower geological material make it possible for water that continues its migration towards the aquifers to become free from exogenic compounds.

As a result of industrial activities spills commonly occur and water-insoluble organic contaminants seep into the soil. These compounds, named NAPLs (non-aqueous phase liquids), have been classified into two types: those lighter than water (known as LNAPLs) and those denser than water (DNAPLs). Typical LNAPLs are petroleum hydrocarbons, their combustible products (such as gasoline, diesel and jet-fuel), benzene, toluene, ethylbenzene and xylenes (BTEX) used as industrial solvents. Other chlorinated industrial solvents, such as tetrachlorethylene, trichlorethylene, chloroform, carbon tetrachloride and methylene chloride, are examples of DNAPLs.

It is important to take into account the classification of contaminants

in terms of their density because when LNAPLs reach the water table they tend to float on the water and to spread radially, whereas the DNAPLs continue their downwards path until bedrock is reached to arrest their movement. In terms of contamination effects, DNAPLs are more hazardous because they are capable of polluting the whole aquifer.

1.2 SURVIVAL OF MICRO-ORGANISMS IN ADVERSE CONDITIONS

Microorganisms possess wide biochemical versatility, which enables them to readily adapt to different microenvironmental conditions of pH, temperature and pressure, even extreme variations. The presence of high pollutant concentrations can be also regarded as extreme conditions. In these cases, contaminants induce a toxic effect on microbial activity to such a degree that the vital functions are inhibited. Microorganisms are capable of developing a certain tolerance to these adverse conditions and to become energy yielding for survival purposes. However, this happens only when microorganisms have the genetic information available, or if they are capable of developing it, to allow the synthesis of enzymes that participate in the transformation of contaminants.

The organic-type contaminants can be used as carbon sources, thus achieving a redution in their concentration and quite probably complete mineralization, i.e. degradation reaches the generation of carbon dioxide. This is highly desirable, although the presence of a cosubstrate might be required to support the energy-yielding activity while biotransformation of contaminants is achieved in different, and hopefully less toxic, chemical entities.

Inorganic contaminants can only be transformed into different molecular entities; some are retained by the cells without becoming degraded. Reduction in the concentration of inorganic contaminants can be observed only when the microbial activity occurs in water where compounds move from the aqueous phase to the inside of the cells.

When reference is made to biodegradation, rather than simply referring to the microorganisms, it is advisable to consider the enzymes which act as catalysts of the transformation reactions induced by the energy-yielding process. For this mechanism to occur the presence of an electron donor and of an electron acceptor is required as well as microenvironmental conditions suitable for synthesis and for the expression of catabolic enzymes. In the case of heterotrophic microorganisms, the

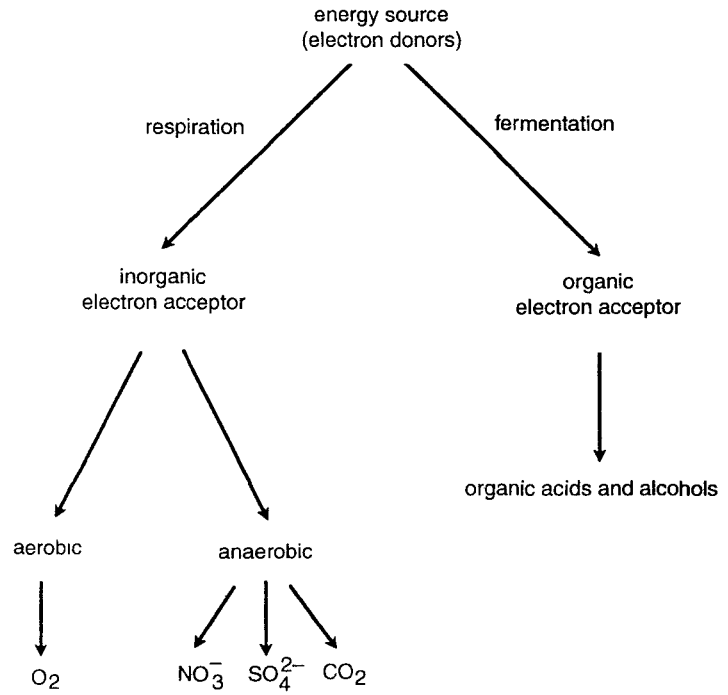


Figure 1.1 : Route of metabolism depending on electron acceptor

electron donor will be the compound used as a source of carbon and energy, most likely the organic contaminant. Its role is to supply energy required by metabolism through electron transfer during the oxidation-reduction reactions to the electron acceptors at the completion of the energyyielding cycle.

Two types of metabolism exist, depending on the type of electron acceptor: if it has an organic origin, fermentation occurs; for inorganic compounds, the process will be respiration. In turn, there are two kinds of respiration: aerobic, when the molecular oxygen becomes the electron acceptor; and anaerobic, when oxidized inorganic compounds such as nitrates, sulphates or carbon dioxide are used. This is why denitrification, sulphate reduction or methanogenesis processes become available.

Traditionally, the initial pathway through which microorganisms start the energyyielding process is glycolysis, also known as the Emden-Meyerhof-Pamas pathway, which carries glucose or other sugars to an

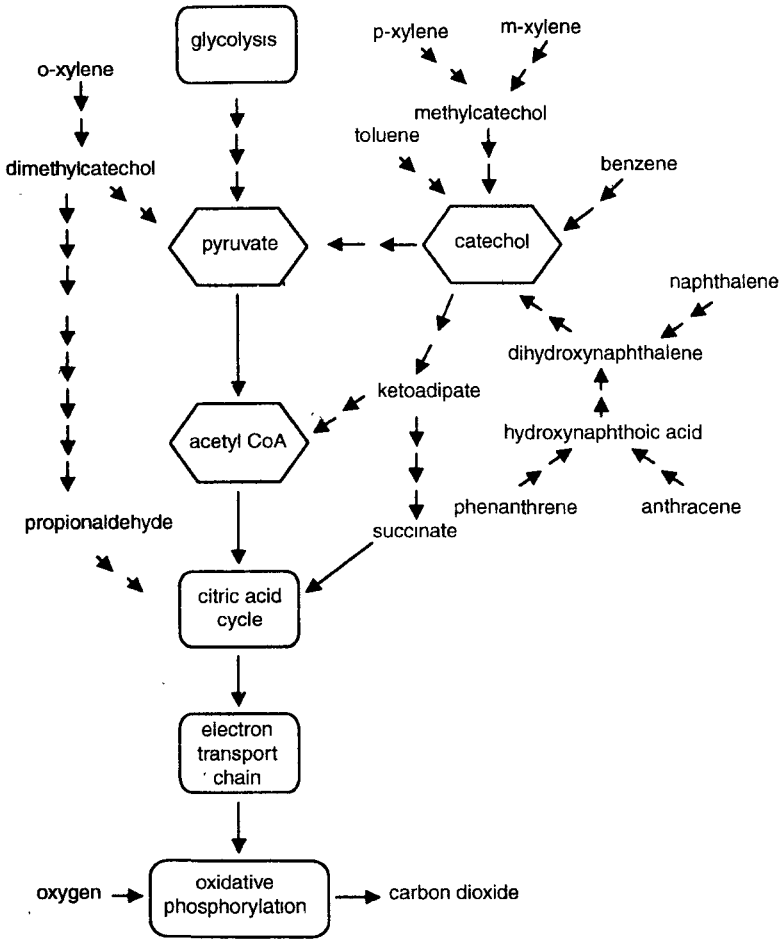


Figure 1.2 : Integration of degradative pathways for some contaminants

intermediary such as pyruvate. Other compounds with different chemical characteristics (such as amino acids and fatty acids) have different degradation pathways, but all arrive at the same intermediary (acetyl CoA), from which the pathway to follow can be defined. When the respiration pathway is aerobic, three more degradative pathways are then followed: the citric acid cycle, electron transport chain, and oxidative phosphorylation. The presence of molecular oxygen is imperative for the last, which constitutes the most important mechanism for energy supply to the cellular activity; carbon dioxide is generated from the reaction. If this happens, it is assumed that the organic

contaminants have become fully mineralized.

For degradation of organic compounds, the presence of enzymes is necessary; however, these are synthesized only when the cells have the specific genetic information and the substrate with which they interact are present in the required concentrations. As an example, mention can be made of the degradation of monoaromatic compounds indicative of contamination, such as gasoline, benzene, toluene and xylene isomers. An intermediary catechol is formed during degradation, which can be degraded through fermentation or respiration, by means of which it can reach other intermediaries of the classic pathway that leads to the generation of carbon dioxide, such as pyruvate or succinate. The integration of degradative pathways for some contaminants with the traditional pathway of glycolysis to oxidative phosphorylation.

Degradation of polyaromatic hydrocarbons (PAHs) such as phenanthrene and anthracene, which are indicators of contamination by diesel. These molecular entities are formed by three aromatic rings which become transformed through several enzymatic reactions to produce catechol. Naphthalene, which has only two aromatic rings, is also transformed into catechol and processed by the mineralization pathway.

Actually, most contaminants are found as very complex combinations; for example, gasoline and diesel contain more than 120 different chemical entities. This situation demands the presence of microbial consortia, i.e. mixed cultures of species that can co-exist without harming each other or which otherwise mutually help each other for the degradation of contaminants. Those native microorganisms which have survived the adverse effects of contamination play a leading role in the biodegradation of contaminants and represent the backbone of bioremediation.

1.3 ADVANTAGES OF BIOREMEDIATION

Bioremediation is a versatile process because it can be adapted to suit the specific needs of each site. Biostimulation can be applied, but only when the addition of nutrients is necessary; bioaugmentation is used when the proportion of degradative microbial flora to contaminant needs to be increased; bioventing is needed when it becomes necessary to supply oxygen from air. Futhermore, bioremediation can be performed off-site when contamination is superficial, but it will have to be *in situ* when contaminants have reached the saturated zone.

One important feature of bioremediation is its low cost compared

with other treatment technologies. According to Alper, bioremediation is at least six times cheaper than incineration, and three times cheaper than confinement. It should however be mentioned that all cost comparisons cannot be generalized because they are only applicable to each particular case.

1.4 CONTAMINATED SITE

Not all contaminated sites are suitable for treatment with bioremediation techniques; it will be necessary to demonstrate their efficacy, reliability and predictability in advance. To this objective site characterization should be performed to obtain information about three closely related aspects: the chemical nature of contamination, the geohydrochemical properties, and the biodegradation potential for the site.

1.4.1 Pollutant Characterization

It will be necessary to determine the composition, concentration, toxicity, bioavailability, solubility, sorption and volatilization of all pollutants.

1.4.2 Geohydrochemical Characterization

The physical and chemical properties of the geological material should be determined to be able to learn if the microenvironment is suitable for the biodegradative activity. In addition, the geohydrological conditions of the site as well as direction and velocity direction of underground flow are of fundamental interest, particularly when contamination has reached the water table.

1.4.3 Microbiological Characterization

It is convenient to analyse the microbial flora in respect to degradative capacity and to the size of the native population with degradative potential.

Integration of the physicochemical and microbiological characterizations should correspond to the results of biofeasibility tests, from which it should be determined whether or not a certain biological treatment is applicable.

Once the site characterization is completed, it is important to proceed almost immediately with the activities leading to its clean-up because contaminants are not static. This is particularly true when pollutants are found in an aquifer.

The characterization of a contaminated site is of utmost importance because a better knowledge of it will facilitate the outlining of an *ad hoc* strategy for its bioremediation. The characterization should be

performed in a logical sequence according to a previously established programme, to be able to respond to questions such:

- What chemical compounds are found as contaminants?
- Is the contamination superficial or has it affected the subsoil?
- Are there any records to prove that the contaminants are biodegradable?
- What is the depth and extension of the contaminant plume?
- What is the depth to the water table?
- Is the permeability of the geological material high or low?
- Are there microorganisms capable of degrading the contaminants?
- Is the environment suitable for microbial activity?
- Is it possible to 'build' a bioreactor at the site to be treated?

If answers are affirmative, then bioremediation could be applied, then it will be necessary to carry out biotreatability studies and the evaluation of a bench-scale or pilot-scale project, from which the full-scale process operation will finally be developed.

Something that is commonly encountered in practice is free contaminant in the aquifer; this must be removed before a bioremediation process is applied because the contaminants are toxic to the microorganisms. The latter are capable of tolerating certain concentrations, and some species show a higher tolerance than others, but it is not definitively possible for microorganisms to develop within pure pollutants. This sort of detail should be taken into account when scale-up of the process is being outlined; otherwise a complete failure of bioremediation can be expected.

1.5 BIOTREATABILITY TESTS

Once information has been gathered on the characteristics of the contaminated site it will be possible to identify its specific requirements. The fact of detecting contaminants at ground surface, in the aquifer or at the mid part of the unsaturated zone, will suggest a strategy of specific bioremediation for each particular case. Therefore, the biotreatability testing procedure will have to be suited to each specific site.

Biotreatability tests are generally performed at a mesocosm level, trying to maintain the environmental conditions that will prevail during treatment in the field. If shallow strata are to be treated with an off-site process, large trays or jars to hold several pounds of soil could be used; it will therefore be necessary to keep humidity and homogeneity

constant so that the microbial activity takes place in the whole volume of soil to be treated. If contamination is detected at the water table, columns packed with contaminated soils will be the preferred experimental model for biotreatability tests. It will be necessary in this case to determine the groundwater flow rate to be able to define the operating mechanism of the columns.

When characterization studies indicate that the microbial population with degradative potential is limited or practically zero, it will become important to add exogenous microorganisms. For very practical cases such as bioremediation two alternatives exist. The most common solution is to add commercial compounds; this way is more accessible and faster but it has the least likelihood of success. The other more, interesting, solution is to isolate the rather few degradative microorganisms that were obtained during site characterization and to promote their growth to obtain a culture that can be used for inoculation purposes. This method is safer although it has the shortcoming of requiring a longer time to increase the microbial biomass.

The purpose of biotreatability studies is to predict the behavior of the process and to determine the nutritional requirements for microorganisms to perform biodegradation. The following measurements are thus required:

- Oxygen consumption
- Carbon dioxide generated
- Exhaustion of added nutrients (particularly nitrogen and phosphate sources)
- Contaminant removal

It is necessary to include biotic and abiotic controls to ensure that the contaminant is removed by a microbial activity. When tests are properly carried out it will be possible to predict the behaviour of bioremediation and the time for large-scale application.

For the biotreatability tests to actually represent the field conditions it will be necessary to adopt microenvironmental conditions as close as possible to those encountered at the site to be treated; otherwise the benefits obtained will be minimal.

Because biotreatability tests are time consuming and costly, when bioremediation is applied at a commercial scale these tests are not always performed; nutrients are incorporated empirically, based on past experience. The results are eventually satisfactory but most of the applications are bound to become a complete failure. This should be taken into account because a bioremediation failure may lead to further problems.

1.6 FROM LABORATORY TO FIELD

Even though techniques for growing microorganisms at a commercial level have been perfectly established, additional studies on technologies for bioremediation of soils and aquifers to promote, facilitate or expedite microbial activity in the field are needed. The more relevant aspects are related to bioavailability, the concept of bioreactor, the supply of oxygen, and mass transfer.

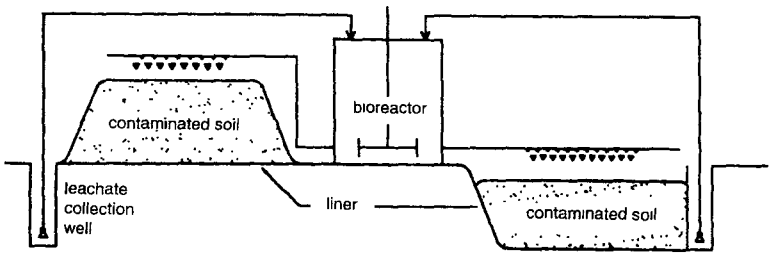


Figure 1.3 Composite diagram for off-site bioremediation techniques: left, a biopile; right, a biocell. Both have collection wells for le4chate recirculation

1.6.1 Bioavailability

For the biotransformation reactions to be carried out it is necessary for the contaminants to be made ready for attack by the microorganisms. Enzymatic reactions generally occur in aqueous solution but when pollutants are insoluble in water - such as in the case of petroleum hydrocarbons - their energy yielding will be slow due to surface tension between the aqueous and the organic phases. To solve this problem special attention has been paid to the production of surfactants that could be incorporated into the medium and improve the bioavailability. A surfactant acts by 'solubilizing' organic contaminants in the aqueous medium, thus making them accessible to the enzymes responsible for their biodegradation. However, the uncontrolled use of surfactants courses many problems; these are referred to later.

1.6.2 The Concept of Bioreactor

As opposed to conventional biotechnological processes, in which previously built reactors are used, it is necessary for bioremediation purposes to 'build' the bioreactor at the contaminated site. When pollution covers shallow soil layers, it is recommended to excavate the material and to carry it somewhere else to build a biopile or a biocell. The difference between these two concepts resides in the fact that in the biopile the contaminated material is piled on the ground surface whereas for the biocell it is necesssary to perform an excavation

in a clean site to deposit the material for further treatment. It will be necessary in both cases to place liners to confine the contaminated material and to prevent leachates from seeping toward the clean soil during treatment.

When contamination has reached the water table, the bioreactor can be built through the bore-holes. These are always drilled in even numbers; half are used for extraction and half for injection. Depth and location of wells is determined from geohydrological characterization of the site, from which the configuration of the contamination plume and the direction of the underground flow can be determined. The number of wells can be chosen once the radius of influence of each well has been established in terms of porosity and permeability of geological material, and the flow rate and direction of underground flow. The operating procedure of the wells determines the dimension and control of the bioreactor.

As it occurs in wastewater treatment procedures, it is not possible for bioremediation to work under sterile conditions. The diversity of the microbial population is set in terms of the kinetic characteristics of each of the species involved; it usually happens that native microorganisms that have developed a biodegradative capability in the same microenvironment where contaminants are present become the most widespread.

1.6.2.1 Oxygen Supply

Although the degradation of contaminant compounds under anaerobic conditions has been reported elsewhere, aerobic metabolism is the preferred choice to apply bioremediation techniques because reactions occur at a faster rate and complete mineralization is achieved.

When compounds used as a source of carbon lack oxygen in their molecule, such as in the case of hydrocarbons, oxygen demands for their biodegradation are higher than those required by oxidized compounds such as sugars.

Oxygen supply is one of the major engineering challenges for bioremediation applications in soils because a solid medium is encountered; in aquifers, it becomes more difficult to solubilize molecular oxygen as the depth of the subsoil increases. This has promoted the design of *ad hoc* aeration equipment and the search for alternatives to supply oxygen through highly oxidized inorganic compounds not related to alternate energy-yielding pathways so that full mineralization is achieved. As examples of these oxidized compounds mention can be made of peroxides.

1.6.2.2 Mass Transfer

The concept of mass transfer in bioremediation basically refers to the homogeneity of the system - i.e. that in all of the points inside the bioreactor the same microenvironmental conditions exist to promote microbial activity. This includes nutrient concentration, humidity, pH, and concentration of oxygen available.

Bioremediation becomes more difficult when the geological material is basically a clay because its low permeability prevents mass transfer in the system. This is important when contamination has reached the water table and an *in-situ* treatment is available. For superficial soils this problem can be overcome if sand or agroindustrial residues are added to increase the permeability.

In contaminated aquifers, the most popular auxiliary technique is the pump-treatinjection that involves the extraction of groundwater, its treatment at the surface, and its subsequent recharging into the aquifer. For bioremediation purposes treatment is carried out in a bioreactor where the degradative microorganisms may be, confined; alternatively, the organisms can be recycled and leave the reactor at the surface to become reactivated. Although this technique is widely used, there are certain aspects that determine its successful aplication; for instance, contaminants can be heavily adsorbed by geological material or may be present in low-permeability zones thus restraining the mass transfer. In other cases, it becomes difficult to reach the clean-up levels required because the low concentrations of contaminants that microorganisms use as a substrate are not sufficient to support their microbial activity and they start to die. If this occurs, treatment becomes very costly because of the power requirements demanded by pumping.

If the previous concepts are taken into account, it will be possible to establish that in bioremediation it is not sufficient to work with ideal cases in the laboratory; it is important to maintain the conditions that prevail at the site and to use suitable experimental models so that the results of the study are representative of the scope intended for the field. This is the true objective of performing biotreatability tests before scaling-up is carried out.

1.7 BIOREMEDIATION MONITORING IN THE FIELD

Monitoring of a bioremediation process is essential to determine two fundamental aspects: the degree of contaminant removal and the catabolic microenvironment. For the former it suffices to determine the residual contaminant concentration by any of the available analytical

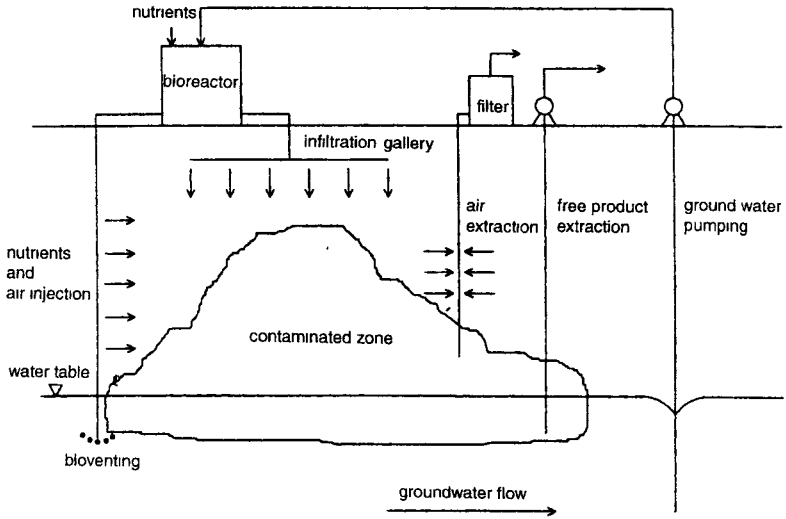


Figure 1.4 : Composite diagram for in-situ bioremediation techniques, including nutrient percolation, vapour and air extraction, bioventing, pump-treat-injection and free product recovery

methods. The second aspect demands the measurement of several parameters such as microbial count of degradative microorganisms, concentration of residual nutrients, pH and humidity (for superficial geological materials), among others.

If the microenvironmental conditions where biodegradation is taking place are periodically determined, it will be possible to maintain each of the parameters within the levels in which the highest metabolic activity can be reached.

1.8 BIOREMEDIATION AS A CLEAN TECHNOLOGY

The global market of bioremediation is becoming increasingly wider and more successful because it is regarded as a clean technology. This is mainly due to the fact that contaminants can be transformed into environmentally harmless compounds and some can be fully mineralized. It is also important because microorganisms die when there are no more pollutants to use as substrate.

Something that favours the image of bioremediation is that the soil, once the treatment is over and the degree of pollution very small, can be used for growing plants with the purpose of reintegrating it to its original biological functions.

On the other hand, it is a well known fact that many of the

technological advances have been accompanied by environmental deterioration; if bioremediation is carelessly handled, it can lead to failures and to even worse environmental disasters. To keep the concept of bioremediation as a clean technology it will be necessary to perform a rigorous analysis based on ethical environmental concepts and on principles of sustainability that in many cases oppose the economic interests of commercial enterprises.

The increasing number of bioremediation companies on a world-wide scale is mainly due to the fact that this technology is economically feasible. These corporations offer not only environmental services but also related consumables that are basically microbial products, nutrients and commonly biodegradable surfactants. If these products are applied only as recommended, in the minimum necessary amounts and under treatment control, the technology will be indeed successful; otherwise, the surroundings of the site being treated can be adversely affected.

Some of the risks involved in a careless application of bioremediation can be described as follows:

- The uncontrolled addition of surfactants to aquifers can help the dispersion of contaminants rather than their full degradation.
- The excessive use of inorganic compounds used as nitrogen and phosphate sources that can be transported to lakes or lagoons favours the growth of undesirable species; an example is provided by eutrophication.
- Where a native microbial population with degradative capabilities exists, it is better to stimulate its activity *in situ* rather than applying exogenous microorganisms that will eventually die from competition in the natural environment.
- An additional aspect is the increasing interest to apply genetically engineered microorganisms (GEMs) to expedite bioremediation. It is convenient in this case to mention that all the mechanisms that govern a natural environment such as an aquifer are not yet fully understood; furthermore, many factors are involved in the stability of the genetic information within the cells.

1.9 MANAGEMENT TECHNOLOGY

A large number of bioremediation technologies are now being developed and successfully implemented in countries that share certain environmental factors. However, when these technologies are transferred to countries with different environmental characteristics, the results

are far from successful. It shall be considered in this respect that successful application of any type. of technology depends on the need to perform studies for its implementation and innovation that could even result in further developments.

In the case of bioremediation technologies it should be understood that every soil has different characteristics and that no general rule exists for the microorganisms to readily adapt to any habitat. The soils in various parts of the world has distinctive physical, chemical and biological characteristics that make them different from each other.

Literature on new bioremediation technologies is being published every day; competition among big companies who have realized that bioremediation is a profitable moneymaking opportunity has also become evident. This competitiveness has been focused on the generation of increasingly efficient, but at the same time more sophisticated and expensive, technologies. The secret of a good bioremediation technology is to suit the know-how to every particular problem.

2

Microbial Fertilizers

The reduction of nitrogen to ammonia gives out energy:

$$3H_2 + N_2 \rightarrow 2NH_3 \qquad \Delta G^\circ = -33\,39 \text{ kJ Mol}$$

Why then do these gases not react together when mixed at normal temperatures and pressures? The answer lies in the great stability of the dinitrogen molecule; both industrial chemists and prokaryotic organisms have to provide energy before the reaction can proceed. Industrially, high temperatures (300-600 °C) and pressures (20-80 MPa) are used (the Haber-Bosch process) : cells couple the reaction to the chemical energy source, adenosine triphosphate (ATP). ATP may be produced by photophosphorylation in photosynthetic cells, by substrate-level phosphorylation in all cells, and by oxidative phosphorylation in all cells with a suitable terminal electron acceptor. Oxygen is the most widespread of these (although some organisms can use other substances such as NO_3^-), and this immediately places many nitrogen-fixing organisms in a dilemma. How can a reductive process such as nitrogen fixation occur in an oxidizing environment, especially when nitrogenase is irreversibly inactivated by oxygen? We shall consider some of the mechanisms adopted to provide sufficient ATP energy to reduce nitrogen, but first we must assess this and other requirements of the nitrogenase enzyme complex.

2.1 ENERGY REQUIREMENTS

Nitrogenase has been purified from all known types of nitrogen-fixing organism, with the exception of the non-legume nodules, although a recent report of an active homogenate from non-legume nodules suggests that this deficiency will soon be made good. In all cases there are two distinct components to the active enzyme complex. One, usually called the Fe protein (mol. wt. about 60 000 Daltons) consists of two apparently identical sub-units together with a cluster of 4 iron and 4 labile sulphur

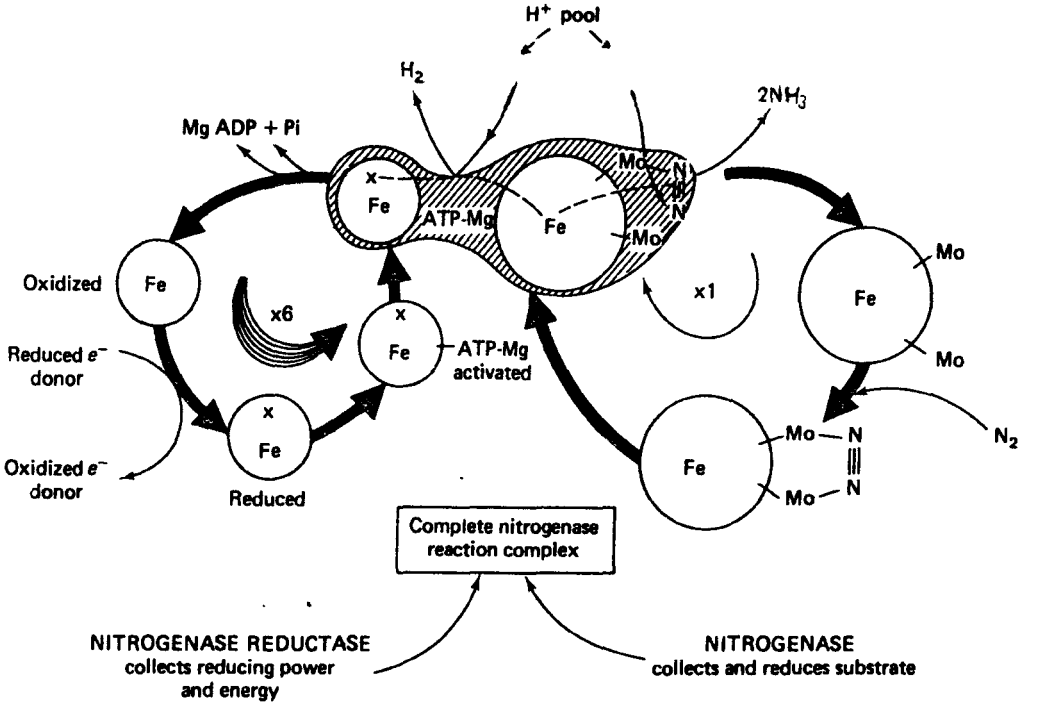


Figure 2.1 : Diagram to show how the two parts of the nitrogenase enzyme may combine to effect the reduction of nitrogen.

atoms.' The whole arrangement is somewhat similar to, but larger than, bacterial ferredoxin. The second component, the Mo-Fe protein (mol. wt. about 200 000 Daltons), consists of 4 sub-units, possibly in 2 slightly different pairs. The detailed structure is less clear than for the Fe protein, with estimates of 1-2 Mo, 12-32 Fe and 24 labile S atoms per complete molecule. Some of the Fe and S atoms are grouped in 4 + 4 clusters, as in the Fe-protein, but others have different structural arrangements. Because the reduction of nitrogen takes place on the Mo-Fe protein, it has been suggested that this component on its own should be called nitrogenase and the Fe-protein called nitrogenase reductase. This terminology will be adopted in the following description of how the reduction of nitrogen to ammonia is currently thought to proceed. A low (about -430 mV) redox potential reductant molecule donates an electron to nitrogenase reductase, which enables it to react with Mg-ATP. Meanwhile, the nitrogen molecule that is to be reduced has combined with the Mo on nitrogenase. The two components now join to form the active enzyme complex. Electrons flow singly from nitrogenase reductase to nitrogenase with the concomitant hydrolysis of 2 ATP molecules. Between the transfer of each electron, the nitrogenase reductase: nitrogenase complex dissociates. Electron flow then proceeds to the nitrogen, which is thought to be reduced in three two-electron steps, thus requiring 6 nitrogenase reductase-Mg-ATP complexes. Protons to complete the formation of the two resultant ammonia molecules are derived from the cellular pool.

The process of reduction of nitrogen to ammonia appears to be only about 80 per cent efficient so that a figure of 15 ATP per molecule of nitrogen reduced maybe more realistic than the 12 accounted for above. In terms of overall cellular energy, the reducing power is equivalent to at least a further 9 ATP per N2-assuming that, if not used for N_2 reduction, the electrons would be available for oxidative phosphorylation. It is possible that ATP energy may actually be used in the production of the very electro-negative electron donor, by a reversal of the usual pathway where electrical energy is used to make ATP. It has similarly been suggested that the reduced nitrogenase reductase uses ATP energy to become 'superreduced' so that it can act as an even stronger reductant.

2.1.1 Hydrogen Production

One way in which the nitrogenase enzyme complex is inefficient is that it permits electrons to by-pass the nitrogen and, instead, to combine directly with protons to form hydrogen gas, which may be evolved. This not only wastes reducing power, but also energy, since ATP is utilized

before the reduction of protons occurs. Burns and Hardy' have likened the nitrogenase complex to a biological battery powering the flow of electrons by hydrolysis of ATP. If nitrogen happens to be around, it becomes reduced, otherwise electrons flow to H^+ and H_2 is evolved. The extent of hydrogen evolution by nitrogen-fixing organisms varies widely. Data for legume and non-legume root nodules are given by Schubert *et al.*, who defined relative efficiency as:

$$\frac{\text{energy used for nitrogen reduction}}{\text{total energy flux available for nitrogenase activity}}$$

Experimentally, this can be calculated as:

$$1 - \frac{H_2 \text{ evolved in air}}{H_2 \text{ evolved in } N_2 \text{ - free air}}$$

In the absence of nitrogen (as, for example, in a gas mixture in which nitrogen is replaced by argon), all the electrons may be available for

TABLE 2.1 Reactions involving hydrogen in nitrogen fixing organisms

1. Hydrogen evolution associated with nitrogenase action. Wasteful of ATP and reductant. Uni-directional.
2. Hydrogen uptake using an irreversible hydrogenase. Found in root nodules and other aerobic nitrogen-fixing systems. May generate ATP and help offset energy and reductant wastage of 1.
3. Reversible hydrogenase in anaerobes, used principally for proton reduction (hydrogen evolution),
   (a) in obligate anaerobes probably constitutive,
   (b) in facultative anaerobes probably induced in anaerobic conditions.

Inhibited by oxygen. May serve to reoxidize NADH, or to remove excess electrons or protons produced in fermentation reactions.

hydrogen production. In a fully efficient system, in the presence of nitrogen, all electrons would be used for ammonia production, no hydrogen would be evolved, and the efficiency index of Schubert and Evans would be 1. Values ranging from 0 99 (cowpea, *Vigna unguicu/ata)* to 0·40 (subterranean clover, *Trifolium subterraneum)* were found. Hydrogen evolution can also have a major effect upon the 'apparent' ATP requirement for nitrogen fixation by free-living species, for example, the anaero *Klebsiella pneumoniae.*

However, the situation may be partly redeemed in aerobic nitrogen fixing systems, many of which possess a hydrogenase, which acts only

in the uptake direction, and which may oxidize some or all of the hydrogen produced by the nitrogenase complex. The exact yield of ATP in this reaction is uncertain, but it is less than the waste incurred during hydrogen evolution: reductant will also be generated. Organisms with an uptake hydrogenase may give no *net* hydrogen evolution, and thus be fully efficient in the Schubert and Evans's sense. However, in overall energetic terms they are less efficient than they would have been if they had *produced* no hydrogen.

To make matters even more complicated, anaerobic nitrogen-fixing organisms may also have a reversible hydrogenase, working mainly in the direction of hydrogen evolution.

2.1.2 Substrate Versatility of Nitrogenase

In addition to nitrogen and protons, nitrogenase can catalyse the reduction of a variety of other substrates, cleaving N—N, N=O and C≡N bonds. One particularly important in experimental work is acetylene, which is reduced to ethylene. This forms the basis for the well-known acetylenereduction assay, and in this and later chapters you will often find nitrogenase activity expressed in moles of ethylene produced. Details of the acetylene-reduction assay and other techniques used for the measurement of nitrogen fixation are to be found in the Appendix. There is evidence that not all of the substrates are bound to nitrogenase at the same site; for example, nitrogen and acetylene may occupy different sites. This is one of several reasons why care should be taken in extrapolating from acetylene reduction to nitrogen fixed, although good correlations are often found." In theory, since two electrons are used to reduce an acetylene to an ethylene molecule, and six are used to reduce a nitrogen molecule to two ammonia molecules, the ratio of acetylene to nitrogen reduced should be 3:1. This, however, takes no account of two facts: first that in the presence of nitrogen, but much less so (if at all) in the presence of acetylene, electrons are diverted to proton reduction; second, that acetylene may inhibit hydrogen uptake. Since net hydrogen evolution varies with organism and conditions, it is clear that frequent checks by independent assays should be made.

2.1.3 Supply of Reductant

It is known that a reductant with low redox potential (about—430 mV) is essential; over the last few years, two types of substance have emerged as the most likely reductants, ferredoxins and flavodoxins. The former have molecular weights varying from 5600 to 24 000 Daltons, and the latter from 14 000 to 22 800 Daltons. Both are

TABLE 2.2 Reductions catalysed by nitrogenase

Substrate	*Product(s)*
N_2	$2NH_3$
N_3^-	N_2, NH_3
N_2O	N_2, H_2O
HCN	CH_4, NH_3, CH_3NH_2
CH_3CN	C_2H_6, NH_3
CH_3NC	CH_3NH_2, CH_4, C_2H_6, C_2H_4, C_3H_8, C_3H_6
CH_2CHCN	C_3H_6, NH_3, C_3H_8
C_2H_2	C_2H_4
$2H^+$	H_2

common in nitrogen-fixing organisms, but whether both donate electrons to nitrogenase is not clear. If we assume that they do, from where do these electrons come? In anaerobic organisms, there is much evidence in favour of pyruvate. *Azotobacter* and blue-green algae probably use NADPH generated from the isocitrate dehydrogenase reaction and photosystem I, respectively. This leaves us with a question mark over the rhizobia. Whether they use NADH, NADPH, or something else is by no means certain. The whole situation with respect to reductants is generally unclear; nitrogenase activity in cell-free extracts with any of the donors thought to act *in vivo* is always less than that using the non-physiological inorganic reductant, dithionite ($S_2O_4^{2-}$), It cannot be ruled out that efficient functioning of the physiological donors requires an intact chemiosmotic membrane, which is disrupted during preparation of cell-free extracts.

2.1.4 General Considerations

In order to synthesize nitrogenase, both iron and molybdenum are necessary. It has been estimated' that 1 kg wet cell paste of *Clostridium pasteurianum* has at least 20 pmol of nitrogenase, which alone requires 20 μmol of Mo and about 400 pmol of Fe. In spite of this, there are few reports of nitrogen fixation being limited by supplies of these ions. On the other hand, supplies of ATP are much more likely to be limiting. Not only does the nitrogenase reaction itself use a considerable quantity of ATP, the enzyme is inhibited by ADP: an ATP/ADP ratio of 10:1 has been suggested as essential for maximum nitrogenase activity. If [ATP]/[ADP] [Pi] controls respiratory rate, the apparently high requirement for Pi of nitrogen-fixing systems may be to maintain the required high [ATP]. Nitrogenase is a relatively slow enzyme,

turnover number about 3 s^{-1}, but even so, with 12-15 ATP being required for each nitrogen molecule reduced, 36-45 ATPs will be used per second for each molecule of active nitrogenase. In order to maintain this supply, it is likely that the ATP generating system needs (a) to be located near to the nitrogenase, and (b) to have either a high turnover number and/or be present in rather high concentrations. This brings us back to the various solutions which organisms have adopted to reconcile the production of ATP with the reduction of nitrogen. This is inevitably coupled with the protection of nitrogenase from oxygen inactivation. It has been suggested that oxygen sensitivity results from the evolution of nitrogenase in the days before oxygen was present in the atmosphere. An alternative proposal is that, in order to reduce such a stable substrate as nitrogen gas, nitrogenase is inevitably able to pass electrons to stronger oxidizing agents of comparable size, such as oxygen. Thus, nitrogenase-oxygen complexes would be expected and, if stable, these would prevent the reduction of nitrogen.

2.1.5 The Solutions

2.1.5.1 Living without oxygen

In the absence of oxygen, the only ways in which non-photosynthetic anaerobes can obtain energy for ATP synthesis are substrate level phosphorylation, or specialized electron transport pathways using something other than oxygen as terminal electron acceptor. Attempts have been made to estimate the ATP utilized in nitrogen fixation in cells of *Klebsiela* (a facultative anaerobe) and *Clostridium* (obligate). To estimate this ATP, cells are grown under carbon-limiting conditions, using substrates such as glucose for which the metabolic pathways are known. Two sources of nitrogen are used; ammonium ions, or nitrogen gas. Since the latter is reduced to ammonium during nitrogen fixation, the difference in ATP utilization between cultures grown on ammonium and nitrogen gas should be a measure of the ATP requirement for nitrogen fixation. If more ATP is used for nitrogen fixation than for ammonia assimilation, less will be available for growth; and hence fewer bacteria will be produced. Thus, the relatively simple experiment of comparing the yield of cells per mole of substrate consumed on ammonium with the yield on nitrogen gas can give an estimate of the ATP requirement for nitrogen fixation.

Data for *K. pneumoniae-which* is known to ferment glucose by the Embden-Meyerhof-Parnas pathway (and hence make ATP by substratelevel phosphorylation) when grown on either ammonium or nitrogen gas-showed a molar growth yield, i.e.,

$$\frac{\text{g dry wt. of organism}}{\text{mol. glucose consumed}}$$

of 25.4 ± 2.1 for ammonium grown and 11.2 ± 1.0 for nitrogen-grown cells." From these figures it was calculated that 10.9 ± 1.5 g of cells were produced per mole of ATP, when ammonium was the nitrogen source, and only 4.2 ± 0.2 when cells were grown on nitrogen. Clearly, considerable energy is utilized for nitrogen fixation; 29 moles of ATP per mole of nitrogen are fixed in this case. However, this includes energy used for the incorporation of ammonium into amino-compounds which may occur by a different route in nitrogen-fixing and ammonium-assimilating organisms," as we shall see later. Although *K. pneumoniae* grows well anaerobically, it fixes nitrogen better when grown micro-aerophilically, i.e., with small amounts of oxygen." Thus, it appears that substrate level phosphorylation on its own may not be a particularly good solution to our problem. It has to suffice for *Clostridium,* because it is a strict anaerobe. This may have nothing to do with its nitrogen -fixing ability, but may reflect the fact that it does not possess the enzymes catalase and superoxide dismutase, which enable aerobic organisms to cope with toxic peroxides and other oxygen-containing radicals."

Photosynthetic anaerobes Of these, *Rhodospirillum rubrum* is probably the most widely studied. Although a facultative anaerobe when grown on combined nitrogen, it has generally been regarded as a strict anaerobe when grown on nitrogen gas in pure culture. However, a recent report 17 has shown fixation in 'semi-aerobic' cultures (i.e., those incubated under air, but not stirred) to be similar to that of anaerobic cultures. On stirring, activity falls off rapidly. This means that some oxygen protective mechanism for nitrogenase may be necessary. Since bacterial photosynthesis does not involve oxygen evolution, there is no problem with internally generated oxygen.

Light is necessary for synthesis of nitrogenase, and for the production of ATP and reductant for nitrogenase. *R. rubrum* is capable of growing photoheterotrophically (i.e., on combined carbon and combined nitrogen), photo-trophically (CO_2, combined nitrogen) and photo-autotrophically (CO_2, N_2). Such an adaptable organism might be expected to have complex metabolic control systems, and this is indeed the case in the control of both carbon dioxide and nitrogen fixation by ammonium. In many organisms, ammonium depresses nitrogen fixation, but in *R. rubrum* it is essential for maintaining high rates of photo-reduction of carbon dioxide, because the products of carbon dioxide

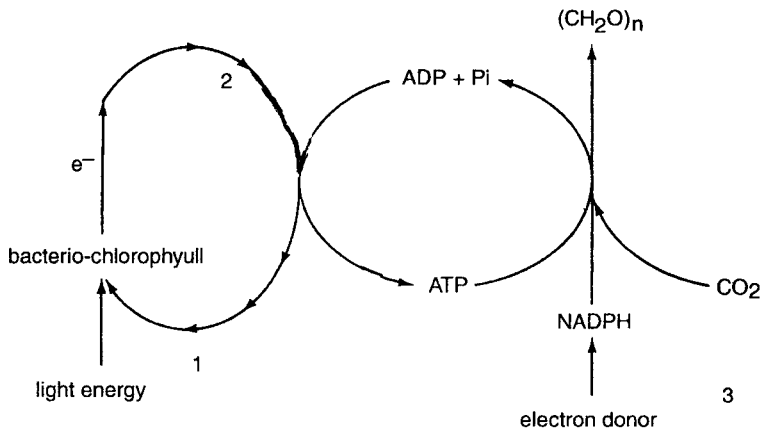


Figure 2.2 : Outline of bacterial photosynthesis. Water is never used as the ultimate electron donor, therefore oxygen is not evolved. 1, Absorption maxima vary with species from 725-1025 am. 2, Cyclic photophosphorylation with 5 intermediate e carriers. 3, H_2S, thiosulphate, H_2, or organic compounds used, according to species.

fixation undergo reductive amination to make amino-acids. Thus, nitrogen fixation and photosynthesis are coupled, in that the product of the former is used to drive the latter. When combined nitrogen is available, nitrogen fixation is not necessary for carbon photo-reduction and, when combined carbon is available, photo-reduction of carbon is not necessary for nitrogen fixation. In the latter case, light is used solely for ATP synthesis. The particular situation described here for *R. rubrum* may be linked to the fact that the species uses protein as a carbon store, rather than the more usual lipid or carbohydrate. This fact serves to emphasize the complexities of the interactions between the processes of carbon and nitrogen reduction in photosynthetic nitrogen-fixing bacteria.

2.1.5.2 Spatial separation of oxygen-evolving photosynthesis and nitrogen fixation

A range of blue-green algae, both free-living and symbiotic, can fix nitrogen: some possess heterocysts, and others do not. It is generally agreed that non-heterocystous species can only fix nitrogen micro - aerophilically, and most workers think that vegetative cells of heterocystous forms need similar conditions. Of the species that fix rapidly under air, most possess heterocysts; and the consensus of opinion is that these heterocysts are then the main sites of nitrogen fixation. Vegetative cells have photosystems I and II, and thus evolve oxygen. Heterocysts lack that part of photosystem II that includes Mn: they

cannot evolve oxygen, nor can they reduce carbon dioxide. However, they can carry out photophosphorylation and produce sufficient ATP to support maximum nitrogenase activity. They can also reduce ferredoxin using a photosynthetic electron transport chain. Further supplies of reduced ferredoxin may be obtained via electrons from high levels of NADPH produced either by the oxidative pentose phosphate pathway or from pyruvate. Carbon skeletons required for amino-acid synthesis can be obtained from vegetative cells through intercellular pores. By keeping the carbon dioxide reducing system, as well as the oxygen evolving system, away from the nitrogenase, competition for supplies of both ATP and reductant is reduced.

The evolution of heterocysts thus provides a neat way of keeping internally produced oxygen from nitrogenase. Conformational protection and thick layers of mucilage might, in addition, help to protect cells from exogenous sources of oxygen. The recent report that heterocysts of Anabaena cylindrica contain twice as much superoxide dismutase as vegetative cells suggests that even if oxygen enters cells and is used in reactions yielding the toxic superoxide radical O_2^-, this can be dealt with.

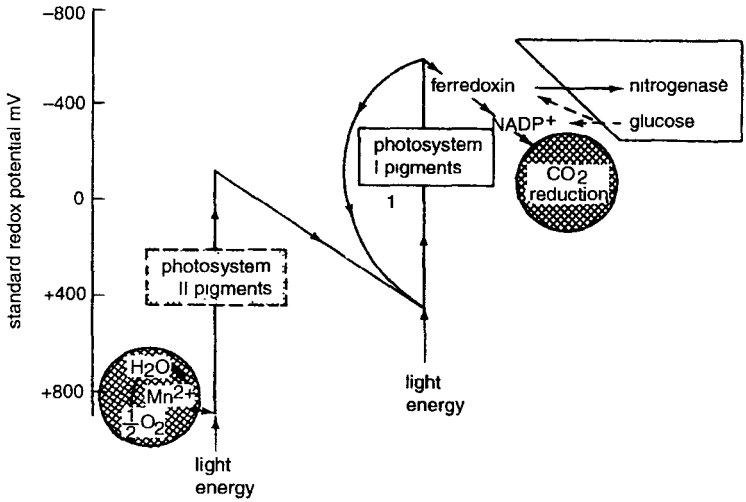


Figure 2.3 : Electron transport chains (indicated by arrows) of Anabaena cylindrica, and their probable relationship to nitrogen reduction. The photosynthetic pathway in heterocysts lacks the parts circled; thus oxygen is not evolved nor carbon dioxide reduced. Pigments of photosystem Hare also thought to be modified. Cyclic photophosphorylation occurs in region], supplying ATP for nitrogenase. An alternative source of electrons may operate from glucose to $NADP^+$ (broken lines), using the enzyme glucose 6-phosphate dehydrogenase.

2.1.6 Temporal Separation of Photosynthesis and Nitrogen Fixation

The unicellular blue-green alga *Gloeothece (Gloeocapsa)* fixes nitrogen aerobically, unlike the filamentous non-heterocystous species. How it solves the oxygen problem has recently been revealed[22] by

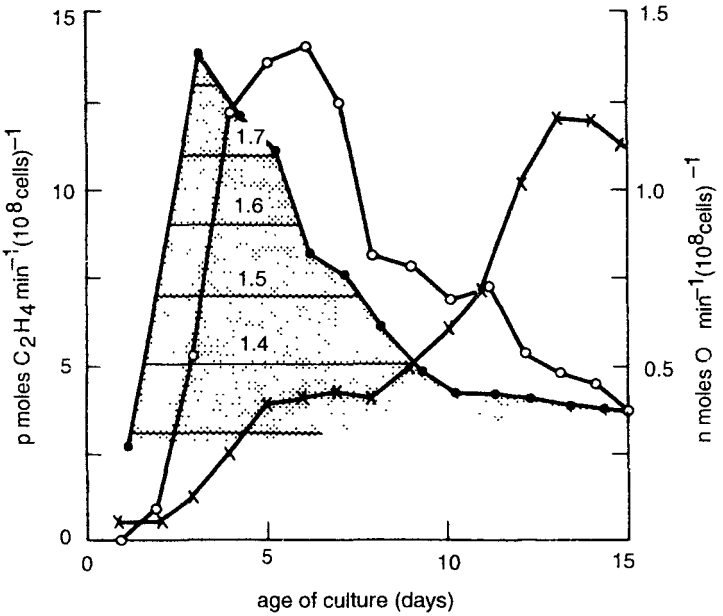


Figure 2.4 : Nitrogenase activity (0-0), oxygen evolution (x - x) and chlorophyll a 675 nml phycocyanin 615 am (shaded area) of Gloeocapsa sp (Gloeothece) cells at different ages.

studying the relative amounts of photosynthesis and nitrogen fixation carried out at various stages in the growth of a culture. Nitrogenase activity is highest during the period of maximum cell division (4-8 days under the conditions used). At this time, oxygen evolution from photosynthesis is kept at a very low level. The latter is achieved by generally low pigment levels; of those pigments present, there is a high ratio of chlorophyll a (not involved in oxygen evolution) to phycocyanin (which is involved in oxygen evolution and which can also be used as a protein store). The pigment content during the rapid division phase gives the culture a bright green appearance, compared with the blue-green appearance of older cultures, when phycocyanin contents are high. Towards the end of cell division, nitrogenase activity drops and O_2 evolution rises. Thus, there appears to be a temporal

instead of a spatial separation of photosynthesis from nitrogen fixation. In *Gloeocapsa,* photosynthetically produced reductant supplies are supplemented with NADPH reduced ferredoxin, which may be associated with either or both of isocitrate or malate dehydrogenases. Unlike other blue-green algal nitrogenases, that of *Gloeocapsa* is tightly particle-bound. It is suggested that this could result from it being located in the thylakoid membranes, thus giving it good access to photosynthetically produced ATP and reductant, which might enable it to function effectively at low photosynthetic rates.

2.1.7 Learning to Live With Oxygen

Aerobic species need oxygen in order to grow, but must keep it away from their nitrogenase. To a free-living organism of bacterial dimensions, this presents a major problem, to which *Azotobacter* at least has evolved a number of solutions." *Azotobacter* is the fastest

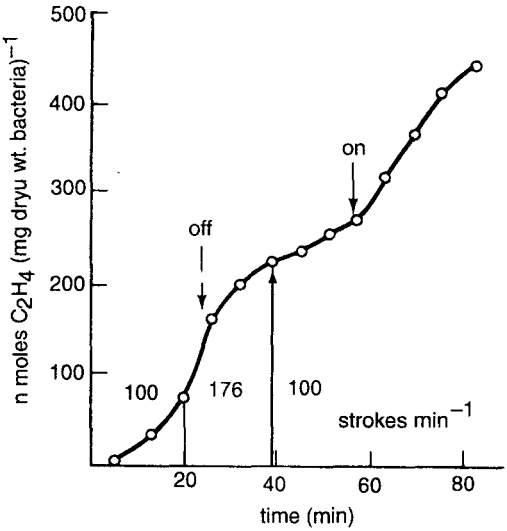
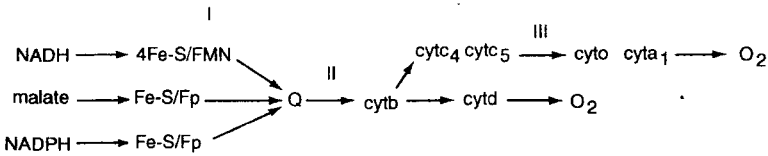


Figure 2.5 : Nitrogenase activity in Derxia gummosa, as affected by rate of shaking of the culture (strokes min'). Higher shakıng rates lead to better aeration. The rate of ethylene production falls after the shaking rate ıs increased, and rises agaın after ıt has been reduced, indıcating a reversible 'off' and 'on' switching of ntrogenase.

growing nitrogen-fixing bacterium known and is capable of extremely high rates of oxygen uptake; this capability indicates that it is well-adapted to an aerobic environment. It has an extensive internal membrane system, which may divide each cell into separate aerobic and nitrogen-fixing regions." Nitrogenase itself may be associated with membrane or other particulate fragments, which protect it (and also

inactivate it) when oxygen concentrations are too high; for example, in young low-density aerated cultures. This 'conformational' protection effectively covers the oxygen-sensitive sites on the enzyme, and must be removed when the enzyme is switched on at reduced oxygen concentrations. Similar switching mechanisms may occur in all free-living aerobic nitrogen fixers: evidence has been published for *Derxia gummosa* and *Mycobacterium flavum,* in addition to *Azotobacter.*

A second way in which *Azotobacter* may deal with excess oxygen is by a process called respiratory protection. This process was originally suggested by observations showing that the optimum partial pressure of oxygen (pO_2) for nitrogen fixation is related to the pO_2 at which the cells are grown: in other words, cells can adapt their metabolism to different oxygen tensions." At about the same time it was found that the cytochrome complement of *Azotobacter* changes with age; in particular, cytochrome of increases as pO_2 decreases: coupled with



FMN flavin momonucleotide
Fe-S iron-suphur centres
Fp flavoprotein
Q ubiquinone
cyt cytochrome
I, II and III approximate sites of energy coupling (ATP production)

Figure 2.6 : Electron transport pathways in Azotobacter.

this is an increase in phosphorylating efficiency. On the basis of these and other observations, the branched respiratory pathway shown in Fig. 2.6[26] was proposed. Under conditions of low pO_2, electrons flow along the upper route, where there are three phosphorylating steps (I, II, and III). The major route, however, is probably from NADH via cytochrome d to oxygen, bypassing phosphorylating step III. Under high pO_2 the pathways from malate and NAD PH maybe followed, neither of which has an equivalent to phosphorylating step I. This allows great flexibility of respiratory pathway according to oxygen supply, so that at high pO_2 oxygen may be taken up rapidly and inactivated (reduced to H_2O).

Under very low pO_2 (down to the limits of detection), oxygen may limit phosphorylation in *Azotobacter.* Carbon compounds and reducing equivalents (NAD(P)H) then tend to accumulate. This leads

to a further metabolic switch in that acetyl coenzyme A, instead of entering the tricarboxylic acid cycle, is used, together with NAD(P) H, for the synthesis of storage polymer poly-$-hydroxybutyrate (PHB). PHB may account for up to 70 per cent of the dry weight of *Azotobacter* cells, and it can be degraded and used when oxygen supplies are restored. It is a common bacterial product, found also in some species of *Rhizobium.*

Derxia gummosa produces two types of colony when grown aerobically on agar. Only the 'massive' type of colony shows nitrogenase activity, probably because oxygen concentrations in the interior of the colony may be sufficiently low.

2.1.8 Living With an Oxygen Consumer-Associative Symbiosis

Much publicity has recently been given to the possibility of growing grasses (particularly cereals) in association with nitrogen-fixing bacteria.

The possible ecological and agricultural significance of this will be considered later. For the present, we shall consider such associations as merely one *way* for a nitrogen-fixing organism to acquire a low oxygen environment—by living with another oxygen-consuming species. Such associations may be very widespread and crop up in such unlikely places as between *Klebsiella* and the surface of nitrogen-fixing root nodules. Associations between two bacteria have been reported; they may enable photosynthetic bacteria to fix nitrogen under conditions which are apparently aerobic. For example, *Rhodopseudomonas capsulata* fixes nitrogen under aerobic conditions when grown in association with *Bacillus megaterium,* from which it obtains pyruvate as a carbon source. The intimacy of bacterial associations may be gauged by the fact that the 'organism' once classified as *Chooropseudomonas ethy/icum* now appears to be a mixture of at least two species, of which the nitrogen-fixing member is probably a *Chlorobium sp.*

Higher plants associate with a number of nitrogen-fixing bacteria, and in some cases the association is highly specific. For example, *Azotobacter paspali* only occurs in association with the roots of the tropical/subtropical grass, *Paspa/um notatum.* This association could be related to the ability of the grass to exude a suitable energy source and/or provide an environment of suitable pH. *Derxia* is a largely tropical genus, which is more oxygen sensitive and acid tolerant than *Azotobacter.* It is generally very closely associated with plant roots and is most common in flooded soils.

Spirillum is a recent addition to the nitrogen-fixing genera. *S.*

lipoferum can associate with and invade the roots of several grasses, where it fixes nitrogen. It can utilize succinate, lactate, or pyruvate as carbon and energy sources, normally obtaining these from the grass, which also provides the necessary micro-aerophilic conditions. In pure culture the optimum pO_2 for nitrogen fixation is in the region of 0 6 kPa-the exact value depends upon the density of bacterial colonies and, consequently, on the rate of oxygen utilization. Under optimal conditions, about 8-10 mg N are fixed and assimilated per g of carbon substrate used : at higher or lower pO_2 less N is fixed per unit of C. Rates of growth on nitrogen gas (micro-aerophilic) are only about one sixth of those on ammonium (aerobic), and the nitrogen content of cells is also much lower (4-5 per cent compared with 9-10 per cent). These facts, together with the high PH B content of nitrogen-grown cells, suggest that, although micro-aerophilic conditions are necessary to obtain optimum nitrogenase activity, they may limit oxidative phosphorylation (see discussion on *Azotobacter,* above). Thus, perhaps *S. lipoferum* has not completely solved its nitrogen-fixing problems, but this is not altogether surprising, since it can apparently carry out all the other reactions of the nitrogen cycle as well. Nitrogenase activity may be restricted to young spiral motile cells.

2.1.9 Letting a Eukaryote Cell Solve The Problem

2.1.9.1 Legumes

We saw how rhizobia are incorporated into legume cells to form the nitrogen-fixing root nodules. This is an extremely well-balanced symbiotic system, in which the host supplies the microsymbiont with all its carbon compounds and receives ammonium in exchange, leaving the rhizobia free to concentrate on fixing nitrogen. In order to do this efficiently, a carefully controlled oxygen supply to the active bacteroids is required. This is achieved with the aid of leghaemoglobin, cytochrome P_{450}, and by branched respiratory pathways. We shall consider each of these in turn.

Leghaemoglobin has been known to be associated with nitrogenf-ixation ability in legumes for many years, and its properties and functions have recently been reviewed. Structurally and functionally, it has much in common with animal myoglobin, including the property of binding oxygen. Leghaemoglobin facilitates the diffusion of oxygen so that the *flux* of bound oxygen exceeds that of free dissolved oxygen by a factor of 10^4-10^5. In other words, even if the *concentration* of oxygen near the bacteroid surfaces is low, facilitated diffusion ensures

that it remains constant-oxygen used for ATP production is immediately replaced.

The efficiency of leghaemoglobin in facilitating oxygen transport is easily seen from its effects on nitrogenase activity and on the ratio of ATP to ADP. Soybean bacteroids incubated under oxygen tension of 2 7 kPa reduced 1 33 nmols acetylene min^{-1} mg dry weight $bacteroid^{-1}$. When 0 5 mM oxyleghaemoglobin was added this value increased to 4 93 nmol. At the same time ATP/ADP increased from 1 2 to 3 6. Doubling the oxygen tension (in the absence of oxyleghaemoglobin) only increased acetylene reduction to 2 2 nmol min^{-1} and ATP/ADP to 2 1. Work with intact soybean nodules has shown that energy charge:

$$\frac{1}{2}\left(\frac{[ADP] + 2[ATP]}{[ATP] + [ADP] + [AMP]}\right)$$

is closely coupled with nitrogenase activity. Thus, the effects of oxyleghaemoglobin in stimulating ATP production by bacteroids could clearly be vital *in vivo*.

Before its role can be exactly quantified, the conflict over the cellular location of leghaemoglobin needs to be resolved. Is it only in the host cytoplasm, only in the solution bathing bacteroids within their membrane sacs, or in both places? It has been pointed out that in order to fulfil a role in facilitating oxygen diffusion, leghaemoglobin itself must be free to diffuse.

Bacteroids contain a number of auto-oxidizable cytochromes, including P_{450} and possibly cytochrome a_1, which are absent from air-grown rhizobia. These have been implicated in altered respiratory pathways, but also, additionally, it has been suggested that cytochrome P_{450} could act as an intracellular oxygen carrier rather than as a terminal oxidase. Cytochrome P_{450}, mainly studied in animal tissues, is now known to be widespread in plants and micro-organisms. In animals, the principal function of a detoxification mechanism has been ascribed to it, and generally it acts as a mono-oxygenase. That it can bind oxygen is unquestioned, but the suggestion that it might act as an oxygen carrier is an interesting and novel one. It is based upon experiments with Nphenylimidazole, which inhibits cytochrome P_{450} This inhibitor decreases aerobic nitrogenase activity by bacteroids, although it has no effect on the ATP-supported activity of the purified enzyme. This infers an effect on respiration, and this was confirmed by reduced ATP levels and ATP/ADP ratios in inhibited bacteroids.

Evidence was cited against the inhibitor acting by uncoupling of oxidative phosphorylation since, for example, it had little effect on nitrogenase activity by the aerobe *Azotobacter vinelandii,* which does not contain cytochrome P_{450} These and other data are most easily explained if cytochrome P_{450} acts as an oxygen carrier, possibly transferring oxygen between oxyleghaemoglobin and a high efficency terminal oxidase. As in *Azotobacter*, rhizobia appear to contain two terminal oxidases, varying in their oxygen affinity. The high affinity oxidase has an optimum of about 0·1 μM free oxygen and is effective in providing ATP to support nitrogenase action. The low affinity oxidase, which has little activity below 1 pM free oxygen, supports neither ATP generation, nor nitrogenase activity: it may act as an oxygen scavenger.

2.1.9.1.1 Non-legumes

The situation for non-legumes is obscure, to say the least. Only recently has even indirect evidence been put forward that the nitrogenase is located in the micro-symbiont, although this was quickly followed by a report of nitrogenase activity in endophyte suspensions. One of the problems associated with the extraction of enzymes from these plants is the high level of inhibitory phenolic substances. Non-legume nodules do not possess haemoglobin. It is not known how their oxygen supply is controlled, although it has been suggested that the enzyme tyrosinase may act as carrier. This enzyme is involved in the oxidation of phenols when cells are damaged, but it has also been shown to form reversible complexes with oxygen. It is structurally similar to the animal copper-containing respiratory pigment haemocyanin, and it is possible that it has a similar oxygen-carrying role in intact non-legume nodules.

2.1.9.2 Ammonium assimilation

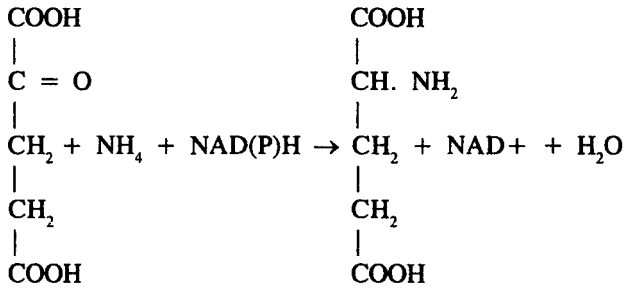
Finally in this chapter we must consider how the ammonium produced by nitrogenase is further assimilated into amino-acids. Since the pK for the reaction

$$NH_3 + H^+ \rightleftharpoons NH_4$$

is 9.25, ammonium is the predominant species at physiological pHs. We shall therefore assume that NH_4 takes part in the various reactions, rather than NH_3—although complete proof of this is lacking.

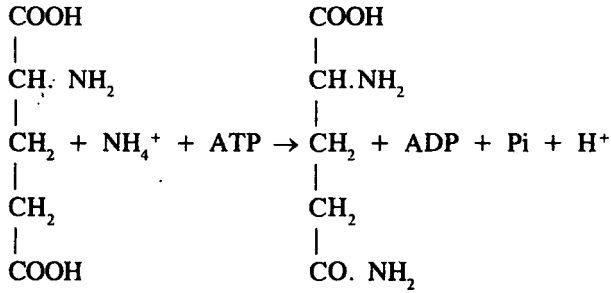
In most nitrogen-fixing species, ammonium represses nitrogenase synthesis, making rapid assimilation imperative. All non-symbiotic nitrogen-fixing species that have been tested (and most have) produce glutamate as their major product of ammonium assimilation. There

are two ways of doing this. First, by reductive amination of 2-oxoglutaric acid by NH' and NADH or NADPH, thus

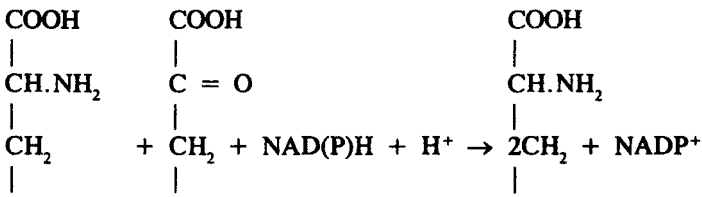


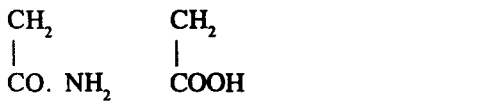
The enzyme concerned, glutamate dehydrogenase (GDH), has a rather low affinity for ammonium (K_m around 14 nM for plant enzyme and 4 mM for bacterial enzyme).

The alternative way is a two-step process. Initially ammonium is added to glutamic acid by an ATP-requiring enzyme, glutamine synthetase (GS),



This enzyme has a much greater affinity for NH4 (K_m around 2×10^5 M) than G DH, and it appears to be the one principally used in the assimilation of products of nitrogen fixation (GDH may be the important one for organisms fed on, rather than producing, NH_4). To make glutamic acid from glutamine we need a second reaction, which is catalysed by an enzyme known commonly as glutamate synthase or GOGAT, short for glutamine-2-oxoglutarate-amino-transferase.





The K_m for glutamine as substrate is 0·25 mM for bacterial enzymes, and 0·3 mM for chloroplast enzymes: for oxoglutarate as substrate the K_m values are 7·3μM (bacteria) and 150μM (plant). The cost of the GS/GOGAT pathway is one ATP per glutamate formed, but the benefit is rapid assimilation of ammonium from pools of low concentration, thus preventing ammonium accumulating in sufficient quantity to repress nitrogenase synthesis.

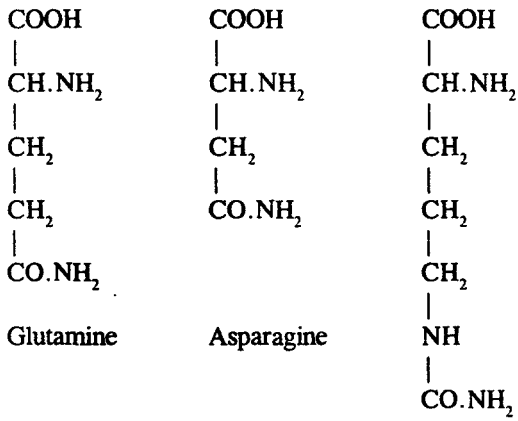


Figure 2.7 : Structures of some of the major compounds exported in xylem sap of nodulated plants.

2.1.9.3 Symbiotic systems

We have seen that many blue-green algal symbionts excrete ammonium, which is assimilated by the macrosymbiont. A similar situation apparently obtains in legume nodules, and the higher plant carries out the ammonium assimilation reactions using GS/GOGAT; however, in addition to glutamine, many legumes export from their nodules either asparagine, allantoin (or allantoic acid), or in some woody species (as well as in many non-legumes), citrulline. The structures of these substances is given in Figure 2.7. Of these, asparagine has been the most widely studied, and a pathway for its synthesis in lupin nodules has been worked out.

The critical enzyme in asparagine synthesis in lupins is a glutamine dependent asparagine synthetase, which catalyses the transfer of the

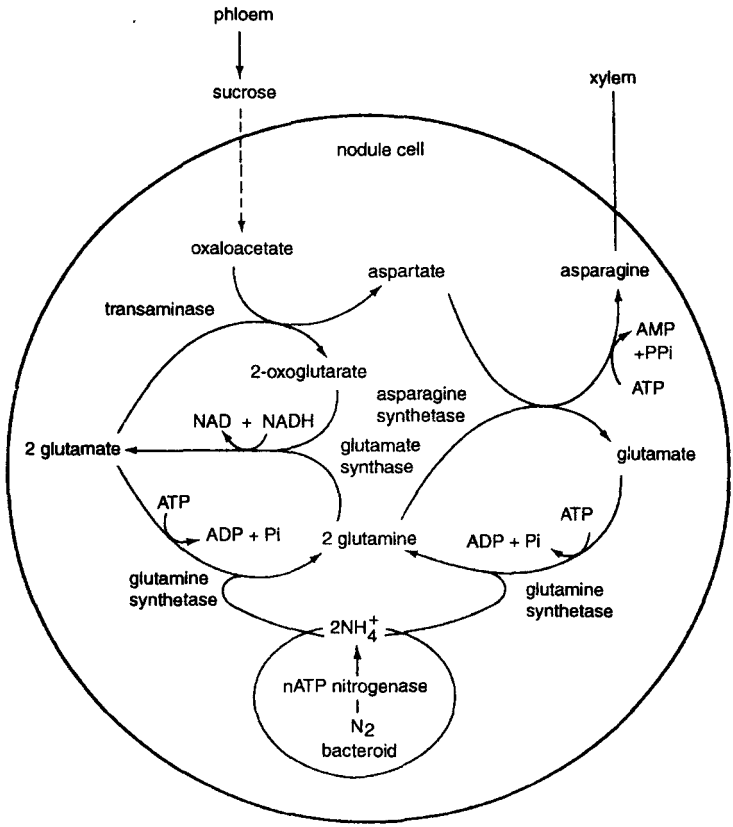
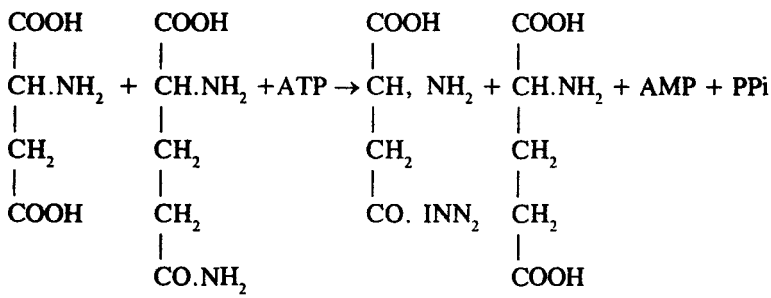


Figure 2.8 : Proposed mechanism of asparagine syntesis in lupin nodules. Enzymes located in plant fraction.

amide group from glutamine to aspartic acid to produce asparagine and glutamic acid

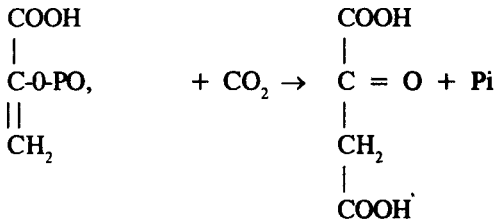


Once again ATP is used, but this time even more expensively

than for GS, since the cleavage occurs between the first and second phosphate groups, yielding AMP and pyrophosphate instead of ADP and orthophosphate. Why bother to assimilate NH. into asparagine? Higher plants have a transport problem not shared by simple organisms like prokaryotes. It is efficient to transport as much N as possible for the minimum C, to avoid sequestering too much in the transport system. Asparagine has 2N to 4C better than glutamine where the proportion is 2 to 5. Citrulline has 3N : 6C, the same ratio as asparagine, but allantoin and allantoic acid are best of all, having 4N to 4C. They are found in many legumes of warmer areas, such as soybean and *Phaseolus vulgaris*. Nodules of soybean possess the enzymes to make allantoin from the purine, xanthine, but how this relates to ammonium produced from fixed nitrogen is not yet clear.

In case you are wondering why plants do not transport ammonium in quantity, the answer is simple-it is toxic. However, its assimilation into amino-acids presents another problem: net production of about 0.3 H^+ per N fixed, which may have to be excreted by the nodulated root system.

There is a tendency to assume that the carboxylic acids to which ammonium is attached-principally oxaloacetate and 2-oxoglutarateare derived from sugars imported to the nodule and processed by Krebs cycle. However, there is increasing evidence that nodules can incorporate carbon dioxide into organic acid molecules for use as carbon skeletons in ammonium assimilation. The principal carboxylating enzyme may be phosphoenol pyruvate (PEP) carboxylase.



This reaction is cytoplasmic, rather than mitochondrial, and thus the product, oxaloacetate, is more accessible to GS/GOGAT. More than enough carbon dioxide appears to be produced by nodule respiration, and its fixation could mean a considerable saving of photosynthate.

By now, you will have realized some of the pitfalls of trying to work out the ATP requirements for nitrogen fixation. Let us end this chapter by attempting some sums for atypical legume, which are based on the scheme for asparagine synthesis in lupin. The net requirement

for asparagine synthesis from oxaloacetate is 3 ATP for every 2 NH_4^+. To make the 2 NH_4^+ probably uses about 15 ATP. Assuming these 18 ATP are made by oxidative phosphorylation following complete oxidation of glucose, and that the latter process yields about 36 ATP, the assimilation of one mol of N_2 needs 0.5 mol glucose for ATP production alone. Oxidation of 0.5 mol glucose uses 3 mol of oxygen. Thus this part of the operation would require 3 mol of O_2 for every N_2 reduced, which is exactly the value found for a different legume, Pisum sativum.50 However, these calculations are probably oversimplified, since energy may also be needed for loading asparagine into the transport system and other associated processes. Also, in addition to photosynthate required for ATP production, further supplies are necessary to provide reductant for nitrogenase and carbon skeletons for amino-acid synthesis.

3

Microbial Fixation of Nitrogen

Seventy eight per cent of the air we breathe in is nitrogen. Unfortunately, 78 per cent of the air we breathe out is also nitrogen. In this respect we are typical members of the plant and animal kingdoms-or eukaryotes-i.e., we cannot convert this very stable gas into a biologically useful form. All the organisms that can utilize nitrogen belong to the kingdom known as the prokaryotes, and the basic reaction appears to be the same in all cases, i.e., the reduction of nitrogen to ammonia with the aid of the enzyme complex, *nitrogenase*. Although this may sound simple, it is beset with all sorts of problems which may be dealt with in a variety of ways. This book aims to describe the organisms which can reduce or fix nitrogen, to show you how they tackle and solve their various problems, and to try to assess their overall significance and potential.

But first, let us put nitrogen fixation into perspective. Figure 3.1 is a simple version of the familiar nitrogen cycle, envisaged as three compartments. Outside we have the atmosphere, next the (soil + water) environment, and inside the organisms. If we accept (and not everybody does) that the world is short of protein, then clearly we need to increase the input into and decrease the output from the inner compartment. Taking the last point first: we know that some soils, particularly under climax vegetation such as forest, contain substances which inhibit specific stages in nitrification and thus tend to conserve nitrogen.' Some synthetic inhibitors are now being produced and introduced into agriculture.' These also reduce losses by leaching, since nitrate is the most easily leached of the nitrogen compounds.

In order to enhance input, nitrogen compounds acceptable to living

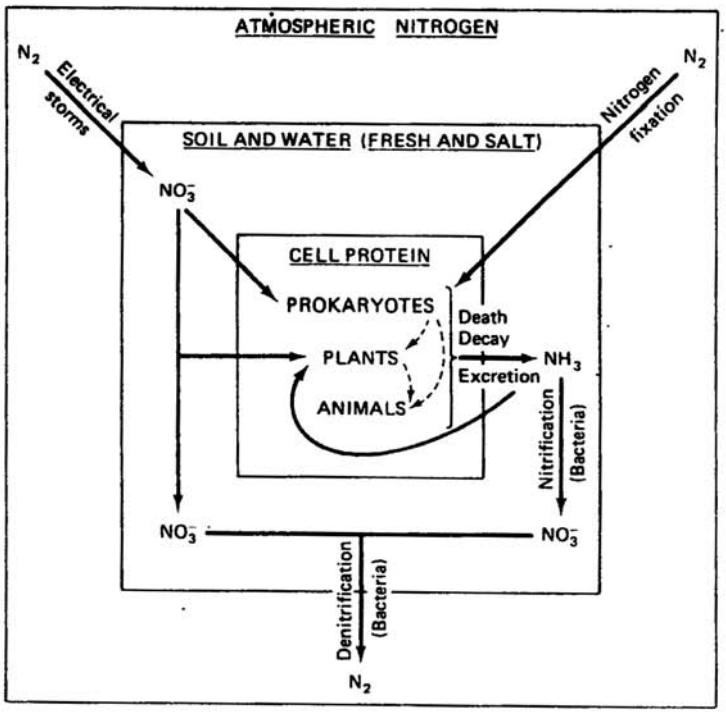


Figure 3.1 : The principal reactions of the nitrogen cycle.

cells must be produced, and this requires a considerable amount of energy. This energy may arise from atmospheric electrical discharge (or from the internal combustion engine), which causes nitrogen and oxygen to combine. Fortunately, we cannot yet control thunderstorms and although in some areas of the world significant nitrogen 'fixation' occurs in this way-values in the region of 30 kg ha^{-1} have been reported-it is likely to remain unpredictable in the foreseeable future. Nitrogen and hydrogen may be combined by the Haber-Bosch process to form ammonia which can be used for fertilizer but only at the expense of fossil fuel. About 1 5 kg of fuel oil is needed to manufacture and deliver 1 kg of fertilizer N to the farm.'

Nitrogen-fixing organisms use light energy, directly or indirectly, to produce ammonia and, since manufacture occurs on site, distribution costs do not arise. This energy-saving potential of biological nitrogen fixation may rank in importance with the end product, ammonia. A third aspect, which has not yet been fully evaluated, stems from the fact that biological nitrogen fixation is tailored to the needs of the organism,

unlike applied fertilizer, which cannot all be used immed-iately. Loss of fertilizer by leaching not only wastes energy and money, but is also a source of environmental pollution, particularly by contamination of fresh water supplies. Economic and sociological aspects of nitrogen fixation have thus added impetus to research in recent years.

It has been estimated that about 130 Tg Na^{-1} may be fixed by biological processes, compared with less than 50 Tg Na^{-1} by industrial and atmospheric fixation.'

3.1 THE RANGE OF NITROGEN-FIXING ORGANISMS

Over the last century, many reports of nitrogen fixation have appeared in the literature. The organisms concerned have been both free living and in symbiotic association with other organisms. With a few exceptions, they have all been prokaryotic, that is, belonging to the bacteria and related groups. The exceptions almost all concern fungi, principally yeasts, but also some more complex fungi and higher plant/fungal symbioses. None of these reports has survived critical rechecking." One of the problems has been that some microorganisms are excellent scavengers of traces of combined nitrogen (e.g., NH_3, NO_3^-) which may be present as impurities in culture media or as atmospheric contaminants (particularly in laboratories inhabited by smokers of tobacco). Unless or until we are successful in transferring genes for nitrogen fixation to eukaryotic organisms, it is probably safe to assume that all nitrogenfixing organisms are prokaryotic. What are the distinguishing features of this group? Basically,

1. that their nucleoplasm is never separated from their cytoplasm by a nuclear membrane and is not associated with basic protein;
2. that their plasma membranes are frequently complex with intrusions into the cytoplasm; and
3. that they rarely have cytoplasmic organelles independent of the plasma membrane and, when they do (gas vacuoles and granules of various types), they are not enclosed by a unit membrane.

The prokaryotes are now frequently classified as a separate kingdom, divided into two divisions, Cyanobacteria and Bacteria. The taxonomy of both of these is hotly disputed. Take the Cyanobacteria, or blue-green bacteria-these are still more commonly known as the blue-green algae or Cyanophyceae and often placed in the plant kingdom. However, they possess all the attributes of prokaryotes, listed above. The main

feature which distinguishes them from 'other' bacteria is that they carry out the higher plant type of photosynthesis, i.e., they use water as an electron donor and consequently evolve oxygen. They also show more cellular differentiation. In this book, they will be regarded principally as prokaryotes, and the argument as to whether they are bacteria or plants left to others. We shall, however, adopt the common usage of 'blue-green algae'. Within the division 'Bacteria', the classification which has been adopted as far as possible is that of the latest (8th) edition of Bergey's *Manual of Determinative Bacteriology.*

3.2 BACTERIA

The species currently known to fix nitrogen. The first and largest section (a) includes the non-photosynthetic, non-filamentous forms, and even a quick glance shows the wide range of bacterial families

TABLE 3.1 Genera of nitrogen-fixing bacteria (a) Non-photosynthetic, non-filamentous forms.

Family, Genus, Species	*General Comments*	
Pseudomonadaceae	Soil, fresh water, and salt water:	
Pseudomonas azotogensis	exact classification uncertain: fixation of N_2 anaerobic?	
Azotobacteraceae	Soil, water, leaf, and root surface: all species fix N_2 aerobically, but generally more efficiently at low pO_2.	
Azotobacter	Alkaline soils: generally produce	
Azomonas	some extracellular slime. *Azotococcus*	
Beijerinckia	Acid soil: not temperate regions,	
Derxia	produce abundant slime.	
Rhizobiaceae		
Rhizobium	All spp fix N_2 in symbiosis with legumes, micro-aerophilically: some strains inactive: grow in soil on combined N, but have been induced to fix N_2 apart from legume in laboratory.	
leguminosarum	Rapid growth	
phaseoli	on yeast extract	Detailed
trifolii	media.	classification
japonicum	Slow growth on	controversial.
lupini	yeast extract media.	

Table Contd.

'cowpea miscellany'	Possibly the more primitive rhizobia.
Bacillaceae	
Bacillus	Widespread in occurrence: aerobic or facultatively anaerobic.
polymyxa	Most strains fix N_2 anaerobically.
megaterium	Fixation less common.
macerans	
Clostridium pasteurianum	*Soil,* fresh water, salt water, sedi
butyricum and other spp	ments, intestines, faeces: some strains fix N_2 anaerobically or micro aerophilically: some reduce Fe.
Desulfotomaculum	Intestines, rumens: strict anaerobes:
mesophilic sppd	convert SO_4^{2-} to S^-: some strains fix N_2.
Enterobacteriaceae	All originally isolated from intestinal flora, now reported from various habitats: only a few strains actively fix N2. Nitrogenase synthesis and activity anaerobic or micro-aerophilic.
Klebsiella pneumoniae	Leaf and nodule surfaces, faeces,
and other spp of uncertain classification	rumen.
Enterobacter aerogenes	
cloacae	
Erwinia herbicola	
Citrobacter freundii	
intermedius	
Escherichia coli	
intermedia	
Spirillaceae	
Spirillum lipoferumb	Obligate aerobe associated with roots of grasses, etc., where it may fix N_2 micro-aerophilically.
Genera of uncertain family	
Desulfovibrio vulgaris	Wet soils, fresh water and salt water
desulfuricans	with high organic content. Not all
gigas	strains actively fix N_2.

Methylosinus trichosporium	*Soils,* water: utilizes methane: growth and N_2 fixation aerobic.
Thiobacillus ferroxidans	Acid waters with high iron content. Chemolithotrophic, oxidizing Fe^{++} + and S compounds: growth aerobic and N_2 fixation microaerophilic

TABLE 3.1 ***Genera of nitrogen fixing bacteria*** **(continued)**

(b) Photosynthetic forms. These genera have photosynthetic pigments associated with intracytoplasmic membranes formed from the plasmalemma and observed either as vesicles or lamellar stacks. All are classified in the order Rhodospirillales which contains three families, members of each of which fix N_2.

Family, Genus, Species	*General Comments*
Rhodospirillaceae	Predominantly aquatic, facultative
Rhodospiril/um rubrum	anaerobes, purple, non-sulphur
Rhodomicrobium sp	bacteria. Generally micro-aerophilic
Rhodopseudomonas capsulata	when grown on combined N. Require
spheroides	light if grown anaerobically and under these conditions some strains fix N_2.
Chromatiaceae	Moist and muddy soils, salt and fresh
Chromatium sp	water where sulphide high. Strict anaerobes: purple sulphur bacteria: fixation of N_2 rare.
Chlorobiaceae	Habitats as for Chromatiaceae: strict
Ch/orobium limicola	anaerobes: green sulphur bacteria: fixation of N_2 rare.

Table 3.1 Genera of nitrogen fixing bacteria (continued) (c) Actinomycetes and associated spp. Filamentous forms.

Family, Genus, Species	*General Comments*
Mycobacterium f/avuma and other spp	Acid soils
Corynebacterium autotrophicuma	Soils, mud: several strains shown to fix N_2 autotrophically (H_2, CO_2), as well as heterotrophically (sucrose) in presence of O_2.

Frankiaceae	
Frankiab	Fix N_2 in symbiosis with nonleguminous angiosperms.

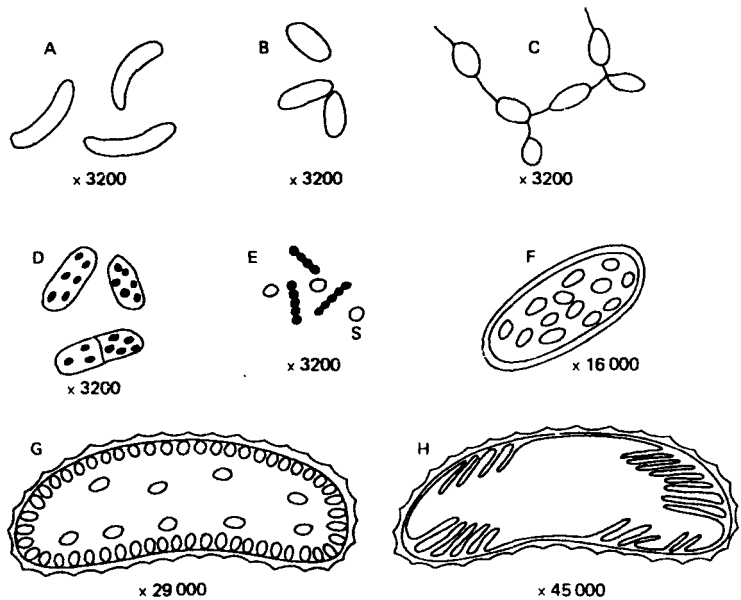


Figure 3.2 : Variation in form and structure in genera of free-living nitrogen-fixing bacteria. A, Rhodospirillum molischianum; B, Rhodopseudomonas acidophila; C, Rhodomicrobium vannielii; D, Chromatium vinosum with internal sulphur granules; E, Chlorobium limicola with extracellular sulphur globules (S); F, Azotobacter chroococcum (note extensive internal membrane system found in cells grown on nitrogen gas); G and H, Rhodospirillum rubrum and R. molischianum grown anaerobically in the light showing, respectively, vesicles and lamellar stacks in intracytoplasmic membranes.

represented. By the time this book is published, additional species may have been discovered: others may be reclassified or even deleted. The reasons for this apparently haphazard occurrence of the nitrogen-fixing habit may be evolutionary. Most of the genera are only active under anaerobic or micro-aerophilic (i.e., low concentrations of oxygen) conditions and only a few strains of a particular species show activity. The exceptions to this among the freeliving forms are the members of the Azotobacteraceae, which have evolved several ways of living in an aerobic environment and yet carry out the strictly anaerobic process of nitrogen fixation.

All these non -photosynthetic genera are only indirectly dependent

upon light energy for nitrogen fixation. They must be supplied with suitable carbon compounds, such as sugars or organic acids containing energy originally derived from photosynthesis but made available by respiratory pathways. The symbiotic genera *Rhizobium* and *Frankia* obtain these compounds from their hosts, but free-living forms must obtain it in open competition with other micro-organisms in their particular environment. If supplies of combined nitrogen are good, nitrogen-fixing species have poor competitive ability because a considerable portion of the energy in the available carbon compounds is used to fix nitrogen, leaving less for growth. Thus, free-living nitrogen-fixing genera are generally confined to habitats rich in organic carbon but low in combined nitrogen. Such conditions occur sporadically in almost any environment soil, salt and fresh water, guts of animals; indeed nitrogen-fixing bacteria have been recorded from just about everywhere. Thus, *Citrobactercan* be found in the gut of termites, where it has plenty of carbohydrate but very little nitrogen, since a termite's principal diet is wood; (it is effectively a symbiont and not a free-living form). The members of the Enterobacteriaceae are normally thought of as inhabitants of the mammalian gut, where combined nitrogen supplies are good. However, there is an increasing number of reports of their occurrence elsewhere, for example, in paper-mill effluent where there is plenty of carbohydrate, but little combined nitrogen. It is from this type of habitat that reports of nitrogen fixation are most common and likely to be added to in the future.

Most photosynthetic bacteria are autotrophic, and thus have no problems of carbohydrate supply. When grown on nitrogen gas as sole nitrogen source they generally live strictly anaerobically, and this considerably restricts their range of habitats. Bacterial photosynthesis does not result in oxygen evolution, since compounds other than water are used as the ultimate electron donors. The reductant source is usually highly characteristic. *Chromatium,* for example, may use sulphide, in which case sulphur granules are deposited inside the cell.

The principal filamentous bacterial genus to fix nitrogen is the symbiotic *Frankia.* In the one isolate so far obtained in pure culture, the branched, septate hyphae form sporangia containing spores of complex structure, presumed to be the propagating stage of the organism. The classification of *Mycobacterium* and *Corynebacterium is* somewhat tentative.

The biological interest in free-living nitrogen-fixing bacteria stems largely from their varied metabolism and habitats. Structurally, they

conform to the usual bacterial range; rods, cocci, curved types, various numbers and positions of flagella, etc..

3.3 BLUE-GREEN ALGAE

Nitrogen fixation by a blue-green alga was reported in 1889, shortly after the significance of legume root nodules was first mooted. Subsequent research was spasmodic, but the pace gradually accelerated until in the last 15 years very rapid progress has been made and several excellent reviews have appeared. The blue-green algae have been classified on morphological and anatomical grounds-type of cell, branched or unbranched filaments, etc.

At first, all the nitrogen-fixing species were thought to belong to the filamentous heterocyst-bearing members of the Nostocales and Stigonematales. Of these, *Anabaena cylindrica* has been the most widely studied and has even been referred to as the *E. coil* of blue-green algae. Its structure and physiology has been summarized by Stewart." Essentially, it has unbranched filaments containing up to three types of cell. Vegetative cells are typical of those of the group as a whole, possessing all the requirements for the higher plant type of photosynthesis, i.e., use of water as an ultimate source of reductant, carbon dioxide fixation by Calvincycle enzymes, and evolution of oxygen. Photosynthetic pigments are located in thylakoids in the outer regions of the cells and-in addition to chlorophyll a, carotenes, and xanthophylls-they have characteristic accessory pigments called phycobiliproteins. In addition to a role as light interceptors, phycobiliproteins can apparently function as a readily usable nitrogen store. Vegetative cells may also contain polyhedral bodies, now called carboxysomes, which are molecular aggregates of the Calvin-cycle enzyme, ribulose bisphosphate carboxylase. As carbon stores, polyglu-coside bodies may be found among the thylakoids. Excess nitrogen is stored in structured granules as a 1 :1 co-polymer of aspartic acid and arginine. In addition, vegetative cells may contain gas vacuoles, polyphosphate bodies, and lipid bodies. They are thus well equipped to accumulate essential metabolites. The various inclusions are most abundant in slowgrowing phases of growth and become depleted during rapid growth.

Heterocysts, which may be spaced at intervals along the vegetative filaments, or be terminal, are characteristically colourless, and apparently empty under the light microscope. They cannot fix carbon dioxide or evolve oxygen, but can make ATP by photophosphorylation and can fix nitrogen aerobically. Thus, the processes of nitrogen and carbon reduction in heterocystous blue-green algae are physically

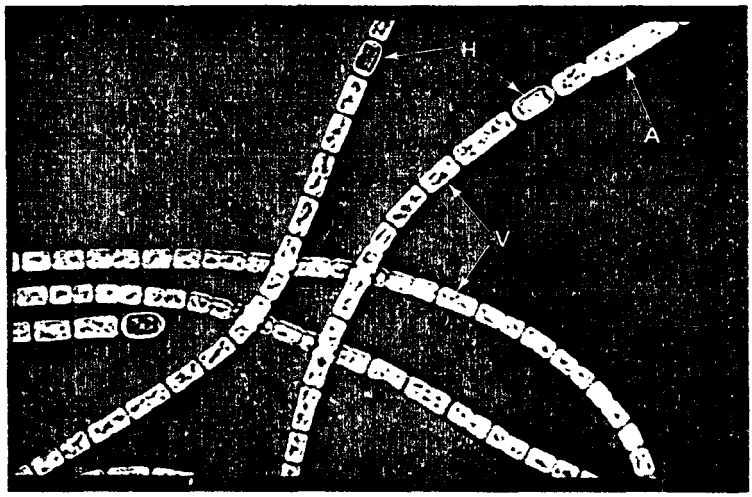


Figure 3.3 : Anabaena cylindrica, a filamentous heterocystous blue-green alga. Note vegetative cells (V), heterocysts (H) and developing akinete (A).

separated and we shall discuss the advantages of this in the next chapter.

The third type of cell, the akinete, is a perennating spore and will not be considered further. *Anabaena* and many other blue-green algae reproduce vegetatively by means of short filaments called hormogonia, which are often motile. The general biology of the blue-green algae has recently been discussed in detail.

Among heterocystous forms, the ability to fix nitrogen appears to be universal. The situation in other blue-green algae is much less clear. Some strains of *Gleothece (Gleocapsa) a/pico/a, a* unicellular species, fix nitrogen aerobically. How they achieve this is not fully understood. However, there is a range of non-heterocystous filamentous forms, that can synthesize nitrogenase anaerobically and fix nitrogen microaerophilically, a situation rather similar to that of the free-living bacteria other than the Azotobacteraceae. Clearly these species must synthesize nitrogenase in vegetative cells, since they have no others. This raises the interesting question as to whether heterocystous forms can fix nitrogen in vegetative cells under micro-aerophilic conditions. This question has not yet been resolved, and the various arguments for and against it have been discussed elsewhere.1*B*

The habitat range of blue-green algae is even wider than the nitrogenfixing bacteria2'-from sub-Arctic to Antarctic. Blue-green algae inhabit hot springs, soil, salt and fresh water. They are generally resistant to extremes of temperature and desiccation, require only low

TABLE 3.2 Genera of blue-green algae with nitrogen-fixing species.

Classical classification	*Physiological types of Kenyon et al.*
Heterocystous forms	
Nostocales	
Nostocaceae	
Anabaena	Anabaena type
Anabaenopsis	*A. circularis* of Anabaena type
Aphanizomenon (*flos-aquae* only)	
Aulosira	
Cylindrospermum	
Heterocystous forms (continued)	
Nostocales (continued)	
Nostocaceae (continued)	
Nodularia	*N. sphaerocarpa* of Calothrix type
Nostoc	Microchaete type
Scytonemataceae	
Scytonema	
Tolypothrix	*T. tenuis* of Calothrix type
Microchaetaceae	
Microchaete	Microchaete type
Rivulariaceae	
Calothrix	Calothrix type
Dichothrix	
Stigonematales	
Stigonemataceae	
Fischerella	
Hapalosiphon	
Stigonema	
Westelliopsis	
Mastigocladaceae	
Mastigocladus	
Chlorococcales	
Chlorococcaceae	
Chlorogloea (*fritschii* only)	Chlorogloeopsis type[b]

Non-heterocystous forms	
Nostocales	
Nostocaceae	
Raphidiopsis	
Rivulariaceae	
Gleotrichia Rivularia	
Oscillatoriaceae	
Oscillatoria	Oscillatoria type
Lyngbya	Lyngbya type
Trichodesmium	
Plectonema	
Phormidium	Lyngbya type
Chlorococcales	
Chlorococcaceae	
Gloeothece	

levels of light, and prefer regions of comparatively high pH (usually above 7). There are also many symbiotic forms as we shall see later.

The morphological and anatomical characters of the blue-green algae are very variable and depend greatly on environment and nutrient availability. This has led to ambiguity and mis-classification. To try to resolve this, an alternative classification, based partly on physiological characters, is being attempted. One of the criteria used is the ability to synthesize nitrogenase aerobically or anaerobically. Further details are outside the scope of this book. However, it seems likely that a classification of the blue-green algae similar to that of the bacteria will emerge in time, i.e., one based on both visual and physiological/biochemical characters. Whether or not this will reduce confusion remains to be seen. A list of nitrogen-fixing genera classified in both the traditional way and, where possible, according to the physiological scheme, with further details on the physiological criteria in the given Table.

3.4 BACTERIAL SYMBIOSES

There are two major types of bacterial symbiosis, both with angiosperms as the macro-symbiont, and in both types swellings (nodules) housing the bacteria are found on the root systems. Large numbers of species in the family Leguminosae are infected with different species of *Rhizobium. A* more heterogeneous group of angiosperms, for convenience called the 'non-legumes', is infected with an actinomycete-

TABLE 3.3 Physiological/biochemical classification of the blue-green algae.

Type	*Characteristic*									
	Heterocysts	Akinetes	Filaments	Filament motility	Hormogonia motility	False branching	Sheath	Uniform thickness of filament	Nitrogenase synthesis aerobic	Nitrogenase synthesis anaerobic
Oscillatoria	–	–	+	+	+	–	–	+	–	+
Lyngbya	–	–	+	–	±	–	+	?	–	+
Plectonema	–	–	+	–	±	+	+	?	–	+
Anabaena	+	±	+			–	±	+	+	
Calothrix	+	–	+			+	+	–	+	
Microchaete	+	–	+			+	+	–	+	
Chlorogloeopsis	+	+	–				+		+	

like organism, tentatively put in the genus *Frankia.* In addition to root nodules, there have been many reports of leaf nodules that fix nitrogen on host plants such as *Psychotria* and *Ardisia.* Recent evidence suggests that fixation does not occur regularly and even if it does it is not sufficient to be of material benefit to the host plant. All leaves have the possibility of bearing epiphytic bacteria on their surfaces, but these are casual associations, not symbioses. We shall assume that leaf nodules do not fix nitrogen, although they may confer other benefits on the host, possibly hormonal. For a survey of the literature on the subject as well as for negative evidence on leaf nodule fixation, the review of van Hove should be consulted.

A number of *bacteria-Azotobacter, Azospirillum* and others-may live on the surface or in the cortex of plant roots, where they may fix nitrogen. These 'associative symbioses' have no common structural features, and the mode of entry of the bacteria into root is not known. However, they may be of considerable importance, as we shall see later.

3.4.1 Legume Root Nodules

Legumes are classified in various ways by different authors: most workers on nitrogen fixation put them all in the family Leguminosae with sub families Caesalpinioideae, Mimosoideae, Papilionoideae, and

Scwartzioideae, and we shall use this system. The extent of nodulation in the family is not known exactly, but is least in the Caesalpinioideae, where only about 30 per cent of species are nodulated, and greatest in the Papilionoideae. Within certain genera, for example, *Cassia,* not all species may be nodulated, possibly reflecting heterogeneity in the genus.

TABLE 3.4 Incidence of nodulation in sub families of the Leguminosae

	Mimosoideae	Caesalpinioideae	Swartz-ioideae	Papili-onoideae
Number of spp	1500	1300	80	10 000
Percentage woody spp	95	97	100	38
Percentage spp in tropics/ subtropics (balance temperate)	90	96	100	37
spp examined	146	115	0	1024
Percentage spp nodulated	87	23	-	93

Nodulation does not normally occur in the field until the first leaves are unfolding and nitrogen fixation does not normally commence until the plant can safely divert a proportion of its photosynthate to this end. Clearly this requires synchronism of plant and rhizobial development at a number of stages, beginning in the soil. The overall sequence of events, discussed below.

TABLE 3.5 Host-rhizobia interactions during legume nodule formation

'Developmental' stage	*Genetical/ physiological requirements*	*Comments*
Rhizobium multiplication in soil	Host secretions may stimulate or inhibit	Degree of specificity variable
Root hair curling and branching	Compatible host and *Rhizobium*	Degree of specificity variable
Attraction between compatible host/rhi-	? Matching of cross-reacting groups.	Found by some to be highly specific: others

zobial cells	? Bridging by lectins	less so
Entry of bacteria	Dissolution or stretching of host cell wall	Temperature affects m strain of *Rhizobium*
Growth of infection thread	Matched development of rhizobia with host cell wall material	entering and thread development. Plants may have genetically controlled anti nodulating factors
Formation of nodules	Correct balance of growth factors from both partners	*Rhizobium* mutants known which cause development to stop at
Formation of mature infected cells	Release of rhizobia from threads. Matched growth of rhizobia and host membranes	any stage. Correct matching of host and rhizobial genotypes essential
Formation of bacteroids, development of nitrogenase and haemoglobin, etc.	Specific interactions involving both partners	Host may act by providing correct environment
Maintenance of bacteroid tissue	Correct interchange of materials between symbionts	Not well understood. Strongly conditioned by environment

3.4.1.1 Multiplication of rhizobia. Root hair curling

When a legume seed germinates in soil containing a natural population of micro-organisms, it has an effect on any rhizobia present and vice versa. Because of the ease of handling and the rapidity with which nodules are formed, most studies on the effects of rhizobia on germinating seedlings have been carried out on small seeded legumes, especially species of *Trifolium.* However, recent studies using the large seeded legumes, particularly *Pisum sativum,* have shown how the host may manipulate the rhizobial population. In common with most large seeded legumes, *P. sativum* has a prominent tap root. This does not favour rhizobial multiplication, and its exudates may even be inhibitory. However, as soon as lateral roots begin to emerge, various compounds are released including the amino-acid, homoserine, known for many years to be present in pea seeds in unusually high amounts. Homoserine preferentially stimulates growth of *Rhizobium legumino-Sarum,* which results in sufficient bacteria being present to nodulate the plant at the most propitious time. The first nodules, in common with those of most

similar legumes, are formed on the tap root. There is no detailed information available for other legumes, but it is reasonable to suppose that they may also encourage growth of compatible rhizobia, i.e., those capable of nodulating that particular host. A recent report of chemotactic movement of rhizobia to plant roots needs confirmation, especially in view of some of the unusual host-rhizobial combinations reported to produce nodules.

Once a sufficient population of rhizobia has been built up, these can begin to influence and possibly infect the host plant. The first obvious symptom is the deformation of root hairs. Three categories of deformation have been described (1) branching; (2) moderate curling, i.e., tip of hair curled through 90—360 (3) marked curling with the tip curled through at least 360°. In the host-endophyte combinations studied, homologous pairings (i.e., host-rhizobia that form effective nodules) usually produced the most curling. Not all hairs on a root respond and sometimes growth right through a rhizobial colony can occur without turning a hair.

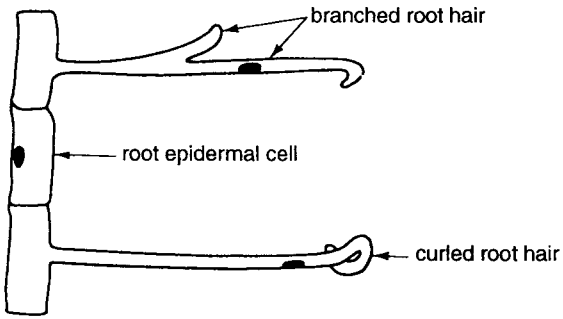


Figure 3.4 : Curling and branching of root hairs as seen in response to presence of rhizobia.

The nature of the factors responsible for root hair curling are not completely understood. The suggestion that auxin (indole acetic acid) is an essential component is not supported by recent evidence discussed by Yao and Vincent. These workers concluded that root hair branching and, usually, moderate curling may be caused by a partially dialysable heatstable product or products of rhizobia. Marked curling requires contact between viable virulent, but not necessarily effective rhizobia, i.e., those capable of forming a nitrogen-fixing nodule. Markedly curled root hairs may reflect a highly localized reaction, possibly including the early stages of infection. A curl could delimit an area where some vital substance or substances' accumulate. It is also possible

that entry holes made in the crook of a curl may be more easily plugged to prevent entry of pathogens. Although there have been several reports of bacteria other than rhizobia in root nodules, the incidence of pathogenic species appears to be low.

However, we are getting ahead of ourselves. Before entry can be effected, bacteria of the right type must be correctly aligned on the hair. At this stage, bacteria are frequently seen to be embedded in mucigel, a material of host origin, which may cover root tips and hairs, help bind roots to soil, and also possibly provide substrates for bacterial growth.

Various suggestions have been put forward to account for the required highly specific interaction between bacteria and root hair surfaces, the most recent of which envisages an attraction similar to that between antigens and antibodies in animals. What substances produced by the symbiotic partners could be sufficiently specific? Legumes are well known for the production of proteins that can agglutinate red blood cells and cause symptoms of poisoning in some people. The proteins, which used to be called phytohaemagglutinins, are now usually referred to as lectins; many workers believe that they can contribute to host-rhizobial specificity.

Lectins produced by the host are thought to be capable of binding to both the root hair surface and also to the capsule of the compatible *Rhizobium*. The binding sites on the two partners are structurally and antigenically similar polysaccharides." Microscopic observations have shown that attachment of the rhizobia to the root hair and other surfaces is polar (i.e., end-on); it has been suggested that the initial attachment is cemented by cellulose fibrils of rhizobial origin." Although this is an attractive hypothesis (as have been many that have gone before), it should be noted that not all workers have been able to find the requisite high degree of binding between lectins and rhizobia, and it may only represent one part of the recognition process." Good supporting evidence for the lectin scheme is the recent report[35] that mutants of *Rhizobium* lacking exopolysaccharides do not form nodules.

3.4.1.2 Formation of infection threads

The coming together of homologous bacteria and roots involves only the first group of a series of complex interactions between the host and endophyte genomes. The next problem is how the bacteria gain entry into host cells. Strictly speaking they never do, because they remain enclosed by a host membrane derived by invagination of the plasmalemma.

When compatible bacteria are firmly attached to the root hair, a refractile (or hyaline) spot becomes visible in the root hair wall, which indicates structural changes. These may involve auxin, which is known to be present in higher concentrations in infected than uninfected roots, and is known to be associated with cell wall loosening and growth. Great metabolic activity on the part of the host cell is indicated by more opaque cytoplasm, increased cytoplasmic streaming, and the doubling in size of the nucleus. Only a proportion of the root hairs becomes infected (about 2 8 per cent for *Pisum sativum),* and only a proportion of these gives rise to nodules. One of the factors affecting this last proportion is the particular host-rhizobial combination.

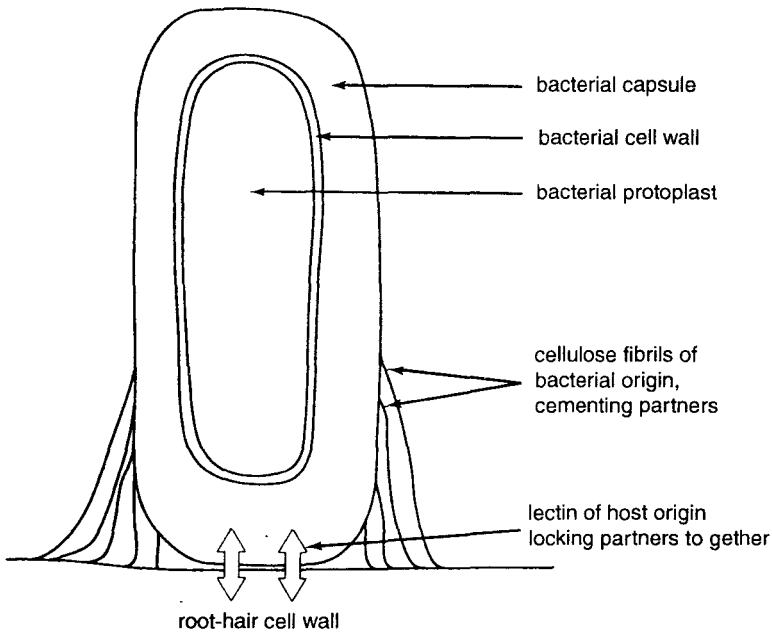


Figure 3.5 : Possible mechanism for binding rhizobia to root hairs.

A few hours after hyaline spots are seen, infection threads become visible. Growth of these is a very finely balanced process-it might almost be regarded as a competition between the bacteria attempting to effect entry and the host, which secretes cell wall material around the bacteria to contain them. If the bacteria achieve early victory, they may be seen swimming freely inside the root hair. If the plant achieves early victory, it may cut off the infection thread by deposition of cell wall material before the bacteria have had time to divide

extensively. Cooperative combatants are indicated by coordinated bacterial division and cell wall deposition. A functional infection thread results. This consists of one or more rows of bacteria aligned within a sheath of host cell wall origin; it grows down the root hair. In successful infections, the root hair nucleus remains close to the growing end of the infection thread. The infection thread grows from the root hair cell between and through cells of the root cortex, often with considerable branching.

Although infection via root hairs is most common, it is not universal. In the peanut *(Arachis hypogea),* root hairs are only formed at the points of emergence of lateral roots. These hairs are induced to curl by compatible rhizobia, but entry is at the junction of the root hair and epidermal cells. No infection threads are formed; instead, cells separate at the middle lamella, and the intervening spaces become filled with bacteria. In due course, masses of bacteria enter cortical cells of the lateral root; both the bacteria and the host cells divide repeatedly to form a nodule."

3.4.1.3 Nodule development

As with the mode of infection, there are variations in the pattern of nodule development. Very broadly, we may recognize two types: those in which infection of new cells occurs by means of infection threads, and those where the bacteria are mainly spread by division of preinfected cells. The latter may either lack infection threads as in the peanut, or have threads that do not branch and ramify extensively (soybean, *Glycine max).* Generally those nodules without extensive infection threads are more or less spherical, lack a persistent meristem (hence have determinate growth) and have a closed vascular system.

Nodules with a longer-lived meristem are indeterminate, generally elongate and often branched; they need persistent infection threads to infect newly formed cells. This indeterminate type of nodule has been widely studied, particularly in peas *(P. sativum)* and clovers *(Trifolium spp).* One of the most detailed studies of the initial stages of nodule development is that of Libbenga and Harkes using *P. sativum,* modern techniques of fluorescence microscopy, and physiological/anatomical experiments. The following description is based largely on their work.

After the infection thread passes through the outer 6 layers of cortical cells, the cells of the inner root cortex show increased RNA contents and begin to divide in the anticlinal plane (i.e., a longitudinal section must be cut to see the divisions). Cell divisions are generally opposite protoxylem points in the root stele. Apparently these divisions

occur when the correct balance of growth substances is present and these are thought to arise from (a) cells containing the infection threads producing auxin, and (b) cytokinins from the stele, together with some as yet unidentified factor.

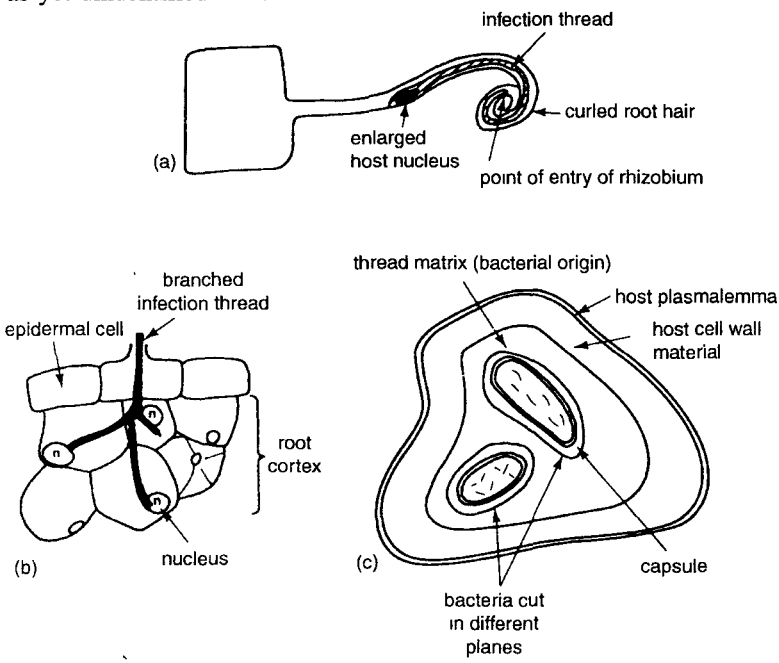


Figure 3.6 : Infection of legume roots by rhizobia. A, Young infection thread; B, Transverse section of outer region of root to show branching of infection thread and its association with host cell nuclei (n); C, Transverse section of infection thread.

The picture may be further complicated in peas by a phloem-translocated inhibitor from the cotyledons; in other species, cotyledons may produce nodule-promoting substances. The inner cortical cells, together with those of the endodermis and possibly the pericycle, continue to divide in various planes and become infected by the advancing infection thread. Rhizobia are released from the confines of the thread wall and develop into nitrogen-fixing bacteroids by a sequence of events described later: at this time, the host cells lose their meristematic activity. To take over, adjacent cortical cells become meristematic and form the more or less persistent nodule meristem, which divides repeatedly and pushes out through the cortex. At the same time, the infection thread branches and infects the growing cells cut off by the nodule meristem: to do this, it has to turn round and follow the progress of the growing nodule back through the root towards

the epidermis. In this way, a continuous sequence of newly infected cells is formed. The process continues until long after the nodule has emerged from its subtending root. The outer layers of the nodule do not become infected; instead they form the cortex, within which form vascular traces, which connect to the stele of the subtending root. There is usually a distinct nodule endodermis continuous with the root endodermis and, additionally, an endodermis around each vascular trace.

For many years, it was thought that nodules arise only from tetraploid cells, which divide when they become invaded by an infection thread. This theory is made untenable by the fact that divisions occur in cells before penetration by the infection thread, and the reason for the frequent occurrence of tetraploid cells in nodules remains obscure. A second theory, which is not supported by recent evidence, is that nodules arise from incipient meristematic foci that might otherwise have given rise to lateral roots. This implies that the number of (nodules + roots) is a constant. The arguments against this are: (a) that nodules arise in peas from the inner cortex, whereas lateral roots arise from the pericycle; (b) that lateral roots and nodules develop at different maturity zones on main roots; and (c) that the lateral roots always arise opposite to protoxylem poles, whereas nodules do not.

In nodules such as those of pea, we have all the bacteroid-containing cells formed from cells individually infected by a branch of the original infection thread. Some cells remain uninfected. The time that the meristem on such a nodule persists is somewhat variable and may be related to the cytokinin content." Some may persist for a whole season or more, others for just a few weeks. Another factor that may affect this time is the rate at which the meristematic cells become infected. Once they are all infected, division must cease. Sometimes (in peanut, for example), roots grow out of nodules. In *Sesbania grandiflora,* a woody legume, they arise from within the endodermis of older (8-12 month) nodules."

A detailed light and electron microscope study of later development of pea nodules has been made by Newcomb[42] and the reader is strongly encouraged to refer to his excellent illustrations. When an infection thread penetrates a cell it contains bacteria embedded in a matrix, probably a mucopolysaccharide of bacterial origin, surrounded by a cell wall of host plant origin. This in turn is surrounded by plasmalemma continuous with that of the invaded cell. At the tip of the infection thread, the matrix appears to break through the host wall, releasing 'unwalled droplets' containing bacteria, which are

bounded by the host cell membrane. In turn, the bacteria break free from their own matrix (which may be digested by the host cell)

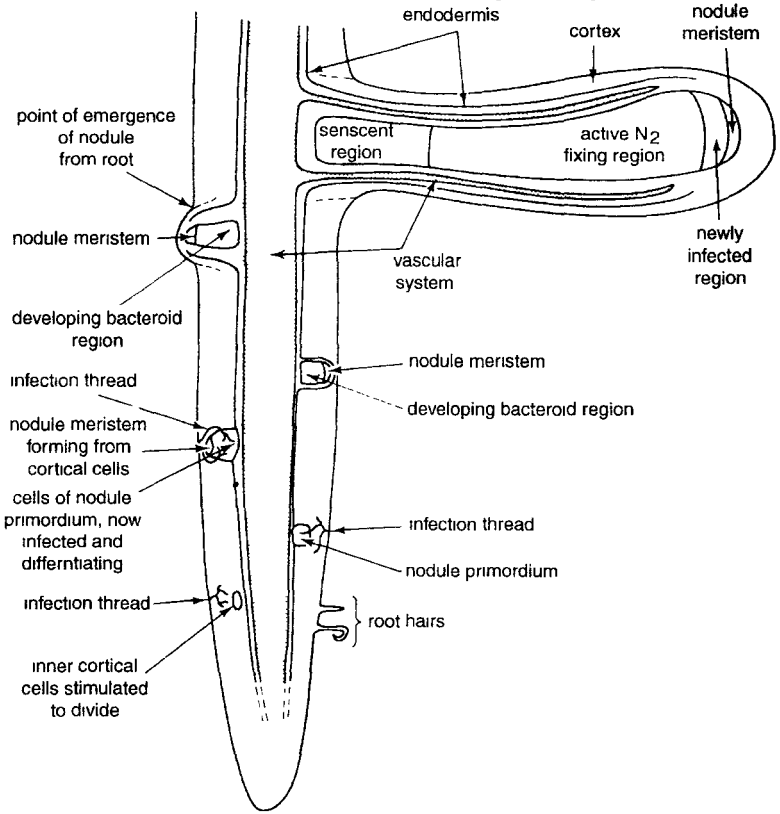


Figure 3.7 : Diagram to illustrate the sequence of stages in the development of an indeterminate root nodule.

and become individually surrounded by membrane envelopes (but in reality still extracellular) in the host cell cytoplasm, where they enlarge and differentiate into the nitrogen-fixing stage known as bacteroids.

In the determinate type of nodule, a cluster of cells is infected, either by an infection thread, or by direct infection as in peanut. Bacteria may be released from the infection thread, as has been described for pea. The cells then divide repeatedly, the bacteria also divide, and membrane material is increased in rough proportion. Bacterial divisions may persist beyond the onset of nitrogen fixation, thus disproving the suggestion that bacteria lose the ability to divide when they produce nitrogenase. The final number of bacteroids per

membrane envelope varies from 1 to 8, depending on species and nodule age. Normally, the meristem of these nodules, which is also spherical, is short lived. The infected cells do not vary greatly in age, and the nodules are of a more limited life-span than those with a persistent meristem. There is some evidence that further growth by cell enlargement can occur if determinate nodules are formed under sub-optimal conditions and then transferred to improved conditions. Unlike some of the indeterminate nodules examined, determinate nodules do not have vascular transfer cells in the pericycle of their vascular traces. Further discussion of variations in nodule shape may be found in Corby." The type of nodule produced is at least partly controlled by the host: *R. lupini* strain D25 forms effective, but differently shaped, nodules on *Lupinus luteus* and *Ornithopus sativum.*

3.4.1.4 Detailed structure of infected cells

Clearly there must be close cooperation between the metabolic activities of both symbionts. This will be considered in more detail in the next chapter. Structurally, co-operation is facilitated by the large surface area of membrane through which interchange can occur. Tu[47] has described endo- and exocytotic nips on the envelopes containing the bacteroids, and on the plasma membranes of infected cells; he suggests that these reflect the extensive exchange of materials between cells and environment. During growth and differentiation, the membranes surrounding the bacteria become altered in protein, lipid and enzyme composition; this suggests a change in function associated with the nitrogen-fixing state."

The bacteria when differentiating into the nitrogen-fixing bacteroids show increasing invagination of the surface membrane and loss of mesosomes. The cell wall becomes altered in structure and much more permeable, largely due to changes in its lipopolysaccharide fraction." These changes may facilitate metabolite exchange, and make it difficult to separate bacteroids, together with their full metabolite complement for physiological and biochemical study. Alterations in bacterial cell wall structure may be under host cell control, and in peanut the wall may eventually disappear completely to leave protoplasts. During formation of lupin bacteroids, alterations in wall structure may be associated with loss of viability." However, little loss of viability was found in bacteroids isolated from nitrogen-fixing protoplasts of clover root nodules.

An invariable correspondence between nitrogen-fixing activity and the presence of haemoglobin" has been found. Haemoglobin is

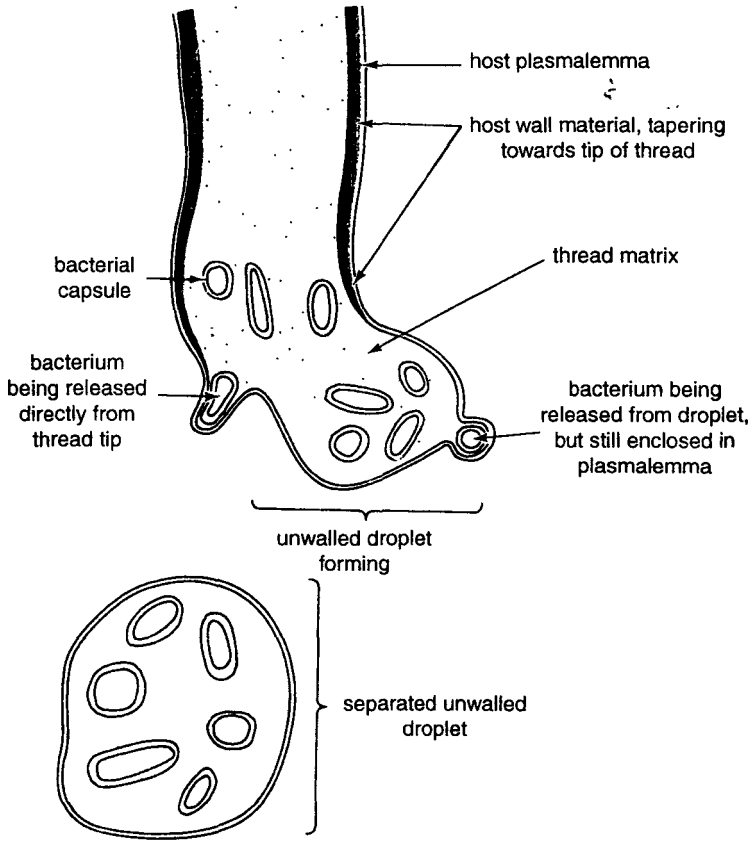


Figure 3.8 : Release of bacteria from an infectin thread.

particularly interesting because of its rarity in the plant kingdom and also because each of its two components appears to be coded by a different symbiont; haem by the *Rhizobium,* and apoprotein by the host. The exact location of haemoglobin is still a matter of dispute, but evidence is accumulating that it may be in the host cytoplasm, rather than in the 'no-man's land' inside the membrane envelope, as was once thought." Its metabolic role, to transport oxygen, will be discussed else where in the text.

3.4.1.5 Senescence

Nodule life is limited. Sooner or later, the infected cells degenerate, becoming greenish brown in the process. This has generally been attributed to the breakdown of haemoglobin into bile pigments (as in

blood), but the evidence is against this.", Studies from several laboratories show that membrane envelopes break down and cell contents degenerate as senescence proceeds. Cells may become invaded by freeliving rhizobia'51, which may later be killed by the action of phytolysosomes.

3.4.1.6 Nodule surfaces and aeration

Since the function of nodules is to fix nitrogen, apart from respiratory considerations, an efficient gaseous exchange system is clearly necessary. Pathways for aeration of soybean nodule have been described, proceeding from the soil, via lenticels on the surface to intercellular spaces in the centre of the nodule, with a possible diffusion barrier somewhere in the cortex. Lenticels seem to be a common feature of the surfaces of determinate nodules such as soybean, *Phaseolus vulgaris*, and cowpea, *Vigna unguiculata;* although the extent of their development is strongly conditioned by water supply. Much less is known about aeration of the indeterminate type of nodule, which apparently does not have lenticels, but which may have gaps allowing gaseous exchange over the whole surface.

Although many nodules respond rapidly to reduced aeration by lowered activity and a switch to more anaerobic metabolism,[66] there is evidence that soybean nodules at least can within a few hours adapt to altered oxygen concentrations and fix nitrogen at the normal rate."

3.4.1.7 Ineffective nodules

Nodules vary widely in effectiveness, i.e., in their ability to fix nitrogen. The factors causing this are poorly understood, but are known to be influenced by environment. Completely ineffective nodules maybe found. These are white in colour, since they do not contain haemoglobin: they are a drain on the host plants' resources, since they use fixed carbon for their metabolism, but do not fix nitrogen in return. In some cases, the bar to effectiveness is a single metabolite-riboflavin in the case of one type of ineffective nodule on clover." Developmentally, ineffective nodules may have a blockage in the final differentiation of bacteroids. Effectiveness and all other stages in nodule development are under the genetic control of both host and *Rhizobium.*

This description of legume nodules has of necessity been very restrictive. The reader is referred to the general literature list, and also to the excellent chapter by Dart, for fuller coverage. It is likely that, as the range of legumes investigated widens, more variations on the common themes will be found.

3.4.2 Non-Legume Root Nodules

The range of non-legumes known to be nodulated increased considerably during the International Biological Programme. It can be seen that a number of families are involved, with no obvious underlying common feature.

3.4.2.1 Infection and development

The whole process of describing nodule initiation and development in non-legumes has been hampered by technical problems. First, the fact that the endophyte has only recently been obtained in pure culture has made work under sterile conditions impossible. Second, many of the plants studied produce quantities of tannins, which obscure the finer structural elements in cells under the light microscope. Third, the presence of suberized host cells and capsulated endophyte makes fixation for electron microscopy difficult, and has led to conflicting reports in the literature.

The following description may need to be modified as new data become available. It is principally based on *Myrica gale. Anus gutinosa, Casuarina cunninghamii, and Comptonia peregrina* (also known as *Myrica asplenifolia).* In experimental work, crushed nodules are normally used as an inoculum, although it is often possible to obtain nodulation by growing plants in soils from areas where the nodulated host plant has been found. The first response to inoculation may be proliferation of the endophyte in the rhizosphere, possibly stimulated by products excreted from roots. Lalonde showed micrographs of *Alnus glutinosa* hairs and epidermal cells, where the cell wall is covered by a membrane-bound external layer, which evaginates to produce material that in due course encapsulates the endophyte prior to infection and forms an exoencapsulation thread.

There is general agreement that root hair deformation and sometimes branching occurs similar to that in legumes although, until recently, direct observations of infection threads have been few. Torrey,78 for example, could find no infection threads in *Casuarina cunninghamii,* he pointed out though that, if only one thread was necessary per nodule, the odds were against it being found by chance. On the other hand, infection threads were found in *Comptonia peregrina.* Of the many thousands of exoencapsulation threads found by Lalonde only one penetrated a deformed root hair; it was suggested that, once a successful penetration has been made, a second penetration in the same region becomes impossible because of some induced host defence mechanism such as tannin deposition.

TABLE 3.6 Genera of non-leguminous angiosperms reported to be nodulated

Genus	*Family*	*Species reported to be nodulated out of total*	*Comments*
Alnus	Betulaceae	22/35	Nodulation a generic attribute
Arctostaphylos	Betulaceae	1/70	Nodules reported from Alaska only
Casuarina	Casuarinaceae	24/45	Nodulation rather variable
Ceanothus	Rhamnaceae	31/55	
Cercocarpus	Rosaceae	4/20	
Colletia	Coriariaceae		One report on *C. paradoxa*
Comptonia	Myricaceae		*C. peregrina* also known as *Myrica asplenifolia*
Coriaria	Coriariaceae	13/5	Nodulation a generic attribute?
Discaria	Rhamnaceae	2/10	
Hippophac	Eleagnaceae	1/3	
Dryas	Rosaceae	3/4	Nodulation a generic at tribute? Nodules found only in Alaska and parts of Canada, not Europe or Japan
Eleagnus	Eleagnaceae	1/3	
Myrica	Myricaceae	26/35	Nodulation a generic attribute?
Purshia	Rosaceae	2/2	
Rubus	Rosaceae		One report on *R. ellipticus,* based on nodule structure only
Shepherdia attribute?	Eleagnaceae	2/3	Nodulation a generic
Trevoa	Rhamnaceae	1 /6	One report on *T. trinervis*

Using plants grown aeroponically or hydroponically, root hair deformation followed by infection and nodulation has been observed with *Comptonia peregrina, Casuarina cunninghamii, Myrica gale,* and *M. cerifera,* all inoculated with *Comptonia endophyte.* Penetration of endophyte occurred at a point where host cell wall folding occurred. Once inside the hair, the endophyte, which is surrounded by host cell wall material, can be seen to penetrate cells of the root cortex. Some associated cell divisions and enlargements are seen just in advance of the penetrating threads. The resultant slight swelling has been termed the primary nodule: it does not develop further. Instead, meristematic activity begins in the root pericycle, presumably due to some hormonal changes analogous to those described for legume nodule initiation. This is the beginning of the true nodule primordium. In *Alnus,* it can be distinguished from a lateral root primordium by its more rounded (i.e., less conical) appearance: in *Casuarina,* the nodule primordium is conical, but always near to infected cells, which in turn are surrounded by tannin-filled and then suberized cells.

There is no evidence that the presence of nodule primordia diminishes the number of lateral roots formed. The nodule primordia remain in an arrested state for some days, and then develop rapidly. Vascular tissue is formed in the centre, and this connects to the stele of the subtending root. Cortical cells become invaded by the endophyte and, in due course, nitrogen fixation begins. In some plants, such as *Alnus,* the nodule primordia never become completely free to develop into roots; the coralloid structure is the result of repeated production of primordia, each developing to a limited extent. In *Casuarina,* on the other hand, the primordia grow free of the endophyte and develop into nodule roots. These nodule roots are generally similar to other ones, except that they grow upwards.

Bond suggested that nodule roots on *Myrica* may assist aeration in boggy habitats, and physiological confirmation of this has recently been obtained." *C. peregrina* is intermediate in the sense that only some of the nodule lobes produce roots, and occasionally these roots develop into true lateral ones. This species also has very rapid development of clusters of nodule lobes and multiple primary nodule primordia. This has been shown in an elegant study using time-lapse photography to study the development of individual clusters of nodule lobes."

Growth of non-legume nodules normally ceases during the dormant season and recommences at the onset of favourable conditions, so nodules are perennial and may reach several centimetres in diameter.

The active tissue is confined to the current season's growth; the older inactive tissue usually becoming lignified.

The above description shows that non-legume root nodules are developmentally and anatomically more related to roots than to the nodules of legumes.

3.4.2.2 Fine structure of infected regions

Initially the infected cells are characterized by a mass of endophyte filaments, which displaces the host cell nucleus to one side. Infections spread from cell to cell in thin threads, and later by more massive crossings until finally the wall between two cells of a radial file may be completely obliterated to give one large cavity. The extent of further differentiation of these filaments appears to be rather varied and partly dependent upon the strain of endophyte used. The arrangement of endophyte regions for *Alnus*.

TABLE 3.7 Comparison of legume and non-legume root nodulation

	Legume	*Non-legume*
Endophyte	*Rhizobium spp*	*Frankia spp*
Reaction to presence of endophyte	Root hair curling and branching	Root hair curling and branching
Method of invasion	Via root hairs or (rarely) epidermis	Via root hairs
Site of nodule initiation	Root cortex	Primary nodule from cortex, nitrogen fixing nodule from pericycle
Location of nodule	Cortical	Central vascular tissue
Location of infected	Central	Cortical tissue
Nitrogen-fixing structure	Bacteroid	? Endophyte vesicle
Pigmentation	Haemoglobin	Anthocyanins/tannins
Longevity	Few weeks to perennial	Perennial

The branched endophyte filaments are septate, contain much apparently nuclear material, and are well supplied with mesosomes. The plasmalemma is surrounded by a well defined cell wall. The endophyte is enclosed by a membrane of plant origin, which is almost certainly derived from the plasmalemma, as in legume nodules. Two zones have been described between this membrane and the endophyte. One, situated immediately inside the host membrane, is generally

termed the capsule and thought to be a layer of pectic material, possibly containing enzymes, and deposited by the host cell. This might be analogous to the infection thread wall of legumes. If so, a further difference between legume and non-legume nodules is that the endophyte in non-legumes never escapes from the thread wall. The second zone or layer is characterized by being electron transparent, and studies using freeze-etching techniques suggest that it may be an artifact resulting from shrinkage of the endophyte during the fixation process."

Young, infected host cells show a reduction in the number of starch grains and a rise in the number of mitochondria and other cell organelles, which suggest considerable metabolic activity. In older infected cells, the tips of the endophyte filaments swell to form vesicles. When the filaments are much branched, the appearance of a cluster of vesicles is similar to that of a bunch of grapes. Akkermans showed that these vesicles have very high reducing power, one of the requirements for nitrogen fixation; more recently, Akkermans and coworkers have isolated a vesicle preparation and shown that it has nitrogenase activity. This provides the first *direct* information that the bacterial partner is responsible for nitrogen fixation in non-legume nodules. It also associates activity with vesicles. This is consistent with the observation that ineffective nodules of *Myrica faya* lack vesicles." However, nitrogen fixation was detected in *Comptonia* before vesicles had visibly developed."

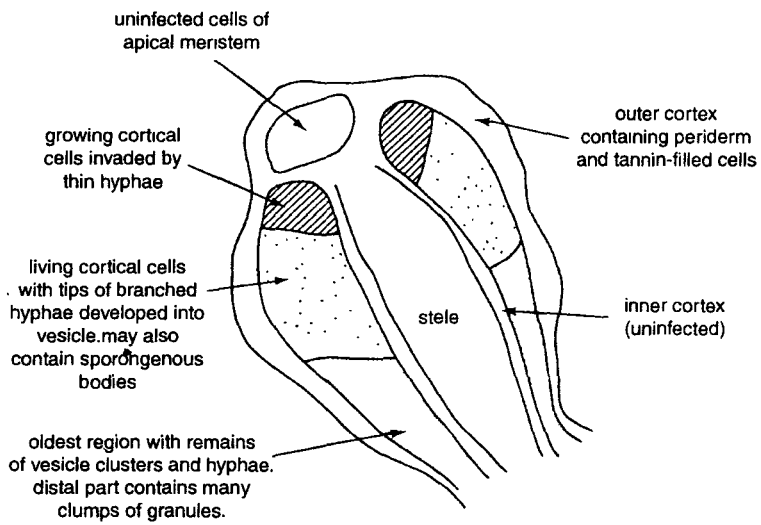


Figure 3.9 : Diagram of endophyte regions in the lobe of an Alnus nodule.

Descriptions of later stages in endophyte development within the nodule are hampered by a great deal of confusion in terminology. Old (senescent) infected cells apparently die; the endophyte is no longer sheathed by capsular material and is seen to be more or less completely divided into fragments which have been variously described as 'bacteroids', 'granules' and 'spores'. The exact appearance of these structures may be related to the method of preparation and, also, to the strain of endophyte used. Whether or not they act as resting or as reproductive structures is not yet clear, although rapid progress may be expected now that spore-bearing filaments have been obtained in pure culture." For further discussion of these points, the papers of van Dijk and Mercus, and Gardner should be consulted.

3.4.2.3 Infection of a non-legume by Rhizobium

One of the most interesting and exciting discoveries of recent years has been the work of Trinick, who found root nodules on a member of the Ulmaceae. The host was originally identified as *Trema aspera,* then reidentified as *T. cannabina,* and now re-classified as *Parasponia sp.* The nodules contained, not an Actinomycete-like organism, but *Rhizobium.* They reduced acetylene; nodulated plants grew well in nitrogen-free medium, and isolates from them were able to nodulate two species each of the legume genera *Vigna* and *Macroptilum.* Structurally, the nodules were similar to those of a typical non-legume, having a centrally located vascular strand, an apical meristem and an infected zone within an enlarged cortex.[93] Squashes of nodules showed that many of the bacteria were aggregated into threads from which structures resembling bacteroids were released. Haemoglobin could not be detected. It will be interesting to see if more reports of this type of association appear in the future.

3.4.3 Blue-Green Algal Symbioses

Blue-green algae associate with almost all groups of plants, forming a range of nitrogen-fixing symbioses. Most of these are less rigid than in the nitrogen-fixing root nodules in that, with a few exceptions, the blue-green algal component can grow and fix nitrogen independently of its partner. Although the blue-green algae fix both carbon and nitrogen when free living, in symbiotic association the ability to reduce carbon is often lowered or even entirely suppressed, so that its main role is to fix sufficient nitrogen for both partners. Parallel with this change in metabolic emphasis, an increased heterocyst frequency and a much reduced rate of growth are often found.

Another general difference between blue-green algal symbioses and

root nodules is that the endophyte in the former usually invades structures such as leaf cavities, that are normal parts of the host plant's morphology, rather than inducing the formation of specialized structures (nodules). Sometimes, as in coralloid roots of cycads and the cavities of *Blasia,* invasion is followed by anatomical modifications.

Of the symbioses between members of the plant kingdom, those between fungi and algae which form lichens are unique in having a characteristic morphology for the symbiotic state, which is classified as a separate 'organism'.

3.4.3.1 Lichens

Without exception, the nitrogen-fixing lichens have a blue-green algal component and a fungal symbiont, and they may, additionally,

TABLE 3.8 Genera of lichens where nitrogen fixation has been established using either C_2H_2 reduction or ^{15}N uptake

Lichen	*Blue-green genus*	*No. of species tested*
Collema	*Nostoc*	9
Dendriscocaulon	*Scytonema*	1
Ephebe	*Stigonema*	1
Leptogium	*Nostoc*	4
Lichina	*Calothrix*	2
L obaria	*Nostoc*	2
Massalongia	*Scytonema*	1
Nephroma	*Nostoc*	2
Pannaria	*Nostoc*	2
Parmeliella	*Nostoc*	2
Peltigera	*Nostoc*	7
Placopsis	not established	1
Placynthium	*Dichothrix*	2
Polychidium	*Scytonema*	1
Pseudocyphellaria	*Nostoc*	1
Solorina	*Nostoc*	3
Stereocaulon	*Nostoc*	2
Sticta	Nostoc	2

have a green algal symbiont. Thus, we have 2 and 3 membered symbioses. In the 2 membered nitrogen-fixing species, the blue-green filaments may be interspersed throughout the fungal mycelium, as in *Collema*; or they maybe confined to a distinct layer, usually just beneath

the surface, as in *Peltigera horizontalis*. In 3 membered symbioses, such as *P. aphthosa,* the blue-green symbiont is restricted to more or less spherical structures, known as cephalodia, where they are confined by a fungal layer. A full list of the established nitrogen-fixing lichens is given in Millbank. Within each lichen genus, only one blue-green genus has been found; although in several cases, such as the *Peltigera* spp mentioned above, the location of the bluegreen alga may vary between species.

In 2 membered symbioses, the blue-green algal member has the dual roles of carbon and nitrogen fixation. The heterocyst frequency in these is about 4 per cent. Where there are three members, the green algal symbiont photosynthesizes, and the blue-green algal symbiont has 20-30 per cent heterocysts, presumably reflecting a concentration on nitrogen fixation rather than carbon fixation. There is evidence that most, if not all, of the nitrogen fixed is released into the lichen thallus to become more-or-less evenly distributed. The rates of fixation per unit of blue-green algal protoplasm appear much higher in 3 than in 2 membered lichens. Morphological and biochemical modifications take place in *Nostoc* when in symbiotic association in *Peltigera canina* and *P. aphthosa.* The symbiotic cells show much lower inhibition of nitrogenase by combined nitrogen than in the free-living forms, and there is a virtual absence of stored nitrogen. These features may be fairly general to symbiotic bluegreen algae and will be discussed later.

One of the important properties of lichens is their ability to withstand extremes of environment-heat, desiccation, and in some species salinity. Their ecological importance is, thus, in pioneer habitats and in inhospitable regions of the earth, such as Arctic tundra and deserts. They may be important in some forest situations. All these aspects will be considered later. The slow-growing nature of most lichens (1 mm a year being quite common, 2 cm exceptional) and their great sensitivity to atmospheric pollutants, such as sulphur dioxide, make research on them difficult without denuding the environment of natural material.

3.4.3.2 Bryophytes

Although five genera of liverworts have been reported to contain bluegreen algal symbionts, only three *Anthoceros, Blasia,* and *Cavicularia* have been studied with respect to nitrogen fixation. The following description is based on *Blasia pusilla* and *Anthoceros punctatus,* both of which have *Nostoc* sp, probably *N. sphaericum,* as endophyte. These thalloid liverworts have mucilaginous cavities on the ventral

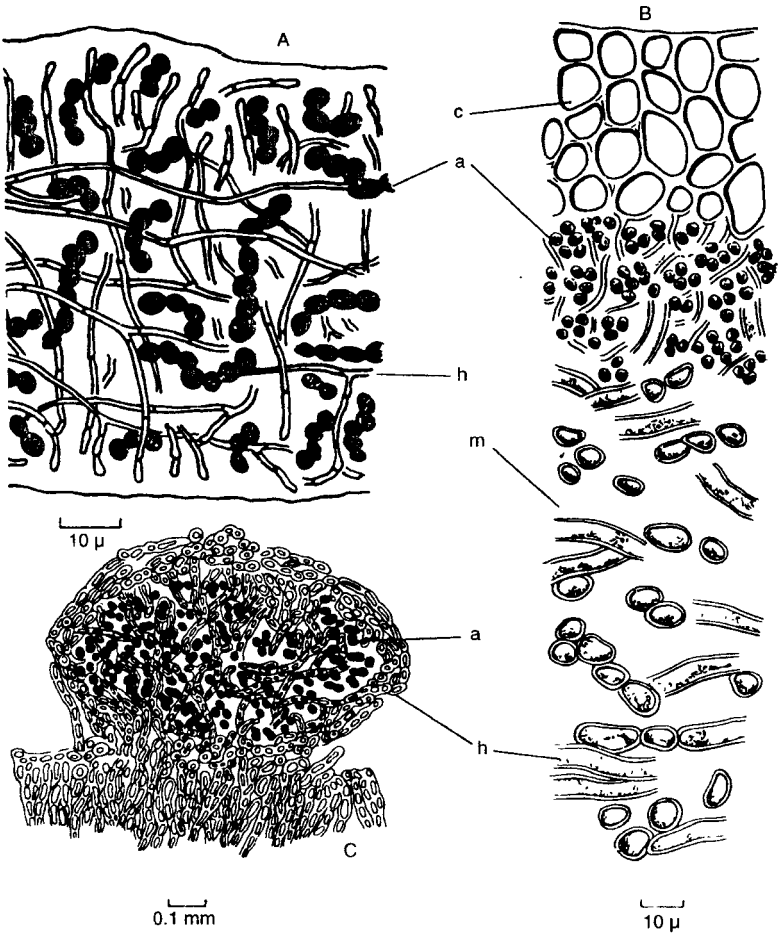


Figure 3.10 : Structure of some lichens. A, simple two-membered lichen as exemplified by Collema; B, more advanced two-membraned lichen with algae confined to a central layer, as in Peltigera horizontalis; C, section through a cephalodium, a structure found in some three-membered lichens, containing the blue-green symbiont: c, cortex; h, hypha; m, mucilage; a, algal cells. Reproduced from Millbank.

side that become infected in the presence of suitable *Nostoc*. The endophyte may obtain entry by means of motile hormogonia. Once inside, the *Nostoc* develops extensively, but produces far more heterocysts (30-40 per cent) than in the free-living state (4-5 per cent). In response to the presence of the *Nostoc*, the liverwort produces branched septate filaments from a region opposite the cavity pore. These grow through the algal colony and greatly increase contact

between the two symbionts. The filaments have wall ingrowths typical of transfer cells, which area type of cell found in regions of rapid movement of solutes. In both liverworts, only those regions which contain *Nostoc* show nitrogenase activity and this is generally related quantitatively to heterocyst frequency.

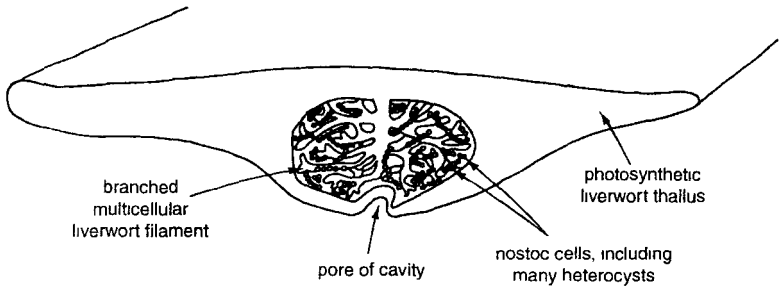


Figure 3.11 : Diagram to show the arrangement of Nostoc cells in a cavity of Blasia thallus.

The observation that the endophyte lacks the structured granules typical of the free-living forms suggests that it does not store fixed nitrogen. The isolated endophyte excretes large quantities of ammonia, whereas freeliving cells do not. All these data suggest a role for the *Nostoc* rather similar to that of rhizobia, i.e., they reduce nitrogen to ammonia, which is assimilated in the host cells. Excised *Nostoc* does not fix $^{14}CO_2$ in the light, although it has polyhedral bodies similar to those known to contain the key Calvin cycle enzyme, ribulose bisphosphate carboxylase. Can one infer from this that fixed carbon is supplied to the endophyte by the host liverwort? Critical confirmation of this point comes from a series of elegantly simple experiments using *Anthoceros,* which forms sporophytes easily in culture. These sporophytes are photosynthetic but do not contain *Nostoc* colonies. ^{15}N fixed by the *Nostoc is* transported to both the gametophyte and sporophyte. By darkening the gametophyte and sporophyte in turn, both were found to provide photosynthetically fixed carbon, almost certainly as sucrose, to the *Nostoc,* supporting nitrogenase activity.

3.4.3.3 Pteridophytes-Azolla

Azolla is a genus of aquatic ferns, a fact which is in itself unusual. In addition, it is the only fern genus known to be symbiotic with a nitrogenfixing endophyte, namely *Anabaena azollae.* There are about 6 extant species of *Azolla,* and many more fossil species have been recorded. The distribution of *Azolla is* world wide, and it is variously considered to be a weed which blocks waterways, and a valuable source

of nitrogen for agriculture. The classic description of modern forms of *Azolla* can be found in the monograph of Strasburger, and his drawing of a leaf containing *Anabaena is* still widely reproduced. The sporophyte generation consists of a small floating rhizome with branches and simple roots at some of the nodes. The apex has specialized hairs to which *Anabaena* filaments adhere. Young leaf primordia invaginate on the lower side of the developing dorsal lobe, forming a cavity that traps *Anabaena* and rapidly becomes sealed off. The cavity is thought to be filled with mucilage (like those of liverworts, described above), and it is certainly lined with hairs. These hairs are multicellular with the dense contents and pronounced wall ingrowths typical of transfer cells. It is doubtful whether these hairs are formed in response to endophyte infection, as in the liverworts, since they were found to be present in uninfected fronds grown on combined nitrogen. Somewhat unexpectedly, the hairs were found to be separated from the leaf vascular traces by a layer of highly vacuolate cells with relatively few plasmodesmatal connections-a situation which does not seem ideal for transport purposes.

Although at least two research groups have obtained *Anabaena-free Azolla,* there are no fully substantiated reports of growth of *Anabaena azollae* on its own. In symbiosis, *Anabaena* growth is at least partly synchronized with leaf growth; division of *Anabaena* ceases long before the cavity is filled. As growth proceeds, the number of heterocysts increases, the proportion of vegetative cells decreases, and the individual vegetative cells increase slightly in length and markedly in width. Hill suggested that the heterocyst frequency increases due to nitrogen depletion and the nitrogen fixed in consequence is used by the fern to maintain *Anabaena* filaments at the apex in a vegetative (nonheterocystous) form, ready for incorporation into new leaf cavities. Increasing heterocyst frequency with leaf maturity is accompanied by an increase in nitrogenase activity until senescence suddenly sets in at about the 20th leaf from the apex.

The isolated symbiont and the intact symbiotic leaves show light saturation curves for nitrogenase activity. All cells show relatively high reducing potential, which possibly indicates high respiratory activity. Under aerobic conditions in the light, ammonia does not inhibit nitrogenase activity, a feature that has already been mentioned in connection with nitrogen-fixing lichens as a possible adaptation to the symbiotic state.

3.4.3.4 Cycads

The Cycadales are an order of the gymnosperms that form an

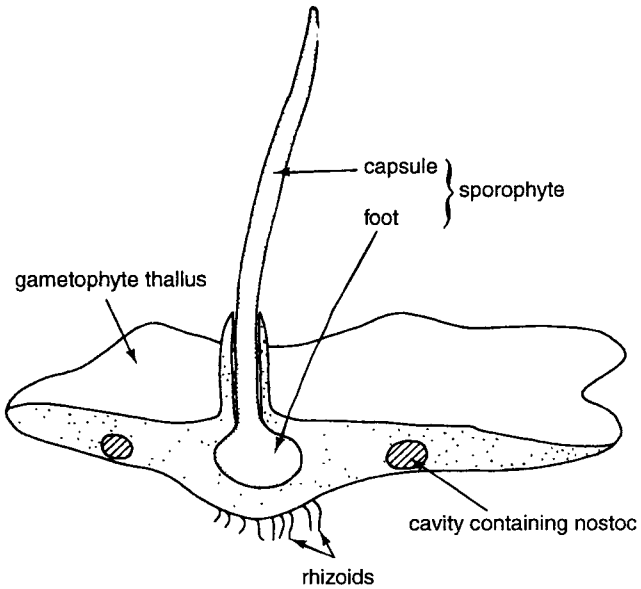


Figure 3.12 : Diagram of Anthoceros punctatus to show the arrangement of sporophyte and gametophyte generations.

appreciable part of the flora of southern Africa, South America, and Australia. For example, 26 species of the genus *Encephalartos* have been found in South Africa. All cycads produce apogeotropic roots, in addition to the usual root system. These apogeotropic roots frequently occur near the surface of the soil and extend above ground, branching dichotomously. They may also be found below ground. Their origin lies in the pericycle of 'true' roots, and in all respects they appear to be modified lateral roots. Infection of the roots, which are usually referred to as coralloid roots, by blue-green algae is *very* common in the field.

Like the endosymbionts of bryophytes, those of cycads can be cultured. Although their taxonomy has not been completely worked out, there is general agreement that species of *Nostoc* and *Anabaena* are involved. In spite of much speculation over nearly a century, the mode of infection of the roots by the endophyte is still not understood. This may seem surprising, but an indication of the time and patience required to study the mode of infection can be gauged from the report of Nathanielsz and Staff who examined 800 root branches from 80 seedlings of *Macrozamia communis* and found only one entry point. This entry point occurred under an external mat of endophyte filaments, and was very small. Once inside the root, a large mass of filaments had separated

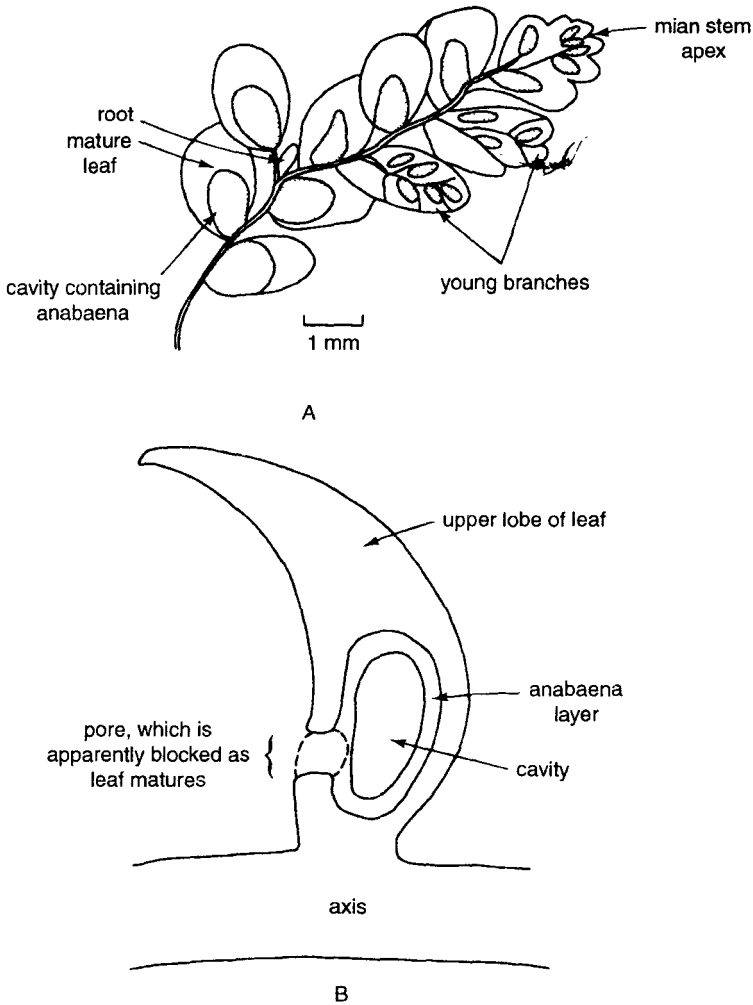


Figure 3.13 A, Drawing of Azolla frond from the undersurface to show location of Nostoc. After Peters B, Diagram of section through a dorsal lobe of a single leaf of Azolla.

and destroyed some of the cells of the outer layers. The general assumption is that filaments must penetrate through natural gaps in the outer layers and/or through breaks caused by the emergence of roots. Lenticels, which are common on coralloid roots, are now not considered to be a likely point for invasion, since they develop too late.

Interpretation of the structure and growth of coralloid roots after

infection has also been rather confused. The brief description that follows is based on Milandasuta, who worked with 7 species from 5 different genera. Milandasuta distinguished young coralloid roots from normal roots by their less massive root cap (only a few cells deep). The root cap gives rise to a secondary cortex (also called the outer cortex), which sheathes the root. The innermost layer of the root cap, the protoderm, gives rise to the endophyte layer. This is a very distinctive layer of cells, which contains nearly all of the endophyte. In the young parts of coralloid roots, transverse division and cell enlargement in the protoderm keep pace with growth of the root cap and secondary cortex. Subsequently, the protoderm enlarges radially, while the secondary cortex enlarges longitudinally. Cells of the protoderm thus separate along their radial and transverse walls, leaving narrow tangential walls in contact with adjacent layers. The corresponding layer in endophyte-free roots is poorly developed. Thus, the host seems to react to the presence of its symbiont. As the spaces between the protoderm cells develop, they become invaded by endophyte.

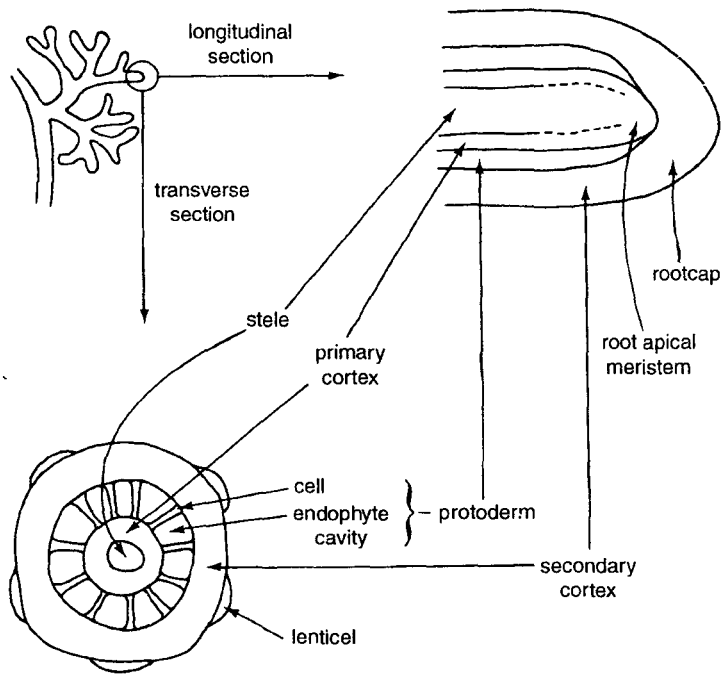


Figure 3.14 : Diagram to illustrate the general appearance and structure of coralloid roots of cycads.

In this work by Milandasuta, as in all other reports except one, the endosymbiont is always stated to be extracellular. The exception associated all extracellular symbionts with the presence of mucilage (as found in liverworts and *Azolla),* and noted that some intercellular spaces in the secondary cortex and the primary cortex contained endophyte; additionally, in a few roots only, intracellular endophyte was found, and an invasion sequence was described in which the initial act was entry of granular mucous material of endophyte origin, which was followed by endophyte cell penetration. If this report is confirmed, then the cycad coralloid roots represent an intermediate stage in evolution-from strictly extracellular location of endophyte in bryophytes and *Azolla,* to intracellular infection in the angiosperm, *Gunnera.* The association of infected regions with mucus suggests a possible nutritional relationship and it may be that the host plant provides carbon sources for the blue-green alga, as in other associations. There are many reports of nitrogen fixation by coralloid roots.

3.4.3.5 Angiosperms-Gunnera

The only genus of angiosperm recorded with a blue-green algal symbiont *is Gunnera,* of which all 40 species may be infected with *Nostoc punctiforme.* The genus varies from small creeping plants, to herbs with leaves over a metre in length. Both symbiotic partners are capable of independent existence. *Gunnera* produces excretory glands just behind the shoot apex, and the cells of these become invaded by *Nostoc. As* the shoot develops, *Nostoc* forms dense nodules within the stem. These nodules are grouped in threes, apparently associated with the insertion of a petiole in the stem.

The details of nodule-cell fine structure have not been fully resolved: there is no doubt that *Nostoc* cells are intracellular, which means that this symbiosis is unique among those involving blue-green algae. There is a clear developmental sequence in nodules taken from different positions along an advancing stolon of *G. albocarpa.* Those in the sub-apical region contain most nitrogen and have the highest nitrogenase activity. Compared with the situation in free-living *Nostoc,* heterocyst frequency is high throughout, but increases with age. By 15 cm, the cytoplasm of the host cell containing *Nostoc* disintegrates, and extra cell wall material of host origin surrounds the endophyte cells, which in turn begin to disintegrate.

As with the liverwort and *Azolla* symbioses, there is a general relationship between heterocyst frequency and nitrogenase activity in young nodules, but the relationship is less clearly defined in older

nodules. Silvester considers that conditions in nodules may be such that the vegetative cells are functionally heterocysts. The nodules are normally buried by the stem cell layer, by leaf bases, and often by soil as well. Not surprisingly, therefore, in spite of the dark green colour of the *Nostoc,* photosynthesis does not occur, the endophyte living heterotrophically on carbohydrate presumably produced by the host plant. Excised nodules do respond to light by increased nitrogenase activity, but this is not accompanied by appreciable $^{14}CO_2$ fixation and is thought to result from photo-production of *ATP.*

TABLE 3.9 Attributes of Gunnera albocarpa-Nostoc punctiforme nodules of different ages

Distance from advancing stolon apex, cm	*n mol C_2H_4 mg protein^{-1} min^{-1}*	*Percentage N in nodules*	*Percentage heterocysts*
0	12	25	23
5	36	40	32
10	26	28	41
15	02	27	63
20	01	20	65

Free-living *N. punctiforme* had 52 per cent heterocysts. Uninfected host tissue had 1 5 per cent N.

Although apparently unique, the *Gunnera-Nostoc* symbiosis is very efficient. The advanced nature of the association is indicated by the intracellular location of the endophyte, and the well developed vascular supply to the nodules. Fixed nitrogen is rapidly transferred to the macrosymbiont and apparently in sufficient quantity to supply all of its needs (72 kg ha^{-1} a^{-1}).

3.4.3.6 Diatoms

Reports of blue-green algae contained within diatoms have appeared from time to time. Little is known of their physiology except that they are able to fix nitrogen.

3.4.3.7 General comments on the blue-green algal symbioses

It will be apparent by now that the different blue-green algal symbioses have many features in common. Perhaps the biggest variation is in the choice of symbiotic partner. As has been mentioned earlier, the association is in some ways looser than that between root nodule symbionts, in that (a) in nearly all cases the endophyte can grow and fix nitrogen independently, and (b) that the nitrogen-fixing symbiont

TABLE 3.10 Modifications to blue-green algae in symbiotic association with other plants

	Symbiotic form	Free-living form
Morphology and anatomy		
Heterocyst frequency	high	low
N-reserve granules	rare	common
Physiology and biochemistry		
Photosynthetic activity		
carbon fixation	low	high
photophosphorylation	high	high
Nitrogenase activity	very high	not so high
Ammonia excretion	high	low
Ammonia assimilation	by alternate symbiont, using GDH (?)	using GS
Nitrogenase inhibition by combined nitrogen	low	high

invades regions that are already present, sometimes producing modifications, rather than inducing the formation of special structures (nodules). This invasion, particularly of cavities such as those in liverworts, is almost certainly facilitated by the gliding movements characteristic of hormogonia.

Functional modifications to the blue-green algal symbionts are summarized in Table.One of the interesting areas of current research is the cause of these modifications. Does the alternate partner produce substances that alter the metabolism of the endophyte, or does the endophyte become modified in response to an altered environment? Perhaps both mechanisms are operative. It is likely that components of the physical environment, such as low oxygen tension, are involved; this is one factor that might be common to the great diversity of symbiotic associations.

4

Microbial Yeasts

Medically important yeasts are those fungi of primarily unicellular growth habit that are capable of producing or contributing to diseases of humans and animals. Of 361 species in 39 genera of yeasts in the second edition of Lodder's The Yeasts, and of 434 species recorded by Barnett and Pankhurst, only about 24 now meet the above definition. Although *Candida lusitaniae* and *Hansenula polymorpha* have recently been added to the list of species of yeasts that are able to cause human disease, the number of pathogens does not appear to be expanding dramatically. The percentage of pathogenic yeasts among, the totality of species remains quite small. However, these newcomers to the species reported to cause human disease do require that clinical laboratory personnel better understand the characteristics that delimit yeast species and the methods that can be used to detect those characteristics.

A unicellular growth habit is maintained when these yeasts are actively growing (log phase) under standard temperatures, aeration, pressures, and humidity. Their colonies therefore are glabrous, moist, creamy or membranous in texture, and lack the aerial hyphae that impart a fluffy or velvety texture to the colonies of filamentous fungi (molds). It is true that many yeasts can form filaments as either true or pseudohyphae, but these are produced under diminshed O_2 such as may prevail in submerged portions of solid culture media, in an atmosphere of 5 to 10% CO_2, or in the tissues of a parasitized host.

Normal vegetative cells of yeasts are round or oval, 2.5 to 6 μm in diameter, and reproduce asexually by budding, bud fission, or fission. Buds (blastoconidia) can remain attached to the mother cell and continue to bud, producing branching clusters of blastoconidia. Individual blastoconidia adhering to their neighbors in a chain can elongate to

produce filaments called pseudohyphae. True septate hyphae result from germination of "transitional cells" (rounded or flattened blastoconidia, or chlamydospores).

According to their method of sexual reproduction, yeasts can be divided into three main groups, Ascomycetes *(Saccharomyces, Endomycopsis, Pichia, Hansenula,* and *Nematospora),* Heterobasidiomycetes *(Leucosporidium, Filobasidiella,* and *Syringospora),* and Deuteromycetes, or Fungi Imperfecti *(Candida, Cryptococcus, Rhodotorula, Torulopsis,* and *Trichosporon).* Ultrastructural and biochemical criteria may be used to suggest affinities of species to higher taxa. DNA hybridization techniques, used increasingly often, may help to clarify relatedness among different isolates.

Unicellular, budding fungi whose cell walls contain melanins are arbitrarily excluded from the species discussed here. A complete discussion of these "black yeasts" (classified in the *Dematiaceae*).

4.1 CLINICAL SIGNIFICANCE

Yeast infections are among the most common fungal infections affecting humans; their incidence has greatly increased with the advent of broad-spectrum antibiotics, immunosuppressive corticosteroids, and antitumor agents. Their severity ranges from benign and localized (either transient or chronic) to disseminated and sometimes fatal. Yeast fungemia occurs frequently in patients with indwelling catheters and can result in endocarditis or pyelonephritis in recipients of organ transplants, artificial heart valves, or other prosthetic devices. *Candida* infections of the eye result from injury to the cornea (keratitis) or can involve the retina (endophthalmitis) as one manifestation of candidiasis spread by hematogenous dissemination. *Candida* infection of the eye is serious and can result in the loss of visual function or enucleation. Yeast endocarditis is also frequent in drug addicts who inject themselves intravenously using nonsterile syringes or needles.

Yeasts exist in nature in a wide variety of substrates, including fruits, vegetables, and homemade fermented beverages. *Candida albicans,* the most frequent pathogen, is a normal inhabitant of the alimentary tract; various surveys show an incidence of 20 to 40% in asymptomatic individuals. Another yeast, *Rhodotorula glutinis, is* a common inhabitant, in the tropics, of moist skin; its role as a pathogen is questionable, although it has been reported to cause fungemia.

Because of the frequent association of yeasts with the internal and external environment of humans, their incidence in clinical specimens is rather high; e.g., yeast colonies grew in cultures from 2,017 of

7,685 clinical specimens received at the mycology laboratory of Columbia-Presbyterian Medical Center during the 1982-1983 fiscal year. During this time, 2,271 cultures were positive for fungi of all types, making yeasts the predominant type of fungi isolated. Most of these yeast isolates were recovered from urine cultures. These numbers may denote increased interest in or awareness of urinary tract infections by the medical staff during this period. Nevertheless, it is estimated that less than 10% of patients examined actually had systemic yeast infections. Hence, collaboration between physicians and laboratory is essential in assessing the possible etiological role of a yeast growing in a clinical culture. Besides clinical and laboratory evaluation of the patients, critical studies should include microscopic examination of fresh specimens and serological tests for specific antigens and antibodies, as well as cultural isolation of such yeasts.

4.2 DIRECT EXAMINATION

Identification of yeasts in a clinical specimen begins with the direct microscopic examination of stained or unstained samples of the specimen. Such examination does not permit exact species identification, but does provide early clues which aid the physician in making a presumptive

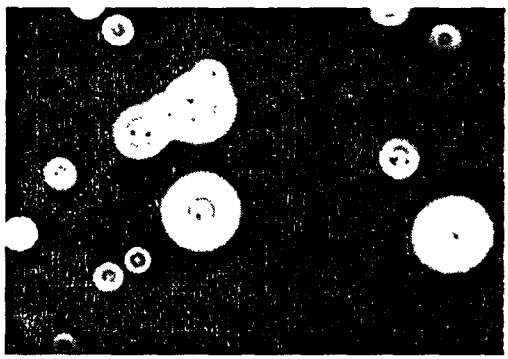


Figure 4.1 : *Cryptococcus neoformans* Nigrosin-stained wet preparation of sediment from centrifuged cerebrospinal fluid. Note encapsulated cells, some with wide capsules which probably originated in vivo; narrow capsules are probably on daughter cells "budded out" in vitro. × 1,200.

diagnosis and aids the laboratorian in deciding which further cultural and biochemical tests are needed. The presence of yeasts in properly collected specimens of body fluids which are normally sterile is immediately significant; however, the significance of yeasts in naturally contaminated specimens such as sputum, feces, and urine depends on other considerations. The important findings from direct microscopy

TABLE 1. Cultural and biochemical characteristics of yeasts frequently isolated from clinical specimens[a]

Species	Growth at 37°C	Pellicle in broth	Pseudo/true hyphae	Chlamydospores	Germ tubes	Capsule, India ink	Assimilation of[b]: Glucose	Maltose	Sucrose	Lactose	Galactose	Melibiose	Cellobiose	Inositol	Xylose	Raffinose	Trehalose	Dulcitol	Fermentation of: Glucose	Maltose	Sucrose	Lactose	Galactose	Trehalose	Cellobiose	Urease	KNO_3 utilization	Phenol oxidase	Ascospores
Candida albicans	+	−	+	+	+	−	+	+	+	−	+	−	−	−	+	−	+	−	F	F	−	−	F	F	−	−	−	−	−
C. guilliermondii	+	−	+	−	−	−	+	+	+	−	+	+	+	−	+	+	+	+	F	−	F	−	F*	F	−	−	−	−	−*
C. krusei	+	+	+	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	F	−	−	−	−	−	−	+*	−	−	−
C. lipolytica[c]	+	+	+	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
C. lusitaniae[d]	+	−	+	−	−	−	+	+	+	−	+	−	+	−	+	−	+	−	F	−	F	−	F	F	F	−	−	−	−
C. parapsilosis	+	−	+	−	−	−	+	+	+	−	+	−	−	−	+	−	+	−	F	−	−	−	−	−	−	−	−	−	−
C. pseudotropicalis (*C. kefyr*)	+	−	+	−	−	−	+	−	+	+	+	−	+*	−	+*	+	−*	−	F	−	F	F*	F	−	−	−	−	−	−
C. rugosa	+	−	+	−	−	−	+	−	−	−	−	−	−	−	+*	−	−	−	−	−	−	−	−	−	−	−	−	−	−
C. stellatoidea	+	−	+	+	+	−	+	+	−	−	+	−	−	−	+	−	+	−	F	F	−	−	−	−	−	−	−	−	−
C. tropicalis	+	+	+	−	−	−	+	+	+	−	+	−	+	−	+	−	+	−	F	F	F	−	F	F	−	−	−	−	−
Torulopsis candida	+	−	−	−	−	−	+	+	+	+*	+	+	+	−	+	−	+	+*	W	−	W	−	−	W	−	−	−	−	−
T. glabrata	+	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	+	−	F	−	−	−	−	F	−	−	−	−	−
T. pintolopesii	+[e]	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	F	−	−	−	−	−	−	−	−	−	−
Cryptococcus neoformans	+	−	R	−	−	+	+	+	+	−	+	−	+	+	+	+*	+	+	−	−	−	−	−	−	−	+	−	+	−
C. albidus	−*	−	−	−	−	+	+	+	+	+*	+*	+	+	+	+	+	+	+*	−	−	−	−	−	−	−	+	+	−	−
C. gastricus	−	−	−	−	−	+	+	+	+*	−	+	−	+	+	+	−	+	−	−	−	−	−	−	−	−	+	−	−	−
C. laurentii	+*	−	−	−	−	+	+	+	+	+	+	+*	+	+	+	+*	+	+	−	−	−	−	−	−	−	+	−	−	−
C. luteolus	−	−	−	−	−	−	+	+	+	−	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	+	−	−	−
C. terreus	−*	−	−	−	−	+	+	+*	−	+*	+*	−	+	+	+	−	+	−*	−	−	−	−	−	−	−	+	+	−	−
C. uniguttulatus	−	−	−	−	−	+	+	+	+	−	−*	−	−*	+	+	+*	−*	−	−	−	−	−	−	−	−	+	−	−	−
Rhodotorula glutinis	+	−	−	−	−	−*	+	+	+	−	+*	−	+	−	+	+	+	−	−	−	−	−	−	−	−	+	+	−	−
R. rubra	+	−	−	−	−	−*	+	+	+	−	+	−	+*	−	+	+	+	−	−	−	−	−	−	−	−	+	−	−	−
Saccharomyces cerevisiae	+	−	−*	−	−	−	+	+	+	−	+	−	−	−	−	+	+*	−	F	F	F	−	F	F*	−	−	−	−	+
Trichosporon beigelii	+*	+	+	−	−	−	+	+	+	+	+	+*	+*	+	+	+*	+*	+*	−	−	−	−	−	−	−	+*	−	−	−
T. pullulans	+*	+	+	−	−	−	+	+	+	+*	+	+*	+	+*	+*	+*	+	−	−	−	−	−	−	−	−	+	+	−	−
Geotrichum candidum	−*	+	+	−	−	−	+	−	−	−	+	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−
G. capitatum	+	+	+	−	−	−	+	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
Hansenula anomala	+*	−	−	−	−	−	+	+	+	−	+	−	+	−	+	−	+	−	F	F*	F	−	F	−	−	−	+	−	+

[a] *, Strain variation, R, rare, F, the sugar is fermented (i.e., gas is produced); W, weak fermentation. Based on data from Ahearn and Schlitzer (3) and Kreger-Van Rij (34).

[b] +, Growth greater than that of the negative control.

[c] *C. lipolytica* assimilates erythritol, *C. krusei* does not.

[d] *C. lusitaniae* assimilates rhamnose; *C. tropicalis* does not.

[e] *C. pintolopesii* is a thermophilic yeast capable of growth at 40 to 42°C.

[f] Occasional strains of *C. tropicalis* produce teardrop-shaped chlamydospores.

include: (i) the presence or absence of encapsulated yeasts, (ii) the presence or absence of pseudohyphae, (iii) the presence or absence of true hyphae and arthroconidia, (iv) the size and shape of the yeast; and (v) the number of buds and nature of their attachment to the mother cell.

4.3 COLLECTION, TRANSPORT, AND STORAGE OF SPECIMENS

Specimens to be examined for yeasts of possible etiological significance can be collected in the same way as bacteriological specimens, with certain precautions and exceptions. A minimum of 5 ml of cerebrospinal fluid is essential for adequate sampling when culturing for *Cryptococcus*.

Whenever possible, specimens for yeasts should be examined and cultured at the bedside, or a portion of the specimen should be fixed as a dry smear or preserved in 10% Formalin for microscopic evaluation of the abundance and morphology of the yeast at the moment of collection. For bedside cultures, Sabouraud agar slants in screw-capped tubes or bottles should be available wherever patients are examined.

To facilitate handling, specimens received by the laboratory should be grouped according to type of material submitted.

4.3.1 Smears or Scrapings

Smears or scrapings on glass slides can be examined by adding a drop of saline solution or 10% NaOH. The latter should be used with fecal smears, thick exudates, or skin scrapings. Stains are seldom necessary, but yeasts can be detected in smears stained by the Gram, Wright, Ziehl-Neelsen, Giemsa, Papanicolaou, periodic acid-Schiff, or Gomori methenamine silver methods. However, it is needless to duplicate efforts by preparing special smears for yeasts from specimens that will also be examined routinely for bacteria and cytological abnormalities. Gram-stained smears from the vagina are quite acceptable for detecting yeasts, and yeasts can readily be observed in specimens submitted for cytological examination and stained with Papanicolaou stain. Yeasts are also effectively revealed by staining heat-fixed smears with Loeffler methylene blue or with Calcofluor white.

4.3.2 Cerebrospinal Fluid or Thoracentesis Specimens

Cerebrospinal fluid or thoracentesis specimens should be centrifuged or passed through a membrane filter to concentrate solids and increase the probability of finding yeasts. A drop of sediment or a piece of the

membrane filter (cut and handled aseptically) can then be examined. India ink or nigrosin may be added by capillarity under the cover slip as a negative "capsule stain" to reveal *Cryptococcus,* if present. Familiarity with the cytological features of *Cryptococcus* is essential to avoid confusing this yeast with blood cells, lipid globules, starch granules, or other artifacts. Examination of India ink stains of living cultures of *Cryptococcus* suspended in body fluids can contribute to this experience. Besides wet preparations, dry smears can be stained as described above, and alcian blue and mucicarmine staining are particularly useful for demonstrating *Cryptococcus neoformans* in smears of sediment from cerebrospinal fluid.

4.3.3 Urine specimens

To assess the significance of *Candida spp. in* urine while handling the specimen appropriately for detecting C. *neoformans* and other fungi, it is necessary to proceed differently than for bacterial cultures of urine. It is generally accepted that 10^4 CFU of *Candida* spp. per ml of urine is suggestive of *Candida* pyelonephritis. Although this value is lower than that accepted as being significant for bacterial pyelonephritis (10^5 CFU/ml), numbers of *Candida* colonies lower than 10^4/ml are regarded as being due to contamination from the lower genitourinary tract and not indicative of infection. Hence one must quantitate yeasts in urine by streaking the specimen with a quantitative loop on an appropriate medium and to examine a sample of the uncentrifuged

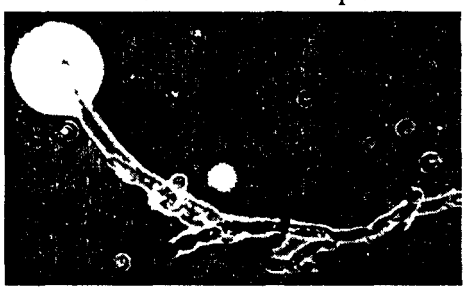


Figure 4.2 : *Cryptococcus neoformans.* Unusual strain which produced pseudohyphae. Nigrosin-stained wet preparation of sediment from centrifuged cerebrospinal fluid.

urine microscopically. On the other hand, a single CFU of C. *neoformans, Histopl-asma sp.,* or other systemic mycotic agent in the entire urine specimen would suggest infection. This necessitates microscopic examination of centrifuged urinary sediment as well as plating some sediment on isolation medium when a fungus culture is specifically requested, although this practice is contraindicated with bacterial cultures.

Figure 4.3 : *Trichosporon beigelii* in fresh unstained preparation of sputum. ×600. Note arthrospores and blastospores characteristic of the genus *Trichosporon*.

4.3.4 Sputum or other Mucous Secretions

It has been suggested that identification procedures used with yeasts isolated from sputum specimens be limited to those essential for detecting C. *neoformans*. However, the isolation of the same species from multiple specimens (sputum, urine, skin lesions, etc.) obtained from an immunocompromised patient suggests infection by that species. For that reason, we recommend that all yeasts isolated from seriously ill patients be identified, whatever the specimen source.

4.3.5 Biopsy or Autopsy Tissue

Slice tissue with a sterile scalpel and examine all surfaces to detect ulcerations or granulomatous areas which usually contain the fungi.

Because of the possible occurrence of yeasts and yeastlike fungi as part of the transient or resident flora of the body, and because they may proliferate rapidly in vitro, microscopic examination of clinical specimens must be done soon after collection. All specimens should be delivered to the diagnostic laboratory as soon as possible, preferably within minutes of collection.

4.4 CULTURE AND ISOLATION

A diversity of agar media which contain salts of metals such as bismuth and molybdenum, or which contain a tetrazolium dye, are commercially available and provide early clues toward identification of yeast colonies based on color reactions. Precaution: bacteria can produce similar color reactions, and therefore microscopic examination of suspensions made from suspicious colonies is essential. Such media

are especially useful for detecting individual colonies in a mixed population of more than one yeast. An excellent differential medium for presumptive identification of C. *neoformans* on primary isolation plates can be prepared by adding to Littman oxgall agar (or to other commonly used media) an extract of seeds from the Indian thistle plant, *Guizotia abyssinica* (a seed common in birdseed mixtures, hence, "birdseed agar"). C. *neoformans* characteristically produces brown colonies on this medium within 1 week when incubated at 30°C, a property not shown by other yeasts including other *Cryptococcus* species. "Birdseed agar" is a general name for the medium first described by Staib and later modified by others to make it more selective or more chemically defined. The color reaction of C. *neoformans* colonies on this medium is mediated by a phenol oxidase which leads to the formation of melanin pigment in the yeast cell walls. Substitute formulations for birdseed agar have been proposed, and optimal conditions for production of melanin pigment by C. *neoformans* have recently been detailed.

Ability to form brown colonies on birdseed agar is one indication of phenol oxidase activity. Various phenols and diaminob-enzenes including caffeic acid and dihydroxyphenylalanine are good substrates for the phenol oxidase produced by C. *neo formans;* however, the oxidase does not react with tyrosine. A rapid pigmentation test that requires only 6 h has been described. A commercial medium containing dihydroxyphenylalanine that is designed to screen yeast isolates for phenol oxidase activity was recently described and has been shown to be reliable for this purpose. Most companies dealing in prepared culture media now offer one or another formulation of birdseed agar.

As stated there, and in our experience, the lysis-centrifugation method for recovering yeasts from blood is recommended.

Whichever isolation medium is employed, human pathogenic yeasts will form pasty, opaque colonies at both 37°C and room temperature. For early presumptive identification of yeasts producing pseudohyphae, specimens positive on direct microscopic examination can be cut directly into cornmeal agar as well as planted onto other isolation media.

4.5 IDENTIFICATION

4.5.1 Recommended Procedures

Methods for identifying yeasts are presented in Figure elsewhere in this chapter; the cultural and biochemical characteristics of the species most often encountered in clinical specimens are summarized

in Table elsewhere in this chapter. Initial steps include: (i) making a wet mount and stained smears for microscopic observation; (ii) making an India ink preparation; (iii) performing a germ tube test; and (iv) inoculating Wolin-Bevis or cornmeal agar to detect pseudohyphae and chlamydospores. While these tests are being done, each distinct colony type should be restreaked on bacteriological medium, such as brain heart infusion agar with or without 10% blood and incubated at 37°C, or Sabouraud agar incubated at 25°C, to ensure obtaining pure cultures for further biochemical testing. Media for subculturing yeasts before biochemical testing should not contain antibiotics. Inoculation of biochemical test media directly from isolation media is not recommended.

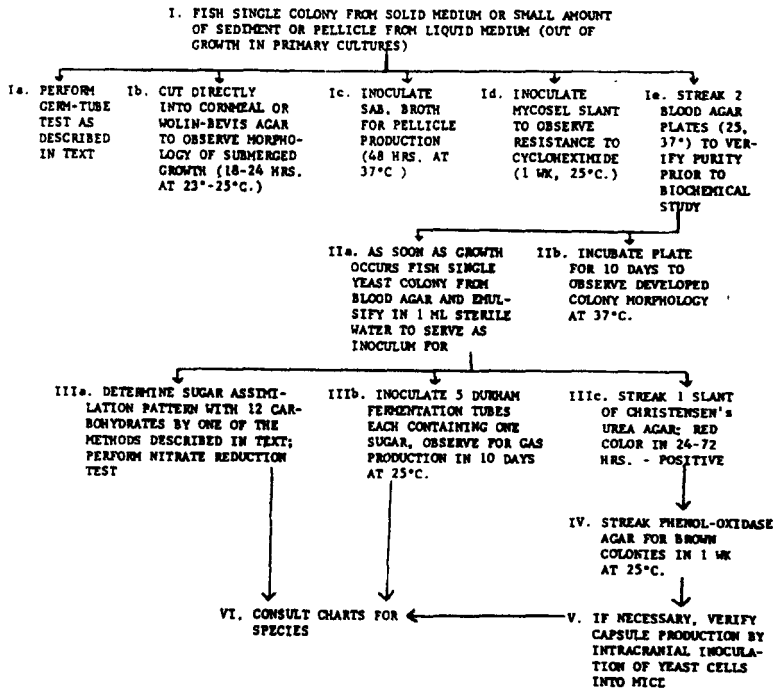


Figure 4.4 : Schema for a step-by-step procedure that can be followed in the isolation and identification of yeasts from primary cultures of clinical specimens.

Species identification based on phenotypic characteristics of anamorphic species is often fraught with difficulty. Some species are separable on the basis of a single biochemical test. A case in point is identification of *Candida stellatoidea* by inability to utilize sucrose as a carbon source. Many yeast taxonomists regard C. *stellatoidea* as a

variant of *Candida albicans* that lacks an a-glucosidase. For simplicity, we retain C. *stellatoidea* as a separate species while recognizing that this species validity is questionable. It does illustrate the dilemma that arises when we try to identify anamorphic form-species.

To complicate matters, mutations affecting phenotypic expression of certain characters in yeasts can be induced by cytotoxic cancer medications. Yeasts isolated from patients receiving such chemotherapy may be suppressed in metabolizing certain carbohydrates, or may require biotin or other growth-factor supplementation to behave appropriately in biochemical tests. Fortunately, correct biochemical reactions often will be obtained upon repeated subculture of the isolate.

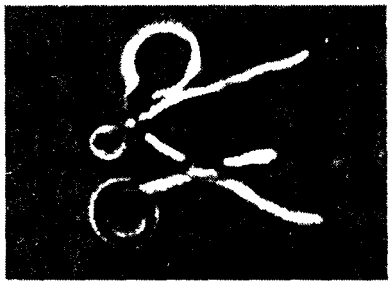


Figure 4.5 : *Candida albicans*. Germ tubes formed after incubation in serum for 2 h at 37°C. × 2,100

Devices combining several biochemical tests into a single minaturized plate or plastic strip make the use of a battery of tests very convenient. Such devices encourage more extensive biochemical testing of isolates and, in some instances, provide faster results. Automation of biochemical testing of yeasts now provides the capability for carrying out extensive biochemical testing with minimal effort and within as little as 20 to 24 h. The accuracy of identifications achieved with commercial manual and automated devices is, in general, very acceptable. Still, it is unwise to abandon the study of microscopic morphology of yeast isolates grown on Wolin-Bevis agar or a similar medium as part of the identification process. Most commercial devices rely upon the germ tube test as an additional safeguard. Moreover, microscopic morphology on media designed to elicit typical morphology of yeasts will usually aid in interpreting biochemical tests. Some species (e.g., *Trichosporon* and *Geotrichum spp.)* are difficult to distinguish without careful examination of microscopic morphology.

4.5.2 Cultural Characteristics

Most yeasts grow well on the common mycological and bacteriological

media, producing visible colonies within 48 to 72 h. Most pathogens grow readily at both 25 and 37°C, whereas most saprophytic yeasts encountered in clinical specimens fail to grow at 37°C. The ability to grow at 37°C is an important characteristic for differentiating species. Some species are inhibited by cycloheximide or chloramphenicol. In certain species, ascospor-ulation may be enhanced by the use of special media. The growth and temperature requirements for the several genera and the indications for using special media are summarized later on in this chapter.

Pellicle formation on broth. When yeasts are cultivated in tubes of liquid media such as Sabouraud broth or malt extract broth, growth takes the form of (i) a sediment, (ii) a ring around the circumference of the broth surface, (iii) a film on the surface, or (iv) a surface pellicle. These characteristics depend largely on O_2 requirements and surface tension. In the past, emphasis was placed on the ability to form a surface pellicle on liquid media, but more recent evidence has demonstrated this to be a variable characteristic, and it is no longer stressed as much as formerly. However, the aspect of growth in liquid media can help identify species; e.g., pellicle-forming yeasts often isolated from clinical specimens include *Candida tropicalis, Candida krusei,* and especially *Trichos-poron spp.*

4.5.3 Microscope examination of culture growth

Wet preparations of primary cultures in distilled water, India ink, nigrosin, or lactophenol cotton blue, or dried smears stained by the Gram or Ziehl-Neelsen method, provide material for microscopic examination of morphology and ascertaining the presence or absence of capsules, the presence or absence of ascospores, and the purity of cultures. Further morphological study demands special media. Incubation of yeast cells in blood serum or egg white at 37°C for 1 to 4 h permits observation of germ tube production *(C. albicans).* Subculture by cut-streaking cornmeal infusion agar or Wolin-Bevis agar plates, followed by incubation at 23 to 25°C for 18 to 48 h, permits detection of pseudomycelium, true mycelium, and chlamydospores. Special sporulation media may be necessary to detect production of ascospores, whose presence and characteristic morphology are essential for identification of ascosporogenous genera, such as *Saccharomyces, Pichia, Endomycopsis, Hansenula,* and *Nematospora,* occasionally found in clinical specimens.

Most of these morphological studies can be carried out directly with primary cultures before purification, and even with the original specimen.

4.5.4 Germ Tube Test

One of the most valuable tests for rapid presumptive identification of *C. albicans is* the germ tube test. One need only make a dilute suspension of a yeast colony in 0.5 to 1.0 ml of serum, which is accomplished by touching the tip of a sterile Pasteur pipette to a yeast colony and then gently emulsifying the cells which adhere to the pipette in the serum. The mixture is then incubated at 37°C for 2 to 4 h, after which a drop of the mixture is examined microscopically for germ tubes. For convenience, the pipette can be left in the serum during incubation and used to transfer a drop of the serum to a slide for examination. When so incubated, individual cells of *C. albicans* and *C. stellatoidea* produce short lateral hyphal filaments called germ tubes which are not produced by other *Candida* species. We have been unable to verify recent reports of germ tube formation by *C. tropicalis* and regard this species as being germ tube negative under the test conditions described here. *C. stellatoidea* and *C. albicans* are easily differentiated by a sucrose assimilation test and by susceptibility to cycloheximide. *C. albicans* assimilates sucrose and grows on commercial cycloheximide-containing media, i.e., Mycosel; *C. stellatoidea* does neither. Human and bovine sera including commercially available fetal calf serum, bovine serum albumin, ovalbumin, dilute oxgall, or peptone can be used with success in the germ tube procedure. Dehydrated and liquid media for use in the germ tube test are available commercially. Known strains of *C. albicans* and *C. tropicalis* should be included as positive and negative controls, respectively.

4.5.5 Chlamydospores

Another standard technique for identifying yeasts is to test for production of chlamydospores. Various media are available for this test, including chlamydospore agar, Wolin-Bevis agar, cornmeal-polysorbate (Tween 80) agar, and rice-Tween 80 agar. A medium was described recently which permits detection of germ tubes when incubated at 37°C for 3 h, demonstration of typical yeast morphology after incubation for 18 to 24 h at 25°C, and the phenol oxidase reaction of *C. neoformans,* all on the same medium. Choice of medium is based largely on individual experience. The technique described here utilizes Wolin-Bevis or cornmeal-Tween 80 agar. Spherical chlamydospores are produced by *C. albicans* and *C. stellatoidea*, thus requiring further differentiation as described above. *C. tropicalis* occasionally produces distinct oval- to teardrop-shaped chlamydospores on the media mentioned, which are also useful for inducing the formation

of pseudohyphae or true hyphae in most species of *Candida* and of true hyphae and arthroconidia in *Trichosporon.*

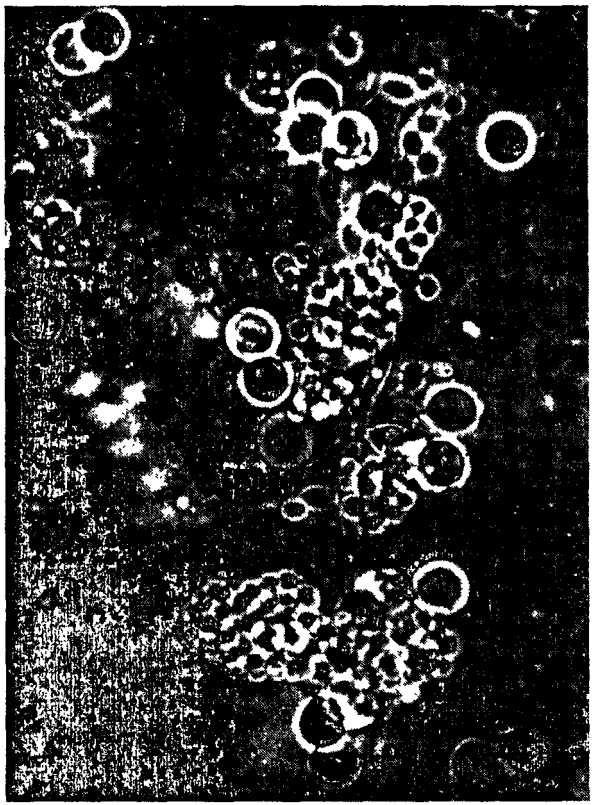


Figure 4.6 : *Candida albicans* Submerged mycelium with spherical clusters of blastoconidia and double-walled, spherical chlamydospores on cornmeal agar × *1,200.*

For the test, cut the yeast colony or original specimen into the agar along two or three parallel lines, no less than 1 cm apart. Alternatively, the Dalmau technique, in which the inoculum is streaked along two shallow scratches in the agar and then covered with a sterile cover slip, can be used. Incubate the plates at room temperature (23 to 25°C) for 18 to 48 h, by which time most strains of *C. albicans* and *C. stellatoidea will* have formed typical chlamydospores. This technique is also useful for demonstrating giant pseudohyphae of *C. parapsilosis.* The experienced eye can also recognize the typical morphological patterns of other *Candida* species. Confirmation by biochemical reactions of purified cultures is recommended.

4.5.6 Biochemical Reactions

Before any biochemical reactions are measured, it is essential to purify yeast cultures as described above, even when contam-ination is not immediately apparent, for a hidden contaminant can significantly affect results. Biochemical tests should not be inoculated from media containing antibiotics because the antibiotics may suppress, but not kill, bacterial contaminants. Once the culture is transferred to biochemical test media, the growth of bacterial contaminants may no longer be inhibited and they will multiply rapidly, confounding interpretation of results.

Biochemical reactions useful in yeast identification include: (i) assimilation (or utilization) of carbohydrates; (ii) assimilation of KNO_3 (equivalent to nitrate reduction); (iii) fermentation reactions (i.e., gas production); (iv) ureas production; (v) ability to produce brown colonies on birdseed agar (or other suitable phenol oxidase detection medium); and (vi) resistance to cycloheximide. Other biological characteristics tested are: (i) ability to grow at 37°C; (ii) ability to form a pellicle on broth; and (iii) animal pathogenicity.

Besides the above routine tests, certain other characteristics are currently exploited by research investigators as taxonomic criteria for yeasts. Some of the following procedures may eventually be adapted for use in clinical laboratories: (i) analysis of cell wall components by gas-liquid chromatography, nuclear magnetic resonance spectra, and methylation techniques; (ii) determination of the guanine-plus-cytosine ratio in DNA of a given strain; (iii) amino acid sequencing of cytochrome c or other proteins; (iv) somatic cell or DNA hybridization techniques; (v) determination of optimal temperature, temperature tolerance limits for growth, and other nutritional imbalances; and (vi) detection of polyols (e.g., arabinitol) produced in liquid cultures and quantitated by gas-liquid chromatography.

On the flow sheet for identification procedures and in the subsequent discussion, a description is given of the methods we recommend as routine procedures. Alternative methods are also listed.

4.5.7 Urease

Urea is split by yeasts having urease, including all species of *Cryptococcus* and *Rhodotorula* and some species of *Trichosporon;* indeed, this reaction is shown by all heterobasidiomycetes. Occasional strains of C. *krusei* are urease positive. The test is performed by transferring with a sterile loop a portion of a purified yeast colony to a slant of Christensen urea agar. The slant is incubated at 25 to 30°C and

examined daily for 4 days. Change in the amber color of the phenol red indicator of the medium to pink or red denotes a positive reaction due to alkalinization of the medium via production of ammonia from the urea. A rapid urease test (15 min) has recently been described for use with yeasts.

4.5.8 Carbohydrate Assimilation Tests

Tests to determine the ability of a yeast species to utilize a carbohydrate as sole source of carbon in a chemically defined medium have long been a mainstay of yeast taxonomists and have become an essential step in yeast identification in the clinical mycology laboratory. Most commercial products now available for identifying yeasts rely heavily on carbohydrate assimilation tests.

Assimilation tests provide a clear biochemical basis and shorten the time required for identifying yeasts to the species level. Several methods are available as follows.

(i) Auxanography. In Beijerinck's auxanographic technique, small amounts of dry carbohydrates are placed on the surface of a heavily seeded synthetic agar medium. Noticeably greater growth encircles the spot where an assimilated compound was placed as compared with background growth. Lack of enhanced growth indicates a lack of enzymes for utilizing the test carbohydrate; the pattern elicited by utilizable carbohydrates is an auxanogram. Although mutations involving enzymes essential to carbohydrate utilization are well known for yeasts, the auxanogram is a depen-dable criterion for identification of each species when used in conjunction with other tests. Modification of Beijerinck's method include: (i) streaking the surface, rather than seeding the agar plate; (ii) using filter-paper disks impregnated with the carbohydrates; (iii) placing drops of carbohydrate solutions on the agar; and (iv) placing the carbon sources in wells cut out of the agar in a petri dish.

Auxanography is the method currently employed for assimi-lation tests. Large plastic petri dishes (150 mm in diameter) are filled in advance with 30 ml each of reconstituted Wickerham yeast nitrogen base solidified with 2% Noble agar and refrigerated. For the test, the plate is streaked heavily, evenly, and confluently with a cotton swab, transferring the entire growth from a 24-h Sabouraud agar slant culture of the yeast to be tested. Immediately after the plate is streaked, 12 "carbohydrate disks" (glucose, maltose, sucrose, lactose, galactose, melibiose, cellobiose, inositol, xylose, raffinose, trehalose, and dulcitol) are applied to the surface of the plate with a BBL SensiDisc dispenser.

Cartridges containing small disks of the 12 carbohydrates are obtained from the same company. Additional carbohydrates may be needed for certain species, e.g., rhamnose utilization to separate *Candida lusitaniae* and C. *tropicalis*.

A modified auxanographic method makes use of an indicator dye (bromocresol purple) for early, sensitive detection of growth. Here a lower concentration of YNB (0.1 ×) than is usually recommended is employed, and the. concentration of each carbohydrate is adjusted for optimum utilization. Correlation between yeast growth with utilization of a test carbohydrate and change in the color of the indicator due to acid production is also demonstrated.

(ii) Wickerham broth technique. The technique originally recommended by Wickerham and Burton uses the same chemically defined medium as discussed above, but in liquid form. The broth tubes containing individual sugars are inoculated either with a depleted suspension of the yeast (grown for 48 h in basal medium without a carbon source or with a drop from a dilute suspension

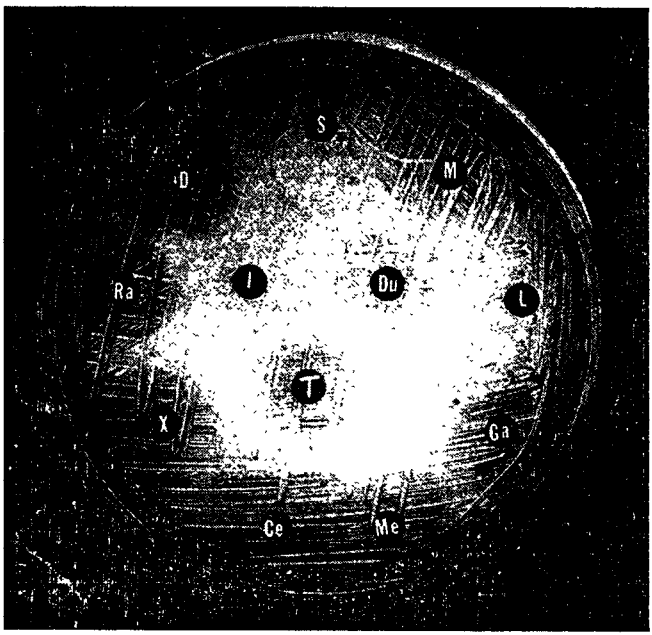


Figure 4.7 : *Candida guilliermondii*. Auxanographic plate (15 cm in diameter) showing growth around disks of glucose (D), sucrose (S, partly rendered transparent by moisture), maltose (M). galactose (Ga), cellobiose (Ce), xylose (X), raffinose (Ra), dulcitol (Du), and trehalose (T) (two-thirds normal size).

of yeast colony in distilled water). Twelve tubes, each containing one of the sugars, are inoculated for each isolate, the tubes are incubated at room temperature (25°C), and growth is evaluated turbidimetrically. When a carbon source is assimilated, growth clouds the medium so that black lines or letters will not show through. Ahearn and Schlitzer recommend shaking the tubes during incubation. When screw-capped tubes are used for this technique, the caps should be loosened slightly (one-half turn) to prevent anaerobiosis. A disadvantage of liquid medium is that a falsepositive reaction may occur if progeny from a single mutant cell render the entire broth turbid upon prolonged incubation. Such a mutant would appear as a single distinct colony on the surface of an agar plate and thus would not lead to a false-positive reading. By the same token, latent or delayed utilization of carbohydrates can be detected more easily with the broth method than on agar media.

(iii) Assimilation agar slant technique. Individual carbohyd-rates can also be incorporated in agar slants of Wickerham YNB adjusted to pH 7.0. Addition of an indicator (bromocresol purple) draws attention to early growth by a change in pH to the acid side (indicator turns yellow). For inoculation, a dilute suspension of the yeast to be tested is prepared by suspending a single colony in 9 ml of sterile distilled water. This suspension is inoculated in 0.1-ml volumes onto each assimilation slant and then incubated at 30°C and observed for growth. Assimilations are considered positive when abundant growth appears on a test medium (and the indicator changes to yellow) with negligible or no growth on the control slant (indicator remains purple); results may be reported in 72 h, but slants should be held for 14 days to allow for delayed reactions. Assimilations are considered negative if a carbohydrate medium is not different from the control medium, i.e., the indicator does not change to yellow.

4.5.9 Nitrate assimilation

Tests of ability of a yeast to utilize nitrate as sole nitrogen source can be carried out by any one of these same procedures, except that Wickerham yeast carbon base (Difco) is used as the basal medium and potassium nitrate (final concentration, 0.078 g/100 ml) and peptone (1%) are used as the test nitrogen sources. The peptone serves as a positive growth control.

A rapid (15-min) test for nitrate utilization by yeasts has proved reliable at Columbia-Presbyterian Medical Center. This test detects nitrite, the end product of nitrate reductase, by exposing the yeast to the dehydrated ingredients of a special nitrate broth on a pretreated

cotton swab, subsequently dipped in a few drops of an equal mixture of 0.5% a-naphthylamine and 0.8% sulfanilic acid, each in 5 N acetic acid. Standard cotton swabs are pretreated by dipping in a 5 × concentration of the special nitrate medium (KNO3, 2 g; NaH2PO4 • H20,11.7 g; Na_2HPO_4, 1.14 g; Zephiran chloride, 1.2 ml of a 17% solution; 200 ml of water). The saturated swabs are dried in a desiccator overnight, autoclaved, and stored in a closed container until needed. For the test, the tip of a pretreated swab is swept over several colonies of the yeast in question, and then the swab is inserted in a test tube and pressed against the side of the tube to increase contact between the yeast and the reagents in the swab. The tube and swab are incubated for 10 min at 45°C (or for 2 h at 37°C). At the end of the incubation period, the swab is immersed in 6 drops of the anaphthylamine-sulfanilic acid mixture (3 drops each), and the color reaction, if any, is compared with that developed by two similarly treated and incubated swabs, one using a yeast known to be nitrate positive and the other with a nitrate-negative yeast. The reaction of the test organism is graded as follows, when compared with the negative control: 0 = no color change; 1+ = faint pink; 2+ = pink; 3+ = dark pink to light red; 4+ = dark red.

In our experience the rapid nitrate and traditional nitrate assimilation tests correlate very well for commonly isolated, yeasts. However, some infrequently isolated species, including *Rhodot-orula spp.*, may give a positive rapid nitrate test and yet be negative for nitrate assimilation. This may be a reflection of a cryptic nitrate reductase enzyme which is detected by the rapid test but poorly expressed in the assimilation test.

4.5.10 Fermentation Reactions

Carbohydrate fermentation tests are familiar, useful tests for yeast identification. However, these tests are more prone to variation and are less dependable than carbohydrate assimilation tests. The only reliable evidence for carbohydrate fermentation by yeasts is by production of gas; therefore, Durham tube inserts should be employed for gas detection. The basal medium employed includes peptone, yeast extract, and bromocresol purple indicator; the pH is adjusted to 7.0. Carbohydrates commonly tested include glucose, maltose, sucrose, lactose, galactose, and trehalose. For inoculation, a dilute suspension of yeast cells in sterile distilled water is prepared (to match a McFarland no. 1 standard), and a single drop is added to each tube. After inoculation each tube should be sealed with 1 ml of molten

Vaspar (equal parts of paraffin and white petrolatum) placed directly onto the top of the broth. Readings are made after incubation at 25°C for 10 to 14 days. Positive fermentation is indicated by accumulation of gas (CO_2) in the Durham tube or underneath the Vaspar seal (the indicator will also change to yellow). All fermented carbohydrates will also be assimilated; however, certain carbohydrates will be assimilated, but not fermented.

A very few isolates have been reported that were capable of fermenting a particular carbohydrate in the peptone-based fermentation medium, but were unable to grow in the, chemically defined YNB used in assimilation tests because biotin or other vitamins were not present in adequate amounts for their growth. Supplementation of YNB with 0.01% yeast extract usually will permit utilization of the carbohydrate in question without yielding false-positive results with other carbohydrates.

4.5.11 Serological Methods

Fluorescent-antibody techniques with specific conjugates are extremely useful in those instances in which cultures have failed and yet yeasts are observed in histological sections, body fluids, or exudates. However, at present these specific conjugates are available only at the Centers for Disease Control and at certain other reference laboratories. These techniques, along with agar gel precipitin tests with soluble antigenic extracts, slide or tube agglutination, or immunoelectrophoresis with soluble extracts, also are useful in specific differentiation of yeasts, provided suitably absorbed, species-specific antisera are employed.

A serological identification of C. *neoformans* has recently been described which utilizes a soluble extract of colonies allowed to react with latex particles coated with antibody. The same commercial kit used for detecting cryptococcal antigen in patients' body fluids can serve for this test and has been shown to be reliable for confirming the identification of a yeast isolate as C. *neoformans*.

4.5.12 Differentiation from Related Species

Since yeasts, as stated, are a heterogeneous conglomerate of various fungal taxa, this chapter by necessity points out differences among yeast genera and species. It is also important to mention certain other organisms encountered in clinical specimens which produce yeastlike colonies on commonly used isolation media. For the most part, morphological differences between these microorg-anisms and yeasts will be observed and must be identified by methods other than those used for yeasts. The following are examples of such microorganisms.

Geotrichum. *Geotrichum is* a filamentous fungus whose initial growth may be glabrous and creamy to pasty, but becomes velvety or fuzzy with aging and repeated subculture. *Geotrichum* produces true hyphae and arthroconidia, but neither pseudohyphae nor blastoconidia, thus differing morphologically from the yeast genus *Trichosporon.* Further differentiation can be obtained by physiological and biochemical tests. Fluffy strains of *Geotrichum* superficially resemble the systemic pathogen *Coccidioides immitis* and must be differentiated from it.

Geotrichum and *Trichosporon spp.* have recently been carefully reevaluated by Weijman, who analyzed their extractable carbohydrates by gas-liquid chromatography. Species having similarities to ascomycetes were retained in the genus *Geotrichum,* whereas those which contained xylose, a basidiomycetous characteristic, were placed in the genus *Trichosporon.* Weijman thereupon renamed *Trichosporon capitatum* as *Geotrichum capitatum.* Recently Salkin et al. proposed the new genus and species *Blastoschizomyces pseudotrichosporon* as the correct epithet for 13 isolates that had been identified as *T. capitatum,* but formed annelloconidia. It is unclear at this writing whether all isolates of *T. capitatum* should be renamed *B. pseudotrichosporon* or whether some isolates are named *G. capitatum.* It may be necessary to retain all three names! Recent reports have documented the pathogenicity of this (these) species. It will be difficult for most laboratories to distinguish between annelloconidia and arthroconidia in these fungi with the microscopes usually available in clinical laboratories. The most practical solution might be to accept *B. pseudotrichosporon* as the correct name for isolates that were formerly called T. *capitatum.* Many mycologists separate *Geotrichum* spp. from *Trichosporon spp.* based on the failure of *Geotrichum spp.* to form blastoconidia.

The colorless alga *Prototheca,* which is also encountered in clinical specimens (and may be a pathogen), produces colonies resembling those of *Cryptococcus* on Sabouraud agar. Although one species of *Prototheca* produces a capsule resembling that of *Cryptococcus,* other *Prototheca* species differ from *Cryptococcus* in not having a capsule, in having a less refringent cell wall, and especially in producing characteristic endospores by internal division into eight (or more) compartments. *Prototheca spp.* can also be differentiated by staining with specific immunofluorescent conjugates and by carbohydrate assimilation tests.

Ustilago. The basidiomycetous genus *Ustilago is* another filamentous fungus which can occasionally grow from sputum specimens and produce

yeastlike colonies. One isolate was reported to cause chronic meningitis in a human. That identification, however, was based solely on the morphology of the fungus in tissue; it was not isolated in culture. At first unicellular, *Ustilago spp.* soon produce short hyphae with clamp connections, and their colonies become finely powdery or velvety. Even at the unicellular stage, the elongated cells characteristic of this genus differentiate it from yeasts. These elongated cells resemble arthroconidia. Most isolates that we have encountered from clinical specimens are nitrate positive and give physiological reactions similar to those of *Cryptococcus albidus*.

Besides the species listed above, occasional isolates of *Sporothrix schenckii* and of the black yeasts *Aureobasidium pullulans* and *Exophiala werneckii* produce white to tan yeastlike colonies on primary isolation media. Careful examination of the microscopic morphology of these isolates growing on cornmeal agar will facilitate their correct identification and avoid the confusion that may result from carrying out biochemical tests with them.

4.5.13 Animal Inoculation

Demonstration of pathogenicity as an aid to differentiation of *C. neoformans* from other cryptococci can be accomplished by intracranial inoculation of 0.05 ml of saline suspension containing a 0.2% concentration of yeast cells (1:500 by volume of packed cells) into each of two 20- to 30-g white mice. Alternatively, 0.5 to 1 ml of the yeast cell suspension can be injected intravenously or intraperitoneally into mice. Most strains of *C. neoformans* kill mice within 4 to 5 days after intracranial injection; examination of brain tissue from infected mice clearly demonstrates encapsulated yeast cells.

The pathogenicity of *C. albicans* can be demonstrated by injecting 0.2 to 0.8 ml of a 1% suspension of packed cells into the marginal ear vein of a 1- to 2-kg rabbit, or a similar volume of a 0.2% suspension into the tail vein of a 20- to 30-g mouse. Death of the animal usually occurs within 1 week; at autopsy miliary abscesses in the kidneys, and sometimes in the spleen and liver, will be observed. Pathogenicity of C. *albicans* can also be demonstrated by chicken embryo inoculation.

4.6 CHARACTERISTICS OF IMPORTANT YEAST GENERA

4.6.1 Genus *Candida*

Candida is a heterogeneous genus presently classified within the

family *Cryptococcaceae,* Fungi Imperfecti (Deuteromycetes). *Candida spp.* are frequently present as members of the normal flora of the mouth, throat, large intestine, vagina, and skin and are often contaminants in exudates or other specimens taken from these areas. In patients whose immune defenses have been compromised by disease or by the secondary effects of drugs used to treat their diseases, normal flora organisms may invade deeper tissues, producing severe, life-threatening infections. *C. albicans,* the principal pathogenic species, causes mild to severe or chronic superficial infections of skin, nails, and mucous membranes in individuals with normal immune defenses, as well as serious systemic infections in debilitated patients. *Candida parapsilosis, C. tropica*lis, and *Candida guilliermondii* have become important causes of endocarditis, pyelonephritis, arthritis, and disseminated candidiasis in patients with indwelling intravenous catheters, patients undergoing cardiovascular surgery, and drug addicts.

C. lusitaniae has recently been reported as a cause of human disease. It appears to be an important pathogen in immuno-compromised patients because it readily develops resistance to amphotericin B. It has been confused with *C. tropicalis* and C. *parapsilosis* because of similarities in biochemical reactions. *C. lusitaniae* differs from *C. parapsilosis* by assimilating cellobiose and from *C. tropicalis* by fermenting as well as assimilating cellobiose and by assimilating rhamnose.

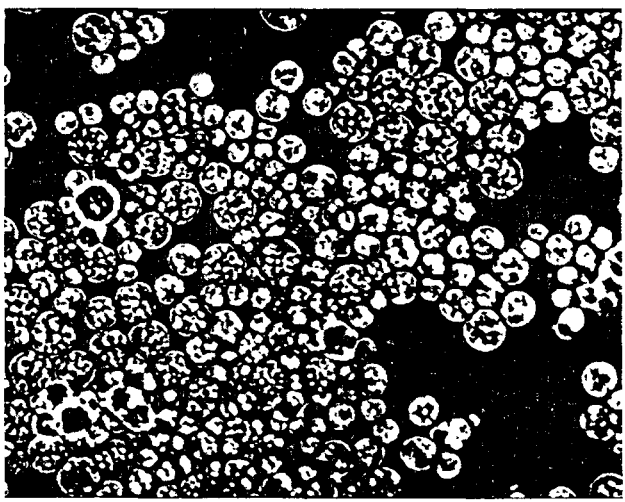


Figure 4.8 : *Prototheca* sp. grown on cornmeal agar for 72 hat 30°C (ca. × *800).*

Candida paratropicalis, described by Baker et al., has been isolated from several different clinical specimens. It varies only slightly in physiological activities from sucrose-negative C. *tropicalis,* and based on DNA homology studies, it is considered to be cospecific with C. *tropicalis.*

4.6.2 Cultural Characteristics

Budding cells (blastoconidia) are round, oval, or oblong, 2.5 μm by 3 to 14 μm, occurring singly or in clusters or chains. Members of the genus *Candida* have been defined on the basis of pseudomycelium formation "by all or most strains of all species and varieties"; this was the main distinction between *Candida* and *Torulopsis,* as the latter was believed not to produce pseudohy-phae. However, Yarrow and Meyer reexamined neotype strains of *Torulopsis* and concluded that the distinction between the two genera did not suffice to retain *Torulopsis* as a separate genus. They proposed amending the definition of *Candida* to include the statement "pseudohyphae absent, rudimentary, or well developed," and they renamed all *Torulopsis* species as *Candida* species. This amended definition for *Candida* has been accepted for the third edition of *The Yeasts.* However, because *Torulop*sis was described before *Candida* and therefore takes precedence over it, this, as well as other factors, has caused great disagreement among taxono-mists. The two genera will be treated separately in this chapter until a more final decision on proper taxonomy can be made.

Production of germ tubes and spherical chlamydospores by C. *albicans* and C. *stellatoidea is* a useful diagnostic characteristic. Ascospores have been reported in certain species later assigned to perfect genera; e.g.; ascospore-forming strains of C. *guillier-mondii* have been assigned to the genus *Pichia. Kluyveromyces fragilis (Kluyveromyces marxianus)* is the ascospore-forming stage of C. *pseudotropicalis (Candida kefyr)* and has been reported by that name as a cause of human disease. The teleomorph of C. *krusei* has recently been designated as *Issatchenkia orientalis.*

Growth is aerobic. Tiny colonies may be visible as early as 24 to 36 h and reach 1.5 to 2 mm in 5 to 7 days on Sabouraud agar. Colonies are usually stark white, but may become cream colored or tan with age. They are glabrous, creamy, or membranous and may have a fringe of submerged hyphae. Optimal growth temperature is 25 to 37°C.

4.6.3 Biochemical characteristics

Carbohydrates are utilized in oxidative (assimilation) or fermentative patterns, or both, that are helpful in differentiating species. Urease is

negative except for occasional strains of C. *krusei.* KNO3 is not utilized by frequently encountered clinical isolates.

4.6.4 Animal Pathogenicity

C. albicans is lethal to mice and rabbits when injected intravenously, causing miliary abscesses in the kidneys and other organs. Other *Candida* spp. may also produce lesions but are seldom lethal for experimental animals that have not been pretreated with corticosteroids or other immunosuppressive drugs.

4.7 GENUS *CRYPTOCOCCUS*

Formerly classified in the same family as *Candida, Cryptoc-occus,* like *Candida, is* now believed to be a heterogeneous genus. Its only recognized pathogenic species, C. *neoformans,* and probably also C. *albidus* have teleomorphs that belong to the genus *Filobasidiella* of the Ustilagenales. Other species of *Cryptococcus* have occasionally been reported to cause human disease. Evaluation of earlier reports is sometimes difficult because the isolates were not always carefully identified. However, some of the recent reports which are accompanied by demonstration of yeasts that are morphologically compatible with the isolate in tissue suggest that species of *Cryptococcus* besides C. *neoformans* may cause disease in certain debilitated patients.

Five serotypes (A, B, C, D, and AD) of C. *neoformans* have been described, and these have been related to two teleomorphs that produce basidiospores. The demonstration of clamp connections in two isolates by Shadomy provided the first convincing evidence of a taxonomic affinity between C. *neoformans* and basidiomycetes. Subsequent mating experiments by Kwon-Chung with isolates from many different sources demonstrated basidiospores with A and D serotype isolates, and the teleomorph was named *Filobasidiella neoformans.* It was later shown that serotype B and C isolates produced a filobasidiella state that formed rod-shaped basidiospores; that teleomorph was named *Filobasidiella bacillaspora.* Subsequently, renaming of the anamorph of B and C serotypes as C. *bacillasporus* was proposed. *F. neoformans* then became the teleomorph of C. *neoformans* (serotypes A, D, and AD), and *F. bacillaspora* became the teleomorph of C. *bacillasporus* (serotypes B and C). However, mating studies with appropriate isolates then revealed that *F. neoformans* and *F. bacillaspora* produced a high percentage of fertile basidiospores. For that reason, *F. bacillaspora is* now regarded as a variant of *F. neoformans* and the anamorph C. *bacillasporus is* considered a variant of C. *neoformans.* The two varieties differ in serotype, utilization of creatinine and certain dicarboxylic acids,

morphology of yeast cells, temperature tolerance, and virulence for mice. DNA hybridization revealed a 60% homology between the two variants. C. *bacillasporus* has also been shown to be identical to the C. *neoformans* subsp. *gatti* described by Vanbreuseghem and Takashio in 1970 using the same criteria given above. It thus becomes a later synonym of C. *neoformans* subsp. *gatti* and is an invalid name. Some regard the two variants as identical and C. *neoformans* (teleomorph *F. neoformans)* as the only valid name necessary to designate all serotypes and variants of the species.

The usual biochemical tests do not differentiate the two varieties. To avert confusion, C. *neoformans is* retained here as the sole name to represent both varieties. At present there are no known differe-nces in disease symptoms or severity associated with the two variants; both appear to respond similarly to antifungal agents. There is, however, a distinct difference in their geographic distribution. C. *neoformans* subsp. *gatti* (serotype B) is found almost exclusively on the West Coast in the United States, whereas most infections elsewhere in the country are caused by C. *neoformans* subsp. *neoformans* (serotype A).

Infections caused by C. *neoformans* are exogenous; the yeast lives naturally on soil contaminated with bird droppings, notably from pigeons and other seedeating birds. Meningitis is the most frequently recognized type of cryptococcal infection, followed in frequency by localized abscesses or granulomas (cryptococcoma or toruloma) in lungs, brain, lymph nodes, skin, or bones. Diffuse pulmonary infection is perhaps the most common type of cryptococcal infection, though it is often asymptomatic and unrecognized. The respiratory tract is believed to be the portal of entry for most if not all cryptococcal infections.

Blastoconidia are chiefly spherical and exhibit a wide range of diameters (from 5 to 30 μm). A polysaccharide capsule is a constant feature although it is much more prominent in infected tissues or when suspended in immune sera than in cultures. The capsule can be demonstrated in wet preparations by negative staining with India ink or nigrosin or by mounting the cells in a drop of normal serum with a drop of 1% acetic acid, or it can be demonstrated in immune serum. The capsule also is revealed by mucin stains such as mucicarmine or alcian blue. This feature permits the presumptive identification of C. *neoformans* in histopathological sections, since other yeasts which invade human tissues are not revealed by these dyes.

Colonies are often mucoid, becoming dull and drier with age. Colonies on Sabouraud agar initially are pale buff, changing to tan or

brown with age. This genus does not ordinarily form mycelium, though occasional isolates have been encountered which produce pseudomycelium. Hypha-forming strains of C. *neoformans* which produce clamp connections have been observed, reflecting the basidiomycetous affinity of the species.

Growth is aerobic. *Cryptococcus* spp. grow well on ordinary bacteriological and mycological media, attaining colony diameters of 1, to 3 mm in less than 1 week. They are inhibited by cycloheximide. Bacteria, serum factors, or tissue inhibitors in clinical specimens may delay the appearance of colonies by as much as 2 to 4 weeks. The optimal temperature varies with species.

4.7.1 Biochemical Characteristics

Metabolism is strictly oxidative; assimilation of certain sugars and of KNO_3 is useful for differentiating species. Utilization of inositol and the usual absence of carotenoid pigments distinguish this genus from *Rhodotorula;* both genera may have capsules, produce starchlike compounds, and are urease positive. C. *neoformans* can be differentiated from other *Cryptococcus* by (i) ability to grow at 37°C, (ii) production of brown colonies on birdseed agar (positive phenol oxidase reaction), (iii) characteristic assimilation pattern, and (iv) pathogenicity for experimental animals. Phenol oxidase may have to be induced in certain isolates (particularly serotype C). Such isolates will not produce brown colonies on birdseed agar when first isolated.

4.7.2 Animal Pathogenicity

C. neoformans, but not other species of *Cryptococcus,* is lethal for mice by invasion of the brain, reached by direct intracranial inoculation or by neurotropic extension from the intravenous or intraperitoneal route.

4.8 GENUS *HANSENULA*

Hansenula anomala and *Hansenula polymorpha* have been reported to cause human disease. All members of the genus are nitrate positive and produce ascospores that may be hat shaped, hemispheroidal, spherical, or Saturn shaped. *H. anomala* ferments as well as assimilates carbohydrates. Colonies vary considerably in macroscopic appearance, from smooth and glistening to white, opaque, and wrinkled, and could easily be confused with colonies of *Cryptococcus* or *Candida spp.* Pseudohyphae may or may not be produced.

4.9 GENUS *MALASSEZIA (PITYROSPORUM)*

Malassezia is now considered the correct name for the genus of lipophilic yeasts that reproduce by unipolar bud fission and have been

associated with tinea versicolor, blepharitis, dacryocystitis, dandruff, and seborrhea. Recent reports of folliculitis, isolation from the upper respiratory tract and lungs, and isolation of lipophilic yeasts from a case of peritonitis suggest that *Malassezia furfur* can cause systemic human infections. *M. furfur* is now considered the correct name under which both *Pityrosporum ovale* and P. *orbiculare* are combined. *M. pachydermatis,* the second species in this genus, has been isolated from dogs and other animals, but rarely from humans.

Bottle-shaped, small budding cells are 1 to 2 by 2 to 4 μm. Buds are produced successively from an invaginated end of the cell and are separated from the parent cell by a transverse septum (bud fission). Short hyphae have been observed in special media but are more typically seen in tinea versicolor skin scales. Sexual spores have not been observed.

M. furfur grows well at 37°C when incubated aerobically on media containing olive oil, coconut oil, or other lipid sources. Colonies are creamy and punctiform when young, becoming membranous and confluent when older. *M. pachydermatis* produces similar colonies on complex media without supplementation of lipids, but grows poorly on chemically defined media such as YNB-glucose.

4.9.1 Biochemical Characteristics

Malassezia spp. have no fermentative ability. *M. furfur* requires lipids or specific fatty acids, including myristic and palmitic acids, for growth.

4.10 GENUS *PICHIA*

Pichia contains the perfect form of *Candida guilliermondii* and belongs to the subfamily *Saccharomycetoideae* of the family *Saccharomycetaceae* of the Ascomycetes. Of the many strains of C. *guilliermondii* tested, only a few produced ascospores. These were hat shaped and ranged from one to four per ascus. Growth characteristics, biochemical reactions, and other laboratory reactions of the teleomorph should be the same as those of the anamorph, which are listed in Table 1.

4.11 GENUS *RHODOTORULA*

Rhodotorula (family *Cryptococcaceae,* subfamily *Rhodoto-ruloideae)* resembles *Cryptococcus* in rate of growth, colony topography, cell size and shape, occasional rudimentary pseudomycelium, presence of capsule, ability to split urea, and absence of fermentation. Its nonpathogenicity, different serotype, and conspicuous carotenoid pigment

have maintained *Rhodotorula* as a distinct genus. *Cryptococcus*, furthermore, can use inositol, invade human and animal tissues, and kill *(C. neoformans)*. *Rhodotorula is* a normal symbiont of moist skin (due to climate or to the patient's abnormal physiology); *R. rubra (mucilaginosa)* and *R. glutinis* have been reported rarely to cause septicemia, meningitis, and a chronic skin infection. Recently, *R. pallida* and *R. marina* were implicated in an invasive infection of a leukemia patient. These two species had not been reported to cause human disease before this. Occasionally fungemia associated with the contamination of plastic tubing left in place during prolonged intravenous therapy will resolve without antifungal therapy when the catheter is removed. *Rhodotorula spp.* are also cultured readily from shower curtains, bathtub-wall junctions, and' the rubber tips and handles of toothbrushes, if not properly aerated.

4.12 GENUS *SACCHAROMYCES*

Saccharomyces is a member of the family *Saccharomycetaceae* (ascosporogenous). *S. cerevisiae is* responsible for occasional cases of thrush and vulvovaginitis; it has also been reported from urine specimens in diabetics and has been reported to cause fungemia.

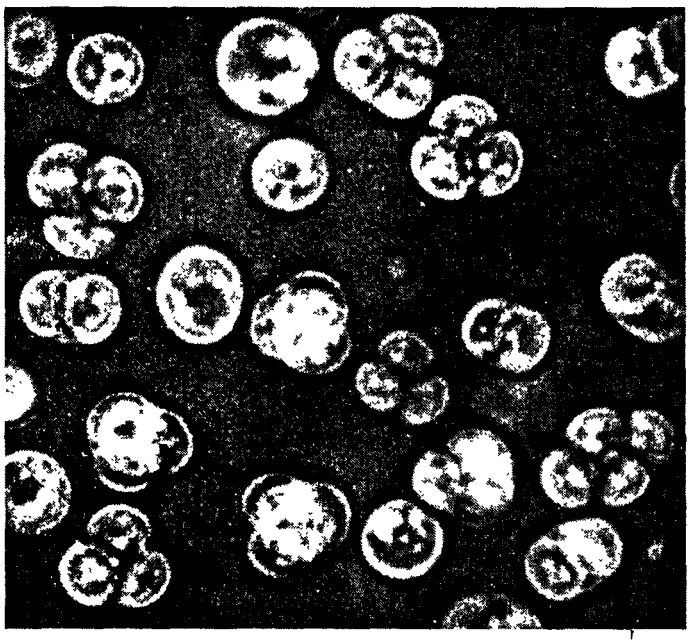


Figure 4.9 : Ascospores *of Saccharomyces cerevisiae. Un*stained. × *2,500.*

Cells are oval to spherical, 3 by 5 μm, and may exist as budding cells as either haploids or diploids (after fusion). Cells may form short chains and elongate as rudimentary pseudohyphae. Ascospores, one to four in number, are in either tetrahedral or linear arrangement and are gram negative (vegetative cells are gram positive). An excellent sporulation medium is the potassium acetate-yeast extract-glucose medium (ascospore medium) of McClary modified from Adams.

Occasional production of rudimentary pseudomycelium occurs; this genus is the teleomorph of *Candida robusta,* a nonpathogen. Growth is rapid on most media, both aerobically and anaerobically; optimal temperature is *25* to *37°C.*

4.12.1 Biochemical Reactions

Saccharomyces spp. assimilate and ferment sugars in a pattern useful for identification. S. *cerevisiae* does not assimilate KNO_3, split urea, or grow on creatine.

4.12.2 Serological Examination

A slide agglutination method with specific antisera can be used in identifying species.

4.13 GENUS *TORULOPSIS*

Torulopsis glabrata is a common isolate from urine and other specimens. It is considered to be a symbiont of humans, but it has been documented as the cause of pyelonephritis, pneumonia, septicemia, and meningitis in immunocompromised patients. Documentation of pathogenicity can be accomplished only by demonstration by biopsy of tissue invasion. In stained tissues it closely resembles *Histoplasma capsulatum,* from which it must be differentiated. *T. pintolopesii is* a thermophilic yeast capable of growth at 40 to 42°C. It is rarely isolated from humans, but is a symbiont of the intestinal tract of mice. This should be kept in mind when animal pathogenicity studies are performed. *T. candida is* occasionally isolated from specimens as a contaminant, but it has not been shown to produce human disease. It closely resembles *C. guilliermondii* and is not easily differentiated from it with the usual biochemical tests.

4.14 GENUS *TRICHOSPORON*

Trichosporon is a member of the subfamily *Trichosporoideae,* family *Cryptococcaceae. T. beigelii* (cutaneum) causes superficial nodules (white piedra) on the distal portion of hair in the scalp, beard, axilla, and pubic area. Occasional opportunistic invasion of the mucous membranes, skin, and deeper tissues by this species has also been

reported. *T. capitatum, T. fermentans,* and *T. penicillatum* have been reclassified in *Geotrichum,* and more recently *T. capitatum* has been renamed *Blastoschizomyces pseudotrichosporon.*

Trichosporon has well-developed hyphae and pseudohyphae, reproducing also by blastoconidia and arthroconidia; no chlamydospores or sexual spores have been reported.

Growth is aerobic and occurs on all of the usual culture media. Smooth, shiny colonies, 3 to 6 mm, appear within 1 week; later, they become membranous, dry, and cerebriform. The optimal temperature is 25 to 37°C; some species grow equally well at 37°C.

4.14.1 Biochemical Characteristics

Most *Trichosporon spp.* are oxidative; an occasional species is also fermentative. A heavy surface pellicle is formed in broth media. Assimilation of KNO_3 is generally absent, but present in *T. pullulans; T. beigelii* and *T. pullulans* are urease-positive species.

Trichosporon can be distinguished from *Geotrichum* by its production of blastoconidia as well as arthroconidia, its rapid growth at 37°C, and its assimilation of a greater number of the usual sugars.

5

Laboratory Techniques

Parasitic diseases continue to cause significant morbidity and mortality in the world, particularly in lessdeveloped tropical and subtropical countries. In the United States, indigenous malaria was eradicated long ago, and indigenous nematode infections such as ascariasis, trichuriasis, and hookworm infection have markedly decreased in both incidence and severity. Some other infections are increasing. Giardiasis is a frequent public health problem, with outbreaks related to water supplies and day care centers for children. *Giardia,* ameba, and *Cryptosporidium* infections are increasing in male homosexuals. *Pneumocystis carinii, Cryptosporidium* species, *Strongyloides stercoralis,* and *Toxoplasma gondii* are increasingly important causes of serious infections in immunoco-mpromised hosts, especially those with AIDS (acquired immune deficiency syndrome).

In addition to infections which are indigenous to the United States, a wide variety of infections may be seen in U.S. citizens who have traveled or worked in foreign countries or in foreign nationals who are visiting or now residing in the United States. Many of these people, such as persons infected with malaria, may be asymptomatic for months or years before disease develops or relapses occur. Some people are recognized as having malaria only when a recipient of their blood develops transfusion-induced malaria or when a baby develops congenital malaria. Other diseases such as echinococcosis may require years before becoming clinically evident.

Efforts to eradicate parasite infections have had variable success. Sanitary fecal disposal, improved water supplies, and improved hygiene in food production and preparation have aided in the control of intestinal

parasites. However, much of the earlier enthusiasm for the eradication of malaria has been tempered by the realization that malaria eradication is going to be difficult because parasites are becoming resistant to chemotherapeutic agents, mosquito vectors are becoming resistant to common insecticides, and the use of some insecticides may harm the environment. Human modifications of the environment, such as the building of dams and irrigation systems, have provided an appropriate environment for vectors such as snails and thus allowed diseases such as schistosomiasis to flourish in areas where these diseases had been uncommon. In addition, immunization programs for parasite infections have developed more slowly than was anticipated.

5.1 HOST-PARASITE RELATIONS

A knowledge of parasite life cycles is crucial in the understanding of the ways infection is acquired and spread, the pathogenesis of disease, and the ways in which disease might be controlled. Some parasites which infect only humans, such as *Enterobius vermicularis* (pinworm), have a narrow host specificity, whereas others such as *Trichinella spiralis* infect numerous species. When other animals harbor the same parasite stage as humans, these animal species may serve as reservoir hosts. Humans infected with a parasite stage usually seen in other animal species are referred to as accidental hosts.

In the simplest life cycle, the parasite stage from humans is immediately infective for other humans, as in pinworm infection or giardiasis. In other infections such as ascariasis or trichuriasis, a maturation period outside the body is required before the parasite is infective. However, for many parasite infections, a second or even a third host is required for completion of the life cycle. Hosts are defined as intermediate hosts if they do not contain the sexual stage and as definitive hosts if they do contain the sexual stage. Some protozoa, such as the amebae, flagellates and hemoflagellates, do not have a recognized sexual stage. In the intermediate host, there may be a massive proliferation of organisms, as occurs in humans harboring malaria parasites or snails harboring schistosome intermediate stages, or there may be no proliferation, as in mosquitoes which harbor microfilaria undergoing maturation. There may be proliferation in definitive hosts, as in mosquitoes harboring the sexual stage of malaria in which thousands of sporozoites are produced, or there may be no proliferation, as in helminth infections in which one adult is developed from each infective larva. However, in the latter, the adult helminths do produce numerous eggs or larvae.

TABLE 5.1 : Incidence of intestinal parasites in 322,735 fecal specimens examined by state health department laboratories, 1978

Parasite	No. of examinations	% of positive specimens
Protozoa		
Giardia lamblia	12,947	4.0
Entamoeba histolytica	2,409	0.8
Dientamoeba fragilis	1,880	0.6
Balantidium coli	7	
Isospora spp.	1	
Nonpathogenic	21,120	6.5
Nematodes		
Trichuris trichiura	5,481	1.7
Ascaris lumbricoides	4,630	1.4
Enterobius vermicularis	4,344	1.4
Hookworm	2,035	0.6
Strongyloides stercoralis	602	0.2
Trichostrongylus spp.	14	
Trematodes		
Clonorchis-Opisthorchis	205	0.06
Schistosoma mansoni	48	
Fasciola hepatica	1	
Paragonimus westermani	1	
Cestodes		
Hymenolepis nana	1,068	0.3
Taenia spp.	251	0.08
Diphyllobothrium latum	20	
Hymenolepis diminuta	12	
Dipylidium caninum	7	

In some helminth infections, a migration through various body tissues is essential for maturation, as in ascarasis or schistosomiasis, whereas in other infections, the larva leaves the egg and simply matures in the intestinal tract, as in trichuriasis and enterobiasis. Host tissues involved vary depending upon the parasite. In severely immunocompromised patients, sites may be involved that are not involved in normal hosts.

Parasites of humans proliferate tremendously at certain stages, with thousands or even millions of forms being produced for every one

that survives to perpetuate the parasite. Parasites may be quite hardy. For example, certain stages, particularly eggs and cysts, may survive for weeks or months in the environment.

Parasites have often developed unique ways of protection from the defense mechanisms of the host.

The survey does not include laboratories in Guam, Puerto Rico, or Virgin Islands. One or more parasites were found in 14.7% of specimens. Percentages are not calculated for parasites identified less than 100 times.

These mechanisms include the ability to change antigenic characteristics so that although the host forms antibody, the antibody does not react with the modified parasite, or the parasite may be coated with host immunoglobulins, as in schistosomiasis, so that the host does not recognize the parasite as foreign. Macrophages and both cell-mediated and humoral immunities appear to be important in the host response to infection. Eosinophils are particularly important in the defense against tissue-invading helminths.

5.2 DIAGNOSIS OF PARASITIC INFECTIONS

The diagnosis of most parasitic infections is dependent upon the laboratory. For intestinal and blood parasites, morphologic demonstration of diagnostic stage(s) is the principal means of diagnosis, whereas for tissue infections, immunodiagnostic techniques are generally more important. During the early stages before diagnostic forms are produced (prepatent period), patients may be symptomatic. For example, patients may have pulmonary symptoms and eosinophilia due to ascaris larva migration at a time when eggs are not produced. In such patients the physician may suspect parasite infection, but the actual diagnosis must be based on a clinical impression or immunodiagnostic tests, or diagnosis must await the production of diagnostic stages.

In establishing a diagnosis, the clinician places a great deal of trust in the laboratory. This trust can be misplaced if laboratory personnel are not competent to identify or exclude parasites. The literature clearly documents instances in which outbreaks have been overlooked due to incompetent laboratory diagnosis or in which inflammatory cells or other objects have been identified as parasites and outbreaks have been diagnosed when none existed. The results of proficiency-testing programs also suggest that laboratories have difficulty with the identification of some parasites, especially intestinal protozoa.

Identification may be by gross examination for adult helminths or,

more commonly, by microscopic examination for protozoa, helminth eggs, and larvae. The diagnostic forms of some parasites, such as the eggs of Ascaris spp., are present on a regular basis. Other forms, such as malaria parasites, *Taenia* eggs, or *Giardia* cysts, vary from day to day.

Most immunodiagnostic tests used today for parasitic infections detect antibody. In recent years, the sensitivity and specificity of many such tests have improved. A number of antigen detection tests have recently been described and show promise, but none of these tests are currently available commercially.

5.3 LABORATORY PROCEDURES

Many methods for diagnostic parasitology have been described. There are advantages and disadvantages to each method. Some are

TABLE 5.2 : Participant performance.

No. of Parasite	Avg correct specimens	identification (%)
Formalin-fixed fecal specimens		
Ascaris lumbricoides eggs	6	90
Hookworm	6	92
Strongyloides stercoralis larvae	4	88
Trichuris trichiura eggs	6	93
Diphyllobothrium latum eggs	6	81
Hymenolepis diminuta eggs	5	91
Taenia sp. eggs	6	87
Paragonimus westermani eggs	5	83
Giardia lamblia cysts	8	65
Entamoeba coli cysts	9	88
No parasite seen	6	92
PVA-fixed specimens		
Entamoeba histolytica	5	73
Entamoeba coli	4	52
Endolimax nana	4	51
Negative for parasites	3	77

particularly valuable for epidemiologic studies or for evaluations of new therapeutic agents, whereas other methods are used primarily for laboratory diagnosis. In this chapter we emphasize the diagnostic procedures. From the numerous methods, we have selected those which are widely used in this country and which are sensitive and relatively easy to perform. These methods should prove adequate for most laboratories. For additional procedures, laboratory manuals or parasitology books should be consulted. When alternative methods or methods for specific parasites are indicated, references will be given, but the methods will not be described.

5.4 PROCEDURES FOR INTESTINAL PARASITES

Intestinal and biliary parasites are generally diagnosed by finding diagnostic stages in feces or other intestinal material such as duodenal or sigmoidoscopic aspirates. Studies have shown that the eggs of most parasites are uniformly distributed in the fecal mass due to the mixing action of the colon, although some, such as schistosome eggs, which originate in the distal colon, may be more numerous on the surface of formed fecal specimens. The distribution of protozoan forms is more variable. There may be fewer protozoan trophozoites in the first part of an evacuation than in the last because they have deteriorated while in the lower colon.

5.4.1 Collection and Handling of Fecal Specimens

The numbers and times of collection for fecal specimens depend somewhat on the diagnosis suspected. As a routine, because some organisms are shed in a variable pattern, it is advisable to examine multiple specimens before excluding parasites. The general recommendation is to collect a specimen every second or third day, for a total of three specimens. From a hospitalized patient, one specimen each day for three days may be more cost effective.

A number of substances may interfere with stool examination. Particulate materials such as barium, antacids, kaolin, and bismuth compounds interfere with morphologic examination, and oily materials such as mineral oil create small, refractile droplets that make examination difficult. Antimicrobial agents, particularly broad-spectrum antimicrobial agents, may suppress amebae. If any of these substances have been used, specimens should not be submitted until the substances have been cleared (generally 5 to 10 days). A fecal specimen may appear satisfactory by gross examination when there is still barium, etc., which can interfere with microscopic examination.

Fecal specimens are best collected into widemouthed, water-tight

containers with tight-fitting lids such as waxed, pint-sized ice cream cartons or plastic containers. Usually patients can defecate directly into such containers. Urine should not be allowed to contaminate specimens, as it is harmful to some parasites. If specimens are to be collected in a bed pan, the patient should micturate into a separate container before the specimen is collected. Toilet paper should not be included with the specimen. Stool should not be retrieved from toilet bowl water, as various freeliving protozoa in water might be confused with the parasites. In addition, water is harmful to some parasites such as schistosome eggs and amebic trophozoites. If the patient is producing formed specimens, stool may be collected by having the patient squat over waxed paper to defecate.

Purgation with sodium sulfate or buffered phosphosoda may be helpful in the diagnosis of amebiasis in some patients. Purgation is usually done after a series of fecal specimens have been negative, and it requires the order of a physician. Prior arrangements must be made with the laboratory, and specimens must be collected during regular laboratory hours. The patient is given the appropriate salt solution orally. In approximately 1 to 1.5 h, the patient will begin to pass stool specimens, and each specimen should be promptly transported to the laboratory for examination.

Clinical information such as the suspected diagnosis, travel history of the patient, and clinical findings should be included on the requisition. In addition, the time the specimen was passed and the time it was placed in fixative should be noted. If the specimen is in fixative, the consistency of the original specimen should be stated, or a portion of unfixed specimen should be included with the fixed specimen.

A laboratory may have specimens placed in fixatives in the home or patient care area immediately after passage, may place portions of specimen in fixatives at the time they are received in the laboratory, or may examine the specimen unfixed. Many laboratories use a combination of these methods depending on the location of the patient, consistency of the specimen, time of day, and laboratory work load. Prompt examination or fixation is particularly important for soft, loose, or watery specimens, which are most likely to contain protozoan trophozoites.

Formed specimens, which are likely to contain protozoan cysts or helminth eggs or larvae, can remain satisfactory for a number of hours at room temperature or overnight in a refrigerator. Soft and liquid specimens should be examined or placed in fixatives promptly (within

1 h). Specimens which cannot be examined or fixed promptly should be either refrigerated or left at room temperature. They should not be incubated, as incubation speeds the deterioration of the organisms. Feces for parasite examination must not be frozen and thawed.

The fixative system generally used is a two-vial technique with one vial containing 5 to 10% buffered Formalin and the other vial containing polyvinyl alcohol (PVA) fixative. A portion of the specimen is added to the fixative in a ratio of approximately 3 parts fixative to 1 part specimen and thoroughly mixed to ensure adequate fixation. An alternative to Formalin is Merthiolate-iodine-formaldehyde (MIF), which fixes and stains at the same time. If unfixed specimens are processed in the laboratory, fecal films may be prepared and immediately fixed in Schaudinn fixative.

5.4.2 Gross Examination of Feces

Specimens should be examined grossly to determine the consistency (hard, formed, loose, or watery), color, and presence of gross abnormalities such as worms, mucus, pus, or blood. It may be profitable to examine flecks of mucus, pus, or blood for parasites. If adult worms or portions of tapeworms are sought, the feces may be carefully washed through a screen. (Small worms may be difficult to see if gauze is used.) The identification characteristics of adult worms are not discussed in this chapter, so parasitology books should be consulted.

5.5 PROCEDURES FOR MICROSCOPIC EXAMINATION

The three principal microscopic examinations performed on stool specimens are direct wet mount, wet mount after concentration, and permanent stain. Although each examination can contribute to diagnosis, the yield of some methods is small with certain kinds of specimens. As a minimum, formed specimens should be examined by a concentration procedure. Soft specimens should be examined by concentration and permanent stain, and, if submitted fresh, by direct wet mount. Loose and watery specimens should be examined by wet mount and permanent stain. If specimens are received in fixative and the consistency is not known, concentration and permanent stain should be performed. Other examinations may be helpful. Special procedures which may assist in the diagnosis of specific parasites are noted below in discussions of the parasites.

5.5.1 Calibration and Use of an Ocular Micrometer

Size is important in the differentiation of parasites and is most accurately determined with a calibrated ocular micrometer;, thus, each

laboratory performing diagnostic parasitology must have such a micrometer.

An ocular micrometer is a disk on which is etched a scale in units from 0 to 50 or 100. To determine the micrometer value of each unit in a particular eyepiece and at a specific magnification, the unit must be calibrated with a stage micrometer. A stage micrometer has a scale 2 mm long ruled in fine intervals of 0.01 mm (10 μm).

5.5.2 Calibration of the Ocular Micrometer

1. Insert the micrometer in the eyepiece so that the micrometer rests on the diaphragm, with the etched scale facing the eye. In many new microscopes, the micrometer can be dropped in and secured with a ring retainer. (It is helpful to have an extra ocular in which the micrometer may be left.)

2. Place the stage micrometer on the microscope stage.

TABLE 5.3 : Laboratory examinations for various types of fecal specimens

Type of specimen	Method		
	Direct wet mount	Concentration	Permanent stain
Unpreserved			
Formed	+	++	±
Soft ++	++	++	
Loose and watery	++	±	++
Preserved			
Formalin			
Formed or soft	+	+ +	–
Loose or liquid	++	+	–
PVA fixative			
Formed	–	–	±
Soft, loose, or liquid	–	–	++

++, Essential for basic examination; +, recommended for basic examination; ±, optional for basic examination.

Concentration is recommended if *Cryptosporidium* sp. is suspected.

3. Focus on the etched scale. Since the micrometer must be calibrated for each objective, begin with the lowest magnification (e.g., × 10).

4. Align the two scales so that the zero points are superimposed.

5. Find a point far down the scales at which a line of the stage micrometer coincides with a line of the ocular micrometer. Count the number of ocular units and the number of stage units from zero to these coinciding lines.

6. Multiply the number of stage micrometer units by 1,000 to convert millimeters to micrometers.

7. Divide the product of step 6 by the number of ocular units to determine the value of an ocular unit.

Repeat the calibration for each objective. Keep a record of the unit value for each objective for each microscope used. Calibration must be done separately for each microscope and must be repeated if an ocular or objective is changed.

Use of the micrometer. Insert the eyepiece containing the calibrated ocular micrometer in the microscope. Count the number of ocular units which equal the structure to be measured. Multiply the number by the micrometer value of the ocular unit for the objective being used. If an ocular micrometer is properly used, parasites which are similar in appearance but different in size can be readily differentiated.

5.5.3 Direct Wet Mount

The direct wet mount made from unconcentrated fresh feces is most useful for the detection of the motile trophozoites of intestinal protozoa and the motile larvae of *Strongyloides spp.* It is also useful for the detection of protozoan cysts and helminth eggs. For fixed feces, the direct wet mount may allow the detection of parasites which do not concentrate well. This method is also useful for the examination of specific portions of feces, such as flecks of blood or mucus.

Direct wet mounts are prepared by placing a small drop of 0.85% saline toward one end of a glass slide (2 by 3 in. [ca. 5 by 7.5 cm]) and a small drop of appropriate iodine solution (see below) toward the other end. With an applicator stick, a small portion of specimen (1 to 2 mg) is thoroughly mixed in each diluent, and a no. 1 cover slip (22 mm) is added. The density of fecal material should be such that newspaper print can be read with difficulty through the smear. The material should not overflow the edges of the cover slip. Grit or debris may prevent the cover slip from seating and may be removed with a corner of the cover slip or an applicator stick. Mounts may be sealed with Vaspar (50% petroleum jelly, 50% paraffin) which is melted on a hot plate (not over an open flame). A cotton applicator or small brush is used to apply small drops of Vaspar to opposite corners to

attach the cover slip and then to seal it with even strokes. The amount of Vaspar on top of the cover slzip should be minimal. Sealing slows drying and allows oil immersion magnification to be used. Alternatively, drying can be slowed by placing wet gauze or paper toweling in a petri dish, laying portions of applicator sticks or glass rods on the moist material, laying the slide on the sticks or rods, and replacing the lid of the dish.

The iodine solution should be that of Dobell and O'Connor (1%) or a 1:5 dilution of Lugol iodine. Iodine solution, if too weak, will not stain organisms properly, and if too strong, it will cause clumping of fecal material. Stock iodine solution should be stored in a tightly stoppered brown bottle away from sunlight. Keep the iodine and saline solutions in small dropper bottles, and replace (don't replenish) the solutions weekly. Iodine solution keeps longer if it is refrigerated. Iodine stain solution can be quality controlled by the observation of appropriate staining in positive clinical specimens or Formalin-fixed specimens kept for that purpose.

For the examination of wet mounts, the light of the microscope must be properly adjusted. To achieve optimal resolution, the condenser should be centered and focused for Kohler illumination (racked up). To achieve contrast of the objects in the field, light intensity is diminished with the iris diaphragm of the condenser rather than by lowering the condenser.

The entire saline wet mount cover slip should be systematically scanned with × 100 to × 200 magnification. Suspicious objects are confirmed at higher magnification. In addition, the preparation should be scanned at higher power (×400 to ×500) for a couple of passes across the cover slip to look for protozoan cysts which might be missed with lower power. Screening a slide should take an experienced microscopist about 10 min. If debris is covering a suspicious object, the debris may be removed by pressing or tapping on the cover slip with an applicator stick. This pressure may also help in reorienting an egg, as when one is looking for an operculum. The saline wet mount is best for the detection of helminth eggs and larvae, and it is especially good for protozoan cysts, which appear refractile. The principal usefulness of the iodine mount is to study the morphology of protozoan cysts, as this stain shows nuclear detail and glycogen masses (but does not stain chromatoid material). If suspicious objects are seen, they can be examined under oil immersion (× 1,000). If definite or possible protozoan cysts or trophozoites are detected which cannot be identified

in wet mounts, permanent stains are required. A solution of buffered methylene blue (pH 3.6) may be used as a vital stain for the examination of fresh specimens for protozoa. The wet mount is prepared as described above, with buffered methylene blue substituted as diluent and 5 to 10 min allowed for the dye to become incorporated in the organisms before examination. Organisms become overstained in 20 to 30 min.

5.5.4 Concentration Procedures

Concentration procedures are used to separate parasites from fecal detritus. These procedures are based on differences in the specific gravity of parasite forms and fecal material. In sedimen-tation, the parasite forms are heavier than the solution and are found in the sediment, whereas in flotation, solutions of high specific gravity are used, and parasite forms float to the surface. An initial washing step removes some of the soluble or finely particulate material, and straining removes larger portions of debris. A wide variety of sedimentation and flotation methods have been described. The Formalin-ethyl acetate modification of the Formalin-ether sedimentation technique and a zinc sulfate flotation technique are widely used and are the only methods described in this chapter. Both methods require that centrifugation be performed in centrifuges with free-swinging carriers. Squeeze bottles for Formalin, saline, or water simplify the processing of large numbers of specimens.

Formalin-ethyl acetate centrifugal sedimentation. The original procedure from which the Formalinethyl acetate centrifugal sedimentation technique was adapted was the Formalin-ether concentration method of Ritchie. The Formalin-ethyl acetate procedure avoids problems with the flammability and storage of ether. This procedure can be performed on specimens which have been fixed in Formalin for a time or on specimens with Formalin added during the processing. The procedure can also be performed on material fixed in MIF.

The procedure with Formalin-preserved specimens is as follows.

1. Thoroughly mix the formalinized specimens.

2. Depending on the density of the specimen, strain a sufficient quantity through gauze into a 15-ml conical centrifuge tube to give the desired amount of sediment. (Wet gauze in a 4-oz [ca. 120-ml] conical paper cup with the tip cut off can be used for straining.)

3. Add tap water or saline, mix the solution thoroughly, and centrifuge it at 650 × g for 1 min. The amount of the resulting sediment should be about 1 ml. The amount of sediment may be adjusted

by the addition of more feces and centrifugation again or by the addition of water, suspension again, the removal of an appropriate amount of material, and then recentrifugation.

4. Decant the supernatant, and wash it again with tap water, if desired.

5. To the sediment, add 10% Formalin to the 9-ml mark, and mix the solution thoroughly.

6. Add 4 ml of ethyl acetate, stopper the tube, and shake the tube vigorously in an inverted position for 30 s. Remove the stopper with care.

7. Centrifuge the solution at 450 to 500 × g for 1 min. Four layers should result: ethyl acetate, plug of debris, Formalin, and sediment.

8. Free the plug of debris from the sides of the tube by ringing the tube with an applicator stick, and carefully pour the top three layers into a discard container. With the tube still tipped, use a swab to remove debris from the sides of the tube. This step is very important, for lipid droplets which reach the sediment make examination difficult.

9. Mix the remaining sediment with the small amount of fluid that drains back down from the sides of the tube (or add a drop of saline or Formalin). If mounts are to be prepared later, a small amount of Formalin may be added to the sediment and the tube may be stoppered.

10. Prepare wet mounts as described above, and examine them.

The procedure for Formalin-ethyl acetate centrifugal sedimen-tation with fresh specimens is as follows.

1. Comminute a portion of stool about 1.5 cm in diameter in 10 ml of saline or water.

2. Strain about 10 ml of the fecal suspension into a 15-ml conical centrifuge tube.

3. Centrifuge the suspension at 650 × g for 2 min. This step should provide about 1 ml of sediment. If not, adjust the amount of sediment as described above.

4. Wash the sediment again if desired.

5. To the sediment, add 10% buffered Formalin to the 9-ml mark, mix thoroughly, and allow the mixture to stand for 5 min or longer.

6. Proceed as for step 6 of the procedure for fixed specimens.

(Note that either saline or water can be used. Tap water will lyse *Blastocystis hominis*. If schistosomiasis is suspected, the specimen

should be preserved in Formalin before concentration to prevent hatching.)

Zinc sulfate centrifugal flotation. The zinc sulfate concentration method originally described by Faust et al. may be performed on unfixed or Formalinfixed specimens, although the specific gravity of the zinc sulfate solution required differs. The disadvantages of the zinc sulfate concentration are: (i) dense schistosome eggs do not concentrate well; (ii) opercula often pop, and thus operculate eggs may be missed; and (iii) larvae and cysts may collapse. The modified procedure with Formalin-fixed feces slows the collapse of larvae and cysts and largely prevents the popping of opercula. The advantages are that it leaves a rather clean background, has less grit than the sedimentation procedure, and is better for the concentration of some parasites, such as *Giardia* cysts.

The procedure for Formalin-preserved specimens is as follows. The specific gravity of the zinc suflate must be 1.20. Centrifugation must be performed in round-bottomed tubes such as 16- by 100-mm disposable tubes.

1. The Formalin-preserved fecal material is mixed, strained through one layer of cheesecloth into a conical paper cup, poured into the tube to a level about 1 cm from the top, and then centrifuged.

2. The tubes are centrifuged for 3 min at about 650 × g. There should be 1 to 1.5 cm of sediment.

3. Decant the supernatant from each tube, and drain the last drop against a clean section of paper towel.

4. To the packed sediment of each tube, add zinc sulfate to within 1 cm of the rim.

5. Insert two applicator sticks, and thoroughly mix the packed sediment.

6. Immediately centrifuge the suspension at 500 rpm for 1.5 min.

7. Very carefully transfer the tubes to a rack of the proper size, so that the tubes remain vertical. Do not disturb the surface films, which now contain the parasites. Allow the tubes to stand for 1 min to compensate for any movement. The countertop must be vibration free.

8. With a loop which is bent at a right angle, transfer to a slide (2 by 3 in.) two loops of surface material beside 1 drop of saline and two loops beside 1 drop of iodine. With the heel of the loop, mix first the saline and then the iodine with the surface material. Cover each mixture with a 22-mm no. 1 cover slip. The slide should be made

within 20 min.

9. To retard drying, place each prepared slide on a bent glass rod or portions of applicator sticks in a petri dish containing a damp paper towel. Petri dishes may be placed in the refrigerator if examination will be delayed. Alternatively, cover slips may be sealed with Vaspar.

The procedure with fresh specimens is as follows.

1. Comminute a fecal specimen about 1 cm in diameter in a tube (16 by 100 mm) half filled with tap water. Add additional water to within 1 to 2 cm of the top.

2. Centrifuge the tube at 650 × g for 1 min.

3. Discard the supernatant, and add a zinc sulfate solution of specific gravity 1.18 to within 1 cm of the rim.

4. Proceed as from step 5 above.

Sheather sugar flotation. Sheather sugar flotation is recommended for the concentration of *Cryptosporidium* cysts. Although these oocysts will concentrate when the Formalin-ethyl acetate or zinc sulfate technique is used, they are more readily detected with the Sheather sugar flotation, for they stand out sharply from the background in this solution of high specific gravity. This procedure may be performed on unfixed or Formalin-fixed feces. The procedure for Sheather sugar flotation is outlined below.

1. (a) *Formed stool.* Place approximately 0.5 g of stool in a tube (16 by 100 mm) about half full of Sheather sugar solution. Mix the solution thoroughly, and then add more sugar solution to within 1 cm of the rim.

(b) *Watery stool.* Centrifuge the fecal specimen and mix 0.5 to 1 ml of sediment with Sheather solution as described above.

2. Centrifuge the solution at 400 × g for 5 to 10 min.

3. Remove the top portion of the sample with a wire loop bent at a right angle. Place several loopfuls on a glass slide (2 by 3 in.). Cover the specimen with a 22-mm cover slip, and examine the slide with a × 40 objective. Oocysts are found just beneath the cover slip and are refractile. Saline or iodine is not used in the preparation of these mounts.

Baermann concentration. The Baermann concentration technique has greater sensitivity for the detection of strongyloides larvae than do the standard concentration techniques described above. This technique is useful clinically for the diagnosis and monitoring of therapy of

strongyloides infections, and it is useful epidemiologically for the examination of soil for the larvae of nematode parasites.

A funnel with a clamped rubber tube on the stem is placed in a ring stand. A circular mesh screen is placed across the funnel approximately one-third from the top, a portion of coarse fabric such as muslin is placed on the screen, and feces is added. Tap water at 37°C is added so that the water just touches the feces. Let the specimen stand 1 h, remove 2 ml of fluid from the stem, and centrifuge the sample at 300 × g for 3 min. Prepare a wet mount of sediment, and examine it for larvae.

Hatching technique for viable schistosome eggs. Place a large amount of feces (5 to 10 g) in a large flask (1 to 2 liters), and add water while mixing to break up the feces to a fine suspension. Bring the water level to 2 to 5 cm from the top of the flask. Cover the sides of the flask with foil or other material to shield all but the top of the liquid from light. Allow the flask to stand at room temperature for several hours. With a hand lens, examine the material at the top of the flask neck for swimming miracidia. Remove the miracidia with a Pasteur pipette for examination with a x 10 objective. It is not possible to determine the species of schistosome from the miracidia.

Other concentration procedures. Concentration procedures have been described for feces preserved in MIF, sodium acetate-Formalin, or PVA fixative. MIF- or sodium acetate-Formalin-fixed feces may be used in place of Formalin-fixed feces in the Formalin-ethyl acetate concentration procedure. Some workers feel that organisms do not concentrate as well from material fixed in PVA fixative or from material which has been in MIF for extended periods.

If large amounts of specimen are to be concentrated, as when specimens of eggs are prepared for teaching, gravity sedimentation is usually used. The feces is thoroughly mixed in liquid (water, saline, or 10% Formalin) and allowed to settle in a sedimentation jar or funnel for several hours or overnight. Supernatant fluid is discarded, and the sediment is again suspended and allowed to settle. This procedure can be continued if desired until the supern-atant is clear.

5.5.5 Permanent Stains

Permanent stains of fecal smears are most needed for the detection and identification of protozoan trophozoites, but they are also used for the identification of cysts. Wet mounts of fresh feces, even with stains such as methylene blue, are not as sensitive for trophozoites and therefore do not substitute for permanent stains. It is sometimes difficult

to identify cysts which are detected in wet mounts; thus, for each specimen, regardless of consistency, it may be worthwhile to fix a portion in PVA fixative or to prepare two fecal films fixed in Schaudinn fixative so that permanent stains can be performed if needed. Permanent stains also provide a permanent record and are easily referred to consultants if there are questions on identification.

A number of staining procedures have been described. Some stains, such as chlorazol black, require fresh specimens and are not widely used. A variety of stains for fecal smears preserved by Schaudinn or PVA fixative have been described, including various hematoxylin stains. The stain most widely used in the United States is the Wheatley trichrome stain, which is the only permanent stain described in this chapter. The trichrome staining procedure uses reagents with a relatively long shelf life and is easy to perform. Note that there are differences in staining times depending on whether the specimen is fixed in Schaudinn or PVA fixative, as penetration is slower in the latter.

Preparation of smears. (i) Unpreserved specimens with Schaudinn fixative.

1. To prepare thin, uniform smears, place a drop of saline on a glass slide (1 by 3 in. [ca. 2.5 by 7.5 cm]). With an applicator stick, transfer a small, representative portion of the specimen to the drop of saline, and mix the two. Spread the solution into a film by rolling the applicator stick along the surface. Remove any lumps.

Before watery specimens are smeared, apply an adhesive such as serum or albumin to the slide. Liquid specimens may be centrifuged, and the sediment may be used for smear preparation.

2. Place fresh smears immediately into Schaudinn fixative. Do not allow the smears to dry at any time before they are stained. Smears should fix for at least I h at room temperature or for 5 min at 50°C; however, they can be left in fixative for several days. After fixation, slides may be kept in 70% alcohol indefinitely before they are stained.

(ii) Unpreserved specimens with PVA fixative.

1. On a slide (1 by 3 in.), thoroughly mix 1 drop of unfixed specimen with I drop of PVA fixative.

2. Spread the specimen as described below.

3. Allow the smear to dry, preferably overnight, before it is stained.

(iii) PVA fixative-preserved specimens.

1. Preserve 1 part specimen in 3 parts PVA fixative. Mix

thoroughly. Fix for at least I h. Specimens keep indefinitely.

2. Add 1 drop of PVA-fixed specimen to a slide.

(a) If there is little sediment, remove a portion of the sediment with a Pasteur pipette.

(b) If there is abundant sediment, mix the specimen thoroughly, and add 1 drop of specimen to a slide with applicator sticks or a Pasteur pipette.

3. Spread the material over the center third of the slide by rolling the specimen with an applicator stick. Remove any lumps. The film should extend to both the top and bottom edges of the slide, as this helps prevent peeling.

4. Allow the slide to dry overnight at room temperature or 35°C. In an urgent situation, the slide can be dried for 4 h at 35°C and then stained.

Trichrome staining procedure. Table 4 outlines the steps in the trichrome staining procedure.

Permanently stained slides may be mounted with a cover slip or may be air dried and examined after oil is added. Slides should be examined at a magnification of ×400 to ×500 or greater after they are scanned under lower power to find optimal areas. A ×50 oil immersion objective is particularly helpful, as it allows the easy use of a × 100 oil immersion objective for the detailed examination of organisms while allowing more rapid screening with a ×50 objective. Oculars of × 5 or × 6 can provide the same result. A ×20 dry objective may also assist in screening.

Permanently stained slides should be kept for 2 years.

Stain reactions. In an ideal stain, the cytoplasm of cysts and trophozoites is blue-green tinged with purple. *Entamoeba coli* cyst cytoplasm is often more purple than that of other species. Nuclear chromatin, chromatoid bodies, erythrocytes, and bacteria stain red or purplish red. Other ingested particles such as yeasts often stain green. Parasite eggs and larvae usually stain red. Inflammatory cells and tissue cells stain in a fashion similar to that of protozoa. Color reactions may vary from the above.

Incompletely fixed cysts may stain predominantly red, and organisms which have degenerated before fixation often stain pale green. Poor fixation due to an inadequate mixing of the specimen in fixative may result in both of these,, appearances. In some specimens, degeneration has occurred before the specimen is placed in fixative, either in the patient before the specimen was evacuated or because of delay in

fixing the specimen.

Troubleshooting the trichrome stain. Except for problems with delayed or inadequate fixation as noted above, problems with the trichrome stain are usually related to reagents other than the stain. If crystalline material is apparent after the specimen is stained, the crystals are probably mercuric chloride in the fixative which was not adequately removed because the iodine in the alcohol-iodine solution was too weak or because the slide was in this solution too short a time. If crystals are present after treatment with proper-strength iodine-alcohol, they are present in the specimen, which is thus unsatisfactory, and another specimen should be requested.

If the stain appears washed out, it is likely that the slide was destained too much. This washed-out appearance can be either because the specimen was left too long in the acid-alcohol destain or because the alcohol wash after the acid-alcohol destain had become acidic as a result of transfer by previous slides.

The trichrome may become diluted by carry-over alcohol if more than 10 slides per day are stained in one Coplin jar. To restore the stain, the lid may be left off for several hours to allow alcohol to evaporate, and then the volume is replaced with new stain.

Control slides should be used to monitor the staining. Specimens containing protozoa are best for controls; however, feces containing inflammatory cells or added buffy-coat leukocytes also are satisfactory.

Restaining. Should the stain be unsatisfactory, the slide can be destained by placing it in xylene to remove the cover slip or immersion oil and then placing it in 50% alcohol for 10 min to hydrate the slide. Destain the slide in 10% acetic acid in water for several hours, and then wash it thoroughly first in water and then in 50 and 70% alcohols. Place the slide in stain for 8 min, and then complete the stain procedures. It is helpful to eliminate or shorten the destain step.

Acid-fast stain for *Cryptosporidium* sp. Acid-fast staining for *Cryptosporidium sp.* has recently become important because this parasite is now recognized as a cause of severe diarrhea in immunodeficient patients such as those with AIDS, and it can cause transient diarrhea in immunocompetent individuals. The modified acid-fast stain recommended is similar to that used to stain *Nocardia* spp. in that it uses milder acid decolorization. A variety of acid-fast and fluorochrome staining procedures have been described for *Cryptosporidium spp.*, and all the procedures appear to work.

The following procedure is useful for staining *Nocardia* species as

well as *Cryptosporidium* species. This procedure may be used on fresh or Formalinfixed material or on material from concentration procedures. If the specimen is liquid, centrifuge it, and use the sediment to prepare a smear.

1. Pick a portion of material with an applicator stick, mix the material in a drop of saline, spread it on a glass slide (1 by 3 in.), and allow it to dry.

2. Fix the dried film in absolute methanol for 1 min, and air dry the slide.

3. Flood the slide with Kinyon carbol-fuchsin, and stain the smear for 5 min.

4. Wash the slide with 50% ethyl alcohol in water, and immediately rinse it with water.

5. Destain the smear with 1% sulfuric acid for 2 min or until no color runs from the slide. 6. Wash the slide with water.

7. Counterstain the smear with Loeffler methylene blue for 1 min.

8. Rinse the slide with water, dry it, and examine the smear with oil immersion.

The results are that *Cryptosporidium* oocysts stain bright red, and background materials stain blue or pale red.

5.5.6 Egg Counts

Egg-counting methods are used in clinical studies to assess the intensity of infections (especially infections by intestinal nematodes) and the efficacy of therapeutic agents, and these methods are commonly used for epidemiologic studies. Methods used for scientific studies, such as Kato thick smear or Stoll egg counting, require greater accuracy than methods used for patient care. The simplest, most practical method is to use a standard fecal suspension which contains approximately 2 mg of feces mixed in a drop of saline and covered with a cover slip. The entire cover slip is examined at a magnification of 100×, field by field, and the number of eggs is counted. For research work, the density of the smear can be standardized with a light meter, but this standardization is not essential for patient care. The number of eggs per cover slip provides a rough index of the severity of the infection.

5.5.7 Duodenal Material

The examination of duodenal fluid or duodenal biopsy material may be useful for the diagnosis of giardiasis, strongyloidiasis, or other

upper intestinal parasite infections in patients in whom parasites cannot be detected in the feces. In addition, duodenal fluid occasionally can be useful in showing whether helminth eggs are originating in the biliary or intestinal tract. Duodenal material may be obtained by passing a tube through the nose and stomach into the upper small intestine and then aspirating enteric fluid. As an alternative, a string test may be used. A weighted gelatin capsule attached to a string is swallowed, and the proximal end of the string is taped to the face of the patient. Over a period of several hours, helped with small sips of water, the string reaches the upper small intestine. After 4 to 5 h, the string is retrieved, and the material on the bilestained portion is stripped from the string and examined for parasites with direct wet mounts or with permanent stains when wet-mount findings are questionable. Aspirated duodenal fluid is examined in a similar fashion. The material for permanent stains can be fixed in Schaudinn or PVA fixative, although the latter may adhere better to the slide. If question able organisms are seen in the direct wet mount, the coverslip can be removed, the material can be mixed with a drop of PVA fixative, and a film can be made for later permanent staining.

A duodenal biopsy can be used to demonstrate *Giardia* organisms. A biopsy is usually obtained by a swallowed biopsy capsule. In searching for *Giardia* spp., it is generally preferable to make both impression smears and sections of biopsy tissue. *Giardia spp.* are usually present in mucus or attached to epithelium rather than in tissue. Biopsies occasionally can confirm a diagnosis of strongyloidiasis or cryptosporidiosis.

5.5.8 Sigmoidoscopic Material

Materials obtained by sigmoidoscopy may be helpful in the diagnosis or monitoring of amebiases, schistosomiasis, or cryptosporidiosis. Patients suspected of having amebiasis may have ulcerations of the colon which can be visualized by sigmoidoscopy or colonoscopy. Scrapings or aspirates of material from ulcers can be examined by direct wet mounts and permanent stain as described above. The finding of typical, erythrophagocytic, motile trophozoites in direct wet mcunts or in permanently stained preparations allows a diagnosis of amebiasis. Material is best aspirated with a pipette or scraped with an instrument. Swabs should not be used, as the parasites may be killed or trapped by swab material.

Biopsy material for amebiasis should be processed for surgical pathology and then examined for ulcers containing amebae. The periodic

acid-Schiff stain counterstained with hematoxylin is particularly helpful because amebae stain more positively with periodic acid-Schiff stain than do inflammatory cells, and amebae show typical amebic nuclei. Of course, there are no amebic cysts in tissue.

Biopsy material for schistosomiasis is better examined in teased preparations than in sections, as the entire thickness can be examined at once, and the viability of eggs can be determined by observation of the movement of the larvae within the eggs.

In cryptosporidiosis, biopsy material shows organisms at the luminal surface of the epithelial cells, but the organisms are small, and the study of structural detail requires electron microscopy.

5.5.9 Abscess Material

Abscesses suspected of being caused by *Entamoeba histolytica* may be aspirated, and the material may be submitted to the laboratory. The last material aspirated is most likely to contain amebae. Material may be examined microscopically in wet mounts and permanent stains, and in addition, it can be cultured for amebae if bacteria are also added to the culture as described below. Abscess material is often thick and difficult to examine. It may be treated with streptokinase and streptodonase enzymes to liquefy the specimen.

1. Reconstitute streptokinase and streptodonase per the instructions of the manufacturer.
2. Add 1 part enzyme solution to 5 parts aspirated material.
3. Incubate the mixture at 35 to 37°C for 1 h. Shake the mixture at intervals.
4. Centrifuge the mixture at 300 to 400 × g for 5 min.
5. The sediment may be used for microscopic examinations for amebae (wet mounts and permanent stains) and for the culture of amebae.

5.5.10 Cellophane Tape

Cellophane tape is used for finding the eggs of *Enterobius vermicularis* or *Taenia* species from the perianal area. The tape used must be clear cellophane and not slightly cloudy or opaque. Alternatively, a Vaspar swab may be used. Specimens from more than 1 day may be required to diagnose light infections.

5.5.11 Examination of Cellophane Tape

1. If the specimen is difficult to examine, raise the tape from the front of the glass slide, add a drop of toluene to the slide, and replace the tape smoothly with an applicator stick. (Remember, *Enterobius*

spp. and *Taenia solium* eggs are infective!)

2. Examine the entire tape, including the edges, with × 100 magnification (× 10 objective).

3. Confirm suspicious objects with high dry objectives (×40 to ×50).

5.5.12 Culture for Amoebae

Cultures for amoebae have improved detection in some studies, but they are not widely used. Although *Giardia spp.* have been cultured in research laboratories, cultures are not useful for diagnosis.

A variety of culture media for amebae have been described, and some may be purchased from commercial medium manufacturers. The method described here uses the modified charcoal agar slant diphasic medium described by McQuay.

1. Place 3 ml of sterile 0.5% phosphate-buffered saline on a charcoal agar slant.

2. Add approximately 30 mg of sterile rice starch to the tube.

3. Warm the medium to 35°C before it is inoculated.

4. (a) Inoculate the medium with fecal specimen (approximately 0.5 ml of liquid specimen or a 0.5-cm sphere of formed specimen) which is mixed with the saline overlay.

(b) If abscess material is cultured, bacteria must be inoculated into the culture in addition to the inoculation with 0.5 ml of specimen. A heavy inoculum with *Clostridium perfringens* or *Escherichia coli is* satisfactory.

5. Incubate the culture at 35°C.

6. At 24 and 48 h, remove 1 drop of liquid from the lowest point of the overlay, and prepare a wet mount.

7. Examine the wet mount for amebae.

8. Permanent stains can be prepared by the fixation of sediment in PVA fixative, with the subsequent preparation of smears and staining.

5.5.13 Larval Maturation

Larval maturation studies, sometimes referred to as cultures, can be performed on fecal specimens applied to wet filter paper. Nematode larvae such as *Strongyloides spp.* or hookworm mature to the filariform stages in the culture container and migrate from feces into water, where they are detected microscopically. The procedure can be performed in a petri dish with a square of filter paper or in a large test tube with a strip of filter paper.

1. Smear approximately 0.5 g of feces on the filter paper.

2. (a) For the tube method, insert the filter paper strip into the tube so that the bottom of the strip is in 3 ml of water. The feces-smeared portion of the strip need not be immersed in the water.

(b) For the petri dish method, place feces on one half of a piece of filter paper. Lay the feces-bearing end of the filter paper on a glass rod or a portion of an applicator stick in the petri dish. Add approximately 3 ml or sufficient water so that the feces-free end of the filter paper is in the liquid.

3. Leave the tube or dish at room temperature in the dark. Add water as needed to ensure that the filter paper is in contact with the water.

4. Examine the liquid for larvae either by direct microscopic examination with an inverted microscope or by examination of a wet mount of sediment from the liquid. With the petri dish method, the surface of the feces also may be examined with a dissecting microscope.

5. Examine the specimen on days 3, 5, and 7. Strongyloides filariform larvae are found on days 2 and 3, and hookworm larvae are found on days 5 through 7. Larvae are identified by their morphological characteristics.

5.5.14 Adult Worms

Adult worms, or objects suspected to be adult worms, may be submitted to the laboratory. The laboratory must determine if these are helminths and, if so, if they are parasites. Identification characteristics are described in standard references. Tapeworm proglottids, particularly those of the *Taenia* species, are difficult to differentiate grossly unless they are cleared so that the internal structure can be seen and the number of lateral uterine branches can be counted. One procedure for clearing the proglottids is outlined below.

Clearing *Taenia* proglottids and other helminths. Proglottids are first relaxed by placing them in warm saline (56°C) for 1 h and then clearing them in carbolxylene while they are kept flat. They may be kept flat in a number of ways. One way is to press the proglottid between two glass slides held together with membrane clips or string. Clearing takes from several hours to overnight.

The proglottid is examined under a dissecting microscope or with a hand lens, and the uterine branching is observed. Glycerine and beechwood creosote can also be used with good results. Cleared proglottids may be mounted or stained if desired.

Small nematodes may also be cleared in carbolxylene or beechwood creosote and mounted in permount or balsam. This method is particularly good for hookworm adults.

5.6 BLOOD AND TISSUE PARASITES

Blood and tissue parasites whose diagnostic forms circulate in the peripheral blood are generally diagnosed by the demonstration of parasites in Giemsastained thick or thin films of blood. Special concentration techniques may be helpful for the diagnosis of some diseases such as filarial or trypanosomal infection. Other tissue parasites which do not circulate in the blood may be diagnosed by the detection of parasites in skin snips, lesion scrapings, body fluids, or biopsy material or by the detection of antibody or antigen in serum or other body fluids.

5.7 COLLECTION AND HANDLING OF BLOOD SPECIMENS

The timing of the collection of blood specimens depends on the parasite disease suspected. For example, for certain filarial infections, specimens are best obtained between 10:00 p.m. and midnight, whereas for other infections, specimens are best obtained during the day. In malaria, the numbers and stages of parasites in the peripheral blood vary with different parts of the cycle.

Blood films are best made from blood which is not anticoagulated, such as that obtained from finger stick or ear lobe puncture. Anticoagulants may interfere with parasite morphology and staining. Care should be taken that the alcohol disinfectant is allowed to dry before the area is punctured, or there may be fixation of erythrocytes, which will interfere with the preparation and staining of thick films. Both thick and thin films can be prepared from blood obtained by venipuncture, although it is best if the blood remaining in the needle of the venipuncture device is used, because it is anticoagulant free. Thick and thin films can *be prepared* from blood that is anticoagulated, but the staining characteristics are not as good. EDTA-anticoagulated blood is better for staining than citrate- or heparin-anticoagulated blood.

Both thick and thin blood films are useful. Thick films are more sensitive because the same amount of blood can be examined in a thick film in 5 min as can be examined in a thin film in 30 min. However, thin films allow the study of the effects of parasites on erythrocytes and provide better parasite morphology.

Thick and thin films may be prepared on separate slides or on the

same slide, with the thick film at one end and the thin film at the other end.

The thick film is prepared by spreading 1 drop or puddling several small drops of blood into an area approximately 1.5 cm in diameter. A properly prepared thick film should be thin enough so that newspaper print can barely be read through it. If the film is too thick, it will fragment and peel, and if the film is too thin, the increased sensitivity will be lost. Thick films should be allowed to dry overnight and should be stained within 3 days. They must not be heated, and they should be protected from dust. If the erythrocytes are fixed, they will not dehemoglobinize. If prompt examination is required, prepare a slightly thinner thick film, dry it for 1 h, and stain it.

The thin film is prepared in the same manner as a film for a differential leukocyte count. A small drop of blood is placed on one end of a microscope slide. A second slide held at an acute angle of 30 to 45° is backed into the drop of blood, which spreads along the junction of the slides. The spreader slide is then pushed along the slide, and it pulls the drop of blood along behind the angled edge of glass. A properly prepared thin film should have a significant area near the end which is only one erythrocyte thick and in which the erythrocytes show good morphology. The angle and speed of the spreader slide and the size of the drop of blood will influence the thickness and size of the film.

Slides with only a thin film can be fixed by being immersed in absolute methyl or ethyl alcohol for 1 min and allowed to air dry. If the thick film is on one end of the slide and the thin film is on the other end, the thin film is fixed by a brief flooding or by immersion in alcohol and allowed to air dry, while the thick film is protected from alcohol or alcohol fumes. In a well-ventilated area, the slide may be dried vertically with the thick film up or horizontally after the thick film is covered with a dry paper towel.

Thick and thin films are best stained with Giemsa stain, as it provides the most detailed and intense staining of parasites. Wright stain can be used for thin films but not for thick films, as it contains alcohol, which will fix the erythrocytes. Wright stain does not stain parasites as well as Giemsa stain. The staining procedure is outlined below.

5.7.1 Tissue

Biopsy or necropsy tissue may be examined by histology sections or impression smears.

To prepare impression smears, tissue should be blotted to remove as much blood or other fluid as possible and then pressed against glass slides (1 by 3 in.) to make a series of impressions. Tissue should stick to the slide slightly and leave an irregular film on the slide. Similar impressions may be made on multiple slides from the same portion of tissue. Portions of biopsy tissue with different gross appearances can be used with the impressions from each portion placed in a longitudinal row. Impressions must be close together, preferably with slight overlapping to make slide scanning easier. Impressions from small fragments may be placed in a small area (1 cm in diameter). After being dried, the area with impressions is circled with a diamond marker to facilitate the location and scanning of the material. Fixatives and stains appropriate for the parasites suspected are used. If amebiasis is suspected, impression smears must be fixed promptly in Schaudinn fixative and not allowed to air dry. For most other parasites, the slides are allowed to dry before fixation in methyl alcohol. Giemsa is the usual stain, but other stains such as Gram-Weigert or hematoxylin may be used depending on the parasite suspected.

5.7.2 Aspirates of Bone Marrow or Spleen

Aspirates of bone marrow or spleen may be useful in the diagnosis of infections such as leishmaniasis, trypanosomiasis, and occasionally malaria. In such instances, Giemsa stains of alcohol-fixed bone marrow films are most useful. Splenic aspiration is rarely performed in the United States because it is dangerous.

5.7.3 Fluids

Fluids such as tissue aspirates, cyst fluid, bronchial washings, cerebrospinal fluid, pleural fluid, and peritoneal fluid can be examined directly, or they can be centrifuged and the sediment examined by wet mounts or stains (or both), depending on the parasite suspected, as described above for abscesses or tissue.

5.7.4 Skin Snips

Skin snips may be useful in the diagnosis of microfilarial infections such as onchocerciasis in which the parasites circulate in the skin and not the blood. A small (2-mm) skin snip is taken with a needle and a knife. The needle point is stuck into the skin, and the skin is raised. With a sharp knife or razor blade, the skin is excised just below the needle. Alternatively, a scleral punch may be used. The skin snip is then placed in a small volume (0.2 ml) of saline in a tube or a microtiter well, teased, and allowed to stand for 30 min or more. The

microfilariae migrate from the tissue into the saline, which is then examined microscopically to demonstrate the wiggling microfilariae.

5.7.5 Concentration Procedures for Blood

A number of procedures have been described for the concentration of blood specimens. Most of these procedures have been developed to diagnose filarial infections.

The three most widely used methods are membrane filter, saponin lysis, and Knotts concentration. Procedures for the first two methods will be given here, as these methods are the most sensitive. Membrane filter techniques use 5- or 3-μm filters. Both filters give satisfactory results, but the procedures with the Nuclepore filters do not require the lysis of erythrocytes. Parasites on filters are often not as suitable for morphologic study as are those in thick films.

5.7.6 Membrane Filter Concentration for Filariae

1. Collect approximately 7 ml of blood in EDTA.
2. With a syringe and firm pressure, pass 5 to 7 ml of blood through a 5-μm Nuclepore filter held in a Swinney adapter.
3. Wash the membrane several times with a small amount of distilled water or physiologic saline.
4. The moist filter may be examined directly or fixed and stained in the usual fashion for a thin blood film.

5.7.7 Saponin Lysis Concentration for Filariae

The saponin lysis method (32) can be performed on either EDTA- or citrate-anticoagulated blood. Saponin solution to lyse erythrocytes is available in most laboratories for use with automated hematology instruments.

1. Centrifuge up to 10 ml of blood at 150 x g for 10 min.
2. Remove and discard the plasma.
3. Mix the packed erythrocytes with 50 ml of 0.5% saponin solution in 0.85% saline.
4. Mix the solution at intervals for 15 min.
5. Centrifuge the solution at 650 x g for 10 min.
6. Decant and discard the supernatant (there should be about 1 ml of sediment).
7. Spread several drops of sediment on a glass slide (1 by 3 in.), and examine two such uncovered wet mounts for motile microfilariae. Allow the wet mounts to dry before they are fixed and stained.
8. Prepare four or five similar wet mounts and examine them as

described above. To each slide, immediately add 2 drops of 1% acetic acid solution and mix it well (microfilariae will be killed and straightened). Allow the slide to air dry.

9. Dip the dried slides in buffered methylene bluephosphate solution.

10. Rinse the slides in distilled water, and let them air dry.

11. Stain the mounts for 10 min in a 1:20 dilution of Giemsa stain in buffered water.

12. Examine the slides microscopically.

5.8 STAINING PROCEDURES

5.8.1 Giemsa Stain Procedure

The procedures for staining thick and thin films differ. Staining is usually done in a Coplin jar. The stain must be made fresh each day.

Stain slides with only a thin film as follows.

1. Fix and dry the blood film as described above.

2. Prepare a 1:40 dilution of stock Giemsa stain in neutral buffered water, pH 7.0 to 7.2 (generally, 2 ml of Giemsa stock plus 38 ml of buffered water with 0.01% Triton X-100).

3. Stain the film for approximately 60 min (the time, which will vary slightly with different lots of stock Giemsa stain, can be determined by the staining of leukocytes and erythrocytes).

4. Wash the slide briefly by dipping it in buffered water.

5. Air dry the slide in a vertical position.

Note that, alternatively, a 1:20 dilution for 20 to 30 min may be used.

Stain slides with only a thick film as follows.

1. Do not fix the slide.

2. Prepare a 1:50 dilution of stock Giemsa stain in neutral buffered water (pH 7.0 to 7.2).

3. Stain the film for approximately 50 min (the optimal time may vary with different lots of stain).

4. Wash the slide by placing it in buffered water for 3 to 4 min.

5. Air dry the slide in a vertical position.

For combination thick and thin films, the procedure is as follows.

1. Fix the thin film but not the thick film as described above.

2. Stain the film in a 1:50 dilution of Giemsa stain in neutral buffered water (pH 7.0 to 7.2) for approximately 50 min.

3. Rinse the thin film briefly by dipping it in buffered water.

Wash the thick film by immersing it in buffered water for 3 to 5 min.

4. Dry the slide in a vertical position with the thick film down.

5.8.2 Gram-Weigert Stain Procedure

The Gram-Weigert stain is used to stain the cyst walls of P. *carinii* cysts. It also stains fungi and many bacteria. Impression smears, sediment smears, or sections are fixed in methanol and air dried. For sections, when reagents are added, flood the slide gently from the end opposite the section, and rinse the slide carefully so that the tissue is not washed from the slide.

The stain procedure is as follows.

1. Stain the slide with eosin Y for 5 min.
2. Wash the slide with water.
3. Stain the slide with crystal violet for 5 min.
4. Rinse the crystal violet from the slide with Gram iodine solution.
5. Leave the iodine solution on the slide for 5 min.
6. Rinse the slide with water.
7. Blot the smears carefully (do not blot the sections).
8. Wipe the reverse of the slide.
9. Air dry the slide completely.
10. Decolorize the smear in aniline-xylene, agitating the slide gently until no purple runs from it (the use of a second Coplin jar of aniline-xylene after the majority of blue stain has been removed aids the visual assessment of decolorization).
11. Rinse the slide in xylene.
12. Air dry the slide, add immersion oil to it, and examine it.

P. carinii cysts and fungi stain dark blue and somewhat irregularly. Cell nuclei may stain blue if they are inadequately decolorized, but they are not as dark as *P. carinii* cysts.

5.8.3 Culture Procedures for Blood and Tissue Parasites

Culture procedures have been developed for a number of blood and tissue parasites, but these procedures are used primarily in research. The culturing of *Leishmania* spp. and *Trypanosoma cruzi* may be helpful for diagnosis, and the procedures are easy to use.

Biopsy or blood specimens may be cultured for *Leishmania* spp. or *T. cruzi* with Novy-MacNeal-Nicolle (NNN) medium. Biopsy specimens are ground in a small amount of saline. Biopsies from skin lesions or other tissues which may contain bacteria may have penicillin

(0.1 ml of 1,000 U/ml) added to the medium with the inoculum. The inoculum is 1 drop of ground tissue or blood. Incubate the culture at room temperature (22°C), and at days 3 and 7, examine a direct mount of liquid from the bottom of the slant at ×400 magnification. These cultured organisms are potentially infective for humans.

5.9 URINE

Urine specimens usually are examined for the eggs of *Schistosoma haemotobium* or the trophozoites of *Trichomonas vaginalis,* although occasionally the larvae of *Strongyloides stercoralis* may be found in patients with hyperinfection syndrome. Urine is the usual specimen for the diagnosis of *Trichomonas* infection in males. See below (Vaginal Material) for culture method. Urine is centrifuged, and the sediment is examined microscopically.

5.10 SPUTUM

Sputum may be examined to diagnose *Paragonimus* infection or hyperinfection due to *Strongyloides stercoralis*. Occasionally an amebic abscess or hydatid cyst may rupture, and amebic trophozoites or hydatid sand, respectively, may be found in sputum. *Entamoeba histolytica* must be differentiated from *Entamoeba gingivalis,* which may be found in the oral cavity of over 30% of people. Occasionally, the migrating larvae of ascarids, strongyloides, or hookworm can be found. Sputum may be examined directly by wet mount or treated with a mucolytic agent such as Nacetyl-cysteine and then concentrated by simple centrifugation, with subsequent examination of the sediment.

5.11 VAGINAL MATERIAL

T. vaginalis frequently infects the vagina, and *Enterobius vermicularis* adults or eggs occasionally may be found. Direct wet mounts of vaginal material for typical, tumbling *T. vaginalis* organisms are widely used and generally allow the diagnosis of symptomatic infection, but wet mounts are not as sensitive as culture methods.

Vaginal material is best submitted as liquid in a tube, although swabs submitted in a small amount of saline may be used. A drop of the material is covered with a cover slip and examined with reduced light. To culture, 1 or 2 drops of urine sediment or vaginal exudate are inoculated into tubes of warmed, modified Diamond medium. If vaginal swabs are submitted, the swab is immersed in the medium and pressed against the side of the tube to express material. Tubes are incubated at 35°C, and drops of culture are examined by wet mount at 48 and 72 h for motile trophozoites.

5.12 REFERRAL OF MATERIALS

Few laboratories perform complete parasitological examination, whereas many perform limited studies, and some perform none. Referral laboratories may provide services not available in the individual laboratory and can provide consultation on specimens with questionable laboratory findings. Referral laboratories with a special interest and competence in parasitology may be found in major cities, university medical centers, and state public health laboratories. The major national resource is the Centers for Disease Control in Atlanta, Ga. Specimens for the Centers must be sent via the state health laboratory, and appropriate clinical information must be provided. Of course, the recommendations of the specific referral laboratory should supersede these guidelines.

5.13 SAFETY

The parasitology laboratory has infection hazards for personnel. Blood, feces, and other body materials as well as parasite cultures may be infective. Eggs of *Ascaris* spp. can survive and embryonate even in Formalin, and *Cryptosporidium* oocysts are hardy. In fresh fecal specimens, the cysts of *Entamoeba histolytica* and *Giardia spp.*, the oocysts of *Cryptosporidium* spp., the eggs of *Enterobius vermicularis, Taenia solium,* and *Hymenolepis nana,* and the larvae of *Strongyloides stercoralis* may be infective. In addition, feces may contain other infectious agents such as hepatitis A, rotavirus, *Salmonella spp., Shigella spp.,* and *Campylobacter* spp. Blood and tissue specimens can be infectious for trypanosomes, *Leishmania* spp., malaria, and *Babesia spp.,* as well as for non-A, non-B hepatitis, hepatitis B, and possibly AIDS.

TABLE 5.5 : Handling of specimens for referral

Specimen	Handling
Feces, for	
Helminths	**Fix in 10% buffered Formalin.**
Protozoa	**Fix a portion in 10% buffered Formalin and either fix a portion in PVA fixative or prepare three Schaudinn-fixed fecal films.**
***Cryptosporidium* spp.**	**Fix a portion in 10% buffered Formalin.**
Material from suspected amebic abscess	**Place the last material aspirated in a sterile tube and send it on ice for culture (do not freeze). Prepare Schaudinn-fixed**

	fecal films, or fix a portion in PVA fixative. Obtain serum for serology.
Duodenal aspirate	Centrifuge, and remove the supernatant. Prepare two films from sediment. Fix in Schaudinn or PVA fixative. Preserve the remainder of sediment in 10% Formalin.
Urine, for	
Trichomoniasis	Centrifuge. Cover the sediment with sterile saline and send it on ice (not frozen) for direct mounts and culture.
Schistosomiasis	Centrifuge entire midday urine. Add an equal volume of 10% buffered Formalin to the sediment.
Sputum, for	
Nematode larvae or Paragonimus eggs	Break up mechanically or digest 1 part sputum plus 5 parts 3% NaOH for 1 h. Centrifuge, and preserve the sediment in an equal volume of 10% buffered Formalin.
Amebae	Prepare films fixed in Schaudinn fixative, or fix a portion in PVA fixative.
Blood	
Malaria and babesiasis	Send unstained and, if available, Giemsa-stained thick and thin films. Fix thin film (but not thick) in alcohol before it is sent.
Filariasis	Send 5 ml of citrate- or EDTA-anticoagulated blood on ice (not frozen). Unfixed thick films may be sent in addition. Send serum for serologic tests.
Trypanosomiasis	Send 5 ml of anticoagulated blood as for filariasis (above).
Send fixed thin films.	
Cerebrospinal fluid	
Trypanosomes, toxoplasma, leishmania, trichinella	Send on ice (not frozen). Send in a sterile container without refrigeration.

Free-living amebae	
Sigmoidoscopic material	**Fix films in Schaudinn fixative or mix material with PVA fixative.**
Tissue	**For impression smears when E. *histolytica* is suspected, fix in Schaudinn or PVA fixative. When toxoplasma, leishmania, *Pneumcoystis* spp., or *Trypanosoma cruzi* is suspected, prepare multiple impression smears and fix in methyl alcohol. For surgical pathology, fix the tissue in buffered Formalin.**
Whole worms or proglottids	**Wash debris from the specimen and send it in saline. If there are multiple worms or proglottids, some may be fixed in Formalin.**

Reagents such as mercury-containing fixatives may be toxic, and solvents such as ether may be flammable. These materials must be handled and discarded properly.

5.14 QUALITY ASSURANCE

The parasitology laboratory must have an up-todate procedure manual and appropriate reference materials which might include color atlases or 35-mm slide collections permanently stained glass slides, wet fecal material containing parasites, and one or more standard reference books on laboratory methods or general medical parasitology. The persons performing parasitic examinations must be competent in the identification of parasites which might be found in patients from whom they receive specimens. Methods should allow the ready use of outside consultants, if there is a question of diagnosis. Personnel may maintain proficiency through participation in formal courses or workshops, review of self-study sets, and periodic review of known positive materials. Participation in external survey programs is particularly valuable, as the performance of the laboratory in the identification of unknown specimens can be compared with the performance of other laboratories.

If a laboratory is unable to do accurate parasitology because of either the types of procedures offered or the quality of personnel available, it should arrange to have specimens appropriately prepared and submitted to a reference laboratory.

6

Microbes in Milk and its Products

6.1 NORMAL FLORA OF MILK

Milk secreted in to the udder of healthy cows is sterile. It has a pH of about 6.8. Some saprophytic bacteria of the outside environment, such as species of Micrococcaceae, Bacillacease, Escherichieae, Corynebacteriaceae and Lactobacillaceae, are able to grow a short way up into the milk duct of the teat, so that the first milk drawn usually contains from hundreds to thousands per milliliter. This milk should be discarded. Except in cases in which extra precautions are taken at the time of milking, the milk receives contributions of organisms from the pail or mechanical milker and other dairy utensils, from soil and dust in the air, from the flanks, tail and udder of the cow, and from the hands of milkers. Yeasts, molds and numerous other sapro-phytes find their way into the milk. These constitute the *normal flora* of market milk.

The presence of these nonpathogenic bacteria in milk is usually not a serious matter, but if they are allowed to multiply they can and will cause the milk to sour quickly, putrefy, or develop undesirable flavours or conditions like bitter milk (*Streptococcus cremoris*), blue milk (*Pseudomonas syncyanea*) red milk (*Serratia marcescens*), ropy (slimy) milk (*Alcaligenes viscolactis, Klebsiella aerogenes* and others). Their presence in very large numbers shows the milk to be stale or dirty.Entrance of these organisms into milk in large numbers can be prevented only by clean handling and routine *effective* sanitization of milk-handling equipment. Their development can be retarded by prompt refrigeration and by pasteurization.

Figure. 6.1: Sanitary milk production. Note the cleanliness of the cows, the absence of dung and straw, the good lighting, ventilation and milking machines.

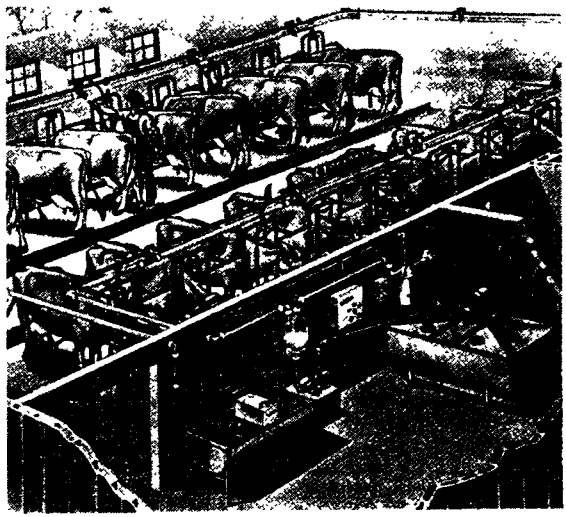


Figure. 6.2 : A two-unit, barn-type combine milker. Milk is conveyed, *directly from the cow's udders* in a completely enclosed glass, plastic or stainless steel pipe, through a filter and into a refrigerated bulk cooler. The milk is never open to contamination from the air or environment. After milking, the assembled units are connected with a manifold washer in the milk room. Detergent is placed in a special compartment in the automatic wash. By pressing a button the entire system is pre-rinsed, washed and rinsed again automatically. Surely is far cry from the romantic (but insanitary) "pretty milk maid"!

6.1.2 Pasteurization

Pasteurization consists in holding the milk in tanks at 145° F. (63°C.) for 30 minutes ("low-temperature holding" or LTH) and immediately refrigerating. *Disinfection* of the milk is accomplished;

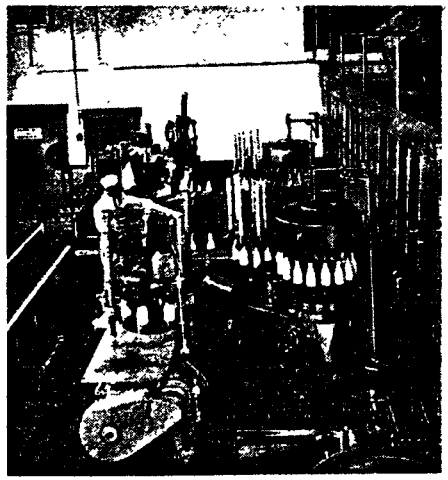


Figure. 6.3: Bottling milk in a sanitary dairy. All of the piping can be demounted in a few minutes for steam sterilization and the floors and walls hosed down. In some plants light, transparent, plastic tubing is used. The bottles in this picture have been steam sterilized just before filling.

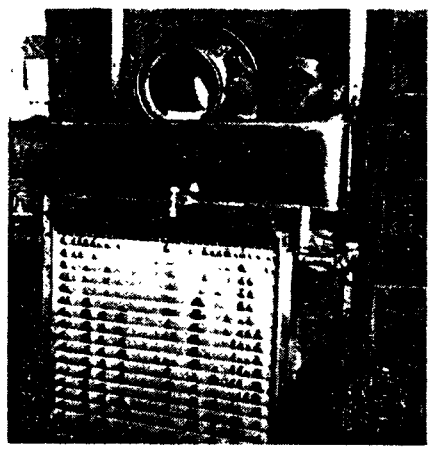


Figure. 6.4: Conditions contributing to high bacterial counts in milk. Milk cooler open to dust and dirt; not readily cleaned.

not sterilization. In many dairies the same result is achieved by heating the milk rapidly in a tube or in thin layers between metal plates, to 71.6° to 80°C. and holding at that temperature for 15 to 30 seconds, then cooling. These high-temperature short-time (HTST or *flash*) methods save time and money and are effective so far as sanitation of milk is concerned.

6.2 CHANGES IN FLORA OF MILK

Since milk is an excellent medium for bacterial growth, the numbers of bacteria in it increase steadily the longer it stands, even if pasteurized and refrigerated. Even if milk is refrigerated so that growth of thermophiles and heat-resistant sporeformers is retarded, psychrophilic species will grow. Under storage at 3°C. some of these can soon cause discolouring, ropiness, off-flavours and other undesirable condition. Common dairy psychrophiles are *Pseudomonas, Achromobacter, Alcaligenes, Flavobacterium* and *Micrococcus* species. Not all are killed by pasteurization and the survivors can multiply rapidly.

It is important that refrigeration be at very near 0°C. Much commercial refrigeration is at about 10°C., a temperature ideal for many psychrophiles. The temperature of the average household refrigerator (not freezer) is about 5°C. At best, such refrigeration is effective for not much over 24 hours.

6.2.1 Raw Milk

If allowed to stand at about 22°C., the flora in raw market milk rapidly undergo a series of changes. Numbers of bacteria increase to almost astronomical figures within 24 hours. Enterobacter-iaceae, lactic streptococci (e.g., *S. lactis*). *Micro-coccus,* some sporeformers such as *Bacillus polymyxa* and other species of saprophytes that thrive at a pH near neutrality grow rapidly and dominate the picture at first. The lactose is fermented. As acidity increases, these species are inhibited. The aciduric lactic organisms then gain the ascendancy, especially *Lactobacillus* and species of *Leuconostoc.* Many *Clostridium* species will also ferment the lactose, but the presence of large numbers of *Clostridium* spores in fresh milk is not usual and indicates excessive contamination of the milk with soil or dung.

When the acidity reaches a pH of about 4.7, curdling occurs. The curd shrinks and settles out. Eventually, organisms capable of attacking lactic acid develop, especially aciduric yeasts and molds growing on the surface. These lower the acidity of the milk by destroying the acid and by producing alkaline products of protein decomposition: amines, ammonia and the like. Since the carbohydrates (lactose) have been decomposed by this time, fermentation does not reoccur.

Organisms capable of hydrolyzing the fat and casein now thrive: eucaryotic fungi, Bacillaceae, both aerobic and anaerobic,

Pseudomonadaceae and many other lipolytic and proteolytic saprophytes. As the oxidation-reduction potential of the milk is reduced, species of *Clostridium* and other anaerobes, both obligate and facultative, gain the ascendancy and the odours (ammonia, odouriferous amines, mercaptans, hydrogen sulfide, rancid odours) and effects of putrefaction become evident. The casein is hydrolyzed; the milk is darkened. After the situation has somewhat stabilized, as more prolonged decomposition continues, mainly by eucaryotic fungi and various microbial enzymes.

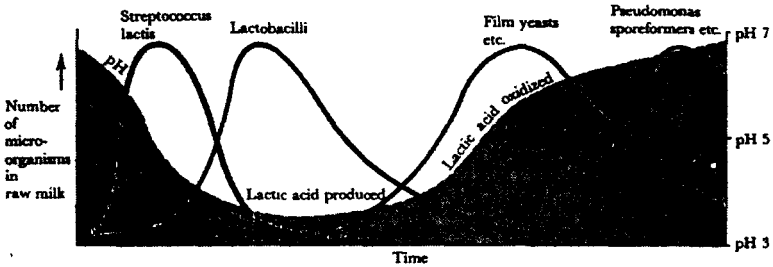


Figure. 6.5 : Changes in numbers and types of microorganisms in raw milk held at summer temperature. Note the initial decline in pH (increase in acidity) during fermentation of the lactose, with later rise in pH as the resulting lactic acid is metabolized and alkaline products of casein putrefaction accumulate.

6.2.2 Pasteurized Milk

Pasteurized milk does not promptly undergo souring because many of the lactose-fermenting species, being nons-poreformers, are killed by the heat of the process. The milk may then undergo *sweet curdling* caused by rennet formation by bacteria, especially proteolytic streptococci and aerobic sporeformers (*Bacillus*). Often pasteurized milk does not sour, but the casein undergoes digestion and, later, putref-action by the proteolyic enterococci, sporeformers and other thermoduric, proteolytic saprophytes that survive pasteu-rization.

6.2.3 Significance of Coliform Organisms in Milk

Coliform organisms are always present in market milk *Before* pasteurization. They are derived from hay, soil, dust, dung and utensils. They do not necessarily indicate human fecal contamination. However, these organisms do not surv-ive in milk in significant numbers if pasteurization is prop-erly carried out. It is possible to detect coliforms in milk by methods similar to those used for determining coliforms in water. Plating the milk in desoxycholate-lactose agar or violet-red bile agar (both selective media) are especially recommended. Red colonies may be counted as coliforms and transferred

for *completed test* if desired. The details for all important laboratory procedures in connection with dairy products are to be found in "Standard Methods for the Examination of Dairy Products".

Unless almost-surgical precautions are used, a few coliform organisms gain entrance to milk *after* pasteur-ization, during cooling or bottling. Small numbers are of little significance and are not inimical to health, but their numbers give a good indication as to post-pasteurization cleanliness and refrigeration or staleness. Coliform organisms are partic-ularly undesirably in milk to be used for cheese, since they cause rapid souring, gassy fermentation, and undesirable odours and flavours in cheese and other products made from the milk.

If considerable numbers of coliforms (more than about 1 to 5 per milliliter) are found in pasteurized milk it is evident that: (a) the milk has not been properly pasteurized; or (b) it has been excessively contaminated after pasteurization by unclean conditions, possibly by sewage, feces or dung; or (c) the milk has been held unduly long above about 15°C. after pasteurization; or (d) raw milk was mixed with it after pasteurization.

6.2.4 The Phosphatase Test

The presence of excessive numbers of coliform bacteria in pasteurized milk suggests staleness or improper pasteur-ization or illegal adulteration with raw milk. Concerning pasteurization, more definite and accurate information can rapidly be obtained by means of the *phosphatase test.*

This test is based on the destruction by pasteurization of the heat-sensitive enzyme, *phosphatase,* normally present in fresh milk. The test for presence or absence of phosphatase if based on the power of phosphatase to liberate phenol from phosphoric-phenyl ester added to a sample of the milk.

In this test, 0.5 ml. of milk sample is added to 5.0 ml. of *buffered substrate* (disodium phenyl phosphate buffered with $NaHCO_3$ Na_2CO_3 $2H_2O$) and held at 40°C. for 15 minutes. If the milk is unpasteurized or insufficiently pasteurized, the phosphate, liberating phenol. The phenol turns blue if 2,6-dichloroquinone chloroimide (CQC), with $CuSO_4$ as catalyst, is added to the mixture. The appearance of blue colour indicates the presence of free phenol liberated by undestroyed phosphatase in the milk. The colour of the tested sample of milk is then compared with the colour of standards containing known amounts of phenol and treated with the same reagent.

When milk is pasteurized at 143°F. for 30 minutes, 96 per cent of the enzyme is destroyed. Only a trace of blue colour should appear. Heating above 145°F. for 30 minutes insures complete inactivation of the phosphatase. No blue colour should develop.

When milk has been underheated (in respect to either temperature or time) or when there is an admixture of raw milk afterward, the phosphatase will be present in larger amounts than when the milk is properly processed, and a definite blue colour appears in the phosphatase test.

The phosphatase test can be made quantitative by comparing colour with known standards. This test will detect 0.5 per cent raw milk mixed with pasteurized milk, or one degree below standard temperature, or five minutes of underheating during pasteurization. Colour values (in the Sharer rapid method) greater than 0.5 μg of phenol per ml. of milk indicate progressive degrees of improper handling of milk.

6.2.5 Sanitary Significance of Phosphatase Test

In pasteurization, *Mycobacterium tuberculosis* (the most resistant of the nonsporeforming pathogens commonly found in milk) is destroyed more quickly than phosphatase. Therefore, a heat treatment

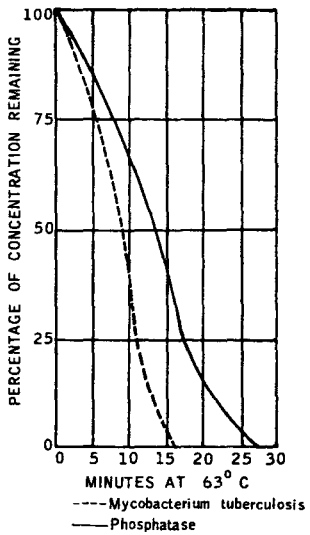


Figure. 6.6 : Representative comparative time-concentration curves for destruction of pathogenic bacteria (here represented by *Mycobacterium tuberculosis*) (--------) and phosphatase (--------) in milk during pasteurization at 63° C.

adequate to inactivate the enzyme likewise kills this organisms and all other common pathogenic microorganisms. A sample of milk that does not have more phosphatase present than the standard allows can be regarded as both safely pasteurized and free from subsequent contamination with raw milk.

Sources of Error. In some case the phosphatase seems to become reactivated after proper pasteurization.

Certain bacteria can give falsely positive results in properly pasteurized milk because they produce a thermo-stable phosphatase before pasteurization. This remains active even after proper pasteurization.

6.3 ENUMERATION OF BACTERIA IN MILK

In order to have some measure of the conditions under which milk has been produced and handled and to have a legal control over its sanitary quality, health departments and dairymen have set up various standards by which to judge milk. Important among these standards is the number of bacteria present.

6.3.1 Plate Colony Count

Numbers may be determined by one or more of several methods, legally recognized procedures for which are detailed in literature cited. One of these, the plate (colony) count, is closely analogous to the plate count procedure.

For counting bacteria in milk the plates may be incubated at 32° or 35°C. for 48 hours for routine work, or at 7°C. for 7 to 10 days to enumerate psychrophiles; at 55°C. for 48 hours to enumerate thermophiles. The plating medium officially recognized for all of these, as well as for enumerating the bacteria in drinking water and widely used for counting bacteria in other materials (e.g., foods) contains: agar 1.5 per cent; yeast extract, 2.5 per cent; protein digest, 5.0 per cent; glucose, 1 per cent. This medium is generally called *Standard Methods Agar* and must conform in bacteriological quality to that of a standard lot of medium specified by the American Public Health Association in 1963.

6.3.2 Direct (Breed) Count

Bacteria may also be enumerated by the direct microscopic examination of milk in a smear. The smear is prepared by spreading exactly 0.01 ml. of the sample on a slide over an area of exactly

one square centimeter (100 mm.^{2}). After staining with specially prepared methylene blue solutions designed to remove fat globules, the smear is examined by means of a microscopy calibrated with a stage micrometer so that the field is exactly 0.206 mm. in diameter. Such a field represents approximately 1/300,000 ml. of milk. The number of bacteria seen in it must therefore be multiplied by the reciprocal of this fraction (the *microscopic factor* or MF) to determine the numbers of bacteria per milliliter:

$$MF = \frac{10{,}000}{\pi r^2}$$

Numerous large *clumps* of bacteria indicate unclean utensils. Many *pus cells* indicate infected udders. Streptococci and staphylococci indicate mastitis. It is common practice to report results in terms of *clump counts* and, in reporting, to translate the microscopic observations directly into such terms as "good," "bad" and "mastitis." The method is quick and inexpensive. It is applied mainly to unpasteurized milk, since heating causes many of the bacteria to die and lose their staining properties.

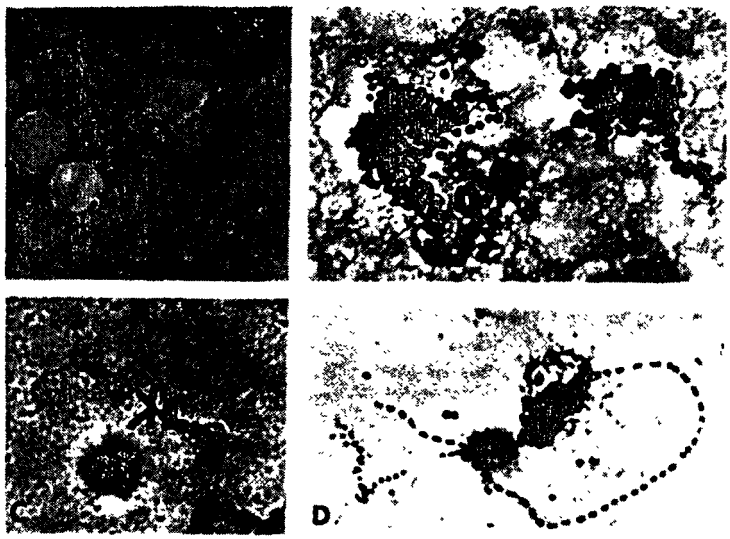


Figure. 6.7 : Examples and probable sources of microorganisms detectable by direct microscopic examination of milk: *A* cocci, milk probably improperly refrigerated; *B*, large masses of bacteria , improperly cleaned milk-handling utensils; *C*, spore-forming bacilli, probably from soil and dust contamination; *D*, pyogenic streptococci and a pus cell from cow with mastitis.

6.3.4 Numerical Relationship Between Counts

The *total* direct microscopic count is usually five to ten or more times as high as the plate count. This is because the total direct count enumerates individual cells, even those in clumps, and also dead bacteria. As in the plate count applied to water and soil, the plate count applied to milk enumerates only live bacteria capable of developing in the medium and under the environmental condition used. In the plate count each clump, even though it may contain scores of live cells, counts only as a single bacterium since each clump forms only one colony. The plate count is usually closer to the clump count than to the total direct count.

6.4 QUALITY AND OXIDATION-REDUCTION POTENTIAL

6.4.1 The Reduction Test

Most actively growing bacteria case a lowered oxidation-reduction (O-R) potential in their medium. The presence in milk of large enough numbers of growing bacteria to produce a significantly lowered O-R potential can be detected by the use of methylene blue because, with a lowering of the O-R potential to about -0.01 volt, this dye becomes colourless. The reduction (so-called "*reductase*") test is used principally with raw milk, and furnishes a rough but useful approximation of the number and kinds of living bacteria present. In performing the test, 10 ml. of milk sample are pipetted into a sterile tube and 1 ml. of a standard methylene blue solution (final concentration about 1 : 250,000) is added. The tube is closed with a rubber stopper and slowly inverted three times to mix. It is placed at 35.0°C. in the water bath imme-diately. At the end of each hour during the test the tube is inverted once. Observations are made after 30 minutes, 1 hour, and later.

6.4.2 Mbrt

The *methylene-blue reduction time* (MBRT) is the interval between the placing or the tubes in the water bath at 35.0° C. and the disappearance of the blue colour from the milk.

The shorter the MBRT, the greater the number of active bacteria in the milk and the lower its bacteriological quality. However, MBRT does not give an accurate *count* of bacteria present; only an overall measure of its bacteriological quality. Milk with an MBRT of six hours is very good.

6.4.3 Resazuri

A dye related to methylene blue, resazurin undergoes a *series* of colour changes depending on O-R potential changes, where as methylene blue changes only from blue to colou-rless. Resazurin, therefore, permits readings of *degrees* of reduction at shorter intervals than does methylene blue. In one procedure (the *1-hour test*), colour of the milk, initially blue, is compared after one hour at 37°C. with several exactly described (*Munsell*) colour standards designated as 5P7/4 (mauve), PBP7/ 5.5 (purple), and so on. Milk showing no grea-ter change during the hour than from blue to OBO7/5.5 is grade 1; to PRP7/8 (lavender), grade 2; and so on to grade 4, complete decolourization. In the *triple reading* or *three-hour* test, three successive readings are made at one-hour intervals to see how long it takes to reach the colour 5P7/4. High grade (acceptable) milk requires at least three hours (resazurin reduction time, or RRT, three hours).

6.5 FACTORS AFFECTING BACTERIOLOGICAL QUALITY

6.5.1 Bacteriophages in Milk

A factor of importance, especially to manufacturers of dairy products that depend on early, rapid and luxuriant growth of certain bacteria such as *Streptococcus lactis*, is the presence of bacteriophage lytic for that species. These phages are common in dust in and around dairies. They are known to interfere with many sorts of dairy work (cheese-making) that is dependent on bacterial growth. Very rigid aseptic technique in preparing and handling the pure *starter* cultures is necessary to eliminate the phages. Steel filters with triple layers of fiberglass have been successfully used to remove phages from air of dairy laboratories and work rooms.

6.5.2 Antibiotics and Disinfectants

Another important factor in bacteriological studies of milk is the possibility that the results have been influenced by (1) preservatives illegally added; (2) residues of disinfectants used to sanitize the dairy equipment; (3) antibiotics or other drugs used to control udder infections, and otherwise admin-istered to cattle. Antibiotics, disinfectants and preservatives can interfere with growth of the bacteria used in the manufac-ture of cheese and butter, cultured milks (e.g., yoghurt), and is such bacteriological controls as reduction tests and plate counts. In most manufacturing procedures, uninhibited growth of *S. cremoris, Leuconostoc citrovorum, Streptococcus lactis,*

Lactobcillus species and other bacteria is essential.

Penicillin is especially undesirable in milk because it can produce severe allergic reactions in persons hypersensitive to it.

6.6 GRADES OF MILK

The actual numbers of bacteria permissible in milks of various grades vary in different cities. A good guide is the standard milk ordinance and code of the Public Health Service of the U. S. Department of Health, Education and Welfare. Various localities may have some what different standards.

6.6.1 Grade A Raw Milk for Pasteurization

Grade A raw milk for pasteurization is raw milk from properly supervised producer dairies conforming to standards of sanitation of workers, cattle, premises and equipment as prescribed in the ordinance. Cattle and personnel must be free from diseases transmissible in milk. The bacterial plate count or the microscopic clump count of the milk, as delivered to the pasteurizing plant for pasteurization, shall not exceed 200,000 organisms per milliliter as determined by standard methods of the American Public Health Association or must have a MBRT of less than 5.5 hours or RRT of less than 2.75 hours.

6.6.2 Grade A Pasteurized Milk

In all cases Grade A pasteurized milk shall show efficient pasteurization as evidenced by satisfactory phosphatase test, and at no time after pasteurization and before delivery shall the milk have as bacterial plate count exceeding 30,000 per milliliter, or a coliform count exceeding 10 per milliliter, as determined by Standard Methods of the American Public Health Association.

6.6.3 Grade B Pasteurized Milk

Grade B pasteurized milk is pasteurized milk that does not meet the bacterial-count standard for grade A pasteurized milk, and certain other sanitary requirements. Such milk may be used in some commercial processes.

Most communities now permit the sale only of grade A milk, pasteurized.

6.6.4 Criteria of Good Milk

Quality tests such as bacterial enumerations and reduction tests, as well as others, have a distinct value and usefulness from the standpoint of *cleanliness* but not neccessarily with regard to *infection.*

The present status of microbiology of milk was well summarized by Robertson, who said.

None of the routine laboratory procedures for estimating the number of bacteria in milk will determine whether or not infectious bacteria are present. The best assurance of freedom from infectious bacteria is that provided by proper pasteurization of the milk. The best assurance of pasteur-ization is that demonstrated by a satisfactory phosphatase test on the bottled pasteurized milk. The best assurance of freedom from recontamination in freshly bottled milk after pasteurization is a satisfactory coliform test in 1.0 ml. portions of the bottled product.

Certified milk

If milk is to be offered for sale unpasteurized, it is often required that it be produced only under very carefully supervised conditions. The American Association of Medical Milk Commissions has established rules and regulations concerning veerinary inspection of cows, especially for tuberculosis, brucellosis and infectious mastitis, and sanitation of barns and utensils. These very rigid regulations are often used by health departments and milk dealers in certifying qualified farms to produce such milk. It is usually called *certified milk* or *baby milk*. The use of Certified milk has much to recommend it, especially its cleanliness. The coliform standard for certified milk–raw, is 10 per milliliter; for certified, pasteurized, 1 per milliliter. It is said to contain a larger proportion of certain vitamins essential for infacts than milk which has been heated.

Most cities and states, as well as the A.A.M.M.C., require that all persons occupied in preparing Certified milk, or, indeed, and food for the public, be examined periodically for typhoid, paratyphoid and dysentery bacilli. Examinations for presence of organisms that cause diphtheria, tuberculosis, scarlet fever and other transmissible diseases are also required for Certified milk handlers.

6.7 CONCENTRATED AND DRIED MILKS

Milks from which part or all of the water has been withdr-awn are termed *concentrated* or *dried*. Assuming that appro-ved standards of cleanliness, freshness, sanitation and chem-ical content (fat, solids and so on) have been met in selecting the milk to be dehydrated, the microbiological quality of the finished product is determined largely by: (a) temperatures and time of storage (it any) prior to processing, (b) times and temperatures of processing; (c) cleanliness of the apparatus and final containers, (d) time and temperature of final proces-sing of canned milks.

6.7.1 Evaporated Milk

The raw milk is first cooled and may be clarified. Fat and solids contents are "adjusted" to meet required standards. The milk is then heated to boiling or nearly so (94 to 100°C.) for about 20 minutes. This kills all but the most heat-resist-ant microorganims. Water is driven off *in vacuo* at about 55°C., a temperature which favours development of undesirable thermophiles, e.g., *Bacillus stearothermophilus* and which therefore requires careful bacteriological control.

After homogenization the product is cooled and canned. It is "commercially sterilized" at about 117°C. for about 15 minutes. Spoilage problems can result from inadequate heat-ing and subsequent storage at unduly high temperatures that favour growth of spores of thermophilic molds and Bacillaceae.

6.7.2 Sweetened Condensed Milk

Milk can be doubly preserved by the addition of about 20 per cent of sucrose or glucose, or both, to whole milk, and by subsequent heating at temperatures near boiling and by partial dehydration as with evaporated milk.

6.7.3 Nonfat Dried Milk

Nonfat dried milk is prepared by preliminary steps similar to those used for evaporated milk. The fat is removed by high-speed cream separators. The water is partly removed by preliminary heating (about 85°C.) and then by: (a) spraying the milk as a mist into a current of hot (about 120°C.) air in a closed chamber; or (b) by spreading the milk as a thin film on hot rollers or drums—at about 145°C. if no vacuum is used, or at less than 100°C. in a vacuum chamber.

The dried *flakes* from hot drums or the *powder* from sprayed milk are packaged to prevent access of moisture. Dried milks should be kept dry and cool at all times or they may spoil since they are not sterile. As previously pointed out, dry heat is an inefficient sterilant. Microorganisms surviving the process are mainly Bacillaceae and thermo-duric streptococci, lactobacilli, micrococci and species of *Microbacterium.* Nonsporing pathogens are eliminated. The presence of coliform bacteria or of pathogenic streptococci or other heat-sensitive microorganisms is of the same sinister significance that it is in pasteurized whole milk.

6.7.4 Reconstitution

In reconstituting dried or evaporated milk it is clearly desirable to use clean, cool and hygienically acceptable water and utensils. The

reconstituted milk should be kept and handled under the same conditions of sanitation and refrige-ration as are recommended for whole milk.

6.7.5 Standards of Quality

Quality standards for dewatered milks are similar to those for whole milk. bearing in mind the changes in flora caused by heating. The standards are established by the U. S. Department of Agriculture, The American Dry Milk Institute and the Evaporated Milk Association (Evaporated Milk Industry Sanitary Standards Code, Chicago, Ill.) Standards for various grades of dried milk, for example, are 50,000 per gram for "extra" quality and 100,000 per gram for "standard" quality. Counting procedures are the same as for fresh milk after the dried milk is reconstituted with sterile water.

6.8 SOME MANUFACTURED DAIRY PRODUCTS

Market milk contains numerous species of microorganisms in varying numbers whose uncontrolled action is too unrelia-ble to serve as a basis for commercial operations requiring uniformity of product. Pure-culture inocula (called *mother* or *stock cultures*) of constant properties are essential to continued success in this highly competitive field. For these purpose *Lyophilized* ("freeze-dried"), or other pure, stock cultures of desired organisms may be maintained in the dairy if a competent bacteriologist and adequate laboratory facilities are available. Otherwise it is best to obtain stock cultures from dairysupply houses. Cultures of species of lactic streptococci, *Leuconostoc* and *Lactobacillus* are used especially.

6.8.1 Starter Cultures

In practice mother cultures in about 2 per cent volume are added as nearly aseptically as possible to about 600 ml. of sterile or very-low-count milk (previously heated 30 minutes at about 88°C. and cooled to 21°C.), and incubated. The lactic organisms soon outgrow other species, of any are present. This culture is called a *starter*. It may be used to inoculate a tank-size batch of milk or cream for butter or cheese, or to inoculate a still larger lot of starter.

Fresh, high-quality (low-count) pasteurized milk is brought *quickly* to the desired incubating temperatures. A large, *pre-emptive,* virtually pure, starter inoculum of vigorously growing young cells of the desired lactic organism is added and thoroughly mixed with the milk or cream. Before the other bacteria in the milk have time to recover from their previous refrigeration or pasteurization and overcome their lag phase, the acidity quickly produced by the actively metab-olizing

added lactic starter suppresses them.

6.8.2 Butter

Butter is generally made by churning cream that has been soured by lactic acid bacteria.

Two species of bacteria, each with a distinct function, are added together to the cream simultaneously. *Leuconostoc citrovorum* is depended on for flavour, *Streptococcus cremoris* or *S. lactis* is selected primarily for *rapid*, initial lactic acid production. If high acidity (pH 4.3) is not produced promptly, numerous undesirable contaminants may grow excessively. When the pH reaches about 4.3 *Leuconostoc* ceases growth, but its enzymes attack the citrates in the milk and produce diacetyl. This substance gives butter and similar products their characteristic buttery flavour and aroma. Neither *S. cremoris* nor *Leuconostoc* alone can produce the desired result in commercial practice.

6.8.3 Cheese

Cheeses may be divided into three general types: (a) soft- or cottage-type cheese, and cream cheese (these are eaten in a fresh or unripened state); (b) hard- or rennet-curd cheese, including Roquefort, American cheddar-type ("rat-bait"), Edam, and Swiss (these are *ripened* by the enzymes and slow growth of bacteria or molds or both, which cause some, but not extensive proteolysis); (c) soft or semisoft rennetcurd cheese, of which Camembert, Limburger and Liederkranz are types (these are ripened by proteolytic and lipolytic organisms which soften the curd and give it flavours). The hardness of cheese depends to some extent on moisture and fat content as well as on heating and acidity of the curd, draining, salting and conditions of storage.

6.8.4 Soft, Acid-curd Cheese

In making cottage cheese, starters containing mixtures of *Leuconostoc citrovorum, L. dextranicum, S. lactis* and the like are added to pasteurized milk. These ferment the lactose. *Leuconostoc* adding flavour and aroma. The lactic acid coagulates the casenin. Rennet may be added to hasten the coagulation and make the curd firmer. The curd is cut into small cubes. To firm the curd and separate it from the whey, the mass is heated slowly to about 50° C. and held so for 30 minutes. Water is added; the curd settles. The water, with the whey, is drained off and the curd is pushed into heaps to drain. It is washed a second time with water and drained. About 0.5 per cent salt is added. Just before packaging

TABLE 6.1: TYPES OF NATURAL CHEESE

Representative cheeses	Distinctive Organisms in Ripening Flora
Soft:	
Cottage	
Cream	not ripened
Liederkranz	*Strep. liquifaciens, Brevibacterium*
Camembert	*Penicillium camemberti, Brevibacterium*
Semisoft:	
Blue (or Bleu)	
Roqueforet	*Penicillium* strains such as
Gorgonzola	*P. roqueforti*
Hard	
Swiss	*Propionibacter* species
Cheddar	Lactic group, *Geotrichum*
Parmesan	Lactic group (brine cured)

Not processed artificially. (Pasteurized and processed cheeses and cheese spreads are not included since they are made almost entirely from the natural cheeses such as those listed above.)

many manufacturers add a little cream.

6.8.5 Hard-Curd Cheese

In the preliminary stages, nearly all natural (i.e., not "processed") cheese are much alike. Differences result from different methods of rating the curd: degree of acidity, addi-tion of different amounts of salt, special ripening microor-ganisms, moisture, temperature and humidity of ripening and other factors.

For yellow cheeses of the Cheddar type, colour is added. After a slight acid it has developed, rennet is added to make an elastic, rubbery curd which is later cut into pieces about one inch in diameter and warmed to about 35°C. The curd becomes firmer and the whey separates and is drained of and may be used for stock feed. The clumped masses of firm curd are chopped (*milled*) again, and piled up to press out whey. This is called *cheddaring* in Cheddar cheese making. The curd is again milled, and then is salted, drained and pressed in hoops to cure. Curing of Cheddar cheese proceeds at about 15°C. It becomes "sharper" with aging.

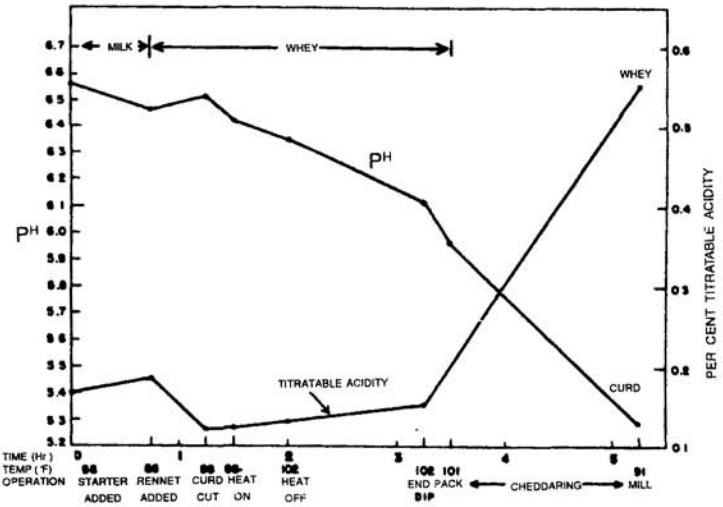


Figure 6.8 : Cheddar cheese: changes in temperature and in acidity as measured by determination of pH and by titration of total acid, during successive early phases, from addition of starter (flow left) to milling (lower right); about is fermented acidity quickly increases. At pH about 6.46, rennet is added and the curd forms. It is soon cut, heated, and the whey is drawn off (dipped). Cheddaring and milling then proceed. Further increases in acidity occur during the ripening process.

6.8.6 Curing or Ripening of Hard Cheeses

During the curing process of Cheddar and other hard cheeses, various microorganisms, the varieties depending on the kind of cheese, continue a slow fermentative, lipolytic and proteolytic action, the products of these processes yielding the characteristic flavours, textures and aromas of various cheese. Prominent among these flavours are diacetyl, lactic, butyric, caproic and acetic acids, and various amines, as well as various esters such as those that give flavours to ripe fruit juices. In addition, since many of these organisms synthesize vitamins, especially nicotinic acid and vitamins of the B complex, the nutritive value of the cured (or ripened) cheeses is increased.

Gas formers such as species of *Clostridium* and Escherichiae are undesirable because they produce gassy cheeses and off flavours; they may be especially active in the early stages. They generally occur in milk of poor quality.

Swiss cheese is heated to 50°C. after cutting the curd. Starters therefore usually contain the *thermoduric* lactics: *Streptococcus*

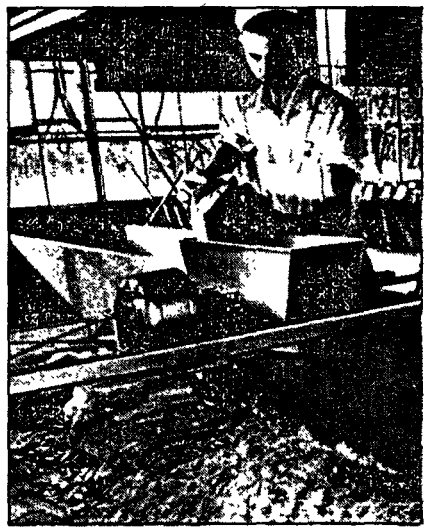


Figure. 6.9 : Making cheddar-type cheese Liquid milk becomes curd and whey after starter and rennet are added. The milk forms into soft curd much like "junket." This is firmed by heating gently. When it has reached the correct firmness, it is cut in small cubes (non-fourth inch) by special wire knives. The whey are heated while slowly agitated by revolving paddles. The cubes of coagulated casein are then washed with water and drained The picture shows the cutting seperation

thermophilis and *Lactobacillus bulgaricus* or *L. lactis,* as well as *Propionibacterium shermanii* and *P. freudenreichii.*

The cheese is soaked in 23 per cent brine for some days at 13° C. Propionibacteria are then favoured by incubation at 22°C. Later the cheese is ripened for months at about 13°C.

The "eyes" in Swiss cheese are a result of the production of carbon dioxide by species of *Propionibacterium*, while its bitter-sweet flavour is caused in part by the formation of glycerol, propionic and succinic acids by *Propionibacterium* species while ripening.

Semisoft cheese, such as Roquefort, Gorgonzola or Blue (or *Bleu*), contains much fat and as high as 5 per cent salt and relatively little moisture. The high salt content prevents continued growth of most bacteria, as does the low ripening temperature (7° to 8°C.) and humidity (60 per cent). A species of mold (*Penicillium roqueforti*) is introduced by the inoculation of spores into the milk or into the curd as it is put into hoops for ripening. The mold grows in the interior, producing the masses of blue green conidia and the sharp flavour so characteristic of this type of cheese. As the mold is aerobic, perforations are made in the cheese to aerate the interior.

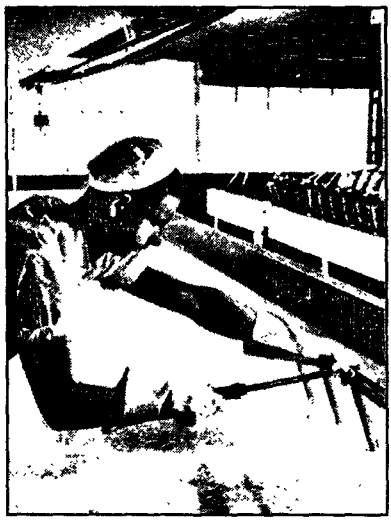


Figure 6.10 : Making cheddar-type cheese. When cheddaring (firming and draining) is complete (judged by the firmness of the curd, its texture and acidity) the curd is milled (cut into small pieces) in preparation for salting and pressing. About one and one-half pounds of salt are used to 100 pounds of curd. After salting, the curd is placed in cheese clother cellophane-lined "hoops" of the desired size and shape for pressing.

6.8.7 Soft Cheeses

Limburger, Liederkranz and Camembert cheeses are cured mainly by the growth of organisms in a red-orange, slimy coating on the outer surface. Numbers of microorga-nisms in this slime sometimes exceed 10 billion per gram. In Limburger cheese, yeasts (*Geotrichum* species) begin to grow on the surace after subsidence of the initial acidity caused by lactic organisms. They persist in the surface slime for about a week. *Brevibacterium linens* and *B. erythrogenes* then grow all over the surface, forming a reddish brown coating, or "*smear*," commonly seen on soft and semisoft cheese (Brick, Camembert). Camembert cheese is sometimes inoculated on the outer surface (or before curd-formation) with a pure culture of the mold. *P. camemberti.* The enzymes of the various micrococci, yeasts or molds in the slime penetrate into the interior of the cheese, producing the flavours, softening, and the famous aroma of Limburger and similar cheese.

In certain countries lactobcilli have been used for centuries in combination with certain yeasts and streptococci to produce foods of fermented milk. The *yoghurt* of eastern central Europe (now available in all grocery stors), the *busa* of Turkestan, the *kefir* of the Cossacks,

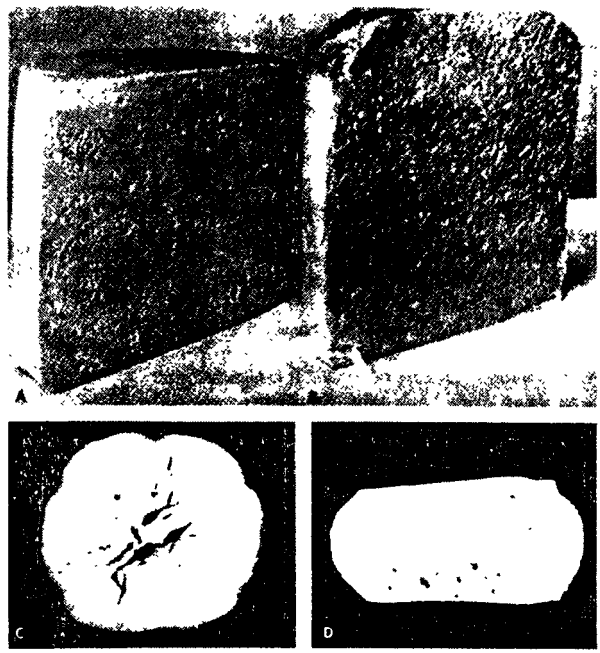


Figure 6.11 : Defects in cheese due to gas-forming (aerogenic) microorganisms. *A* and *B, Gas* formed by Klebsiella aerogenes in early ripening of Cheddar cheeses (*A*, yellow; *B*, white). (Courtesy of The Borden Co.) *C, Gas* formed late in ripening of Provolone, and Italian cheese similar to Cheddar. Compare with "stormy fermentation," Figure. 32 7. *D*, Gas formed by lactose fermenting yeast in early ripening of Brick cheese. (*C* and *D* from Foster et al., Dairy Microbiology, Prentice Hall, Inc.)

the *koumiss* of central Asia and the *leben* of Egypt are examples of these. Formerly, the microbial nature of these processes was unknown. In all of these fermented milks lactobacilli act in company with other microorganisms: yeasts, lactic streptococci and various rods. For example, kefir, made from milk of various domestic animals, is prepared by putting *kefir grains* (small, cauliflower-like masses) into the milk. These grains consist of dried masses of lactobacilli (*L. brevis*), yeasts (*Sacch. delbruckii*), *Streptococcus lactis* and probably other lactic organisms held together in a matrix of coagulated casein and bacterial polysaccharide gum. The kefir grains increase in size and break apart as the fermentation proceeds. The combined growth of the mixed flora yields a characteristically flavoured, soured milk containing small amounts of alcohol. The kefir gains are found in the bottom of the vessels of fermented milk.

6.8.8 Yoghurt

In the United States this is made of pasteurized milk, and soured by species of *Lactobacillus*.

6.8.9 Acidophilus Milk

Milk sourced by *Lactobacillus acidophilus* is thought by some to have medicinal properties in the intestinal tract. *Bulgarian buttermilk*. a similar beverage, is prepared with pure-culture starters of *Lactobacillus bulgaricus*.

6.8.10 Buttermilk

Much of the product commonly sold in the United States as buttermilk is in reality pasteurized skim or whole milk soured mainly by *Streptococcus cremoris* with *Leuconostoc citrovorum* and then beaten so as to produce a smooth, creamy beverage. It is a pleasant, nourishing drink. Addition of 0.15 per cent of citric acid to the milk results in formation of increased flavour because diacetyl is produced by *Leucono-stoc citrovorum*.

7

Allergy and Hypersensitivity

IgE is produced by plasma cells located in lymph nodes draining the site of antigen entry or locally, at the sites of allergic reactions, by plasma cells derived from germinal centers developing within the inflamed tissue. IgE differs from other antibody isotypes in being located predominantly in tissues, where it is tightly bound to the mast-cell surface through the high-affinity IgE receptor known as FcRI. Binding of antigen to IgE cross-links these receptors and this causes the release of chemical mediators from the mast cells, which may lead to the development of a type I hypersensitivity reaction. Basophils and activated eosinophils also express FcRI; they can therefore display surface-bound IgE and also take part in the production of type I hypersensitivity reactions. The factors that lead to an antibody response dominated by IgE are still being worked out. Here we will describe our current understanding of these processes before turning to the question of how IgE mediates allergic reactions.

7.1 ALLERGENS

There are certain antigens and routes of antigen presentation to the immune system that favor the production of IgE. CD4 T_H2 cells can switch the antibody isotype from IgM to IgE, or they can cause switching to IgG2 and IgG4 (human) or IgG1 and IgG3 (mouse). Antigens that selectively evoke T_H2 cells that drive an IgE response are known as allergens.

Much human allergy is caused by a limited number of inhaled small-protein allergens that reproducibly elicit IgE production in susceptible individuals. We inhale many different proteins that do not

induce IgE production; this raises the question of what is unusual about the proteins that are common allergens. Although we do not yet have a complete answer, some general principles have emerged. Most allergens are relatively small, highly soluble proteins that are carried on desiccated particles such as pollen grains or mite feces. On contact with the mucosa of the airways, for example, the soluble allergen elutes from the particle and diffuses into the mucosa. Allergens are typically presented to the immune system at very low doses. It has been estimated that the maximum exposure of a person to the common pollen allergens in ragweed (*Artemisia artemisiifolia*) does not exceed 1 mðg per year! Yet many people develop irritating and even life-threatening T_H2-driven IgE antibody responses to these minute doses of allergen. It is important to note that only some of the people who are exposed to these substances make IgE antibodies against them.

It seems likely that presenting an antigen transmucosally and at very low doses is a particularly efficient way of inducing T_H2-driven IgE responses. IgE antibody production requires T_H2 cells that produce interleukin-4 (IL-4) and IL-13 and it can be inhibited by T_H1 cells that produce interferon-gð (IFN-gð). The presentation of low doses of antigen can favor the activation of T_H2 cells over T_H1 cells, and many common allergens are delivered to the respiratory mucosa by inhalation of a low dose. The dominant antigen-presenting cells in the respiratory mucosa are myeloid dendritic cells. These take up and process protein antigens very efficiently and become activated in the process. This in turn induces their migration to regional lymph nodes and differentiation into professional antigen-presenting cells with co-stimulatory activity that favors the differentiation of T_H2 cells.

7.1.1 Role of Enzymes in Allergy

Several lines of evidence suggest that IgE is important in host defense against parasites. Many parasites invade their hosts by secreting proteolytic enzymes that break down connective tissue and allow the parasite access to host tissues, and it has been proposed that these enzymes are particularly active at promoting T_H2 responses. This idea receives some support from the many examples of allergens that are enzymes.

The major allergen in the feces of the house dust mite (*Dermatophagoides pteronyssimus*), which is responsible for allergy in approximately 20% of the North American population, is a cysteine protease homologous to papain, known as Der p 1. This enzyme has been found to cleave occludin, a protein component of intercellular tight

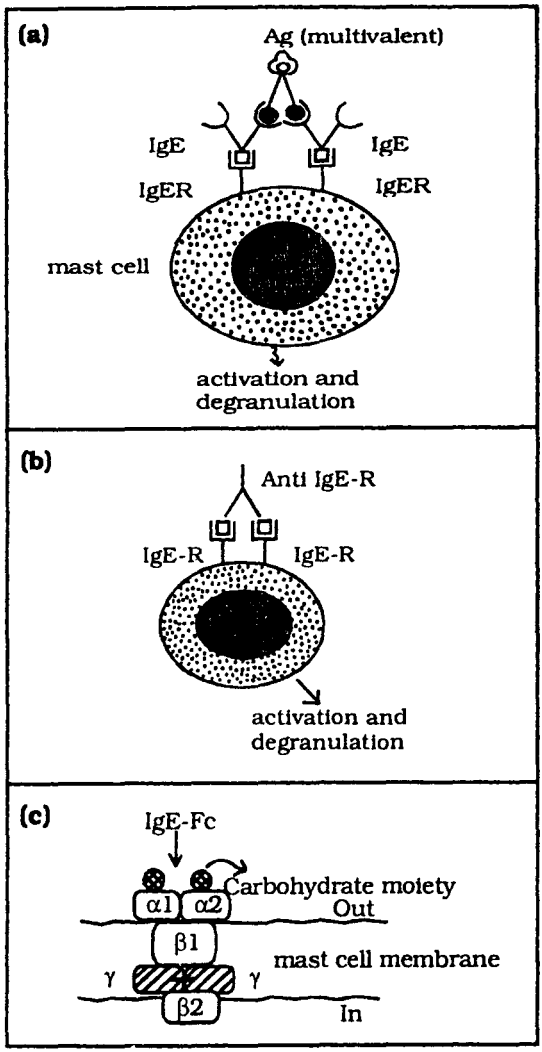


Figure 7.1 : Immunological triggering of mast cell activation and degranulation. (a) IgE-Fc binds to IgE-R and IgE-R are crosslinked by antigen binding (b) Anti IgE-R Abs also do a similar thing (c) The native hexameric IgE-R assembly.

junctions. This reveals one possible reason for the allergenicity of certain enzymes. By destroying the integrity of the tight junctions between epithelial cells, Der p 1 may gain abnormal access to subepithelial antigen-presenting cells, resident mast cells, and eosinophils.

The allergenicity of Der p 1 may also be promoted by its proteolytic

action on certain receptor proteins on B cells and T cells. It has been shown to cleave the subunit of the IL-2 receptor, CD25, from T cells. Loss of IL-2 receptor activity might interfere with the maintenance of T_H1 cells, leading to a T_H2 bias.

The protease papain, derived from the papaya fruit, is used as a meat tenderizer and causes allergy in workers preparing the enzyme; such allergies are called occupational allergies. Another occupational allergy is the asthma caused by inhalation of the bacterial enzyme subtilisin, the 'biological' component of some laundry detergents. Injection of enzymatically active papain (but not of inactivated papain) into mice stimulates an IgE response. A closely related enzyme, chymopapain, is used medically to destroy intervertebral discs in patients with sciatica; the major (although rare) complication of this procedure is anaphylaxis, an acute systemic response to allergens.

Not all allergens are enzymes, however; for example, two allergens identified from filarial worms are enzyme inhibitors. Many protein allergens derived from plants have been identified and sequenced, but their functions are currently obscure. Thus, there seems to be no systematic association between enzymatic activity and allergenicity.

7.1.2 B Lymphocytes in Specific Signals

There are two main components of the immune response leading to IgE production. The first consists of the signals that favor the differentiation of naive T_H0 cells to a T_H2 phenotype. The second comprises the action of cytokines and co-stimulatory signals from T_H2 cells that stimulate B cells to switch to producing IgE antibodies.

The fate of a naive CD4 T cell responding to a peptide presented by a dendritic cell is determined by the cytokines it is exposed to before and during this response, and by the intrinsic properties of the antigen, antigen dose, and route of presentation. Exposure to IL-4 favors the development of T_H2 cells and to IL-12 favors that of T_H1 cells. IgE antibodies are important in host defense against parasitic infections and this defense system is distributed anatomically mainly at the sites of entry of parasites under the skin, under the epithelial surfaces of the airways (the mucosal-associated lymphoid tissues), and in the submucosa of the gut (the gut-associated lymphoid tissues). Cells of the innate and adaptive immune systems at these sites are specialized to secrete predominantly cytokines that drive T_H2 responses. The dendritic cells at these sites are of the myeloid phenotype; after taking up antigen they migrate to regional lymph nodes where their interaction with naive CD4 T cells drives the T cells to become T_H2 cells, which secrete IL-

4 and IL-10. It is not known how myeloid dendritic cells induce this differentiation. One possibility is that they express a particular set of cytokines and co-stimulatory molecules yet to be characterized. Another is that they activate a specialized subset of CD4 T cells, the $NK1.1^+$ subset, that produce abundant IL-4 that can induce CD4 T cells to differentiate into T_H2 cells following stimulation by antigen. These in turn induce B cells to produce IgE.

Class switching of B cells to IgE production is induced by two separate signals, both of which can be provided by T_H2 cells. The first of these signals is provided by the cytokines IL-4 or IL-13, interacting with receptors on the B-cell surface. These transduce their signal by activation of the Janus family tyrosine kinases JAK1 and JAK3 which ultimately lead to phosphorylation of the transcriptional regulator STAT6. Mice lacking functional IL-4, IL-13, or STAT6 all show impaired T_H2 responses and IgE switching, demonstrating the key importance of these signaling pathways. The second signal for IgE class switching is a co-stimulatory interaction between CD40 ligand on the T-cell surface with CD40 on the B-cell surface. This interaction is essential for all antibody class switching (see Section 9-3); patients with the X-linked hyper IgM syndrome have a deficiency of CD40 ligand and produce no IgG, IgA, or IgE.

The IgE response, once initiated, can be further amplified by basophils, mast cells, and eosinophils, which can also drive IgE production. All three cell types express FcRI, although eosinophils only express it when activated. When these specialized granulocytes are activated by antigen cross-linking of their FcRI-bound IgE, they can express cell-surface CD40L and secrete IL-4; like T_H2 cells, therefore, they can drive class switching and IgE production by B cells. The interaction between these specialized granulocytes and B cells can occur at the site of the allergic reaction, as B cells are observed to form germinal centers at inflammatory foci. Blocking this amplification process is a goal of therapy, as allergic reactions can otherwise become self sustaining.

7.2 GENETIC FACTORS

As many as 40% of people in Western populations show an exaggerated tendency to mount IgE responses to a wide variety of common environmental allergens. This state is called *atopy* and seems to be influenced by several genetic loci. *Atopic* individuals have higher total levels of IgE in the circulation and higher levels of eosinophils than their normal counterparts. They are more susceptible to allergic

diseases such as hay fever and asthma. Studies of atopic families have identified regions on chromosomes 11q and 5q that appear to be important in determining atopy; candidate genes that could affect IgE responses are present in these regions. The candidate gene on chromosome 11 encodes the bð subunit of the high-affinity IgE receptor, whereas on chromosome 5 there is a cluster of tightly linked genes that includes those for IL-3, IL-4, IL-5, IL-9, IL-12, IL-13, and granulocyte-macrophage colony-stimulating factor (GM-CSF). These cytokines are important in IgE isotype switching, eosinophil survival, and mast-cell proliferation. Of particular note, an inherited genetic variation in the promoter region of the IL-4 gene is associated with raised IgE levels in atopic individuals; the variant promoter will direct increased expression of a reporter gene in experimental systems. Atopy has also been associated with a gain-of-function mutation of the að subunit of the IL-4 receptor, which is associated with increased signaling following ligation of the receptor. It is too early to know how important these different polymorphisms are in the complex genetics of atopy.

A second type of inherited variation in IgE responses is linked to the MHC class II region and affects responses to specific allergens. Many studies have shown that IgE production in response to particular allergens is associated with certain HLA class II alleles, implying that particular MHC:peptide combinations might favor a strong T_H2 response. For example, IgE responses to several ragweed pollen allergens are associated with haplotypes containing the MHC class II allele *DRB1*1501*. Many individuals are therefore generally predisposed to make T_H2 responses and specifically predisposed to respond to some allergens more than others. However, allergies to common drugs such as penicillin show no association with MHC class II or the presence or absence of atopy.

There is evidence that a state of atopy, and the associated susceptibility to asthma, rhinitis, and eczema, can be determined by different genes in different populations. Genetic associations found in one group of patients have frequently not been confirmed in patients of different ethnic origins. There are also likely to be genes that affect only particular aspects of allergic disease. For example, in asthma there is evidence for different genes affecting at least three aspects of the disease phenotype IgE production, the inflammatory response, and clinical responses to particular types of treatment.

The prevalence of atopic allergy, and of asthma in particular, is increasing in economically advanced regions of the world, an observation

that is best explained by environmental factors. The four main candidate environmental factors are changes in exposure to infectious diseases in early childhood, environmental pollution, allergen levels, and dietary changes. Alterations in exposure to microbial pathogens is the most plausible explanation at present for the increase in atopic allergy. Atopy is negatively associated with a history of infection with measles or hepatitis A virus, and with positive tuberculin skin tests (suggesting prior exposure and immune response to *Mycobacterium tuberculosis*). In contrast, there is evidence that children who have had attacks of bronchiolitis associated with respiratory syncytial virus (RSV) infection are more prone to the later development of asthma. Children hospitalized with this disease have a skewed ratio of cytokine production away from IFN-gð towards IL-4, the cytokine that induces T_H2 responses. It is possible that infection by an organism that evokes a T_H1 immune response early in life might reduce the likelihood of T_H2 responses later in life and vice versa. It might be expected that exposure to environmental pollution would worsen the expression of atopy and asthma. The best evidence shows the opposite effect, however. Children from the city of Halle in the former East Germany, which has severe air pollution, had a lower prevalence of atopy and asthma than an ethnically matched population from Munich, exposed to much cleaner air. This does not mean that polluted air is not bad for the lungs.

While it is clear that allergy is related to allergen exposure, there is no evidence that the rising prevalence of allergy is due to any systematic change in allergen exposure. Nor is there any evidence that changes in diet can explain the increase in allergy in economically advanced populations.

8

Immunological Tolerance

Immunological Tolerance or simply tolerance to self is acquired by clonal deletion or inactivation of developing lymphocytes. Tolerance to antigens expressed by grafted tissues can be induced artificially, but it is very difficult to establish once a full repertoire of functional B and T lymphocytes has been produced, which occurs during fetal life in humans and around the time of birth in mice. We have already discussed the two important mechanisms of self-tolerance clonal deletion by ubiquitous self antigens and clonal inactivation by tissue-specific antigens presented in the absence of co-stimulatory signals. These processes were first discovered by studying tolerance to nonself, where the absence of tolerance could be studied in the form of graft rejection. In this section, we will consider tolerance to self and tolerance to nonself as two aspects of the same basic mechanisms. These mechanisms consist of direct induction of tolerance in the periphery, either by deletion or by anergy. There is also a state referred to as immunological ignorance, in which T cells or B cells coexist with antigen without being affected by it. Finally, there are mechanisms of tolerance that involve T-cell-T-cell interactions, known variously as immune deviation or immune suppression. In an attempt to understand the related phenomena of autoimmunity and graft rejection, we also examine instances where tolerance to self is lost.

8.1 RECOGNITION OF AUTOANTIGENS

Clonal deletion removes immature T cells that recognize ubiquitous self antigens and in Chapter 8 that antigens expressed abundantly in the periphery induce anergy or clonal deletion in lymphocytes that encounter them on tissue cells. Most self proteins are expressed at levels that are

too low to serve as targets for T-cell recognition and thus cannot serve as autoantigens. It is likely that very few self proteins contain peptides that are presented by a given MHC molecule at a level sufficiently high to be recognized by effector T cells but too low to induce tolerance. T cells able to recognize these rare antigens will be present in the individual but will not normally be activated. This is because their receptors only bind self peptides with very low affinity, or because they are exposed to levels of self peptide that are too low to deliver any signal to the T cell. Such T cells are said to be in a state of immunological ignorance. This state has been demonstrated experimentally using transgenic animals in which ovalbumin was expressed at high or very low concentrations in the pancreas. Lymphocytes reactive to ovalbumin were transferred to these animals. The lymphocytes transferred to animals expressing high levels of ovalbumin proliferated in response to ovalbumin presented by antigen-presenting cells and then died. In contrast, the lymphocytes transferred to animals expressing very low levels of pancreatic ovalbumin did not divide but persisted and could be stimulated normally when exposed to high levels of ovalbumin in vitro.

In the organ-specific autoimmune diseases such as type I IDDM and Hashimoto's thyroiditis, autoimmunity is unlikely to reflect a general failure of the main mechanisms of tolerance clonal deletion and clonal inactivation. For example, clonal deletion of developing lymphocytes mediates tolerance to self MHC molecules. If such tolerance were not induced, the reactions to self tissues would be similar to those seen in graft-versus-host disease. To estimate the impact of clonal deletion on the developing T-cell repertoire, we should remember that the frequency of T cells able to respond to any set of nonself MHC molecules can be as high as 10%, yet responses to self MHC antigens are not detected in naturally self-tolerant individuals. Moreover, mice given bone marrow cells from a foreign donor at birth, before significant numbers of mature T cells have appeared, can be rendered fully and permanently tolerant to the bone marrow donor's tissues, provided that the bone marrow donor's cells continue to be produced so as to induce tolerance in each new cohort of developing T cells. This experiment, performed by Medawar, validated Burnet's prediction that developing lymphocytes collectively carrying a complete repertoire of receptors must be purged of self-reactive cells before they achieve functional maturity; it won them a Nobel Prize.

Clonal deletion reliably removes all T cells that can mount aggressive responses against self MHC molecules; the organ-specific autoimmune diseases, which involve rare T-cell responses to a particular self peptide

bound to a self MHC molecule, are therefore unlikely to reflect a general failure in clonal deletion; nor are they likely to be caused by a random failure in the mechanisms responsible for anergy. Rather, the lymphocytes that mediate autoimmune responses seem not to be subject to clonal deletion or inactivation. Such autoreactive cells are present in all of us, but they do not normally cause autoimmunity because they are activated only under special circumstances.

A striking demonstration that autoreactive T cells can be present in healthy individuals comes from a strain of mice carrying transgenes encoding an autoreactive T-cell receptor specific for a peptide of myelin basic protein bound to self MHC class II molecules. The autoreactive receptor is present on every T cell, yet the mice are healthy unless their T cells are activated. As the level of the specific peptide:MHC class II complex is low except in the central nervous system, a site not visited by naive T cells, the autoreactive T cells remain in a state of immunological ignorance. When these T cells are activated, for example, by deliberate immunization with myelin basic protein, as in EAE, they migrate into all tissues, including the central nervous system, where they recognize their myelin basic protein:MHC class II ligand. Recognition triggers cytokine production by the activated T cells, causing inflammation in the brain and the destruction of myelin and neurons that ultimately causes the paralysis in EAE.

It is likely that only a small fraction of proteins will be able to serve as autoantigens. An autoantigen must be presented by an MHC molecule at a level sufficient for the antigen to be recognized by effector T cells, but must not be presented to naive T cells at a level sufficient to induce tolerance. Many self proteins are expressed at levels too low to be detected even by effector T cells. It has been estimated that we can make approximately 10^5 proteins of average length 300 amino acids, capable of generating about 3×10^7 distinct self peptides. As there are rarely more than 10^5 MHC molecules per cell, and as the MHC molecules on a cell must bind 10–100 identical peptides for T-cell recognition to occur, fewer than 10,000 self peptides (<1 in 3000) can be presented by any given antigen-presenting cell at levels detectable by T cells. Thus, most peptides will be presented at levels that are insufficient to engage effector T cells, whereas many of the peptides that can be detected by T cells will be presented at a high enough level to induce clonal deletion or anergy. However, a few peptides may fail to induce tolerance yet be present at high enough levels to be recognized by effector T cells. Autoreactivity probably arises most frequently when the antigen is expressed selectively in a

tissue, as is the case of insulin in the pancreas, rather than ubiquitously, because tissue-specific antigens are less likely to induce clonal deletion of developing T cells in the thymus. The nature of such peptides will vary depending on the MHC genotype of the individual, because MHC polymorphism profoundly affects peptide binding. This argument leaves aside the crucial issue of how T cells specific for such autoantigens are activated to become effector T cells.

If it is true that only a few peptides can act as autoantigens, then it is not surprising that there are relatively few distinct autoimmune syndromes, and that all individuals with a particular autoimmune disease tend to recognize the same antigens. If all antigens could give rise to autoimmunity, one would expect that different individuals with the same disease might recognize different antigens on the target tissue, which does not seem to be the case. Finally, because the level of autoantigenic peptide presented is determined by polymorphic residues in MHC molecules that govern the affinity of peptide binding, this idea could also explain the association of autoimmune diseases with particular MHC genotypes.

8.1.1 Induction of a Tissue-specific Response

Only antigen-presenting cells that express co-stimulatory activity can initiate clonal expansion of T cells an essential step in all adaptive immune responses, including graft rejection and, presumably, autoimmunity. In tissue grafts, it is the donor antigen-presenting cells in the graft that initially stimulate host T cells, starting the direct allorecognition response that leads to graft rejection. Antigen-presenting cells bearing both graft antigens and co-stimulatory activity travel to regional lymph nodes. Here they are examined by large numbers of naive host T cells and can activate those that bear specific receptors. Grafts depleted of antigen-presenting cells are tolerated for long periods, but are eventually rejected. This rejection is due to the recipient's T cells responding to graft antigens, both MHC and minor H antigens, after they have been processed and presented by recipient antigen-presenting cells.

The ability of the recipient's antigen-presenting cells to pick up antigens in tissues and initiate graft rejection may be relevant to the initiation of autoimmune tissue damage as well. Transplantation experiments show that host antigen-presenting cells can stimulate both cytotoxic T-cell responses and inflammatory T_H1 responses against the transplanted tissue; thus, tissue antigens can be taken up and presented in conjunction with both MHC class I and class II molecules by antigen-

presenting cells. In autoimmunity, tissues may be similarly attacked by MHC class I-restricted cytotoxic T cells or injured by inflammatory damage mediated by T_H1 cells, as a consequence of the uptake and presentation of tissue antigens by such antigen-presenting cells.

To induce a response to tissue antigens, the antigen-presenting cell must express co-stimulatory activity. The expression of co-stimulatory molecules in antigen-presenting cells is regulated to occur in response to infection. Transient autoimmune responses are seen in the context of such events, and it is thought that one trigger for autoimmunity is infection.

8.1.2 Induction of Tolerance in the Absence of Co-stimulation

Activation of naive T cells requires interaction with cells expressing both the appropriate peptide:MHC complex and co-stimulatory molecules; in the absence of full co-stimulation, specific antigen recognition leads to partial T-cell activation, leading to T-cell anergy or deletion. Tissue cells are not known to express B7 or other co-stimulatory molecules, and can therefore induce tolerance. Experiments with transgenes show that the expression of foreign antigens in peripheral tissues can in some cases induce tolerance, whereas in other cases the foreign antigen seems not to be presented to naive T cells at a sufficient level and is ignored. Autoimmunity can be induced by coexpression of a foreign antigen and B7 in the same target tissue, but as B7 expression on peripheral tissue cells is not by itself a sufficient stimulus for autoimmunity, it is clear that the loss of tolerance to self tissues requires the coexpression of both a suitable target antigen and co-stimulatory molecules. Antigens that are unable to induce clonal anergy or deletion, but that can nonetheless act as targets for effector T cells, can serve as autoantigens; these antigens are likely to be tissue-specific and relatively few.

By analogy with graft rejection, it seems likely that autoimmunity is initiated when a professional antigen-presenting cell picks up a tissue-specific autoantigen and migrates to the regional lymph node, where it is induced to express co-stimulatory activity. Once an autoantigen is expressed on a cell with co-stimulatory potential, naive ignorant T cells specific for the autoantigen can become activated and can home to the tissues, where they interact with their target antigens. At this stage, the absence of co-stimulatory molecules on tissue cells that present the autoantigen can again limit the response. Armed effector T cells kill only a limited number of antigen-expressing tissue cells if

these lack co-stimulatory activity; after killing a few targets, the effector cell dies. Thus, not only can responses not be initiated in the absence of co-stimulatory activity, they also cannot be sustained. Therefore, in addition to the question of how autoimmunity is avoided, we must ask: Why does it ever occur? How are responses to self initiated, and how they are sustained?

8.2 DOMINANT IMMUNE SUPPRESSION

In some models of tolerance, it can be shown that specific T cells actively suppress the actions of other T cells that can cause tissue damage. Tolerance in these cases is dominant in that it can be transferred by T cells, which are usually called suppressor T cells or regulatory T cells. Furthermore, depletion of the suppressor T cells leads to aggravated responses to self or graft antigens. Although it is clear that immune suppression exists, the mechanisms responsible are highly controversial. Here, we will examine the phenomenon in several animal models.

Neonatal rats can be rendered tolerant to allogeneic skin grafts by prior injection of allogeneic bone marrow. This tolerance is highly specific and can be transferred to normal adult recipient rats. This shows that tolerance in this model is dominant and active, as the transferred cells prevent the lymphocytes of the recipient from mediating graft rejection. In order to transfer this tolerance, cells of both the allogeneic graft donor and the neonatal tolerized host must be transferred. Removal of either cell type abolishes the transfer of tolerance.

This finding is reminiscent of Medawar's studies of tolerance in neonatal bone marrow chimeric mice. In both cases, even injection of massive numbers of normal syngeneic lymphocytes, which would react vigorously against the foreign cells in the normal environment of the cell donor, did not break tolerance. Tolerance could be broken only by alloreactive cells from an animal that had been immunized with cells from the allogeneic donor before transfer; such cells probably break tolerance by killing the allogeneic donor cells. Thus, an active host response prevents graft rejection in this model. The tolerance is specific for cells of the original donor, and so the suppression must also be specific.

When mice transgenic for a T-cell receptor specific for myelin basic protein are crossed with RAG mice, they spontaneously develop EAE. TCR transgenic mice that have functional RAG genes are able to rearrange their endogenous TCRa chain genes. Since TCRa chain expression is not allelically excluded, many T cells in such TCR

transgenic mice nevertheless express receptors containing endogenous TCRαδ chains and have a diverse repertoire. In the $RAG^{-/-}$ mice, no such rearrangements can occur, and the only T-cell receptor expressed is encoded by the transgenes. The observation that, when the background population of diverse T cells is lost, the mice develop an autoimmune disease suggests that this population contains cells normally capable of suppressing the autoimmune disease. Such cells have been shown in an increasing number of autoimmune diseases, and their isolation as cloned T-cell lines is a major goal for people who study the induction of autoimmunity. The reason for this renewed interest in T cell-mediated regulation is that the intentional induction of such cells could be a major advance in the prevention of autoimmune disease.

The mechanisms of tolerance in these animal models are not fully understood. There is evidence for the existence of CD4-positive regulatory cells that can inhibit autoimmune disease. These cells have an activated phenotype and can be identified by the expression of CD25, the IL-2 receptor αδ chain. Depletion of these cells from the peripheral immune system of normal mice leads to the development of insulin-dependent diabetes, thyroiditis, and gastritis. Similarly, if thymocytes depleted of $CD25^+$ CD4 cells are adoptively transferred into athymic recipients, autoimmune disease results. It is not known how these cells mediate their effects, but one possibility is by secreting cytokines that inhibit other lymphocytes.

In the NOD mouse model for type I IDDM, transfer of a particular insulin-specific T-cell clone can prevent the destruction of pancreatic βδ cells by autoreactive T cells. This suggests that the insulin-specific T cells can suppress the activity of other autoaggressive T cells in an antigen-dependent manner. They do this by homing to the islet, where they react with insulin peptides presented on the macrophages or dendritic cells. This stimulates the secretion of cytokines, prominent among which is TGF-βδ, a known immunosuppressive cytokine. There are interesting hints that such cells naturally affect the course of the autoimmune response that causes human diabetes; βδ-cell destruction in humans occurs over a period of several years before diabetes is manifest, yet when new islets are transplanted from an identical twin into his or her diabetic sibling, they are destroyed within weeks. This suggests that, in the normal course of the disease, specific T cells protect the βδ-cells from attack by effector T cells and the disease therefore progresses slowly. It might be that after the host islets have been destroyed, these protective mechanisms decline in activity but the effector cells responsible for b-cell destruction do not.

If specific suppression of autoimmune responses could be elicited at will, autoimmunity would not be a problem. Feeding with specific antigen is known to elicit a local immune response in the intestinal mucosa, and responses to the same antigen given subsequently by a systemic route are suppressed. This response has been exploited in experimental autoimmune diseases by feeding proteins from target tissues to mice. Mice fed with insulin are protected from diabetes; mice fed with myelin basic protein are resistant to EAE.

EAE is usually caused by T_H1 cells that produce IFN-gð in response to myelin basic protein. In mice fed this protein, CD4 T cells found in the brain instead produce cytokines such as TGF-bð and IL-4. TGF-bð, in particular, suppresses the function of inflammatory T_H1 lymphocytes. In these cases, the protection seems to be tissue-specific rather than antigen-specific. Thus, feeding with myelin basic protein will protect against EAE elicited by immunization with other brain antigens. Feeding with antigen might induce the production of T cells producing TGF-b and IL-4 because these cytokines are also required for IgA production against antigens ingested in food. If feeding with antigen works as a clinical therapy, it would have the advantage over treatments with immunosuppressive drugs that it does not alter the general immune competence of the host. Unfortunately, early studies of this approach to treatment in humans with multiple sclerosis or rheumatoid arthritis have shown minimal, if any, benefit. These strategies may prove more successful in preventing the onset of disease than reversing established disease. However, this approach requires the identification of patients at the very onset of disease or those who are at extremely high risk of developing disease and adds impetus to studies to identify the disease susceptibility genes for the development of autoimmune diseases.

Another strategy for controlling immune responses is to manipulate the cytokine balance that determines whether a CD4 T-cell response is predominantly T_H1 or T_H2. It is possible experimentally to switch T_H1 to T_H2 responses by administration of IL-4, and vice versa using IFN-g. This is known as immune deviation.

Unlike human diabetes, which follows a chronic progressive course in humans, multiple sclerosis is a chronic relapsing disease with acute episodes followed by periods of quiescence. This suggests a balance between autoimmune and protective T cells that can alter at different stages of the disease. However, it remains to be proven whether the specific suppressive cells discussed in this section exist naturally and

contribute to self-tolerance, or whether they arise only upon artificial stimulation or in response to autoimmune attack. Nevertheless, because they can play an active, dominant part in self-tolerance, they are particularly attractive potential mediators for immunotherapy of autoimmune disease.

8.2.1 Antigens at Immunologically Privileged Sites

Tissue grafts placed in some sites in the body do not elicit immune responses. For instance, the brain and the anterior chamber of the eye are sites in which tissues can be grafted without inducing rejection. Such locations are termed immunologically privileged sites. It was originally believed that immunological privilege arose from the failure of antigens to leave privileged sites and induce immune responses. Subsequent studies have shown, however, that antigens do leave immunologically privileged sites, and that these antigens do interact with T cells; but instead of eliciting a destructive immune response, they induce tolerance or a response that is not destructive to the tissue. Immunologically privileged sites seem to be unusual in three ways. First, the communication between the privileged site and the body is atypical in that extracellular fluid in these sites does not pass through conventional lymphatics, although proteins placed in these sites do leave them and can have immunological effects. Naive lymphocytes, similarly, may be excluded by the tissue barriers of privileged sites, such as the blood-brain barrier. Second, humoral factors, presumably cytokines, that affect the course of an immune response are produced in privileged sites and leave them together with antigens. The anti-inflammatory cytokine TGF-bð seems to be particularly important in this regard: antigens mixed with TGF-bð seem to induce mainly T-cell responses that do not damage tissues, such as T_H2 rather than T_H1 responses. Third, the expression of Fas ligand by the tissues of immunologically privileged sites may provide a further level of protection by inducing apoptosis of Fas-bearing lymphocytes that enter these sites. This mechanism of protection is not fully understood, as it appears that under some circumstances the expression of Fas ligand by tissues may trigger an inflammatory response by neutrophils.

Paradoxically, the antigens sequestered in immunologically privileged sites are often the targets of autoimmune attack; for example, brain autoantigens such as myelin basic protein are targeted in multiple sclerosis. It is therefore clear that this antigen does not induce tolerance due to clonal deletion of the self-reactive T cells. Mice only become diseased when they are deliberately immunized with myelin basic protein,

in which case they become acutely sick, show severe infiltration of the brain with specific T_H1 cells, and often die.

Thus, at least some antigens expressed in immunologically privileged sites induce neither tolerance nor activation, but if activation is induced elsewhere they can become targets for autoimmune attack. It seems plausible that T cells specific for antigens sequestered in immunologically privileged sites are more likely to remain in a state of immunological ignorance. Further evidence for this comes from the eye disease sympathetic ophthalmia. If one eye is ruptured by a blow or other trauma, an autoimmune response to eye proteins can occur, although this happens only rarely. Once the response is induced, it often attacks both eyes. Immuno-suppression and removal of the damaged eye, the source of antigen, is frequently required to preserve vision in the undamaged eye.

It is not surprising that effector T cells can enter immunologically privileged sites: such sites can become infected, and effector cells must be able to enter these sites during infection. Effector T cells enter most or all tissues after activation, but accumulations of cells are seen only when antigen is recognized in the site, triggering the production of cytokines that alter tissue barriers.

8.2.2 Regulation of B cells with receptors specific for peripheral autoantigens.

During B-cell development in the bone marrow, B-cell antigen receptors specific for self molecules are produced as a consequence of the random generation of the repertoire. If a self molecule is expressed in the bone marrow in an appropriate form, clonal deletion and receptor editing can remove all of these self-reactive B cells while they are still immature. There are, however, many self molecules available only in the periphery whose expression is restricted to particular organs. An example is thyroglobulin (the precursor of thyroxine), which is expressed only in the thyroid and at extremely low levels in plasma. Back-up mechanisms exist to ensure that B cells reactive to these self molecules do not cause autoimmune disease. When a mature B cell in the periphery encounters self molecules that bind its receptor, four proposed mechanisms could bring about the observed nonreactivity. Failure of any one of these mechanisms could lead to autoimmunity.

First, B cells that recognize a self antigen arrest their migration in the T-cell zone of peripheral lymphoid tissues, just like B cells that bind a foreign antigen. However, in contrast to the response to foreign antigens, in which activated effector CD4 T cells are present,

B cells binding self antigens will not be able to interact with helper CD4 T cells because normally no such cells exist for self antigens. This lack of interaction prevents the B cells from migrating out of the T-cell zones into the follicles; instead, the trapped self-reactive B cells undergo apoptosis.

A second mechanism is the induction of B-cell anergy, which is associated with downregulation of surface IgM expression and partial inhibition of the linked signaling pathways. B-cell anergy can be induced by exposure to soluble circulating antigen; if mice are inoculated intravenously with protein solutions from which all protein aggregates have been rigorously removed to eliminate multivalent complexes, the

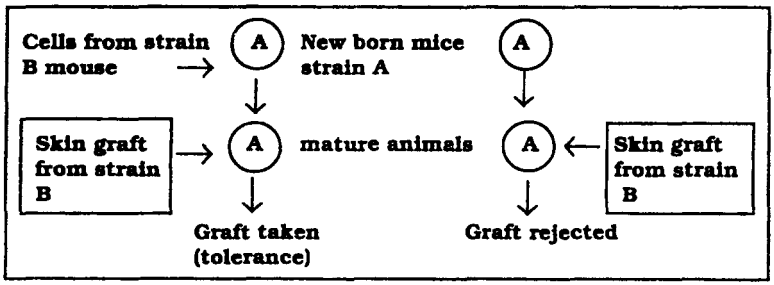


Figure 8.1 : Experiment showing that neonatal expossure to an antigen generates tolerance to it.

peripheral B cells that bind these proteins can be inactivated. The lifespan of anergic B cells is short as they are usually eliminated after failing to enter the primary lymphoid follicles or after interacting with antigen-specific T cells, as described below. This form of B-cell tolerance can therefore be viewed as a form of delayed B-cell deletion, with the significant difference that there may be autoimmune diseases such as SLE in which anergic cells can be rescued.

The third mechanism depends on the presence of T cells that are specific for the self antigen and express Fas ligand. In the rare instances when such an autoreactive T cell matures and is activated, it is able to kill autoreactive anergic B cells in a Fas-dependent manner. In the absence of the normal pathways of co-stimulation, anergized B lymphocytes that have been chronically exposed to self antigen show enhanced sensitivity to apoptosis after ligation of Fas by Fas ligand. They are therefore not subject to the stimulatory signals that oppose apoptosis in B cells whose surface receptors have just been ligated by foreign antigen. The importance of this mechanism is nicely illustrated by the consequences of mutation in the genes for Fas or Fas ligand.

Mice deficient in Fas or Fas ligand develop severe autoimmune disease, similar to SLE, associated with the production of a similar spectrum of autoantibodies.

Finally, there is evidence for a distinct mechanism for dealing with B cells that develop self-reactive specificities as a result of somatic hypermutation during a response to a foreign antigen. At a crucial phase at the height of the germinal center reaction, an encounter with a large dose of soluble antigen causes a wave of apoptosis in germinal center B cells within a few hours. Thus B cells that become reactive for abundant soluble self antigens could be removed.

All these mechanisms reemphasize the fact that the mere existence in the body of some B lymphocytes with receptor specificities directed against self is not in itself harmful. Before an immune response can be initiated they need to receive effective help, the B-cell receptors must be ligated, and their intracellular signaling machinery must be set to respond normally.

8.2.3 Autoimmunity Triggered by Infection

Human autoimmune diseases often appear gradually, making it difficult to find out how the process is initiated. Nevertheless, there is a strong suspicion that infections can trigger autoimmune disease in genetically susceptible individuals. Indeed, many experimental autoimmune diseases are induced by mixing tissue cells with adjuvants that contain bacteria. For example, to induce EAE, the spinal cord or myelin basic protein used for immunization must be emulsified in complete Freund's adjuvant, which includes killed Mycobacterium tuberculosis. When the mycobacteria are omitted from the adjuvant, not only is no disease elicited, but the animals become refractory to any subsequent attempt to induce the disease by antigen in complete Freund's adjuvant, and this resistance can be transferred to syngeneic recipients by T cells. Infection is important in the induction of disease in several other known cases. For example, transgenic mice that express a T-cell receptor specific for myelin basic protein often develop spontaneous autoimmune encephalo-myelitis if they become infected. One possible mechanism for this loss of tolerance is that the infectious agents induce co-stimulatory activity on antigen-presenting cells expressing low levels of peptides from myelin basic protein, thus activating the autoreactive T cells.

It has also been suggested that autoimmunity can be initiated by a mechanism known as molecular mimicry, in which antibodies or T cells generated in response to an infectious agent cross-react with self antigens.

9

Transplantation and Rejection

The transplantation of tissues to replace diseased organs is now an important medical therapy. In most cases, adaptive immune responses to the grafted tissues are the major impediment to successful transplantation. Rejection is caused by immune responses to alloantigens on the graft, which are proteins that vary from individual to individual within a species, and are thus perceived as foreign by the recipient. In blood transfusion, which was the earliest and is still the most common tissue transplant, blood must be matched for ABO and Rh blood group antigens to avoid the rapid destruction of mismatched red blood cells by antibodies. Because there are only four major ABO types and two Rh blood types, this is relatively easy. When tissues containing nucleated cells are transplanted, however, T-cell responses to the highly polymorphic MHC molecules almost always trigger a response against the grafted organ. Matching the MHC type of donor and recipient increases the success rate of grafts, but perfect matching is possible only when donor and recipient are related and, in these cases, genetic differences at other loci still trigger rejection. In this section, we will examine the immune response to tissue grafts, and ask why such responses do not reject the one foreign tissue graft that is tolerated routinely the mammalian fetus.

9.1 GRAFT REJECTION: AN IMMUNOLOGICAL RESPONSE MEDIATED PRIMARILY BY T CELLS

The basic rules of tissue grafting were first elucidated by skin transplantation between inbred strains of mice. Skin can be grafted

with 100% success between different sites on the same animal or person (an autograft), or between genetically identical animals or people (a syngeneic graft). However, when skin is grafted between unrelated or allogeneic individuals (an allograft), the graft is initially accepted but is then rejected about 10-13 days after grafting. This response is called a *first-set rejection* and is quite consistent. It depends on a T-cell response in the recipient, because skin grafted onto *nude* mice, which lack T cells, is not rejected. The ability to reject skin can be restored to *nude* mice by the adoptive transfer of normal T cells.

When a recipient that has previously rejected a graft is regrafted with skin from the same donor, the second graft is rejected more rapidly (6-8 days) in a *second-set rejection*. Skin from a third-party donor grafted onto the same recipient at the same time does not show this faster response but follows a first-set rejection course. The rapid course of second-set rejection can be transferred to normal or irradiated recipients by transferring T cells from the initial recipient, showing that graft rejection is caused by a specific immunological reaction.

Immune responses are a major barrier to effective tissue transplantation, destroying grafted tissue by an adaptive immune response to its foreign proteins. These responses can be mediated by CD8 T cells, by CD4 T cells, or by both. Antibodies can also contribute to second-set rejection of tissue grafts.

9.1.1 Matching of Donor and Recipient at the MHC

When donor and recipient differ at the MHC, the immune response, which is known as an alloreactive response as it is directed against antigens (alloantigens) that differ between members of the same species, is directed at the nonself allogeneic MHC molecule or molecules present on the graft. In most tissues, these will be predominantly MHC class I antigens. Once a recipient has rejected a graft of a particular MHC type, any further graft bearing the same nonself MHC molecule will be rapidly rejected in a second-set response. As we learned in Chapter elsewhere in this book, the frequency of T cells specific for any nonself MHC molecule is relatively high, making differences at MHC loci the most potent trigger of the rejection of initial grafts; indeed, the major histocompatibility complex was originally so named because of its central role in graft rejection.

Once it became clear that recognition of nonself MHC molecules is a major determinant of graft rejection, a considerable amount of effort was put into MHC matching between recipient and donor. Although HLA matching significantly improves the success rate of

clinical organ transplantation, it does not in itself prevent rejection reactions. There are two main reasons for this. First, HLA typing is imprecise, owing to the polymorphism and complexity of the human MHC; unrelated individuals who type as HLA-identical with antibodies against MHC proteins rarely have identical MHC genotypes. This should not be a problem with HLA-identical siblings: because siblings inherit their MHC genes as a haplotype, one sibling in four should be truly HLA-identical. Nevertheless, grafts between HLA-identical siblings are invariably rejected, albeit more slowly, unless donor and recipient are identical twins. This rejection is the result of differences between minor histocompatibility antigens, which is the second reason for the failure of HLA matching to prevent rejection reactions.

Thus, unless donor and recipient are identical twins, all graft recipients must be given immunosuppressive drugs to prevent rejection. Indeed, the current success of clinical transplantation of solid organs is more the result of advances in immunosuppressive therapy than of improved tissue matching. The limited supply of cadaveric organs, coupled with the urgency of identifying a recipient once a donor organ becomes available, means that accurate matching of tissue types is achieved only rarely.

9.1.2 MHC-identical Grafts Rejection

When donor and recipient are identical at the MHC but differ at other genetic loci, graft rejection is not so rapid. The polymorphic antigens responsible for the rejection of MHC-identical grafts are therefore termed minor histocompatibility antigens or *minor H antigens*. Responses to single minor H antigens are much less potent than responses to MHC differences because the frequency of responding T cells is much lower. However, most inbred mouse strains that are identical at the MHC differ at multiple minor H antigen loci, so grafts between them are still uniformly and relatively rapidly rejected. The cells that respond to minor H antigens are generally CD8 T cells, implying that most minor H antigens are complexes of donor peptides and MHC class I molecules. However, peptides bound to MHC class II molecules can also participate in the response to MHC-identical grafts.

Minor H antigens are now known to be peptides derived from polymorphic proteins that are presented by the MHC molecules on the graft. MHC class I molecules bind and present a selection of peptides derived from proteins made in the cell, and if polymorphisms in these proteins mean that different peptides are produced in different members

of a species, these can be recognized as minor H antigens. One set of proteins that induce minor H responses is encoded on the male-specific Y chromosome. Responses induced by these proteins are known collectively as H-Y. As these Y chromosome-specific genes are not expressed in females, female anti-male minor H responses occur; however, male anti-female responses are not seen, because both males and females express X-chromosome genes. One H-Y antigen has been identified in mice and humans as peptides from a protein encoded by the Y-chromosome gene Smcy. An X-chromosome homologue of Smcy, called Smcx, does not contain these peptide sequences, which are therefore expressed uniquely in males. The nature of the majority of minor H antigens, encoded by autosomal genes, is unknown, but one, HA-2, has been identified as a peptide derived from myosin.

The response to minor H antigens is in most ways analogous to the response to viral infection. However, an antiviral response eliminates only infected cells, whereas all cells in the graft express minor H antigens, and thus the entire graft is destroyed in the response against these antigens. Thus, even though MHC genotype might be matched exactly, polymorphism in any other protein could elicit potent T-cell responses that would destroy the entire graft. It is no wonder that successful transplantation requires the use of powerful immunosuppressive drugs.

9.1.3 Ways of Presenting Alloantigens on the Transplant to the Recipient's T Lymphocytes

Alloreactive effector T cells that bind directly to allogeneic MHC class I molecules in mismatched organ grafts are an important cause of graft rejection; this is called direct allorecognition. Before they can cause rejection, naive alloreactive T cells must be activated by antigenpresenting cells that both bear the allogeneic MHC molecules and have co-stimulatory activity. Organ grafts carry with them antigen-presenting cells of donor origin, known as passenger leukocytes, and these are an important stimulus to alloreactivit). This route for sensitization of the recipient to a graft seems to involve donor antigen-presenting cells leaving the graft and migrating via the lymph to regional lymph nodes. Here they can activate those host T cells that bear the corresponding T-cell receptors. The activated alloreactive effector T cells are then carried back to the graft, which they attack directly. Indeed, if the grafted tissue is depleted of antigen-presenting cells by treatment with antibodies or by prolonged incubation, rejection occurs only after a much longer time. Also, if the site of grafting lacks lymphatic drainage, no response to the graft results.

A second mechanism of allograft recognition leading to graft rejection is the uptake of allogeneic proteins by the recipient's own antigen-presenting cells and their presentation to T cells by self MHC molecules. The recognition of allogeneic proteins presented in this way is known as indirect allorecognition, in contrast to the direct recognition by T cells of allogeneic MHC class I and class II molecules expressed on the graft itself. Among the graft-derived peptides presented by the recipient's antigen-presenting cells are the minor H antigens and also peptides from the foreign MHC molecules themselves, which are a major source of the polymorphic peptides recognized by the recipient's T cells.

The relative contributions of direct and indirect allorecognition in graft rejection are not known. Direct allorecognition is thought to be largely responsible for acute rejection, especially when MHC mismatches mean that the frequency of directly alloreactive recipient T cells is high. Furthermore, a direct cytotoxic T-cell attack on graft cells can be carried out only by T cells that recognize the graft MHC molecules directly. Nonetheless, T cells with indirect allospecificity can contribute to graft rejection by activating macrophages, which cause tissue injury and fibrosis, and are likely to be important in the development of an alloantibody response to a graft.

9.1.4 Antibodies Reacting with Endothelium Cause Hyperacute Graft Rejection

Antibody responses are also an important potential cause of graft rejection. Preexisting alloantibodies to blood group antigens and polymorphic MHC antigens can cause rapid rejection of transplanted organs in a complement-dependent reaction that can occur within minutes of transplantation. This type of reaction is known as hyperacute graft rejection. Most grafts that are transplanted routinely in clinical medicine are vascularized organ grafts linked directly to the recipient's circulation. In some cases, the recipient might already have circulating antibodies against donor graft antigens, produced in response to a previous transplant or a blood-transfusion. Such antibodies can cause very rapid rejection of vascularized grafts because they react with antigens on the vascular endothelial cells of the graft and initiate the complement and blood clotting cascades. The vessels of the graft become blocked, causing its immediate death. Such grafts become engorged and purple-colored from hemorrhaged blood, which becomes deoxygenated. This problem can be avoided by cross-matching donor and recipient. Cross-matching involves determining whether the recipient has antibodies that react with the

white blood cells of the donor. If antibodies of this type are found, they are a serious contraindication to transplantation, as they lead to near-certain hyperacute rejection.

A very similar problem prevents the routine use of animal organs xenografts in transplantation. If xenogeneic grafts could be used, it would circumvent the major limitation in organ replacement therapy, namely the severe shortage of donor organs. Pigs have been suggested as a potential source of organs for xenografting as they are a similar size to humans and are easily farmed. Most humans and other primates have antibodies that react with endothelial cell antigens of other mammalian species, including pigs. When pig xenografts are placed in humans, these antibodies trigger hyper-acute rejection by binding to the endothelial cells of the graft and initiating the complement and clotting cascades. The problem of hyperacute rejection is exacerbated in xenografts because complement-regulatory proteins such as CD59, DAF (CD55), and MCP (CD46) work less efficiently across a species barrier; the complement-regulatory proteins of the xenogeneic endothelial cells cannot protect them from attack by human complement. A recent step toward xenotransplantation has been the development of transgenic pigs expressing human DAF. Preliminary experiments have shown prolonged survival of organs transplanted from these pigs into recipient cynomolgus monkeys, under cover of heavy immunosuppression. However, hyperacute rejection is only the first barrier faced by a xenotransplanted organ. The T lymphocyte-mediated graft rejection mechanisms might be extremely difficult to overcome with present immunosuppressive regimes.

9.2 GRAFT-VERSUS-HOST DISEASE

Allogeneic bone marrow transplantation is a successful therapy for some tumors derived from bone marrow precursors, such as certain leukemias and lymphomas. It may also be successful in the treatment of some primary immunodeficiency diseases and inherited bone marrow diseases, such as the severe forms of thalassemia. In leukemia therapy, the recipient's bone marrow, the source of the leukemia, must first be destroyed by aggressive cytotoxic chemotherapy. One of the major complications of allogeneic bone marrow transplantation is graft-versus-host disease (GVHD), in which mature donor T cells that contaminate the allogeneic bone marrow recognize the tissues of the recipient as foreign, causing a severe inflammatory disease characterized by rashes, diarrhea, and pneumonitis. Graft-versus-host disease occurs not only when there is a mismatch of a major MHC class I or class II antigen but also in the context of disparities between minor H antigens. Graft-

versus-host disease is a common complication in recipients of bone marrow transplants from HLA-identical siblings, who typically differ from each other in many polymorphic proteins encoded by genes unlinked to the MHC.

The presence of alloreactive T cells can easily be demonstrated experimentally by the mixed lymphocyte reaction (MLR), in which lymphocytes from a potential donor are mixed with irradiated lymphocytes from the potential recipient. If the donor lymphocytes contain alloreactive T cells, these will respond by cell division. The MLR is sometimes used in the selection of donors for bone marrow transplants, when the lowest possible alloreactive response is essential. However, the limitation of the MLR in selection of bone marrow donors is that the test does not accurately quantitate alloreactive T cells. A more accurate test is a version of the limiting-dilution assay, which precisely counts the frequency of alloreactive T cells.

Although graft-versus-host disease is usually harmful to the recipient of a bone marrow transplant, there can be some beneficial effects. Some of the therapeutic effect of bone marrow transplantation for leukemia can be due to a graft-versus-leukemia effect, in which the allogeneic bone marrow recognizes minor H antigens or tumor-specific antigens expressed by the leukemic cells, leading the donor cells to kill the leukemic cells. One such minor H antigen, HB-1, is a B-cell lineage marker that is expressed by acute lymphoblastic leukemia cells, which are B-lineage cells, and by B lymphocytes transformed with Epstein-Barr virus (EBV). One of the treatment options for suppressing the development of graft-versus-host disease is the elimination of mature T cells from the donor bone marrow *in vitro* before transplantation, thereby removing alloreactive T cells. Those T cells that subsequently mature from the donor marrow *in vivo* in the recipient are tolerant to the recipient's antigens. Although the elimination of graft-versus-host disease has benefits for the patient, there is an increase in the risk of leukemic relapse, which provides strong evidence in support of the graft-versus-leukemia effect.

9.2.1 Chronic Organ Rejection

The success of modern immuno suppression means that approximately 85% of cadaveric kidney grafts are still functioning a year after transplantation. However, there has been no improvement in rates of long-term graft survival: the half-life for functional survival of renal allografts remains about 8 years. The major cause of late organ failure is chronic rejection, characterized by concentric arteriosclerosis of

graft blood vessels, accompanied by glomerular and tubular fibrosis and atrophy.

Mechanisms that contribute to chronic rejection can be divided into those due to alloreactivity and those due to other pathways, and into early and late events after transplantation. Alloreactivity may occur days and weeks after transplantation and cause acute graft rejection. Alloreactive responses may also occur months to years after transplantation, and be associated with clinically hard-to-detect gradual loss of graft function. Other important causes of chronic graft rejection include ischemia-reperfusion injury, which occurs at the time of grafting but may have late adverse effects on the grafted organ, and later-developing adverse factors such as chronic cyclosporin toxicity or cytomegalovirus infection.

Infiltration of the graft vessels and tissues by macrophages, followed by scarring, are prominent histological features of late graft rejection. A model of injury has been developed in which alloreactive T cells infiltrating the graft secrete cytokines that stimulate the expression of endothelial adhesion molecules and also secrete chemokines such as RANTES, which attracts monocytes that mature into macrophages in the graft. A second phase of chronic inflammation then supervenes, dominated by macrophage products including interleukin (IL)-1, TNF-aδ and the chemokine MCP, which leads to further macrophage recruitment. These mediators conspire to cause chronic inflammation and scarring, which eventually leads to irreversible organ failure. Animal models of chronic rejection also show that alloreactive IgG antibodies may induce accelerated atherosclerosis in transplanted solid organs.

9.2.2 Organs Routinely Transplanted in Clinical Medicine

Although the immune response makes organ transplantation difficult, there are few alternative therapies for organ failure. Three major advances have made it possible to use organ transplantation routinely in the clinic. First, the technical skill to carry out organ replacement surgery has been mastered by many people. Second, networks of transplantation centers have been organized to ensure that the few healthy organs that are available are HLA-typed and so matched with the most suitable recipient. Third, the use of powerful immunosuppressive drugs, especially cyclosporin A and FK-506, known as tacrolimus, to inhibit T-cell activation, has increased graft survival rates dramatically. Some of different organs are transplanted routinely with a very high

success rate. By far the most frequently transplanted solid organ is the kidney, the organ first successfully transplanted between identical twins in the 1950s. Transplantation of the cornea is even more frequent; this tissue is a special case, as it is not vascularized, and corneal grafts between unrelated people are usually successful even without immunosuppression.

There are, however, many problems other than graft rejection associated with organ transplantation. First, donor organs are difficult to obtain; this is especially a problem when the organ involved is a vital one, such as the heart or liver. Second, the disease that destroyed the patient's organ might also destroy the graft. Third, the immunosuppression required to prevent graft rejection increases the risk of cancer and infection. Finally, the procedure is very costly. All of these problems need to be addressed before clinical transplantation can become commonplace. The problems most amenable to scientific solution are the development of more effective means of immunosuppression, the induction of graft-specific tolerance, and the development of xenografts as a practical solution to organ availability.

9.3 THE FETUS: AN ALLOGRAFT

All of the transplants discussed so far are artefacts of modern medical technology. However, one tissue that is repeatedly grafted and repeatedly tolerated is the mammalian fetus. The fetus carries paternal MHC and minor H antigens that differ from those of the mother, and yet a mother can successfully bear many children expressing the same nonself MHC proteins derived from the father. The mysterious lack of rejection of the fetus has puzzled generations of reproductive immunologists and no comprehensive explanation has yet emerged. One problem is that acceptance of the fetal allograft is so much the norm that it is difficult to study the mechanism that prevents rejection; if the mechanism for rejecting the fetus is rarely activated, how can one analyze the mechanisms that control it?

Various hypotheses have been advanced to account for the tolerance shown to the fetus. It has been proposed that the fetus is simply not recognized as foreign. There is evidence against this hypothesis, as women who have borne several children usually make antibodies directed against the father's MHC proteins; indeed, this is the best source of antibodies for human MHC typing. However, the placenta, which is a fetus-derived tissue, seems to sequester the fetus away from the mother's T cells. The outer layer of the placenta, the interface between fetal and maternal tissues, is the trophoblast. This does not express classical

MHC class I and class II proteins, making it resistant to recognition and attack by maternal T cells. Tissues lacking class I expression are, however, vulnerable to attack by NK cells. The trophoblast might be protected from attack by NK cells by expression of a nonclassical and minimally polymorphic HLA class I molecule HLA-G. This protein has been shown to bind to the two major inhibitory NK receptors, KIR1 and KIR2, and to inhibit NK killing.

The placenta may also sequester the fetus from the mother's T cells by an active mechanism of nutrient depletion. The enzyme indoleamine 2,3-dioxygenase (IDO) is expressed at a high level by cells at the maternal-fetal interface. This enzyme catabolizes, and thereby depletes, the essential amino acid tryptophan at this site. T cells starved of tryptophan show reduced responsiveness. Inhibition of IDO in pregnant mice, using the inhibitor 1-methyltryptophan, causes rapid rejection of allogeneic but not syngeneic fetuses. This supports the hypothesis that maternal T cells, alloreactive to paternal MHC proteins, may be held in check in the placenta by tryptophan depletion.

It is likely that fetal tolerance is a multifactorial process. The trophoblast does not act as an absolute barrier between mother and fetus, and fetal blood cells can cross the placenta and be detected in the maternal circulation, albeit in very low numbers. There is direct evidence from experiments in mice for specific T-cell tolerance against paternal MHC alloantigens. Pregnant female mice whose T cells bear a transgenic receptor specific for a paternal alloantigen showed reduced expression of this T-cell receptor during pregnancy. These same mice lost the ability to control the growth of an experimental tumor bearing the same paternal MHC alloantigen. After pregnancy, tumor growth was controlled and the level of the T-cell receptor increased. This experiment demonstrates that the maternal immune system must have been exposed to paternal MHC alloantigens, and that the immune response to these antigens was temporarily suppressed.

Yet another factor that might contribute to maternal tolerance of the fetus is the secretion of cytokines at the maternal-fetal interface. Both uterine epithelium and trophoblast secrete cytokines, including transforming growth factor (TGF)-β, IL-4, and IL-10. This cytokine pattern tends to suppress T_H1 responses. Induction or injection of cytokines such as interferon (IFN)-γ and IL-12, which promote T_H1 responses in experimental animals, promote fetal resorption, the equivalent of spontaneous abortion in humans.

protected by a nonimmunogenic tissue barrier, and it promotes a local immunosuppressive response in the mother. We will see later that several sites in the body have these characteristics and allow prolonged acceptance of foreign tissue grafts. They are usually called immunologically privileged sites.

10

Immunodeficiency

Immunodeficiencies occur when one or more components of the immune system is defective. The commonest cause of immune deficiency worldwide is malnutrition; however, in developed countries, most immunodeficiency diseases are inherited, and these are usually seen in the clinic as recurrent or overwhelming infections in very young children. Less commonly, acquired immunodeficiencies with causes other than malnutrition can manifest later in life. Although the pathogenesis of many of these acquired disorders has remained obscure, some are caused by known agents, such as drugs or irradiation that damage lymphocytes, or infection with measles or HIV. By examining which infections accompany a particular inherited or acquired immunodeficiency, we can see which components of the immune system are important in the response to particular infectious agents. The inherited immunodeficiency diseases also reveal how interactions between different cell types contribute to the immune response and to the development of T and B lymphocytes. Finally, these inherited diseases can lead us to the defective gene, often revealing new information about the molecular basis of immune processes and providing the necessary information for diagnosis, for genetic counseling, and eventually for gene therapy.

10.1 A HISTORY OF REPEATED INFECTIONS

Patients with immune deficiency are usually detected clinically by a history of recurrent infection. The type of infection is a guide to which part of the immune system is deficient. Recurrent infection by pyogenic bacteria suggests a defect in antibody, complement, or phagocyte function, reflecting the role of these parts of the immune system in host defense against such infections. By contrast, a history of recurrent viral infections is more suggestive of a defect in host defense mediated by T lymphocytes.

To determine the competence of the immune system in patients with possible immunodeficiency, a battery of tests is usually conducted; these focus with increasing precision as the nature of the defect is narrowed down to a single element. The presence of the various cell types in blood is determined by routine hematology, often followed by FACS analysis of lymphocyte subsets, and the measurement of serum immunoglobulins. The phagocytic competence of freshly isolated polymorphonuclear leukocytes and monocytes is tested, and the efficiency of the complement system is determined by testing the dilution of serum required for lysis of 50% of antibody-coated red blood cells (this is denoted the CH50).

10.1.1 Inherited Immunodeficiency Diseases

Most of the gene defects that cause these inherited immunodeficiencies are recessive and, for this reason, many of the known immunodeficiencies are caused by mutations in genes on the X chromosome. Recessive defects cause disease only when both chromosomes are defective. However, as males have only one X chromosome, all males who inherit an X chromosome carrying a defective gene will manifest disease, whereas female carriers, having two X chromosomes, are perfectly healthy. Immunodeficiency diseases that affect various steps in B- and T-lymphocyte development have been described, as have defects in surface molecules that are important for T- or B-cell function. Defects in phagocytic cells, in complement, in cytokines, in cytokine receptors, and in molecules that mediate effector responses also occur. Thus, immunodeficiency can be caused by defects in either the adaptive or the innate immune system. Individual examples of these diseases will be described in later sections.

The main effect of low levels of antibody is an inability to clear extracellular bacteria.

Pyogenic, or pus-forming, bacteria have polysaccharide capsules that are not directly recognized by the receptors on macrophages and neutrophils that stimulate phagocytosis. They therefore escape immediate elimination by the innate immune response and are successful extracellular pathogens. Normal individuals can clear infections by such bacteria because antibody and complement opsonize the bacteria, making it possible for phagocytes to ingest and destroy them. The principal effect of deficiencies in antibody production is therefore a failure to control this class of bacterial infection. In addition, susceptibility to some viral infections, most notably those caused by enteroviruses, is also increased, because of the importance of antibodies

in neutralizing infectious viruses that enter the body through the gut. The commonest humoral immune defect is the transient deficiency in immunoglobulin production that occurs in the first 6-12 months of life. The newborn infant has initial antibody levels comparable to those of the mother, because of the transplacental transport of maternal IgG. As the transferred IgG is catabolized, antibody levels gradually decrease until the infant begins to produce useful amounts of its own IgG at about 6 months of age. Thus, IgG levels are quite low between the ages of 3 months and 1 year and active IgG antibody responses are poor. In some infants this can lead to a period of heightened susceptibility to infection. This is especially true for premature babies, who begin with lower levels of maternal IgG and also reach immune competence later after birth.

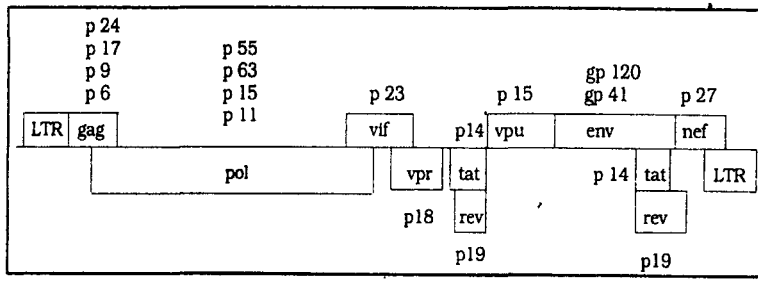


Figure 10.1 : The HIV Genome.

The most common inherited form of immunoglobulin deficiency is selective IgA deficiency, which is seen in about 1 person in 800. Although no obvious disease susceptibility is associated with selective IgA defects, they are commoner in people with chronic lung disease than in the general population. Lack of IgA might thus result in a predisposition to lung infections with various pathogens and is consistent with the role of IgA in defense at the body's surfaces. The genetic basis of this defect is unknown but some data suggest that a gene of unidentified function mapping in the class III region of the MHC could be involved. A related syndrome called common variable immunodeficiency, in which there is usually a deficiency in both IgG and IgA, also maps to the MHC region.

People with pure B-cell defects resist many pathogens successfully. However, effective host defense against a subset of extracellular pyogenic bacteria, including staphylococci and streptococci, requires opsonization of these bacteria with specific antibody. These infections can be suppressed with antibiotics and periodic infusions of human immunog-

lobulin collected from a large pool of donors. As there are antibodies against many pathogens in this pooled immunoglobulin, it serves as a fairly successful shield against infection.

10.1.2 T-cell Defects can Result in Low Antibody Levels

Patients with X-linked hyper IgM syndrome have normal B- and T-cell development and high serum levels of IgM but make very limited IgM antibody responses against T-cell dependent antigens and produce immunoglobulin isotypes other than IgM and IgD only in trace amounts. This makes them highly susceptible to infection with extracellular pathogens. The molecular defect in this disease is in the CD40 ligand expressed on activated T cells, which therefore cannot engage the CD40 molecule on B cells; the B cells themselves are normal.

10.1.3 Defects in Complement Components Cause Defective Humoral Immune Function

Not surprisingly, the spectrum of infections associated with complement deficiencies overlaps substantially with that seen in patients with deficiencies in antibody production. Defects in the activation of C3, and in C3 itself, are associated with a wide range of pyogenic infections, emphasizing the important role of C3 as an opsonin, promoting the phagocytosis of bacteria. In contrast, defects in the membrane-attack components of complement (C5-C9) have more limited effects and result exclusively in susceptibility to *Neisseria* species. This indicates that host defense against these bacteria, which are capable of intracellular survival, is mediated by extracellular lysis by the membrane-attack complex of complement.

Another set of diseases are caused by defects in complement control proteins. People lacking decay-accelerating factor (DAF) and CD59, which protect a person's own cell surfaces from complement activation, destroy their own red blood cells. This results in the disease paroxysmal nocturnal hemoglobinuria. A more striking consequence of the loss of a regulatory protein is seen in patients with C1-inhibitor defects. C1-inhibitor irreversibly inhibits the activity of several serine proteinase enzymes. These include two enzymes that participate in the contact activation system, factor XIIa (activated Hageman factor) and kallikrein, in addition to the two enzymes that together initiate the classical pathway of complement, C1r and C1s. Deficiency of C1-inhibitor leads to failure to regulate these two pathways. Their unregulated activity results in the excessive production of vasoactive mediators that cause fluid accumulation in the tissues and epiglottal swelling that can lead to suffocation. These mediators are bradykinin, produced by the cleavage

of high molecular weight kininogen by kallikrein and the C2 kinin, produced by the activity of C1s on C2a. This syndrome is called hereditary angioneurotic edema.

10.1.4 Defects in Phagocytic Cells Permit Widespread Bacterial Infections

Defects in the recruitment of phagocytic cells to extravascular sites of infection can cause serious immunodeficiency. Leukocytes reach such sites by emigrating from blood vessels in a tightly regulated process consisting of three stages. The first is the rolling adherence of leukocytes to endothelial cells, through the binding of a fucosylated tetrasaccharide ligand known as sialyl-Lewis to E-selectin and P-selectin. Sialyl-Lewis^x is expressed on monocytes and neutrophils, whereas E-selectin and P-selectin are expressed on endothelium activated by mediators from the site of inflammation. The second stage is the tight adherence of the leukocytes to the endothelium through the binding of leukocyte $b\delta_2$ integrins such as CD11b:CD18 (Mac-1:CR3) to counterreceptors on endothelial cells. The third and final stage is the transmigration of leukocytes through the endothelium along gradients

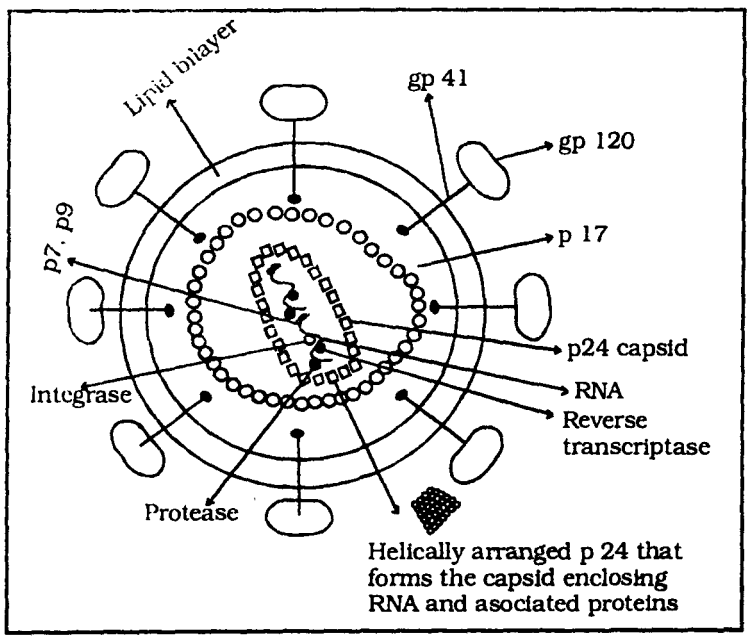


Figure 10.2 : Ultrastructure of HIV. HIV Diameter about 100 nm. Number of gp 120 molecules per virus : about 80. Virus lipid layer may also contain host HLA.

of chemotactic molecules originating from the site of tissue injury. Deficiencies in the molecules involved in each of these stages can prevent neutrophils and macrophages from reaching sites of infection to ingest and destroy bacteria.

Most of the other known defects in phagocytic cells affect their ability to kill intracellular and/or ingested extracellular bacteria. In chronic granulomatous disease, phagocytes cannot produce the superoxide radical and their antibacterial activity is thereby seriously impaired. Several different genetic defects, affecting any one of the four constituent proteins of the NADPH oxidase system, can cause this. Patients with this disease have chronic bacterial infections, which in some cases lead to the formation of granulomas. Deficiencies in the enzymes glucose-6-phosphate dehydrogenase and myeloperoxidase also impair intracellular killing and lead to a similar, although less severe, phenotype. Finally, in *Chediak-Higashi syndrome*, a complex syndrome characterized by partial albinism, abnormal platelet function, and severe immunodeficiency, a defect in a gene encoding a protein involved in intracellular vesicle formation causes a failure to fuse lysosomes properly with phagosomes; the phagocytes in these patients have enlarged granules and impaired intracellular killing.

10.1.5 Defects in T-cell Function Result in Severe Combined Immunodeficiencies

Although patients with B-cell defects can deal with many pathogens adequately, patients with defects in T-cell development are highly susceptible to a broad range of infectious agents. This demonstrates the central role of T cells in adaptive immune responses to virtually all antigens. As such patients make neither specific T-cell dependent antibody responses nor cell-mediated immune responses, and thus cannot develop immunological memory, they are said to suffer from *severe combined immunodeficiency (SCID)*.

Several different defects can lead to the SCID phenotype. In X-linked SCID, which is the commonest form of SCID, T cells fail to develop because of a mutation in the common gδ chain of several cytokine receptors, including those for the interleukins IL-2, IL-4, IL-7, IL-9, and IL-15. The commonest forms of autosomally inherited SCID are due to *adenosine deaminase (ADA) deficiency* and purine nucleotide phosphorylase *(PNP) deficiency*. These enzyme defects affect purine degradation, and both result in an accumulation of nucleotide metabolites that are particularly toxic to developing T cells. B cells are also somewhat compromised in these patients.

Finally, in patients with DiGeorge's syndrome the thymic epithelium fails to develop normally. Without the proper inductive environment T cells cannot mature, and both T-cell dependent antibody production and cell-mediated immunity are absent. Such patients have some serum immuno-globulin and variable numbers of B and T cells. As with all the severe combined immuo-deficiency diseases, it is the defect in T cells that is crucial. These diseases abundantly illustrate the central role of T cells in virtually all adaptive immune responses. In many cases B-cell development is normal, yet the response to nearly all pathogens is profoundly impaired.

X-linked lymphoproliferative syndrome is associated with fatal infection by Epstein-Barr virus and with the development of lymphomas.

Epstein-Barr virus is a herpes virus that infects the majority of the human race and remains latent in B cells throughout life after primary infection. EBV infection can transform B lymphocytes and is used as a technique for immortalizing clones of B cells in the laboratory. This does not normally happen *in vivo* in humans because EBV infection is actively controlled and maintained in a latent state by cytotoxic T cells with specificity for B cells expressing EBV antigens. In the presence of T-cell immunodeficiency, this control mechanism can break down and a potentially lethal B-cell lymphoma may develop. One of the situations in which this occurs is the rare immuo-deficiency, X-linked lymphoproliferative syndrome, which results from mutations in a gene named SH2-domain containing gene 1A (*SH2D1A*). Boys with this deficiency typically develop overwhelming EBV infection during childhood, and sometimes lymphomas. EBV infection in this condition is usually fatal and is associated with necrosis of the liver. Thus SH2D1A must play a vital, nonredundant role in the normal control of EBV infection.

10.1.6 Bone Marrow Transplantation or Gene Therapy can be Useful to Correct Genetic Defects

It is frequently possible to correct the defects in lymphocyte development that lead to the SCID phenotype by replacing the defective component, generally by bone marrow transplantation. The major difficulties in these therapies result from MHC polymorphism. To be useful, the graft must share some MHC alleles with the host. As we learned in Section 7-20, the MHC alleles expressed by the thymic epithelium determine which T cells can be positively selected. When bone marrow cells are used to restore immune function to individuals with a normal thymic stroma, both the T cells and the antigen-presenting

cells are derived from the graft. Therefore, unless the graft shares at least some MHC alleles with the recipient, the T cells that are selected on host thymic epithelium cannot be activated by graft-derived antigen-presenting cells. There is also a danger that mature, post-thymic T cells in donor bone marrow might recognize the host as foreign and attack it, causing *graft-versus-host disease (GVHD)*. This can be overcome by depleting the donor bone marrow of mature T cells. Bone marrow recipients are usually treated with irradiation that kills their own lymphocytes, thus making space for the grafted bone marrow cells and minimizing the threat of host-versus-graft disease *(HVGD)*. In patients with the SCID phenotype, however, there is little problem with the host response to the transplanted bone marrow, as the patient is immunodeficient.

10.2 ACQUIRED IMMUNE DEFICIENCY SYNDROME

The first cases of the acquired immune deficiency syndrome (**AIDS**) were reported in 1981 but it is now clear that cases of the disease had been occurring unrecognized for at least 4 years before its identification. The disease is characterized by a susceptibility to infection with opportunistic pathogens or by the occurrence of an aggressive form of Kaposi's sarcoma or B-cell lymphoma, accompanied by a profound decrease in the number of CD4 T cells. As it seemed to be spread by contact with body fluids, it was early suspected to be caused by a new virus, and by 1983 the agent now known to be responsible for AIDS, called the human immunodeficiency virus (**HIV**), was isolated and identified. It is now clear there are at least two types of HIV HIV-1 and HIV-2 which are closely related to each other. HIV-2 is endemic in West Africa and is now spreading in India. Most AIDS worldwide, however, is caused by the more virulent HIV-1. Both viruses appear to have spread to humans from other primate species and the best evidence from sequence relationships suggests that HIV-1 has passed to humans on at least three independent occasions from the chimpanzee, *Pan troglodytes*, and HIV-2 from the sooty mangabey, *Cercocebus atys*.

HIV infection does not immediately cause AIDS, and the issues of how it does, and whether all HIV-infected patients will progress to overt disease, remain controversial. Nevertheless, accumulating evidence clearly implicates the growth of the virus in CD4 T cells, and the immune response to it, as the central keys to the puzzle of AIDS. HIV is a worldwide pandemic and, although great strides are being made in understanding the pathogenesis and epidemiology of the disease, the

number of infected people around the world continues to grow at an alarming rate, presaging the death of many people from AIDS for many years to come. Estimates from the World Health Organization are that 16.3 million people have died from AIDS since the beginning of the epidemic and that there are currently around 34.3 million people alive with HIV infection, of whom the majority are living in sub-Saharan Africa, where approximately 7% of young adults are infected. In some countries within this region, such as Zimbabwe and Botswana, over 25% of adults are infected.

10.2.1 Most Individuals Infected with HIV Progress Over Time to AIDS

The initial infection with HIV generally occurs after transfer of body fluids from an infected person to an uninfected one. The virus is carried in infected CD4 T cells, dendritic cells, and macrophages, and as a free virus in blood, semen, vaginal fluid, or milk. It is most commonly spread by sexual intercourse, contaminated needles used for intravenous drug delivery, and the therapeutic use of infected blood or blood products, although this last route of transmission has largely been eliminated in the developed world where blood products are screened routinely for the presence of HIV. An important route of virus transmission is from an infected mother to her baby at birth or through breast milk. In Africa, the perinatal transmission rate is approximately 25%, but this can largely be prevented by treating infected pregnant women with the drug zidovudine (AZT). Mothers who are newly infected and breastfeed their infants transmit HIV 40% of the time, showing that HIV can also be transmitted in breast milk, but this is less common after the mother produces antibodies to HIV.

Primary infection with HIV is probably asymptomatic in 50% of cases but often causes an influenza-like illness with an abundance of virus in the peripheral blood and a marked drop in the numbers of circulating CD4 T cells. This acute viremia is associated in virtually all patients with the activation of CD8 T cells, which kill HIV-infected cells, and subsequently with antibody production, or *seroconversion.* The cytotoxic T-cell response is thought to be important in controlling virus levels, which peak and then decline, as the CD4 T-cell counts rebound to around 800 cells $m\delta l^{-1}$ (the normal value is 1200 cells mol^{-1}). At present, the best indicator of future disease is the level of virus that persists in the blood plasma once the symptoms of acute viremia have passed.

Most patients who are infected with HIV will eventually develop

AIDS, after a period of apparent quiescence of the disease known as clinical latency or the asymptomatic period. This period is not silent, however, for there is persistent replication of the virus, and a gradual decline in the function and numbers of CD4 T cells until eventually patients have few CD4 T cells left. At this point, which can occur anywhere between 2 and 15 years or more after the primary infection, the period of clinical latency ends and opportunistic infections begin to appear.

10.2.2 HIV is a Retrovirus that Infects CD4 T Cells, Dendritic Cells, and Macrophages

HIV is an enveloped retrovirus. Each virus particle, or virion, contains two copies of an RNA genome, which are transcribed into DNA in the infected cell and integrated into the host cell chromosome. The RNA transcripts produced from the integrated viral DNA serve both as mRNA to direct the synthesis of the viral proteins and later as the RNA genomes of new viral particles, which escape from the cell by budding from the plasma membrane, each in a membrane envelope. HIV belongs to a group of retroviruses called the lentiviruses, from the Latin *lentus*, meaning slow, because of the gradual course of the diseases that they cause. These viruses persist and continue to replicate for many years before causing overt signs of disease.

Drugs that block HIV replication lead to a rapid decrease in titer of infectious virus and an increase in CD4 T cells.

Studies with powerful drugs that completely block the cycle of HIV replication indicate that the virus is replicating rapidly at all phases of infection, including the asymptomatic phase. Two viral proteins in particular have been the target of drugs aimed at arresting viral replication. These are the viral reverse transcriptase, which is required for synthesis of the provirus, and the viral protease, which cleaves the viral polyproteins to produce the virion proteins and viral enzymes. Inhibitors of these enzymes prevent the establishment of further infection in uninfected cells. Cells that are already infected can continue to produce virions because, once the provirus is established, reverse transcriptase is not needed to make new virus particles, while the viral protease acts at a very late maturation step of the virus, and inhibition of the protease does not prevent virus from being released. However, in both cases, the released virions are not infectious and further cycles of infection and replication are prevented.

Most of the HIV present in the circulation of an infected individual is the product of rounds of replication in newly infected cells, and that

virus from these productively infected cells is released into, and rapidly cleared from, the circulation at the rate of 10^9 to 10^{10} virions every day. This raises the question of what is happening to these virus particles: how are they removed so rapidly from the circulation? It seems most likely that HIV particles are opsonized by specific antibody and complement and removed by phagocytic cells of the mononuclear phagocyte system. Opsonized HIV particles can also be trapped on the surface of follicular dendritic cells, which are known to capture antigen:antibody complexes and retain them for prolonged periods.

10.2.3 Lymphoid Tissue is the Major Reservoir of HIV Infection

Although viral load and turnover are usually measured by detecting the viral RNA present in viral particles in the blood, the major reservoir of HIV infection is in lymphoid tissue, in which infected CD4 T cells, monocytes, macrophages, and dendritic cells are found. In addition, HIV is trapped in the form of immune complexes on the surface of follicular dendritic cells. These cells are not themselves infected but may act as a store of infective virions.

HIV infection takes different forms within different cells. As we have seen, more than 95% of the virus that can be detected in the plasma is derived from productively infected cells, which have a very short half-life of about 2 days. Productively infected CD4 lymphocytes are found in the T-cell areas of lymphoid tissue, and these are thought to succumb to infection in the course of being activated in an immune response. Latently infected memory CD4 cells that are activated in response to antigen presentation also produce virus. Such cells have a longer half-life of 2 to 3 weeks from the time that they are infected. Once activated, HIV can spread from these cells by rounds of replication in other activated CD4 T cells. In addition to the cells that are infected productively or latently, there is a further large population of cells infected by defective proviruses; such cells are not a source of infectious virus.

Macrophages and dendritic cells seem to be able to harbor replicating virus without necessarily being killed by it, and are therefore believed to be an important reservoir of infection, as well as a means of spreading virus to other tissues such as the brain. Although the function of macrophages as antigen-presenting cells does not seem to be compromised by HIV infection, it is thought that the virus causes abnormal patterns of cytokine secretion that could account for the wasting that commonly occurs in AIDS patients late in their disease.

HIV infection leads to low levels of CD4 T cells, increased susceptibility to opportunistic infection, and eventually to death.

There are three dominant mechanisms for the loss of CD4 T cells in HIV infection. First, there is evidence for direct viral killing of infected cells; second, there is increased susceptibility to the induction of apoptosis in infected cells; and third, there is killing of infected CD4 T cells by CD8 cytotoxic lymphocytes that recognize viral peptides.

When CD4 T-cell numbers decline below a critical level, cell-mediated immunity is lost, and infections with a variety of opportunistic microbes appear. Typically, resistance is lost early to oral *Candida* species and to *Mycobacterium tuberculosis*, which shows as an increased prevalence of thrush (oral candidiasis) and tuberculosis. Later, patients suffer from shingles, caused by the activation of latent herpes zoster, from EBV-induced B-cell lymphomas, and from Kaposi's sarcoma, a tumor of endothelial cells that probably represents a response both to cytokines produced in the infection and to a novel herpes virus called HHV-8 that was identified in these lesions. Pneumonia caused by the fungus *Pneumocystis carinii* is common and often fatal. In the final stages of AIDS, infection with cytomegalovirus or *Mycobacterium avium* complex is more prominent. It is important to note that not all patients with AIDS get all these infections or tumors, and there are other tumors and infections that are less prominent but still significant. Rather, this is a list of the commonest opportunistic infections and tumors, most of which are normally controlled by robust CD4 T cell-mediated immunity that wanes as the CD4 T-cell counts drop toward zero.

10.2.4 Vaccination Against HIV is an Attractive Solution but Poses Many Difficulties

A safe and effective vaccine for the prevention of HIV infection and AIDS is an attractive goal, but its achievement is fraught with difficulties that have not been faced in developing vaccines against other diseases. The first problem is the nature of the infection itself, featuring a virus that proliferates extremely rapidly and causes sustained infection in the face of strong cytotoxic T-cell and antibody responses. HIV evolves in individual patients by the selective proliferative advantage of mutant virions encoding peptide sequence changes that escape recognition by antibodies and by cytotoxic T lymphocytes. This evolution means that the development of therapeutic vaccination strategies to block the development of AIDS in HIV-infected patients

will be extremely difficult. Even after the viremia has been largely cleared by drug therapy, immune responses to HIV fail to prevent drug-resistant virus from rebounding and replicating at pretreatment levels.

The second problem is our uncertainty over what form protective immunity to HIV might take. It is not known whether antibodies, cytotoxic T lymphocyte responses, or both are necessary to achieve protective immunity, and which epitopes might provide the targets of protective immunity. Third, if strong cytotoxic responses are necessary to provide protection against HIV, these might be difficult to develop and sustain through vaccination. Other effective viral vaccines rely on the use of live, attenuated viruses and there are concerns over the safety of pursuing this approach for HIV. Another possible approach is the use of DNA vaccination. Both of these approaches are being tested in animal models.

The fourth problem is the ability of the virus to persist in latent form as a transcriptionally silent provirus, which is invisible to the immune system. This might prevent the immune system from clearing the infection once it has been established. In summary, the ability of the immune system to clear infectious virus remains uncertain.

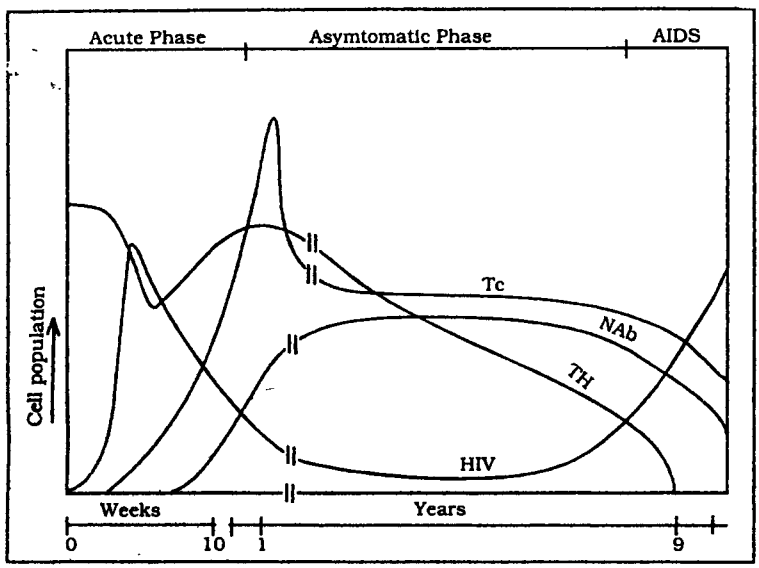


Figure 10.3 : Profile (in separate scales) of the population of Tc, TH, HJIV and NAb (neutralizing Abs) during progression of HIV infection.

However, against this pessimistic background, there are grounds for hope that successful vaccines can be developed. Of particular interest are rare groups of people who have been exposed often enough to HIV to make it virtually certain that they should have become infected but who have not developed the disease. In some cases this is due to an inherited deficiency in the chemokine receptor used as co-receptor for HIV entry. However, this mutant chemokine receptor does not occur in Africa, where one such group has been identified. A small group of Gambian and Kenyan prostitutes who are estimated to have been exposed to many HIV-infected male partners each month for up to 5 years were found to lack antibody responses but to have cytotoxic T lymphocyte responses to a variety of peptide epitopes from HIV. These women seem to have been naturally immunized against HIV.

Finally, there are difficult ethical issues in the development of a vaccine. It would be unethical to conduct a vaccine trial without trying at the same time to minimize the exposure of a vaccinated population to the virus itself. However, the effectiveness of a vaccine can only be assessed in a population in which the exposure rate to the virus is high enough to assess whether vaccination is protective against infection. This means that initial vaccine trials might have to be conducted in countries where the incidence of infection is very high and public health measures have not yet succeeded in reducing the spread of HIV.

10.2.5 Prevention and Education are One Way in which the Spread of HIV and AIDS can be Controlled

The one way in which we know we can protect against infection with HIV is by avoiding contact with body fluids, such as semen, blood, blood products, or milk from people who are infected. Indeed, it has been demonstrated repeatedly that this precaution, simple enough in the developed world, is sufficient to prevent infection, as health-care workers can take care of AIDS patients for long periods without seroconversion or signs of infection.

For this strategy to work, however, one must be able to test people at risk of infection with HIV periodically, so that they can take the steps necessary to avoid passing the virus to others. This, in turn, requires strict confidentiality and mutual trust. A barrier to the control of HIV is the reluctance of individuals to find out whether they are infected, especially as one of the consequences of a positive HIV test is stigmatization by society. As a result, infected individuals can unwittingly infect many others. Balanced against this is the success

of therapy with combinations of the new protease inhibitors and reverse transcriptase inhibitors, which provides an incentive for potentially infected people to identify the presence of infection and gain the benefits of treatment. Responsibility is at the heart of AIDS prevention, and a law guaranteeing the rights of people infected with HIV might go a long way to encouraging responsible behavior. The rights of HIV-infected people are protected in the Netherlands and Sweden. The problem in the less-developed nations, where elementary health precautions are extremely difficult to establish, is more profound.

11

Specific or Adaptive Immunity

The macrophages and neutrophils of the innate immune system provide a first line of defense against many common microorganisms and are essential for the control of common bacterial infections. However, they cannot always eliminate infectious organisms, and there are some pathogens that they cannot recognize. The lymphocytes of the adaptive immune system have evolved to provide a more versatile means of defense which, in addition, provides increased protection against subsequent reinfection with the same pathogen. There is a delay of 4-7 days before the initial adaptive immune response takes effect, the innate immune response has a critical role in controlling infections during this period.

11.1 MOST INFECTIOUS AGENTS INDUCE INFLAMMATORY RESPONSES BY ACTIVATING INNATE IMMUNITY

Microorganisms such as bacteria that penetrate the epithelial surfaces of the body for the first time are met immediately by cells and molecules that can mount an innate immune response. Phagocytic macrophages conduct the defense against bacteria by means of surface receptors that are able to recognize and bind common constituents of many bacterial surfaces. Bacterial molecules binding to these receptors trigger the macrophage to engulf the bacterium and also induce the secretion of biologically active molecules. Activated macrophages secrete cytokines, which are defined as proteins released by cells that affect the behavior of other cells that bear receptors for them. They also release proteins known as chemokines that attract cells with

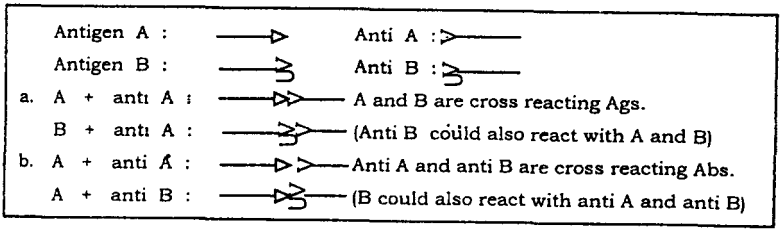


Figure 11.1 : Immunochemical cross reaction.

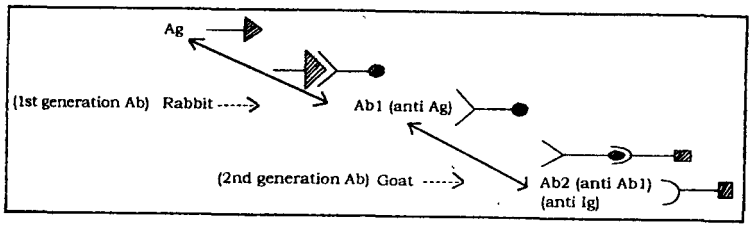


Figure 11.2 : First and second generartion antibodies Ab1 recognizes Ag and Ab2 recognizes Ab1. Ag and Ab2 are not complimentary.

chemokine receptors such as neutrophils and monocytes from the bloodstream. The cytokines and chemokines released by macrophages in response to bacterial constituents initiate the process known as inflammation. Local inflammation and the phagocytosis of invading bacteria may also be triggered as a result of the activation of complement on the bacterial cell surface. Complement is a system of plasma proteins that activates a cascade of proteolytic reactions on microbial surfaces but not on host cells, coating these surfaces with fragments that are recognized and bound by phagocytic receptors on macrophages. The cascade of reactions also releases small peptides that contribute to inflammation.

Inflammation is traditionally defined by the four Latin words *calor*, *dolor*, *rubor*, and *tumor*, meaning heat, pain, redness, and swelling, all of which reflect the effects of cytokines and other inflammatory mediators on the local blood vessels.

The innate immune response makes a crucial contribution to the activation of adaptive immunity. The inflammatory response increases the flow of lymph containing antigen and antigen-bearing cells into lymphoid tissue, while complement fragments on microbial surfaces and induced changes in cells that have taken up microorganisms provide signals that synergize in activating lymphocytes whose receptors bind to specific microbial antigens. Macrophages that have phagocytosed bacteria and become activated can also activate T lymphocytes. However, the

cells that specialize in presenting antigen to T lymphocytes and initiating adaptive immunity are the dendritic cells.

11.1.1 Activation of Specialized Antigen Presenting Cells: First Step for Induction of Adaptive Immunity

The induction of an adaptive immune response begins when a pathogen is ingested by an *immature dendritic cell* in the infected tissue. These specialized phagocytic cells are resident in most tissues and are relatively long-lived, turning over at a slow rate. They derive from the same bone marrow precursor as macrophages, and migrate from the bone marrow to their peripheral stations, where their role is to survey the local environment for pathogens. Eventually, all tissue-resident dendritic cells migrate through the lymph to the regional lymph nodes where they interact with recirculating naive lymphocytes. If the dendritic cells fail to be activated, they induce tolerance to the antigens of self that they bear.

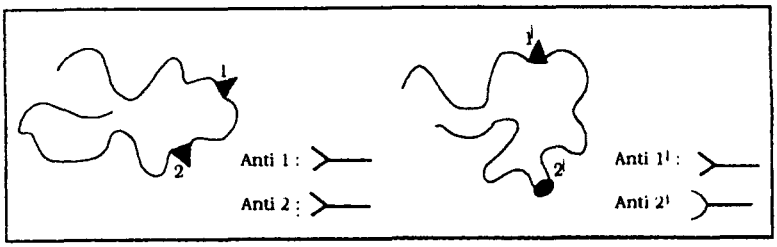


Figure 11.3 : Similar and dissimilar Ag determinants (a) Similar determinants; Abs raised against one would recognize the other (b) Dissimilar determinants : anti 1 and 2 posses different specificities.

The immature dendritic cell carries receptors on its surface that recognize common features of many pathogens, such as bacterial cell wall proteoglycans. As with macrophages and neutrophils, binding of a bacterium to these receptors stimulates the dendritic cell to engulf the pathogen and degrade it intracellularly. Immature dendritic cells are also continually taking up extracellular material, including any virus particles or bacteria that may be present, by the receptor-independent mechanism of macropinocytosis. The function of dendritic cells, however, is not primarily to destroy pathogens but to carry pathogen antigens to peripheral lymphoid organs and there present them to T lymphocytes. When a dendritic cell takes up a pathogen in infected tissue, it becomes activated, and travels to a nearby lymph node. On activation, the dendritic cell matures into a highly effective *antigen-presenting cell (APC)* and undergoes changes that enable it to activate pathogen-specific

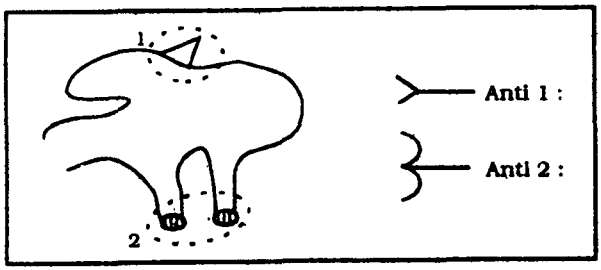


Figure 11.4 : Continuous and discontinous Ag determinants.

lymphocytes that it encounters in the lymph node. Activated dendritic cells secrete cytokines that influence both innate and adaptive immune responses, making these cells essential gatekeepers that determine whether and how the immune system responds to the presence of infectious agents We shall consider the maturation of dendritic cells and their central role in presenting antigens to T lymphocytes in Chapter elsewhere in this book.

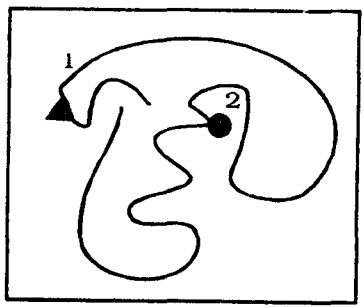


Figure 11.5 : External and Internal Ag determinants.

11.1.2 Lymphocytes Activated by Antigen Give Rise to Clones of Antigen-Specific Cells that Mediate Adaptive Immunity

The defense systems of innate immunity are effective in combating many pathogens. They are constrained, however, by relying on germline-encoded receptors to recognize microorganisms that can evolve more rapidly than the hosts they infect. This explains why they can only recognize microorganisms bearing surface molecules that are common to many pathogens and that have been conserved over the course of evolution. Not surprisingly, many pathogenic bacteria have evolved a protective capsule that enables them to conceal these molecules and thereby avoid being recognized and phagocytosed. Viruses carry no invariant molecules similar to those of bacteria and are rarely recognized

directly by macrophages. Viruses and encapsulated bacteria can, however, still be taken up by dendritic cells through the nonreceptor-dependent process of macropinocytosis. Molecules that reveal their infectious nature may then be unmasked, and the dendritic cell activated to present their antigens to lymphocytes. The recognition mechanism used by the lymphocytes of the adaptive immune response has evolved to overcome the constraints faced by the innate immune system, and enables recognition of an almost infinite diversity of antigens, so that each different pathogen can be targeted specifically.

Instead of bearing several different receptors, each recognizing a different surface feature shared by many pathogens, each naive lymphocyte entering the bloodstream bears antigen receptors of a single specificity. The specificity of these receptors is determined by a unique genetic mechanism that operates during lymphocyte development in the bone marrow and thymus to generate millions of different variants of the genes encoding the receptor molecules. Thus, although an individual lymphocyte carries receptors of only one specificity, the specificity of each lymphocyte is different. This ensures that the millions of lymphocytes in the body collectively carry millions of different antigen receptor specificities the lymphocyte receptor repertoire of the individual. During a person's lifetime these lymphocytes undergo a process akin to natural selection; only those lymphocytes that encounter an antigen to which their receptor binds will be activated to proliferate and differentiate into effector cells.

This selective mechanism was first proposed in the 1950s by *Macfarlane Burnet* to explain why antibodies, which can be induced in response to virtually any antigen, are produced in each individual only to those antigens to which he or she is exposed. He postulated the preexistence in the body of many different potential antibody-producing cells, each having the ability to make antibody of a different specificity and displaying on its surface a membrane-bound version of the antibody that served as a receptor for antigen. On binding antigen, the cell is activated to divide and produce many identical progeny, known as a clone; these cells can now secrete clonotypic antibodies with a specificity identical to that of the surface receptor that first triggered activation and clonal expansion. *Burnet* called this the clonal selection theory.

11.1.3 Clonal Selection of Lymphocytes is the Central Principle of Adaptive Immunity

Remarkably, at the time that **Burnet** formulated his theory, nothing

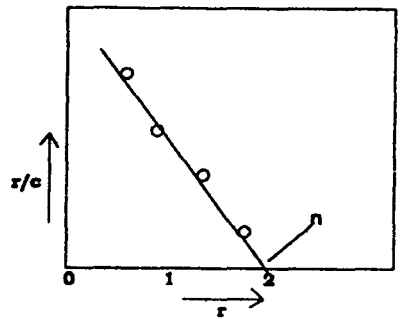


Figure 11.6 : Scatchard plot for an ag-Ab reaction whenthe Ab preparation is homogenous with respect to affinity.

was known of the antigen receptors of lymphocytes; indeed the function of lymphocytes themselves was still obscure. Lymphocytes did not take center stage until the early 1960s, when James Gowans discovered that removal of the small lymphocytes from rats resulted in the loss of all known adaptive immune responses. These immune responses were restored when the small lymphocytes were replaced. This led to the realization that lymphocytes must be the units of clonal selection, and their biology became the focus of the new field of cellular immunology.

Clonal selection of lymphocytes with diverse receptors elegantly explained adaptive immunity but it raised one significant intellectual problem. If the antigen receptors of lymphocytes are generated randomly during the lifetime of an individual, how are lymphocytes prevented from recognizing antigens on the tissues of the body and attacking them? *Ray Owen* had shown in the late 1940s that genetically different twin calves with a common placenta were immunologically *tolerant* of one another's tissues, that is, they did not make an immune response against each other. *Sir Peter Medawar* then showed in 1953 that if exposed to foreign tissues during embryonic development, mice become immunologically tolerant to these tissues. *Burnet* proposed that developing lymphocytes that are potentially self-reactive are removed before they can mature, a process known as clonal deletion. He has since been proved right in this too, although the mechanisms of tolerance are still being worked out.

Clonal selection of lymphocytes is the single most important principle in adaptive immunity. It has four basic postulates.. The last of the problems posed by the clonal selection theory how the diversity of lymphocyte antigen receptors is generated was solved in the 1970s when advances in molecular biology made it possible to clone the genes encoding antibody molecules.

11.1.4 The Structure of the Antibody Molecule in Relation to Adaptive Immunity

Antibodies, as discussed above, are the secreted form of the B-cell antigen receptor or BCR. Because they are produced in very large quantities in response to antigen, they can be studied by traditional biochemical techniques; indeed, their structure was understood long before recombinant DNA technology made it possible to study the membrane-bound antigen receptors of lymphocytes. The startling feature that emerged from the biochemical studies was that an antibody molecule is composed of two distinct regions. One is a constant region that can take one of only four or five biochemically distinguishable forms; the other is a variable region that can take an apparently infinite variety of subtly different forms that allow it to bind specifically to an equally vast variety of different antigens.

The two variable regions, which are identical in any one antibody molecule, determine the antigen-binding specificity of the antibody; the constant region determines how the antibody disposes of the pathogen once it is bound.

Each antibody molecule has a twofold axis of symmetry and is composed of two identical heavy chains and two identical light chains. Heavy and light chains both have variable and constant regions; the variable regions of a heavy and a light chain combine to form an antigen-binding site, so that both chains contribute to the antigen-binding specificity of the antibody molecule.

11.1.5 Each Developing Lymphocyte Generates a Unique Antigen Receptor by Rearranging its Receptor Genes

How are antigen receptors with an almost infinite range of specificities encoded by a finite number of genes? The genes for immunoglobulin variable regions are inherited as sets of gene segments, each encoding a part of the variable region of one of the immunoglobulin polypeptide chains. During B-cell development in the bone marrow, these gene segments are irreversibly joined by DNA recombination to form a stretch of DNA encoding a complete variable region. Because there are many different gene segments in each set, and different gene segments are joined together in different cells, each cell generates unique genes for the variable regions of the heavy and light chains of the immunoglobulin molecule. Once these recombination events have succeeded in producing a functional receptor, further rearrangement is prohibited. Thus each lymphocyte expresses only one receptor specificity.

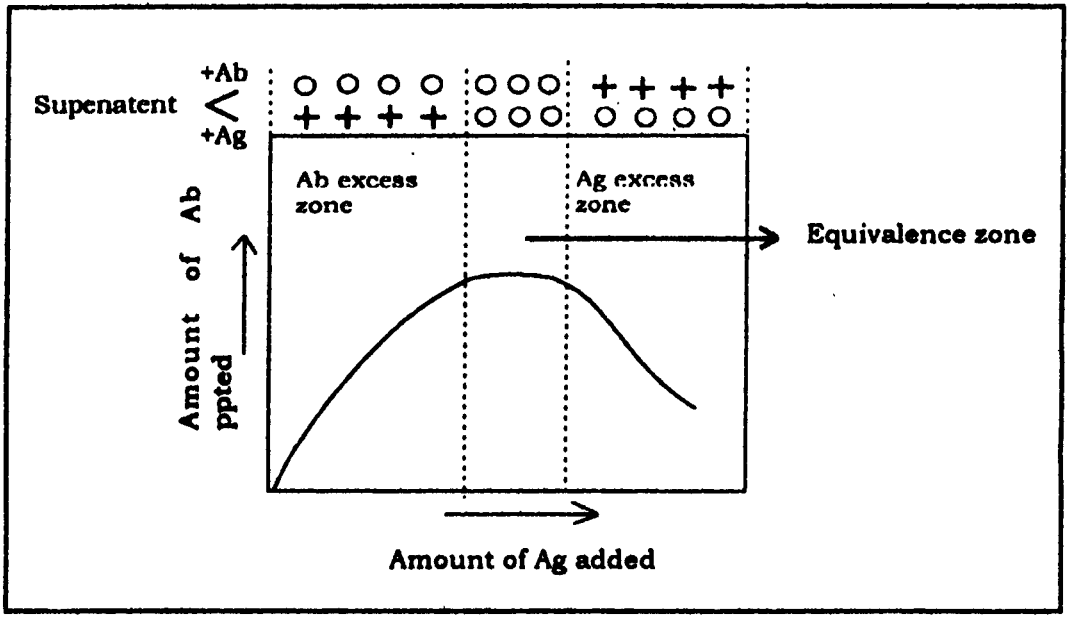


Figure 11.7 : Precipitin curve of a macromolecular Ag and its Ab. Pattern of precipitate obtained for a monospecific reaction when supernatants in each zone are mixed with either Ab or Ag is given at the top of the figure.

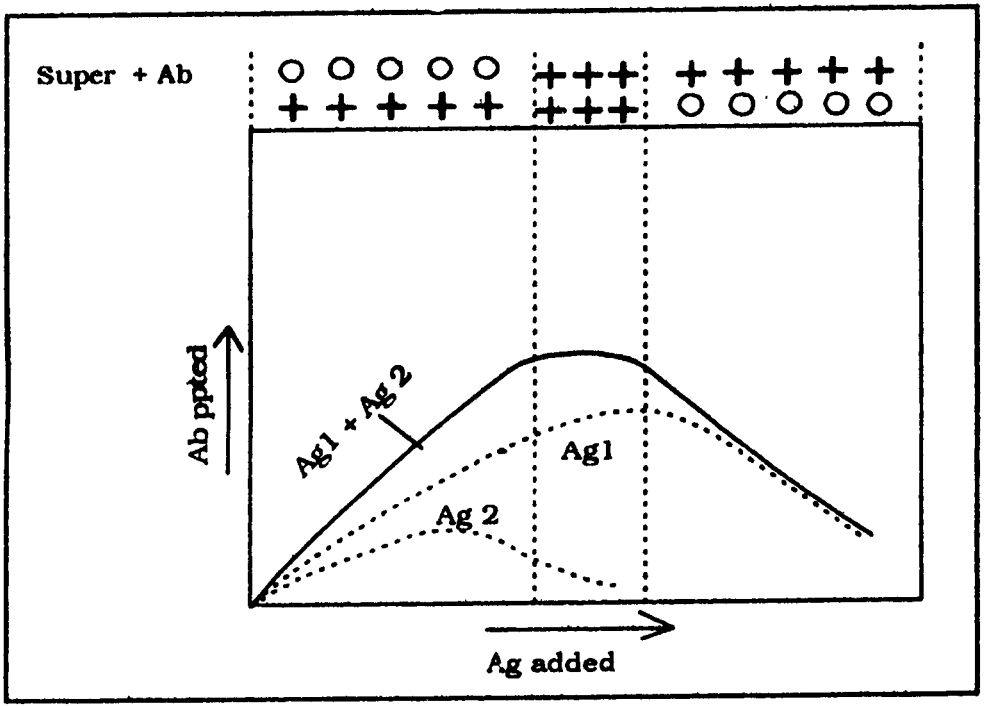


Figure 11.8 : A multispecific precipitin reaction. The top curve is the sum of individual precipitin reactions taking place in the system. Individual precipitin reactions are shown by dotted lines. The top curve would be obtained in teh actual experiment. Pattern of precipition obtained when supernatents in each zone are mixed with either heterogenous Ab or Ag is shown at the top of the figure.

This mechanism has three important consequences. First, it enables a limited number of gene segments to generate a vast number of different proteins. Second, because each cell assembles a different set of gene segments, each cell expresses a unique receptor specificity. Third, because gene rearrangement involves an irreversible change in a cell's DNA, all the progeny of that cell will inherit genes encoding the same receptor specificity. This general scheme was later also confirmed for the genes encoding the antigen receptor on T lymphocytes. The main distinctions between B- and T-lymphocyte receptors are that the immunoglobulin that serves as the B-cell antigen receptor has two identical antigen-recognition sites and can also be secreted, whereas the T-cell antigen receptor has a single antigen-recognition site and is always a cell-surface molecule. We shall see later that these receptors also recognize antigen in very different ways.

The potential diversity of lymphocyte receptors generated in this way is enormous. Just a few hundred different gene segments can combine in different ways to generate thousands of different receptor chains. The diversity of lymphocyte receptors is further amplified by junctional diversity, created by adding or subtracting nucleotides in the process of joining the gene segments, and by the fact that each receptor is made by pairing two different variable chains, each encoded in distinct sets of gene segments. A thousand different chains of each type could thus generate 10^6 distinct antigen receptors through this combinatorial diversity. Thus a small amount of genetic material can encode a truly staggering diversity of receptors. Only a subset of these randomly generated receptor specificities survive the selective processes that shape the peripheral lymphocyte repertoire; nevertheless, there are lymphocytes of at least 10^8 different specificities in an individual at any one time. These provide the raw material on which clonal selection acts.

11.2 LYMPHOCYTE DEVELOPMENT AND SURVIVAL

Equally amazing as the generation of millions of specificities of lymphocyte antigen receptors is the shaping of this repertoire during lymphocyte development and the homeostatic maintenance of such an extensive repertoire in the periphery. How are the most useful receptor specificities selected, and how are the numbers of peripheral lymphocytes, and the percentages of B cells and T cells kept relatively constant? The answer seems to be that lymphocyte maturation and survival are regulated by signals received through their antigen receptors. Strong signals

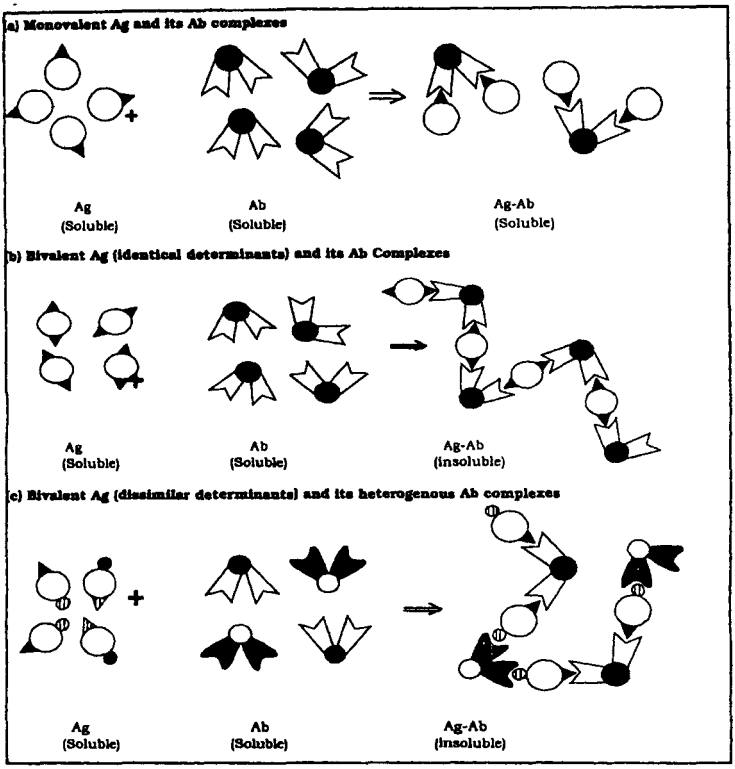


Figure 11.9 : Soluble and insoluble Ag-Ab complexes : Minimum Ag valency required for insoluble Ag-Ab complex formation is two which makes the Ag-Ab crosslinking possible.

received through the antigen receptor by an immature lymphocyte cause it to die or undergo further receptor rearrangement, and in this way self-reactive receptor specificities are deleted from the repertoire. However, a complete absence of signals from the antigen receptor can also lead to cell death. It seems that in order to survive, lymphocytes must periodically receive certain signals from their environment via their antigen receptors. In this way, the body can ensure that each receptor is functional and regulate the number and type of lymphocytes in the population at any given time. These survival signals appear to be delivered by other cells in the lymphoid organs and must derive, at least in part, from the body's own molecules, the self antigens, as altering the self environment alters the life-span of lymphocytes in that environment. Developing B cells in the bone marrow interact with stromal cells, while their final maturation and continued recirculation

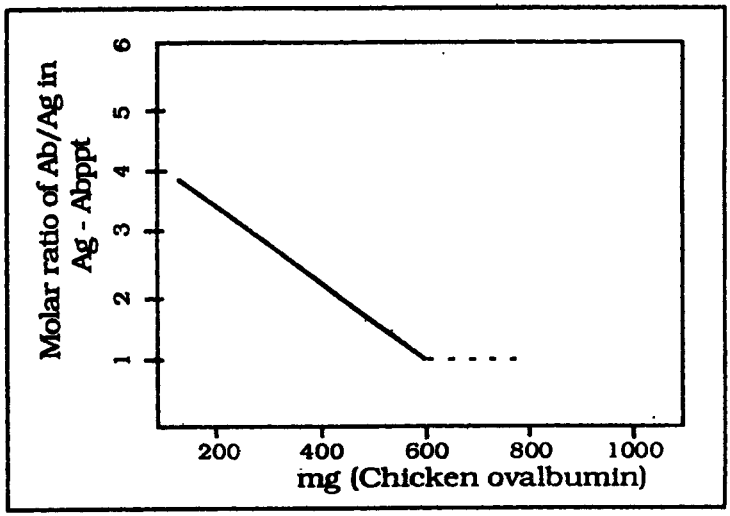


Figure 11.10 : Dependence of composition of Ab-Ab precipitate on Ag (Chicken ovalbumin) concentration.

appears to depend on survival signals received from the B-cell follicles of peripheral lymphoid tissue. T lymphocytes receive survival signals from self molecules on specialized epithelial cells in the thymus during development, and from the same molecules expressed by dendritic cells in the lymphoid tissues in the periphery. The self ligands that interact with the T-cell receptor to deliver these signals are partially defined, being composed of known cell-surface molecules complexed with undefined peptides from other self proteins in the cell. These same cell-surface molecules function to present foreign intracellular antigens to T cells. They select only a subset of T-cell receptors for survival, but these are the receptors most likely to be useful in responding to foreign antigens.

Lymphocytes that fail to receive survival signals, and those that are clonally deleted because they are self-reactive, undergo a form of cell suicide called apoptosis or programmed cell death. Apoptosis, derived from a Greek word meaning the falling of leaves from the trees, occurs in all tissues, at a relatively constant rate in each tissue, and is a means of regulating the number of cells in the body. It is responsible, for example, for the death and shedding of skin cells, the turnover of liver cells, and the death of the oldest intestinal epithelial cells that are constantly replaced by new cells. Thus, it should come as no surprise that immune system cells are regulated through the same mechanism. Each day the bone marrow produces many millions

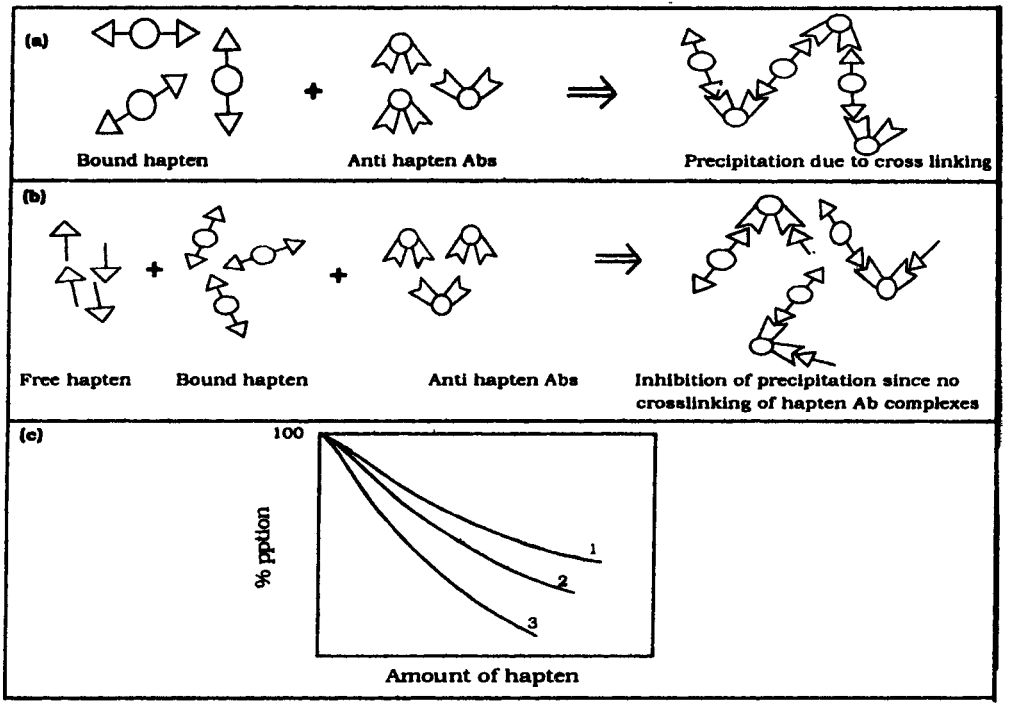


Figure 11.11 : Hapten inhibition of precipitation. Under standard conditions in which amount of bound hapten and antihapten Abs is kept constant, the degree of precipitation is inversely proportional to the amount of free hapten added and can be used to measure amount of hapten is unknown sample.

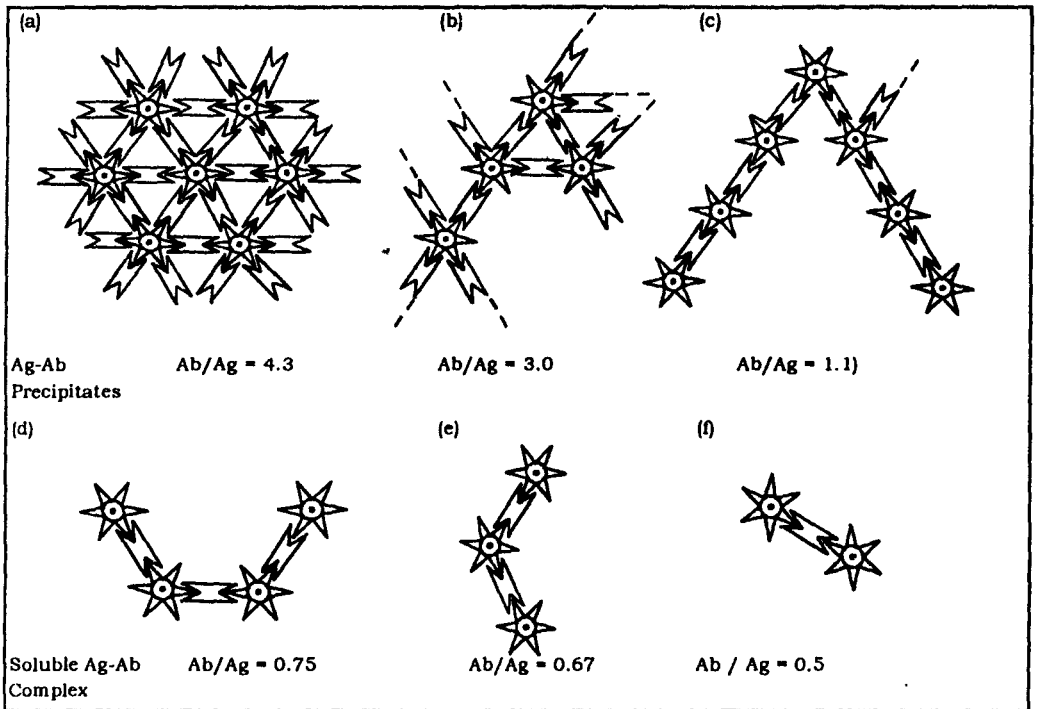


Figure 11.12 : Lattice theory of formation of Ag-Ab complexes of macromolecular soluble Ags.

of new neutrophils, monocytes, red blood cells, and lymphocytes, and this production must be balanced by an equal loss of these cells. Regulated loss of all these blood cells occurs by apoptosis, and the dying cells are finally phagocytosed by specialized macrophages in the liver and spleen. Lymphocytes are a special case, because the loss of an individual naive lymphocyte means the loss of a receptor specificity from the repertoire, while each newly matured cell that survives will contribute a different specificity. The survival signals received through the antigen receptors appear to regulate this process by inhibiting the apoptosis of individual lymphocytes, thus regulating the maintenance and composition of the lymphocyte repertoire.

11.2.1 Lymphocytes Proliferate in Response to Antigen in Peripheral Lymphoid Organs, Generating Effector Cells and Immunological Memory

The large diversity of lymphocyte receptors means that there will usually be at least a few that can bind to any given foreign antigen. However, because each lymphocyte has a different receptor, the numbers of lymphocytes that can bind and respond to any given antigen is very small. To generate sufficient antigen-specific effector lymphocytes to fight an infection, a lymphocyte with an appropriate receptor specificity must be activated to proliferate before its progeny finally differentiate into effector cells. This clonal expansion is a feature common to all adaptive immune responses.

Lymphocyte activation and proliferation is initiated in the draining lymphoid tissues, where naive lymphocytes and activated antigen-presenting cells can come together. Antigens are thus presented to the naive recirculating lymphocytes as they migrate through the lymphoid tissue before returning to the bloodstream via the efferent lymph. On recognizing its specific antigen, a small lymphocyte stops migrating and enlarges. The chromatin in its nucleus becomes less dense, nucleoli appear, the volume of both the nucleus and the cytoplasm increases, and new RNAs and proteins are synthesized. Within a few hours, the cell looks completely different and is known as a lymphoblast.

The lymphoblasts now begin to divide, normally duplicating themselves two to four times every 24 hours for 3 to 5 days, so that one naive lymphocyte gives rise to a clone of around 1000 daughter cells of identical specificity. These then differentiate into effector cells. In the case of B cells, the differentiated effector cells, the plasma cells, secrete antibody; in the case of T cells, the effector cells are able to destroy infected cells or activate other cells of the immune system.

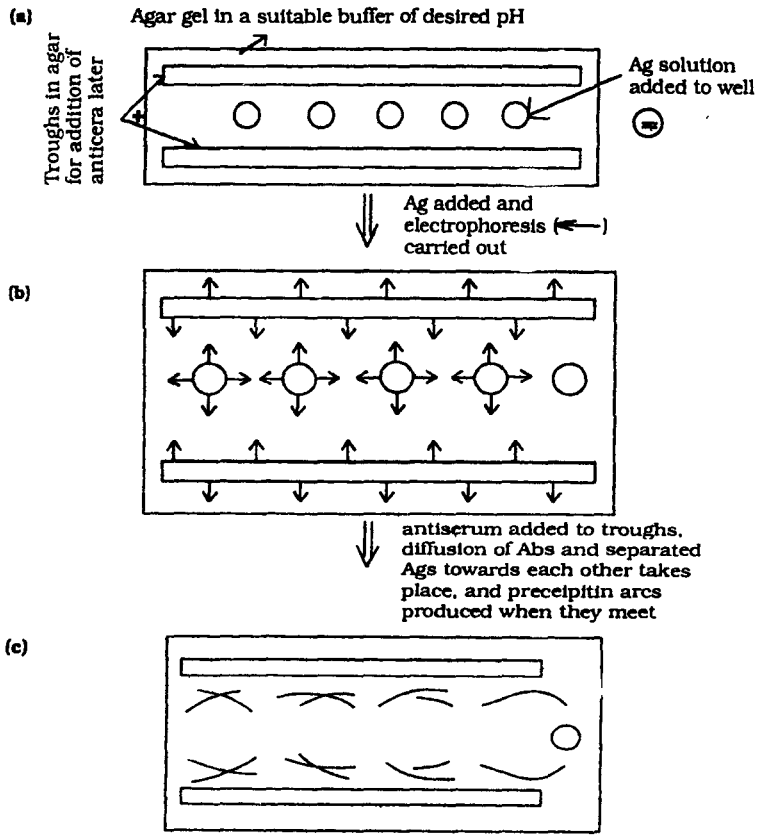


Figure 11.13 : Immuno electrophoresis.

These changes also affect the recirculation of antigen-specific lymphocytes. Changes in the cell-adhesion molecules they express on their surface allow effector lymphocytes to migrate into sites of infection or stay in the lymphoid organs to activate B cells.

After a naive lymphocyte has been activated, it takes 4 to 5 days before clonal expansion is complete and the lymphocytes have differentiated into effector cells. That is why adaptive immune responses occur only after a delay of several days. Effector cells have only a limited life-span and, once antigen is removed, most of the antigen-specific cells generated by the clonal expansion of small lymphocytes undergo apoptosis. However, some persist after the antigen has been eliminated. These cells are known as memory cells and form the basis of immunological memory, which ensures a more rapid and

effective response on a second encounter with a pathogen and thereby provides lasting protective immunity.

The characteristics of immunological memory are readily observed by comparing the antibody response of an individual to a first or *primary immunization* with the response elicited in the same individual by a *secondary* or booster immunization with the same antigen. The secondary antibody response occurs after a shorter lag phase, achieves a markedly higher level, and produces antibodies of higher affinity, or strength of binding, for the antigen. The cellular basis of immunological memory is the clonal expansion and clonal differentiation of cells specific for the eliciting antigen, and it is therefore entirely antigen specific.

It is immunological memory that enables successful vaccination and prevents reinfection with pathogens that have been repelled successfully by an adaptive immune response. Immunological memory is the most important biological consequence of the development of adaptive immunity, although its cellular and molecular basis is still not fully understood

11.2.2 Interaction with Other Cells as well as with Antigen is Necessary for Lymphocyte Activation

Peripheral lymphoid tissues are specialized not only to trap phagocytic cells that have ingested antigen but also to promote their interactions with lymphocytes that are needed to initiate an adaptive immune response. The spleen and lymph nodes in particular are highly organized for the latter function.

All lymphocyte responses to antigen require not only the signal that results from antigen binding to their receptors, but also a second signal, which is delivered by another cell. Naive T cells are generally activated by activated dendritic cells but for B cells, the second signal is delivered by an armed effector T cell. Because of their ability to deliver activating signals, these three cell types are known as professional antigen-presenting cells, or often just antigen-presenting cells. Dendritic cells are the most important antigenpresenting cell of the three, with a central role in the initiation of adaptive immune responses. Macrophages can also mediate innate immune responses directly and make a crucial contribution to the effector phase of the adaptive immune response. B cells contribute to adaptive immunity by presenting peptides from antigens they have ingested and by secreting antibody.

Thus, the final postulate of adaptive immunity is that it occurs on

a cell that also presents the antigen. This appears to be an absolute rule *in vivo*, although exceptions have been observed in *in vitro* systems. Nevertheless, what we are attempting to define is what does happen, not what can happen.

12

Cells Involved in Immunity

The cells responsible for the immune system originate in the bone marrow, where many of them also mature. Immune responses are mediated by leukocytes, which derive from precursors in the bone marrow. They then migrate to guard the peripheral tissues, circulating in the blood and in a specialized system of vessels called the lymphatic system.

12.1 ORIGIN OF WHITE BLOOD CELLS

The cellular elements of blood, including the red blood cells that transport oxygen, the platelets that trigger blood clotting in damaged tissues, and the white blood cells of the immune system, derive ultimately from the same progenitor or precursor cells the hematopoietic stem cells in the bone marrow. They are often known as pluripotent hematopoietic stem cells. Initially, they give rise to stem cells of more limited potential, which are the immediate progenitors of red blood cells, platelets, and the two main categories of white blood cells. The present chapter covers all the cells derived from the common lymphoid progenitor and the myeloid progenitor, apart from the megakaryocytes and red blood cells.

12.2 THE IMMUNE CELLS

12.2.1 Myeloid Progenitor

The myeloid progenitor is the precursor of the granulocytes, macrophages, dendritic cells, and mast cells of the immune system.

Macrophages are one of the three types of phagocyte in the immune

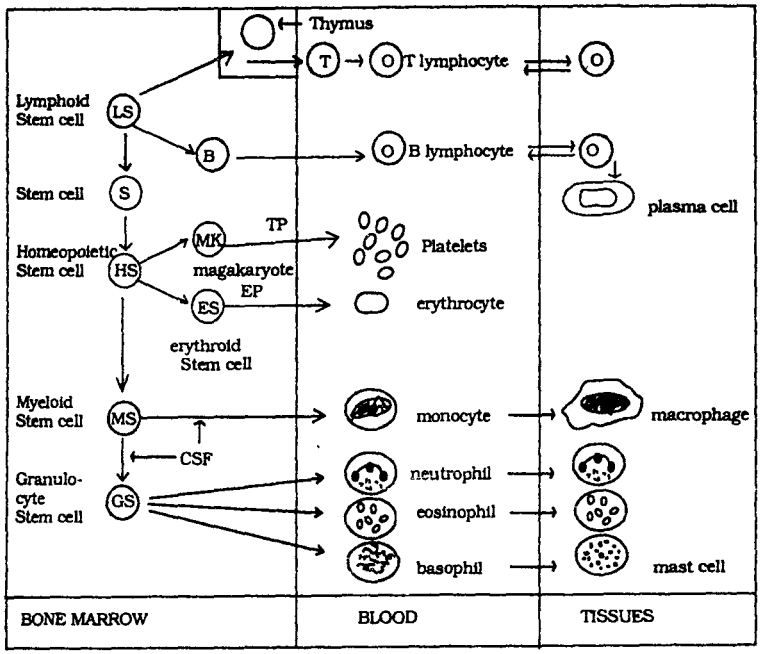


Figure 12.1 : Generation of immunological cells from stem cells in mammals TP thrombopoeitin; EP erythropoietin; CSF: colony stimulating factor.

system and are distributed widely in the body tissues, where they play a critical part in innate immunity. They are the mature form of monocytes, which circulate in the blood and differentiate continuously into macrophages upon migration into the tissues.

Dendritic cells are specialized to take up antigen and display it for recognition by lymphocytes. Immature dendritic cells migrate from the blood to reside in the tissues and are both phagocytic and macropinocytic, ingesting large amounts of the surrounding extracellular fluid. Upon encountering a pathogen, they rapidly mature and migrate to lymph nodes.

Mast cells, whose blood-borne precursors are not well defined, also differentiate in the tissues. They mainly reside near small blood vessels and, when activated, release substances that affect vascular permeability. Although best known for their role in orchestrating allergic responses, they are believed to play a part in protecting mucosal surfaces against pathogens.

The *granulocytes* are so called because they have densely staining granules in their cytoplasm; they are also sometimes called *poly-*

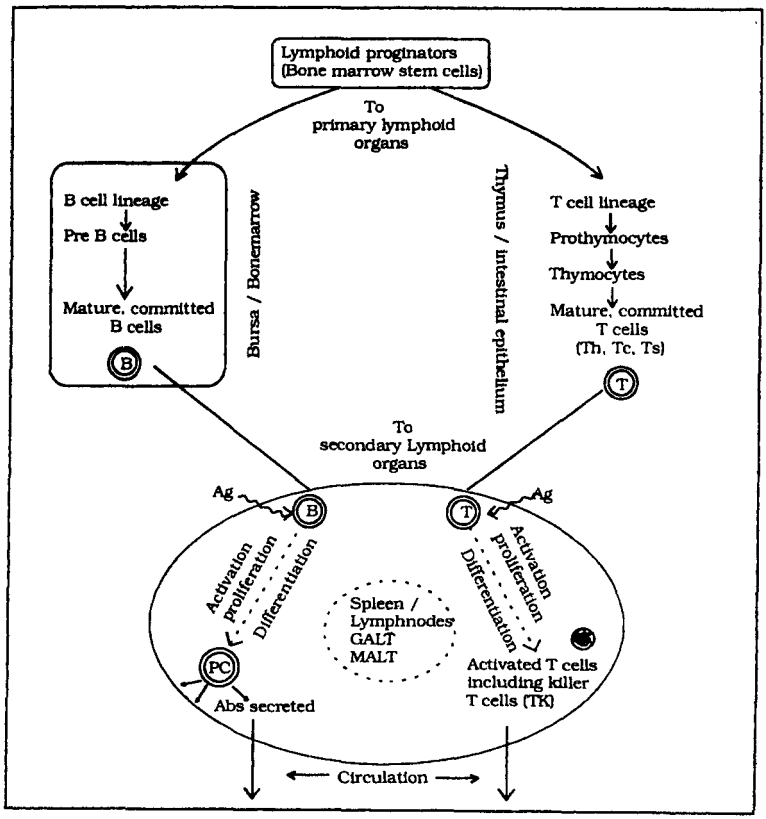


Figure 12.2 : Differentiation of B and T-cell proginators. Th: helper T-cells; cytotoxic T-cells; Ts: Suppressor T-cells; Pc: Plasma cells.

morphonuclear leukocytes because of their oddly shaped nuclei. There are three types of granulocyte, all of which are relatively short lived and are produced in increased numbers during immune responses, when they leave the blood to migrate to sites of infection or inflammation.

Neutrophils, which are the third phagocytic cell of the immune system, are the most numerous and most important cellular component of the innate immune response: hereditary deficiencies in neutrophil function lead to overwhelming bacterial infection, which is fatal if untreated.

Eosinophils are thought to be important chiefly in defense against parasitic infections, because their numbers increase during a parasitic infection.

The function of *basophils* is probably similar and complementary

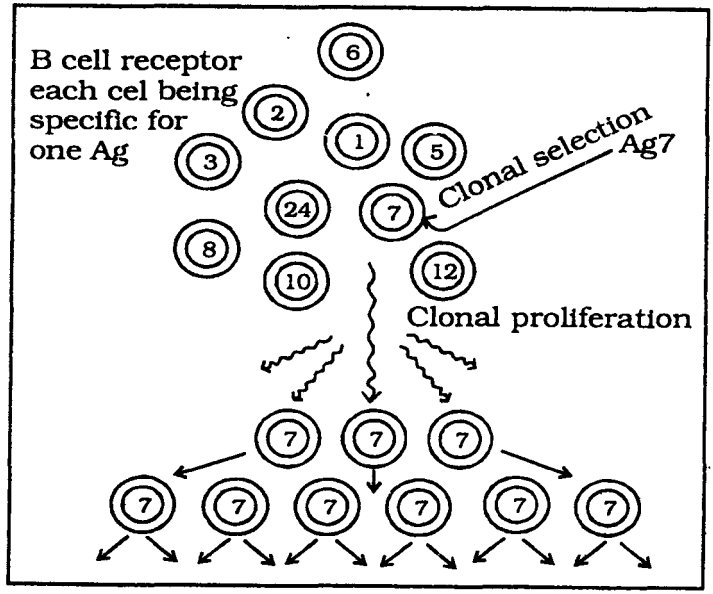


Figure 12.3 : Burnet's clonal selection theory.

to that of eosinophils and mast cells.

The common lymphoid progenitor gives rise to the lymphocytes. There are two major types of lymphocyte:

1. B lymphocytes or **B cells**, which when activated differentiate into plasma cells that secrete antibodies; and

2. T lymphocytes or **T cells**, of which there are two main classes. One class differentiates on activation into cytotoxic T cells, which kill cells infected with viruses, whereas the second class of T cells differentiates into cells that activate other cells such as B cells and macrophages.

Most lymphocytes are small, featureless cells with few cytoplasmic organelles and much of the nuclear chromatin inactive. This appearance is typical of inactive cells and indeed, these small lymphocytes have no functional activity until they encounter antigen, which is necessary to trigger their proliferation and the differentiation of their specialized functional characteristics.

Lymphocytes are remarkable in being able to mount a specific immune response against virtually any foreign antigen. This is possible because each individual lymphocyte matures bearing a unique variant of a prototype antigen receptor, so that the population of T and B

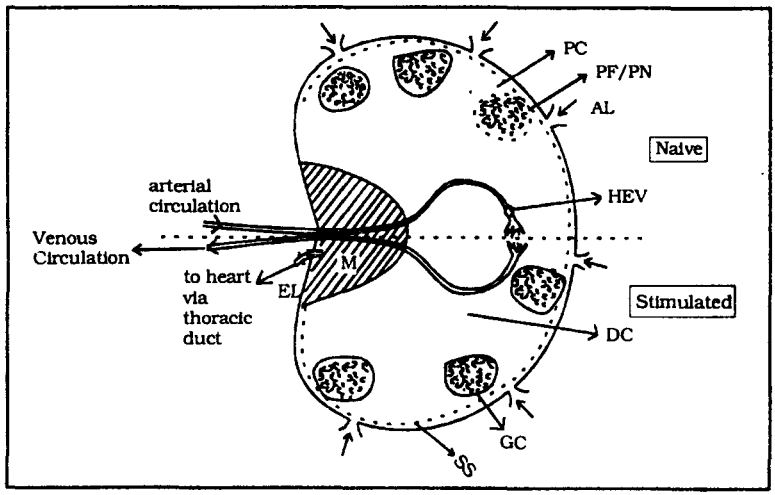


Figure 12.4 : The lymph node: Pc: Primary cortex contains (primary nodule PN/primary follicle PF); Gc: Germinal center (contains secondary nodules SN/secondary follicle); Al: afferent lymphatic; EL: efferent lymphatic; HEV: High endothelial venule; M: Medulla; DC: Difluse cortex (para cortex); SS: Subcapsular sinus.

lymphocytes collectively bears a huge repertoire of receptors that are highly diverse in their antigen-binding sites.

The **B-cell antigen receptor (BCR)** is a membrane-bound form of the antibody that the B cell will secrete after activation and differentiation to plasma cells. Antibody molecules as a class are known as **immunoglobulins**, usually shortened to **Ig**, and the antigen receptor of B lymphocytes is therefore also known as membrane immunoglobulin

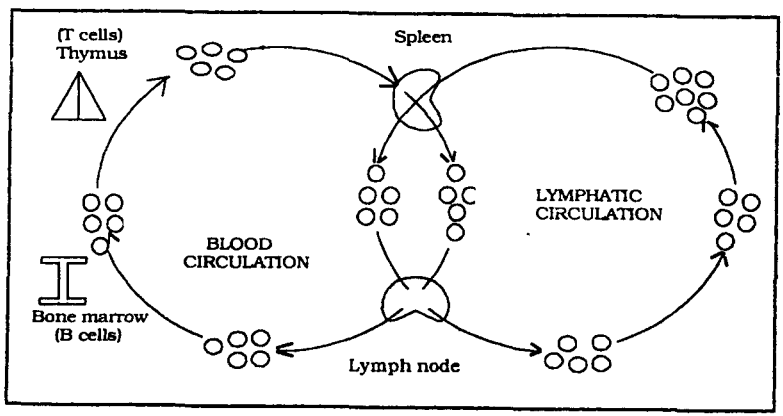


Figure 12.5 : Lumphocyte circulation.

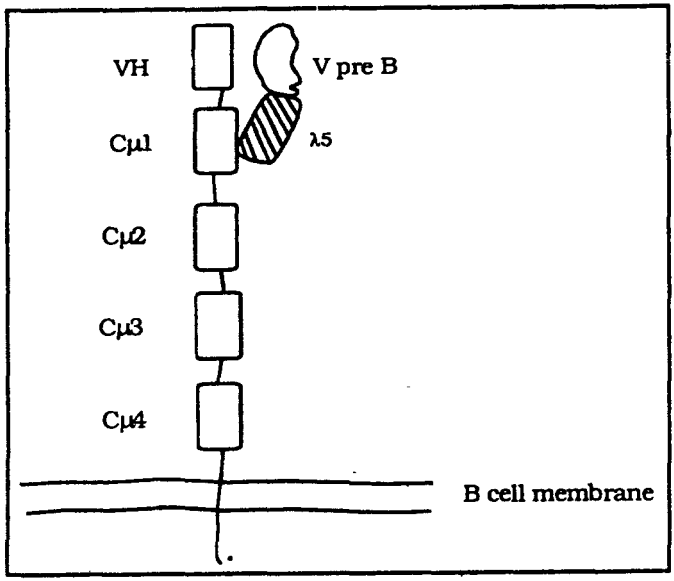


Figure 12.6 : A possible composition of the surrogate B-cell receptor. λ5 is covalently linked with cμ1, λt and V pre-B are non covalently linked.

(mIg). The *T-cell antigen receptor (TCR)* is related to immunoglobulin but is quite distinct from it, as it is specially adapted to detect antigens derived from foreign proteins or pathogens that have entered into host cells.

A third lineage of lymphoid cells, called *natural killer cells*, lack antigenspecific receptors and are part of the innate immune system. These cells circulate in the blood as large lymphocytes with distinctive cytotoxic granules. They are able to recognize and kill some abnormal cells, for example some tumor cells and virus-infected cells, and are thought to be important in the innate immune defense against intracellular pathogens.

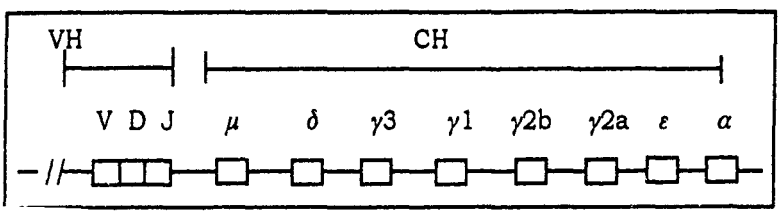


Figure 12.7 : Primary RNA transcript of murine Heavy chain (H) gene. Posttranscriptional splicing results in formation of a gene for a given heavy chain type. Note that each CH is in turn a split gene at the DNA level.

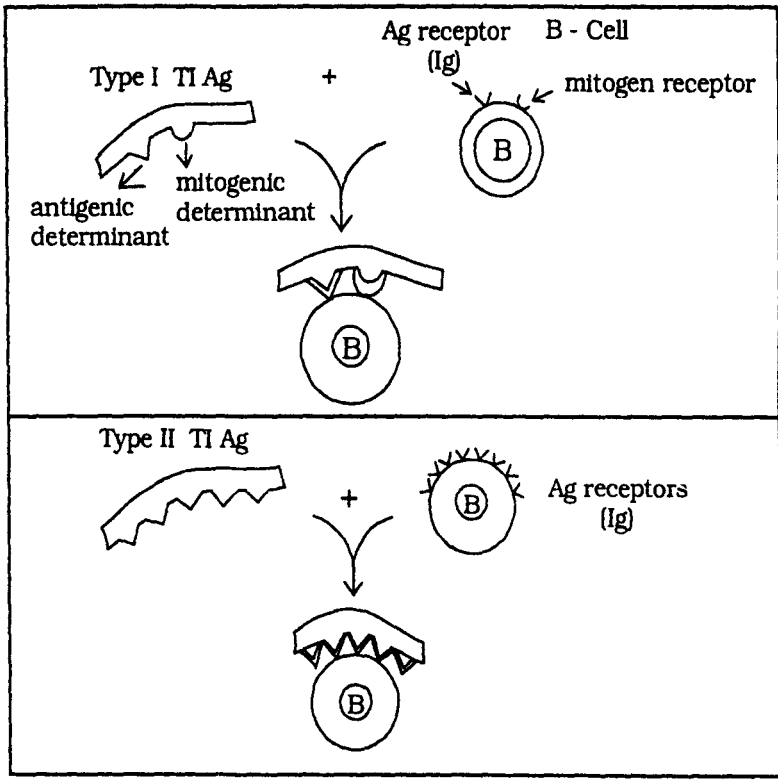


Figure 12.8 : T-cell independence antigens (TI Ags)

12.2.2 The Maturation of Lymphocytes

The lymphoid organs are organized tissues containing large numbers of lymphocytes in a framework of non-lymphoid cells. In these organs, the interactions lymphocytes make with non-lymphoid cells are important either to lymphocyte development, to the initiation of adaptive immune responses, or to the sustenance of lymphocytes. Lymphoid organs can be divided broadly into *central or primary lymphoid organs*, where lymphocytes are generated, and *peripheral or secondary lymphoid organs*, where adaptive immune responses are initiated and where lymphocytes are maintained. The central lymphoid organs are the bone marrow and the thymus, a large organ in the upper chest.

Both B and T lymphocytes originate in the bone marrow but only B lymphocytes mature there; T lymphocytes migrate to the thymus to undergo their maturation. Thus B lymphocytes are so-called because they are bone marrow derived, and T lymphocytes because they are

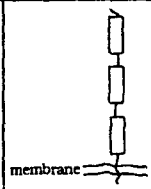
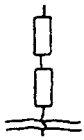
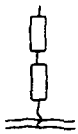
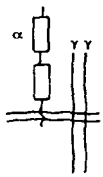
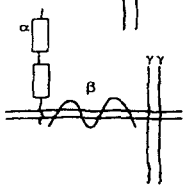
FCR type	Structure	Affinity	Cellular expression
FcγRI		High : for monomeric IgG Higher : for aggregated IgG	monocytes, mφ, neutrophils
FcγRII		Low : mostly binds aggregated IgG Two isoforms (A and B) identified. FcγRIIB is present on B-cells FcγRIIA is present on other cells.	neutrophils, eosinophils, platelets B-cells.
FcγRIIIB		Low : binds aggregated IgG.	neutrophils no cytoplasmic tail.
		Low : binds aggregated IgG. γ chain is homologous with TCR γ and FCεRI γ chain.	monocytes mφ, NK cells
FCεRI		Very High	mast-cells

Figure 12.9 : The Fc receptors (FcRs).

thymus derived. Once they have completed their maturation, both types of lymphocyte enter the bloodstream, from which they migrate to the peripheral lymphoid organs.

12.2.3 Trapping of Antigen for the Initiation of Adaptive Immune Responses

Pathogens can enter the body by many routes and set up an infection anywhere, but antigen and lymphocytes will eventually encounter each other in the peripheral lymphoid organs the lymph nodes, the spleen, and the mucosal lymphoid tissues. Lymphocytes are continually recirculating through these tissues, to which antigen is also carried

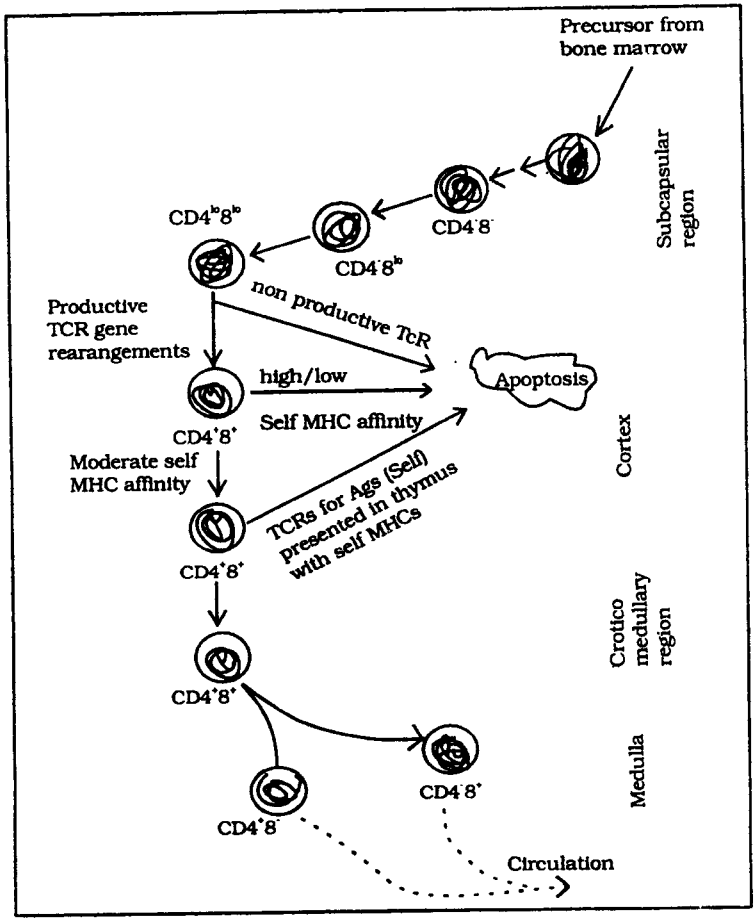


Figure 12.10 : Generation of mature $CD4^+$ and CDS^+. T-cells in the thymus.

from sites of infection, primarily within macrophages and dendritic cells. Within the lymphoid organs, specialized cells such as mature dendritic cells display the antigen to lymphocytes.

The lymph nodes are highly organized lymphoid structures located at the points of convergence of vessels of the lymphatic system, an extensive system of vessels that collects extracellular fluid from the tissues and returns it to the blood. This extracellular fluid is produced continuously by filtration from the blood, and is called lymph. The vessels are lymphatic vessels or lymphatics. Afferent lymphatic vessels drain fluid from the tissues and also carry antigen-bearing cells and antigens from infected tissues to the lymph nodes, where they are

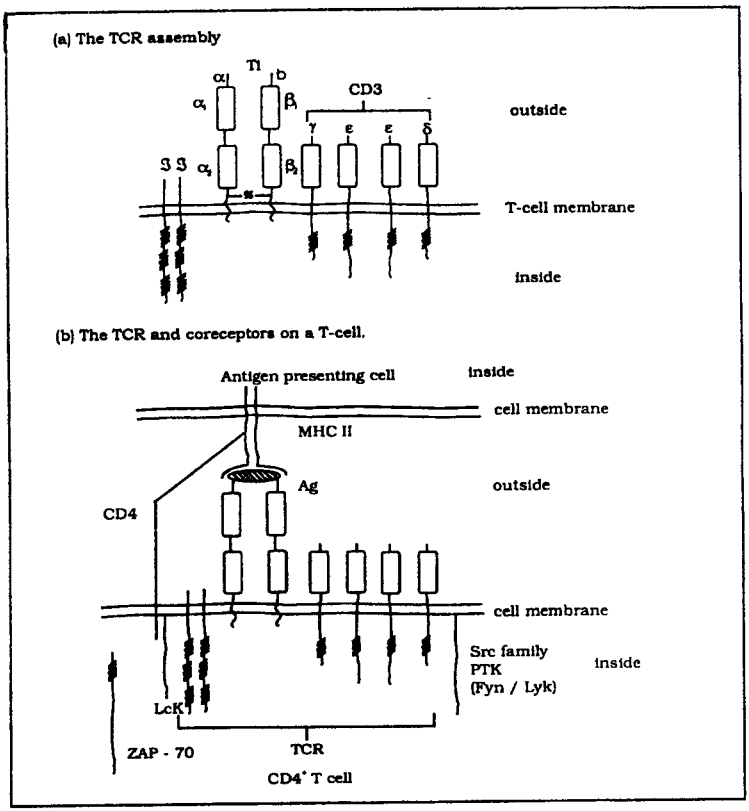


Figure 12.11 : The cell receptor (TCR).

trapped. In the lymph nodes, B lymphocytes are localized in follicles, with T cells more diffusely distributed in surrounding paracortical areas, also referred to as T-cell zones. Some of the B-cell follicles include germinal centers, where B cells are undergoing intense proliferation after encountering their specific antigen and their cooperating T cells. B and T lymphocytes are segregated in a similar fashion in the other peripheral lymphoid tissues, and this organization promotes the crucial interactions that occur between B and T cells upon encountering antigen.

The spleen is a fist-sized organ just behind the stomach that collects antigen from the blood. It also collects and disposes of senescent red blood cells.

The gut-associated lymphoid tissues (GALT), which include the tonsils, adenoids, and appendix, and specialized structures called Peyer's patches in the small intestine, collect antigen from the epithelial surfaces

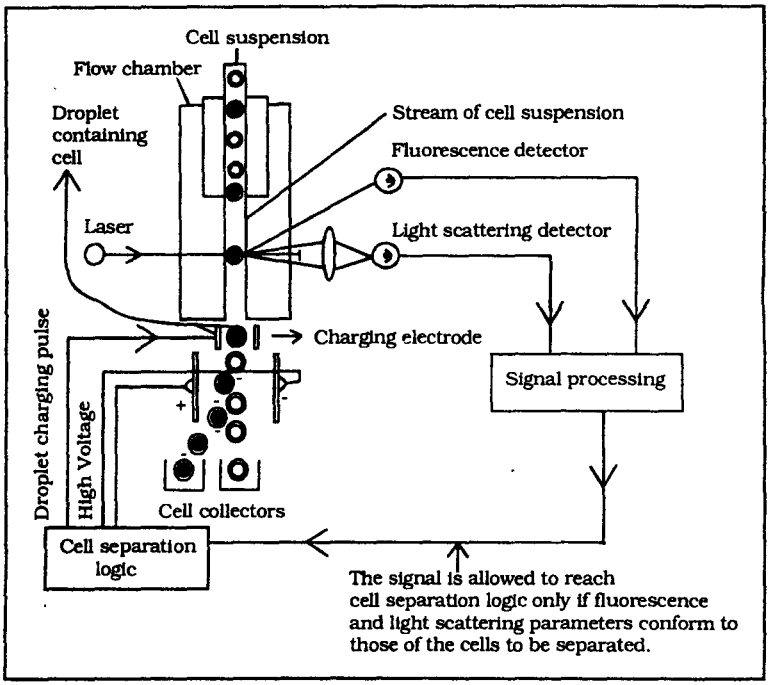


Figure 12.12 : Fluorescence Activated Cell sorting (FACS).

of the gastrointestinal tract. In Peyer's patches, which are the most important and highly organized of these tissues, the antigen is collected by specialized epithelial cells called multi-fenestrated or M cells. The lymphocytes form a follicle consisting of a large central dome of B lymphocytes surrounded by smaller numbers of T lymphocytes. Similar but more diffuse aggregates of lymphocytes protect the respiratory epithelium, where they are known as bronchial-associated lymphoid tissue (BALT), and other mucosa, where they are known simply as mucosal-associated lymphoid tissue (MALT). Collectively, the mucosal immune system is estimated to contain as many lymphocytes as all the rest of the body, and they form a specialized set of cells obeying somewhat different rules.

Although very different in appearance, the lymph nodes, spleen, and mucosal-associated lymphoid tissues all share the same basic architecture. Each of these tissues operates on the same principle, trapping antigen from sites of infection and presenting it to migratory small lymphocytes, thus inducing adaptive immune responses. The peripheral lymphoid tissues also provide sustaining signals to the

lymphocytes that do not encounter their specific antigen, so that they continue to survive and recirculate until they encounter their specific antigen. This is important in maintaining the correct numbers of circulating T and B lymphocytes, and ensures that only those lymphocytes with the potential to respond to foreign antigen are sustained.

12.2.4 Lymphocytes Circulation Between Blood and Lymph

Small B and T lymphocytes that have matured in the bone marrow and thymus but have not yet encountered antigen are referred to as naive lymphocytes. These cells circulate continually from the blood into the peripheral lymphoid tissues, which they enter by squeezing between the cells of capillary walls. They are then returned to the blood via the lymphatic vessels or, in the case of the spleen, return directly to the blood. In the event of an infection, lymphocytes that recognize the infectious agent are arrested in the lymphoid tissue, where they proliferate and differentiate into effector cells capable of combating the infection.

When an infection occurs in the periphery, for example, large amounts of antigen are taken up by dendritic cells which then travel from the site of infection through the afferent lymphatic vessels into the draining lymph nodes. In the lymph nodes, these cells display the antigen to recirculating T lymphocytes, which they also help to activate. B cells that encounter antigen as they migrate through the lymph node are also arrested and activated, with the help of some of the activated T cells. Once the antigen-specific lymphocytes have undergone a period of proliferation and differentiation, they leave the lymph nodes as effector cells through the efferent lymphatic vessel.

Because they are involved in initiating adaptive immune responses, the peripheral lymphoid tissues are not static structures but vary quite dramatically depending upon whether or not infection is present. The diffuse mucosal lymphoid tissues may appear in response to infection and then disappear, whereas the architecture of the organized tissues changes in a more defined way during an infection. For example, the B-cell follicles of the lymph nodes expand as B lymphocytes proliferate to form germinal centers, and the entire lymph node enlarges, a phenomenon familiarly known as swollen glands.

13

Regulation of the Immune Response

Antibody and cell-mediated immune responses must be closely regulated following antigenic stimulation. Otherwise, our bodies would be overwhelmed by proliferating clones of lymphocytes and their products.

Immune responses to foreign antigens are self-limiting and diminish as the antigens are eliminated, leading to a return of the immune system to a resting state.

13.1 ANTIGEN

- antigen is at the center of the immunoregulatory process since immune responses (humoral and cell-mediated) wane as antigen concentration decreases.
- the presence of foreign antigen results in signals from antigen receptors, costimulatory molecules, and cytokine receptors that promote survival of reactive lymphocytes. At the same time, there is decreased expression of anti-apoptotic Bcl-2. Effector cells that develop in the course of an adaptive immune response function to eliminate foreign antigen, thereby depriving activated effector cells of essential survival signals. Cessation of activating signals therefore causes effector cells that are no longer needed to die by apoptosis. On the other hand, memory cells are quiescent, and therefore survive.
- antigenic competition can sometimes occur, i.e., the presence of a competing antigen can regulate the immune response to an unrelated antigen. One example involves cytokine cross-regulation whereby an immunodominant antigen which

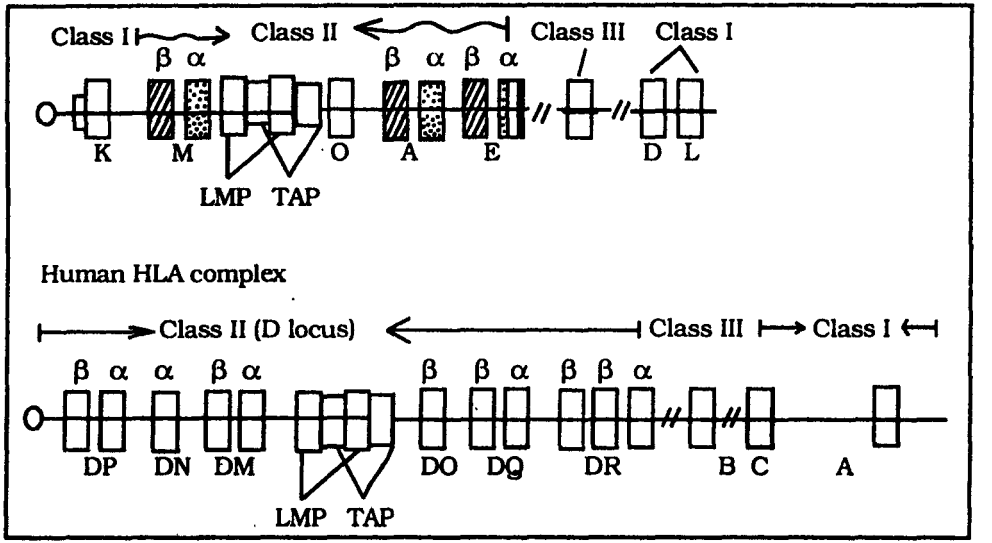


Figure 13.1 : The major histocompatibility gene complex (MHC) in mouse (H-2) (chromosome No. 17) and man (HLA). (Chromosome No. 6) - A simplified picture.

stimulates a Th1 response with attendant IFN-gamma synthesis that would down-regulate a less potent Th2 response to an unrelated antigen.

13.2 ANTIBODY

- simple clearance of antigen by freely circulating antibody (immune complexes are phagocytosed and removed by macrophages) eventually results in less antigen being available to activate B cells and T cells.
- secreted IgG antibodies inhibit continuing B cell activation by forming complexes with multivalent antigen that cross-link surface immunoglobulin and IgG Fc receptors on B cells. This is called antibody feedback and involves the activation of a phosphatase that prevents further signaling through the B cell receptor.

13.3 ACTIVE REGULATORY MECHANISMS

CTLA-4, which is not detectable on resting T cells but is expressed after several days of T cell activation, has a greater avidity for B7 molecules than CD28 and therefore favorably competes with CD28 for its ligands. The interaction of CTLA-4 with B7 results in an inhibitory signal (believed to involve the activation of a phosphatase) that down-regulates T cell activation. Mutant mice that are deficient in CTLA-4 expression exhibit signs of massive T cell proliferation and die within 4 weeks of birth.

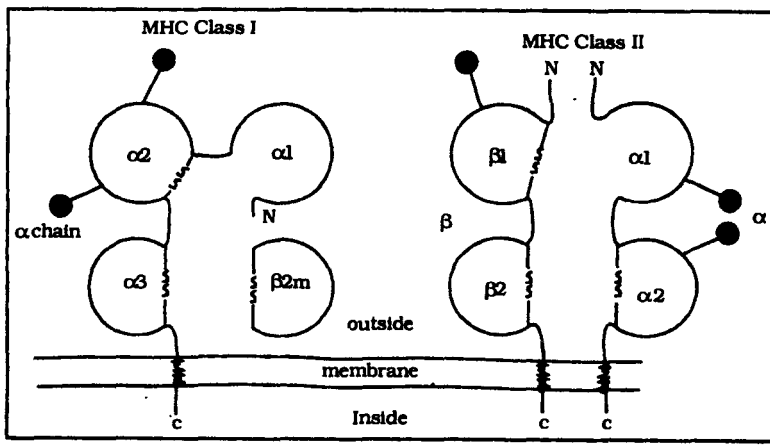


Figure 13.2 : MHC class I and class II molecules, β-2 m : β-2 microglobulin represents oligosaccharide moiety.

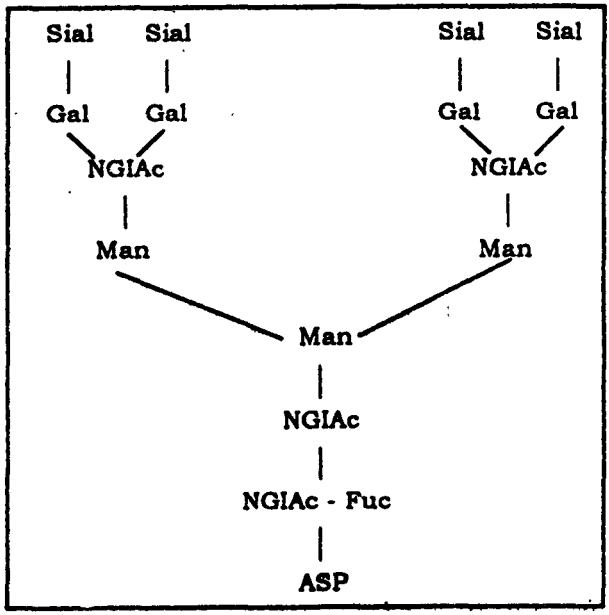


Figure 13.3 : A typical sugar moiety atached to MHC Class I molecule.

Activation-induced cell death (AICD) is an important mechanism for the maintenance of T cell homeostasis. T cell activation due to persistant or repeated stimulation by antigen leads to the expression of Fas and Fas ligand. Trimerization of Fas by Fas ligand on a neighboring cells leads to T cell death by apoptosis. Mutant mice that lack functional Fas or Fas ligand have excessive numbers of T cells and spontaneously develop autoimmune disease.

$CD4^+CD25^+$ regulatory T cells may arise after persistant antigenic stimulation of lymphocytes. These regulatory cells, which are antigen-specific, can inhibit the proliferation of other T cell populations via a cell contact-dependent mechanism. It may be possible to increase the activity of regulatory T cells to diminish unwanted immune responses (eg., autoimmune disease, allograft rejection) or decrease the activity of regulatory T cells to promote desirable immune responses (eg., anti-tumor responses).

Natural suppressor cells are antigen-nonspecific inhibitory cells typically found in bone marrow and other sites of hematopoiesis. These large granular lymphocytes are activated by IFNgamma and release nonspecific suppressor factors (eg., transforming growth factor-ß) that inhibit IL-2 production and IL-2R expression.

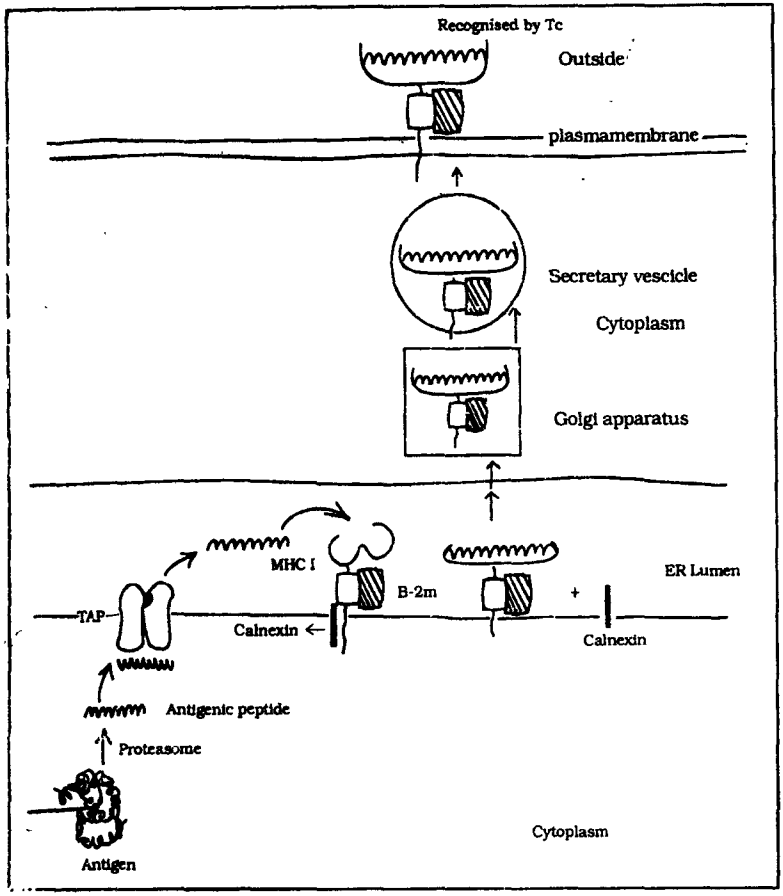


Figure 13.4 : Processing of indegenous antigen and its presentation by MHC class I molecule.

Under certain conditions macrophages produce prostaglandins that are able to inhibit immune responses in a nonspecific manner by promoting cAMP accumulation in T cells, thereby blocking IL-2 production and utilization.

13.4 IDIOTYPE NETWORKS

The antigen binding site (formed by the light and heavy chain variable domains) of an antibody molecule has a characteristic structure or conformation which can be recognized by anti-idiotypic antibodies. Each antigenic determinant in the variable region is called an idiotope. An idiotope may be identical to the paratope (antigen-binding site) or may be outside of the paratope. The sum of the individual idiotopes of

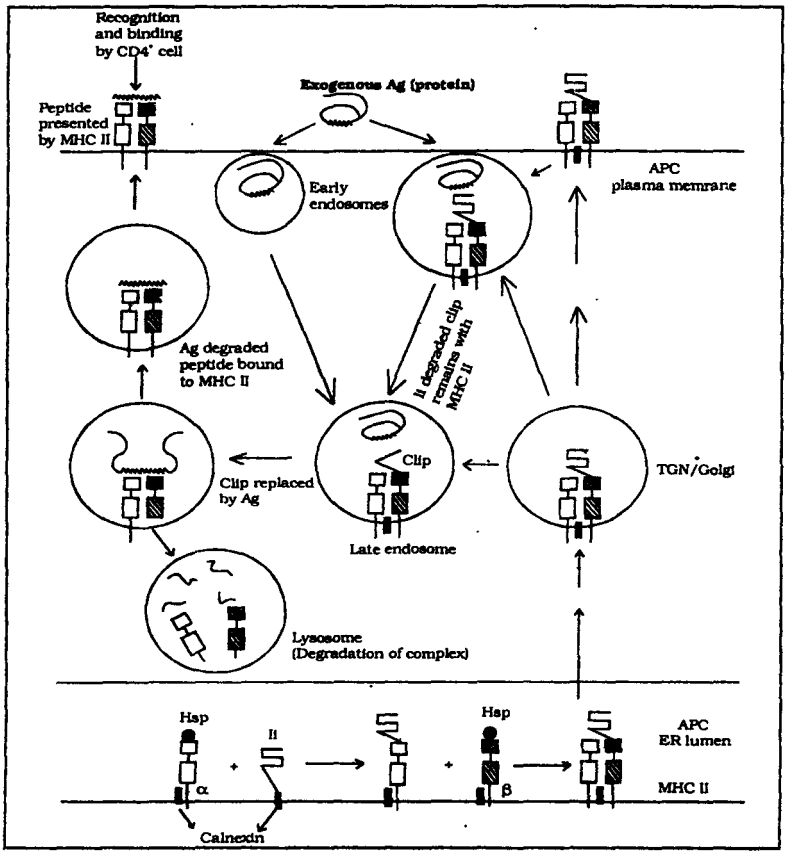


Figure 13.5 : Processing of exogenous antigen and its presentation by MHC II by APC. Ii : invariant chain; TGN : transgolgi network; HSP : Heat shock protein.

an antibody molecule is called the idiotype.

Since anti-idiotypic antibodies directed against the paratope appear as the internal image of the original epitope on the antigen, a network of idiotype-anti-idiotype interactions can be formed.

Antigen induces Ab_1

Ab_1 induces Ab_2 (anti-idiotypic antibody)

Ab_2 induces Ab_3 (Ab_2 recognizes determinants on both Ab_1 and Ab_3)

And so on...

When a foreign antigen enters the body, the idiotypes of responding clones of lymphocyte are expanded and anti-idiotypic responses are

triggered that shut off the antigen-specific (idiotype-expressing) lymphocytes. Complementary interactions between idiotypes and anti-idiotypes reach a steady state at which the immune system is at homeostasis.

Since T cell receptor molecules also possess idiotypic determinants, idiotype-anti-idiotype T cell networks also exist in parallel with idiotype-anti-idiotype B cell networks. Idiotype networks may serve to amplify as well as suppress immune responses. An idiotype network could also provide the basal stimulation needed for the maintenance of memory B and T cells.

13.3 NEUROENDOCRINE REGULATION

Lymphoid tissues are innervated by autonomic and primary sensorial neurons while individual lymphocytes possess receptors for a variety of neuropeptides and hormones.

Neuroendocrine factors can suppress or enhance immune responses. For example, during periods of stress, IL-1 stimulates glucocorticoid synthesis by the adrenal cortex, resulting in inhibition or down-regulation of antibody production, NK activity, and cytokine synthesis. On the other hand, prolactin synthesis enhances macrophage activation and IL-2 production.

14

Tumor Immunology

Cancer is one of the three leading causes of death in industrialized nations. As treatments for infectious diseases and the prevention of cardiovascular disease continue to improve, and the average life expectancy increases, cancer is likely to become the most common fatal disease in these countries. Cancers are caused by the progressive growth of the progeny of a single transformed cell. Therefore, curing cancer requires that all the malignant cells be removed or destroyed without killing the patient. An attractive way to achieve this would be to induce an immune response against the tumor that would discriminate between the cells of the tumor and their normal cell counterparts. Immunological approaches to the treatment of cancer have been attempted for over a century, with tantalizing but unsustainable results. Experiments in animals have, however, provided evidence for immune responses to tumors and have shown that T cells are a critical mediator of tumor immunity. More recently, advances in our understanding of antigen presentation and the molecules involved in T-cell activation have provided new immunotherapeutic strategies based on a better molecular understanding of the immune response. These are showing some success in animal models and are now being tested in human patients.

14.1 T LYMPHOCYTES CAN RECOGNIZE SPECIFIC ANTIGENS ON HUMAN TUMORS

Tumor rejection antigens are peptides of tumor-cell proteins that are presented to T cells by MHC molecules. These peptides can become the targets of a tumor-specific T-cell response because they are not displayed on the surface of normal cells, at least not at levels sufficient to be recognized by T cells. Six different categories of tumor rejection antigens can be distinguished. The first category consists of antigens

that are strictly tumor specific. These antigens are the result of point mutations or gene rearrangements, which often arise as part of the process of oncogenesis. Point mutations may evoke a T-cell response either by allowing *de novo* binding of a peptide to MHC class I molecules or by creating a new epitope for T cells by modification of a peptide that already binds class I molecules. A special class of tumor-specific antigen in the case of B- and T-cell tumors, which are derived from single clones of lymphocytes, are the idiotypic sequences unique to the antigen receptor expressed by the clone.

The second category comprises proteins encoded by genes that are normally expressed only in male germ cells, which do not express MHC molecules and therefore cannot present peptides from these molecules to T lymphocytes. Tumor cells show widespread abnormalities of gene expression, including the activation of these genes and thus the presentation of these proteins to T cells; hence, these proteins are effectively tumor specific in their expression as antigens.

The third category of tumor rejection antigen is comprised of differentiation antigens encoded by genes that are only expressed in particular types of tissue. The best examples of these are the differentiation antigens expressed in melanocytes and melanoma cells; a number of these antigens are proteins involved in the pathways of production of the black pigment, melanin. The fourth category is comprised of antigens that are strongly overexpressed in tumor cells compared with their normal counterparts. An example is HER-2/neu (also known as c-Erb-2), which is a receptor tyrosine kinase homologous to the epidermal growth factor receptor. This receptor is overexpressed in many adenocarcinomas, including breast and ovarian cancers, where it is linked with a poor prognosis. MHC class I-restricted, CD8-positive cytotoxic T lymphocytes have been found infiltrating solid tumors overexpressing HER-2/neu but are not capable of destroying such tumors *in vivo*. The fifth category of tumor rejection antigens is comprised of molecules that display abnormal posttranslational modifications. An example is underglycosylated mucin, MUC-1, which is expressed by a number of tumors, including breast and pancreatic cancers.

Proteins encoded by viral oncogenes comprise the sixth category of tumor rejection antigen. These oncoviral proteins are viral proteins that may play a critical role in the oncogenic process and, because they are foreign, they can evoke a T-cell response. Examples of such proteins are the human papilloma type 16 virus proteins, E6 and E7, which are expressed in cervical carcinoma.

Although each of these categories of tumor rejection antigen may

evoke an anti-tumor response *in vitro* and *in vivo*, it is exceptional for such a response to be able to spontaneously eliminate an established tumor. It is the goal of tumor immunotherapy to harness and augment such responses to treat cancer more effectively. In this respect, the spontaneous remission occasionally observed in cases of malignant melanoma and renal cell carcinoma, even when disease is quite advanced, offers hope that this goal is achievable.

In melanoma, tumor-specific antigens were discovered by culturing irradiated tumor cells with autologous lymphocytes, a reaction known as the mixed lymphocyte-tumor cell culture. From such cultures, cytotoxic T lymphocytes could be identified that would kill, in an MHC-restricted fashion, tumor cells bearing the relevant tumor-specific antigen. Melanomas have been studied in detail using this approach. Cytotoxic T cells reactive against melanoma peptides have been cloned and used to characterize melanomas by the array of tumor-specific antigens displayed. These studies have yielded three important findings. The first is that melanomas carry at least five different antigens that can be recognized by cytotoxic T lymphocytes. The second is that cytotoxic T lymphocytes reactive against melanoma antigens are not expanded *in vivo*, suggesting that these antigens are not immunogenic *in vivo*. The third is that the expression of these antigens can be selected against *in vitro* and possibly also *in vivo* by the presence of specific cytotoxic T cells. These discoveries offer hope for tumor immunotherapy, an indication that these antigens are not naturally strongly immunogenic, and also a caution about the possibility of selecting, *in vivo*, tumor cells that can escape recognition and killing by cytotoxic T cells.

Consistent with these findings, functional melanoma-specific T cells can be propagated from peripheral blood lymphocytes, from tumor-infiltrating lymphocytes, or by draining the lymph nodes of patients in whom the melanoma is growing. Interestingly, none of the peptides recognized by these T cells derives from the mutant proto-oncogenes or tumor suppressor genes that are likely to be responsible for the initial transformation of the cell into a cancer cell, although a few are the products of mutant genes. The rest derive from normal genes but are displayed at levels detectable by T cells for the first time on tumor cells. Antigens of the MAGE family are not expressed in any normal adult tissues, with the exception of the testis, which is an immunologically privileged site. They probably represent early developmental antigens reexpressed in the process of tumorigenesis. Only a minority of melanoma patients have T cells reactive against the MAGE antigens, indicating

that these antigens are either not expressed or are not immunogenic in most cases. The most common melanoma antigens are peptides from the enzyme tyrosinase or from three other proteins gp100, MART1, and gp75. These are differentiation antigens specific to the melanocyte lineage from which melanomas arise. It is likely that overexpression of these antigens in tumor cells leads to an abnormally high density of specific peptide:MHC complexes and this makes them immunogenic. Although in most cases tumor rejection antigens are presented as peptides complexed with MHC class I molecules, tyrosinase has been shown to stimulate CD4 T-cell responses in some melanoma patients by being ingested and presented by cells expressing MHC class II.

Tumor rejection antigens shared between most examples of a tumor, and against which tolerance can be broken, represent candidate antigens for tumor vaccines. The MAGE antigens are candidates because of their limited tissue distribution and their shared expression by many melanomas. It might seem dangerous to use tumor vaccines based on antigens that are not truly tumor-specific because of the risk of inducing autoimmunity. Often, however, the tissues from which tumors arise are dispensable; the prostate is perhaps the best example of this. With melanoma, however, some melanocyte-specific tumor rejection antigens are also expressed in certain retinal cells, in the inner ear, in the brain, and in the skin. Despite this, melanoma patients receiving immunotherapy with whole tumor cells or tumor-cell extracts, although occasionally developing vitiligo a destruction of pigmented cells in the skin that correlates well with a good response to the tumor do not develop abnormalities in the visual, vestibular, and central nervous systems, perhaps because of the low level of expression of MHC class I molecules in these sites.

In addition to the human tumor antigens that have been shown to induce cytotoxic T-cell responses, many other candidate tumor rejection antigens have been identified by studies of the molecular basis of cancer development. These include the products of mutated cellular oncogenes or tumor suppressors, such as Ras and p53, and also fusion proteins, such as the Bcr-Abl tyrosine kinase that results from the chromosomal translocation (t9;22) found in chronic myeloid leukemia. It is intriguing that, in each of these cases, no specific cytotoxic T-cell response has been identified in cultures of autologous lymphocytes with tumor cells bearing these mutated antigens. However, cytotoxic T lymphocytes specific for these antigens can be developed *in vitro* by using peptide sequences derived either from the mutated sequence or

from the fusion sequence of these common oncogenic proteins; these cytotoxic T cells are able to recognize and kill tumor cells. In chronic myeloid leukemia, it is known that, after treatment and bone marrow transplantation, mature lymphocytes from the bone marrow donor infused into the patient can help to eliminate any residual tumor. At present, it is not clear whether this is a graft-versus-host effect, where the donor lymphocytes are responding to alloantigens expressed on the leukemia cells, or whether there is a specific anti-leukemic response. The ability to prime the donor cells against leukemia-specific peptides offers the prospect of enhancing the anti-leukemic effect while minimizing the risk of graft-versus-host disease. It is a challenge for immunologists to understand why these mutated proteins do not prime cytotoxic T cells in the patients in which the tumors arise. They are excellent targets for therapy, as they are unique to the tumor and have a causal role in oncogenesis.

14.2 TUMORS REJECTION PREVENTION

The ability of the immune system to detect tumor cells and destroy them is know as 'immune surveillance.' However, it is difficult to show that tumors are subject to surveillance by the immune system; after all, cancer is a common disease, and most tumors show little evidence of immunological control. The incidence of the common tumors in mice that lack lymphocytes is little different from their incidence in mice with normal immune systems; the same is true for humans deficient in T cells. The major tumor types that occur with increased frequency in immunodeficient mice or humans are virus-associated tumors; immune surveillance thus seems to be critical for control of virus-associated tumors, but the immune system does not normally respond to the novel antigens deriving from the multiple genetic alterations in spontaneous tumors. The goal in the development of anti-cancer vaccines is to break the tolerance of the immune system for antigens expressed mainly or exclusively by the tumor.

It is not surprising that spontaneously arising tumors are rarely rejected by T cells, as in general they probably lack either distinctive antigenic peptides or the adhesion or co-stimulatory molecules needed to elicit a primary T-cell response. Moreover, there are other mechanisms whereby tumors can avoid immune attack or evade it when it occurs. Tumors tend to be genetically unstable and can lose their antigens by mutation; in the event of an immune response, this instability might generate mutants that can escape the immune response. Some tumors, such as colon and cervical cancers, lose the expression

of a particular MHC class I molecule, perhaps through immunoselection by T cells specific for a peptide presented by that MHC class I molecule. In experimental studies, when a tumor loses expression of all MHC class I molecules, it can no longer be recognized by cytotoxic T cells, although it might become susceptible to NK cells. However, tumors that lose only one MHC class I molecule might be able to avoid recognition by specific CD8 cytotoxic T cells while remaining resistant to NK cells, conferring a selective advantage *in vivo*.

Yet another way in which tumors might evade rejection is by making immunosuppressive cytokines. Many tumors make these, although in most cases little is known of their precise nature. TGF-bδ was first identified in the culture supernatant of a tumor (hence its name, transforming growth factor-bδ) and, as we have seen, tends to suppress inflammatory T-cell responses and cell-mediated immunity, which are needed to control tumor growth. A number of tumors of different tissue origins, such as melanoma, ovarian carcinoma, and B-cell lymphoma, have been shown to produce the immunosuppressive cytokine IL-10, which can reduce dendritic cell development and activity. Thus, there are many different ways in which tumors avoid recognition and destruction by the immune system.

14.3 CONTROL OF TUMOR GROWTH BY MONOCLONAL ANTIBODIES AGAINST TUMOR ANTIGENS

The advent of monoclonal antibodies suggested the possibility of targeting and destroying tumors by making antibodies against tumor-specific antigens. This depends on finding a tumor-specific antigen that is a cell-surface molecule. So far there has been limited success with this approach, although, as an adjunct to other therapies, it holds promise. Some striking initial results have been reported in the treatment of breast cancer with a humanized monoclonal antibody, known as Herceptin, which targets a growth factor receptor, HER-2/neu, that is overexpressed in about a quarter of breast cancer patients. As we discussed elsewhere in this chapter, this overexpression accounts for HER-2/neu evoking an antitumor T-cell response, although HER-2/neu is also associated with a poorer prognosis. It is thought that Herceptin acts by blocking interaction between the receptor and its natural ligand and by downregulating the level of expression of the receptor. The effects of this antibody can be potentiated when it is combined with conventional chemotherapy. A second monoclonal antibody that has promise for the treatment of non-Hodgkin's B-cell lymphoma binds to

CD20 and is known as Rituximab. Ligation and clustering of CD20 transduces a signal that causes lymphocyte apoptosis. Monoclonal antibodies coupled to gō-emitting radioisotopes have also been used to image tumors, for the purpose of diagnosis and monitoring tumor spread.

The first reported successful treatment of a tumor with monoclonal antibodies used anti-idiotypic antibodies to target B-cell lymphomas whose surface immunoglobulin expressed the corresponding idiotype. The initial course of treatment usually leads to a remission, but the tumor always reappears in a mutant form that no longer binds to the antibody used for the initial treatment. This case represents a clear example of genetic instability enabling a tumor to evade treatment.

Other problems with tumor-specific or tumor-selective monoclonal antibodies as therapeutic agents include inefficient killing of cells after binding of the monoclonal antibody and inefficient penetration of the antibody into the tumor mass. The first problem can often be circumvented by linking the antibody to a toxin, producing a reagent called an *immunotoxin*; two favored toxins are ricin A chain and *Pseudomonas* toxin. Both approaches require the antibody to be internalized to allow the cleavage of the toxin from the antibody in the endocytic compartment, allowing the toxin chain to penetrate and kill the cell.

Two other approaches using monoclonal antibody conjugates involve linking the antibody molecule to chemotherapeutic drugs such as adriamycin or to radioisotopes. In the first case, the specificity of the monoclonal antibody for a cell-surface antigen on the tumor concentrates the drug to the site of the tumor. After internalization, the drug is released in the endosomes and exerts its cytostatic or cytotoxic effect. Monoclonal antibodies linked to radionuclides concentrate the radioactive source in the tumor site. Both these approaches have the advantage of also killing neighboring tumor cells, because the released drug or radioactive emissions can affect cells adjacent to those that actually bind the antibody. Ultimately, combinations of toxin-, drug-, or radionuclide-linked monoclonal antibodies, together with vaccination strategies aimed at inducing T cell-mediated immunity, might provide the most effective cancer immunotherapy.

14.4 ENHANCING THE IMMUNOGENICITY OF TUMORS HOLDS PROMISE FOR CANCER THERAPY

Although vaccines based on tumor antigens are, in principle, the ideal approach to T cell-mediated cancer immunotherapy, it may be

many decades before the dominant tumor antigens for common cancers are identified. Even then, it is not clear how widely the relevant epitopes will be shared between tumors, and peptides of tumor rejection antigens will be presented only by particular MHC alleles. To be effective, a tumor vaccine may therefore need to include a range of tumor antigens. MAGE-1 antigens, for example, are recognized only by T cells in melanoma patients expressing the HLA-A1 haplotype. However, the range of MAGE-type proteins that has now been characterized encompasses peptide epitopes presented by many HLA class I and II molecules.

Until recently, most cancer vaccines have used the individual patient's tumor removed at surgery as a source of vaccine antigens. These cell-based vaccines are prepared by mixing either irradiated tumor cells or tumor extracts with bacterial adjuvants such as BCG or *Corynebacterium parvum*, which enhance their immunogenicity. Such vaccines have generated modest therapeutic results in melanomas but have, in general, been disappointing.

Where candidate tumor rejection antigens have been identified, for example in melanoma, experimental vaccination strategies include the use of whole proteins, peptide vaccines based on sequences recognized by cytotoxic T lymphocytes (either administered alone or presented by the patient's own dendritic cells), and recombinant viruses encoding these peptide epitopes. A novel experimental approach to tumor vaccination is the use of heat-shock proteins isolated from tumor cells. The underlying principle of this therapy is that one of the physiological activities of heat-shock proteins is to act as intracellular chaperones of antigenic peptides. There is evidence for receptors on the surface of professional antigen-presenting cells that take up certain heat-shock proteins together with any bound peptides. Uptake of heat-shock proteins via these receptors delivers the accompanying peptide into the antigen-processing pathways leading to peptide presentation by MHC class I molecules. This experimental technique for tumor vaccination has the advantage that it does not depend on any prior knowledge of the nature of the relevant tumor rejection antigens, but the disadvantage that the heat-shock proteins purified from the cell carry very many peptides, so that any tumor rejection antigen might constitute only a tiny fraction of the peptides bound to the heat-shock protein.

A further experimental approach to tumor vaccination in mice is to increase the immunogenicity of tumor cells by introducing genes that encode co-stimulatory molecules or cytokines. This is intended to

make the tumor itself more immunogenic. A tumor cell transfected with the gene encoding the co-stimulatory molecule B7 (see Section 8-5) is implanted in a syngeneic animal. These B7-positive cells can activate tumor-specific naive T cells to become armed effector T cells able to reject the tumor cells. They are also able to stimulate further proliferation of the armed effector cells that reach the site of implantation. These T cells can then target the tumor cells whether they express B7 or not; this can be shown by reimplanting nontransfected tumor cells, which are also rejected.

The second strategy, that of introducing cytokine genes into tumors so that they secrete the relevant cytokine, is aimed at attracting antigen-presenting cells to the tumor and takes advantage of the paracrine nature of cytokines. In mice, the most effective tumor vaccines so far are tumor cells that secrete granulocyte-macrophage colony-stimulating factor (GM-CSF), which induces the differentiation of hematopoietic precursors into dendritic cells and attracts these to the site. GM-CSF is also thought to function as an adjuvant, activating the dendritic cells. It is believed that these cells process the tumor antigens and migrate to the local lymph nodes, where they induce potent anti-tumor responses. The B7-transfected cells seem less potent in inducing anti-tumor responses, perhaps because the bone marrow-derived dendritic cells express more of the molecules required to activate naive T cells than do B7-transfected tumor cells. In addition, the tumor cells do not share the dendritic cells' special ability to migrate into the T-cell areas of the lymph nodes, where they are optimally placed to interact with passing naive T cells.

The potency of dendritic cells in activating T-cell responses provides the rationale for yet another strategy for vaccinating against tumors. The use of antigen-pulsed autologous dendritic cells to stimulate therapeutically useful cytotoxic T-cell responses to tumors has been developed in experimental models, and there have been initial trials in humans with cancer.

Clinical trials are in progress to determine the safety and efficacy of many of these approaches in human patients. What is uncertain is whether people with established cancers can generate sufficient T-cell responses to eliminate all their tumor cells under circumstances in which any tumor-specific naive T cells might have been rendered tolerant to the tumor. Moreover, there is always the risk that immunogenic transfectants will elicit an autoimmune response against the normal tissue from which the tumor derived.

"This page is Intentionally Left Blank"