Toxicology and Risk Assessment: A Comprehensive Introduction

Edited by

HELMUT GREIM, M.D.

Institut für Toxikologie und Umwelthygiene, Technische Universität München

ROBERT SNYDER, Ph.D.

Environmental and Occupational Health Sciences Institute, Rutgers, The State University of New Jersey

John Wiley & Sons, Ltd

Toxicology and Risk Assessment

Toxicology and Risk Assessment: A Comprehensive Introduction

Edited by

HELMUT GREIM, M.D.

Institut für Toxikologie und Umwelthygiene, Technische Universität München

ROBERT SNYDER, Ph.D.

Environmental and Occupational Health Sciences Institute, Rutgers, The State University of New Jersey

John Wiley & Sons, Ltd

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England
Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Ltd, 6045 Freemont Blvd, Mississauga, Ontario L5R 4J3, Canada

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data

Toxicology and risk assessment : a comprehensive introduction / edited by Helmut Greim and Robert Snyder.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-470-86893-5 (cloth : alk. paper)
1. Toxicology. 2. Poisons–Risk assessment. 3. Health risk assessment.
I. Greim, Helmut. II. Snyder, Robert, 1935-
[DNLM: 1. Poisoning. 2. Hazardous Substances–toxicity. 3. Risk Assessment. QV 600 T75566 2007]
RA1211.T6348 2007
615.9–dc22 2007021098

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 978-0-470-86893-5

Typeset in 10/12 pt Times by Thomson Digital, India
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire, England
This book is printed on acid-free paper responsibly manufactured from sustainable forestry
in which at least two trees are planted for each one used for paper production.

This book is dedicated to Professor Herbert Remmer (1919–2003), late Director of the Institute of Toxicology of the University of Tübingen. He was a pathfinder in the biomedical sciences, a dedicated teacher, and an inspiration to all of his friends and colleagues. His work established a standard of excellence for generations of scientists to come. He is sorely missed.

Contents

Preface

About 30 years ago the need for trained toxicologists in the German chemical industry prompted Professors Herbert Remmer and Helmut Greim to organize a 3-year toxicology training program for 20 chemists. Using this experience the German Society of Pharmacology and Toxicology developed criteria to receive a certificate designating the 'Fach-Toxikologe' and initiated a broad training program to provide the information required. It became obvious that a text book was needed to accompany the classroom work to meet the needs of the students. The book* was published in the German language in 1995 and subsequently in Italian.** When time came for a new edition, the publishers, who were interested in expanding the market, suggested that a new edition, which could service a broader representation of the community of scholars in toxicology, should be written in English. The editors, Helmut Greim and Robert Snyder, decided to prepare a completely new book to ensure that recent achievements in toxicology are covered and each chapter produced by the faculty contained essential knowledge for a toxicologist. The authors and editors hope that the book proves useful to students and provides information at a level that will enable them to successfully study toxicology.

The current book is intended for people with a broad range of toxicological interests, including both practical and mechanistic subjects. References at the end of each chapter will allow the reader to go beyond this book into the toxicological literature. Furthermore, coverage extends into areas such as the 'omics,' biomarkers, molecular and cell biology, and newer approaches to risk assessment. It stresses the important need for overlap between mechanistic studies and safety assessment.

There are two famous admonitions in toxicology. The first, by Paracelsus, appears in the introduction. The second has been credited to any of several of our colleagues: 'Toxicology can be learned in two lessons, each 20 years long.' We hope that this book can start students down the path toward an exciting and productive career in toxicology.

Helmut Greim and Robert Snyder

**Toxikologie. Eine Einführung für Naturwissenschaftler und Mediziner, H. Greim und E. Deml eds, (Wiley)-Verlag Chemie, Weinheim, 1995.*

***Tossicologia, H. Greim and E. Deml eds, Zanichelli, Bologna, 2000.*

List of Contributors

Ilse-Dore Adler Institut für Genetik, GSF-Forschungszentrum für Umwelt & Gesundheit, Neuherberg, Germany

Marion W. Anders Department of Pharmacology, University of Rochester, NY, USA

Ulrich Andrae Institut für Toxikologie, GSF-Forschungszentrum für Umwelt & Gesundheit, Neuherberg, Germany

Rüdiger Bartsch Institut für Toxikologie und Umwelthygiene, Technische Universität München, Germany

Michael G. Bird Toxicology & Environmental Sciences Division, ExxonMobil Biomedical Sciences, Inc., Annandale, New Jersey, USA

Paul J. A. Borm Centre of Expertise in Life Sciences, Zuyd University, Heerlen, The Netherlands

Jeroen T. M. Buters Zentrum Allergie & Umwelt, Klinik & Poliklinik für Dermatologie & Allergologie der Technischen Universität München, Germany

Peter Calow Department of Environmental, Social and Spatial Change, Roskilde University, Denmark

György Csanády Institut für Toxikologie, GSF-Forschungszentrum für Umwelt & Gesundheit, Neuherberg, Germany

Gisela H. Degen Institut für Arbeitsphysiologie an der Universität Dortmund, Germany

Wolfgang Dekant Institut für Pharmakologie und Toxikologie der Universität Würzburg, Germany

Gunter P. Eckert Pharmakologisches Institut deo Universität Frankfurt, Germany

Thomas Efferth Deutsches Krebs Forschungs-Zentrum, Heidelberg, Germany

Victor J. Feron TNO Quality of Life, Business Unit Toxicology & Applied Pharmacology, Zeist, Netherlands

Johannes G. Filser Institut für Toxikologie, GSF-Forschungszentrum für Umwelt & Gesundheit, Neuherberg, Germany

Valery E. Forbes Department of Environmental, Social and Spatial Change, Roskilde University, Denmark

Kristian W. Fried Deptartment of Pharmacology, Toxicology and Therapeutics, University of Kansas Medical Center, Kansas City, USA

Carol R. Gardner Department of Pharmacology and Toxicology, Ernest Mario School of Pharmacy, Rutgers, The State University of New Jersey, Piscutaway, NJ, USA

Helmut Greim Institut für Toxikologie und Umwelthygiene, Technische Universität München, Germany

Stefan Halbach Institut für Toxikologie, GSF-Forschungszentrum für Umwelt & Gesundheit, Neuherberg, Germany

Franz Hofmann Institut für Pharmakologie und Toxikologie, Technische Universität München, Germany

Diana Jonker TNO Quality of Life, Business Unit Toxicology & Applied Pharmacology, Zeist, Netherlands

Bernd Kaina Institut für Toxikologie, Universität Mainz, Germany

Herrmann Kappus Klinik für Dermatologie, Charite, Campus Virchow-Klinikum, Berlin, Germany

Peter Kasper Bundesinstitut für Arzneimittel und Medizinprodukte, Bonn, Germany

Ines M. Lanzl Augenklinik und Poliklinik Rechts der Isar der Technischen Universität München, Germany

Debra L. Laskin Department of Pharmacology and Toxicology, Ernest Mario School of Pharmacy, Rutgers, The State University of New Jersey, Piscutaway, NJ, USA

Stephan Madle Bundesinstitut für Risikobewertung, Berlin, Germany

Pearl Moy ChemRisk, Inc., San Francisco, USA

Walter E. Müller Pharmakologisches Institut für Naturwissenschaftler, Johann Wolfgang Goethe-Universität, Frankfurt am Main, Germany

J. William Owens The Procter and Gamble Company, Central Product Safety, Cincinnati, Ohio, USA

Ulrike Pabel Bundesinstitut für Risikobewertung, Berlin, Germany

Jürgen Pauluhn Bayer AG, Institut für Toxikologie IC, Wuppertal, Germany

Dennis J. Paustenbach ChemRisk, Inc., San Francisco, USA

Karl K. Rozman Deptartment of Pharmacology, Toxicology and Therapeutics, University of Kansas Medical Center, Kansas City, USA

Tibor A. Rozman Deptartment of Pharmacology, Toxicology and Therapeutics, University of Kansas Medical Center, Kansas City, USA

Jens Schlossmann Institut für Pharmakologie und Toxikologie, Technische Universität München, Germany

Leslie Schwarz Institut für Toxikologie, GSF-Forschungszentrum für Umwelt & Gesundheit, Neuherberg, Germany

Michael Schwenk Medizinische Hochschule Hannover, Germany

Andreas Seeber Institut für Arbeitsphysiologie an der Universität Dortmund, Germany

Thomas Singer F. Hoffmann-La Roche AG, Basel, Switzerland

Robert Snyder Associate Dean for Research and Professor of Pharmacology and Toxicology, Ernest Mario School of Pharmacy and the Environmental and Occupational Health Sciences Institute, Rutgers the State University of New Jersey, Piscataway, USA

Günter Speit Institut für Humangenetik, Universität Ulm, Germany

Horst Spielmann Bundesinstitut für Risikobewertung, Berlin, Germany

Lesley Stanley Consultant in Investigative Toxicology, Fife, Scotland

Myriam Straube Bundesinstitut für Arzneimittel und Medizinprodukte

Karl-Heinz Summer Institut für Toxikologie, GSF-Forschungszentrum für Umwelt & Gesundheit, Neuherberg, Germany

Laura Suter-Dick F. Hoffmann-La Roche AG, Basel, Switzerland

Ladislaus Szinicz Institut für Pharmakologie und Toxikologie, Sanitätsakademie der Bundeswehr, München, Germany

Jürgen Timm Institut für Statistik der Universität Bremen, Germany

Kurt Ulm Institut für Medizinische Statistik und Epidemiologie des Klinikums Rechts der Isar der Technischen Universität München, Germany

John B. Watkins Medical Sciences, Indiana University School of Medicine, Bloomington, IN, USA

Thomas Zilker Toxikologische Abteilung der II. Medizinischen Klinik und Poliklinik Rechts der Isar der Technischen Universität München, Germany

1

Introduction to the Discipline of Toxicology

Helmut Greim and Robert Snyder

"In all things there is a poison, and there is nothing without a poison. It depends only upon the dose whether a poison is poison or not."

Paracelsus (1493–1541)

1.1 Introduction

The discipline of Toxicology is concerned with the health risks of human exposure to chemicals or radiation. According to Paracelsus' paradigm, Toxicology is charged with describing the adverse effects of chemicals in a qualitative sense, and with evaluating them quantitatively by determining how much of a chemical is required to produce a given response. Taking these two together, we can describe the intrinsic properties of an agent (hazard identification) and we can estimate the amount of the chemical required to produce these properties (risk characterization).

Fundamental to understanding Toxicology are the definitions of hazard, exposure, and risk:

Hazard: Intrinsic toxic properties [Expression of hazard depends upon conditions of use or exposure.]

Toxicology and Risk Assessment: A Comprehensive Introduction Edited by H. Greim and R. Snyder

Dose: Concentration of the chemical and the time of exposure
Risk: Likelihood of an adverse effect resulting from a given exposure

Since humans or organisms in the environment can be exposed via inhalation, skin contact, or oral intake, the concentrations in the different environmental compartments, which result in human or environmental exposure, must be determined. It is obvious from this that risk characterization comprises the following elements:

- Hazard identification, i.e. a description of the agent's toxic potential.
- Dose–response, including information on the concentration above which the agent induces toxic effects to identify the no-observable-effect level (NOEL).
- Exposure assessment, in which the concentration of the agent in the relevant medium and time of exposure are evaluated.

The sensitivity of measurements in analytical chemistry has advanced to the point where infinitesimally small amounts of chemicals can be detected and identified in the various media that characterize our environment. Detection of a chemical does not mandate that a toxicological effect in exposed people will be observed. Since the dose makes the poison, effects only occur when exposure exceeds the NOEL.

People may be exposed to chemicals in the air, water, food, or on the skin. The external dose at which a chemical exerts its toxic effects is a measure of its potency, i.e. a highly potent chemical produces its effects at low doses. Ultimately, the response to the chemical depends upon duration and route of exposure, the toxicokinetics of the chemical, the dose–response relationship, and the susceptibility of the individual.

It is necessary to establish toxicological profiles of each chemical, either pre-existing or newly developed, to insure that it can be utilized safely either by the public or under specific conditions of use such as in the workplace. Toxicological evaluations may take different forms for new and existing chemicals. In the case of newly developed drugs, pesticides, or new chemicals a stepwise procedure is used starting from simple *in vitro* and *in vivo* short-term tests. Depending on the hazardous potential of the agent, studies can be extended to evaluate long-term effects by repeated dose studies, toxicokinetics, and toxic mode of action. For existing chemicals the available information can be collected and a risk assessment, based on exposure data, knowledge of the dose–response relationship, and the mode of action, can be performed.

The parameters which determine toxic potential and potency are discussed in the following chapters. Here they are briefly discussed to indicate their importance for the risk-characterization process.

1.2 The Risk-Assessment Process

1.2.1 Hazard Identification

Acids or bases can be direct-acting agents which cause irritation or corrosion at the site of exposure. However, most chemicals induce systemic effects such as embryotoxicity,

hepatotoxicity, neurotoxicity, etc., after absorption from the gastrointestinal tract, through the skin, or via the lungs. Depending on exposure, concentration and time of exposure, acute or chronic effects may result. Acute intoxication usually occurs in response to large doses. Chronic effects are seen after repeated exposure during which time the chemical reaches critical concentrations at the target organ and the result is persistent accumulated damage. Exposure to some chemicals, such as 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), can result in retention and long-lasting effects even after a single high exposure. This is because TCDD is lipophilic and not well metabolized, which results in very slow elimination. The consequence is accumulation in adipose tissue. In humans the half-life of excretion is about 8 years. In laboratory animals and humans TCDD induces tumors in various organs. Since TCDD does not induce DNA damage or mutations, the carcinogenic effect is considered to have a threshold, i.e. there are doses below which no adverse effects will be observed. Induction of sensitization and of allergic responses by sensitizing agents is also considered to require reaching a threshold dose, even if this dose is very low, and the NOELs of these effects are rarely known. When establishing acceptable exposure standards, thresholds are not considered to be a property of the dose–response curves for genotoxic carcinogens because any genotoxic event is considered irreversible. [A more detailed discussion of this concept appears below.]

The Toxic Potential

Depending on reactivity, solubility, and metabolism, the chemical, or its metabolites, can reach critical target organs. Irritation or corrosion may occur when the reactant comes into contact with the skin or mucous membranes of the eye, the gastrointestinal tract, or the respiratory system. Distribution and metabolism of the chemical can result in various systemic effects upon interaction at targets in the critical organ. The organ-specific effects are described in the various chapters on organ toxicity, e.g. liver, kidney, the central and peripheral nervous system.

Historically, histopathological and biochemical changes have been the major parameters used to detect organ toxicity. Increasing availability of sensitive methods in analytical chemistry and molecular-biological approaches including toxicokinetics and the various 'omics' have significantly improved the understanding of alterations in cellular and sub-cellular function and the reaction of the cell to toxic insults. The result is a better understanding of toxic mechanisms, species differences, and the consequences of exposures at high and low concentrations for different times.

1.2.2 Dose–Response and Toxic Potency

The Paracelsian admonition teaches us that the occurrence and intensity of toxic effects are dose dependent. His paradigm addresses the concept of threshold effects, which implies knowledge of the dose–response relationship. Animal or human exposure is usually defined as the dose, e.g. in mg of the chemical/kg body weight/day. This daily dose may result from oral, inhalation, or dermal exposure or as a sum thereof. The external dose leads to a specific internal dose, which depends on the amount absorbed via the different routes. Absorption rates via the different routes can vary significantly, although oral and inhalation exposure usually lead to the highest internal dose. For

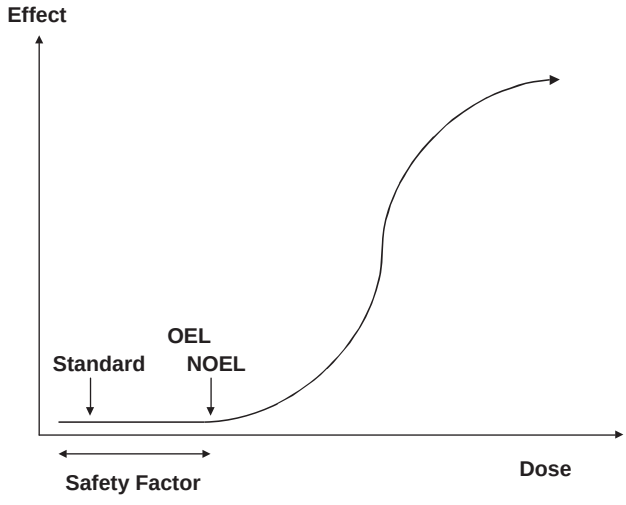


Figure 1.1 *Dose–response curve showing log dose on the X-axis and % response (Effect) on the Y-axis. The figure illustrates the location of regulatory values such as the NOEL, Occupational Exposure Levels (OELs), or environmental standards such as Acceptable Daily Intake (ADI). Note that a doubling of dose in the lower or upper part of the S-shaped curve results in small increases of effects, whereas it is much more prominent in the steep part [Modified from Greim and Deml, Toxikologie, Ch.1, Abt. 1-2. Copyright (1996), with permission from Wiley-VCH].*

example, about 50% of cadmium in inhaled air, e.g. in tobacco smoke, is absorbed in the lung, whereas cadmium absorption from the gastrointestinal tract is about 10%. Ultimately, it is the dose that reaches the cellular target over a given time period that results in the toxicological response. The dose that defines the toxic potency of a chemical is the product of the interrelated external, internal, and target doses. No toxic effects will be seen if the dose is below the NOEL, whereas effects increase with increasing exposure. The dose–response curve may be expressed using a variety of mathematical formulas. Using the semi-logarithmic form of the dose–response relationship the curve is sigmoidal in shape and varies in slope from chemical to chemical. Thus, if the curve is shallow a doubling of the dose results in a small increase of effects, whereas effects increase several-fold when the slope is steep (see Figure 1.1). The log of the dose is plotted on the abscissa (X axis) and increases toward the right. The location of the curve on the abscissa is a measure of the potency of the chemical.

1.2.3 Exposure Assessment

Since toxic effects are dose dependent, knowledge of the extent and duration of exposure is an integral part of the risk-assessment process. Exposure defines the amount of a chemical to which a population or individuals are exposed via inhalation, oral, and dermal routes. Animal or human exposure is commonly defined by mg of the chemical/kg body weight per day.

Toxicologists are concerned with exposure to any chemical, by any route, which may lead to adverse health effects. Workplace exposure may occur via dermal contact, by inhalation, even inadvertently, by oral ingestion. Permissible levels of occupational exposure are determined for an 8-hour work day over a 5-day work week extrapolated to an estimated working lifetime of 40 years. Monitoring of exposure usually requires measurement of the chemicals in the air of the workplace and/or the use of personal monitoring equipment. Whereas occupational exposure is regular and repetitive, exposure to consumer products is more difficult to assess. These methods are not usually applied in the home or the ambient environment. Exposure in the home can be estimated by the use of appropriate modeling techniques. Data useful for the determination of exposure to chemicals in consumer products include frequency, duration and site of exposure, concentration and weight of substance in the product, and the amount of product used per contact. Children represent a special case of exposure. For example, they may be exposed to chemicals that are released from toys during mouthing or via skin contact. Accurate exposure determinations for children are difficult to achieve. Exposure can be modeled based on data such as information on frequency of mouthing, migration rates of the specific compound from the toy during mouthing, and absorption rates from the oral cavity and gastrointestinal tract. The rate of absorption through the skin will also influence the body burden of the chemical. Use of these parameters to assess exposure is plagued by many uncertainties, which often lead to overestimation of the actual exposure. This external exposure may not necessarily correlate with internal exposure.

Biomonitoring of the compound or its reaction products in the exposed individuals provides the most reliable estimate of internal exposure. However, dose–response curves usually provide a correlation between external dose and effects. Therefore, risk assessment of an internal exposure requires either knowledge of the dose–response of internal exposure versus adverse effects or information to which extent external and internal doses correlate. The estimation of exposure is more complicated when mixtures of chemicals are the source of exposure.

1.2.4 Risk Characterization

The risk-assessment process requires differentiation between reversible and irreversible effects. The dose–response curves for chemicals that induce reversible effects display a region below which no effects can be observed. The highest dose at which no effects are seen is called the 'no-observable-effects level' (NOEL). The point at which effects become observable is called the 'lowest-observable-effect level' (LOEL). A threshold is not the equivalent of an NOEL, since it describes concentration or exposure at which the slope of the dose–response curves changes.

If damage is not repaired the effect persists and accumulates upon repeated exposure. In such cases a NOEL cannot be determined and every exposure can be related to a defined risk. Reversibility depends on the regenerative and repair capacity of cells, subcellular structures, and macromolecules during and after exposure. Epithelial cells of the intestinal tract or the liver have a high regenerating capacity and rapidly replace damaged cells by increased cell replication. The highly specialized cells of the nervous system have lost this capacity during natal and postnatal development. Consequently, damaged cells are not replaced, at least in the adult.

For chemicals which induce reversible effects the NOEL of the most sensitive endpoint is determined and compared with the human exposure to describe the Margin of Exposure (MOE) (or Margin of Safety: MOS). If the NOEL is derived from animal experiments an MOE of 100 or greater is desirable. An MOE of at least 10 is sufficient if the NOEL is derived from human data (see Figure 1.4).

The covalent binding of genotoxic mutagens and carcinogens to DNA is considered an irreversible event despite the availability of repair processes. Although there is increasing knowledge about DNA-repair mechanisms, the role of tumor-suppressor genes and apoptosis, their interactions, and dose–responses are not sufficiently understood to allow us to conclude whether genotoxic effects exhibit a threshold at low exposure or even a NOEL. So far, the general agreement remains that the potency of genotoxic carcinogens increases with increasing dose so that the risk at a given exposure needs to be estimated by linear extrapolation from the dose–response data obtained from experimental studies in animals or from data obtained from studies in humans.

1.3 Toxicological Evaluation of New and Existing Chemicals

The various toxic effects which chemicals may exert and the different applications for which chemicals are designed require in-depth understanding of the cause-and-effect relationship, i.e. knowledge of the chemical and the specific organs upon which it impacts. As a result toxicologists tend to focus on specific organs, specific applications (e.g., pesticides or drugs), specific compounds like metals or solvents, or specific effects like carcinogenicity. Chapters in this book are devoted to specific organ toxicity, and specific effects of compounds such as carcinogenicity and mutagenicity.

1.3.1 General Requirements for Hazard Identification and Risk Assessment

Toxicological evaluation of chemicals requires knowledge on the health consequences of acute, subchronic, and chronic exposure via routes relevant to the common use of the chemical. Therefore, all elements of risk assessment: hazard identification, dose–response, exposure, and risk have to be evaluated.

Organ specificity and other relevant endpoints like fertility, pre- and postnatal toxicity or carcinogenicity, their dose–response, and determination of the NOEL can only be identified by appropriate repeated-dose studies in animals. The use of in vitro testing can contribute important pieces of information, but so far cannot replace whole animal experimentation.

To obtain sufficient information on the hazardous properties of a chemical requires investigation of:

- acute, sub-chronic and chronic toxicity (oral, inhalation, dermal);
- irritation (skin, mucous membranes, eye) and phototoxicity;
- sensitization and photosensitization;

- genotoxicity (in vitro and in vivo methods);
- carcinogenicity (lifetime studies);
- reproductive toxicity;
- toxicokinetics;
- mode and mechanism of action.

In all studies information on the dose–response of effects is essential to identify the slope of the dose–response curves, possible thresholds, and the NOEL, LOEL, and the maximal tolerated dose (MTD).

Acute Toxicity, Subchronic Toxicity, and Chronic Toxicity

Acute toxicity studies describe toxic effects assessed after a single administration of the chemical to rodents and are primarily aimed at establishing a range of doses in which the chemical is likely to produce lethality. After dosing, the animals are observed over a period of one to two weeks to determine immediate or delayed effects. It is possible to plan studies in which other endpoints are examined as well.

Having established the lethal dose range the chemical may be examined for effects produced upon repeated administration. Common practice of such **repeated-dose studies** is to treat animals each day for a few weeks or months. These studies usually include rodents, but larger species such as dogs, and in the case of new drugs, monkeys or apes, may be employed. The animals must be observed for effects on general, as well as specific! organ toxicity. At the termination of these studies the animals are usually examined for gross and microscopic pathology.

Chronic studies, usually in rodents, involve treatment of animals for several months up to a lifetime. Their intent is to examine the likelihood of the development of pathology after long-term exposure to low levels of chemicals and in the case of lifetime studies are focused on cancer.

There is an ongoing discussion regarding the extent to which in vitro studies and consideration of structure–activity relationships provide sufficient information to waive repeated in vivo exposure studies. From a toxicological point of view it has to be stressed that this discussion is primarily concerned with cost reduction and protection of animals. It is necessary to insure that in this climate protection of human health and the environment do not become secondary considerations.

In vitro studies allow identification of hazardous properties of substances, but only those which can be detected by the specific test system. Even when the test system has a metabolic capacity, its appropriateness must be verified in intact organisms. Consequently, identification of all relevant endpoints, their dose–response, thresholds, and NOELs can be determined in the intact animal only by repeated-dose studies. In the absence of such information hazard identification is incomplete and without information of the dose–response relationship obtained from these studies there is no basis for appropriate assessment of the risk of human exposure.

Irritation and Phototoxicity

Dermal irritation of compounds is evaluated by studies in animals and humans prior to testing for sensitization. These are usually performed by using a single occluded patch

under the same conditions as applied when testing skin sensitization. Phototoxicity and photoallergic reactions have to be expected when compounds show significant absorption in the ultraviolet range (290–400 nm). Using the test strategy for irritation, an additional patch site is irradiated immediately after application of the test substance or after patch removal. Phototoxicity can also be tested by validated in vitro tests, such as uptake of Neutral Red by 3T3 cells. If such a test is negative further in vivo testing may not be necessary.

Sensitization and Photosensitization

For detection of sensitizing potential of products the choice of a relevant animal is crucial. However, in many cases animal models may be inappropriate for detection of a sensitizing potential so that most dermatologists prefer studies in humans. An acceptable alternative may be studies with nonhuman primate species like cynomologus or rhesus monkeys. Generally the Buehler guinea pig test and the popliteal lymph node assay (PLNA) in mice are used in the preclinical testing program. The PLNA received great attention because it is the only reliable test for screening compounds that cause sensitization via routes other than the skin. So far the test has been successfully applied to determine relative potencies of contact allergens and has been reported to closely correlate with NOELs established from human repeat patch testing. When the animal data indicate a weak contact-sensitizing potential, human skin-sensitizing testing is conducted usually by a human repeated-insult patch test (HRIPT). In any case, detection of antibodies in the serum during the studies using specific ELISA methods or bioassays to measure antibodies may be appropriate.

Genotoxicity (Carcinogenicity)

Test systems and test strategies to evaluate possible genotoxicity of a compound are described in detail in Chapter 4.3.1–3. Generally, a bacterial mutation assay and an in vitro cytogenetic assay are performed. The results are often verified by the mouse bone marrow micronucleus test, a reliable and widely used test system, which detects aneugens as well as clastogens. Chemicals that yield positive responses to these tests frequently do not undergo further development. However, those that appear to lack genotoxicity may be carried forward and evaluated in lifetime carcinogenicity studies in rodents.

Toxicity for Reproduction and Development

Studies to evaluate reproductive and developmental effects may only be needed if there are indications that the chemical, or critical metabolites, can reach the embryo and/or fetus and could cause either teratological, fetotoxic, or developmental effects. In such cases tests such as a reproduction/developmental toxicity screening test (OECD 421), a combined repeated dose toxicity study with the reproduction/developmental toxicity screening test (OECD 422), or the appropriate standard tests to evaluate effects on reproduction (One-generation reproduction toxicity, OECD 415) and prenatal developmental study (OECD 414) may be performed.

Toxicokinetics

Toxicokinetics describe Absorption, Distribution, Metabolism, and Elimination (ADME) of a chemical in humans, experimental animals, or cellular systems. Of specific importance for interpretation of animal studies and for extrapolation of hazards between species is the comparative information on the exposure and the dose that reaches the critical target.

A chemical may enter the body via food, air, or the skin. The amount absorbed depends on the concentration in the different media, on physical-chemical parameters such as solubility in water and fat, on its stability, and on the route of exposure (Figure 1.2).

Upon inhalation or skin penetration the compound directly enters the circulation and distributes into the organs. When absorbed from the gastrointestinal tract the chemical enters the liver via the portal vein. The epithelial cells of the gut wall and the liver demonstrate a large capacity for metabolizing chemicals so that a compound may be extensively metabolized by this 'first-pass effect' before entering the general circulation. Larger molecules, e.g. the glucuronosyl-conjugates, can be excreted via the biliary system into the duodenum where the conjugates may be hydrolysed so that the original compound is reabsorbed and reenters the liver. This process is defined as **enterohepatic circulation**. Inhalation or dermal exposure to a chemical, or its intravenous or

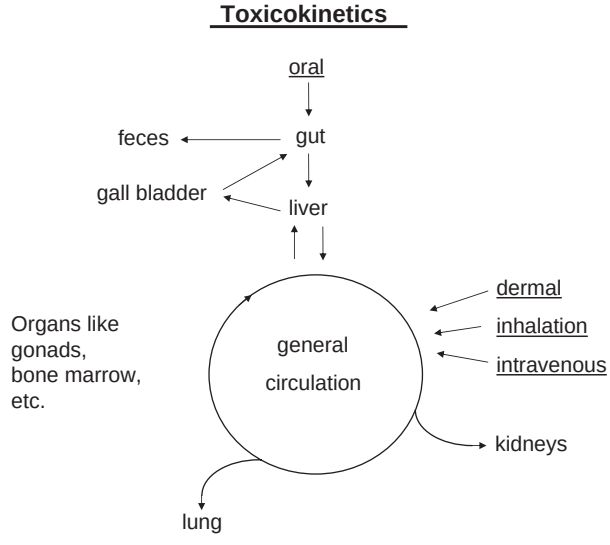


Figure 1.2 *Routes of exposure and systemic distribution of a compound within the organism. After oral ingestion the compound reaches the liver, where it can be extensively metabolized. Upon inhalation or dermal exposure and intravenous application the compound reaches the circulation without major metabolism [Modified from Greim and Deml, Toxikologie, Ch.1, Fig. 1-1. Copyright (1996), with permission from Wiley-VCH].*

intraperitoneal injection, may result in different effects than after oral exposure because of the 'first-pass effect.'

After entering the general circulation the chemical or its metabolites distribute to the organs where they can accumulate in organs such as fat or bones, or are further metabolized. Reactive metabolites will interact with tissue components and may induce cellular damage. This 'tissue dose,' i.e. the concentration of a chemical or its metabolite at the critical target over a given time, is an important factor that helps us to understand the correlation between internal exposure and external (environmental) exposure in relation to toxicity. By comparing tissue doses in different species at similar exposures it also helps us to understand species differences in the sensitivity to chemicals, as well as inter-individual variations.

The chemical or its more water-soluble metabolites are primarily excreted via the kidneys or the biliary system. Volatile compounds may be exhaled. The great variety of processes observed during absorption, metabolism, distribution, and excretion cannot be predicted by modeling or by in vitro experiments without confirmatory data from animals and man.

Mode and/or Mechanism of Action

Identification of the possible modes or mechanisms by which a chemical induces toxicity and the dose–response relationship are essential to our understanding of species specificities, species differences, sensitive populations, or the interpretation of data regarding threshold or nonthreshold effects. They also help us to evaluate the relevance of the toxic effects to humans when the data are derived from experimental animals. Whereas the toxic mechanism is often not known in detail, modes of action, which can be described in a less restrictive manner, are undergoing consideration for inclusion in the risk-assessment process.

There is an array of mechanisms by which chemicals or any other stressors like heat or radiation can lead to toxicity. They may be differentiated as follows:

Physiological changes are modifications to the physiology and/or response of cells, tissues, and organs. These include mitogenesis, compensatory cell division, escape from apoptosis and/or senescence, inflammation, hyperplasia, metaplasia and/or preneoplasia, angiogenesis, alterations in cellular adhesion, changes in steroidal estrogens and/or androgens, and changes in immune surveillance.

Functional changes include alterations in cellular signaling pathways that manage critical cellular processes (e.g. modified activities for enzymes involved in the metabolism of chemicals such as dose-dependent alterations in phase I and phase II enzyme activities, depletion of cofactors and their regenerative capacity), alterations in the expression of genes that regulate key functions of the cell (e.g. DNA repair, cell cycle progression, post-translational modifications of proteins), regulatory factors that determine rate of apoptosis, secretion of factors related to the stimulation of DNA replication and transcription, or gap–junction-mediated intercellular communication.

Molecular changes include reversibility or irreversibility of changes in cellular structures at the molecular level, including genotoxicity. These may be formation of DNA

adducts and DNA strand breaks, mutations in genes, chromosomal aberrations, aneuploidy, and changes in DNA methylation patterns.

As indicated in the chapter on Toxicogenomics, data derived from gene expression microarrays or from high-throughput testing of agents for a single endpoint will become increasingly available and need to be evaluated for suitability for use in the hazard and risk-assessment process. As long as the information is not related to functional changes their applicability is poor and there is the possibility to overinterpret the effects observed, although they might be useful in assessing mechanisms. High-throughput data on specific endpoints may aid in the identification of common mechanisms of multiple agents.

Mechanistic information is most relevant for the evaluation and classification of carcinogens. If the carcinogenic effect is induced by a specific mechanism that does not involve direct genotoxicity, such as hormonal deregulation, immune suppression, or cytotoxicity, a detailed search for the underlying mode of action may allow identification of an NOEL. This can also be considered for materials, such as poorly soluble fibers, dusts, and particles, which induce persistent inflammatory reactions as a result of their long-term physical presence, which ultimately lead to cancer.

1.3.2 General Approach for Hazard Identification and Risk Assessment

Before starting any evaluation, structural alerts and physical-chemical parameters like water/lipid solubility and volatility need to be identified as well as the purpose of the hazard identification. To screen for specific effects such as relative cytotoxicity, mutagenicity, or hormonal effects, simple *in vitro* tests may be appropriate. This allows identification of specific wanted or unwanted effects and by that selection of useful compounds for further studies or their elimination.

For a more detailed evaluation (Figure 1.3) the stepwise procedure usually starts with the determination of the dose range for Lethality and the evaluation of genotoxicity by an *in vitro* bacterial test system (Ames test) and cytogenicity in mammalian cells. In the case of positive results the results are verified *in vivo* usually by the mouse bone marrow micronucleus test. In the case of structural alerts or questionable results the compound needs further evaluation by additional tests, including studies on toxicokinetics or potential genotoxic mechanisms.

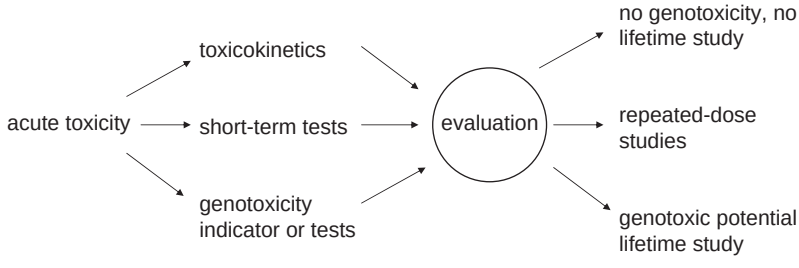


Figure 1.3 *Stepwise procedure to evaluate the toxic potential of a chemical [Modified from Greim and Deml, Toxikologie, Ch. 1, Fig. 1-3. Copyright (1996), with permission from Wiley-VCH].*

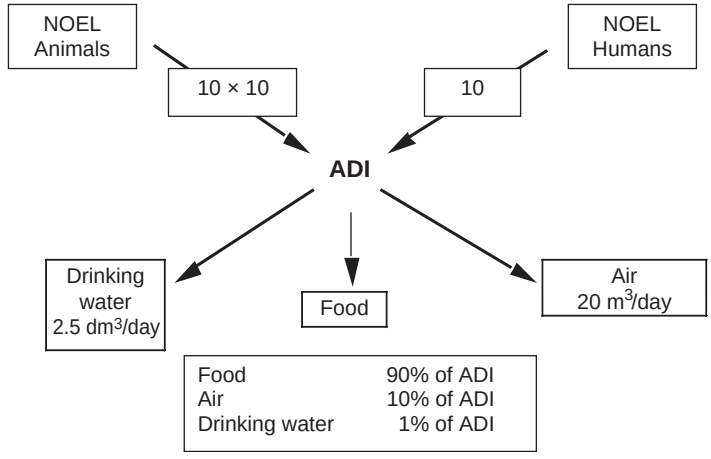


Figure 1.4 *Acceptable daily intake (ADI) and maximal acceptable concentration in air, food, and drinking water [Modified from Greim and Deml, Toxikologie, Ch. 25, Abt. 25-1. Copyright (1996), with permission from Wiley-VCH].*

The information so far collected provides information on the reactivity of the test compound, its absorption and distribution in the organism, and possibly on critical targets. This allows the decision to be made as to whether the database is appropriate for further testing by repeated-dose studies in animals for 28 and 90 days, which, depending on their outcome and the intended use of the chemical, are followed by a 6-months or lifetime study to evaluate potential effects upon long-term exposure, including carcinogenicity.

This information finally allows appropriate risk assessment for a potential human exposure or to set acceptable exposure limits for risk management. For example, when the detailed toxicological evaluation can exclude genotoxic and carcinogenic effects the NOEL in long-term studies in experimental animals can be determined. This NOEL is the starting point to set the ADI, which is usually 100-fold below the NOEL (Figure 1.4). This factor considers a 10-fold difference between the sensitivity of the experimental animals and humans and another factor of 10 to take into account possible inter-individual differences among the human population. If the NOEL is derived from studies in man a factor of 10 is used to cover possible individual differences. These factors can be reduced if specific information is available to conclude that the species-species or intra-species differences are less than 10. From the ADI, permissible concentrations in food, drinking water, consumer products, indoor and outdoor air, and other environmental compartments may be established.

1.3.3 Toxicological Issues Related to Specific Chemical Classes

Jurisdictions and regulatory agencies around the world have established a variety of guidelines for risk assessment and permissible exposure standards for chemicals in the workplace, the home, and the general environment. Regulatory decision-making depends upon the estimation of health risks from chemical exposure.

Health risks of chemicals designed for specific applications, e.g. consumer products, drugs, or pesticides, must be assessed when people are exposed in the many types of

environment in which people can be found. Therefore all elements of risk assessment: hazard identification, dose-response, exposure, and the risk, have to be thoroughly evaluated.

Data requirements for **new and existing chemicals** usually depend on annual production rate and the extent of human exposure. When there is considerable exposure regulatory requirements demand an extensive toxicological evaluation of the potential adverse effects of the specific chemical and the likelihood of their expression under the conditions of use or exposure and the definition of the MOE or the health risk under defined conditions of exposure.

For **drugs** special emphasis must be placed on efficacy, therapeutic index, potential side effects, and the effects of overdosage.

For **pesticides** the relative impacts of the chemical on the target versus on people is a critical requirement. Thus, the NOEL for people must be established, and an ADI must be determined because of the possibility of contamination of food and other consumer products with the pesticide, and the margin of safety needs to be established.

Exposure to **chemicals at the workplace** is, according to law, controlled by the Occupational Safety and Health Administration (OSHA) in the United States and by the Chemicals Law Act (1992) in Europe. Various governmental and nongovernmental institutions are involved in setting occupational exposure standards. Since the institutions publish the complete toxicologically relevant information and a justification for the proposed limit value these documentations are valuable sources for the toxicological database of the compounds. Institutions that publish these documents are listed in Table 1.1.

1.3.4 Existing Chemicals

In 1992 the European Commission estimated that about 100,000 chemicals are in use. They are produced in quantities ranging from less than a ton to several million tons per year. Except for drugs and pesticides, toxicity data requirements for existing or new chemicals have not been regulated. Although it is the responsibility of the producer and downstream user to release safe products, there are high-volume products with a relatively small database. Several programs have been launched to obtain knowledge at least for compounds with high annual production rates. In the US the Environmental Protection Agency (EPA) has initiated an HVP (High Production Volume) program. In an international cooperation the Organization for Economic Cooperation and Development (OECD) has launched the ICCA program, which evaluates and documents the available information on environmental and human health hazards and risks for about 1000 chemicals. In Europe, Risk Assessment Reports under the Existing Chemical Program of about 150 compounds are being produced and the REACH regulation (see below) will be set into action by 2008. A list of institutions that evaluate toxicity of chemicals and publish the results in comprehensive English documentations is given in Table 1.2.

REACH

REACH (Registration, Evaluation and Authorization of Chemicals) of the European Union is to identify hazardous properties of substances and to evaluate the risks of human and environmental exposure. The regulation will become effective by 2008. It is the responsibility of the producer or downstream user to provide the necessary information to the 'Agency'. As indicated in Table 1.3, the extent of toxicological information largely depends on the annual production rate of a chemical. As long as there is no indication

Table 1.1 *The new DFG/MAK criteria for the classification of carcinogenic chemicals compared with those from international bodies.*

EU	DFG/MAK	ACGIH/TLV	IARC
1 Substances known to be carcinogenic to man	**1** Substances which cause cancer in man	**A1** Confirmed human carcinogen	**1** The agent is carcinogenic to humans
2 Substances which should be regarded as if they are carcinogenic to man	**2** Substances which are considered to be carcinogenic for man	**A2** Suspected human carcinogen	**2A** The agent is probably carcinogenic to humans
3 Substances which cause concern for man owing to possible carcinogenic effects	**3** Substances which cause concern that they could be carcinogenic for man but which cannot be assessed conclusively because of lack of data	**A3** Animal carcinogen	**2B** The agent is possibly carcinogenic to humans
3b Substances which are insufficiently investigated			
3a Substances which are well investigated	**4** Substances with carcinogenic potential for which genotoxicity plays no or at most a minor role. No significant contribution to human cancer risk is expected, provided that the MAK value is observed.	**A4** Not classifiable as a human carcinogen	**3** The agent is not classifiable as to its carcinogenicity to humans
	5 Substances with carcinogenic and genotoxic potential, the potency of which is considered to be so low that, provided that the MAK value is observed, no significant contribution to human cancer risk is to be expected		
		A5 Not suspected as a human carcinogen	**4** The agent is probably not carcinogenic to humans

Abbreviations: DFG/MAK: Deutsche Forschungsgemeinschaft (German Commission for the Investigation of Health Hazards of Chemical Compounds in the Work Area); EU: European Union (EU); IARC: International Agency for Research on Cancer (IARC); ACGIH/TLV: American Conference of Governmental Industrial Hygienists / Threshold Limit Values Committee (ACGIH). For contacts see Table 1.2.

of a specific risk the chemicals will be registered for the intended use. Special attention will be paid to carcinogens, mutagens, and reproductive toxins (CMR compounds) and to other serious toxic effects as well as to chemicals which show bioaccumulation, persistence, and toxicity (BPT compounds) in the environment. The specific use of such compounds needs to be authorized. Within the first 3 years after implementation the >1000 t/a chemicals need to be submitted, during the 2nd three-year phase the <100 t/a, and within the 3rd phase the 10–100 t/a chemicals.

Table 1.2 *International institutions that publish documentations on chemicals.*

Institution	Contact
ACGIH (American Conference of Governmental Industrial Hygienists)	http://www.acgih.org/TLV/
ATSDR (The Agency for Toxic Substances and Disease Registry)	http://www.atsdr.cdc.gov/
BUA – Advisory Committee on Existing Chemicals (of the GDCh, the German Chemical Society)	http://www.gdch.de/taetigkeiten/bua__e.htm; http://www.gdch.de/taetigkeiten/bua/berichte__e.htm
The Canadian Centre for Occupational Health and Safety	http://www.ccohs.ca/
Dutch Expert Committee on Occupational Standards (DECOS),	http://www.gr.nl/adviezen.php?Jaar=2006
Environmental Protection Agency (EPA)	http://www.epa.gov/
European Centre for Ecotoxicology and Toxicology of Chemicals	http://www.ecetoc.org/
HSE (UK Health and Safety Executive)	http://www.hse.gov.uk/
International Programme on Chemical Safety	http://www.inchem.org/ International Agency for Research on Cancer (IARC) Summaries and Evaluations Concise International Chemical Assessment Documents (CICADS) Environmental Health Criteria (EHC) monographs Health and Safety Guides (HSG) JECFA (Joint Expert Committee on Food Additives) - monographs and evaluations JMPR (Joint Meeting on Pesticide Residues) - monographs and evaluations. OECD Screening Information Data Sets (SIDS)
The Japanese Association of Industrial Health	http://joh.med.uoeh-u.ac.jp
MAK Commission (German Research Foundation)	http://www.mak-collection.com http://www.dfg.de/mak
NIOSH	http://www.cdc.gov/niosh/homepage.html
The Nordic Expert Group	http://www.nordicexpertgroup.org//
OSHA	http://www.osha.gov/
SCOEL - EC Scientific Committee on Occupational Exposure Limits	http://ec.europa.eu/employment_social/health_safety/docs_en.htm

The Globally Harmonized System (GHS) contains classification criteria and hazard communication elements and is applicable for REACH (http://www.unece.org/trans/danger/publi/ghs/ghs_rev00/00files_e.html).

Table 1.3 *Use- [tons of annual production(t/a) + import] dependent data requirements of the European REACH regulation. LLNA: local lymph node assay.*

Annual production (Years)	1–10 t/a	10–100 t/a (2014–2017)	<100 t/a (2011–2013)	>1000 t/a (2008–2010)
Toxicokinetics		assess	assess	assess
Acute toxicity		oral, dermal, inhalation	oral, dermal, inhalation	oral, dermal, inhalation
Irritation	skin/eye *in vitro*	skin/eye *in vivo*	skin/eye *in vivo*	skin/eye *in vivo*
Sensitization	LLNA	immunotoxicity	immunotoxicity	immunotoxicity
Repeated dose		28 days	90 days	≥12 mo
Genotoxicity	bacterial test	cytogenetic, gene	*in vitro/in vivo*	mutagenicity,
Carcinogenicity	*in vitro*	mutation *in vitro*	mutagenicity	carcinogenicity
Reproduction		reproduction/ development screening test, developmental toxicity	development 2nd-generation	development 2nd-generation

The TTC Concept

The Threshold of Toxicological Concern (TTC) is a concept to establish a level of exposure for chemicals, regardless of their chemical-specific toxicity data, below which there is no appreciable risk to human health. The concept is based on knowledge of the chemical structure for evaluation of structural alerts, the amount of a specific chemical in a product, and the daily human exposure.

So far the TTC is applied for chemicals in food. It is defined as a nominal oral dose which poses no or negligible risk to human health after a daily lifetime exposure. At a mean dietary intake below the level of the TTC, toxicology safety testing is not necessary or warranted. By that the TTC concept can contribute to a reduction in the use of animals for safety tests. The TTC concept may also represent an appropriate tool to evaluate or prioritize the need for toxicological testing. There is ongoing discussion on its general applicability for safety assessment of substances that are present at low levels in consumer products such as cosmetics of impurities or of degradation products.

1.3.5 Classification of Carcinogens

The systems for classification of carcinogens used by various national or international institutions were developed in the 1970s. Classification is based on qualitative criteria, and reflects essentially the weight of evidence available from animal studies and epidemiology. Classification is usually based on the certainty with which a carcinogenic potential for a chemical can be established. Generally, three categories, the definitions of which slightly differ, are used (Table 1.3):

- human carcinogens;
- animal carcinogens, reasonably anticipated to be human carcinogens;
- not classifiable because of inadequate data.

For classification, the mode of action and potency of a compound are either not taken into account, or at best are used as supporting arguments. The advancing knowledge of reaction mechanisms and the different potencies of carcinogens has initiated a reevaluation of the traditional concepts.

The International Agency for Research of Cancer (IARC 2006) and the OECD propose to use data on the carcinogenic mechanism and potency in decision-making. Similarly, the General Directorate of Employment of the European Union is currently discussing the application of information on carcinogenic mode of action and potency as criteria for a revised classification. The European Food and Safety Authority has proposed to determine the margin of exposure (MOE) between the effect level obtained from animal studies and human exposure assuming that carcinogenicity is 'of minor concern' below a certain dose. Recently, the US Environmental Protection Agency (EPA) and a committee of the German Research Foundation (Deutsche Forschungsgemeinschaft) recommended consideration of the mode of action and have published modified concepts for classification. These activities in part originate from the recognition that one can distinguish between mechanisms of carcinogenicity caused by nongenotoxic and genotoxic carcinogens. Thus, it is possible to identify a NOEL for nongenotoxic carcinogens, provided that there is sufficient information on the primarily nongenotoxic mechanism. The American Conference of Governmental Industrial Hygienists (ACGIH 1997) has used a concept that considers carcinogenic potency for classification since 1995.

To determine the potency of genotoxic carcinogens and cancer risk at a given exposure a linear or sub-linear extrapolation from the high-dose effects observed in animals to the usually lower human exposure is requested by regulatory agencies. Recently, the European Food and Safety Authority (EFSA) recommended avoiding this extrapolation because of its inherent uncertainties. Instead, the margin of exposure (MOE) between a benchmark dose, or the T25 (dose at which 25% of the animals developed tumors) calculated from a carcinogenicity study in animals and human exposure should be determined. An MOE of 10,000 and more would reduce concern for carcinogenicity to be that of only a minor concern. The advantage is that neither a debatable extrapolation from high to low doses needs to be performed nor are hypothetical cancer cases calculated.

1.4 Summary

Toxicology is charged with describing the adverse effects of chemicals in a qualitative sense, and with evaluating them quantitatively by determining how much of a chemical is required to produce a given response. Fundamental to understanding Toxicology are the definitions of hazard, exposure, and risk. Since humans or organisms in the environment can be exposed via different routes, the concentrations in the different environmental compartments are a prerequisite for appropriate risk assessment. This needs to be specifically recognized, since the sensitivity of measurements in analytical chemistry has advanced to the point where infinitesimally small amounts of chemicals can be detected and identified in the various media of human environment. When one understands the principles of toxicology it is obvious that the presence of a chemical does not necessarily imply a health hazard. Since the dose makes the poison, effects only occur when exposure exceeds the NOEL. This applies for chemicals that induce reversible

effects. If damage is not repaired the effect persists and accumulates upon repeated exposure. In such cases every exposure represents a defined risk, which needs to be quantified.

There is an array of testing procedures to determine the hazardous properties of a chemical, such as acute, subchronic, and chronic toxicity, irritation and phototoxicity, sensitization and photosensitization, genotoxicity, carcinogenicity, or toxicity to reproduction. Information on the toxicokinetics and mechanisms of the toxic effects improves the relevance of the findings for man. More recent methodologies like toxicogenomics, or high-throughput testing of agents for a single endpoint, will become increasingly available and may improve hazard identification and may aid in the identification of common mechanisms of multiple agents.

Altogether, Toxicology describes the intrinsic properties of an agent (hazard identification) by applying conventional and substance-specific test procedures to estimate the amount of the chemical required to produce these effects (risk characterization).

References

Greim, H. and Deml, E. (1996), Toxikologie, Eine Einführung für Naturwissenschaftler und Mediziner, VCH Verlagsgemeinschaft mbH, Weinheim.

ParaceIsus, Selected Writings, ed. J. Jacobi, transl. N. Guterman (1951), p. 95, Princeton University Press, Princeton, NJ.

For Further Reading

EFSA (European Safety Authority): Opinion of the Scientific Committee on a request from EFSA related to A Harmonized Approach for Risk Assessment of Substances Which are both Genotoxic and Carcinogenic.

http://www.efsa.europa.eu/en/science/sc_commitee/sc_opinions/1201.html

IARC (International Agency for the Research of Cancer): Monographs on the evaluation of carcinogenic risks to humans, Geneva.

Kroes, R., Kleiner J., and Renwick A. (2005): The Threshold of Toxicological Concern Concept in Risk Assessment, *Toxicol. Sci.*, **86**, 226–230.

http://ecb.jrc.it/tgddoc/ Technical Guidance Document, European Chemicals Bureau.

US EPA Assessing Health Risks from Pesticides (Pesticides: Topical and Chemical Fact Sheets).

http://www.epa.gov/pesticides/factsheets/riskassess.htm

2

Principles

2.1 Toxicokinetics

Johannes G. Filser

2.1.1 Definition and Purpose

The object of toxicokinetics is the investigation of absorption, distribution, and elimination of toxicants and their toxicologically relevant metabolites as functions of dose and time.

In toxicokinetic studies, concentrations of the administered substance and its relevant metabolites in body fluids, organs, and excrement are determined in a time-dependent manner. If substances are volatile, concentration–time courses are monitored in the exhaled air. In vitro, concentration–time courses are examined in organs, cells, and cell fractions, and distribution and binding studies are carried out.

For toxicological studies, experimental animals are used. Since the probability of detection of a toxic or carcinogenic effect is a function of the dose, very high doses of the toxicant are used so that even with small numbers of animals adverse effects can be seen. The interpretation and extrapolation of the obtained dose–response curves to low doses and concentrations relevant in the human environment require knowledge of the ultimate active chemical species (the parent compound or a metabolite) and of its tissue burden as a function of dose and time. Therefore, toxicokinetic studies must be carried out with experimental animals, and very wide concentration ranges of the toxicant must be used. On the other hand, to allow appropriate extrapolation of the dose–response curves to man, it is also necessary to know the systemically available dose of the ultimately active

Toxicology and Risk Assessment: A Comprehensive Introduction Edited by H. Greim and R. Snyder

substance. For ethical reasons we are limited in the toxicokinetic studies that can be performed in people. However, it is possible to perform toxicokinetic experiments in vitro using organ and tissue fractions from experimental animals, and, when available, from man. Thus, animal studies, both in vivo and in vitro, and human in vitro studies, plus the use of physiological toxicokinetic models that take into account physiological???

2.1.2 Absorption, Distribution, and Elimination

Absorption means uptake of a substance into the lymph and the bloodstream. Distribution comprises both the transport of the substance with the circulating blood and its accumulation in organs and tissues. Elimination describes the removal of the substance from the organism. It covers two processes: the biotransformation (metabolism) of the substance into other products (metabolites) and the excretion from the organism (see Figure 2.1).

Once formed, metabolites can also be distributed in the bloodstream and finally eliminated.

The basic physiology of these processes is summarized below.

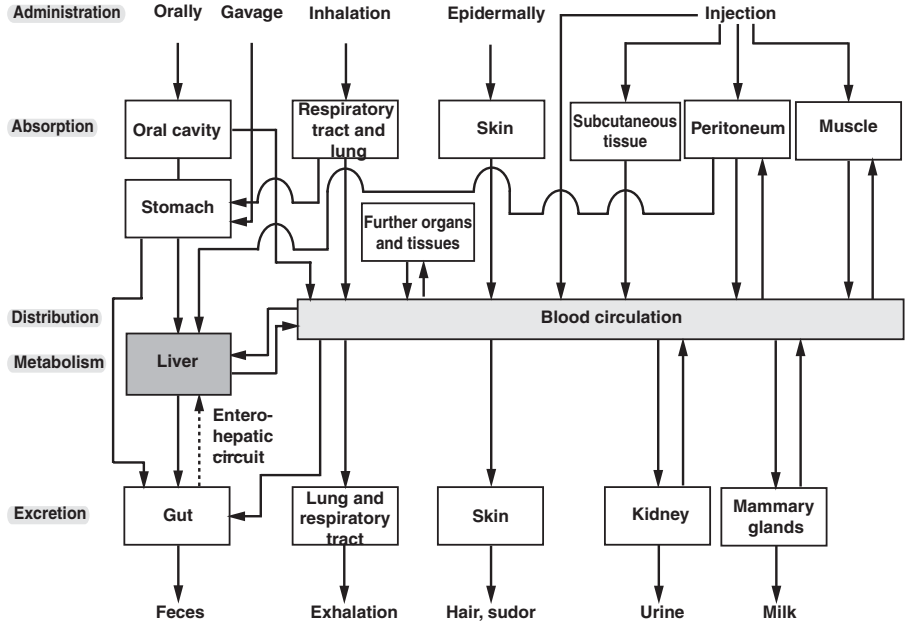


Figure 2.1 *Routes taken by substances in the organism: absorption, distribution, and elimination (metabolism and excretion).*

Passage Through Membranes

During the processes of absorption, distribution, and elimination, substances must pass through biological membranes.

Biological membranes separate morphologically and functionally differing entities (e.g., cells, nuclei, mitochondria, Golgi apparatus, endoplasmic reticulum). They consist primarily of a protein-containing phospholipid bilayer. The proteins serve to maintain the membrane structure and may also function as enzymes. Carriers that transport sub-strates through the membrane span the entire membrane. Some proteins form 'pores' that permit water to cross the membrane. The biological membrane is selectively permeable. Lipophilic and small polar molecules can easily diffuse through it. Others require specific transport mechanisms.

In decreasing order of importance, mechanisms for the transport of substances through a biological membrane include: passive diffusion, convective transport, active transport, facilitated diffusion, and pinocytosis.

Passive Diffusion Passive diffusion through membranes is most important for the transfer of substances and takes place in both directions. To be able to be transported by passive diffusion, the permeating substance must be in solution. The rates of diffusion are directly proportional to the concentrations on the two sides of the membrane, and to the surface area of the membrane, and are inversely proportional to the thickness of the membrane. At equilibrium, the amounts of substance diffusing per time unit through the membrane from the two sides are the same, but the concentrations on the two sides of the membrane need not necessarily be the same: the concentrations of a substance which accumulate in individual compartments of cells and tissues are determined by hydrophilic and hydrophobic interactions between the substance and its biological surroundings. Lipophilic substances, for example, accumulate in the lipids of an organism. A measure of the maximum possible accumulation of a substance in the various tissues is given by the partition coefficients. They are thermodynamic constants, which express the ratio of the concentrations of a substance in two phases in equilibrium. In Table 2.1 the partition coefficients muscle:blood and fat:blood for some hydrocarbons and the partition coefficients blood:air for their vapors are shown.

Ionized organic substances (e.g., quaternary ammonium compounds), which cannot normally diffuse through biological membranes, can nonetheless be absorbed from the gastrointestinal tract. It is assumed that these compounds combine with endogenous polyionic substances to form ion pairs, which are externally uncharged and so can diffuse passively through the membranes. Subsequently, the complexes are thought to dissociate. Such 'ion pair transport' would also be a passive diffusion process.

Convective Transport Convective transport or 'filtration' is understood to mean the permeation, through membranes, of substances dissolved in water, which flow with the water through pores of 7–10 Å diameter. This route is especially important for ionized

Table 2.1 *Partition coefficients muscle:blood and fat:blood for some hydrocarbons, and partition coefficients blood:air for their vapors; data for human tissues.*

	Partition coefficient		
Substance	muscle:blood	fat:blood	blood:air
n-Pentane	1.84	104	0.38
n-Hexane	6.25	130	0.80
n-Heptane	5.25	162	2.38
Benzene	2.23	51.5	7.37
Toluene	2.31	63.7	15.1
p-Xylene	1.61	51.9	38.9

Values were taken or calculated from a collection of blood:air and tissue:air partition coefficients (Meulenberg and Vijverberg, 2000). Partition coefficients fat:blood and muscle:blood were obtained by dividing the corresponding tissue:air partition coefficients by the respective blood:air partition coefficients.

substances, for small hydrophilic molecules such as urea, and for substances such as the lower alcohols with properties like those of water. In general, spherical and thread-like molecules with molecular weights up to about 150 and 400, respectively, can cross biological membranes by convective transport. Convective transport is another kind of passive transport; the driving force is mainly the difference in hydrostatic pressure between the two sides of the membrane. In addition, osmotic pressure plays an important role, and the charge on the walls of pores can influence the transport of charged particles.

Active Transport When a substance is transported through the membrane with the help of a carrier molecule in an energy-requiring process, the mechanism is described as active transport. A carrier is a membrane-bound enzyme, generally an adenosine triphosphatase, which binds the substance more or less specifically and transfers it to the other side of the membrane, where it is released.

Being a process catalysed by an enzyme, active transport obeys saturation kinetics (see below). This means that an increase in the dose is associated with an increase in the rate of transport only within a certain dose range. Active transport can also take place against a concentration gradient as it involves energy consumption.

Among the enzyme-catalysed processes is exchange diffusion, an energy-consuming process in which Na^+ is actively transported out of the cell in exchange for K^+, which enters passively. This sodium pump is responsible for reducing intracellular Na^+ in exchange for extracellular K^+.

Facilitated Diffusion Facilitated diffusion is also carrier-mediated and saturable. However, in this case the concentration gradient is the driving force. The energy reserves of the cell are not used.

Pinocytosis Pinocytosis is an active-transport mechanism by which larger aggregates are moved into the cell. The aggregate is enclosed in the cell membrane and incorporated by the cell in a vesicle. In the cytoplasm, the vesicle membrane is broken down and the aggregate released. In this way even solid objects such as plastic particles can be taken up.

Absorption

Of most toxicological significance for the uptake of chemicals by the human organism is absorption via the respiratory tract, the gastrointestinal tract, and the skin. In animal studies other routes also play a role.

Respiratory Tract

Foreign compounds (xenobiotics) can be inhaled not only when they are in gaseous form but also when they are particles, solid or liquid (aerosols), suspended in the air.

Gaseous substances can be absorbed throughout the respiratory tract and most especially in the alveoli (surface area of the alveolar epithelium is about 90 m^2 in an adult human and 3000 cm^2 in a rat of 250 g body weight). The alveoli are surrounded by a dense capillary net, which is perfused with the blood from the pulmonary circulation. The gas exchange between the alveolar space and the capillary blood proceeds extremely rapidly because the separating alveolocapillary membrane is less than 1 µm thick. This membrane consists of the alveolar epithelium, an interstitium with elastic fibers, and the capillary endothelium. Consequently, when neglecting any uptake in the upper airways, the inhalation and exhalation processes can be described by considering only the mass transfer between air and capillary lung blood (Figure 2.2). The amount of a gaseous substance taken up from the capillary blood per time unit

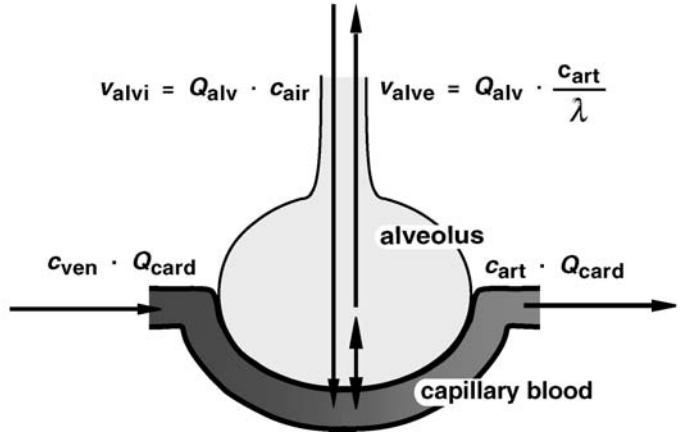


Figure 2.2 *Processes describing the fate of an inhaled gaseous substance in lung capillaries. c_{art}: concentration in oxygen-rich (arterial) blood; c_{air}: concentration in the inhaled air; c_{ven}: concentration in oxygen-poor (venous) blood; λ: partition coefficient blood:air; Q_{alv}: alveolar ventilation; Q_{card}: blood flow through the lung; v_{alve}: rate of alveolar elimination (exhalation); v_{alvi}: rate of absorption in the capillary blood (inhalation); ↕: immediate distribution between oxygen-rich capillary lung blood and alveolar air according to λ.*

(v_{alvi}) is equal to the product of the alveolar ventilation (Q_{alv}) and the concentration of the substance in the inhaled air (c_{air}) [Equation (2.1)]:

$$v_{alvi} = c_{air} \cdot Q_{alv} \tag{2.1}$$

The alveolar ventilation at rest is about 120 ml/min and 5 l/min in a rat of 250 g body weight and an adult human, respectively.

The amount of gaseous substance exhaled per time unit into the alveolar space (v_{alve}) is equal to the product of Q_{alv} and the concentration of the substance in the alveolar air that is given by the ratio of the substance concentration in the oxygen-rich blood (c_{art}) leaving the lung capillaries to the partition coefficient blood:air (λ) [Equation (2.2)]:

$$v_{alve} = Q_{alv} \cdot c_{art}/\lambda \tag{2.2}$$

The quantity of the substance entering the lung capillaries per time unit (v_{ven}) is equivalent to the product of the blood flow through the lung (Q_{card}) and the concentration of the substance in the oxygen-poor blood (c_{ven}) entering the lung capillaries [Equation (2.3)]:

$$v_{ven} = Q_{card} \cdot c_{ven} \tag{2.3}$$

Q_{card} is identical with the cardiac output - the total volume of blood pumped by the ventricle per time - which is about 83 ml/min in a rat (250 g body weight) and 6.2 l/min in a human (70 kg body weight) at rest.

The absorbed amount of the substance that leaves the lung capillaries per time unit (v_{abs}) in the oxygen-rich blood and enters the residual body is equivalent to the product of Q_{card} and c_{art} [Equation (2.4)]:

$$v_{abs} = Q_{card} \cdot c_{art} \tag{2.4}$$

Considering all four processes and taking into account that the amounts entering the capillaries equal those leaving them per time unit, Equation (2.5) is obtained:

$$Q_{card} \cdot c_{ven} + c_{air} \cdot Q_{alv} = c_{art} \cdot Q_{alv}/\lambda + c_{art} \cdot Q_{card} \tag{2.5}$$

$Q_{card} \cdot c_{ven}$ is zero during the first inhalation process when there is still none of the xenobiotic in the organism. Under this condition the absorption of the inhaled substance into the oxygen-rich blood occurs with maximum rate (v_{absmax}). From Equations (2.4) and (2.5) v_{absmax} is obtained as in Equation (2.6):

$$v_{absmax} = c_{air} \cdot [Q_{alv} \cdot \lambda \cdot Q_{card}/(Q_{alv} + \lambda \cdot Q_{card})] \tag{2.6}$$

From this equation it may be seen that, for large values of λ, v_{absmax} becomes the product of only c_{air} and Q_{alv}, and at small values of λ, however, it becomes independent of Q_{alv}, (resulting in the product of c_{air}, λ, and Q_{card}). Only at high values of λ is the maximum rate of absorption of gaseous substances limited by the alveolar ventilation. If λ is small, estimation of v_{absmax} from only the alveolar ventilation yields values which are too high (Figure 2.3). During intensive physical exercise, which leads to an increase especially in Q_{alv}, the v_{absmax} for substances with small λ increases less than that for those with large λ.

With the duration of exposure at constant c_{air}, the concentration of the substance in the oxygen-poor venous blood (c_{ven}), which flows from the heart to the lung, increases to a plateau determined not only by λ but also by the rate of metabolic elimination of the

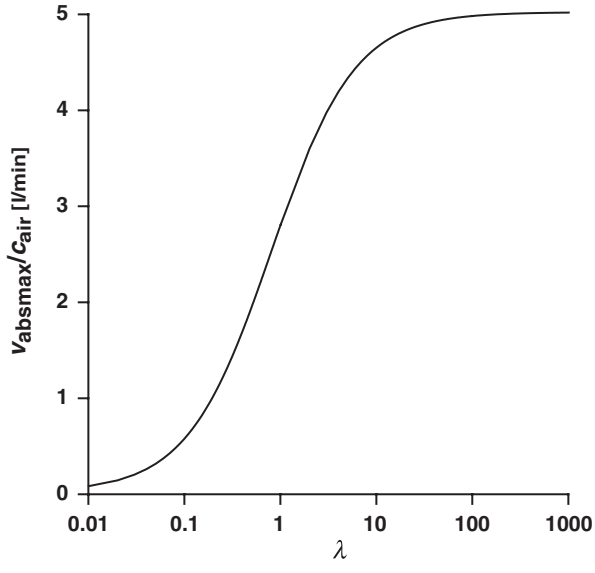


Figure 2.3 *Ratio of the maximum rate of absorption (v_{absmax}) of an inhaled gaseous substance to the concentration of the substance in the inhaled air (c_{air}) as a function of the blood:air partition coefficient (λ) in persons at rest. Only when λ is larger than about 10 is v_{absmax} given by the product of the alveolar ventilation (5 l/min at rest) and c_{air}. At small values of λ, v_{absmax} can be much smaller than this product.*

substance. The effectiveness of metabolic elimination can be very different for different substances. During the time to the plateau of c_{ven} the absorption rate v_{abs} decreases from the initial v_{absmax} and then remains constant till the end of exposure.

Particles can become airborne when they are smaller in size than 100 µm. When inhaled by man, the absorption of particles depends strongly upon their size and the kind of breathing (nose or mouth). Particles larger than about 2.5 µm are almost exclusively filtered by the nose under conditions of nose breathing. In mouth breathing, even large particles reach the bronchi. Generally, particles can penetrate into the alveoli when they are smaller than about 2 µm. Some of the particles, which are deposited in the tracheobronchial tree, are transported by the ciliated epithelium into the throat where they are swallowed and can then be absorbed in the gastrointestinal tract. Soluble particles can be absorbed directly through the epithelium of the respiratory tract into the bloodstream.

Oral Cavity, Gastrointestinal Tract

In man, the absorption of toxic substances through the oral cavity and from the gastrointestinal tract takes place when persons have eaten contaminated food or accidentally ingested the toxin. It is also relevant for toxicological studies with laboratory animals to which emulsified or dissolved xenobiotics are administered by gavage.

In man, uncharged and sufficiently lipophilic substances can readily enter the systemic circulation through the oral mucosa, which is well supplied with blood vessels and has an effective surface area of about 0.02 m^2 in adults. The administration of chemicals into the stomach by means of a tube (gavage) is a standard method in animal studies. As a rule, substances are taken up rapidly through the mucous membrane of the stomach, which has an acid milieu, and that of the small intestine, which is weakly alkaline, and transferred with the blood via the portal vein to the liver, the main metabolic organ of the body. However, some substances, especially some heavy metals, are only poorly absorbed from the gastrointestinal tract. Substances which are not stable in acid, such as epoxides, can be rapidly hydrolysed in the stomach, resulting in reduced absorption. Many compounds are subjected to a first biotransformation in the intestinal walls and, even more so, in the liver (**first-pass effect**). Not until they have passed through the liver can they, or at least the fraction which remains unmetabolized, enter the blood of the systemic circulation. Because of the first-pass effect, usually only part of the dose of substances administered orally becomes available systemically. This phenomenon is described as reduced **bioavailability**. The bioavailability expresses the fraction of the dose which reaches the systemic blood circulation. When the substance is absorbed through the lungs or the skin, the bioavailability is usually not or is only slightly reduced.

Skin

The rate of absorption of a substance through the skin (surface area about 300 cm^2 for a rat of 250 g and 1.8 m^2 for an adult man of 70 kg body weight) depends on the physicochemical properties of the substance and the state of the skin. The stratum corneum, the outermost layer of the skin, made up of stacked interconnected dead corneocytes embedded in a lipohilic matrix, forms the main barrier and represents the rate-determining step for the passage of xenobiotics through the skin.

There are four possible routes of percutaneous diffusion of substances: through the cells (transcellular) or between the cells (intercellular) of the stratus corneum, through the excretory ducts of the sebaceous and sweat glands (transglandular), and along the hair shafts through the hair follicles (transfollicular). The last two possibilities are routes through pores in the skin. As the pores account for only 0.1% of the surface area of the human skin, these routes are of less importance than those through the stratum corneum. Of these, hydrophilic compounds prefer the transcellular and lipophilic substances the intercellular route. In general, the hydro- or lipophilicity of a substance and its molecular size determine its ability to penetrate the skin. Additionally, the permeability of the skin can be altered by the xenobiotic, for example as a result of swelling or defatting.

Water penetrates the stratum corneum only very slowly. Other substances may penetrate relatively rapidly, especially lipophilic substances such as organophosphates and polychlorinated biphenyls. For most gaseous substances, percutaneous absorption is only a few percent of that from inhalation. High rates of transdermal uptake are found for vapors of some volatile amphiphiles which have very high partition coefficients between tissue:gas phase. For example, under conditions of whole body exposure the vapors of 2-butoxyethanol and of dimethylformamide are taken up more rapidly through the skin than by inhalation.

The rate of absorption through the skin (i.e., the stratum corneum) may be described in a simplified way in terms of Fick's first law [Equation (2.7)]:

$$J = k_p \cdot \Delta c \tag{2.7}$$

J is designated as the flux. It expresses the quantity of substance that penetrates the skin per unit area and time. Δc is the difference between the concentrations of the substance on the two sides of the stratum corneum; k_p is the permeability constant (dimension: distance/time). Normally the concentration under the stratum corneum is assumed to be negligible. This yields the generally used Equation (2.8):

$$J = k_p \cdot c \tag{2.8}$$

Here, c is the concentration of the substance on the surface of the skin. The value of k_p depends on the properties of the substance and the state of the stratum corneum.

Intravascular, Subcutaneous, Intramuscular, and Intraperitoneal Injection

In animal studies, toxicants are often injected into a blood vessel (intravascular), directly under the skin (subcutaneous), into the muscles (intramuscular), or into the abdominal cavity (intraperitoneal).

Intravascular, mostly intravenous, injection of a substance bypasses the absorption phase so that the administered dose is systemically available at once.

Substances administered by subcutaneous or intramuscular injection are transferred relatively slowly into the blood so that, although the bioavailability is the same, the initial concentrations in the blood are lower than after intravascular injection. Substances injected intraperitoneally are absorbed through the peritoneum, a membrane that consists of a thin epithelium covering a layer of well vascularized connective tissue and which lines the abdominal cavities (peritoneum parietale) and covers the abdominal viscera (peritoneum viscerale). In man and rat the peritoneum has a surface area of about $1\,m^2$ and $500\,cm^2$, respectively. The movements of the intestines distribute the injected substance rapidly over the peritoneum. Lipophilic substances are absorbed very rapidly. After passage through the peritoneum, the absorbed substance can be transported in the lymph ducts or bloodstream in two directions: via the lymph ducts and blood vessels of the membrane covering the walls of the abdomen (peritoneum parietale) the absorbed substance is transported to the heart and so directly into the **systemic circulation**, and via the blood vessels of the membrane covering the peritoneum viscerale the absorbed substance is transported through the portal vein to the liver. Because of this passage through the liver, after intraperitoneal application a first-pass effect may be seen.

Distribution

The substances are distributed in the organism by the bloodstream.

At continuous uptake, substances accumulate in the organism until **steady state** is reached, i.e. until the concentration-dependent rate of elimination (see below) equals

the rate of uptake. The length of time required until steady state depends on the physicochemical and biochemical properties of the substance and can be very different for different substances; in particular, if the substance has a high affinity for poorly vascularized tissues and if it is slowly or not metabolized the time until steady state can be very long. Lipophilic substances accumulate in the adipose tissue, and heavy metals such as strontium or lead with chemical properties similar to those of calcium are stored in bones in the form of poorly soluble salts. Many substances can bind reversibly to plasma and tissue proteins.

After a single dose or a brief exposure, redistribution of a substance between tissues can take place. The substances are first distributed in the well perfused organs such as brain, heart, liver, lungs, spleen, and kidneys. The accumulation in tissues which are less well perfused occurs more slowly, and continues despite the decline in the concentration in the blood and in well perfused organs exposure ceases [Figure 2.4(a)]. When a substance is absorbed continuously at a constant rate, as it is during exposure to a constant concentration in the inhaled air or during continuous infusion under constant conditions, the tissues become saturated with the substance in sequence. Then, after the end of exposure or infusion, redistribution does not take place. Instead the concentrations decrease in all tissues, first more rapidly in the blood and in the better perfused organs. Finally, in the terminal elimination phase, the concentrations decrease in a parallel manner in all tissues. This phase is often determined by the slowest of the processes which release the substance into the blood. With well metabolized lipophilic substances, e.g. styrene, this is generally the release of the substance from the poorly perfused adipose tissue that serves as a storage organ [Figure 2.4(b)].

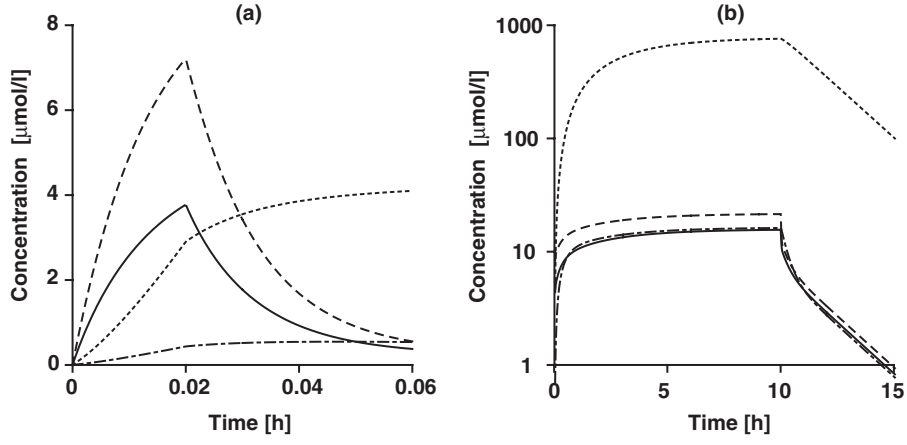


Figure 2.4 *Modeled concentration–time courses of inhaled styrene in various tissues of a rat during and after exposure to styrene vapor at a concentration of 100 ppm in the inhaled air. (a) Duration of inhalation exposure 0.02 h; post-exposure period 0.04 h; linear plot. (b) Duration of inhalation exposure 10 h; post-exposure period 5 h; half-logarithmic plot. —: Oxygen-poor (venous) blood; – – –: richly perfused organs; –·–·–: moderately perfused tissues (e.g. muscles); ··· : adipose tissue. Curves were calculated with a physiologically based toxicokinetic model for styrene (Csanády et al., 1994).*

Elimination

The two elimination routes are conversion of substance into metabolites and excretion of the unchanged substance and of metabolites from the organism.

Metabolism Metabolism or biotransformation is generally catalysed by enzymes. It can take place in all organs and tissues. For the biotransformation of most xenobiotics, however, the liver plays the quantitatively most important role.

Metabolites are other substances than the initial one; they have their own chemical and toxicological properties and their own toxicokinetic profiles. Frequently it is a metabolite that is considered to be responsible for the toxic or carcinogenic effect of a substance.

Saturation kinetics

Enzyme-catalysed metabolism obeys saturation kinetics which, in the simplest case, can be described by Michaelis–Menten kinetics.

The Michaelis–Menten equation expresses the rate of an enzyme-mediated substance conversion (dc/dt; dimension: concentration/time) as a function of the concentration of the substrate (c) at the enzyme, the maximum change of substance concentration per time (S_{max}), and the so-called Michaelis constant (K_m) [Equation (2.9)]:

$$dc/dt = (S_{max} \cdot c)/(K_m + c) \tag{2.9}$$

K_m equals that value of c at which half of S_{max} is attained and is a measure of the affinity of the substrate for the enzyme: the smaller the value of K_m, the larger is the affinity. In the low concentration range where c is much smaller than K_m, the reaction obeys **first-order kinetics**. This means that dc/dt is directly proportional to c as may be seen from Equation (2.9), when $c \ll K_m$, so we have Equation (2.9a):

$$dc/dt \cong (S_{max}/K_m) \cdot c \qquad (\text{when } c \ll K_m) \tag{2.9a}$$

Here the proportionality factor is identical with the ratio S_{max} / K_m.

With increasing c, direct proportionality is no longer found; dc/dt increases more slowly than c and reaches 91% of S_{max} when c is 10 times the K_m. Further increases in c have only a little effect on dc/dt. When S_{max} is attained, dc/dt obeys **zero-order kinetics**. In a zero-order reaction, dc/dt is independent of c and is constant [Equation (2.9b)]:

$$\lim_{c \to \infty} dc/dt = S_{max} = \text{constant} \tag{2.9b}$$

The value of S_{max} is proportional to the total concentration of the enzyme involved.

Simultaneous presence of other substances which can interact with the enzyme can reduce the rate of conversion. The most important example of such enzyme inhibition is **competitive inhibition** in which a second substance competes with the substrate for the binding site of the enzyme. This process causes a reduction in the value of dc/dt below that for the substrate on its own. The properties of the enzyme itself are not changed; however, the apparent K_{mih} values obtained are higher than the real K_m value determined in the

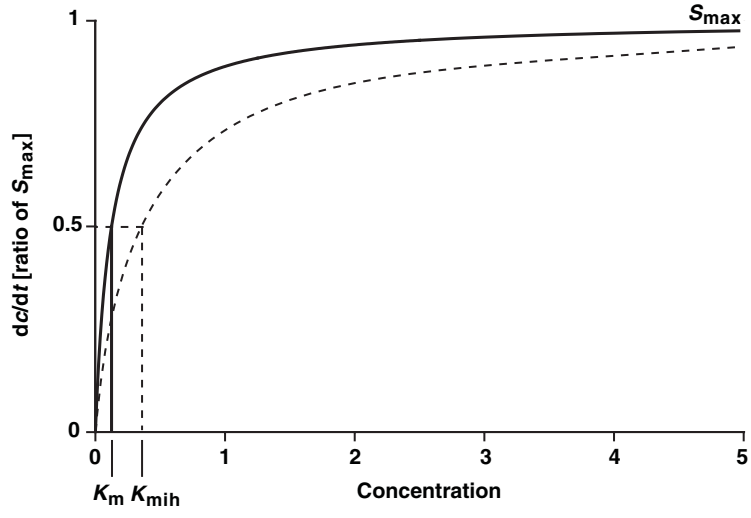


Figure 2.5 *Rate of conversion of a substrate as a function of its concentration: Michaelis–Menten kinetics in the presence and absence of a competitive inhibitor (linear plot). —: Without inhibitor; – – –: with inhibitor. K_m: Michaelis constant: substrate concentration at half-maximum conversion rate; K_{mih}: apparent K_m in the presence of a certain concentration of the inhibitor; S_{max}: maximum change of substance concentration per unit of time.*

absence of the inhibitor. In Figure 2.5, the relationship between $\mathrm{d}c/\mathrm{d}t$ and c is shown for Michaelis–Menten kinetics in the absence and presence of a competitive inhibitor.

Excretion

Excretion takes place mainly via the kidneys, the intestinal tract and the lungs, but the skin and mammary glands may also be involved.

Kidneys

Substances below a certain molecular size are excreted via the healthy kidney; in man these are mostly substances with molecular weights below 300.

Many substances are converted by Phase I and Phase II reactions into more water-soluble compounds, which are excreted in the urine. Both passive- and active-transport processes can be involved in the excretion from the blood plasma via the kidney into the urine. In the passive processes, the average rate of excretion with the urine (dimension: amount/time) is directly proportional to the concentration of the free substance in the blood plasma (c_P). This is also the case for active processes if the concentration of the substance is markedly smaller than the K_m value of the active, enzyme-mediated transport process (see above). The proportionality factor between rate of urinary

excretion and c_P is called the **renal clearance** (Cl_{ren}). To determine Cl_{ren}, c_P is measured in the middle of a relatively short urine collection period. In addition, the quantity of substance eliminated with the urine during this period (Δt) is determined. It is the product of the measured urine volume (V_{ur}) and the concentration of the substance in the urine (c_{ur}). Thus, Cl_{ren} is given by Equation (2.10).

$$Cl_{ren} = (V_{ur} \cdot C_{ur})/(\Delta t \cdot c_P) \tag{2.10}$$

In toxicokinetics, clearance (Cl) is given the dimension volume / time. It links a rate (v, dimension: amount/time), with a concentration (c, dimension: amount/volume); as shown in Equation (2.11).

$$Cl = v/c \tag{2.11}$$

Cl is a constant over the concentration range in which the elimination obeys first-order kinetics and thus v is directly proportional to c. When saturation kinetics applies, the clearance becomes smaller with increasing concentration because the rate of the active process tends gradually towards a maximum rate and thus becomes independent of the concentration. The concentration-dependence of the clearance when kinetics according to Michaelis–Menten apply is given by Equation (2.12) [cf. Equations (2.9) and (2.11)],

$$Cl = V_{max}/(K_{mapp} + c) \tag{2.12}$$

where V_{max} is the maximum rate (dimension: amount/time) and the concentration K_{mapp} is called the apparent Michaelis constant. (The abbreviation K_{mapp} demonstrates that it is not the true enzyme-specific K_m).

Generally, a clearance is a measure of the elimination of a substance. Clearances can be defined not only for the kidney but for any eliminating organ or tissue and for the sum of such organs and tissues. As the clearance is related to a concentration, in all cases the medium in which the concentration was determined, mostly plasma or blood, must be given. Frequently, metabolic clearance is determined; this parameter enables the calculation of the rate of metabolism in a certain organ (e.g., liver) or the whole organism in association with the concentration of the substance in the medium. The sum of all clearances from the organism, if related to the concentration in a single medium (e.g., plasma), is the so-called total clearance (Cl_{tot}).

Intestinal tract

The intestinal tract includes the small and large intestines. Substances which are excreted with the feces can have entered the intestinal tract from the liver via the bile or have been excreted directly through the intestinal membrane.

Whereas low-molecular-weight substances are mostly excreted via the kidneys (see above), substances with molecular weights greater than 300 are mostly transferred into the intestines with the bile. The bile is actively secreted by the parenchymal cells of the liver. To be excreted with the bile, substances must not only be of a certain minimum size but must also contain a polar group. Both conditions may be fulfilled after the substance has undergone Phase II conjugation reactions. Of most quantitative significance is the conjugation with glucuronic acid, which is catalysed by the enzyme

UDP-glucuronyltransferase. The concentration of xenobiotics in the bile can be a multiple of that in plasma and so high concentrations of toxic substances can be found in the hepatobiliary system. Some of the substances that enter the intestines with the bile are metabolized by bacterial enzymes. For example, glucuronic acid conjugates can be cleaved by bacterial β-glucuronidases. The substances so released can be absorbed and transferred via the portal vein to the liver and, after conjugation with glucuronic acid, are excreted with the bile back into the intestine where they are released again. This is the so-called **enterohepatic circulation** (Figure 2.1).

Substances can enter the intestines directly through the intestinal mucosa, as has been demonstrated both for the cardiotonic glycoside digoxin and for mercury. This route is particularly significant for the elimination of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin and its congeners and for the very slowly metabolized biphenyls. These very lipophilic substances, which are excreted in feces lipids, probably enter the lipids in the intestinal contents by passive diffusion.

Lung

Volatile substances and gases can be eliminated via the lung by exhalation.

Not only volatile metabolites but also inhaled gases and vapors can be excreted by respiration. The rate of alveolar elimination [v_{alve}; Equation (2.2)] depends not only on Q_{alv} and λ but also on the concentration of the substance in the oxygen-rich blood (c_{art}) which flows from the lung via the heart into the arteries of the body. It is obtained by reformulating Equation (2.5) to Equation (2.13):

$$c_{art} = (c_{ven} \cdot Q_{card} + c_{air} \cdot Q_{alv})/(Q_{card} + Q_{alv}/\lambda) \qquad (2.13)$$

In its turn c_{art} depends on the rate of alveolar absorption from the air [v_{alvi}; Equation (2.1)], v_{alve}, Q_{card}, and c_{ven}, the concentration of the substance in the oxygen-poor blood which enters the lung capillaries [Equation (2.3) and Figure 2.2]. The extent of metabolic elimination has a considerable influence on c_{ven} and, consequently, also on v_{alve}.

The ratio of the amount of a volatile substance eliminated by exhalation to that eliminated metabolically is substance-specific and concentration-dependent if saturation kinetics occurs. It can have very different values. Some substances (e.g., 1,1,1-trichloroethane) are exhaled mainly unchanged; others (e.g., styrene inhaled at concentrations of less than 200 ppm) are eliminated predominantly by the metabolic route.

Skin

Perspiration is also an elimination route.

The final product of normal protein catabolism, urea, and a number of drugs have been shown to be eliminated by this route. Arsenic and thallium can be readily detected in hair because they accumulate in the skin and its appendages.

Mammary glands

The elimination of xenobiotics with contaminated breast milk can result in an internal exposure of the baby.

Passive diffusion is the main mechanism by which xenobiotics enter mother's milk, the pH of which (6.6) is slightly lower than that of the organism (7.4). Not only many medicines (e.g., antibiotics) but also everyday drugs such as nicotine, ethanol, and environmental substances such as heavy metals, polyhalogenated biphenyls, polychlorinated dibenzodioxins, and polychlorinated dibenzofurans (PCDD/PCDF) are eliminated with the milk. In particular, the elimination of the slowly metabolized, highly lipophilic PCDD/PCDF in milk fat has the effect of markedly reducing the body burden of these substances in the breast-feeding woman. Reversely, in breast-fed babies, higher PCDD/PCDF levels have been determined than in those who were not breast-fed.

2.1.3 Toxicokinetic Models

Toxicokinetic models describe the fate of substances (absorption, distribution, elimination) by mathematical functions. Mostly 'compartment models' or 'physiologically based toxicokinetic models' are used.

In both kinds of model, 'open compartments' are defined, which are characterized by their volumes and current concentrations of substance. The number of compartments in a model depends on the physicochemical and biochemical properties of the substance investigated and on the problems that are dealt with.

Compartment Models

In compartment models (Figures 2.6 and 2.13), the compartments are usually imaginary entities, which do not need a physiological basis.

In the low-concentration range, invasion and elimination processes generally obey first-order kinetics. A kinetic is first order when a rate is directly proportional to a concentration. It is therefore also described as linear kinetics. In such a process (absorption, distribution, metabolism, or excretion) an exponential function of the general form given in Equation (2.14)

$$c_{(t)} = \mathrm{A}_1 \cdot \mathrm{e}^{-k \cdot t} + \mathrm{A}_2 \tag{2.14}$$

describes the substance concentration in dependence of time. The concentration in the compartment at any time t is given by $c_{(t)}$. A_1 and A_2 are constants, their sum being the initial substance concentration $c_{(0)}$ at time $t = 0$; k is a rate constant with the dimension time^{-1}. If deviations from first-order kinetics are observed, like those always seen at

higher concentrations in saturable processes, the kinetics are said to be nonlinear (see heading *Metabolism*, above). To indicate the direction of a process in a particular tissue, its rate (v; dimension amount/time) is given a positive sign if the amount of substance in the compartment is increasing, and a negative sign if it is decreasing.

The One-compartment Model

The one-compartment model is the most straightforward toxicokinetic model. It represents the whole body as a single compartment. A one-compartment model (Figure 2.6) may be used when the distribution processes are much more rapid than the absorption and elimination processes.

In the section below, the reader will be familiarized with the basic principles in toxicokinetics exemplified by the one-compartment model.

Apparent volume of distribution The absorbed substance is often not evenly distributed in the organism but concentrated in certain depots (e.g., adipose tissue, protein binding) according to its physicochemical properties. The use of a factor (V_d) makes it possible to associate the concentration $c_{(t)}$ measured in a defined body fluid at a certain time point t with the amount of substance $N_{(t)}$ present at this time in the body; as given by Equation (2.15).

$$N_{(t)} = V_d \cdot c_{(t)} \tag{2.15}$$

V_d has the dimensions of a volume and is called the 'apparent volume of distribution'. Usually, it is a purely mathematical quantity. In a one-compartment model for which it is assumed that the distribution processes take place so rapidly that they have no effect on the toxicokinetics, the value of V_d is constant (which is not the case in models that have more than one compartment). V_d can be calculated following intravascular administration of the substance from the dose D_{iv} [the administered amount; see Figure 2.6(a)], and the measured concentration $c_{(0)}$ at time point $t = 0$ [Equation (2.16)]:

$$V_d = D_{iv}/c_{(0)} \tag{2.16}$$

Because most substances are not distributed homogeneously in the various tissues, the size of V_d depends on the medium in which c is measured. For example, if the substance-specific partition coefficient blood:blood plasma (P_{BP}), which can be determined in vitro, is not equal to 1, then the value of V_d obtained when c is measured in blood plasma (c_P) is different from that obtained when c is measured in blood (c_B). The apparent volume of distribution when the concentration of the substance is measured in plasma ($V_{dP)}$) is given by Equation (2.17):

$$V_{dP} = D_{iv}/c_{P(0)} \tag{2.17}$$

Considering that in a one-compartment model there is no distribution phase, that is, Equation (2.18) holds,

$$c_{B(t)} = c_{P(t)} \cdot P_{BP} \tag{2.18}$$

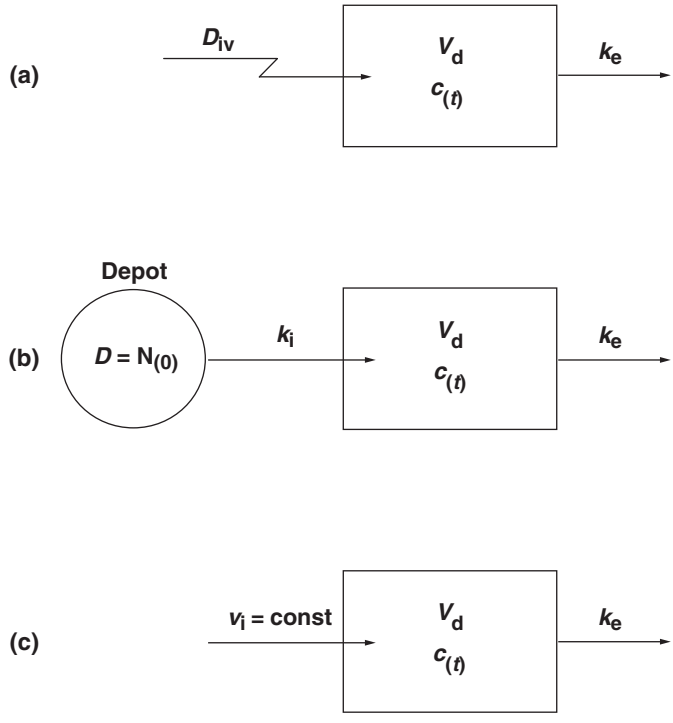


Figure 2.6 *One-compartment model for the description of the toxicokinetics of a substance administered by various routes and eliminated according to first-order kinetics. (a) Intravascular injection. (b) Absorption from a depot after extravascular administration. (c) Inhalation of a substance at constant concentration in the air or continuous intravascular infusion.* $c_{(t)}$*: Concentration at time t;* D_{iv}*: intravascular dose; D: dose;* k_e*: elimination rate constant;* k_i*: invasion rate constant (first-order kinetics);* V_d*: apparent volume of distribution;* v_i*: constant invasion rate.*

the apparent volume of distribution related to the concentration in the blood (V_{dB}) is given by Equation (2.19):

$$V_{dB} = D_{iv}/(c_{P(0)} \cdot P_{BP}) \tag{2.19}$$

or, considering Equation (2.17), by Equation (2.20):

$$V_{dB} = V_{dP}/P_{BP} \tag{2.20}$$

In some few cases, V_{dP} is identical with the volume of the blood plasma, which accounts for about 4% of the body weight of an adult human. This applies for example, for heparin or Evan's Blue, macromolecules which can penetrate neither the vessel walls nor the cell membranes of the erythrocytes and which are therefore distributed only in the intravascular plasma volume. For ethylene oxide, an epoxide, and the solvents ethanol, ethylene glycol, and acetone, V_{dP} is slightly larger than the volume of the total body water, which is about 60% of the body weight in adult men.

Many xenobiotics bind to plasma proteins, which reduces the concentration of the free substance in the plasma. This phenomenon can result in very large values for V_{dP} because

it is related to the concentration of the free substance. In general, the ratio of free to bound substance is constant over a wide concentration range. However, with increasing concentration the binding sites at the plasma proteins are increasingly occupied, and finally the fraction of the total substance which is free also increases. In such cases, V_{dP} becomes concentration-dependent, having a smaller value at high concentrations than at lower ones.

Intravascular administration; elimination; total clearance; half-life The easiest toxicokinetic problem is the elimination kinetics of an intravascularly, as a single dose D_{iv} administered substance which is distributed much more rapidly than it is eliminated, and for which the elimination obeys linear kinetics and so the amount eliminated per time unit (dN_e/dt) is directly proportional to the amount of the substance in the compartment, which is given by the product of its apparent volume V_d with the actual concentration $c_{(t)}$ [see Figure 2.6(a)]. Consequently, the rate of elimination is described by Equation (2.21):

$$dN_e/dt = -k_e \cdot V_d \cdot c_{(t)} \tag{2.21}$$

The proportionality factor k_e is the rate constant (dimension: time^{-1}) for the elimination process, and the product of k_e and V_d represents the total clearance of elimination from the organism (Cl_{tot}; compare heading Excretion, above); where Cl_{tot} is given by Equation (2.22).

$$Cl_{tot} = k_e \cdot V_d \tag{2.22}$$

The value of Cl_{tot} can be calculated not only from Equation (2.22); there are also other possibilities (see below).

Since dN_e/dt equals the product of the concentration change in the compartment with V_d, Equation (2.21) can be rewritten as Equation (2.21a).

$$dc_{(t)}/dt \cdot V_d = -k_e \cdot V_d \cdot c_{(t)} \tag{2.21a}$$

Canceling V_d in the differential Equation (2.21a) and solving it for $c_{(t)}$ we obtain Equation (2.23).

$$c_{(t)} = c_{(0)} \cdot e^{-k_e \cdot t} \tag{2.23}$$

The constant $c_{(0)}$ stands for the initial concentration when the whole dose D_{iv} is still present in the organism. Consequently, Equation (2.23) can also be expressed by considering Equation (2.16) as Equation (2.23a).

$$c_{(t)} = (D_{iv}/V_d) \cdot e^{-k_e \cdot t} \tag{2.23a}$$

Concentration–time curves corresponding to the function given in Equation (2.23) are shown in Figures 2.7(a) and 2.8. By plotting $c_{(t)}$ in the log scale versus t in the linear scale (half-logarithmic plot) this function yields a straight line with a slope of $-k_e/\ln 10$ [Figure 2.7(b)]. After taking the logarithm of Equation (2.23), k_e can be expressed as Equation (2.24).

$$k_e = \ln(c_{(0)}/c_{(t)})/t \tag{2.24}$$

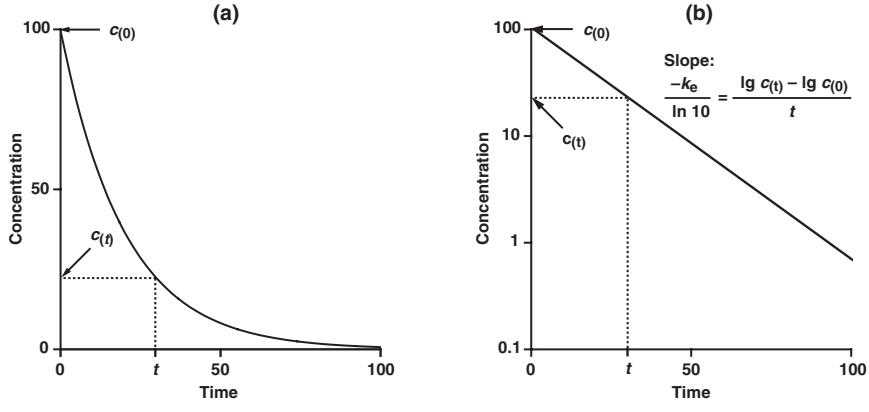


Figure 2.7 *Concentration–time course of a substance administered as a single dose and eliminated according to first-order kinetics as shown by a one-compartment model. (a) Linear plot. (b) Half-logarithmic plot.* $c_{(0)}$, $c_{(t)}$*: Concentration at time 0 and time t, respectively;* k_e*: elimination rate constant; ln 10: factor for the conversion of logarithm base 10 into natural logarithm.*

Substituting $c_{(0)}/2$ for $c_{(t)}$ in Equation (2.24) yields the elimination half-life ($t_{1/2}$) at which the initial concentration has been reduced to one half [Equation (2.25)]:

$$t_{1/2} = \ln 2/k_e \approx 0.693/k_e \tag{2.25}$$

This equation demonstrates that $t_{1/2}$ depends only on k_e and is independent of the concentration of the substance in the organism in the exposure range in which the elimination follows first-order kinetics. The value of $t_{1/2}$ gives the time span required to halve a given $c_{(t)}$.

Extravascular administration; Bateman function The most straightforward model to describe the absorption of a substance administered as a single oral, intraperitoneal, intramuscular, epicutaneous, or subcutaneous dose (D) is shown in Figure 2.6(b). It is assumed that the substance moves quantitatively in only one direction, from the application site into the compartment that represents the organism. In analogy to Equation (2.23), the quantity of substance $N_{i(t)}$ at the application site (intestinal tract, peritoneum, muscle, skin, or subcutaneous tissue) is given at any time by Equation (2.26).

$$N_{i(t)} = D \cdot e^{-k_i \cdot t} \tag{2.26}$$

Because the substance is taken into the compartment from the application site, the rate constant for the invasion process is designated as k_i.

The derivative of Equation (2.26) with respect to time yields the rate of disappearance of the substance at the application site [Equation (2.27)].

$$dN_{i(t)}/dt = -k_i \cdot D \cdot e^{-k_i \cdot t} \tag{2.27}$$

If the substance is taken up quantitatively (no first-pass effect!), the rate of absorption into the organism is the same as the rate of disappearance of substance at the application site. Thus, taking into account the volume of distribution V_d, the

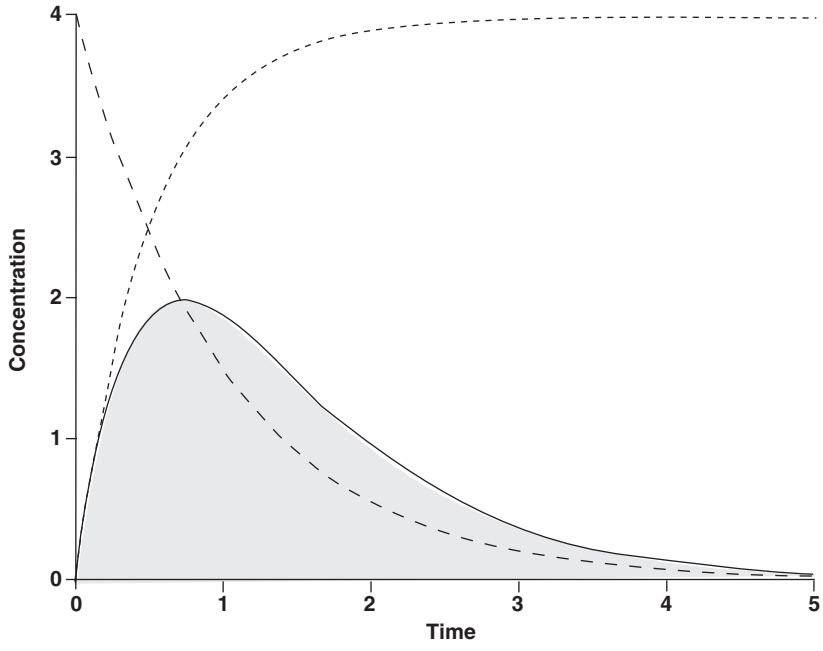


Figure 2.8 *Concentration–time courses of a substance during absorption and elimination obeying first-order kinetics as shown by a one-compartment model. – – –: Elimination after intravascular injection [Equation (2.23a)]; - - - - - : invasion from a depot without elimination [Equation (2.29)]; —: interaction of invasion and elimination (Equation (2.31), Bateman function); gray marked: area under the Bateman curve.*

increase in the concentration ($c_{(t)}$) of the absorbed substance in the compartment is given by Equation (2.28):

$$dc_{(t)}/dt = +(k_i \cdot D/V_d) \cdot e^{-k_i \cdot t} \tag{2.28}$$

Integrating Equation (2.28) over time and considering that $c_{(0)}$ at time point $t = 0$ is zero, Equation (2.29) is obtained:

$$c_{(t)} = D/V_d \cdot [1 - e^{-k_i \cdot t}] \tag{2.29}$$

This function describes a curve as shown as a dotted line in Figure 2.8. If there is no elimination from the organism ($k_e = 0$ in the model), the concentration of the absorbed substance increases rapidly at first and finally reaches a plateau when the whole dose has been taken up.

In general, substances are not only taken up by the organism but also eliminated ($k_e > 0$)! The elimination of a substance begins simultaneously with its absorption. In the one-compartment model [Figure 2.6(b)] the concentration change per unit time is written, when both processes absorption [Equation (2.28)] and elimination [Equation (2.21a)] are considered, by Equation (2.30):

$$dc_{(t)}/dt = +(k_i \cdot D/V_d) \cdot e^{-k_i \cdot t} - k_e \cdot c_{(t)} \tag{2.30}$$

Given that the initial concentration in the body $c_{(0)}$ is zero, solution of this differential equation yields a function describing the substance concentration in the one-compartment organism in dependence of time which is called the Bateman function (Figure 2.8); shown in Equation (2.31).

$$c_{(t)} = (D/V_d) \cdot [k_i/(k_e - k_i)] \cdot [e^{-k_i \cdot t} - e^{-k_e \cdot t}] \tag{2.31}$$

The half-life of the final concentration–time course depends on that expression $e^{-k \cdot t}$ in the Bateman function, which has the smaller value of k. The values of the expression with the smaller k approaches zero more slowly than the values of that with the larger k. In the rare cases where the terminal elimination phase is determined by k_i and not by k_e, one speaks of 'flip-flop' kinetics. In such cases the absorption is slower than the elimination. It must be stressed again that Equations (2.26)–(2.29) are true only if there is complete absorption of the extravascularly administered substance.

Area under the concentration–time curve; bioavailability The probability of chronic damage and the extent of such damage are often correlated with the area under the concentration–time curve (AUC) of the active substance in blood or plasma.

The AUC (dimension: concentration · time) for the period between $t = 0$ and $t = \infty$ is given by Equation (2.32):

$$AUC_0^\infty = \int_0^\infty c_{(t)} \cdot dt \tag{2.32}$$

For concentration ranges in which linear kinetics apply, $c_{(t)}$ is generally given by a function of the form shown in Equation (2.33):

$$c_{(t)} = A_1 \cdot e^{-k_1 \cdot t} + A_2 \cdot e^{-k_2 \cdot t} + \ldots + A_n \cdot e^{-k_n \cdot t} \tag{2.33}$$

A_1 to A_n (dimension: amount/volume) are concentration constants and k_1 to k_n (dimension: time^{-1}) are rate constants [cf. Equations (2.14) and (2.31)].

For such a function, AUC_0^∞ is given by Equation (2.34):

$$AUC_0^\infty = A_1/k_1 + A_2/k_2 + \ldots + A_n/k_n \tag{2.34}$$

According to Equations (2.33) and (2.34), AUC_0^∞ obtained following intravascular administration [Equation (2.23a)] is given by Equation (2.35):

$$AUC_0^\infty = D_{iv}/(k_e \cdot V_d) \tag{2.35}$$

The AUC_0^∞ of the Bateman function (grey area marked in Figure 2.8) is given by Equation (2.36):

$$AUC_0^\infty = (D/V_d) \cdot [k_i/(k_e - k_i)] \cdot [1/k_i - 1/k_e] \tag{2.36}$$

which can be cancelled to Equation (2.35a).

$$AUC_0^\infty = D/(k_e \cdot V_d) \tag{2.35a}$$

Obviously, the AUC_0^∞ of the Bateman function is identical with the AUC_0^∞ following intravascular administration when the same doses are given. This is because it was

shown that the extravascularly administered dose D is completely taken up (see also below).

As derived from Equations (2.35), (2.35a), and (2.22) it can be seen from Equation (2.37) that Cl_{tot} can also be obtained from the administered dose and AUC_0^∞:

$$Cl_{tot} = D_{iv}/AUC_0^\infty \tag{2.37}$$

and, if D is completely taken up, Equation (2.37a) holds:

$$Cl_{tot} = D/AUC_0^\infty \tag{2.37a}$$

Because the value of V_d depends on the medium in which the substance concentration is measured (see above), this holds true also for Cl_{tot}. A simple way of determining Cl_{tot} in blood (Cl_{totB}) or blood plasma (Cl_{totP}) is based on the experimental determination of $AUC_{0\,iv}^\infty$ in blood ($AUC_{0\,ivB}^\infty$) or blood plasma ($AUC_{0\,ivP}^\infty$) after intravenous injection of a dose D_{iv}. Intravenous administration has to be chosen because only by this method is complete bioavailability (see heading ***Absorption***, above) ensured.

If the partition coefficient P_{BP} is known, Cl_{totP} and Cl_{totB} can be interconverted; as shown in Equation (2.38):

$$Cl_{totB} = Cl_{totP}/P_{BP} \tag{2.38}$$

When determining Cl_{tot} via an experimentally determined $AUC_{0\,iv}^\infty$ no use of a toxicokinetic model is required. It is emphasized again that Cl_{tot} is a concentration-independent constant only if the elimination kinetics is linear (see heading *Excretion*, above).

After extravascular administration of a substance, often only a fraction (F) of the administered dose reaches the systemic blood circulation (e.g., because of a first-pass effect). The bioavailability expresses this fraction (see heading ***Absorption***, above). For example, under conditions of first-order kinetics the bioavailability F of an orally administered dose (D_{po}) can be calculated from the corresponding dose-normalized $AUC_{0\,po}^\infty$ and from the dose-normalized $AUC_{0\,iv}^\infty$ determined after intravenous injection of a dose (D_{iv}) in a second experiment, as given in Equation (2.39):

$$F = (AUC_{0\,po}^\infty/D_{po})/(AUC_{0\,iv}^\infty/D_{iv}) \tag{2.39}$$

F is a dimensionless number. When $F = 1$ a substance is completely bioavailable and there is no first-pass effect (applicable for first-order kinetics).

Continuous administration The one-compartment model for continuous administration is shown in Figure 2.6(c). During continuous administration of a substance, e.g. during intravascular infusion or exposure to a constant concentration of a gas in the inhaled air, the absorption of the substance into the compartment representing the organism can be modeled as a process obeying zero-order kinetics: the substance enters the compartment with a constant absorption rate v_i (dimension: amount / time). When the elimination from the organism obeys first-order kinetics [Equation (2.21)] the change in the concentration in the organism in dependence of time is given by Equation (2.40).

$$dc_{(t)}/dt = +v_i/V_d - k_e \cdot c_{(t)} \tag{2.40}$$

Solving this differential equation for $c_{(t)}$ considering that $c_{(0)} = 0$, yields Equation (2.41).

$$c_{(t)} = [v_i/(k_e \cdot V_d)] \cdot [1 - e^{-k_e \cdot t}] \tag{2.41}$$

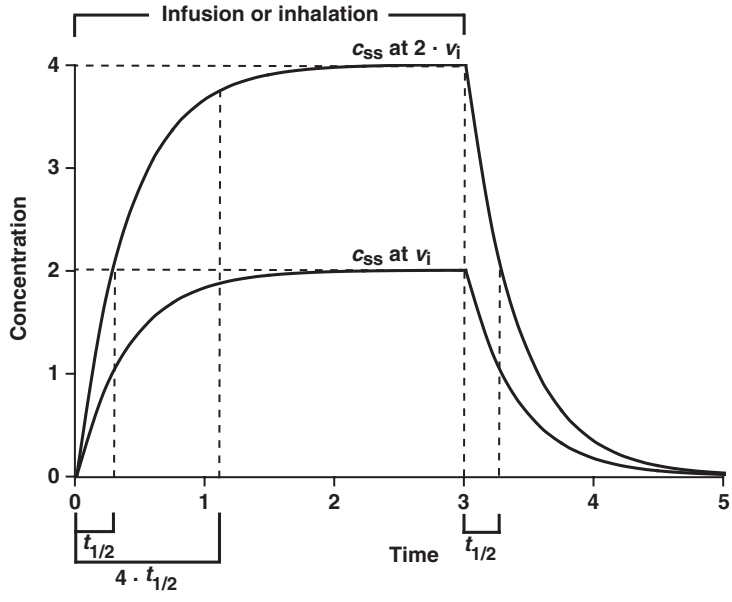


Figure 2.9 *Concentration–time courses of a substance during continuous absorption (e.g., by intravenous infusion or inhalation) and thereafter as shown by a one-compartment model (linear plot). Elimination follows first-order kinetics. c_{ss}: Concentration at equilibrium; $t_{1/2}$: half-life; v_i: invasion rate (constant).*

A plot of the concentration–time curve as given by this function is shown in Figure 2.9. With increasing time $c_{(t)}$ approaches a plateau concentration (c_{ss}) at which steady state is attained; that is, the elimination rate [$k_e \cdot V_d \cdot c_{ss}$; see Equation (2.21)] is exactly equal to the constant rate of absorption (v_i). Thus the change in concentration of the substance in the organism at the steady-state concentration (c_{ss}) is equal to zero [Equation (2.42)]:

$$dc_{ss}/dt = 0 \tag{2.42}$$

Inserting Equation (2.42) into Equation (2.40) and solving for c_{ss} yields Equation (2.43):

$$c_{ss} = v_i/(k_e \cdot V_d) \tag{2.43}$$

Equations (2.22) and (2.43) show that the plateau concentration c_{ss} (the maximum accumulation of a substance) is determined only by the ratio of v_i to Cl_{tot}. As may be seen from Figure 2.9, increasing v_i by a factor of 2 doubles the value of c_{ss}. The time until c_{ss} is attained does not depend on v_i; it is determined only by k_e. After four elimination half-lives, about 94% of c_{ss} has been attained. In the elimination phase after the end of administration, the course of the reduction in concentration is also given by an exponential function with $t_{1/2} = \ln 2/k_e$ (Figure 2.9).

Repeated administration The simplest case of repeated administration via a depot (for example, oral or intraperitoneal doses; bioavailability $F = 1$; linear kinetics) is shown schematically in Figure 2.10. Regular administration of the dose D with a constant time interval τ between doses is equivalent to a constant dose rate D/τ. In analogy to

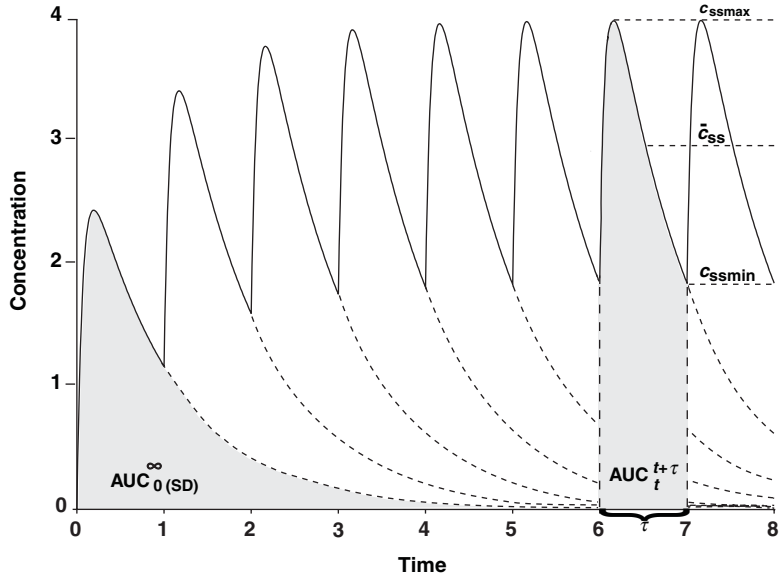


Figure 2.10 *Concentration–time courses of a substance administered repeatedly at constant intervals (e.g., oral or intraperitoneal administration) with complete absorption and with elimination, both obeying first-order kinetics, as shown by a one-compartment model (linear plot). —: Concentration–time course resulting from multiple dosing; – — –: theoretical concentration–time courses for each dose; gray marked: areas under the curves following the first dose and a dose given when steady-state conditions are reached;* $AUC^{\infty}_{0\,(SD)}$*: Area under the concentration–time curve from time 0 to infinity for a single dose;* $AUC^{t+\tau}_{t}$*: area under the concentration–time curve for a dose interval at the steady state;* $\bar{c}_{ss}$*: mean concentration at equilibrium;* c_{ssmax}*: maximum concentration during a dose interval at equilibrium;* c_{ssmin}*: minimum concentration during a dose interval at equilibrium;* τ*: dose interval.*

Equation (2.43), the mean concentration $\bar{c}_{ss}$ which is attained at steady state is given by Equation (2.44):

$$\bar{c}_{ss} = (D/\tau)/(V_d \cdot k_e) \tag{2.44}$$

Thus, just as in the case of continuous administration, the concentration of the substance accumulates to a maximum value. Whereas during constant continuous administration of a substance a constant c_{ss} is attained, during repeated administration this concentration varies between a maximum (c_{ssmax}) and a minimum value (c_{ssmin}). The mean concentration $\bar{c}_{ss}$ is not identical with the arithmetic mean of these two values because the concentration–time course obeys first-order kinetics. The value of $\bar{c}_{ss}$ can be calculated from the area under the concentration–time curve from $t = 0$ to infinity obtained after administration of a single dose ($AUC^{\infty}_{0\,SD}$). When steady state is reached under repeated dosing conditions, the area $AUC^{t+\tau}_{t}$ under a dose interval τ is identical with $AUC^{\infty}_{0\,SD}$ (see Figure 2.10). Therefore $\bar{c}_{ss}$ is given by Equation (2.45):

$$\bar{c}_{ss} = AUC^{t+\tau}_{t}/\tau = AUC^{\infty}_{0\,SD}/\tau \tag{2.45}$$

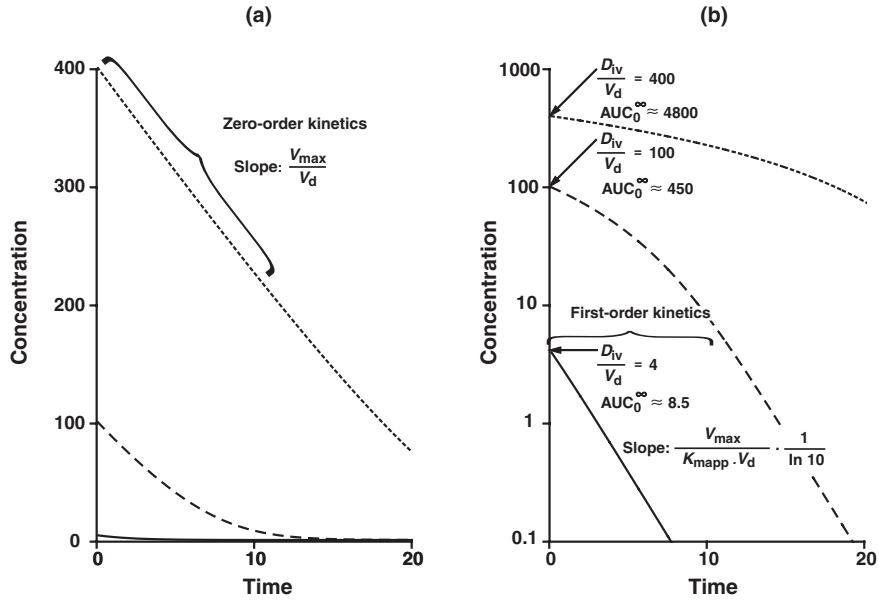


Figure 2.11 *Concentration–time courses of a substance administered as single variously sized doses and eliminated according to saturation kinetics as shown by a one-compartment model. (a) Linear plot. (b) Half-logarithmic plot. AUC_0^∞: Area under the concentration–time curve from time 0 to infinity; D_{iv}: intravenously administered dose; K_{mapp}: apparent Michaelis constant; V_d: apparent volume of distribution; v_{max}: maximum elimination rate; ln 10: factor for the conversion of logarithm base 10 into natural logarithm.*

Saturation kinetics For a first-order process, the AUC_0^∞ is proportional to the administered dose [cf. Equations (2.37) and (2.37a)] provided that the administration route is not changed or the bioavailability F is equal to 1. In toxicological studies, however, it is often observed that as the dose is increased the AUC_0^∞ increases more than in proportion with the administered dose. AUC_0^∞ is then no longer given by Equation (2.34) and derived equations. Often concentration–time curves are observed like the two upper ones shown in Figure 2.11 for administration by intravenous injection. Such curves are characteristic for a process that obeys saturation kinetics of the kind shown for enzyme-catalysed metabolism (see heading *Metabolism*, above). In such cases, the rate of elimination is usually described according to a Michaelis–Menten equation. In the one-compartment model saturation kinetics of a metabolic elimination can then be allowed for by replacing the concentration-independent Cl_{tot} [given as product of k_e and V_d in Equation (2.21a)] by the concentration-dependent clearance presented in Equation (2.12). According to Equation (2.21a), the obtained rate of elimination is then expressed by Equation (2.46):

$$dc_{(t)}/dt \cdot V_d = -[V_{max}/(K_{mapp} + c_{(t)})] \cdot c_{(t)} \qquad (2.46)$$

To determine $c_{(t)}$ as a function of t from this differential equation, iterative calculations must be used.

Frequently it is not the absorbed substance itself which is of toxicological relevance but a metabolite. The toxicokinetics of such a metabolite can also be described by the

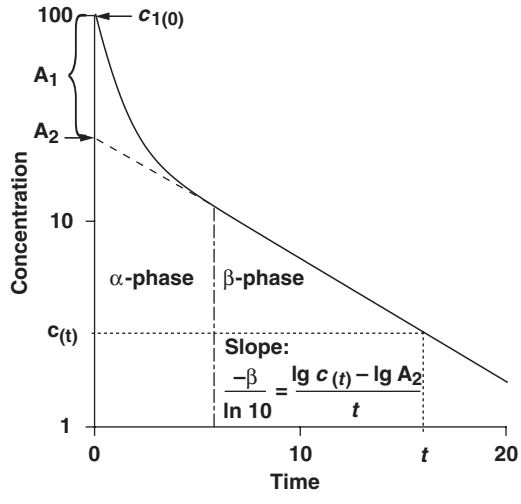


Figure 2.12 *Biphasic concentration–time curve as shown by a two-compartment model for first-order kinetics: concentration–time curve in the central compartment after administration of a single dose of a substance into the central compartment (half-logarithmic plot). —: Concentration in the central compartment; – – –: backwards extrapolation of the β-phase curve to time 0; – - –: end of α-phase and begin of β-phase. $c_{1(0)}$: Initial concentration ($A_1 + A_2$); β: rate constant of the β-phase; ln 10: factor for the conversion of logarithm base 10 into natural logarithm.*

processes given above. Its rate of formation (corresponding to the rate of absorption) can be determined from the rate of metabolic elimination of its precursor.

The Two-compartment Model

Linear kinetics, elimination from the central compartment When, after intravascular injection, the concentration–time course of the substance is plotted on a half-logarithmic scale, it is often possible to recognize a first phase (α-phase in Figure 2.12) in which the concentration in blood or plasma decreases rapidly, mainly as a result of distribution of the substance between organs and tissues. This distribution phase is followed by an elimination phase (β-phase in Figure 2.12) which is characterized by a slower decrease in concentration which obeys an exponential function. The whole time-course of the concentration changes is described by Equation (2.47):

$$c_{(t)} = A_1 \cdot e^{-\alpha \cdot t} + A_2 \cdot e^{-\beta \cdot t} \tag{2.47}$$

where α and β are two rate constants.

Such a concentration–time course cannot be reflected by a one-compartment model, which is based on the assumption that distribution is instantaneous. Therefore the model must be extended by including a further 'peripheral' or 'deep' compartment which takes into account the distribution phase.

Figure 2.13 shows a frequently used open two-compartment model with a 'central' compartment and a peripheral compartment. First-order kinetics describes the elimination from the central compartment and the distribution process between both compartments. In

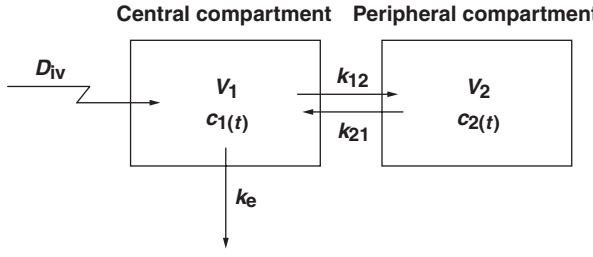


Figure 2.13 *Two-compartment model for the description of the toxicokinetics of a substance administered as a single intravascular injection, distributed from the central into the peripheral compartment and eliminated from the central compartment, in each case according to first-order kinetics. $c_{1(t)}$: Concentration in the central compartment at time t; $c_{2(t)}$: concentration in the peripheral compartment at time t; D_{iv}: intravenous dose; k_{12}, k_{21}: rate constants of the processes of distribution between the central and peripheral compartments; k_e: elimination rate constant; V_1, V_2: volumes of the central and peripheral compartments, respectively.*

this model, the changes in amount per time unit $dN_{1(t)}/dt$ in the central and $dN_{2(t)}/dt$ in the peripheral compartment are described by the differential Equations (2.48) and (2.49):

$$dN_{1(t)}/dt = -(k_{12} + k_e) \cdot V_1 \cdot c_{1(t)} + k_{21} \cdot V_2 \cdot c_{2(t)} \quad (2.48)$$

$$dN_{2(t)}/dt = -k_{21} \cdot V_2 \cdot c_{2(t)} + k_{12} \cdot V_1 \cdot c_{1(t)} \quad (2.49)$$

where $c_{1(t)}$ is the concentration of the substance in the central compartment with volume V_1, and $c_{2(t)}$ is that in the peripheral compartment with volume V_2. The rate constants k_{12} and k_{21} describe the processes of distribution between the two compartments, and k_e is the elimination rate constant for the central compartment. After a dose D_{iv} of the substance has been injected into the central compartment, which generally represents the blood and the well perfused organs (lung, brain, heart, liver, spleen, kidneys), the substance is immediately evenly distributed in this compartment. From here it enters the peripheral compartment, which generally represents the less well perfused tissues (muscle, skin, adipose tissue). There the accumulation of the substance is determined by the rates of distribution between the two compartments ($V_1 \cdot k_{12} \cdot c_{1(t)}$ and $V_2 \cdot k_{21} \cdot c_{2(t)}$). These rates are not constant because they are determined by the concentrations $c_{1(t)}$ and $c_{2(t)}$ which change with time.

In the β-phase the ratio of the concentrations $c_{2(t)}$ to $c_{1(t)}$ is constant and the elimination from both compartments has the same half-life, $t_{1/2} = \ln 2/\beta$.

The area under the concentration–time curve AUC_0^∞ may then be calculated from Equations (2.33), (2.34) and (2.47) as shown in Equation (2.50):

$$AUC_0^\infty = A_1/\alpha + A_2/\beta \quad (2.50)$$

The rate constants α and β (dimensions: time^{-1}) are complex functions of the rate constants k_{12}, k_{21}, and k_e (see Figure 2.13); and are given by Equations (2.51) and (2.52):

$$\alpha = 0.5 \cdot \left\{(k_{12} + k_{21} + k_e) + [(k_{12} + k_{21} + k_e)^2 - 4 \cdot k_e \cdot k_{21}]^{0.5}\right\} \quad (2.51)$$

$$\beta = 0.5 \cdot \left\{(k_{12} + k_{21} + k_e) - [(k_{12} + k_{21} + k_e)^2 - 4 \cdot k_e \cdot k_{21}]^{0.5}\right\} \quad (2.52)$$

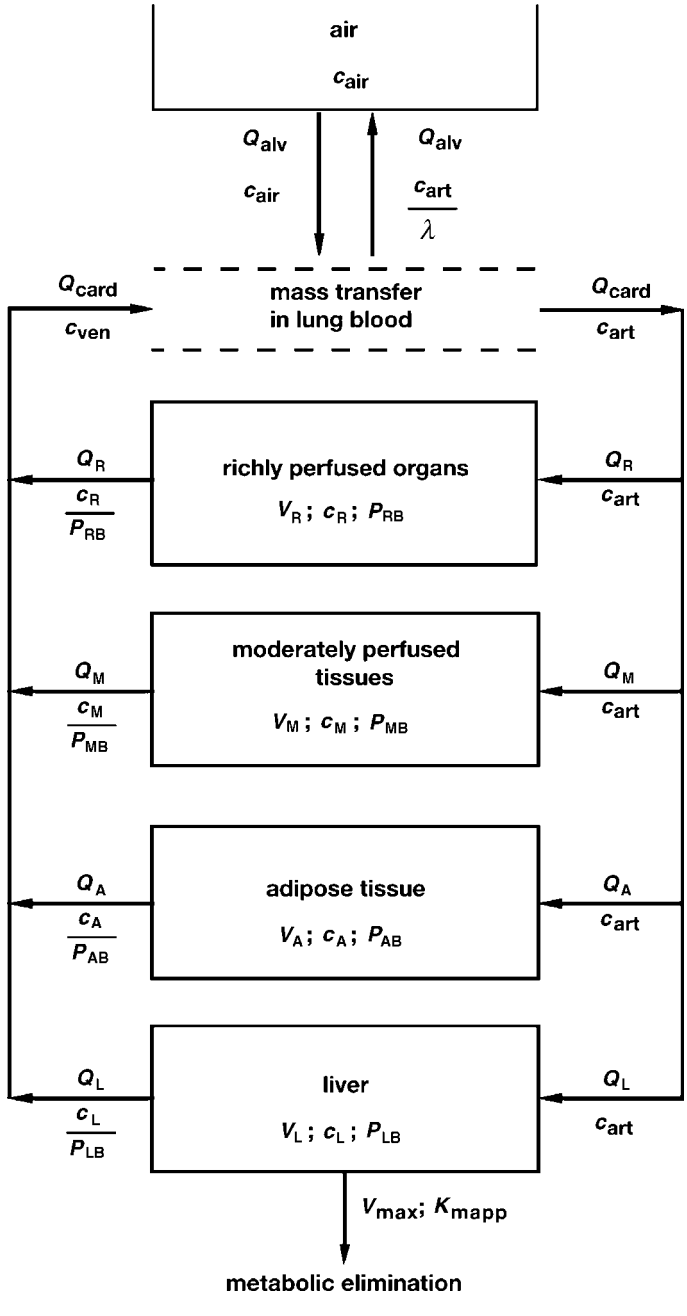

air
c_{air}
Q_{alv}
c_{air}
Q_{alv}
$\frac{c_{art}}{\lambda}$
Q_{card}
c_{ven}
mass transfer in lung blood
Q_{card}
c_{art}
Q_R
$\frac{c_R}{P_{RB}}$
richly perfused organs
V_R; c_R; P_{RB}
Q_R
c_{art}
Q_M
$\frac{c_M}{P_{MB}}$
moderately perfused tissues
V_M; c_M; P_{MB}
Q_M
c_{art}
Q_A
$\frac{c_A}{P_{AB}}$
adipose tissue
V_A; c_A; P_{AB}
Q_A
c_{art}
Q_L
$\frac{c_L}{P_{LB}}$
liver
V_L; c_L; P_{LB}
Q_L
c_{art}
V_{max}; K_{mapp}
metabolic elimination

Figure 2.14

The apparent volume of distribution V_d related to the central compartment is not constant in the two-compartment model, unlike in the one-compartment model, but increases in the time-course of the α-phase until the β-phase begins. In the β-phase, the apparent volume of distribution ($V_{d\beta}$) may be determined from the intravenous dose (D_{iv}), the resulting area under the concentration–time curve in the central compartment $AUC_{0\,iv}^{\infty}$, and the rate constant β, as shown in Equation (2.53):

$$V_{d\beta} = D_{iv}/(AUC_{0\,iv}^{\infty} \cdot \beta). \tag{2.53}$$

The total clearance (Cl_{tot}; related to the concentration in the central compartment) equals the expression $D_{iv}/AUC_{0\,iv}^{\infty}$ [see Equation (2.37)]. It is independent of the

$$\text{(a)}\quad c_{ven} = \frac{Q_{R}\cdot\frac{c_R}{P_{RB}} + Q_M\cdot\frac{c_M}{P_{MB}} + Q_A\cdot\frac{c_A}{P_{AB}} + Q_L\cdot\frac{c_L}{P_{LB}}}{Q_{card}}$$

$$\text{(b)}\quad c_{art} = \frac{c_{ven}\cdot Q_{card} + c_{air}\cdot Q_{alv}}{Q_{card} + \frac{Q_{alv}}{\lambda}}$$

$$\text{(c)}\quad \frac{dc_R}{dt}\cdot V_R = +Q_R\cdot c_{art} - Q_R\cdot\frac{c_R}{P_{RB}}$$

$$\text{(d)}\quad \frac{dc_M}{dt}\cdot V_M = +Q_M\cdot c_{art} - Q_M\cdot\frac{c_M}{P_{MB}}$$

$$\text{(e)}\quad \frac{dc_M}{dt}\cdot V_A = +Q_A\cdot c_{art} - Q_A\cdot\frac{c_A}{P_{AB}}$$

$$\text{(f)}\quad \frac{dc_L}{dt}\cdot V_L = +Q_L\cdot c_{art} - Q_L\cdot\frac{c_L}{P_{LB}} - \frac{V_{max}\cdot c_L}{K_{mapp} + c_L}$$

Figure 2.14 *Straightforward physiologically based toxicokinetic model for lipophilic gaseous substances, enabling the description of absorption by inhalation, distribution by the blood flow in tissues and organs, metabolic elimination obeying saturation kinetics in the liver, and exhalation.*

Equations: (a) concentration of substance in oxygen-poor 'venous' blood, (b) concentration of substance in oxygen-rich 'arterial' blood [obtained by reformulating Equation (2.5)]. Mass-balance equations giving the mass changes in: (c) the richly perfused organs, (d) the moderately perfused tissues, (e) adipose tissue, and (f) the liver.

Symbols:

c_{air}: Concentration of substance in the air at time t; c_{art}, c_{ven}: concentrations of substance in the oxygen-rich 'arterial' blood leaving the lung and in the oxygen-poor 'venous' blood entering the lung at time t, respectively; c_R, c_M, c_A, c_L: concentrations of substance in richly perfused organs, moderately perfused tissues, adipose tissue, and liver at time t, respectively; K_{mapp}: apparent Michaelis constant for the concentration of substance in the liver; λ: substance-specific partition coefficient blood:air; $P_{RB}, P_{MB}, P_{AB}, P_{LB}$: substance-specific partition coefficients richly perfused organs:blood, moderately perfused tissues:blood, adipose tissue:blood, and liver:blood, respectively; Q_{alv}: alveolar ventilation; Q_{card}: cardiac output blood (equals the blood flow through the lung); Q_R, Q_M, Q_A, Q_L: blood flows through the richly perfused organs, the moderately perfused tissues, the adipose tissue, and the liver, respectively; V_R, V_M, V_A, V_L: volumes of richly perfused organs, moderately perfused tissues, adipose tissue, and liver, respectively; V_{max}: maximum rate of metabolism.

toxicokinetic model. If the value of V_1 is known, the elimination rate constant k_e can be calculated in various ways; for example, as shown in Equation (2.54):

$$k_e = V_{d\beta} \cdot \beta / V_1 = C|_{tot} / V_1 = D_{iv} / (AUC_{0\,iv}^{\infty} \cdot V_1) \tag{2.54}$$

Appropriate equations can be derived for the calculation of the constants and functions $A_1, A_2, \alpha, k_{12}, k_{21}$ and the function for the time-dependence of V_d.

Physiologically Based Toxicokinetic Models

In physiologically based toxicokinetic models, compartments represent organs, tissues, or groups of tissues with their actual volumes. All rates are described by means of physiological, physiochemical, and biochemical parameters. For instance, the transport of substances between the tissues takes place via the bloodstream, and their accumulation is dependent on tissue:blood partition coefficients and on tissue-specific metabolic parameters.

A straightforward model for exposure of an animal or a human to lipophilic gaseous substances is exemplified in Figure 2.14. Gas exchange between atmosphere and lung blood is modeled as described in detail earlier (see heading *Lung* in section *Excretion*, above). The organism is subdivided into several compartments representing organs and tissues. Adipose tissue (the storage tissue) and liver are represented by their own compartments. Metabolism follows saturation kinetics according to Michaelis and Menten and takes place exclusively in the liver. The compartment 'richly perfused organs' summarizes primarily lung, brain, kidney, spleen, heart, and intestines, the compartment 'moderately perfused tissues' represents mainly muscle and skin. The scarcely perfused bones and cartilage are disregarded. The model is a so-called perfusion-limited model, i.e. it is assumed that the substance in each tissue is always in equilibrium with the blood leaving the tissue. A series of differential equations describes the mass changes in each compartment (Figure 2.14); computer programs are available for the solution of these equations.

Physiologically based toxicokinetic models have the advantage over the classical compartment models that they permit knowledge of the fate of a substance in individual tissues like, for instance, in a tissue which is the target of an adverse effect. They are useful for the extrapolation of tissue burdens between species ('species scaling') when sufficient anatomical, physiological, and biochemical information is available. For these reasons, the use of such models is continuously growing. However, there are also certain disadvantages: physiologically based toxicokinetic models require a large number of experimental in vivo and in vitro data which are essential for model calibration and validation.

2.1.4 Summary

Toxicokinetics is fundamental to the understanding of the quantitative relationships between the amount of a substance administered or taken up and its toxic effects. It describes by means of mathematical functions the processes of the absorption of a substance into an organism, its distribution and accumulation in various organs and tissues, and of its elimination from the organism. Usually, substances are absorbed via the lung, the skin, or the gastrointestinal tract. In the latter case, a first-pass effect may occur due to metabolism of the substance in the liver before it enters the systemic blood circulation. The most common absorption mechanisms include diffusion through biological membranes and active transport catalysed by enzymes located in cell membranes. The absorption phase can

be circumvented experimentally by direct injection of the substance into the bloodstream. Elimination processes are also mostly mediated by enzymes and so are subject to saturation kinetics, which becomes especially relevant at high substrate doses or concentrations as used in animal studies. Many toxicants only become harmful when they are metabolically activated. Some absorbed substances are excreted unchanged. Excretion of xenobiotics takes place mainly via the kidneys, the intestinal tract, and the lungs, but the skin and mammary glands may also be involved. For the mathematical analysis of experimentally determined changes in substance concentration with time, compartment models and physiologically based toxicokinetic models are used. The most frequent toxicokinetic expressions are defined by means of a one-compartment model which is presented in detail. Additionally, an introduction is given into the basic principles of a two-compartment model and a physiologically based toxicokinetic model.

References

Csanády, GA, Mendrala, AL, Nolan, RJ, Filser, JG, A Physiologic Pharmacokinetic Model for Styrene and Styrene-7,8-oxide in Mouse, Rat and Man. *Arch. Toxicol.* (1994), **68**, 143–157.

Mevlenberg, CJW, Vijverberg, HPM, Empirical Relations Predicting Human and Rat Tissue: Air Partition Coefficients of Volatile Organic Compounds. *Toxicol. Appl. Pharmacol.* (2000), **165**, 206–216.

For Further Reading

Gibaldi, M, Perrier, D, Pharmacokinetics, 2nd edition, revised and expanded, Swarbrick, J (ed.), Marcel Dekker, New York (1982).

Ritschel, WA, Kearns, GL, Handbook of Basic Pharmacokinetics ... Including Clinical Applications, 6th edition, APhA (American Pharmaceutical Association) Publications (2004).

Reddy, M, Yang, RS, Andersen, ME, Clewell, HJ, Physiologically Based Pharmacokinetic Modeling: Science and Applications, John Wiley & Sons, Ltd, Weinheim, and Hoboken, New Jersey (2005).

2.2A Phase I Metabolism

Jeroen T.M. Buters

2.2A.1 Introduction

Living systems are composed of endogenous chemicals, large and small, which account for both the morphology and physiology of the organism. In daily life we are exposed to xenobiotic chemicals such as therapeutic agents, food contaminants, or chemicals at the workplace or by environmental exposure. All these chemicals may affect body functions either per se or by reactive metabolites or they are detoxified and eliminated. The complex mechanisms involved in metabolic activation or inactivation of chemicals are discussed to

describe their role in the physiological disposition of chemicals which includes absorption, distribution, metabolism, and excretion (ADME).

Absorption, Distribution, Metabolism, and Excretion (ADME) are physiological processes by which chemicals enter, are disseminated, metabolically altered, and eliminated. The body eliminates water-soluble compounds that display polarity through the kidneys, bile, sweat, and intestinal secretions. The excretion of lipophilic compounds is hampered because:

1. **They can be sequestered in body fat. At equilibrium the ratio of lipophilic chemicals in blood versus fat is extremely low.**
2. **Although they are excreted into the gut or into kidney tubules they are rapidly reabsorbed and reenter the circulation.**

However, metabolism to more polar compounds enhances their elimination from the body.

Maintenance of homeostasis requires a balance between the entry and excretion of xenobiotic chemicals from the body. Without excretion the organism will accumulate a compound and with each repeated dose the concentration increases. The chemical would remain in the body indefinitely and after repeated doses, toxic levels could be achieved. Consequently, there is a need for mechanisms which promote elimination of these compounds.

Once absorbed into the circulation chemicals are transported by the blood to all body compartments. Chemicals can leave the blood by passing through the membranes of small blood vessels, enter extra-cellular fluid, and can then enter cells in various organs. Entry into cells can be through passive transport, which is largely dependent on the polarity of the chemical, or active transport, which requires metabolic energy and is usually structurally specific. Thus, to a large extent the entry of chemicals into any individual organ or cell is a function of the chemical and physical properties of the compound.

The metabolic processes involved in absorption, distribution, metabolism, and excretion (ADME) are described as Phase 0, Phase I, Phase II, and Phase III metabolism.

Phase 0	**describes the passage of a chemical through the cellular wall into the cytosol.**
Phase I	**metabolism introduces reactive groupings into chemicals.**
Phase II	**enzymes transfer polar groups to products of the Phase I reactions.**
Phase III	**describes transports of water-soluble metabolites of Phase I and Phase II metabolism out of the cells.**

Table 2.2 *Phase I and phase II are metabolic processes. Hydrophilic compounds do not need metabolism to be excreted*

		Hydrophilic compound	Lipophilic compound
Absorption		−[a]	+
Distribution		+	+
Metabolism	Phase I	−	+
	Phase II	−	+
Excretion		+	−

[a]Often require transporters for their absorption.

Chemicals that are sufficiently polar may go through the ADME process without metabolism. If not, they have to be rendered more water-soluble before excretion can take place. The two-step process usually involves an initial metabolic step, which most frequently involves oxidation or reduction (Phase I metabolism) of the compound to which a highly polar group may be attached (Phase II metabolism). The different steps within ADME to metabolize hydrophilic or lipophilic compounds are illustrated in Table 2.2 and Figure 2.15. Table 2.3 contains a list of enzymes responsible for the various phases of xenobiotic metabolism.

The metabolic pathway followed by a chemical is determined by its structure and polarity. Figure 2.16 shows that all pathways are available, but the number of steps taken by a given chemical is determined by its structure and polarity.

Phase I reactions introduce groups like -OH into molecules to make the chemicals more polar and more reactive to Phase II metabolism. Conjugation with more polar

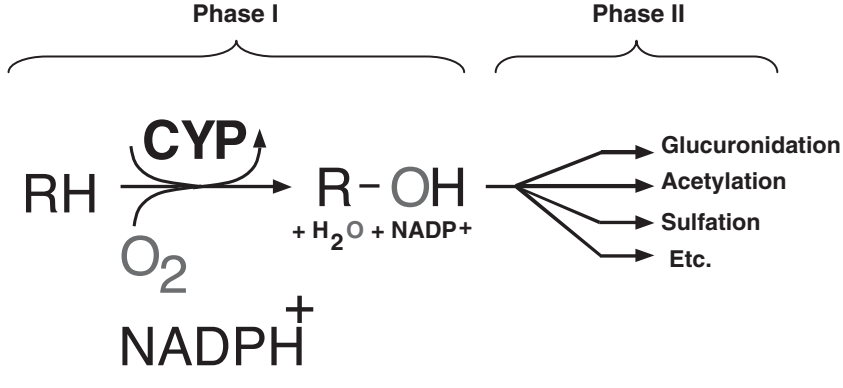


Figure 2.15 *General reaction scheme for Phase I metabolism by cytochrome P450 and subsequent Phase II metabolism.*

Table 2.3 *Phase I, Phase II, and Phase III enzymes*

Phase I Enzymes
Oxidoreductases
Cytochromes P450 (CYP)
Flavin-containing monooxygenases (FMO)
Monoamine oxidases (MAO)
Cyclooxygenases (COX)
Dihydrodioldehydrogenases
DT-Diaphorases (NQOR)
Alcohol dehydrogenases (ADH)
Aldehyde dehydrogenases (ALDH)
NADPH P450 oxidoreductase (OR)
Hydrolases
Esterases
Amidases
Glucuronidases
Epoxide hydrolases (EH)
Phase II Enzymes
Transferases
Glutathionetransferases (GST)
UDP-Glucuronosyltransferases (UGT)
Sulfotransferases (SULT)
Acetyltransferases (NAT)
Methyltransferases
Aminoacyltransferases
Phase III Enzymes
Transporters
ATP Binding Cassette transporter family B1 (ABCB1)[a]
ATP Binding Cassette transporter family C1 (ABCC1)[b]
ATP Binding Cassette transporter family G2 (ABCG2)[c]

[a]Multidrug resistance proteins (MDR).
[b]Multidrug resistance-associated proteins (MRP).
[c]Breast Cancer Resistance Protein (BCRP).

moieties such as sulfate, glucuronic acid, or glutathione usually results in sufficiently water-soluble complexes for excretion to be possible.

The major enzymes that catalyse Phase I metabolism are the cytochromes P450 (CYP enzymes). This large class of enzymes is usually located in the endoplasmic reticulum and in some cases in mitochondria of cells throughout the body. The highest concentrations of these enzymes are found in hepatocytes, but lower concentrations are found throughout the body. The activity of these enzymes is dependent upon the availability of molecular oxygen and NADPH. In humans there are 57 CYP enzymes but the number varies by species. In some plants, where CYP enzymes play key synthetic roles, there are more than 100 CYP enzymes. As a general rule substrates for CYP enzymes are relatively nonpolar.

Phase I enzymes also include other oxygenases such as flavin-containing monooxygenases (FMO), which act primarily on amines and thiols; monoamine oxidases (MAO); dehydrogenases; and cyclooxygenases (see Table 2.3). Other Phase I enzymes are hydrolases, which add water across epoxide bonds (epoxide hydrolases) or ester bonds (esterases).

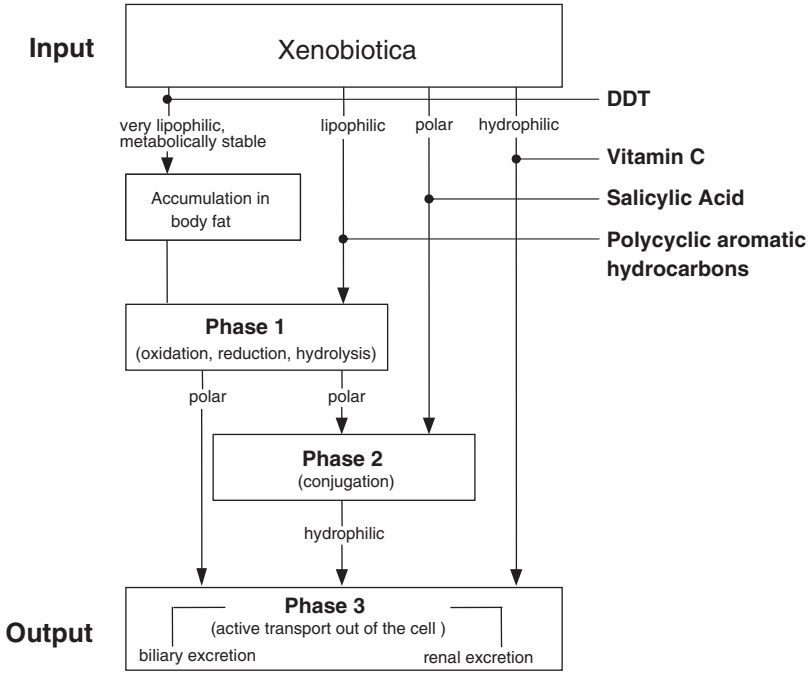
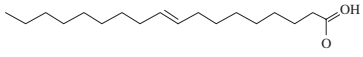
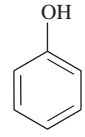
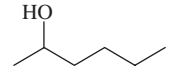
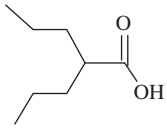
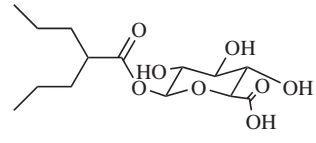
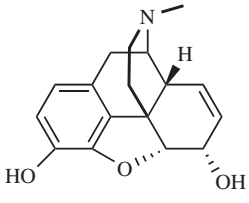
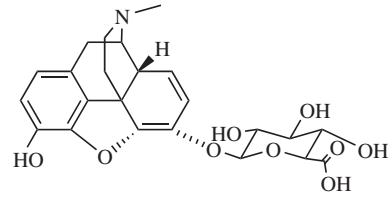
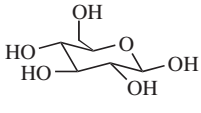


Figure 2.16 *Depending on its polarity a compound follows different pathways of elimination. Although a single elimination pathway can dominate, all pathways can work simultaneously. Reproduced (in adapted form) from Schober (2000), with permission from Dr. W. Schober.*

Table 2.3 also shows a list of Phase II enzymes, which transfer polar groups to products of the Phase I reactions. Glucuronic acid and sulfate conjugates are the most water soluble, but other conjugates such as acetyl, methyl, and aminoacyl groups can also be transferred. Conjugation with a glutathionyl group is the first step in a series of reactions that lead to the formation of a mercapturic acid derivative of a compound.

Table 2.4 demonstrates how metabolism can increase the polarity of a chemical, expressed as Log *P*, where *P*, the partition coefficient, equals the ratio of lipid solubility/water solubility. A decrease in Log *P* denotes greater polarity. The least soluble compound in the table is oleic acid which is not truly water soluble. Glucose is the other extreme as it is sufficiently water soluble and has no need for conjugation to increase its polarity. Polarity and excretion of compounds can be increased by Phase I metabolism, exemplified by benzene and hexane which are water insoluble but where hydroxylation by Phase I metabolism increases their polarity. However, Phase II metabolism is more efficient than Phase I in changing the polarity of compounds, exemplified by valproic acid and morphine where conjugation with glucuronic acid changes their *P* value more than oxidative Phase I metabolism does with benzene and hexane.

Table 2.4 *Changes of lipophilicity of compounds by metabolism. Phase II is more efficient in changing lipophilicity than phase I.*

Compound	Log P^a	Metabolite	Log P^a
Oleic acid	7.73		
			
Benzene	2.1	Phenol	1.5
			
Hexane	3.9	2-Hexanol	1.76
			
Valproic acid	0.20	Valproic acid-glucuronide	−1.85
			
Morphine	0.21	Morphium-glucuronide	0.001
			
β-D-Glucose	−3.3		
			

[a]Octanol/water partition coefficient. Logarithm of partition coefficient, P = [Organic]/[Aqueous]. Octanol models body fat, water models body water.

2.2A.2 The Enzymes Involved in Phase I Metabolism

Cytochrome P450

Phase I metabolism is dominated by CYP enzymes.

1. **The vast majority of compounds metabolized in Phase I are processed by CYP enzymes, which demonstrate broad substrate specificity. However, other Phase I enzymes can have a high turnover for specific substrates.**
2. **CYP enzymes can generate reactive metabolites that are more toxic than the original substrate.**

History Although the metabolism of nutrients such as carbohydrates, proteins, and lipids was recognized early in the 20th century, the appreciation that xenobiotic chemicals also underwent biotransformation was slower to develop. Key figures in developing our understanding regarding xenobiotic metabolism were R.T. Williams (1959) who studied the metabolism of many chemicals and their excretion products in a variety of animal species ranging from rodents to elephants, and B.B. Brodie (1950), among whose many contributions were the discovery of techniques to isolate and quantify the metabolites of many chemicals.

CYP enzymes were identified as hemeproteins and named cytochrome P450 by Omura and Sato (1964), and were shown to be the oxygen-activating enzymes in xenobiotic metabolism by Estabrook, Cooper, and Rosenthal (1965). CYP enzymes are present in high concentration in the hepatic endoplasmic reticulum, which after homogenization of the liver and differential centrifugation can be recovered in the microsomal pellet.

The CYP concentration of a microsomal sample can be measured by ultraviolet–visible light (UV-VIS) difference spectroscopy. A sample of microsomes suspended in a physiological medium, to which reducing agents have been added, is divided between 2 cuvettes and placed into the light path of a split-beam recording spectrophotomer, equipped with end-on photomultiplier tubes. Carbon monoxide is bubbled through the sample compartment but not through the reference compartment. The difference spectrum will show a highly absorptive peak at 450 nm, reflecting the binding of CO to the heme group of the CYP enzymes. Thus, the enzyme was given the name cytochrome P450.

Up to the 1970s it was thought that cytochrome P450 was one, liver-specific enzyme. When the enzyme was then purified, several enzymes could be separated. It was not until the isolation of the cDNA of the cytochromes in the late 1980s that it became clear that there were at least a dozen CYP in humans; only after the human genome was sequenced do we now know the complete number of the human cytochrome P450 superfamily, i.e. 57 CYP (see Table 2.5).

The Cytochrome P450 Superfamily

The CYP enzymes are divided into families and subfamilies. If the similarity in cDNA-derived amino acid sequence is >40%, then the enzymes are in the same family. If the similarity is >55%, then the enzymes are in the same subfamily.

Table 2.5 *A complete list of human and mouse putative full-length CYP genes. Currently 57 Human CYP genes are known (and 58 pseudogenes). This number is expected to be complete. Each gene additionally has its allelic variants*

Human	Mouse	Human	Mouse	Human	Mouse
CYP1A1	Cyp1a1	CYP2E1	Cyp2e1	CYP4F2	Cyp4f13
CYP1A2	Cyp1a2	CYP2F1	Cyp2f2	CYP4F3	Cyp4f14
CYP1B1	Cyp1b1		CYP2g1[b]	CYP4F8	Cyp4f15
CYP2A6	Cyp2a4	CYP2J2	Cyp2j5	CYP4F11	Cyp4f16
CYP2A7	Cyp2a5		Cyp2j6	CYP4F12	Cyp4f17
CYPA13	Cyp2a12		Cyp2j7	CYP4F22	Cyp4f18
	Cyp2a22		Cyp2j8		Cyp4f37
CYP2B6	Cyp2b9		Cyp2j9		Cyp4f39
	Cyp2b10		Cyp2j11		Cyp4f40
	Cyp2b13		Cyp2j12	CYP4V2	Cyp4v3
	Cyp2b19		Cyp2j13	CYP4X1	Cyp4x1
	Cyp2b23	CYP2R1	Cyp2r1	CYP4Z1	[a]
CYP2C8	Cyp2c29	CYP2S1	Cyp2s1	CYP5A1	Cyp5a1
CYP2C9	Cyp2c37		Cyp2t4[b]	CYP7A1	Cyp7a1
CYP2C18	Cyp2c38	CYP2U1	Cyp2u1	CYP7B1	Cyp7b1
CYP2C19	Cyp2c39	CYP2W1	Cyp2w1	CYP8A1	Cyp8a1
	Cyp2c40		Cyp2ab1[b]	CYP8B1	Cyp8b1
	Cyp2c44		CYP2aca-ps[b]	CYP11A1	Cyp11a1
	Cyp2c50	CYP3A4	Cyp3a11	CYP11B1	Cyp11b1
	Cyp2c54	CYP3A5	Cyp3a13	CYP11B2	Cyp11b2
	Cyp 2c55	CYP3A7	Cyp3a16	CYP17A1	Cyp17a1
	Cyp2c65	CYP3A43	Cyp3a25	CYP19A1	Cyp19a1
	Cyp2c66		Cyp3a41	CYP20A1	Cyp20a1
	Cyp2c67		Cyp3a44	CYP21A2	Cyp21a1
	Cyp2c68		Cyp3a57	CYP24A1	Cyp24a1
	Cyp2c69		Cyp3a59	CYP26A1	Cyp26a1
	Cyp2c70	CYP4A11	Cyp4a10	CYP26B1	Cyp26b1
CYP2D6	Cyp 2d9	CYP4A22	Cyp4a12a	CYP26C1	Cyp26c1
	Cyp2d10		Cyp4a12b	CYP27A1	Cy27a1
	Cyp2d11		Cyp4a14	CYP27B1	Cyp27b1
	Cyp2d12		Cyp4a29	CYP27C1	[a]
	Cyp2d13		Cyp4a30b	CYP39A1	Cyp39a1
	Cyp2d22		Cyp4a31	CYP46A1	Cyp46a1
	Cyp2d26		Cyp4a32	CYP51A1	Cyp51a1
	Cyp2d34	CYP4B1	Cyp4b1		
	Cyp2d40				

[a]Subfamily absent in mice.
[b]Subfamilies in mice with only pseudogene orthologs in humans. The gene names are listed numerically and alphabetically by subfamilies, and pairing does not necessarily denote orthologous genes, e.g. it is not known whether human CYP2A6 is the ortholog of mouse Cyp2a4. Genes having the identical combination of numbers and letters are orthologs between the two species (adapted from Nelson et al., 2004).

All of the CYP enzymes contain the same heme group but differ in the amino acid sequence of the protein portion of the enzyme. Their relationship to one another is based on the degree of similarity of their respective amino acid sequences. Ordinarily the amino acid sequence of a protein can be determined by isolating the protein and determining the sequence of amino acids using classical techniques. However, because of the difficulty of

obtaining purified CYP enzymes, the fact that some are expressed in very small quantities, and because there are so many enzymes, comparisons are based on cDNA-derived amino acid sequence.

When working with cDNAs it is clearly understood that a given amino acid may be represented by different codes, e.g., GTT and GTC both code for valine. If the amino acid sequence is the same, the function of the protein is the same, and the CYP are given the same name, despite noncoding sequence differences. If the cDNA-derived amino acid sequence is not exactly identical, the proteins are named differently. Based on amino acid sequence, families are composed of CYP enzymes having 40% identity in amino acid sequence and 55% for subfamilies.

In comparisons between species, if two CYPs have similar sequences and catalytic functions they are given the same name. For example, CYP1A1 appears in almost all known mammals and catalyses similar reactions. Human and mouse CYP1A1 are 80% similar in cDNA-derived amino acid sequence. Although not similar in sequence, because of catalytic similarities they were given the same name. Mouse CYP1A1 metabolizes many similar, but not always the same, substrates as does human CYP1A1. Thus, mouse CYP1A1 is considered the counterpart of human CYP1A1, despite the incomplete similarity.

A comparison of base sequences in two samples of DNA which code for the same CYP may display mutations which appear as single nucleotide polymorphisms (SNPs). An asterisk and an Arabic number after CYP indicates an allelic variant, which can have one or more SNPs. Thus, CYP2D6*4A denotes an allelic variant of CYP2D6 which differs from the wild type because of mutations. For example, in CYP2D6*4A at position 100 in the sequence a cytosine was replaced by a thymine (100C>T). Other SNPs in this sequence were: 974C>A; 984A>G; 997C>G; 1661G>C; 1846G>A; 4180G>C. The CYP nomenclature is an attempt to group similar cytochromes within one species together and to differentiate them from others. Although the method is not perfect, in practice it is indispensable.

Structure of Cytochrome P450

Cytochrome P450 is a heme-containing protein with an active site deep inside the protein.

The three-dimensional structure of proteins is best examined using X-ray crystallography. Mammalian CYP enzymes are bound into microsomal membranes and are not crystallizable. Some bacterial CYP enzymes, e.g., Cytochrome P450cam (CYP101) and cytochrome P450BM (CYP102) are water soluble. They were crystallized and their structures were determined. Recently, truncated human cytochromes CYP2A6, CYP2C8, CYP2C9, CYP2D6, and CYP3A4 were generated by removing the N-terminal domain which anchors the enzyme to the membrane. These proteins were soluble and could be crystallized and analysed. These structures provide a model for human cytochrome P450 (see Figure 2.17).

Figure 2.17 shows the protein ribbon three-dimensional model of CYP2D6 and the surface of the molecule. The alpha helices are shown in red, the beta strands in green, and the loops in blue. The heme group is shown as a ball-and-stick representation. Above the heme group there is a cavity in the shape of a ‘right foot’ (shown as a three-dimensional

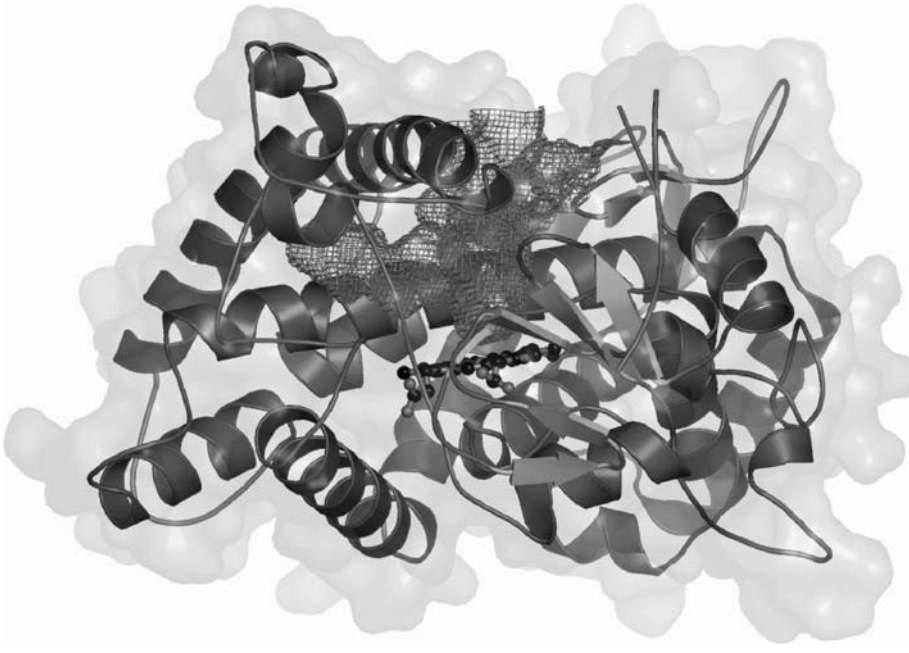


Figure 2.17 *Protein ribbon three-dimensional model of CYP2D6.*

mesh in dark blue), which extends down into the protein from a small opening at the surface. Access to the cavity probably involves some movement of the F-helix (at the top around the foot 'ankle') which would help to expand the size of the entrance to allow substrates in. It is not exactly clear how CYP2D6 embeds in the membrane. The cytochrome P450 reductase is believed to bind to the underside of the protein (as in the view shown in the Figure, the largely basic proximal face) (Rowland et al., 2006).

Although the exact structure of human cytochrome P450 embedded in the endoplasmic reticulum is not known, it is deduced that the channel to the active site determines the substrate specificity. The channel is sufficiently flexible to permit various substrates to approach the active site. The substrate binds to that site, somewhere between the heme group and a specific site on the protein.

Cytochromes P450 are heme-containing proteins that are bound to the endoplasmic reticulum and associated with electron-donating enzymes such as cytochrome P450, oxidoreductases, and cytochrome b_5.

Only intact cells have xenobiotic-metabolizing capacity because CYP is bound to the endoplasmic reticulum. There is no soluble CYP in serum, lymph fluids, or cytosol. The reaction occurs when the electron-donating substrate is positioned in proximity to oxygen which is activated at the heme group, and electrons necessary for the reaction

are donated from NADPH, via another membrane-bound enzyme called cytochrome P450 NADPH-dependent oxidoreductase. In some cases another microsomal cytochrome, cytochrome b_5 (not a cytochrome P450), increases the catalytic activity of CYP several-fold.

Broad Substrate Specificity

> Any electron-donating substrate that is appropriately positioned will be metabolized. The channel in the CYP protein leading to the active site determines the substrate specificity. The channel owes its flexibility to the 'breathing motion' of the protein, which permits a wide range of chemical structures to enter. The price paid for this broad substrate specificity is the relatively slow rate of enzyme activity.

Substrate specificity of the CYP enzymes is quite broad. Given sufficiently high substrate concentration, CYP enzymes can metabolize many different chemicals. The substrate overlap is a consequence of the structure of the system. Substrate specificity is controlled at the point of entry of the substrate into the channel of the protein leading to the active site of the cytochrome and depends upon whether the compound is properly positioned and whether there are enough electrons to activate the oxygen.

CYP2D6 has a preference for compounds with a basic nitrogen located 5–7 Å away from the site of oxidation. CYP1A1 prefers planar 'flat' molecules like polycyclic aromatic hydrocarbons; CYP2E1 prefers small molecules like ethanol and other solvents. In contrast, CYP3A4 has a very large active site that can accommodate two or more substrates simultaneously, e.g. midazolam can be metabolized at two different sites or both midazolam and α-naphthoflavone can be metabolized simultaneously. Alternatively, it can metabolize very large molecules like the polypeptide cyclosporin, an anti-immunogenic compound. Examples of the substrate specificity for CYP enzymes are given in Table 2.6.

The price paid for this broad substrate specificity is the relatively slow rate of metabolism by CYP enzymes. At maximum velocity at high substrate concentrations one molecule of cytochrome P450 metabolizes 1 to 20 molecules of substrate per minute. Compared with other enzymes this is very slow: acetylcholinesterase is very fast with a turnover of 1,500,000/min; generally, enzymes have a turnover of 60–600,000/min. An exception is a bacterial cytochrome P450, P450cam (CYP101) with a turnover of about 1600/min for camphor. Almost all compounds that are metabolized by cytochrome P450 are eliminated slowly.

Table 2.6 *In-vivo substrates for cytochrome P450 enzymes*

CYP	Marker substrate	CYP	Marker substrate
CYP1A2	Caffeine	CYP2C9	Tolbutamide
CYP1B1	Not known	CYP2C19	S-Mephenytoin
CYP2A6	Nicotine	CYP2D6	Dextromethorphan
CYP2B6	Bupropion	CYP2E1	6-Chlorzoxazone
CYP2C8	Taxol	CYP3A4	Midazolam

Remark: CYP1A1 is only present after induction.

Organ Specificity of Cytochrome P450

CYP enzymes are localized primarily in the hepatic endoplasmic reticulum. Many other organs express low levels of specific CYP enzymes. Little CYP activity is found in the brain or fat tissue.

The liver contains about 0.5–1.0 nmol cytochrome P450/mg microsomal protein, which represents about 5% of total liver protein. Most other organs express less than 0.1 nmol cytochrome P450/mg microsomal protein. An exception is the gut wall, which expresses quantities of CYP3A4 at levels similar to the liver. Because other CYP enzymes in the gut are not expressed at the same level as the liver the total amount of cytochrome P450 in the gut is lower than in the liver. Since the liver weighs about 1.7 kg in a 70 kg male, it is the main site of Phase I metabolism. Several cytochromes are predominantly expressed extrahepatically. CYP2S1 is expressed in epithelial tissues such as skin and mucous membranes. CYP1A1 and CYP1B1 are found in many organs throughout the body but not in the liver. The brain contains low concentrations of some CYP enzymes, such as CYP2D6, but it remains unclear whether they are involved in synthesis of endogenous compounds, in detoxification, or in metabolic activation.

Cytochrome P450 Forms Relevant to Humans

Of the 57 known human CYP enzymes only about 13 of them are predominantly active in xenobiotic metabolism. They are present for the most part in the liver in concentrations that exceed those of the other CYPs. Most known xenobiotics that are metabolized by CYP enzymes are metabolized by one of these 13 isoenzymes (Table 2.7).

Table 2.7 shows a list of 13 key CYP enzymes and indicates the relative amounts of each in the first column. The second column shows the estimated percentage of drugs that are metabolized by that CYP. Of the drugs known to be metabolized by CYP enzymes, CYP2C9, CYP2D6 and, especially, CYP3A4 dominate. CYP3A4 has the broadest substrate specificity of all CYP enzymes, and is the most abundant cytochrome P450 in the body, accounting for about 50% of total liver CYP. Owing to its low turnover of substrates it is not always the most dominant enzyme in all reactions. In some cases more efficient enzymes determine the rate of metabolism. For instance, CYP2D6 represents only about 2–5% of total liver CYP but is characterized by a high turnover rate for 22% of all drugs. Such generalization must be made with caution because metabolism of new chemicals may show that other CYP enzymes are more important than previously recognized.

The catalytic activity of CYP enzymes does not change with age. However, liver volume and liver blood flow decrease approximately 30% with age, resulting in a reduced clearance of drugs.

Table 2.7 *Quantitation of and quantitative role in xenobiotica metabolism of cytochrome P450 isoenzymes. Amount of cytochrome P450 in liver was determined by immunoquantitation. Participation of cytochrome P450 enzymes in metabolism of drugs is estimated. CYP2D6 is overrepresented in drug metabolism.*

Isoenzyme	% Of hepatic cytochrome P450	Estimated % of drugs metabolized
CYP1A1	<1	0
CYP1A2	8	3
CYP1B1		
CYP2A6	13	2
CYP2B6	7	3
CYP2C8	12	2
CYP2C9	18	18
CYP2C18	<1	
CYP2C19	4	9
CYP2D6	2	22
CYP2E1	9	2
CYP3A4	20	50
CYP3A5	0.2	1
Others	6.8	

From the moment of birth an infant is exposed to by-products of cellular metabolism, environmental toxins, and dietary constituents. In response the pattern of CYP enzyme expression changes. For instance, CYP3A7, present in fetal liver, is down regulated and CYP3A4, which is not present in fetal liver, quickly reaches adult levels. Rapid induction of CYP enzymes after birth is observed for CYP1A2, CYP2C8, CYP2C9, CYP2C18, CYP2D6, and CYP2E1. By the age of 6 months most CYP enzymes are equally expressed in infant and adult livers. During a lifetime the expression of CYPs and their catalytic activity is maintained. Clearance of drugs tends to be higher in children, in part because of the larger liver-to-body weight ratio and a higher blood flow through the liver. In old age the liver volume decreases up to 30% and liver blood flow decreases up to 35%, resulting in reduced drug metabolism.

Polymorphisms of Cytochrome P450

Almost all Phase I enzymes exhibit genomic polymorphisms. Some polymorphisms of CYP enzymes influence the rate of elimination of chemicals and the polymorphism of CYP2D6, 2C9, and 2C19 are clinically relevant. Polymorphisms can result in relevant variations in the susceptibility to drugs, chemicals at the workplace, or accidental ingestion. Whether polymorphisms are relevant at much lower environmental exposures, which do not lead to enzyme saturation, is debatable.

In the human genome SNPs can be detected at a rate of about 1 in about 500 base pairs. The open reading frame (ORF) of a CYP enzyme is approximately 1500 bp long and

Table 2.8 *Functional polymorphisms of cytochrome P450 enzymes.*

CYP enzyme	Wild-type allele	Most frequent genotype to yield poor metabolism
CYP1B1	CYP1B1*1	CYP1B1*6 and *7
CYP2A6	CYP2A6*1	CYP2A6*4
CYP2B6	CYP2B6*1	CYP2B6*5
CYP2C9	CYP2C9*1	CYP2C9*2 or *3
CYP2C19	CYP2C19*1	CYP2C19*2
CYP2D6	CYP2D6*1 or *2	CYP2D6*4 and *10 (Asians)

almost all show allelic variants. When the frequency of an allelic variant exceeds 1% it is called a polymorphism. Thus, many allelic variants of CYP enzymes exist, but their prevalence may be low or they exert no influence on catalytic activity. Silent variants describe cases when the base sequence between two genes is different whereas the translated protein is identical. For example, ATT and ATA both code for isoleucine so that the expressed proteins will be identical. Moreover, slight differences in amino acid sequence of CYP enzymes may not affect catalytic activity. Polymorphisms are clinically relevant when at equal dose and time intervals the null-genotypic individual is less capable of metabolizing a compound and therefore eliminates substrates for that enzyme more slowly than the wild-type individual. Such functional polymorphisms of cytochrome P450 are listed in Tables 2.8 and 2.9 and the influence of a functional polymorphism in drug elimination is depicted in Figure 2.18. Note that all these

Table 2.9 *Polymorphism of cytochromes P450 with clinical relevance and their probe substrate. Almost all phase I enzymes exhibit genomic polymorphisms. Most phase I enzymes exhibit clinically relevant genetic polymorphisms. However, the polymophisms listed below are generally accepted and more frequently encountered.*

Cytochrome[a]	Phenotype with mutation	Substrate[c]	In-vivo probe[d]
CYP1B1	Glaucoma	Unknown	Unknown
CYP2C9	Reduced elimination[b]	Tolbutamide, Warfarin, Phenytoin	Tolbutamide
CYP2C19	Reduced elimination	S-Mephenytoin, Omeprazole, Propranolol Proguanil	S-Mephenytoin
CYP2D6	Reduced elimination	Codeine β-Blockers, i.e. propafenone Antidepressants, i.e imipramine Antipsychotics, i.e. haloperidol	Dextromethorphan

[a]For more polymorphisms see http://www.sciencemag.org/feature/data/1044449.dtl and http://www.imm.ki.se/CYPalleles/, accessed January 2006.
[b]Of xenobiotica.
[c]For more substrates see http://www.sciencemag.org/feature/data/1044449.dtl and http://www.pharmgkb.org, accessed January 2006.
[d]Several other probes exist.

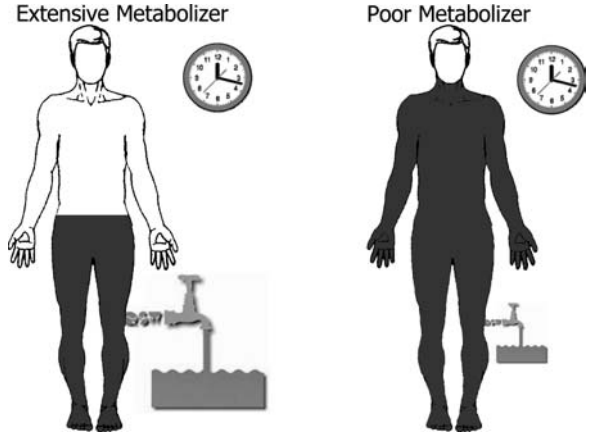


Figure 2.18 *Influence of polymorphism of Phase I enzymes on compound elimination. Given equal dose and time interval the null-genotypic individual will have more remaining compound in the body than the wild-type individual because the null-genotype for an enzyme eliminates compounds for that enzyme more slowly.*

examples are drugs, which need to be given at high doses as compared with human environmental exposure to chemicals.

Table 2.8 shows that the nomenclature for allelic variants is the name of the gene combined with an asterisk and the number of the variant. For example, CYP2D6*10 is a variant with reduced catalytic activity and stability of the protein and is usually found in people of Asian descent. The number is determined by convention and often by order of discovery.

Table 2.9 demonstrates several polymorphisms of clinical relevance because a reduction in the rate of metabolism of the specific substrate has been shown.

Although allelic variants can be detected in most CYP enzymes, few appear to be functional polymorphisms. Currently, CYP2D6, CYP2C9, and CYP2C19 are best known to influence drug clearance and allow differentiation between individuals who are poor, intermediate, extensive, or ultra-rapid metabolizers. Further research on other CYP enzymes such as CYP2A6 and CYP2B6 is likely to demonstrate clinically relevant polymorphisms.

CYP2D6 metabolizes many compounds and displays a number of allelic variants. Figure 2.19 shows the results of a study of polymorphic metabolism by CYP2D6 in 316 subjects. The ordinate shows the genotype of the subjects and the abscissa (MR_s) is a measure of the metabolic rate. The tetra-modal distribution of metabolic rates can be classified into poor metabolizers ($MR_s > 15$), intermediate metabolizers (MR_s 1–10), extensive metabolizers (MR_s 0.2–10), and ultra-rapid metabolizers ($MR_s < 0.2$). At least 89 allelic variants of CYP2D6 are known, with some allelic variants observed more frequently than others. Poor metabolism is explained by the presence of two 2D6*4

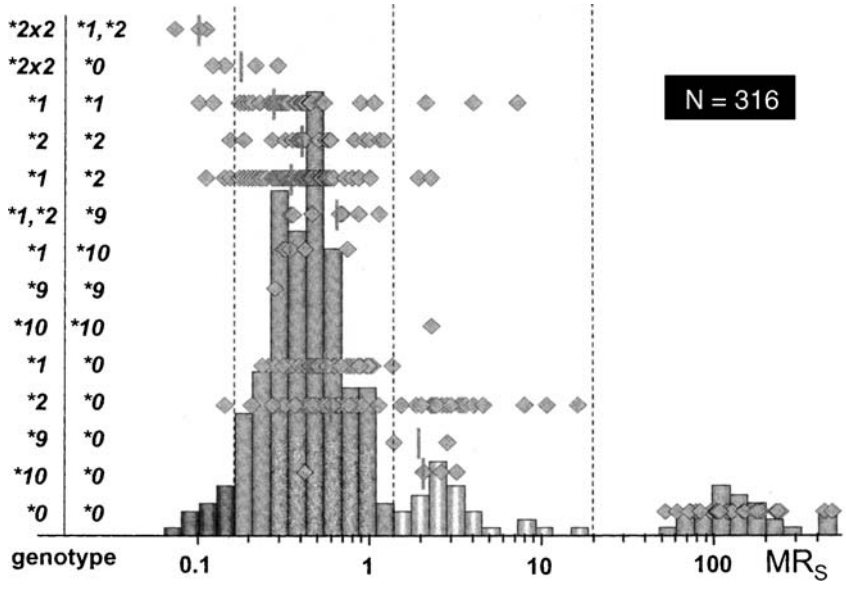


Figure 2.19 *4-Modal distribution of the metabolic ratios of a CYP2D6 substrate. Depending on their genotype, individuals either are poor (metabolic ratio > 15), intermediate, extensive, or ultra-rapid metabolizers (metabolic ratio < 0.2). The intermediate metabolizers often have a combination of CYP2D6*2 with a newly discovered allele CYP2D6*41. CYP2D6*0 denotes all genotypes with no catalytic activity, mainly CYP2D6*4. Inducers do not influence CYP2D6, and genotype determines phenotype [adapted and reprinted from Zanger et al., Naunyn-Schmiedeberg's Arch. Pharmacol. (2005),* **369***, 23–37].*

alleles, intermediate metabolism by a combination of CYP2D6*41 with *4 alleles, extensive metabolism by CYP2D6*1 (36% allele frequency) or *2 (24% allele frequency), and the ultra-rapid metabolism by the presence of multiple copies of the CYP2D6*1 or *2 alleles. Ultra-rapid metabolizers had up to 13 copies of the gene and were found especially in Ethiopians. CYP2D6 has no known inducer and is always constitutively expressed at the same level. Only by making gene duplicates can the catalytic activity be increased in a population exposed to dietary compounds metabolized by CYP2D6.

In clinical practice the polymorphisms of CYP2D6, CYP2C9, and CYP2C19 have a relevant impact on the clearance of drugs. Treatment with anticoagulants such as warfarin reduces the likelihood of intravascular thrombus formation, which can be manifested as cardiac infarctions or strokes. Warfarin, which is primarily metabolized by CYP2C9, has a small therapeutic window and overdose can lead to excessive bleeding. As a preventive measure clotting time is determined regularly during therapy. Blood levels of anti-con-vulsants such as mephenytoin, which are primarily metabolized by CYP2C19, should also be monitored.

Human CYP 2C enzymes metabolize many drugs. Several polymorphic variations of these enzymes have been described. People who express CYP2C19*2 are poor metabolizers of mephenytoin, and people who express CYP2C9*2 or *3 are poor

metabolizers of warfarin. When treated with otherwise acceptable doses of these drugs poor metabolizers are likely to respond as if they were treated with toxically active overdoses.

To prevent these events it would seem appropriate to genotype potential users of these drugs to detect poor metabolizers. However, genetic prediction of catalytic activity by genotyping is seldom performed because we do not completely understand the impact of single or multiple allelic variants on phenotype and because of the superimposed influences of regulation of expression on phenotype.

Genotyping to prevent adverse drug reactions is seldom mandated, although it was recently recommended in patients about to be treated with 6-mercaptopurine. For people lacking the enzyme thiopurine methyltransferase (TPMT), which is not a CYP, but which would normally metabolize 6-mercaptopurine, chemotherapy with this anticancer drug would be life threatening. There is as yet no mandatory phenotyping for any CYP enzyme.

Although many SNPs have been identified in the CYP3A4 gene they do not lead to significant variability in metabolism. Nevertheless, females express more CYP3A4 than do males and display greater catalytic activity than do men for CYP3A4 substrates such as in the metabolism of verapamil to norverapamil. The gender-based difference in metabolism has not been reported to result in adverse effects in patients undergoing verapamil therapy.

During the development of new drugs, compounds are preferred that are not eliminated by a single CYP. Elimination dominated by a single enzyme harbors the danger that null-genotypes for this enzyme show higher blood levels than wild-type genotypes, which could lead to a higher prevalence of side effects in the null-genotypes. Thus, there is a preference for newly developed drugs which undergo metabolism via multiple competing pathways.

Noncytochrome P450 Phase I Enzymes

Epoxide Hydrolases

Epoxides are detoxified to diols by epoxide hydrolases, which protect cells from the formation of both DNA and protein adducts by these very reactive species found in diet or formed during Phase I metabolism. There are microsomal and soluble epoxide hydrolases with different substrate specificities. At present, 3 epoxide hydrolases are known: Cholesterol epoxide hydrolase (chEH), microsomal epoxide hydrolase (mEH, EPHX1), and soluble epoxide hydrolase (sEH, EPHX2).

Epoxides are ubiquitously found in the environment and are also generated during Phase I by enzymes such as the CYPs and cyclooxygenases (COX). Not all CYP metabolites are epoxides. However, CYP oxidations at carbon–carbon double bonds lead to epoxide formation. Most epoxides are of intermediate reactivity and do not present an acute danger to cells. However, some epoxides are highly reactive electrophilic compounds, which are able to react with nucleophilic macromolecules like DNA or proteins. These highly and intermediate reactive epoxides require detoxification.

Spontaneous rearrangement of epoxides results in the formation of phenolic metabolites or glutathione conjugates. Epoxide hydrolases rapidly hydrolyse epoxides (oxiranes) and yield vicinal diols. Since the hydrolysis of the epoxide bond is energetically favourable, only water is needed as a co-substrate. Epoxide hydrolases are expressed in nearly all tissues but their concentration is highest in the liver.

Cholesterol epoxide hydrolase is a little known enzyme and metabolizes cholesterol 5,6-oxide to its diol.

Microsomal epoxide hydrolase (mEH, EPHX1) is a key enzyme involved in the hydrolysis of xenobiotic epoxides. It has wide substrate specificity with a high apparent affinity. The turnover of substrates by the enzyme is low. These contradicting features are accomplished by a high affinity for binding substrates but a low substrate turnover. The enzyme absorbs the substrates like a sponge, removing toxic intermediates from the cytosol, but is slow in metabolizing them to the corresponding diols. Since low substrate concentrations are usual, this is an efficient way to detoxify epoxides.

Polymorphisms in microsomal epoxide hydrolase have been described but seem to have limited impact on catalytic activity. However, an association of a polymorphism with low epoxide hydrolase activity was reported to be protective in lung cancer patients. Mice with a genetic deletion of EPHX1 were resistant to the carcinogenic effects of 7,12-dimethylbenz[*a*]anthracene.

Soluble epoxide hydrolase is involved in the metabolism of endogenous epoxides like arachidonic epoxides (epoxyeicosatrienoic acid, EETs) and linoleic acid epoxides (leukotoxins). EETs regulate blood pressure. Inhibition of EPHX2 could be a potential target for the treatment of hypertension in man. Leukotoxins elicit inflammatory symptoms, and inhibition of soluble epoxide hydrolase could have anti-inflamatory effects. Polymorphisms in the EPHX2 were associated with differences in lung cancer susceptibility.

Soluble epoxide hydrolase, in contrast to microsomal epoxide hydrolase, is not inducible by classic Phase I inducers. However, it can be induced in experimental animals by peroxisome proliferators, resulting in tumors in, for instance, mice. The human receptor for peroxisome proliferation (PPAR-α) is about 10-fold less sensitive than the rodent receptor for peroxisome proliferation, explaining the low induction of soluble epoxide hydrolase in humans. This probably explains the absence of peroxisome proliferators-induced tumors in humans.

Flavin-containing Monooxygenases (FMO)

FMO and CYP enzymes metabolize similar substrates, but FMO cannot oxidize C–H bonds. Unlike CYP, FMO reactions do not produce reactive metabolites, but its substrate specificity is smaller than that of CYPs. Defective FMO3 mutants, the main FMO in humans, are responsible for the 'Fish Odor Syndrome'.

Flavin-containing monooxygenases resemble cytochrome P450 in that both utilize NADPH and oxygen to function, and are membrane-bound enzymes. FMO and CYP accept similar substrates, but FMO accepts weaker nucleophiles than does CYP, and does not oxidize C–H bonds like CYP does. Both accept N-, S-, Se-, and P-containing

compounds as substrates, forming N-, S-, Se-, and P-oxides. Both have broad substrate specificity. In contrast to CYP there are few inhibitors and inducers of FMO, and consequently drug interaction due to FMO interactions is rare. Furthermore, adverse drug reactions due to FMO metabolism are rare since FMO reactions are mostly detoxifying and do not generate reactive intermediates. FMO metabolizes chemicals as soon as the substrate binds because its active site contains activated organic peroxide. Compounds that can react with peroxides are substrates for FMO.

There are 5 FMO genes (FMO1–5) and 6 pseudogenes. FMO3, the most important, is found primarily in the liver but in other tissues as well. It is found in the liver at a concentration of about 60% of that of CYP3A4. FMO1 is the main FMO found in the kidney. FMO2 is found in lungs. Several mutations have been found which code for inactive FMO3. The result is an inability to metabolize endogenously generated trimethylamine to its *N*-oxide. Trimethylamine is then exhaled unchanged and is responsible for the 'fish odor syndrome'.

In-vitro the role of FMO in drug metabolism is often underestimated due to the instability of the enzyme and inadequate handling of the enzymatic preparations.

Alcohol Dehydrogenases

Alcohol dehydrogenases oxidize alcohols to aldehydes, which are subsequently detoxified by aldehyde dehydrogenases to carboxylic acids. Ethanol is the main biological substrate.

Alcohol dehydrogenases (ADH) are a group of zinc-containing enzymes that metabolize alcohols to aldehydes. They are cytosolic enzymes of about 40 kDa and are not inducible. In humans 5 different classes of alcohol dehydrogenases have been identified. The liver expresses the highest levels of ADH1 and ADH2. ADH3 is ubiquitous and ADH4 is expressed in the aero-digestive tract. Alcohol dehydrogenase is a homodimer formed by subunits coded by the genes ADH2–4. ADH1 can form homo- or heterodimers of ADH1A–1C. The main substrates of toxicologic interest are methanol, ethanol, and ethylene glycol for which ADH1 and ADH2 are quantitatively the main metabolizing enzymes.

Ethanol is mainly metabolized by ADH1 and ADH2. A genetic polymorphism of the ADH2 exists with a normal, ADH2*1 and 40-times more active ADH2*2 phenotype toward ethanol. 90% of East Asians express the more active ADH2*2 enzyme; in the Caucasian and African population only 20% express this enzyme. 40% of East Asians concomitantly express the defective aldehyde dehydrogenase ALDH2*2 and the formed acetaldehyde is more slowly eliminated than in ALDH2*1 phenotypic individuals. Acetaldehyde is responsible for the facial flushing observed after drinking alcoholic beverages. ADH also converts acetaldehyde back into ethanol at higher acetaldehyde levels. Thus, accumulation of acetaldehyde explains both the increased sensitivity toward ethanol and the facial flushing in East Asians. Individuals with expression of both ADH2*2 and ALDH2*2 are less prone to heavy drinking and alcoholism, perhaps due to the discomfort experienced after imbibing.

The K_m of ADH for ethanol is 0.2–2 mM. Ingestion of 100 ml beer into an empty stomach leads to a blood concentration of 0.1‰ (1.7 mM), which is above the K_m.

Ethanol above 0.1‰ is eliminated by zero-order kinetics, with the consequence that the elimination rate of ethanol is independent of its blood concentration. Ethanol-metabolizing capacity of an adult is equivalent to about 300 ml beer/h.

ADH has a higher affinity for ethanol (175 mg/kg/h) than for methanol (25 mg/kg/h) and ethylene glycol. Methanol toxicity results from high levels of formic acid formed from methanol by ADH via formaldehyde. Ethanol blocks the formation of formaldehyde from methanol by competitive inhibition of ADH and is used as an antidote in methanol intoxication. Instead of being metabolized the methanol is exhaled. Ethylene glycol itself is non-toxic but the oxalic acid formed by ADH and ALDH forms oxalate crystals in the kidney after ingestion of high doses. Preventing metabolism of ethylene glycol by blocking the ADH with massive doses of ethanol (0.5–1‰) prevents kidney toxicity. Fomepizol (4-methylpyrazole) is a modern blocker of ADH used in the treatment of ethylene glycol and methanol intoxications. ADH3 (glutathione-dependent formaldehyde dehydrogenase) catalyses detoxification of formaldehyde.

Aldehyde Dehydrogenases

Aldehyde dehydrogenases metabolize aldehydes to their corresponding carboxylic acids. Acetaldehyde is the most important substrate. 40% of East Asians are deficient in ALDH2, the main acetaldehyde-metabolizing enzyme.

Aldehyde dehydrogenases are involved in the oxidative degradation of aldehydes to their corresponding carboxylic acids. Examples of substrates include acetaldehyde, retinal, aldehydes derived from fatty acid peroxidation such as 9-hydroxynonenal, and aldehyde intermediates in the metabolic degradation of GABA, folate, and cyclophosphamide.

Aldehyde dehydrogenases are divided into 11 different families separated by <40% homology in their cDNA-derived amino acid sequence. A total of 19 aldehyde dehydrogenase genes and 3 pseudogenes, which express 26 polymorphic forms, have been reported. Several mouse strains, which exhibit genomic deletions of ALDH forms, display lethal phenotypes, exemplifying the importance of this class of enzymes. Inborn errors in the ALDH6A1 gene are responsible for developmental delays; errors in the ALDH3A2 gene (trivial name ALDH10) explain the Sjögren–Larsson Syndrome, i.e., mental retardation and skin defect due to defective fatty acid metabolism in skin; mutations in the ALDH4A1 explains type II hyperprolinemia resulting in mental retardation; and the ALDH2*2 form with defective acetaldehyde oxidation capacity explains the flushing syndrome and pronounced responses to alcohol in many Asians. The trivial names do not always correspond to their gene abbreviations (see www.aldh.org).

The main organ of expression is the liver. ALDH1 and ALDH3 are cytosolic enzymes but the most important ALDH2 is a mitochondrial enzyme. For ALDH2 an E487K (glutamic acid-to-lysine) mutation codes for an isoform without catalytic activity. Persons bearing this inactive form show alcohol intolerance. ALDH2 is the major aldehyde dehydroxygenase that metabolizes acetaldehyde with a K_m of $< 5\,\mu M$. Disulfiram (Antabuse) is an inhibitor of ALDH2 and was used as an adjuvant in the therapy of alcoholics. It increases the concentration of acetaldehyde and induces ethanol avoidance. Alcoholics who continued to drink despite Antabuse therapy experienced

sudden and severe side effects, induced by the exaggerated levels of acetaldehyde. The therapy is now considered obsolete.

Esterases

Esterases hydrolyse esters or amides. Esterases and amidases are the same enzymes. They are targeted by nerve gasses and many pesticides. Many drugs are ester pro-drugs that become active after de-esterification.

Esters are ubiquitous in the diet and in products of daily use, and many drugs are delivered as pro-drugs in the form of esters to improve solubility, absorption, and bioavailability. Generally, hydrolysis rates for esters are higher than for amides, although both reactions are catalysed by the same enzymes. Esterases are grouped into several subfamilies of the superfamily of α, β-hydrolases, which encompasses about 450 members. This family includes many subfamilies such as haloperoxidases, thioesterases, epoxide hydrolases, lipases, cholesterol esterases, carboxylesterases, cholinesterases, and paraoxonases and others (see Table 2.10 and http://bioweb.ensam.inra.fr/esther).

Esterases are the target of deliberate poisoning by nerve gasses like Tabun, Sarin, VX, and others and accidental poisoning by pesticides like parathion, paraoxon, and malathion. Because of the toxicological importance of these organophosphates some authors classify esterases as follows:

- A-esterases, which metabolize organophosphates,
- B-esterases, which are targets of organophosphates but do not accept them as substrates (e.g., they inhibit but are not metabolized by acetylcholinesterase),
- C-esterases, which do not interact with organophosphates.

Unlike mammals, insects have less A-Esterases than B-Esterases and are, therefore, more sensitive to these nerve poisons.

Table 2.10 *Subdivision of esterases as subfamilies of α, β-hydrolases.*

Subfamily	Enzyme	Abbreviation	Substrate
Cholinesterases	Acetylcholinesterase	AChE	Acetylcholine
(B-Esterases)	Butyrylcholinesterase	BuChE	Suxamethonium
Carboxylesterases	Carboxylesterase 1	hCE-1, CES1	Cocaine, Cilazapril, Methylphenidate
(B-esterases, alkylesterases)	Carboxylesterase 2	hCE-2, CES2	Heroin, procaine
	Carboxylesterase 3	hCE-3	
	Carboxylesterase 4	hCE-4	
	Neuropathy Target Esterase	NTE	Organophosphates
Paraoxonases		PON1	Organophosphates
(A-esterases,		PON2	(paraoxon)
Arylesterases)		PON3	Spironolactone, lovastatin

More commonly, esterases are regarded as subfamilies of the superfamily α, β-**hydrolases.** The subfamily of choline esterases includes acetylcholinesterase (AChE), an enzyme with an extremely high turnover for acetylcholine (1,500,000/min), but few other substrates. Organophosphates target this enzyme. Butyrylcholinesterase (BuChE, pseudocholinesterase) has a more promiscuous substrate specificity. A polymorphism in this gene is responsible for the unusual persistence of the action of suxamethonium (succinylcholine) that is used as a muscle relaxant in anaesthesia. In individuals deficient in this enzyme the normal duration of relaxation or paralysis is prolonged. Sensitivity to succinylcholine was one of the first genetic polymorphisms discovered.

Carboxylesterases (hCE, CES) are expressed everywhere, but the highest levels are found in the endoplasmic reticulum of the liver. No carboxylesterase activity is found in plasma. hCE-1 and hCE-2 are the major forms expressed in humans. hCE-1 metabolizes cocaine to its main metabolite benzoylecgonine. Methylphenidate (Ritalin®), an important stimulant prescribed to treat attention-deficit hyperactivity disorder, is metabolized to the inactive ritalinic acid by hCE-1A1. hCE-1 prefers esters with a small alcohol group like ethyl, and larger acyl groups. hCE-2 prefers the opposite, i.e., larger alcohols but smaller acyl groups. A special form of carboxylesterase is NTE (Neuropathy Target Esterase), a carboxylesterase located in neurons that is targeted by organophosphates, where an unknown mechanism leads to long axon degradation.

Paraoxonases (arylesterases, lactonases) have no sequence analogy with carboxylesterases and cholinesterases and are a different class of esterase. They prefer lactones and aryl esters like lovastatin, simvastatin, and spironolactone as substrates. Paraoxonase, PON1, metabolizes paraoxon, the active metabolite and toxic ingredient of parathion, to inactive metabolites. PON1 is expressed in blood on HDL vesicles, and reduces oxidized low-density lipoprotein (LDL) and cholesterol, which play a role in atherosclerosis and coronary heart disease.

Mice with a genomic deletion of the PON1 gene develop atherosclerosis, and high PON1 activity in humans protects against pesticide toxicity, atherosclerosis, and coronary heart disease.

2.2A.3 Phase III (Metabolism)

Phase III metabolism describes efflux of compounds out of cells by energy-dependent transporters. The removal of negatively charged glucuronides, sulfates, and glutathione conjugates generated in Phase II metabolism is predominantly carried out by the transporters ABCC2 (old nomenclature MRP2, multidrug resistance related protein 2) and ABCG2 (old nomenclature BCRP, breast cancer resistance protein). The term Phase III metabolism is erroneous, as, in contrast to Phase I and Phase II metabolism, it describes transport of compounds across membranes without altering their chemical structure.

Transporters take care of selective uptake and efflux of components through the cell membrane. Uptake transporters like OATs (organic anion transporters), OCTs (organic cation transporters), and OATPs (organic anion transporter proteins) are usually bidirectional but at physiologic concentrations uptake dominates.

Table 2.11 *Selection of human transporter proteins with a function in xenobiotica elimination.*

Enzyme	Trivial name	Full trivial name
Uptake transporters		
SLC22A1[a]	OCT1	Organic Cation Transporter 1
SLC22A7	OAT2	Organic Anion Transporter 2
SLCO1A2[b]	OATP1A2	Organic Anion Transporter-Polypeptide 1A2
SLCO1B1	OATP1B1	Organic Anion Transporter-Polypeptide 1B1
SLCO1B3	OATP1B3	Organic Anion Transporter-Polypeptide 1B3
SLCO2B1	OATP2B1	Organic Anion Transporter-Polypeptide 2B1
Efflux transporters		
ABCB1[c]	MDR1, PGP	Multidrug Resistance Protein 1, P-glycoprotein
ABCB4	MDR3	Multidrug Resistance Protein 3
ABCB11	BSEP	Bile Salt Export Pump
ABCC1	MRP1	Multidrug Related Protein 1
ABCC2	MRP2, cMOAT	Multidrug Related Protein 2
ABCC3	MRP3	Multidrug Related Protein 3
ABCC4	MRP4	Multidrug Related Protein 4
ABCC5	MRP5	Multidrug Related Protein 5
ABCC6	MRP6	Multidrug Related Protein 6
ABCG2	BCRP	Breast Cancer Resistance Protein

[a]Solute-linked carrier protein.
[b]Solute carrier organic anion protein.
[c]ATP-binding cassette transporter. MDR2 is the rodent homolog of human MDR3.

The transporters (see Table 2.11) are grouped into the superfamilies ABC transporter (ATP-binding cassette proteins), SLC (Solute Linked Carrier Proteins) and SLCO/OATP (Solute Linked Organic Anion Transporter Protein). Uptake transporters fall into the SLC and SLCO families. Uptake is sometimes called Phase 0, and is not discussed here, but some major uptake transporters are also depicted in Figure 2.20.

Efflux transporters limit the concentration of organic ions inside the cell, and transport metabolites formed by Phase I and Phase II metabolism out of the cell. This prevents accumulation of hydrophilic Phase II metabolites like glucuronides, sulfates, and glutathione derivates inside the cell. Most efflux transporters are members of the ABC transporter family. They require the energy of ATP to function as pumps for transporting compounds out of the cell.

There are 48 known ABC transporters. Important ABC transporters in efflux detoxification are depicted in Figure 2.20 and in Table 2.11. Additional pumps can perform more specialized tasks. ABCB1 and ABCB3 are the most dominant organic cation efflux pumps for xenobiotica like alkaloids, steroids, anthracyclines, digoxin, paclitaxel, cyclosporin, tacrolimus, and others. Glucuronides, sulfates, and glutathione conjugates are negatively charged and are not substrates for the cation pumps ABCB1 or ABCB3. In the liver, which is the major site for the generation of these conjugates, Phase II conjugates are transported into the bile by ABCC2 (old nomenclature MRP2) and ABCG2 (BCRP) and at the basolateral side (sinusoidal side, liver blood side) by ABCC3 (MRP3) and ABCC4 (MRP4) transporters.

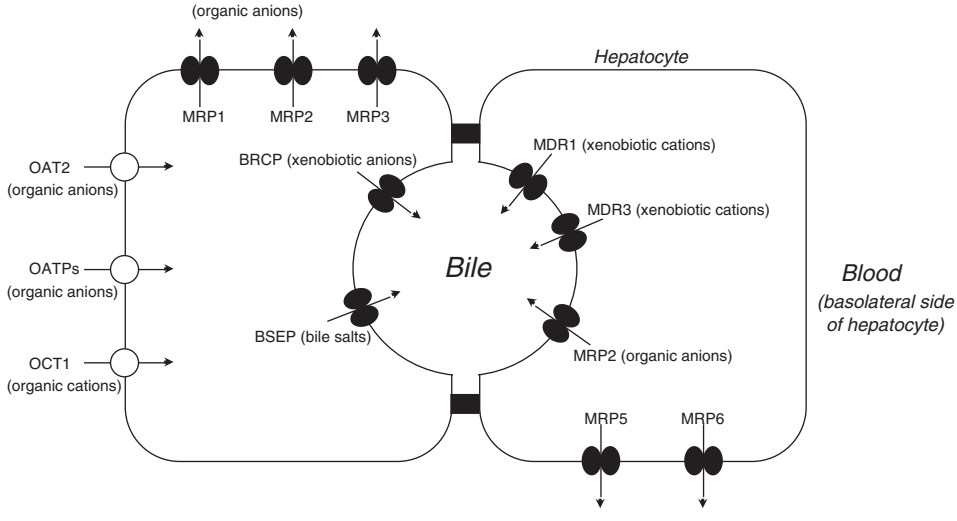


Figure 2.20 *Location and function of hepatocyte transporters for xenobiotica. Efflux transporters (solid symbols): MDR (multidrug resistance protein: ABCB); MRP (multidrug resistance-associated protein: ABCC); BCRP (breast cancer resistance protein: ABCG2); ABCB11 (bile salt export pump: BSEP); Uptake transporters (open symbols): OATP (organic anion-transporting polypeptide); OCT (organic cation transporter); OAT (organic anion transporter).*

Sulfate conjugates are generally transported into bile by ABCC2 (MRP2) and ABCG2 (BCRP). Glucuronides are mostly formed in the endoplasmic reticulum of the liver. They are excreted into bile by ABCC2 and ABCG2 also, but the main route of elimination is into the blood by ABCC3 (MRP3) from where they are carried by the blood to the kidneys, where they are excreted into urine by glomerular filtration. Glutathione conjugates are predominantly exported by ABCC2 into the bile and by ABCC1 (MRP1) into the blood. In the brain, ABCB1 and ABCC1 are important efflux transporters maintaining the blood–brain barrier.

Efflux transporters are major players in the process of eliminating xenobiotics and thus serve to detoxify or limit the action of compounds that are substrates. To eliminate compounds across a cellular barrier like from blood to bile, from brain to blood, or from intestine into blood, uptake transporters regulate the uptake into cells and efflux pumps of Phase III metabolism export out of the cell into the lumen. In this process many transporters are involved but ABCB1, ABCC2, and ABCG2 are the major pumps involved in Phase III excretion.

Especially in drug therapy, enhanced transport out of the cells may have adverse consequences. For instance, ABCB1 is responsible for the removal of many compounds out of the brain across the blood–brain barrier. ABCB1 expression can also confine resistance to chemotherapeutic agents. Cancer cells during treatment with chemotherapeutic agents like vinblastine can develop over-expression of ABCB1, which then prevents intracellular build-up of sufficient concentrations of the chemotherapeutic agent and makes cancer cells resistant against this therapy.

Triglitazone – a blockbuster treatment for type 2 diabetes - was withdrawn from the market because of severe hepatotoxicity. Among the reasons was that triglitazone and especially its sulfate inhibited the bile salt transporter BSEP (ABCB11). Thus bile salts, instead of being stored as nontoxic micelles in the bile, deployed their detergent action inside the hepatocytes, which led to severe hepatotoxicity.

2.2A.4 Summary

Phase I metabolism chemically converts lipophilic compounds into more hydrophilic compounds, which then can be excreted directly. More often these chemical changes provide the 'handles' to which Phase II enzymes conjugate very water-soluble groups, making excretion simple. The reactive metabolites formed by Phase I metabolism very often interact with macromolecules and lead to toxicity.

Phase I metabolism is dominated by the CYP enzymes, which are the oxygen-activating enzymes in drug metabolism. For many substrates, however, other Phase I enzymes play a role in oxidative metabolism. To excrete a compound many enzyme systems work simultaneously and, depending on the characteristics of a substrate, no pathway, a few pathways, or even a single pathway dominates.

Cytochrome P450 can be divided into families and subfamilies. This subdivision makes Phase I metabolizing enzymes more understandable than a telephone book of single enzymes. Because of the prevalence of Single Nucleotide Polymorphisms which are found in the genome once in about every 500 base pairs, many polymorphisms exist in Phase I enzymes, some of which were shown to influence drug-excretion rates.

References

Nelson, DR, Zeldin, DC, Hoffman, SM, Maltais, LJ, Wain, HM, Nebert, DW. Comparison of Cytochrome P450 (CYP) Genes from the Mouse and Human Genomes, including Nomenclature Recommendations for Genes, Pseudogenes and Alternative-splice Variants. *Pharmacogenetics* (2004), **14**, 1–18.

Rowland, P, Blaney, FE, Smyth, MG, Jones, JJ, Leydon, VR, Oxbrow, AK, Lewis, CJ, Tennant, MG, Modi, S, Eggleston, DS, Chenery, RJ, Bridges, AM. Crystal Structure of the Human Cytochrome P450 2D6. *J. Biol. Chem.* (2006), **281**, 7614–7622.

Schober, W. Thesis for a PhD Degree in Natural Sciences, Medical Faculty, Institute of Toxicology and Environmental Hygiene, Technical University Munich (2000) (ISBN 3-932108-10-8).

Zanger, UM. Personal Communication (2005).

http://bioweb.ensam.inra.fr/esther

http://www.aldh.org

http://www.imm.ki.se/CYPalleles/

http://www.pharmgkb.org

http://www.sciencemag.org/feature/data/1044449.dtl

For Further Reading

Cashman, JR. Some Distinctions between Flavin-containing and Cytochrome P450 Monooxygenases. *Biochem. Biophys. Res. Commun.* (2005), **338**, 599–604.

Evans, WE, Relling, MV. Pharmacogenomics: Translating Functional Genomics into Rational Therapeutics. *Science* (1999), **286**, 487–91.

Ingelman-Sundberg, M. Human Drug Metabolizing Cytochrome P450 Enzymes: Properties and Polymorphisms. *Naunyn-Schmiedeberg's Arch. Pharmacol.* (2004), **369**, 89–104.
Meyer, UA. Pharmacogenetics - Five Decades of Therapeutic Lessons from Genetic Diversity. *Nature Rev. Genet.* (2004), **5**, 669–676.
Morisseau, C, Hammock, BD. Epoxide Hydrolases. Mechanisms, Inhibitor Designs, and Biological Roles. *Annu. Rev. Pharmacol. Toxicol.* (2005), **45**, 311–333.
Satoh, T, Taylor, P, Bosron, WF, Sanghani, SP, Hosokawa, M, La Du, BN. Current Progress on Esterases: From Molecular Structure to Function. *Drug Metab. Dispos.* (2002), **30**, 488–493.
Vasiliou, V, Pappa, A, Estey, T. Role of Human Aldehyde Dehydrogenases in Endobiotic and Xenobiotic Metabolism. *Drug Metab. Rev.* (2004), **36**, 279–299.
Zanger, UM, Raimundo, S, Eichelbaum, M. Cytochrome P450 2D6: Overview and Update on Pharmacology, Genetics, Biochemistry. *Naunyn-Schmiedeberg's Arch. Pharmacol.* (2004), **369**, 23–37.

2.2B Drug Metabolism

Leslie Schwarz and John B. Watkins

2.2B.1 Conjugation

Conjugation of xenobiotics typically leads to more water soluble and therefore more easily excretable compounds.

In conjugation reactions, molecules or molecule moieties formed during intermediary metabolism are transferred to xenobiotics. Basically two types of reactions can be distinguished, i.e. conjugation of electrophilic or nucleophilic compounds. The conjugating agent for electrophilic compounds is principally the tripeptide glutathione (GSH, γ-glutamylcysteinylglycine). Conjugation partners for nucleophilic compounds include UDP-glucuronic acid (UDPGA), 3′-phosphoadenosine-5′-phosphosulfate (PAPS), acetyl-coenzyme A, *S*-adenosylmethionine, glycine, and glutamine.

Glutathione S-Transferases and Their Reactions

Many reactive electrophilic compounds form conjugates with glutathione (GSH). The conjugation reactions are catalysed by GSH *S*-transferases. Strong electrophiles may also react nonenzymatically with GSH. All cells contain GSH in concentrations that may reach 10 mmolar in some tissues.

Glutathione has a free SH-group associated with the amino acid cysteine that reacts with xenobiotics. GSH is also involved in the inactivation of reactive oxygen, the maintenance of membrane functions, and the redox state of the cell. Oxidation of GSH leads to glutathione disulfide (GSSG) (Figure 2.21).

In humans, at least 5 different classes of GSH *S*-transferases with overlapping substrate specificities for electrophilic substrates and high specificity for GSH have been detected. Four classes of the enzyme are located in the cytosol; in the liver of humans and rats, GSH *S*-transferases amount to about 3% and 10% of the soluble proteins, respectively.

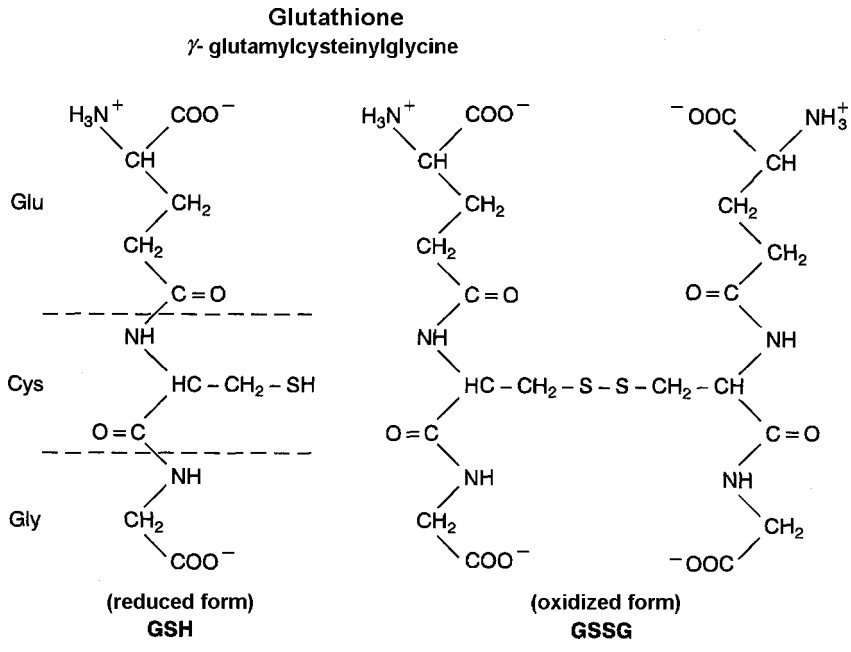


Figure 2.21 *Structure of glutathione (GSH) and glutathione disulfide (GSSG).*

The microsomal GSH *S*-transferase is located in the membranes of the endoplasmic reticulum and other cell organelles. Its substrate specificity differs from that of the cytosolic enzyme. Membrane-bound GSH *S*-transferase is inducible and plays an important role in the detoxification of reactive metabolites that are formed in the endoplasmic reticulum.

GSH *S*-transferases catalyse the formation of a thioether bond between an electrophilic C-atom of the substrate and the sulfhydryl group of GSH. Typical examples for such reactions with chemicals are shown in Figure 2.22. An example for the participation of GSH *S*-transferases in the metabolism of endogenous substances is the conjugation of the 5,6-epoxide of arachidonic acid in the synthesis of leukotrienes.

GSH-conjugates of xenobiotics are degraded to N-acetylated cysteine thioethers (mercapturic acids), which are subsequently excreted.

The reaction sequence for the formation of mercapturic acids and the enzymes that are involved are shown in Figure 2.23. The thioether bond of the GSH-conjugates is fairly stable. GSH-conjugation and subsequent transformation to the easily excretable mercapturic acid derivatives are thus regarded as detoxification reactions. Cysteine conjugates may, however, be split between the β-carbon atom of cysteine and the sulfur by cysteine

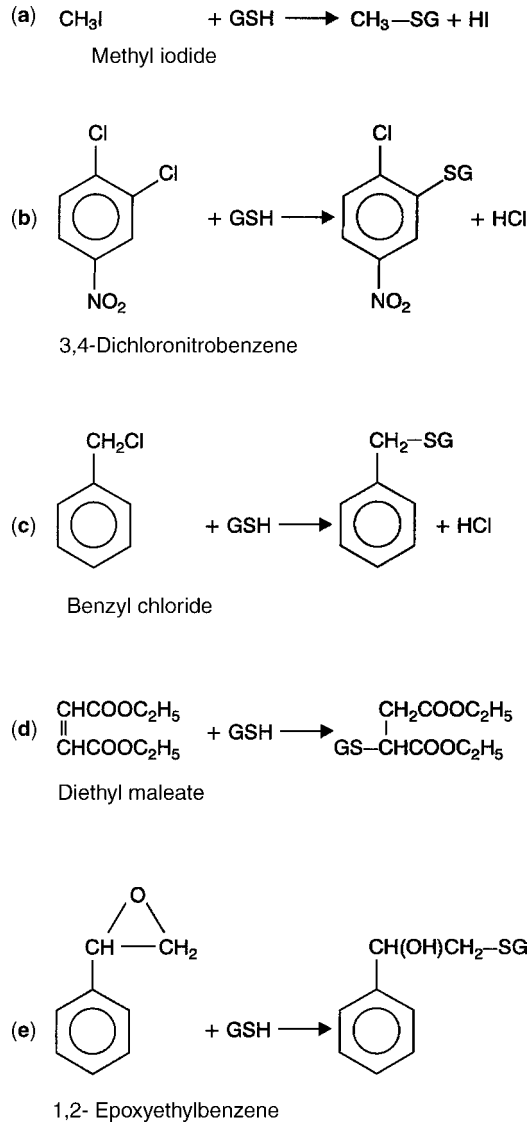


Figure 2.22 *Typical reactions catalysed by GSH S-transferases. (a) Nucleophilic substitution at a saturated C-atom. (b) Nucleophilic substitution at an aromatic C-atom. (c) Nucleophilic substitution at a saturated C-atom of a side chain of a ring. (d) Addition reaction at a C=C double bond (Michael addition). (e) Addition reaction with ring opening of an epoxide.*

β-lyase, giving rise to the formation of unstable, reactive thiol compounds, as well as pyruvate and ammonia [Equation (2.55)]:

$$\underset{\text{cysteine-conjugate } \beta\text{-lyase-mediated reaction}}{\text{CYS-S-R} + H_2O \rightarrow \text{R-SH} + CH_3COCOOH + NH_3} \qquad (2.55)$$

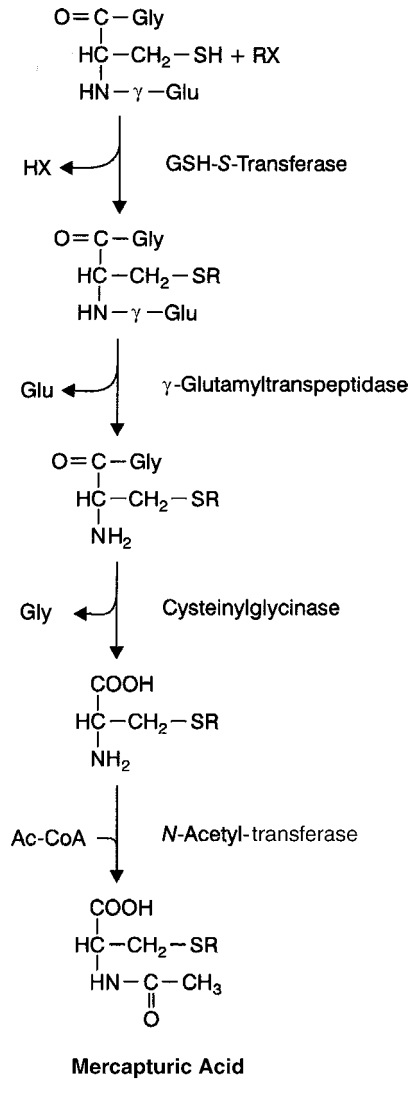


Figure 2.23 *Biosynthesis of mercapturic acids (R = electrophilic moiety).*

Cysteine-conjugate β-lyase occurs preferentially in kidney, intestinal microflora, and liver.

Some GSH-conjugates are nephrotoxicants, owing to their activation to toxic metabolites in the kidney by cysteine-conjugate β-lyase. For example, after initial transformation into mercapturic acids in the liver, polyhalogenated alkenes such as hexachlorobutadiene and trichloroethylene, upon arrival in the kidney, undergo deacetylation and further metabolism to electrophilic, DNA-reactive vinyl thiols by cysteine β-lyase (Figure 2.24; see also Chapter 6.5). This metabolic pathway is also responsible for the kidney tumors in animals dosed with tri- and perchloroethylene or

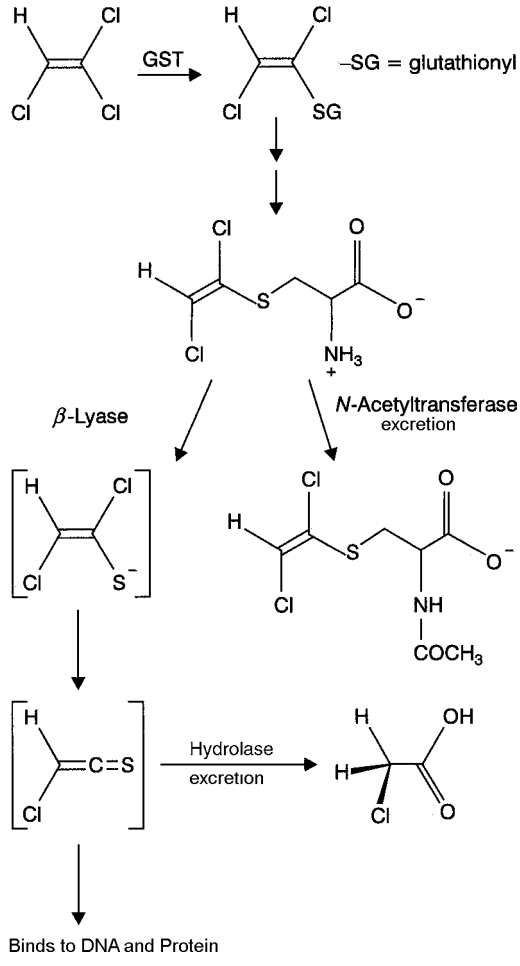


Figure 2.24 *Bioactivation of trichloroethylene by GSH conjugation, formation of a cysteine-conjugate, and its conversion by cysteine ß-lyase in the kidney. Reprinted from Koob and Dekant Chem.-Biol. Interact.,* **77***, 107–136. Copyright (1991), with permission from Elsevier.*

dichloroacetylene. Mercapturate formation may also explain the increased incidence of kidney tumors in workers after exposure to trichloroethylene.

There are also other examples for the toxification of xenobiotics by GSH conjugation. Vicinal dihalogen compounds, quinones, isothiocyanates, and α,β-unsaturated compounds may be activated by GSH and GSH *S*-transferase-catalysed metabolism. Thus, conjugation of dibromoethane and dichloroethane with GSH initially forms *S*-(2-halogenoethyl) glutathione and subsequently, after displacement of the second halogen, the mutagenic and carcinogenic episulfonium or thiiranium ion (Figure 2.25). Toxic quinones may be taken up into γ-glutamyltranspeptidase-rich tissues as hydroquinone-GSH-conjugates where they can be oxidized again to toxic quinones. Finally, owing to the reversibility of GSH-conjugation with toxic isothiocyanates and α,β-unsaturated

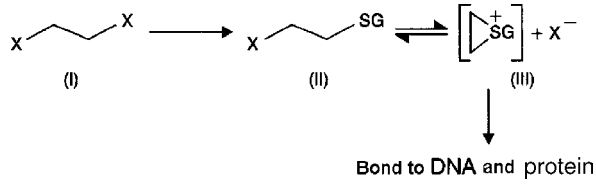


Figure 2.25 *Metabolism of 1,2-dihalogenoethane (I) to S-(2-halogenoethyl)GSH (II) and an episulfonium ion (III).*

compounds, the toxic parent compounds may be released, especially in tissues with low GSH-content.

UDP-Glucuronosyltransferases and Their Reactions

UDP-Glucuronosyltransferases catalyse the transfer of D-glucuronic acid from UDP-glucuronic acid to functional groups of aliphatic and aromatic alcohols, carboxylic acids, amines, hydroxylamines, amides, and thiols, giving rise to O-, N-, S- and C-glucuronides (Table 2.12). In this reaction a β-glycosidic bond between the aglycone (xenobiotic) and the glucuronic acid is formed. Humans and larger primates are the only species that conjugate tertiary amines to quaternary ammonium glucuronides.

The co-substrate for this reaction is UDP-glucuronic acid, which is synthesized from D-glucose-1-P in two steps (Figure 2.26). Cellular concentrations of UDP-glucuronic acid are relatively high, i.e. approximately 0.3–0.5 mmolar in liver. Since D-glucose-1-P is in equilibrium with glycogen and D-glucose-6-P, the co-substrate of glucuronidation is only exhausted under extreme conditions in the liver. Some inhalation anesthetics such as diethyl ether, halothane, and isoflurane may, however, significantly decrease the UDP-glucuronic acid level in the liver.

Glucuronidation is catalysed by a group of isoenzymes.

Similar to other xenobiotic-metabolizing enzyme systems, the UDP-glucuronosyltransferases comprise a group of isoenzymes with partly overlapping substrate specificities. Some isoenzymes are inducible by the same inducers as distinct CYP isoenzymes. The pretreatment of laboratory animals with 3-methylcholanthrene or β-naphthoflavone increases the activity of a UDP-glucuronosyltransferase, which catalyses the glucuronidation of phenolic substrates with planar structures such as 1-naphthol and 3-hydroxybenzo[*a*]pyrene. Conjugation of bulky molecules, such as morphine, 4-hydroxybiphenyl, and chloramphenicol, can be stimulated by pretreatment with phenobarbital. Enzyme activities of the first group are already detectable in the late fetal period, while enzymes of the second group are expressed only after birth. Accordingly the antibiotic chloramphenicol is toxic in newborn children and during the first months of life due to insufficient glucuronidation of the antibiotic.

Table 2.12 *Examples of the different classes of glucuronide conjugates. Data are taken from Casarett and Doull (1986).*

Type of Glucuronide[a]	Functional Group	Example
C-O-Glucuronide		
—C(—)—O—G	Alcohol Aliphatic Alicyclic Benzylic Phenolic	Trichloroethanol Hexobarbital Methylphenylcarbinol 3-Hydroxy benzo[*a*]pyrene
—C(=O)—O—G	Carboxylic acid Aliphatic Aromatic	α-Ethylhexanoic acid α-Aminobenzoic acid
—CH=C(—)—O—G	α-,β-Unsaturated ketone	Progesterone
N-O-Glucuronide		
—N(—)—O—G	*N*-Hydroxy group	*N*-Acetyl-*N*-phenyl hydroxamine
N-Glucuronide		
—O–C(=O)—N(H)—G	Carbamate	Meprobamate
—Ar—N(H)—G	Arylamine	2-Naphthylamine
—$(R)_3$—N^+—G	Aliphatic tertiary amine	Tripelannamine
—R—SO_2—N(H)—G	Sulfonamide	Sulfadimethoxine
S-Glucuronide	Arylthiol	Thiophenol
—Ar—S—G		
—C(=S)—S—G	Dithiocarbaminic acid	*N*,*N*-Diethyldithiocarbaminic acid
C-Glucuronide		
—C(—)—G	1,3-Dicarbonyl group	Phenylbutazone

[a]G-Glucuronide. (after *Casarett and Doull's Toxicology. The Basic Science of Poisons, 6th Edition,* 2001)

The cDNAs of several UDP-glucuronosyltransferases are known and similar to the CYP-dependent monooxygenases, permitting the naming and systematic organization of the individual isoenzymes on the basis of their derived amino acid sequence. The abbreviation UGT (UDP-glucuronosyltransferase) is followed by an Arabic number, which denotes the enzyme family. The subfamily is indicated by the addition of one letter, e.g. UGT2B. Members of a subfamily show a sequence homology of more than 60%. The last number denotes the individual gene or protein, e.g. UGT2B1.

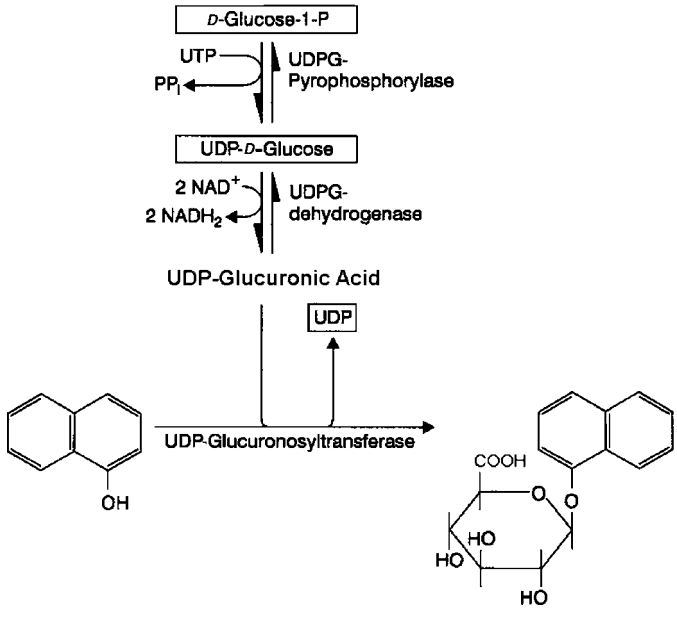


Figure 2.26 *The synthesis of UDP-glucuronic acid and the glucuronidation of 1-naphthol.*

Besides liver, which contains the greatest variety and highest concentrations of UDP-glucuronosyltransferases, intestine and kidney express significant enzyme activities. Lung, skin, spleen, brain, and other organs of the body also have demonstrable capacity to glucuronidate xenobiotics.

The splitting of glucuronides in the intestine by bacterial β-glucuronidase may markedly prolong the half-life of xenobiotics in the body.

Glucuronic acid has a carboxylic group, which is predominantly ionized at physiologic pH, thus increasing the water solubility of the aglycone. Excretion of the conjugates is mediated by anion-transport systems of the tubular cells of the kidney and hepatocytes. Humans preferentially eliminate molecules with molecular weights lower than about 500 Da via the kidneys. In contrast, conjugates exceeding this molecular mass are excreted preferentially into the bile. In the intestine the aglycone can be set free by bacterial β-glucuronidases, which hydrolyse the β-glycosidic bond. Following their transport through the epithelium of the intestine, xenobiotics are again transported to the liver via portal blood. This enterohepatic circulation can take place repeatedly and may cause a considerable delay of the elimination of endogenous and exogenous compounds.

In specific cases, glucuronides act as transport metabolites of toxic compounds.

For example, the bladder carcinogen 2-naphthylamine is N-hydroxylated and N-glucuronidated in the liver and subsequently the conjugate is transported to the bladder via blood

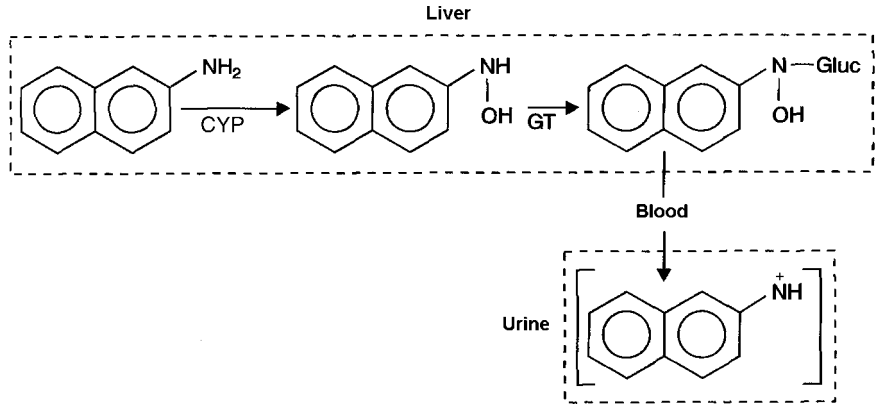


Figure 2.27 *The role of glucuronidation in the activation of 2-naphthylamine. (CYP = cytochrome P450; GT = UDP-glucuronosyltransferase; Gluc = glucuronic acid.)*

and kidney. The glucuronide is unstable in urine owing to the low pH-value and hydrolyses to the hydroxylamine, which displaces water and forms a reactive nitrenium ion (Figure 2.27).

In contrast, glucuronidation contributes to the detoxification of reactive metabolites, interrupting the oxidation/reduction cycles that occur during the metabolism of quinones to hydroquinones. Quinones can be reduced to semiquinones and hydroquinones, which may convert back to quinones, transferring electrons to oxygen. In this process toxic oxygen species and semiquinone radicals are formed. This cycle can be interrupted by the glucuronidation of the hydroquinones (Figure 2.28).

Sulfotransferases and Their Reactions

Sulfotransferases and UDP-glucuronosyltransferases frequently compete for the same substrates. Sulfoconjugates may be formed with alcohols, phenols, hydroxylamines, and arylamines (Table 2.13). The sulfonates produced exist preferentially in the ionized form at physiological pH, which markedly increases the water solubility of the xenobiotic.

The co-substrate of the reaction is 3′-phosphoadenosine-5′-phosphosulfate (PAPS). It is formed from inorganic sulfate in two ATP-dependent enzyme-catalysed reactions (Figure 2.29). The inorganic sulfate originates from food or the oxidative metabolism of the sulfur-containing amino acid cysteine. High doses of substrates of the sulfotransferases may consume so much of the intracellular sulfate supply that the synthesis of PAPS and, consequently sulfation, decreases. The concentration of PAPS amounts to only about 50 μmolar in the liver of rats; PAPS can be rapidly synthesized at rates up to $100\,\text{nmol} \times \text{min}^{-1} \times \text{g}^{-1}$ liver, which readily restores normal concentrations. Substrates of sulfonation, such as phenols and alcohols, are often substrates of UDP-glucuronosyltransferases. At lower concentrations phenols are preferentially sulfated, while at higher concentrations conjugation with glucuronic acid prevails.

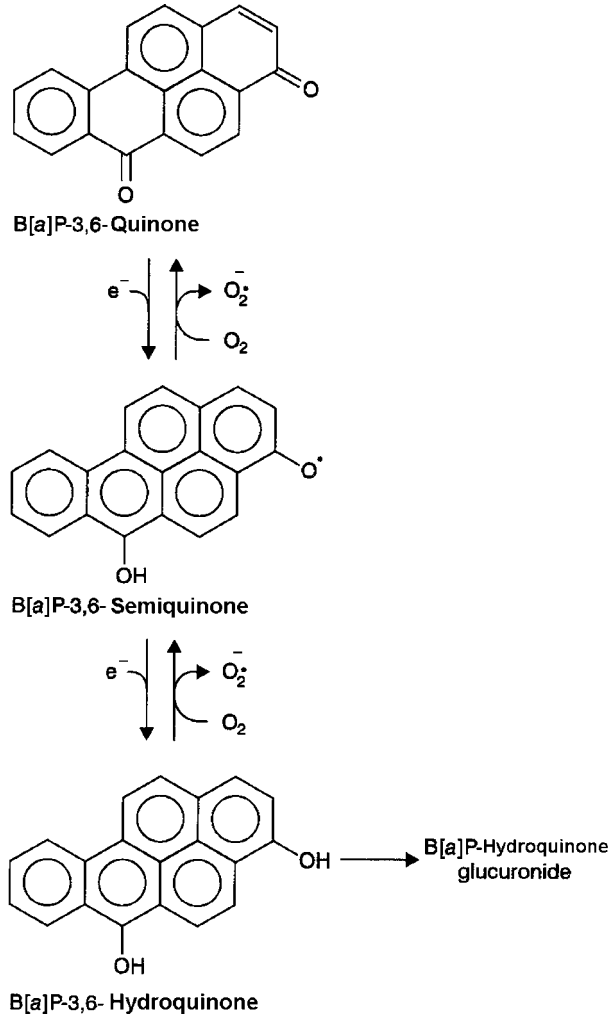
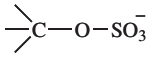
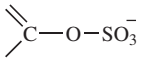
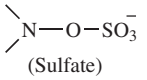
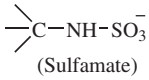


Figure 2.28 *Formation of the reactive superoxide radical during the redox-cycling of quinones. This toxic metabolic pathway is interrupted by the glucuronidation of the hydroquinone. (B[a]P = Benzo[a]pyrene.)*

Sulfoconjugates are frequently eliminated via urine since the $SO_4^{=}$ increases the molecular mass of the conjugate only slightly. Thus, sulfoconjugates often do not reach a molecular mass sufficient for excretion into bile. Transport proteins excrete sulfoconjugates from liver and kidney cells into blood and urine, respectively.

Sulfotransferases are cytosolic enzymes, whose concentration cannot be increased by known inducers of xenobiotic enzymes.

Table 2.13 *Examples of different classes of Sulfoconjugates.*

Type of sulfoconjugate	Functional group	Example
	Alcohol	Hydroxydimetridazol
	Phenol	Phenol
 (Sulfate)	Hydroxylamine	*N*-Hydroxy-2-acetylaminofluorene
 (Sulfamate)	Amine	Aniline, 2-Naphthylamine

The formation of sulfoconjugates is catalysed by a group of enzymes termed sulfotransferases. Several isoenzymes have been detected with partly overlapping substrate specificity that catalyse the sulfonation of alcohols, phenols, benzylic alcohols, hydroxysteroids, amines, and hydroxylamines; the products of the latter two reactions are described as sulfates ($R\text{-}NH\text{-}OSO_3^-$) or sulfamates ($R\text{-}NH\text{-}SO_3^-$). Sulfoconjugates are not only formed in the liver but also in many other tissues of the body, e.g. intestine, kidney, lung, and platelets. In this context sulfation by cells of the intestinal mucosa is of great importance. Thus, orally ingested phenols are already sulfated in the intestine to a large extent.

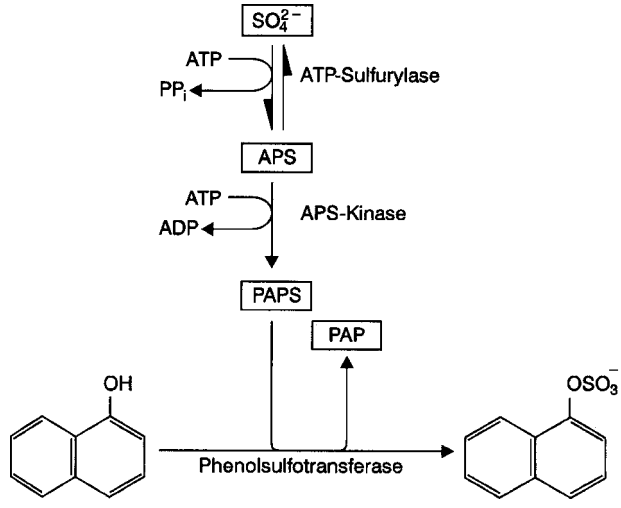


Figure 2.29 *The synthesis of PAPS (3′-phosphoadenosine-5′-phosphosulfate) and sulfonation of 1-naphthol. (APS = adenosine-5′-phosphosulfate; PAPS = 3′-phosphoadenosine-5′-phosphosulfate; PAP = 3′-phosphoadenosine-5′-phosphate.)*

Sulfonation may also give rise to toxic reactive intermediates.

Sulfoconjugates are usually less stable than glucuronides of the same substrate, because the sulfate is a better leaving group than is glucuronic acid, which forms a β-glycosidic bond. This is especially true for sulfates of hydroxylamines, which have a half-life of less than 1 minute. During the bioactivation of these substances a reactive nitrenium or carbonium ion is formed (Figure 2.30). The electrophilic molecule binds covalently to proteins and DNA.

Acetyltransferases and Their Reactions

Substrates of *N*-acetyltransferases are aromatic amines, some aliphatic primary amines and hydroxylamines, hydrazines, hydrazides, and sulfonamides.

Acetylation of amino and hydroxylamino groups of xenobiotics represents an important type of conjugation reaction. The masking of functional groups by acetylation generally results in decreased water solubility of xenobiotics. Acetyl-coenzyme A (Acetyl-CoA), which is formed in intermediary metabolism by the acetylation of the sulfhydryl group of CoA, is the activated co-substrate for acetylation (Figure 2.31). Substrates of this reaction are aromatic amines, some aliphatic primary amines, hydrazines, hydrazides, and sulfonamides as well as the N-hydroxyl derivatives of arylamines, the arylhydroxylamines. Acetylation of arylhydroxylamines may take

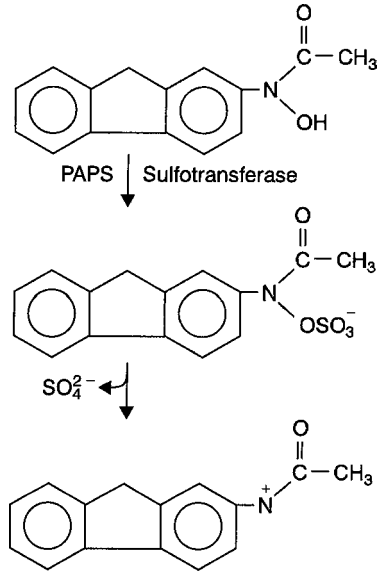


Figure 2.30 *Activation of 2-(N-hydroxy-N-acetylamino)fluorene by sulfonation. The sulfoconjugate hydrolyses spontaneously. Reprinted from Mulder. Copyright (1981) with permission from CRC Press.*

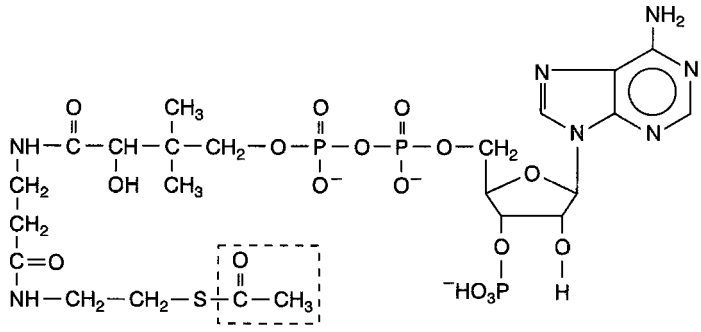


Figure 2.31 *Acetyl-Coenzyme A (Acetyl-CoA).*

place at the nitrogen or at the hydroxy group, producing arylhydroxamic acids and acetoxyarylamines, respectively. An intramolecular transfer of the acetyl group of a hydroxamic acid to its hydroxy group may also form the acetoxy compound.

Acetoxyarylamines are of particular toxicological significance, since they are unstable and decay to a reactive arylnitrenium ion. These reactions are catalysed by *N*-acetyltransferases which are cytoplasmic and mitochondrial enzymes, found in many tissues.

N- and O-acetylation of arylamines and arylhydroxylamines takes place in two steps. First the acetyl group of the co-substrate, acetyl-CoA, is transferred to the enzyme and then in a second reaction to the amino, or hydroxylamino, group of the substrate. In liver, intestine, and the epithelium of the bladder, both monomorphic and polymorphic *N*-acetyltransferase, which differ in their substrate specificities, are expressed. While the monomorphic enzyme is similarly distributed throughout the human population, the second enzyme exhibits a polymorphism, i.e. different humans will acetylate specific substrates, such as isoniazid, either slowly or rapidly. Depending on the substrate, the monomorphic and polymorphic forms of the enzyme catalyse not only the N-acetylation of arylamines but also the O-acetylation of arylhydroxylamines

The concentration of the acetylated metabolites of an amine depends both on the velocity of the acetylation of the xenobiotic and the extent of its deacetylation. In this context the dog and man represent extremes. While dogs are poor acetylators of aromatic amines, but good deacetylators, man shows a high activity of *N*-acetyltransferase towards arylamines but only a poor capacity to deacetylate. Deacetylases and amidases are located in the endoplasmic reticulum, in mitochondria, and the soluble fraction of cells.

Acetylation and deacetylation play an important role in the activation of arylamines.

It has been known since the end of the 19th century that aromatic amines may cause cancer. After N-hydroxylation, very different metabolic pathways may lead to the activation of these substances. In this context acetylation and deacetylation represent fundamental reactions, but also sulfation, peroxidation, and glucuronidation can contribute to the activation of aromatic amines. The significance of these metabolic

pathways for the development of arylamine-induced tumors depends on the species and the target tissue, and both may differ markedly in their enzymatic make-up. The role of acetylation and deacetylation in the activation of arylamines is shown in Figure 2.32. Oxidation at the nitrogen is essential to the activation of these substances, thereby

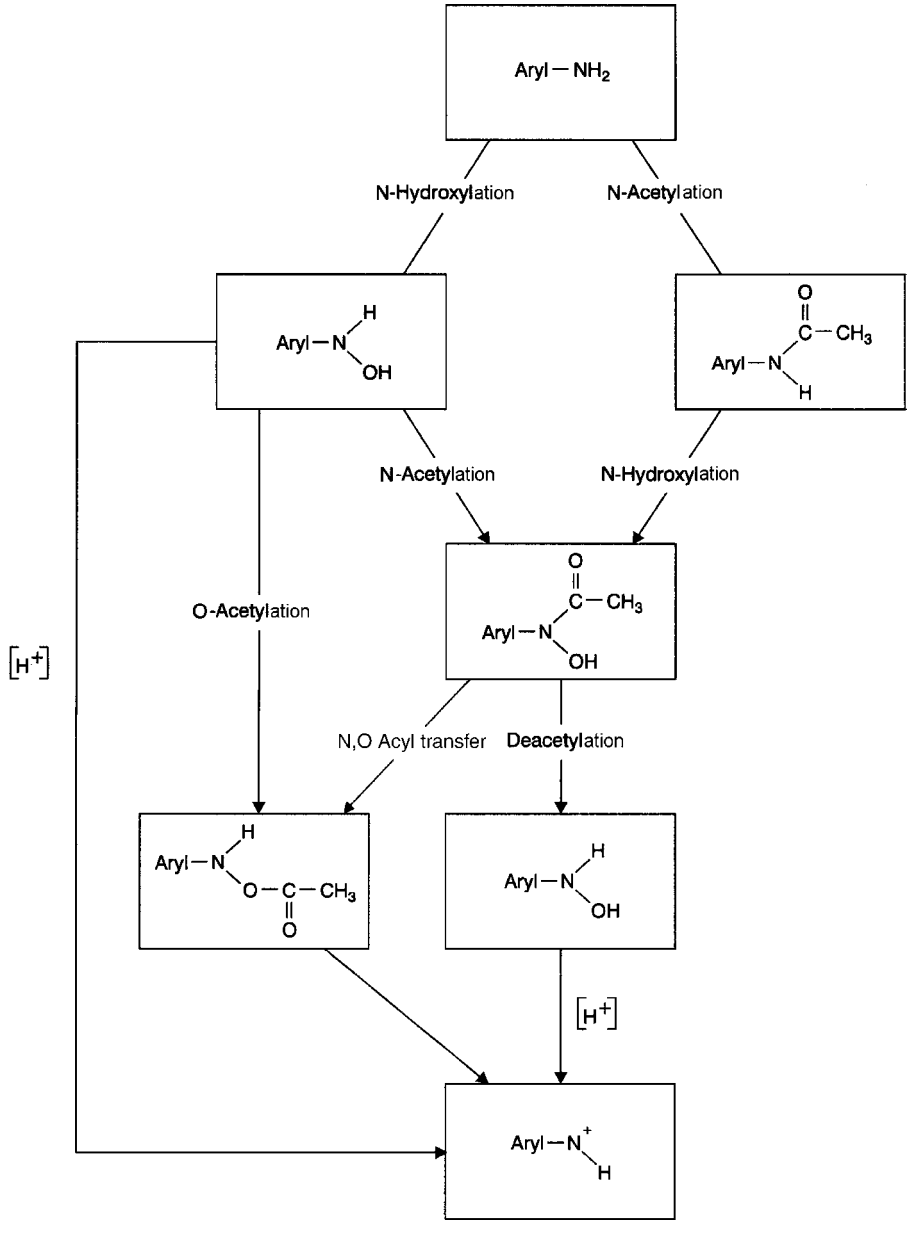


Figure 2.32 *Metabolic activation of aromatic amines by acetylation, N-hydroxylation, and deacetylation. The Scheme does not show all possible reactions.*

forming an arylhydroxylamine that can undergo N-acetylation forming an arylhydroxamic acid. Ultimate carcinogens may then result from the deacetylation to yield arylhydroxylamines, or from the O-acetylation or intramolecular acyltransfer to yield acetoxyarylamines). From these unstable products an electrophilic nitrenium ion, which can bind to the DNA, is produced nonenzymatically.

The risk of inducing cancer by arylamines probably depends on the expression of the polymorphic *N*-acetyltransferase. Epidemiologic studies have noted that slow acetylators, as opposed to rapid acetylators, have a higher risk to develop bladder tumors when exposed to benzidine. The polymorphic distribution of *N*-acetyltransferase in the population may also have consequences for the activation of distinct carcinogenic heterocyclic amines that result from the cooking of meat (Figure 2.33). The polymorphic acetyltransferase N-acetylates these substances poorly, but activates the N-hydroxyl derivatives of the heterocyclic amines very efficiently by O-acetylation.

Methyl Transferases and Their Reactions

Methylation masks functional groups of various chemicals thereby increasing the lipophilicity of these substances.

Functional groups that may be methylated include alcoholic or phenolic hydroxy groups, e.g. the hydroxy group of pyrogallol; primary, secondary, and tertiary amino

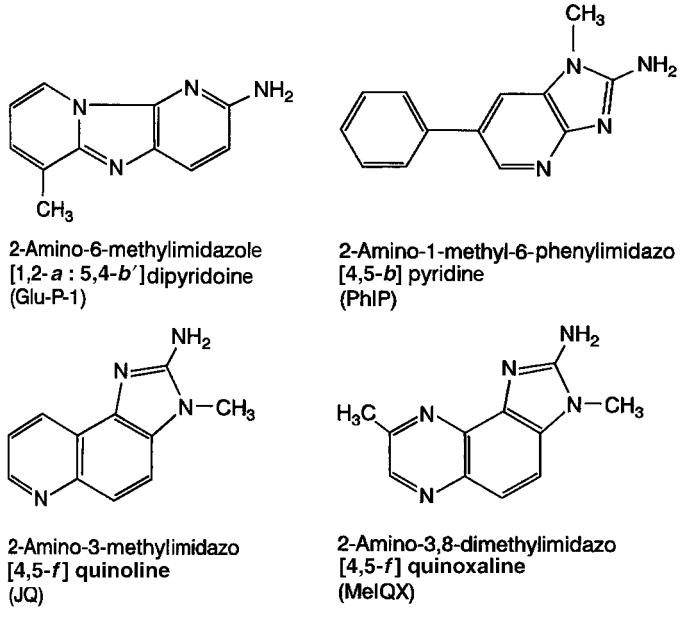


Figure 2.33 *Examples of heterocyclic carcinogenic amines, which are formed during cooking (pyrolysis of proteins).*

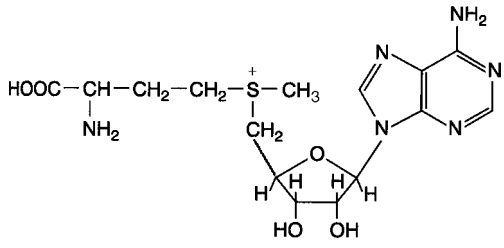


Figure 2.34 *S-Adenosylmethionine.*

groups such as the N-atom of pyridine; and sulfhydryl groups of endogenous and foreign molecules, such as the SH-group of methanethiol. In quantitative terms, however, methylation is only of secondary importance in the metabolism of xenobiotics. The 'activated' co-substrate *S*-adenosylmethionine (Figure 2.34) donates the methyl group, generating *S*-adenosylhomocysteine. The reaction is catalysed by several substrate-specific and -nonspecific *O*-methyl-, *N*-methyl-, and *S*-methyltransferases. The enzymes are expressed in the endoplasmic reticulum or cytosol of several cells of the body, including liver, intestine, lung, and kidney.

Methylation is involved in the detoxification of toxic H_2S, which is formed by anaerobic bacteria. In the first reaction, the produced methanethiol is one tenth as toxic as the parent compound. Dimethyl sulfide is then formed in a second methylation reaction. Additional substrates of the S-methyltransferases can arise during the degradation of glutathione conjugates via the β-lyase pathway (see the latter half of the section **glutathione S-Transferases and Their Reactions** above).

Conjugation with Amino Acids

Several carboxylic acids are metabolized to amides by conjugation with amino acids.

Xenobiotics with a carboxylic group may form conjugates with the amino acids glycine and glutamine in humans. Substrates of this reaction include aromatic carboxylic acids, aromatic substituted acetic acids, β-arylpropionic acids, and β-substituted arylic acids.

First the carboxylic acid is activated to the adenylate in an ATP-consuming reaction and then to a thioester derivative of coenzyme A; next the acyl moiety is transferred to the amino acid, creating an amide bond (Figure 2.35). The reactions described are catalysed by ATP-dependent acid:CoA ligases and *N*-acetyltransferases. These enzymes are located in the soluble cell fraction and the mitochondria. Occasionally substances may form conjugates with several amino acids, such as a metabolite of DDT with asparagine and serine.

Besides being substrates for amino acid conjugation, carboxylic acids may also undergo glucuronidation. Which of these two metabolic pathways is taken depends both on the substance and the species under investigation.

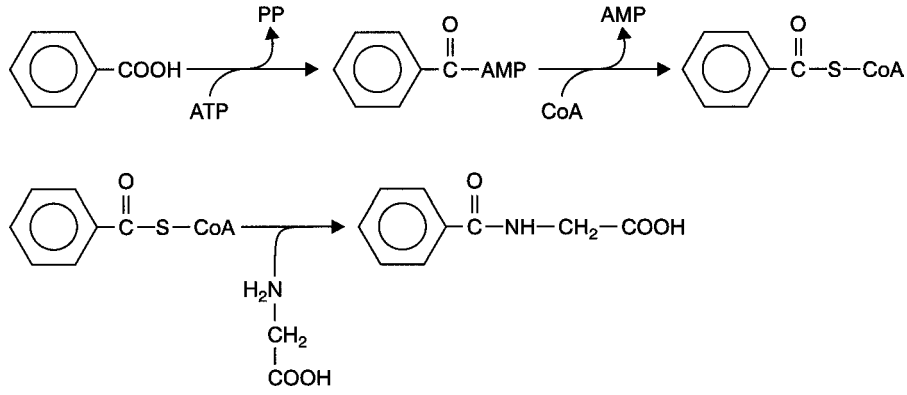


Figure 2.35 *Conjugation of benzoic acid with the amino acid glycine.*

2.2B.2 Hydrolysis

Hydrolytic reactions take place not only in the cells of the body but also in blood (unspecific esterases) and the microflora of the intestine. Of great importance are epoxide hydrolases metabolizing reactive epoxides to dihydrodiols, and esterases/amidases converting esters and amides into acids and alcohols or amines. Sulfatases and glucuronidases are hydrolytic enzymes that split conjugates arising in the biotransformation of xenobiotics.

Epoxide Hydrolases and Their Reactions

In the course of the oxidative metabolism of olefins and aromatic compounds, the reactive epoxides that are formed are substrates of epoxide hydrolase.

Electrophilic epoxides (oxiranes) may rearrange to phenols spontaneously, form GSH-conjugates, or undergo hydrolysis resulting in vicinal dihydrodiols. The latter reaction is catalysed by epoxide hydrolase. The enzyme is regioselective, i.e. the hydroxide ion is taken up at that carbon of the oxirane which has the least steric hindrance. The enzyme catalysis follows most likely a base-catalysed nucleophilic reaction mechanism. The dihydrodiols formed show a *trans*-configuration and, as exemplified with stilbene oxide (Figure 2.36), *cis*- and *trans*-epoxides are converted into the corresponding *threo*- and *erythro*-diols.

Xenobiotic-metabolizing epoxide hydrolases are located in the membrane of the endoplasmic reticulum as well as in the cytosol, primarily in peroxisomes. The microsomal enzyme plays an important role in xenobiotic metabolism. Owing to its spatial proximity to the microsomal cytochrome P450-dependent monooxygenases, the enzyme contributes to an efficient detoxification of epoxides close to their site of generation. Both the cytosolic and the microsomal isoenzymes are present in very high concentrations in liver, and demonstrable amounts are expressed in organs such as kidney, lung, intestine, brain, spleen, testicles, and ovaries. Marked species differences can be demonstrated by comparing the hydrolysis of *trans*-ethylstyrenel oxide and *cis*-stilbene oxide, model substrates of the cytosolic and microsomal epoxide hydrolases,

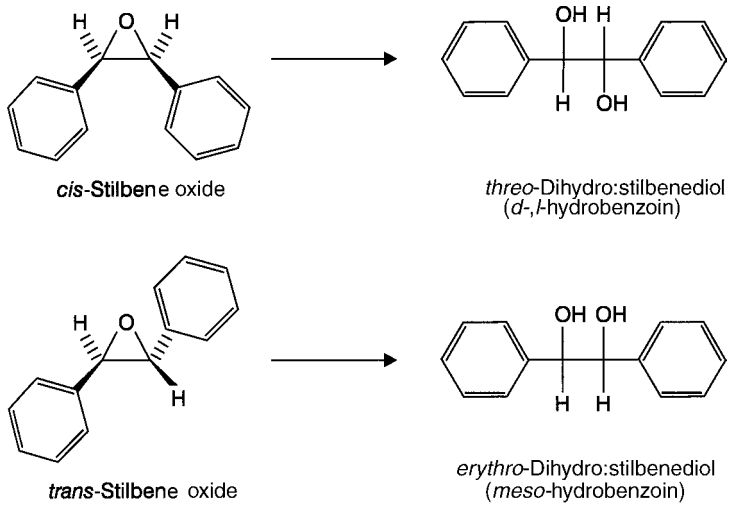


Figure 2.36 *Hydrolysis of cis- and trans-stilbene oxide by epoxide hydrolase.*

respectively. The activity of the cytosolic epoxide hydrolase is significantly lower in humans and rats when compared with mice, whereas activity of the microsomal enzyme is higher in humans than in mice and rats.

Enzyme inducers of xenobiotic metabolism increase the content of both enzymes (Table 2.14).

Generally, the hydrolysis of epoxides represents a detoxification reaction in xenobiotic metabolism. In the case of many epoxides of polycyclic aromatic hydrocarbons, however, the enzymatic reaction enhances their toxicity.

The role of epoxide hydrolase in the toxification of polycyclic **aromatic hydrocarbons** will be demonstrated using benzo[*a*]pyrene as an example. In the metabolism of benzo[*a*]pyrene-7,8-diol 9,10-epoxide, the 7,8-epoxide of benzo[*a*]pyrene is formed first. This epoxide is a good substrate for the microsomal epoxide hydrolase and is readily transformed into the dihydrodiol. This dihydrodiol still shows a considerable lipophilicity and, therefore, may be a substrate for further monooxygenation by CYP to benzo[*a*]pyrene-7,8-diol 9,10-epoxide. Diol epoxides are poor substrates for the microsomal epoxide hydrolase. Accordingly, the ultimate carcinogenic benzo[*a*]pyrene metabolite is not detoxified significantly by epoxide hydrolase.

Carboxylesterases and Amidases and Their Reactions

The body has numerous unspecific esterases and amidases, which hydrolyse esters and amides.

Table 2.14 *Enzyme inducers and inhibitors of microsomal and cytosolic epoxide hydrolase.*

	Epoxide hydrolase	
	Microsomal	Cytosolic
Enzyme inducers		
trans-Stilbene oxide	+++	−
2-(Acetylamino)fluorene	+++	−
Butylated hydroxyanisole	+++	−
Phenobarbital	++	−
Methylcholanthrene	+	
Clofibrate	−	+
Di-(2-ethylhexyl) phthalate	−	+
Inhibitors		
3,3,3-Trichloropropene oxide	+	−
Cyclohexene oxide	+	−
4-Phenylchalcone	−	+

Hydrolysis of esters, thioesters, and amides leads to a carboxylic acid and depending on the substrate, to an alcohol, thioalcohol, or amine, respectively (Figure 2.37).

The enzymes involved in this reaction show both esterase and amidase activity. Because of the lack of specific criteria the individual esterases and amidases are categorized into A-, B- and C-esterases according to their reactivity with phosphoric acid esters: A-esterases hydrolyse organic phosphoric acid esters, B-esterases will be inhibited by them, and C-esterases do not interact with them. As shown in Table 2.15 the enzymes of the distinct groups have a preference for certain substrates. However, the substrate specificities of some of the enzymes overlap considerably. Esterases/amidases are located in the cytosol, endoplasmic reticulum, and mitochondria of

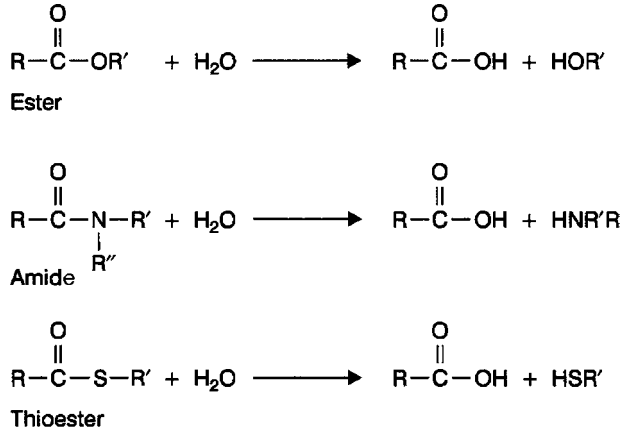


Figure 2.37 *Hydrolytic splitting of xenobiotics by esterases and amidases.*

Table 2.15 *Classification of esterases according to their inhibition by phosphoric esters and carbamates.*

Esterase	Interaction with organic phosphates	Preferred substrate
A-Esterases	Substrate	Aryl ester
B-Esterases	Inhibition	
Carboxylesterases /Amidases		Aliphatic and aromatic ester/amide
Acetylcholinesterase		Acetylcholine
Pseudocholinesterase		Carboxylic acid ester or amide, ester, quaternary ammonium ions
C - Esterases	No interaction	Acetyl ester

many cells; they may be also found in blood. Plasma pseudocholinesterases catalyse the hydrolysis of numerous carboxylic esters and amines. A peculiarity of the pseudocholinesterases is that they show a polymorphic distribution in the population.

Unlike pseudocholinesterase and many other esterases/amidases, acetylcholinesterase has a relatively high substrate specificity. Its most important physiologic task involves the inactivation of the neurotransmitter acetylcholine. Poisoning with organic phosphoric acid esters, which inhibit the B-esterase enzyme acetylcholinesterase, is life threatening. Inhibition of the acetylcholinesterase is caused by the phosphorylation of the hydroxy group of a serine moiety in the active center of the enzyme. Carbamates bind to the enzyme via a similar mechanism; however, the bond is weak and the reaction is readily reversed.

Glucuronidases, Glycosidases, Sulfatases and Their Reactions

Hydrolytic enzymes may split glucuronide and sulfoconjugates of xenobiotics.

Glucuronidases and sulfatases, which catalyse the hydrolysis of conjugates of xenobiotics, are present in many cells of the body, as well as the intestinal microflora. The β-glucuronidase located in the lysosomes of mammalian cells splits glucuronide conjugates into glucuronic acid and the aglycone. This reaction is more rapid with O- and S-glucuronides as substrates than with N-glucuronides. The various sulfatases known today are located partly in the endoplasmic reticulum and partly in the lysosomes.

The toxicokinetic behavior of xenobiotics may be influenced by their hydrolysis in the intestine. Thus, the splitting of glucuronides can markedly increase the half-life of xenobiotics in the body owing to the absorption of the aglycone.

The intestinal microflora express glycosidases that hydrolyse the glycosidic bonds of sugar molecules. Bacterial glycosidases play an essential role in the activation of the carcinogenic glycoside cycasin (Figure 2.38). The methylazoxymethanol

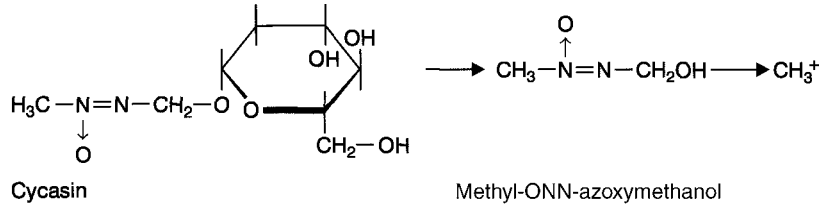


Figure 2.38 *Metabolism of cycasin by microbial glycosidases.*

that is formed by hydrolysis is further transformed to an electrophilic methylating species.

2.2B.3 Summary

Lipophilic compounds are excreted only very slowly from the body. An important function of xenobiotic metabolism is, therefore, to transform lipophilic compounds into hydrophilic molecules that are more easily secreted.

Xenobiotic metabolism takes place particularly in liver, as well as in intestine, kidney, and lung, and to a lesser extent in other organs of the body. Furthermore, the intestinal microflora contribute to biotransformation primarily by hydrolysing and reducing xenobiotics. It is remarkable that organisms are capable of metabolizing and eliminating xenobiotics displaying widely different structures. One reason for this is the broad substrate specificity of many of the enzymes and the expression of isoenzymes exhibiting different substrate specificities.

Xenobiotic metabolism is often subdivided into two phases. In the reactions of phase I, i.e. oxidation, reduction, and hydrolysis, functional groups are inserted or uncovered in the parent compound. In the reactions of phase II, functional groups are joined (conjugated) to molecules formed in intermediary metabolism such as glucuronic acid, inorganic sulfate, carboxylic acids, methyl groups, amino acids, and glutathione. The products of the phase II reactions are more water soluble as a rule. Exceptions include the products of acetylation and methylation.

Xenobiotic metabolism does not necessarily lead to biologically inactive compounds. Several drugs and xenobiotics are transformed into the ultimate active substance or to toxic metabolites. These toxic chemicals may react directly with cellular macromolecules such as DNA, proteins, and lipids and cause toxicity. Indirect effects include the generation of reactive oxygen species during the redox cycling of radical-generating metabolites, the depletion of GSH, and the disturbance of the redox status of GSH and NAD(P)H.

The enzymes of xenobiotic metabolism are distributed in different cellular compartments. CYP-dependent monooxygenases as well as glucuronosyltransferases and one form of epoxide hydrolase are located in the membranes of the endoplasmic reticulum and nuclear envelope. Consequently, hydroxylation products may conjugate with glucuronic acid or form vicinal dihydrodiols in the case of epoxides, close by their place of generation. Other enzymes of xenobiotic metabolism that are found in the cytoplasm

include sulfotransferases and nearly all GSH *S*-transferases, generating sulfoconjugates and glutathione-conjugates, respectively.

Enzymes of xenobiotic metabolism such as the CYP-dependent monooxygenases and UDP-glucuronosyltransferases can be induced by certain xenobiotics including substances contained in plants. Inducing compounds are often substrates that can cause an increased synthesis of, and, in some cases, reduced degradation of, these enzymes. Owing to the increase of the enzyme the elimination rate of a xenobiotic is enhanced when similar concentrations at the enzyme are considered.

Xenobiotic-metabolizing enzymes differ in their occurrence, specificity, and activity in different species. The expression of the enzymes may also differ significantly in the individuals of a species. Thus, the activity of distinct enzymes may differ by a factor of 100 in humans. In contrast, the activity of inbred strains of laboratory animals, e.g. mice and rats, shows only minor differences. Knowledge of differences in xenobiotic metabolism between experimental animals and man helps in the extrapolation of animal data to humans and thus in the assessment of the health risk due to exposure to xenobiotics.

References

Koob, M and Dekant, W. Bioactivation of Xenobiotics by Formation of Toxic Glutathione Conjugation. *Chem. -Biol. Interact*., **77**, 107–136 (1991).

Casarett and Doull's Toxicology. The Basic Science of Poisons, ed. Klaassen, CD, Amdur, MO, Doull, J, 3rd edition (1986), Macmillan Publishing Company, New York.

Casarett and Doull's Toxicology. The Basic Science of Poisons. ed. Klaassen, CD, 6th edition (2001), MacGraw-Hill, New York.

Mulder, GJ (ed.), Sulfation of Drugs and Related compounds (1981), CRC Press, Boca Raton.

Further Reading

Denisov, IG, Makris, TM, Sligar, SG, Schlichting, I. Structure and Chemistry of Cytochrome P450. *Chem. Rev*., **105**; 2253–2277 (2005).

Dupret, JM, Rodrigues-Lima, F. Structure and Regulation of the Drug-metabolizing Enzymes Arylamine *N*-acetyltransferases. *Curr. Med. Chem*., **12**; 311–318 (2005).

Dorne, JL, Walton, K, Renwick, AG. Human Variability in Xenobiotic Metabolism and Pathway-related Uncertainty Factors for Chemical Risk Assessment: a Review. *Food Chem. Toxicol*., **43**; 203–216 (2005).

Hayes, JD, Flanagan, JU, Jowsey, IR. Glutathione Transferases. *Annu. Rev. Pharmacol. Toxicol*., **45**; 51–88 (2005).

Kauffman, FC. Sulfonation in Pharmacology and Toxicology. *Drug. Metab. Rev*., **36**; 823–843 (2004).

Kiang, TK, Ensom, MH, Chang, TK. UDP-glucuronosyltransferases and Clinical Drug-Drug Interactions. *Pharmacol. Ther*., **106**; 97–132 (2005).

Tukey, RH, Strassburg, CP. Human UDP-glucuronosyltransferases: Metabolism, Expression, and Disease. *Annu. Rev. Pharmacol. Toxicol.*, **40**; 581–616 (2000).

Zhou, J, Zhang, J, Xie, W. Xenobiotic Nuclear Receptor-mediated Regulation of UDP-glucuronosyltransferases. *Curr. Drug Metab*., **6**; 289–298 (2005).

2.3 Toxicogenetics

Lesley Stanley

2.3.1 Introduction

Toxicogenetic variation is determined by the genes, each of which is found at a specific chromosomal locus. A polymorphism is defined as a sequence variant at a particular locus which is present in at least 1% of the population. An understanding of genetic polymorphisms facilitates the evaluation of variability in toxic responses and may make it possible to identify and protect susceptible subpopulations.

Inter-individual variability in responsiveness to toxic agents can occur as a result of both biochemical and genetic variation. Toxicogenetics addresses genetic variability by considering the role of an individual's genetic makeup in determining the effects of xenobiotics to which they may be exposed.

Toxicogenetic variation is determined by the genes which encode proteins responsible for the metabolic processing of, and response to, xenobiotics. Each gene is found at a specific *locus*. At each locus there may be one or more possible sequence variants, or *alleles*. The existence of more than one possible allele at a particular locus is called a *polymorphism*, and loci where multiple alleles exist are described as polymorphic. A polymorphism is defined as a sequence variant, which is present in at least 1% of the population. Humans, being diploid, have two copies of each locus, i.e. they carry two alleles of each gene. If both copies are the same, the individual is said to be *homozygous* at this locus; if two different alleles are present, the individual is said to be *heterozygous*.

The consequences of polymorphic variation at a hypothetical allele are illustrated in Figure 2.39. In the case of a detoxifying enzyme, which has two polymorphic variants, one which has high activity and one which has low activity, a homozygous individual who has two high activity alleles will be protected against toxicity whereas a homozygous individual who has two low-activity alleles might be susceptible to toxicity. In a heterozygous individual, roughly half the enzyme molecules will have high activity and half will have low activity; one would predict that this individual would have intermediate susceptibility. In the case of a polymorphic gene that is responsible for metabolic activation of a toxic chemical, the predicted pattern of susceptibility would be the reverse of that for a detoxifying enzyme: a homozygous individual who has two high-activity alleles will be susceptible to toxicity whereas a homozygous individual who has two low activity alleles might be resistant to toxicity.

An understanding of genetic polymorphisms affecting either the toxicokinetics or toxicodynamics of toxic agents facilitates the evaluation of variability in response within the population and may make it possible to identify and protect particularly susceptible subpopulations. This chapter will introduce the technologies used in toxicogenetics and exemplify the impact of toxicogenetics on susceptibility to chronic and acute toxic effects. The current status of the field will be evaluated and possible future developments will be identified.

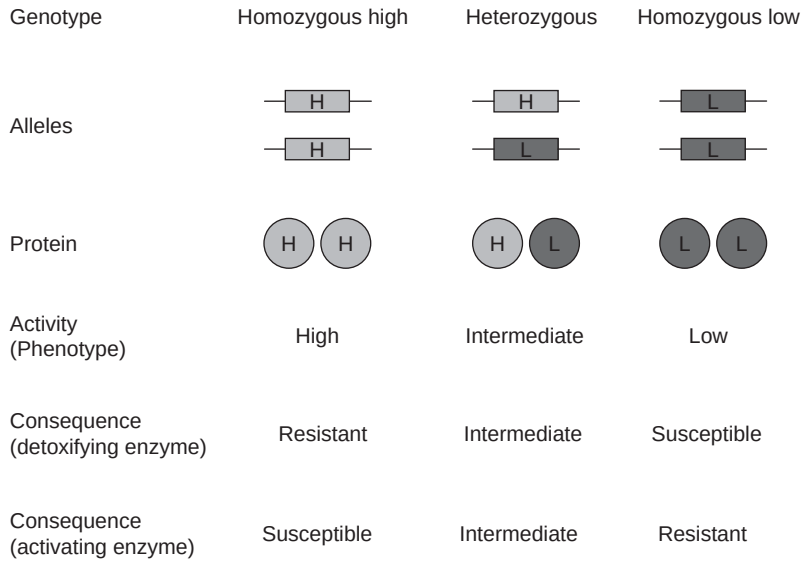


Figure 2.39 *Genotype and phenotype and their consequences*
The possible consequences of polymorphism in a hypothetical xenobiotic-metabolizing enzyme that has both a high-activity variant and a low-activity variant are shown. In the case of an individual who is homozygous for the high-activity variant, two high-activity alleles will be present (genotype) and these will be transcribed and translated to form entirely high-activity enzyme molecules. Thus the individual's liver will have high enzyme activity (phenotype). If the enzyme in question is a detoxifying enzyme, the individual is predicted to have reduced susceptibility to toxicity, whereas if it is a metabolic activating enzyme the individual may have increased susceptibility. The reverse is the case for an individual who is homozygous for the low-activity variant, while a heterozygous individual has one high- and one low-activity allele, a mixture of high- and low-activity enzyme molecules in the liver, and hence an intermediate phenotype.

2.3.2 Genotyping and Phenotyping

Genetic characteristics, i.e. the allelic composition, define the genotype. Phenotype refers to the biological expression of the genotype. Genotyping involves examining the altered nucleotide sequences of polymorphic variants in the DNA itself, whereas phenotyping involves measuring biological functions, e.g. enzyme activity.

Many polymorphisms affect the amino acid sequence of a protein, but some affect the noncoding regulatory regions of genes. Polymorphisms which affect the amino acid sequence may be detected either at the level of nucleotide sequence or amino acid sequence. Those which affect function may be detected directly by measuring the function of the protein, e.g. its enzyme activity. Polymorphisms in noncoding regions may affect the level of expression of a particular protein and can be detected at the nucleotide sequence level or by looking at mRNA or protein expression. The complete absence of a gene, called a 'null genotype,' may also be classified as a polymorphism.

Techniques which look at the nucleotide sequence directly are called genotyping methods, whereas those which address expression and function are known as phenotyping approaches. Genotyping involves examining the altered nucleotide sequences of polymorphic variants in the DNA itself. This approach traditionally involved cloning and sequencing the gene of interest from different individuals, but recent technological developments mean that genotyping is now usually undertaken using polymerase chain reaction-based methods (PCR).

Traditional methods for phenotyping xenobiotic metabolizing enzymes involve measuring enzymatic activity using diagnostic substrates. The advantage of this approach is that it directly addresses the actual function of the enzyme and does not require assumptions about genotype–phenotype correlations, but it has become less popular in recent years because it is more labor intensive than genotyping.

The genotype can be determined by looking at any cell type because all of an individual's cells, except for red blood cells, contain the same genomic DNA sequence. However, in order to look at levels of expression and function, it is important to identify and access the tissue in which the phenotype is actually expressed. In the case of hepatic enzymes, this would entail obtaining a liver biopsy sample, which is clearly impracticable when examining healthy populations. Instead, white blood cells are usually used for this purpose since they may be obtained by relatively noninvasive methods. Variable CYP expression has been detected in these cells, but it must be acknowledged that this might not reflect the situation in the liver or the target tissue.

Many investigators now use mRNA-based techniques such as TaqMan 'real time' PCR, which is a simple and rapid method for quantitating the expression of a particular mRNA rather than using labor-intensive protein-based approaches. However, the results of these studies must be examined critically because of the potential that differences in mRNA levels may not be reflected at the protein, and hence functional, level.

2.3.3 Correlating Genotype and Phenotype

In order to use genotype as an indicator of phenotype it is necessary to demonstrate a clear correlation between genotype and phenotype. In some, but by no means all, cases a clear correlation between genotype and phenotype has been established. The use of validated assays is critical, and ideally one should also be confident that all possible allelic variants of the gene in question have been identified.

The drawback of genotyping methods in toxicogenetics is that they provide only an indirect measure of the actual activity of the enzyme or protein of interest. To interpret the results it is necessary to make the assumption that in a given individual the phenotype is a direct function of the genotype. In order to justify this assumption it is necessary to demonstrate a clear correlation between genotype and phenotype. In some cases, such as the aromatic-amine-metabolizing enzyme NAT2, a clear correlation between genotype and phenotype has been established. However, in other cases the relationship between genotype and phenotype is more complex.

A good example of the difficulty of relating genotype to *in vitro* and *in vivo* phenotype is serum paraoxonase (PON1), an enzyme that is involved in the metabolism of organophosphate pesticides such as diazoxon. There is a large variation in activity levels of this enzyme between individuals. Two polymorphic sites, at codons 55 (Leu^{55} to Met^{55}) and 192 (Gln^{192} to Arg^{192}), have been identified within the human PON1 gene. These occur relatively frequently and lead to structural changes in the protein. The literature on the consequences of these polymorphisms is contradictory due to an artifact generated by the method used to assay the activity. When assayed under published conditions, i.e. in the presence of 2 M NaCl, the Gln^{192} isoform was more active towards diazoxon than was the Arg^{192} isoform, but when injected into PON1 knockout mice both isoforms gave similar levels of protection against a diazoxon challenge. Further experimentation, using more physiological salt concentrations, revealed that high concentrations of NaCl differentially inhibit the Arg^{192} isoform of PON1. It also proved important to use concentrations of substrate which are as similar as possible to those to which cells might be exposed *in vivo*. This emphasizes the importance of determining activities under conditions that reflect the *in vivo* situation as closely as possible.

Another point arising from the example of PON1 is the importance of considering actual function, as well as genetics, when evaluating the relevance of metabolic polymorphisms as risk factors in susceptibility to the adverse consequences of exposure to xenobiotics. In the case of PON1, there was more variation in activity between individuals having the same genotype than there was between the activities associated with different genotypes, meaning that genotyping alone is insufficient to determine the role of PON1 in risks associated with exposure to organophosphates.

To summarize, if genotyping methods are to be used as a way of evaluating the role of polymorphisms in toxic processes, it is very important that a good correlation has been established between genotype and phenotype. In addition, the use of validated assays is critical. Ideally one should also be confident that all the possible allelic variants of the gene in question have been identified.

2.3.4 The Role of Polymorphisms in Influencing Susceptibility to Toxic Agents

Great effort has been devoted to trying to understand the roles of polymorphic xenobiotic-metabolizing enzymes in determining susceptibility to toxic agents. Following the identification of polymorphisms in drug transporters and xenobiotic receptors, it has become apparent that these, too, may contribute to toxicogenetic variation. The current status of work in these areas is summarized briefly in this section.

Polymorphic Xenobiotic-metabolizing Enzymes

Polymorphisms exist among both Phase I and Phase II xenobiotic metabolizing enzymes. The genes affected include several CYP isozymes, xanthine oxidase, alcohol dehydrogenase, aldehyde dehydrogenase, glutathione *S*-transferases (GSTs), *N*-acetyltransferases (NATs), UDP-glucuronyltransferases (UGTs), choline esterase, phenylacetate esterase, methyltransferase, and paraoxonase.

The nomenclature and consequences of polymorphisms within the CYP family are introduced in Chapter 2.2A. Several polymorphic CYP isozymes, including CYP2C9, CYP2C19, and CYP2D6, have a significant impact on the metabolic clearance of therapeutic agents, as discussed in detail in Chapter 2.2A.

The topic of this chapter is the role of polymorphisms in the toxic effects of xenobiotics. A number of metabolic polymorphisms are known to have a significant impact on individual susceptibility to toxic and carcinogenic compounds, although the literature in this area can often be contradictory. In this chapter, rather than presenting a long list of polymorphisms which may or may not be involved in susceptibility to chemically induced disease processes and toxicity, a few of the better characterized examples will be presented.

Cancer Susceptibility

The role of polymorphic xenobiotic metabolizing enzymes in cancer susceptibility is well illustrated by the induction of bladder cancer by exposure to aromatic amines, and of colorectal cancer by heterocyclic amines. The current state of knowledge in this area is summarized in Table 2.16.

Aromatic Amine-induced Bladder Cancer

Bladder cancer is associated with occupational exposure to aromatic amines, and smoking is also a major risk factor. Aromatic amines are activated by CYP1A2-mediated N-hydroxylation and possibly by O-acetylation of *N*-hydroxy aromatic amines in situ in the transitional epithelium. They are detoxified by N-acetylation, glutathione conjugation, sulfation, and glucuronidation. Thus, individuals who are slow for NAT2, rapid for CYP1A2, GSTM1 null, and rapid for NAT1 would be expected to have the highest risk of developing bladder cancer. Certainly, slow acetylators (at least those who smoke) can be shown to have an increased risk, and GSTM1-null individuals are also at increased risk. There is also evidence that individuals with the slow NAT2 in association with a rapid CYP1A2 phenotype have a further increase in their risk of bladder cancer if they are smokers.

Aromatic amines such as 4-aminobiphenyl are used in a number of industrial processes and are also found in cigarette smoke. Bladder cancer is associated with occupational exposure to aromatic amines in the rubber, textile, dye, and chemical industries. Smoking is also a major risk factor (approximately 66% of bladder cancers in Western countries are attributable to cigarette smoking) and this is thought to be due to the presence of aromatic amines as well as polycyclic aromatic hydrocarbons and nitrosamines in cigarette smoke.

Carcinogenic aromatic amines are metabolized by acetylation, which is catalysed by *N*-acetyltransferases (NATs). Humans have two functional NAT genes: NAT1, which was originally called the ‘monomorphic’ NAT (although it is now known to be polymorphic) and NAT2, which was originally identified as the ‘polymorphic’ NAT. The expression

Table 2.16 *Influence of metabolic polymorphisms on susceptibility to bladder and colon cancer.*

Enzyme	Bladder cancer		Colon cancer	
	Effect	Proposed mechanism	Effect	Proposed mechanism
CYP1A2	High CYP1A2 activity in association with the slow NAT2 phenotype may confer increased risk, especially in smokers	Increased metabolic activation of aromatic amines by hepatic N-hydroxylation	High CYP1A2 activity may be associated with increased risk, especially in association with the fast acetylator phenotype and regular consumption of well cooked red meat	Increased metabolic activation of heterocyclic amines by N-hydroxylation
NAT2	Slow acetylation is associated with increased risk in smokers	Reduced detoxification of aromatic amines by N-acetylation	Fast acetylation may be associated with increased risk, especially in association with rapid CYP1A2 phenotype and regular consumption of well cooked red meat	Increased metabolic activation of *N*-hydroxy heterocyclic amines by O-acetylation
NAT1	Not yet clear	May activate *N*-hydroxy aromatic amines by in situ O-acetylation in the transitional epithelium	The high-activity variant NAT1*10 may be associated with increased risk	Increased metabolic activation of *N*-hydroxy heterocyclic amines by *in situ* O-acetylation in the colon
GSTs	The GSTM1 null genotype may be associated with increased risk	Reduced detoxification of *N*-hydroxy aromatic amines by glutathione conjugation	The low-activity variant GSTA1*B may be associated with increased risk	Reduced detoxification of *N*-hydroxy heterocyclic amines

pattern of NAT2 is typical of a drug-metabolizing enzyme, the highest levels of expression being observed in the liver and small intestine with only very low levels being found in other tissues.

One of the reasons that the role of the NAT2 polymorphism in mediating the effects of foreign compounds is relatively well understood is that convenient methods exist for measuring the activity of this enzyme both *in vitro* and *in vivo*. The most commonly used *in vitro* model substrate for NAT2 is sulphamethazine, but for *in vivo* phenotyping it is convenient to use caffeine since volunteers may be administered this compound without inciting concerns regarding toxicity.

In addition to the availability of convenient and reliable phenotyping methods, a number of methods are available for the determination of NAT2 genotype by PCR. This has facilitated epidemiological analysis of the NAT2 polymorphism, facilitated by the fact that a clear correlation between genotype and phenotype was quickly established for this enzyme. Humans have traditionally been classified as either 'fast' or 'slow' acetylators depending upon their NAT2 phenotype; in most Caucasian populations approximately 60% of individuals are slow acetylators. Now that genotyping methods are readily available, it is also possible to identify heterozygotes (i.e. people with one rapid and one slow allele) as an 'intermediate' group.

The majority of studies have not addressed the potential role of individual variation in the kinetics of the different NAT2 variants in determining the consequences of exposure to hazardous chemicals (which may be at very low levels) in terms of toxic effects or predisposition to specific diseases. The kinetic characteristics of several NAT2 variants, however, have been investigated *in vitro* by expressing them in *E. coli* and measuring the acetylation of model substrates. The *in vivo* acetylation phenotype, with respect to a particular variant, will depend on the concentration of the substrate at the enzyme active site (which is expected to be a function of the plasma concentration) relative to the values of K_m and V_{max}. For example, when the K_ms of two NAT2 variants (NAT2*4 and NAT2*7) were compared using the drugs sulphamethazine and dapsone, the K_m values for NAT2*4 were much lower than those for NAT2*7 (12.8 vs 123 μM for sulphamethazine and 22 vs 117 μM, for dapsone). In the case of sulphamethazine, this would not be predicted to affect the rate of metabolism *in vivo* since the plasma concentration of this drug in clinical use is ~100 μM. However, in the case of dapsone, plasma concentration is only ~10 μM and, therefore, the difference in K_m between NAT2*4 and NAT2*7 might have therapeutic consequences.

In an influential study by Cartwright published in 1982, it was found that 22/23 (95.7%) dye factory employees (or ex-employees) with bladder cancer were slow acetylators, whereas only ~60% of controls were slow acetylators. Although subsequent studies found somewhat lower slow acetylator frequencies in bladder cancer patients, the observation that slow acetylators have an increased risk of bladder cancer is now well established, and has been confirmed in a series of meta-analyses. This effect seems to be specific to smokers: NAT2 slow acetylators are especially susceptible to the adverse effects of smoking on bladder cancer risk, whereas slow acetylation does not increase bladder cancer risk among subjects who had never smoked.

The role of NAT2 should not be considered in isolation, since susceptibility to bladder cancer is also modulated by interaction with other xenobiotic-metabolizing enzymes. The metabolic activation of aromatic amines is illustrated in Figure 2.40. The first step in this process is hepatic N-hydroxylation by the CYP isozyme CYP1A2. The resulting

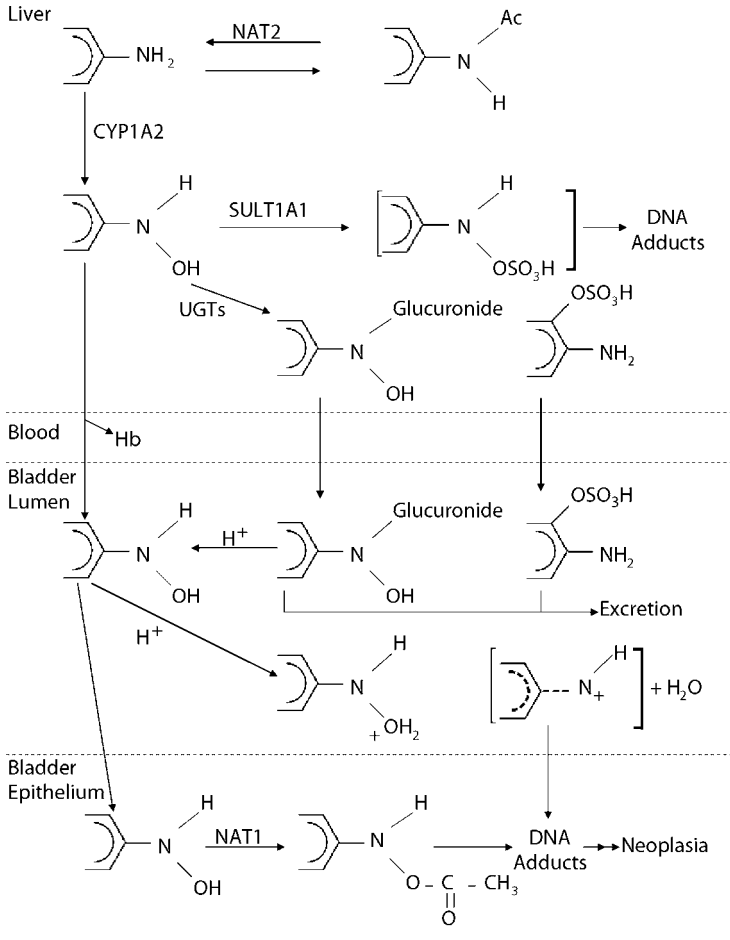


Figure 2.40 *Metabolism of aromatic amines*
In the liver, aromatic amines are either acetylated by NAT2 or N-hydroxylated by CYP1A2. The acetylation reaction is a detoxifying step. The N-hydroxy aromatic amine may be further metabolized by SULT1A1 or UGTs. The metabolites travel via the systemic circulation to the bladder lumen, where the conjugated metabolites can be excreted. However, the N-hydroxylated metabolite, which can enter the bladder as it is or be regenerated from the glucuronide under acidic conditions, is able to enter the transitional epithelial cells. Within these cells, O-acetylation by NAT1 may lead to metabolic activation and hence to the formation of DNA adducts and initiation of bladder tumours. Figure kindly provided by Prof. F.F. Kadlubar, National Center for Toxicological Research, USA.

N-hydroxy aromatic amines form glutathione, glucuronide, or sulfate conjugates (catalysed by GSTs, UGTs, and sulfotransferases, respectively). These soluble metabolites are transported to the bladder via the systemic circulation and may then be taken up by transitional epithelial cells. The conjugates may be hydrolysed, either under acidic conditions in the urine itself or as a result of intracellular hydroxylase activity, regenerating the original *N*-hydroxyl amine.

The other NAT isozyme (NAT1!) is expressed in epithelial tissues, including the urothelium, and is thought to play a key role in cellular homeostasis. Like NAT2, it can metabolize aromatic amines, but it is as yet unclear what the consequences of this polymorphism are for bladder cancer susceptibility. It is currently believed that NAT1 O-acetylates *N*-hydroxy aromatic amines in situ in the transitional epithelium, forming highly reactive N-acetoxy esters, which are able to bind directly to DNA, potentially leading to the initiation of carcinogenesis.

The GST isozyme GSTM1 may also contribute to the risk of bladder cancer. The GSTM1 polymorphism takes the form of a null allele, i.e. a 'missing gene,' designated GSTM0, and individuals may carry two, one, or no copies of the GSTM1 gene. Case-control studies and meta-analyses indicate that the risk of bladder cancer is increased by ~50–70% in homozygous GSTM1 null-individuals. This effect, unlike that observed in relation to NAT2, was similar in both smokers and nonsmokers.

Thus, with respect to bladder cancer, the toxicogenetic situation is relatively clear-cut. Activating and detoxifying enzymes in the liver compete for the aromatic amine substrate, leading to the generation of metabolites, which are subsequently delivered to the bladder. In this scenario, low levels of hepatic N-acetylation (NAT2) allow N-hydroxylation by CYP1A2 to predominate. Following further metabolism by GSTs and transport to the bladder, hydrolysis regenerates a chemically reactive *N*-hydroxy aromatic amine, which is susceptible to O-acetylation by NAT1, leading to the initiation of carcinogenesis. The hypothesis based upon this scheme is that individuals who are slow for NAT2, rapid for CYP1A2, GSTM1-null, and rapid for NAT1 have the highest risk of developing bladder cancer. Certainly, slow acetylators (at least those who smoke) can be shown to have an increased risk, as are GSTM1-null individuals. There is also evidence that individuals with the slow NAT2 in association with a rapid CYP1A2 phenotype have a further increase in their risk of bladder cancer if they smoke.

Heterocyclic Amine-induced Colorectal Cancer

The determination of colorectal cancer susceptibility involves both genetic and environmental factors. Heterocyclic amines, generated during the cooking of red meat, are thought to be key carcinogens in the colon. The main CYP isozyme involved in the Phase I metabolism of heterocyclic amines is CYP1A2. Both the parent heterocyclic amines and their Phase I metabolites also undergo Phase II metabolism catalysed by GSTs, UGTs, sulfotransferases, and NATs. There is evidence that the risk of colorectal cancer is increased in individuals who are current or ex-smokers, prefer red meat well cooked, and have the rapid phenotypes of both CYP1A2 and NAT2. The GSTA1 isozyme may also play a role in colorectal cancer susceptibility.

It is unusual for studies on toxicogenetic factors in cancer to yield the kind of relatively clear-cut results seen in the case of bladder cancer. Attempts to elucidate the role of the metabolic polymorphisms in colorectal cancer illustrate some of the difficulties in obtaining definitive answers.

An individual's risk of colorectal cancer is modified by both genetic and environmental/lifestyle factors, including a first degree family history of colorectal cancer,

overeating, physical inactivity, a high intake of red meat, alcohol use, smoking and a low intake of vegetables. Heterocyclic amines are generated during chemical reactions which occur during the cooking of red meat. Thus, people who frequently consume cooked meat are exposed to heterocyclic amines on a regular basis, and these compounds are thought to be the central carcinogens that mediate the carcinogenic process in the colon. The heterocyclic amine 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine (PhIP) is considered to be of particular significance with respect to colorectal cancer because it is usually the predominant heterocyclic amine found in cooked meat, it is mutagenic in bacterial and mammalian cell-based assays, and it is a colon carcinogen in rats. The pathways involved in the metabolic activation of PhIP are illustrated in Figure 2.41.

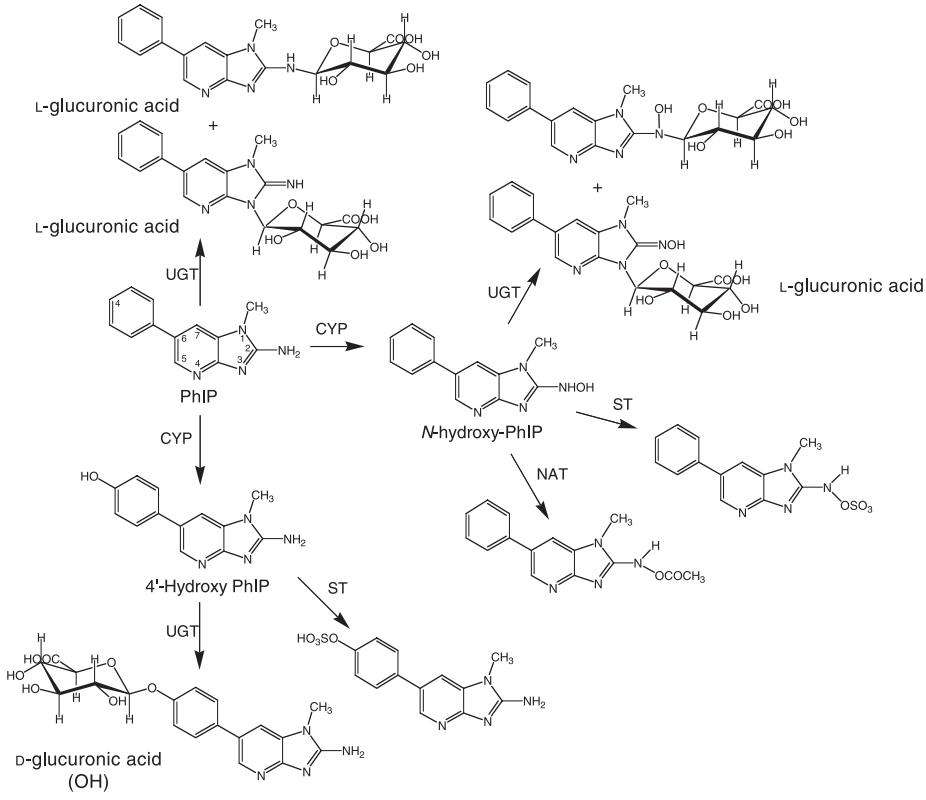


Figure 2.41 *Metabolism of the heterocyclic amine 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP)*
The parent molecule, PhIP, may undergo either glucuronidation (UGT-mediated), 4′-hydroxylation (CYP-mediated), or N-hydroxylation (CYP1A2-mediated). The N-hydroxylation step is considered to represent metabolic activation, and N-hydroxy-PhIP can be further metabolized by glucuronidation (UGT-mediated), sulfate esterification (ST-mediated), or O-acetylation (NAT-mediated). Both sulfate esterification and O-acetylation lead to the generation of highly reactive intermediates, which are thought to play a key role in the induction of colorectal cancer. Figure kindly provided by Dr N.J. Gooderham, Imperial College, London.

Heterocyclic amines such as PhIP are subject to both ring- and N-oxidation mediated by CYP; the N-hydroxy metabolites are direct-acting mutagens whereas the ring-hydroxylated metabolites are not mutagenic. The main CYP isozyme involved in the Phase I metabolism of heterocyclic amines is CYP1A2. Both the parent heterocyclic amines and their Phase I metabolites also undergo Phase II metabolism catalysed by GSTs, UGTs, sulfotransferases, and NATs. The key metabolites implicated in colorectal cancer induction are the N-acetoxy derivatives, which may be generated by the action of NAT1 and NAT2. The Phase I metabolism of heterocyclic amines is primarily hepatic, but O-acetylation can occur in either the liver or in the colon.

In humans, the generation of reactive genotoxic metabolites via N-oxidation is the primary route of oxidative metabolism for heterocyclic amines. In experimental animals, on the other hand, both activation and detoxification of these compounds occur, and this may lead to differences in susceptibility to, and the tissue specificity of, their carcinogenic effects. The species differences observed in the oxidative metabolism of heterocyclic amines illustrate the difficulty of extrapolating from animal to human, especially where polymorphic enzymes are involved.

The pattern of metabolism of heterocyclic amines suggests that the rapid NAT2 phenotype would be likely to confer an increased risk of colorectal cancer, especially when combined with the rapid CYP1A2 and regular consumption of well cooked red meat. Phenotyping studies suggest that fast acetylators do have increased risk of colorectal cancer and this is exacerbated by consumption of well cooked red meat. It has also been suggested that the NAT1*10 variant of NAT1, which is associated with high acetylation activity in colon tissue, may be a risk factor. However, the results of studies using genotyping alone have been contradictory. This may be a function of the fact that the phenotype–genotype correlations for NAT1 and NAT2 are still not fully understood.

It has recently been established that *N*-acetoxy-PhIP is a substrate for detoxification by the GST isozyme GSTA1. This is the only human GST that can detoxify *N*-acetoxy-PhIP efficiently and is polymorphic in the human population. Its role as a factor in colorectal cancer susceptibility provides a mechanism-based example of the toxicogenetic influences of GSTs. Two GSTA1 alleles, GSTA1*A and GSTA1*B, exist in the human population. The GSTA1*B allele is an example of a regulatory polymorphism since it has three point mutations in the 5′ regulatory region, leading to reduced hepatic expression of the corresponding protein. Individuals who are homozygous for GST1A*B have approximately twice the risk of colorectal cancer compared with heterozygotes and GSTA1*A homozygotes. The pattern of expression of GSTA1 (it is expressed at high levels in the liver but only at very low levels in the colon) provides a possible explanation for the tissue specificity of carcinogenesis induced by heterocyclic amines such as PhIP: one can hypothesize that the liver is protected by its high level of GSTA1 whereas the colon is susceptible because it lacks the necessary enzyme. In individuals with high hepatic levels of GSTA1, PhIP is efficiently detoxified and very little escapes to be transported to the colon. In individuals with low hepatic levels of GSTA1, however, more PhIP escapes hepatic detoxification and can be transported to the colon where it is able to exert its carcinogenic effect.

It is evident that the determination of colorectal cancer susceptibility is multifactorial, involving both genetic and environmental factors. There is some evidence that the risk of colorectal cancer is increased in individuals who are current or ex-smokers, prefer red

meat well cooked, and have the rapid phenotypes of both CYP1A2 and NAT2. Large epidemiology studies are required to dissect this complex interaction, illustrating the difficulty of dissecting the role of toxicogenetic factors in disease susceptibility. Overall, it is likely that polymorphic metabolic factors such as CYP1A2 and NAT1/2 do affect susceptibility to colorectal cancer, but there is currently insufficient evidence to justify the inclusion of metabolic factors in population screening programs.

2.3.5 Acute Toxicity

Acrylonitrile is known to be metabolically activated and detoxified via oxidative (CYP2E1) and reductive (GST) routes, respectively. Slower CYP2E1-mediated metabolism of acrylonitrile caused by a polymorphism in some individuals could lead to lower levels of metabolic activation. Less efficient detoxification may occur in individuals who carry variant forms of GSTP1, leading to higher blood levels of acrylonitrile and potentially increased susceptibility to toxicity. It has been suggested on the basis of these findings that the biomonitoring of industrial exposure to acrylonitrile should be supported by genotyping for GSTP1.

Relatively few well characterized examples of acute toxicity affected by polymorphic xenobiotic-metabolizing enzyme currently exist. One of the best examples to date is acrylonitrile, an industrial solvent which is, in addition to being carcinogenic, acutely toxic. There is significant inter- and intraspecies variability in response to the acute toxicity of acrylonitrile.

Acrylonitrile is metabolized via oxidative and reductive routes. The glutathione-dependent detoxification pathway leads via the primary metabolites *S*-cyanoethylglutathione and *S*-methylglutathione to the mercapturic acids *N*-acetyl-*S*-cyanoethylcysteine and *N*-acetyl-*S*-cyanomethylcysteine, which are the final urinary excretion products. The alternative oxidative pathway of acrylonitrile metabolism, mediated by the isozyme CYP2E1, generates a reactive epoxide metabolite, cyanoethylene oxide (CEO), which is further metabolized to cyanide. Both acrylonitrile and CEO also react with tissue thiols, leading to glutathione depletion. The acute toxicity of high doses of acrylonitrile is thought to be 'gated' by the depletion of hepatic glutathione, and this has led to the recommended use of *N*-acetylcysteine as an antidote to acrylonitrile poisoning.

It is difficult to extrapolate directly from the results of animal experiments to the effects of acrylonitrile in humans because humans have an active epoxide hydrolase pathway that leads to the generation of cyanide from CEO, whereas this is not the case in rodents. The oxidative metabolism, which leads to the formation of cyanide, seems to be much less important in animals than it is in humans. The acute toxicity of acrylonitrile in animals seems to be mediated by GSH depletion, whereas in humans acute toxicity is largely determined by the metabolic formation of cyanide.

Humans are known to have a higher level of CYP2E1-mediated oxidative metabolism of acrylonitrile than rodents. The formation of an *N*-(cyanoethyl)valine adduct at the *N*-terminus of hemoglobin serves as a biomarker that is indicative of blood levels of acrylamide. By measuring levels of *N*-(cyanoethyl)valine in relation to acrylonitrile exposure, higher adduct levels have been demonstrated in individuals with at least one copy of a CYP2E1 promoter variant ($A_{-316}G$). The biological effects of this polymorphism require clarification, but there

are reports of individual variation in the inducibility of CYP2E1. Slower CYP2E1-mediated metabolism of acrylonitrile in some individuals could lead to lower levels of metabolic activation (to CEO and cyanide) and increased blood levels of the parent compound.

Studies on individual cases of acrylonitrile intoxication have provided some clues about the role of metabolic polymorphisms in the detoxification of acrylonitrile in humans. When the effects of acute acrylonitrile exposure were compared in two individuals, one with low and one with high GST activity, the individual with low activity experienced headache, nausea, and vomiting. Furthermore, the level of hydrocyanic acid in his blood was within the lethal range, although fortunately he recovered following treatment with the antidote, *N*-acetylcysteine. This was consistent with the hypothesis that, particularly in individuals with low GST activity, toxicity is 'gated' by GSH depletion. If insufficient GSH is available for conjugation (or the activity of GST is too low), free acrylonitrile may become available to enter the CYP2E1-mediated oxidative pathway leading to toxicity.

Biomonitoring studies have suggested a role for the GST isozyme GSTP1-1 in the human metabolism of acrylonitrile. Individuals with the wild-type form of GSTP1 (GSTP1*A) appear to have a lower level of *N*-(cyanoethyl)valine adducts compared with those who are polymorphic at this locus (GSTP1*B and GSTP1*C). This was thought to be due to altered affinity of the enzyme for electrophilic substrates: the wild-type GST1*A enzyme efficiently conjugates glutathione to electrophilic substrates (i.e. it has a low K_m and a high K_{cat}/K_m ratio) whereas the GSTP1*B and GSTP1*C variants have higher K_m and lower K_{cat}/K_m values. This would be expected to lead to less efficient detoxification in the individuals who carry variant forms of GSTP1 and potentially to increased susceptibility to acrylonitrile toxicity.

It has been suggested on the basis of these findings that the biomonitoring of industrial exposure to acrylonitrile should be supported by genotyping for GSTP1. This illustrates the way in which toxicogenetic data could, in the future, be used to identify and protect susceptible subgroups within the exposed subpopulation.

2.3.6 Polymorphic Drug Transporters

Drug transporters influence the oral bioavailability, disposition, and excretion of xenobiotics. Inter-individual variation in the expression of drug transporters may affect the excretion of cationic xenobiotics and metabolites, thus altering their toxicity.

Drug transporters are membrane proteins, which mediate the cellular uptake and efflux of molecules, and play a key role in toxicokinetics because they influence the oral bioavailability, disposition, and excretion of xenobiotics (see Phase III metabolism in Chapter 2.2A). There is a marked degree of inter-individual variation in the expression of the drug transporter protein P-glycoprotein (MDR1) in the small intestine and liver, and furthermore there is polymorphic variation in the sequence of the MDR1 gene. By the beginning of 2004, 29 single nucleotide polymorphisms (SNPs) had been identified in the MDR1 gene, but the role of MDR1 polymorphisms in responsiveness to drug and toxic agents remains to be

elucidated. Other drug transporters, including the uptake carriers OATP and OCT, have also been shown to be polymorphic. OATP is localized in the liver, kidney, brain, and intestine. The OATP-A transporter, which is also polymorphic, is localized in the capillary endothelial cells of the brain and may therefore play a role in the CNS action and toxicity of drugs and toxicants. Similarly, polymorphisms have been identified in the OCT1 transporter, which mediates hepatic uptake of cationic substrates. These may affect the hepatic excretion of cationic xenobiotics and metabolites, thus altering their hepatotoxicity.

2.3.7 Polymorphic Xenobiotic Receptors

The Ah receptor, Ah-receptor nuclear translocase (Arnt), and the Ah-receptor repressor (AHRR) are thought to be involved in the endocrine disrupting effects of chemicals such as dioxin. The study of polymorphisms in xenobiotics' receptors and their role in toxicological responsiveness is still in its infancy, but there is already evidence that polymorphisms in the AHRR may be involved in male reproductive problems such as micropenis and infertility.

Our current understanding of the role of receptor polymorphisms in determining susceptibility to toxic agents is much less advanced than in the case of drug-metabolizing enzymes. There are a number of reasons for this; in particular, genotype–phenotype correlations are difficult to determine for receptors because, in order to characterize their function, it is necessary to use an experimental system in which one can demonstrate:

- Ligand binding
- Conformational change to the DNA-binding state
- Interaction with co-activators and co-repressors
- Binding to enhancer sequences
- Transcriptional activation

In practice, this makes it necessary to use viable intact cell systems (e.g. cultured lymphocytes or recombinant cell systems) rather than simple biochemical methods in order to analyse receptor function, with all the technical challenges this implies.

One of the most important receptors involved in the response to toxic agents, including polycyclic aromatic hydrocarbons and dioxins, is the Ah receptor (AhR) (see Chapter 6.1). Three key proteins are involved in the regulation of AhR responsive genes: the AhR itself, the Ah-receptor nuclear translocase (Arnt), and the Ah-receptor repressor (AHRR). All three of these are polymorphic in human populations, but the biological consequences of these polymorphisms are not yet fully understood. However, by analogy with the situation in the mouse, in which genetic variation in the AhR has powerful phenotypic effects, it is reasonable to predict that polymorphisms in the human AhR and its associated proteins will be of toxicological significance. This is an active area of research; in particular, the role of the AhR and its associated proteins, and polymorphisms therein, in mediating susceptibility to the endocrine disrupting effects of chemicals such as dioxin is the subject of intense research activity.

Dioxins are known to have hormone-like (oestrogenic and anti-oestrogenic) activities, which are mediated via the AhR and its associated proteins, and it has been predicted that

polymorphisms in these proteins will modulate individual susceptibility to such effects. A concrete example exists in the case of the AHRR, which plays a key role in the feedback regulation of AhR function. The AHRR is polymorphic at codon 185: approximately 30% of Japanese men are homozygous for the Pro^{185} form of this protein, whereas 70% are either heterozygous or homozygous for the Ala^{185} form of the AHRR. This polymorphism is associated with the condition known as micropenis and may also be linked to male infertility. It is thought that the oestrogenic effects of dioxin exposure may be exaggerated in men with the homozygous Pro^{185}/Pro^{185} genotype because the Pro^{185} form of the protein exerts weaker feedback effects on dioxin signaling than does the Ala^{185} form.

At the time of writing, the study of polymorphisms in xenobiotics' receptors and their role in toxicological responsiveness is still in its infancy. However, it is reasonable to anticipate that over the next few years this aspect of toxicogenetics will be elucidated, leading to a much better understanding of the consequences of exposure to potent chemicals such as dioxin.

2.3.8 Summary

The strength of the toxicogenetic approach is that it makes it possible to examine susceptibility to xenobiotics in human populations, but interpretation of the data generated depends upon a sound knowledge of the range of polymorphisms which exist at any given locus. A degree of variability and uncertainty will always be present, and this must be borne in mind during the risk assessment process.

The emerging field of toxicogenetics promises to enhance our understanding of susceptibility to toxic agents in ways which could not have been imagined even a few years ago. The availability of high-throughput PCR-based methods has made it possible to examine large numbers of samples for many different polymorphisms simultaneously, generating huge volumes of data, which should be of great value in performing risk assessment. However, in order to ensure that the data generated are of high quality and the conclusions drawn are valid, a number of key issues must be borne in mind.

- **Genotyping and phenotyping:** Genotyping is becoming increasingly popular as a means of evaluating toxicogenetic variation, because of the advent of high-capacity automated PCR-methods. However, it is important to remember that it is the phenotype that determines function and will mediate any observed effects on susceptibility. Genotype is determined by the DNA sequence and is unchanging, whereas phenotype may change during life due to developmental changes in gene expression and, potentially, xenobiotic-induced effects such as induction or enzyme inhibition.
- **Study size and design:** The majority of studies on polymorphic variation and risk of toxicity or disease have, for scientific and resource reasons, concentrated on one or two key genes. However, in practice, susceptibility to toxic agents almost always involves many factors. Redundancy within xenobiotic-metabolizing pathways means that, if the activity of one enzyme is reduced, potential toxic agents may be directed down other pathways. For example, in smokers, various CYPs, GSTM, GSTT, GSTP, NAT1, and NAT2 may exert combined effects on DNA adduct formation and cancer

susceptibility. The epidemiological studies required in order to understand the roles of all the possible polymorphic variants of these enzymes would need to be enormous.

Many preliminary studies aimed at the identification of risk factors are relatively small, with only 100–300 cases and controls. This number is sufficient to detect common polymorphisms which double risk, but will not detect rare polymorphisms or those which cause less than a doubling in risk. Much larger studies and/or meta-analysis are required to reveal small increases in risk. It is also important to note that polymorphic variants may be present at different frequencies within ethnic groups; for example, 45% of Caucasians but only 10% of Japanese individuals are NAT2 slow acetylators. This means that the implications of the identification of a particular variant may differ depending on the ethnic makeup of the exposed population. It may not always be possible to extrapolate risk assessments from one ethnic group to another.

Most of the existing studies on the effects of toxicogenetic factors on, for example, cancer susceptibility have indicated only small effects on risk. Usually any observed increase in risk is less than 2-fold, which is insufficient to justify genetic screening or make it possible to offer genetic counseling. It now appears that a more fruitful approach might be to use epidemiological information to try to identify subgroups for whom a specific variant is actually a risk factor. In order to do this, it will be necessary to undertake very large studies with thousands, rather than hundreds, of subjects. With this in mind, the International Project on Genetic Susceptibility to Environmental Carcinogens (GSEC), a world-wide collaboration which collates data on frequencies of genetic polymorphisms in genes associated with carcinogen metabolism, was initiated in 1999. The project has collected data on approximately 15,000 control subjects from a variety of ethnic backgrounds. These individuals have been genotyped for CYP1A1, CYP2E1, CYP2D6, GSTM1, GSTP, GSTT1, NAT2, and epoxide hydrolase.

In order to understand the effects of polymorphisms on processes in toxicity, it is essential to generate the highest quality data using the best possible study design and methodology. The interpretation of the data generated depends upon a sound knowledge of the range of polymorphisms that exist at any given locus, and considerable work is still required in order to complete this database. Initiatives such as the GSEC project will improve this situation and hopefully encourage the generation of more interpretable data.

The strength of the toxicogenetic approach is that it makes it possible to examine susceptibility to xenobiotics in human populations, thus answering concerns relating to differences in susceptibility between humans and animals. However, the nature of toxicogenetic analysis, including the fact that, for ethical reasons, chemicals may not be administered to humans at known toxic doses and invasive methods may not be used to obtain relevant tissue samples, means that a degree of variability and uncertainty will always be present.

References

Cartwright, R.A., Rogers, H.J., Barhamhall, D., Glashan, R.W., Ahmad, R.A., Higgins, E., and Nahn, M.A. Role of N-Acetyltransferase phenotype in Bladder Carcinogenesis - a Pharmacogenetic Epidemiological Approach to Bladder-Cancer. *Lancet*, 1982, 2(8303), 842–846.

Gooderham, N.J., Murray, S., Lynch, A.M., Yadollahi-Farsani, M., Zhao, K., Boobis, A.R., and Davies, P.S. Food-derived Heterocyclic Amine Mutagens: Variable Metabolism and Significance carcinogen - DNA to Humans. *Drug Metab. Dispos.*, 2001, **29**, 529–534.

Kadlubar, F.F. and Badawi, A.F. Genetic Susceptibility and Carcinogen - DNA Adduct Formation in Human Urinary Bladder Carcinogenesis. Toxicol. Lett., 1995, **82–83**, 627–632.

For Further Reading

Daly, A.K., Development of Analytical Technology in Pharmacogenetic Research. *Naunyn-Schmiedeberg's Arch. Pharmacol*., 2004, **369**, 133–140.

Garcia-Closas, M., et al., NAT2 Slow Acetylation, GSTM1 Null Genotype, and Risk of Bladder Cancer: Results from the Spanish Bladder Cancer Study and Meta-analyses. *Lancet*, 2005, **366**, 649–659.

Garte, S., et al., Metabolic Gene Polymorphism Frequencies in Control Populations. *Cancer Epidemiology Biomarkers and Prevention*, 2001, **10**, 1239–1248.

Gerloff, T., Impact of Genetic Polymorphisms in Transmembrane Carrier-systems on Drug and Xenobiotic Distribution. *Naunyn-Schmiedeberg's Arch. Pharmacol*, 2004, **369**, 69–77.

Harper, P.A., Wong, J.M.Y., Lam, M.S.M., and Okey, A.B. Polymorphisms in the Human AH Receptor. *Chem-Biol Interact*, 2002, **141**, 161–187.

O'Leary, K.A., Edwards, R.J., Town, M.M., and Boobis, A.R. Genetic and Other Sources of Variation in the Activity of Serum Paraoxonase/Daizoxonase in Humans: Consequences for Risk from Exposure to Dioxin. *Pharmacogenetics Genomics*, 2005, **15**, 51–60.

Thier, R., Lewalter, J., Selinski, S., and Bolt, H.M. Possible Impact of Human CYP2E1 Polymorphisms on the Metabolism of Acrylonitrile. *Toxicol. Lett*., 2002, **128**, 249–255.

2.4 Cytotoxicity

Leslie Schwarz and John B. Watkins

2.4.1 Introduction

The term 'cytotoxic' may have different meanings. Frequently this term is used to indicate that a toxic chemical induces apoptosis or necrosis. In the present chapter effects of chemicals are also called cytotoxic when they disturb cellular functions in excess of normal physiological fluctuations, i.e. when they damage the cell but do not cause its death. Nevertheless, the consequence may be fatal for the organism.

Targets of Toxic Chemicals and Consequences for the Tissue

Generally, all cellular functions and structures may be targets of toxic chemicals.

Elucidation of the mechanisms of action of toxic chemicals is closely connected with the progress of biochemistry and cell biology during the past 80 years. On the one hand, knowledge of the physiologic functions of various cell structures is a necessary prerequisite for the analysis of the toxic action of chemicals on these structures. On the other hand, basic discoveries, such as elucidation of the function of nerve membranes or mitochondria, were only possible by studying an experimentally caused disturbance of

Table 2.17 *Criteria for the assessment of the meaning and possible consequences of cell damage.*

Meaning and function of the cellular target (qualitative)
Extent of the damage (quantitative)
Possibility of repair of the damage
Reversibility of the damage
Possibility of regeneration of the tissue

these functions by poisons. The extent of the damage depends on the function and meaning of the affected cell constituents, as well as on the possibility of their repair by the cell (Table 2.17). Repair often occurs via the degradation and resynthesis of the damaged cellular components.

Irreversible damage of essential functions inevitably causes death of the cell (Table 2.17). If the necrotic areas are small or only single cell necrosis is present, most tissues are capable of replacing the affected cells by regeneration. However, this mode of repair is not possible for all cell types, e.g. nerve cells of the brain cannot be replaced owing to their inability to divide. When larger areas of a tissue are damaged, many organs often replace the necrotic areas with connective tissue, resulting in scarring. The healing process results in a defective organ with diminished capacity.

Frequently, however, the disturbance induced by toxic chemicals does not result in destruction of the cell, but only in a temporary adjustment of cellular metabolic pathways to a new steady state. These disturbances may be manifested as a change in cell size (atrophy or hyperplasia) or in the storage of cellular contents such as the accumulation of triglycerides in liver cells (Chapter 3.2). In this context it is important to note that the impairment of cellular functions that are not involved in the survival of the cell may have serious consequences for the total organism. Examples are the disturbance of the function of the nerves or the heart muscle.

Structure-Effect Relationships at the Target

The toxicity of xenobiotics or their metabolites is frequently due to their chemical reactivity (i.e. covalent binding, radical attack); in contrast many animal, plant, or bacterial toxins act via their high structural specificity for distinct cellular targets (noncovalent interactions).

A characteristic feature of many toxic natural compounds is that they disturb specific cellular functions by highly selective interaction with distinct components of cells owing to their specific structural features (Table 2.18). The mode of interaction is mostly noncovalent. Animal, plant, and bacterial toxins inhibit or activate enzymes; interfere with the synthesis, storage, or secretion of autocoids, hormones, neurotransmitters; block receptors; inhibit the transfer of electrons by the mitochondrial respiratory chain; or interact with components of the cytoskeleton.

In contrast, relatively few synthetic chemicals exert their toxic actions via a disturbance of distinct cellular functions (Table 2.19). Many synthetic chemicals damage

Table 2.18 *Disturbance of specific cell functions by animal and plant poisons.*

Poison	Occurrence	Type of compound	Target
Amanitin	Mushroom: *Amanita*	Octapeptide	Inhibition of RNA-polymerases, in particular DNA-dependent RNA-polymerase II
Antimycin A	Bacteria: *Streptomyces griseus*		Inhibition of the mitochondrial electron transport at cytochrome b/c1
Atractyloside	Plant: *Atractylis gumnifera*		Inhibition of the ATP/ADP exchange at the inner mitochondrial membrane
Botulinus toxin	Bacteria: *Clostridium botulinum*	Protein (M 150,000)	Inhibition of the release of acetylcholine from presynaptic neurons
Cholera toxin	Bacteria: *Vibrio cholerae*	Protein (M 87,000)	ADP-ribosylation of the stimulatory G-protein of the adenylate cyclase causing an increased c-AMP-production
Colchicine	Plant: meadow saffron	Alkaloid	Inhibition of the aggregation of the microtubules affecting mitotic spindle formation
Curare	Plant: *Strychnos* and *Chondrodendrons*	Alkaloids	Inhibition of the neural-muscular signaling due to the blockage/ occupation of cholinergic receptors
Digitoxin	Plant: foxglove	Glycoside	Inhibition of the Na^+/K^+-ATPase
Diphtheria toxin	Bacteria: *Corynebacterium diphtheriae*	Protein	Inhibition of the protein synthesis by ADP-ribosylation of the elongation factor (EF-2)
Phalloidin	Mushroom: *Amanita*	Heptapeptide	Stabilization of actine filaments

the cell because of their chemical reactivity. These chemicals and their metabolites react nonspecifically with various cellular structures. Electrophilic metabolites bind covalently to nucleophilic positions in proteins and lipids as well as RNA and DNA. Radicals, which arise during the metabolism of a compound, may undergo addition or abstraction reactions with cellular molecules (Chapters 3 and 4).

Table 2.19 *Disturbance of specific cell functions by chemicals.*

Chemical	Target
Hydrocyanic acid	Inhibition of the mitochondrial electron transport due to binding to cytochrome oxidase
Fluoroacetate	Fluoroacetate is converted into fluoroacetyl-CoA, which condenses with oxaloacetate to form fluorocitrate; the latter inhibits the enzyme aconitase and thereby the citric acid cycle
Iodoacetate	Inhibition of glycolysis via the carboxylation of the SH-group of glyceraldehyde-3-phosphate dehydrogenase

For most chemicals, it is not clear which of the numerous possible cellular targets of reactive chemicals and metabolites are responsible for their toxic action. An answer to this question requires an analysis and toxicological assessment of the various targets. This has not yet been done sufficiently. Determining the extent of covalent binding to cellular macromolecules offers a crude measurement of the reactivity of chemicals and their metabolites, but does not specify their locus of toxicological action.

In Section 2.4.3 some important targets of toxic chemicals will be described. As this topic requires basic knowledge of the structure and function of the cell, an introduction in cell biology will be given here first.

2.4.2 The Cell

More than 200 different cell types can be found in the various organs of the body. They may exist separately or form closely connected cell layers such as the epithelial cells of the skin. Their morphology and that of the organs in which they are found are the result of the process of differentiation which prepares them for the specific physiological functions they must perform, e.g. production of mechanical energy by muscle cells, absorption of nutrients by epithelial cells of the intestine, or reception of light signals by the neural epithelium of the retina of the eye.

All cells contain the same genome but differentiation results from the fact that the various cell types only express distinct parts of the total genetic information. Nevertheless, cells of man and animals show similar structural elements, i.e. the outer membrane, the cell nucleus and the **cytoplasm**, which is a colloidal fluid containing numerous cell organelles. The structural elements of the cell will be described briefly below (Figure 2.42).

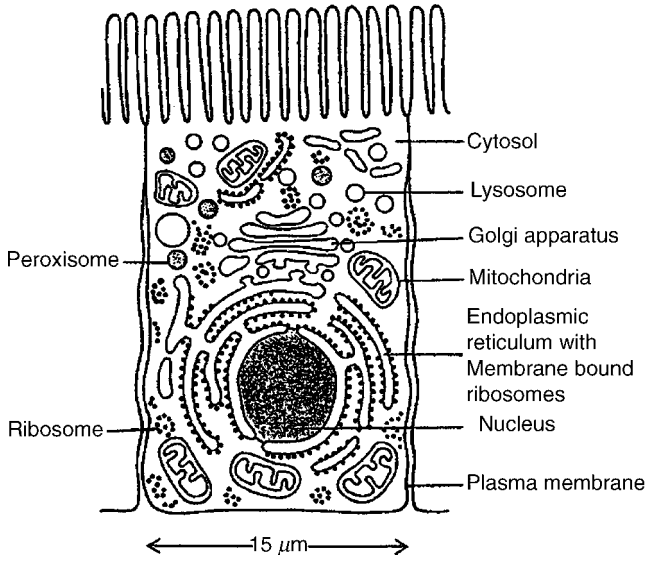


Figure 2.42 *Schematic drawing of a cell. (Reprinted from Alberts, copyright (1983), with permission from Garland Publishing.*

The Membrane System of the Cell

Membranes enclose the cell and subdivide it into various compartments.

Cell membranes consist of a double layer of polar lipids containing proteins (Figure 2.43). The **lipid bilayer** is a diffusion barrier for many compounds. Both nonpolar molecules, such as O_2, N_2, or benzene, and small polar molecules, such as H_2O, urea, and CO_2, can diffuse through the lipid phase of the cell membrane. Membrane permeability is minimal for larger uncharged polar molecules, such as glucose, whereas the membrane is almost impermeable to charged molecules. Some polar molecules, e.g. certain inorganic ions, sugars, amino acids, and numerous metabolites, which are formed in intermediary metabolism, can traverse the cell membrane, due to the presence of highly specific transport proteins.

The transport of small nonpolar molecules through the membrane may be driven by the concentration gradient of the chemical on either side of the barrier. Transporting charged chemicals against a concentration gradient often requires the cell to utilize mechanisms which require energy in the form of ATP (adenosine 5′-triphosphate). As a consequence of selective permeability and active-transport processes cells develop characteristic environments within their membrane-enclosed compartments. The creation of specific reaction areas in the cell is a prerequisite for the well ordered course of cellular metabolism. To give an impression of the great differences in the composition of extracellular and intracellular fluids, the concentrations of some inorganic ions are listed in Table 2.20. The concentration gradients between the cytosol and the outer liquid and between the cytosol and distinct intracellular organelles are maintained by energy-dependent transport processes, e.g. the Na^+/K^+-ATPases and Ca^{2+}-ATPases.

The differential conductivity of membranes for Na^+, K^+, and Cl^- is responsible for the generation of the electric membrane potential. As a basic feature of living cells, the membrane potential is of central importance for the transport and distribution of many molecules. Besides control of the transport of molecules, the cell membrane also plays an essential role in the reception and forwarding of the signals of many messenger molecules, such as hormones, which are indispensable for coordination of different

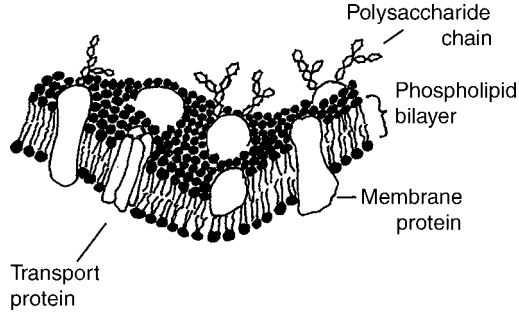


Figure 2.43 *Schematic depiction of a cell membrane. The phospholipid bilayer contains (glyco)proteins.*

Table 2.20 *Comparison of ion concentrations inside and outside a typical mammalian cell.*[a]

Ion	Intracellular concentration (mmol l^{-1})	Extracellular concentration (mmol l^{-1})
Na^+	5–15	145
K^+	140	5
Mg^{2+}	30	1–2

[a]Reprinted from Alberts, et al. Copyright (1983), with permission from Garland Publishing.

tissues of the body. Proteins, peptides, steroids, amino acids, and fatty acid derivatives with hormonal activity are either produced by special gland tissues of the body and reach the target cells via the bloodstream, or they are synthesized close to their place of action, often in the same tissue. It is even possible for cells to secrete growth factors for their own stimulation. Such mechanisms are termed autocrine and play an important role in several malignant tumors. Many hormones bind to specific receptor glycoproteins and lipids within the membranes of their target cells and set off an ordered sequence of cellular reactions, resulting in specific adjustment of cell metabolism.

Components of the Cytoplasm

The cytoplasm consists of the cytosol and the organelles of the cell.

Many important reactions in cellular metabolism occur within the cytosol, which is the fluid portion of the cytoplasm, e.g. the synthesis of fatty acids and cytosolic proteins, metabolic pathways such as glycolysis, and the pentose phosphate pathway, and ribosomal protein synthesis.

The cytosol contains a complex network of protein filaments, i.e. the **cytoskeleton**, which consists of actin filaments, intermediary filaments, and microtubules. The actin filaments and the microtubules contribute to the shape as well as to the stability and placement of surface structures of the cell. Microtubules are also components of the mitotic spindles, which move the chromosomes during cell division. Furthermore they form the centrioles (cylindrical organelles) which serve as the start of the spindle during mitosis.

The cytosol contains several cell organelles that are surrounded by a membrane: **endoplasmic reticulum (ER), Golgi apparatus, lysosomes, peroxisomes and mitochondria.** With the exception of the mitochondria these cell organelles develop from the ER. The ER is a network of tubules and stacks of flattened cisternae that are continuous with the outer nuclear membrane. Functionally, the ER is the site of the synthesis of lipids, membrane proteins, and secretory proteins. The latter are synthesized by the so-called rough ER (rER), which is covered with ribosomes.

The Golgi apparatus (GA) consists of stacks of disc-shaped cisternae from which small vesicles bud, which then may fuse with the plasma membrane and mediate secretory processes. In the GA oligosaccharides of glycoproteins, which were formed in the ER, achieve their final structure.

Lysosomes are small membranous organelles containing hydrolytic enzymes such as peptidases, lipases, and phophatases, and are the location of the controlled intracellular digestion of cellular components.

Peroxisomes are cell organelles in which several oxidative reactions requiring consumption of molecular oxygen occur. These reactions form toxic H_2O_2, which is largely detoxified owing to the high concentration of catalase.

Mitochondria differ significantly from other cell organelles. They are surrounded by two membranes and contain their own DNA. The outer membrane is relatively permeable. Mitochondria can grow and divide; for this they need both their own genetic information and that of the cell nucleus. The mitochondria perform several vital reactions of cellular metabolism such as **the citric acid cycle, the degradation of fatty acids, and gluconeogenesis.**

Furthermore, mitochondria play an essential role in maintaining the energy balance of most cells. In the degradation of energy-rich substrates, such as carbohydrates and fats, hydrogen atoms (reducing equivalents) are transferred to nicotinamide adenosine dinucleotide (NAD) and a flavoprotein in the citric acid cycle. These hydrogen atoms and their electrons are fed into the electron-transport chain of the inner mitochondrial membrane. The electron-transport chain consists of a sequence of oxidation–reduction (redox) systems, through which substrate electrons pass until they reduce molecular oxygen to water (Figure 2.44). In this process the considerable energy of the oxyhydro-

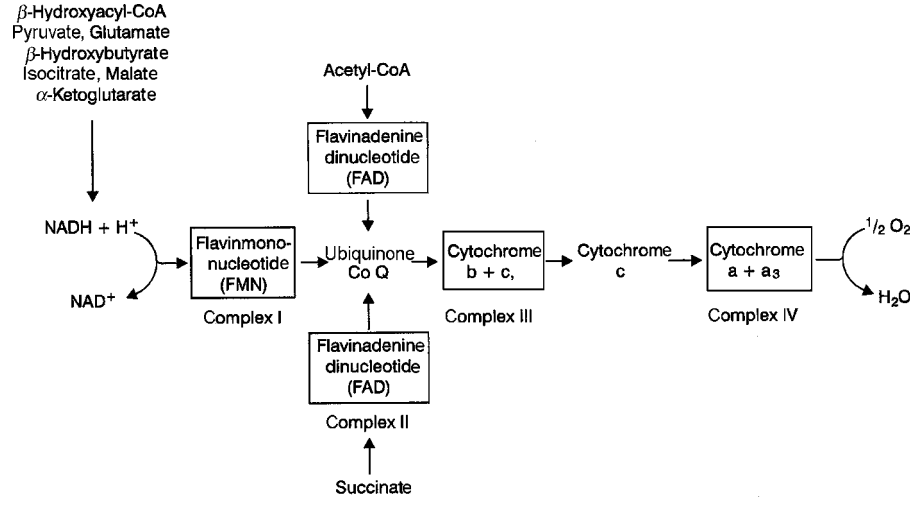


Figure 2.44 *Schematic drawing of the mitochondrial electron-transport chain. The reduction equivalents of the substrates move through three different protein complexes (redox systems) before they are transferred to oxygen. The energy of the redox reactions builds an electrochemical proton potential, which is utilized by the enzyme ATP-synthase for the synthesis of ATP, the most important molecular form of cellular energy.*

gen intermediate formed during this reduction is fractionated and finally used for the synthesis of energy-rich ATP molecules. This process has been termed **oxidative phosphorylation.**

The cell nucleus is enclosed by the nuclear envelope, a double membrane with numerous pores. Compounds may exchange between the nucleus and the cytoplasm through these pores. The sequence of the four bases of DNA (deoxyribonucleic acid) in the nucleus determines the genetic information which is packaged into segments of hereditary factors which are termed the 'genes'. DNA covered with proteins is termed chromatin and represents the basic structure of the chromosomes. The somatic cells of the human body contain 46 chromosomes. The information of small DNA sequences is transcribed (**transcription**) into so-called messenger ribonucleic acid (**mRNA**) and subsequently translated (**translation**) into the amino acid sequence of distinct proteins at the ribosomes. Some cell types lose the nucleus prior to their maturation (erythrocytes, thrombocytes) and cannot divide or synthesize RNA.

2.4.3 Cellular Targets of Toxic Chemicals

Environmental chemicals react with very different components of the cell. Although only single mechanisms of toxicity will be discussed, it is important to remember that in almost every case toxic chemicals act simultaneously on several key functions of the cell. The simultaneous disturbance of several cell functions can cause a synergistic increase in toxicity. For example, this may be the case if cellular structures and repair processes are affected.

Cell Membrane

Chemicals may affect membrane fluidity and the function of membrane proteins.

Cellular membranes contain many important metabolic enzymes that deliver their products specifically into distinct compartments of the cell. Xenobiotics may affect both the permeability of cellular membranes and the activity of the enzyme proteins in the membranes. The disturbance may be caused by either radical attack as well as covalent bonding of a reactive metabolite or by a noncovalent interaction of the xenobiotic with membrane components, e.g. lipid phase, proteins. In the case of lipid peroxidation the membrane is destroyed (see Section **Lipid Peroxidation**, below). A further mechanism of damage involves the oxidation of functional thiol groups of membrane proteins (see Section **Disturbance of Redox Systems**, below).

Depending on their physicochemical characteristics lipophilic and amphipathic chemicals may concentrate in cellular membrane systems and affect the package density and mobility of the membrane lipids. Chemicals may increase or decrease the mobility of the fatty acid chains and, as a consequence, the membrane may become more rigid or fluid and the activity of distinct membrane enzymes and transport proteins may change. Very high concentrations of lipophilic chemicals can cause a complete disintegration of cellular membranes.

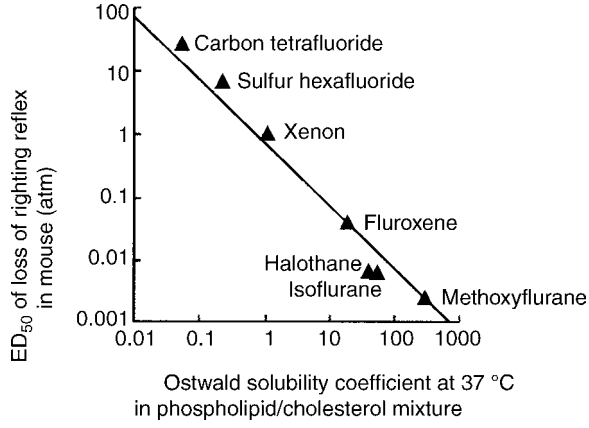


Figure 2.45 *Comparison of the narcotic effect and the solubility of chemicals in artificial lipid membranes, which show similar phospholipid/cholesterol ratios to those of a neuron membrane. Reprinted in modified form from Miller, KW (1979), Weekly Anesthesiology Update,* **2**, *1–8. Copyright (1979), with permission from Princton University Press.*

Historically, the narcotic action of various narcotics and solvents was thought to be due to their interaction with the lipid phase of the plasma membrane of neurons. This notion was supported by the correlation of narcotic effectiveness of these lipophilic chemicals with their distribution coefficient in octanol/water or a phospholipid–cholesterol/water mixture (Figure 2.45). Recent studies, however, show that the narcotic action is most likely caused by a direct interaction between lipophilic chemicals and the hydrophobic part of channel proteins in the membrane.

A direct interaction of chemicals with membrane proteins underlies the action of DDT, pyrethroids, and some other insecticides (Figure 2.46). These chemicals bind to a hydrophobic section of sodium channels and prevent channel closure after an action potential. Consequently, the chemicals cause over-excitability at low concentrations and paralysis at high concentrations.

Lipid Peroxidation

Oxidative attack on unsaturated membrane lipids results in lipid peroxidation as a result of which ethane and pentane are released and can be determined in exhaled air.

In contrast to the interactions of chemicals with membranes described above, components of the membrane may be severely damaged during lipid peroxidation. Oxidative destruction of unsaturated membrane lipids has been studied primarily in liver (Figure 2.47). In a monooxygenase-dependent reaction CCl_4 is cleaved homolytically to yield a chlorine ion and a trichloromethyl radical ($CCl_3^{\bullet}$) in the liver cell; the free radical can react with unsaturated lipids, via H-abstraction, to produce a lipid radical and $CHCl_3$. The weakly bound hydrogen of the methylene group of the divinylmethane grouping of multiple unsaturated lipids is particularly prone to H-abstraction, allowing generation of a

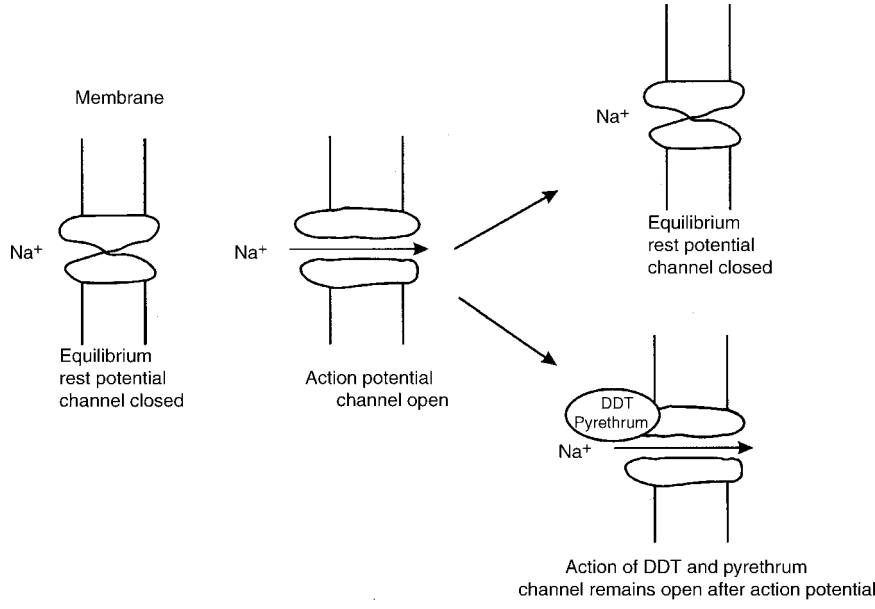


Figure 2.46 *Effect of DDT and pyrethrum on sodium channels of neurons. Under physiological conditions sodium channels are closed immediately after an action potential. In contrast, the sodium channels remain open for a longer period after binding the insecticide.*

pentadienyl radical, which bonds with oxygen. It is likely that the ultimate location of the free electron is at the second oxygen atom. The result of the uptake of the oxygen by the isolated double bond is the energetically driven formation of a conjugated series of double bonds. The lipid peroxyl radicals can now react with unsaturated fatty acids to generate lipid hydroperoxides and new fatty acid radicals via a chain reaction. The final products of the destruction of membrane lipids include inter alia malonic dialdehyde, alkanes, and alkenes. These products can be easily determined and are used for the detection of lipid peroxidation.

Lipid peroxidation is not only produced by strong oxidants but may take place in principle whenever the redox balance of the cell is disturbed (see Section ***Disturbance of Redox Systems***, below). The evidence that a chemical induces lipid peroxidation does not necessarily mean that this process plays a central role in its cytotoxic action. Often it is not possible to decide whether lipid peroxidation is a primary cause or only a consequence of cell damage. In the case of CCl_4, however, lipid peroxidation appears to be causally connected with the toxicity of the halogenocarbon.

Disturbance of Redox Systems

Disturbance of cellular redox homeostasis may lead to cytotoxicity by various means.

The cell contains various redox systems, such as the hydrogen-transferring coenzymes NAD, NADP, and riboflavin, the substrate–product pairs lactate/pyruvate, acetaldehyde/

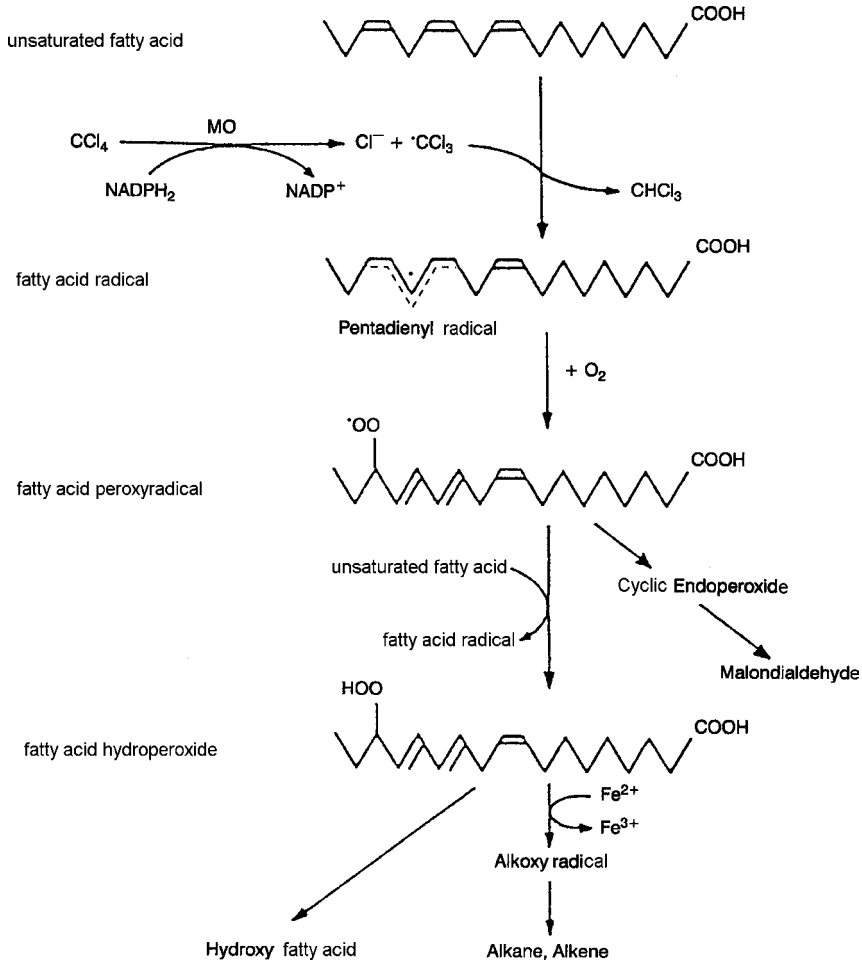


Figure 2.47 *Schematic representation of lipid peroxidation induced by CCl_4. MO = Monooxygenase.*

acetic acid, and ascorbic acid/dehydroascorbic acid, the glutathione system, and protein thiols. Enzymes catalysing one-electron transitions can change the redox state of these paired molecules. Both the concentration and the relative proportions of oxidized and reduced coenzymes and substrates determine the redox state of the cell. Its continued disturbance has serious effects on the vitality of cells as has been demonstrated by

Table 2.21 *Possible consequences of oxidative stress.*

Decrease of the GSH/GSSG ratio
Lipid peroxidation
Oxidation of protein thiols and loss of the activity of (enzyme) proteins
DNA-strand breaks

numerous *in vitro* studies (Table 2.21). The significance of some of the following mechanisms for toxicity in vivo has not been sufficiently clarified yet.

The disturbance of cellular redox homeostasis induced by the redox cycling of chemicals, so-called **oxidative stress,** has gained particular interest. During redox cycling, catechols, quinones, nitroaromatics, or aromatic azo dyes are reduced due to one-electron transitions and may subsequently react with oxygen, producing a superoxide radical anion (Figure 2.48). Superoxide radical anion is cytotoxic and genotoxic. These effects, however, are most likely not due to the superoxide radical anion itself but to other toxic oxygen species derived from superoxide, such as $H_2O_2^{\bullet}$, OH, and $^1O_2$. The cell protects itself against reactive oxygen species and their cellular products, such as lipid peroxides, by means of enzymes such as superoxide dismutase, catalase, and glutathione peroxidase as well as antioxidants such as β-carotene and α-tocopherol. If these systems are overwhelmed, the reactive oxygen species may damage the cell directly by an oxidative attack or indirectly by the disturbance of important redox systems.

The cellular targets of reactive oxygen species are lipids, proteins, and DNA. Lipid peroxidation, as exemplified by CCl_4, has been described above. Similarly, the reactive trichloromethyl radical may also directly react with the double bond of multiple unsaturated lipids and, thereby, initiate the destruction of lipids. In proteins, thiol groups are preferred targets of toxic oxygen species. Frequently enzymes have reduced SH-groups, which are essential for their catalytic activity. Thus, the oxidation of thiol groups is often accompanied by a loss of protein function (Figure 2.49). Thus, the inhibition of the catalytic activity of Ca^{++} transport proteins of the plasma membrane and the endoplasmic reticulum may play a role in cell damage because these carrier proteins contribute to the maintenance of a low concentration of cytoplasmic calcium. Furthermore, nonoxidative mechanisms, such as thiol acylation and arylation by reactive chemicals, can produce a loss of protein function (Chapter 4) (Figure 2.49). Finally, reactive oxygen species can produce cytotoxicity and arrest cell growth by reacting with DNA to cause **DNA-strand breaks**.

H_2O_2 is detoxified by either catalase or glutathione peroxidase. GSSH generated during the glutathione peroxidase reaction is in turn reduced to GSH by the enzyme glutathione reductase at the expense of $NADPH_2$, and the formation of $NADP^+$. When the formation of $NADPH_2$ becomes rate limiting, the cellular concentrations of $NADP^+$ and GSSG increase.

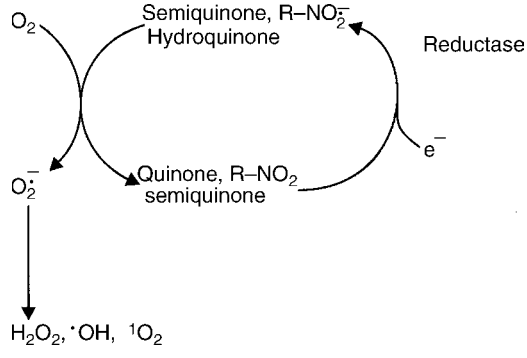


Figure 2.48 *Formation of the superoxide radical anion and further reactive oxygen species by redox cycling of chemicals. Examples of reductases are NADPH-cytochrome P450 reductase, aldehyde reductase, and ketone reductase; R-NO_2 represents an aromatic nitro compound.*

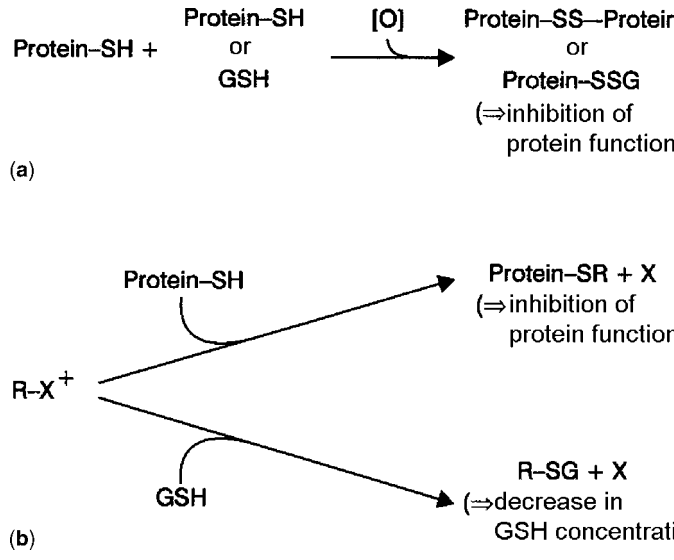


Figure 2.49 *Reactions of reactive oxygen species [O] (a) or electrophilic chemicals ($R\text{-}X^+$) (b) with GSH and protein thiols.*

Damage to both redox systems can contribute to cytotoxicity in several ways. Some important cellular functions are very sensitive to changes in the GSH/GSSG ratio. If the portion of oxidized glutathione increases, protein synthesis and cell division may be inhibited. The cell protects itself against a disturbance of thiol redox status by the active export of GSSG from the cell. Under extreme conditions this can lead to a decrease in glutathione concentration. Moreover, the concentration of reduced glutathione can decrease under a critical value by a change in the thiol redox ratio that can result from the extensive conjugation of electrophilic chemicals or their metabolites with GSH (Figure 2.50). This can cause a loss of vitality or cell death, since the capacity of the GSH system is no longer sufficient to inactivate reactive, electrophilic chemicals and toxic oxygen species as well as their by-products.

Another factor that may contribute to cell damage is an increase in the concentration of oxidized pyridine nucleotides in the cytosol and the mitochondria. Elevation of the intramitochondrial oxidized pyridine nucleotides, at the expense of reduced pyridine nucleotides, may result in an excessive efflux of mitochondrial Ca^{++} into the cytosol, a key event that may lead to cell death.

Inhibition of ATP-formation

Inhibitors of ATP-formation cause arrest of cellular metabolism.

Most of the cells of the body form their ATP (adenosine triphosphate) via both **glycolysis and mitochondrial oxidative phosphorylation** (Figure 2.44). In principal,

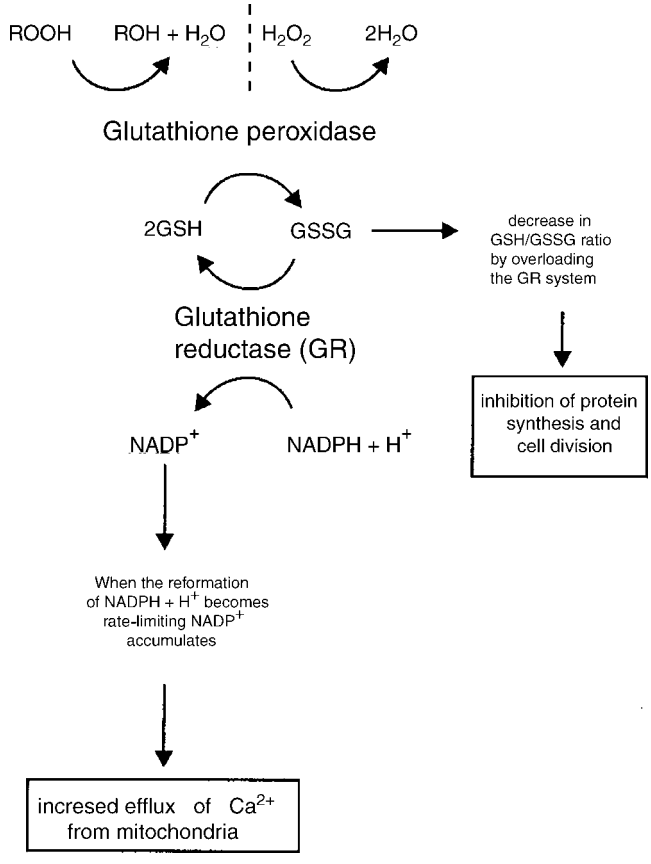


Figure 2.50 *Possible consequences of a too great strain on the systems metabolizing H_2O_2 and hydroperoxides. ROOH = organic hydroperoxides; ROH = alcohols; GR = glutathione reductase.*

chemicals can inhibit glycolysis as well as mitochondrial ATP synthesis (Tables 2.18 and 2.19). Cyanide poisoning is a well known example of a blockage of mitochondrial electron transport. Cyanide binds to the central Fe^{3+} of the last component of the respiratory chain, cytochrome a/a_3, and, thereby, inhibits the electron flow to oxygen.

A second important mechanism for the damage to the mitochondrial ATP-generating system is the uncoupling of ATP synthesis from electron transport (Figure 2.51). According to the 'chemiosmotic hypothesis' the energy of the redox reactions that accompany electron transport produces an electrochemical proton potential at the mitochondrial inner membrane, an essential prerequisite for which is that the inner membrane of the mitochondria is impermeable to protons and hydroxide ions. The so-called proton-motive force that is generated is used for the synthesis of ATP. Uncouplers cause a passive transport of protons through the inner membrane and, thereby, decrease the electrochemical proton potential. With the loss of the proton-motive force, the mitochondria lose their capability to synthesize ATP.

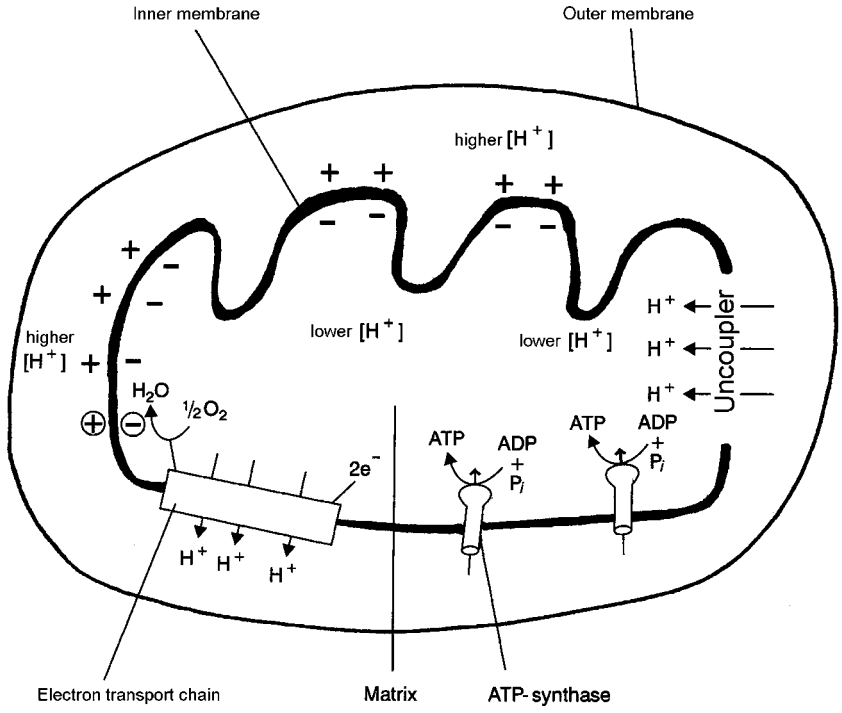


Figure 2.51 *Schematic and simplified drawing of the chemiosmotic hypothesis and the action of uncouplers. During mitochondrial electron transport, H^+-ions are pumped through the inner mitochondrial membrane, which is basically impermeable to protons, generating an electrochemical proton potential. This potential is the driving force for the synthesis of ATP (protonmotive force). Uncouplers increase the permeability of the inner mitochondrial membrane for protons and thus cause the breakdown of the electrochemical proton potential and a disturbance of the ATP-synthase.*

Pentachlorophenol and polychlorinated biphenyls have been shown to be uncouplers *in vitro*. However, it is not clear whether toxicity of these chemicals in humans is also related to their effect on mitochondrial energy metabolism.

Total inhibition of the cellular energy supply, derived from glycolysis and mitochondrial ATP synthesis, for a sufficient period of time causes the death of the cell. All energy-dependent cellular metabolic reactions and active processes that are responsible for the maintenance of homeostasis in the cell will cease. In contrast, partial inhibition of the energy supply will initially cause a decrease in cellular metabolism, but not necessarily cell death. A consequence, however, of decreased repair capacity is that damage to cell structures can be synergistically increased. It is not sufficiently clear whether the ATP that is formed from glycolysis and oxidative phosphorylation is compartmentalized and, therefore, whether the inhibition of the two pathways of ATP synthesis have different consequences and meanings.

Inhibition of the Synthesis of Macromolecules

Inhibition of the synthesis of macromolecules can have various causes.

DNA, RNA, and protein synthesis are complex processes that proceed only in the presence of metabolic energy, protein factors, t-RNA (transfer-RNA), ribosomes, enzymes, and various anabolic pathways. Hence, it would be expected that disturbances in macromolecular synthesis is a frequent effect of cytotoxic chemicals. A decreased synthesis of macromolecules, however, may cause an altered steady state of cellular metabolism and may not necessarily lead to cell death. Even temporary complete inhibition of protein synthesis may not cause cell death. For example, cells may readily survive a short-term treatment with the antibiotic cycloheximide despite the inhibition of protein synthesis.

2.4.4 Mechanisms Underlying Cell Death

Apoptosis is an active process leading to cell death via an ordered sequence of events.

The cellular events that ultimately lead to cell death are an area of active research and are incompletely understood. In principle, cell death may be caused by the accumulation of injurious events resulting in loss of the activity of many cellular functions. Alternatively, cell death may be the consequence of active processes. The first scenario may contribute to distinct forms of necrosis; the second may be the case in apoptosis, which is an active and often physiological form of cell death having characteristic morphologic features (see Section ***Morphology of Necrosis and Apoptosis***, below) achieved via specific biochemical mechanisms. Apoptosis plays an important role in the elimination of embryonic tissue, the regulation of the size of organs, and the elimination of precancerous cells. Chemically induced cell death resulting from either necrosis or apoptosis depends on the dose of the cytotoxic chemical used. For example, diethylnitrosamine induces apoptosis in the liver at low doses, whereas it causes necrosis at higher doses.

Two general pathways leading to apoptosis have been described. One originates from specific 'death receptors' at the cell membrane and the interaction with their ligands (Figure 2.52). The other involves the mitochondrial compartment and the release of cytochrome c. In both pathways caspases, i.e. cysteine-containing proteases, which are characterized by their ability to split peptides or proteins at aspartate residues, play a central role. Recently, various 'death receptors' that belong to the tumor necrosis factor (TNF) receptor gene superfamily have been characterized. Fas, otherwise known as CD95 or APO1, is a cell surface receptor that belongs to the death receptor family. Ligand binding to Fas causes three of these receptor molecules to form trimers, which can then attract FADD (Fas-associated death domain) which, in turn, binds caspase 8 via the so-called 'death effector domain (DED)'. The result is the activation of the caspase, which then activates other caspases downstream, ultimately leading to activation of so-called 'executor caspases' such as caspase 3, which cleave essential cellular proteins leading to the death of the cell (Figure 2.52). Distinct cytotoxic chemicals may induce apoptosis via this pathway. One example is bleomycin, which induces the expression of the CD95 receptor and its ligand in a hepatoma cell line.

In the other apoptotic pathway the mitochondria of the cell are compromised. Cytochrome c, a component of the electron-transport chain that is localized on the outside of the inner mitochondrial membrane, is released into the cytoplasm (Figure 11) and participates, with ATP, in the apoptosis protease activating factor 1

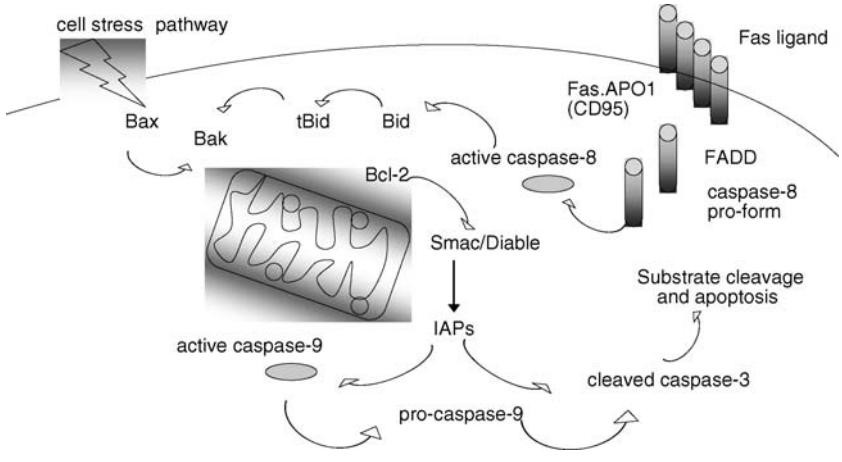


Figure 2.52 *Schematic drawing of two general pathways leading to apoptosis, one originating from 'death receptors' such as the CD95 receptor recruiting and activating caspases. This process is mediated by adapter proteins such as FADD (Fas-associated death domain). Similarly the other pathway leads to the activation of caspases but it originates from disturbed mitochondrias releasing cytochrome c. The cytochrome c forms, together with caspase 9, ATP, and Apaf-1 (apoptosis protease activating factor-1), an 'apoptosome' that catalyses the activation of the caspase. There are numerous other proteins that play a role in apoptosis that are not included in the Figure.*

(Apaf-1)-mediated activation of caspase 9. The result is the onset of a caspase cascade ultimately leading to cell death. If the loss of cytochrome c from mitochondria is massive, necrosis, but not apoptosis, is induced, owing to an interruption of the mitochondrial electron-transport chain and consequently of ATP production.

During the apoptotic process various pro-apoptotic and anti-apoptotic proteins play a regulatory role. In this context the members of the bcl-2 multigene family have attracted special attention. Thus, bcl-2 and bcl-x_L exert anti-apoptotic activity, whereas bax, bad, and bcl-x_s promote apoptosis. Most likely the relative concentrations of anti-apoptotic and pro-apoptotic proteins determine whether the cell lives or dies. The binding of bcl-2 or bcl-x_L to bax results in mutual inactivation of their biological activities.

Morphology of Necrosis and Apoptosis

The morphologic appearance of apoptosis is very different to that of necrosis.

Apoptosis shows very distinct morphologic changes in the nuclear area. In particular the condensation of the chromatin, which has been termed pyknosis, is prominent; other cell organelles remain largely intact. In general a progressive contraction of cellular

volume occurs. Subsequently, apoptotic bodies are formed, which undergo phagocytosis by other cells such as macrophages.

Apoptosis is a complex system requiring the interaction of oasis proteins that have pro-apoptotic or anti-apoptotic actions. The cristae (invaginations) of the mitochondria vanish; flocculent and crystalline materials consisting of denatured proteins and hydroxyapatite accumulate in the mitochondria. In contrast to apoptosis, the nucleus usually does not show significant changes and only in special cases does caryolysis, i.e. disintegration of the nucleus, take place. The final stage of necrosis is characterized by widespread disintegration of cellular membranes and the appearance of the organelles in the extracellular fluid. The results are prompt immunologic reactions and inflammation that do not occur in apoptosis, since the plasma membrane remains intact.

The Role of Disturbed Calcium Homeostasis in Cell Death

An unphysiologic increase of cytosolic calcium concentration can either cause necrosis or contribute to apoptosis.

Under physiologic conditions the calcium concentration in the cytosol amounts to only 50–300 nmolar versus about 1 mmolar in the plasma. This high concentration gradient is maintained by various mechanisms (Figure 2.53):

1. Ca^{2+}-ATPases and Na^+/Ca^{2+}-exchange transport systems, which continuously pump calcium out of the cell;

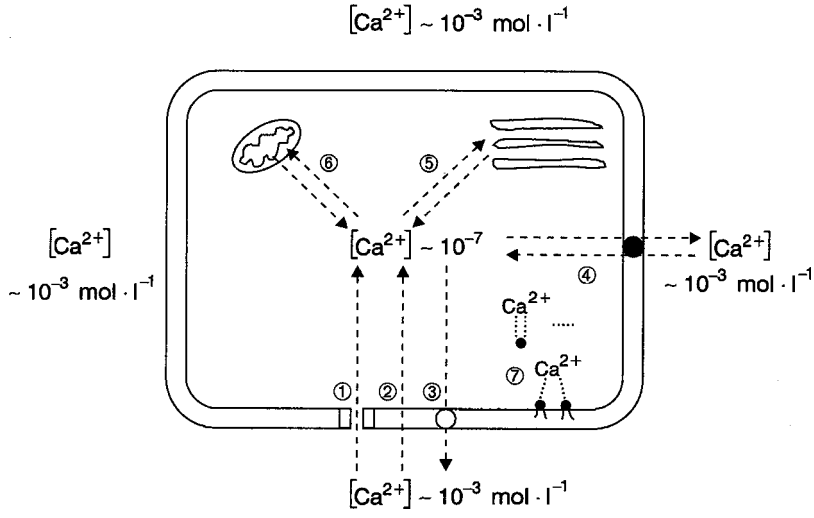


Figure 2.53 *Schematic drawing of the processes, that affect or regulate the concentration of free calcium in the cell. 1. Calcium channels; 2. Permeability of the cell membrane; 3. Ca^{2+}-ATPases (transport proteins); 4. Na^+/Ca^{2+}-exchange systems (transport proteins); 5. Sequestration into the ER by Ca^{2+}-ATPases as well as release from the ER; 6. Release and uptake by mitochondria (energy-dependent uptake systems as well as Ca^{2+} /H^+-exchange systems); 7. Binding to proteins and anions as well as to the polar head groups of membrane lipids.*

2. transport proteins, which catalyse the uptake of calcium into intracellular compartments, i.e. endoplasmic reticulum and mitochondria;
3. binding of calcium to molecules such as the polar head groups of phospholipids.

The free calcium concentration in cytoplasm plays a central role in the regulation of many essential cell functions such as growth, differentiation, secretion, change of cell shape, and transfer of hormonal signals. Exact control of calcium homeostasis is, therefore, of vital importance for the cell. As already described in Section ***Disturbance of Redox Systems*** transport proteins can be targets of toxic chemicals. For example, calcium transport proteins contain functional SH-groups that can be oxidized as well as alkylated or arylated. The loss of the ability to pump calcium either out of the cell and/or into intracellular storage sites becomes critical under conditions where other mechanisms cause an increase of cytoplasmic calcium. In this context, it is important to remember that damage of the cell membrane results in increased influx of calcium into the cell, and that oxidation of pyridine nucleotides (Section ***Disturbance of Redox Systems***) leads to increased efflux of calcium from the mitochondria. Furthermore, inhibition of ATP synthesis decreases the activity of transport proteins. All of these mechanisms result in unphysiological increases of cytoplasmic calcium concentration, which induce cell

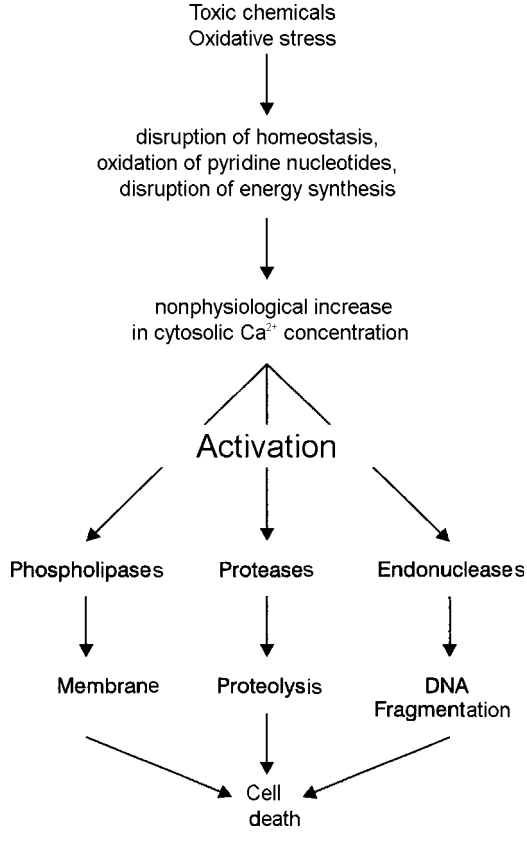


Figure 2.54 *Possible contribution of calcium activated catabolic processes to cell death.*

death (Figure 2.54). Thus, Ca^{2+}-stimulated activation of phospholipase A_2 enzymes leads to the formation of lysophosphatides and release of arachidonic acid. The altered turnover of phospholipids and the formation of toxic lysophosphatides can compromise the integrity of the plasma membrane. Activation of proteases by calcium can lead to the degradation of essential cell structures, e.g. proteins of the cytoskeleton are substrates of calcium-activated proteases. Last but not least, calcium-dependent endonucleases may contribute to cell death by the fragmentation of DNA.

2.4.5 Summary

All cell structures and functions may be targets of cytotoxic chemicals. Damage may be due to covalent or noncovalent interaction with the xenobiotic. The consequences of the damage depend on the structure and function of the affected cellular component and the cell's ability to repair the damage. Important cellular targets of toxic chemicals are membrane systems, the generation of metabolic energy, the synthesis of critical macromolecules, and the control of redox homeostasis. The plasma membrane and intracellular membrane systems create specific reaction compartments due to their selective permeability, which is a prerequisite of the ordered metabolism of the cell. Damage to the integrity of the cell membrane results in serious consequences because enzymes may be inhibited, the fluidity of the lipid phase of the membrane may be altered, and lipid peroxidation may ensue. An extended disturbance of the synthesis of ATP or macromolecules is not compatible with life, whereas a brief interruption of these central cell functions will cause only a transient alteration in the steady state of the metabolic pathways of the cell. A disturbance of the redox homeostasis of the cell can be caused by the redox cycling of chemicals. The resulting inactivation of enzyme proteins due to the oxidation of their functional thiol groups may lead to cell death.

Cell death induced by cytotoxic chemicals may take the form of necrosis or apoptosis, depending on the cellular context and the dose of the toxic chemical. Necrosis is characterized by an increase in cell volume, marked changes in cell organelles, and ultimately by a disintegration of cell membranes. Owing to the emergence of cell constituents in the extracellular fluids immunological reactions and inflammation occur. In contrast, apoptosis is an active and well ordered process, which may be induced by the action of 'death receptors'. In apoptosis, cell organelles remain largely intact. There is a progressive contraction of cellular volume and a condensation of the chromatin of the nucleus. The resulting apoptotic bodies are phagocytized and immunological reactions do not occur.

References

Alberts, B. (editor), Molecular Biology of the Cell (1983), Garland Publishing, Inc., New York and London.

Miller, KW. Physical Theories of General Anaesthesia. *Weekly Anesthesiology Update*, **2**, 1–8 (1979).

For Further Reading

Artal-Sanz, M, Tavernarakis, N. Proteolytic Mechanisms in Necrotic Cell Death and Neurodegeneration. *FEBS Lett.*, **579**, 3287–3296, (2005).

Ashton-Rickardt, PG. The Granule Pathway of Programmed Cell Death. *Crit. Rev. Immunol.*, **25**, 161–82, (2005).

Blaise, GA, Gauvin, D, Gangal, M, Authier, S. Nitric Oxide, Cell Signaling and Cell Death. *Toxicology*, **208**, 177–192, (2005).

Carmody, RJ, Cotter, TG. Signalling Apoptosis: a Radical Approach. *Redox Rep.*, **6**, 77–90, (2001).

Lee, AG. How Lipids Affect the Activities of Integral Membrane Proteins. *Biochim. Biophys. Acta.*, **1666**, 62–87, (2004).

Los, DA, Murata, N. Membrane Fluidity and its Roles in the Perception of Environmental Signals. *Biochim. Biophys. Acta*, **1666**, 142–157, (2004).

Orrenius, S, Zhivotovsky, B, Nicotera, P. Regulation of Cell Death: the Calcium-Apoptosis Link. *Nature Rev. Mol. Cell. Biol.*, **4**, 552–565, (2003).

Owuor, ED, Kong, AN. Antioxidants and Oxidants Regulated Signal Transduction Pathways. *Biochem. Pharmacol.*, **64**, 765–770, (2002).

Pinkoski, MJ, Green, DR. Apoptosis in the Regulation of Immune Responses. *J. Rheumatol. Suppl.* **74**, 19–25, 2005.

Segal-Bendirdjian, E, Dudognon, C, Mathieu, J, Hillion, J, Besancon, F. Cell Death Signalling: Recent Advances and Therapeutic Application. *Bull. Cancer*, **92**, 23–35, (2005).

Xu, JJ, Diaz, D, O'Brien, PJ. Applications of Cytotoxicity Assays and Pre-lethal Mechanistic Assays for Assessment of Human Hepatotoxicity Potential. *Chem.-Biol. Interact.*, **150**, 115–28, (2004).

Zimber, A, Nguyen, QD, Gespach, C. Nuclear Bodies and Compartments: Functional Roles and Cellular Signalling in Health and Disease. *Cell. Signal.*, **16**, 1085–1104, (2004).

2.5 Receptor-Mediated Mechanisms

Jens Schlossmann and Franz Hofmann

'Dosis facit venenum'

2.5.1 Introduction

Receptors are components of an organism which bind molecules of diverse chemical structures. These molecules are ligands that activate or inhibit the receptor function and thereby elicit a physiological response. Ligands that activate a response are agonists; those that block the response are antagonists.

The classical concept of receptors was developed by John Newport Langley and Paul Ehrlich. Receptors can be considered to be locations within the body with which drugs can react to provoke a biological response. Some drugs act in a generalized way without reacting with a specific biological site, e.g. osmotic diuretics increase urine flow by increasing the osmotic pressure within the kidney tubule, thereby preventing the reabsorp-

tion of water. Most drugs, however, act at localized sites to produce specific responses. It was recognized in the mid 19th century that there must be some relationship between their chemical structures and the responses that they induce. As early as 1878 John Newport Langley suggested that atropine and pilocarpine altered secretion of saliva in dogs by reacting with physiological substances within the gland, and in 1905 in his report on the effects of nicotine and curare on muscles he described their locus of action as a 'receptive substance'. Independently, Paul Ehrlich, in his pioneering studies in immunology, suggested that the specificity of immunological interactions required an interaction of specific chemical structures to form antibodies. He formulated the sentence 'Corpora non agunt nisi fixata' (Substances do not act unless bound).

Today we know that receptors are rather large proteins located at specific sites on or within cells, at which chemicals (agonists) react to produce responses. When other chemicals react at the receptor to inhibit the effect of the agonists we call them antagonists. The agonist is also often called the ligand. The receptor–ligand interaction follows the law of mass action and its kinetics are similar to the Michaelis–Menten equilibrium, except that the products of the Michaelis–Menten type of interaction are metabolites, whereas interactions of the agonist at the receptor usually do not result in a change of chemical structure of the agonist. The reversible binding of an inhibitor to the agonistic receptor site leads to competitive antagonism. The allosteric inhibition of the agonist action results in noncompetitive inhibition. These effects on the receptor action can be measured and defined by a dose–response curve. Therefore, only agonists or antagonists in sufficient concentrations based on the dose–response curve can modulate receptor-mediated functions.

2.5.2 Ligand–Receptor Interactions

Classes of receptors are hormone-, neurotransmitter-, growth factor-, and cytokine-receptors, ion channels, carriers, enzymes, transcription factors, structural proteins, lipids, mRNA, or DNA. Most ligands bind to protein-receptors. The specific binding of a ligand at its receptor is a prerequisite for its action and is based on the law of mass action. The interaction between a ligand and its receptor can trigger a cascade of events.

The term receptor usually refers to receptors that are membrane bound, or cytosolic and respond to hormones, neurotransmitters, growth factors, or cytokines. As suggested above, there are a few nonreceptor-mediated mechanisms (e.g., acid–base neutralization, osmosis, high or low temperature, alteration of the redox potential of a cell, exhaustion of essential cellular molecules).

The ligand–receptor interaction is based on the law of mass action, which was first applied to this interaction by Alfred Joseph Clark in the 1920s. The kinetic analysis is derived from the proposal of Michaelis and Menten for the kinetics of enzymatic reactions. The traditional receptor theory was developed for the ligand-binding to membrane-bound hormone- and neurotransmitter-receptors. These receptors were

thought to be metabolically inactive. Therefore, this interaction was mathematically described as a bimolecular reaction; [Equation (2.56)];

$$\text{L (Ligand)} + \text{R (Receptor)} = \text{L}{-}R\,\text{(Ligand}{-}\text{Receptor complex)} \tag{2.56}$$

Based on this theory, the receptor itself is inactive and ligand binding leads to the active conformation which transduces the signal. The ligand binding is determined by binding curves from which the amount of bound ligand at a given concentration can be calculated. Rearrangement of the equation to [Equation (2.57)] yields the dissociation constant K_D, which defines the affinity of the ligand to the receptor. The value of K_D is often between 0.1 and 1000 nM. A smaller K_D value represents a higher receptor affinity.

$$[\text{L}] \times [\text{R}]/[\text{L}{-}\text{R}] = k_{-1}/k_{+1} = K_\text{D} \tag{2.57}$$

In the latter half of the 20th century a receptor model evolved which provides a better understanding of ligand–receptor mechanisms. The basis of this model is an inactive conformation of the receptor R and an active conformation R* (Leff, 1995) (Figure 2.55).

Binding of the ligand to its receptor can induce different reactions:

a) The ligand binds with high affinity to the active conformation R* and shifts the equilibrium between R and R* towards R*. This increases receptor signalling and is characteristic of receptor binding by an agonist.
b) The ligand binds with identical affinity to R and R*. It does not change the existing equilibrium between inactive (R) and active (R*) receptor, but occupation of the binding site prevents binding of the endogenous ligand and, thereby, activation of receptor signalling. This type of binding is characteristic of an antagonist. (Note that some antagonists react covalently with, and inactivate, the receptor.)
c) The ligand binds with slightly higher affinity to the R* than the R conformation and prevents binding of the endogenous ligand. This type of binding is characteristic of a partial agonist.

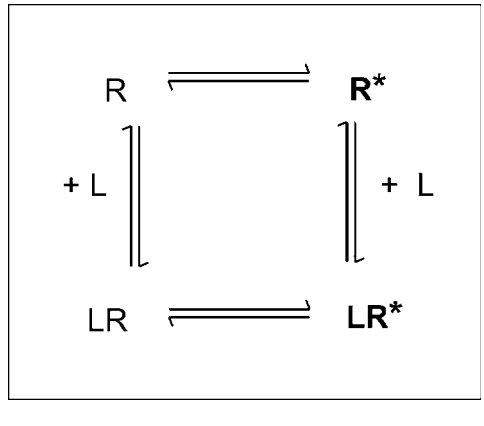


Figure 2.55 *Two-state model of ligand (L)–receptor (R) interaction.*

d) The ligand binds with high affinity to the R conformation and shifts the equilibrium towards the R conformation. This type of binding is characteristic of an inverse agonist.

Thus, we can define the following types of ligands based on binding affinity: the agonist, an antagonist, a partial agonist, and an inverse agonist.

Biological Consequences of Ligand–Receptor Interactions

Toxic agents can act as agonists or antagonists of receptors. The effect of ligand binding upon the receptor activation is termed efficacy. Various types of agonists can be differentiated. Full agonists elicit a maximal tissue response upon receptor binding. In contrast, partial agonists elicit a submaximal response.

Agonistic Effects The 'receptor occupancy theory' has been useful for helping to interpret the results of studies of ligand–receptor interactions. The theory suggests that the magnitude of the response is directly related to the number of receptor sites occupied. Thus, in the absence of ligand there is no response, when 50% of the receptors are occupied we observe 50% of maximal activity, and when all receptors are occupied there is maximal response. These responses over a range of doses give rise to the dose–response curve in its various forms. Empirical examination of receptors often demonstrates that it is not necessary for all receptors to be occupied by agonists to elicit a maximal response, which suggests the presence of spare receptors.

The agonistic effect of a ligand is measured by the dose–response curve seen in Figure 2.56, which permits the measurement of the maximal response of the ligand, the dose needed to evoke 50% of a maximal response (ED_{50}), and the dose required to produce any other fractional response. In most situations the ED_{50} is numerically

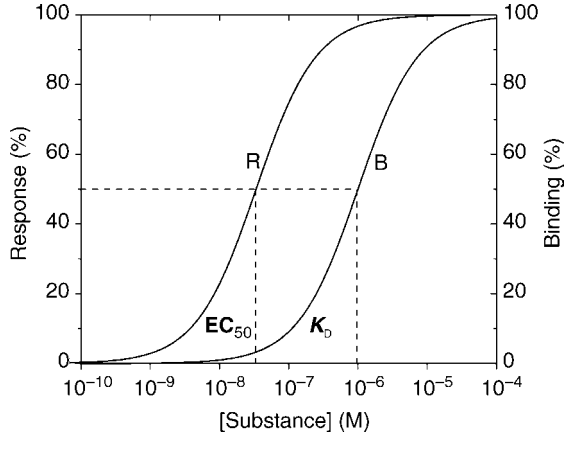


Figure 2.56 *Dose–response curve demonstrating the correlation between Binding (B) and Response (R) of a substance at 90% spare receptor action (K_D: 1 μM; EC_{50}: 50 nM). If the receptor is known, EC_{50} is replaced by K_a.*

equal to the K_a, the activating concentration of an agonist. The availability of these values permits us to compare the potencies of different ligands. Some full agonists are capable of eliciting a maximal response at low occupancy of the available receptor binding sites because the investigated response is already maximal when, e.g., 20% of the receptors are activated, which indicates the availability of, but not the requirement for, the spare receptors.

Antagonistic Effects Various subclasses of antagonists can be distinguished by dose–response curves. The inhibitory concentration of an antagonist is standardized by an ID_{50} value or, if the receptor is known, by a K_i value. Antagonists that bind reversibly to the agonist binding site on R and R* are competitive antagonists. Raising the concentration of the agonist at a given concentration of a competitive antagonist restores the maximal tissue response because the agonist competes with the antagonist for the binding site of R*. In contrast, the maximal response is reduced by an antagonist that binds irreversibly to the agonist receptor interaction site. Furthermore, the maximal response of the agonist is also diminished by a noncompetitive antagonist, which binds to a second site on the receptor that interacts allosterically with the agonist receptor interaction site.

Replacement of a physiological agonist by a competitor depends on its relative affinity to the receptor and its concentration. For example, replacement of the physiological ligand from the receptor by a compound of 1000-fold lower affinity requires a compound concentration that is 1000-fold higher than that of the physiological ligand. This demonstrates that information on the relative binding affinities of the compounds and their concentration in the organism are important criteria to evaluate possible effects of compounds.

When investigating the dose–response relationship for toxicity it is possible to select from a variety of endpoints. Historically, determination of lethality involved treating groups of animals with different doses of a chemical, plotting the dose–response curve, as discussed above, and determining the dose that killed 50% of the animals, i.e. the LD_{50}. Although the method was highly accurate and precise for a given group of animals of the same strain, the data could not always be transferred to other strains of mice, rats, or other species. Furthermore, it required many animals. More practical methods are available today, which provide a reasonable estimate of the lethality of chemical and require few animals.

When studying drugs it is often important to estimate the potential for adverse effects by comparing the effective dose range with the range of doses likely to cause some form of toxicity. In the past a term called the 'therapeutic index' was used and defined as the ratio LD_{50}/ED_{50} in animal studies. Obviously, the higher the value, the safer the drug. However, one must be certain that there is no overlap between the efficacy and lethality curves. An alternative would be to use the ratio LD_{01}/ED_{99} (or a value of LD as low as can be determined.) In the clinical situation toxicity endpoints such as nausea, rash, headache, etc,. i.e. effects which may be unpleasant but not necessarily life threatening, are more preferably used.

2.5.3 Receptor-signal Transduction

The receptors are sensing elements, which induce endogenous signal transduction pathways. They are divided in several classes: (Figure 2.57)

1. **G-protein-coupled receptors**
2. **Ion channels**
3. **Enzymes**
4. **Nuclear receptors/transcription factors**

G-Protein-coupled Receptors (GPCR)

These are heptahelical receptors located at the plasma membrane. They represent a large family of about 500 proteins and are expressed by about 5% of all invertebrate genes. Many ligands act via the GPCR. In addition to known and well characterized GPCRs, a group of GPCRs exist that have no defined ligands and functions. These GPCRs

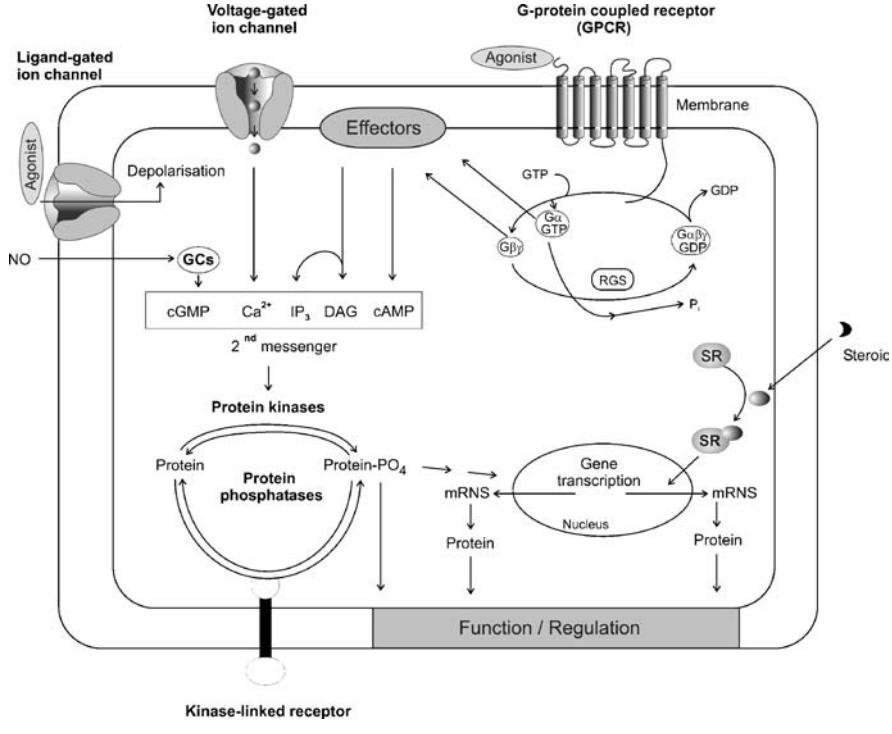


Figure 2.57 *Summary of receptor-mediated cellular regulation by G-protein-coupled receptors (GPCR), voltage-gated ion channels, ligand-gated ion channels, guanylyl-cyclases (GCs), kinase-linked receptors, and steroid receptors (SRs). Further abbreviations: NO, nitric oxide; cGMP, cyclic guanosine monophosphate; IP_3, inositol trisphosphate; DAG, diacylglycerol; cAMP, cyclic adenosine monophosphate; RGS, regulator of G-protein signalling.*

are identified as orphan receptors. Upon ligand binding, a trimeric G-protein is activated. G-proteins consist of three subunits, a Gα-subunit that binds and hydrolyses GTP, a β-, and a γ-subunit (Wettschureck and Offermanns, 2005). The rate of GTP hydrolysis can be accelerated by RGS proteins (regulator of G-protein signalling). The G proteins are separated into the different Gα protein classes: $G\alpha_s$, $G\alpha_i$, $G\alpha_q/\alpha_{11}$, and $G\alpha_{12}/\alpha_{13}$. Activation of these different G-proteins leads to diverse cellular signals. $G\alpha_s$ stimulates the adenylylcyclase which synthesizes the second messenger cyclic adenosine monophosphate (cAMP), $G\alpha_i$ inhibits the adenylylcyclase, $G\alpha_q/\alpha_{11}$ stimulates phospholipase C and, thereby, the synthesis of the second messengers inositol trisphosphate (IP_3) and diacylglycerol (DAG), and $G\alpha_{12}/\alpha_{13}$ activate the small GTP-binding protein Rho.

Ion Channels

These comprise cation-selective or nonselective channels (e.g. for Na^+, Ca^{2+}, K^+) and anion-channels (e.g. for Cl^-), and can be divided into ligand-gated and voltage-gated channels (Catterall, 1996). For example, a ligand-gated channel is the nicotinic acetylcholine receptor which, upon acetylcholine binding, is permeable to sodium ions. The voltage-gated calcium channels are activated by depolarization. Ion channels are large proteins consisting of a channel protein with several trans-membrane segments and often of several additional subunits. The activity of these channels is altered by phosphorylation/dephosphorylation through kinases/phosphatases and by intracellular signals such as Ca^{2+} and ATP.

Enzymes as Receptors

Tyrosine- and serine/threonine-phosphorylation by kinases is a common theme in intracellular signalling cascades. Kinase-linked receptors mediate the actions of a variety of signalling molecules including growth factors (e.g., EGF), cytokines (e.g., erythropoetin), and hormones (e.g., insulin). An important class of these receptors are the tyrosine kinases (RTKs) (Schlessinger, 2000). RTKs are integrated into the plasma membrane, are often dimeric, and comprise a large extracellular ligand-binding and an intracellular effector domain. Intracellular autophosphorylated tyrosine allows the binding of the SH2 domain of an adaptor protein that activates a signalling cascade, usually leading to gene transcription. Guanylylcyclase-linked receptors synthesize the second messenger cyclic guanosine monophosphate (cGMP). They mediate the actions of nitric oxide (NO) and that of distinct peptides, e.g. the atrial natriuretic peptide (Schlossmann et al., 2005).

Nuclear Receptors

These are intracellular proteins that act as transcription factors regulating nuclear-based transcription mechanisms (Gronemeyer et al., 2004). Their ligands include steroid hormones (e.g., gluco- and mineralo-corticoids and sex hormones), thyroid hormones, vitamin D and retinoic acid, lipid lowering and anti-diabetic drugs. The nuclear receptors are assembled in three independent domains: The carboxy terminal domain binds to the ligand, the central domain mediates the specific interaction with hormone-responsive elements in the nuclear DNA located upstream from the regulated gene, and the amino-terminal variable

domain modulates the receptor function and allows the binding of co-transcription factors. The nuclear receptors act as dimers, but they are additionally modulated by other regulating molecules. Activation of the nuclear receptors by the different ligands induces or represses specific gene transcription and, thereby, regulates protein synthesis. The selection of the interacting co-transcription factor depends on the tissue and the conformation of the ligand-bound transcription factor. This conformation may vary with the ligand, as best exemplified by the specific estrogen receptor modulators (SERM). The physiological effects of the nuclear receptors are delayed but occur within hours or days.

The *duration of a receptor signal* is regulated at several levels. Desensitization mediated at the level of transcription, translation, and degradation alters long-term receptor function. A short-term change of receptor activity is induced by phosphorylation through kinases and through endocytosis of membrane-bound receptors. Furthermore, subunits and cofactors of receptors are regulators of receptor function. Therefore, it is essential to define each receptor function in its given physiological background.

References

Catterall, WA. Molecular Properties of Sodium and Calcium Channels. *Bioenerg. Biomembr.,* 1996, **28**, 219–230.

Gronemeyer, H, Gustafsson, JA, and Laudet, V. Principles for Modulation of the Nuclear Receptor Superfamily. *Nature Rev. Drug. Discov.,* 2004, **3**, 950–964.

Leff, P. The Two-state Model of Receptor Activation. *Trends Pharmacol. Sci.,* 1995, **16**, 89–97.

Schlessinger, J. Cell Signaling by Receptor Tyrosine Kinases. *Cell*, 2000, **103,** 211–25.

Schlossmann, J, Feil, R, and Hofmann, F. Insights into cGMP Signalling Derived from cGMP Kinase Knockout Mice. *Front. Biosci.,* 2005, **10,** 1279–1289.

Wettschureck, N and Offermanns, S. Mammalian G Proteins and their Cell Type Specific Functions. *Physiol. Rev.*, 2005, **85**, 1159–1204.

2.6 Mixtures and Combinations of Chemicals

Victor J. Feron and Diana Jonker

2.6.1 Introduction

Toxicity testing and risk assessment have traditionally focused on single chemicals. However, humans are exposed to complex and ever-changing mixtures and combinations of chemicals in the air they breathe, the water and beverages they drink, the food they eat, the surfaces they touch, and the consumer products they use. Insight into the health consequences of such complex exposures is urgently needed to bring safety evaluation of chemicals and real life into harmony. To this end a conceptual framework that defines three types of joint action (viz. dissimilar and similar joint action, and interaction) has been introduced.

In real life, humans are exposed simultaneously or sequentially to large numbers of chemicals via multiple routes. Given this reality, insight into the health consequences

of such complex exposures is needed to judge whether current approaches to safety evaluation of chemicals offer adequate protection to humans. A key question concerning the toxicity of a chemical mixture is whether the toxic effects of a mixture can be predicted on the basis of the toxic effects and the underlying mechanisms of toxicity of the individual components. To answer this question a conceptual framework that defines three types (modes) of joint action (viz. dissimilar and similar joint action, and interaction) has been introduced. In addition to information on the toxicity and mechanism of action of the individual chemicals in a mixture, knowledge about the mode of joint action of the chemicals is often critically important for properly assessing the safety (the potential health risk) of a mixture. The consistent use of terms that describe mixed exposures is also a key issue for understanding the toxicological consequences of such exposures.

In this chapter, the various types of mixed exposures and joint action are defined, designs for mixture toxicity studies are briefly described, and methods for the safety evaluation of mixed exposures are discussed.[1]

2.6.2 Types of Mixed Exposures

In the absence of internationally harmonized terminology, types of mixed exposures should always be defined. The major types are: mixture, simple mixture, complex mixture, combination of chemicals, specified combination of chemicals, and cumulative exposure. These terms are described in this section.

A *mixture* is any combination of chemicals characterized by simultaneity of exposure to the constituents as a result of their joint occurrence. A *simple mixture* is a mixture consisting of a relatively small number of chemicals, say ten or fewer, the composition of which is qualitatively and quantitatively known (for example, a cocktail of three pesticides or a combination of two medicines taken simultaneously). A *complex mixture* is a mixture that consists of tens, hundreds, or thousands of chemicals, and its composition is qualitatively and quantitatively not fully known (for example, diesel exhaust, welding fume, drinking water).

A *combination* is any combination of chemicals regardless of whether or not they occur as a mixture. In a *specified combination* all components are known. An example of a specified four-component combination is exposure to the air contaminants sulfur dioxide and formaldehyde, the food additive butylated hydroxytoluene, and the pain killer acetylsalicylic acid (aspirin). This example involves various independent exposures, via different routes, which may, but do not necessarily fully, overlap in time. *Cumulative* exposure refers to exposure to multiple chemicals and cumulative risk from such exposures.

[1]Combined exposure to initiators (mutagens), promoters, converters, and/or co-carcinogens in chemical carcinogenesis, as well as sequential exposure to toxicants and antidotes, are not discussed in this chapter because such combined exposures are specific areas in toxicology that have been extensively studied and are hardly recognized as examples of combination toxicology.

2.6.3 Types of Joint Actions and Their Role in Safety Evaluation

Three types of joint (or combined) action of chemicals in a mixture or combination have been defined: dissimilar joint action, similar joint action, and interaction. Dissimilar and similar joint action are noninteractive: the chemicals in the mixture or combination do not affect each others' toxicity. With dissimilar joint action the chemicals in the mixture act independently. With similar joint action the chemicals in the mixture induce similar effects because they act similarly regarding primary physiological processes. With interaction one chemical influences the toxicity of one or more other chemicals by direct chemical–chemical interaction (a 'new' chemical is formed), toxicokinetic interaction (alteration in metabolism or disposition of chemicals), or toxicodynamic interaction (interactive effects at target sites). The type of joint action in combination with the margin of safety (= the margin between the actual exposure level and the No-Observed-Adverse-Effect Level) of the individual chemicals in the mixture determines to a high degree whether or not exposure to a mixture is of health concern.

Types of Joint Actions

The toxicity of mixed exposures can be estimated when the toxicity of the individual chemicals as well as types of joint action of the constituents are known.

Joint (or combined) action describes any outcome of exposure to multiple chemicals, regardless of the source or spatial or temporal proximity. Three different types of joint action are distinguished:

Dissimilar joint action (or independent joint action, or simple independent joint action, or response addition) is noninteractive; namely, the chemicals in the mixture do not affect each others' toxicity. In other words, the chemicals act independently, so that the body's response to one of the chemicals is the same regardless of whether or not the other chemicals are present. The modes of action and probably the nature and the site of the toxic effects differ among the chemicals in the mixture. An important characteristic of independent joint action is that a chemical is assumed not to contribute to the (adverse) effect of the mixture when that chemical is present at a level below its own individual (adverse) effect threshold. An example: in a 4-week toxicity study, rats were exposed to a combination of eight different chemicals (sodium metabisulfite, Mirex, Loperamide, metaldehyde, di-n-octyltin dichloride, stannous chloride, lysinoalanine, and potassium nitrite) that were arbitrarily chosen regarding target organ and mechanism of action. The study produced no (convincing) evidence for increased risk from exposure to the combination when each chemical was administered at or below its own No-Observed-Adverse-Effect Level (NOAEL).

Similar joint action (or simple similar action, or dose addition, or concentration addition) is also noninteractive. The chemicals induce similar effects because they act similarly in terms of mode of action and primary physiological processes (uptake, metabolism, distribution, elimination). They differ only in their potency. The toxicity of the mixture can be calculated using summation of the doses of each individual chemical

after adjusting for the differences in potency. The addition of doses implies that toxicity can be expected if the summed dose is high enough to exceed the threshold of toxicity of the mixture, even when the dose level of each individual chemical is below its own effect threshold. An example: in a 4-week toxicity study, rats were exposed to a combination of four different but similarly acting nephrotoxicants (tetrachloroethylene, trichloroethylene, hexachloro-1:3-butadiene, and 1,1,2-trichloro-3,3,3-trifluoropropene). Kidney effects of the mixture were seen at dose levels not showing renal toxicity of the individual compounds. Thus, the study provided support for the assumption of dose additivity for mixtures of similarly acting systemic toxicants under conditions of concurrent, repeated exposure at dose levels below the toxicity thresholds of the individual constituents.

Interaction is characterized by one chemical influencing the toxicity of one or more other chemicals in the mixture, resulting in an effect/response to the mixture that deviates from that expected from (dose or response) addition. Interactive joint actions can be less than additive (commonly termed, for example, antagonistic, inhibitive, infra- or sub-additive) or greater than additive (for example, synergistic, potentiating, supra-additive). Antagonistic and synergistic interaction indicate in which direction a response to a mixture differs from what is expected under the assumption of additivity. An interaction may be based on direct chemical–chemical interaction, toxicokinetic interaction, or toxicodynamic interaction**.**

Direct chemical–chemical interaction means that one chemical directly reacts with another, resulting in a 'new' chemical that might be more or less toxic than the original chemicals. In the case of decreased toxicity, this mechanism is one of the common principles of antidote treatment. An example of increased toxicity is the formation of carcinogenic nitrosamines in the stomach through the reaction of noncarcinogenic nitrite (from food or drinking water) with secondary amines (for example, from fish proteins).

Toxicokinetic interactions involve alterations in the metabolism (biotransformation) or disposition of a chemical. These alterations are often divided in changes in absorption, distribution, metabolism, and excretion. Essentially, toxicokinetic interactions alter the amount of the toxic agent(s) reaching the cellular target site(s) without qualitatively affecting the toxicant–receptor site interaction. With respect to their potential toxicological consequences at low doses, interactions in the process of metabolism (enzyme induction or inhibition) are considered most relevant. The toxicological impact of an alteration in absorption, distribution, or excretion is expected to be small at low dose levels.

Toxicodynamic interactions involve interactive effects occurring at or among target sites. Interactions at the same site resulting in antagonism have been described as receptor antagonism (for example, the antagonistic effect of oxygen on carbon monoxide). Interactions resulting from different chemicals acting on different sites and causing opposite effects have been described as functional antagonism (for example, opposing effects of histamine and noradrenaline on vasodilatation and blood pressure). Another type of toxicodynamic interaction is alteration of a tissue's response or susceptibility to toxic injury (for example, immunomodulation, depletion of protective factors, and changes in tissue repair or haemodynamics).

Finally, an example of synergistic interaction: combined exposure to noneffective doses of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) and the synthetic glucocorticoid hydrocortisone (HC) induced cleft palate, an irreversible structural change, in all exposed

mouse embryos. This marked synergistic interaction was explained by mutual upregulation of the receptors which mediate the biological activity of TCDD and HC (the aryl hydrocarbon receptor and the glucocorticoid receptor, respectively).

Role in Safety Evaluation

Approaches for the safety evaluation (risk assessment) of mixtures and combinations of chemicals will be discussed in detail in the last section of this chapter. Nevertheless, the consequences of using the aforementioned types of joint action as guidance in the process of the safety evaluation are briefly discussed here because insight into these consequences is also of relevance for designing mixture toxicity studies.

Whether a mixed exposure constitutes a safety concern mainly depends on two key factors; one, the type of joint action (similar or dissimilar joint action, or synergistic or antagonistic interaction) and two, the margin of safety (the margin between the actual exposure level and NOAEL) of the individual constituents of the mixed exposure. Moreover, experience has shown that *as a rule* exposure to mixtures at *nontoxic exposure levels* of the individual chemicals in the mixture is of no health concern. However, there are exceptions to this rule. When a mixture consists of chemicals with a *similar mode of action* (and the same target organ) *dose addition* is to be expected and should be taken into account in the safety evaluation of such a mixture. Furthermore, since *synergistic interaction* can never be fully excluded, its likelihood of occurrence (to be estimated on the basis of data on the mechanism of toxicity of the individual chemicals) and its toxicological relevance in the light of *actual exposure levels* should be estimated on a case-by-case basis. If a case of potential health risk emerges, a wide variety of methods for hazard identification and risk assessment is available. Mixtures with very small (or no) margins of safety for the individual chemicals (for instance, heavily polluted outdoor air containing ozone, nitrogen dioxide, sulfur dioxide, ammonia, and ultrafine particles with the lung as a common target organ) are to be regarded as priority mixtures because adverse effects due to combined exposure are likely to occur. When, on the other hand, the margins of safety for the individual chemicals of a mixture are large, say a factor of 5 or 10 or greater, additive or interactive adverse effects are unlikely to occur, and thus, such mixtures hardly need any further consideration.

2.6.4 Designs for Toxicity Studies of Mixtures or Combinations

Factors that influence study design include the number of chemicals in a mixture or combination, the availability of a mixture for testing it in its entirety, the extent to which the toxicity of a mixture or combination needs to be characterized in terms of dose–response relationship or departure from additivity, and available resources. Major designs include whole-mixture studies, full factorial designs, fractional factorial designs, the additivity approach, the isobolographic method, effect/response-surface analysis, and mechanistic methods including physiologically based toxicokinetic modelling. In this section, the various designs are briefly discussed, including their strengths and limitations.

Given the diversity of exposure scenarios, the questions to be addressed in toxicity studies of mixtures and combinations of chemicals vary widely and, consequently, there is no 'one size fits all' design.

In *a whole-mixture study*, a given mixture or combination is examined as a whole; the mixture may be viewed as a 'single chemical.' This approach has been used for complex mixtures with poorly characterized but stable composition (for example, industrial effluents; welding fumes), and for specially designed mixtures with completely characterized composition. Obviously, whole-mixture studies do not provide information about possible interactions between components of the mixture, nor do they enable identification of the most toxic components or fractions.

A *full factorial design* is a design in which each dose level (concentration) of each chemical is combined with each dose level (concentration) of every other chemical in the mixture or combination. The number of groups in a factorial design is m^k where k is the number of chemicals and m the number of dose levels (concentrations) of each chemical. The simplest form of a factorial design is a 2×2 *design*, which measures the effects/responses to the control situation (concentration zero for both chemicals), to one dose level of each of two chemicals, and, to the same dose levels of these two chemicals combined. This design can provide, for example, evidence that the chemicals interact antagonistically (when the response to the two chemicals combined is smaller than that to the individual chemicals). The number of treatment groups in a full factorial design increases so drastically with the number of chemicals that such a design is unfeasible when a mixture or combination contains more than four or five chemicals, especially in the case of costly studies in experimental animals. However, the number of groups can be strongly limited by considering only part of the possible combinations by using a so-called *fractional factorial design.* An illustrative example of the use of such a design is the application of a fractional two-level factorial design to examine the toxicity of a combination of nine different chemicals (widely varying in chemical nature) in a four-week toxicity study in rats. In this study, a full factorial design would have required 2^9 ($= 512$) treatment groups; instead, the 1/32 fraction ($= 16$ groups exposed to various carefully selected combinations of the individual chemicals) plus four groups exposed to different dose levels of the entire combination were used. The investigators were able to identify cases of nonadditivity as well as the components responsible for adverse effects of the combination on specific endpoints.

The *additivity approach* uses the dose/response curves of the individual chemicals to calculate the expected response for a given combination (or mixture), under the assumption of additivity. Then the predicted response is compared with the experimentally obtained response to that combination. For a combination of k chemicals, the number of treatment groups would amount to $(m \times k) + 1$ (m dose levels of k chemicals + the combination) which is much less than the m^k groups required for a full factorial design. A disadvantage of this approach is that it can only detect whether the response to the whole combination or mixture deviates from additivity; it cannot identify the chemicals that cause the interaction.

The *isobolographic method* is the classical approach to determine whether or not two chemicals interact. An isobologram is a graphical representation of the joint effect of two

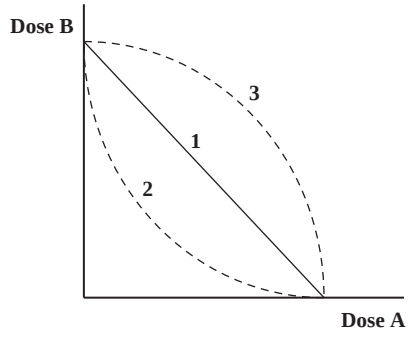


Figure 2.58 *Isoboles of two chemicals, indicating different types of joint action. On the axes are the doses of chemical A and B that produce an equal effect (for example, 60% increase in liver weight relative to body weight). Isobole 1 represents dose additivity, isobole 2 supra-additivity or synergism, and isobole 3 infra-additivity or antagonism.*

chemicals in which the doses of the chemicals A and B are given on the *x*- and *y*-axis, respectively, and the experimentally determined dose combinations of these two chemicals which all elicit the same effect are plotted and then connected by a line, the *iso-effect line* or the *isobole* (Figure 2.58). The experimentally determined line is then compared with the theoretical iso-effect line based on the assumption of additivity. Differences between these lines indicate departure from additivity. Isoboles below the line of additivity indicate synergism (in the presence of chemical A, less of chemical B is required to generate the specified response than would be the case under the assumption of additivity), those above it indicate antagonism. A major disadvantage of the isobolographic method is its extensive data demand. Also, because of its graphical nature and the use of perpendicular axes, isobolograms are unsuitable for mixtures of more than three chemicals.

Effect/response-surface analysis yields a statistically based mathematical relationship between the doses of each chemical in a mixture and the effect parameter. This method is much easier to perform but conceptually more difficult to interpret than isobolographic methods. The advantage of effect/response-surface analysis is that it includes all data points obtained and does not require a dose–effect equation for each individual chemical per se. In a study on the cytotoxicity to nasal epithelial cells of mixtures of aldehydes (formaldehyde, acrolein, and crotonaldehyde), effect-surface analysis revealed that two- and three-factor interactions were significantly different from zero, although the magnitude of the interaction was relatively small compared with the effect of single aldehydes. Besides the above empirical methods, *mechanistic methods* can be used to study the toxicity of mixtures or combinations of chemicals. For example, physiologically based toxicokinetic modelling approaches are potentially useful for predicting quantitatively the consequences of interactions, and to conduct extrapolations (high to low dose, route to route, binary to multichemical mixtures) of the occurrence and magnitude of interactions from laboratory *in vitro* or *in vivo* studies to human real life exposure scenarios.

2.6.5 Safety Evaluation

There are three possibilities to evaluate the safety of a mixture (or combination) of chemicals: the mixture can be approached as

- **a single entity**;
- **a number of fractions**;
- **a number of individual constituents.**

The safety evaluation based on the individual constituents can be approached by considering

- **all constituents**;
- **the 'top** n' **constituents**;
- **the 'pseudo top** n' **constituents.**

Then, depending on the type of joint action of the chemicals in a mixture (or combination), the safety evaluation can be performed using

- **the dose additivity/Hazard Index method**;
- **the response additivity/Hazard Index method; and/or**
- **the Hazard Index/Weight-of-Evidence method.**

The three possibilities to evaluate the safety of a mixture (or combination) of chemicals are: the mixture can be approached as a single entity, a number of fractions, or a number of individual constituents (Figure 2.59).

Mixture as a Single Entity

Once the NOAEL of a mixture has been established, the possibility of recommending an exposure limit can be examined, using methods similar to those for individual substances. It is common practice to accept toxicity data on a mixture as an entity, although the intention may not always be to recommend exposure limits and to set standards. If a mixture has not been examined toxicologically but another mixture with a reasonably similar composition has, one may consider basing the safety evaluation on data on the similar mixture. In the simplest situation, the data can be adopted directly for the mixture concerned. In more complex situations, sophisticated techniques (such as pattern recognition, principal-component analysis, and partial least square projections to latent structures) can be used to compare mixtures physico-chemically and toxicologically. Such techniques, for example, have been used to compare the dioxin composition of sludge and to determine the toxicity of diesel exhaust fumes.

Mixture as a Number of Fractions

If a mixture cannot be tested in its entirety, it may be possible to divide it into fractions according to its chemical or physical properties, and to study the toxicity of these fractions. The data obtained may be useful for risk assessment and there is also, at least theoretically, still the option of calculating an exposure limit through conversion and extrapolation of the

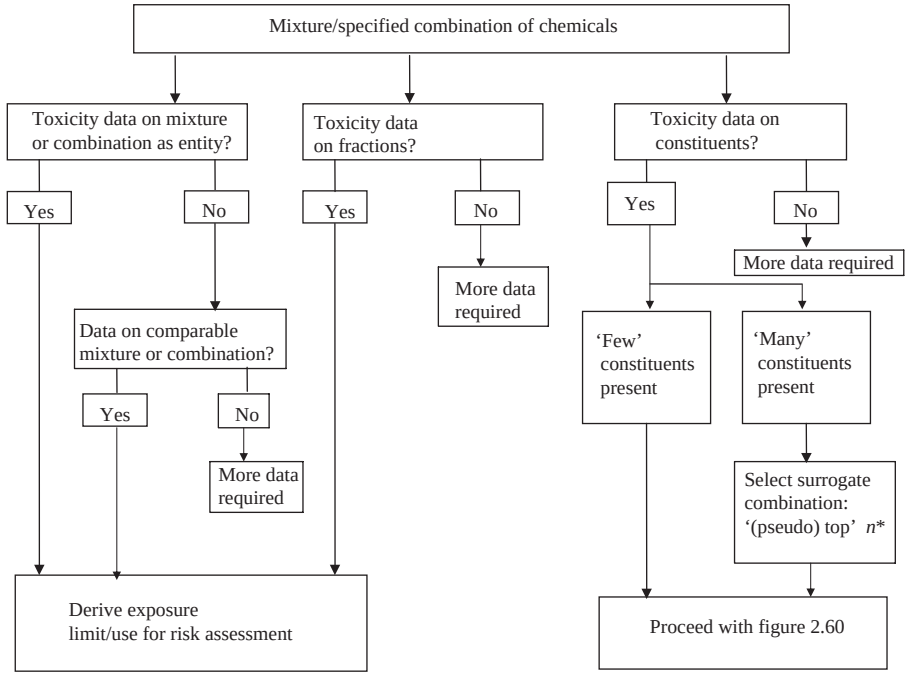


Figure 2.59 *Flow chart for the safety evaluation of mixtures (or combinations) of chemicals. * = '(pseudo) top' n: see text for explanation.*

NOAEL for the most harmful fraction. However, when taking this approach, one needs to be prepared for changes in toxicity due to chemical changes that fractionation may cause. Fractionation has been successfully applied to diesel exhaust fumes; the particulate matter fraction appeared to be entirely responsible for the respiratory tract inflammation and cancer discovered in animal studies; neither the free constituents nor the gaseous constituents adsorbed to particles contributed to these effects.

Mixture as a Number of Individual Constituents

The third possibility to evaluate the safety of a mixture (or combination) of chemicals is to take the individual constituents as the starting point. Depending on the number of chemicals present in a mixture, the safety evaluation can be approached by considering all constituents, the 'top *n*' constituents or the 'pseudo top *n*' constituents.

For mixtures (or combinations) consisting of a small number of chemicals, say ten or fewer, the safety evaluation should include *all individual constituents*; they should all be considered in detail (see the next section '*Type of Joint Action as the Basis for Safety Evaluation*'). However, for mixtures or combinations consisting of a large or very large number of constituents the evaluation has to be reduced to manageable proportions. In such cases, the evaluation should focus on the most risky chemicals, viz. the chemicals that constitute the highest health risks based on toxic potential and (estimated) exposure.

The joint toxicity of the most risky chemicals is assumed to represent that of the entire mixture or combination, or at least to closely approximate it.

At what number of constituents does the safety evaluation require a surrogate mixture or combination? Ten has been suggested because risk assessment was considered to become too difficult and too complex above this number. However, it is more appropriate to decide the precise number on a case-by-case basis and therefore *n* has been introduced. Above *n*, a surrogate combination (of no more then *n* constituents) has to be selected and assessed for combined toxicity. In the case of a not overly complex situation, it is best to select the '*top* n'. In case of 'very many' constituents, the '*pseudo top* n' is the best option: first placing each constituent in a (chemically or physiologically) related class followed by identification of the *n* most risky classes. Then, for each class a representative chemical (a true chemical or a pseudo-chemical representing the fictional average of a class) is identified. Once the '*top* n' or '*pseudo top* n' *constituents* have been identified, they will be approached as a simple (defined) mixture or combination. Toxicity testing should first involve two or three chemicals and later deal with a gradually increasing number and thus a gradually increasing *n.* The combination issue must be realistic and the evaluation must extend to determining which health-protection measures are advisable. Working in this way, the number *n* can be determined, incorporating both feasibility and scientific soundness. Experience has shown that $n = 1$ is sometimes sufficient: in the case of polycyclic hydrocarbons, benzo[*a*]pyrene could represent the total. A similar approach has successfully been used for more then a decade (in the United States) to identify the priority substances (the 'top *n*' or 'pseudo top *n*' chemicals) that are released from hazardous-waste sites. A slightly different, though in principle comparable, procedure has recently been developed for the safety evaluation of natural flavour complexes obtained from botanical sources and intended to be used as flavouring substances for foods and beverages.

Type of Joint Action as the Basis for Safety Evaluation

Figure 2.60 shows a flow chart for the safety evaluation based on individual constituents and type of joint action. The following question has to be answered: is the type of joint action expected to be similar or dissimilar, or interaction? If this question cannot be answered more data are required to perform a safety evaluation.

In the case of ***similar joint action*** dose addition can determine the effect of the combined exposure using empirically obtained exposure-effect data. To assess the potential health risk of such combined exposures the dose additivity/Hazard Index (HI) method should be used. This method involves summation of the ratios of the actual exposure concentration to the established exposure limit value (hazard quotients) for each individual chemical. If the sum exceeds unity, the combined exposure is considered of health concern. An example: an atmosphere contains 400 ppm acetone (exposure limit is 500 ppm) and 150 ppm sec-butyl acetate (exposure limit is 200 ppm). Both chemicals are central nervous system depressants. Application of the dose additivity/HI method shows that the sum exposure exceeds 1 (400: 500 + 150: 200 = 0.8 + 0.75 = 1.55), indicating an exposure situation of concern.

Another method based on dose additivity is the Toxic Equivalency Factor (TEF) method that was developed as a procedure to assess the toxicity of complex mixtures of

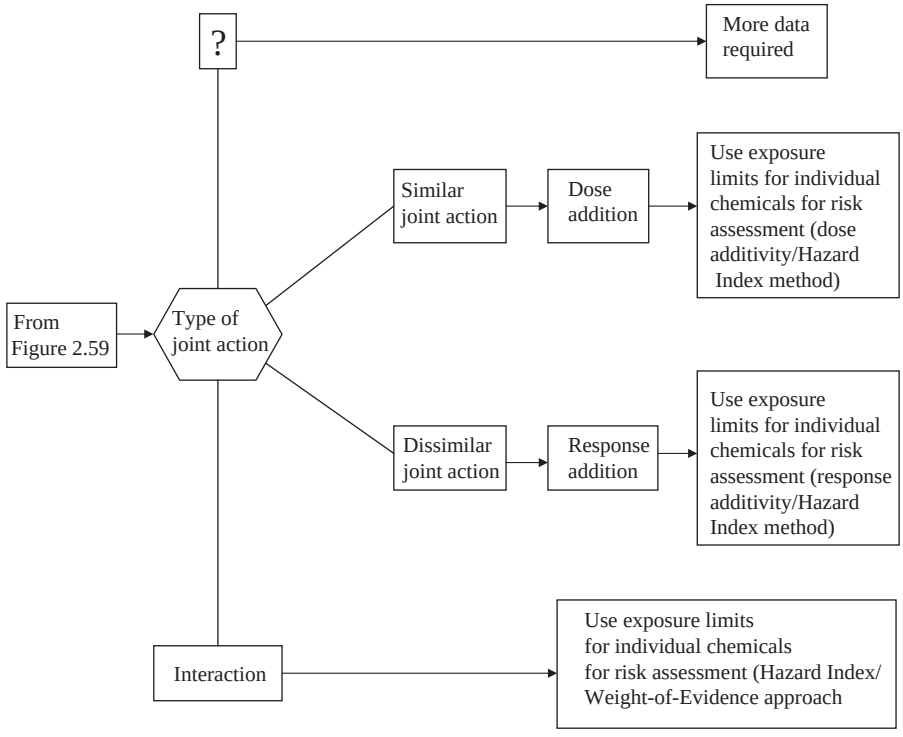


Figure 2.60 *Flow chart for the constituent-based safety evaluation based on type of joint action.*

polychlorinated dibenzo-*p*-dioxins and dibenzofurans. The TEF method assumes a common mode of action. In this method, the dose of each component is normalized against the dose of one of the components, usually the most potent and best-studied one (called the index chemical), to derive a relative potency for each component. The relative potencies are then summed to estimate the toxicity of the combination or mixture.

In the case of ***dissimilar joint action***, response addition is the designated method for determining the combined effect. As long as the various exposure concentrations do not exceed the respective exposure limits, there is adequate protection. An example: an atmosphere contains 0.04 mg/m^3 of the neurotoxicant lead (exposure limit is 0.05 mg/m^3) and 0.9 mg/m^3 of the irritant sulfuric acid (exposure limit is 1 mg/m^3). Application of the response additivity/Hazard Index (HI) method shows that the exposure situation is nonhazardous because the Hazard Index (the hazard quotient) of both substances is smaller than 1, viz. 0.8 (0.04: 0.05) for lead and 0.9 (0.9: 1) for sulfuric acid. Only if at least one of the components has a hazard quotient that exceeds unity is exposure to the combination deemed to be of health concern. Indeed, here the exposure limits of the individual chemicals constitute the reference point. There is one (important) exception: response addition in the case of substances for which the effect has no zero risk exposure level (for example, genotoxic carcinogens). If the effect measure is an incidence

(or likelihood) and the substances act dissimilarly without interaction but have similar health effects (for example, different types of cancer), the combined effect can be calculated through response addition. In that case, cancer (not the type of cancer) is the reference point, and combined exposure to the two substances, each below its own recommended exposure limit, may exceed the acceptable cancer risk level. If, indeed, the acceptable cancer risk level is exceeded, the exposure concentrations of the individual chemicals (carcinogens) must be reduced to maintain the level of protection.

In the case of ***interaction*** it may or may not be possible to draw (semi-)quantitative conclusions depending on the data. Since neither the dose additivity/HI method nor the response additivity/HI method takes into account interactions, these methods cannot be used as such. A method that takes into account both possible synergistic or antagonistic interactions is the so-called Hazard Index/Weight-of-Evidence (HI/WOE) method. With this method, a WOE classification is used to estimate the joint actions (additivity, antagonism, synergism) for all pairs of chemicals in a simple mixture or specified combination, based on information on the individual chemicals; as shown in

$$\mathrm{HI_A} = \mathrm{HI} \times \mathrm{UF_I^{WOE}} \tag{2.58}$$

Equation (2.58), where $\mathrm{HI_A}$ stands for the HI adjusted for the uncertainty for interactions, $\mathrm{UF_I}$ stands for a 'fixed' (chosen) uncertainty factor for interactions, e.g. 10, and WOE represents the Weight-of-Evidence score (ranging from -1 for the highest possible confidence in significant antagonistic interaction to $+1$ for the highest possible confidence in significant synergistic interaction, with 0 as full confidence in absence of any interaction). In the WOE score, several weighting factors are taken into account, such as mechanistic understanding and toxicological significance of the binary interactions, route, duration, and sequence of exposure, and whether data are from *in vitro* or *in vivo* studies. The better the data set on the individual chemicals, the more precisely the joint action of the combined exposure can be predicted.

2.6.6 Summary

Definitions are given for the following types of mixed exposures: mixture, simple mixture, complex mixture, combination and specified combination of chemicals, and cumulative exposure. Three types of joint action of chemicals in a mixture or combination have been defined: dissimilar joint action, similar joint action, and interaction. With respect to interaction, three types are often distinguished: direct chemical–chemical, toxicokinetic, and toxicodynamic interaction. Given the wide diversity of exposure scenarios, the questions to be addressed in toxicity studies of mixtures (or combinations) of chemicals vary widely and, consequently, there is no 'one size fits all' design for toxicity studies of mixtures. Major factors that influence study design as well as the method of analysis of the results include the number of chemicals in a mixture or combination, the availability of a mixture for testing it in its entirety, and the extent to which the toxicity of a mixture or combination needs to be characterized in terms of dose–response relationship or departure from additivity, and available resources. To evaluate the safety of mixtures, they can be approached as a single entity, a number of fractions, or a number of individual constituents. The safety evaluation based on the individual constituents can be approached by considering all constituents, the 'top *n*'

constituents or the 'pseudo top *n*' constituents. Depending on the type of joint action of the chemicals in a mixture (or combination) the safety evaluation can be performed using the dose additivity/Hazard Index method for similarly acting chemicals, the response additivity/Hazard Index method for dissimilarly acting chemicals, and/or the Hazard Index/Weight-of-Evidence method for chemicals showing interactive toxicity.

For Further Reading

Feron, VJ, Bolt, HM (1996). Combination Toxicology. *Food Chem. Toxicol*., **34**, 1025–1185.

Feron, VJ, van Vliet, PW, Notten, WRF (2004). Exposure to Combinations of Substances: a System for Assessing Health Risks. *Environ. Toxicol. Pharmacol.*, **18**, 215–222.

Jonker, D, Freidig, AP, Groten, JP, de Hollander, AEM, Stierum, RH, Woutersen, RA, Feron, VJ (2004). Safety Evaluation of Chemical Mixtures and Combinations of Chemical and Non-chemical Stressors. *Rev. Environ. Health*, **19**, 83–139.

Jonker, D, Woutersen, RA, Feron, VJ (1996). Toxicity of Mixtures of Nephrotoxicants with Similar or Dissimilar Mode of Action. *Food Chem. Toxicol*., **34**, 1075–1082.

Organisation for Economic Co-operation and Development (2001). Harmonised Hazard Classification Criteria for Mixtures. In: Harmonised Integrated Classification System for Human Health and Environmental Hazards of Chemical Substances and Mixtures, Report ENV/JM/MONO(2001)6, Paris, France.

Simmons, JE (1995). Chemical Mixtures and Quantitative Risk Assessment, *Toxicology*, **105**, 109–441.

Suk, WA, Olden, K, Yang, RSH (2002). Application of Technology to Chemical Mixtures Research. *Environ. Health Perspect*., **110**, Suppl. 6, 891–1036.

US Environmental Protection Agency (2003). Framework for Cumulative Risk Assessment, EPA/600/P-02/001F, Office of Research and Development, National Center for Environmental Assessment, Washington DC, USA.

Yang, RSH (1994). Toxicology of Chemical Mixtures, Academic Press, San Diego, p. 720.

2.7 Chemical Carcinogenesis: Genotoxic and Nongenotoxic Mechanisms

Thomas Efferth and Bernd Kaina

2.7.1 Introduction

Cancer is a multi-step process that can be dissected into tumor initiation, tumor promotion, and tumor progression (Figure 2.61). Tumors result from altered expression of genes that regulate cell proliferation, cell–cell communication, cell migration, and cell adherence. The most important genes are oncogenes and tumor-suppressor genes. Alteration in their function or their expression may be caused by mutations. Change in expression can also be caused by epigenetic mechanisms such as gene silencing. Mutations, however, appear to be the predominant mechanism underlying the process of carcinogenesis.

Mutations arise from DNA damage that may occur spontaneously or that is induced by exogenous insults, e.g. DNA-adduct formation by reactive metabolites of carcinogenic chemicals or radiation (Figure 2.62). Therefore, it is important to consider the mechanism of DNA-adduct formation, as well as the potency of metabolically activated

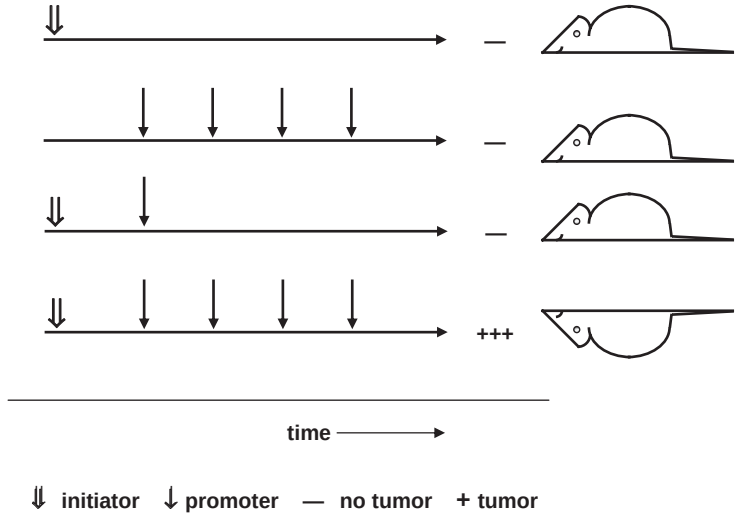


Figure 2.61 *Staging of tumor formation on mice skin. All DNA-damaging agents can be considered to be initiators. Tumor promoters stimulate cell proliferation, notably of initiated cells. They also may prevent apoptosis, thus causing initiated cells to survive and propagate in the tissue. Tumor promoters must be applied repeatedly in order to provoke tumor promotion; their effect is reversible. This is in contrast to initiators, which exert irreversible cellular changes.*

carcinogens. Because DNA adducts are subject to repair, DNA repair at the cellular level is a major defence mechanism against cancer. However, mutations in DNA-repair genes evoke genomic instability and predispose the affected individuals to tumor formation. Furthermore, during the process of DNA replication DNA damage may be erroneously processed by low-fidelity DNA polymerases that would result in generation of mutations. Thus, the DNA-repair process may result in either removal of DNA damage or the fixation of altered DNA in the form of mutations.

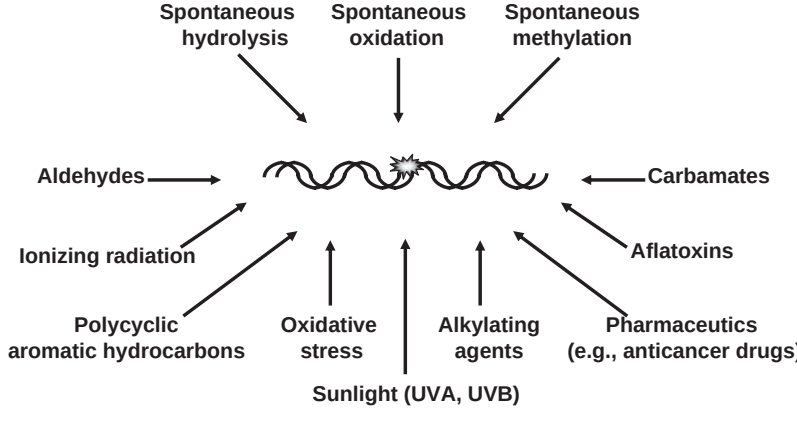


Figure 2.62 *Spontaneous and induced insults damaging DNA.*

2.7.2 Mechanisms of DNA Damage, Repair, and Carcinogenesis

DNA damage can result from loss of any of the bases, modification of base structure via adduct formation or oxidation by reactive oxygen species, impact of UV light, cross-linking of DNA strands, cross-linking between DNA and proteins, or by strand breakage. Xenobiotics and endogenous DNA damaging agents such as reactive oxygen species may be detoxified either by metabolic inactivation or by antioxidant mechanisms such as superoxide dismutases, catalase, and glutathione-related enzymes. DNA damage can be repaired by a number of sophisticated DNA repair pathways including damage reversal, base-excision repair, nucleotide excision repair, mismatch repair, nonhomologous end joining, and homologous recombination. Alternatively, DNA damage can be inadvertently fixed by template switching and lesion bypass.

DNA-damaging Agents

Damage to DNA is not necessarily mutagenic unless it is fixed to form permanent mutations that are starting points for carcinogenesis. The main types of DNA damage are:

- **Damage to DNA bases:** Typical examples are O^6-methylguanine, thymine glycol, fragmented bases by oxidative stress, UV-light-induced cyclobutane dimers and 6-4-photoproducts. Alkyl adducts are formed by alkylating agents and large, bulky adducts by polycyclic hydrocarbons.
- **Lesions at the DNA backbone** are abasic sites and DNA single- and double-strand breaks.
- **Cross-links:** Bifunctional agents such as nitrogen mustard, cisplatinum, psoralen, etc., form interstrand adducts between two complementary DNA-strands and DNA–protein adducts.
- **Aberrant nonduplex DNA forms** such as DNA bubbles, recombination structures, and aberrant fork structures, which are generated during disrupted replication, recombination, or DNA repair.

Exogenous DNA-damaging Agents DNA lesions may result from the metabolic activation of xenobiotic carcinogens which by virtue of their electrophilicity tend to form covalent adducts with DNA. Exposure to carcinogens may arise from release of industrial chemicals in the occupational setting or in the general environment. Furthermore, foods and beverages also contain compounds such as N-nitrosamines that can damage DNA and which are potentially mutagenic.

Among the plethora of carcinogenic compounds five large classes with some of the main representatives are listed here:

- **Polycyclic aromatic hydrocarbons**, e.g. benzo[*a*]pyrene, 3-methylcholanthrene, and benzanthracene
- **Aromatic amines**, e.g. β-naphthylamine, 2-acetylaminofluorene, and *o*-toluidine
- **N-Nitroso compounds**, e.g. dimethylnitrosamine and *N*-nitroso-*N*-methylurea

- **Carcinogenic natural products**, e.g. aflatoxin B1, safrole, and aristolochic acid
- **Cross-linking and alkylating agents**, e.g. sulfur and nitrogen mustards and ethylene oxide.

The most important chemical reactions leading to adduct formation of carcinogenic molecules with DNA are the transfer of alkyl, arylamine, or aralkyl groups. Alkylated or aralkylated compounds are generated by the oxidation of carbon atoms by the CYP enzymes, whereas the oxidation or reduction of nitrogen atoms leads to arylaminated substances.

Carcinogens form adducts with DNA bases. Alkylating agents frequently bind to exocyclic oxygen or ring nitrogen atoms, i.e. at the O^6 position of guanine and the N7 position of guanine, respectively. Arylaminated substances bind to nitrogen atoms at the N7 position. Subsequently, rearrangements can take place and C8-deoxyguanosine adducts can be generated. Polycyclic aralkylated compounds form various adducts by binding to exocyclic nitrogen atoms of adenine or guanine. The formation of adducts can lead to the substitution of DNA bases.

Endogenous DNA Damage The hydrolysis of N-glycosidic bonds of bases in the DNA is frequently observed. It is estimated that about 10,000 purines are lost per day in the human genome. In comparison with purines, only 5% of the pyrimidines are affected. Apurinic/apyridinic sites (AP-sites) are generated, which are cytotoxic and mutagenic since they block DNA replication, cause base misincorporation, and may lead to DNA strand breaks. The hydrolytic separation of exocyclic amino groups in cytosine and 5-methylcytosine leads to the formation of uracil and thymine, respectively, which then occur as mismatches in the DNA. The rate of cytosine deamination is high, estimated to be as much as 500 per cell per day. In a comparable manner, adenine can be deaminated to hypoxanthine, and guanine to xanthine.

Another cause of spontaneous DNA damage is oxidation by reactive oxygen species (ROS) and reactive nitrogen species (RNS), which are generated as by-products of oxidative metabolism, during inflammation processes, and after ionizing radiation. The mitochondrial electron-transport chain is not only the main source of ATP production by oxidative phosphorylation, but also the main source for the generation of ROS (H_2O_2 and $O_2^{\bullet -}$). The spatial proximity of ROS generation and mitochondrial DNA explains why oxidative damage is more frequently found in mitochondrial DNA than in nuclear DNA. During chronic inflammation, activated macrophages and neutrophilic leukocytes release NO, $O_2^{\bullet -}$, $^{\bullet}OH$, and HOCl, which damage the DNA of neighboring cells. Fatty acid radicals, aldehydes, and other compounds are generated during lipid peroxidation. They cause etheno-adducts of pyrimidines and purines. The oxidation of bases, which occurs mainly at electrophilic carbon centers, leads to very mutagenic bases in DNA such as 8-oxoguanine, formamidopyrimidine, and thymine glycol. Endogenous products of normal metabolism such as estrogens, heme precursors, amino acids, and glycol-oxidation products may also damage DNA.

Prevention and Repair of DNA Damage

The cell is equipped with several defense mechanisms to avoid mutations caused by endogenous and exogenous DNA-damaging agents.

There are several defense mechanisms that protect the cell from DNA damage:

- Mechanisms of oxidative stress response for the avoidance of oxidative DNA lesions by endogenous or exogenous reactive oxygen species.
- Multiple control mechanisms of the DNA replication machinery to minimize the error rate, i.e. < 1 misincorporation/10^6 nucleotides.
- Mechanisms to regulate cell-cycle progression to guarantee error-free duplication and segregation of chromosomes during cell division. Cell-cycle checkpoints control this function.
- DNA-repair mechanisms: over 130 DNA-repair genes have thus far been identified.
- If DNA cannot be repaired after massive damage, cells undergo apoptosis, i.e. programmed cell death. Hence, from the viewpoint of the entire organism, death of damaged cells can be looked upon as a protective mechanism.
- DNA lesions can evoke transcriptional changes characterized by up- or down-regulation of DNA-repair-gene expression. This can lead to an elevated level of defense due to increased repair enzyme activity.

If mutations cause the activation of proto-oncogenes and inactivation of tumor-suppressor genes, malignant transformation of cells can take place.

Antioxidant Mechanisms

ROS attack cell membrane lipids, generating DNA-reactive compounds and damage DNA directly. Antioxidant response genes (SODs, GSTs, CAT) counteract the deleterious effects of ROS.

The interaction of ROS with lipids is deleterious, since a single ROS molecule can generate many toxic reaction products due to autocatalytic dispersion. Examples are hydrogen peroxide, peroxyradicals, alkoxyradicals, and unsaturated aldehydes. The reaction of ROS with DNA is rated as most relevant for carcinogenesis. It is mainly based upon the induction of DNA strand breaks and base damage and the induction of certain cellular functions. Both events can finally lead to apoptosis. Organisms have, therefore, developed various protective mechanisms to fight oxidative stress induced by ROS and lipid peroxidation products, e.g. catalase, superoxide dismutases, and glutathione-associated proteins (glutathione reductase, glutathione peroxidase, glutathione S-transferases) as well as nonenzymatic molecules such as glutathione, or α-tocopherol. Antioxidant enzymes protect against the development of cancer.

Various isoenzymes of **superoxide dismutase** (SOD) have been identified, e.g. iron-dependent FeSODs in the cytosol, mitochondria, and chloroplasts. The extracellular EC-SOD appears to be membrane-associated in the extracellular space. The copper- and zinc-dependent CuZnSOD of eukaryotes is localized cytoplasmatically. Moreover, a manganese-dependent SOD (MnSOD) has been observed in eukaryotes. SODs catalyse the dismutation of superoxide anions ($2\ O_2^-$) in the presence of protons ($2\ H^+$) to molecular oxygen (O_2) and hydrogen peroxide (H_2O_2).

Subsequently, hydrogen peroxide is transformed into water by **catalase**. **Glutathione *S*-transferases** (GSTs) catalyse the nucleophilic reaction of glutathione (GSH) to many hydrophobic xenobiotics, which are then less toxic and easier to excrete. Oxidized

glutathione is returned to the reduced form by **glutathione reductase**. Furthermore, GSTs can participate on the repair of oxidative damage at membrane lipids and DNA.

Selenium-dependent **glutathione peroxidases** protect against lipid peroxidation by termination of the lipid peroxidation cascade through reduction of fatty acid hydroperoxides (FS-OOH) and phospholipid hydroperoxides (PL-OOH). Hydroxyl radicals are eliminated by a glutathione peroxidase-mediated reaction and are converted into water, while GSH is oxidized to GSSG. The cellular GSH level is regenerated by the NADPH-dependent glutathione reductase from GSSG. **Glucose-6-phosphate dehydrogenase**, which maintains the cellular NADPH level, can, therefore, also be considered an antioxidant enzyme.

DNA Repair

DNA repair is essential for maintenance of the genome. Most DNA-repair processes are error-free, and clearly protect against cancer induced by chemical agents. The main mechanisms of DNA repair encompass damage reversal by MGMT and ABH proteins, which repair O^6-alkylguanine and N1-methyladenine, respectively. Small lesions are repaired by base-excision repair, while larger adducts are repaired by nucleotide-excision repair (NER). Mismatches are subject to mismatch repair, and cross-links to cross-link repair that involves components of NER and recombination repair. DNA double-strand breaks are repaired by nonhomologous end joining (NHEJ) and homologous recombination (HR). Both are complex pathways involving a different set of proteins. NHEJ is an error-prone pathway, leading to gross chromosomal changes.

DNA has a special role as a critical target molecule compared with other macromolecules (lipids, proteins, RNA), since in case of damage it cannot simply be replaced. Therefore, the repair of DNA-lesions is of eminent importance for life, and a plethora of different DNA repair pathways have been developed during evolution. The main mechanisms (Figure 2.63) are discussed below.

Damage reversal

Damage-reversal mechanisms are rapid repair processes that remove specifically highly premutagenic and pretoxic lesions. Among the best known examples is the repair protein MGMT that clearly protects against cancer initiation and progression.

Alkylating agents produce N- and O-alkylated purines and pyrimidines as well as DNA phosphotriesters. O^6-Alkylguanines (O^6-methylguanine and O^6-ethylguanine) are critical lesions, which cause GC-to-AT transition mutations. Likewise, O^4-methylthymine is critical, forming TA–CG mutations. These lesions are repaired by O^6-methylguanine-DNA methyltransferase (MGMT, also designated as alkyltransferase). The protein removes the methyl group from the O^6-position of guanine and, less efficiently, from the O^4 position of thymine, and transfers it to an own internal cytosine residue. This transfer reaction leads to irreversible inactivation of MGMT. MGMT is, therefore, not an

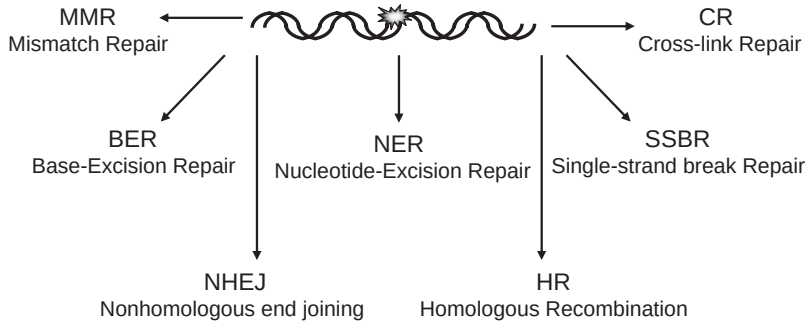


Figure 2.63 *Mechanisms of repair of DNA damage.*

enzyme in the classical sense; it has been termed a suicide enzyme. Since O^6-methylguanine is a major promutagenic (Figure 2.64A) and procarcinogenic lesion, MGMT clearly protects against cancer formation, which has experimentally been shown in MGMT-overexpressing mouse models.

A second recently discovered damage reversal mechanism is represented by ABH proteins (ABH = alkB homologues of *E. coli*). At least three of these proteins are known to occur in human cells: ABH1, ABH2, and ABH3. ABH2 and ABH3 proteins belong to a superfamily of

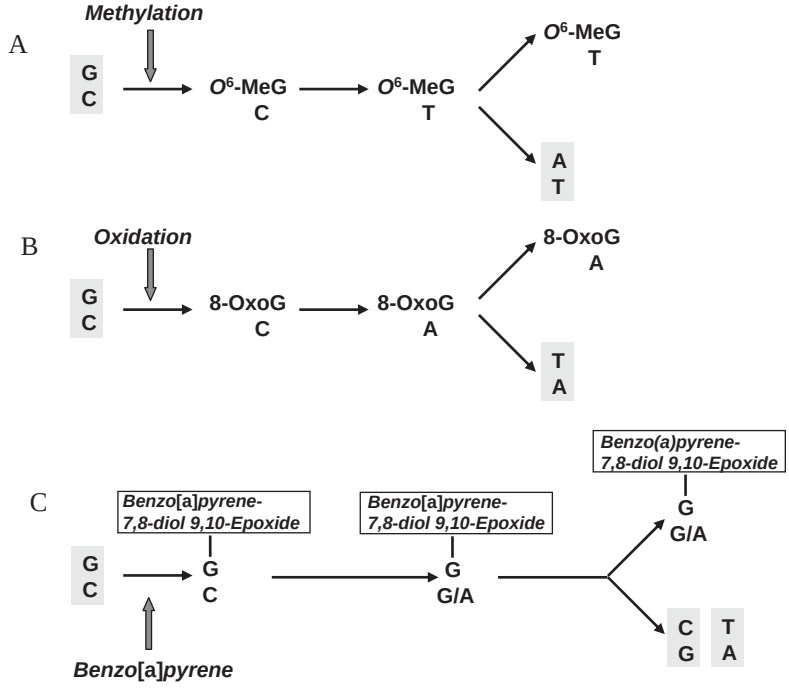


Figure 2.64 *Mechanism of mispairing mutagenesis by instructive lesions O^6-methylguanine (A), 8-oxoguanine (B), and the noninstructive benzo[a]pyrene–guanine adduct that requires low-fidelity DNA polymerase action (C).*

ketoglutarate- and Fe II-dependent dioxygenases that repair *N*1-methyladenine, *N*3-methylcytosine and *N*1-ethyladenine in a reversal reaction involving oxidative demethylation and restoration of the undamaged base. ABH3 repairs single-stranded DNA and even RNA. The repair mechanism contributes to cellular protection against alkylation damage.

Base-excision repair (BER)

Many modified bases in DNA are removed by BER. Although several mouse strains deficient in BER have been created, human disorders characterized by a defect in BER are not known. Nevertheless, it is believed that BER plays a crucial role in the body's defense against cancer, chronic inflammation, and malformations.

'Small' lesions in DNA, i.e. nonbulky adducts, which do not cause clear distortion of the DNA helix, are repaired by BER (Figure 2.65). The process is highly specific because of recognition of lesions by specific enzymes, the DNA glycosylases. BER results in

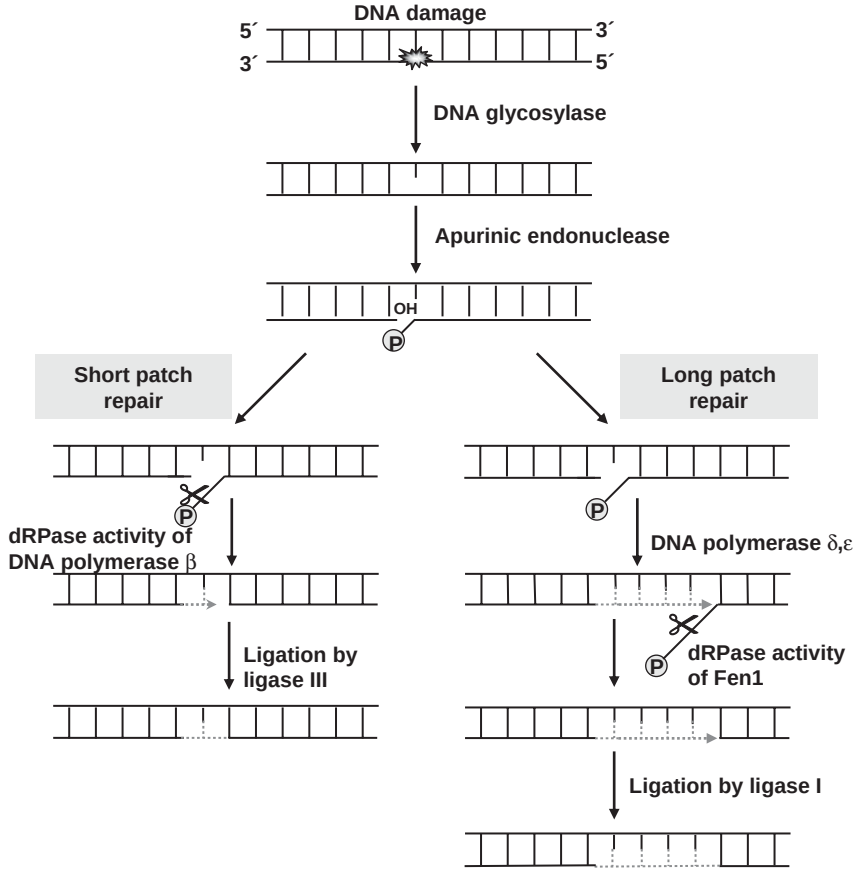


Figure 2.65 *Mechanism of base excision repair. For explanation see text.*

elimination of incorporated uracil (a base normally present in RNA, but not DNA), fragmented pyrimidines, N-alkylated purines such as *N*7-methylguanine and *N*3-methyladenine, 8-oxoguanine, thymine glycol, and others. From these lesions, 8-oxoguanine is presumably the most important oxidative DNA lesion. It is mutagenic because of its ability to mispair with adenine (Figure 2.64B).

BER can be subgrouped into single-nucleotide and long-patch BER. The steps of BER are as follows:

1. Damage recognition, base removal, and incision: First, **DNA glycosylases** recognize and remove a damaged base by hydrolysis of the N-glycosidic bond, thereby generating apurinic or apyrimidinic sites (**AP-sites**). Subsequently, **AP endonucleases** cut the phosphodiester bond at the AP-site, generating a DNA single-strand break.
2. Nucleotide insertion: **DNA polymerase-β** (Polβ) inserts a nucleotide at the AP-site.
3. Depending on the nature of the base damage, **short-patch** or **long-patch repair** takes place. During long-patch repair **proliferation-dependent nuclear antigen (PCNA)** cuts out several nucleotides close to the lesion.
4. Strand displacement: During *long-patch* repair, up to 10 nucleotides are excised by endonucleases.
5. Ligation: **DNA ligases I and III** ligate the newly synthesized strand with the DNA. Ligase I is active in *long-patch* base-excision repair, ligase III in *short-patch* repair.

Nucleotide-excision repair (NER)

NER is a highly complex system for the repair of bulky DNA lesions. It discriminates between active and inactive genes in order to repair preferentially the DNA that is in the process of transcription. Several genetic diseases are characterized by defects in NER. Xeroderma pigmentosum displays a defect in global genomic repair, while Cockayne's syndrome is defective in transcription-coupled repair.

A common feature of BER and NER is that the complementary DNA strand is used as a matrix for the new synthesis after elimination of the damage. Therefore, both pathways are error-free. BER and NER differ from each other in that:

- BER is used for small base adducts (**nonbulky lesions**), while NER is activated for larger damage (**bulky lesions**).
- About 30 nucleotides are replaced in NER, but only 1–10 in BER.
- In BER, the damaged base is removed as free base, whereas in NER the damaged site is cut out as a part of a longer single-stranded fragment.

Bulky adducts are induced in guanine by *N*-acetylaminofluorene, benzo[*a*]pyrene, aflatoxin, and cisplatinum. Psoralen and ultraviolet light induce thymine cross-links (CPDs) and 6-4-photoadducts that inhibit DNA replication. Error-prone DNA polymerases can bypass the lesion. This causes mutations, such as G → T transversions during DNA replication.

NER represents a complex pathway involving approximately 30 proteins. Two main pathways are **global genomic repair (GGR)** and **transcription-coupled repair (TCR)**.

GGR is independent of the transcriptional activity of affected gene sequences. This pathway removes lesions in transcriptionally active and inactive DNA regions with similar efficacy.

XP proteins (XPA–XPG) contribute to GBR and TCR. The name XP derives from the UV-hypersensitive hereditary disease Xeroderma pigmentosum. Mutations in XP genes cause extreme sensitivity to sunlight and development of skin cancer that may occur already early in life. For unknown reasons, Cockayne's syndrome patients defective in TCR are not cancer prone.

DNA repair by NER occurs in four steps:

1. DNA damage recognition: The protein complexes XPC-HR23B and RPA-XPA recognize DNA lesions, such as 6-4-PPs or cisplatinum–DNA adducts, which are characterized by distortion of the DNA. Another protein complex consists of the damaged binding proteins XPE and XPG. Binding of these proteins at the site of the lesion further distorts the DNA and invites repair.
2. DNA unwinding: A complex of the transcription factor TFIIH and 9 other proteins bind to the damaged site. This causes unwinding of the DNA around the damaged site.
3. Excision of the DNA lesion: XPD induces incisions at the 3′-position and the ERCC1-XPF complex at the 5′-position flanking the damaged site. A 27–29-bp sequence containing the damaged site binds to the protein complexes and is removed.
4. Repair synthesis: The gap is filled by DNA polymerases POLδ and POLε and sealed by DNA ligase I and other accessory factors.

Transcription-coupled repair (TCR) is less well understood. DNA lesions in constitutive active (housekeeping) genes are more efficiently repaired than are lesions in noncoding sequences. Within genes, the transcribed 5′-strand is more efficiently repaired than the nontranscribed 3′-strand. TCR takes place only in genes that are transcribed by RNA polymerase II. Lesions in RNA polymerase II-dependent ribosomal genes are scarcely repaired. The same is true for mitochondrial genes.

Mismatch repair (MMR)

With the finding that a sub-group of colon carcinomas is defective in MMR, this repair pathway attracted much interest as a tumor-defense strategy. MMR not only removes misincorporated bases from the daughter strand after replication, it is also considered as a sensor of DNA damage involved in the process of DNA damage-triggered apoptosis. MMR defective cells are highly resistant to the killing effect of methylating carcinogens that further contributes to the high mutation frequency and genomic instability in the affected cells.

MMR repairs single-base mismatches and small insertion and deletion mismatches of multiple bases. Base mismatches can occur spontaneously or by chemically induced base deamination, oxidation, and methylation as well as by replication errors. Therefore, MMR has also been termed replication error repair (RER). Examples of chemically induced base mismatches are alkylation-induced O^6-methylguanine matched with thymine, cisplatinum-induced 1,2-intrastrand cross-links, UV-induced 6-4-PPs, purine adducts of benzo[*a*]pyrene or 2-aminofluorene, and 8-oxoguanine mispaired with adenine.

MMR can be divided into four steps:

1. DNA-damage recognition: Mismatches are recognized by the dimeric proteins MSH2/MSH6 or MSH2/MSH3, which bind to the site of the lesion and distort the DNA.
2. Strand discrimination: Two mechanisms have been proposed. The molecular-switch model claims that ADP binds to the MutSα complex, thereby recognizing the mismatch. The MutSα–ADP complex corresponds to the active state. By binding to the mismatch, an ADP → ATP transition and intrinsic ATPase activity is stimulated. This causes a conformational change and the docking of the MutLα-complex (MLH1-PMS2). In the hydrolysis-driven translocation model, the ATP hydrolysis induces a translocation of the MutSα complex.
3. Excision: After binding to the mismatch, MutSα associates with MutLα. The excision of the mismatched is mediated by exonuclease I.
4. Repair synthesis: The new synthesis is performed by POLδ.

DNA double-strand break (DSB) repair

DSBs are considered to be the most lethal DNA lesions. Therefore, their repair is of utmost importance for cellular survival. In most cases DSB repair accurately replaces damaged DNA. However, a subpathway of DSB repair is error-prone, leading to the formation of chromosomal rearrangements. The cells may survive, but at the expense of chromosomal mutations that may ultimately result in cancer.

DSBs occur as a result of ionizing radiation or exposure to DNA topoisomerase II inhibitors. DSBs can induce either genotoxic effects or apoptosis. DSBs are repaired via two pathways: error-free homologous recombination (HR) and error-prone nonhomologous end joining (NHEJ). HR prevails in primitive eukaryotes such as yeast, while NHEJ is the predominant pathway in mammals. NHEJ takes place mainly in the G_0/G_1 phase whereas HR occurs in the late S and G_2 phase of the cell cycle.

NHEJ is a mechanism for repairing DSBs which permits direct ligation of the broken ends of DNA without the need for a homologous template to guide the joining process. The first step is the binding of a heterodimer termed Ku70, Ku80 to the broken ends, which then facilitates binding to DNA ligase IV, which mediates the ligation. During homologous recombination (HR) the damaged chromosome attracts an undamaged strand of DNA with a homologous sequence that serves as a template for the repair. HR starts with nucleolytic resection of the DSB in 5′-3′-orientation. Two 3′-single-stranded DNA ends are generated. A protein called RAD52 catalyses the annealing of homologous strands of DNA. RAD51 is a recombinase, which is responsible for the DNA strand exchange of damaged DNA with a homologous region of the undamaged DNA.

Tolerance to DNA Damage

Bulky DNA lesions block DNA replication. Unless they are removed by NER, they may cause misreplication and mutations because of error-prone DNA polymerases that read over the lesion. Thus, bulky DNA lesions may go unrepaired and result in

mutations. For many chemical carcinogens that form large adducts on DNA such as benzo[*a*]pyrene and aflatoxins, error-prone translesion synthesis is the major pathway of mutagenesis. Therefore, error-prone polymerases are presumably crucially involved in carcinogenesis induced by genotoxins.

Considering the high frequency at which DNA lesions occur, e.g. 10 000 depurinations per cell per day, only a small but nevertheless significant portion escape error-free DNA repair. Since persistant DNA damage impedes DNA replication and is cytotoxic, mechanisms in which DNA damage is tolerated have developed that are termed template switching and lesion bypass.

Template switching Despite the inhibition of DNA synthesis on the damaged strand, DNA synthesis can take place to some extent on the nondamaged strand. The newly synthesized daughter strands serve as templates. After dissociation of both newly synthesized daughter strands, re-annealing to the original parental strands and semi-conservative replication takes place. Thereby, the damaged site can be bypassed and the replication machinery can proceed with normal DNA synthesis. Since the damaged site is bypassed, DNA synthesis continues in a error-free manner.

Lesion bypass In contrast to template switching, the damaged DNA strand is used in the lesion-bypass mechanism. Nucleotides are incorporated opposite to the damaged strand (**translesions synthesis**). Then, DNA synthesis is extented, and the replication machinery continues DNA synthesis. An error-free lesion bypass takes place if the right nucleotide is incorporated. However, error-prone incorporation can also occur. Lesion bypass is performed by specific DNA polymerases, which use the damaged DNA strand as a template. Polη catalyses error-prone DNA synthesis with UV-induced damage, adducts of benzo[a]pyrene (Figure 2.64C) or GG-cisplatinum intrastrand adducts. The human **DNA polymerase** (Polι) represents an exception, since this enzyme does not follow the common Watson–Crick base pairing. Polι frequently incorporates G instead of A at sites opposite to a T in undamaged DNA regions. As a consequence, DNA elongation truncates opposite to Ts (T-stop). The enzyme displays less catalytic efficacy opposite to a C. Hence, Polζ catalysis is cumbersome, and only short DNA sequences are synthesized. The enzyme promotes somatic hypermutation during immunoglobulin development. Somatic hypermutations in heavy and light antibody chains occur with 100 000-fold higher frequency than in other genes.

Halting of the Cell Cycle as a Means to Provide Time for DNA Repair or the Termination of Mitosis

Halting cell-cycle progression allows a cell to repair DNA damage. Defined cell-cycle checkpoints (G1/M, G2/M, S/M) maintain quality control of DNA and, hence, guarantee genomic integrity. Upon DNA damage, master genes (ATM, ATR, p53) induce cell-cycle arrest by regulation of effector genes (cyclin-dependent kinases, cyclins).

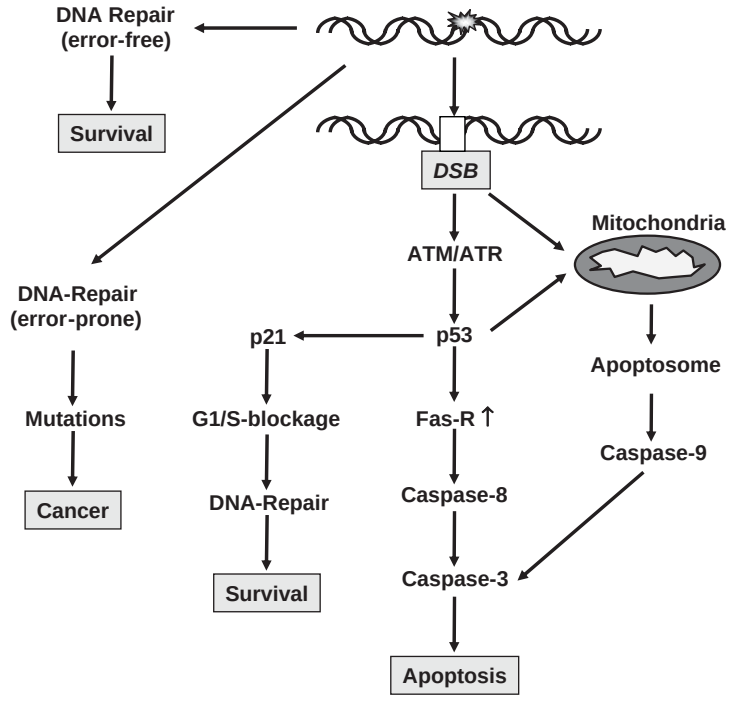


Figure 2.66 *Fate and consequences of DNA lesions. For explanation see text.*

Cell division is one of the utmost importance in biology. Correct and error-free repair processing during cell cycling is crucial to a cell's life. Defined checkpoints serve the maintenance of genomic integrity. These checkpoints allow halting of the cell cycle and DNA repair. Upon massive damage, the cell can undergo either apoptosis or permanent cell cycle arrest (also called **cellular senescence**).

ATM and **ATR** are very important sensor proteins, which activate cell-cycle arrest upon DNA damage (Figure 2.66). Both proteins switch on the G1/S and G2/M checkpoints by phosphorylation of Chk1, Chk2, and p53. Another checkpoint at the S boundary of the cell cycle is driven by the ATR-interacting protein (ATRIP). Therefore, ATM, ATR, and ATRIP are *master proteins* for checkpoint regulation.

At the G1/M DNA damage checkpoint the cell cycle can be halted by wild-type p53 in early and mid G1 phase as a result of DNA damage. Mutations of p53 result in loss of the ability to control the cell cycle. At the G2/M DNA damage checkpoint DNA damage by oxidative stress or ionizing radiation leads to G2-arrest. At the replication checkpoint (S/M checkpoint) mitosis is inhibited when errors occur during DNA replication, i.e. after depletion of nucleotide pools or DNA damage by alkylating agents (MMS). If the replication fork collides with a damaged site, replication is stopped, and the transition from S- to M-phase is inhibited.

2.7.3 Cancer Development

Multiple-step Carcinogenesis

Persistant DNA damage, which has escaped the DNA-repair defence mechanisms, may initiate tumor development. This process is traditionally divided into three steps: initiation, promotion, and progression. During the initiation phase, cells develop mutations in one or several critical genes. During tumor promotion, proliferative stimuli lead to clonal expansion, notably of the initiated cell. The progression phase is characterized by further accumulation of genetic changes that may be driven by chromosomal instability. Deregulated apoptosis and telomerase activity as well as neoangiogenesis support the carcinogenic process.

The development of cancer takes place in several steps: initiation, promotion, and progression (Figure 2.61). During **initiation**, DNA damage leads to replication errors and mutations such as base exchanges or frameshift mutations. If mutations appear in coding regions of genes that are relevant for cell growth and differentiation, malignant tumor formation can be initiated.

Initiation Several cellular processes are important for initiation:

- Metabolic activation of carcinogens;
- Errors in DNA repair;
- Transfer of mutations to daughter cells by cell proliferation.

The initiation phase is irreversible. Nevertheless, not all initiated cells end up as tumors, since most damaged cells undergo apoptosis. An initiated cell is not yet a tumor cell, since it does not grow autonomously. Many DNA lesions can be accumulated during the life span of an organism without the generation of a tumor. Therefore, additional events are required in order to fully transform a cell.

Apart from the inactivation of DNA repair genes, **proto-oncogenes** are activated and **tumor-suppressor genes** are inactivated. Furthermore, **epigenetic events** contribute to the transformation of cells from a benign to a malignant state. The minimal constellation for the development of cancer also consists of a constitutive expression of **telomerase** that is responsible for the immortalization of normal cells as a precondition for cancer development. Immortalization precedes transformation. Telomerase circumvents the erosion of telomers at the chromosomal ends. The permanent shortening of telomeres during each replication round leads to senescence of healthy cells.

Promotion During the second phase of cancer development (promotion), initiated cells are stimulated to grow. Initiated cells start to replicate, and clonal cell expansion takes place. Chemical compounds that do not damage DNA and which are not genotoxic, but which stimulate the proliferation of initiated cells, are termed tumor promoters. Tumor development starts with pre-neoplastic precursors, many of which regress spontaneously via clonal expansion. Phenobarbital and tetradecanoylphorbol acetate (TPA) are

examples of compounds with cancer-promoting properties. TPA has structural similarities to diacylglycerol (DG), which is a signal molecule for the activation of protein kinase C (PKC). PKC stimulates cell growth. TPA can substitute for DG in the corresponding signal-transduction pathway. While the signaling effect of DG is controllable, TPA (upon repeated application) permanently activates PKC, which stimulates continual tumor cell proliferation.

Carcinogens can be classified based on their carcinogenic function:

- **Primary carcinogens** act directly without prior metabolic activation.
- **Secondary carcinogens (pre-carcinogens)** are activated by biotransforming enzymes, e.g. cytochrome P450 monooxygenases.
- **Co-carcinogens** boost the effect of carcinogens by induction of biotransformation enzymes.
- **Promoters** indirectly increase the carcinogenic effects by the stimulation of cellular proliferation and/or suppression of apoptosis.

Progression Cancer progression is characterized by the **invasion** of tumor cells into the neighboring normal tissues, the spread of **metastases** and **genetic instability** of the tumor. Genetic instability does not mean only the acquisition of point mutations (base substitutions) as indicated above, but also larger structural changes at the chromosomal level (translocations, amplifications, deletions, etc.) as well as gain or loss of entire chromosomes (aneuploidy). The availability of nutrients and oxygen is an important condition for progressive tumor growth. Therefore, tumor cells excrete growth factors and other substances, which foster the generation of novel blood vessels in the tumor tissue (**neoangiogenesis**).

Other theories for multistage carcinogenesis The **classical three-step model of carcinogenesis** (initiation, promotion, progression) was launched in the 1950s and is still useful. Subsequently, further theories for multistage carcinogenesis have been developed. For example, as shown by Vogelstein and colleagues, the **development of colon cancer** necessitates mutations of at least four genes on different chromosomes. Colon cancer represents a suitable model to study tumor progression, since histologically distinguishable tumor stages can easily be separated from each other. The functional loss of the tumor suppressor p53 by deletion of the corresponding chromosomal locus on chromsomal 17 or by point mutations in the TP53 gene causes a malignant colon carcinoma.

The **two-hit hypothesis** has been postulated by Knudsen for retinoblastoma, a childhood tumor of the eye. According to this hypothesis, tumor-suppressor genes reveal a tumorigenic effect, if both alleles are inactivated by deletion or mutation. If one allele of the *RB* (*retinoblastoma*) gene on chromosome 13 has already been inactivated in the germ cells, a hereditary predisposition for retinoblastoma can be observed. This mutation is recessive. In somatic cells, an inactivation of the second RB allele causes the development of retinoblastoma. **Loss of heterozygosity (LOH) occurs when** a mutated allele is inherited from both parents, from the paternal and the maternal chromosome. This also leads to the onset of retinoblastoma.

Mutator Phenotype

A mutator phenotype is generated by mutations in genes responsible for the maintenance of genetic integrity. It has been postulated to occur in normal or premalignant cells as an early event during carcinogenesis.

In normal cells, DNA replication and chromosomal segregation are processes of highest precision. In cell culture, about 10^{-10} single base substitutions per nucleotide per cell division and 2×10^{-7} mutations in genes can be observed. Stem cells have even lower frequencies. The fidelity of these processes is also important for cells of the germ line. Curiously, tumor cells contain many thousands of mutations. This speaks against the assumption that rarely occuring, spontaneous mutations in normal cells are sufficient to explain the entire magnitude of genetic changes in tumor cells. Therefore, a mutator phenotype has been postulated as an early event during carcinogenesis. This phenotype is generated by mutations in genes, which are responsible for the maintenance of genetic integrity in normal cells. Mutations in genes encoding DNA polymerases and DNA-repair enzymes increase the susceptibility of DNA for further mutations. They cause the erronous incorporation of bases or the insufficient repair of DNA lesions. This results in genetic instability of tumors at an early stage of carcinogenesis. Thus, the mutator phenotype is linked to mutations in certain target genes. Generally speaking, caretaker and gatekeeper genes have to be considered to fulfil such a function.

Caretakers are proteins, whose encoded genes ensure the fidelity of DNA synthesis or proper and efficient DNA repair. Hereditary defects in **DNA repair genes** are suitable models to study mutator phenotypes. Examples are:

- Xeroderma pigmentosum: Inherited mutations in XP genes, which repair UV-induced DNA lesions by nucleotide-excision repair, and increase the susceptibility for UV-induced skin cancer.
- Hereditary nonpolyposis colon cancer (HNPCC) is associated with mutations in mismatch-repair genes.

The **DNA polymerases** Polα, Polδ, Polε are also caretakers. Polß is associated with base excision repair. Polγ is responsible for mitochondrial DNA synthesis. Polδ plays a major role for DNA replication and $3' \rightarrow 5'$ exonuclease activity. It can remove erronously incorporated noncomplementary bases.

DNA helicases belong to caretaker genes too. They unwind double-stranded DNA before it is read by DNA polymerases. Hereditary mutations in DNA helicases cause syndromes with increased cancer susceptibility:

- **Bloom syndrome**: The DNA synthesis is defective, and many sister chromatid exchanges can be found. This is due to defects in the BLM gene and an insufficient repair of DNA double-strand breaks.
- **Werner syndrome**: Homozygous mutations in the WRN (RecQL1) gene disturb the repair of double-strand breaks and the maintenance of telomeres.

Gatekeeper genes regulate the cell cylce and apoptosis. Typical diseases of this group are:

- **Ataxia telangiectasia** (AT): Cells show defective DNA repair, which is due to mutations in the ATM gene (*ataxia telangiectasia mutated*). ATM is involved in cell-cycle control (see above). AT patients frequently develop a lymphoma.
- **Li Fraumeni syndrome**: Heterozygous mutations in the tumor-suppressor gene *TP53* are inherited over the germ line. If the remaining *TP53 w*ild-type allele is lost in somatic cells, mesenchymal or epithelial tumors develop in early ages.

Clonal selection is another important feature contributing to the mutator phenotype. Mutated cells obtain a selective growth advantage, which leads to the spread of this cell population. DNA lesions cannot become manifest without cellular proliferation. Clonal expansion drives tumor progression and causes further mutations and more aggressive cell types. The clonal selection of mutations in specific genes allows the overcoming of limitations in the host tissue set by restricted oxygen supply, missing growth factors, etc. The mutation frequency and genetic instability increase with every selection round. At this stage, **heterogeneity** of cellular subpopulations, which is typical for tumors, emerges. Tumor progression is irreversible. The magnitude of mutations indicates that the fidelity of DNA replication may decrease to a degree where the limit of cell survival is reached. Indeed, the fraction of apoptotic cells in tumors is much higher than in normal tissues.

Chromosomal Instability

There are various types of chromosomal instability that occur spontaneously and are induced by chemical carcinogens: **translocations, inversions, amplifications**, and **deletions** are changes within one or even several chromosomes. The gain or loss of entire chromosomes can also occur. Changes in the number of chromosomes are termed **aneuploidy**.

Translocations are characterized by the transfer of chromosomal material from one chromosome to another. Two patterns can be distinguished: **Idiopathic translocations** occur by chance and vary from patient to patient regardless of the tumor type. Their number increases in the course of tumor progression. It is unclear to which extent they contribute to the aggressiveness of tumors. More interesting are **specific translocations**. They appear consistently in certain tumor types. They have a causative role for the development of these tumor types. They appear predominately in leukemia, lymphoma, some sarcoma types, but only rarely in the most common solid epithelial tumors types. In contrast, idiopathic translocations can be found in solid epithelial tumors. Important specific translocations are:

- **t(8;14)(q24;q32)**, which activates the *c-MYC* proto-oncogenes in Burkitt's lymphoma;
- **t(14;18)(q32;q21)**, which activates the anti-apoptotic *BCL-2* gene in follicular lymphoma;
- **t(9;22)(q34;q11)**, which generates a fusion protein consisting of the proto-oncogene *c-ABL* and the break point cluster (BCR) region. BCR/ABL is an important signal transducer in acute lymphoblastic and chronic myeloid leukemia.

Inversions Translocations originate from nonrepaired DNA double-strand breaks. Frequently, this is due to a functional loss of proteins involved in double-strand break repair or cell-cycle control, i.e. ATM, ATR, BRCA1, BRCA2, p53, etc. Very likely, nonhomologous end joining of double-strand breaks is involved. Idiopathic and specific forms are also known for inversions. Inversions are characterized by the unhingement of chromosomal segments and the reinstallment in reverse orientation.

Amplifications The **amplification** of proto-oncogenes increases the probability of a cancer cell to survive. Healthy cells recognize gene amplifications as genetic instability, and the tumor suppressor induces apoptosis in such cells. Since many tumors carry mutated *TP53* genes, apoptosis is not induced. Instead, the amplified DNA sequence (**amplicon**) is further amplified in subsequent replication rounds. Gene amplifications are typical for tumors in the progression phase. Examples are the amplification of the epidermal growth factor receptor genes *HER1(EGFR/cErbB1)* and *HER2(neu/cErbB2)* in breast cancer and other tumor types.

Deletions **Deletions** are characterized by a loss of genetic material. The size of a deletion can vary considerably from single bases to long chromosomal parts. Small deletions can cause a shift of the open reading frame during transcription (frameshift mutations) and the translation of inactive or truncated proteins. Parts of genes, entire genes, or several genes can be lost by larger deletions. An example represents the tumor suppressor gene *CDKN2A*, which codes for p16INK4a. Since p16 has an important role in the regulation of cell growth, the deletion of the coding gene causes uncontrolled growth of tumor cells.

Oncogenes and Tumor-suppressor Genes

The targets for carcinogen-induced mutations are proto-oncogenes and tumor-suppressor genes. Mutations critical for carcinogenesis may also occur in DNA-repair genes. Defects in DNA repair predispose to tumor formation similar to mutations in proto-oncogenes and tumor-suppressor genes. Chemical carcinogenesis is a multistep process based on the accumulation of mutations in more than one critical target gene.

Oncogenes Initially, oncogenes were identified in cancer-causing viruses. There are two types of tumor viruses:

- DNA viruses with linear, double-stranded DNA;
- RNA viruses, which carrry a reverse transcriptase to rewrite RNA into DNA.

Viral oncogenes (v-onc) are viral genes that have a sequence similiar to cellular proto-oncogenes. Proto-oncogenes are DNA sequences involved in the regulation of cellular differentiation and proliferation. Mutations in proto-oncogenes may lead to 'true' oncogenes which promote dedifferentiation and uncontrolled proliferation common to cancer. Upon invasion of a cell by RNA-based retroviruses, the viral reverse transcriptase generates DNA, which may be inserted into the DNA of the host cell. Transcription of the incorporated viral genome may lead to escape from regulated growth.

Alternatively, the reverse transcriptase-generated DNA may activate an otherwise latent oncogene to initiate excessive cellular proliferation. The net effect is transduction of a normal into a malignant cell.

The proteins encoded by proto-oncogenes (oncoproteins) are involved in the cellular regulation of proliferation, apoptotis, and differentation signals. These signals interact with the cell via surface rceptors, which transmit the information to the nucleus via signal transduction. The constitutive activation of proto-oncogenes by mutation and genomic instability causes a permanent and uncontrollable activation of the corresponding signaling pathways. According to their cellular localization and biochemical features, oncoproteins can be classified as follows:

- growth factors (examples: Sis, Int-1);
- growth factor receptors (examples: Kit, Met, ErbB1/Her1, ErbB2/Her2);
- cytoplasmic protein tyrosine kinases (examples: Bcr-Abl, Src);
- cell membrane-associates guanine nucleotide-binding proteins (examples: H-Ras, K-Ras, N-Ras);
- soluble cytoplasmic serine threonine protein kinases (examples: Raf, Mos);
- nuclear proteins (examples: c-Myc, c-Fos, c-Jun, Ets);
- anti-apoptotic proteins (example: Bcl-2).

Tumor-suppressor Genes

In contrast to oncogenes, *tumor-suppressor genes* protect from carcinogenesis. Whereas oncogenes act in a dominant manner, tumor-suppressor genes are recessive.

Mutations and deletions inactivate tumor-suppressor genes. While proto-oncogenes are activated sporadically in somatic cells, the inactivation of tumor-suppressor genes can occur sporadically in somatic cells as well as in the germ cells. The origin of familial tumors, as well as inherited diseases with predisposition for cancer, is frequently caused by mutated tumor-suppressor genes. Examples of important tumor-suppressor genes are shown in Table 2.22.

Table 2.22 *Tumor-suppressor genes.*

Gene	Tumor	Hereditary syndrome
TP53	Diverse	Li Fraumeni syndrome
RB1	Retinoblastoma	Retinoblastoma
APC	Colon carcinoma	Familiar adenomatosis polyposis
hMSH2	Colorectal carcinoma	Hereditary non-polypous colon carcinoma
NF1	Fibroma	Neurofibromatosis type 1
NF2	Schwannoma, Menningioma	Neurofibromatosis type 2
MEN1	Insulinoma	Multiple endocrine neoplasia 1
MEN2	Pheochromocytoma	Multiple endocrine neoplasia 2
WT1	Wilms' tumor	Wilms' tumor
VHL	Kidney carcinoma	von Hippel Lindau syndrome

Of the plethora of oncogenes and tumor-suppressor genes, the ras-oncogenes and the TP53 tumor suppressor will be discussed in more detail here.

Ras proto-oncogenes There are four different Ras proteins (H-Ras, N-Ras, K-Ras4A, and K-Ras4B), which share 90% homology. Ras proteins are associated with the inner side of the cell membrane. Their ability to transduce external signals from surface receptors is regulated by the binding of GDP, which inactivates RAS, and by GTP, which activates RAS. Activated Ras molecules bind and activate effector molecules. Mutations in *ras* genes occur at codon 12, 13, or 61. Approximately 30% of all human tumors carry *ras* mutations, which cause permanent uncontrollable activity. Aberrant *ras* activation induces carcinogenicity by various mechanisms, e.g. increased cell proliferation, loss of cell cycle control, decreased apoptosis, increased angiogenesis, invasion, and metastasation.

Another Ras-dependent signal-transduction pathway represents the phosphoinositide-3′-OH-kinase (PI3K)-mediated activation of Akt and Rac. After activation by Ras, PI3K phosphorylates the signal molecules phosphatidylinositol 4,5-bisphosphate to form phosphatidylinositol 3,4,5-trisphosphate (PIP_3). PIP_3 activates Akt, which upregulates the anti-apoptotic transcription factor NF-κB.

Tumor-suppressor gene TP53 *TP53* is a gene that codes for the transcription factor p53, which is a tumor suppressor protein. Mutations in *TP53* result in loss of p53 function. Roughly half of all human tumors contain mutations in the *TP53* gene. Other tumors switch off p53 by amplification of the *MDM2* gene, which is an inhibitor of p53. Thus, the inactivation of p53 promotes carcinogenesis. Various hypotheses, which complement one another, help to explain the anti-carcinogenic action of this gene:

- Termed the 'guardian of the genome,' wild-type p53 protects the genome from mutations causing cancer provoked by genotoxic stress such as UV-radiation, food-borne carcinogens, ionizing radiation, and xenobiotic compounds. The strongest evidence for this hypothesis comes from *TP53* knockout mice that all develop tumors, and patients with inherited *TP53* germ line mutations, who develop a cancer-linked disease known as the Li Fraumeni syndrome.
- Wild-type p53 induces apoptosis, which is a strong deterrent to the development of cancer. Tumors with mutated, inactivated, p53 are resistant towards apoptosis. Tumor cells with mutated p53 exhibit a survival advantage compared with tumors with wild-type p53, which explains the high frequency of *TP53* mutations among human tumors.

Tumor cells can grow more rapidly than blood vessels are formed, which supply them with oxygen and nutrients because they can function better than normal cells under relatively anerobic conditions. In normal cells hypoxia activates wild-type p53, which induces apoptosis. Because mutations in p53 inhibit apoptosis, tumor cells survive under hypoxic conditions.

Apoptosis

Upon treatment with chemical carcinogens, cells may die by apoptosis, necrosis, autophagy, or may undergo 'premature senescence.' Apoptosis appears to be the

major route of cell death triggered by DNA damage. It is executed via a specific cellular program, involving the death receptor and/or the mitochondrial pathway. A key player in DNA-damage-triggered apoptosis is p53. Apoptosis is thought to eliminate genetically damaged cells and, therefore, counteracts carcinogenesis.

The initial cellular reaction to DNA damage is cell-cycle arrest. This provides the cell with time to repair lesions before they are transmitted to the next cell generation. At extremely high amounts of lesions, which supersede the cellular repair capacity, the cell switches to **programmed cell death (apoptosis)**. The death of a damaged cell prevents its malignant deterioration. This keeps the entire organism in a healthy state. Many tumors show a reduced ability to induce apoptosis, i.e. by mutations in apoptosis-inducing genes or permanent activation of apoptosis-repressing genes.

The switch from DNA repair to apoptosis (see the simplified scheme in Figure 2.66) is regulated by the cellular energy balance. Long-running DNA repair consumes much energy in the form of ATP. This energy is not further available for ATP-dependent ion pumps, which maintain low concentrations of Ca^{2+} ions intracellularly. Rising Ca^{2+} levels activate the apoptotic cascade. Since apoptosis is also an energy-consuming process, apoptosis causes the cleavage of ATP-dependent DNA-repair enzymes, i.e. ATM, DNA-PK, and PARP. Thereby, DNA-repair processes are switched off. Since PARP inhibits p53 by poly(ADP)adenylation, the cleavage of PARP activates p53, which in turn induces apoptosis.

With time, high apoptosis rates lead to the selection of apoptosis-resistant cell clones. Apoptosis resistance and insufficient DNA repair cause an accumulation of mutations, an increase of genomic instability, and tumor progression.

Regulation of Apoptosis by p53 The half life of p53 protein is usually short but it may be extended by stabilization, or an increased rate of translation, in the face of genotoxic stress. After DNA damage, p53 arrests the cell cycle in the G1 or G2 phase by transcriptional activation of cell-cycle-regulating proteins. DNA lesions result in extension of p53 activity via the phosphorylation of p53 and the inhibition of the binding of Mdm2 to p53. When DNA repair is completed Mdm2 binds to p53 and the cell cycle continues. However, if cellular repair capacity is saturated, apoptosis may proceed via activation of the *Bax* gene, direct interaction of p53 with mitochondria, degradation of Bcl-2, release of cytochrome c and activation of the apoptosome, or by a nonmitochondrial pathway by activating apoptosis-inducing death receptors.

There are both intrinsic and extrinsic pathways to apoptosis. The intrinsic pathway is activated by mitochondrial membrane damage, whereas the extrinsic pathway involves the activation of death receptors localized at the outer cell membrane. Both pathways can be activated by genotoxic stimuli.

Intrinsic Pathway of Apoptosis The intrinsic pathway is initiated by **mitochondria** and proteins of the **Bcl-2 family**. They are mainly localized in the outer mitochondrial membrane, but to a lesser extent also in the endoplasmic reticulum, and the nuclear membrane. The family is composed of a series of both anti-apoptotic, e.g. Bcl-2, and pro-apoptotic, e.g. Bax, proteins. The Bax protein forms pores in the mitochondrial membrane, which allow the release of **cytochrome *c*** from the mitochondria into the

cytosol. Cytochrome *c* stimulates the formation of the **apoptosome**, a protein complex consisting of Apaf-1 and procaspase-9. Procaspase-9 is autocatalytically cleaved in the apoptosome to caspase-9, which in turn activates downstream caspases required for apoptosis. The permeabilization of the mitochondrial membrane activates further pro-apoptotic proteins and AIF, the apoptosis-inducing factor.

Caspases are cysteine aspartyl proteinases. **Initiator caspases** (caspase-8, -9, -10) cleave and activate downstram **executioner caspases** (caspase-3, -6, -7) to amplify pro-apoptotic signals. Executioner caspases break down many cellular substrates. Laminin cleavage leads to chromatin condensation and nuclear shrinkage. The degradation of some DNA-repair enzymes has already been mentioned. Caspase-3 finally degrades ICAD, the inhibitor of caspase-activated DNAase that in turn becomes activated and cleaves nuclear DNA in nucleosomal size fragments. The degradation of the cytoskeleton proteins actin and plectrin causes cellular fragmentation and the release of **apoptotic bodies**, which are attacked by phagocytes**.** Phosphatidylserine, which is localized at the inner side of plasma membranes of healthy cells, relocates to the outer surface of apoptotic cells and provides a point of attack for phagocytosis. Furthermore, there is a caspase-independent pathway of cell death. It is termed as **autophagy** or **type II cell death**, and is regulated by death-associated proteins (DAPs).

Extrinsic Pathway of Apoptosis This pathway is initiated by the binding of specific death ligands to death receptors. There are at least six death receptors and a number of death ligands.

The ligand called TRAIL binds to the functional receptors TRAIL-R1 and TRAIL-R2, which transduce signals for apoptosis induction. Interestingly, TRAIL also binds to nonfunctional receptors called 'decoy receptors' that may antagonize TRAIL-induced apoptosis.

The formation of a multimer protein complex is crucial for the transduction of apoptosis signals. After binding of a ligand, trimerization of receptors occurs. The adapter molecule ***Fas-associated death domain protein*** **(FADD/Mort1)** as well as procaspase-8 and caspase-10 are involved in complex formation. This complex is termed ***death-inducing signal complex*** **(DISC)**. Procaspase-8 activates itself by autocatalytic cleavage. Active caspase-8 can transduce the signals by two ways:

- Large amounts of caspase-8 are activated in the DISC. Caspase-8 cleaves procaspase-3, and caspase-3 degrades several target proteins by proteolysis.
- Small amounts of caspase-8 are activated in the DISC, and signal transduction is performed via the mitochondria. Caspase-8 cleaves another member of the Bcl-2 family, Bid. Cleavage of Bid allows translocation of Bax to the mitochondria, which permeabilizes the mitochondrial membrane, and cytochrome *c* is released into the cytosol. The apoptosome is generated and further caspases such as caspase-9 and -3 are activated.

Carcinogen-induced Immortalization of Cells

A property of carcinogen-transformed cells is their ability to propagate indefinitely. This phenotype is designated as immortalization. A main mechanism of immortalization rests on upregulation of the enzyme telomerase that provokes telomere lengthening. Usually, tumor cells exhibit high telomerase activity and

long telomeres. Whether telomerase gets upregulated as a direct consequence of carcinogen exposure is an open question.

Normal diploid cells only have a limited doubling capability. They pass over to cellular ageing (senescence) after 50–70 cell divisions. The number of cell doublings depends on the length of telomeres. **Telomeres** have a size of 5–10 kb and consist of repetitive sequences $(TTAGGG)_n$. They cover the ends of chromosomes to protect DNA. During every cell division the telomeres shorten by 50–100 bp, since the DNA-replication machinery cannot replicate chromosomal ends. This phenomenon has been termed the **end-replication problem**. If the telomeres reach a critical length, p53 is activated, and a permanent cell-cycle arrest or apoptosis is induced.

Telomerase counteracts telomere shortening by new synthesis of repetitive TTAGGG sequences. The enzyme consists of a RNA component (TERC), which serves as template for the synthesis of telomere sequences, and a reverse transcriptase (TERT). The activity of telomerase is highly restrictive. Telomerase is mainly active during embryogenesis. In the adult organism, telomerase activity is found in germ cells, stem cells, and activated lymphocytes. Most other mitotically active tissues (gastrointestinal epithelial cells, skin cells, etc.) do not show telomerase activity. A continued telomerase shortening happens as a consequence in these tissues. The telomere length remains constant in mitotically inactive tissues (brain, heart, etc.). Factors such as oxidative stress, epigenetic chromatin alterations, and others accelerate telomerase shortening and thus contribute to faster ageing.

Telomerase is active in over 80% of human tumors. The expression of telomerase in normal cells causes cellular transformation. **Transformation** confers unlimited growth capability on cells. The inhibition of telomerase activity prevents carcinogenesis. Though transformed cells are immortalized, they are not cancer cells. Hence, telomerase is not an oncogenic factor. Nevertheless, **telomerase supports the malignant process of carcinogenesis**. Telomere shortening contributes to cancer initiation. Chromosomes with shortened telomeres show higher genetic instability than do chromosomes with long telomeres. Furthermore, telomerase is expressed as a consequence of p53 inactivation. Telomerase activity also contributes to tumor progression. Since telomere shortening can be induced by chemical carcinogens, e.g. as a result of adduct formation in telomeric sequences and their misrepair, it can be a mechanism for chemical carcinogenesis.

Chromosome Segregation and Aneuploidy

Normal human somatic cells possess 46 chromosomes (diploid chromosome set). Germ cells are haploid and posses 23 chromosomes. Most tumors have more than 46 chromosomes; the numbers vary between 60 and 90. These abnormal chromosome numbers are termed aneuploidy.

Aneuploid tumor cells are frequently more than diploid but less than tetraploid. The explanation is that hyperdiploid sets of chromosomes derive from a duplication of a diploid set of chromosomes (tetraploidy), and a gradual loss of single chromosomes takes place subsequently.

Aneuploidy contributes to the development of an aggressive phenotype during tumor progression, especially by polysomy of chromosomes with activated oncogenes or mutated tumor-suppressor genes. Although most tumors are aneupolid, there is a small fraction of diploid tumors. They reveal specific defects in DNA repair pathways. The mechanisms of aneupolidy are different from those, which lead to the accumulation of DNA mutations. DNA-repair defects lead to increased mutational rates in microsatellite sequences. This phenomenon is termed **microsatellite instability (MIN)**. By contrast, **chromosomal instability (CIN)** causes aneuploidy. It is possible that CIN accelerates loss of heterozygosity of tumor-suppressor genes and duplication of chromosomes harboring oncogenes. Hence, tumor cells may get a growth advantage during carcinogenesis.

Abnormal mitotic spindles: Aneuploid tumors possess supernumerous centromeres. An amplification of chromosomes increases the probability for abnormal mitosis and erronous chromosomal segregation. STK15/Aurora2 represents a centromere-associated serine-threonine-kinase, which is overexpressed in aneuploid tumors. The coding gene can be amplified. Aurora kinases phosphorylate histone proteins and foster the condensation of chromosomes.

Mitotic spindle checkpoints: These checkpoints guarantee the correct adjustment of chromosomes during metaphase and their fixation to the mitotic spindle. Two checkpoints are known:

- During the G_2M checkpoint, microtubule-dependent processes are controlled, i.e. the separation of duplicated centromeres during G_2 phase and delay of the G_2M transition in the presence of tubulin poisons.
- The metaphase checkpoint controls the fixation of the mitotic spindle to the kinetochors. If single kinetochors are not fixed to the spindle, the separation of daughter chromatids is stopped to enable fixation.

Two mutations can be found in genes that regulate the mitotic checkpoints: BUB1, BUBR1, and MAD2.

Neoangiogenesis in Chemical Carcinogenesis

The formation of new blood vessels in tumor tissues is termed neoangiogenesis. It is necessary for tumor growth and for metastasis. Neoangiogenesis is an essential element of tumor progression. Tumors and normal tissues require oxygen and nutrients.

Early in carcinogenesis, tumor growth is limited to about 0.5–1 mm, until such time as neoangiogenesis provides nearby blood vessels for a supply of oxygen and nutrients. The avascular phase terminates and the vascular phase begins when new blood vessels reach as far as 150–200 μm from the tumor. In the absence of neoangiogenesis the tumor undergoes apoptosis.

Tumor neoangiogenesis is regulated by various pro- and anti-angiogenetic factors. If the balance is switched to pro-angiogenic, it is activated by:

- Angiogenic oncoproteins, e.g. Ras, which upregulate the expression of pro-angiogenic proteins such as VEGFs (vascular endothelial growth factors), FGFs (fibroblast growth

factors), PDGF (platelet-derived growth factor), and EGF (epidermal growth factor). Oncoproteins can also downregulate angiogenesis inhibitors such as thrombospondin, endostatin, angiostatin, and tumstatin.
- Mutations in tumor-suppressor genes, e.g. *TP53*. Mutated p53 downregulates the expression of thrombospondin and upregulates vascular endothelial growth factor A. Tumor cells with p53 mutations are selected under hypoxic conditions, since they are apoptosis-resistant and survive.

Tumors also usurp existing vessels in normal tissues by displacement of normal cells. This process is called **vessel cooption**. Whether neoangiogenesis can be driven by chemical carcinogens, thus contributing to tumor progression, is still a matter of dispute.

2.7.4 Nongenotoxic Mechanisms of Carcinogenesis

In recent years, the relevance of nongenetic, so-called epigenetic mechanisms for carcinogenesis has been recognized. Aberrant methylation of cytosines and deacetylation of histones are central epigenetic mechanisms. Furthermore, agents have been unravelled that do not directly damage DNA but promote tumor development by diverse alternative mechanisms, e.g. stimulation of signalling pathways triggering cell-cycle progression or inhibition of apoptosis of initiated cells. So-called nongenotoxic carcinogens, e.g. asbestos and classical tumor promoters (e.g. TPA), may, however, also indirectly damage DNA by endogenous radical formation. DNA-damaging radical burst can also be provoked during inflammation and infection. Sustained oxidative stress together with a chronic stimulus of cell division is an important driving force of tumor formation.

Epigenetics

Apart from genetic lesions, there are epigenetic mechanisms of carcinogenesis. Epigenetic alterations are base-sequence-unrelated genomic changes that are transmitted to the next generation of cells. They are, however, not related to mutations by base substitutions or deletions and translocations, but rather affect the transcriptional cellular program. Epigenetics deals with the propagation of information by regulatory mechanisms of temporally and spatially different gene activities.

The adaptation of organisms to changing environmental conditions by adaptive and hereditary epigenetic alterations challenges the central dogma of biological evolution, which dictates that mutation and selection are the sole driving forces of evolution. Epigenetic mechanisms are **alterations of the chromatin structure** via posttranslational histone modification, nucleosome adjustment, and DNA chromatin complexes as well as **methylation of cytosines** in the DNA. DNA methylation is also involved in ***imprinting***, where homologous genes are differentially expressed depending on their maternal or paternal inheritance. This is in contrast to classical Mendelian genetics.

About 4% of cytosines are methylated in mammalian DNA. The main function of DNA methylation represents the transcriptional regulation of gene expression during development, in different tissues and organs, for imprinting, and for the switch-off of transposon elements. Generally, DNA methylation causes repression of gene expression. Typically, methylated promoter regions suppress gene transcription. In some cases, hypermethylation, which is associated with increased gene expression, occurs in non-promoter regions, The maintenance of different methylation patterns is important for the regulation of normal gene expression.

On the one hand, methylation of promoter regions with high CG contents (***CpG islands***) inhibits the binding of transcription factors. On the other hand, other proteins can bind to these DNA areas and displace the actual transcription factors. Additionally, methylation of CpG islands is coupled with the **deacetylation of histones**. Histones are 'packaging' proteins, which wrap DNA to repetitive nucleosomal units and fold them to chromatin fibers of higher order. Histones are modified at the posttranslational level via acetylation, methylation, phosphorylation, and ubiquitination. The acetylation is controlled by histone acetyltransferases and histone deacetylases. Acetylation is linked to nucleosomal rearrangement processes and transcriptional activation. Deacetylation causes transcriptional repression by chromatin condensation.

DNA methylation and histone deacetylation support each other in terminating gene transcription. The **DNA methyltransferases** DNMT1 and DNMT3b are involved in DNA methylation. *S*-Adenosylmethionine serves as a source for methyl groups. DNMTs recruit histone deacetylases. After the binding of the methyl-CpG-binding protein, MeCP2, to methylated DNA, the SIN3 protein, which is a part of the deacetylase complex, binds to MeCP2. This leads to histone deacetylation of methylated DNA.

Hypomethylated CpG islands are a characteristic feature of many tumor types. DNA hypomethylation during carcinogenesis indicates that carcinogenic chemicals can alter methylation patterns of genes that contribute to the toxic effects of xenobiotics or are oncogenic. For example, cadmium inhibits DNMT1 and thereby causes hypomethylation. Arsenic induces hypomethylation of the *ras* oncogene. Altered DNA methylation and histone acetylation have been observed in nickel-induced carcinogenesis.

The genetic and the epigenetic models of carcinogenesis mutually supplement each other. The epigenome as well as the genome contributes to carcinogenesis by specific alterations in gene expression. Epigenetic changes fit to the classical multiple-step model of carcinogenesis consisting of initiation, promotion, and progression. Potentiation of genetic lesions by epigenetic alterations contributes to the explanation of the age-related increase of cancer incidence. Since epigenetic alterations accumulate during one's life span, they may lead to a transient erosion of DNA-methylation patterns in specific genes involved in carcinogenesis.

Nongenotoxic Carcinogens and Tumor Promoters

The term nongenotoxic carcinogens is a designation for a heterogeneous group of agents that attack DNA indirectly. These agents do not form DNA adducts. Their carcinogenic potential is highly variable; upon acute exposure it is usually lower than the potential of genotoxic carcinogens that act as tumor initiators. However,

looking more closely at DNA damage, it becomes more and more evident that many (if not all) nongenotoxic carcinogens are able to damage DNA indirectly by generating reactive oxygen species (ROS) and reactive nitrogen species (RNS) or by inhibition of DNA-metabolizing enzymes, such as DNA-repair proteins. Indirect induced DNA damage may occur not only in the affected cell, but also in the surrounding cell population (bystander effect). For example, asbestos and wood dust (nanoparticles) may provoke a chronic radical burst in lung macrophages that cause genetic damage in the neighboring cells. If these cells are stimulated to undergo DNA synthesis, mutations may be induced in cells occupying the inflammatory area. Thus, nongenotoxic carcinogens may drive tumor formation by genetic damage they induce indirectly. The effect may even be enhanced by chronic stimulation of cell propagation that supports fixation of mutations. Nongenotoxic carcinogens are, therefore, often at the same time tumor promoters and vice versa.

Since point mutations and chromosomal aberrations have the potential to activate proto-oncogens and inactivate/delete tumor-suppressor genes, all DNA-damaging agents should be considered to act as genotoxic carcinogens. Nongenotoxic carcinogens, some examples of which are given in Figure 2.67, do not directly bind to and damage DNA. However, even without any reactivity towards DNA, DNA damage may be provoked by them indirectly. The most common way in which this may occur is radical formation. During chronic inflammation, because of infections or exposure to particles or fibers, macrophages and granulocytes respond with an oxygen burst that is strong enough to kill infectious particles. ROS and RNS produced in inflammatory tissues may damage cellular DNA, forming, among other lesions, 8-oxoguanine that leads to mutations (due to mispairing with adenine). Asbestos permanently activates lung macrophages, causing chronic inflammation and presumably DNA damage. Another example is the

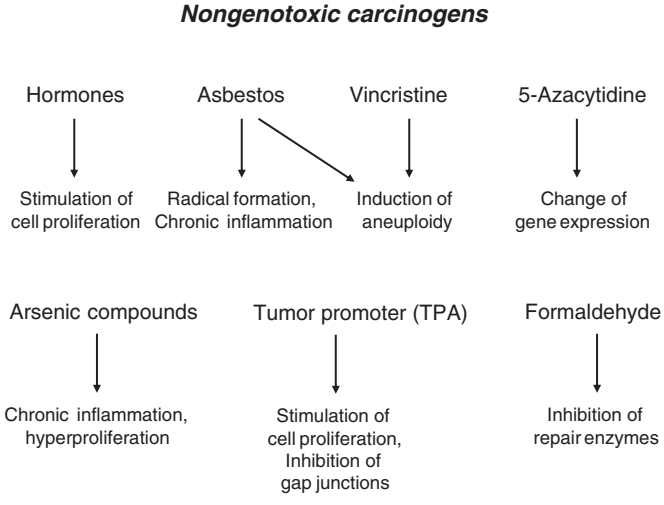


Figure 2.67 *Examples for nongenotoxic carcinogens and tumor promoters.*

tumor promoter TPA that promotes cell division by stimulation of the PKC-driven ERK pathway targeting c-Fos/AP-1. However, TPA was also shown to be able to induce chromosomal aberrations. The mechanism is unknown; it is likely to be related to the oxygen burst induced by TPA in the cell. Some carcinogenic metals such as arsenic do not directly bind to DNA. However, they are able to inhibit DNA-repair enzymes and, therefore, may increase the 'spontaneous' level of DNA damage. Given the fact that several hundred 8-OxoG and ~10 000 apurinic sites are formed per cell per day, inhibition of DNA-repair enzymes by carcinogenic heavy metals will clearly contribute to the endogenous genotoxic stress level. These examples demonstrate that a clear distinction between genotoxic and nongenotoxic carcinogens is often arbitrary.

It should be noted that agents that are potent in inducing point mutations are clearly positive as tumor initiators, whereas most nongenotoxic carcinogens rather act as tumor promoters. Therefore, agents and exposures that do not directly attack DNA and have a low potency to induce point mutations should be considered as tumor promoters. This group encompasses not only the classical tumor promoters TPA, phenobarbital, and chlorinated hydrocarbons (DDT, TCDD) but also hormones (estradiol, diethylstilbestrol) and peroxisome proliferators. All they have in common is that they stimulate cell proliferation, which is required for fixation of mutations and the expansion of initiated cells.

Not all genotoxic agents are potent tumor initiators. A classical example is methyl methanesulfonate (MMS), which alkylates DNA to form predominantly *N*7-methylguanine and *N*3-methyladenine. These types of alkylating lesions are not mispairing. They give rise, however, to apurinic sites and, therefore, MMS is a very weak mutagen but rather a strong clastogen. Interestingly, MMS is quite potent in promoting tumor formation in mouse skin. Thus, given after a single dose of 7,12-dimethylbenz[*a*]anthracene (DMBA), MMS is as effective as TPA in tumor promotion. Both TPA and MMS induce chromosomal aberrations in skin keratinocytes. This has been taken to suggest that chromosomal aberrations together with stimulation of proliferation may drive tumor promotion, e.g. by loss of tumor-suppressor genes. MMS provides an example to show that even clear genotoxic agents are not necessarily tumor initiators and, if they were applied on their own, would be only very weakly carcinogenic.

Should tumor promoters be considered as nongenotoxic carcinogens? If an agent has been experimentally proven not to be able to induce tumors on its own, it is, per definition, a tumor promoter and should not be considered as 'carcinogen.' However, chronic application of a promoter may also cause tumor formation, although at a low frequency compared with the effect obtained if it is applied in combination with an initiator. Whether the slight tumorigenic effect of tumor promoters is due to fixation and expression of DNA lesions that were brought about by endogenous DNA-damaging metabolites or to the indirect genotoxic effect of 'promoters' (e.g. by endogenous radical formation) remains an open question.

2.7.5 Implications of Initiation and Promotion for Risk Assessment

Notably for purposes of risk assessment it is important to distinguish between genotoxic and nongenotoxic carcinogens. In contrast to genotoxic carcinogens, for nongenotoxic carcinogens (tumor promoters) the effect is thought to be reversible and detectable only above a given dose threshold. Even if nongenotoxic carcinogens induce genetic changes

indirectly, the level might be quite low and insufficient to induce cancer if the promoting agent were to be applied on its own. Therefore, for practical purposes most, if not all, nongenotoxic carcinogens should be considered as tumor promoters that exhibit a dose threshold in order to induce, upon chronic exposure, malignant transformation of cells.

2.7.6 Summary

Cancer-causing agents are termed carcinogens, and carcinogenic compounds that damage DNA are designated as being genotoxic. DNA alterations transform normal cells into tumor cells. Changes in RNA, proteins, and lipids are not causatively related to the development of cancer. Under certain circumstances they can, however, support malignant growth.

The knowledge of molecular mechanisms of carcinogenesis has increased dramatically over the past years. Nevertheless, despite the sequencing of the human genome we do not understand the full range of molecular mechanisms of carcinogenesis. Apart from DNA alterations, transcriptional and translational mechanisms (alternative splicing) and post-translational modifications (phosphorylation, glycosylation, acetylation, methylation, ubiquitinylation, etc.) must be taken into account in elucidating the spectrum of carcinogenic and tumor-promoting mechanisms.

Although many carcinogenic and co-carcinogenic compounds have been identified in the past few decades and successful protection is possible now, personal risk assessment remains difficult to perform.

Pharmacogenetics (single nucleotide polymorphisms) and the novel '-omics' technologies (genomics, proteomics, metabonomics) may help to define patterns specifically linked to carcinogenic processes or individual risks. It is to be expected, however, that this new dimension of molecular analyses will raise even more questions than it can answer.

A new discipline at the horizon is systems biology. It attempts to integrate the vast amount of data generated by the '-omics' technologies and to generate not only models to describe mechanisms and diseases, but also to develop predictive bioinformatic models for carcinogenic risk assessment at the level of each single individual.

For Further Reading

Benigni, R. Structure-Activity Relationship Studies of Chemical Mutagens and Carcinogens: Mechanistic Investigations and Prediction Approaches. *Chem. Rev.*, 2005, **105**, 1767–1800.

Christmann, M., Tomicic, M.T., Roos, W.P., and Kaina, B. Mechanisms of Human DNA Repair: an Update. *Toxicology*, 2003, **193**, 3–34.

Dixon, K., Kopras, E. Genetic Alterations and DNA Repair in Human Carcinogenesis. *Semin. Cancer Biol.*, 2004, **14**, 441–448.

Goodman, J.I., Watson, R.E. Altered DNA Methylation: a Secondary Mechanism Involved in Carcinogenesis. *Annu. Rev. Pharmacol. Toxicol.*, 2002, **42**, 501–525.

Kim, Y.I. Folate, Colorectal Carcinogenesis, and DNA Methylation: Lessons from Animal Studies. *Environ. Mol. Mutagen.*, 2004, **44**, 10–25.

Klaunig, J.E., Kamendulis, L.M. The Role of Oxidative Stress in Carcinogenesis. *Annu. Rev. Pharmacol. Toxicol*., 2004, **44**, 239–267.
Roos, W.P., Kaina, B. DNA Damage-induced Cell Death by Apoptosis. *Trends Mol. Med*., 2006, **12**, 440–450.
Valko, M., Rhodes, C.J., Moncol, J., Izakovic, M., Mazur, M. Free Radicals, Metals and Antioxidants in Oxidative Stress-induced Cancer. *Chem. Biol. Interact*., 2006, **160**, 1–40.
Williams, G.M. Mechanisms of Chemical Carcinogenesis and Application to Human Cancer Risk Assessment. *Toxicology*, 2001, **166**, 3–10.

2.8 Reproductive Toxicology

Horst Spielmann

2.8.1 Introduction

Reproductive toxicology covers a wide spectrum of toxic effects on all segments of the reproductive cycle (Figure 2.68) starting with female and male fertility, pre- and postnatal effects, and also late manifestations that can only be detected in the next generation. Thus, in contrast to other field of toxicology, reversible and irreversible effects of exposure to toxicants may occur not only in the individuals who were exposed but also in their offspring. Owing to the Thalidomide/Contergan® disaster in Germany in 1960 the main interest in reproductive toxicology is focused predominantly on adverse effects on the embryo; these side effects are covered by **embryo-/fetotoxicity** and **teratogenicity**. Moreover, malformations and intrauterine death are easier to detect and are raising more concern than is reduction of female and male fertility in the offspring.

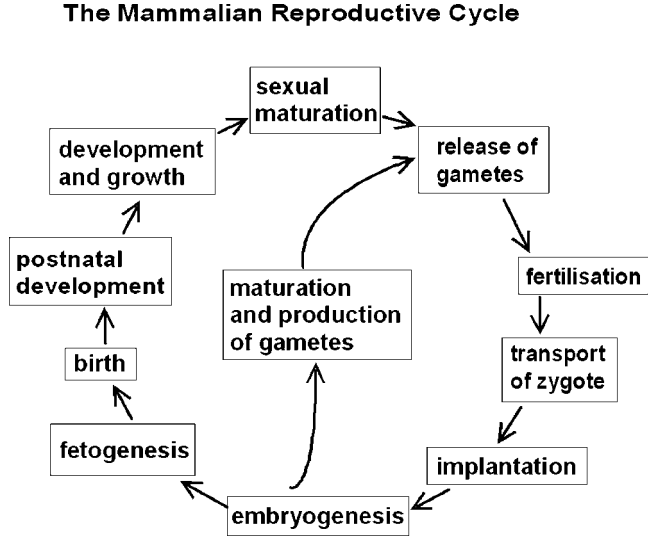


Figure 2.68 *The mammalian reproductive cycle.*
Effects of toxic agents on each stage of the reproductive cycle can be analysed in laboratory animal studies. This is more difficult in humans, where the generation time is 20–25 years.

To evaluate the effects of exposure to chemicals and other agents on all segments of the reproductive cycle two approaches have proven successful, viz. epidemiological investigations in exposed human populations and studies in experimental animals, which are conducted according to internationally harmonised guidelines.

Experience gained over more than 40 years confirms for all areas of reproductive toxicology that clear cut dose–response relationships exist similar to all other areas of toxicology.

2.8.2 Characteristics of Reproductive Toxicology

Reproduction is essential to maintain the survival of all species. Mammalian reproduction is characterised by a rather complex reproductive cycle, which has allowed the genetic material of individual species to adapt to a continuously changing environment. Human reproduction is characterised by the same essential features and has allowed adult individuals around the globe to survive under quite complex environmental conditions. The mammalian reproductive cycle is in principle, identical in all mammalian species; the essential steps are depicted in Figure 2.68. To adapt to the environment, all species have developed species-specific deviations from the general pattern.

As a general rule physical and chemical agents can adversely affect each of the essential steps of the mammalian reproductive cycle.

Some essential features of reproductive toxicology are quite different from all other areas of toxicology, e.g.:

- Adverse effects may only occur in the next generation/offspring;
- Adverse effects on the vital functions of all organs may also indirectly induce adverse effects on the reproduction.

Agents are termed 'reproductive toxicants,' when they interfere with the reproductive cycle at a dose range that does not induce adverse effects in any other organ systems.

In the field of reproductive toxicology, risk assessment is challenging due to the complexity and unusually long time frame/duration of the reproductive cycle, which covers at least two generations. To assess the potential of chemicals to interfere with reproduction, test methods in experimental animals must cover the essential steps of reproduction:

- Fertilisation, e.g. fusion of oocyte and sperm, resulting in a complete, diploid set of chromosomes;
- Normal cleavage divisions, implantation, intrauterine development, birth and postnatal development throughout the period of lactation;

- Normal development of the offspring to fertile adult animals, which then produce a second healthy generation.

For systematic reasons the essential stages of the reproductive cycle, which are important for female and male fertility, e.g. maturation and development of ooctes and sperm up to fertilisation, will be described in this chapter separately from pre- and postnatal toxicology.

2.8.3 Adverse Effects on Female and Male Fertility

Spermatogenesis

Spermatogenesis covers the maturation of sperm from its immature progenitor cells.

Spermatogenesis, which is dependent on sex hormones, starts at puberty and it is localised in the tubuli of the testis. The stem cells of mammalian spermatogenesis, the spermatogonia, are attached to the wall of the seminiferous tubuli of the testis. At puberty these cells start to cleave into primary spermatocytes, which migrate to the centre of the seminiferous tubuli and contain a normal, diploid set of chromosomes. During the next step of maturation the diploid chromosome set is reduced to a haploid set in secondary spermatocytes. In the subsequent stage spermatids continue to differentiate into mature sperm, which in contrast to the earlier stages are characterised by neck and tail. The mature sperm is located in the centre of the seminiferous tubuli and migrates to the epididymis, where it is stored and will leave the testis via the seminiferous tubuli.

Sertoli cells are also located in the tubuli of the testis and are attached to its wall. They not only secrete hormones and proteins essential for the maturation of sperm but also provide the morphological substrate of the blood–testis barrier, which prevents the transfer of large and polar molecules into the seminiferous tubuli. Leydig cells, which are located in the interstitial tissue between the seminiferous tubuli, produce the male sex hormone testosterone under the control of luteinising hormone (LH) of the hypophysis. In contrast to oocyte maturation, spermatogenesis is a continuous process and thus, in principle, men are fertile from puberty up to very old age.

The high rates of cellular division and metabolic activity associated with spermatogenesis are the basis for the susceptibility to certain types of damage. During the duplication of genetic material and cell division, DNA is particularly vulnerable to damage. In addition, many specialised cellular proteins and enzymes are needed during spermatogenesis. Therefore, chemicals that may cause DNA damage or interfere with cellular protein functions are of particular concern in rapidly dividing tissues, e.g. reactive electrophilic chemicals such as alkylating agents and ionising radiation.

Maturation of Oocytes (Oogenesis)

At birth the human ovary contains around 300,000 ovarian follicles and by puberty they will have decreased to around 50% of that number. During the 30 fertile years of a

woman between puberty and menopause ca. 400 mature oocytes will develop, while after menopause no follicles containing oocytes are left in the ovary.

During the four weeks of a normal ovarian cycle only a single oocyte will develop into a primary follicle. Oocyte maturation is stimulated under the control of follicle-stimulating hormone (FSH) secreted by the hypophsis. FSH also stimulates the maturation of the uterine mucosa to provide nutrition for implanting of the preimplantation embryo that has developed from the fertilised oocyte and which migrates into the uterus. After ovulation, growth of the uterine mucosa is stimulated by LH of the hypophysis and, if the oocyte has not been fertilised, it is rejected/expelled every month with the uterine mucosa during the periodic bleeding.

The synchronised maturation cycles of ovary and uterine mucosa are controlled by gonadotropin-releasing hormone (GnRH) of the hypothalamus.

The small proliferative cell population in the ovary and the intermittent nature of the proliferative stage makes the ovary less susceptible to disruptions of cell division. The most significant type of toxicants for female reproductive function are those that interfere with the dynamic endocrine balance required for folliculogenesis and oogenesis. However, the impact of endocrinologically active chemicals through environmental exposure (‘*endocrine disruptors*’) on human reproduction is not yet clear.

Fertilisation

In mammals, fertilisation, the fusion of female and male gametes – oocyte and sperm – , occurs in the ovarian tube. At fertilisation both oocyte and sperm contain only a haploid set of chromosomes (50% of the number found in ‘normal’ somatic cells).

The reduction of the diploid set (100%) of chromosomes to a haploid set (50%) during the two maturation divisions of each oocyte is termed ‘**meiosis**’, and this term is also used for the two maturation divisions during spermatogenesis, when spermatocytes develop into spermatids (see above). The meiotic divisions of oocyte and sperm ensure that mother and father are each contributing 50% of the chromosomes to the fertilised zygote and thus to the embryo. The sex of the embryo is controlled by the distribution of the X- and Y-sex chromosomes; in humans XX individuals are female and XY are male. The basic physiological parameters controlling human fertilisation are so well established that for more than 20 years in vitro fertilisation has been the standard therapy for many infertile couples.

Cleavage Divisions and Implantation

After fertilisation the zygote undergoes cleavage divisions of preimplantation development while migrating from the ovarian tube to the uterus. During cleavage divisions all

cells of the preimplantation embryo are not yet committed; they are omnipotent embryonic stem cells and can differentiate into all tissues of the embryo. After implantation the number of embryonic stem cells decreases while they are differentiating into specific cells and tissues.

During early embryonic development before implantation in the uterine mucosa the embryo is rather insensitive to toxic agents, in particular since the embryonic stem cells forming the preimplantation embryo have a high capacity to replace damaged embryonic cells.

When exposed to high doses of toxic agents preimplantation embryos are not able to repair damaged cells, and they die. At implantation into the uterine mucosa on day 6–7 after fertilisation the early human embryo has reached the blastocyst stage and consists of 80–160 cells, among which only the inner cell mass still contains omnipotent embryonic stem cells.

Adverse Effects on Female and Male Fertility

As a general rule all chemicals that interfere with the balance of female and male sex hormones can adversely affect the maturation of female and male germ cells and thus reduce fertility in animals of both sexes.

Morphological parameters are routinely used to study adverse effects on fertility, both in humans and in animal studies. Genotoxic chemicals and physical agents may reduce fertility by interfering with the maturation of sperm and oocytes, in particular during meiotic division. Most of the early embryos exposed to toxic doses of genotoxic chemicals, X-rays and radioactive irradiation will die early in pregnancy. Such adverse effects will be detected in the dominant-lethality test, which is described in Chapter 4.2A on mutagenesis.

Evidence from epidemiological studies conducted on men and women exposed for several years to industrial chemicals at safe levels recommended for occupational exposure suggests that female and male fertility were not impaired.

To assess the reproductive toxicity of chemicals and drugs, a series of animal tests have been developed, which have been harmonised at the international level and which show some differences depending on the field of application. In drug testing, specific segments of the reproductive cycle are evaluated, e.g. segment 1, segment 2, and segment 3 studies as outlined in Table 2.23 and Figure 2.69. These tests have been harmonised by the International Conference on Harmonisation (ICH), which is formed by the drug industry and regulatory agencies of the three major economic zones Europe, Japan, and USA. For all other chemicals, e.g. industrial chemicals, cosmetics, food additives, pesticides, and

Table 2.23 *International Test Guidelines in Reproductive Toxicology (EU, OECD, ICH).*

Drugs (ICH)	Industrial and other Chemicals (EU, OECD)
Segment 1 Combined fertility and embryotoxicity study	Teratology study
Segment 2 Teratology study	1-Generation study
Segment 3 Pre- and postnatal study	2-Generation study
	Developmental neurotoxicity study

biocides, the Organisation for Economic Cooperation and Development (OECD) has harmonised the reproductive toxicity tests, which include an embryotoxicity test, 1- and 2-generation studies (Figure 2.70), and more recently also a developmental neurotoxicity study, as indicated in Table 2.23.

2.8.4 Pre- and Postnatal Toxicology

Principles of Drug Effects in Pregnancy

More than 40 years after the Thalidomide/Contergan® disaster in Germany the principles of the action of drugs and other chemicals in pregnancy are well established. They were derived from animal experiments by James Wilson in 1977 and, meanwhile, they were confirmed by epidemiological studies for human pregnancy. These principles will, therefore, be outlined in more detail.

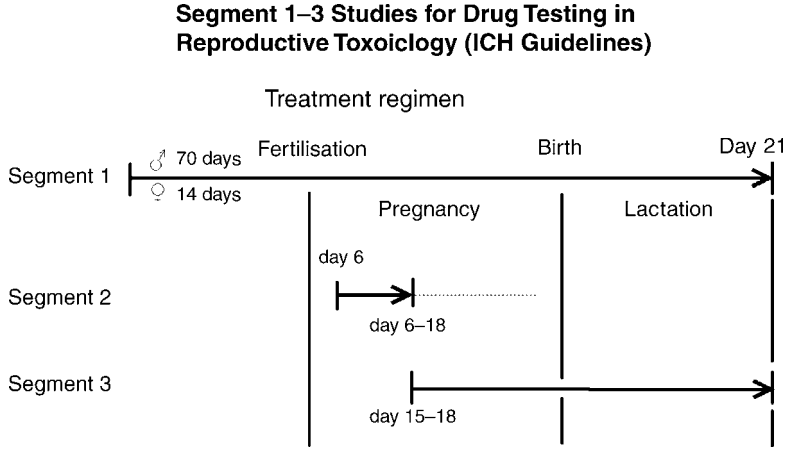


Figure 2.69 *Reproductive toxicity testing of drugs - ICH segment 1–3 studies. Arrows indicate period of treatment (days). ICH (International Conference on Harmonisation), Segment 1 = combined fertility and embryotoxicity study; Segment 2 = embryotoxicity/teratology study; Segment 3 = peri- and postnatal study.*

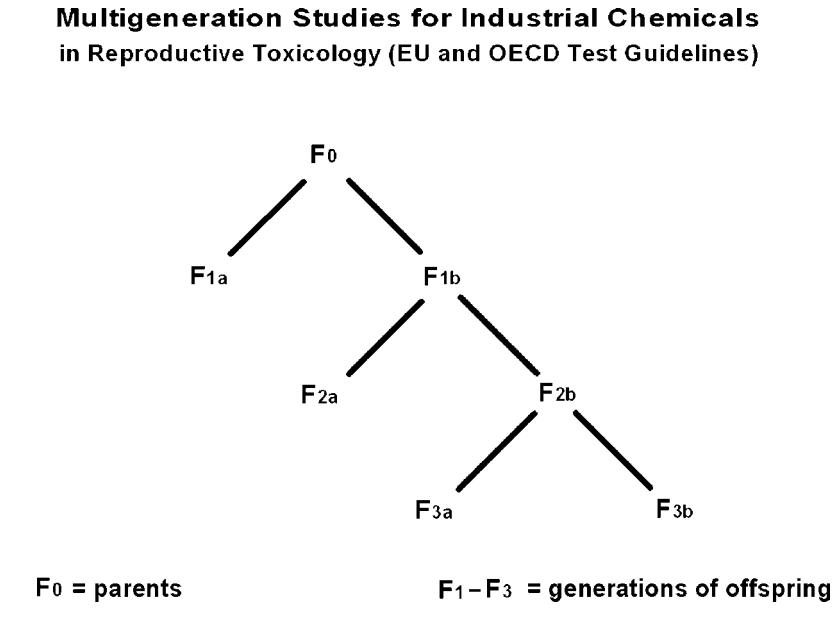


Figure 2.70 *Multigeneration studies in chemicals testing (OECD). F_0 = parental generation; F_{1b} to F_{2b} = first and second generations, which are continuously exposed to toxic agents; F_{1a}, F_{2a}, F_{3a}, and F_{3b} = first to third generations, which are examined for morphological and functional effects on development.*

First principle: *The susceptibility to embryotoxic agents depends on the genotype of the conceptus and the manner in which this interacts with environmental factors.*

In simple terms the genetic background and extrinsic factors contribute to the outcome of pregnancy.

Second Principle: *The susceptibility to embryotoxic agents varies with the developmental stage at the time of exposure.*

According to this principle the sensitivity of the embryo/fetus changes during pregnancy in the following manner: As described earlier, the preimplantation embryo is rather insensitive to toxic agents. However, after implanting in the uterus the sensitivity of the embryo increases during the period of organogenesis, when organs and limbs are developing. Thereafter the sensitivity decreases continuously until birth. Therefore, malformations are generally induced during organogenesis while the organs are developing. In humans this period covers the time frame from day 15 to day 60 after fertilisation. Thereafter, during the period of fetal development functional defects are induced, e.g. functional defects of brain development may result in behavioral anomalies.

Third Principle: *There are only a few final manifestations of abnormal development both in laboratory animals and in man:*

- *Normal development, due to the active repair capacity of the fast growing embryo*;
- *Intrauterine death of the embryo*;
- *Growth retardation without any sings of abnormality*;
- *Malformations*;

- *Functional disorders without any signs of abnormality*;
- *Tumors induced via 'transplacental' carcinogenesis.*

The first point has been confirmed in animal experiments, since even during organogenesis embryos have a high capacity to repair damaged cells and thus at birth no signs of abnormal development can be detected. The sixth point refers to unsuccessful attempts to prevent abortion by treatment with diethylstilbestrol, a synthetic female sex hormone, which unexpectedly induced cancer of the vagina after puberty in daughters of mothers who had been treated with diethylstilbestrol during pregnancy.

Fourth Principle: *The access of adverse environmental influences to embryonic tissues and organs depends on the physical and chemical nature of the agent.*

It has to be realised that the maternal organism is protecting the embryo/fetus against most of toxic chemical and physical agents except X-rays. The chemical properties of a toxic chemical and its metabolites in the maternal organism will have an important impact on the amounts that will in the end reach the embryo/fetus.

Fifth Principle: *Manifestations of abnormal development increase in degree as dosage increases from the no-effect level to lethality.*

Thus, increasing doses of an embryotoxic agent will increase the adverse effects in the embryo/fetus in a dose-related manner. Depending on the dose to which the mother is exposed, adverse effects are induced; they increase continuously from the nontoxic dose level at the lower end to embryo- and fetotoxic effects and embryolethality at the highest dose levels (for details see Figure 2.71).

The effects of toxic agents in pregnancy are dose-related as usual in pharmacology and toxicology.

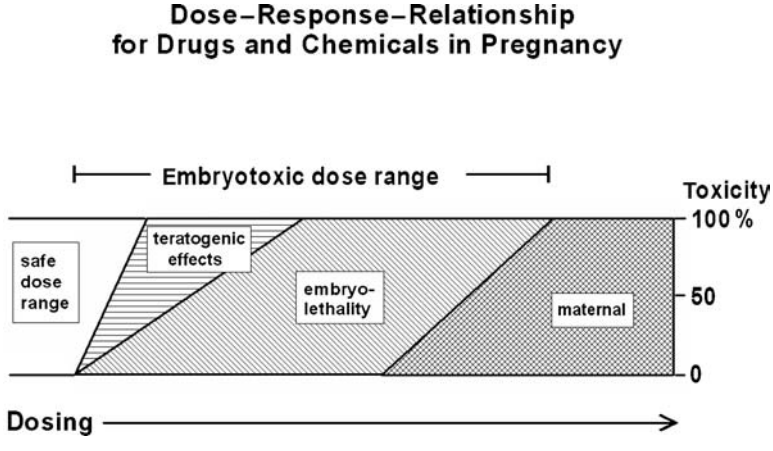


Figure 2.71 *The dose–response relationship in prenatal toxicology. Depending on the dose to which the mother is exposed, adverse effects are induced in the embryo/fetus in a dose-related manner. They increase from the nontoxic dose level to embryotoxic effects and embryolethality at the highest dose levels, which finally will also induce maternal toxicity.*

At low doses neither embryo or mother will be damaged. When increasing the dose the embryotoxic/teratogenic range is reached first and malformations may be induced. When increasing the dose level even higher, the embryo's lethal dose range is reached and the embryo/fetus will die and not survive to term. At the highest doses maternal side effects will be observed.

The embryotoxic risk of drugs and other chemicals is high when adverse effects are induced in the embryo/fetus at or even below the therapeutic dose range of the mother. This situation occurred in the Thalidomide/Contergan® disaster, since the drug had no side effects in adults and was specifically prescribed by doctors as a sleeping pill in pregnancy. It was, therefore, highly unexpected that a single dose of the drug taken in week four or five of pregnancy could induced severe limb malformations (phocomelia) in the children.

Metabolism of Drugs and other Chemicals in Pregnancy

Drug metabolism in pregnancy is more complex than during any other period of life.

The following factors influence metabolism of a drug or chemicals and its metabolites:

- Pharmacokinetics and metabolism in the mother, who is exposed first;
- Placental transfer;
- Distribution in the embryo/fetus;
- Drug metabolism in the embryo/fetus;
- Excretion by the embryo/fetus;
- Excretion by the mother.

In the past it was assumed that the placenta is a barrier protecting the embryo against toxic effects induced by drugs and chemicals. Advanced analytical methods have shown that almost all drugs and chemicals as well as their metabolites will reach the circulation of the embryo/fetus. The passage of toxic agents through the placenta is quite similar to transfer through the intestinal wall, since lipid-soluble materials will easily pass through this barrier, while a much smaller percentage of water-soluble, ionised materials will pass into the blood of the embryo/fetus.

According to current knowledge, drugs that are readily absorbed after oral application will easily pass the placenta and reach the embryo/fetus, while drugs that have to be injected will pass the placenta at a significantly lower rate.

Molecular weight also has a high impact on placental transfer of drugs and chemicals. Up to a molecular weight of about 800 daltons they will easily pass the placental barrier. Since the majority of drugs and their metabolites belong to this group, most of them will reach the embryo/fetus and may induce abnormal development.

On the other hand the placenta is a tight barrier to complex molecules, which may, therefore, not reach toxic concentrations in the embryo/fetus, e.g. steroid and peptide

hormones, insulin, and even the mushroom toxin α-amanitin. The placental transfer of drugs with a high binding affinity to maternal serum proteins is slow. Drug metabolism activity of the placenta is low and does not, therefore, significantly contribute to drug metabolism in pregnancy.

Regulatory Testing in Pre- and Postnatal Toxicology The internationally harmonised regulatory tests in reproductive toxicology are summarised in Table 2.23. The table shows that for industrial and other chemicals a teratogenicity study has to be conducted according to EU and OECD guidelines, while for drugs according to ICH guidelines in addition to the segment 2 (teratogenicity) study, a pre- and a postnatal study (segment 1 and segment 3 studies) have to be conducted. On the other hand, for industrial chemicals 1- and 2- generation studies have to be performed in order to detect postnatal effects. More recently an additional developmental neurotoxicity study is proposed for industrial chemicals and a new OECD guideline has been drafted for these tests.

Effects of Drugs and other Chemicals in Pregnancy

Epidemiological studies suggest that drugs and other chemicals, including food additives and recreational/life style drugs, account for only 4 to 5% of malformations in humans (Table 2.24).

Thus despite intensive research in prenatal toxicology, today 60–70% of malformations in humans are due to unknown intrinsic or extrinsic etiology. The lack of an accepted mechanistic hypothesis for drug/chemical-induced malformations in humans and laboratory animals has so far not allowed us to sufficiently predict the outcome of

Table 2.24 *Estimates of causes of developmental disorders in man (%).*

Unknown spontaneous developmental disorders	65
Genetic diseases	20
Chromosomal disorders	5
Anatomical factors Uterus anomalies Twin pregnancy Oligohydramny	2
Chemical and physical agents Medicinal products Drugs of abuse Ionising radiation Environmental chemicals	4
Maternal diseases Epilepsy Diabetes melitus Hypothyroidism Phenylketonuria	4
Infectious diseases (cytomegaly, listeriosis, syphilis, rubella, toxiplasmosis, varicella)	

Table 2.25 *Drugs and chemicals with proven embryotoxic/fetotoxic potential in man (%).*

Agent	Features
ACE inhibitors	Anuria
Alcohol	Fetal alcohol syndrome
Androgens	Masculinisation
Antimetabolites	Multiple malformations
Benzodiazepines	Floppy infant syndrome
Lead	Cognitive developmental retardation
Carbamazepine	Spina bifida, multiple malformations
Coumarin anticoagulants	Coumarin syndrome
Diethylstilbestrol	Vaginal dysplasia and neoplasms
Ionising radiation	Microcephaly, leukemia
Iodine overdose	Reversible hyperthyroidism
Cocaine	CNS, intestinal and kidney damage
Methylmercury	Cerebral palsy, mental retardation
Misoprostol	Reduction of limbs
Polychlorinated biphenyls	Mental retardation, immune defects
Penicillamine	Cutis laxa
Phenobarbital/primidone	Multiple malformations
Phenytoin	Multiple malformations
Retinoids	CNS, skeletal, cardiovascular disorders
Tetracycline	Discoloration of teeth
Thalidomide	Malformations of the limbs
Trimethadione	Multiple malformations
Valproic acid	Spina bifida, multiple malformations
Vitamin A ($>25\,000$ IU/day)	see retinoids

pregnancy. To improve this situation, teratology information centres have been established in many countries. They have set up networks of collaboration, e.g. in Europe ENTIS (European Network of Teratology Information Services) and worldwide OTIS (Organisation of Teratology Information Services). Most of these institutions are conducting multicentre studies on patients who have been interviewed during counselling in order to obtain robust data on the risk of exposure to drugs and other chemicals in pregnancy.

Table 2.24 summarises the estimation of the causes of abnormal development in man 45 years after the Thalidomide/Contergan® disaster. The table shows that inheritable genetic disorders account for around 20% of malformations and that disease of the mother, including infections, accounts for an additional 5–7%. In man the percentage of malformations that may be induced by drugs and other chemicals is fairly low, at about 4%.

Table 2.25 provides for a more detailed analysis of individual drugs and other chemicals with proven embryo-/fetotoxic potential in humans. Most of the proven embryo-/fetotoxic chemicals in this table have been identified by carefully designed epidemiological studies on pregnant women, who have been exposed in pregnancy, e.g. to alcohol, cocaine, diethylstilbestrol, methylmercury, as well as vitamin A and the retinoids.

The percentage of malformations that are induced by drugs and other chemicals has not increased during the past 40 years. This proves that the current regulatory test methods in reproductive toxicology are sufficiently sensitive to protect the human embryo/fetus against risks arising from adverse chemicals.

Although this result is quite welcome and assuring, it must be taken into account that due to the long generation times in humans in comparison with experimental animals, we usually have only limited information on adverse effects in the next two generations. The situation for Thalidomide/Contergan® is a special one, since the drug does not induce genetic disorders. However, until today all of the malformed Thalidomide/ Contergan® children have had normal children of their own.

Effects of Drugs and other Chemicals on Lactation

The concentration of most drugs and chemicals in human milk does not reach therapeutic or even toxic dose levels.

Today concentrations of drugs and other chemicals as well as their metabolites can be determined in human milk due to the high sensitivity of advanced analytical methods. However, only in very few exceptional cases have toxic concentrations been measured. Newborns and premature babies are very sensitive to toxic agents, since the liver is immature and the rate of drug metabolism is low while they are getting adapted to extrauterine life. In particular, the clearance of toxic agents from the blood by metabolising enzymes (cyps) of the liver and excretion by the kidney are insufficiently established in these babies. Moreover, it takes several days in normal babies until the function of the blood–brain barrier is established, which protects the brain against the penetration of intrinsic and foreign molecules that may induce irreversible damage.

During the first days of life human milk provides the newborn with specific immunoglobulins, which are not destroyed in the immature gastrointestinal tract but are instead absorbed and protect the newborns against infections during the first weeks of life, while the immune system of the baby is immature and not yet able to produce protective antibodies.

Chemicals are secreted from the blood into the milk and vice versa. The transfer of molecules into the milk is favored by high lipid solubility, low molecular weight (< 200 daltons), and a low binding to maternal serum proteins.

Most organochlorine compounds, which are persistent in the environment, e.g. pesticides, polychlorinated biphenyls, and dioxins, and which are stored in the human fat tissue are mobilised during lactation and accumulate in the mother′s milk. Although the newborn is particularly sensitive to these toxic agents, epidemiological studies have shown that breastfeeding had no adverse effects on the development of children, including their intellectual performance.

The advantages of breastfeeding outweigh the disadvantages, in particular as far as the transfer of anti-infectious immunoglobulins is concerned. Therefore, today around the globe mothers are encouraged to breastfeed their babies.

Endocrine Disruption

There is little doubt that exposure of certain wildlife to high concentrations of synthetic chemicals has produced reproductive and developmental effects. However, scientists disagree as to which chemicals or environmental factors may be responsible and whether the effects are due to hormonal mechanisms or other types of toxicity. There is even greater controversy as to whether adverse effects are also occurring in humans and wildlife species at low exposure levels (*'endocrine disruptor' hypothesis*). Despite positive results in laboratory assays, few chemicals – e.g., organochlorine pesticides, dioxins, and phthalate and phenolic plasticizers – have been shown to produce adverse developmental outcomes in exposed humans. Another highly disputed consequence of endocrine disruption in humans is reduced sperm counts in men in industrialised countries. More definitive laboratory studies and risk assessments for a number of these chemicals indicate little or no potential for adverse effects in humans at environmentally relevant exposure levels.

2.8.5 Summary

Reproductive toxicology (RT) covers all adverse effects on the reproductive cycle, including those on female and male fertility, on development of the embryo, such as growth retardation, malformations, and death, and also on the induction of adverse postnatal effects during the period of lactation as well as impaired development of the next generation. Therefore, specific characteristics of sperm and oocyte maturation were described, as well as the specific stages of prenatal development of the mammalian embryo. Differences in sensitivity during prenatal development and the fertility risks to women and men from toxic agents have been outlined. Chemicals that are specifically toxic during human pregnancy and lactation were described, as well as international test methods in reproductive toxicology.

To assess RT it has to be taken into account that all stages of the mammalian reproductive cycle are controlled by the endocrine system. Tests in laboratory animals to assess the effects of drugs and other chemicals on reproduction and on pre- and postnatal development are, therefore, rather complex and time consuming.

For historical reasons the testing approach in RT is significantly different for drugs, for pesticides, and for industrial chemicals. All drugs have to be tested in segment 1–3 studies in animals before they can undergo clinical testing in human patients. For pesticides, RT testing is also quite extensive, since it includes 1- and 2-generation studies in rodents. In contrast, RT testing requirements for industrial chemicals are dependent on the production volume. Thus most of the high production-volume chemicals have been tested sufficiently for RT, while the vast majority of existing industrial chemicals

have either not or only marginally been tested for RT. Therefore, the new EU chemicals policy REACH will focus in particular on RT testing of approximately 30,000 existing chemicals.

Most of the drugs currently in use have proven to be safe and can be prescribed to women during pregnancy and lactation. During the past 2 decades international clinical surveillance programs of pregnant women who were exposed to drugs have shown that the majority of drugs are safe in human pregnancy and during lactation. This experience has provided a list of drugs that may safely be used by women during pregnancy and lactation. In addition, epidemiological studies have shown that, in general, pesticides and industrial chemicals have no serious side effects on human pregnancy and lactation, while effects on fertility have to date not sufficiently been studied in women and men who are either exposed at the workplace or as consumers to finished products.

Reference

Wilson, J.D. Embryotoxicity of Drugs to Man, *in* Handbook of Teratology, Vol. 1., Eds. J.D. Wilson and C.J. Frazier, Plenum Press, New York, 1977, pp. 309–355.

For Further Reading

EU - European Commission. Annex V of the EU Directive 86/906/EEC for Classification and Labelling of Hazardous Chemicals. EU Test Guideline B 31 (Teratogenicity Test - Rodent and Non-rodent), EU Test Guideline B 34 (One-Generation Reproduction Toxicity Test), EU Test Guideline B 35 (Two-Generation Reproduction Toxicity Test). Official Journal of the European Communities, updated since 1986.

Christian, M.S. Test Methods for Assessing Female Reproductive and Developmental Toxicology, *in* Principles and Methods of Toxicology, 4th Edition. Ed. A.W. Hayes, Taylor & Francis, Philadelphia, 2001, pp. 1301–1382.

Clegg, E.D., Perreault, S., Klinefelter, G.R. Assessment of Male Reproductive Toxicity, *in* Principles and Methods of Toxicology, 4th Edition. Ed. A.W. Hayes, Taylor & Francis, Philadelphia, 2001, pp. 1263–1300.

ICH - International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guidelines - Test Guideline S5A (Detection of Toxicity to Reproduction for Medicinal Products - Segment I, II and III studies), *in* Proceedings of the Second International Conference on Harmonisation, Orlando, 1993. Eds. P.F. D'Arcy and D.W.G. Harron, Greystone Books, Antrim, UK 1994, pp. 557–577.

Koren, G. (Ed.). Maternal-Fetal Toxicology, 3rd Edition, Marcel Dekker, New York and Basel, 2001.

OECD - Organisation for Economic Co-operation and Development. Guidelines for Testing of Chemicals, Guideline 415 (One-Generation Reproduction Toxicity Study), Guideline 416 (Two-Generation Reproduction Toxicity Study), Guideline 421 (Reproduction/Developmental Toxicity Screening Study), Guideline 421 (draft) (Developmental Neurotoxicity Study). OECD, Paris, updated since 1983.

Schaefer, C. Drugs during Pregnancy and Lactation. Elsevier Science B.V., Amsterdam, 2001.

Schaefer, C., Spielmann, H., Vetter, K. Arzneiverordnung in Schwangerschaft und Stillzeit. 7th Edition, Elsevier GmbH, München, 2006.

2.9 Ecotoxicology: Not just Wildlife Toxicology

Peter Calow and Valery E. Forbes

2.9.1 Introduction

Ecotoxicology was first defined explicitly by Truhaut in the 1970s as 'the branch of toxicology concerned with the study of toxic effects, caused by natural or synthetic pollutants, to the constituents of ecosystems, animal (including human), vegetable and microbial, in an integral context'. It has subsequently been refined to: the description and understanding of impacts of industrial and agricultural chemicals on ecological systems. This includes the study of natural chemicals used in industrial processes, such as metals and oils, and synthetically produced chemicals that do not occur naturally in the environment, such as PCBs, detergents, most pesticides, and most pharmaceuticals. Ecotoxicology has its roots in toxicology but the targets of interest are different: toxicology is concerned with impacts on individual human beings, and ecotoxicology on ecological systems. So ecotoxicology is not just toxicology applied to wildlife (Calow and Forbes, 2006).

In what follows we first define the protection targets for ecotoxicology and demonstrate how they differ fundamentally from the protection targets for toxicology. We then consider the consequences of this in terms of the practicalities of carrying out ecotoxicological tests and then in using the information in assessing the risks posed by chemicals for the ecological targets.

2.9.2 Protection Targets

The focus for toxicology is on the individual, whereas the focus for ecotoxicology is on collective groups of individuals of the same species (populations) and/or groups of individuals of different species (communities/ecosystems). Individuals may be removed from populations without impact on population persistence; and similarly, some species may be lost from communities without impact on community persistence or ecosystem processes.

Toxicologists are interested in the extent to which chemicals impact the health and survival of individual human beings. Ecotoxicologists by analogy often refer to protection of ecosystem health. Yet this betrays a fundamental misunderstanding of the structure and organization of ecosystems. Within the human body, cells, organs, and systems operate in concert for the well-being of the whole. By contrast there is no cooperation between the individuals and the species within ecosystems for the well-being of the whole. Natural selection favors selfishness, i.e. the ability of individuals and the genes that they contain to reproduce and spread at the expense of others. Cooperation of the kind seen between the constituents of organisms to ensure the

survival of the individual is not an appropriate concept when referring to the different parts of ecosystems. Wherever cooperation appears to occur in Nature, such as in the social insects, it can be explained in terms of selfish gene theory (Dawkins, 2006).

Objectively, the only characteristic that can be said to be important in ecological systems is persistence; i.e. of individuals in populations and species in communities and ecosystems. The persistence of populations depends on the survival and reproduction of individuals within them, but not all individuals need survive and reproduce to ensure the persistence of the whole. Similarly, communities and ecosystems depend on the species within them; but again not all. It follows that whereas toxicologists are interested in detecting impairment of any aspect of the individual organism since it is likely to contribute to health and survival, ecotoxicologists ought to be interested only in detecting those critical impairments that relate to the persistence of populations and communities through space and time. In this context it should be noted that assessing persistence ought to involve more than simply estimating numbers of individuals or species. For populations, changes in genetic composition or age/size structure may also be of concern, and long-term persistence of communities seems to depend not only on the number of species they contain but also on their relative abundances and niches.

2.9.3 Necessary Information

Estimating the adverse effects and likely exposures from chemicals on ecological systems follows the procedures in toxicology but there are some complications. The assessment of effects involves standard tests but with species and multi-species systems supposed to be relevant for the protection of ecosystems. So plants, invertebrates, and fish are more frequently used than laboratory rodents. Exposure assessments rarely consider doses but exposure concentrations, and here the challenges are not only in determining what these may be but also in how they relate to bioavailability and effects.

Assessment of Effects

Some Principles Both toxicology and ecotoxicology are concerned with making predictions about effects on their targets from limited observations in the laboratory over limited time periods. In this context it is useful to make a distinction between dose/concentration–effect relationships that are measured at high concentrations over short periods (acute) and that usually involve lethal effects from those measured at lower concentrations over long periods (chronic) and that are usually measured as sublethal responses, including impacts on development, growth, and reproductive performance.

In toxicological tests, exposure to the chemical toxicant is expressed in terms of dose which typically may be precisely applied by one of several routes (e.g., oral, dermal, inhalation). By contrast, in ecotoxicological tests exposure to chemical toxicants is expressed in terms of environmental concentration relating to an appropriate

environmental compartment (e.g., water, sediment, food). This important distinction also leads to some differences in the way that endpoints are expressed. For lethal effects, toxicologists often express the test result as an LD_{50}, i.e. the dose required to kill 50% of the test subjects. Ecotoxicologists, in contrast, express the equivalent test result as an LC_{50}, i.e. the environmental concentration required to kill 50% of the test subjects. For sublethal endpoints toxicologists might express responses in terms of no-observed-effects level (NOEL) or lowest-observed-effects level (LOEL), where level refers to the applied dose. In contrast, ecotoxicologists refer to effects as no-observed-effects concentrations (NOEC) or lowest-observed-effects concentrations (LOEC).

An important difference between toxicology and ecotoxicology is that the former is generally concerned with extrapolating effects from a few standard test species to a single target (i.e. humans), whereas the latter involves extrapolating from a few laboratory test species to the many in natural ecosystems. Toxicologists invariably focus on single-species tests; ecotoxicologists, as well as using single-species tests, may also use multi-species test systems, which are potentially important in picking up indirect effects due to species interactions (below).

Strictly, ecotoxicology is interested in adverse effects at the population level and above. So an important question is the extent to which changes that manifest themselves within and at the level of the individual organism (e.g. impairments of survival indicated by LC_{50}s and of reproduction indicated by LOECs) do indeed have impacts at higher ecological levels. In principle, the effects at lower levels might either disappear at higher levels because of damping or be exacerbated at higher levels because of cascading effects.

Damping effects would include the within-organism homeostatic responses that might be induced by the presence of chemical stressors, the ability of populations to persist (e.g., by increased recruitment) despite loss of individuals from them, and finally the ability of communities and ecosystems to maintain processes as a result of redundancy despite losses of species from them.

Cascading effects might involve situations in which small changes in key metabolic processes within individuals, key individuals within populations, or key species within communities lead to major and potentially catastrophic changes in the biological levels above. When cascading effects occur as a result of interactions between species they are referred to as indirect effects, and these are particularly challenging for ecotoxicology. This is because the ecotoxicological tests are most often on individual species and so may miss indirect effects due to, for example, a chemical destroying the food of the consumer rather than impacting the consumer directly as a toxicant. Other examples of indirect effects may involve interference with competitive interactions and removing predators. Clearly these examples illustrate that indirect effects can lead to positive as well as negative effects on species.

Ecotoxicology has barely started to take all these complications into account in assessing the effects that chemicals have on ecologically relevant endpoints but some progress has been made (Walker et al., 2006).

More Details on Test Systems

Test systems have been developed for freshwater and marine organisms, and for terrestrial systems. These include single-species tests as well as multi-species tests that are representative of all major environmental compartments.

Key requirements of test systems are that they should be not only realistic but also replicable so that the results can be defended scientifically and in legal situations. For this reason tests are subject to standardization. The OECD (Organization for Economic Cooperation and Development) plays an important role in publishing standard test protocols. In addition, for regulatory purposes tests are carried out according to Good Laboratory Practice (GLP, see Chapter 4.2). Good texts on testing systems are Walker et al. (2006) and Rand (1995).

Freshwater systems Test systems are more developed for freshwater organisms than for other biota because rivers have been a major conduit of sewage and industrial wastes. In Europe, for example, the focus has been on test systems involving unicellular algae, invertebrate zooplankton (*Daphnia*), and fish as providing not only representative organisms but also representative feeding (trophic) groups: photosynthetic algae (primary producers) fed on by herbivorous zooplankton (primary consumers) fed on by predatory fish (secondary consumers). Short-term, acute tests seek to define the concentrations that cause fifty percent reduction in: population growth rates of the algal cells; the number of individual daphnids moving in test vessels; survival in fish–each observed usually over periods of no more than 24–96 hours. The daphnid endpoint is used because observing when individuals stop moving is relatively straightforward, and immobility is a precursor of death in these animals. For the algae and the daphnids the endpoint is expressed as an effect concentration and so should be referred to as an EC_{50} rather than LC_{50} (cf. above). Long-term, chronic tests (over several weeks) are carried out on the daphnids and fish to determine effects on individual growth and reproductive performance. The algal tests do not readily fit into these kinds of classification: they are short term on a human scale but long term in the life of an algal cell and involve changes in the balance of cell division rates and cell death. They are in fact population responses, but in unicellular organisms under laboratory conditions.

All of the above tests represent open-water systems. There are some specific tests for organisms that inhabit the benthic sediments of lakes and rivers (e.g., midge larvae) but often the results from open waters are extrapolated to sediments on the presumption that sediment organisms are most exposed through the pore water (in the interstices of the sediments) that surrounds them. However, there is some evidence that the exposure route for sediment-dwelling organisms might also include the food they eat (i.e., the sediment particles). This is an area of active research.

Marine systems There are some equivalent marine tests for open-water and sediment organisms (e.g., involving mysid shrimps, sea urchin embryos, and sediment-dwelling annelids–and Table 2.26). However, often it is presumed that marine organisms will have similar sensitivities to their freshwater equivalents. And indeed there would seem to be

Table 2.26 *Marine 'tox kits'.*

Owing to their ability to form dormant cysts, rotifers and brine shrimps have been developed into 'toxicant testing kits'. They can be stored for long periods in the dormant stage without the need to maintain expensive cultures.
For brine shrimps, hatching is initiated by adding cysts to seawater a few days before tests are to be performed. Freshly hatched larvae are rinsed and then put into wells in a multi-well tray, with the wells containing a series of replicated test solutions. The tray is covered to prevent evaporation and incubated at the desired temperature. After a prescribed time (usually 24 hours) live and dead larvae are counted and an LC_{50} calculated.
Two kinds of quality control can be applied. First, if more than 10% of larvae in the control well(s) die the test is considered invalid. Second, standard toxicants can be used to check endpoint fidelity.
Rotifer cysts are used in similar ways.

no *a priori* physiological or ecological reasons for this not being the case. However, there are marine taxa that are not found in freshwaters at all–such as jellyfishes, echinoderms, and cephalopods–that could have unusual ecotoxicological features. This is used as an argument for treating these extrapolations, from observations on freshwater organisms to expectations for marine biota, with caution in marine risk assessments (see below).

Terrestrial systems For the terrestrial environment, tests have been most thoroughly developed to gauge the effects of biocides applied to crops on biota in abutting ecosystems such as field margins, sometimes on the biota intermingled with the crops (Table 2.27) and, of course, on the soil organisms. This is because agricultural applications of biocides are an important source of deliberate exposure for terrestrial systems. These tests include ones on beneficial arthropods, earthworms, springtails, plants, and vertebrates.

Multi-species tests As well as single-species tests, multi-species tests have been constructed as representative of all major environmental compartments. Depending on size and level of complexity they are referred to as micro-, meso- or macrocosm studies (Table 2.28). Very occasionally, and usually for agricultural applications, field studies are carried out. On the one hand all these multi-species studies bring the advantage of more ecological realism, but drawbacks include cost, knowing what to measure and how to interpret it and, for - cosms, difficulties in setting up and maintaining complex systems. Moreover the complexity in itself may mean that the characteristics and responses of such systems are somewhat unique and not always easily translated to ecosystems in nature.

Table 2.27 *Brief description of some laboratory ecotoxicity tests on beneficial (i.e. not pests) arthropods that live in and around crops.*

Types	Exposure	Endpoint
Mites, spiders, various insects including beetles, dipterans, hymenopterans	Natural foliage pretreated, or glass slides onto which chemical has been applied	Most often mortality, occasionally emergence from eggs or pupae and fecundity

Table 2.28 *The major classes of multi-species tests and their properties.*

Description	Replication	Endpoint
Flasks or flow-through chemostats with cultures of bacteria and/or protozoans	Many possible	Abundance, diversity
Large flask, bench systems with periphyton and metazoans–may be constructed or naturally derived communities	Some	Relative abundance, production, respiration, and nutrient uptake
Artificial streams ranging from small indoor to large outdoor systems	Occasional; system before treatment often used as control	Abundance of key species, diversity, production
Artificial ponds ranging from bowls to large outdoor systems	As above	As above
Enclosures of natural systems; e.g. tubes in lakes, field plots	As above	As above
Manipulation of whole ecosystems	Rarely	As above

Exposure Assessments

As much as for toxicology, it is the dose that determines effect. And as in toxicology this depends on how much gets where in the organism and how long it remains there in toxic form. However, there are added layers of complexity in ecotoxicology.

The exposure concentration (EC) of a chemical might either be predicted (PEC) from models that take account of its release, distribution (e.g. partition between air, water, soil, and sediment), and persistence in toxic form in the natural environment (all processes that determine the fate of the chemical in the environment) or else may be measured (MEC), for example in monitoring programmes. Either way, neither PEC nor MEC need express the effective concentration for the targets since this will depend on other factors such as the extent to which the material present in the environment can enter the organism (bioavailability)–e.g., it might be present in soils or sediments but bound so tightly to the materials there that it is not taken up by organisms.

An important characteristic of chemicals in these terms is how they partition between water and organic phases, since this will determine their likelihood of binding to organic particles in soil and sediments and/or passage into the bodies of organisms. A widely used technique for assessing this measures the partitioning of the chemical between octanol and water under standard conditions (shorthand $= K_{ow}$)

The uptake of toxicants from their surrounding medium and food is referred to generally as bioaccumulation (bioaccumulation factor = concentration in organism/concentration in surrounds *and* food). Bioconcentration, on the other hand, refers more specifically to the capacity of aquatic organisms to take up toxicants from the surrounding water (bioconcentration factor = concentration in organism after specified time under standard conditions/concentration in surrounding water).

Exposure through feeding is important because it opens the possibility of the biomagnification of toxicants. For example, there may be limited uptake of a toxicant in organisms at one trophic level, but feeders on these may eat many individuals and accumulate the toxicant in their tissues, so increasing the level of exposure to animals that feed on them. Thus mercury may occur in very low concentrations in seawater and be absorbed by algae. These are eaten by invertebrates that are unable to excrete the mercury at a fast rate so it accumulates. The same goes for the fish that feed on them. Top predators such as sharks and swordfish may become exposed to dangerously high concentrations of the mercury. The commonality of biomagnification effects in Nature is open to controversy; but biomagnification could be a potential source of problems for humans eating fish and game from Nature.

Finally, a feature that is of concern in both ecotoxicology and human health studies is the extent to which exposure to mixtures, rather than single substances, is important. Mixtures may be more complex and even less easily defined in ecosystems as compared with human exposures. In principle the effects of chemicals in mixtures can either be additive, or more or less than additive–but additivity seems to be predominant, particularly as the number of chemicals in the mixture increases (Warne and Hawker, 1995).

A straightforward description of the principles and practice of fate and exposure assessment is given in Connell et al. (1999).

2.9.4 Risk Assessment

Fundamentally, ecological risk assessment is similar to human health risk assessment in that it involves comparing likely exposure with likely effect concentrations. Both involve uncertainties especially concerned with extrapolation from observations in test systems to the real thing. However, human health assessments often involve more expert judgment than their ecological equivalents. Stepwise approaches leading to management decisions are a feature of both.

Risk assessment is concerned with predicting the extent to which exposure from a source of chemical pollution is likely to lead to adverse effects in the target (van Leeuwen and Hermens, 1995). For both toxicology and ecotoxicology this amounts to comparing likely exposures with doses/concentrations known to cause adverse effects. The key challenges for assessing risks to human health and ecological systems from these data are therefore in terms of taking account of the uncertainties in using appropriate extrapolation techniques. As already mentioned the extrapolations involved in human health risk assessments are from observations on a few laboratory test species to human beings. On the other hand, the extrapolation for ecological risk assessment is from information on a few species to all those species in ecosystems that we seek to protect.

For both there is usually the common challenge of extrapolating from effects observed at high concentrations over short time periods (the most frequent test situation) to effects of low concentrations over long time periods (often the most realistic exposure situation).

In toxicology predicted dose is compared with the NOEL and uncertainties are taken into account by considering if the difference is greater than a defined margin of safety (MOS). These MOSs can be explicitly defined but often it is left open to expert judgment to decide if the differences are big enough for safety.

For ecological risk assessments dealing with pesticides and herbicides a similar technique is used. The endpoint is compared with exposure in a toxicity exposure ratio (TER) and these have to exceed specified values for them to be considered acceptable. On the other hand for industrial chemicals PECs and/or MECs are compared with predicted no-effect concentrations (PNECs) to give a risk quotient (RQ = PEC or MEC divided by PNEC). The PNEC is calculated from whatever endpoint is available divided by uncertainty factors that are designed to take account of the extent of extrapolation that is involved. There is a lot of variability in the way this extrapolation is applied in ecological risk assessment, but one interpretation of the factors used within European Union legislation is summarized in Table 2.29.

Sometimes sufficient data are available on the effects of a chemical on a large enough number of species to express variability between species in the form of a statistical distribution (species sensitivity distribution = SSD) and from this to predict that exposure concentration which is likely to have minimum effect (usually defined as affecting <5% of species) and using this as the PNEC. Were the sample of species that make up the SSD representative of the ecosystem(s) under consideration this would be the preferred approach, but often it is not and this can obscure uncertainties compared with the simpler RQ approach (Forbes and Calow, 2002b).

Clearly the MOS approach allows expert judgment to be applied in weighing the difference between likely exposure and effects in each case, whereas predefined uncertainty factors as used for RQs limit the scope for expert judgment, at least in

Table 2.29 *Simplified version of extrapolation (assessment) factors suggested in the Technical Guidance currently used in EU legislation; but the allowances are our interpretation of the way the extrapolation factors have been compiled. Thus a factor of 10 allows for uncertainty between acute and chronic effects, a factor of 10 allows for uncertainty in interspecies chronic effects, and a factor of 10 allows for the uncertainty between laboratory single-species tests to effects in multi-species or field tests. Extra factors might be added for other uncertainties; e.g. when using observations on freshwater biota to predict the effects on marine. After Forbes and Calow, 2002a.*

	Extrapolation factors	Allowances
At least one acute L(E)C_{50} from each of three trophic levels	1000	10
One chronic NOEC	100	10
At least three chronic NOECs from each of three trophic levels	10	10
Field data or model ecosystems	Reviewed on a case-by-case basis	

principle. This appears paradoxical in the sense that there should be less uncertainty in extrapolating from effects in a few species to one than in extrapolating from few to many. On the other hand the more uncertainties of the latter case, representing the ecological situation, mean that expert judgments are themselves likely to be more variable and hence the need for a more restrictive and transparent procedure.

Like human health risk assessment, ecological risk assessment is usually performed using a tiered approach, with lower tiers designed to represent relatively realistic worst-case scenarios. If these give no cause for concern, then no more analyses are required. On the other hand if they do indicate likely risks then analyses involving more species and possibly 'cosm' or field tests would be required. The idea is to get most information with minimum time, effort, and hence costs.

2.9.5 Fast-track Approaches

With concerns about the time, cost and number of organisms used in ecotoxicological work there has been pressure for fast-track and short-cut techniques. Simplification as a general rule introduces uncertainties so these methods have to be treated with some caution.

Examples of short-cut approaches similar to those used in human health risk assessment with attendant problems are models that link the structure of chemicals with their likely uptake and effect in organisms [so-called (Q)SARs] and *in-vitro* systems that are supposed to simulate whole-organism (e.g. fish cell/tissue cultures) responses to chemicals (see Chapters 2.4, 4.1, and 4.2B).

In-organism measures - on genes, enzymes, cells, and tissues - have also been used as signals of exposure and effect in organisms sampled from circumstances where there was suspected pollution. These are broadly referred to as biomarkers and should be treated with some caution in an ecological context (Forbes et al., 2006). First, and similar to the concern in human health biomarkers (see Chapter 4.4), one effect might potentially be caused by various factors so the signal does not lead unequivocally to a conclusion about the actual cause. Second, it follows from the discussions above that even if the biomarker is signaling a specific exposure to a chemical the effect may not be relevant in terms of harm to the ecological system concerned. Early warnings are often invoked as a rationale for taking these signals seriously but that should be set against the costs of any management that follows and the likelihood of false-positive results.

Another increasingly used method of fast-tracking chemicals for restriction is based on the argument that those that are persistent (P), bioaccumulative (B), and have high toxicity (T) are likely to cause most problems for ecosystems. PBT criteria can be ascertained from simple and hence cheap methods and so can potentially identify problems for rapid action. However, it is important to recognize that all these criteria indicate potential to cause harm–i.e. are so-called hazard criteria (see Chapter 1)–and may not always be good indicators of risk. For example, the PBT criteria say nothing about the amounts of a substance likely to escape into the environment and hence about likely exposure concentrations. Hence, management on the basis of PBT criteria should be modulated with an understanding of the costs of likely restrictions based on them.

2.9.6 Summary

Ecotoxicology aims to describe and understand how the structures and processes of ecological systems are affected by exposure to chemicals as a result of human activities. The responses of individual organisms are less interesting in this context than the responses of populations, communities, and ecosystems. However, it is easier to make observations on individuals in test systems than on populations, communities, and ecosystems. Extrapolations from one level to another involve a number of uncertainties that are most often brought into consideration through use of uncertainty factors. Owing to the inherent complexity of ecological systems, methods promising short-cuts in assessments should be treated with caution.

References

P. Calow and V.E. Forbes, Ecotoxicology, in Encyclopedia of Life Sciences, John Wiley & Sons Ltd., Chichester, www.els.net, January 2006.

D. Connell, P. Lam, B. Richardson, and R. Wu, Introduction to Ecotoxicology, Blackwell Science Ltd, Oxford, 1999.

R. Dawkins, The Selfish Gene - 30th Anniversary Edition, Oxford University Press, Oxford, 2006.

V.E. Forbes and P. Calow, Extrapolation in Ecological Risk Assessment - Balancing Pragmatism and Precaution in Chemical Controls Legislation. *BioScience*, **52**, 249–257, 2002a.

V.E. Forbes and P. Calow, Species Sensitivity Distributions Revisited: a Critical Appraisal. *Hum. Ecol. Risk Assess.*, **8**, 473–492, 2002b.

V.E. Forbes, A. Palmqvist, and L. Bach, Use and Misuse of Biomarkers in Ecotoxicology, *Environ. Toxicol. Chem.*, **25**, 272–280, 2006.

C.J. van Leeuwen and J.L.M. Hermens, Risk Assessment of Chemicals: An Introduction, Kluwer Academic Publishers, Dordrecht/London, 1995.

G.M. Rand, Fundamentals of Aquatic Toxicology, 2nd edn, Taylor & Francis, London, 1995.

C.H. Walker, S.P. Hopkin, R.M. Sibly, and D.B. Peakall, Principles of Ecotoxicology, 3rd edn, Taylor & Francis, London, 2006.

M. Warne and D.W. Hawker, The Number of Components in a Mixture Determines Whether Synergistic and Antagonistic or Additive Toxicity Predominate - the Funnel Hypothesis. *Ecotoxicol. Environ. Saf.*, **31**, 23–28, 1995.

3

Organ Toxicology

3.1 The Gastrointestinal Tract

Michael Schwenk

3.1.1 Introduction

Xenobiotic chemicals gain direct entry into the body via the gastrointestinal tract, which is the first organ that comes into contact with food-born foreign compounds. The major organs of the GI tract are the stomach, the small intestine, and the large intestine.

The gastrointestinal tract (GI-tract) is the route by which food enters the body, and which provides mechanisms for the digestion and absorption of nutrients.

Food enters the mouth where enzymatic degradation of carbohydrates is initiated and the process of chewing and mixing with saliva facilitates swallowing. The food then passes down the esophagus to reach the stomach. Movement of food through the esophagus, stomach, and intestines, and eventual discharge of waste products, is enabled by the process of peristalsis. Smooth muscle throughout the GI tract propels the food, while also mixing the intestinal contents to enhance contact between digestive enzymes and other secretions with the food.

The superficial cell layer (epithelium) of the gastric mucosa secretes enzymes that initiate the digestion of proteins, that of the intestine is specialized to digest and absorb nutrients and that of the colon to keep the water balance. The epithelial cells are protected against mechanical and chemical damage by a thin mucous layer. If cells are injured, small defects may be rapidly repaired by migrating neighboring cells (hours) and

Toxicology and Risk Assessment: A Comprehensive Introduction Edited by H. Greim and R. Snyder

subsequently by proliferation (days). The normal lifetime of intestinal cells is only 2–3 days.

The direct contact of foreign compounds contained in food has several consequences for the toxicology of the GI-tract:

- The GI-tract is potentially exposed to higher concentrations of foreign compounds in food than other organs.
- The extent of absorption of foreign compounds is influenced by the varying conditions, e.g. solubilization, osmolarity, pH, transit time within the GI-tract.
- Foreign compounds can be modified by acid catalysis, or by enzymatic biotransformation in the epithelial cells or intestinal bacteria.
- In case of acute poisonings, direct access to the GI-tract permits irrigation of the stomach, and administration of charcoal to absorb toxins and of laxatives to expel them.

3.1.2 Structure and Function

Structure

The gastrointestinal tract extends as a hollow organ from the mouth to the anus and is composed of several cell types in contiguous layers. The innermost superficial tissue is formed by the mucosa, the epithelial cell layer of which is in direct contact with the intestinal contents and has the main function to further digest and to absorb nutrients. The underlying tissue (submucosa) is rich in blood vessels, lymphatic vessels, nerve fibers, endocrine cells, and cells of the immune system (Figure 3.1). The mucosa is wrapped by muscle fibers (muscularis), responsible for peristalsis, which promote movement of the GI-contents. The outermost layer of the intestine, the serosa, holds the intestine in place.

Function

After ingestion, degradation, and liquefaction of the food in the mouth, the food mash glides through the esophagus into the ***stomach***. It is mixed there with gastric juice, which contains hydrochloric acid, pepsin, and mucous slime. Toxicologically relevant chemical reactions can take place in the acidic environment of the stomach. For example, carcinogenic nitrosamines can be formed in the stomach by the interaction of secondary amines and nitrite (Figure 3.2).

Normally, the stomach is resistant to the irritant effects of HCl secreted by the gastric mucosal cells because of the secretion of a protective mucus, which coats the inner gastric lining. The secretions of the gastric mucosa are controlled by a number of cell types and mediators. Among these are prostaglandins, which help to regulate the excretion of acid and protective mucus. The gastric contents enter in small portions the ***small intestine***, where they are mixed with bile from the liver and digestive enzymes from the pancreas. Bile contains a number of excretory products, including bile pigments largely the result of the breakdown of hemoglobin, but also derived from the cytochromes and myoglobin. Bile is also rich in bile salts and bile acids, which act in the intestine as detergents and help to emulsify fats, an important step in lipid digestion. Pancreatic juice contains a number of important digestive enzymes including trypsin and chymotrypsin,

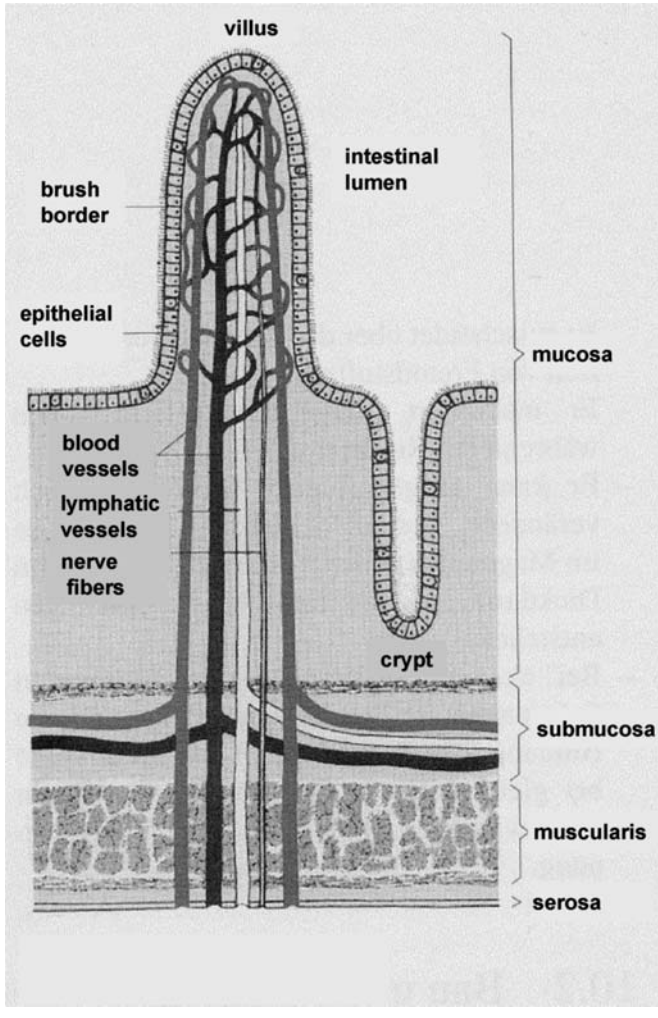


Figure 3.1 *Structure of the intestine wall.*

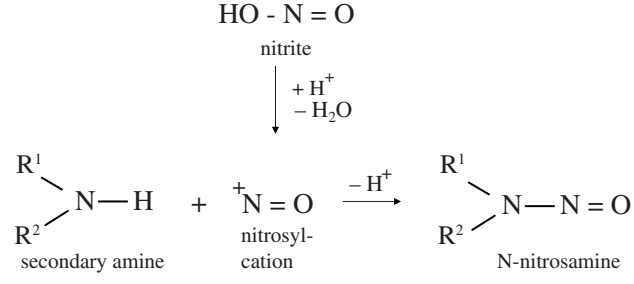


Figure 3.2 *The acidic environment of the stomach favors formation of nitrosamines from nitrite and secondary amines.*

which hydrolyse large proteins; lipase, which degrades triglycerides and phospholipids; and amylase, which initiates the digestion of carbohydrates.

The small intestine has a very large surface due to numerous convolutions of the epithelial cell layer (villi and crypts) and a brush-like surface of the epithelial cells (brush border). The surface area corresponds approximately to the size of a tennis court. The brush border membrane generates several digestive enzymes. These include peptidases, which release amino acids from proteins partially digested by trypsin and chymotrypsin. Amylase breaks down carbohydrates into disaccharides, i.e. units containing 2 sugar moieties bound by a glycosidic linkage, e.g. sucrose, lactose, maltose, etc. The brush border membrane carries disaccharidases, which hydrolyse these to simple sugars. Thus, sucrase converts sucrose to glucose plus fructose and lactase converts lactose to glucose and galactose.

A number of transport proteins are also present in the brush border membrane. They have high affinities to essential nutrient molecules, such as sugars, amino acids, fatty acids, nucleosides, vitamins, and minerals, which they transport across the membrane into the epithelial cells. The nutrients are processed through the cells and released across the basolateral membrane, where they enter the bloodstream. The transit time of the food mash through the small intestine is about 3–8 hours. Undigested and unabsorbed components are passed into ***the colon***. The daily water secretions of about 8 liters of digestive juices are largely reabsorbed in the colon. Disturbances of reabsorption will cause diarrheas with loss of water and electrolytes.

While the upper small intestine contains only few bacteria (100–1000 per ml), the colon contains up to 10^{12} bacteria per ml. The bacteria decompose some undigested carbohydrates and synthesize some nutrients (e.g., vitamin K and folic acid). In the anoxic environment of the colon, they produce various toxic substances, such as ammonia, nitrite, phenols, endotoxin, and nitrosamines. These may accumulate to clinically relevant levels if they evade detoxication in the liver; for example, in the event that the liver fails to degrade them (hepatic failure).

3.1.3 Fate of Xenobiotics in the GI Tract

Absorption

Most food-born xenobiotics can be absorbed from the GI-tract to the blood and travel to body organs to produce toxic effects. In contrast to water-soluble nutrients, which are absorbed with the help of transport proteins, transport proteins seem not to be involved in absorption of most foreign compounds, but rather in their intestinal elimination. The rate and extent of absorption depend on various processes such as: Solubilization in the lumen, transport rate across the epithelial barrier, and intestinal transit time.

Hydrophilic compounds tend to be freely dissolved in the intestinal fluid. Moderately lipophilic compounds are solubilized with the aid of bile. Highly lipophilic agents, such as liquid paraffin and compounds which are strongly bound to a matrix (e.g. soil) tend to be poorly solubilized.

Generally, xenobiotics may be rapidly transfered across the epithelial barrier, if they are dissolved or solubilized, uncharged, partly lipophilic, and of small size. The

mechanism by which they are transfered depends on their physicochemical properties. If they are highly polar, water-solubility dominates, and they may traverse across hydrophilic clefts in the membrane to the cytoplasma and from there across the basolateral membrane to the blood. If fat-solublility dominates, they diffuse into the lipophilic cell membrane and move by lateral diffusion to the blood-sided pole of the cell. Molecules that are unable to cross the cell membrane may be absorbed to a certain extent through the molecular sieve-like connections (*tight junctions*) between neighboring cells; an osmotic water drag is their driving force.

P-glycoprotein is an ATP-dependent transporter which transports a broad variety of xenobiotics out of the epithelial cells, back into the intestinal lumen. It can be induced by substrates and various substrates may compete for transport. P-glycoprotein is believed to be responsible for individual variations of bioavailabilities and for drug-drug and drug-nutrient interactions.

Small xenobiotics with moderate lipophilicity (e.g. ethanol, dimethylnitrosamine, or glycerol trinitrate) are rapidly absorbed in all segments of the GI-tract. Xenobiotics with high lipophilicity (e.g. liquid paraffin) are water-insoluble, which restricts their molecular access to the cell membrane; they are often absorbed to the extent of less than 10%. Large water-soluble compounds (e.g. strophanthin G, atropine) are rejected by the lipophilic cell membrane and are often absorbed to the extent of less than 10%. Macromolecules are believed not to be absorbed. Nevertheless, the highly toxic Botulinus toxin B, a protein with a molecular mass of 150 000 daltons, is absorbed to about 0.01%, which is sufficient to lead to severe poisoning.

Metabolism

The activity of monoamine oxidase (MAO) in intestinal cells is higher than in liver cells. The activity of glucuronosyltransferases, sulfotransferases, acetyltransferases and glutathione *S*-transferases is comparable with that in liver cells. In general cytochrome P, (CYP)-monooxygenases display only about 10% of the activity in liver tissue, but some isozymes can have activities comparable to those in the liver (e.g. CYP3A4).

In some cases, biotransformation already occurs in the epithelial cells during absorption (intestinal first-pass effect). Thus, tyramine, which is abundant in fermented food (cheese, salami, herring) in amounts that would cause a life-threatening rise of blood pressure, is efficiently (>90%) detoxicated by monoamine oxidases of the epithelial cells (Figure 3.3). Likewise, many phenolic substances are conjugated, often to the extent of 50%, in the intestinal epithelial cells with sulfate and glucuronic acid and thus detoxicated. In general less than 10% of the dose is metabolized during absorption by intestinal CYP enzymes.

The levels of drug-metabolizing enzymes differ between the various sections of the gastrointestinal tract. The activities of CYP and phase II enzymes are low in the

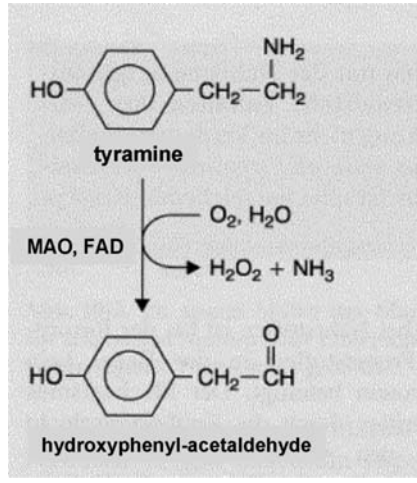


Figure 3.3 *Inactivation of tyramine by the enzyme monoamine oxidase (MAO) in intestinal epithelial cells. FAD is the coenzyme. The product hydroxyphenylacetaldehyde is biologically inactive and is further oxidized to p-hydroxyphenylacetic acid.*

esophagus and in the stomach, rise at the transition to the small intestine, and then decreases gradually towards the colon. The enzymes are predominantly located in the villi and less in the crypts. The CYP isoenzyme-pattern differs from that in the liver. Some forms of intestinal CYP enzymes and glucuronosyltransferases can be induced by food ingredients (such as ethanol, polyaromatic compounds) 2-fold or more. There are considerable variations depending on species, age and nutrition.

While biotransformations in the intestinal wall are oxidations (phase I) and conjugations (phase II) the microflora of the colon acts in an anaerobic and substrate-deficient environment and, therefore, reduces foreign compounds, (e.g. azo-compounds) and cleaves conjugates (e.g. phenol glucuronide) in order to gain energy sources. Both bacterial reaction types may have adverse consequences: in some cases, chemical reductions lead to toxic products such as nitrite and ammonia. Cleavage of conjugates liberates nonconjugated parent compounds, which are usually more lipophilic and can be reabsorbed. Intestinal bacteria can form nitrosamines and other mutagens that are detectable in the feces. The bacterial flora may thus contribute to the human cancer risk.

Enterohepatic Circulation

After absorption highly lipophilic compounds are bound to chylomicrons and released into the lymphatics, but the majority of foreign compounds are transported to the liver drained via the portal vein. During the passage through the liver, many metallo-organic compounds, organochlorine compounds, and pharmaceutical drugs are taken up by efficient transport systems into the liver cells and then secreted into bile either unchanged or following conjugation. They are then delivered with the bile into the intestine, from where they may be reabsorbed again either unchanged or after bacterial deconjugation. This enterohepatic circulation may trap xenobiotics in the organism (Figure 3.4). Thus,

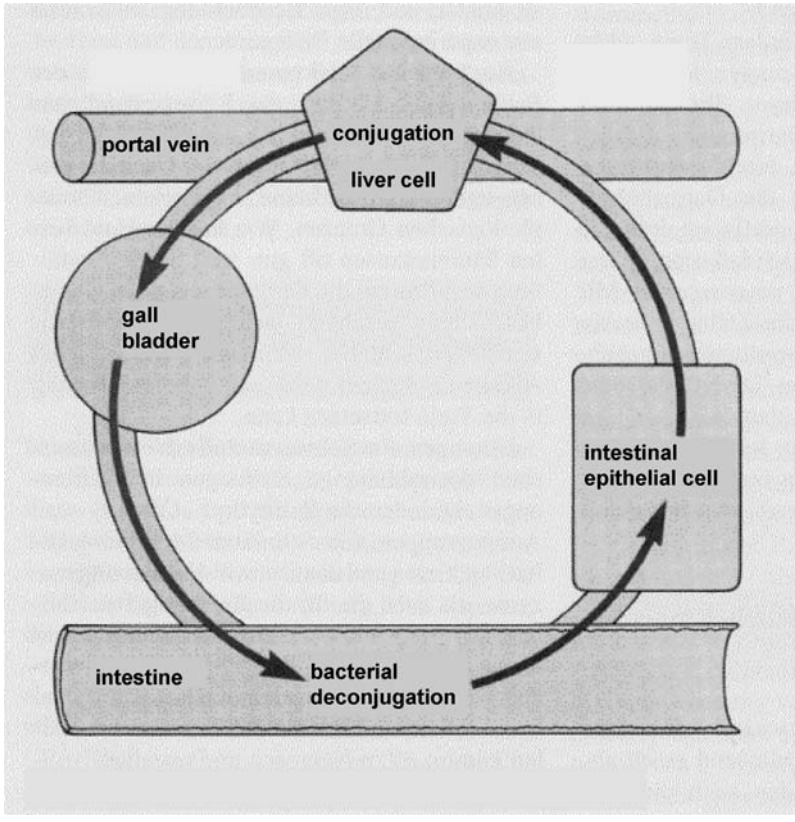


Figure 3.4 *Scheme of the enterohepatic circulation of xenobiotics.*

about 90% of an oral dose of pentachlorophenol is absorbed and then eliminated into bile. Only 3% appears in the feces due to enterohepatic circulation; the rest is reabsorbed.

3.1.4 Toxicology

Disturbance of Function

The function of the gastrointestinal tract is regulated by a large number of hormones, neurotransmitters, and tissue mediators. If foreign compounds interfere with these regulatory processes, this leads to disturbances of motility and water absorption/secretion usually without a morphological sign of damage. Such functional disturbances express themselves nonspecifically as pain in the stomach, cramps, constipation, diarrhea, or vomiting. These symptoms are often the first signs of an acute poisoning.

In poisonings with organic phosphate insecticides (e.g. parathion) the synaptic cholinesterases in the GI-tract are inhibited. This leads to a flooding of the synapses

with the neurotransmitter acetylcholine, causing increased motility and increased salivation, cramps of the stomach and intestines, as well as diarrhea. Atropine, the poison of the deadly nightshade, exhibits the opposite effect. It blocks the synaptic acetylcholine receptors and, thereby, acetylcholine effects. The first symptoms of an atropine poisoning are a dry mouth, decreased intestinal motility, and constipation. During severe poisonings with atropine or parathion, however, the effects on the central nervous system dominate.

Even the highly toxic cholera toxin acts 'solely' via functional disturbances. This bacterial toxin is a glycoprotein. It is released from cholera bacteria into the intestinal lumen, from where it is taken up by a specific transport process into the epithelial cells. There it activates a regulatory protein (Gs) which generates overshooting levels of the second messenger cyclic-AMP, which induces excessive ion- and water-secretion into the intestinal lumen. This causes extreme water loss, which may be in excess of 10 liters per day. This is an acutely life-threatening situation demanding that water and electrolytes be administered in large amounts.

Cytotoxic Effects

Compounds that disturb the protective mucous layer, or substances that induce the release of mediators that inhibit the local blood supply of the mucous membrane favor ulcer formation. Ulcers are mainly found in the stomach and upper small intestine because the acidic gastric juice rapidly attacks a damaged surface.

If epithelial cells die as a consequence of toxic influences, they detach from the tissue by a process termed erosion. If deeper tissue layers including the muscularis are also affected, an ulcer arises. This is a deep, round-edged tissue defect that often bleeds.

A common side effect associated with the therapeutic use of non-steriodal anti-inflammatory agents (NSAIDs) is gastric ulceration and bleeding. These drugs inhibit prostaglandin synthesis. Reduced gastric prostaglandin levels lead to decreased secretion of protective mucus and bicarbonate secretions, reduced local blood circulation, and overshooting secretion of the aggressive gastric acid. Ethanol and smoking also favor gastric ulcers in humans. The experimental compounds cysteamine, propionitrile, and various toluene derivatives induce duodenal ulcers in experimental animals.

Necroses of the mucous membrane of mouth and GI-tract may be caused by accidental intake of strong acids, bases, or other aggressive chemicals. The necroses don't hold on to predefined morphological borders. Acid necroses often heal well because acids precipitate tissue proteins, which protect the deeper cell layers. In contrast, lye necroses heal poorly because the tissue is liquidified by alkali and the process continues into deeper tissue layers.

Solutions of heavy metal salts cause dose-dependent disturbances in the GI-tract. They block sulfhydryl, carboxy, and amino groups of cellular macromolecules and thus inhibit digestive enzymes, transport proteins, and enzymes involved in healing. Moderate poisonings are associated with temporary disturbances, such as malabsorption, intestinal cramps, diarrhea, and vomiting. Severe poisoning leads to massive mucosal cell death with local necroses and even complete destruction of the mucous membrane.

Immunologic Reactions

The intestine possesses an immune defense system against penetrating agents and microorganisms, called MALT (mucosa associated lymphoid tissue). M-cells are specialized to present antigens from the GI-lumen to the immune system. Cells of the immune system synthesize antibodies [immunoglobulin gamma A (IgA)] that are secreted into the GI-lumen. The immune defense may be directed against food antigens (e.g. in kiwi, banana).

Following the consumption of the individually intolerable nutrient, an allergic reaction is initiated, involving the release of local mediators such as prostaglandins, leukotrienes, histamine, and cytokines. The reaction may be restricted to the GI-tract with symptoms of a full stomach and stomach pain. It also may include other organs, such as skin (rash), respiratory tract (asthma), or circulation (fall of blood pressure). Celiac disease is a severe incompatibility towards the grain protein gluten, which causes local inflammation and villous atrophy accompanied with diarrhea and malabsorption.

Cancer of the Stomach and Large Intestine

While the colon is a predominant cancer location in many western countries, cancers in other parts of the GI-tract are more common in other countries, such as the oral cavity in India, esophagus in China, or stomach in Japan. This may in part be due to genetic factors, but observations suggest that it is also associated with lifestyle.

Japanese immigrants to the USA have cancer locations as in Japan, but their descendants in the second generation have less gastric cancer (common in Japan) and more colon cancer (common in the USA). It has been assumed that carcinogenic chemicals play a role. Stomach cancer in Japan has been explained by the intake of nitrosamine-rich food, and colon cancer in the western world has been associated with low-fiber and high-fat diets, cancer of the oral cavity in India with betel-nut chewing, and cancer of the esophagus with zinc deficiency, alcohol, smoking, and consumption of unpeeled grain.

Animal experiments suggest that the GI-tract is sensitive to chemical carcinogenesis. In the rat, *N*-methyl-*N*-nitrosobenzylamine produces cancer of the esophagus, *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine cancer of the stomach, and 1,2-dimethylhydrazine cancer in the colon (Figure 3.5). The ultimate carcinogen of all three compounds is an alkyl cation, which reacts with DNA. In the case of the metastable *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine the alkyl cation forms spontaneously, in the case of *N*-methyl-*N*-nitrosobenzylamine a CYP-dependent hydroxylation is involved, and in the case of 1,2-dimethylhydrazine biotransformation proceeds through several enzymatic steps (Figure 3.6). One of the chemical intermediates of the 1,2-dimethylhydrazine metabolism is methyl-*ONN*-azoxymethanol. This compound occurs in Nature as a glycoside (cycasin) in the edible parts of the Asian cycas palm. After consumption the glycoside is hydrolysed by bacterial glycosidases of the colon. Cycasin is carcinogenic in normal rats, but harmless in germ-free rats where methyl-*ONN*-azoxymethanol cannot be released from the glycosidic bond.

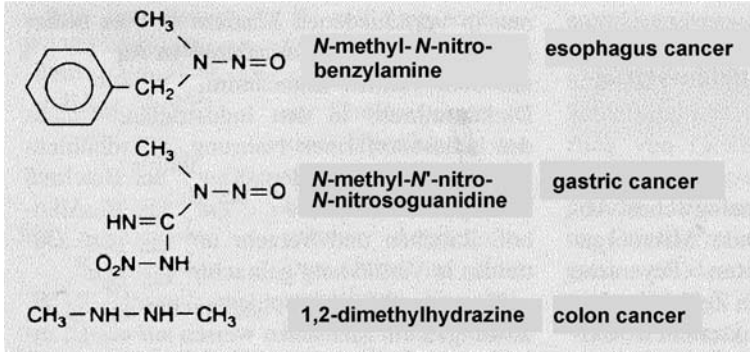


Figure 3.5 *Some carcinogenic compounds that act on the GI-tract. These compounds are degraded by different mechanisms to methyl cations, which react with the DNA in the cell nucleus.*

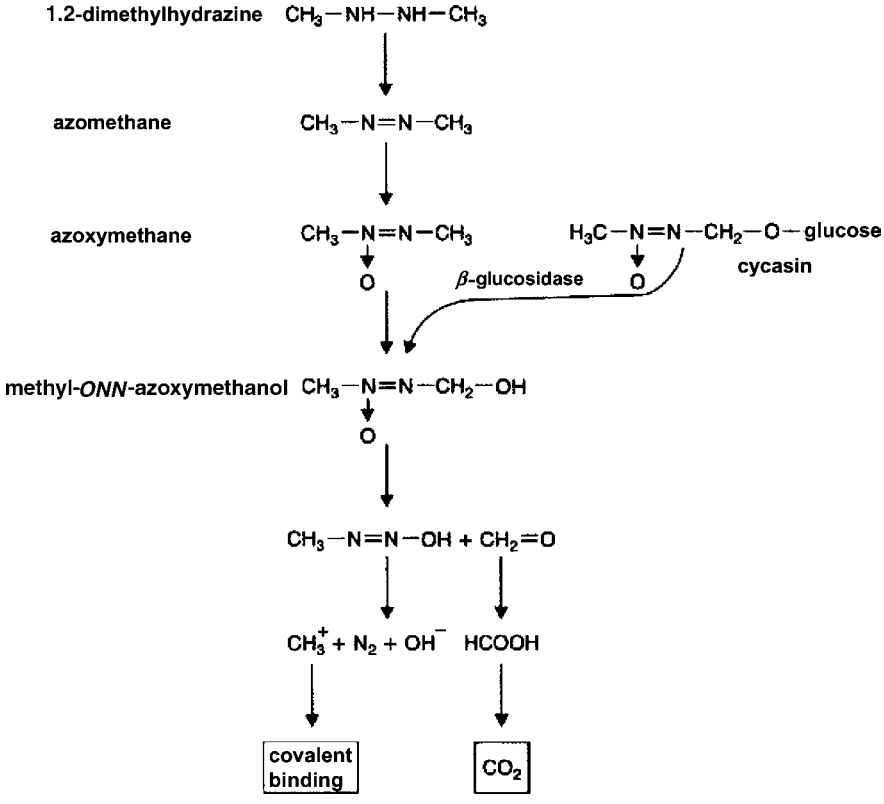


Figure 3.6 *Degradation of 1,2-dimethylhydrazine in the colon epithelium. Monoxygenases and possibly other enzymes are involved in the formation of methyl-ONN-azoxymethanol from 1,2-dimethylhydrazine. Bacterial β-glucosidases are involved in its liberation from the naturally occurring glycoside cycasin. Methyl-ONN-azoxymethanol decomposes spontaneously and forms reactive methyl cations.*

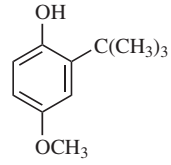


Figure 3.7 *tert-Butylated hydroxyanisole is used as antioxidant. Results of carcinogenicity studies are controversial.*

The difficulties of interpreting animal experiments with regard to their relevance for man can be exemplified with the antioxidant *tert*-butylated hydroxyanisole (Figure 3.7). Antioxidants usually inhibit the carcinogenic and toxic effects of various chemicals. However, when *tert*-butylated hydroxyanisole is added in very high concentrations of 2% to the feed of rats, it induces tumors. These are found exclusively in the forestomach, an organ that does not exist in humans. This tumor formation is assumed to be the result of local tissue irritation, which results in tumor promotion not being a primary genotoxic effect. Owing to the specific tumor location and the missing genotoxic potential of the compound, it is assumed that low concentrations of the antioxidant in food would not be associated with a cancer risk in humans.

3.1.5 Summary

The gastrointestinal tract shows organ-specific reactivities to toxic compounds. These are compiled in Table 3.1 in accordance with the examples given in the text. Considering the

Table 3.1 *Types of toxic damage in the gastrointestinal tract.*

Type of damage	Compound	Occurrence	Biological effect
Functional disturbances	Atropine	Deadly nightshade	Motility inhibition
	parathion	Insecticide	Motility increase
	cholera toxin	Cholera infection	Secretion of water
Erosions, ulcers	Alcohol, coffee	Semi-luxury items	Stomach irritation
	acids, lyes heavy metals	Cleaning agents	Necrosis of mouth esophagus, stomach
	cytotoxic agents	Chemicals	Necrosis
		Drugs, toxins, chemicals	Necrosis
Hypersensitivity reactions	Allergens (e.g. nutritional proteins	Food ingredients	Allergy (local or Systemic)
	Nonallergens (e.g. salicylates)	Food ingredients	Allergy-like symptoms
Cancer	Betel nut	Semi-luxury item	Oral cancer
	Cigarettes	Smoking	Esophagus cancer,
	Alcohol		Colon cancer
	Nitrosamines	Semi-luxury item	Esophagus cancer
	Vitamin	Food	GI cancer?
	deficiency	Food	Gastric cancer?
	Low-fiber intake	Food	Colon cancer?

potentially high concentrations of foreign compounds in the food, the gastrointestinal tract exhibits an amazing resistance to toxic damage and has the ability to detoxify and neutralize many compounds at their place of entry. The importance of chemicals in GI-carcinogenesis is not clear at present.

Reference

Comprehensive Toxicology, I.G. Sipes, C.A., McQueen, and A.J., Gandolfi, (editors), Vol. 9, Hepatic and Gastrointestinal Toxicology, Elsevier, Amsterdam, 1997.

For Further Reading

M. J. Ellenhorn and G. Barceloux, Medical Toxicology, Elsevier, New York, 1988.
Intestinal Toxicology, C.M. Schiller (editor), Raven Press, New York, 1984.
Gastrointesinal Toxicology, K. Rozman and O. Hänninen (editors), Elsevier, New York, 1986.
Comprehensiv Toxicology (eds. Sipes IG. McQueen CA, Gandolfi AK), Vol 9: Hepatic and Gastrointestinal Toxicology, Elsevier (Amsterdam), 1997.
S. C. Gad: Toxicology of the Gastrointestinal Tract. CRC-Press (2007).

3.2 The Liver

Leslie Schwarz and John B. Watkins

3.2.1 Introduction

The liver fulfils numerous vital functions associated with the endogenous metabolism of the organism such as the maintenance of constant levels of amino acids and glucose in the blood, synthesis and control of plasma proteins including several blood coagulation factors, the degradation of porphyrins, and the synthesis of creatinine, urea, and uric acid. As the largest glandular organ appended to the intestine, it produces bile, which is essential for the intestinal digestion of fat. The liver is also of paramount importance for the detoxification of foreign compounds. Hepatocytes contain a great variety of xenobiotic-metabolizing enzymes, which are often expressed in significantly higher concentrations than in other organs. To facilitate these diverse metabolic functions the liver is not only supplied with oxygen-rich blood via the hepatic artery, but also with hormone- and nutrient-rich blood from the intestine via the portal vein**.** In view of these and many other important functions it is understandable that massive damage of the liver can be life threatening.

3.2.2 Structure and Function

The morphological unit of the liver is the hexagonal lobule, the functional unit is the acinus.

Organ-specific metabolism takes place in the hepatocytes, which make up about 80% of the volume and about 60% of the cell number of the human liver. The remaining cells are mostly sinusoidal endothelial cells, cells of connective tissue, and bile ductular epithelial cells.

Close contact between the bloodstream and the liver cells is essential for intensive and quick exchange of substances between the bloodstream and hepatocytes. Hepatocytes are organized into single-layer plates in which the cells are joined by tight junctions which surround canaliculi through which bile flows to the gall bladder and eventually the small intestine. The cellular plates are separated by sinusoids through which blood flows from the hepatic artery and the portal vein to the central vein. The sinusoids (Figure 3.8 a,c) are lined with a permeable layer of endothelial cells. Between the endothelial and liver cells, there is a supporting scaffolding of connecting fibers. In addition the endothelium contains fixed macrophages called Kupffer cells, which belong to the defense system of the body. They take up and digest particles such as bacteria, senescent blood cells, and cell debris by phagocytosis. In the sinusoids the blood flows slowly under mild pressure, which makes an intensive exchange of substances possible.

The structure of the liver can be described in morphological or functional terms (Figure 3.8 a,b). The morphological unit, the **liver lobule**, is formed by the liver cell trabeculae and the interspersed sinusoids, which are oriented radially to the central vein (Figure 3.8 a). The lobule has a diameter of 1 to 2 mm and is hexagonal in cross-section. On its periphery at the points where three lobules meet is the portal space consisting of the portal vein, an hepatic arteriole, and a bile duct.

In contrast the functional unit of the liver, the so-called **hepatic acinus,** comprises the neighboring sectors of two to three liver lobules bordering on each other, which are supplied by a common vessel (Figure 3.8 b). With increasing distance from these vessels, the oxygen and nutrient content of the blood decreases. Accordingly, the oxygen and nutrient supply of the cells at the periphery of the acinus, i.e. in the vicinity of the central vein, is considerably lower than it was at the point of entry into the acinus. As a result, three zones can be distinguished in the acinus based on availability of oxygen and nutrients: zone 1 comprises the well provided cells in the center, zone 3 contains the nutrient-poor cells at the periphery, and zone 2 lies between zones 1 and 3.

Xenobiotic-metabolizing Enzymes

The liver cell possesses high concentrations and a great variety of xenobiotic-metabolizing enzymes.

Each zone of the acinus contains a different complement of enzymes. For example, zone 3, which is closest to the central vein, is rich in cytochrome P450-dependent monooxygenases (CYP), which in most cases perform the first metabolic step (Phase I) in the metabolism of xenobiotics. Zone 3 also exhibits high concentrations of GSH *S*-transferases and UDP-glucuronosyltransferases. High concentrations of glutathione and phenolsulfotransferases have been found in zone 1. Chemicals may cause liver damage to different zones of the acinus due to the heterogeneous distribution of xenobiotic-metabolizing enzymes. CYP-rich zone 3 is the most frequent primary target. Typical liver poisons that are activated by CYP-dependent monooxygenases and cause damage to zone 3 are CCl_4, bromobenzene, and acetaminophen (Figure 3.9 a). There are, however,

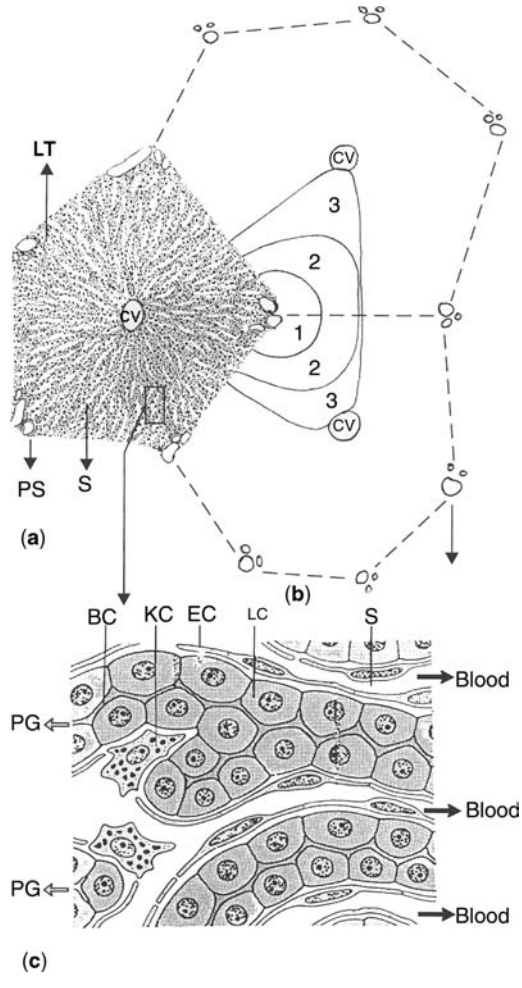


Figure 3.8 *Schematic representation of the hexagonal lobule and the hepatic acinus. (a) The hexagonal liver lobule is the morphological unit of the liver; liver trabeculae (LT) and sinusoids (S) are oriented radially to the central vein (CV). At the corners, the periportal space (PS) contains a branch of the portal vein, an hepatic arteriole, and a bile duct. (b) The functional unit is the acinus, which has been divided into three zones, depending on their distance from the supplying terminal vascular branch. (c) The magnification of this clipping shows liver trabeculae and sinusoids. In the sinusoids the blood flows to the central vein; they are lined with endothelial cells (EC) and sporadic Kupffer cells (KC). Between the liver cells run the bile canaliculi (BC) that carry primary bile (PG). [Reprinted from dtv-Atlas der Anatomie, Vol. 2, p. 229. Copyright (1984), with permission from Thieme Verlag.]*

also examples of damage in other zones, e.g. the formation of necrosis in zone 1 by allyl alcohol. Allyl alcohol is metabolized by alcohol dehydrogenase to the reactive metabolite, acrolein, immediately upon entry into zone 1, with the result that periportal cells are the principal sites of allyl alcohol-induced liver damage.

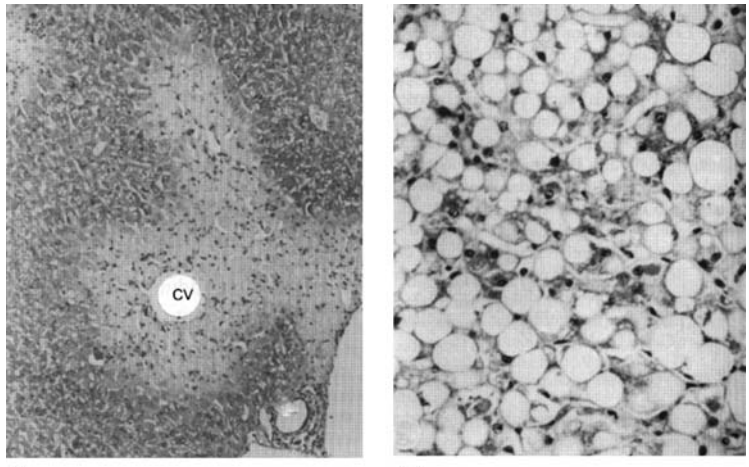


Figure 3.9 *Toxic damage. (a) Centrizonal hepatic necrosis in the liver of mice after a fatal dose of acetaminophen (× 80); CV central vein. [Reprinted from H.J. Zimmerman, p. 67. Copyright (1999), with permission from Lippincott Williams and Wilkins.] (b) Steatosis of macrovesicular type in alcoholic liver disease; the big fat drops are clearly visible (× 300). [Reprinted from H.J. Zimmerman, p. 76. Copyright (1999), with permission from Lippincott Williams and Wilkins.*

Formation of Bile and Detoxification

The formation of bile as well as the uptake and excretion of chemicals represent essential functions of the liver.

Bile is formed by the liver and actively excreted across a small specialized region of the cell membrane into fine channels called canaliculi that run between two rows of neighboring hepatocytes and flow into the bile ducts (Figure 3.8 c). The human liver produces 800–1000 ml of bile daily; about half of the bile is delivered directly via a common duct into the duodenum. The rest is concentrated via removal of water and is stored in the gallbladder. Bile acids within bile act in the intestine to emulsify fat and assist in fat digestion and absorption.

Several xenobiotics and their metabolites are excreted into the bile. In the case of polar and hydrophilic chemicals, specific transport systems at the sinusoidal and biliary (canalicular) membranes are a prerequisite for this elimination route. Some of the liver-specific carrier systems have wide substrate specificity. The selective uptake of compounds into the hepatocytes may explain their organotropic action. Thus, it is thought that the organ-specific toxicity of the mushroom poison phalloidin is due to the selective uptake of the toxin into the liver cell.

For the excretion of organic molecules into the bile a certain molecular mass and polarity is required. These conditions are often only fulfilled after the biotransformation of the chemical. For example, xenobiotic conjugation with glutathione or glucuronic acid results in a considerable increase in both the molecular mass and the polarity, due to the

introduction of the anionic moieties. In humans, organic anions will be excreted via bile only if their molecular mass is greater than about 500 Da. This threshold value is gender- and species-specific. Compounds that have been excreted via bile into the intestine may be subject to **enterohepatic circulation** (Chapters 1 and 2.1): For example, β-glucuronidases and sulfatases present in the intestinal microflora may hydrolyse glucuronides and sulfate conjugates, enabling the reabsorption of the lipophilic compounds and their subsequent reconjugation in the liver. Enterohepatic circulation can proceed several times, which results in a substantial extension in the half-life of xenobiotics in the body.

3.2.3 Toxicology

The liver is the organ of the body that is primarily responsible for the detoxification of xenobiotic chemicals. All substances absorbed from the gastrointestinal tract enter the portal circulation and are deposited in the liver. Here potentially dangerous chemicals can be transformed into harmless water-soluble metabolites before they reach the general circulation. The liver is unique in that it displays an exceptionally high capacity to detoxify xenobiotics. Lower molecular weight xenobiotic metabolites enter the circulation and are excreted in urine. The liver excretes higher molecular weight foreign compounds via the bile into the intestine from which they are excreted via deposition in the feces.

The liver can also be the target for the toxic effects of chemicals (Table 3.2). A wide range of diverse substances have been demonstrated to act as hepatotoxins including (1) halogenated aliphatic hydrocarbons such as chloroform, carbon tetrachloride, 1,2-dichloroethane, and tetrachloroethane, (2) substituted aromatic hydrocarbons such as dibenzodioxins, dibenzofurans, and bromobenzene, (3) mushroom poisons such as phalloidin and amanitin, and (4) inorganic compounds such as arsenic and white phosphorus.

Chemically induced liver injury may be manifested as a series of increasingly severe effects ranging from accumulation of fat in hepatocytes (**steatosis**) or disturbances of porphyrin metabolism (**porphyria**), through occlusion of the central and sublobular veins (**veno-occusive disease**), inhibition of bile flow (**cholestasis**) as well as inflammation of the liver (**hepatitis**) and cell death (**necrosis** or **apoptosis**). These syndromes may develop separately or together and may be the result of exposure to any of a large variety of chemicals. Massive damage of the organ can result from a single large dose of a hepatotoxic agent or after repeated small doses. The liver will display extended fibrosis, i.e. increasing formation of connective tissue, as it attempts regeneration of its architecture. Cirrhosis represents a terminal stage in liver disease in which much of the functional organ has been replaced by fibrotic scar tissue. Similar to many other organs, carcinogenic compounds may induce **tumors** in the liver.

Steatosis

Fatty liver is a frequent and unspecific side effect of liver damage induced by many chemicals.

Steatosis arises as a consequence of excessive consumption of alcohol or poisoning with chemicals such as CCl_4, dimethylnitrosamine, polyhalogenated biphenyls, or

Table 3.2 *Some chemicals inducing liver damage in man.*

Chemical	Damage	Exposure
Arsenicals	Steatosis, necrosis[a] angiosarcomas	Used as pesticide in wine-growing
Aflatoxins	Carcinoma	Consumption of moldy food stuff; of the aflatoxins, aflatoxin B_1 exhibits the strongest carcinogenic potential in animal experiments
Amanitin	Necrosis[a]	Consumption of *Amanita phalloides*
CCl_4	Steatosis, Necrosis[a]	In the 1930s used as vermicide; used as solvent
4,4′-Diaminodiphenylmethane (4,4′-DPM)	Cholestasis	1965 mass poisoning in Epping (GB) by bread contaminated with 4,4′-DPM
Dimethylnitrosamine	Steatosis, Necrosis,[a] carcinoma	Used for the production of 1,1-dimethylhydrazine (rocket propellant) in the 1940s
Hexachlorobenzene	Liver malfunction, porphyria	In the 1950s, a mass poisoning occurred in Turkey, where seeds treated with the fungicide hexachlorobenzene had been ingested
Acetaminophen	Necrosis[a]	Ingestion intending to commit suicide
Pyrrolidizine alkaloids	Occlusion of the central and lobular veins; necrosis[a]	Ingestion of *Senecio* plants and related species
Elemental phosphorus	Steatosis, necrosis[a]	Poisonings mostly in the past: ingestion of phosphorus-containing matches as well as rat and roach poison
Trinitrotoluene	Necrosis[a]	In munition factories
Tetrachloroethane	Steatosis, necrosis[a]	Employed as a solvent in aircraft construction during World War I and II
Vinyl choride	Fibrosis, angiosarcoma	Industrial production of poly(vinyl chloride)

[a]Compounds that cause liver cell necrosis may also induce fibrosis and cirrhosis upon chronic exposure. These syndromes are not listed separately in the Table.

phosphorus. It is characterized by cytoplasmic storage of triglyceride-rich fatty droplets (Figure 3.9 b), which are likely to result from increased synthesis of triglycerides, as well as a disturbance in the synthesis and transport of lipoproteins containing triglycerides, phospholipids, and glycoproteins. Triglycerides are primarily released as very-low-density

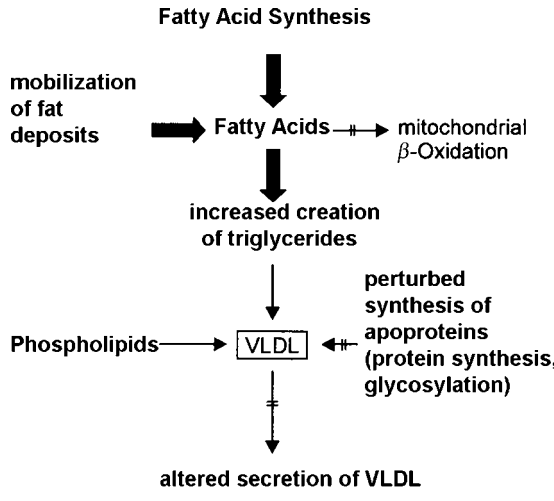


Figure 3.10 *Possible mechanisms causing steatosis. Chemically induced steatosis may develop after a disturbance in triglyceride synthesis and degradation or in the formation or release of very-low-density lipoprotein (VLDL).*

lipoprotein (VLDL) into the blood. Incomplete lipoprotein particles cannot be secreted from the cells into the blood (Figure 3.10). In most cases, it is reasonable to assume that the fatty liver induced by toxic compounds is reversible and that the accumulation of fat is not per se cell damaging.

Porphyria

Porphyrias are a genetically based series of metabolic disturbances of heme synthesis.

The liver is an important site for the synthesis and degradation of porphyrins (Figure 3.11). Porphyrins represent the backbone of heme, which plays a critical role in the actions of both mitochrondrial and microsomal cytochromes. Disorders of heme synthesis may cause markedly increased concentrations of specific porphyrinogens (precursors of heme) or porphyrins in the liver, urine, or feces. There are several forms of the disease, each of which is characterized by the degree to which inflicted individuals may display abdominal pain, photosensitivity, or neuropsychiatric symptoms.

Chemicals such as hexachlorobenzene, polychlorinated biphenyls, and dibenzodioxins may cause porphyria. They are inducers of CYP enzymes and increase not only the synthesis of the protein moiety of the monooxygenases, but also the formation of the heme prosthetic group. The porphyrogenic chemicals increase the activity of the first and rate-limiting enzyme in heme synthesis, δ-aminolevulinic acid synthase. In addition they decrease the activity of one or more enzymes that transform intermediates of heme synthesis in the further course of the metabolic pathway. Porphyria cutanea tarda, which is caused by ethanol and the xenobiotics mentioned above, is characterized by a

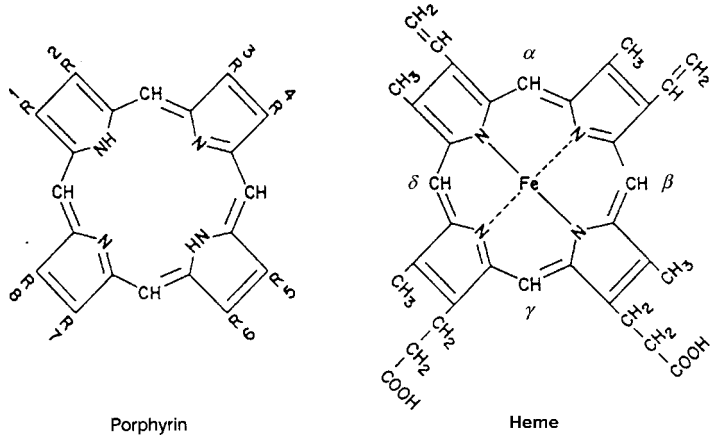


Figure 3.11 *Chemical structures of porphyrins.*

decreased activity of the enzyme uroporphyrinogen decarboxylase (Figure 3.12). Consequences of this disturbance of heme metabolism include increased concentrations of toxic porphyrinogens in liver and in blood. The hallmark symptom of porphyria cutanea tarda is photosensitivity of the skin. Although alcohol abuse is the most frequent cause of this disease, less than 1% of alcoholics develop a porphyria.

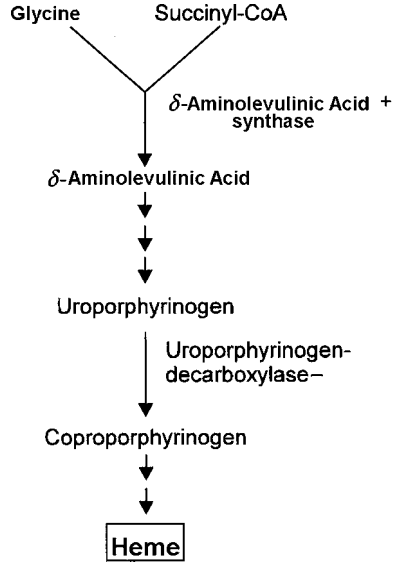


Figure 3.12 *Possible targets of chemicals in heme synthesis, causing porphyria of the type porphyria cutanea tarda; +, − increase or decrease of enzyme activity, respectively.*

Cholestasis

Disturbance of bile excretion causes excessive concentrations of bilirubin and bile acids in liver and blood.

Chemicals can cause cholestasis, i.e. an inhibition of bile flow, via very different mechanisms, such as an increased viscosity of the bile, blockage of the bile canaliculi by precipitates, an increase in the permeability of the tight junctions, and membrane effects such as disturbance of the transport of bile components. As a result of cholestasis, back-up of bile pigments into blood occurs and jaundice develops. The causative agents are almost exclusively drugs, e.g. the synthetic sexual steroid ethinylestradiol or the neuroleptic compound chloropromazine. An example of a xenobiotic causing cholestasis is 4,4′-diaminodiphenylmethane, a hardening agent used with epoxy resins (Table 3.2).

Necrosis

The increase in activity of liver-specific enzymes in blood points to necrotic processes in the liver.

Necrosis, i.e. cell death, is the most severe form of cell damage. The underlying events are only partially known. Dependent on the toxic compound and the extent of damage, liver cell necrosis may be zonal or diffuse throughout wide areas of the liver. Owing to the zonal distribution of CYP many hepatotoxic compounds cause necrosis in zone 3 of the acinus (Figures 3.8 and 3.9 a).

As a consequence of necrotic processes, soluble enzymes from damaged cells reach the blood. Measurement of distinct liver-specific enzymes in the serum offers a convenient diagnostic procedure to assess liver damage. The most frequently determined enzymes used to evaluate liver function are aspartate aminotransferase (AST), alanine aminotransferase (ALT), glutamate dehydrogenase (GLDH) (shown in Figure 3.13), ornithine-carbonyl-transferase, γ-glutamyl transpeptidase and sorbitol dehydrogenase.

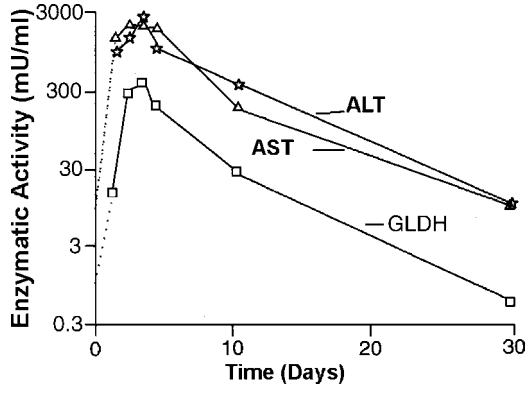


Figure 3.13 *Determination of increased activity of cellular enzymes in serum after poisoning by CCl_4. The enzyme activities are plotted logarithmically. ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; GLDH: Glutamate dehydrogenase.*

The injured liver can compensate for minor cell losses by an increased proliferation of the surviving cells. If wide areas of the tissue are damaged by poisoning, the resulting liver failure floods the body with toxic products of metabolism, such as ammonia, leading to liver coma, which is characterized by unconsciousness and convulsions. The latter represents the immediate cause of death in liver failure.

Hepatitis

Hepatitis is an inflammation of the liver.

The most common forms of hepatitis are those induced by viruses. Equally prevalent is alcohol-induced hepatitis. Drugs such as the anesthetics halothane and methoxyflurane (Figure 3.14), the antihypertensive methyldopa, the antituberculosis drugs isoniazid and rifampicin, the NSAIDS ibuprofen and indomethacin, the antifungal ketoconazole, and others have been shown to induce hepatitis very rarely. However, for most hepatotoxic chemicals such as CCl_4, phosphorus, and dialkylnitrosamines no or only marginal inflammation in the liver is observed. Exceptions include the amino sugar galactosamine (Figure 3.14), which has been extensively used in experimental liver research. In contrast to galactosamine, whose dose- and time-dependent action is highly predictable, the drugs mentioned above cause a liver inflammation only in rare cases. It is most likely that immunologic processes play an essential role in drug-induced hepatitis. For some of the drugs it has been shown that haptens are formed during their metabolism. Dependent on the causative chemical, the prognosis for hepatitis may be fatal; the mortality of a liver inflammation caused by halothane amounts to about 50%.

Cirrhosis

Liver cirrhosis is characterized by the replacement of hepatocytes with connective tissue and the destruction of normal liver architecture.

Massive cell loss due to acute poisoning or persistent repeated damage, e.g. by alcohol abuse or an infection with hepatitis B-virus, can irreparably destroy the architecture of the liver lobule. Normal liver tissue is replaced by connective tissue. Subsequently, blood vessels develop along the bridges of connective tissue between portal vein and central

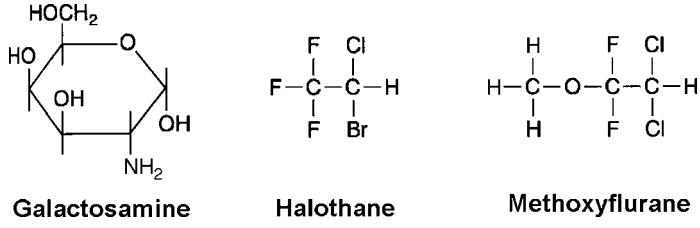


Figure 3.14 *Chemical structures of galactosamine, halothane, and methoxyflurane.*

vein and shortcircuit the bloodstream. The remaining liver cells are no longer arranged as liver lobules, but are present in disordered structures in which blood flow is severely impaired. Blood backs up into the portal system, leading to the development of portal hypertension, and into the circulation of the esophagus and stomach, resulting in the formation of varices that may rupture and produce internal hemorrhaging. Since the liver is responsible for the synthesis of a number of clotting factors, liver failure decreases the capacity of the blood to clot, which exacerbates internal bleeding. Furthermore, patients with cirrhosis have a significantly higher risk of developing a liver carcinoma. Development of cirrhosis takes years and leads to death, if it cannot be stopped. The most frequent causes of death are internal hemorrhaging and liver coma.

Carcinoma

The incidence of malignant liver tumors is very different in various regions around the world.

Hepatocellular carcinoma is the most frequent malignant liver tumor of humans. Tumors of the connective tissue and the endothelium of the sinusoids (angiosarcoma) or of the bile duct cells (cholangiocellular carcinoma) are rarer. On the basis of animal experiments a great number of xenobiotics are known to cause liver tumors. This applies to tumor-initiating as well as to tumor-promoting chemicals such as phenobarbital, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin, polychlorinated biphenyls, and distinct synthetic sexual steroids. In contrast only a few chemicals have been identified with certainty as human liver carcinogens. In this context vinyl chloride and the arsenicals are known to induce angiosarcomas, and aflatoxins cause liver carcinomas. Of the various aflatoxins that are produced by molds, aflatoxin B_1 has the strongest carcinogenic activity. Liver carcinomas are a rare event in Europe and North America. Liver cancer amounts to only about 2–3% of all cancer illness in Europe and North America. In contrast, in parts of Africa and Asia, liver carcinomas amount to 60%, representing one of the most frequent tumor types. These findings have been associated with the consumption of moldy aflatoxin-containing foodstuffs by poor people and the common occurrence of viral hepatitis in these countries.

In Europe and North America, about 12–15% of all fatal cases of cirrhosis, most which can be related to alcohol abuse, are associated with liver carcinomas. Indeed the livers of up to 80% of these patients show a cirrhotic rebuilding when the tumor is observed. These epidemiological findings suggest that the occurrence of cirrhosis and liver tumors is causally related. However, a mechanistic and molecular explanation of the assumed correlation has yet to be established.

3.2.4 Summary

The liver fulfils numerous essential functions in intermediary and xenobiotic metabolism. It shows the greatest variety and the highest concentration of xenobiotic-metabolizing enzymes. The liver produces bile, which is required for fat digestion.

Xenobiotics, which are absorbed in the gastrointestinal tract, reach the liver via the portal blood and most undergo hepatic metabolism. Xenobiotics and their metabolites are excreted into the blood or via bile into the intestine.

Hepatotoxic compounds may cause metabolic disorders of the liver (steatosis, porphyria), occlusion of the central and lobular veins, and inhibition of bile flow (cholestasis) as well as hepatitis or necrosis. Massive necrosis and repeated damage of the liver parenchyma can give rise to cirrhosis. This syndrome is characterized by the replacement of liver cells by cells of connective tissue and by destruction of the architecture of the liver tissue. Carcinogenic compounds may cause liver tumors. Patients with cirrhosis and people suffering from viral hepatitis have a higher risk for liver carcinoma.

References

dtv-Atlas der Anatomie, Deutscher Taschenbuch-Verlag, München, 1984, Vol. 2, p.229
Hepatotoxicity: *The Adverse Effects of Drugs and Other Chemicals on the Liver.* H.J. Zimmerman (ed.), Lippincott Williams & Wilkins, Baltimore, 1999.

For Further Reading

Toxicology of the Liver (G. L. Plaa and W.R. Hewitt, eds), 2nd edn., Taylor & Francis, Washington, DC, 1998.
B. K. Park, N. R. Kitteringham, J. L. Maggs, M. Pirmohamed, and D. P. Williams. The Role of Metabolic Activation in Drug-Induced Hepatotoxicity. *Annu. Rev. Pharmacol. Toxicol*. 2005, **45**, 177–202.

3.3 The Respiratory System

Jürgen Pauluhn

3.3.1 Introduction

The respiratory system of mammals is responsible for oxygenating the blood via inhalation and removing excess of carbon dioxide via exhalation. It acts by drawing air into the nasal passages, distributing the air through the tracheal and bronchiolar conducting airways, and into the alveoli where gas exchange occurs. The adult human inhales about 20 m^3 of air daily. The system contains a surface area of approximately 140 m^2 over which inhaled agents can be deposited, retained, and/or absorbed. The respiratory system is second in area only to the small intestine (approximately 250 m^2) and considerably larger than the skin (approximately 1.75 m^2), the two other organs that are in contact with the environment.

The anatomy of the respiratory tract is designed to humidify and warm the air while filtering out dust particles and germs. The lining of the nasal epithelium is covered with fine hairs that capture these foreign particles and prevent them from passing into the

lungs where they may cause local inflammation or infection. The lung efficiently absorbs many types of potentially hazardous materials, which may be present as gas, vapours, liquid (mist), or solid dust aerosols. The site of major deposition within the respiratory tract may differ from one substance to another depending on its shape and size, physicochemical properties, chemical reactivity, or blood:air partition coefficient. The most critical site in the lung is the alveolus, which is designed to facilitate the passage of gases through the alveolar membranes between the lumen of the alveolus and the associated capillaries.

Toxicants that reach the gas-exchange region may pass the air–blood barrier and are distributed to other organs. This can be very rapid as all the blood returning from the lung to the left ventricle is subsequently distributed throughout the body. Not all materials deposited and retained in this region pass this barrier to any appreciable extent. For example, chronic exposure to poorly soluble particulates (PSP) may lead to particle accumulation in the gas-exchange region by virtue of their engulfment by resident and recruited macrophages. Alveolar macrophages are a mobile cell population designed to maintain this region free of particles, bacteria, cellular debris, etc., by the process of phagocytosis. This cleansing function may decrease in efficiency with increasing accumulation of foreign materials, which constitutes a long-term threat to the health of the individual. Likewise, the unsuccessful digestion of phagocytized PSPs by alveolar macrophages may lead to the discharge of inflammatory mediators or even self-digestion of cells often associated with chronic inflammatory tissue reactions and lung tumors. Thus, although PSPs may be chemically inert, specific dispositional factors may lead to disease.

For many poorly soluble and poorly reactive materials the lung is the portal of entry for systemic absorption. For other materials, such as particles of metal oxides or silica fibers such as asbestos, and reactive materials such as chemical intermediates, the respiratory tract and especially the gas-exchange region is often the primary target for toxic effects.

The thickness of the blood–air barrier is approximately 0.5 μm. This thin tissue barrier provides an extremely efficient means of gas transfer over a large surface area but is an area vulnerable to toxic insult. The integrity of the delicate architecture of alveolar septae is maintained by a highly specialized structural network consisting of epithelial, interstitial, and endothelial components. In mammals, these include structures that scrub the inhaled air to remove particulates, and specific clearance mechanisms to remove deposited and trapped materials from the airways.

3.3.2 Structure and Function

The principal function of the respiratory tract is gas exchange through alveolar ventilation for the oxygen supply of blood and removal of carbon dioxide by exhalation. This requires highly specialized anatomical structures characterized by a thin membrane with a very large surface area, the alveoli (Figure 3.15). Efficient systems remove inhaled particles and protect the lung from the adverse effects of inhaled toxic agents. Protection is achieved by effective clearance mechanisms and biochemical pathways preventing oxidative stress and proteolytic degradation of the structures.

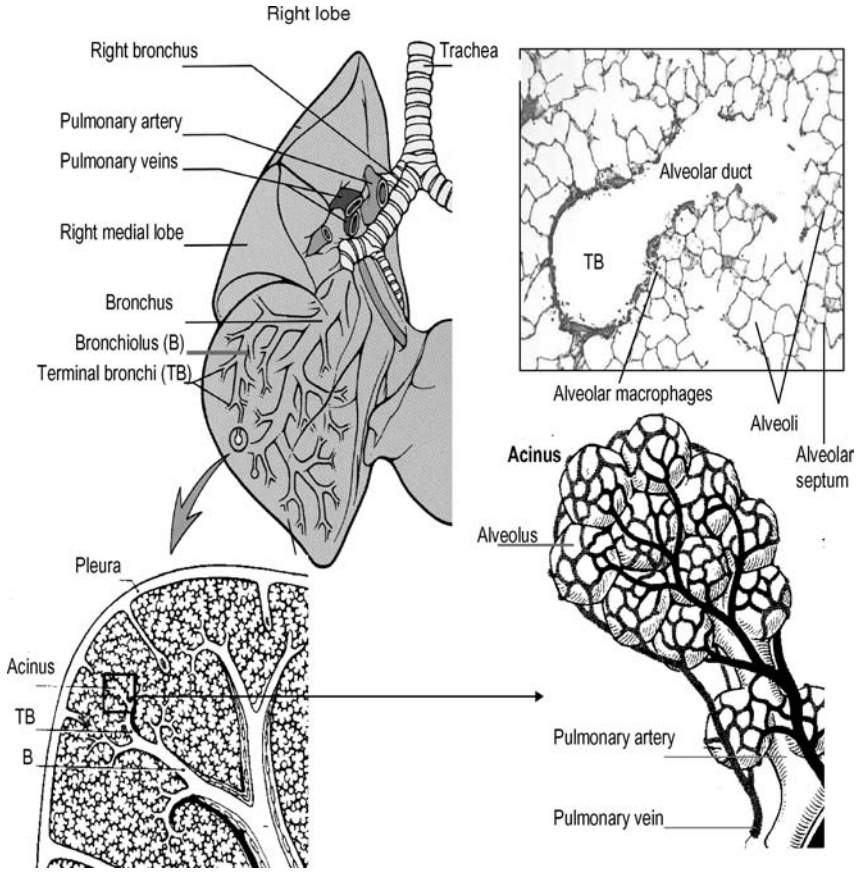


Figure 3.15 *Illustration of the lung, showing details of the lobar structure, conducting airways, and vessels. The acinus includes a terminal bronchiole and its respiratory bronchioles, alveolar ducts, and alveolar sacs (alveoli). The histology photograph details these structural components in a transsectional manner.*

The inhaled air passes from the nose to the nasopharynx. At this point, the glottis is open, with the epiglottis permitting air flow through the larynx into the trachea. However, during swallowing, food passes through the oral cavity and is prevented from entering the trachea by the action of the epiglottis which closes to cover the entrance of the glottis. The highly branched conducting airways that convey the air to the acinus (Figure 3.15) are lined by ciliated epithelium and mucus-producing cells. These features are designed to capture particles based on their aerodynamic properties and to convey volatile agents by convection.

The air–blood barrier separates the air from the large vascular capillary bed. The efficacy of gas exchange depends on the integrity of this membrane and is *diffusion* limited. The transport of gases in and out of the alveoli depends on lung mechanics and is *ventilation* controlled. The transport of gases from the blood to the alveoli is *perfusion* limited and controlled by the right ventricular cardiac output that regulates the blood flow

through the lung capillary system. The alveolar membranes separate air, at low pressure in the alveoli, from the blood-containing capillaries, in which the pressure is high. The separation is maintained by specific cell packaging, architectural organization, and physiological mechanisms that permit rapid, high-volume, gas exchange.

The *pulmonary circulation,* which is composed of the pulmonary veins containing blood emanating from the right ventricle, carries approximately 70–80 cm^3 of blood per heartbeat to the lung. The blood returns to the heart via the pulmonary arteries into the left heart and from there into the *systemic circulation*. Thus, the lung may be exposed to substantial amounts of endogenous and exogenous agents carried in blood. During one circulation the entire blood volume passes the lung, resulting in the necessity for the lung to perform a number of housekeeping functions. These include filtration of microthrombi from the circulation and maintenance of a protease-transport system to eliminate proteases released from dead bacteria, neutrophils, and lysosomes of alveolar macrophages. The most deleterious of these enzymes is elastase, which destroys structural elastin fibers, leading to a disruption of the alveolar septum. Vasoactive substances are also efficiently cleared by the vascular endothelium of the lung.

The lung must inevitably deal with inhaled toxic materials. In addition to a potent antioxidative system to protect from oxidative stress, protective functions are carried out by different mechanisms at the various locations of the respiratory tract. In the nasal cavity, the larynx, and the tracheobronchial tree, deposited material is removed by mucociliary clearance mechanisms. In the alveolar region the majority of the protective functions are carried out by different epithelial cell types: the secretory type II pneumocytes and the barrier-maintaining squamous type I pneumocytes. Water-soluble reactive vapors/gases may be inactivated by glutathione or proteins contained in mucus or fluids lining the airways.

Structure

The respiratory tract is divided into three regions: the nasopharyngeal (NP), the tracheobronchial (TB), and the pulmonary (P). Each contains region-specific anatomical features and mechanisms of deposition and clearance.

The *NP region* begins at the anterior nares and extends to the pharynx and larynx. The nasal passages of small laboratory rodents are highly tortuous and in most animal species are lined with four distinct nasal epithelial populations. These include squamous, transitional, and pseudostratified respiratory epithelium in the anterior part of the main nasal chamber, and the olfactory epithelium, which is metabolically the most active nasal epithelium, located in the dorsal region of the nasal cavity. The olfactory epithelium appears to be particularly vulnerable to metabolically induced lesions, e.g. the impact of acids released from esters via the action of esterases.

Metabolism by the olfactory epithelium may play a role in providing or preventing access to inhalants directly to the brain. For example, inhaled xylene may be converted into metabolites that move to the brain by axonal transport, while cadmium or mercury oxides may follow the same pathway without undergoing metabolism. In the NP region, poorly soluble agents are subject to physical clearance by mucociliary transport to the

throat for subsequent swallowing. Relatively soluble material is rapidly cleared into the blood. The anatomical structures, the relative abundance and distribution of specific cell types, and the air-flow rate-dependent surface dose of inhaled xenobiotics differ markedly from one species to another.

The *TB region* begins at the larynx and includes the trachea and the ciliated bronchial airways down to and including the terminal bronchioles (Figure 3.15). The TB region is both ciliated and equipped with mucus-secreting elements. The rate of mucus movement is slowest in the lower airways and increases toward the trachea.

A relatively small fraction of all sizes of particles, which pass through the NP region, are deposited in the TB region (Figure 3.16). Here deposition occurs through inertial impaction at bifurcations, sedimentation, and, for small particles, Brownian diffusion. Interception can be an important deposition mechanism for fibrous dusts. During mouth breathing of aerosols the benefits of the collection of larger particles in the nose are lost and larger particles deposit in the TB region with high efficiency. PSP are rapidly cleared by mucociliary transport to the throat where they are swallowed. Relatively soluble material deposited in this region may rapidly enter the blood circulation. The local damage to cells, especially in the lower airways, may lead to lingering inflammatory reactions such as bronchitis or asthma, with associated pathophysiological responses including bronchoconstriction or airway hyperresponsiveness

The *P region* includes the functional gas-exchange sites. The most prominent structure is the *alveolus* (Figures 3.15 and 3.17). Each alveolus opens directly into an alveolar duct or sac. Alveoli and alveolar ducts arising from a single conducting airway constitute a *pulmonary acinus*. A thin tissue barrier, consisting of type I and type II alveolar

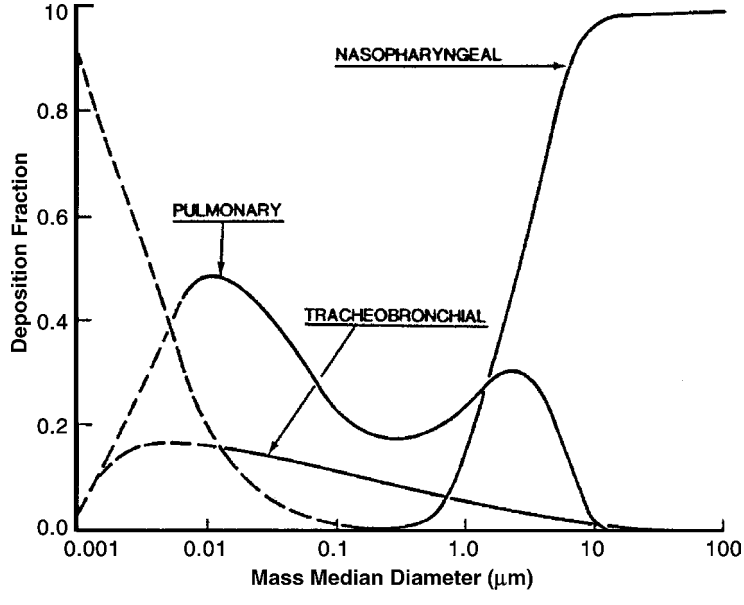


Figure 3.16 *The range of regional deposition fractions for various sizes of inhaled airborne particles in humans. This figure presents a generalization for regional deposition patterns in mammalian species.*

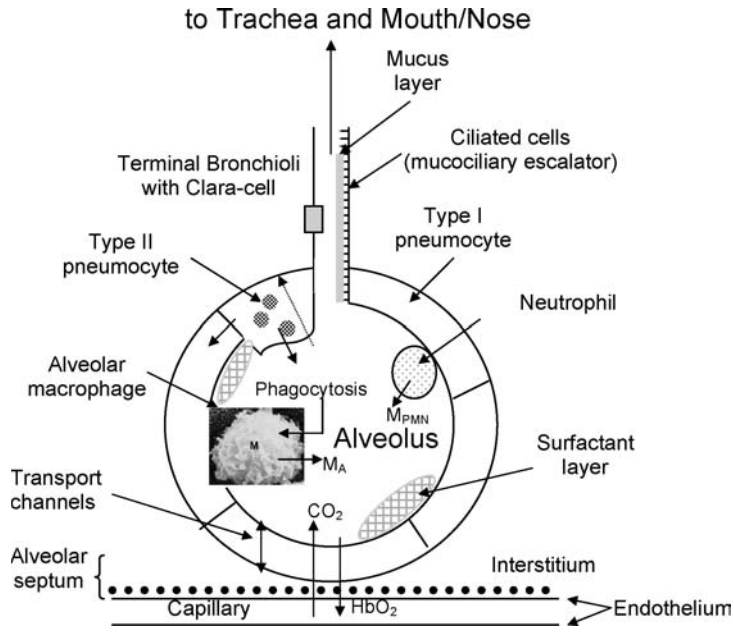


Figure 3.17 *Diagrammatic representation of the terminal air space of the lung (terminal bronchial airways, alveolus with migratory cells, interstitium, and vascular space). At the active air-exchange site the air–blood barrier constitutes a three-barrier system, which is the capillary endothelium, the basal membrane, and the type I pneumocyte. At its passive site it is stabilized by interstitial cells. Pneumocyte type I cells make the large alveolar wall cells, and pneumocyte type II cells are the stem cells of type I cells; they synthesize, store, and recycle surfactant and produce an osmotic gradient to remove excess of fluid from the alveolus (edema). Damaged surfactant is incompatible with a proper function of the alveolus due to its collapse (atelectasis) and is degraded in the lysosomes of alveolar macrophages. Clara-cells are metabolically active and produce also mucus and surfactant cofactors. M_A from alveolar macrophages and M_{PMN} from neutrophilic granulocytes (PMN = polymorphonuclear neutrophils) are mediators either orchestrating early inflammatory responses or being directly cytotoxic. HbO_2 oxygen is reversibly bound to the hemoglobin contained in red blood cells, the erythrocytes. Alveolar macrophages are responsible for keeping this space clean by phagocytosis. Alveolar macrophages have a ruffled surface and when loaded with inhaled material or debris are translocated from the alveolus via the mucociliary escalator toward the larynx. This translocation is impaired if this cell is 'overloaded', leading eventually to its 'self-digestion' and further amplification of inflammatory stimuli, as illustrated later in Figure 3.19.*

pneumocytes, which represent approximately 25% of all the cells in the alveolar septum, provides an extremely efficient means of gas transfer over a large surface area. A typical human alveolus with a surface area of 200 000 to 300 000 μm^2 contains 32 type I cells and 51 type II cells. A typical rat alveolus of 14 000 μm^2 surface area contains an average of two type I cells and three type II cells. *Type I cells* cover approximately 90% of the alveolar surface and are vulnerable targets because of their large surface area in relation to cell mass. *Type II cells* are cuboidal cells that remove extravasated fluid from the

alveoli. Type II cells are metabolically active and produce surfactant. In the event that Type I pneumocytes are damaged, Type II cells undergo mitotic division and differentiation to replace damaged Type I cells. The usually slow turnover of alveolar epithelium is speeded up considerably after injury and the number of Type II cells in the lung doubles within a short period of time.

The integrity of the delicate alveolar septa is supported by a network of mesenchymal interstitial cell populations, which produce collagen and elastin fibers. *Clara-cells* are located in the terminal bronchioles and have a high content of xenobiotic-metabolizing enzymes.

Alveolar *macrophages* represent the most variable lung-cell type. These cells help maintain a sterile environment within the lung. Inhaled foreign matter is removed through phagocytic uptake to form primary lysosomes and phagolysosomes. The phagocytized material is subsequently degraded by enzymatic digestion or removed from the lungs. Macrophages can move out of the lungs via the mucociliary escalator of the tracheobronchial tree or migrate into the lung interstitium and lymphatics. The number of alveolar macrophages may increase in the lungs by in-situ cell division of resident alveolar macrophages or by the egress of interstitial macrophages or blood-borne monocytes, which have matured into macrophages in the interstitium, into alveolar air spaces. Comparative studies indicate that alveolar macrophage number is significantly higher in 'normal' human lungs compared with other species.

The number and size of alveolar macrophages within the air spaces of the lungs are influenced by a variety of conditions that may produce large fluctuations within a short period of time. These fluctuations in cell number may lead to lung injury or reflect a diverse number of disease processes occurring in the lungs.

Protein transfer across the epithelial barrier occurs mainly by passive diffusion through *paracellular* water-filled porous channels in the tight junctions of the alveolar epithelium (Type I pneumocytes). The permeability of the tight junction is dynamic and regulated. Tight junctions play an important role in maintaining any gradients created by active *transcellular* mechanisms, e.g. endocytosis or pinocytosis. Furthermore, the permeability properties of the tight junctions vary between different epithelia that have specific physiological requirements for transepithelial solute transport.

Protective Systems

The protective functions at different locations of respiratory tract are carried out by different mechanisms. Clearance and filtration of particles require secretory products or the recruitment of phagocytically active cells to facilitate the trapping and removal of biopersistent particles. Water-soluble reactive vapours/gases may readily be inactivated by binding to glutathione or proteins contained in the mucus or lining fluids of the conducting airways.

Clearance of Inhaled Materials Clearance from the pulmonary region occurs via a number of mechanisms and pathways. The mechanisms involve either absorptive or nonabsorptive processes, which may occur simultaneously or with temporal variations. Deposited material not cleared will be retained within a specific structure of the lung

and from there may be translocated to other organs, such as the lung-associated lymph nodes.

In the NP and TB regions, poorly soluble agents are preferentially removed by mucociliary transport to the throat. The effective removal of insoluble particles from the NP region may require 1–2 days.

In the alveolar region nonadsorptive clearance processes prevail for deposited particulates that provide a chemotactic stimulus to alveolar macrophages. These cells move freely on the epithelial surfaces and phagocytize, transport, and detoxify deposited material, which they contact by random motion or via directed movement under the influence of chemotactic factors. Alveolar macrophage-mediated clearance is a slow process, which can be approximated by a monoexponential decay function with a clearance rate of ~1% per day for the rat, equivalent to an alveolar retention halftime of ~60–70 days. For humans, estimates of up to 10-fold longer alveolar retention halftimes have been given. A volumetric overloading of the macrophages with particles will result in a failure to actively move macrophages toward the mucociliary escalator.

The large alveolar surface area promotes the absorption of inhaled substances by diffusion. Surface tension at the air–water interface produces forces that tend to reduce the area of the interface, leading eventually to a collapse of alveoli. This is prevented by the presence of surface-active material, the *surfactant*. It also reduces the pressure gradient between the vascular system, which exhibits high hydrostatic pressure and the alveolus, where the pressure is subatmospheric, thus preventing edema caused by extravasation of plasma into the alveolus.

Pulmonary surfactant is highly insoluble and floats on the surface of the alveolar cells. High-capacity lipid-metabolism is required to maintain surfactant homeostasis. Surfactant turnover, which may be stimulated by interaction of surfactant with various agents, is a function of the macrophage. Chemicals that interfere with the synthesis and catabolism of phospholipids may cause alveolar phospholipoproteinosis in which catabolism of surfactant–drug complexes may be inhibited and macrophages retain toxic levels of excessive lipids.

Normal turnover of these essential lipids contributes to surfactant insufficiency and the resulting surfactant dysfunction may lead to lung edema.

Proteases, such as elastases derived from inhaled dead bacteria or leachates from phagocytic cells, damage structural proteins of the lung, and cause subsequent disruption of the alveolar septa and lung emphysema. These enzymes may also be released from activated neutrophils that undergo neurosis (abnormal pathway) rather than apoptosis and phagocytosis by alveolar macrophages (physiological pathway). Proteases released into the alveoli are either inactivated by conjugation with the plasma proteins known as α1-antitrypsin, or are cleared through the mucociliary escalator up to the larynx. Conjugated proteases are then removed by pulmonary blood and lymph and conugated with α2-macroglobulin to a form that is destroyed in the liver.

Many enzymatic processes, especially those taking place in the mitochondria and lysosomes of phagocytic cells, may generate highly reactive oxygen and nitrogen species; these include superoxide, hydroxyl radicals, and nitroperoxides. Highly reactive oxygen metabolites can react with protein, carbohydrates, lipids, and nucleic acids and can be a significant threat to the integrity of all cells. Some redox-active transitional metals, e.g. Fe^{2+}–Fe^{3+}, may produce reactive oxygen species by redox cycling via the Haber–Weiss

reaction. Such *vicious cycles* may contribute to the consumption of the reducing agents glutathione and NADPH, resulting in oxidative stress. Protection against oxygen radical attack in mammalian cells can be mediated by a variety of antioxidant enzymes, such as superoxide dismutase, and catalase, and enzymes that increase the local concentrations of reduced glutathione. The enzyme glutathione peroxidase uses the reducing power of glutathione to convert hydrogen peroxide into water. Oxidized glutathione is reduced back to its original state by glutathione reductase using NADPH as the reducing agent. NADPH is replenished via the hexose monophosphate shunt. In addition to these enzymatic systems, cells possess a variety of agents that interrupt free-radical chain reactions or scavenge reactive oxygen metabolites. These include vitamin E, ascorbic acid, and uric acid. The relative abundance of these antioxidants is different in the lower and upper respiratory tract and differs also across mammalian species.

3.3.3 The Respiratory System as Target for Toxicity

The sites of localization of inhaled gases, vapors, and particulates in the respiratory tract define the pattern of toxicity. Damage to tissues may be caused by surface dose-related factors, by the presence of localized enzyme systems leading to increases or decreases in the toxicity of inhaled materials, or by exceptional susceptibility of specific cell types at critical locations within the respiratory tract.

The distribution of inhaled substances does not occur uniformly along all airways. Specific patterns of enhanced localization within the respiratory tract imply that the initial dose delivered to specific sites may be greater than that occurring if equal distribution along the pulmonary epithelial surface is assumed. This is especially important for inhaled irritant particles and vapours that affect the tissue on direct contact. Moderate concentrations of a highly reactive or water-soluble gas do not penetrate farther than the nose and would be relatively nontoxic following nasal exposure. When the same gas is associated with particles or aerosols, it can bypass the nasopharynx, penetrate to the deep lung, and then elicit toxic responses.

After passing the NP and TB regions, vapours and particles interact with the alveolar surfaces by settling, diffusion, or interception. Direct damage or disruption of the blood–air barrier leads to inflammation and an increased extravasation of plasma into the alveoli. When damaged excessively, alveolar flooding, i.e. lung edema, occurs. Isocyanates, phosgene, or chlorine are examples of irritant aerosols or reactive vapors that cause death due to acute alveolar edema formation.

The increased permeation of protein across the air–blood barrier most likely results from a loss of the size selectivity of the epithelial barrier due to an enlargement of paracellular pores located between the cells, or the appearance of nonrestrictive transepithelial leaks. Thus, the extravasation of serum components into the alveoli usually accompanies functional perturbation or injury of the blood–air barrier.

The pathophysiological basis for PSP-induced pulmonary fibrosis has been shown for crystalline silica. The process is initiated when particles small enough to penetrate the alveolar region are ingested by alveolar macrophages. Apparently, as part of the cytotoxic

response to silica phagocytosis, the macrophage releases inflammatory cytokines and other substances that cause fibroblasts to replicate and collagen synthesis to increase. In the rat it was demonstrated that short-term inhalation exposure to respirable quartz dust did not elicit acute responses. However, within months after exposure a marked, self-perpetuating, and self-amplifying increase of inflammatory mediators was detected in the bronchoalveolar lavage fluid (BALF) (Figure 3.18). Thus, inhaled particulates that penetrate the alveolar region and are phagocytized by alveolar macrophages, but do not undergo lysosomal degradation, give rise to a *vicious cycle* of inflammatory events, leading to fibrosis (expressed as pneumoconiosis) and eventually tumor formation as a result of chronic oxidative stress and reactive cell proliferation (Figure 3.19).

In experimental animals, increased transepithelial/transmucosal permeability is conveniently detected by the measurement of the constituents of BALF. This technique is suitable and sensitive enough to provide integrative but not site-specific information on dose–response or time-course changes from initiation to resolution. It is a suitable technique to quantify the relative potency of agents that reach and damage the alveolar region. The endpoints shown in Figure 3.18 address changes in the soluble collagen, the lysosomal enzyme *β-N*-acetylglucosaminidase (NAG), the enzyme lactate dehydrogenase (LDH), and protein. NAG reflects the derangement in lysosomes due to the

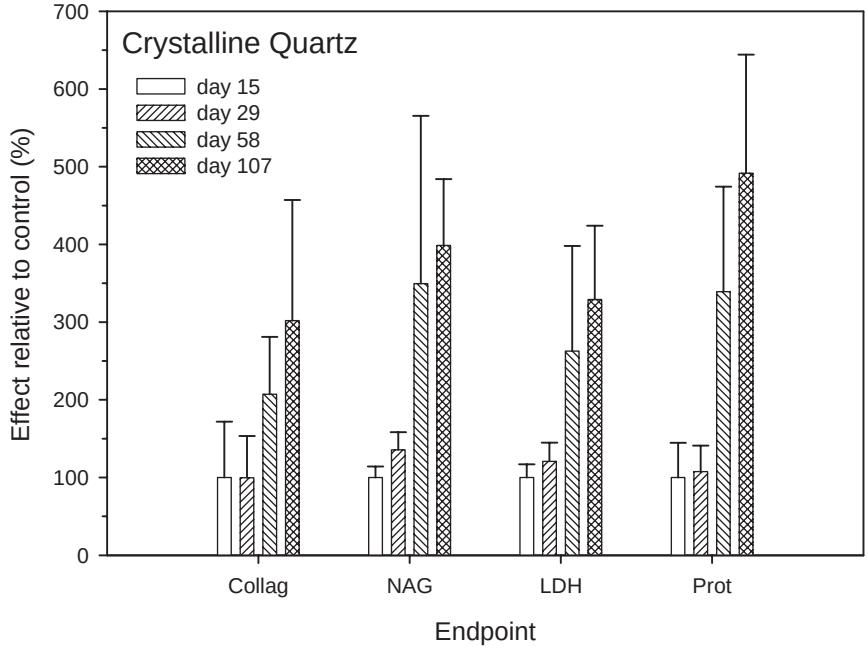


Figure 3.18 *Time-course of inflammatory responses in rats exposed by inhalation to crystalline quartz dust (up to day 14). During a 3-month postexposure period lungs were analysed by bronchoalveolar lavage for the structural protein collagen (Collag), the lysosomal enzyme N-acetylglucosaminidase (NAG), the enzyme lactate dehydrogenase (LDH), and protein (Prot). Data points represent means ± SD.*

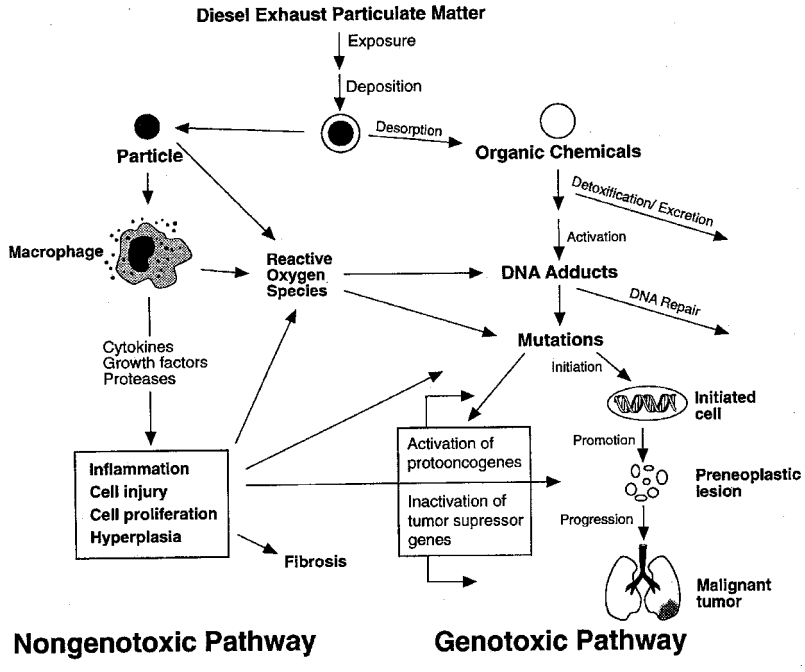


Figure 3.19 *Diagrammatic representation of possible mechanisms for PSP (Diesel exhaust)-induced carcinogenesis. The nongenotoxic pathway may be generating endogenous reactive oxygen or nitrogen species leading to genotoxic changes so that both pathways cross over at conditions causing frank pulmonary inflammation.*

unsuccessful digestion of PSPs, whereas LDH is indicative of local cytotoxicity with protein as marker of increased functional disturbance of the blood–air barrier.

Particles

In the average adult human, most particles larger than 10 μm in aerodynamic diameter are deposited in the nose or oral pharynx and are unlikely to penetrate to tissues distal to the larynx. Very fine particles (0.01 μm and smaller) are also trapped relatively efficiently in the upper airways by diffusion. Particles that penetrate beyond the upper airways are deposited in the bronchial region and the alveoli and terminal airways (Figure 3.16).

Particles deposit where the airways are small and the velocity of airflow is low, preferentially in the smaller bronchi, the bronchioles, and the alveolar spaces. As an airborne particle moves downward, buoyancy and the resistance of air act on the particle in an upward direction while gravitational force acts on the particle in a downward direction. Eventually, the gravitational force equilibrates with the sum of the buoyancy and the air resistance, and the particle continues to settle with a constant velocity known as the terminal settling velocity. Sedimentation is not significant when the aerodynamic

diameter is below 0.5 μm, when diffusion becomes increasingly important in the deposition of submicrometer particles.

The clearance of deposited particles is an important aspect of lung defense. Rapid removal lessens the time available to cause direct tissue damage. Particles are cleared by the mucociliary escalator from the airways or may be phagocytized by alveolar macrophages and are then transported to the mucociliary escalator. This mode of clearance, however, is particle-size-dependent. The fate of ultrafine or nanoparticles after their deposition may be very different from that of larger particles deposited at the same location. It appears that deposited ultrafine particles are not as readily phagocytized by alveolar macrophages as are larger particles, and that they consequently penetrate much more rapidly to interstitial sites and the endothelium. They may even enter the blood circulation, resulting in their translocation to extrapulmonary tissues. Even moderately soluble particles dissolve rapidly in the lining fluids of the lung.

Particle Overload

It was estimated that a phagocytized particle volume of about 6–8% of the normal macrophage volume signals the beginning of impaired motility, and, when reaching 60%, clearance function ceases to exist (Figure 3.20). The lung burden at which 'overloading'

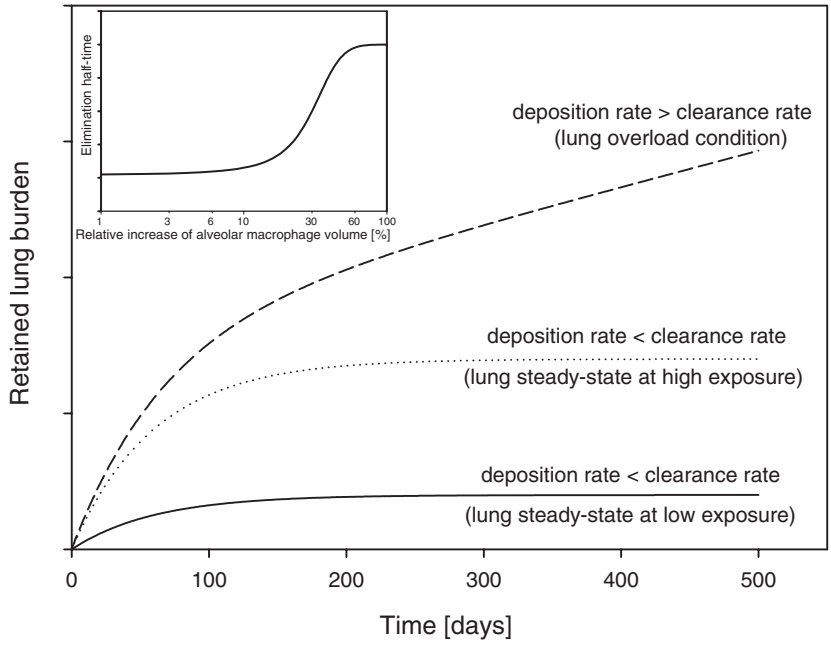


Figure 3.20 *Accumulation of poorly soluble particles in the rat lung during long-term inhalation exposure. The impairment of alveolar macrophage-mediated clearance is due to a volumetric overloading of the macrophages, resulting eventually in a failure to actively move particle-laden alveolar macrophages toward the mucociliary escalator. It was estimated that a phagocytized particle volume of about 6% of the normal macrophage volume signals the beginning of impaired motility, whilst at an attainment of 60% its clearance function ceased to exist (offset).*

starts to occur is approximately 1 mg of particles/g lung. The excessive particle overload triggers a series of pulmonary events in small laboratory animals, which is conceptualized in Figure 3.19. They include the influx and activation of inflammatory cells and tissue remodeling, commonly orchestrated by a complex interrelationship of cytokines, chemokines, and factors involved in maintaining lung homeostasis through various systems. These include the protease–antiprotease system and a balance between the formation of reactive oxidant species and the consumption of water or lipid-soluble antioxidants. Inflammatory responses are commonly accompanied by chronic remodeling of the lung structures characterized by cell proliferation, focal fibrosis, localized emphysematous responses, and ultimately benign and/or malignant lung tumors. It has been shown that the smaller the particles, the greater is the inflammatory potency. For ultrafine particles ($<0.1\,\mu m$) the surface area appears to be the unifying dose metric of exposure and response. The response is best characterized as both the influx of neutrophilic granulocytes into the alveolar space and eventually tumor formation following chronic exposure. The results of exposure can vary depending upon the toxic agent, e.g. transitional metals, endotoxins, or polyaromatic hydrocarbons, and their properties such as size, surface area, and surface-loading properties, etc.

Gases

The mechanisms affecting the transport and distribution of gases involve convection, diffusion, absorption, dissolution, and chemical reactivity. It is possible to differentiate between 3 categories of gases based on their water solubility and reactivity.

Category I gases are highly water soluble and/or reactive and thereby interact with the surface-liquid/tissue of the upper respiratory tract. The deposition of these gases is ventilation-dependent and the fraction exhaled is relatively low. Examples of Category I gases are hydrogen fluoride, chlorine, formaldehyde, and volatile organic acids and esters.

Category II gases are moderately water soluble and slowly metabolized in respiratory tract tissue. Category II gases include ozone, sulfur dioxide, xylene, and propanol.

Category III gases are relatively water insoluble and nonreactive in the extrathoracic and tracheobronchial surface-liquid and tissue. Therefore, relatively small doses reach these regions. The uptake of Category III gases is predominantly in the pulmonary region and is *perfusion* rather than *ventilation* limited. Depending on the mechanism of gas uptake and pattern of exposure (intermittent vs. steady state), the estimated default retention of nonreactive volatile agents is about 50%. Category III gases include chlorofluoroalkanes, propellants, and gases used for inhalation anesthesia.

Retention factors of gases may differ appreciably from one gas to another and may be dependent on ventilation and body activity, the duration of exposure, and gas concentration. Therefore, default assumptions are difficult to make for gases and caution is advised when doing so.

Thus, the localization of gases within the respiratory tract may vary due to the factors explained above. Among the most typical responses to reactive, irritant gases is damage to the conducting airways. The result may be protracted and sustained damage within the

airways, exemplified by inflammation and overproduction of mucus leading to loss of airway patency. In the absence of damage to the alveoli, acinary perfusion continues but gas exchange does not occur, and blood returning to the heart is not adequately oxygenated. Systemic hypoxemia results when this blood enters the general circulation. The net effect mimics the interruption of diffusional processes observed in acute alveolar edema.

Agents causing Pulmonary Toxicity

The pulmonary region can be damaged by effects other than direct chemical reactivity. Indirect effects can be mediated by cytotoxic factors liberated from alveolar macrophages upon phagocytosis (e.g., crystalline quartz) or physical factors, such as agent-induced changes in surface tension of pulmonary surfactant (e.g., detergents, emulsifiers). For instance, polyalkene glycols, which are completely nontoxic by other common routes of administration, cause toxicity upon direct entry into the lung. In some cases the molecular weight may also be an important variable of acute inhalation toxicity. It has been shown that the acute toxic lethal potency of aerosols of polyalkene glycols increases with increasing molecular weight up to a critical maximum.

Pharmaceutical agents similar to those of dicationic amphiphilic drugs (CADs) induce storage disorders caused by phospholipids and/or mucopolysaccharides in a variety of animal species and in humans. CADs or CAD–phospholipid complexes are concentrated in lysosomes and inhibit the intralysosomal breakdown of phospholipids that interfere with the anabolism and catabolism of constituents of the surfactant system in lysosomes of alveolar macrophages. Bleomycin is a systemically administered chemotherapeutic agent that combines with soluble iron (Fe^{2+}), promoting the formation of reactive oxygen species (ROS), which in turn stimulate the production of collagen and extensive fibrosis in the lung. Ingested agents, such as the bipyridylium compound paraquat, produce extreme, progressive, and eventually fatal, lung injury that is characterized by diffuse interstitial and intra-alveolar fibrosis. The extreme susceptibility of the lung to paraquat may be explained by its accumulation in the lung via the polyamine-uptake system. Once inside the cells, paraquat undergoes redox cycling with the concomitant formation of active oxygen species (Figure 3.21), depletion of NADPH, and extensive lipid peroxidation.

Upon inhalation, some xenobiotics are either specifically bioactivated in the metabolically active cells of the nasal cavities, e.g. olfactory epithelium, or the lower respiratory tract, e.g. Clara-cells in the bronchiolar epithelium, or they are sequestered locally and cause site-specific cellular damage. Many soluble transitional metal ions that are essential nutrients at homeostatic concentrations may cause marked inflammatory responses in the lung when inhaled as dust. The lung is protected to some degree from metal-induced redox cycling by proteins capable of forming chelates. For example, Hg^{2+} and Cd^{2+} are bound to metallothionein, and ionic iron may be bound to ferritin, transferrin, and lactoferrin.

Respiratory Allergy and Asthma

Allergic asthma, the etiology of which appears to be multifactorial, is a complex chronic inflammatory disease of the airways. It involves the recruitment and activation of many

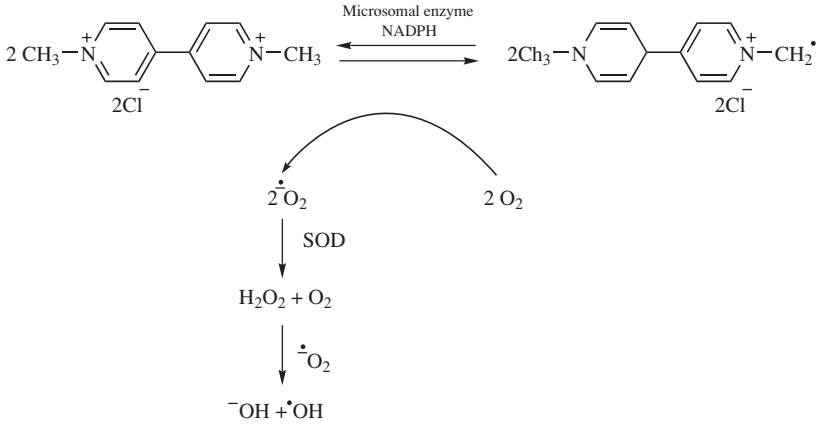


Figure 3.21 *Redox cycle of paraquat and formation of active oxygen species (SOD = superoxide dismutase).*

inflammatory and structural cells, all of which release inflammatory mediators that result in typical pathological changes. Several features of asthma can suitably be investigated in animal models (See Bice, et al., 2000): cellular infiltrations in the lung, antigen-specific IgE production, and a predominant Th2 type of immune response characterized by elevations in the levels of typical cytokines seen upon allergen (hapten) sensitization and challenge. The number of mediators involved in the sensitization process to an allergen and/or the development of a chronic inflammatory process in the mucosa of the lower airways, including airway remodeling, tends to confer an image of overwhelming complexity.

Airway hyperreactivity is defined as an exaggerated acute obstructive response of the airways to one or more nonspecific stimuli, often associated with airway epithelial damage and disruption, a common feature of even mild asthma.

In experimental models, animals are commonly sensitized by two successive treatments with the antigen (or hapten) followed by a challenge dose. The primary allergen challenge results in an asthmatic phenotype. However, to more closely resemble the human disease, secondary allergen challenges, after prolonged lapses of time, are often used. The route, method, and dose of allergen exposure also determine the phenotype of the allergen response. The response to short-term, high-level exposures causes different airway lesions from low-level, chronic antigen challenge. The methods used to assess immediate or delayed bronchoconstriction, the kind and extent of airway inflammation, and tissue remodeling have an impact on the outcome of studies involving allergen (hapten)-sensitized animal models of asthma. The typical pathology of asthma is associated with reversible narrowing of airways, with prominent features that involve structural changes in the airway walls and extracellular matrix remodeling. Features include abnormalities of bronchial smooth muscle, eosinophilic and neutrophilic inflammation of the bronchial wall, and hyperplasia and hypertrophy of mucous glands.

Asthmagenic compounds include many high-molecular-weight proteins (flour, animal dander, dust mite constituents, proteolytic enzymes used as detergents) and low-molecular-weight compounds, e.g. platinum salts, diisocyanates, organic anhydrides, and reactive dyes. Experimental models have shown that dermal contact to such agents can lead to asthma following re-exposure by inhalation.

Lung Cancer

The primary cause of lung cancer in the human is cigarette smoking. The latency interval from first exposure to invasive carcinoma is usually in excess of 20 years. A number of other agents, including many occupational agents, have been causally associated with human lung cancer. These include arsenic and arsenic compounds, asbestos fibers, beryllium compounds, bis(chloromethyl) ether, cadmium, nickel and chromium(VI) compounds, coal tar, mustard gas, metallic dusts, and welding fumes, man-made fibers, silica, soot (carbon black, Diesel exhaust), talc containing asbestiform fibers, tobacco smoke, radon and other radioactive materials, vinyl chloride, and industrial processes, such as aluminum production and coke production.

The pathogenesis of human lung cancer is a multistep process involving the sequential accumulation of alterations in two groups of genes: the tumor-suppressor genes and the proto-oncogenes. These genetic changes, in combination with increased growth-factor expression, result in the evolution of clones of cells with a selective growth advantage over adjacent normal cells. Especially in the lung, *initiation* by direct genetic damage may occur as a result of oxidative stress and a derangement of antioxidant control mechanisms (Figure 3.19), associated with a chronic inflammation-driven sustained cell proliferation and lung remodeling (promotion). In the rat, a continuum of morphologic changes has been described, ranging from hyperplasia of the alveolar epithelium through adenoma to adenocarcinoma, and this progression in changes is reminiscent of the development of the squamous-cell carcinomas observed in humans. The acute inflammatory and proliferative responses to high PSP burdens seen in the rat under overloading conditions appear to result in clearance inhibition and are causative factor for tumor formation. With the exception of Diesel emissions, which contain genotoxicants, exposure to PSPs has not been convincingly linked to human lung cancer.

3.3.4 Test Systems to Detect Toxic Effects of Inhaled Materials

Inhalation test systems are designed to detect and quantify the toxic hazards associated with inhaled materials in experimental models. The provision of a stratified test system generates data for regulatory purposes, as well as data for their safe production and use.

Inhalation studies are designed to detect and quantify the toxic hazards caused by inhaled materials under highly standardized conditions that maximize pulmonary exposure in the experimental system. Internationally recognized protocols for inhalation

toxicity studies have been devised by agencies such as the Organization for Economic Cooperation and Development (OECD) to assure that studies are reproducible from laboratory to laboratory.

However, the results from inhalation studies have to be analysed cautiously due to the large number of experimental variables involved. These are technical issues such as generation of test atmospheres or specific features of the animal model employed. Although inhalation exposure of animals mimics human exposure best, and route-to-route extrapolations are not necessary, not all results may produce data relevant to real-life exposure conditions. The choice of animal models is usually based on guideline requirements and practical considerations, such as available exposure technology, expediency, and previous experience, rather than by consideration of the relevance for human beings. Moreover, differences in routes of exposure, associated with intermittent exposure regimens, can result for any given agent in dissimilar toxicokinetics and substance-specific toxicodynamics. Being aware of these technical and dispositional variables, inhalation studies provide a tool to evaluate adverse effects by using regimens and routes similar to those prevailing for humans.

3.3.5 Summary

The principal function of the respiratory system is gas exchange. This requires the presence of a membrane with a very large surface area, but with minimal thickness. The surface area of this membrane is several times that of the integumentary surface of an individual. The barrier that separates the air surface of this membrane from the large vascular bed associated with it is limited in thickness by the requirement of gas diffusion. The existence of such a large surface area creates a tremendous number of problems in terms of packaging and architectural organization to optimize such a large gas flow and volume change. Another entire set of problems concerns the mechanisms that need to be developed to protect such a membrane from injury. Evaluation of the pulmonary toxicity of chemicals in animal models is possible because the respiratory tract provides the initial interface between the organism and the external air. Nevertheless, there is a significant amount of interspecies variability in both pulmonary architectural organization and cellular composition.

Toxic agents may reach the lung both via the blood circulation or via the inhaled air. The latter is associated with a nonhomogeneous localization of inhaled agent largely due to physical and physicochemical factors characteristic of the agent. The lung is designed to prevent dust from entering the alveoli. Inhaled biological substances, such as bacteria, are destroyed by (bio)chemical processes, which require tight control. Any mismatch of these intricate control systems by inhaled toxicants, including the impairment of the clearance of inhaled particles from the lung, may initiate chronic disease.

References

Bice, D.E., Seagrave, J.C., and Green, F.H.Y. (2000). Animal Models of Asthma: Potential Usefulness for Studying Health Effects of Inhaled Particles. *Inhal. Toxical.*, **12**, 829–862.

For Further Reading

Gardner, D.E., Crapo, J.D., and McClellan, R.O. (1993). *Toxicology of the Lung*, 2nd Edition, Raven Press, New York, 1993.

McClellan, R.O. and Henderson, R.F. (1989). *Concepts in Inhalation Toxicology*, Hemisphere Publishing Corporation, New York, pp. 193–227.

Phalen, R.F. (1984). Inhalation Studies: Foundations and Techniques, Animal Models, CRC Press, Inc., Boca Raton, Florida, pp. 211–241.

Schlesinger, R.B. (1985). Comparative Deposition of Inhaled Aerosols in Experimental Animals and Humans: a Review. *J. Toxicol. Environ. Health*, **15**, 197–214.

Snipes, M.B. (1989). Species Comparison for Pulmonary Retention of Inhaled Particles. In: *Concepts in Inhalation Toxicology*, Hemisphere Publishing Company, New York, pp. 193–227.

Witschi, H. and Nettesheim, P. (1982). *Mechanisms in Respiratory Toxicology*, Vol. I and II, CRC Press, Inc., Boca Raton, Florida.

3.4 The Nervous System

Gunter P. Eckert and Walter E. Müller

3.4.1 Structure and Function of the Nervous System

The nervous system (Figure 3.22), which consists of the peripheral nervous system (PNS) and the central nervous system (CNS), is an extremely complex network that provides the structural basis for electrochemical processes that are responsible for its multiple functions. These are cognition, sensation, memory, speech, locomotion, and unconscious hormonal and autonomic processes. The nervous system has an absolute requirement for mechanisms that enable it to transmit information within individual nerves and between nerve cells.

Transmission of Neuronal Information

The principal function of a nerve is to transmit information. In general the structure of nerves, which will be discussed in detail below, consists of the nerve cell body from which extend relatively short processes, called dendrites, which act to receive information, and a longer process, called the axon, which extends away from the cell body and acts as a mechanism for sending information to other nerves or to muscles or glands. Information is transmitted from the region of the cell body down the axon via an electrical signal called the 'action potential.' Intercellular transmission of information is accomplished by the passage of specific chemical mediators between cells. The locus of the chemical transfer of information is called the **synapse**.

The synapse called the myoneural junction lies between the foot of the axon of a motor neuron and the motor end-plates, which are specific receptors of chemical mediators on the muscle fiber. Upon release of the chemical mediator from the presynaptic neuron the neurotransmitter binds to these receptors on the postsynaptic membrane and induces an action potential at the end-plate which results in a muscle contraction. Similar synapses control interactions between nerves. For example, in the sympathetic nervous system, which is a subdivision of the autonomic nervous system, axons of nerves emerging from

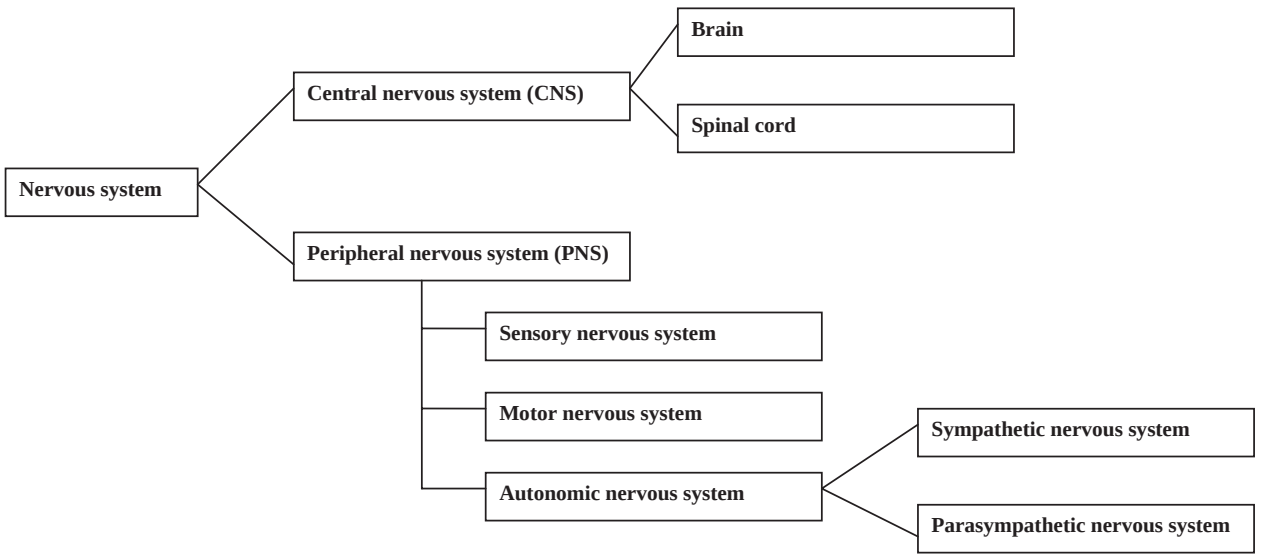


Figure 3.22 *Schematic partition of the nervous system. The nervous system is divided into the central nervous system (CNS) and the peripheral nervous system (PNS). The CNS consists of the brain and spinal cord. The PNS consists of all other nerves and neurons that do not lie within the CNS. The peripheral nervous system is divided into the somatic nervous system and the autonomic nervous system.*

the spinal column enter an organelle called a ganglion where they come in contact with the cell bodies of nerves that extend to various organs of the body. Thus, we have preganglionic presynaptic fibers entering the synapse and transfering information, via chemical mediators, to postsynaptic postganglionic fibers, which then direct the activity of various muscles or glands.

The Peripheral Nervous System (PNS)

The PNS consists of both sensory neurons, which transmit information from peripheral receptors to the CNS, and motor neurons, which transmit information from the CNS to the muscles and glands, otherwise termed effector organs. It is subdivided into the somatic sensory and motor systems and the autonomic nervous system (Figure 3.22). All spinal nerves contain both sensory and motor neurons. Sensory neurons transmit information from sensory organs, such as eyes, skin, ears, or tongue to the CNS; motor neurons transmit information from the CNS to skeletal muscles.

The autonomic nervous system consists of sensory neurons and motor neurons that run between the central nervous system, particularly the hypothalamus and medulla oblongata, and various organs. It is responsible for monitoring and regulating the internal environment (Figure 3.23). In contrast to the sensory-somatic system, the

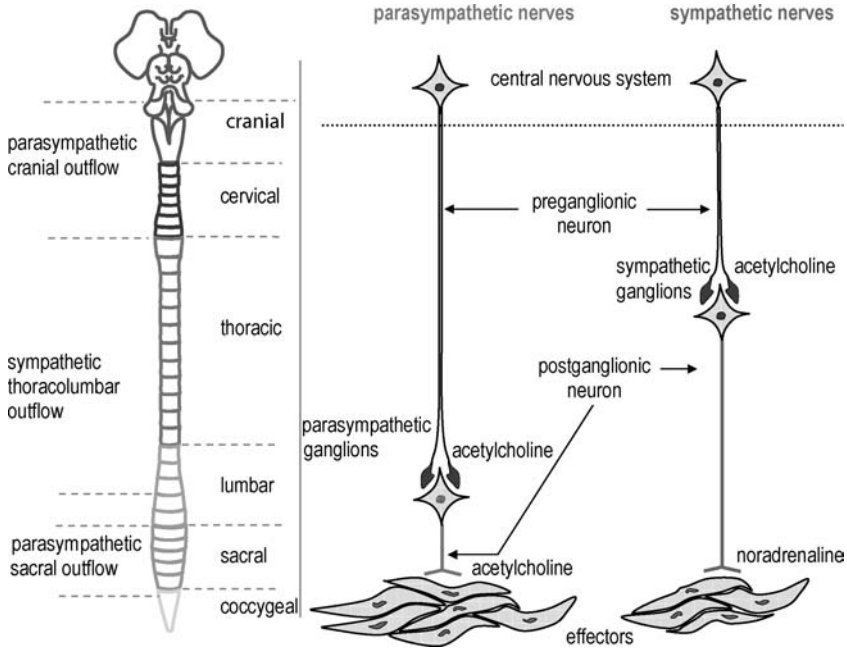


Figure 3.23 *Schematic diagram of the autonomic nerve system. The autonomic nervous system is commonly divided into two generally antagonistic subsystems: the sympathetic and parasympathetic nervous system. It controls such vital functions as heart rate, dilation of the bronchioles, and dilation and constriction of the pupil; the digestive tract is controlled almost completely by autonomic mechanisms.*

actions of the **autonomic nervous system** are largely involuntary and use two groups of motor neurons to stimulate the effectors. Efferent axons leaving the CNS, except those innervating skeletal muscle, belong to the autonomic nervous system. The axons of lower motor neurons run without interruption to the neuromuscular junctions in skeletal muscles. However, the autonomic axons leaving the CNS make synaptic connections with peripheral neurons, in ganglia, which in turn innervate the effector-cells. The axons that form synapses in ganglion cells are called preganglionic autonomic fibers, while axons that innervate the effector cells are called postganglionic autonomic fibers.

The autonomic nervous system is subdivided into the **sympathetic nervous system** and the **parasympathetic nervous system**, based on the location of the cell bodies of the preganglionic autonomic axons in the CNS (Figure 3.23). The cell bodies that give rise to the preganglionic axons of the sympathetic division of the autonomic nervous system are located in the lateral horns of the thoracic and the upper two or three lumbar segments of the spinal cord. Synaptic transmission in the sympathetic ganglia and the effector organs is mediated by acetylcholine and norepinephrine (noradrenaline), respectively. Activation of the sympathetic nervous system generally primes the body to respond to stresses that might require rapid physical and/or mental activity and often demand increased aerobic metabolism, heart rate, and muscular activity.

The preganglionic axons of the parasympathetic division emerge from the cranial and sacral region of the central nervous system. The synaptic transmissions in the para-sympathetic ganglion and at the parasympathetic-innervated organs are also mediated by acetylcholine (Figure 3.23). Functionally, parasympathetic innervations are generally opposed to those of the sympathetic nervous system. For example, activation of parasympathetic nerves in the eyes causes miosis, whereas sympathetic stimulation causes mydriasis; parasympathetic nerves increase and sympathetic nerves decrease gastrointestinal activity; bronchial smooth muscles are contracted by the parasympathetic nerves, whereas they are relaxed by sympathetic nerves, etc. Both systems are controlled by the central nervous system.

The Central Nervous System (CNS)

The **brain** and the **spinal cord** form the basic elements of the CNS (Figure 3.22). The **brain** receives sensory input mainly via spinal cord and cranial nerves and uses most of its volume and computational power to process the various sensory inputs and to initiate appropriate and coordinated motor outputs as well as cognitive functions. The brain is an assembly of interrelated neural systems that regulate their own functions, as well as the functions of other body organs, in a dynamic, complex fashion, largely through inter-cellular chemical neurotransmission. Anatomically, the brain is divided into the forebrain, midbrain, and hindbrain (Table 3.3). These are composed of several specialized structures, each responsible for defined functions of the brain. The cerebral cortex, the limbic system, and the **diencephalon** all belong to the forebrain. The **cerebral cortex** consists of two hemispheres that represent the largest division of the brain. The specialized functions of a cortical region arise from the interplay between connections from other regions of the cortex and noncortical areas of the brain and a basic intracortical processing module of vertically connected cortical columns. Cortical areas called association areas process

Table 3.3 *Divisions and macro-functions of the brain.*

Parts of the Brain	Location	Function
Brain stem	The brain stem is located at the juncture of the cerebrum and the spinal column. It consists of the midbrain, medulla oblongata, and the pons	Alertness, arousal, breathing, blood pressure; contains most of the cranial nerves, digestion, heart rate, other autonomic functions; relays information between the peripheral nerves and spinal cord to the upper parts of the brain
Hindbrain (Rhombencephalon)	The rhombencephalon is the inferior portion of the brain stem. It comprises the metencephalon, the myelencephalon, and the reticular formation	Attention and sleep, autonomic functions, complex muscle movement, conduction pathway for nerve tracts, reflex movement, simple learning
Midbrain (Mesencephalon)	The mesencephalon is the most rostral portion of the brain stem. It is located between the forebrain and brain stem	Controls responses to sight, eye movement, pupil dilation, body movement, hearing
Forebrain (Prosencephalon)	The prosencephalon is the most anterior portion of the brain. It consists of the telencephalon, striatum, diencephalon, lateral ventricle, and third ventricle	Chewing; directs sense impulses throughout the body, equilibrium, eye movement, vision, facial sensation, hearing, phonation, intelligence, memory, personality, respiration, salivation, swallowing, smell, taste

information from primary cortical sensory regions to produce higher cortical functions such as abstract thought, memory, and consciousness. The cerebral cortices also provide supervisory integration of the autonomic nervous system and integrate somatic and vegetative functions.

The **limbic system** is a term for an assembly of brain regions clustered around the subcortical borders of the underlying brain core to which a variety of complex emotional and motivational functions have been attributed. The main structures of the diencephalon are the thalamus and the hypothalamus. The **thalamus** lies in the centre of the brain, beneath the cortex and basal ganglia and above the hypothalamus. The neurons of the hypothalamus are arranged into distinct clusters or nuclei. These nuclei act as relays between the basal ganglia and the associated regions of the cerebral cortex. Moreover, the thalamic nuclei and the basal ganglia exert regulatory control over visceral functions. The **hypothalamus** is the principal integrating region for the entire autonomic nervous system and regulates body temperature, water balance, intermediary metabolism, blood pressure, sexual and circadian cycles, secretion of the adenohypophysis, sleep, and emotion.

The **midbrain and brain stem** are considered to be bridges to the spinal cord and are primarily composed of the mesencephalon, pons, and medulla oblongata. These bridging areas of the CNS contain most of the nuclei of the cranial nerves, as well as the major inflow and outflow tracts from the cerebral cortices and spinal cord. They contain the reticular activating system, an important region of the gray matter linking peripheral sensory and motor events, with higher levels of nervous system integration. The reticular activating system is essential for the regulation of sleep, wakefulness, and level of arousal, as well as for coordination of eye movement. These structures of the midbrain and the brain stem represent the points of central integration for coordination of essential reflexive acts, such as swallowing and vomiting, and those that involve the cardiovascular and respiratory systems. They also include the primary receptive regions for most visceral afferent sensory information.

The **cerebellum** arises from the posterior pons behind the cerebral hemispheres. In addition to maintaining postural regulation achieved by the proper tone of the musculature and providing continuous feedback during voluntary movements of the trunk and extremities, the cerebellum may also regulate visceral function, and plays a significant role in learning and memory.

The spinal cord extends from the caudal end of the medulla oblongata to the lower lumbar vertebrae and transmits sensory information from the peripheral nervous system, both somatic and autonomic, to the brain and routes motor information from the brain to various effectors, e.g. skeletal muscles, cardiac muscle, smooth muscle, or glands. The spinal cord is divided into anatomical segments that correspond to divisions of the peripheral nerves and the spinal column (Figure 3.23). Within local segments of the spinal cord, autonomic reflexes can be elicited.

The Blood–Brain Barrier (BBB)

The BBB is a protective network that filters blood flow to the brain. The BBB is composed of dense and tightly packed layers of endothelial cells around the brain's blood vessels.

The BBB regulates the access of physiological substances and xenobiotics to the CNS. It effectively impairs the transfer of charged, hydrophilic, and high-molecular-weight compounds from circulation to nervous tissue, but does not provide protection against small, lipid-soluble compounds. Several transport mechanisms, like monocarboxylic-acid transporters or organic-anion transporters present in endothelial cells, mediate active transport through the BBB. Other transporters, like ABC-transporters, mediate the excretion of compounds out of the endothelial cells, back into the peripheral blood stream. Leaking of the BBB makes the brain vulnerable to toxins. Such areas, which are not completely protected by the blood–brain barrier, include those regions of the brain involved in neuroendocrine activity, such as the area postrema, hypothalamus, and pineal body, as well as places where the barriers are fenestrated, e.g. in autonomic ganglia, and motor and sensory nerve terminals. All these sites are potential points of entry for toxins.

Cells of the Nervous System

Neurons are the fundamental unit of the nervous system. The electrical charge across neuronal membranes allows the conduction of signals. They are supported by glial cells, which build a sustentacular tissue in the nervous system. Astrocytes, oligodendrocytes, and microglia are glial cells of the CNS. Schwann cells are glial cells in the PNS.

Although the fine structure of the organelles of the CNS and PNS is not exclusive to these tissues, the interactions between cell types, such as synaptic contacts between neurons and myelin sheaths around axons, are unique. Diverse cell types are organized into assemblies and patterns such that specialized components are integrated into the nervous system. Neurons are cells within the nerve systems that integrate and conduct information; glial cells have supporting functions for the neurons, as well as immunological and signalling functions.

Neurons significantly differ from other cell types. Most neurons are highly branched cells that consist of cell bodies and appendices. The appendices are categorized as axons and dendrites (Figure 3.24). The **axons** represent appendices of the nerve cell that are usually not branched. Parts of the axons, which transmit information from neurons to other nerve cells, organs, or tissues, are surrounded by concentric myelin layers that have high electric impedance. The myelin layers are disconnected at the nodes of Ranvier, which occur at 1.5 mm intervals along the axon. During fast axonal transmission, electrical signals jump from one node of Ranvier to another. In nonmyelinated axons, depolarization moves slowly down the axon. **Dendrites** are highly branched and are a point of communication to other neurons. They receive signals from other neurons, integrate the data, and transmit the information to the rest of the neuron.

Neurons of the adult brain are terminally differentiated cells that have not been thought to undergo proliferative responses to damage. However, recent evidence suggests the possibility of neuronal stem cell proliferation as a natural means for replacement of neurons in selected areas of the CNS. As a result, neurons of the CNS have evolved other adaptive mechanisms to maintain function following injury. In the PNS damaged cells can be resubstituted. This requires surviving neurons, which expand their territory by axonal branching, overtaking territory vacated by dead neuronal cells. This process is ineffective in the brain and spinal cord.

Other adaptive mechanisms provide the brain with considerable capacity for structural and functional modification well into adulthood. The brain is highly resistant to damage or toxic effects and is functionally flexible. Dysfunction in selected brain areas can be compensated for by other areas of the brain. Another reason for *the maintenance of function in the face of injury* is the strong homeostatic protection of the brain milieu based, in part, on the specialized vascularization of that organ.

Neurons are very active cells and have a high metabolic demand. Almost all neurons are highly branched. Long and energy-consuming routes of transport are necessary to provide the nerve terminals with nutrients and essential compounds. Purkinje cells of the cerebellum, or motor neurons in the PNS, both of which have very long axons, represent examples of highly effective systems for distributing metabolites between the cell body, the dendrites, and the

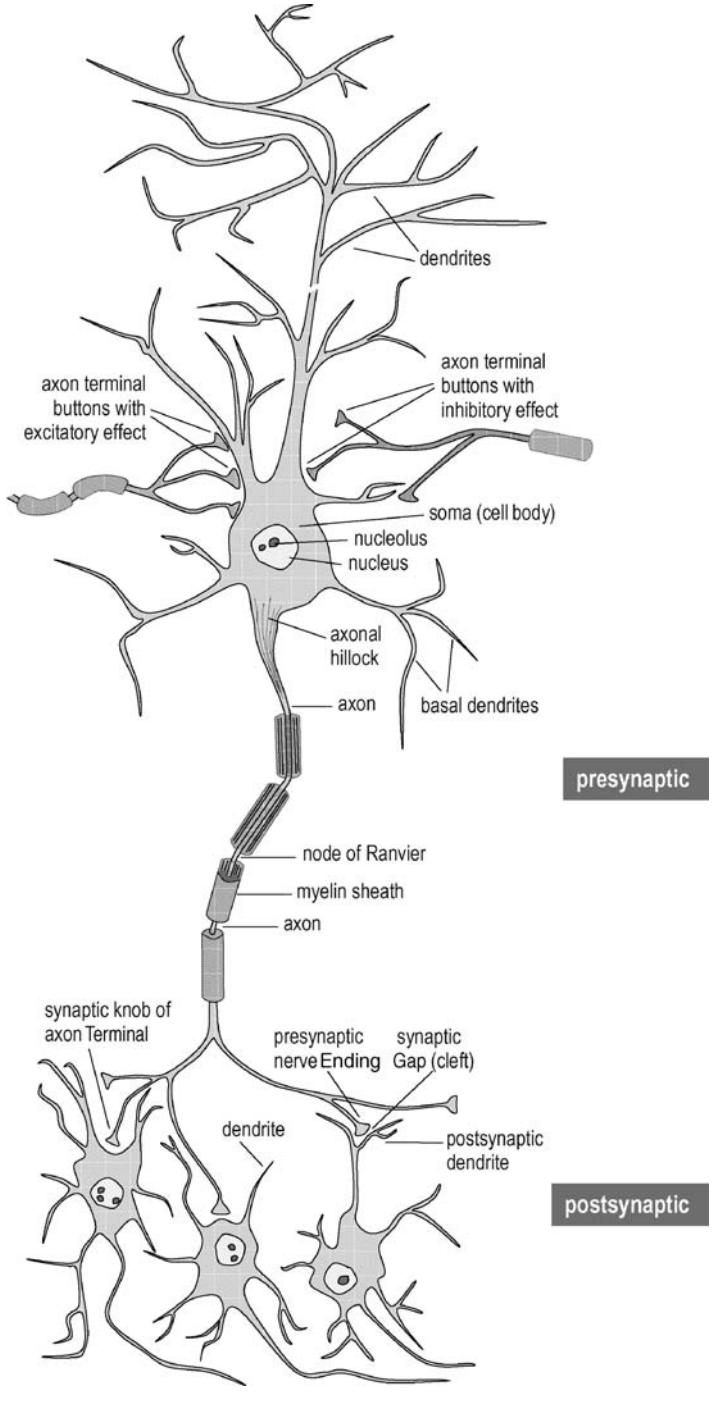


Figure 3.24 *Schematic diagram of neurons. Neurons are a major class of cells in the nervous system. Their primary role is to process and transmit neural information.*

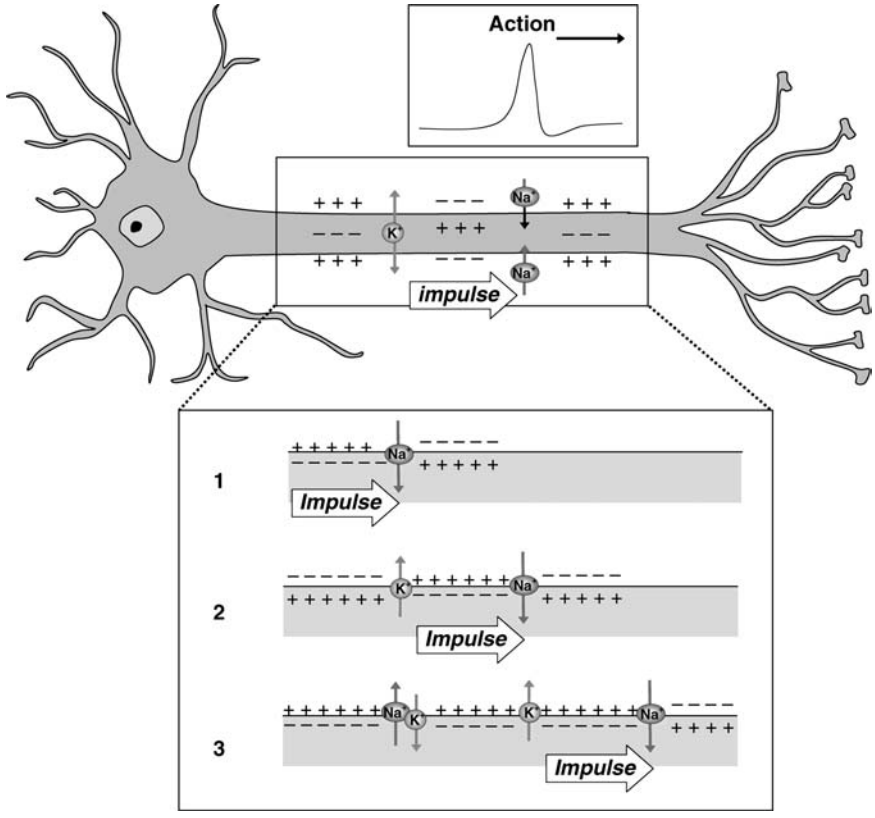


Figure 3.25 *Depolarisation of the neuronal membrane due to opening of gated Na^+ channels. The membrane voltage changes during an action potential as a the result of changes in the permeability of the membrane to sodium and potassium ions. In resting neurons, non-gated K^+-channels are open, but the more numerous gated Na^+ channels are closed. The movement of K^+ ions outward establishes the inside-negative membrane potential characteristic of neuronal cells. Opening of gated Na^+ channels permits-an influx of sufficient Na^+ ions to cause a reversal of the membrane potential (1). The impulse moves along the neuronal membrane, thousands of Na^+ channels open and that prolong the action potential (2). Meanwhile, K^+ channels open and K^+ ions are excreted out of the cell leading to a inside-negative characteristic of the neuronal cell (2). Finally, the Na^+/K^+ ion pump restores the original ionic conditions within the respected segment of the axon (3).*

axons of the neuron. Hence, neurons are especially vulnerable to chemicals that destroy the myelin layers or the cytoskeleton, or interfere with the supply of energy.

Neuronal cells express a high density of ion channels to reliably assure the transmission of action potentials from one neuron to another (Figures 3.25 and 3.26). The maintenance of the ion gradient requires continuous and active ion transport, during both normal and stimulated neurological activity. Glucose is the energy-producing fuel for the brain. Adenosine triphosphate (ATP) is generated during the oxidation of glucose to water and carbon dioxide via the sequential pathways known as glycolysis and the Krebs Cycle. Early in the course of neuronal activation the rapid requirement for energy results in a disproportionate reliance on glycolysis. Ultimately, however, oxidative phosphorylation is the major source of energy

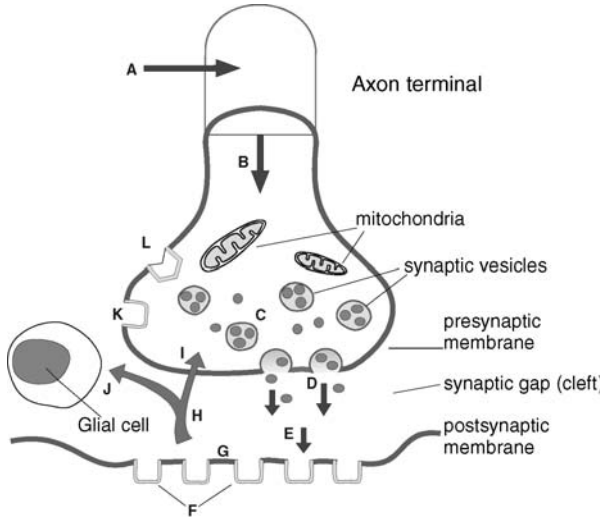


Figure 3.26 *Communication between two nerve cells. Synapses are specialized junctions through which cells of the nervous system signal to one another and to nonneuronal cells such as muscles or glands. The synaptic cleft separates the plasma membrane of the pre-synaptic and post-synaptic cells. Arrival of action potentials at a synapse causes release of neurotransmitters by pre-synaptic cell their diffusion across the synaptic cleft, and their binding by specific receptors on the plasma membrane of the postsynaptic cell. Generally these signals depolarize the postsynaptic membrane, tending to induce an action potential in it. (A) Neuronal uptake of transmitters or pre-stage compounds by specific mechanisms. (B) Axonal transport of the absorbed or synthesized transmitters to the nerve ends and (C) storage in vesicles. (D) Exocytosis of the transmitters into the synaptic cleft. This process is initiated by an Ca^{2+}-influx due to an axonal action potential. (F) After diffusion through the synaptic cleft, the transmitter interact with specific receptors in the postsynaptic membrane. The inactivation of transmitters results from different mechanisms: (H) decomposition or uptake at the post-synaptic site (G), re-diffusion and (I) re-uptake into the pre-synaptic neuron or (J) in glial cells.*

production for the brain. The vulnerability of neurons to disturbances in these functions varies considerably, and is dependent upon time, functional status, or cell type.

Other brain cell types include astrocytes, which assist in neurotransmission and neuronal metabolism by controlling the development of the nervous system in early developmental stages of life and sustain vascular, ionic, and metabolic homeostasis. These neuronal cells synthesize components essential for the regeneration and maintenance of the nervous system. The **microglia** are part of the immune system in the CNS. **Oligodendrocytes** separate nerves in the CNS by the formation of boundaries around axons. In the PNS that function is assigned to **Schwann cells.**

The PNS contains two types of nerve fibers: Thin, nonmyelinated fibers are embedded in Schwann cells. Medullated fibers are composed of Schwann cells enveloped by thick layers of lipids and proteins called the medullary nerve sheath. Myelin represents the main lipid component of oligodendrites and Schwann cells. Gliosis refers to the characteristic reaction of astrocytes after nerve damage, which involves the production of a dense fibrous network of neuroglia in areas of damage, resulting in scars in the CNS.

Neurotransmission

Neurotransmission describes the transmission of information between nerve cells or from nerve cells to other cells. Signals are transmitted (i) electrically within neurons and activated by a function termed the action potential, and (ii) chemically from one neuron to another. Chemical transmission involves compounds called neurotransmitters, which are stored in synaptic vesicles. Synapses are the junctions between nerve terminals of different neurons. Upon electrical excitation of one neuron, neurotransmitters are released in the synaptic cleft. They diffuse to the postsynaptic site and bind to postsynaptic receptors. These receptors are specialized proteins embedded in the postsynaptic membrane. Binding of neurotransmitters induces a biochemical response in the postsynaptic neuron, such as new action potentials.

Nerve impulses elicit responses in neurons by liberating specific chemical neurotransmitters. The sequence of events involved in neurotransmission is of particular importance because pharmacologically active substances and neurotoxins modulate the individual steps. The arrival of an action potential at the axonal terminals initiates a series of events that trigger transmission of an excitatory or inhibitory impulse across the synapse. Synapses between neurons and smooth, cardiac, and skeletal muscles or exocrine glands are called neuroeffector junctions (Figures 3.25 and 3.26). The synaptic cleft separates the plasma membrane of the presynaptic and postsynaptic neurons. **Neurotransmitters** are synthesized by nerve cells and stored in vesicles at the presynaptic membrane (Figure 3.26). Neurotransmitters of the PNS include acetylcholine, norepinephrine (noradrenaline), epinephrine (adrenaline), serotonin, and histamine. Acetylcholine, serotonin, and histamine are also abundant in the CNS. However a wider variety of chemical transmitters exists in the CNS.

Excitatory synapses are characterized by a densely thickened postsynaptic membrane and contain round vesicles with neurotransmitters such as glutamate and acetylcholine. In contrast, vesicles in inhibitory synapses are usually flattened and postsynaptic membranes of these synapses are only partly thickened. Inhibitory neurotransmitters of the CNS include gamma-aminobutyric acid (GABA) and glycine. The arrival of action potentials at a synapse causes the release of neurotransmitters out of the synaptic vesicles in the presynaptic cell (Figure 3.26). Released neurotransmitters diffuse across the synaptic cleft and bind to specific receptors on the plasma membrane of the postsynaptic site. Transmission may be modulated through inactivation of ion channels or internalization of receptors in the postsynaptic membrane, as well as binding of neurotransmitters to inhibitory presynaptic receptors. Neurotransmission is terminated by enzymatic degradation, neuronal re-uptake, or uptake of neurotransmitters in extraneuronal cells (Figure 3.26).

Binding of neurotransmitters to specific receptors on the postsynaptic neuron generally induces signals that depolarize the postsynaptic membrane, tending to induce an **action potential**. The depolarization of the neuronal membrane due to opening of gated Na^+ channels is described in Figure 3.25. In resting neurons, nongated K^+ channels are open, but the more numerous gated Na^+ channels are closed. The movement of K^+ ions outwards establishes the inside-negative membrane potential characteristic of neuronal cells. Opening of gated Na^+ channels permits an influx of sufficient Na^+ ions to cause a reversal of the membrane potential. The impulse moves along the neuronal membrane, thousands of Na^+

channels open, and that prolongs the action potential. Meanwhile, K^+ channels open and K^+ ions are excreted out of the cell, leading to an inside-negative characteristic of the neuronal cell. Finally, the Na^+/K^+ ion pump restores the original ionic conditions within the respective segment of the axon. The action potential spreads out over the whole neuron and is able to induce neurotransmitter release at a synaptic site of that particular nerve cell.

3.4.2 The Nervous System - Site of Attack for Toxins

Neurotoxicology investigates adverse effects of various agents (drugs, chemicals, biologicals) on the structure or function of the nervous system. All parts of the CNS and the PNS are potential points of attack for toxins. The nervous system is especially vulnerable to toxins, since it is exceptionally complex and undergoes a prolonged period of development characterized by cellular migration, differentiation, and synaptic pruning. Neurotoxic effects are usually classified based on the structure of the nerve cell that is impaired. Hence, neuronopathy, axonopathy, or demyelination can be differentiated based on neurotoxic insults.

Potential neurotoxins can be absorbed by inhalation, contact with the skin, or ingestion. Absorbed compounds enter the circulation and can undergo metabolism in the liver, while sublingual, rectal, or direct venous application circumvents the initial liver passage. Exposure to neurotoxins mostly occurs through breathing, ingestion along with food or water, or contact through the skin. Uptake into the circulation from these sources will always be highly variable, and it is impossible to calculate intuitively the toxic dose unless suitable biomarkers are available to assess the body burden of the toxin. The parts of the nervous system that are affected by chemical substances depend on the distribution of the compounds in the body. Ionic and hydrophilic toxins mainly impair the PNS, since they do not cross the BBB. Lipophilic agents and substances that are carried by special transport mechanisms will enter the brain and affect the CNS. Toxicological effects in the CNS can vary extremely depending on brain region, cell type, and residence time. Thus, only a minor number of compounds are neurotoxic in general, and most of the toxins exhibit highly selective effects.

Compounds that harm endothelial cells, such as nitropropionate, dinitrobenzene, or metronidazole, lead to malfunction of the BBB, rendering it porous and permitting the entry of toxins to the CNS. Although not neurotoxic per se, the damage that they cause may ultimately result in neurotoxic effects.

Intoxication depends on the dose of compound and the duration of exposure. In terms of duration of exposure, acute and chronic intoxication that lead to reversible or irreversible effects have to be distinguished. For example, acute exposure of organic solvents causes a reversible CNS depression. Chronic exposure, to small doses, may result in peripheral nerve degeneration. Hence, it is important to distinguish between acute and chronic exposures resulting in acute and chronic effects.

Cellular Damage Resulting in Neurotoxicity

Impairment of neurons in defined regions of the nervous system results in characteristic disturbances grossly observed as symptoms of specific neuronal diseases. The resulting neuronal damage can be classified based on the subcellular target of the toxic agent.

Neuropathy The susceptibility of neurons to neurotoxins varies; few chemicals affect all types of neuronal cells in the brain (Table 3.4A). Neuropathy is subdivided into cytoplasmic, nuclear, and postsynaptic neuropathies. Damage of the cell soma is followed by the loss of axons and dendrites. Neuropathies in the PNS are generally axonopathies and damage to neurons by chemicals can ultimately lead to cell death. The loss of neurons in the adult brain is irreversible. A peripheral neuropathy may have its origin in the neuron, causing either cell death or dysfunction. There may be degeneration of the axon or disruption of neuronal or axonal function caused by damage to the myelin sheath. Alcoholism is the most common cause of peripheral neuropathy. In contrast to the CNS, regeneration and reinnervation of surviving neurons is possible in the PNS.

Specific Neurotoxins **Methylmercury** is the best known neurotoxin associated with neuropathy, although the mechanisms of toxicity are poorly understood. The cerebellar granular cells are extremely sensitive to methylmercury. Small neurons of the visual and the cerebellar cortex die after intoxication with methylmercury, leading to central neuropathy. Symptoms include sensory deficits followed by ataxia and impaired motor coordination. Experimental data indicate that methylmercury also causes peripheral neuropathies in rats, where accumulation in spinal ganglia leads to death of sensory neurons. However, this type of special peripheral neuropathy following methylmercury intoxication has not been observed in humans.

MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) MPTP damages neurons in the brain. The compound was formed as a contaminant during the synthesis of MPPP (1-methyl-4-phenyl-4-propionoxypiperidine), a ''designer'' drug of abuse in the late 1970s as an alternative to heroin, which exhibits opioid analgesic effects. MPTP results in an irreversible neurodegenerative syndrome, similar to that of Parkinsonism. The active metabolite of MPTP, methylphenylpyridinium (MPP^+), selectively induces neuropathy in dopaminergic neurons of the substantia nigra, an area of the brain important for motor coordination.

Trimethyltin is a potent biocide that selectively causes neuropathy in the limbic system of the brain. Trimethyltin poisoning is characterized by a marked augmentation of excitability of hippocampal neurons that results in cell death.

Axonopathy Nerve damage classified as a central or peripheral distal axonopathy is produced by a variety of chemicals (Table 3.4B). Distal axonopathy, also known as 'dying-back' neuropathy, is typical of peripheral neuropathy resulting from metabolic or toxic derangement of the PNS. Distal swelling and degeneration of axons in the CNS and PNS is the most common response of nerves to neurotoxins. Axonopathy is typically gradual in onset, affecting first the distal regions of long axons and advancing slowly towards the nerve's cell body. Sensory loss tends to precede loss of motor function. The symptoms spread proximally as the axon 'dies back' for as long as exposure lasts. Loss of ankle reflexes is an early indication of axonopathy. Depending on the severity of the damage, neurons may survive, but loss of neuronal connections occurs. If the noxious stimulus is removed, regeneration is possible, though prognosis depends on the duration and severity of the stimulus.

Exposure to **acrylamide** produces a distal axonopathy in both humans and experimental animals. Neurons of both the CNS and PNS are affected. The mechanism by which this distal axonopathy is produced remains unknown, although experimental data

Table 3.4 *Neurotoxins that cause structural damages in the nervous system.*

A	**Neuropathies**	
Compound	Mechanism	Clinical signs and symptoms
Cyanide	Inhibition of mitochondrial cytochrome oxidase; breakdown of oxidative phosphorylation	Cell respiration failure (hypoxia)
Carbon monoxide	High affinity for hemoglobin, to form carboxyhemoglobin	Insufficient supply of oxygen (hypoxia)
Methanol	Metabolic production of formic acid	Destruction of visual nerves, Blindness
MPTP	Active metabolite MPP^+, inhibition of mitochondrial respiration chain (complex I)	Destruction of dopaminergic neurons, Parkinson's syndrome
Hydroxydopamine	Generation of free radicals	Oxidative stress, destruction of dopaminergic neurons, Parkinson's syndrome
Lead	Replaces calcium > NMDA-receptor interactions	Glutamatergic failure, encephalopathy
Triethyllead	Excitatory compound	Swelling of Golgi-apparatus and endoplasmic reticulum, accumulation of lysosomes; lost of peroxisomes, necrosis specific for the limbic system (hippocampus), delirium
Methylmercury	High affinity for SH-groups > inactivation of proteins	Neuronal failure, encephalopathy
Trimethyltin	Similar to triethyllead	In contrast to triethyltin, no edema of the myelin layers, delirium
B	**Axonopathies**	
Compound	Mechanism	Clinical signs and symptoms
n-Hexane	Metabolites (γ-diketones) > interaction with lysylaminoacids > Destruction of neurofilamentes > disruption of intra-axonal transport mechanisms	Sensor and motor malfunction
Carbon disulfide	Cross-linking of proteins > disturbances of neurofilaments > dying-back axonopathy	Peripheral neuronopathy
Ethanol	Neuronal membrane interactions > membrane fluidity changes > disturbed ion permeability	Encephalopathy

(*continued*)

Table 3.4 *(Continued)*

B	**Axonopathies**	
Compound	Mechanism	Clinical signs and symptoms
Organic phosphates	Inhibition of neuropathy target esterase > aging reaction > irreversible polyneuropathies	Paralysis
2-Propenamide (acrylamide)	Inhibition of retrograde axonal transport mechanisms > dying-back axonopathy	Neuronal dysfunction
Colchicine, taxol	Disruption of microtubule polymerization > antimitotic > disturbances of neurofilaments	Encephalopathy
Arsenic	High affinity for SH-groups > inactivation of proteins > neuronal failure	Polyneuropathy
C	**Demyelination**	
Compound	Mechanism	Clinical signs and symptoms
Hexachlorophene	Membrane interactions > disturbed ion gradient > swelling of myelin > demyelination	Ischemia (central)
Cuprizon	Copper chelator > enzyme inhibition > malfunction of oligodentrocytes > demyelination	Ischemia (central)
Lead	Schwann's cell destruction > endoneuronal edema; no penetration into CNS	Peripheral neuropathy
Triethyltin	Membrane interactions > enhanced chloride and hydroxide exchange > myelin swelling	Nausea, vertigo, convulsions
Tellurium	Squalene epoxidase inhibition > reduced cholesterol synthesis > reduced myelin production	Nausea, vomiting
Toluene	Metabolic activation to reactive products, like benzaldehyde, cresols > diffuse demyelinisation	Amentia, nausea, headache

suggest several theories. Acrylamide interferes with axonal retrograde transport mechanisms essential for the survival of the axon.

Diketone neuropathy, of which n-hexane-induced neuropathy is an example, is characterized by a pathological process of neurofilament accumulation and degeneration of the axon. In humans, this pathology is manifested as a clinical peripheral neuropathy. Electrophysiological changes consist of decreased motor and sensory nerve conduction, with an increased distal latency period. n-Hexane-induced axonopathy studies in experimental animals have shown that the metabolite of n-hexane, i.e. 2,5-hexanedione, a gamma-diketone, is the toxic metabolite. The mechanism of n-hexane-mediated neuropathy is speculated to involve covalent binding of diketones to lysine residues to form pyrrole derivatives, which are important in protein–protein cross-linking between neurofilaments.

Demyelination Glial cells are nonneural support cells in the nervous system and are possible points of attack by neurotoxins. Oligodendrocytes of the CNS and Schwann cells of the PNS produce myelin, which forms layers that electrically isolate the axons. Damage to the myelin layer may, therefore, result in failures of nerve conduction. Myelin damage may involve swelling of myelin, primary demyelination, and secondary demyelination. The swelling of myelin may result from edema following disturbances in the myelin homeostasis or metabolic damage to glia. Functional consequences are a disruption in nerve conduction and increased water content of the tissue. The swelling is partly reversible.

Primary demyelination occurs as a consequence of a direct attack of chemicals on myelin or myelin-producing cells, resulting in loss of myelin or in myelin destruction (Table 3.4C). Again, this damage is partly reversible.

Secondary demyelination is an active process that ultimately leads to the destruction of axons. It occurs as a consequence of a traumatic destruction or toxic axonopathy, where myelin is digested by macrophages. Demyelinating neuropathies of the PNS result from damage to the Schwann cell or to the myelin sheath of the internodes. **Diphtheria toxin** causes segmental demyelination by damaging the Schwann cell. **Inorganic lead** is a compound that causes peripheral demyelination, and motor nerves are specially affected. **Hexachlorophene** and **perhexilene** have also been implicated in myelin disruption. Recovery depends on the replication and activation of surviving Schwann cells. In regenerated axons the internodes tend to be shorter and myelin sheaths thinner than in the axons prior to injury, and conduction velocity is usually slower in remyelinated axons. Hexachlorophene also reaches the CNS and damages the sensitive long nerve fibres. **Triethyltin** leads to myelin swelling. Its neurotoxic effects are almost entirely restricted to the CNS.

Effects on neurotransmission The transmission of information within the neuronal network of the nervous system is extremely complex. Consequently, neurotransmission is especially susceptible to attack by chemicals at several sites. Synapses are particularly sensitive to chemicals. Synaptic function requires the synthesis, storage, and release of transmitters, the reaction of transmitters with receptors and, ultimately, transmitter inactivation. Each of these steps represents a possible point of attack. Reduced inactivation results in an excess of transmitters, which causes continuous depolarization of the nerve cell. An alteration in synthesis, storage, or release of neurotransmitters also profoundly disrupts neurotransmission. Compounds that are structurally related to transmitters interact with their receptors. Chemicals that activate receptors by mimicking transmitters may have agonistic effects. Antagonists occupy receptors and inhibit the

binding of neurotransmitters to their receptors, thus preventing neuronal activation. Mechanisms and clinical symptoms of various other neurotoxins are given in Table 3.5.

Organophosphate derivatives are used in agriculture and in household applications as pesticides. Toxic nerve agents used for chemical warfare are often organophosphate compounds. Many organophosphates are potent neurotoxins, functioning by irreversibly inhibiting acetylcholinesterase. Inhibition of the inactivation of acetylcholine leads to excessively high concentrations of that neurotransmitter at the synapse. Excessive acetylcholine causes continuous depolarization of the postsynaptic cholinergic receptor, thereby blocking neurotransmission.

The effects of organophosphate poisoning are salivation, lacrimation, urination, defecation, gastrointestinal upset, and emesis. Atropine acts as an antagonist and diminishes the effects of acetylcholine on postsynaptic muscarinic cholinergic receptors. Atropine can be used as an antidote to organophosphate poisoning, but requires very high doses.

Excitatory amino acids like **domoic acid**, **kainic acid**, or **glutamate** act as false neurotransmitters. Activation of glutamate receptors results in excessive influx of calcium ions into nerve cells. Although calcium flowing into cells is normal, the uncontrolled increase of internal calcium causes the cell to degenerate.

Domoic acid, also called amnesic shellfish poison, is an algal aminoacid toxin associated with certain algal blooms. The chemical can bioaccumulate in marine organisms that feed on phytoplankton, such as shellfish, anchovies, and sardines. In mammals, including humans, domoic acid acts as a neurotoxin, causing short-term memory loss, brain damage, and death. In the brain, domoic acid damages the hippocampus and amygdaloid nucleus.

Kainic acid, an analogue of glutamate, is present in some algae. Kainate is a specific agonist for the ionotrophic receptor and mimics the effects of glutamate.

Glutamic acid, or the anionic form glutamate, is one of the amino acids used by all organisms in their proteins. Many Asian dishes are characterized by the taste of glutamate which comes both from natural origins such as soy or fish sauce, and added glutamate as a flavor enhancer. Flavor enhancers, such as monopotassium glutamate, calcium diglutamate, monoammonium glutamate, and magnesium diglutamate, have been implicated as a cause of the Chinese restaurant syndrome, which often causes headache, flushing, sweating, or sense of facial swelling after eating in a Chinese restaurant. However, it has not been proven that glutamate is the causative agent.

Other compounds like **muscimol** act as agonists on inhibitory receptors sensitive to the neurotransmitter GABA. **Strychnine** acts as an antagonist at glycine receptors and compensates for the inhibitory effect of glycine in spinal ganglions, which results in seizures (convulsions).

Muscimol is a specific agonist of $GABA_A$ receptors and one of the toxic psychoactive compounds present in the mushroom *Amanita muscaria.* It originates from decarboxylation of ibotenic acid. Both substances produce the same effects, but muscimol is more potent than ibotenic acid. Symptoms of poisoning generally occur within 1–2 hours after ingestion of the mushrooms, and may include drowsiness and dizziness, followed by a period of hyperactivity, excitability, illusions, and delirium.

Strychnine is a very toxic, colorless crystalline alkaloid from *Strychnos nux-vomica.* It is used as a pesticide, particularly for killing small vertebrates such as rodents. Strychnine acts as an antagonist at the inhibitory glycine receptor, a ligand-gated chloride channel in the spinal cord and the brain. Strychnine causes muscular contractures and eventually

Table 3.5 *Assorted neurotoxins: mechanism of action and symptoms of intoxication.*

Compound	Mechanism	Clinical signs and symptoms
Aluminium	Al^{3+}ions > disturbed cellular ion homeostasis	Encephalopathy
Atropine	Agonist for acetylcholine receptors > anticholinegic and parasympatholytic effects	Agitation, anxiety, central respiratory paralysis
Botulinum toxin (Clostridial toxins)	Motoneurons and vegetative neurons of the PNS > inhibition of exocytosis acetylcholine > flag of horizontally striped musculature / loss of vegetative functions	Palsy
Cocaine (alkaloid of *Erythroxylon coca*)	Inhibition of presynapatic neurotransmitter > re-extended and accelerated effect on postsynaptic receptors	Paranoid psychopathy
Domoic acid (excitatory amino acid; produced by dinoflagellate)	Agonists for excitatory glutamate and glycine receptors	Headache, confusion, memory deficits
Glutamate	Agonist for excitatory glutamate receptors > overstimulation of glutamate receptors > e.g., increase of intracellular Ca^{2+} and H_2O levels > cell swelling	Apoplexy, epilepsy
Kainic acid (excitatory amino acid; produced by dinoflagellate)	Agonist for excitatory glutamate receptors > overstimulation of glutamate receptors > e.g., increase of intracellular Ca^{2+} and H_2O levels > cell swelling	Apoplexy, epilepsy
Tetanus toxin (Clostridial toxins)	Interneurons of the PNS > inhibition of exocytosis of inhibitory neuro-transmitters > loss of inhibitory effects on motoneurons > continued contraction	Convulsion
Thallium	Tl^{+} ions > exchange of K^{+} / e.g., intracellular oxidation to Tl^{3+} > damage to mitochondria	Encephalopathy, polyneuropathy

asphyxia and sheer exhaustion. It produces some of the most dramatic, terrifying, best known, and painful symptoms, which are manifested just a few minutes after exposure. Every muscle in the body will start to simultaneously contract. Strychnine poisoning may result from inhalation, swallowing, or absorption through eyes or mouth, and can be fatal.

Botulinum toxin is the toxic compound produced by the bacterium *Clostridium botulinum.* The toxin attacks one of the fusion proteins at the neuromuscular junction, preventing neurotransmitter vesicles from anchoring to the synaptic membrane to release acetylcholine. By inhibiting acetylcholine release, the toxin interferes with nerve impulses and causes the paralysis of muscles seen in botulism. The toxin is the most potent acute toxic substance known, with a lethal dose of about 200–300 pg/kg. Food-borne botulism usually results from ingestion of food that has become contaminated with *Clostridium* spores, which in an anaerobic environment may germinate, grow, and produce the toxins.

Tetrodotoxin is one of those compounds that selectively inactivates sodium channels in nerve cell membranes, inhibits the transmission of action potentials and interrupts electrical signalling within nerves (Figure 3.25). Although tetrodotoxin was discovered in pufferfish and found in several other animals, it is actually a bacterial product. Fish poisoning produced by consumption of members of the order Tetraodontiformes is one of the most violent intoxications caused by marine species. The gonads, liver, intestines, and skin of pufferfish can contain levels of tetrodotoxin sufficient to produce rapid and violent death. Tetrodotoxin poisoning produces symptoms including shortness of breath, numbness, tingling, light-headedness, paralysis, and irregular heartbeat. This potent neurotoxin does not cross the blood–brain barrier. Hence, the brain is not affected and it is not unusual for the patient to remain conscious while paralysed.

3.4.3 Clinical Signs and Symptoms Induced by Neurotoxins

The mature nervous system is remarkably vulnerable to toxin-induced damage, and because any damage may disrupt the extensive communication systems that characterize the brain, neurotoxins have the capacity to affect gait and posture, sensation, behavior, and cognition, and produce a complex pattern of clinical signs and symptoms. These signs and symptoms may be expressed in the central, the peripheral, and the autonomic nervous systems. Poisoning with neurotoxins is often associated with pain, and changes in the special senses of taste and smell, as well as changes in visual acuity and hearing.

Hypoxia describes a state of oxygen deficiency that causes an impairment of function. Reasons for hypoxia in the brain are mainly inadequate oxygen transport, the inability of the tissues to use oxygen, or a reduction in the oxygen carrying capacity of the blood. **Carbon monoxide** and **cyanide** poisoning are responsible for histotoxic hypoxia, defined as the inability of the tissues to use oxygen (Table 3.4A). Certain narcotics and alcohol also prevent oxygen use by the tissues (Table 3.4B). Symptoms of mild cerebral hypoxia include inattentiveness, poor judgment, memory loss, and a decrease in motor coordination. Brain cells are extremely sensitive to oxygen deprivation. Long-lasting hypoxia can cause coma and even brain death. Whereas hypoxia is a general term denoting a shortage of oxygen, **ischemia** refers to a specific condition where there is shortage of blood supply to a particular organ. Since oxygen is mainly bound to hemoglobin in red blood cells, insufficient blood supply can cause tissues to become hypoxic, or, if no oxygen is supplied

at all, anoxic, which will ultimately cause necrosis and cell death. The brain is especially sensitive to ischemia, triggering a process in which proteolytic enzymes, reactive oxygen species, and other harmful chemicals damage and may ultimately kill brain tissue. Hypoxia or ischemia may cause **encephalopathy** which, depending on its type and severity, may manifest neurological symptoms such as progressive loss of memory and cognitive ability, subtle personality changes, inability to concentrate, lethargy, and progressive loss of consciousness. Acute **encephalopathies**, most of which are mild and resolve within a few days, are common. Numerous compounds, including aluminium, cannabis, cocaine, domoic acid, lead, organic solvents, and trimethyltin have been demonstrated to be potential causative agents (Table 3.4C and 3.5). The transition from acute mild to chronic severe encephalopathy with associated loss of cognition and psychomotor function is relatively uncommon, but has been reported after acute severe poisoning by domoic acid, aluminium, cadmium, and lead, and following chronic abusive use of alcohol or organic solvents. Although acute signs are usually rapidly resolved, persistent problems may have **serious adverse effects** that necessitate long-term psychiatric and psychological assessments and monitoring.

Cerebellar and basal ganglion dysfunctions result in **disorders of movement**, characterized by ataxia, intention tremor, and loss of coordination. This condition is best known as a feature of chronic exposure to mercury; however, overdose with a variety of potentially toxic drugs and chemicals such as MPTP, 5-fluorouracil, lithium, and acrylamide have also been implicated. Extrapyramidal syndromes such as Parkinsonism, dystonias, dyskinesias, and tics are relatively well known toxic syndromes. Tics are repeated, uncontrolled or impulsive actions that appear almost reflexive in nature. The toxic mechanisms are poorly understood but are usually reversible, although problems may recur many years after the original onset of the disorder. In most cases, the psychiatric abnormalities are relatively mild, but major cases of dementia and a Parkinsonism/dementia syndrome have been associated with aluminium toxicity, cerebellar ataxia with lithium overdose, and severe psychotic disorders related to use of psychedelic drugs such as lysergic acid diethylamide (LSD).

Common clinical complaints caused by neurotoxins are the loss or changes in the perception of taste and smell. Organic solvents are frequently implicated although the pathophysiological basis is not yet fully understood. A perception of change in taste is common following the use of numerous therapeutic agents, but is usually reversible. The **loss of hearing** has been attributed to the abuse of organic solvents, e.g. toluene. It is more often associated with the use of well known ototoxic drugs such as the aminoglycosides. Direct attack on the neuronal components of the **visual system** is less common. Mydriasis and miosis are obvious signs of exposure to cholinolytic and cholinomimetic agents, such as atropine or muscarine respectively (Table 3.5). The optic nerve can be damaged by toluene, which causes demyelination, or hexachlorophene which causes myelin deformation (Table 3.4C). Alcohol abuse is also associated with widespread damage to the neuronal components of the visual system, but it is suspected that the aetiology is confounded by the poor nutritional status of many chronic abusers of alcohol.

3.4.4 Summary

The peripheral and the central nervous systems are extremely complex networks that provide the structural basis for electrochemical processes that are responsible for

multiple functions such as cognition, sensation, memory, or speech. Neurons are the fundamental unit of the nervous system. The electrical charge across neuronal membranes allows the conduction of signals. For proper ion gradients, the supply of energy is essential for neurons. Hence, neurons are highly dependent on aerobic metabolism. Neurons are supported by glial cells, which build a sustentacular tissue in the nervous system. All parts of the nervous system are potential points of attack for toxins. The nervous system is especially vulnerable to toxins, since it is exceptionally complex and undergoes a prolonged period of development. Neurotoxic effects are usually classified based on the structure of the nerve cell that is impaired. Neurotoxins typically target the neuron, the axon, or myelinating cells. Synthetic chemicals, as well as naturally occurring toxins, may interrupt neurotransmission, including inhibition of transsynaptic communication, inhibition of neurotransmitter re-uptake, or interference with second-messenger systems. The mature nervous system is remarkably vulnerable to toxin-induced damage, and, because any damage may disrupt the extensive communication systems that characterize the brain, neurotoxins have the capacity to affect gait and posture, sensation, behavior, and cognition, and produce a complex pattern of clinical signs and symptoms. Poisoning with neurotoxins is often associated with pain, changes in the special senses of taste and smell, as well as changes in visual acuity and hearing.

For Further Reading

R. C. MacPhail, Principles of Identifying and Characterizing Neurotoxicity, *Toxicol. Lett.*, **64–65**, 209–215 (1992).

T. Narahashi, M. L. Roy, and K. S. Ginsburg, Recent Advances in the Study of Mechanism of Action of Marine Neurotoxins, *Neurotoxicology*, **15**, 545–554 (1994).

L. Vecsei, G. Dibo, and C. Kiss, Neurotoxins and Neurodegenerative Disorders, *Neurotoxicology*, **19**, 511–514 (1998).

R. M. LoPachin, Redefining Toxic Distal Axonopathies, *Toxicol. Lett.*, **112–113**, 23–33 (2000).

D. H. Overstreet, Organophosphate Pesticides, Cholinergic Function and Cognitive Performance in Advanced Age, *Neurotoxicology*, **21**, 75–81 (2000).

M. A. Philbert, M. L. Billingsley, and K. R. Reuhl, Mechanisms of Injury in the Central Nervous System, *Toxicol. Pathol.*, **28**, 43–53 (2000).

W. Zheng, Neurotoxicology of the Brain Barrier System: New Implications, *J. Toxicol. Clin. Toxicol.*, **39**, 711–719 (2001).

J. B. Harris and P. G. Blain, Neurotoxicology: What the Neurologist Needs to Know, *J. Neurol. Neurosurg. Psychiatry*, **75**, iii, 29–34 (2004).

D. Zemke, J. L. Smith, M. J. Reeves, and A. Majid, Ischemia and Ischemic Tolerance in the Brain: an Overview, *Neurotoxicology*, **25**, 895–904 (2004).

S. C. Bondy and A. Campbell, Developmental Neurotoxicology, *J. Neurosci. Res*, **81**, 605–612 (2005).

D. R. Wallace, Overview of Molecular, Cellular, and Genetic Neurotoxicology, *Neurol. Clin.*, **23**, 307–320 (2005).

F. E. Bloom. Neurotransmission and the Central Nervous System, in *Goodman and Gillman's the Pharmacological Basis of Therapeutics*, L.L. Brunton, L.S. Lazo, and K. L. Parker (Eds), McGraw Hill, New York (2006).

T. C. Westfall and D. P. Westfall, Neurotransmission – The Autonomic and Somatic Motor Nervous System, in *Goodman and Gillman's the Pharmacological Basis of Therapeutics*, L.L. Brunton, L.S. Lazo, and K.L. Parker (Eds), McGraw Hill, New York (2006).

D. C. Anthony, T.J. Monine, W.M. Valentine and D.G. Graham, Toxic Responses of the Nervous System, in: *Casarett and Doull's Toxicology*, C.D. Klaassen (Ed.), McGraw Hill, New York (2001).

3.5 Behavioral Neurotoxicology

Andreas Seeber

3.5.1 Introduction

Neurotoxicity is the consequence of adverse effects on the structures and functions of the nervous system. Behavioral toxicology refers to the functional consequences of neurotoxicity and includes cognitive, executive, and sensory effects, the personality domain, as well as peripheral nerve functions and electrophysiologically detectable effects, which are not elaborated here.

To assess the variability of neurobehavioral effects certain methodological requirements have to be met.

1. The exposure has to be measured using acute or chronic testing procedures and past exposures should be estimated.
2. The methodological approach should be supported by information on the underlying neurotoxic mechanism or the brain structures affected by the toxicant.
3. The design of a study and the statistical evaluation of the data should enable the identification of neurobehavioral effects as a function of exposure. Confounders such as age, education, verbal intelligence, etc. that may affect the outcome of the study should be excluded.
4. Significant associations between exposure and neurobehavioral test results do not necessarily imply impairment of health. A careful check of criteria is necessary to evaluate whether an observed effect is adverse.

3.5.2 Exposure Assessment

At workplaces data on exposures of a substance usually are air concentrations and represent time-weighted average (TWA) values of an 8-h shift. If exposure is Concentration (C) $\times$ Time (T), the total exposure after repeated exposure would be: as given in Equation (3.1), assuming that the biological effect depends on the total dose including estimates of prior workplace exposures.

$$CT_1 + CT_2 + CT_3 + \ldots = CT_{\text{total}} \tag{3.1}$$

Calculations of 'hygienic effects' (HE) estimate exposure to mixtures of chemicals with regard to the limit values of the substances in the mixture. Although measures of exposure are usually expressed as mean values over an 8-h day, it is appreciated that excursions resulting in high peak exposures could impact on the observed toxic effects.

Time-weighted averages of air concentrations at workplaces or data from biological monitoring of exposed workers provide data for investigations into associations between exposure and potential alterations of neurobehavioral function. Complete exposure

assessment for all relevant substances over all periods of exposure is necessary to establish dose–effect relationships.

To estimate past exposures, job exposure matrices are used taking into account technological changes in a plant during specific time periods at different workplaces. For example, in the European printing industry over a 20-year period exposures to toluene were reduced approximately 4-fold and in boat-building plants exposures to styrene were reduced at least by a factor of 3. Therefore, estimation of the individual lifetime-weighted average exposure (LWAE) can only be accomplished after reviewing the work history and cumulative lifetime exposure (CE in ppm years) of the worker. LWAE and current information on workplace exposure should be considered independently when estimating total lifetime exposure. Table 3.6 shows how the exposure indices HE, CE, and LWAE are calculated, using Equations (3.2)–(3.4).

The HE is a summarized index of threshold limit value (TLV) weighted concentrations of the compounds in chemical mixtures, such as solvents. When searching for associations between HE and neurobehavioral effects, either equal neurotoxic actions or no interaction or potentiation between the compounds is assumed. HE values provide relative information on exposure depending on the threshold limit value for a substance. Usually only those compounds exceeding 10 % of the limit value are included in the calculation. HE values increase if threshold limit values are reduced and if the number of compounds in the mixture is elevated. HE < 1 indicates an exposure level below the summarized limit of the compounds in the mixture, whereas HE > 1 indicates greater exposure. A comparison between studies of different countries or time periods requires information on corresponding exposure limits.

Establishment of no-observed-adverse-effect level (NOAEL) or lowest observed-adverse-effect level (LOAEL) values in behavioral toxicology is facilitated when the broadest possible range of exposures to the chemical(s) in question are investigated.

Table 3.6 *Exposure indices used in neurobehavioral studies.*

HE, hygienic effect: Sum of TLV-weighted concentrations of the individual solvents

$$\text{HE (Job)} = \sum_{i=1}^{N} \frac{C_i}{\text{TLV}_i} \tag{3.2}$$

C_i = Concentration of the solvent i
TLV_i = Threshold limit value of solvent i

CE, cumulative lifetime exposure for different jobs in an individual working life

$$\text{CE (person)} = \sum_{i=1}^{N} C_i\, t_i \tag{3.3}$$

C_i = Concentration of the substance for the job i (ppm),
t_i = Time in job i (years)

LWAE, lifetime-weighted average exposure derived from CE

$$\text{LWAE (person)} = \frac{1}{t} \sum_{i=1}^{N} C_i\, t_i \tag{3.4}$$

t = Exposure time of the working life (years).

3.5.3 Methods

Ideally, the relationship between behavior and its underlying morphological and physiological mechanisms can be established. However, significant gaps exist between the biological and behavioral bases of neurotoxicity. Experimental or epidemiological studies in exposed people using neurobehavioral methods can be done side by side with morphological and physiological analyses. Estimating the manner in which these data are linked unambiguously is a challenge. However, it is possible to view the results as self-contained data obtained by a broad spectrum of standardized methods upon which cause-and-effect hypotheses can be conceptualized.

Neuromorphological and Neuropathological Methods

Neuromorphological and neuropathological methods determine alterations of nerve cell bodies, axons, dendrites, neurofilaments, synaptic terminals, and myelin sheaths. Histopathological and histochemical methods permit detection of pathological lesions in nerve cells and nervous tissue with respect to the types of cell injury, the time course over which damage appears, and the extent of injury. Neuroanatomical data characterize toxin-induced structural damage to central and peripheral nervous structures. Quantitative methods are available to measure how changes in the total number of specific types of neurons may be related to exposure to specific toxicants. The results of such studies can provide the basis for developing hypotheses aimed at linking these lesions to changes in behavior.

Neurophysiological Methods

The primary electrophysiological procedures include studies of membrane ion channels, synaptic transmission, and analysis of sensory-evoked potentials. The latter are intended to localize areas of neurotoxicological impact, and will be outlined in some detail.

The study of sensory functions by sensory-evoked potentials is a procedure to examine complex neural systems both in animals and man. It includes the stimulation of sensory receptors or afferent nerves. Typical stimuli are light flashes or changes of visual patterns aimed at eliciting visual-evoked potentials, sounds intended to elicit auditory-evoked potentials, and stimulation of peripheral nerves with electrical impulses to generate somatosensory-evoked potentials.

The investigation of sensory reactions presents problems due to the small voltage of the evoked potentials. The data consist of positive and negative voltage deflections. Event related potentials are commonly depicted as waveforms representing voltage against time. However, techniques must be used to overcome problems associated with the signal-to-noise ratio. Evoked potentials can be optimized by careful placement of the electrodes, adequate filtration and amplification of the potentials, and repeated measurements. Event-related potentials are commonly depicted as waveforms representing voltage against time. Interpretation of the data requires the use of averaging procedures and statistical analysis. Neural electrical activity can be detected from electrodes placed on the skull or on the surface of the cortex. The signals may arise from excitatory

postsynaptic potentials, inhibitory postsynaptic potentials, and action potentials from an electrically active focus called a neural generator. Surface-mapping techniques permit the localization of neural generators.

Examples of Sensory-Evoked Potentials *Electroretinograms* are evoked responses from the cornea in response to visual stimulation. Negative and positive waves and oscillations of the waves are electrical characteristics, which can be interpreted as responses of rods and cones, of the middle retina, or of other retinal structures.

Flash-evoked potentials represent the cortical responses to visual stimulation mapped by use of surface recordings at the visual cortex and surrounding areas. Positive and negative waves can be associated, over their time course, with different origins in the cortical structure, e.g. a negative wave 40 ms after flash (N_{40}) is thought to be a sign for depolarization processes in the visual cortex lamina 4 following the thalamocortical input.

Brain stem auditory-evoked responses are composed of several peaks occurring between 10 and 1000 ms after stimulation. Early peaks are generated by the auditory nerve and the cochlear nucleus. Subsequent peaks seem to be generated by activity in several structures representing higher-level processing of the auditory signal at the level of brain stem.

Lists of generators of visual-, auditory- or somatosensory-evoked potentials have been offered to support hypotheses on the association between exposure to neurotoxicants and alterations of evoked potentials in neural structures.

Neurobehavioral Methods

Approach Variations of human behavior can be measured using standardized methods. Deviations from expected norms or patterns of tested behavior can have many causes. Chemical exposure is only one of the possible external sources for variations of behavior. Typical internal sources are the level of education, age, the status in the diurnal rhythm, physical conditions during measurement, as well as motivational or emotional states.

Evaluating the likelihood that a change in behavior is caused by exposure to a toxicant must include a comparison with behavioral variations having other causes. Investigation of the causes of behavioral changes can be studied using carefully designed investigations in which there are controls for confounding external and/or internal factors.

A variety of neuropsychological tests are available for neurobehavioral measurement. These include surveys to evaluate cognitive functions, executive functions, and psychomotor functions, as well as personality or emotional changes.

Cognitive functions involve the processing of learned information. Successful performance on tests of cognitive function require speed and accuracy and can be measured within a short period of time using either paper and pencil methods or by using computer-administered exams.

- Receptive function tests measure the selection, acquisition, classification, and integration of information.
- Memory and learning tests measure the storage and retrieval of information.

- Thinking tests measure the mental organization and reorganization of information.
- Expressive functions and mental activity tests involve evaluation of the means by which information is acted upon and cognitive operations by which the level of consciousness is expressed.

Executive functions relate to management skills. If executive functions are impaired, the individual can no longer organize parallel tasks or maintain social relationships satisfactorily. Deficits in executive functions affect different aspects of behavior:

- A defective capacity for self-control or self-direction;
- A tendency to irritability and excitability, impulsivity, and rigidity;
- Difficulties in shifting one's attention;
- An impaired capacity to initiate activity;
- A decreased motivation, and defects in planning and carrying out sequential activities.

It is difficult to test for clear distinctions between cognitive and executive functions.

Psychomotor functions deal with visible behavior. Motor activity represents the interface between behavior and the environment. Measurable endpoints in psychomotor tests include the time, the force, the speed, and the spatial boundaries of the response. Neurobehavioral tests in animals and infants rely on psychomotor responses because verbal reactions are limited. Compared with cognitive and executive examinations, measured deficits, especially in animal psychomotor tests, can be associated with neural structures, e.g. the relationships between:

- Loss of individual or integrated movements to specific cortical structures;
- Tremor, dystonic postures, turning behavior, and changed blinking rates to the basal ganglia;
- Nystagmus, loss of postural reflexes, and uncoordinated movements to the cerebellum;
- Diminished grip strength or leg orientation and changed gait characteristics to peripheral nerves and neuromuscular junctions.

Personality or emotional changes directed preferably at emotional lability are typical signs of brain damage by neurotoxic agents. They are measured using standardized questionnaires or interviews directed at examining:

- Symptoms of impaired mental and physical health;
- Tolerance to frustration;
- Loss of emotional sensitivity and capacity for modulating emotional behavior;
- Episodes of affective changes.

Examples for Cognitive and Psychomotor Tests

Cognitive, executive, and psychomotor functions are measured using studies designed to show the impact of exposure to chemical toxicants compared with other factors that may influence the analysed function. Among these would be verbal intelligence, age, and alcohol consumption. It is necessary to control for these factors because they may mask some of the effects of the toxicant.

Tests such as the National Adult Reading Test (*NART*) are available to estimate verbal intelligence. It is a single-word, oral reading test. The words are short and irregular, not following normal grapheme–phoneme correspondences (e.g., ache, gauche). Previous familiarity with the words supports successful test performance, and makes minimal demands on current cognitive capacity. The other widespread method is *testing of vocabulary* by a multiple-choice method. In a block of several nonsense words and one correct word the last one has to be marked without additional explanations and time limitation. The number of correctly answered blocks of increasing difficulty gives the estimation of verbal abilities used as 'pre-morbid' ability.

A typical attention test very often used in neurobehavioral studies is the test *Symbol Digit Substitution*. It belongs to the performance tests of the Wechsler Intelligence Scale. In the traditional paper–pencil version (Digit Symbol test, DST) the subjects have to draw in a limited time symbols belonging to digits according to a code listed. The number of correct symbols serves as score. The computer-administered version (Symbol Digit Substitution Test, SDST) requires a reversed order of actions. The subjects have to press the keys of digits for the corresponding symbols presented in random order. Translations from symbols to digits – not from digit to symbols – are required. Time and correct reactions are registered. The SDST-version requires more than the DST-version as regards executive functions of cognitive flexibility. Figure 3.27 shows exemplarily the use of SDST in a study on effects of solvent mixtures in paint manufacturing. Raw data

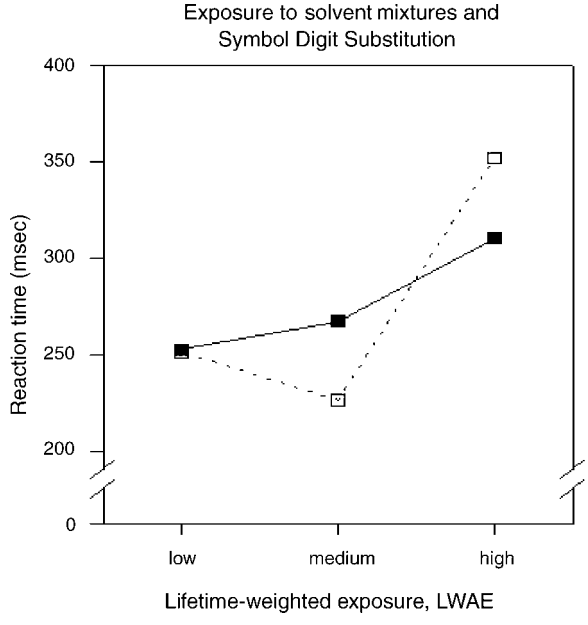


Figure 3.27 *Example for adjustment of data. Reaction time from the Symbol Digit Substitution Test in an epidemiological study with workers exposed to solvent mixtures in paint manufacturing. The open squares represent the raw data, the black squares the adjusted data. [Reprinted from Seeber et al., Food Chem. Toxicol.* **34***, 1113–1120, Copyright (1996), with permission from Elsevier.]*

and the corrected data are presented taking into account lifetime-weighted exposure to solvent mixtures, age, alcohol consumption measured by carbohydrate-deficient transferrin, and verbal intelligence measured by a vocabulary test. After control of the factors age, alcohol, and verbal intelligence a significant association between exposure and performance became apparent.

Meta-analyses on applications of DST and SDST in epidemiological studies of lead- or mercury-exposed workers indicate reproducible associations between biomonitoring data on exposure and test performance.

A typical example for measuring executive functions is the computer-administered test *Switching Attention* covering three sub-tasks. First the subjects have to react compatibly on the left/right position of a square on the screen (block version), then compatibly on the left/right direction of an arrow (arrow version), and finally they have to react either in the block or in the arrow manner depending on the written instruction given just before the stimulus appeared. Speed and errors are measured.

Measurements of *Simple as well as Choice Reaction Time* represent attention tests with low cognitive demands on simple psychomotor stimulus response performances. Simple reaction tests include single visual or auditory stimuli and a simple response, whereas choice reaction tests usually cover two or more stimuli (often four) with fixed corresponding responses. In these tests a first reaction time covers the processing from the appearance of the stimuli up to beginning of the response movement; a second movement time covers the process up to the pressing of the response key. Figure 3.28 shows the results of a choice reaction test (four stimuli with corresponding keys). From two experiments with 3 and 4 h of styrene exposure, respectively, the results for constant

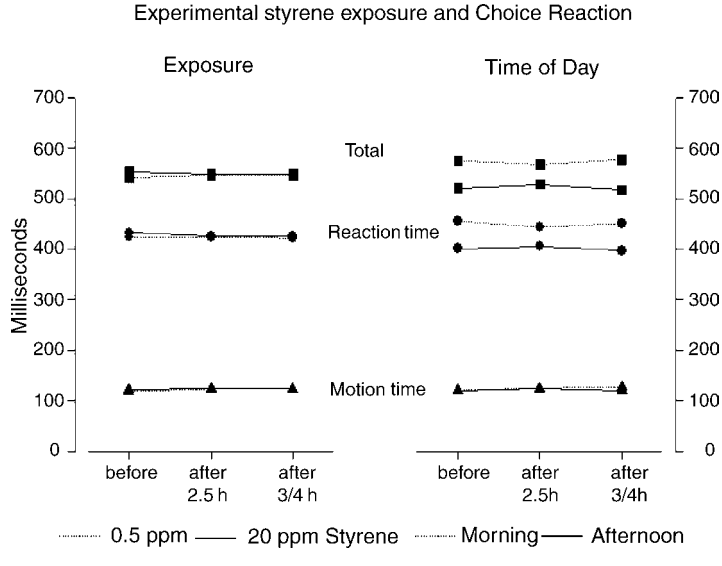


Figure 3.28 *Example for benchmarking of data. Results from a choice reaction test in experiments with two constant styrene exposures combined with two daytime conditions. The curves for exposure do not differ. [Reprinted from Seeber et al., Toxicol Lett.,* **151***, 183–192, Copyright (2004), with permission from Elsevier.]*

0.5 ppm and 20 ppm exposure are shown. In a special design the exposure conditions were combined with the time of day for operating the experiments. The exposure conditions do not provoke different reactions, whereas the time delay of about 6 h does do so significantly. This validates the conclusion that 20 ppm styrene does not prolong the reaction or movement time.

A typical memory test also originating from the Wechsler Intelligence scale is the *test Digit Span*. In the original paper–pencil version standardised series of digits are presented verbally and the response has to be given also verbally. The longest sequence of presented digits reproduced correctly – both forward and backward – is the score for digit span. In the computer administered version the sequence appears on the screen, the response is given on the keyboard. An adaptive mechanism regulates the presented sequences again forward and backward in order to find the individual optimum of the longest sequences for the score digit span.

Another often used memory test with higher portions of visual-spatial processing is the *Benton test,* again as a paper–pencil or computer administered test. More or less complex figures are presented and the subject has to reproduce them either by drawing or by marking the correct pattern out of similar patterns in a multiple choice response.

Measurements of *psychomotor functions* in humans are usually conducted with test batteries for manual dexterity. These include sub-tests for Steadiness (holding a contact pen without contacting the boundary of a hole), Line tracing (directing the contact pen in a pattern line without contacting the boundaries), Aiming (directing the contact pen quickly over contact plates), Tapping (touching of contact plates as soften as possible in a constant time), and Peg board (manipulating pegs in small ports). Speed and precision are measured. The tests are performed subsequently with both hands.

Symptom Measurement and Personality Characteristics Symptom measurements are essential for neurobehavioral studies in humans. Mostly self-administered questionnaires are used with items covering chronic and/or acute neurotoxic effects. For example, EUROQUEST is a common questionnaire especially for investigating solvent effects. Chronic neurological problems, psychosomatic symptoms, mood lability, memory and concentration problems, tiredness, and sleeping problems as well as acute effects of solvents like irritations and intoxications are listed. Besides these, personality characteristics like environmental sensitivity, anxiety trait, and feeling healthy are asked for in order to measure confounding factors. The experienced symptoms have to be scaled between 'seldom/never,' 'sometimes,' 'often,' and 'very often.'

Table 3.7 summarizes typical symptoms identified by questionnaires in 28 neurobehavioral studies on solvent effects. The questionnaires were different, thus the concordance of symptoms is limited.

Figure 3.29 demonstrates the association of symptoms both to data on exposure and to a personality trait, in particular (as an example) the anxiety trait. The symptoms depend predominantly on individual levels of anxiety in a study at low dioxin and furan exposures, whereas in the solvent-mixtures study they depend additively on real existing exposure and anxiety trait.

Sensory Functions Increasingly, sensory functions are being measured in neurobehavioral studies. Especially, colour discrimination and examinations are used in order to evaluate chronic solvent effects.

Table 3.7 *Symptoms of occupational solvent exposure reported in 28 studies using no uniform questionnaires. Significant effects refer to comparisons between exposed and nonexposed groups independently of the level of solvent exposure.*

Symptom	Number of investigations asking for the symptom	Number of significant effects shown
Sleepiness	10	4
Tiredness	10	7
Physical exhaustion	7	5
Mental exhaustion, passivity	12	6
Tension, intolerance of stress	14	5
Headaches	17	6
Nausea	9	3
Dizziness	14	7
Breathing difficulties (including coughing)	8	5
Irritation of eyes, nose, airways	10	5
Feeling of drunkenness, confusion	10	5

Impaired *colour vision* in the blue–yellow-discrimination – classified as acquired dischromatopsia – are predominantly measured by the Lanthony Panel D-15 desaturated test. The D-15d consists of 15 moveable, colored caps chosen to represent perceptually equal steps of hue and to form a natural hue circle. Subjects must arrange the caps in order according to colour. Scoring is done by the Color Confusion Index (CCI), indicating the distance of the individual placements from the perfect cap replacement. In studies on solvent exposures (toluene, styrene, and solvent mixtures) increased CCI related to elevated exposures could be shown.

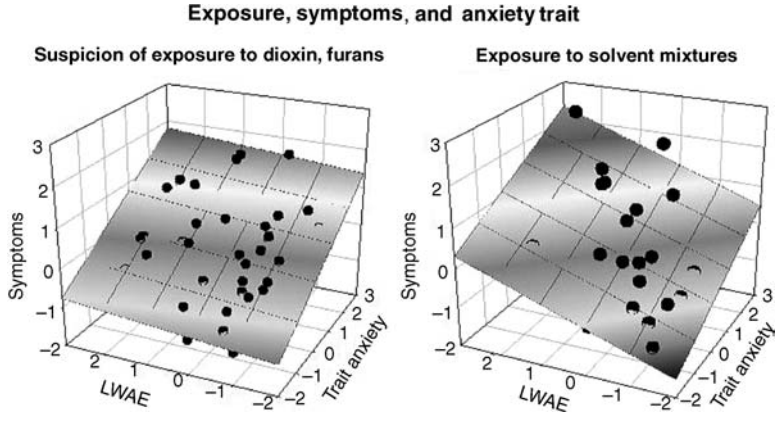


Figure 3.29 *Example for additive effects. Standardized symptom scores related to standardized lifetime-weighted average exposure and anxiety trait. Left: Results from persons assuming to be exposed to dioxins and furans without biological proof of exposure. Right: Results from really exposed workers in paint manufacturing. [Reprinted from Seeber et al.; Neuro Toxicology,* **21***, 677–684, Copyright (2000), by permission from Elsevier.]*

Measurements of *auditory thresholds*, especially of high frequency hearing loss, are part of epidemiological studies on solvent effects. Diminished auditory thresholds were shown repeatedly related to higher levels of solvent exposure and in interaction of solvent, especially toluene, and noise exposure.

Effects on *olfactory functions* are known for numerous chemicals. According to the regeneration cycle of the olfactory system, reversibility of changes of the olfactory threshold and discrimination can be expected in most occupational exposures. Also, measurements of *standing stability* under different conditions, under which the feedback system for the function is changed, are included in neurobehavioral studies.

Neurobehavioral Test Batteries Various recent neurobehavioral test batteries have further developed the traditional paper–pencil tests by using fully computerized systems to reduce diversity. A Neurobehavioral Core Test Battery was recommended by a WHO working group. In most cases, computer-assisted test batteries such as the Swedish Performance Evaluation System (SPES) and the Neurobehavioral Evaluation System (NES) have been used. In particular, the NES has been adapted to newer computer techniques and has been revised for specific demands in epidemiological research.

Table 3.8 lists examples of test applications using paper–pencil tests. The test principles are comparable with new computer-based approaches. Meantime, vocabulary tests are regularly included.

The advancements of the Neurobehavioral Evaluation System (NES) led to the broad application of this computer-administered test battery in recent studies on neurobeha-

Table 3.8 *Examples for methods in neurobehavioral test systems and their usage in neurotoxicological studies. Numbers indicate the application in test batteries analysed until 1990 and in studies analysed until 1994. The years of analyses until 1994 indicate the use of tests dominated by traditional paper–pencil tests as opposed to computer-administered tests.*

Tests	Number of inclusions in 16 test batteries	Number of applications in 210 studies
Cognitive functions		
Visuospatial design (Block Design)	11	31
Symbol-digit/symb.-dig. substitut.	12	51
Visual pattern processing		
Choice reaction		26
Trail making	3	13
Memory span forwards/backwards	13	37
Memory scanning		4
Visual retention (Benton)	8	14
Associate learning		13
Eye–hand coordination		4
Visuomotor coordination	6	19
Verbal reasoning		13
Continuous performance / attention	10	6
Executive functions		
Simple reaction time	9	39
Tracking tasks		3
Finger tapping/aiming	5	19
Vocabulary test ('pre-morbid intelligence')		10

vioral effects in humans. New developments in computer-based testing consider (a) standardized training procedures in order to compensate for missing computer experiences, (b) adaptive computer-based speech communication, (c) adaptive task difficulties optimized for an individually adjusted level.

Evaluation of Neurobehavioral Effects In order to evaluate the harmfulness of results of neurobehavioral studies, evaluation criteria were developed. They discriminate between 'substance-related' and 'adverse' effects. Substance-related effects are statistically significant associations between exposure and a neurobehavioral variable. However, statistical significance does not imply that the result is important for health and well-being. Adverse effects imply a significance for health and well-being because an intolerable loss of capability or sense of well-being is observed. Important criteria for the transition from 'substance-related' to 'adverse' are the extent, the number, the concordance, the severity, and the irreversibility of effects shown in different studies.

3.5.4 Neurobehavioral Effects in Humans

The toxic encephalopathy is a clinical disease pattern associated with chronic and high exposures to solvents, lead, and mercury. For the solvent-induced encephalopathy three degrees were defined that describe the transition from early effects, characterized as reversible slight impairments in well-being and in cognitive and executive functions, up to severe mental deficits and multiple complaints as the latest stage after long and intensive exposures.

Differences between the exposure-related effects of solvents, lead, and mercury are hard to document, whereas similarities are seen regarding effects on attention, memory, and psychomotor functions. For mercury, a dominance of impaired psychomotor functions could be shown.

Solvents

Neurobehavioral methods have been used to document neurotoxic effects due to occupational and environmental exposures to single solvents and solvent mixtures. For occupational solvent exposures a well defined disease pattern is *chronic toxic encephalopathy* (also 'painter's syndrome,' 'solvent-induced psycho-organic syndrome'). In its severest degree (Type 3) – recently very seldom observed – it is characterized as dementia with a marked global deterioration in intellect and memory, often accompanied by neurological signs or neuroradiological findings and is not reversible. The weakest and reversible Type 1 includes nonspecific symptoms, such as fatigue, memory impairment, difficulty in concentration, and loss of initiative without objective evidence of neuropsychiatric dysfunction. Between them, sustained personality or mood changes (Type 2a) and impairments of intellectual function (Type 2b) with questionable reversibility occur. Substances possibly inducing a chronic toxic encephalopathy are:

(a) the aliphatic hydrocarbons n-hexane and n-heptane;
(b) the aromatic hydrocarbons benzene, toluene, styrene, and xylene;
(c) the chlorinated aliphatic hydrocarbons dichloromethane, 1,1,1-trichloroethane, trichloroethylene, and tetrachloroethylene;

(d) the ketones 2-butanone, and 2-hexanone;
(e) the alcohols methanol, ethanol, and 2-methoxyethanol.

So far, substance specificity of the effects has not been accomplished, mostly because substance-specific neurobehavioral symptoms or performance deficits cannot be defined. Reasons for this may be because of (a) the unknown mechanisms that induce behavioral impairments, (b) the human intellectual and emotional capacity to compensate for slight specific cognitive or emotional impairments or (c) the methodology not being appropriate to identify very distinct types of effects.

Exemplarily, some additional neurobehavioral data is given only for the solvent *toluene*. For high occupational exposures ($\gg$ 100 ppm) clear clinical manifestations were described with lists of symptoms experienced during workdays (e.g., drunken feeling, headache, dizziness, etc.) and after 6 months of exposure (nervousness, loss of consciousness, forgetfulness, headache, etc.). Intentional inhalation of toluene vapors (misuse) has been associated with cognitive deficits in attention, memory, visuospatial and complex executive functions, motor disorders, cerebellar dysfunctions (ataxic gait, tremor) as well as in brain stem and cranial nerve functions. In such cases also hearing impairments and optic atrophy were reported.

The regional distribution of toluene in brains of rats and humans showed highest concentrations in regions with more white matter whereas the gray matter of the cerebral cortex and hippocampus had lower concentrations. That seems to reflect the low molecular weight, high lipid solubility, and lack of protein-binding capability of toluene.

In the neurobehavioral literature on toluene some hints of weak cognitive deficits below 40 ppm were published, but past exposures of the workers were higher by up to a factor of 5. Also, findings on deficits in color discrimination at this level and hearing loss at about 50 ppm were published. However, these data from cross-sectional studies were checked in a longitudinal study of 5 years' duration with repeated measurements. It was shown that current toluene exposures of 26 ppm and lifetime-weighted average exposures of 45 ppm did not induce neurobehavioral effects.

Lead

Lead is one of the well investigated neurotoxicants, known as a toxic agent for the cognitive development of children by environmental exposures as well as for the mental abilities of adults by environmental and occupational exposure.

Exposure sources with high occupational risk operations are smelting, welding, and cutting of lead-containing materials, spray painting, or scraping of lead paints. Moderate risks are associated with activities of lead miners, solderers, plumbers, type setters, cable makers, lead glass workers, and automotive repair personnel. Sources of environmental exposure to inorganic lead are air, soil, dust, and food. For organic lead compounds (tetraalkyl-, tetraethyl- and tetramethyl-lead) the emission into the atmosphere dominated before their ban as gasoline additives. At present the soil- and water-related exposures and those by aquatic organisms reaching the food chain are dominant.

Well known neurotoxic effects of lead are:

- Impaired development and function of oligodendrocytes;
- Abnormal myelin formation;
- Altered neurotransmitter release and receptor density;
- Abnormal myelin formation, dendritic branching pattern, and neurotrophic factor expression;
- Disruption of the blood–brain barrier and of thyroid hormone transport;
- Impaired neuropsychological functioning and lowered IQ.

Young pregnant women living in lead contaminated housing or who were lead-poisoned themselves as adolescents can pass lead on to their foetus. Strong correlations between maternal and umbilical cord blood lead levels prove the transfer of lead to the children. Also the lead level in the breast milk correlates with the maternal blood lead level. Pre- and peri-natal blood lead levels of neonates are amongst other factors (reduced body weight, circumference of the head, and length) associated with disturbances in cognitive development.

Although such disturbances are considered to be the most sensitive endpoint of reproduction toxicology there are results of prospective studies in independent cohorts with young children which do not provide a concordant pattern of results. These studies included consideration of confounders such as the maternal intellectual status, the socio-economic status of the family and housing conditions (e.g., scale 'Home Observation for Measurement of the Environment'). In some studies including about 1200 children no association could be shown by standardized test batteries for early cognitive and psychomotor development, whereas in other studies with 3300 children associations between maternally induced exposure data and impaired early development were shown up to the fourth year. With increasing age the postnatal individual exposure determines the strength of lead-induced mental impairments. Without lasting exposure other factors determine the further development, especially the educational conditions; earlier lead impacts can be compensated for.

Critical lead concentration in the pre- or peri-natal blood causing measurable effects is about 100 μg lead/l blood for effects over 2 years, whereas critical blood lead levels due to environmental exposure during later periods of infantile development are less than 100 μg lead/l blood. The overall analysis of all data indicates that increasing blood lead levels from 50 μg/l to 200 μg/l are associated with increasingly impaired functions. An increase of 100 μg lead/l blood is associated with a decrement of up to 3 points in full-scale IQ. A threshold of blood lead levels under which adverse health effects are to be expected could not be derived. A critical discussion of these findings emphasizes the variety of different factors influencing the intellectual development of children, but the congruence of conclusions from different studies is convincing.

Mercury

Toxic effects of exposures to inorganic mercury were already known over many centuries, whereas effects of organic mercury compounds have been known for only the last century. Organic mercury compounds are relevant after methylation of mercury by microorganisms in sediment, soil, and water, which enter the human food chain

through plants and fish. However the use of mercury in manufacturing products is diminishing, and prohibition of mercury-containing products has reduced the occurrence of mercury as a hazard in both the workplace and the environment.

Owing to different rates of absorption and of transfer through the blood–brain barrier the various mercury compounds have different neurotoxic potencies. Acute inhalation of elemental mercury induces inflammation of the upper respiratory tract along with general malaise, headache, vomiting, nausea, fever, and chills leading to the 'metal fume fever syndrome' after very high exposures. After some hours, neurological effects including tremor or delirium occur, which persist whereas other symptoms diminish.

Chronic exposure to elemental mercury induces symptoms of 'mercurialism' including fatigue, general weakness, loss of appetite, diarrhoea, mood changes, insomnia, and bilateral tremor. In severe cases progressive motor neuropathy with fasciculations of muscle and atrophy occurs, associated with distal axonal degeneration and denervation atrophy in muscles.

Chronic exposure to organic mercury in contaminated fish and grain treated with methylmercury led to decreased intellectual abilities, loss of concentration and memory, emotional changes, and depression. Additionally, cerebellar signs of ataxia and stumbling gait, incoordination, paresthesias, sensory loss, and other sensory impairments were noted. Prenatal exposure to methylmercury can result in brain damage to the children. The neurophysiological have been verified by electroencephalography (EEG), somatosensory evoked potentials (SSEP), nerve conduction velocities (NCV), and electromyography (EMG).

In biological monitoring studies 100 μg Hg/g creatinine seems to generate no neuropsychological deficits. Later studies, partly in dental workers in contact with dental amalgam, revealed significant effects in tests measuring attention, perception, reasoning, motor speed, and motor precision at exposure levels of about 25 μg Hg/g creatinine. In other studies effects only in motor functions and not in cognitive functions at exposure levels of about 60 μg Hg/g creatinine were seen. This was supported by the meta-analysis of the available studies. Motor performance tests showed the strongest mercury effects. Less but still significantly affected was the performance shown in memory tests, while the performance in tests for attention was hardly affected.

3.5.5 Summary

Variations of human behavior can be observed and measured with standardized methods, and deviations from expected or predicted behavior patterns can be described. Sources inducing deviations from 'normal' are multifarious, external and internal from person to person. In particular, education, age, the status of the diurnal rhythm, physical conditions during measurement, as well as motivational or emotional states can induce behavioral variations to a measurable extent. In evaluating the neurotoxicity of a chemical the exposure-related variations of behavior have to be compared with variations induced by other conditions also affecting human behavior. In this way the inhalation or another form of exposure to a substance can be seen as an assured source inducing significant variations of behavior. Following this approach, the objective of studies in behavioral toxicology – especially in human neurobehavioral approaches – consists not only of data showing any relationship between parameters of exposures and parameters for behavior

measurements. The objective should be also to collect data showing that confounding conditions of behavioral variations have been controlled statistically or by the study design. Before rash toxicological interpretation of test results are made the methodological background of the data should always be analysed.

References

Seeber, A., Sietmann, B., and Zupanic, M. In Search of Dose-Response Relationships of Solvent Mixtures to Neurobehavioral Effects in Paint Manufacturing and Painters. *Food Chem. Toxicol*., **34**, 1113–1120 (1996).

Seeber, A., Blaszkewicz, M., Golka, K., Hallier, E., Kiesswetter, E., Schaper, M., and van Thriel, C. Neurobehavioral Effects of Experimental Exposures to Low Levels of Styrene. *Toxicol. Lett*., **151**, 183–192 (2004).

Seeber, A., Demes, P., Golka, K., Kiesswetter, E., Schäper, M., van Thriel, C., and Zupanic, M. Subjective Symptoms due to Solvent Mixtures, Dioxin and Toluene: Impact of Exposure versus Personality Factors. *Neuro Toxicology*, **21**, 677–684 (2000).

For Further Reading

Chang, L. W. and Slikker, W. (Eds.).: Neurotoxicology. Approaches and Methods. Part II. Academic Press, San Diego, New York (1995).

Chouanière, D., Cassitto, M.G., Spurgeon, A., Verdier, A. and Gilioli, R. An International Questionnaire to Explore Neurotoxic Symptoms (EUROQUEST). *Environ Res*., **73**, 70–72 (1997).

Feldman, R.G. Occupational and Environmental Neurotoxicology. Lippincott-Raven, Philadelphia, New York (1999).

Iregren, A., Gamberale, F. and Kjellberg, A. SPES: A Psychological Test System to Diagnose Environmental Hazards. *Neurotoxicol. Teratol.*, **18**, 485–491 (1996).

Letz, R. NES3 User's Manual. Atlanta (MA): Neurobehavioral Systems, Inc. Atlanta, GA (2000).

Lezak, M.D. Neuropsychological Assessment, 3rd edition, HDI Publishers, Oxford (1995).

3.6 The Skin

Tibor A. Rozman, Myriam Straube and Karl K. Rozman

The skin and the anterior segment of the eye, with their appendages, represent the outermost boundary of an organism. The surface area of the skin is approximately 1.5–2 m^2 and its weight is 3.5–10 kg depending on body weight and size. The skin and the anterior segment of the eye have developed characteristics that enable them to endure direct exposure to the external environment. In this chapter the anterior part of the eye and its toxicology will be dealt with only to the extent necessary to understand its function as a boundary to the surrounding environment as effected by xenobiotics.

3.6.1 Structure

Skin

The skin consists of three major components: epidermis, dermis, and subcutis.

The epidermis is a stratified squamous epithelium. Its external surface hornifies and forms the stratum corneum (thickness 5–20 microns). Depending on location, age, and gender the thickness of the epidermis varies between 80 and 160 microns. Up to four different cell layers can be distinguished depending on the body area. The cell layer above the basement membrane separating epidermis and dermis is the stratum germinativum where cell proliferation takes place. As the epidermal cells migrate to the surface, forming the stratum granulosum and stratum lucidum, they undergo morphological and physicochemical changes, step by step losing their cell structure. In the outermost part of the stratum corneum, cell structure is no longer discernible and the only histologically distinguishable feature is the desquamating stratum disjunctum. The turn-over time for this process is 4 weeks. Other cell types of the epidermis are melanocytes, Langerhans cells, Merkel cells, and some lymphocytes mostly located in the basal cell layer. The epidermis also contains some nerve endings but no vasculature.

The thickness of the dermis, which is approximately 1–3 mm, cannot be measured accurately because it is contiguous with the subcutaneous tissue, which varies with anatomical site. The dermis consists of an upper papillary and a lower reticular layer not clearly separated. Common cell types occurring in the dermis are histiocytes (active macrophages), mast cells, and fibroblasts producing copious amounts of collagenous and elastic fibers. This structure is embedded in a gelatinous matrix, forming a thick network that provides firmness to the skin. The dermis also possesses a rich vasculature with a high rate of blood flow. A cross-section of the epidermis and dermis is shown in Figure 3.30.

Appendages of the skin include hair follicles, sebaceous glands, which secrete oils, apocrine and eccrine glands, which secrete sweat, and hair and nails extending deep into the dermis.

The subcutis consists of loose connective tissue in between lobes of fat cells, sometimes forming a continuous layer of fatty tissue. The septa of the fat islets bring the nerve and blood supply to the dermis. Some collagen and elastic fibers also permeate into the reticular layer.

Anterior Segment of the Eye

The cornea and adjacent parts of the sclera form the external barrier that protects internal ocular structures.

The cornea is a transparent and avascular tissue with an anterior nonkeratinizing epithelial cell layer. Below the cornea lies the stroma, which is also completely translucent and is composed of water, collagen, and glycosaminoglycans. At the inner surface an endothelial cell layer can be found. Descemet's membrane separates this endothelial cell layer from the substantia propria.

Anterior and posterior chambers containing the transparent aqueous and vitreous humor, respectively, are separated by the iris and delimited towards deeper eye structures

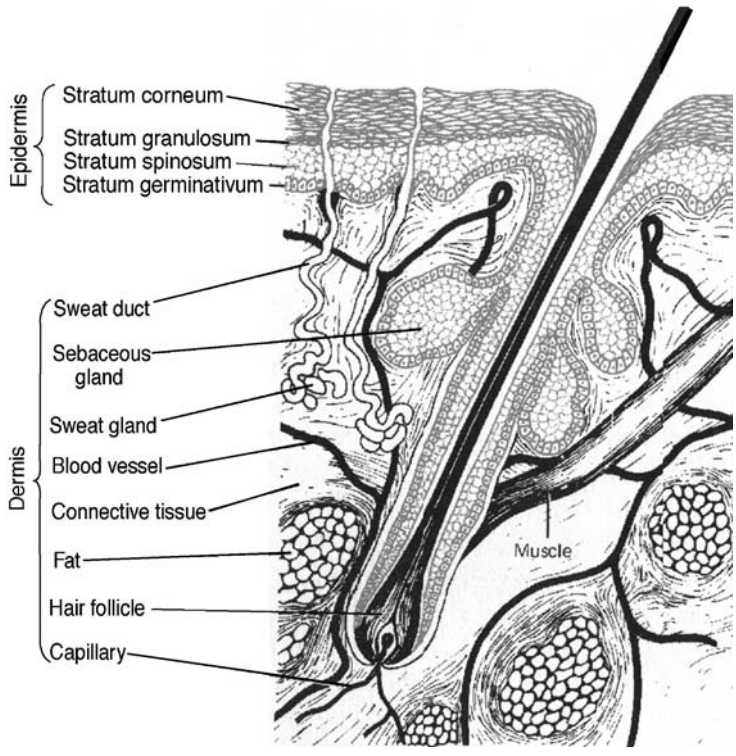


Figure 3.30 *Cross-section of the skin. [Reprinted from Cassarett and Doull. Copyright (2001), with permission from McGraw-Hill.]*

by the lens and ciliary body. The iris itself is a highly vascularized tissue consisting mainly of connective tissue, containing a large number of pigmented cells.

3.6.2 Function

Skin

The skin protects the organism from mechanical, chemical, and immunological insults as well as against dehydration, heat, and excessive light exposure.

The surface of the skin (i.e. the stratum corneum) is covered by a lipid layer produced by the sebaceous glands. Its pH of 5.7 is achieved by the secretion of sweat glands and other acidic components such as free fatty acids produced by resident flora. This combination is particularly suitable as defense against attacks by bacteria but also against some xenobiotica. Langerhans cells provide support for immune responses, whereas melanocytes can diminish potential light-induced damage by producing UV-absorbing melanin.

Thermoregulation by sweating and diffusion of water through skin layers to the outside (perspiratio insensibilis) are both life-supporting functions of the skin. These functions can be carried out effectively only because dermis and subcutis are perfused by quantities

of blood 20 times higher than needed to provide nutrients. Under the influence of light, 7-dehydrocholesterol is metabolized to cholecalciferol in the skin, which is the precursor of the active metabolites of vitamin D3 in both liver and kidney required for calcium and phosphate absorption, deposition, and mobilization. Another important function of the skin is the communication with the external world. This occurs through organelles of different sensory qualities such as sensation of heat, cold, touch, and pain.

Subcutaneous fat serves as insulation against overt loss of heat. At the same time, it is also the site of deposition and storage of persistent, lipophilic xenobiotics.

Anterior Segment of the Eye

Accessory glands on the rim and the posterior surface of the lids together with the lacrimal glands produce the lacrimal film.

Cornea as well as bulbar and palpebral conjunctiva are bathed in a continuously produced tear film consisting of three domains: a hydrophobic lipid layer, an aqueous layer, and a mucoid layer, the latter acting as interface between the hydrophilic aqueous layer and the hydrophobic epithelial cell membranes of the cornea. Eyelids and lacrimal glands are very efficient means of defense to prevent foreign bodies from injuring the cornea. If xenobiotics splash into the eye, lacrimation immediately dilutes and rinses the corneal sac and the cornea.

Dermal Absorption

Penetration through the skin is governed primarily by passive diffusion as defined by Fick's law. Primary determinants are: the thickness of the stratum corneum, its moisture content, and the temperature of the skin as well as molecular weight (MW) and lipid solubility of the compound in question; in addition, properties of the vehicle are also important. Generally, amphiphilic molecules with low MW penetrate the skin most readily, whereas for large hydrophilic molecules it is almost completely impenetrable.

The very thin stratum corneum, consisting of only a few layers of cornified, flattened, dead cells, is the actual rate-determining barrier in transdermal penetration. The fairly homogeneous horny layer consists mainly of keratin and sphyngolipids, in which hydrophilic and hydrophobic layers alternate. This unique structure and composition is ideally suited to prevent polar, ionized and very apolar substances from entering the organism. The horny layer and subsequently the epidermis will be most readily penetrated by compounds of mixed properties, i.e. having some polar as well as apolar characteristics, e.g. many suitably substituted corticosteroids.

MW is another important property, which is inversely related to penetration. Chemicals having MWs above about 6000–7000 are generally unable to penetrate the skin. Once a molecule crosses the horny layer, diffusion takes place at a higher rate. The situation is comparable to a dam and a reservoir behind it, the stratum corneum playing the role of the reservoir and the epidermis the outflow.

In addition to the properties of drugs and other chemicals, the condition of the skin itself is also of importance for transdermal penetration. Rate-limiting conditions are skin temperature

and degree of hydration. The degree of hydration is of particular importance for enhanced penetration. Occlusive bandages greatly increase the moisture content of the stratum corneum, thereby enhancing transdermal penetration, often by a factor of 10 or more.

Partitioning of compounds and their vehicles (or solvents) into lipids of the horny layer is a critical factor determining diffusion through the stratum corneum and epidermis (octanol/water partitioning ratio, K_{ow}). This effect is often used therapeutically for enhanced and controlled drug delivery through the skin (e.g., when propylene glycol is the vehicle). Some detergents and solvents, e.g. dimethyl sulfoxide (DMSO) and dimethylformamide (DMF), render the horny layer more porous and penetrable by dissolving and removing sphyngolipids from the stratum corneum.

The importance of the transdermal pathway as a potential route of entrance into an organism to cause toxicity is exemplified by monochloroacetic acid. It is an intermediate in the synthesis of numerous detergents and many other chemicals used in large quantities by industry. Monochloroacetic acid can cause lethal poisoning if as little as 6% of the skin is exposed for a short period of time.

Ocular Absorption

The lacrimal film and the different layers of the anterior segment of the eye are equally efficient barriers against direct entry of xenobiotics to the inner structures of the eye. The stratum corneum is equally protective for the skin.

If penetration through the cornea does occur, clearance into blood can occur via the aqueous flow through the canal of Schlemm. Molecules that can penetrate the blood/central nervous system (CNS) barrier may be secreted in the process of aqueous humor production. Systemic absorption can also occur through the conjunctiva, in the nasolacrimal duct, and further down in the nose and epipharynx without exhibiting a first-pass effect in the liver. Absorbed compounds can be distributed via blood, partitioning into cells or interstitial fluid, or sequestered in organs. For retinal toxicity to occur, systemic absorption is much more important than penetration through the cornea, aqueous humor, lens, ciliary body, or vitreous body.

Excretory Function of the Skin

Cutaneous appendages (hair follicles, sebaceous glands, eccrine and apocrine sweat glands, hair, and nails) serve primarily excretory functions for compounds that cannot be eliminated from the organism otherwise.

The excretory function of sweat glands for inorganic chemicals is well established. Sebaceous glands can reduce the body burden of lipophilic xenobiotics that are not otherwise amenable to biodegradation. Chemicals can also translocate into the epidermis by passive diffusion and are externalized in the process of cornification and desquamation. Other appendages like hair and nails may also serve as sites of accumulation and storage. For example, there are metals that may be deposited in hair or nails, in some instances leading to topical toxicity in the form of loss of hair or nails. Some toxicants are

deposited and stored in the dermis. The hair follicle, unlike the appendages cited above, provides an avenue for transdermal absorption. Thus, experimental animals having dense fur absorb material through the skin up to 3–4 times more rapidly than do corresponding nude animals, because the epidermal cells in the hair follicle do not cornify, i.e. do not generate a horny layer. However, it is not possible to accurately predict the rate at which dermal absorption will take place in furry animals, because of the large number of hair follicles, and as a result we cannot extrapolate from these data to predict dermal absorption in nude animals or in humans.

Metabolism in the Skin

The skin contains both Phase I and Phase II enzymes active in the metabolism of xenobiotic compounds.

The skin contains most enzymes needed for the metabolism of glucose, lipids, and proteins, as well as those for the synthesis of glycogen, lipids, and proteins. The epidermis appears to be the most active in these processes. The papilla of the hair follicle is one of the most metabolically active tissues in the body. The skin is a constantly regenerating tissue with a turn-over rate of about 4 weeks. It requires a constant energy supply provided by adenosine triphosphate (ATP) derived from different sources.

Xenobiotic-metabolizing enzyme activities are relatively low in the skin as compared with other major organs, e.g. liver. Under normal conditions approximately 2% of total enzymatic activity of the body can be found in skin. However, it has been shown that enzymatic activity in the fully induced state can be as high as 20% or more of the whole body, e. g. after exposure to TCDD, a well known inducer of various cytochrome P (CYP) isozymes. Crude coal tar is another example of an inducer of the CYP system, which is widely used in dermatological therapy. Phase II enzyme activity has been demonstrated in the epidermis. The presence of glucuronidases has been documented in the stratum corneum as well. Hormones such as testosterone, estrone, and drugs such as corticosteroids are metabolized in the epidermis and only approximately 2/3 of a dose applied topically reaches the systemic circulation unchanged. The structures of the anterior segment of the eye contain also all major Phase I and Phase II xenobiotic-biotransforming enzymes.

Typical examples of diseases caused by enzyme activity or lack of it in the skin are male-pattern baldness and xeroderma pigmentosum. In the first, increased 5-alpha-reductase production in keratinocytes of genetically predisposed individuals leads to enhanced transformation of testosterone into 4,5-dihydrotestosterone (DHT) and as a consequence hair loss. Enzyme deficiency in xeroderma pigmentosum will be dealt with in Section 3.6.3, Sub-section ‘Skin Tumors.’

3.6.3 Toxicology of the Skin and the Anterior Segment of the Eye

Owing to its direct exposure to the environment, the skin is a readily available target for the toxic effects of xenobiotic chemicals, ultraviolet light, and extremes of temperature.

In most industrialized countries half or more of reported occupational diseases are related to the skin. Direct injury in the form of burns or corrosions can be caused by alkali and acids as well as oxidizing agents. Most xenobiotics, however, are less aggressive, but penetration through the stratum corneum is a prerequisite for the exertion of toxicity.

Acidic and Alkaline Corrosion

Alkalies, acids, some salts of metals, oxidizing agents, and different organic compounds are in many instances aggressive enough to cause cell death and even complete destruction of skin and/or cornea/conjunctivae.

Injuries at times are severe enough that 'restitutio ad integrum' (complete recovery) is not possible. In such cases epithelial and other cells and cell layers will be replaced by scar formation. This is particularly damaging to the anterior segment of the eye, especially to the cornea. In severe cases the cornea becomes opaque due to vascularization and completely impenetrable for light. Acids precipitate intra- and extra-cellular proteins, creating a barrier that prevents further penetration into deeper structures. Concentrated sulfuric acid denatures protein through dehydration and the development of heat. Alkalies have a different mode of action with even more damaging consequences. They cause so-called colliquation necroses by hydrolysing peptide bonds of proteins, which can result in long-lasting and deepening lysis of tissues. Therefore, alkaline injury to skin or eye may look less damaging at first than acid based injury but delayed effects may lead to the development of a deep, poorly healing ulcer. Only immediate irrigation with copious amounts of water may diminish or prevent further destruction of the respective organs. Table 3.9 lists some chemicals that are corrosive to skin and eye.

Irritation and Contact Dermatitis

Acute primary irritation is defined as a reversible inflammatory response of normal skin or the eye to a single topical injury by an agent of chemical, biological, or physical origin, provided the mechanism is not due to an immunological response.

Contact of an irritant with the skin or the eye resulting in reversible erythema and edema due to extravasation is termed acute primary irritation. Patch testing is a common diagnostic procedure to reveal the causative agent in both contact and allergic contact

Table 3.9 *Compounds with corrosive effects on skin and eye.*

Metal salts	$HgCl_2$, $K_2 Cr_2 O_7$
Acids	HCl, H_2SO_4, chloroacetic acid, formic acid, oxalic acid
Alkalies	NaOH, KOH, NH_4OH
Phenols	Cresol, carbolic acid
Solvents	Trichloroethylene

dermatitis by applying the suspected compound or, if unknown, a set of common allergens to the back of a patient.

In the evaluation of a potential irritant the bioassay usually involves an albino rabbit in which a comparison is made of the effects of applying the material to either intact or abraded skin. The skin is read after 24 and 72 h and the degree of erythema and edema is recorded on a scale of 1 to 4. Testing of cumulative irritation (repeated application) may reveal the irritant potential of compounds not detectable in an acute test.

When measuring ocular irritancy in the eye of the albino rabbit, the test compound is administered to one eye of the animal, with the other eye serving as control. Scores at 1, 24, 48, and 72 h are assigned on the basis of damage inflicted upon the cornea, conjunctivae, and iris. The following criteria are applied: degree of opacity of the cornea, hyperemia of conjunctivae and iris, light reaction of the iris, and chemosis (swelling) of the conjunctivae.

Acute contact dermatitis arises from exposure to a primary irritant, such as those shown in Table 3.10. Old or damaged skin is particularly prone to contact irritation, which is characterized by intense redness, tenderness, heat, and swelling. In severe cases, oozing, blistering, and even necrosis may be observed. If exposure becomes chronic, lichenification (thickened skin), itching, and scaling might become the more predominant symptoms. At times irritant contact dermatitis might not be distinguishable from allergic contact dermatitis.

Allergic Contact Dermatitis

In cases of allergic contact dermatitis, also termed skin sensitization, exposure to a few molecules can produce an intense skin reaction. The underlying pathomechanism is a cell-mediated type-IV hypersensitivity disorder.

The best known example for this skin condition is chronic eczema in masons working with cement treated with potassium dichromate.

Allergic contact dermatitis is initiated by prior exposure to a small reactive molecule, called the hapten, which penetrates the stratum corneum and binds covalently to a carrier protein. The complex, termed the complete antigen, may then be incorporated via pinocytosis in the stratum spinosum into dendritic phagocytes called Langerhans cells for processing. Langerhans cells subsequently migrate to regional lymph nodes to present the processed antigen to the immune system. A complex interplay of cellular and humoral factors leads to the formation of a large number of specifically sensitized T-lymphocytes. Renewed contact with the allergen at any site of the skin triggers an allergic reaction.

Table 3.10 *A few common contact irritants.*

Poison ivy, poison oak
Detergents
Solvents
Adhesives
Soaps
Bleaches
Many other chemicals

There is ample evidence for the role of genetic predisposition in contact sensitization. Moreover, some haptens must be metabolically activated by Phase I and/or Phase II enzymes of the epidermis before forming the hapten–carrier complex. Symptoms of acute and chronic allergic contact dermatitis resemble those of contact dermatitis, the minute amount of substance needed to elicit the skin reaction being the only distinguishing feature. Allergic contact dermatitis can be provoked by a very large number of agents of chemical, plant, or animal origin (Table 3.11).

Table 3.11 displays the most potent sensitizers. However, exposure to thousands of other compounds may also lead to an allergic contact dermatitis. In many instances the sensitized immune system reacts to substances that are similar to, but not identical with, the contact allergen. Such cross-reactivity is called cross-sensitivity.

Toxic Epidermal Necrolysis (Lyell's Syndrome)

Toxic epidermal necrolysis (TEN) is a severe, in some instances life-threatening, usually drug-induced immunological response resulting in detachment and sloughing of significant amounts of skin.

TEN is characterized by severe bullous eruptions on skin and mucous membranes, fever, malaise, conjunctivitis, and general redness of the entire skin. Detachment of epidermis and mucous membranes from adjacent tissues causes skin to slough off in large pieces, leading to infections and causing life-threatening loss of body fluids. Mortality may range from 15–40%. Most frequently, it occurs after drug treatment, when, in genetically susceptible individuals, the drugs or their metabolites accumulate in skin, resulting in an immune response resembling graft-versus-host disease. It has been proposed that T-lymphocytes and macrophages induce an inflammatory response leading

Table 3.11 *Groups of compounds that frequently give rise to allergic contact dermatitis.*

Antiseptics and preservatives	Parabens
	Formalin
	Mercury compounds
	Chlorohexidine
	Dichlorophen
Dyes	Azo compounds
	p-Phenyleneiamine
Metal salts of	Nickel
	Chromium
	Cobalt
	Mercury
Antibiotics	Bacitracin
	Sulfonamides
	Aminoglycosides
Other therapeutic agents	Phenothiazines
	Benzocaine
	Idoxuridine
	Fluorouracil

to significant apoptosis of epidermal cells. The disease seems to be closely related to severe erythema multiforme and Stevens–Johnson syndrome.

Phototoxicity

Phototoxic reactions resemble severe sunburn (acute dermatitis). The distinguishing feature is a very rapid appearance of redness and blisters, sometimes within minutes of exposure to light.

The most frequently involved wavelengths of light in these disorders is UVA (320–400 nm) though UVB (290–320 nm) in some cases may also be effective. There are endogenous and exogenous chemicals listed in Table 3.12 that absorb UV light readily and cause phototoxic responses in the skin.

There are two mechanisms known to initiate phototoxicity. Some compounds, like chlorpromazine and tetracyclines, are excited to a higher energy state by the aborption of light and form cytotoxic free radicals, which can cause cell death. It is possible to take advantage of the light-activation process for therapeutic purposes. For example, one can administer porphyrins, which can accumulate in neoplasms. When the tissue is irradiated by UV light of optimized wavelength the activation of the porphyrins results in tumor cell death. Another mechanism of action in phototoxicity is characteristic of psoralens. After topical or systemic absorption, they intercalate with DNA, resulting in a covalently bound adduct between psoralens and pyrimidine bases. The resulting reaction products inhibit synthesis and repair of DNA, thereby diminishing cell proliferation Thus, administration of 8-methoxypsoralen, followed by exposure to UVA light, can be used in the treatment of hyperproliferative conditions like psoriasis, eczemas, and cutaneous T-cell lymphomas.

Photoallergy

In contrast to phototoxicity, photoallergy is a true allergic reaction in that minute quantities of a photoallergen cause a type-IV delayed hypersensitivity reaction.

Whereas in phototoxicity the activation of chemicals in skin cells via the absorption of UV light leads to reactive metabolites that cause cell death, in photoallergy the active metabolite acts as a hapten which reacts with cell proteins to form photoantigens and

Table 3.12 *Phototoxic agents.*

Psoralens
Polycyclic aromatic hydrocarbons (PAHs)
Tetracyclines
Sulfonamides
Chlorpromazine
NSAIDs
Porphyrins

Table 3.13 *A few typical photoallergens.*

Chlorhexidine
Hexachlorophene
Chlorpromazine
Promethazine
6-Methylcoumarin
Musk ambrette
Salicylates
Sulfanilamides
Thiourea
Triclosan
Topical antibiotics

generate allergic responses much like allergic contact dermatitis. Systemic treatment may lead to phototoxicicty and in some instances also to photoallergy but photoallergic agents usually act following direct application to the skin and exposure to UVA light. The rate of onset of photoallergy is slower than phototoxicity because it is a multistep process. Some typical photoallergens are listed in Table 3.13.

The most predictive animal model for photoallergy is the guinea pig. To distinguish between a phototoxic and a photoallergic response, testing is done on the backs of guinea pigs or human subjects by applying the test compounds in two symmetric rows. One side is uncovered 24 h later and is than irradiated by UV light. Reading is done immediately after irradiation as well as 24, 48, and 72 h later. Allergic reaction on the irradiated side only is proof of true photoallergy.

Urticaria

The appearance of urticaria is comparable to a wheal or hive. It is a more or less circular, red, spongy lesion caused by hyperemia and local edema. The underlying pathomechanism can be immune- or nonimmune-mediated.

Allergic urticaria is a type-I immunoglobulin E (IgE)-mediated hypersensitivity disorder. The severity may vary considerably from a few hives with minimal redness and edema to severe, generalized forms. Other type-I allergic diseases are rhinoconjunctivitis, asthma bronchiale, and insect bites with or without urticaria. Anaphylaxis is the most severe and life-threatening condition attributed to a type-I IgE-mediated hypersensitivity reaction. It has been described occasionally in the context of exposure to latex. Specific, e.g. type-I, hypersensitivity responses due to prior sensitization may result from exposure to a great number of chemicals, foodstuffs, and/or their components, fragrances, and antibiotics. Of particular importance is contact urticaria to latex. Millions of healthcare workers are exposed occupationally every day to latex-containing gloves with sensitization as a frequent consequence.

A common feature of nonallergic urticaria is the release of histamine and vasoactive peptides from mast cells as a result of a large number of physical, mechanical, and chemical causes. The urticarial lesion develops usually within minutes, accompanied by itching, tingling, and burning of varying intensity. Nonallergic causes of urticaria include such diverse

compounds as aspirin, azo dyes, benzoates, and toxins of plant or animal origin. Quite common physical causes are cold or warm temperatures as in cholinergic urticaria.

Loss of Hair

The hair follicle is a metabolically highly active tissue, which can produce up to 100 ft of hair per day. Its proliferative activity is matched only by the bone marrow and some parts of the gastrointestinal tract.

Every hair follicle develops through 3 stages before maturation and shedding occur. These are growth phases collectively called anagen (the germinative phase), catagen, and telogen. Hair loss occurs following the use of chemotherapeutic drugs such as antimetabolites or alkylating agents because proliferation is inhibited by these compounds. A large number of unrelated drugs and other chemicals may impair the mitotic activity of the hair matrix and cause its reduction or complete abolishment. Hair loss follows 2 weeks later and new growth does not commence for 2–4 months after cessation of treatment or other forms of exposure. Hair loss affects rapidly growing areas (scalp and beard) more than those of low growth rate (eyebrows). Some chemicals causing alopecia (hair loss) are listed in Table 3.14.

Accidental poisoning with the rat poison thallium sulfate may occur in households and lead to complete depilation. Large-scale poisonings in the industrial world resulting in hair loss have not been reported. However, many compounds, most of which are drugs, are suspected of causing loss of hair, usually without conclusive proof. One of the reasons may be that taking reliable hair counts is quite difficult because damaged hair broken off at the surface may be mistaken for shedding instead of toxicity to the matrix.

Chloracne (Halogenacne)

The label chloracne originates from the first observation of xenobiotic-induced acne to octachloronaphthalene more than 100 years ago. The term halogenacne would be more appropriate since bromine- and also to some extent iodine-containing compounds have been recognized as being comedogenic.

The basic lesion of chloracne is the comedo(ne) consisting of a dilated hair follicle containing surplus amounts of squamae, sebum, and bacteria (*Corynebacterium acnes*). If open, the surface is usually dark due to pigment containing cell debris. If closed, it has a

Table 3.14 *Some agents causing hair loss.*

Antimetabolites
Alkylating agents
Alkaloids like colchicine
Dixyrazine
Oral contraceptives
Anticoagulants
Thallium and its salts

Table 3.15 *Some of the most potent chloracne-inducing agents.*

Halogenated PAHs
Tar products
Lubricating oils
Plant oils in cosmetics

more yellowish appearance. Resident anaerobic bacteria present in everybody's skin find in such an environment plenty of nourishment and start multiplying.

If outflow is blocked, pressure builds up until the blocked hair follicle bursts and empties its content into the surrounding dermis. *Propionibacteria acnes* decomposes sebum into free fatty acids, which are highly proinflammatory. The result is an inflammatory lesion of various intensity, ranging from a red infiltrated papule to pustules. In severe cases cysts and tunnel systems may be formed in the dermis. Face, neck, and trunk are sites of high proclivity. In severe poisoning, the whole body surface may be affected. Table 3.15 lists some compounds capable of inducing chloracne.

Polycyclic halogenated aromatics such as dioxins, biphenyls, dibenzofurans, and azobenzenes usually have long half-lives and tend to accumulate upon chronic exposure. Therefore, typical acneiform lesions may persist as long as 30 years after their first appearance. The exact mechanism of action is not known but induction of hyperkeratinization and probably excessive formation of sebum in the hair follicle is a common feature of all cases of chloracne.

In systemic poisoning with halogenated aromatic compounds severe conjunctivitis and hyperkeratosis in the Meibomian glands, which results in squamous cysts, may accompany the clinical picture. Topical agents with shorter half-lives without tendency towards accumulation cause less severe and less persistent lesions, with clearance (remission) occurring as early as within several weeks.

Bromism is another skin disease, in which patients develop a skin condition called 'bromoderma' characterized by cuneiform papulous eruptions in the face and on the hands. Until about 50 years ago it was a well known disease entity because of the large number of bromide-containing drugs used in psychiatry. Recently, many cases of intoxication have been reported in workers exposed to solvents like n-propyl bromide and isopropyl bromide with manifestations of neurotoxicity and skin lesions. It was not clear how a compound with a half-life of 20 min to 1 h could cause such severe toxicity until it was determined that bromide was the proximate toxicant, which has a half-life of 12 days, having been liberated via a glutathione-mediated reaction. It is interesting to note that neurotoxicity and dermal toxicity occur usually together, most likely because both tissues are of ectodermal origin.

Disturbances of Pigmentation

Hyper- and hypo-pigmentation can be a specific or nonspecific response of the skin to chemical, physical, and mechanical injury.

Both endogenous and exogenous causes may change the pigmentation of the skin. Table 3.16 shows typical agents causing hyper- or hypo-pigmentation.[a]

Table 3.16 *Factors in hyper- and hypo-pigmentation.*

I. Hyperpigmentation due to endogenous causes
Ultraviolet light exposure
Post-inflammatory changes (melanin and/or hemosiderin depositions)
Hypoadrenalism
Internal malignancy (acanthosis nigricans)
Primary acquired melanosis (conjunctiva)
Hyperpigmentation due to chemicals
Coal tar
Anthracene
Picric acid
Psoralens
Hydroquinone (high dose)
Heavy metal salts of (Hg, Ag, Bi, Pb, Au, As)
Beta-carotene
Canthaxanthin (eye/skin)
Hyperpigmentation due to drugs
Chloroquine
Amiodarone
Bleomycin
Zidovudine
Minocycline
II. Hypopigmentation due to endogenous causes
Postinflammatory loss of pigment
Vitiligo
Hypopigmentation due to chemicals (leukoderma)
Hydroquinone (low dose)
Monobenzyl, monoethyl, and monomethyl ethers of hydroquinone
p-(Butyl)phenols
Mercaptoamines
Phenolic germicides
p-(Butyl)catechols
Butylated hydroxytoluene

[a]Modified from Table 19-9 in Casarett and Doull, 2001.

Endogenous hyperpigmentation is mostly due to increased melanin production and deposition in the basal cell layer or in tumor cells of epithelial origin. The formation and deposition of hemosiderin in the upper dermis or in poorly healing wounds is another mechanism of generation of hyperpigmentation.

Chemically induced hyperpigmentation usually occurs as a result of accumulation of deposition of particles in different cellular elements and appendages of the skin or in mucous membranes. The epidermis most often remains unaffected. The anatomical site of deposition varies from compound to compound. Deposition of lead or mercury results in grayish-black or bluish discoloration of the gingiva. Silver depositions confer a bluish discoloration upon the skin and conjunctivae (argyrosis) in sun-exposed areas. Interestingly, metal salt deposition in skin is often accompanied by stimulation of melanin production.

Loss of pigmentation in vitiligo is a disturbance of melanin production due to genetically determined enzyme defects of the tyrosine pathway. Therefore, it is not surprising that analogs of tyrosine (Table 3.16) are the most common compounds to cause depigmentation of the skin leading to the well known disease entity termed leukoderma. The most potent are substances with an alkyl group in the *para*-position, e.g. the monobenzyl ether of hydroquinone (MBEH), or *p-tert*-butylcatechol.

MBEH was the very first compound 50 years ago to be recognized as the cause of leukoderma, a skin condition characterized by complete loss of pigmentation at the site of contact. Another common property of these compounds is their ability to form semiquinone radicals and, thus, to induce lipid peroxidation.

Skin Tumors

Most of the benign or malignant neoplasms of skin originate in the epidermis. Benign skin tumors such as papillomas, warts, fibromas, or hemangiomas are not life-threatening.

In skin carcinogenesis, three types of malignant neoplasms are important: basal cell carcinoma, squamous cell carcinoma, and melanoma. There are many different causes of skin cancer, ranging from excessive exposure to sunlight to different sources of radioactive irradiation including X-rays, as well as exposure to chemicals such as PAHs and arsenic. Two fundamental mechanisms appear to be operational exemplified by PAHs and arsenic. PAHs require metabolic activation, via CYP isozymes, to epoxides, which form DNA-adducts having long half-lives. Therefore, the accumulated adducts induce mutations in critical genes and/or inhibit the p53 tumor-suppressor gene. Chronic irritation depletes the repair capacity of the skin, facilitating growth of malignant cells. UV light also has immunosuppressive effects that may contribute to increased survival of malignant cells by weakening defense mechanisms.

Skin cancer incidence is highest in the tropics in individuals with fair complexion at sites that are exposed to UV irradiation for long periods of time. Xeroderma pigmentosum best exemplifies the mechanism of action of sunlight-induced cancer. In this disease repair of pyrimidine dimers is genetically deficient in affected individuals and they develop squamous cell carcinomas after minimal exposure to UV light.

Scrotal squamous cell cancer in chimneysweeps was the first occupational cancer recognized as a result from chronic exposure to soot. Lip cancer of heavy pipe smokers is another example of chemically induced skin cancer. Coal tar, pitch, and creosote, rich in PAHs, are also known skin carcinogens. Coal tar in combination with UVB light used for the treatment of severe forms of psoriasis carries an additional risk of skin cancer.

Arsenic is another chemical carcinogen which causes skin tumors. High doses result in precancerous hyperkeratosis followed by the development of squamous cell carcinoma. Arsenic in drinking water has been shown to cause squamous cell carcinomas of the skin in populations living in areas with high soil content of arsenic. Unlike PAHs, arsenic is not mutagenic and, hence, it is considered an epigenetic carcinogen, also called a promoter. In contrast to genotoxic agents the continuous presence of epigenetic carcinogens is required at the site of action because promotion is a reversible effect and irreversibility (cancer) does not occur until the stage of progression is reached.

3.6.4 Summary

Skin and the anterior segment of the eye protect the organism efficiently against toxic insults from chemical, biological, mechanical, and physical injuries. The stratum corneum and the tear film, if intact, are impenetrable for many xenobiotics. When this protective capacity is exceeded or destroyed, absorption occurs with the consequence of potential damage to adjacent cells and/or distant structures. Though epidermis and cornea are not vascularized, the next layers beneath them have an extensive blood supply facilitating not only thermoregulation in the form of 'perspiratio insensibilis' (insensible perspiration) but also quick removal of absorbed molecules. Furthermore, the skin is a highly active and immunocompetent organ. The Langerhans cells of the epidermis play an important role in antigen presentation resulting in an immune response to invading organisms or chemicals. In addition, the skin and its appendages can have an excretory function, which in some instances is the only way to rid the organism of persistent xenobiotics that are not readily metabolized.

Toxic insults to the skin and/or eye by chemical, physical, biological, or mechanical agents cause, first, irritation, followed by an acute local inflammatory response when protective barriers are penetrated or destroyed. When the invader is a protein or hapten–protein complex, an allergic reaction at the local site or at a distant site may occur.

Sunlight is the most potent environmental insult to the skin. Historically, it was believed that exposure to sunlight was beneficial for the health of an individual. Only during the last century has it been recognized that excessive irradiation with sunlight has also detrimental effects, ranging from the destruction of elastic elements of the skin to carcinogenesis. In some cases, a combination of light and the presence of a chemical in the epidermis are required to bring about either a phototoxic or photoallergic reaction, depending on the underlying pathology.

Chronic exposure to irritants or allergens produces chronic inflammation accompanied by tissue proliferation, scaling, and itching. The appendages of the skin are particularly prone to toxic responses, affecting mainly hair follicles and sebaceous glands. Chronic irritation in combination with aggravating factors such as chemical or physical carcinogens is the most common pathway for skin carcinogenesis. Testing of compounds for potential skin irritation involves application of material into the conjunctival sac of test animals. This procedure is standardized and if performed in at least two species it has a reasonably good predictive value for humans. Skin and eye are potential entry portals for xenobiotics and should be always taken into consideration when conducting risk assessments for chemicals.

Reference

Casarett and Doull's Toxicology. The Basic Science of Poisons, ed. C.D. Klaassen, 6th edn, McGraw-Hill, New York, 2001.

For Further Reading

R.L. Bronough and H.I. Maibach, Percutaneous Absorption, Marcel Dekker, New York, 1989.
F.N. Marzulli and H.I Maibach, Dermatotoxicology, Hemisphere Publishing Corp., Washington, New York, London, 1983.

Slaga, T.J., Klein-Szanto, A.J.P., Boutwell, R.K., Stevenson D.E., and Spitzer H.L., Skin Carcinogenesis, Mechanism and Human Relevance. *Prog. Clin. Biol. Res.*, **298**, 1–424, 1989.
Yamagawa K., and Ichikawa, K. Experimentelle Studie ueber die Pathogenese der Epithelgeschwuelste. *Mitt. Med. Fak. Kaiserl. Univ. Tokyo*, **15**, 295–344, 1915.

3.7 The Kidney and Urinary Tract

Helmut Greim

3.7.1 Introduction

The kidney is mainly involved in maintaining the physiological homeostasis (functional, steady state, equilibrium) of the organism and in the elimination (excretion) of (metabolic) waste. 99% of the water of the filtered primary urine is reabsorbed, mainly in the tubular system, leading to an almost 100-fold increase in the concentration of compounds that are not or only poorly reabsorbed, which explains the specific vulnerability of the kidney to toxic chemicals.

The kidney plays a central role in maintaining cellular homeostasis. It regulates body fluid volume, the content and composition of electrolytes, and it eliminates useless or harmful water-soluble substances via the urine, such as urea or water-soluble foreign compounds and their metabolites. These functions are the result of the high blood flow through the glomerulus. The abundant blood supply, together with the absorption of water and substances in the tubular system, may result in high concentrations of toxicants present at only low concentrations in the blood.

Phase I and Phase II enzymes in the kidney present a high capacity for metabolic activation or inactivation of foreign compounds. Presentation of extrarenally formed toxic metabolites in the kidney depends mainly on their water solubility and chemical stability. Although glucuronides, sulfates, and glutathione conjugates usually are less toxic than their parent compounds, several of these conjugates become more toxic by kidney metabolism than the original products (Chapters 2.2B and 6.5).

3.7.2 Anatomy and Function

Anatomy and function of the kidney are focused on filtration of water and water-soluble substances, and absorption of water and essential substances out of the lumen of the tubular system, as well as elimination of metabolic waste into the urine. The functional unit is the nephron.

The kidney consists of an outer part, the cortex, and an inner part, the medulla. The shape of the medulla is pyramidal, the units are the papillae, with the outlets of the collecting ducts on each tip (papillary duct). The papillae project to the renal pelvis, which narrows towards the renal hilus. Here the renal artery enters the organ, and the renal vein and the ureter leave it. Via the ureter the urine is transferred to the bladder, from which it is voided via the urethra.

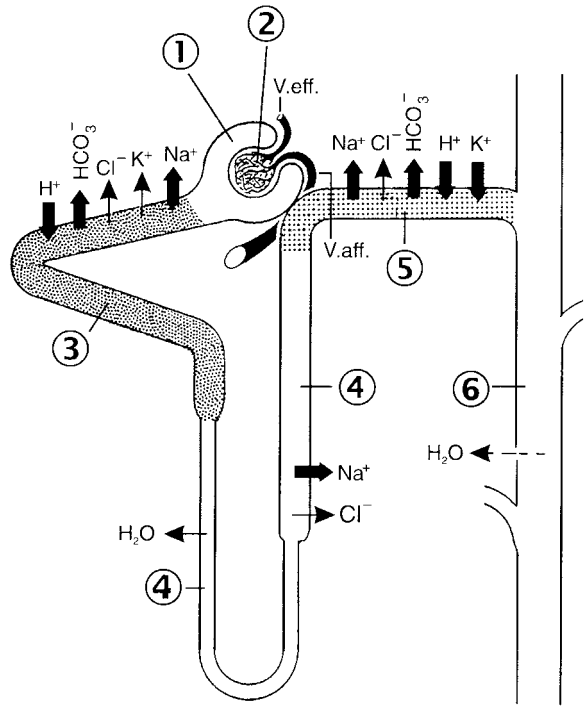


Figure 3.31 *Structure and function of a nephron. 1, Bowman's capsule. 2, Glomerulus with afferent and efferent arterioles. 3, Proximal tubule. 4, Loop of Henle. 5, Distal tubule. 6, Collecting tubule. The efferent arterioles form a dense capillary network around the proximal and distal tubules and the Loop of Henle, and regulate absorption of water and the different ions from the increasingly concentrated ultrafiltrate. [Modified from Greim and Deml, Toxikologie, Ch. 16, Abt. 16-1. Copyright (1996), with permission from Wiley-VCH.]*

Blood Filtration

The nephron (Figure 3.31), which is located in the cortex, consists of the renal corpuscle (or corpuscle of Malpighi) with the glomerulus and the Bowman's capsule, which filters the primary urine. In the renal tubule water and sufficiently lipid-soluble compounds are reabsorbed, while others are transported through the tubular membranes by active processes.

The renal artery supplies the blood for filtration as well as for organ supply. Its branches, the afferent arterioles, enter the Bowman's capsules at the vasculary pole and branch again to form a ball of capillaries, the glomerulus. The capillaries and the inner layer of the Bowman's capsule constitute a semipermeable membrane through which the ultrafiltrate (primary urine) leaves the blood and passes into the lumen of the capsule of Bowman. The ultrafiltrate is practically free of cells and proteins, but contains all soluble substances of the blood plasma, including small molecules with a molecular mass of up to about 15 000 daltons. According to molecular loading and structure, larger molecules

up to a molecular mass of about 70 000 can also appear in the ultrafiltrate (hemoglobin of molecular mass of about 68 000 to about 3%). Normally no proteins, except traces, appear in the urine, because they are almost completely reabsorbed during their passage through the tubular system.

The rate of blood perfusion of the kidney is about 1 L per minute, which corresponds to about one-fourth of the volume that passes the heart per minute. Of this about 120 mL of ultrafiltrate per minute is produced. Thus, most of the blood leaves the Bowman's capsule at the urinary pole unfiltered via the efferent arterioles. The subsequent second capillary system of the kidney, the peritubular capillary net, encircles the tubules and takes up water and substances from the increasingly reduced volume of the ultrafiltrate.

During passage through the tubules 98–99% of the water previously filtered is reabsorbed. Most constituents of the ultrafiltrate are reabsorbed in the proximal convoluted tubule. Here active, adenosine triphosphate (ATP)-consuming processes selectively transport substances from the ultrafiltrate back to the blood, like sodium, bicarbonate, and phosphate ions, amino acids, and glucose. Urea, chloride ions, and water follow passively. About 60–80% of the water diffuses back in the proximal tubule. This passive process is primarily driven by ATP-dependent reabsorption of Na^+. It is obvious that any toxic effects in this part will heavily affect kidney function. Under physiological conditions the pH value of the blood is about 7.38, that of the voided urine about 6.0. Maintenance of this gradient is of great importance for the absorption of acids and bases. Any changes, like an increase of the pH (alkalosis) or a decrease (acidosis), will affect their excretion and may result in toxic effects.

During the passage through the loop of Henle and the distal tubule, the absorption of water and sodium chloride continues. In the last part of the convoluted tubule and in the collecting ducts the fine adjustment in the concentration of electrolytes, the acid-and-base balance, and water takes place. The regulation of electrolytes and water homeostasis is extremely complex and is partly controlled by hormones. For example, angiotensin II and vasopressin regulate blood flow and by that glomerular pressure.

The straight, ascending limb of the distal tubule closely joins the afferent arterioles near the corpuscle of Malpighi (glomerulus and capsule of Bowman). Here a complex feedback mechanism regulates the production of the ultrafiltrate. An increased volume of the solute or concentrations in the solute in the distal tubule reduces the blood flow in the afferent arteriole and by that reduces the glomerular filtration rate.

The human kidneys, with about two million nephrons, produce per day 1–2 L of final urine out of about 180 L of ultrafiltrate. As a consequence, concentrations of compounds that are present in the ultrafiltrate and are not reabsorbed increase by 100-fold and may become toxic.

Collection and Excretion of Urine

The final urine flows through the collecting ducts, the renal pelvis, and the ureter into the urinary bladder. Ureter and urinary bladder are muscular organs lined by a multilayered epithelium.

3.7.3 Toxicology

Intoxication of the Kidney

Owing to the specific physiology of the kidneys toxicity of chemicals may result from increasing concentrations in the lumen of the tubular system or in the tubular cells. Intoxication of the kidneys mostly leads to impaired membrane function of the glomerulus or the tubular cells.

Intoxication of the kidney usually impairs filtration and reabsorption. Some site-specific effects can be distinguished.

Glomerulus Although the glomerulus is the primary site of chemical exposure, disturbance of its function by nondrug chemicals is rare. Several mechanisms may be involved: Impairment of ultrafiltration by vasoconstriction of renal arteriae (amphotericin), interaction with endothelial cell membranes (gentamicin), or direct cytotoxic effects on epithelial cells (cyclosporin). Such effects lead to impaired production of the ultrafiltrate. Another cause of glomerular injury is circulating immune complexes that become trapped in the glomerulus. These attract neutrophils and macrophages, which release reactive oxygen species and mediators for inflammation, contributing to glomerulonephritis, the inflammation of the renal corpuscles. Deposition of antigen–antibody complexes against the glomerular basement membrane and/or thickening of the glomerular membrane will also impair filtration. A reduced ultrafiltrate production due to a reduced or less permeable filter surface or a decline in the filtration pressure causes a diminished excretion of metabolic waste. Accumulation of these compounds in the organism will result in azotemia.

Exposure to volatile hydrocarbons and organic solvents can enhance membrane permeability in the glomerulus. As a consequence, larger molecules like albumin and gamma-globulin, which are normally retained, pass to the ultrafiltrate and are excreted with the urine (proteinuria).

Tubular System Tubular atrophy and thickening of the basement membrane, especially in cells of the proximal segment, impair reabsorption of electrolytes, water, and other substances. Intracellular intoxication can also occur by high concentrations of absorbed substances. Water reabsorption can lead to increasing and finally toxic concentrations of compounds in the tubular solute. This may lead to toxic effects in the surrounding epithelium or, by precipitation, to the formation of crystals and larger precipitates and by that to mechanical effects.

The proximal tubular system is preferentially affected by chemicals. Owing to high reabsorption rates chemicals can accumulate and reach toxic concentrations in the tubular epithelium. Such chemicals are the haloalkene-S-conjugates, and the $\text{alpha}_{2\text{urinary}}$-globulin-($\alpha_{2u}$)-bound chemicals such as limonene, cadmium, and mercury. Of the nephrotoxic drugs, aminoglycosides, β-lactam anitibiotics like penicillin, and the mycotoxin ochratoxin preferentially affect this part.

The loop of Henle and the distal tubular system are less affected. So far, only drugs like methoxyflurane or cisplatin or acidification are known to affect this area. Since this part controls the fine regulation of water and sodium uptake the clinical consequences of functional impairment are polyuria and loss of electrolytes.

As a consequence of tubular membrane damage, enzymes leak out and appear in the urine (enzymuria). Determination of specific enzymes allows identification of the site of the lesions. Enhanced urinary concentration of ligandin points to lesions in the proximal tubuler, increased concentration of lactate dehydrogenase indicates lesions in the distal tubule.

Functional disorders may occur in combination or in temporal sequence. After poisoning with halogenated alkenes, polyuria is observed first, which in case of serious intoxication progresses to oliguria. Important factors are the dose and the time of exposure. A slight proteinuria will be without further consequences when the concentration of enzymes – traces of nearly all serum enzymes are normally present in the urine – is not exceeded dramatically. A severe proteinuria, on the other hand, leads to protein deficits, formation of edema, and renal failure.

Compounds toxic to the tubular system

Hydrocarbons

Hydrocarbons affect the kidneys by different mechanisms. These comprise direct cytotoxicity, interaction with specific proteins (alpha$_{2u}$-globulin), and formation of reactive metabolites of mercapturic acids by β-lyase.

Cytotoxicity.–In humans kidney glomerulonephritis and tubular necroses were observed after exposure to petroleum hydrocarbons, like benzene, solvents, kerosene, or Diesel fuel (Table 3.17). These effects seem to be the direct interaction of the compounds with cellular membranes.

Halogenated alkanes and alkenes are basic substances in industrial production, and are used as solvents and for pest control. They are an important group among the kidney toxins and damage mainly the straight part of the proximal tubule. Allyl chloride enlarges the gap of Bowman's capsule. Some halogenated hydrocarbons exert a toxic potential in the liver as well as in the kidney. This applies for carbon tetrachloride (tetrachloromethane) and chloroform. The reactive metabolites are generated mainly in the liver and attain the kidney via the blood. In the case of chloroform the toxic metabolite phosgene is generated in the proximal tubule by CYP450. It reacts with various cell components and is detoxified by GSH-conjugation.

Alpha$_{2u}$-globulin-mediated effects.–The low-molecular mass protein alpha$_{2u}$-globulin (alpha$_{2urinary}$-globulin) is formed in the male rat under the influence of testosterone. About 60% of the protein is reabsorbed in the middle portion of the tubular system and is degraded within the tubular epithelium by lysosomal enzymes. The nonreabsorbed portion is excreted and constitutes the main component among the proteins found in the urine of this gender. In the aging male rat the content of alpha$_{2u}$-globulin declines and that of albumin rises, pointing to glomerulonephritis, which often spontaneously occurs in the aged rat. It entails necroses, regenerative growth, and kidney tumors.

In kidneys of male rats a sex-specific kidney tumor occurs upon exposure to volatile hydrocarbons and their derivatives like unleaded gasoline, decalin, *p*-dichlorophenol, isophorone, limonene, and methyl isobutyl ketone. These compounds are metabolized in the liver, where the reactive metabolites bind to the aforementioned substrates and are

Table 3.17 *Nephrotoxic chemicals.*

Compounds	Mechanisms	Primary target	Relevance
Allyl chloride		Bowman's capsule	
Solvents	Increased permeability	Bowman's capsule, proximal tubule	
Unleaded gasoline, decalin, *p*-dichlorophenol, isophorone, limonene, methyl isobutyl ketone	Formation of alpha$_{2u}$-globulin	Middle portion of proximal tubule	Male rat specificity
Trichloroethene, tetrafluoroethene, hexachloro-1,3-butadiene	β-Lyase-mediated formation of thioketene	Proximal tubule	Possible human carcinogens
Tetrachloromethane	Trichloromethyl radical metabolite	Proximal tubule	
Chloroform	Biotransformation to phosgene	Proximal tubule	
Ethylene glycol	Precipitation of calcium oxalate	Proximal tubule	
Diethylene glycol	Increase in osmolality	Whole nephron	
Heavy metals	''Various, see Chapter 6.2 on metals''	Proximal tubule	
NTA	Ca** extraction from membranes	Proximal tubule	
Aminoglycosides	Formation of undegradable phospholipids	Proximal tubule	Antibiotic drugs
Cisplatin	DNA-crosslinks	Loop of Henle	Cytostatic drug
N-Nitroso compounds	Alkylating	Bladder	
[4-(5-Nitrofuryl)-2-thiazolyl]hydrazine	Peroxidation by prostaglandin-H-synthase	Bladder, tubule	
2-Naphthylamine, 4-Aminobiphenyl	Metabolism to arylhydroxylamines	Bladder	Contaminants in aniline production
Aromatic amines	Metabolism to arylhydroxylamines	Bladder	
Bladder stones	Mechanical irritation	Bladder	

transported to the kidneys. They are taken up by the epithelial cells of the middle portion of the proximal tubule and accumulate as small crystalline protein droplets because the protein–metabolite complex cannot be degraded by lysosomal enzymes. This chemically mediated accumulation of α_{2u}-globulin results in cytotoxicity, which in turn stimulates cell division. With prolonged chemical exposure, repeated cycles of cytolethality and reparative replication are responsible for the tumorigenic response. Alpha$_{2u}$-globulin is not produced in other

experimental animals and in humans, so that tumors induced by interaction of reactants with alpha$_{2u}$-globulin are species-specific for male rats and have no relevance to man.

β-Lyase-mediated carcinogenicity.–An important mechanism in kidney toxicity is the bioactivation of GSH-conjugates, such as of trichloroethylene or of hexachloro-1, 3-butadiene. Formation of such conjugates by GSH-transferases occurs in the liver and the kidney. By means of gamma-glutamyltransferases or dipeptidases, which split the gamma-glutamyl and the glycine residues, cysteine *S*-conjugates are formed in the intestine or in the kidney. These cysteine *S*-conjugates can be toxic by themselves or after further metabolism. Especially in the kidney, high activity of C-S-lyase (beta-lyase) generates toxic metabolites, which alkylate macromolecules and ultimately lead to cancer (see Chapter 6.5 on solvents).

Heavy metals

Heavy metals especially damage the proximal tubule.

Many metals are reabsorbed by the epithelium of the proximal tubule. Among others, these are cadmium, mercury, arsenic, chromium, platinum, and bismuth. The target can be limited, especially in the case of slight poisoning. Chromium preferentially affects the first portion of the convoluted tubule, mercury the straight part. With increasing mercury exposure the whole tubule becomes affected. Copper salts damage the epithelia of the loop of Henle and of the distal tubule, as known from their use as a fungicide (Figure 3.31). Mercury nephrotoxicity is the consequence of two mechanisms: The contracting effect on blood vessels with concomitant increase in blood pressure, and the cytotoxic effect, which inhibits a series of enzymes and affects cell organelles. Probably, the binding of mercury to the endoplasmic reticulum (ER) and its excretion into the lumen is a detoxifying process. The binding to lipoproteins in lysosomes or, in the case of cadmium and other heavy metals, binding to metallothionein is also protective.

Polyvalent alcohols

Polyvalent alcohols damage the tubule in different ways.

Polyvalent alcohols such as the bivalent alcohols ethylene glycol and diethylene glycol damage the proximal tubule. They are used as solvents and antifreeze agents. Oxalate, the metabolite of ethylene glycol, and probably also intermediate products like glyoxylic acid are nephrotoxic. During reabsorption of water, precipitation of calcium oxalate crystals occurs in the extracellular space, which obstructs the lumina and damages the epithelial cells mechanically, entailing urine retention (ischuria), excessive pressure in the nephrons, and insufficient excretion of toxic waste from the blood. Diethylene glycol increases osmolarity in the kidneys and causes swelling and degeneration of epithelial cells. Formation of oxalate does not occur. Another example is the trivalent alcohol glycerol. After treatment with glycerol, nephrons situated near the surface of the kidney collapse.

Nitrilotriacetic acid (NTA) In long-term repeated-dose studies NTA applied via feed or drinking water induced cytotoxicity, resulting in hyperplasia of the tubular epithelium

and tumors in the kidney, the ureter, and the bladder. In the kidney the primary location of the carcinoma is the tubular cells. The lowest tumor-inducing dose is 100–140 mg/kg body weight (bwt). NTA is readily absorbed from the gastrointestinal tract and excreted unchanged via the kidney. It has a high complexing activity for divalent ions. In the gut it preferentially binds zinc, and NTA–zinc is the major complex excreted. Since there is little tubular reabsorption, NTA increases in concentration during passage through the tubular systems and finally precipitates. At about 200 mg/kg bwt free NTA appears in the ultrafiltrate, which extracts Ca^{++} ions from the tubular epithelium and the epithelial cells of the urinary system. Membrane damage, cytotoxicity, hyperplasia, and tumors are the result of reabsorbed zinc, extraction of calcium from the membranes, as well as physical damage by precipitating NTA.

Other chemicals Among the anthropogenous substances in the environment there are pesticides and industrial chemicals with nephrotoxic potential. In pest control by soil fumigation HCBD (hexachlorobutadiene) and the herbicides Paraquat (1,1′-dimethyl-4,4′-bipyridinium) and 2,4,5-T (2,4,5-trichlorophenoxyacetic acid) are used. They inhibit the active elimination mechanism of organic ions in the proximal tubule and by that inhibit their own excretion. Some of the halogenated hydrocarbons like polychlorinated biphenyls, dibenzodioxins, and dibenzofurans are not directly nephrotoxic. They induce enzymes, which metabolize foreign compounds and at high exposure may metabolically activate other toxicants.

Toxicology of the Urinary Tract

The urinary tract is a relatively insensitive target for toxic foreign compounds.

There is little information on the toxicity of foreign compounds in the ureter and the urethra. It may be that the short stay of the urine in these organs hinders any effective influence from the toxic compounds.

Toxic effects on the urinary bladder by foreign compounds are practically unknown, but there are a number of potentially carcinogenic substances that may affect this organ (Table 3.17).

Carcinogenic Effects

Tumors of the Kidney

The formation of malignant neoplasms in the human kidney is uncommon, but there is a series of different tumors, being dependent on the type of tissue and the age of the person. In the last few years there is increasing evidence that genetic defects play an important role. So in renal cell carcinoma a deletion on chromosome 3, segment p has frequently been found.

There is little known about tumors in the human kidney induced by foreign compounds. Smoking is a factor of risk, but in this case the transitional epithelium of the renal pelvis seems to be affected rather than the renal parenchyma.

In contrast to the human situation, in experimental animals several chemicals are tumorigenic in the kidney. Among them are N-nitroso compounds, e.g. diethylnitrosamine and dimethylnitrosamine with initiating potency in different target organs in rats and mice, prevalently in the liver but also in the kidney. Ethylnitrosurea produces tumors of the brain and the kidney. With methylnitrosurea the kidney is the preferential target of carcinogenesis.

The furan (4-(5-nitrofuryl)-2-thiazolyl)hydrazine (FNT) causes tumors of the transient epithelium of the urinary bladder and tumors of the tubuli.

Naturally occurring substances like the mycotoxin aflatoxin, or methyl-*ONN*-azoxymethanol, the aglycone of cycasin, also induce tumors in experimental animals. In man, after consumption of flour of the Cycas fruit and peanuts contaminated with aflatoxin B_1, tumors of the liver but not of the kidney were observed.

Comparable to carcinogenesis in liver and skin, the steps of initiation and promotion can be distinguished. Besides the already mentioned initiating agents in kidney carcinogenesis, there are also promoting agents, at least experimentally; for example the plasticizer di(2-ethylhexyl) phthalate (DEHP).

Bladder Tumors

The bladder is a specific target for several occupational carcinogens.

Since the end of the 19th century it is known that 2-naphthylamine, which is an intermediate in aniline dyestuffs production, can give rise to tumors of the urinary bladder epithelium in men, dogs, and other mammals (aniline cancer). This is one of the earliest examples of occupational carcinogenesis. Meanwhile there are other aromatic amines known to be genotoxic carcinogens implicated in urinary bladder tumors of mammals, such as 2-(acetylamino)fluorene (2-AAF) in mice. In man the exposure to aromatic amines results in tumors of the urinary bladder, besides nephrotoxicity (Table 3.17).

Analogous to the situation in the renal pelvis the ingredients of tobacco smoke and the roasting products of coffee beans are suspected to enhance the cancer incidence of the transient epithelium in the urinary bladder.

In animal experiments some nitrosamines, urea derivatives, and nitrofurans have been identified as initiating agents. In particular the following substances have been found to be effective initiators: *N*-butyl-*N*-(4-hydroxybutyl)nitrosamine (BBN), *N*-methyl-*N*-nitrosurea (MNU), and *N*-[4-(5-nitro-2-furyl)-2-thiazolyl]formamide (FANFT), which in the kidney is metabolized to 2-amino-4-(5-nitro-2-furyl)thiazole (ANFT). The 5-nitrofurans are metabolized to the ultimate carcinogens by peroxidation via prostaglandin-H-synthase in the urinary bladder, as shown also for benzidine. Enhanced amounts of *N*-nitroso compounds were found in the urine of bilharziasis (schistosomiasis) patients with or without cancer of the urinary bladder. Epidemiologic studies revealed that infections of the kidney and the urinary tract enhance the risk for urinary bladder cancer. Experimentally, FANFT in combination with *E. coli* induces tumors, preferentially of the renal pelvis. In animal studies FANFT or *E. coli* alone are ineffective regarding tumor development.

Promoting agents are saccharin and cyclamate. Promoting stimuli that modulate tumorigenesis are the urine volume, pH-value, and the time a compound remains in the urinary bladder. Such factors stimulate proliferation of epithelial cells and evoke hyperplastic changes. Also, hypothermia or mechanical irritation by bladder stones (calculus vesicalis) stimulate proliferation. Calculi are produced experimentally by high doses of various chemicals, e.g. uracil. An important role is played by the urinary pH-value. In an initiation–promotion experiment salts that increased the pH-value of the urine have promoted carcinogenicity. Those that lower the pH-value inhibit the carcinogenic process.

3.7.4 Summary

The kidneys maintain physiological homeostasis by regulating the electrolyte and water balance. It takes place during blood filtration and reabsorption of essential amounts of water and salts. In the urine, which represents about 1% of the original ultrafiltrate, all water-soluble substances are eliminated. These can be those of excess, not useful for the body or toxic. Toxic effects are either the result of the high blood flow through the kidneys or due to the reabsorption of water in the tubular system. There, concentrations of water-soluble compounds that are not reabsorbed increase by 100-fold and may damage tubular epithelia or precipitate. Compounds that are reabsorbed may accumulate in the tubular epithelia and induce toxic effects. High urinary concentrations of toxic compounds may induce lesions in the epithelium of the urethra and the bladder.

Owing to its high metabolic capacity the kidney is able to metabolize foreign compounds to water-soluble and in some cases toxic derivatives, which also induce toxic effects at various sites of the kidney, the urethra, and the bladder.

References

Greim, H. and Deml, E. Toxikologie, Eine Einführung für Naturwissenschaftler und Mediziner, VCH Verlagsgemeinschaft mbH, Weinheim, 1996.

For Further Reading

Shitara, Y., Horie, T., and Sugiyama, Y. Transporters as a Determinant of Drug Clearance and Tissue Distribution. *Eur. J. Pharm. Sci.*, **27**, 425–446, 2006.

Barbier, O., Jacquillet, G., Tauc, M., Cougnon, M., and Poujeol, P. Effect of Heavy Metals on, and Handling by, the Kidney. *Nephron Physiol.*, **99**, 105–110, 2005.

Van Vleet, T.R., and Schnellmann, R.G. Toxic Nephropathy: Environmental Chemicals. *Semin. Nephrol.*, **23**, 500–5008, 2003.

Zalups, R.K., and Ahmad, S. Molecular Handling of Cadmium in Transporting Epithelia. *Toxicol. Appl. Pharmacol.*, **186**, 163–188, 2003.

Brüning, T. and Bolt, H. Renal Toxicity and Carcinogenicity of Trichloroethylene: Key Results, Mechanisms and Controversies. *Crit. Rev. Toxicol.*, **30**, 253–285, 2000.

Endou, H. Recent Advances in Molecular Mechanism of Nephrotoxicity, *Toxicol. Lett.*, **102–103**, 29–33, 1998.

3.8 Toxicology of the Blood and Bone Marrow

Robert Snyder

3.8.1 Introduction

Erythrocytes, various types of leucocytes, and thrombocytes arise in the bone marrow from stem cells, which undergo differentiation, proliferation, and maturation, and are released into the circulation. The toxicology of the blood and bone marrow involves two areas of concern: (1) the impact of drugs and chemicals on circulating blood cells, and (2) the impact on cells developing in the bone marrow. Thus, hemolytic anemias are events, which arise in the circulating blood as a result of factors such as genetics or chemicals, which affect mature circulating erythrocytes. In contrast most cases of decreases in white blood cells, i.e. leucopenia, result from disturbances in the formation of specific types of leucocytes in the bone marrow.

The bone marrow is the principal source of the cellular components of blood and maintains the proper levels of each cell type despite the fact that many leave the circulation and the life span of each cell type differs. Marrow is made up of fat through the hollow part of the bone with red marrow containing the blood cell-generating system largely at the ends of the long bones. Active marrow can also be observed in the bones of the spinal column, the ribs, the sternum, and in some pathologic conditions, e.g. in the bones of the jaw and skull in the disease called thalassemia major. The total volume of the bone marrow approaches the size of the liver.

The circulatory system in adults normally contains about 5 liters of blood. Each microliter contains 5 million erythrocytes. Therefore, at any given time there are 5×10^{12} erythrocytes in circulation. The life span of the red cell is about 120 days. Each day the marrow must therefore supply 4.2×10^{10} new erythrocytes. There is a need to synthesize various white blood cells and platelets, as well. Thus, the marrow displays a very high capacity for generating blood cells. By the same token, interference with the normal functioning of the marrow may result in failure to properly generate cells, as in aplastic anemia, or in the uncontrolled production of aberrant cells, as in leukemia.

Table 3.18 also contains a list of essential plasma proteins. Although not necessarily made in the bone marrow, they play a critical role in normal physiology. Among these, albumin is essential for maintaining the osmotic pressure of the blood and acts as a carrier for many drugs. The globulins have important immunological functions and also help to transport lipids. Other blood proteins include fibrinogen, which is converted into fibrin in the blood-clotting process, and transferrin and ferritin, both of which are essential for iron transport.

This chapter will focus on the normal sequence of events in marrow function, the roles played by circulating cells in homeostasis, and on the adverse effects caused by exposure to various drugs and other chemicals leading to commonly observed hemopathies.

Table 3.18 *Normal components of blood.*

Component	Normal amount in circulating blood
Blood cells	
Erythrocytes (red blood cells)	4.5 – 5.5 million/μL
Total leucocytes (white blood cells)	5000 – 9000/μL
Neutrophils	60–70%
Eosinophils	2–4%
Basophils	0.5–1.0%
Lymphocytes	20–25%
Monocytes	3–8%
Thrombocytes (platelets)	250 000
Plasma proteins	Approximate concentration (g/100 ml)
Total	7.3
Albumin	4.5
Globulins	2.5
Fibrinogen	0.3
Transferrin	0.25
Ferritin	0.015–0.3

3.8.2 Hematopoiesis

The generation of blood cells in the bone marrow from pluripotential stem cell to the mature circulating cell involves differentiation to yield cells, which are morphologically and physiologically equipped to perform the functions of the mature cells. Differentiation drives the evolution of cells from the primitive state of the stem cell to the mature cell. Amplification involves a series of mitotic events, which ensure that there are a proper number of cells of each type released into the blood daily.

The mechanism by which the bone marrow supplies mature blood cells is termed hematopoiesis and is summarized in Figure 3.32. The process involves the sequential differentiation from stem cells to progenitor cells to precursor cells, and finally, to mature circulating cells. At any given stage the cells may undergo mitosis as well as differentiation. This complicates the identification of specific cell types in the bone based on morphological observations alone because at any given time a smear of bone marrow cells will offer mainly more advanced cells and very few of the early cell types.

Two approaches have been used to facilitate identification of cells in marrow. One is based on the ability to grow individual cell types in colonies by stimulating proliferation using specific growth factors. These have been termed colony forming units (CFUs) and are specifically named for the differentiation pathway in which they are found. Figure 3.32 provides an example of the differentiation pathway outlined in terms of identified CFUs. In addition, the development of antibodies, which react at specific cell-surface receptors in bone marrow cells, provides an additional approach to identification of bone marrow intermediate cells and provides a useful tool for determining differentiation, using changes in cell-surface markers for specific antibodies.

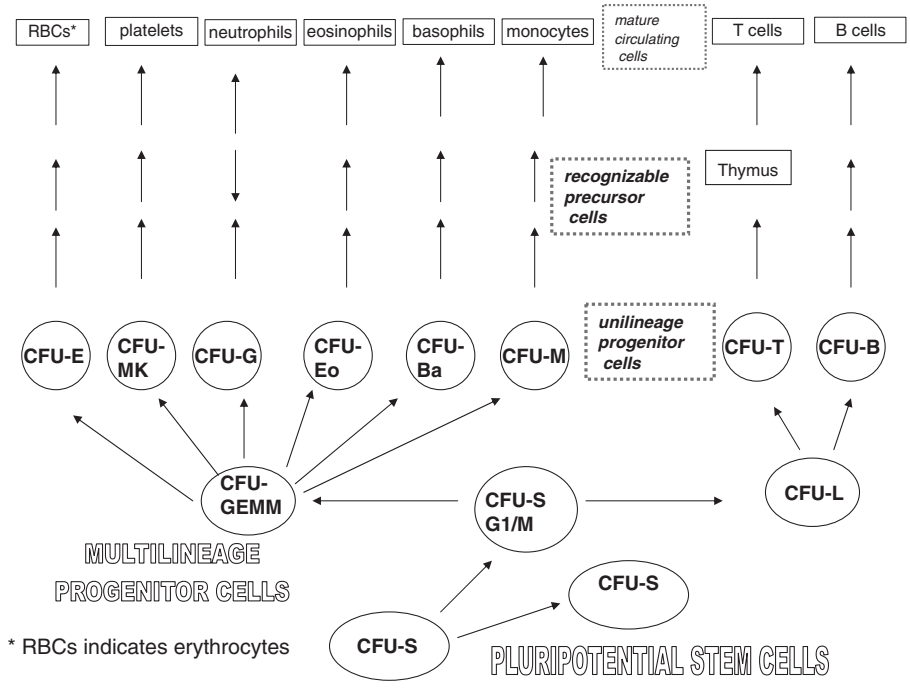


Figure 3.32 *Scheme outlining differentiation of bone marrow cells from pluripotential stem cells to mature circulating blood cells.*

The stem cell model argues that in the bone marrow there are a finite number of stem cells capable of replacing themselves by a mitotic process that also involves the release of a daughter cell that can progress through the differentiation/maturation scheme as influenced by specific growth factors. The first cell in the sequence has been termed the pluripotential, or multipotent, stem cell. The daughter cell (CFU-S, G_1/M), is a stem cell capable of giving rise to the multilineage progenitor cells CFU-GEMM and CFU-L. CFU-GEMM gives rise to all of the myeloid cells and CFU-L gives rise to lymphocytes. Further differentiation leads to unilineage progenitor cells for each of the circulating cell types. Although there may be some intermediary stages, ultimately all give rise to CFUs for each of the cell types in blood. Beyond the CFU stage morphological characteristics of the precursor cells are more readily apparent. The final stage is maturation from the last precursor cell to the mature circulating cell.

The need to produce specific numbers of each cell type at various times and in different amounts requires that the system must be under delicate control. The controlling factors are, for the most part, a series of cytokines. In many cases they are either colony-stimulating factors (CSF) or interleukins (IL). While some of these appear to concentrate their activity on specific cell lines, most others impact on several types of cells. A more complete discussion of the many stimulating factors and the cells upon which they act can be found in resources such as Hoffmann et al. (2000).

3.8.3 Toxicological Features of Circulating Blood Cells

Erythrocytes (Red Blood Cells)

Anemia, which is a decrease in circulating functional erythrocytes, can result from inhibition of red cell development in the bone marrow, or as a result of hemolysis or inhibition of red cell function due to genetic defects or interactions with specific chemicals.

Erythropoiesis Erythrocytes are unique among cells in the body by virtue of both their morphology and their mission. Erythropoietin is a specific cytokine, which directs differentiation of the erythroid cell line. Early recognizable precursor cells in the marrow, in sequence, are termed pronormoblasts, several stages referred to as normoblasts, and reticulocytes. Each of these cell types is capable of synthesizing hemoglobin, the oxygen-carrying protein that defines red cell function. The final steps in maturation of the erythrocyte involve the loss of the nucleus at the stage described as the polychromatic normoblast. The resulting cell, the reticulocyte, continues to synthesize hemoglobin, is released into the circulation, and matures to the erythrocyte, at which point hemoglobin synthesis is complete.

The primary function of the erythrocyte is to transport oxygen from the lung to the tissues of the body. The red cell undergoes mechanical and oxidative damage during its 120-day life time. Eventually, as repair mechanisms fail, the cell is removed from the circulation, usually upon passage through the spleen, and the cell is lysed.

Hemoglobin Function Hemoglobin is a protein composed of 4 peptide chains (globins), 2 of which are termed α chains and 2 are β chains. Each chain carries a heme group, composed of a mole of ferrous (Fe^{++}) ion embedded in protoporphyrin IX and linked to an imidazole group on a histidine in the peptide chain (Figure 3.33). Hemoglobin can reversibly bind a molecule of O_2 to the Fe^{++}. The extent to which hemoglobin is saturated

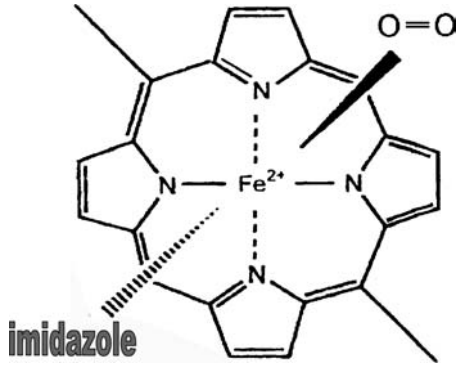


Figure 3.33 *Heme group. Active component of hemoglobin, which transports oxygen from the lung to the tissues. [Modified from Greim and Deml, Toxikologie, Ch. 17, Abt. 17-3. Copyright (1996), with permission from Wiley-VCH.]*

with O_2, i.e. the degree to which the total number of heme groups are bound to O_2, is determined primarily by the oxygen tension. The air in the alveoli of the lung contains a higher concentration of O_2 than the blood circulating through blood vessels surrounding the alveoli. The concentration gradient drives the O_2 from the alveoli into the blood, where it combines with the heme groups of hemoglobin to form oxyhemoglobin. During the course of circulation O_2 is released to tissues having a relatively low O_2 tension. Carbon dioxide is found in the tissues largely as bicarbonate, and is transported through the blood to the lung. Association of bicarbonate with oxyhemoglobin enhances release of oxygen in the tissues whereas release of bicarbonate as CO_2 in the lung promotes oxyhemoglobin formation. The presence of O_2 constantly oxidizes hemoglobin ($HbFe^{++}$) to methemoglobin ($HbFe^{+++}$). The level of methemoglobin in the blood is kept at about 1% by efficient NADH- and NADPH-dependent methemoglobin reductase systems. The key enzyme that provides NADPH is glucose-6-phosphate dehydrogenase (G6PD). G6PD deficiency results in greater sensitivity to methemoglobin-inducing agents.

Hemolytic Anemias are Diseases Characterized by a Deficiency of Effective Erythrocytes Resulting from Lysis of Abnormal Red Cells Hemolytic anemias are usually the result of either genetic defects in the red cell membrane structure or in the structure of hemoglobin. For example, hereditary spherocytosis is a hemolytic disease featuring genetic variants in which there are deficiencies of a number of erythrocyte membrane proteins, leading to the production of small spherical red cells, which are trapped in the spleen and hemolysed. Sickle cell anemia and the thalassemias are genetic diseases in which mutant globin chains are formed, resulting in ineffective red cells, which are readily hemolysed.

In some Cases Hemolysis Results from the Interaction of a Chemical with Normal or Abnormal Red Cells

Glucose-6-phosphate dehydrogenase deficiency Reduced glutathione (GSH) is a critical factor in the maintenance of the red cell membrane. The oxidation of GSH to GSSG can be reversed by the NADPH-requiring glutathione reductase. NADPH is generated during the oxidation of glucose-6-phosphate by G6PD. People genetically deficient in G6PD fail to provide enough NADPH to meet the needs of glutathione reductase and subsequent membrane failure leads to hemolysis.

G6PD deficiency, which is an X-linked recessive disorder, is observed primarily in males. About 10% of African-American males demonstrate a relatively mild form of the disease, which has been called primaquine sensitivity because of its occurrence in black US servicemen treated with prophylactic doses of primaquine to protect against malaria during the Korean War. Favism, which occurs among people who live in the Mediterranean region, is a more severe form of G6PDH deficiency disease in which people who eat fava beans (or inhale its pollen) may undergo hemolysis sufficiently severe to be fatal. Fava beans contain pyrimidine aglycones that cause more excessive oxidation of GSH in G6PD-deficient erythrocytes than is observed in primaquine sensitivity.

Carbon monoxide poisoning Carbon monoxide binds to hemoglobin (Fe^{++}) 200–300-times more avidly than does oxygen. Carboxyhemoglobin formation results in a 'cherry red' complexion with cyanosis. Low concentrations of CO lead to headaches, 40% carboxyhemoglobin results in impaired vision, tachycardia, and hyperpnea, and

60% can be fatal. Treatment involves increasing the ambient oxygen pressure by inhaling 100% oxygen or 95% oxygen: 5% carbon dioxide. The role of carbon dioxide in the latter is to increase breathing frequency. In severe cases providing oxygen at hyperbaric pressures may be required to save the life of the exposed individual.

Methemoglobinemia

Hemoglobin in which the iron has been oxidized from the ferrous to the ferric state is called methemoglobin. It cannot bind and transport oxygen.

During the normal course of oxygen transport an electron from the iron becomes partially associated with the oxygen, which assumes a superoxide-like structure, but returns to the iron upon the release of oxygen to the tissues. The process is not completely efficient, with the result that the electron may not always be restored to the iron. Under these circumstances the iron is converted from the reduced (ferrous) form into the oxidized (ferric) form. Hemoglobin with an oxidized iron is called methemoglobin, and cannot bind and transport oxygen. The stricken individual has a characteristic gray facial appearance at levels below 30% methemoglobin. Above that value the patient becomes cyanotic. Many nitrogenous compounds, sulfonamides, and quinones can induce the formation of methemoglobin. The cell normally maintains the level of methemoglobin at about 1% or less via the activity of an NADH-requiring cytochrome b_5 methemoglobin reductase. In the event of an overdose with a methemoglobin-forming agent additional reductase activity is required. A nascent form of methemoglobin reductase requires NADPH and its activity requires a reducing agent such as methylene blue.

Cyanide is a mitochondrial poison, which binds cytochrome oxidase in its Fe^{+++} configuration and prevents oxygen utilization. However, CN^- binds avidly to methemoglobin. Thus, in the event of cyanide poisoning, sodium or amyl nitrite can be administered to oxidize a small fraction of hemoglobin to methemoglobin, which acts as a trap for the cyanide ion.

The interactions of hemoglobin, methemoglobin, carbon monoxide, and cyanide are summarized in Figure 3.34.

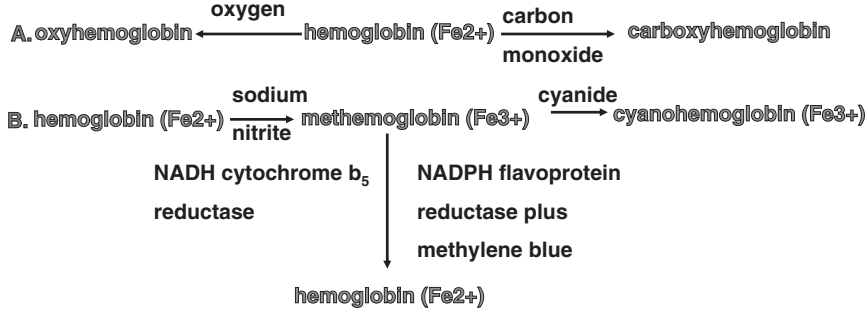


Figure 3.34 *A. Reactions of hemoglobin with oxygen and carbon monoxide. B. Methemoglobin reduction; methemoglobin reaction with cyanide.*

In the event of an overdose of cyanide, advantage can be taken of an enzyme called rhodanese (rhodanase), which is a normal blood component. Upon treatment with thiosulfate, rhodanase catalyses a reaction with cyanide, the products of which are thiocyanate and sulfite. Thiocyanate is a detoxified form of cyanide.

Leucocytes (White Blood Cells)

The term leucocytes, or white blood cells, encompasses several cell types of differing morphology, with diverse physiological functions, i.e. protection against infection by neutrophils and macrophages; mediation of inflammatory reactions by eosinophils and basophils; T lymphocytes concerned with cell-mediated immunity; B lymphocytes, which enhance immunity via the production of antibodies; and platelets, which are essential for blood clotting.

The leucocytes are a heterogeneous group of cells that are distinguished from the erythrocytes primarily because they lack hemoglobin and all but the platelets contain a nucleus. They include the granulocytes (neutrophils, basophils, and eosinophils), the lymphocytes, and the thrombocytes (platelets), as well osmonocytes and macrophages.

Granulocytes are Named on the Basis of their Staining Properties

(a) Neutrophils comprise about 55% of leucocytes in the circulation. Normal neutrophil counts range from 3000–6000 cells/μl of blood. Their primary function is to engulf and destroy bacteria and other foreign material. Neutrophils are critically important in countering bacterial infections. People who exhibit neutropenia have a reduced capacity to resist infections. Agranulocytosis is a condition in which there is a severe reduction in neutrophils and is usually accompanied by high fevers indicative of infection.

Upon maturation of neutrophils in the bone marrow they require an array of stimuli to affect their release into the circulation. These may include endotoxin, androgens, glucocorticoids, and so-called 'CXC' chemokines. Neutrophils leave the blood, in response to a variety of stimuli, with a half life of about 4–10 h. They accumulate at sites of infection or inflammation, age, and die within 72 h.

The bactericidal effect of neutrophils involves first engulfing the organism and then releasing lethal reactive oxygen species via the respiratory burst mechanism. The process is initiated via a cytochrome b-associated NADPH oxidase reaction leading to generation of superoxide, hydrogen peroxide, hydroxyl radicals, and singlet oxygen. The Haber–Weiss reaction plays a key role in this sequence of events.

In addition to the oxidative pathway, neutrophils attack microorganisms by generating a series of proteins, such as the defensins and serprocidins, which alter surface protein activity and results in viricidal, fungicidal, and bactericidal activity.

(b) Eosinophils comprise about 3% of leucocytes or between 50–250 cells/μl of blood. They stain red when exposed to eosin, an acidic dye. Eosinophils are important in protecting against parasitic infections, but excessive accumulation of eosinophils in airways may further impair breathing in asthmatics.

(c) Basophils comprise less than 1% of leucocytes or about 15–50 cells/μl. They are concerned with generating allergic and hypersensitivity reactions. Their granules contain histamine, appear blue in the presence of a basic dye, and release histamine to initiate a variety of immediate hypersensitivity reactions such as asthma, urticaria, rhinitis, and possibly anaphylactic shock.

Lymphocytes Comprise Two Types of Cell, Arising in the Bone Marrow Lymphocytes comprise about 30% of circulating white cells or 1500–3000 cells/μl. The lymphocyte population may be divided into two types of cells, termed B and T cells. Both types arise in the bone marrow. B cells (about 10–15% of circulating lymphocytes) mature in the bone marrow but T cells (about 70–80%) migrate to the thymus to complete their maturation.

Lymphocytes are key components of the immune system. B cells are *lymphocytes* concerned with *humoral immune response*, whereas T lymphocytes function in *cell-mediated immune response* mechanisms.

They react with antigens, which are macromolecules containing protein or polysaccharide structures, which may also have lipoid characteristics. There are B cell and T cell receptors specific for designated antigen structures. Once the cells have taken up and processed the antigen they proceed to undergo mitosis to produce a clone of cells having the same antigen-responsive characteristics, which appear on the cell surface as so-called class II histocompatibility molecules. These bind to helper T lymphocytes, which then generate lymphokines, which in turn stimulate the production of B cell receptors in a soluble form. The cells differentiate to plasma cells, which secrete the soluble receptors, which are now called antibodies and which are capable of binding with and inactivating the antigens.

Alternative mechanisms by which T cells can act relate to the presence of either CD4 or CD8 cell-surface glycoproteins. CD8+ cells transmit cytotoxic molecules to infected cells, which induces apoptosis. CD4+ cells combine with phagocytic cells, such as macrophages, to initiate an inflammatory response, which leads to the death of the antigen-containing cell. Thus, B and T lymphocytes play complex roles in support of the immune system.

The Monocyte/Macrophage System Monocytes develop in the marrow and are released into the circulation where they may function in phagocytosis, chemotaxis, and the killing of microorganisms. They have immunologic and secretory functions. Within a day of their entry into the circulation they leave the blood and migrate to various organs where they appear as in situ macrophages. Thus, there are pulmonary/alveolar macrophages in lung and Kupfer cells in the liver, which perform the functions of macrophages in those organs.

Macrophages have a number of important physiological functions, but when activated may release toxic mediators that have a deleterious effect on the surrounding tissues. Thus, among the many secretions emanating from macrophages there are cytokines, stress proteins, reactive oxygen species (which may be bactericidal or which may damage local tissues), tumoricidal mediators, fever-inducing pyrogens, bioactive lipids and oligopeptides, a variety of degradative enzymes, growth factors, and other materials. The ultimate impact will depend upon the physiological state of the local tissues and effects of other agents.

Platelets (Thrombocytes)

Platelets are essential for initiating the process of blood coagulation. They are formed from a cell type called the megakaryocyte, which develops in the bone marrow in a fashion similar to other cells of the myeloid and lymphoid lines, but instead of maturing and entering the circulation they release small cells that lack a nucleus but are highly adapted to initiate blood coagulation.

Upon damage to blood vessels, collagen is released into the immediate circulation and in response platelets accumulate at the site of injury and release adenosine diphosphate (ADP), which enhances their ability to stick together and form a plug to halt blood loss. Fibrinogen adheres to the surface of the platelet plug and is converted into fibrin, the strong protein backbone of the clot; other cells, such as erythrocytes, are trapped in the plug and reinforce the clot.

Thrombocytopenia is a decrease in blood platelets and can lead to hemorrhages. So-called 'black and blue' marks, which appear as bruises after traumatic injury, are due to subdermal bleeding. In thrombocytopenia a form of excessive hemorrhage termed 'thrombocytopenic purpura' can result in extensive bleeding and can be fatal. Chemicals and drugs that can cause aplastic anemia, such as benzene and a variety of anticancer alkylating agents, depress platelet production and can lead to purpura and other hemorrhages.

Platelets play a key role in the formation of the atherosclerotic plaque. High levels of blood cholesterol are associated with the accumulation of monocytes and macrophages in the vascular epithelium. They take up the cholesterol and become so fat-laden that they are termed foam cells. Platelets are attracted to the site, smooth muscle cells proliferate, and the plaque is formed. The result is occlusion of the lumen of the blood vessel and thrombosis.

3.8.4 Leucocytosis and Leukemia

The effects of many chemicals that impact on the bone marrow, such as benzene and anticancer alkylating agents, is to reduce the level of the various circulating white blood cells by interfering with the processes of differentiation and/or proliferation. Under some circumstances disturbances in differentiation and loss of control of proliferation lead to bone marrow cancers called leukemias.

Drugs and Chemicals may Produce Cytopenia by Impairment of Bone Marrow Function

Among organic chemicals, exposure to benzene in the workplace has been know for over a century to cause decreases in circulating erythrocytes, leucocytes, and/or thrombocytes as a result of chronic exposure. In the late 19th and early 20th century many cases of cytopenia and aplastic anemia were observed in factories where benzene exposure may have been as high as several hundred ppm. In Western Europe and the United States benzene exposure is now controlled at low levels, e.g. 1–5 ppm, but there is evidence of continuing high exposures in developing countries around the world. The mechanism of

benzene-induced aplastic anemia appears to involve both direct damage by benzene metabolites to bone marrow stroma, and inhibition of both differentiation and proliferation of developing cells in the bone marrow.

The drugs that most frequently lead to impairment of bone marrow function are the antineoplastic agents. Both antimetabolites and alkylating agents are intended to inhibit cellular reproduction in cancer cells. Bone marrow cells are ideal targets for chemicals that inhibit proliferation because of their high mitotic rate. Thus, alkylating agents (e.g. mechlorethamine, melphalan, cyclophosphamide, chlorambucil, etc.) can produce decreases in all of the circulating cell types by virtue of their effects on developing cells in the bone marrow. By the same token antimetabolites (e.g., methotrexate, thioguanine, mercaptopurine, fluorouracil, cytarabine, etc.), which impair DNA synthesis in neoplastic cells, can cause similar disruptions of cell replication in the bone marrow.

In the Netherlands between 1974 and 1994 approximately 40 drugs, drug classes, or drug combinations were cited where there was a link between treatment and neutropenia, most of which was agranulocytosis. Drugs most frequently reported to cause agranulocytosis were dipyrone, mianserin, salazosulfapyridine, co-trimoxazole (a combination of trimethoprim and sulfamethoxazole), penicillins, cimetidine, thiouracils, phenylbutazone, and penicillamine.

Historically, many other drugs have been claimed to cause bone marrow depression leading to aplastic anemia, such as the antibiotic chloramphenicol, antithyroid thiourea derivatives such as propylthiouracil and methimazole, phenothiazines, nonsteroidal anti-inflammatory agents (NSAIDs) such as phenylbutazone, sulfonamides, and many other drugs. Depression of the marrow usually begins with some form of leucopenia, i.e. neutropenia, thrombocytopenia, etc., and progresses to agranulocytosis or aplastic anemia. Although these tend to be rare events for any drug, the highest incidences have been observed with chloramphenicol and phenylbutazone. A higher frequency of aplastic anemia has been reported when aspirin, penicillin, acetophenetidin, phenytoin, streptomycin, or sulfisoxazole was given with other drugs than when given alone.

Early forms of leucopenia or anemia can be reversed by withdrawing exposure to the offending agent. Some therapeutic intervention may be necessary. Damage sufficiently severe to result in aplastic anemia would require a bone marrow transplant to restore normal hematopoiesis.

1. Exposure to benzene and many alkylating agents may result in myelodysplasia, a syndrome characterized by abnormal bone marrow cell morphology and chromosome damage to bone marrow cells. Exposure to benzene frequently results in myelodysplastic syndrome (MDS). A similar response may be observed in patients who have been treated with alkylating antineoplastic agents. Among these a significant percentage goes on to acute myelogenous leukemia (see below).
2. Leucocytosis refers to increases in circulating mature leucocytes such as neutrophils. Nonneoplastic leucocytosis has been reported to result from treatment with steroids, beta-agonists, lithium, or tetracyclines.

Leukemias are cancers of the bone marrow manifested by excessive proliferation of transformed immature granulocytes or lymphocytes. Erythroleukemias have also been recorded. The most frequently observed forms of leukemias are referred to as acute or chronic myelogenous leukemia (granulocytic leukemias), acute or chronic lymphatic

leukemias, Hodgkin's disease, nonHodgkin's lymphoma, and multiple myeloma (a B cell leukemia arising from plasma cells.)

Direct associations between exposure to chemicals and leukemogenesis are difficult to discern. Some authorities suggest that if a chemical can cause bone marrow damage any form of leukemia may result. Others point to benzene or the alkylating agents where the evidence demonstrates that acute myeloid leukemia is the form most frequently observed.

Mechanisms by which Chemicals can Induce Leukemia

Leukemias are bone marrow cancers and mechanisms of carcinogenesis in other organs can be invoked to attempt to explain the mechanism of leukemogenesis. Thus, covalent binding of carcinogens or their biological reactive intermediates, e.g. anticancer alkylating agents, to DNA can be envisioned as a trigger mechanism for initiating a mutagenic event. Thus, mutations in the *ras* family of oncogenes have been shown to be associated with some types of leukemia. Inhibition of topoisomerase II has been shown to be an effective chemotherapeutic strategy for some cancers, but frequently results in subsequent leukemia, probably because when the enzyme is inhibited at the cleavable complex stage of the reaction the double-stranded break created by the enzyme fails to re-anneal and the broken DNA chain represents a mutation. DNA repair is a critical mechanism for maintaining the integrity of DNA. Normally, in the absence of DNA repair, damage to DNA triggers cellular apoptosis, which is a mechanism aimed at ensuring that mutated cells do not survive. If DNA repair fails and apoptosis is inhibited the stage is set for neoplastic transformation.

Epigenetic mechanisms of carcinogenesis have been proposed. These frequently involve hormone-mediated excessive cell proliferation. Protein-based carcinogenesis is more difficult to demonstrate because of the prevailing concept that there is no threshold dose for carcinogenesis. Demonstration of a protein-based mechanism of carcinogenesis is hindered because of the many potential protein targets for chemicals, which might influence control of cell physiology and proliferation.

Current studies of chemically induced leukemia are focused on the observation that remission of cancers following treatment for a variety of tumors with either alkylating agents or topoisomerase II inhibitors leads to so-called 'second cancer,' which is manifested as a form of acute myelogenous leukemia termed t-AML. Both treatments resulted in abnormalities in chromosomes 5 and/or 7. Patients with alkylating agent-associated t-AML developed myelodysplasia and experienced a latency period of 5–7 years between treatment and second cancer. Following topoisomerase II treatment the latency period was 1–3 years and no myelodysplasia was observed. Similar effects have not been observed with other forms of leukemia.

3.8.5 Summary

The anatomy of multi-organ animals requires a mechanism for the maintenance of homeostasis. The blood provides a system to maintain adequate cellular oxygen levels, body temperature, the pH of the cells of the body, etc. Furthermore, the cells of the body require anabolic nutrients and an avenue for the removal of the products of catabolism. The introduction of foreign agents, either chemical, physical, or biological, into the body requires routine or emergency response mechanisms, a prime example being the immune

system. Furthermore, the bone marrow provides a milieu in which the system can continually revitalize itself. Clearly, the diverse mechanisms responsible for protection against disease are essential for life. The hematopoietic system, however, is also prone to diseases having a variety of etiologies ranging from genetic impairments, nutritional deficiencies, traumatic injury, or the impact of chemicals or microorganisms. Cancers, i.e. leukemias, are the most insidious threat to life because all of the functions of the system that support the lives of the cells of the body are interrupted. Despite dramatic advances in the therapy of the leukemias, around the world they remain largely refractory to treatment and frequently fatal. The hematopoietic system has been the subject of intensive research for centuries. The discoveries of the circulatory system by William Harvey and the microscope by Anton van Leeuwenhoek in the 17th century, and of oxygen and its biological functions by Joseph Priestley and Antoine Lavoisier in the 18th century opened the door to the study of hematology, which continues to be an ever-expanding area of research into an understanding of the mechanisms by which the bone marrow and blood function in health and disease.

References

Hoffmann, R., Benz, E.J., Jr., Shattil, S.J., Furie, B., Cohen, H.J., Silberstein, L.E., and McGlave P. Hematology: Basic Principles and Practice. 3rd edition, Churchill Livingstone, New York, 2000.

Greim, H. and Deml, E. Toxikologie, Eine Einführung für Naturwissenschaftler and Mediziner, VCH Verlagsgemeinschaft mbH, Weinheim, 1996.

For Further Reading

Swanson, M. and Cook, R. Drugs, Chemicals and Blood Dyscrasias. Drug Intelligence Publications, Inc., Hamilton, Illinois, 1977.

van der Klauw, M.M., Wilson, J.H., and Stricker, B.N. Drug-Associated Agranulocytosis: 20 Years of Reporting in The Netherlands (1974–1994). *Am. J. Hematol.*, **57**, 206–211, 1998.

Snyder, R. Benzene and Leukemia, *Crit. Rev. Toxicol.*, **32**, 155–210, 2002.

Snyder, R. Overview of the Toxicology of Benzene. *J. Toxicol. Environ. Health Part A*, **61**, 339–346, 2000.

Larson, R.S. and Le Beau, M.M. Therapy Related Myeloid Leukemia: A Model for Leukemogenesis in Humans. *Chem.-Biol. Interact.,* **153–154**, 187–195, 2005.

3.9 Immunotoxicology

Carol R. Gardner and Debra L. Laskin

3.9.1 Introduction

The immune system consists of physical, chemical, and cellular barriers that function to recognize and destroy invading pathogens and other foreign substances, and to restore normal structure and function following injury. The innate immune system provides an immediate but nonspecific response, whereas

the specific immune system adapts its response to improve recognition and destruction of foreign substances, and is associated with the development of immunological memory. Immunotoxicology is the study of the adverse effects of xenobiotics on the immune system.

Immune protection is afforded by layered defenses of increasing specificity (Figure 3.35). Physical barriers such as the skin, considered the largest immune organ in the body, and chemical barriers including hydrolytic enzymes present in the gastrointestinal and respiratory tracks, provide the first lines of defense. If a pathogen or other foreign substance breaches these barriers, cells of the innate immune system including macrophages and granulocytes provide an immediate but nonspecific secondary response. Innate immune systems are found in all plants and animals and are highly conserved in evolution. Vertebrates, however, are unique in that they also possess another level of defense; the adaptive or specific immune system. This arm of the immune system adapts its response to improve recognition and destruction of foreign substances, and is associated with the development of immunological memory. This allows the adaptive immune system to mount faster and stronger attacks each time a foreign substance is encountered. The adaptive immune system is composed of highly specialized cells (B and T lymphocytes) that are activated by the innate immune system.

Adverse effects of xenobiotics on the immune system may take the form of immunosuppression or immunostimulation. The consequences of immunosuppression include increased incidence of infections and cancer, and delayed or aberrant wound repair. Adverse effects related to immunostimulation encompass hypersensitivity and allergic reactions, chronic inflammation, and autoimmune diseases. The immune system consists of numerous dynamic cell populations undergoing continuous adaptation, differentiation and proliferation. Because host immunocompetence requires

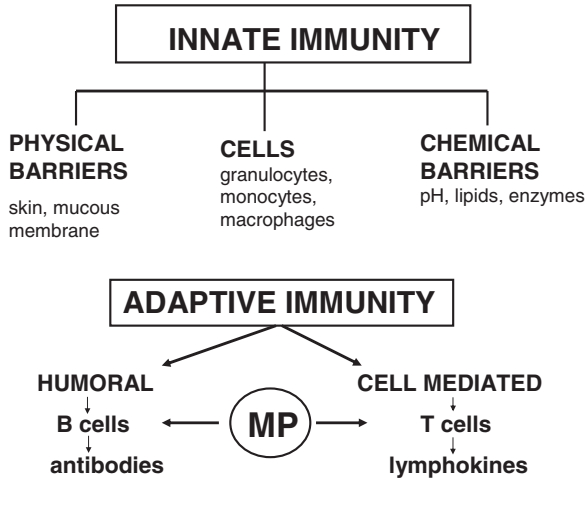


Figure 3.35 *Vertebrate immune system organization.*

coordinated and balanced functioning of the individual cell populations, the immune system is particularly sensitive to the adverse effects of chemical and physical agents. In this chapter the functioning of the various components of the immune system is reviewed, as well as the consequences of altered immune system functioning. Methods for assessing the response of these cells to xenobiotics are also discussed.

3.9.2 Anatomical Organization of the Immune System

Lymphatic System

The lymphatic system is a complex network of lymphoid organs, lymph nodes, lymph ducts, and lymph vessels that produce and transport lymph fluid from tissues to the circulatory system. The lymphatic system functions to remove excess of fluids, certain proteins, and bacteria from body tissues, to transport fat, and to generate mature lymphocytes, monocytes, and antibody-producing plasma cells.

Lymphatic organs include the thymus, spleen, lymph nodes, Peyer's patches, tonsils, and the bone marrow. These organs contain a scaffolding that support circulating B and T lymphocytes, as well as macrophages and dendritic cells. Pathogens or other foreign substances (antigens) that gain access to the body are transported from the tissue to the lymph, which is carried in the lymph vessels to regional lymph nodes. The lymph nodes filter the lymph fluid and remove foreign materials. Macrophages and dendritic cells phagocytize these materials, process them, and present them to antigen-specific lymphocytes. This leads to enlargement of the lymph nodes as additional immune cells are generated to help destroy antigens.

Reticuloendothelial System (RES)

The RES consists of phagocytic monocytes circulating in the blood, and macrophages located in the sinusoids of the spleen, lymph nodes, and bone marrow. Kupffer cells of the liver, alveolar and interstitial macrophages in the lung, and tissue histiocytes are also part of the RES. These cells are collectively known as mononuclear phagocytes.

Mononuclear phagocytes are responsible for engulfing (phagocytosis) and removing cellular debris, old cells, pathogens, and foreign substances from the blood and the lymph. Mononuclear phagocytes are derived from bone marrow precursor cells. In response to specific growth factors (e.g., GM-CSF and M-CSF), they develop into monocytes, which are released into the bloodstream. Most monocytes enter body tissues, developing into much larger and long-lived macrophages. Some mononuclear phagocytes roam through the body and can coalesce into a single cell around a foreign object in order to destroy it.

Cellular Effectors of the Immune System

Innate Immune System

The major functions of the innate immune system include the recruitment of inflammatory cells (e.g. neutrophils and macrophages) to sites of infection and injury, identification and removal of foreign substances, and activation of the adaptive immune system through a process known as antigen processing and presentation.

Cellular effectors that mediate innate immunity are neutrophils and macrophages. These cells have the ability to phagocytize microorganisms and other foreign substances, and to release soluble mediators that can directly destroy invading pathogens. Mediators released by neutrophils and macrophages are also important in recruiting other immune cells to sites of injury and infection.

Neutrophils, along with two other types of white cells found in the blood, **eosinophils** and **basophils**, are known as **granulocytes** or polymorphonuclear cells (PMNs), due to their distinctive lobed nuclei and the presence of intracellular granules. Neutrophils are the most abundant type of phagocyte, normally representing 50–60% of the total circulating leukocytes, and are usually the first cells to arrive at the site of injury or infection. They respond to chemotactic factors released by bacteria, and damaged cells and/or tissues. Neutrophils destroy foreign substances or pathogens by activating a 'respiratory burst' and releasing reactive oxygen intermediates, and by releasing cytotoxic lysosomal enzymes.

Macrophages are phagocytic leukocytes derived from blood monocytes and myelocytic bone marrow precursors. Once localized in tissues, these cells develop specialized functions depending on the needs of the tissue. Thus in the liver, resident macrophages or Kupffer cells develop a high phagocytic capacity, while in the lung, alveolar macrophages develop the capacity to release large quantities of highly reactive cytotoxic oxidants. Macrophages contribute to a number of innate immune responses. As scavengers, they rid the body of worn-out cells and other debris through the process of phagocytosis. They are also one of the most active secretory cells in the body releasing a vast array of mediators that regulate all aspects of host defense, inflammation, and homeostasis including enzymes, complement proteins, cytokines, growth factors, eicosanoids, and oxidants (Table 3.19). Macrophages also possess receptors for cytokines that

Table 3.19 *Macrophage-derived mediators.*

Proteolytic enzymes (e.g. lysozyme, neutral proteases, acid hydrolases)
Endogenous pyrogen (IL-1)
Binding proteins (e.g. transferrin, fibronectin)
Enzyme inhibitors (e.g. alpha-2-macroglobulin, plasmin inhibitor)
Reactive oxygen intermediates (e.g. superoxide, hydrogen peroxide, hydroxyl radical)
Reactive nitrogen intermediates (e.g. nitric oxide, peroxynitrite)
Bioactive lipids (e.g. platelet-activating factor, prostaglandins, leukotrienes, thromboxanes)
Chemotactic factors [e.g. leukotriene B4 (LTB4), interleukin-8 (IL-8)]
Cytokines and growth factors [e.g. interleukin-1 (IL-1), interleukin-6 (IL-6), tumor-necrosis factor (TNF), fibroblast growth factor (FGF), colony-stimulating factor (CSF)]

allow them to be activated and respond more robustly to microbes and tumor cells. Macrophages also participate in adaptive immunity. They are considered professional antigen presenting cells, one of the major cell types involved in initiating specific immune responses of T lymphocytes (see further below).

Adaptive Immune System

The cells of the adaptive immune system include two major types of lymphocytes: B cells and T cells. Whereas the majority of mature T cells are found circulating in the blood, most B cells are localized in lymphatic organs including the spleen, lymph nodes, and bone marrow. B cells exert their biological activity by releasing antibodies or immunoglobulins, which are found in blood plasma and lymph and are responsible for humoral immune responses. In contrast, T cells release soluble protein mediators (lymphokines and cytokines) and function in cell-mediated immune responses.

B cells are named for the bursa of Fabricius, an organ unique to birds, where they were found to develop. However, in nearly all other vertebrates, B cells, as well as T cells, are derived from a common pluripotential stem cell population in the bone marrow. B cell maturation in humans is thought to occur within the bone marrow or in lymphoid organs. During this process, a unique antigen-specific receptor is inserted into the cell membrane of the B cell. These receptors are immobilized antibody molecules (IgM or IgD) that recognize and bind only one specific antigen. After a B cell encounters its cognate (specific) antigen, and it receives additional signals from a helper T cell, it differentiates into a plasma cell that synthesizes antigen-specific antibody, or a B memory cell that remembers it has encountered the antigen. Memory B cells are long-lived, and respond quickly following a second exposure to the same antigen (secondary immune response).

Antibodies or immunoglobulins (Ig) are large Y-shaped glycoproteins that recognize and neutralize foreign substances or antigens. In mammals five types of antibody have been identified: IgA, IgD, IgE, IgG, and IgM, which differ in their physical, chemical, and biological properties (Table 3.20).

IgM is the most complex, consisting of five Y-shaped glycoproteins linked together in a pentamer. It is generated by B cells during a primary immune response; that is the first time a B cell encounters an antigen. IgG is the most common and simplest of the

Table 3.20 *Classes of antibody molecules.*

IgM: involved in primary immune response and Type III (immune complex) hypersensitivity reaction; B cell receptor
IgG: involved in secondary immune response and Type II hypersensitivity reaction; produced mainly by B memory cells
IgA: found in external secretions; produced locally against bacteria and viruses
IgE: involved in Type I hypersensitivity (allergic) reactions
IgD: found in umbilical cord blood; may be important in primitive recognition or regulation; B cell receptor

antibody isotypes. It is produced in large quantities for relatively long periods of time by B memory cells during a secondary immune response. Owing to its complex structure, IgM can bind large amounts of antigen. However, these large antibody–antigen complexes are difficult to remove by the body. More efficient antigen destruction occurs when the antigen is bound to IgG. These relatively small antigen–antibody complexes are readily phaocytized and destroyed by macrophages or lysed by a serum protein known as complement.

The second major type of lymphocyte mediating adaptive immunity is the **T cell or thymocyte**, which plays a central role in cell-mediated immune responses. T cells orchestrate the immune system's response to virally infected cells, to malignant cells, to transplanted cells, and to somatic cells that are recognized as 'nonself.' T cells can be distinguished from other lymphocytes by the presence of a unique cell-surface T-cell receptor. Several different subsets of T cells have been identified, each with a distinct function. These include helper T cells, cytotoxic T cells, and regulatory T cells.

Helper T cells (T_h cells) secrete cytokines that stimulate immune responses. They are identified by a cell-surface glycoprotein called CD4. T_h cells express T-cell receptors that recognize antigen bound to Class II major histocompatability complex (MHC) molecules (genetically determined self markers that allow immune cells to communicate). When T_h cells become activated, they release cytokines that stimulate the activity of antigen-specific B cells, cytotoxic and regulatory T cells, and macrophages. Two types of T_h cells have been identified, T_h1 and T_h2 cells, which are involved in elimination of different types of pathogens. Whereas T_h1 cells release cytokines like interferon-gamma and interleukin-2 (IL-2), which function to activate macrophages and induce cell-mediated immunity, T_h2 cells produce cytokines such as interleukin-4 (IL-4), which activates B cells and humoral immunity. In general, T_h1 responses are more effective against intracellular pathogens (viruses and bacteria that are inside host cells), while T_h2 responses are more effective in eliminating extracellular bacteria and parasites, and other foreign substances.

Cytotoxic T cells (T_{cyt}) destroy virally infected cells and tumor cells, and play a key role in transplant rejection. They express CD8 glycoprotein on their surface. T_{cyt} cells are activated when their T-cell receptor interacts with an antigen bound to MHC class I molecule (protein recognized as 'nonself'). Once activated, T_{cyt} cells undergo clonal expansion during which time they gain functionality and divide rapidly, to produce 'armed'-effector cells. Activated T_{cyt} cells then travel throughout the body in search of cells bearing the specific antigen-MHC Class I marker. When exposed to these infected, dysfunctional, or foreign cells, effector T_{cyt} cells release cytotoxic molecules like perforin and granulysin which induce pores in target-cell membranes, causing them to lyse, as well as proteases which induce apoptosis.

A third group of T lymphocytes are **regulatory T cells** (T_{reg}) or suppressor T cells. These cells are essential for maintenance of immunological tolerance. Their major role is to turn off T cell-mediated immunity towards the end of an immune reaction. Typically, these cells can be distinguished from other T cells by the presence of an intracellular molecule called FOXP3. The majority of T_{reg} cells also express CD4.

T cell development begins in the bone marrow as they develop from pluripotent hemopoietic stem cells. Once released from the bone marrow, T cells travel to the thymus where they mature and become 'educated.' This process involves T cells learning to

recognize specific antigens and to distinguish between 'self' and 'nonself.' This is accomplished through the acquisition of cell-surface CD4 or CD8 molecules and specific T-cell receptors. The distinction between self and nonself is essential for immunologic tolerance and for recognition of foreign antigens by T cells. It is controlled by a group of genes called the MHC, which regulate expression of surface-tissue compatibility proteins. These proteins are the principal determinants of tissue type and transplant compatibility. MHC proteins are divided into three subgroups, each coded for by a distinct group of genes; these are MHC class I and MHC class II proteins, which direct T cell responses, and MHC class III proteins, which include immune mediators such as complement. MHC class I proteins are present on almost all nucleated cells and they give individuals their genetically unique immune identity. These are the classical 'transplantation' antigens. MHC class II proteins are found only on a few specialized cell types (e.g. macrophages, dendritic cells, and some B cells), which have the capacity to act as antigen-presenting cells (APCs), and they are important in immune cell communication. APCs recognize and internalize antigens, process them, and then re-express them on their surface along with MHC class II proteins. These processed antigens are then presented to T_h cells. T cell recognition of antigen and activation is MHC-dependent. Thus, as described above, T_h are only activated after binding to processed antigen bound to MHC class II protein, together with CD4, while T_{cyt} cells are activated after binding a processed antigen bound to MHC class I protein, together with CD8.

Antigens and Antigenicity

The immune system is constantly being challenged by a wide variety of substances that it recognizes as foreign or 'nonself.' These substances are termed antigens and include infectious agents, such as viruses, bacteria, fungi, or parasites, as well as noninfectious materials, including drugs and chemicals, pollen, certain foods, vaccines, venoms, and transplanted tissues.

An antigen is a molecule or complex of molecules that stimulates an immune response. Antigens are usually proteins or polysaccharides, but can be any type of molecule, including small proteins (haptens) coupled to a carrier-protein. Antigenicity is the molecule's innate capacity to react with immune cells or antibody, and is determined by the chemical structure of the antigen molecule. The specific part of the antigen that binds to antibody is the antigenic determinant or epitope.

3.9.3 Effects of Toxicants on the Immune System

Immunotoxic consequences of exposure to xenobiotics include immunosuppression, immunostimulation, hypersensitivity, and autoimmunity. The effects of a toxicant may be specific for a particular cell type within the immune system, or they may be more generalized. According to the United States Food and Drug Administration, immunotoxicity refers to any adverse effect on the structure or function of the immune system.

Immunosuppression

The major adverse consequences of immunosuppression include increased incidence of infection and virally induced malignancies. A number of agents possess general immunosuppressive activity. These include pharmaceuticals (cyclosporine, corticosteroids, chemotherapeutic agents such as cyclophosphamide and methotrexate), environmental/industrial chemicals (glycol ethers, organic solvents, polycyclic aromatic hydrocarbons, halogenated aromatic hydrocarbons, heavy metals, pesticides and some air pollutants), agents of abuse (cannabinoids, opiates, ethanol, tobacco smoke), and physical agents (ultraviolet light, ionizing radiation, physical stress). These diverse agents suppress the immune system by distinct mechanisms. Thus, whereas corticosteroids induce lymphopenia and modulate macrophage gene transcription, thereby effectively suppressing inflammation, dioxin induces thymic atrophy and suppresses lymphocyte proliferation. Other agents, like cyclophosphamide, methotrexate, and ionizing radiation inhibit the immune system by suppressing bone marrow stem cell proliferation. In contrast, ultraviolet radiation causes immunosuppression by inducing the production of interleukin-10, a cytokine known to inhibit the immune response of macrophages. The general response to all of these agents is decreased host resistance.

Immunostimulation

A variety of agents have been identified that stimulate the immune system by inducing activation or increasing the activity of its cellular components. These may be specific agents, which stimulate only one cell population and can induce antigenic specificity in the immune response, or nonspecific immunostimulators, which globally augment immune responses. Examples of immunostimulants include cytokines like interleukin-2 (IL-2), which induces T cell proliferation, and interferon-gamma, which activates macrophages. Vaccines are examples of agents that induce antigen-specific immunostimulation. The consequences of this type of immunostimulation include flu-like symptoms such as fever, malaise, and chills. More general immunostimulants include adjuvants (e.g., lipolysaccharide, muramylpeptides), levamisole, somatotropin, and mitogens [e.g. colony-stimulating factor (CSF)].

Hypersensitivity

Another potential consequence of xenobiotic-induced overactivity of the immune system is induction of hypersensitivity reactions. These reactions are considered to be the most widely recognized manifestations of exposure to toxins. They are undesirable (damaging, irritating, and sometimes fatal) reactions produced by the normal immune system. All hypersensitivity reactions require a presensitized (pre-exposed) state of the host. They most commonly occur in the lungs, skin, and gastrointestinal tract. In the lung, they are manifest as bronchial asthma (dusts, vapors, gases), hypersensitivity pneumonitis (organic dusts, inorganic chemicals), or cell-mediated delayed-type hypersensitivity (beryllium), while in the skin, they appear as allergic contact dermatitis (poison ivy).

Hypersensitivity reactions are divided into four types based on the mechanisms involved and time for the reaction to occur (Table 3.21).

Table 3.21 *Comparison of different types of hypersensitivity reactions.*

	Type I anaphylactic	Type II cytotoxic	Type III immune complex	Type IV delayed type
Antibody	IgE	IgG, IgM	IgG, IgM	None
Antigen	Exogenous	Cell surface	Soluble	Tissues and cells
Response time	15–30 min	min–h	3–8 h	48–72 h
Appearance	Weal, flare	Lysis, necrosis	Erythema, edema	erythema, edema
Cells/mediator	Basophils, eosinophils	Antibody, complement	Complement, neutrophils	Macrophages, T lymphocytes
Example	Allergic asthma, allergic contact dermatitis	Farmer's lung disease	Immune complex disease, serum sickness, Arthus reaction, acute hypersensitivity pneumonitis	Tuberculin test, chronic hypersensitivity pneumonitis
Precipitating agent	Pollen, latex,	Fungi	Cefaclor	Beryllium

Type I hypersensitivity is also known as immediate or anaphylactic hypersensitivity. The reaction, which typically occurs within 15–30 min of exposure to the antigen, may involve the skin (urticaria and eczema), eyes (conjunctivitis), nasopharynx (rhinorrhea, rhinitis), bronchopulmonary tissues (asthma), and/or gastrointestinal tract (gastroenteritis) and can cause a corresponding range of symptoms. Type I hypersensitivity reactions are mediated by IgE, which binds to sensitized basophils (blood) or mast cells (tissue), causing them to release vasoactive mediators including histamine, leukotrienes, and platelet-activating factor (PAF). This type of reaction has been observed after exposure to latex, peanuts, pollen, isocyanates, and metals like cobalt and nickel.

Type II hypersensitivity, also known as cytotoxic hypersensitivity, may affect a variety of organs and tissues. The antigens are normally endogenous, although exogenous chemical haptens that attach to cell membranes can also lead to Type II hypersensitivity. Drug-induced hemolytic anemia, granulocytopenia, and thrombocytopenia are examples. Farmer's lung or hypersensitivity pneumonitis is also an example of a Type II reaction. It is induced by inhalation of biologic dusts (e.g. fungi, actinomyces) arising from moldy hay or other agricultural products. The reaction time is minutes to hours. Type II hypersensitivity is primarily mediated by antibodies of the IgM or IgG classes and complement.

Type III hypersensitivity or immune complex hypersensitivity is observed 3–10 h after exposure to an antigen and is mediated by soluble immune complexes (aggregations of antigens and IgG and IgM antibodies). The reaction may be general (e.g. serum sickness) or may involve specific organs including the skin (e.g. systemic lupus erythematosus, Arthus reaction), kidneys (e.g. lupus nephritis), lungs (e.g. aspergillosis), blood vessels (e.g. polyarteritis), joints (e.g. rheumatoid arthritis), or other organs. Tissue damage is caused by platelets, neutrophils, and complement.

Delayed or **Type IV hypersensitivity** is involved in the pathogenesis of many autoimmune and infectious diseases (e.g. tuberculosis, leprosy, blastomycosis, histoplasmosis, toxoplasmosis, leishmaniasis). The reaction typically occurs 48–72 h following exposure of a sensitized host to the antigen. Damage is due to overactivity of T lymphocytes and/or macrophages, and excessive production of cytokines including monocyte chemotactic factor, IL-2, interferon-gamma, and TNF-alpha. One example of delayed-type hypersensitivity is allergic contact dermatitis, which has been observed following exposure to acids, alkalis, resins, and other chemicals frequently found in drugs, plants, preservatives, and cosmetics, as well as to metals (e.g. beryllium, chromium, mercury, nickel). Typically these agents act as small-molecular-weight haptens that form a complete antigen by covalently binding to proteins in the skin. The complex is then taken up by dendritic phagocytes called Langerhans cells for processing. Langerhans cells subsequently migrate to regional lymph nodes to present the processed antigen to T lymphocytes, which then become 'sensitized.' Subsequent exposure to the antigen leads to a macrophage-mediated inflammatory response. The best known example of an allergic contact hypersensitivity reaction is chronic eczema in masons working with cement treated with potassium dichromate.

In cases of allergic contact dermatitis, also termed skin sensitization, exposure to only a few molecules can produce an intense skin reaction. The underlying pathobiological mechanism is a cell-mediated type IV hypersensitivity reaction.

Autoimmunity

Overactivity of the immune system can also lead to autoimmune diseases. This represents a failure of an organism to recognize its own cells and tissues as 'self.' This leads to an inappropriate immune response against self antigens. While genetic factors contribute to some types of autoimmune disease (e.g. systemic lupus erythematosus; rheumatoid arthritis), others are induced by drugs such as hydralazine, procainamide, isoniazid, and phenytoin, or chemicals, such as trichloroethylene and mercury. These agents induce a lupus-like syndrome in which the ratio of T_h cells to T_{reg} cells increases; this leads to increased B cell activity and the production of autoantibodies. Drugs like tienilic acid and troglitazone have also been shown to induce autoimmune disease by altering self antigen, specifically in the liver. This is associated with hepatitis and in some instances, acute liver failure.

3.9.4 Procedures for Assessing Immunotoxicity

A classic tier approach can be used for assessing immunotoxicity. This involves first evaluating gross pathologic changes in immune organs followed by assessment of specific immune cells and their functional activity. Immunotoxic effects that are typically assessed include immunosuppression, immunostimulation, hypersensitivity, and autoimmunity.

Immunosuppression and Immunostimulation

Immunopathology Immunotoxicologic evaluation of a xenobiotic is most often performed in rodents. Immunopathology predominantly involves assessing necrotic, inflammatory, and autoimmune effects. Following exposure of animals to the test agent, hematology profile (total and absolute differential leukocyte counts), clinical chemistry (globulin levels), lymphoid organ weight, histology, and gross pathology are assessed. Frank lesions and other evidence of injury (e.g. necrosis, apoptosis, fibrosis, and inflammation) are also evaluated. Immunohistochemistry, cell-surface marker expression on isolated peripheral blood white cells, quantitative microscopy, and toxicogenomics can also be used to improve the sensitivity of the evaluation.

Host-resistance Assays General immunosuppressive effects of xenobiotics are assessed using host-resistance assays. These assays are particularly useful for determining if modest histopathologic changes (e.g. small decreases in circulating leukocytes) or small changes in immune function result in clinical impairment such as increased susceptibility to an infectious or tumor cell challenge. Host-resistance assays are also useful as screening tools during the early phases of drug discovery. Typically, a pathogen is used that recruits multiple immune cell lineages to resist the infection. More specific immune function tests can be applied subsequently if information on the specific mechanism of immune suppression is required. A wide variety of bacteria, viruses, and tumor models have been used in rodent models. In these assays, animals are exposed to the potential immunosuppressive agent and then challenged with an infectious agent or with tumor cells. Typical measurements include analysis of the survival rate of the animals over a two-week period. Latency and frequency of infection or tumor development are also measured.

Assays for Immune Cell Functioning

Measurement of humoral immunity T-dependent antibody formation is a screening assay for evaluating overall immune competence; thus it provides information not only on B-cell function, but also on T-cell function, and antigen processing and presentation. In this assay, animals are injected with an antigen such as sheep erythrocytes and sacrificed five days later. Isolated spleen cells are then incubated with sheep erythrocytes and complement on an agar plate. Splenocytes that produce anti-erythrocyte antibodies form hemolytic plaques that can be quantified. Animals that are immunosuppressed because of exposure to a xenobiotic have reduced plaque formation. Another T-dependent antigen that can be used is keyhole limpet hemocyanin (KLH). The specific IgG or IgM produced is then measured by an enzyme-linked immunosorbant assay (ELISA). Additional information on the status of humoral immunity can be obtained by analysis of lymphocyte proliferation in response to B-cell mitogens (e.g. pokeweed mitogen) and expression of various cell-surface markers (e.g. CD20+/CD3−, MHC class II) typically using techniques in immunofluorescence and flow cytometry.

Measurement of cell-mediated immunity To evaluate the effects of a xenobiotic on cell-mediated immunity, specific T-cell function assays can be performed. T cells undergo clonal proliferation after stimulation with specific antigens (e.g., tetanus toxoid), and this can be measured in culture. The proliferative response of lymphocytes to the T-cell mitogens phytohemaglutinin (PHA) or concanavalin A (Con A) or to activation of the

T-cell receptor with anti-CD3 antibody can also be analysed. Cell-mediated immunity can also be assessed using a cytotoxic CD8+ T-lymphocyte assay. In this assay splenocytes from animals treated with a potential immunotoxic agent are co-cultured with mitomycin C-inactivated histoincompatable target cells. Cytotoxicity is quantified based on the release of radioactivity from labeled target cells. Delayed-type hypersensitivity assays are used to measure cell-mediated immunity *in vivo*. This assay provides information on T_h and T_{cyt} activity. Test animals are injected in the foot pad with a sensitizing dose of a known T-dependent antigen (e.g. KLH, dinitrochlorobenzene, or oxazolone) and then with a challenge dose at another site. The response is evaluated by foot pad swelling.

Measurement of nonspecific immunity The activity of macrophages and/or neutrophils is measured as an indicator of nonspecific immunity. Assays of phagocytosis, chemotaxis, or the generation of reactive oxygen intermediates are used, as well as antigen-presenting activity. Release of cytokines such as tumor necrosis factor (TNF)-alpha, IL-1, and transforming growth factor-beta can also be measured.

NK cells are a subclass of lymphocytes involved in nonspecific immunity. NK cells have been shown to play a critical role in the defense against infections and neoplasia. NK cells are enumerated based on surface markers (e.g. CD56) and their cytotoxicity activity assessed *in vitro* using radiolabeled target cells or by flow cytometry. NK cell assays have been reported to be highly sensitive to modulation by toxic substances. However, the clinical relevance of alterations in NK activity has yet to be determined.

Assays for Hypersensitivity and Autoimmunity Assessment of hypersensitivity reactions generally involves analysis of hematology, serum IgE, IgG and C reactive protein levels. Skin tests may also be used to detect IgE-mediated immediate hypersensitivity reactions to allergens as in a TB test. Autoimmune diseases are studied using various genetic rodent models. Serum levels of autoantibodies (e.g. antinuclear antibody, antithyroglobulin antibody, rheumatoid factor) or T_{cyt} cells in the blood are indicative of autoimmune disease.

3.9.5 Summary

The immune system consists of a complex network of cells and mediators that interact to protect the host. Although derived from common precursors within the bone marrow, these cells differentiate and mature into unique cellular effectors that have the capacity to respond to pathogens, transplanted tissues, and other foreign materials including drugs and chemicals (Table 3.22). This response may be highly specific for one particular antigen, and involve unique receptors as in the case of T cells and B cells, or it may be more generalized, as observed with neutrophils and macrophages. Appropriate functioning of these two arms of the immune system is essential for appropriate host defense.

While redundancy in function is well established within the immune system, even minor perturbations in the activity or response of one of the cellular effectors can have major impact on overall host immunocompetence. Clinical and epidemiologic data demonstrate that aberrations in immune-system functioning, including both immunosuppression, and immunopotentiation, are associated with significant morbidity and even mortality. Thus, evaluation of the potential untoward effects of drugs, chemicals, and environmental agents on the immune system is essential. A number of *in vivo* and *in vitro* test systems have been developed to assess the overall immune status of the host, as well

Table 3.22 *Cellular effectors of the innate and adaptive immune systems.*

Type of immunity	Cellular effectors (location)	Function
Innate immunity	Granulocytes Neutrophils (blood) Eosinophils (blood) Basophils (blood) Mast Cells(tissue)	Inflammation/allergic reactions
	Monocytes (blood)	Inflammation
	Macrophages (localized in tissues): Langerhans cells (skin) Alveolar macrophages (lung) Kupffer cells (liver) Microglia (brain)	Inflammation, antigen processing/ presentation, tumor surveillance
	NK cells (CD56+) (blood and tissue)	Cytotoxic lymphocytes; kill tumor cells and virally infected cells
Adaptive immunity	T lymphocytes (mostly blood)	Cell-mediated immunity
	T-helper (CD4+)	Stimulate responses
	T-regulatory (FoxP3+)	Suppress responses
	T-cytotoxic (CD8+)	Destroy antigens
	B lymphocytes (spleen and lymph nodes)	Humoral immunity; produce antigen-specific antibodies

as the activity of individual cellular immune effectors. Using these systems it is also possible to elucidate mechanisms mediating the immunotoxicity of xenobiotics. This is important since it may lead to the development of new therapeutic approaches to preventing tissue injury and limiting morbidity.

Further Reading

Carfi M., Gennari A., Maleba I., Corsini E., Pallardy M., Pieters R., Van Loveren H., Vohr H.W., Hartung T., and Gribaldo L., *In vitro* Tests to Evaluate Immunotoxicity: A Preliminary Study, *Toxicology*, **229**, 11–22 (2007).

Descotes J., Methods of Evaluating Immunotoxicity, *Expert Opin.*, **2**, 249–259 (2006).

Descotes J., Immunotoxicology. Role in the Safety Assessment of Drugs, *Drug Safety*, **28**, 127–136 (2005).

Inadera H., The Immune System as a Target for Environmental Chemicals: Xenoestrogens and Other Compounds, *Toxicol. Lett.*, **164**, 191–206 (2006).

Kimber I. and Dearman R.J., Immune Responses: Adverse versus Non-adverse Effects, *Toxicol. Pathol.*, **30**, 54–58 (2002).

Lawrence A., Toxicology of the Immune System, in *Comprehensive Toxicology, Volume 5*, I.G. Sipes, C.A. McQueen, and A.J. Gandolfi (Eds), Elsevier Science Inc., New York (1997).

Luebke R., House R., and Kimber I., *Immunotoxicology and Immunopharmacology*, CRC Press, New York (2007).

Luster M.I. and Rosenthal G.J., Chemical Agents and the Immune Response, *Environ Health Perspect.*, **100**, 219–236 (1993).

Luster M.I., Simeonova P., Gallucci R., Matheson J., Yucesoy B., and Sugawara T., Overview of Immunotoxicology and Current Applications to Respiratory Diseases, *Immunopharmacology*, **48**, 311–313 (2000).

3.10 The Eye

Ines M. Lanzl

3.10.1 Introduction

The eye and its adnexae represent a highly organized, well protected system, designed to gather and process visual information. The information processed within the rather small organ globe and along the visual system is important enough to occupy the largest portion of all sensory input in the human cortex. Considering its complex formation during fetal development, its structure and function bear resemblances to other differentiated organs such as skin, kidney, and brain, but it has its own unique functions. All these are necessary to understand how the eye is affected by xenobiotics.

3.10.2 Structure and Function of the Eye

The eye is protected from the environment by the lids and tear film. The eye itself consists of three major segments: anterior segment (conjunctiva, cornea, iris, lens), posterior segment (sclera, choroid, retina, vitreous body), and the optic nerve (Figure 3.36). The outer anteroposterior diameter of an adult's globe measures about 24 mm.

The Lids

The outermost layer of the eye is created by the lids, which serve as a mechanical barrier to the outside and as a constant means of producing and spreading the tear film by blinking. The lids comprise skin, muscle, sebaceous glands on the outside and the tarsal plate, conjunctiva, and sebaceous and accessory lacrimal glands on the inside.

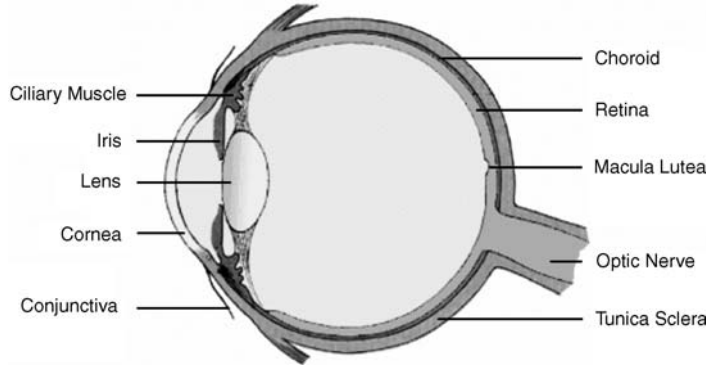


Figure 3.36 *Cross-section of the eye (courtesy of Chibret, Haar, Germany). [Reprinted from http://www.bebig.com, with permission from Echert & Ziegler BEB16].*

The Anterior Segment

The Tear Film The tear film consists of an outer layer of lipid produced by the sebaceous glands of the eyelid and caruncle, a middle layer of aqueous fluid produced by the lacrimal and accessory lacrimal glands, and an inner layer of mucoprotein produced by conjunctival goblet cells. The conjunctival surface of the eyelids is adjacent to and in contact with the tear film layer. During blinking, this layer is distributed over the surface of the corneal epithelium, continually rewetting the entire surface.

The Conjunctiva The conjunctiva is a mucous membrane that lines the entire exposed surface of the eye from the lid margin up to the corneal limbus. In addition to acting as a physical barrier, it houses immune cells and its substantia propria is highly vascular. Its goblet cells produce mucin, which adsorbs to the glycoproteins coating the microvilli of corneal and conjunctival epithelial cells. This precorneal and preconjunctival mucin layer merges gradually with the overlying aqueous tear film to ensure complete wetting of the ocular surface.

The Corneal Limbus The conjunctival and corneal epithelia merge at the limbus, where the subtle transition from the nonkeratinized, stratified columnar epithelium of the conjunctiva to the nonkeratinized, stratified squamous epithelium of the cornea occurs. These two epithelial surfaces together cover the exposed portions of the eye and function as a barrier to chemical insults and to invasion of microorganisms.

The limbal region is of utmost importance for the maintenance of corneal transparency. First intact conjunctival vascular arcades are necessary for nutrition of the anterior and peripheral parts of the cornea. The area also hosts corneal epithelial stem cells, which repopulate the corneal epithelium in a centripetal fashion.

If the limbus region is not intact, conjunctivalization of the corneal epithelium occurs, followed by opacification because of the following mechanism:

The conjunctival epithelium provides a source of cells to repopulate the corneal surface when the entire corneal epithelium has been denuded and the limbal stem cells have been destroyed by severe chemical injuries. After complete re-epithelialization of the cornea with conjunctival epithelium the structure of these cells changes by the process of transdifferentiation, and the histology and biochemical function of the new epithelial surface increasingly resembles that of corneal epithelium. However, in cases of severe chemical injury, vascularization of the tissue occurs, with conjunctivalization of the corneal surface and thus loss of transparency.

The Cornea The cornea is a transparent and avascular tissue, which forms the major refracting surface of the eye in conjunction with the tear film.

The corneal epithelium consists of several cell layers, is capable of rapid renewal, and needs to be intact for the regular spread of the tear film. If not intact it leaves terminal nerve endings free, which induce a feeling of severe pain and increase tearing. The corneal epithelium has the unique function of providing a smooth optical surface and the transparency necessary to transmit images with minimal distortion. The chemically injured corneal epithelium may desquamate, or become irregular and lose its clarity. When the integrity of the epithelium is compromised, exposure of the underlying stroma may result in an alteration of hydration that further compromises corneal transparency.

The corneal stroma is about 450 μm thick. The intricate parallel layering of its main constituents, collagen fibrils, ensures transparency. In the event of damage this property is lost.

The inner corneal lining comprises endothelial cells stemming from neuroectoderm and therefore unable to regenerate. Nutrition of the avascular cornea is provided mostly by the aqueous humor produced by the endothelial cells of the anterior chamber and additionally by the tear film.

The Iris The iris is the anterior extension of the uvea that forms a mobile diaphragm between the posterior chamber and anterior chamber of the eye and is highly vascularized. The central aperture of the iris forms the pupil, through which light enters the eye.

The Anterior Chamber The anterior chamber represents an area between the posterior surface of the cornea, and the anterior surface of the iris and the pupillary portion of the lens, and peripherally by the trabecular meshwork, scleral spur, ciliary body, and iris root. These comprise the anterior chamber angle, where most of the aqueous humor contained in the anterior chamber is drained.

The Ciliary body The ciliary body has three basic functions: production and removal of aqueous humor, accommodation, and the production of vitreous mucopolysaccharides. The ciliary body represents the internal surface of the globe.

The Lens The lens is a transparent biconvex structure located behind the iris and the pupil and in front of the vitreous body. In adults the lens has a diameter of approximately 10 mm and is 4 mm thick. The lens is held in place and shaped according to the accommodative needs by thin zonular fibres radiating into the ciliary body.

The lens continues to accumulate cells throughout life, such that the central portion of the lens becomes less pliable, more compact, sclerosed, and yellow. The increased density eventually reduces accommodation (zooming from far to near objects) and visual acuity. In addition to this sclerosis, the lens can become opaque because of a variety of other factors, including disorganization of fibre membranes or of lens proteins and accumulation of colored, insoluble proteins in the lens as a result of numerous systemic or extrinsic factors.

The ciliary body processes, lens zonules, and posterior lens capsule represent the border to the posterior segment of the eye.

The Posterior Segment

The Sclera The sclera shields the posterior segment of the eye. It forms five-sixths of the outer tunic of the eye, consists of dense fibrous tissue, and contains openings and canals for the various vessels and nerves entering and exiting the globe. Its mechanical properties help contain the intraocular pressure and prevent deformations of the globe. It is relatively avascular and principally composed of collagen with few fibrocytes and ground substance. Externally it is white, whereas internally it is slightly brown because of the presence of pigmented melanocytes.

The Choroid The choroid is a highly vascularized and pigmented tissue and is the extension of the ciliary body up to the optic nerve. The choroid is light to dark brown in color and sponge-like in appearance. It mostly consists of blood vessels. Except for the

central retinal artery that supplies the inner retina the blood supply of the eye is derived from the ophthalmic artery. The main function of the choroid is to nourish the outer two-thirds of the retina.

The Retina The retina is a very thin and transparent membrane, with a surface area of approximately 266 mm^2. It is loosely attached to the choroid via the pigment epithelium. The retina consists of two distinct layers: the neurosensory retina and the retinal pigment epithelium. These two layers extend over the inner surface of the eye from the optic nerve head to the retinal pigment epithelium, which continues into the pigmented epithelium of the ciliary body. The neurosensory retina extends into the nonpigmented epithelium of the ciliary body. The macular region is the neuronal zone close to the optical nerve head and provides the highest visual acuity.

The Vitreous Body The vitreous body is a clear, transparent, gel-like substance caused by a matrix composed of small-diameter collagen fibrils in hyaluronic acid that fills the eye between the lens and retina and contains 99% water.

The Optic Nerve

The optic nerve head is located slightly nasally to the posterior pole of the eye. Here, all the layers of the retina terminate, with the exception of the innermost nerve fiber layer. The ganglion cells of the retinal nerve fiber layers extend all the way to the lateral geniculate in the brain. From there the visual information is projected into the visual cortex.

3.10.3 Routes of Delivery of Xenobiotics to the Eye

Toxic compounds can enter the eye by two routes. The external route has to overcome the mechanical barriers of the lids, conjunctiva, and/or cornea. The internal systemic route is via the blood stream. Since the special feature of the eye consists of its transparent media, light toxicity has to be considered as well.

Ocular Absorption

From the development of topical eye drops, we know that a substance cannot easily overcome the natural barriers of the anterior surface. Substances that come into contact with this surface may, however, penetrate the rather efficient barrier of the tear film and the cornea. The tear film is lipophilic on the outside and lipophobic on the inside due to its three-layer composition. The intact cornea presents a physical barrier to intraocular substance penetration. Active transport has not been demonstrated and substances appear to penetrate by diffusion. This process is characterized by the following factors: the rate of diffusion parallels drug concentration, the process is not saturable, is not retarded by metabolic inhibitors, and penetration will not occur against an electrochemical gradient.

Owing to the limited time that substances in the lacrimal fluid are in contact with the cornea, the first barrier to penetration, the corneal epithelium, is the most important. Cell-membrane lipids limit substance penetration. The greater the lipid solubility of a substance, the greater its diffusion into the corneal epithelium. The corneal stroma is relatively acellular and made up primarily of collagen. It contains only 1% lipids. As a result, substances need to be water soluble to penetrate.

The third component of the cornea, the endothelium, is one cell layer thick. It represents a lipid barrier.

In order to penetrate the cornea, substances need to be both lipid and water soluble, which is called the differential solubility concept.

Vasoconstrictors and vasodilators affect the blood flow in the conjunctiva and, by decreasing or increasing conjunctival absorption, alter both the amount and duration of substance in the tear film. These in turn will affect corneal and intraocular substance penetration.

Systemic Absorption

Lacrimal Drainage Systemic absorption and redistribution ultimately occur through blood vessels. The most obvious superficial ones in the eye are the conjunctival vessels. However, additionally the drainage of tears and substances through the canaliculi can be significant. Per blinking of the lids a volume of up to 2 μl may be pumped into the excretory lacrimal system. Ultimately the substances are reabsorbed by the nasal mucosa, thus bypassing the first-pass metabolism in the liver, and enter the systemic blood flow in rather high concentrations. Also, swallowing into the digestive system may occur. Again, from topical eye drop use it is known that, for example, beta blockers applied for glaucoma therapy may reach systemic therapeutic levels such as used for control of blood pressure.

Blood Flow Most of the blood flow to the eye stems from the inner carotid artery. Retinal vessels are end-stream vessels, they nourish the inner third of the retina, and do not anastomose. The retinal vasculature resembles widely the vasculature in the central nervous system, including the blood–retina barrier, which is equal in its function to the blood–brain barrier. Choroidal vessels, which possess a fenestrated endothelium, nourish the outer two-thirds of the retina including the photoreceptors. Thus, substances present in the blood stream and able to penetrate the blood–brain barrier, such as methanol, reach and damage the eye and its delicate neuronal structures, namely the retina and optic nerve.

3.10.4 Specific Toxicology of the Eye

See also Chapter 3.6, Section 3.6.3.

Chemical Injuries of the Eye

Chemical injuries are among the most urgent of ocular emergencies, often resulting in a dramatic decrease in visual acuity or loss of an eye. The prognosis for a burned eye depends not only on the severity of the injury but also on the rapidity with which therapy is initiated. Instant and copious long-term irrigation of the ocular surface, even if painful to the patient, is mandatory, in order to reduce the contact of toxic substances. In general, alkali injuries are more damaging to the eye than those caused by acids.

Chemical injuries are common in the chemical industry and laboratories, in machine factories, in agriculture, and construction workers. They also are frequently reported from fabric mills, automotive repair facilities, and cleaning and sanitizing crews. With the work environment responsible for about 60% of chemical burns, roughly 35% of those injuries occur at home. Most of the eye injuries at home result from the use of lime and drain cleaners. Injuries caused by caustic chemicals are among the most severe; the prognosis for satisfactory recovery from these injuries is extremely poor. Automotive battery acid burns have become increasingly more common, and can be especially devastating when combined with the shrapnel resulting from explosion. These accidents typically occur in the colder months, almost always after dark, and usually involve young men. During recharging of a lead acid storage battery, which contains up to 25% sulfuric acid, hydrogen and oxygen produced by electrolysis form a highly explosive gaseous mixture. The most common causes of storage battery explosions are lit matches (used to see battery cells) and the incorrect use of jumper cables.

The deployment of automobile air bags involves the conversion of sodium azide into nitrogen gas, and is accompanied by the sudden release of an alkaline gas and powder. Ocular chemical injuries from air-bag inflation must be considered dangerous and as great a threat to vision as other caustic injuries.

Alkalies

In alkali burns hydroxide ions saponify the superficial cell membranes and intercellular bridges, which facilitates rapid penetration into the deeper layers and into the aqueous and vitreous compartments. Possible affected structures include almost the whole eye: lids, conjunctiva, cornea, sclera, iris, lens, and retina.

Cell damage from alkaline agents depends on both the concentration of the alkali, and the duration of exposure. The higher the pH, the greater the damage, with the most significant injuries occurring at a pH of 11 or higher. As the pH rises, destruction of the epithelial barrier becomes progressively more extreme. In the corneal stroma, alkali cations cause damage and necrosis by binding to the mucopolysaccharides and to collagen.

The most common alkalies involved in ocular injury are calcium hydroxide (lime), potassium hydroxide (potash), sodium hydroxide (lye), and ammonium hydroxide (ammonia). Of these, calcium hydroxide penetrates corneal tissue the slowest because the calcium soaps are relatively insoluble, impeding further penetration of the agent. Potassium hydroxide penetrates corneal tissue faster, sodium hydroxide penetrates even faster, and ammonium hydroxide passes through the cornea most rapidly because it destroys the epithelial barrier by saponification of the lipoidal cell walls and because it diffuses fastest through the stroma.

Acids

Acids quickly denature proteins in the corneal stroma, forming precipitates that retard additional penetration. Overall, the rates of penetration for acids at equivalent concentrations and pH vary widely, with sulfurous acid penetrating

more rapidly than hydrochloric, phosphoric, or sulfuric acids. Acid burns are usually limited to the anterior segment. The affected structures typically include: lids, conjunctiva, and cornea.

Free hydrogen ions cause cellular necrosis and the strongest acids have the strongest necrotic potency. Weak mineral acids generally cause less severe ocular damage than do alkalies. Corneal tissue has an inherent buffering capacity that tends to equilibrate local pH to physiological levels, but severe chemical injuries exhaust the cellular and extracellular buffering capacities, allowing extremes of pH that are incompatible with tissue survival.

Hydrofluoric acid is a strong inorganic acid (used in industry for cleaning and etching) that is particularly toxic because it has a complex mode of tissue injury. Along with necrosis from a high concentration of hydrogen ions, this agent causes cellular death because the fluoride anion binds calcium more quickly than the body can mobilize calcium. In addition, the fluoride blocks the Na-K-ATPase of cell membranes, resulting in a fatal loss of potassium from the cells.

Severe chemical burns represent a devastating injury. If the conjunctiva, corneal limbal region, and cornea are severely damaged, opacification of the transparent structures may occur permanently. Even corneal transplants have a guarded prognosis. When the corneal epithelial stem cell area is damaged, neovascularization occurs and transparency of the transplant may not be ensured.

Conjunctival Irritation and Inflammation

As described in Chapter 3.6, irritation not only of the skin of the lids but also of the conjunctiva may occur. An irritated conjunctiva has a large potential for swelling, even to a point when lid closure is no longer possible due to the bulging forward of conjunctival tissue, aggravating the problem by continued desiccation.

Contact inflammation is produced on the basis of a direct chemical effect on tissue; the resulting inflammation is not allergic in origin. Frequent defatting of the skin caused by excessive moisture also plays a role. Toxic reactions are much more common than allergic ones, accounting for about 90% of all reactions to topical ophthalmic medications. Toxic reactions include papillary conjunctivitis, follicular conjunctivitis, keratitis, pseudopemphigoid, and pseudotrachoma.

Contact Allergy

Contact inflammation of the eye may affect the lids, conjunctiva, or cornea, or a combination of these, and can involve allergic or nonallergic (irritant/toxic) mechanisms.

Contact allergic reactions involve exposure to a sensitizing substance that is absorbed through the skin. Sensitization may take weeks to years to develop, depending on the ability of the hapten to act as a sensitizer, the amount applied and the duration, preexisting lid or ocular disease, and individual susceptibility. In sensitized persons, contact allergic reactions can occur within 48 to 72 h upon rechallenge, in keeping with type IV delayed-type hypersensitivity reactions. Allergic contact dermato-conjunctivitis is the second most common type of drug reaction, accounting for 10% of reactions to topical

ophthalmic medications. In contact reactions of the eye and elsewhere, irritant and allergic mechanisms may coexist. Clinically, it is usually difficult to separate the two. Classically, a contact allergic reaction will begin in 48 to 72 h, in keeping with type IV hypersensitivity reactions. An irritant reaction may begin within a few hours of contact or may occur only after prolonged use of a topical medication.

Although sometimes the offending agent may be obvious, often identification is difficult. Patch testing may provide the answer. A true allergic response will occur in 48 to 72 h. An irritant can also cause a positive reaction, but this usually develops within a few hours and can be avoided by using lower doses of testing substances.

The acute lesions of contact allergic dermatitis/blepharitis resemble acute eczema, with erythema, vesicles, edema, oozing, and crusting. The chronic phase is characterized by dryness, crusting, fissuring, and thickening of the skin. Contact allergic conjunctivitis involves conjunctival injection and chemosis; there may be a papillary response and serous or mucoid discharge. Initially, the lower conjunctiva and lid are usually more affected; later, the entire conjunctiva and upper lid may also become involved. Itching can be prominent.

Several topical ocular drugs are known to act as sensitizers, including aminoglycosides (gentamicin, tobramycin, and neomycin), sulfonamides, atropine and its derivatives, topical antiviral agents (idoxuridine, trifluridine), topical anesthetics, echothiophate, epinephrine and phenylephrine, and preservatives (thimerosal, benzalkonium chloride). Other sensitizing agents include lanolin and parabens (cosmetics, skin creams, and lotions), nickel sulfate (jewelry), copper (colored lid shades), chromates (jewelry, leather products, fabrics, industrial chromate steel), and *p*-phenylenediamine (hair sprays, clothing, shoes). Rubbing the eyes after handling soaps, detergents, or chemicals may explain a localized ocular or periocular reaction after a more general exposure.

Lyell's Syndrome

See Chapter 3.6, Section 3.6.3.

Mucocutaneous Disease

Drug-induced ocular cicatricial pemphigoid (pseudopemphigoid) Drug-induced ocular cicatricial pemphigoid (pseudopemphigoid) includes a spectrum of disease ranging from a self-limited toxic form to a progressive immunologic form indistinguishable from true ocular cicatricial pemphigoid. Immunoglobulins bound to the basement membrane can be detected.

Stevens–Johnson syndrome (SJS) The Stevens–Johnson syndrome is an acute vesiculobullous disease characterized by systemic toxicity (fever, malaise, headache) and extensive cutaneous and mucous membrane involvement, including the conjunctiva. Precipitating factors include drugs (e.g. sulfonamides, penicillin) and infections, particularly herpes simplex virus (HSV) and *Mycoplasma.* Toxic epidermal necrolysis is considered to be a severe variant.

The clinical picture resembles that of ocular cicatricial pemphigoid. However, the major difference is that whereas the scarring in ocular cicatricial pemphigoid is chronic and progressive, the conjunctival scarring of SJS occurs as a result of the acute inflammatory episode and is self-limited. Clinical deterioration is a result of the

subsequent tear deficiency and lid malpositions that result from the acute event, rather than chronic inflammation and progressive scarring.

Phototoxicity and Thermal Effects

Cornea

The portions of the human body that are not protected by either melanin or increased thickness of the stratum corneum are the lips, the cornea, and the conjunctiva. Wavelengths of light shorter than 295 nm are absorbed and reflected by the corneal surface. Light with a peak near 288 nm induces fragmentation of epithelial nuclear protein with resultant cell death.

One characteristic of ultraviolet (UV) injury is a latent period of several hours between exposure and effect. Corneal epithelial cells usually begin to die several hours after irradiation and are brushed off by the action of blinking or undergo spontaneous fragmentation. This exposes a number of bare nerve endings around each missing cell and results in excruciating pain. UV damage may arise from artificial sources such as a welding arc, a close lightning strike, or an electrical flash caused by the breaking of a high-tension circuit. The amount of damage depends on the intensity of the light and the period over which it is delivered.

Photokeratitis is also common at high altitudes in the winter, when the reflectance from new, clean snow may be as high as 85%. The very bright light coming from below is not shielded by the eyebrows, forehead, or most headgear. This is most troublesome in high-altitude northern settings where there are few impurities in the air to absorb the short wavelengths of light. It is less troublesome in the desert where the reflectance from sand is only approximately 17%.

Changes caused by UV light in the corneal epithelium are reversible and heal completely. They are not like the effects of sunlight on the skin, which may lead to neoplastic changes.

Conjunctiva Pterygium, a nonneoplastic hypertrophic growth of conjunctiva, occurs most often in persons who work in environments with a high surface reflectance of UV light. Those who live and work at latitudes of less than 30° have a much higher incidence of pterygium than do those living at higher latitudes. Limbal stem cells seem to be activated by chromic UV light exposure to initiate this process. Climatic droplet keratopathy shares these same environmental associations.

Lens Transmittance of light through the ocular media rises rapidly from 400 to 442 nm. The decreased transmittance of shorter wavelengths is due primarily to absorption in the lens. This is an important consideration because sunlight peaks at 550 nm, and the large amounts of energy that are present in the shorter wavelengths of light are readily absorbed by the lens. These shorter wavelengths have been implicated in the formation of dark brown (brunescent) cataracts that occur in higher frequency in areas where the UV

components of sunlight are most intense. It has been suggested that the short-wavelength light results in lens damage by inducing chemical changes in proteins of the epithelial cells. This formation of toxic photoproducts may inhibit growth and important metabolic activities. It has also been suggested that photo-oxidation of tryptophan may either remove this essential amino acid from its intended metabolic pathway or induce other effects.

Premature cataract formation is the ultimate clinical symptom of light-induced changes in the lens.

Retina Damage occurs more commonly with short wavelengths of light near 441 nm falling on the retina and usually requires approximately 48 h to induce cellular proliferation, with mitotic figures in the retinal pigment epithelium and in the choroid. Significant healing and regeneration may occur and visual acuity after a solar eclipse burn often improves. Photochemical lesions are usually produced with light intensities several orders of magnitude below that needed to produce a direct thermal burn.

Thermal injury produces an intense central core of damage surrounded by edema. Nonlinear effects are caused by ultrashort exposure times and are produced by strong electric fields, acoustic signals, shock waves, and other phenomena generated by transient elevations in temperature gradients. Thermal effects are produced primarily by intense bright beams of light focused on the retina, such as the xenon photocoagulator or the argon or ruby laser.

Specific Retinopathies Solar retinopathy is a clinical entity that has been recognized for centuries. Synonymous terms include eclipse blindness, photoretinitis, photomaculopathy, and foveomacular retinitis. Solar retinopathy has been described in military personnel, sunbathers, religious sun gazers, solar eclipse viewers, and people under the influence of psychotropic drugs. Whereas this form of retinal insult is photochemical, it may be enhanced by thermal effects.

Although welder's maculopathy is the most common ocular injury associated with welding arc exposure, retinal damage has been reported to occur. Clinically, welder's maculopathy is similar to solar retinopathy. Clinical correlation between the use of the operating microscope and macular phototoxicity was first suggested in 1977 and has become a commonly recognized clinical syndrome associated with various anterior, posterior, and combined surgical procedures.

Cataract Formation

The solid mass of the lens consists of about 98% protein. These proteins undergo minimal turnover as the lens ages. Accordingly, on aging, they are subject to the chronic stresses of exposure to light or other high-energy radiation and oxygen, and this damage is thought to be causally related to cataractogenesis. The term age-related cataract is used to distinguish lens opacification associated with old age from opacification associated with other causes, such as congenital and metabolic disorders or trauma.

Apart from irradiation-induced changes, alterations due to a change in lens metabolism by toxic substances are well known. Examples are organophosphates, naphthalene, cobalt, gold, silver, and lead. The most widely used iatrogenic substances are steroids, causing a characteristic posterior pole cataract. For other substances see Table 3.23.

Table 3.23 *Site of action in the ocular system of select xenobiotics following systemic exposure (modified from Fox and Boyes, Caserett & Doull's Toxicology, 2001).*

Toxicant	Cornea	Lens	Retina	Optic nerve
Acids	+			
Alkalies	+	+	+	
Acrylamide				++
Carbon disulfide			+	++
Chloroquine	+	+	+	+
Chlorpromazine	+	+	+	
Corticosteroids		++		+
Digoxin, digitoxin	+	+	++	+
Ethambutol			+	++
Hexachlorophene			+	+
Indomethacin	+	+	+	
Isotretinoin	+			
Lead	+	+	++	+
Methanol			++	++
Mercury, methylmercury			+	
n-Hexane			+	+
Naphthalene		+	+	
Organic solvents			+	
Organophosphates		+	+	+
Styrene			+	
Sunlight and welding			+	
Tamoxifen	+		+	+
UV-light	+	+		

Retinal Toxicity

In the retina of the eye, if the macula is involved, examples of chemicals producing retinal toxicity are lead, methanol, mercury, n-hexane, naphthalene, organic solvents, and organophosphates.

Representative medications whose use has given rise to recent reports of retinal toxicity are chloroquine and hydroxychloroquine, isotretinoin, sildenafil, vigabatrin, tamoxifen, and phenothiazines. A very small amount of damage (small fraction of 1%) may be perceived to be a significant abnormality by the patient.

Lead For over 100 years inorganic lead has been known to produce visual symptoms. Occupational lead exposure produces time and concentration-dependent retinal alterations. At higher lead exposure levels the retina and optic nerve seem affected. At lower levels the rods, which are the retinal receptors for black and white information, and their neuronal pathway seem to be the target structure.

Methanol Methanol, which is in wide use as a solvent, fuel source, or antifreeze agent is quickly absorbed after any type of exposure. Since it easily crosses membranes it distributes well to tissues in relation to their water content.

Acute methanol poisoning induces permanent structural alterations in the retina and optic nerve, which results in blurred vision to complete loss of vision. Retinal ganglion

cell edema and optic nerve head swelling may be observed. Methanol is oxidized in the liver to formaldehyde and then to formic acid.

Formate is perceived as a mitochondrial poison that inhibits the oxidative phosphorylation of retinal photoreceptors, glial cells, and ganglion cells.

Organic Solvents Organic solvents are known neurotoxins. Their adverse effects on the retina are not well understood to date. However, color vision deficiencies have been noted in exposed workers.

Organophosphates Again the neurotoxicity of organophosphates is well know with the effect of exposure and retinotoxicity being still under investigation. Retinal degeneration has been described in some individuals chronically exposed to pesticides.

Chloroquine and Hydroxychloroquine Hydroxychloroquine and chloroquine are taken by many patients on a chronic basis for rheumatoid disease. The incidence of retinopathy is very low when patients take less than 6.5 mg/kg/day of hydroxychloroquine or 3 mg/kg/day of chloroquine, and almost unknown within the first 5 years of usage at these dosage levels. These drugs accumulate in the melanotic retinal pigment epithelium (RPE) that lies just behind the retina, and gradually damage both RPE and the overlying retina. If drug exposure continues, the retinal damage spreads and eventually both center and periphery of the retina can be destroyed. There is no known treatment for this retinopathy. Once recognized there will virtually always be some degree of permanent visual loss, and the damage may continue to progress long after stopping the drugs (sometimes for years, perhaps because of residual drug deposits in the body).

Therefore the American Academy of Ophthalmology recommends that all individuals starting one of these drugs should have a complete baseline ophthalmologic exam within the first year of drug usage.

Isotretinoin Ocular side effects are dose related and probably the most frequent adverse reactions associated with these drugs. Retinal changes are rare compared with reports of blepharoconjunctivitis, subjective complaints of dry eyes, and transient blurred vision. Moreover, decreased ability to see at night after taking these agents may occur as early as a few weeks or after taking the drug for 1 to 2 years. Retinal dysfunction is probably due to the competition for binding sites between retinoic acid and retinol (vitamin A).

Sildenafil Sildenafil citrate, an oral therapy for erectile dysfunction (ED), is one of the largest-selling prescription drugs in the world. This inhibitor of phosphodiesterase type 5 (PDE-5) is a unique class of drugs not previously used in humans. PDE-5 is responsible for the degradation of cyclic guanosine monophosphate (cGMP) in the corpus cavernosum. With increased levels of cGMP the smooth muscle in the corpus cavernosum is relaxed, allowing inflow of blood. Retinal side effects may occur because sildenafil, although selective for PDE-5, has a minor effect on PDE-6, an enzyme involved in light excitation in visual cells. The ocular side effects most commonly associated with sildenafil are a bluish tinge to the visual field, hypersensitivity to light, and hazy vision. These reversible side effects may last from a few minutes to hours, depending on drug dose. Visual changes are seen in approximately 3% of men taking the standard 50-mg dose, in 11% of men taking 100 mg, and in 40% at a dose of 200 mg daily.

Vigabatrin Vigabatrin is a drug used in more than 50 countries for the treatment of refractory epilepsy. A 2% incidence of visual field abnormalities after 6 months of

therapy is reported. The main ocular side effect is symptomatic or asymptomatic visual-field constriction, which is usually bilateral and can progress to tunnel vision. More than 80% of these visual-field defects appear to be irreversible.

Tamoxifen This antiestrogen is used primarily in the palliative treatment of breast carcinoma, ovarian cancer, pancreatic cancer, and malignant melanoma.

An acute, debatable form of retinal effect, which is not well defined, may occur after only a few weeks of therapy, with any or all of the following: vision loss, retinal edema, retinal hemorrhage, and optic disc swelling. This may be a result of tamoxifen's estrogenic activity, which may cause venous thromboembolism. These findings are reversible with discontinuation of tamoxifen. The typical crystalline retinopathy reveals striking white-to-yellow perimacular bodies. This finding has been reported for many medicines including nitrofurantoin, canthaxanthin, and methoxyflurane. These occur most commonly after more than 1 year of therapy with at least 100 g or more of the drug. There are, however, a number of cases of minimal retinal pigmentary changes occurring after a few months and only a few grams of tamoxifen. Loss of visual acuity in this chronic form is often progressive, dose-dependent, and irreversible.

Phenothiazines Phenothiazines are used in the treatment of depressive, involutional, senile, or organic psychoses and various forms of schizophrenia. Some of the phenothiazines are also used as adjuncts to anesthesia, as antiemetics, and in the treatment of tetanus.

In patients on phenothiazine therapy for a number of years, a 30% rate of ocular side effects has been reported. If therapy continues over 10 years, the rate of ocular side effects increases to nearly 100%. The most significant side effects are reported with chlorpromazine and thioridazine, probably because they are the most often prescribed. The most common adverse ocular effect with this group of drugs is decreased vision, probably resulting from anticholinergic interference. Retinopathy, optic nerve disease, and blindness are exceedingly rare at the recommended dose levels, and then they are found almost only in patients on long-term therapy.

Retinal pigmentary changes are most often found with thioridazine. A phototoxic process has been postulated to be involved in both the increased ocular pigmentary deposits and the retinal degeneration. The group of drugs with piperidine side chains (i.e. thioridazine) has a greater incidence of causing retinal problems than the phenothiazine derivatives with aliphatic side chains (i.e. chlorpromazine), which have relatively few retinal toxicities reported. The phenothiazines combine with ocular and dermal pigment and are only slowly released. This slow release has, in part, been given as the reason for the progression of adverse ocular reactions even after use of the drug is discontinued.

The patient should avoid bright light when possible. Sunglasses that block out UV radiation up to 400 nm are recommended.

Other Common Drugs Ibuprofen has caused visual evoked response alteration.

Indomethacin has been noted to cause retinal pigment epithelium (RPE) disturbances.

Desferrioxamine, used in the treatment of systemic iron overload, has been noted to cause retinal epithelial alterations.

Clofazimine, used in the treatment of dapsone-resistant leprosy, has been noted to cause a bull's eye maculopathy.

Methanol ingestion can cause visual-field defects, retinal edema, and optic atrophy.

Toxic Optic Neuropathy

Toxic/nutritional optic neuropathy often presents as a painless, progressive, bilateral, symmetrical visual disturbance with variable optic nerve pallor. The patient may manifest reduction in visual acuity, loss of central visual field, and reduced color perception.

Toxic optic neuropathy may result from exposure to neuropoisonous substances in the environment, ingestion of foods, or from elevated serum drug levels. Nutritional deficiencies or metabolic disorders may also cause this disease. In most cases, the cause of the toxic neuropathy impairs the tissue's vascular supply or metabolism.

Among the common offenders is tobacco, which produces metabolic deficiencies as part of the systemic nicotine cascade. Alcohol (methanol as well as ethanol) produces its toxic effects through metabolic means. Chronic exposure typically leads to vitamin B-12 or folate deficiency. Over time, these deficiencies cause accumulations of formic acid. Both formic acid and cyanide inhibit the electron-transport chain and mitochondrial function, resulting in disruption of ATP production and ultimately impairing the ATP-dependent axonal transport system.

Numerous other agents can produce toxic optic neuropathy as can be seen in Table 3.23.

3.10.5 Summary

The eye is well designed by Nature to protect itself against the insults of regular life activities. Lids and tear film constitute a potent barrier towards xenobiotics delivered externally. The blood–retina barrier partially shields the delicate neuronal retinal tissues from insults delivered systemically. However, all protective mechanisms have their limits and so external insults, for example from acids or alkalies, may lead to extremely devastating injuries of the eye, compromising visual acuity by destroying the cornea and its clarity as well as internal ocular structures.

The metabolism of the transparent lens is very vulnerable to systemically delivered insults. Disturbances result in cataract formation.

The retina and optic nerve may be considered as a specific part of the brain so that compounds toxic to the CNS are generally expected to affect the retina and the optic nerve as well. Owing to the transparent optic media of the eye, phototoxic damage to nonsuperficial structures is a unique feature of this organ,.

References

http://www.bebig.com

Fox, D.A., Boyes, W.K. In Casarett and Doull's Toxicology. The Basic Science of Poisons, ed. Klaassen, CD, 6th edition, McGraw-Hill, New York, 2001.

Further Reading

Diamante, G.G., Fraunfelder, F.T. Adverse Effects of Therapeutic Agents on Cornea and Conjunctiva. In Corneal Disorders, Clinical Diagnosis and Management, ed. Leibowitz, HM, Waring, GO. 2nd edition, Saunders, Philadelphia, 1998.

Grant, W.M. Toxicology of the eye, 4th edition, Charles C Thomas, Springfield IL, 1993.
Kaufman, P.L. and Alm, A. Adler's Physiology of the Eye, 10th edition, Mosby, St Louis, Missouri, 2003.
Tasman, W., Jaeger, E.A., Duane's Ophthalmology, Lippincott Williams & Wilkins, Baltimore, Maryland, 2006.

3.11 The Cardiovascular System*

Helmut Greim

3.11.1 Structure and Function

The heart is a muscular hollow organ, which consists mainly of cardiac muscle, the myocardium. The heart is divided into chambers: the right and left *atrium*, and the right and left *ventricle*. The right atrium receives deoxygenated blood from the body via the superior and inferior *venae cavae* and from the heart itself during contraction, and passes it to the right ventricle. The left atrium receives oxygenated blood from the lungs and passes it to the left ventricle. The rate of the heart beat is primarily regulated by a pacemaker system consisting of the *sinoatrial node*, the *atrioventricular node*, both situated in the right atrium, and the *atrioventricular bundles* within the *interventricular septum*.

The Myocardium

The heart, an organ mainly consisting of a single muscle, is the driving force for the pulmonary circulation (right half of heart) and for the greater systemic circulation (left half). The low-oxygen blood flowing back from the body through the veins enters the right half of the heart and is pumped through the lungs. The blood, now saturated with oxygen, flows back through the pulmonary veins to the left half of the heart from where it is transported back into the body through the aorta (Figure 3.37).

The pumping effect of the heart is produced by the rhythmic successions of contraction **(systole)** and dilatation **(diastole)** of the myocardium. The direction of the blood flow is determined by the opening and closing of valves (tricuspid and mitral valves).

The contraction of the myocardial cells is triggered by abruptly changing the direction of the electric potential at their outer membrane (80 mV, negative on the inner side) by means of a transmembrane ion flow **(action potential).** The action potential releases calcium ions from intracellular calcium reserves **(sarcoplasmic reticulum)**, which, in turn, initiate the contraction of the contractile elements actin and myosin in the myocardial cells. The ions flowing during the action potential phase regulate the filling level of the intracellular calcium reserves and consequently – together with other factors – the force of the contraction.

*This chapter is partially based on Seibel: Herz, in Toxikologie, H. Greim and E. Deml, Wiley-VCH, Weinheim, 1996.

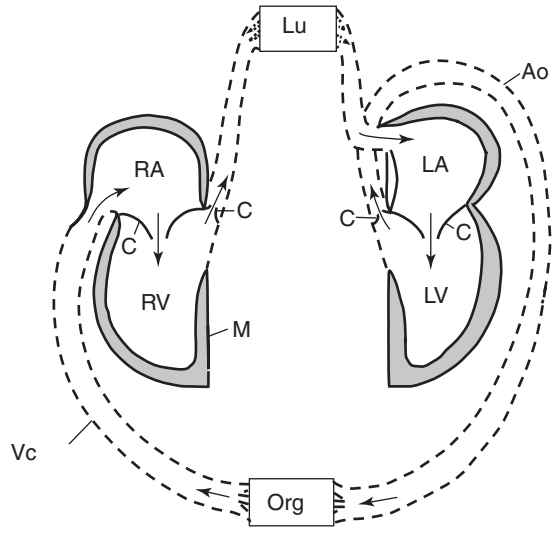


Figure 3.37 *Diagram of cardiac circulation. The arrows show the direction in which blood flows. The left half of the illustration shows low-oxygen blood returning from the organs being pumped out to the lungs. The right half shows oxygen-saturated blood returning from the lungs being pumped out to supply the organs. (M: myocardium; C: cardiac valves; RA: right atrium; RV: right ventricle; LA: left atrium; LV: left ventricle; Ao: Aorta, the main artery, Vc: Vena cava, the collecting vein). [Modified from Greim and Deml, Toxicologie, ch. 13, Abt. 13-1. Copyright (1996), with permission from Wiley-VCH].*

The System of Excitation and Conduction

The action potentials of the heart are normally produced in a specialized area (sinoatrial nodes), are then conducted via the atria to the atrioventricular nodes at the atrioventricular boundary; from here, they spread out into the entire myocardial tissue of the ventricles via a system of fibres specializing in conducting these stimuli (Figure 3.38).

Any form of damage to this system of excitation and conduction, for example due to oxygen deficiency, necroses, scar formation, or direct toxic impairment of the ion pumps and ion flows, results in cardiac arrhythmia, which can pose a threat to the pumping efficiency of the myocardium and thus a threat to life itself.

The Coronary Blood Vessels

The supply of the heart with substrates and oxygen takes place via the coronary arteries, which branch out of the aorta and spread out to cover the entire cardiac organ. The walls of the heart are drained by the cardiac veins.

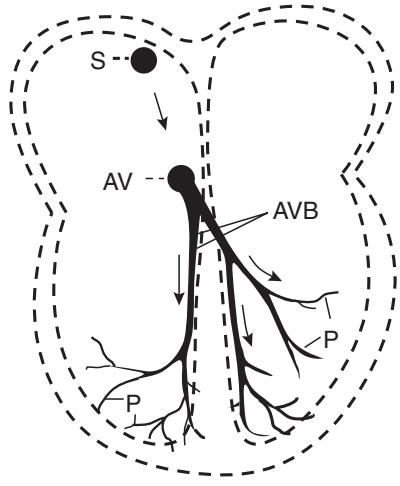


Figure 3.38 *Diagram of the excitation system. The arrows indicate the direction in which the excitation process takes place. S = sinoatrial node, the site from which excitation normally starts; AV = atrioventricular node; AVB: atrioventricular bundles; P = Purkinje fibres. [Modified from Greim and Deml, Toxikologie, Ch. 13, Abt. 13-2. Copyright (1996), with permission from Wiley-VCH.]*

After passing through the myocardial tissue the blood, deprived of its oxygen, is collected in the coronary veins and conducted to the right side of the heart. It is worth noting that the extraction of oxygen from the coronary blood by the heart is already very high in the state of rest when compared with the circulatory systems of other organs. As a greater amount of oxygen is required, for example, on physical or mental strain, the amount needed by the heart cannot simply be met by increasing the level at which oxygen is extracted from the coronary blood; an increase in blood flow must also take place. This makes it necessary for the coronary circulation to have a highly responsive and effective regulatory system. A pathological constriction of one or several coronary arteries therefore results in symptoms of oxygen deficiency in the heart. As a rule, primary toxic damage to the coronary arterial system has serious consequences for the myocardium and the system of excitation and conduction.

The Vascular System (Pulmonary and Systemic Circulation)

The vascular system is a closed system of tubes whereby blood permeates through all the tissues of the body. The pulmonary circulation serves the lung, the systemic circulation all other organs.

The **pulmonary circulation** transports the deoxygenated blood from the heart to the lung and, after gaseous exchange of carbon dioxide and oxygen, transports the oxygenated blood back to the heart.

The **systemic circulation** transports oxygenated blood from the heart into the body and the deoxygenated blood back to the heart. It provides the tissues with oxygen, nutrients, which are absorbed from the alimentary tract, hormones, and enzymes and carries back waste products such as carbon dioxide or urea, which are eliminated via the lungs or the kidneys, respectively. The driving force is the contraction of the left ventricle during systole, which forcibly expels blood into the aorta and from there via the arteries, arterioles, and capillaries to the organs. Venules and veins carry the deoxygenated blood back to the right atrium.

The capillaries are thin-walled blood vessels situated between arterioles and venules. They are the site of the exchange of materials (oxygen, nutrients, carbon dioxide, and other waste products) between capillary blood and surrounding tissues.

3.11.2 Toxicology

In principle, each of the three functional cardiac systems (myocardium, excitation and conduction, and the coronary arteries) can be damaged by toxic substances.

Damage to the Myocardium

Cadmium Cadmium reduces the **contractility** of the myocardium by blocking the calcium flow into the cells during the **action potential**. In addition, the membrane of the myocardial cell possesses an ion-exchange system, which transports sodium and calcium. Cadmium principally competes with calcium at the binding site on the external surface of the cell membrane. By these mechanisms, cadmium reduces the flow of calcium into the heart muscle cell, which means that less calcium is available for contraction.

Cadmium also penetrates into the cell, reducing the available amount of adenosine triphosphate and phosphocreatine as well as the phosphorylation of the contractile protein. Furthermore, cadmium competes more powerfully than other divalent cations with calcium for those binding sites on the contractile protein, which finally trigger contraction.

Cobalt A considerable amount of experience has been gathered as regards the effect of cobalt on the human heart from a poisoning catastrophe. In the USA, Canada, and Belgium during 1965/1966, cobalt was used as a foam-stabilizing additive to beer. This resulted in severe and sometimes fatal damage to the myocardium. Death occurred with the symptoms of severe cardiac insufficiency with circulatory collapse and acidosis. A similar intoxication has occurred after industrial exposure to cobalt. The decisive effect is a disturbance of the myocardial energy metabolism, which is initiated by binding of cobalt to the disulfide groups of lipoic acid. It was noted that, in the case of the catastrophe cited above, persons with insufficient nutritional intake were more liable to suffer from intoxication than those with a normal nutrition. This could be explained by the fact that cobalt binds to the sulfhydryl groups of food components (for example cysteine from proteins), so that binding to the endogenous disulfide groups is reduced.

Arsenic (Arsenic Trioxide) In 1900 a poisoning epidemic occurred with beer to which arsenic had been added. Around 6000 persons suffered, of which 70 cases were fatal. In nearly all cases heart failure was involved. As to whether the effects of the arsenic on the myocardium were the result of the capillary damage known to be produced by this poison or whether they were due to a direct effect on the myocardial cells remains open.

Arsines Poisoning through inhalation of arsines is not infrequent in the metal-processing industry. The most prominent feature of such toxic effects is a **haemolysis** (erythrocytic breakdown). However, additional damage to the heart also seems to take place. In persons who had died from arsine poisoning, a distension of the ventricles and, microscopically, a fragmentation of the myocardial fibres were found.

Lead Cases of acute to subacute lead poisoning have been described that resulted in death within a few days. Symptoms of severe acute **cardiac insufficiency**, which indicate damage to the myocardium, preceded death. The occurrence of severe **cardiac arrhythmia** also shows that, in addition to the myocardium, the system of excitation and conduction is also damaged by lead.

Antimony Cases of poisoning with antimony are no longer as frequent compared with earlier years. Antimony tartrate was used medically for a long time. Disturbances in repolarization, which indicate damage to the myocardium, predominated in larger groups of these patients.

Ethanol Approximately 1% of heavy alcohol drinkers fall prey to alcohol-produced damage to the myocardium (**cardiomyopathy**). In the light of widespread high alcohol consumption, this disease represents one of the most frequent forms of heart damage. As a rule, alcoholic cardiomyopathy develops after a high consumption of alcohol over the preceding 10 years. This alcohol-related damage should not be equated with the cardiomyopathy produced by vitamin B_1 deficiency, which occurs relatively often in alcoholics.

Among the acute effects of alcohol on the heart, the reduction in contractile force is to be emphasized, which already shows extreme effects in isolated hearts at concentrations of around 4% ethanol. In persons with healthy hearts, a single dose of 110 g ethanol reduces the stroke volume of the heart, resulting in a reduction of the contractile power of the myocardium.

Damage to the System of Excitation and Conduction

Halogenated Hydrocarbons Over recent years, a large number of halogenated hydrocarbons have produced cases of fatal poisoning, which were principally caused by damage to the system of excitation and conduction of the heart. However, accurate differentiation of the substances according to their toxicity, let alone differences in their active mechanisms, is not possible. Concerning the frequency at which fatal damage to the heart occurs, the following are to be cited in particular: **trichloroethylene** and **trichloroethane**, formerly used as narcotics and still processed in industry, and **dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane** and **trichlorofluoroethane**, used as propellants for aerosols or as freezing agents.

In the United States in the 1960s it became fashionable for adolescents to inhale solvents for their intoxicating effects. As a result, within a relatively short time, over 100 fatal cases were reported. Fluorinated hydrocarbon propellants and trichloroethane were among the substances most frequently responsible for accidents of this kind. The mortalities occurred shortly after inhalation with such suddenness that fatal cardiac arrhythmia had to be regarded as being the probable cause of death.

Animal experiments demonstrated that halogenated hydrocarbons at low concentrations, at which they were by themselves not able to produce cardiac arrhythmia, intensified the arrythmogenic effect of adrenaline. It is not clear whether the phenomena described are produced by the direct action of the halogenated hydrocarbons on the excitation and conduction system or indirectly via induced reflex impulses of the sympathetic and parasympathetic nerves acting on this system.

In addition, halogenated hydrocarbons restrict the contractibility of the heart and reduce the cardiac output (per minute) and the blood pressure. With some of the substances, relatively low concentrations in the inhaled air are effective within a very short time, e.g. 1% trichlorofluoroethane or 2.5 % dichlorotetrafluoroethane.

Cadmium Apart from blocking the flow of calcium into the cell, cadmium also inhibits excitation transmission in the atrioventricular node. In workers who had been exposed to concentrations of 0.04 to 0.5 mg cadmium per m^3 inhaled air, a significantly disturbed excitation transmission was found.

Mercury Although mercury chloride and organic mercury compounds have effects on the heart similar to those of the highly toxic **digitalis glycosides**, cardiac involvement received little mention compared with the massive damage to the nervous system in the descriptions of epidemic poisoning with organic mercury compounds (in Minamata, Japan, and in Iraq). However, an electrocardiographic investigation in Iraq revealed changes in the electrocardiograms of all poisoned persons. Disturbed repolarization was found in all cases, indicating damage to the myocardium, but also a high percentage of different types of arrhythmias, so that toxic effects on the excitation and conduction system had to be assumed as well.

Barium Acute poisoning with barium chloride, inadvertently taken as a laxative, produced severe cardiac arrhythmias.

Damage to Coronary Arteries and other Blood Vessels

In relatively rare cases, toxic damage to coronary perfusion due to the short-term action of an environmental poison (**acute damage**) occurs. More frequently, an accelerated development of atherosclerosis of the coronary arteries under the long-term action of toxic substances in the environment (**chronic damage**) takes place. In such cases, additional factors are also involved as a rule (**high blood pressure, high blood cholesterol concentrations**).

In many cases, not only the coronary arteries are affected, but also other blood vessels, e.g. those supplying the brain, the kidneys, or the muscles, so that circulatory disturbances can also occur as sequelae there, resulting in reduced cerebral performance, stroke, renal failure, and muscle pains, mostly as a consequence of oxygen deficiency.

Carbon Disulfide After many years' exposure to carbon disulfide, increased blood pressure, angina pectoris, accelerated arteriosclerosis and higher mortality from myocardial infarction can be observed in viscose workers.

In persons exposed to CS_2, an increase of the cholesterol concentration in their plasma has been found. In experimental animals, functional and morphological changes of the heart, including necrosis of the myocardium, have been attributed to both a direct effect of CS_2 on the heart and to increased incorporation of cholesterol and lipoproteins into heart vessels, leading to arteriosclerosis. In this context, a disturbed glucose metabolism caused by CS_2 should be mentioned, as the metabolic disturbances occurring in diabetes also promote the formation of coronary sclerosis.

Compared with this chronic effect, the acute effects of CS_2 on the heart are of little importance. The concentrations at which acute effects have been obtained in animals are relatively high.

Nitrate Esters In human toxicology, cases of acute intoxication with organic nitro compounds are rare. Chronic poisonings, particularly with nitroglycerine and ethylene glycol dinitrate did, however, occur in the past with a relatively high frequency. Nitroglycerine is readily absorbed through the skin and has usually been taken up by this route where poisonings occur. Ethylene glycol dinitrate also readily penetrates the skin. However, as it is volatile, poisoning usually takes place via inhalation.

At higher doses, both substances, similar to the longer acting substances used in medicine, are responsible for dilatation of all blood vessels, whereas at low doses mainly the veins are involved. At low doses, congestion of blood in the veins of the lower extremities and in the abdominal cavity occurs, which results in less blood flowing back to the heart. The filling of the heart decreases and the volume of transported blood becomes smaller. By this means and due to the direct effect on the vessels, the blood pressure drops. The heart rate increases as a reflex. Orthostatic collapse may occur if the body is in an upright position. The characteristic **nitrate headache** is to be seen as the result of vascular distension and also occurs after administration of other vasodilators. In addition, there is a drop in the pressure of the cerebrospinal fluid, which also plays a part in causing headaches.

After intake of higher doses the organism develops tolerance (**tachyphylaxis**) to the described effects of organic nitrate esters within 1–2 days. When exposure to these substances is interrupted, the ability to react is restored after a relatively short time (about 2 days), so that the symptoms described above recur after re-exposure (Monday headache in exposed workers).

A different reaction can occur after contact with ethylene glycol dinitrate and/or nitroglycerine lasting for several years, shown in the form of withdrawal symptoms when exposure is discontinued. In such cases, symptoms of **angina pectoris** or **myocardial infarction** appear 2–3 days after termination of exposure. Fatal cases have also occurred. As sclerotic changes in the coronary arteries are not usually present, the term ''non-arteriosclerotic ischaemic heart disease'' has been coined. In such cases, spastic constrictions of the coronary arteries have been demonstrated, which were reversible after the administration of nitroglycerine.

Carbon Monoxide An increase of the HbCO (haemoglobin–carbon monoxide complex, carboxy haemoglobin) concentration to 5 % causes a rise in blood flow through the

coronary blood vessels similar to hypoxia; this finding is to be interpreted as a reaction to the reduced supply of O_2 to the myocardium. Correspondingly, physical exercise can disturb the relationship between the oxygen supply and the oxygen requirements of the myocardium at mean HbCO concentrations of only 5 % and may lead to an angina pectoris attack.

High concentrations of HbCO in the blood reduce cardiac contractility and the pumping efficiency of the heart. The electrocardiogram indicates ischemia as seen in cases with insufficient blood supply of the myocardium. Cardiac arrhythmias frequently occur.

Necroses of the myocardium are found in fatal CO poisoning. In addition to binding to haemoglobin, recent experimental findings indicate that the binding of CO to intracellular myoglobin and cytochromes plays an additional role. It has been found, for example, that the CO–myoglobin concentration is 2–3 times higher than the corresponding HbCO concentration. The affinity of the CO to cytochrome P450 is greater than that to haemoglobin.

Cigarette Smoking Epidemiologically, it has been proved that smoking promotes the development of atherosclerotic diseases of the heart and other organs. It seems that pipe and cigar smokers are less at risk than cigarette smokers. The explanation for this is probably that cigarette smokers inhale more frequently, so that the substances in the smoke are absorbed to a greater extent. Which ingredient or ingredients of the cigarette smoke are responsible for the development of **arteriosclerosis** is a matter of controversy. Lipid peroxidation caused by radicals probably plays a decisive role.

Besides the chronic effect, cigarette smoke also has acute effects, which can be of importance for a previously damaged heart. By means of increasing the HbCO content of the blood, smoking produces a **reduction in the oxygen supply,** which is significant for the heart with coronary sclerosis. Smoking also, mainly due to nicotine, produces a rise in blood pressure and heart rate due to its effect on the vegetative nervous system and by releasing the endogenous catecholamines.

Water hardness There is no longer any doubt that there is an inverse correlation between the hardness of drinking water and the frequency of coronary heart disease. While it is certain that atherosclerotic coronary heart disease occurs more frequently in areas where the water is soft, the causal relationship is still doubtful.

Lead Patients with chronic lead poisoning frequently complain of palpitations, dyspnoea and precordial pains, which constitute **saturnine angina pectoris**, a set of pathological symptoms which have been known for a very long time. These are probably due to contractions of the coronary arteries.

The question as to whether long-term effects of lead increase the frequency of atherosclerotic coronary heart disease and other atherosclerotic diseases is to be considered. Epidemiological investigations and animal experiments indicate an enhancing effect of chronic lead poisoning on the development of arteriosclerosis, both in the coronary blood vessels and in the aorta.

Cadmium According to several authors there is an increased risk of death due to coronary sclerosis and other atherosclerotic diseases in persons chronically exposed to cadmium. However, the way in which cadmium can effect the coronary blood vessels is

not known. As an increased blood pressure is one of the major factors in the development of atherosclerosis, the effect of cadmium on blood pressure may be involved.

In animal experiments, hypertension can be induced by long-term feeding with small to moderate amounts of cadmium, although this was not reproducible in all of the studies. This contradiction can mostly be explained because the dependence of blood pressure on cadmium dose actually has two phases. The hypertensive effect starts at a dose which is low compared with frequent human exposures, i.e. 0.01 ppm in the drinking water; it reaches a maximum at 1 ppm before dropping steeply, then rapidly turning to a hypotensive effect at 10 ppm.

Fibers and Particles Epidemiological studies have shown that ambient exposure to particles that display a mass mean diameter of 10 μm (PM_{10}) is related to both respiratory and cardiovascular mortality and morbidity (see Chapter 6.3 Toxicology of Fibers and Particles). These cardiovascular effects are explained by either impairment of the blood flow to or in the heart, or interference with autonomic innervation. Moreover, repeated exposure to PM_{10} may, by increasing systemic inflammation, exacerbate the vascular inflammation of atherosclerosis and promote plaque development or rupture of blood vessels. Atherogenesis is an inflammatory process, initiated via endothelial injury, which produces systemic markers of inflammation that are risk factors for myocardial and cerebral infarction. The inflammatory response to particles or the particles themselves may also impact on the neural regulation of the heart, leading to death from fatal dysrhythmia. In support of this hypothesis, studies in humans and animals have shown changes in the heart rate and heart rate variability in reponse to particle exposures.

In particular, combustion and model nanoparticles (NPs) as well as Diesel particles, which also contain NPs, can gain access to the blood following inhalation and enhance experimental thrombosis, but it is not clear whether this is an effect of pulmonary inflammation or of particles translocated to the blood. High exposures of combustion-derived NPs (CDNPs) by inhalation caused altered heart rates in hypertensive rats and dogs, which are interpreted as a direct effect of CDNPs on the pacemaker activity of the heart. Inflammation in distal sites has also been associated with increased progression of atheromatous plaques in rabbits, ApoE−/− mice, and humans.

3.11.3 Summary

The heart is a functional unit comprising the working muscle (myocardium), a system of electric stimulation and conduction, and the system of coronary arteries and veins supplying the heart muscle. Basically, all three systems can be damaged by toxic substances.

Metals cause damage to the myocardium as well as to the heart conduction system, whereas halogenated hydrocarbons produce cardiac arrhythmias, i.e. they can disturb regulation of the heartbeat frequency. The effects on coronary arteries are principally due to chemicals that can either produce an acute constriction of these blood vessels, such as nitro compounds, or constrict them after chronic exposure by sclerosing (coronary sclerosis). In both cases, the consequence is an insufficient supply of oxygen to the myocardium, such as can also be induced in the blood by increased CO–haemoglobin levels. Sclerosing effects are also produced, e.g., by cigarette smoke, carbon disulfide, lead, and cadmium. The association between low levels of water hardness and

arteriosclerosis needs further evaluation. Inflammation can be seen as the main driver of the cardiovascular effects of particulate matter (PM). There is evidence of systemic inflammation following increased exposure to PM, as shown by elevated C-reactive protein, blood leukocytes, platelets, fibrinogen, and increased plasma viscosity.

References

Greim, H. and Deml, E. Toxikologie, Eine Einführung für Naturaissenschaftler and Mediziner, VCH Verlagsgemeinschaft mbH, Weinheim, 1996.

For Further Reading

Burkhoff, D., Mirsky, I., and Suga, H. Assessment of Systolic and Diastolic Ventricular Properties via Pressure-Volume Analysis: a Guide for Clinical, Translational, and Basic Researchers. *Am. J. Physiol. Heart Circ. Physiol.*, **289**, 501–512, 2005.

Clerico, A., Recchia, F.A., Passino, C., and Emdin, M. Cardiac Endocrine Function is an Essential Component of the Homeostatic Regulation Network: Physiological and Clinical Implications. *Am. J. Physiol. Heart Circ. Physiol*., **290**, 17–29, 2006.

Frishman, W.H., Del Vecchio, A., Sanal, S., and Ismail, A. Cardiovascular Manifestations of Substance Abuse: Part 2: Alcohol, Amphetamines, Heroin, Cannabis, and Caffeine. *Heart Disease*, **5**, 253–271, 2003.

Gibbs, C.L. and Chapman, J.B. Cardiac Mechanics and Energetics: Chemomechanical Transduction in Cardiac Muscle. *Am. J. Physiol. Heart Circ. Physiol*., **249**, 199–206, 1985.

Drexler, H., Ulm, K., Hardt, R., Hubman, M., Goen, T., Lang, E., Angerer, J., Lehnert, G. Carbon Disulfide IV. Cardiovascular Function in Workers in the Viscose Industry. *Int. Arch. Occup. Environ. Health*, **69**, 27–31, 1996.

Leone, A. Biochemical Markers of Cardiovascular Damage from Tobacco Smoke. *Curr. Pharm. Des.*, **11**, 2199–2208, 2005.

Malpas, S.C. Neural Influences on Cardiovascular Variability: Possibilities and Pitfalls. *Am. J. Physiol. Heart Circ. Physiol*., **282**, 6–20, 2002.

Meininger, G.A. and Davis, M.J. Cellular Mechanisms Involved in the Vascular Myogenic Response. *Am. J. Physiol. Heart Circ. Physiol*., **263**, 647–659, 1992.

Routledge, H.C. and Ayres, J.G. Air Pollution and the Heart. *Occup. Med*., **55**, 439–447, 2005.

Thornell, L.E. and Eriksson, A. Filament Systems in the Purkinje Fibers of the Heart. *Am. J. Physiol. Heart Circ. Physiol*., **241**, 291–305, 1981.

Young, M.E. The Circadian Clock within the Heart: Potential Influence on Myocardial Gene Expression, Metabolism, and Function. *Am. J. Physiol. Heart Circ. Physiol*., **290**, 1–6, 2006.

Zhang, X., Li, S.-Y., Brown, R.A., and Ren, J. Ethanol and Acetaldehyde in Alcoholic Cardiomyopathy: From Bad to Ugly en route to Oxidative Stress. *Alcohol*, **32**, 175–186, 2004.

Zhang, D.X. and Gutterman, D.D. Mitochondrial Reactive Oxygen Species-mediated Signaling in Endothelial Cells. *Am. J. Physiol. Heart Circ. Physiol*., **292**, 2023–2031, 2007; doi:10.1152/ajpheart. 01283.2006.

Cardiovascular System: http://www.innerbody.com/anim/card.html: and http://www.innerbody.-com/image/cardov.html

4

Methods in Toxicology

4.1 OECD Test Guidelines for Toxicity Tests *in vivo*

Rüdiger Bartsch

4.1.1 Introduction

Animal tests are necessary to discover possible adverse effects of chemical substances and pharmaceuticals to humans. Certain *in vivo* tests are, therefore, required by regulatory agencies and must be carried out according to internationally recognized test guidelines by qualified test laboratories.

Tests, or bioassays, in animals are still a prerequisite for the evaluation of the toxicity of chemicals to humans. Although *in vitro* tests are a valuable tool in assessing certain effects, the behavior of a chemical in the body can not yet be predicted reliably. Uptake via the various routes of administration, distribution, and metabolism can be modeled with so-called 'physiologically based pharmacokinetic' models but the complex interactions of the substance or its metabolites with cell components, e.g. proteins or enzymes, and the resulting toxicity can only be investigated in the living organism. Animal tests have come under criticism because the results supposedly cannot be scaled to humans. While it is true that some effects cannot be reliably detected in animals, e.g. combined effects of medicines, hypersensitivity, etc., most of the adverse reactions caused by chemicals can be recognized using *in vivo* tests on mammals. Therefore, animal tests are necessary and are required by regulatory agencies throughout the world to protect patients, consumers, and workers from the hazards of chemicals to which they are exposed.

Toxicology and Risk Assessment: A Comprehensive Introduction Edited by H. Greim and R. Snyder

In vivo tests are required for a variety of regulatory purposes:

- Authorization for sale and use of pharmaceuticals, pesticides, fertilizers, food additives.
- Registration of work-place chemicals.

These fields demand different tests and requirements but the tests to be carried out can be principally grouped into:

- Acute toxicity
- Skin and eye irritation
- Sensitization
- Toxicity after repeated treatment
- Reproductive toxicity
- Mutagenicity
- Carcinogenicity.

The results of the tests must be reliable and reproducible to be accepted by regulatory agencies. This is ensured by standardized study protocols for each test and by an extensive report of the study director on the conduct of the study, so that any deviation from the study guideline can be reconstructed and assessed to determine whether or not it significantly influenced the outcome of the study. Only qualified laboratories are accredited to perform tests for regulatory purposes and they must adhere to nationally and internationally accepted **Good Laboratory Practice (GLP)** guidelines. These guidelines describe how to report and archive laboratory data and records to avoid manipulation of the results. The GLP guidelines also require Standard Operating Procedures (SOPs), statistical procedures for data evaluation, instrumentation validation, materials certification, personnel qualification, proper animal care, and independent Quality Assurance (QA). GLP regulations were first developed by the US Food and Drug Administration: http://www.fda.gov/ora/compliance_ref/bimo/glp/78fr-glpfinalrule. pdf. Similar procedures have been adopted by the Organization for Economic Cooperation and Development (OECD): http://www.oecd.org/document/63/0,2340,en_2649_201185_2346175_1_1_1_1,00.html.

The OECD originated in 1948, including only European countries. The present 30 member countries comprise an international group for the exchange of information and policies. The tests used for regulatory purposes have been standardized internationally by the OECD since the 1980s and are referred to as OECD test guidelines (OECD TG). Therefore, tests conducted according to an OECD TG are recognized by regulatory agencies of any member country. As a result of this agreement the number of animals used for tests is reduced because each substance need only be tested once in each assay and the results will be accepted in every OECD member country.

An overview of the important OECD TGs for *in vivo* tests is given below. It is not possible to list every detail of the guidelines; instead, the most important aspects are highlighted.

4.1.2 Requirements for *in vivo* Tests

The first prerequisite to initiating a series of tests is that the physicochemical properties and purity of the test substance have been documented. The exposure

route chosen should be the same as for the anticipated (or known) human exposure. Several dose groups are used to detect a possible dose–response relationship including a no-observed-adverse-effect level (NOAEL) and a lowest-observed-adverse-effect level (LOAEL). Test animals must be healthy to ensure that no disease interferes with a substance-related effect. The experiment should include an adequate number of appropriate control animals and a thorough statistical evaluation to ensure that an observed effect is not a chance finding.

Test Substance

The test substance should be of the highest purity to avoid possible influences from impurities. Physicochemical analyses such as mass spectroscopy, NMR, etc., offer measures of both identity and purity of the compound. Normally only pure substances are tested except when there is exposure to technical products, which generally are of a lesser purity.

The stability of the substance in the vehicle employed for the test has to be verified to avoid loss of the compound due to decomposition. If it is to be administered via the animal feed, the homogeneity of the preparation must be controlled. In the case of inhalation exposure the homogeneity of the test chamber atmosphere needs to be verified during exposures.

Exposure Route

The exposure route for animals in the study should be the same route by which human exposure is anticipated. Oral exposure via gavage, i.e. gastric intubation, has the advantage that the administered dose is known with precision. In many cases the test material is administered in the feed or in drinking water, where concentrations must be adjusted according to feed and water intake and body-weight changes during the course of the study. Administration by gavage is less preferable due to the possibility of overwhelming metabolism of a bolus dose. Furthermore, in many cases it does not represent the human pattern of exposure. For exposure via the diet or drinking water it should be ensured that the amounts administered do not interfere with normal nutrition and water balance. For these exposure routes, reduced palatability of feed or water caused by a bad taste of the test substance can lower the feed and water consumption, which in turn can have adverse effects on body weight and functions of some organs such as the kidney.

Other exposure routes used are dermal and inhalation. When dermal exposure is intended it is necessary to assure that the applied substance cannot be ingested by the animal. Application of the test substance can be open, semi-occlusive, or occlusive depending on the volatility of the test substance. An inhalation exposure for rodents can be either whole-body, head-only, or nose-only. The latter is preferable if the test substance is a dust or an aerosol, which can be deposited on the animals' fur and ingested by licking.

Animals

The Test Guidelines require that only healthy animals be used and that the health of the animals is checked before and during the study. The animals used should be free of

infection and should be monitored for disease during exposure to avoid confounding influences that might significantly impair interpretation of the results. Specific-pathogen-free animals (SPF-animals), which are certified to be free of certain known diseases, may be used to help control this variable. Change in body weight is an important, albeit nonspecific, measure of toxicity. Therefore, animals are randomly assigned to study groups according to their body weight at the beginning of a study to control for potential differences in body weight between dose groups.

Control Animals

Except for acute tests, a concurrent control group of animals of the same strain and sex is used to ensure that the observed effect is indeed caused by the chemical. The control group is handled in the same manner as the test group and receives only the vehicle (if any is used) in the highest amount used for the substance-treated animals. For substances administered via the diet and which cause a reduced food intake, the use of a pair-fed control group may be considered. This control group is fed the same amount of diet as was consumed by the dosed animals. In carcinogenicity studies it is especially important to compare the results with those of an 'historical control' (i.e., all control animals of the same sex and strain used for the same type of test and exposure route in the last few years in the testing laboratory) because tumor incidences can vary with time, and the incidences for certain tumor locations display high background variability.

Interpretation of Results and Statistical Evaluation

At the termination of most of the studies the animals are examined for gross and microscopic pathology. Ideally, histopathological examination of the organs of animals from the study should be conducted without knowledge of the treatment group to minimize bias. The physical findings of control and treated animals are compared and evaluated by appropriate statistical methods to exclude the possibilities of chance findings. Only statistically significant results can be regarded as establishing a cause-and-effect relationship between the test substance and the generation of an adverse effect. Evaluation of the data should also include an evaluation of the biological plausibility of the findings. In some cases a statistically significant effect of a chemical is not observed despite a mechanistically plausible expectation of an effect. Therefore, a high degree of expertise is required to interpret marginal or contradictory results.

4.1.3 Acute Toxicity

Acute toxicity tests provide information on lethality (at high doses), symptoms (at lower doses), and target organ toxicity after a single application of graded amounts of a substance via the oral, dermal, or inhalation route. Efforts to reduce animal numbers used in these tests have led to the abolishment of the so-called 'LD_{50}-test' for acute oral toxicity. New test guidelines require relatively few animals to classify substances into predefined acute toxicity classes for regulatory purposes.

Significance of Acute Toxicity Tests

Acute toxicity tests provide information on symptoms of acute overdoses, a hint of general toxicity, and the mechanism of toxicity. They are the starting point for further, more elaborate tests, e.g. genotoxicity *in vivo*, sub-chronic and chronic toxicity. Since the results of LD_{50} tests on the same substance may vary, a grouping of the substance into an acute toxicity class is sufficient for classification and labeling purposes. Moreover, information on acute toxicity is important for emergency planning.

Oral Toxicity

The oral toxicity test is aimed at determining the lethality of a chemical after a single dose given by gavage. Before 2002 the LD_{50} test (OECD TG 401) was employed. Groups of rats were given the test substance in graded amounts and the mortality in each dose group was recorded up to 14 days after application. The dose causing lethality in 50% of the animals (LD_{50}) was calculated using standardized mathematical procedures.

The LD_{50} test has been replaced by 3 new test methods (Table 4.1). The fixed-dose procedure (OECD TG 420), the acute toxic class method (OECD TG 423), and the up-and-down procedure (OECD TG 425), each of which has been validated against the LD_{50} test and shown to produce similar toxicity rankings for various chemicals.

OECD TG 420 Groups of animals (usually 5 female rats) are dosed in a stepwise procedure with 5, 50, 300, or 2000 mg/kg body weight (bw). The starting dose, which is expected to produce some signs of toxicity without causing severe toxic effects or mortality, is selected on the basis of a rangefinding study with single animals. Further groups of animals may be dosed with higher or lower fixed doses, depending on the presence or absence of signs of toxicity or mortality. A period of 3–4 days between dosing at each dose level is recommended (if needed) to allow for the observation of delayed toxicity. This procedure continues until (1) the dose causing evident toxicity is established, (2) no more than one death is identified, (3) until no effects are seen at the highest dose or, (4) when deaths occur at the lowest dose.

OECD TG 423 Groups of animals (usually 3 female rats) are dosed in a stepwise procedure with 5, 50, 300, or 2000 mg/kg bw. Absence or presence of mortality will determine whether further dosing is needed,. The same dose is given to 3 additional

Table 4.1 *OECD Test guidelines for acute mammalian toxicity tests.*

Test (OECD Test Guideline)	Species, Number of animals per dose group	Duration Dosing regime	Observation timepoints
TG 420, 423, 425 (oral)	rat, see text	stepwise procedure	see text
TG 402 (dermal)	rabbit, (rat) 5 (10)	single application, $\geq$ 3 doses	mortality, up to 14 days p.a.
TG 403 (inhalation)	rat, 10	4 h, $\geq$ 3 concentrations	mortality, up to 14 days p.a.

animals, or the next higher or lower dose is given to 3 animals. Treatment of animals at the next dose should be delayed until the survival of the previously treated animals is certain. The method will enable a judgment with respect to classifying a substance according to one of a series of toxicity classes defined by fixed LD_{50} classification values.

OECD TG 425 The test consists of a single ordered dose progression, where animals (usually rats) are dosed, one at a time, at a minimum of 48-h intervals. The first animal receives a dose a step below the best estimate of the LD_{50}. If the animal survives, the dose for the next animal is increased by a factor of 3.2; if it dies, the dose for the next animal is decreased by a similar dose progression. Each animal is carefully observed for up to 48 h before making a decision on whether and how much to dose the next animal. The LD_{50} is estimated with the method of maximum likelihood based on the status of all the animals dosed.

The difference between OECD TG 401 and 420 and the other methods is that the decisive endpoint is not the mortality but rather clear signs of toxicity. The exact LD_{50} cannot be calculated with any of the replacement methods. These methods, however, allow the substance to be ranked and classified in one of the acute oral toxicity categories according to the Globally Harmonized System for the Classification of Chemicals.

Dermal Toxicity

The rabbit is the preferred species for testing the toxicity of a chemical applied by the dermal route. The substance is applied once undiluted for liquids or moistened with water for solids, in at least 3 adequately spaced doses on the shaved backs of 5 animals of the same sex under a semi-occlusive dressing to avoid ingestion of the test material. The treatment period is 24 h, the study is terminated after 14 days, and the LD_{50} is calculated. The test is repeated with 5 additional animals of the other sex except when it has been shown that the sex tested initially is more sensitive.

Inhalation Toxicity

Gases, vapors, aerosols, or dusts suspended in air are administered by inhalation to groups of 5 male and 5 female animals (usually rats) for 4 h. At least 3 adequately spaced concentrations are used. Animals can be exposed whole body, but the nose-only method is often preferred to avoid uptake via the skin or licking of the substance from the fur. In the case of aerosols and dusts it is important to characterize the particle-size distribution of the test sample. Only particles with a size of less than 4 μm (mass median aerodynamic diameter) can reach the lower respiratory tract of rats. However, OECD TG 403 does not prescribe the particle-size distribution. The LC_{50} is then calculated. It should be pointed out that a fixed concentration procedure will also be introduced for acute inhalation toxicity testing (TG 433).

Limit Test (All Exposure Routes) If there is information that the toxicity of a compound will be low, e.g. from tests on chemicals with a similar structure, a single oral or dermal dose of 2000 mg/kg bw (5000 mg/kg body weight if required by regulatory agencies) or an inhalation concentration of 5 mg/l can be tested on 5 animals of each sex. If these doses/concentrations do not produce mortality, a full study may not be necessary.

Examinations Necessary for All Acute Toxicity Tests The animals will be observed for clinical signs at multiple intervals during the course of the study. Survivors will be killed and all animals examined *post mortem* for macroscopically visible signs of organ damage, which may provide evidence of possible mechanisms of toxicity. Histopathology of target organs is optional.

4.1.4 Skin and Eye Irritation

These tests measure the effectiveness of a substance to elicit adverse effects when applied to the skin or eye. The species used is the rabbit.

Acute Dermal Irritation/Corrosion

Dermal irritation is the production of reversible damage of the skin following the application of a test substance for up to 4 h. Dermal corrosion is the production of irreversible damage to the skin observed as visible necrosis through the epidermis and into the dermis, following the application of a test substance for up to 4 h.

A fixed amount of the substance is applied on the clipped skin and held in contact with a semi-occlusive dressing. Solid substances should be moistened to assure good contact, whereas liquids are applied undiluted. Residual material will be removed from the skin without altering the existing response or the integrity of the epidermis. Erythema and edema at the application site are graded and recorded at the timepoints indicated in Table 4.2. If effects last beyond 72 h animals will be observed for up to 14 days. If reversibility is seen before 14 days the experiment should be terminated at that time. This test detects dermal irritation, which is a reversible damage of the skin, as well as dermal corrosion, which is visible necrosis through the epidermis. This TG was updated in 2002 with the aim of reducing both the number of animals used and the extent of pain for the individual animal. A sequential test strategy is proposed where as much information as possible is gathered to avoid unnecessary testing. Substances with one of the following properties need not undergo further testing: (1) high toxicity via the dermal route; (2) known irritant or corrosive properties; (3) chemical reactivity, e.g., pH ≤ 2 or ≥ 11.5; (4) structure–activity relationships that suggest similarities to known dermal irritants; (5) data from validated *in vitro* or *ex vivo* tests, which indicate a potential for severe irritancy or corrosivity. If the above information is not available, an initial *in vivo* test with one rabbit is considered necessary. If the results indicate a corrosive potential then no further testing is needed, otherwise one or two further animals are used for additional testing.

Acute Eye Irritation/Corrosion

Eye irritation is defined as the production of changes in the eye, which are fully reversible within 21 days following the application of a test substance to the anterior surface of the eye. Eye corrosion is the production of tissue damage, in the eye, or serious physical decay of vision, which is not fully reversible within 21 days following application of a test substance to the anterior surface of the eye.

Table 4.2 *OECD Test guidelines for* in vivo *skin and eye irritation and skin sensitization.*

Test (OECD Test Guideline)	Species, Number of animals	Duration, Dosing regime	Observation timepoints
Acute dermal irritation/corrosion			
TG 404	rabbit see text	0.5 ml of liquid or 0.5 g of solid substance, 6 cm^2, 4 h	60 min, 24, 48 and 72 h after patch removal
Acute eye irritation/corrosion			
TG 405	rabbit see text	0.1 ml of liquid or 0.1 g of solid substance, aerosols sprayed from a distance of 10 cm	60 min, 24, 48 and 72 h after application
Skin sensitization			
TG 406	guinea pig		
Maximization test	10 5 controls	**1st Induction**: day 0, intradermal ± FCA **2nd Induction**: day 6–8, epicutaneous, occlusive, 48 h ± FCA **Challenge**: day 20–22, epicutaneous, occlusive, 24 h, naïve application site	48 and 72 h after start of challenge
Buehler test	20 10 controls	**1st Induction**: day 0, epicutaneous, occlusive, 6 h **2nd Induction**: day 6–8, epicutaneous, occlusive, 6 h, on same application site **3rd Induction**: day 13–15, epicutaneous, occlusive, 6 h, on same application site **Challenge**: day 27–29, epicutaneous, occlusive, *24 h, naïve application site*	30 and 54 h after start of challenge

FCA = Freunds Complete Adjuvant.

A fixed amount of substance is placed in the conjunctival sac of one eye of each animal used. The lids are gently held together for about 1 s. The eyes of the test animals should not be washed for at least 24 h following instillation of the test substance, except for solids and in case of immediate corrosive or irritating effects. Local anesthetics may be used on a case-by-case basis. The eyes are examined after the timepoints indicated in Table 4.2 and the grades of the reaction of conjunctiva, cornea, and iris are recorded

according to a standardized evaluation scheme. Animals showing no ocular reactions are not examined beyond day 3. Animals with mild to moderate lesions should be observed up to 21 days to evaluate reversibility of the lesions. This TG was updated in 2002 with the aim of reducing the number of animals used, as well as to reduce pain for the individual animal.

A new testing strategy has been proposed in which substances with one of the following properties need not undergo further testing: (1) known irritant or corrosive properties to the eye; (2) chemical reactivity, e.g. pH ≤ 2 or ≥ 11.5; (3) structure–activity relationships that suggest similarities to known eye irritants; (4) data from validated *in vitro* or *ex vivo* tests, which indicate a potential for severe irritancy or corrosivity; (5) high toxicity via the dermal route; (6) known irritant or corrosive properties to the skin (i.e., if the substance is corrosive or highly irritating to the skin, the same would be assumed for the eye). If none of the above information is available and the substance is not corrosive or highly irritating to the skin *in vivo*, an initial *in vivo* rabbit eye test with one animal is performed. If the results indicate a corrosive potential then no further testing is needed, otherwise one or two further animals are used for additional testing.

4.1.5 Skin Sensitization

In the skin-sensitization test, the immunologically mediated cutaneous reaction to a substance is determined. Usually, a positive reaction is manifested by dermal edema and erythema in the experimental animals.

Validated structure–activity relationships and *in vitro* models may not be available. The test species to be used is the guinea pig. Two tests can be performed: the maximization test of Magnusson and Kligman, in which sensitization is potentiated by injection of Freund's Complete Adjuvant, and the Buehler test, which does not use adjuvant. Other procedures may also be used but only these two are described in detail in the TG. Recently two tests in mice, the 'mouse ear swelling test' and the 'local lymph node assay' have been developed. In the case of a positive result, guinea pig tests are not considered necessary. However, in the case of a negative result a guinea pig test must be carried out.

In both guinea pig tests, the animals will be initially treated with the test substance (induction dose) and after a certain time (induction period), during which an immune reaction may develop, they are treated again (challenge dose). At two defined timepoints after the start of the challenge (Table 4.2) the skin reactions are graded and evaluated against a control group, which will be handled identically but only receives the vehicle. A concurrent positive control is not required. Instead, animals are tested every 6 months with a known sensitizer to prove that the test system is stable and sufficiently sensitive. The concentrations of the test substance have to be chosen in a pre-test such that the induction dose(s) elicits a mild to moderate skin irritation but the challenge dose does not lead to a skin irritation. A substance is considered to be a sensitizer if a positive reaction is seen in at least 30% of the animals in the maximization test or in 15% in the Buehler test. In borderline cases, additional animals can be tested.

4.1.6 Toxicity after Repeated Dosing

In bioassays, which are intended to mimic either brief or extended human exposure to chemicals, animals can be exposed daily for various timeperiods ranging from 2 weeks to 2 years. The oral, dermal, or inhalation route may be selected and groups of animals may be exposed to a range of different doses. The main goal is to identify a NOAEL, which can be used as a starting point for calculating acceptable exposures for humans. Hematological changes, clinical biochemistry, and histopathological alterations of most tissues and organs can be observed. Carcinogenicity studies can be performed in animals to detect potential oncogenic activity in humans. Since the incidence of a given tumor type within a population may be low and there are practical limits to the number of animals that can be studied in a bioassay, high doses, which might lead to earlier detection of neoplasms, have to be used to overcome the problem of low statistical power. All of these studies require careful interpretation of the effects observed in rodents and the extrapolation of the findings to humans.

Bioassays to assess the effects of chemical treatment after multiple exposures are designated as subacute (14–28 days), subchronic (90 days), or chronic (about the average life-span of the species, i.e. for rats and mice about 2 years) studies. The test species are usually rats and mice. For dermal toxicity studies rabbits or hamsters are usually used. For chronic tests of pharmaceuticals in a nonrodent species, dogs are the preferred animals. At least 3 dose groups and an optional satellite group (i.e., additional animals in the highest-dose group that are held untreated for an additional period after termination of the study to check for delayed effects or recovery from effects) are used. Dose selection and dose spacing are critical factors in identifying possible toxic effects of long-term treatment and in revealing an NOAEL.

Ideally the highest dose should lead to clear toxic effects without producing lethality because the death of the animal would preclude observations for the entire intended time period and would impair evaluation of the results. The lowest dose should not result in any adverse effects. Finding an NOAEL in a chronic study is one of the most important and most expensive tasks in the testing of chemicals and, therefore, proper doses have to be selected.

To ensure that the entire effective dose range is adequately encompassed, a chronic study is always preceded by a shorter-term study, the results of which are used to select appropriate doses for the chronic study. A shorter-term assay can thus be regarded as a dose-finding study for a longer-term experiment.

Data Collection During and Following Repeated Treatments with Chemicals

Prior to initiation of the study all animals will be allocated to ‘control’ or ‘treated’ groups with equal distribution of body weights in each group. Animal weights should be measured and recorded throughout the study and at its termination. A record of feed and water consumption should also be kept throughout the study and animals should be monitored for clinical signs of toxicity during the course of the study.

Laboratory studies of samples periodically taken from the animals will include hematology, e.g. measurement of blood cell counts, hemoglobin, hematocrit, and clotting time and clinical biochemistry of plasma or serum components, e.g. liver enzymes, creatinine, and blood urea nitrogen. For subacute and subchronic studies urinalysis is optional. These measurements may indicate treatment-related functional impairment of organs such as the bone marrow, kidney, or liver. In some cases it will also be deemed appropriate to measure methemoglobin, cholinesterase, and/or the levels of various hormones.

At the termination of the study survivors will be killed and all animals, including survivors and those that died during treatment, will be subjected to gross necropsy. The weights of selected organs will be determined (Table 4.3). All organs with gross lesions and additional selected organs of all animals of the control group and of the highest-dose group will be examined histopathologically. These examinations should be extended to all animals of the other dose groups if treatment-related changes are observed in the highest dose group. Finally, measurements of functional neurotoxicity such as grip strength, or motor and sensory activity may be measured in either a subacute or a subchronic study.

Results will be evaluated statistically and if possible the NOAEL will be identified. Study designs of tests for toxicity after repeated treatment are outlined in Table 4.3.

Subchronic Studies Ophthalmological examinations have to be carried out before the study and at its termination.

Chronic Studies For pharmaceuticals both a rodent and a nonrodent species, usually the dog, are used. Exposure route and frequency are not fixed but should be chosen to simulate likely human exposure. Data from toxicokinetic studies can be used to determine exposure frequency. Hematology and urinalyses should be performed every 3 months, clinical chemistry investigations every 6 months on 10 rodents per group and on all nonrodents.

Combined Chronic/Carcinogenicity Studies The test species are usually rats, mice, or hamsters. A satellite group of 20 animals/sex at the highest dose is treated for at least 12 months to evaluate chronic toxicity. Fifty animals/sex/dose are used to study carcinogenicity over a major part of the lifetime of the species chosen. The exposure time is not fixed. Hematology and urinalyses should be performed every 3 months and clinical chemistry investigations every 6 months. For a valid negative test to be acceptable, no more than 10% of any group may be lost to autolysis, cannibalism, or to laboratory-related management problems, and survival in each group has to be no less than 50% at 18 months for mice and hamsters and 24 months for rats.

Carcinogenicity Studies Rats and mice are used because of their relatively short lifespan and the wealth of information on their susceptibility to tumor induction, physiology, and pathology. Moreover, a large historical database on tumor incidence in most strains and tissues exists, which is important in view of the large variability of tumor incidence in untreated animals and among different strains. The animals are exposed after weaning for the major part of their life but the experiment is terminated prior to the natural death of the animals (Table 4.3). The incidence of spontaneous and substance-induced tumors increases in older animals, so it is necessary to terminate the study after a

Table 4.3 *OECD Test guidelines for* in vivo *tests on acute and repeated toxicity and carcinogenicity.*

Test (OECD Test Guideline)	Number of animals per sex and dose	Duration, Dosing regime	Number of organs to be weighted	Tissues/organs to be examined histopathologically
Subacute	5 (high-dose satellite group optional)	7 d/week, can be reduced to 5 d/week		
TG 407 (oral)		gavage, 1/d, 28 d or feed or drinking water usually continuously	9	20
TG 410 (dermal)		clipped skin, 10% of body area, 6 h/d, 21 or 28 d, semiocclusive	4	4
TG 412 (inhalation)		nose-only, head-only, or whole body, 6 h/d, 14 or 28 d	4	6
Subchronic	10 (high-dose satellite group optional)	7 d/week (can be reduced to 5 d/week), 90 days; satellite group: at least 28 days		
TG 408 (oral) - rodents		gavage, 1/d, or feed or drinking water usually continuously	11	30
TG 411 (dermal)		clipped skin, 10% of body area, 6 h/d, semiocclusive	4	27
TG 413 (inhalation)		nose-only, head-only, or whole body, 6 h/d	5	30
Chronic				
TG 452	20 rodents 4 non rodents	$\geq$12 months exposure route and frequency depending on anticipated human exposure	5	35 + respiratory tract if inhalation study
Carcinogenicity				
TG 451	50	duration predefined: mice, hamsters: $\geq$ 18 months, rats: $\geq$ 24 months (depending on strain) or when 25% survival in lower dose groups is reached exposure route and frequency depending on anticipated human exposure		34 + respiratory tract if inhalation study + all grossly visible tumors or lesions suspected of being tumors

Table 4.3 (*continued*)

Test (OECD Test Guideline)	Number of animals per sex and dose	Duration, Dosing regime	Number of organs to be weighted	Tissues/organs to be examined histopathologically
Combined chronic/ carcinogenicity				
TG 452	50 + 20 for satellite group + additional animals for interim sacrifices	see TG 451	5	34 + respiratory tract if inhalation study

defined period (which, however, is not fixed in the TG) to avoid the impact of different life spans on the interpretation of the data.

At least three adequately spaced doses are tested. It is necessary to use relatively large doses/concentrations to maximize the chance of finding a possible increase in tumor incidence because only relatively few animals (50 per sex/dose) can be used for practical reasons. The highest dose used is called the 'maximum tolerated dose' (MTD). It should not cause excessive premature deaths, which would reduce the number of animals at risk. Ideally the top dose should result in no overt signs of toxicity but a decrease in body weight of no more than 10–15% compared with the control group. Careful selection of the MTD is crucial because if the MTD is not reached and no increase in tumor incidence is found, the results of the study might be questioned.

At 12 months, 18 months, and before termination of the study a blood smear is obtained from all animals. A differential blood count is performed on animals of the high-dose group and control animals. Clinical biochemistry and urinalysis is not required. For a valid negative test to be acceptable, no more than 10% of any group can be lost due to autolysis, cannibalism, or management problems, and survival in each group is no less than 50% at 18 months for mice and hamsters and at 24 months for rats. If a significant difference is observed in hyperplastic, pre-neoplastic, or neoplastic lesions between the highest-dose and control groups, microscopic examinations should be made on that particular organ or tissue of all animals in the study. If the survival of the high-dose group is substantially less than the control, animals of the next-lower-dose group should be examined microscopically.

It should be pointed out that in some cases, due to the large doses used, the metabolism of the animal is overwhelmed and toxifying or detoxifying mechanisms may not be operative. When interpreting the data it should also be recognized that a number of tumors are species-specific and, therefore, are not relevant to man. The incidence and type of tumors found have to be evaluated by experts and the aforementioned influences

of metabolism have to be taken into account in the final judgment of whether or not a substance is carcinogenic to man.

Special Consideration for Dermal Studies The concentration of test substance in the vehicle can be reduced if skin irritation is severe. It may be necessary to restart the study with lower concentrations if skin has been badly damaged early in the study.

Limit Tests Subacute and subchronic limit tests via the oral and dermal route with a dose of 1000 mg/kg body weight can be performed for relatively nontoxic substances. The limit test applies except when human exposure indicates the need for a higher dose level to be used.

4.1.7 Reproductive Toxicity

Tests on reproductive toxicity assess the effects on fertility of the parental generation and on development of the offspring. In developmental toxicity studies, pregnant females receive the test substance from implantation to delivery. The soft tissues and the skeleton of the offspring are examined in detail. In fertility tests, male and female animals are exposed to the substance during the pre-mating period and the mating period. The females are further exposed during pregnancy and lactation until weaning of the offspring (one-generation study). These offspring are then further exposed till they reproduce to examine possible effects of exposure from conception to the reproductive stage (two-generation study). These tests serve to determine an NOAEL for toxic effects on fertility and development.

Developmental Toxicity

Studies on developmental toxicity are focused on the effects of chemicals given to pregnant animals between implantation and delivery. The test animals used are mice or (preferably) rats and a non rodent species such as the rabbit. Each test and control group should contain a sufficient number of females to result in approximately 20 animals with implantation sites at necropsy. Groups with fewer than 16 animals with implantation sites may be inappropriate. The exposure route should be the same as the anticipated human exposure. At least 3 dose levels and a vehicle control are used. Ideally the highest dose should induce toxicity but no mortality in the parental animals. Maternal mortality does not necessarily invalidate the study provided it does not exceed approximately 10 percent. The substance is administered daily from the day of implantation to the day before expected delivery when dams are killed and the uterine content is examined. Females are weighed on day 0 (mating day), on the first day of dosing, at least every three days during the dosing period, and on the day of scheduled kill. Food consumption is measured on the same days. The fetuses are examined immediately after the scheduled kill. For rodents half of the fetuses should be examined for skeletal alterations and the remainder for soft-tissue alterations. Each rabbit fetus should be examined for both types of alterations. Numerical results should be evaluated using the litter as the unit for data analysis. Data

from animals that did not survive to the scheduled kill should be reported. Whether or not these data are included in the data analysis should be judged on an individual basis. For some endpoints historical data are useful. NOAELs for maternal and developmental toxicity should be derived if possible. A very important point is to consider whether developmental toxicity is manifested in the presence or absence of maternal toxicity. In the latter case, the developmental effects are clearly caused by the substance whereas in the former case an observed effect may not be attributed unequivocally to the test substance, because maternal toxicity might have contributed to the result. For example, inhibition of weight gain in the dam may lead to retardation of skeletal ossification in the fetus.

Fertility

One- and Two-Generation Studies Reproductive toxicity comprises effects on both parental fertility and developmental toxicity. Effects on pregnancy and lactaton may be observed on the parental generation (F0 generation) and on postnatal development in the offspring (F1 generation). In a two-generation study, the F1-generation is further exposed and also mated to assess possible reproductive effects in this generation as a result of exposure from conception to the reproductive stage.

The test species usually employed are mice and rats because of their high fecundity. Males should be dosed for at least one complete spermatogenic cycle, i.e. 70 days in the rat or 56 days in the mouse. Females should be dosed for at least 2 complete estrus cycles. Animals are then mated and the test substance is administered to both sexes during the mating period and, thereafter, only to females during pregnancy and lactation. The exposure route should be identical to the route by which human exposure is anticipated. At least 3 dose levels and a vehicle control should be used. Ideally the highest dose should induce toxicity but no mortality in the parental animals. The number of animals per sex and dose is not exactly specified but should contain a sufficient number of animals to yield about 20 pregnant females at or near term. Note that failing to achieve this number does not necessarily invalidate the study; the results should be evaluated on a case-by-case basis. One male can be mated to one or two females. The maximum mating period is 3 weeks. Reproductive organs of pairs that fail to mate should be examined microscopically to determine the cause of the apparent infertility. Males are examined after mating. Females are allowed to litter normally and rear their progeny until weaning. All of the offspring in the litter may continue in the study or litters may be standardized on day 4 after birth to contain, as nearly as possible, four males and four females per litter. The study is terminated after weaning of the offspring and dams and pups are then examined.

In a two-generation study, the F1 animals are dosed in the same manner as the F0 animals except that dosage starts after weaning. F1 animals are mated with pups from another litter of the same dose group when the animals reach the age of 13 weeks for rats or 11 weeks for mice. Further conduct of the study is as described above for F0 animals. In certain instances such as treatment-related alterations on litter size or equivocal effects observed in the first mating, it is recommended that the adults be mated again to produce a second litter.

The data to be collected and reported are shown in Table 4.4. TG 416 was updated in 2001 and requires a more detailed report of the data than did TG 415.

Note that offspring in fertility studies are exposed indirectly *in utero* via the blood stream if the substance crosses the blood–placenta barrier, during lactation via the milk if

Table 4.4 *OECD Test guidelines for* in vivo *tests on reproductive toxicity.*

Test (OECD Test Guideline)	Number of animals per sex and dose	Duration, Dosing regime	Parameters measured
Developmental toxicity			
Prenatal developmental toxicity TG 414	>16	Rat, rabbit: from day of implantation to one day before cesarian section (ca. d 6 to 19 after mating)	**Dams**: body weight, food consumption, clinical observations, weight of gravid uterine, number of corpora lutea and implantation sites. Pre- and post-implantation losses, number of dead and viable fetuses, sex and body weight of each fetus, external alterations, skeletal and visceral variations, anomalies and malformations
Toxicity to fertility			
One-generation reproduction toxicity TG 415	>20	Males (rat 70 d, mouse 56 d) + mating period, females: 2 weeks + mating period + pregnancy + lactation until weaning, 7 d/week	**Parental animals**: clinical observations, body weight, food consumption, indices of fertility and gestation, duration of gestation; **offspring**: number of pups, sex ratio, stillbirths, live births, presence of gross anomalies, weight of the litter at birth and on day 4 and 7, thereafter weight of the pups weekly until weaning, viability index **Histopathology**: **parental animals**: All target organs and ovaries, uterus, cervix, vagina, testes, epididymides, seminal vesicles, prostate, coagulating gland, pituitary gland of the high-dose group and controls if not done in other repeated-dose studies
Two-generation reproduction toxicity TG 416	>20	F0-generation: same as in TG 415, F1-generation: after weaning, until 13 (rat) or 11 (mouse) weeks of age + mating period, females: + pregnancy + lactation until weaning of F2-animals	Same as in TG 415 **Additionally**: sperm analysis in parental males, water consumption if test substance is administered in drinking water, estrous-cycle length and normality in parental females, physical or behavioral abnormalities in dams and offspring; offspring: age and body weight at vaginal opening or balanopreputial separation, anogenital distance in F2 pups if triggered by alterations in F1 sex ratio or timing of sexual maturation, functional investigations (motor activity, sensory function, reflex ontogeny), if not included in other studies; additionally ear and eye opening, tooth eruption, hair growth may be recorded. Functional investigations may be omitted in groups showing otherwise clear signs of toxicity, they should not be done on pups selected for mating. Organ weights of uterus, ovaries, testes, epididymides (total and cauda), seminal vesicles with coagulating gland, and their fluids, brain, liver, kidneys, spleen, pituitary, thyroid, adrenal glands, and known target organs. Brain, spleen and thymus from one pup/sex/litter. Calculation of indices of mating, birth and lactation, reporting of time-to-mating, number of implantations, post-implantation loss, and runts

Reproduction/ developmental toxicity screening test TG 421	10	Daily, 7 d/week, 2 weeks prior to mating + during mating period, males: further dosed until 28 days of dosing females: + pregnancy + up to day 3 postpartum (dermal and inhalation: up to day 19 of gestation)	**Parental animals**: body weight (weekly), food consumption, clinical observations, gross necropsy. Weights of testes and epididymides (all animals). Histopathology of testes, epididymides, ovaries, and optionally accessory sex organs and all organs showing macroscopic lesions (all animals); **offspring**: number and sex of pups, stillbirths, live births, runts, gross anomalies, weight of litters on day 0 or 1 post partum and day 4 post partum
Combined repeated dose toxicity study with the reproduction/ developmental toxicity screening test TG 422	10	Daily, 7 d/week, 2 weeks prior to mating + during mating period males: further dosed until 28 days of dosing females: + pregnancy + up to day 3 postpartum (dermal and inhalation: up to day 19 of gestation)	same as in TG 421 **Additionally**: **Parental animals**: Hematology, clinical chemistry, optionally urinalysis, functional observations (each endpoint 5 animals/sex; except when done in a 90-day study). Weights of liver, kidneys, adrenals, thymus, spleen, brain, and heart (5 animals/sex). Histopathology of these 7 organs and of gross lesions, spinal cord, stomach, small and large intestines, thyroid, trachea, lungs, uterus, urinary bladder, lymph nodes, peripheral nerve, bone marrow (5 animals/sex)

the substance is excreted into milk, and the feed as pups start to eat for themselves during the last week of lactation, before scheduled dosing begins.

Screening Tests These tests can be used to provide initial information on possible effects on reproduction and development, either at an early stage of assessing the toxicological properties of chemicals, or for chemicals of concern or for chemicals for which little or no information is available. The tests do not provide complete information on all aspects of reproduction and development. TG 421 is a reproductive/developmental toxicity screening test with shortened exposure of males but histopathological examination of male sex organs is made after 28 days of exposure. TG 422 is the same but includes a more detailed histopathological examination of organs usually evaluated in a sub-acute test (TG 407). Also the optional functional neurotoxicity screen contained in TG 407 is included. At least 3 dose groups and a control are used. Initially, histopathology is performed only in high-dose and control animals but has to be extended to lower-dose groups if triggered by other findings.

Differences of TG 421 and TG 422 to a full fertility test making these screening tests less sensitive as a full test are: males are dosed for a shorter period of time during the pre-mating period, which does not cover the whole spermatogenic cycle; the number of animals and pups is only half, thereby reducing the chance of detecting effects; the tests are terminated already on day 4 postpartum and not at weaning so that delayed effects during the lactation period cannot be registered.

Limit Tests A so-called 'limit test' (applies for testing of toxicity to fertility and developmental toxicity as well as for both screening tests) with a single dose of at least 1000 mg/kg bw can be performed for substances showing no evidence of toxicity at this dose level in repeated-dose studies. Thereby, the number of animals used and the costs compared with a complete test with more dose groups are reduced. The limit test does not apply when human exposure indicates the need for a higher oral dose level to be used. For other types of administration, the physical and chemical properties of the test substance often may indicate the maximum attainable level of exposure (for example, dermal application should not cause severe local toxicity).

4.1.8 Other Test Guidelines

Other regulatory bodies like the European Communities, the US Food and Drug Administration, and the US Environmental Protection Agency have published similar TGs. They are available from the websites listed as references below.

4.1.9 Summary

In vivo tests are required to discover possible adverse effects of chemical substances and pharmaceuticals to humans. They must be carried out according to internationally recognized test guidelines and under the principles of Good Laboratory Practice to ensure validity of the data. Although *in vitro* tests offer a great deal of information on the actions of a substance, *in vivo* tests cannot be replaced, due to the complexity of the organism and the various possibilities of toxification, detoxification, and reactions at the cell surface, within cells, and in sub cellular structures. However, efforts have been

undertaken to reduce the number of animals used for testing purposes. Important *in vivo* tests concern acute toxicity, i.e. toxicity after a single application, skin and eye irritation, skin sensitization, toxicity after repeated administration, and reproductive toxicity which comprises effects on the developing fetus and on fertility of the parent animals. For each of these tests, their most important principles and requirements according to the latest OECD test guidelines are described. Prerequisites are known purity of the test substance, healthy test animals of sufficient number, appropriate control animals and statistical evaluation of the results. The exposure route chosen should be the same as for the anticipated (or known) human exposure. Dose selection is of critical importance. Therefore, long-term studies or full-range studies are preceded by studies with fewer animals to find appropriate doses. Ideally, tests with repeated application should give a dose–response relation and a clear No-observed-adverse-effect level from which safe doses for humans can be extrapolated. A dose that exerts some but not excessive toxicity is necessary for several tests to be valid. Possibilities and limitations of the various tests and the importance of expert judgment when interpreting the findings are highlighted.

References

US Food and Drug Administration GLP Regulations, http://www.fda.gov/ora/compliance_ref/bimo/glp/78fr-glpfinalrule.pdf

Organization for Economic Cooperation and Development GLP Regulations http://www.oecd.org/document/63/0,2340,en_2649_201185_2346175_1_1_1_1,00.html

For Further Reading

European Communities, Directive 67/548/EEC Annex V Part B, http://ecb.jrc.it/testing-methods/

OECD, Guidelines for the Testing of Chemicals, Section 4: Health Effects, http://lysander.sourceoecd.org/vL=3769318/cL=31/nw=1/rpsv/cw/vhosts/oecdjournals/1607310x/v1n3/contp1-1.htm

US Environmental Protection Agency, Office of Prevention, Pesticides and Toxic Substances, OPPTS Harmonized Test Guidelines, Series 870 Health Effects Test Guidelines, http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/index.html

US Food and Drug Administration, Red Book 2000, Toxicological Principles for the Safety Assessment of Food Ingredients, http://www.cfsan.fda.gov/~redbook/red-toca.html

4.2A Mutagenicity Tests *in vivo*

Ilse-Dore Adler

4.2A.1 Introduction

Animal experiments for mutagenicity testing are divided into two groups: analyses of somatic-cell effects and analysis of germ-cell effects. Mutational events in somatic cells are indicative of a carcinogenic potential for the test chemical. Mutational events in germ cells imply a genetic hazard for progeny of exposed individuals.

Basically one can distinguish three classes of genetic alterations: (1) chromosomal aberrations, which are structural alterations of chromosomes and which entail loss or translocation of chromosomal segments; (2) gene mutations, which are changes in the genetic code; and (3) genome mutations, which involve changes in chromosome numbers. Mutagenicity testing is commonly performed for two of these classes, namely chromosomal aberrations and gene mutations. No validated test systems are available for testing genome mutations *in vivo*. However, recent research results on aneuploidy may well lead to recommendations for aneuploidy testing *in vivo*.

The detection of chromosome mutations induced in somatic cells is possible with the **microscopic analysis of structural or numerical chromosome aberrations** in dividing cells or with the **observation of micronuclei in polychromatic erythrocytes of bone marrow**. Occasionally, the micronucleus analysis can be performed with other dividing tissues, e.g. in hepatocytes after stimulation of cell division or in gut cells. During the last few years the peripheral blood micronucleus test has gained increasing popularity.

The detection of **gene mutations in somatic cells** is performed with the **mouse coat-color spot test**. It recognizes mutations in melanocytes of transplacentally treated mouse fetuses heterozygous for a number of coat-color genes.

Additionally, a number of indicator tests are available, e.g. the sister-chromatid exchange (SCE) test in bone marrow cells or the unscheduled DNA synthesis (UDS) assay in hepatocytes. These indicators do not represent genetic endpoints but point indirectly to a possible genetic alteration. They will not be discussed in this chapter.

In germ cells, the induction of **chromosomal aberrations** can be analysed microscopically during various stages of cell division in male germ cells, i.e. **in mitotic divisions of spermatogonia** or in **first meiotic divisions of spermatocytes**. The **dominant lethal test** provides indirect evidence for the induction of chromosomal damage since most of the inherited chromosomal damage leads to embryonic death. The **heritable translocation assay** detects those forms of chromosomal alterations that are compatible with fetal survival, predominantly **reciprocal translocations**, but also Robertsonian translocations and aneuploidies of the sex chromosomes (XO, XXY, and XYY).

The classical method for the analysis of induced **gene mutations in germ cells** is the **mouse specific-locus test**, which detects mutations at seven recessive loci coding for morphological changes in coat color, eye color, and ear size. Tests for other genes have also been developed, e.g. genes for dominant cataracts or charge and activity alterations of specific enzymes. All of these tests can be combined in a '**multiple endpoint approach**' by analysing the progeny of treated parental animals in one experiment for different phenotypes.

Most experiments for the detection of germ-cell mutagenicity are performed with male animals because germ-cell divisions take place during the entire reproductive life span of male animals. The number of available sperm is unlimited and all developmental stages from spermatogonial stem cells to mature sperm can be analysed. In female animals, germ-cell multiplication and some steps in the germ-cell maturation process occur during prenatal development. At birth, the number of oocytes is fixed and only the last steps of germ-cell maturation take place in adult females.

The analysis of **clastogenic** (chromosome breaks) and **mutagenic effects of environmental toxins** is usually performed with **rodents**. Mice of specific genetic constitution

are commonly used and are essential for some of the tests, i.e. for the coat-color spot test and for the specific-locus test. The cytogenetic tests can be performed with other rodents such as rats or Chinese hamsters.

Commonly, animals are treated with the test substances by intraperitoneal injection (ip), gavage (po), or by inhalation. Usually the test substance is not administered in food or drinking water because of the uncertainty of the dose taken up when these routes are used. Intavenous, intramuscular, or subcutaneous injections are seldom used. To maximize the effect, high acute exposure is preferred over subacute or chronic low-dose exposure. Dosing is performed in mg per kg body weight (mg/kg) using a maximum tolerated dose (MTD) in order to avoid systemic toxicity. For dose–response studies, two additional lower doses are employed, e.g. MTD/2 and MTD/5. Sample size and protocols for most of the tests discussed in this chapter are prescribed in the EU Directive Annex V.

4.2A.2 Chromosomal Mutations in Somatic Cells

For the detection of clastogenic effects of a test substance structural chromosomal aberrations are analysed in bone marrow cells, or any other cells, undergoing mitosis.

In principle, the first mitotic division following treatment should be scored for chromosomal aberrations to detect all aberrations, since acentric chromosomal fragments are lost during cell division, often lead to cell death and would, thereby, escape detection in later cell divisions. Microscopically visible aberrations are classified into chromosome-type and chromatid-type aberrations depending on the involvement of both or just one chromatid. Chemicals can have an S-dependent or S-independent mode of action, i.e. they may or may not require a round of DNA synthesis for the translation of the preclastogenic lesion into a visible aberration. A useful description of the classification of chromosomal aberrations was published by Savage (1976). The experimental protocol is described in the EU Directive Annex V.

Any chromosomal analysis requires a certain amount of training. Comparative studies have shown that the judgment on aberrations, especially the discrimination of achromatic lesions (gaps) and true breaks, differs among scorers. Gaps are unstained regions within chromatids, which do not represent DNA discontinuities but instead alterations of chromatin condensation. They should be clearly distinguished from breaks characterized by a dislocation of the resulting fragment (see Figure 4.1).

For the detection of clastogenic effects the micronucleus test with young (polychromatic) erythrocytes is often applied as an alternative to chromosome analyses.

Micronuclei are formed during cell division and can be observed in the daughter cells as a result of acentric fragments or entire misplaced chromosomes. In principle, micronuclei can be scored in any proliferating tissue. However, polychromatic erythrocytes are especially suited for counting micronuclei because these cells have extruded the main nucleus during maturation. This method was developed by Schmid in the sixties and has gained increasing acceptance. Comparative studies in several laboratories have shown that interlaboratory variation is far smaller for the results of micronucleus tests

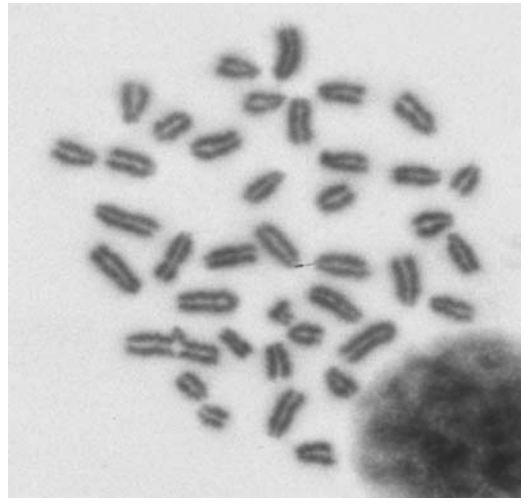


Figure 4.1 *Mouse bone marrow cell in metaphase of mitosis showing a chromatid interchange.*

than for the results of analyses of chromosomal aberrations. Micronuclei are less prone to subjective interpretation than are chromosomal aberrations. The experimental protocol is described in the EU Directive Annex V.

During the past several years, technical modifications of the micronucleus assay have been published, which improved the test considerably. For example, appropriate filtration and centrifugation steps can increase the number of polychromatic erythrocytes in a bone marrow preparation that facilitates automatic scoring by image analysis or flow-cytometric procedures. Furthermore, the fluorescent stain acridine orange is often preferred to the conventional May–Gruenwald–Giemsa stain because the discrimination between young and old erythrocytes is easier in acridine orange-stained slides.

The micronucleus assay has also been modified to recognize genomic mutations since micronuclei are not only formed by acentric fragments but also by entire chromosomes that have lost spindle attachment. To discriminate between the two possible origins of micronuclei, centromere staining by CREST-antibodies or fluorescence-*in situ*-hybridization with centromeric DNA-probes have been applied Figure 4.2).

It was shown also that reticulocytes in peripheral blood can be used for micronucleus scoring. The advantage of using peripheral-blood samples is that they can be obtained from animals in ongoing toxicological studies and several samples spaced by appropriate intervals can be obtained from the same treated animal. Thereby, the numbers of experimental animals can be reduced considerably.

Reactive metabolites of indirect mutagens may be short-lived and may not reach the bone marrow. Since the metabolites are usually formed in the liver, several attempts have been made to perform micronucleus tests with liver cells. For that purpose, liver cells have to be stimulated to undergo cell division either by partial hepatectomy or by treatment with a strong liver toxin. Other variations of the micronucleus test have been used and were reviewed by Hayashi et al. (2000).

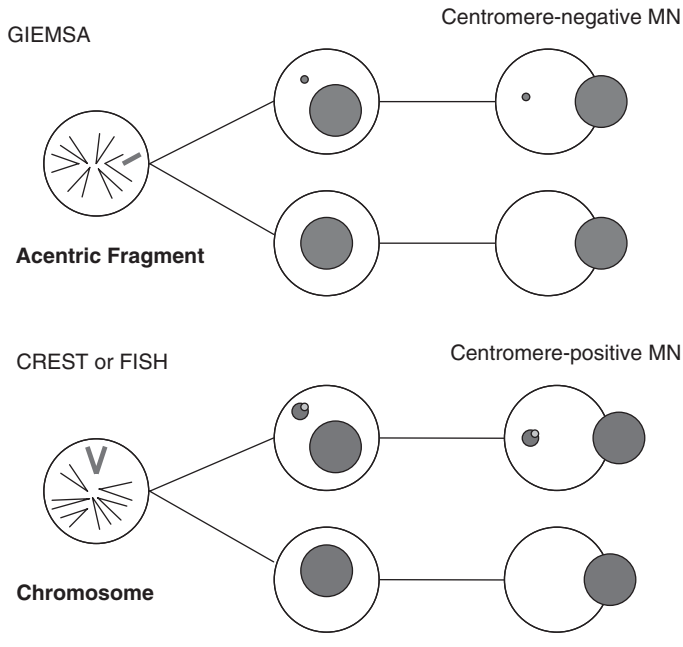


Figure 4.2 *Origin of micronuclei: acentric fragments or entire chromosomes with centromeric region that can be visualized by CREST-antibodies or fluorescence-*in situ*-hybridization (FISH) with pancentromeric DNA-probes.*

In general, an *in vivo* clastogenicity test is indicated to verify a positive result in an *in vitro* test for chromosomal aberrations. The US EPA requires *in vivo* cytogenetic tests in the first stage of testing pesticides and other toxic substances.

4.2A.3 Gene Mutations in Somatic Cells

With the mouse coat-color spot test, gene mutations are detected in pigment-forming melanocytes after transplacental treatment of fetuses. A mutated melanocyte forms a cell clone, which can be recognized in adult animals as a fur patch of different color.

In this test, mice having a specific, genetically determined coat color are crossed. The fetuses are homozygous for nonagouti (black fur color) and heterozygous for 5 or 6 different other recessive fur-color genes. Pregnant females are treated with the test chemical, usually by gavage, between the 8th and 12th day of gestation and are allowed to give birth. At the age of 3 weeks the fur of the progeny is inspected for coat-color spots, which represent mutations in the melanocytes, at the wild-type loci, to the recessive alleles. Microscopic analysis of hair probes allows the recognition of the pigment alterations and, thereby, the identification of the mutated loci. A very good correlation has been found between the results of the mouse spot test and the

mouse specific-locus test. The experimental protocol is described in the EU Directive Annex V.

Owing to the need to perform the mouse spot test with special strains of mice the assay has not been used widely. The newly developed tests with transgenic mice may replace the mouse spot test for the detection of gene mutations in somatic cells *in vivo*. However, the transgenic animal models have not been fully validated and test protocols with transgenic animals have not been introduced in the international guidelines for testing chemicals. Therefore, the mouse spot test *in vivo* remains the only somatic test for gene mutations to verify positive results obtained *in vitro*.

4.2A.4 Chromosome Mutations in Germ Cells

Cytogenetic Analyses in Germ Cells

The analysis of chromosomal aberrations can be performed in mitotic or meiotic cell divisions of male germ cells. The commonly used cells are mitotically dividing spermatogonia.

All mutagenicity tests in germ cells require the exact knowledge of the germ cell's developmental stages of and their timing in the respective experimental animals. They are illustrated for male mice in Figure 4.3 (see also Table 4.1).

Cytogenetic analyses in female germ cells are rarely performed because the mitotic divisions of oocyte propagation and the initial stages of meiosis occur during fetal development of female mammals. At birth, the numbers of oocytes are fixed and they are arrested in a specific stage of meiotic prophase termed the dictiate stage. Oocytes are released from this meiotic arrest in small numbers during each oestrus cycle of the adult female. They undergo meiosis I (MMI) and are ovulated during the second meiotic division (MMII), which is only completed after sperm entry at the time of fertilization. Female germ cells are difficult to collect from the ovarium (MMI) or oviduct (MMII) for cytogenetic analyses and are limited in number so that their use cannot be recommended for routine cytogenetic analyses.

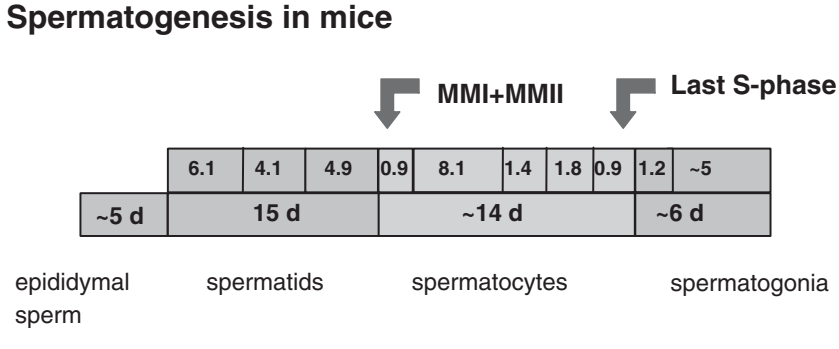


Figure 4.3 *The timing of different stages of spermatogenesis in the mouse. MMI and MMII = first and second meiotic division. The maturation stages progress from right to left.*

The spermatogonial test is best used for cytogenetic analyses in male germ cells. It demonstrates whether or not a chemical or its reactive metabolite reaches the germ cells and causes clastogenic effects. Spermatogonial stem cells divide slowly with a cell cycle of 6–8 days. The resulting differentiating spermatogonia divide every 26–32 h. The majority of analysable mitoses represent type B spermatogonia. Analyses of chromosomal aberrations in mitoses of differentiating spermatogonia resemble cytogenetic analyses in bone marrow mitoses. The same types of aberrations can be scored. The experimental protocol is described in the EU Directives Annex V.

Microscopic analyses in first meiotic metaphase chromosomes is more difficult because paired homologous chromosomes form bivalents with less clearly defined structures. Depending on the interval between treatment and cell sampling, chromosome-type aberrations (reciprocal translocations, Figure 4.4, right) or chromatid-type aberrations (gaps, breaks, fragments, and exchange configurations) can be scored (Figure 4.4, left). Often, unpaired homologous chromosomes, either autosomes or sex chromosomes, are also noted. They are not the result of clastogenicity but represent failures of homologous pairing or premature separation of homologous chromosomes.

At intervals of 30 or more days between treatment of the males and sampling of germ cells, the only types of aberrations analysed at the first meiotic division are reciprocal translocations since any unstable aberrations induced in spermatogonial stem cells are lost during subsequent mitotic divisions. These reciprocal translocations are recognized as multivalents, e.g. as either chains of four chromosomes or rings of four chromosomes (Figure 4.4, right) depending on the number of cross-over events between the involved chromosomes that take place during meiotic prophase. In contrast to ionizing radiation, the majority of chemical mutagens give negative results in this assay. Only a few S-independent chemicals, e.g. bleomycin, adriamycin, or triethylenemelamine, yielded positive results. Therefore, the cytogenetic translocation test is not used routinely for *in vivo* testing. It is applied only when indicated by the observation of an S-independent clastogenic effect in *in vitro* systems.

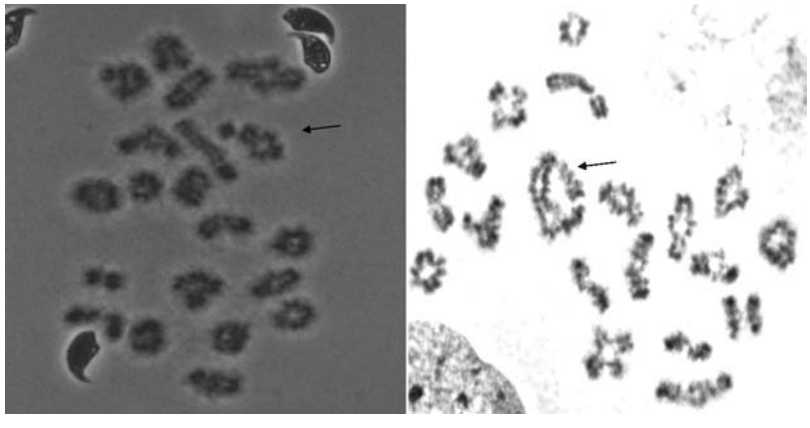


Figure 4.4 *Left: Bivalent with chromatid fragment (arrow). Right: reciprocal translocation (ring of four chromosomes, arrow), in mouse spermatocytes at first meiotic metaphase.*

The analysis of first meiotic metaphases for clastogenic effects induced during meiotic prophase is performed 1–12 days after treatment depending on the duration of the various stages of meiotic prophase (Figure 4.3). Only chromatid-type aberrations can be scored. While ionizing radiation induces aberrations in all prophase stages, the effects of chemical mutagens are usually limited to the premeiotic S-phase. The method of meiotic chromosome analysis for chromatid-type aberrations is too complicated for routine testing because aberration types are difficult to recognize in the complex meiotic bivalents and require that the scorers have considerable experience. The method is used only when scientifically indicated.

Chromosomal aberrations induced in post-meiotic germ cell stages, i.e. spermatids and spermatozoa, can only be analysed during the first cleavage division after fertilization. The preparation technique for the isolation of the first cleavage division zygotes is very tedious, which prevents routine use of this methodology. It is used only when scientifically justified.

The spermatid micronucleus assay determines micronuclei in early spermatids. These micronuclei are induced usually during the meiotic prophase as acentric fragments or during the meiotic divisions as lagging chromosomes. Similar to the somatic micronucleus test, the origin of the micronuclei observed in spermatids can be determined by centromere-specific staining procedures. The spermatid micronucleus test has been used in Europe only by a small number of laboratories and has not been included in the international guidelines for chemical testing.

Information on clastogenic effects of chemicals from cytogenetic germ cell tests leads to categorization in mutagenicity classes and, thereby, to regulation (EU category 2).

Dominant Lethal Test

Dominant lethal mutations are due to structural chromosome aberrations in germ cells, which are not compatible with embryonic survival, and which give indirect evidence of clastogenic effects of the test substance.

Chromosomal aberrations that lead to loss of genetic material cause embryonic death after fertilization at the time of implanation in the uterus. The method was described by Kaplan and Lyon as early as in 1953 and was recommended for mutagenicity testing by Bateman in 1966.

Dominant lethal experiments do not require animal strains with a specific genetic background. However, the litter size of the strains used should be relatively large (9–12 pups per female) and constant. After treatment, male mice or rats are mated to untreated virgin females at intervals of 4–7 days. In mice, successful copulation is indicated by a vaginal plug, whereas in rats vaginal smears have to be taken in order to prove copulation has occurred. The pregnant females are sacrificed at mid-pregnacy and the uterus contents are inspected for live and dead implants. The dominant lethal effect is expressed as 1 minus the ratio between the number of live implants in the treated group divided by the number of live implants in the control group expressed on a percentage basis. These data includes pre- and post-implantation losses. The actual dominant lethal mutations, however, are the post-implantation losses, i.e. resorptions, early deaths, and late deaths,

Table 4.5 *Mating scheme to sample different stages of mouse spermatogenesis after acute exposure.*

Mating intervals (days)	Treated spermatogenic stages
1–7	Spermatozoa
8–14	Late spermatids
15–21	Mid-early spermatids
22–28	Spermatocytes
29–35	Spermatocytes
36–42	Differentiated spermatogonia
43–many months	Spermatogonial stem cells

since pre-implantation losses may be caused by physiologically based effects such as reduced sperm counts of the treated males. Only at high rates of dead implants and normal pregnancy frequencies may pre-implantation loss be due to multiple structural chromosome aberrations per gamete and, thus, represents a true dominant lethal effect. Therefore, data are often presented as percent dead implants or as dead implants per female.

The increases of dead embryos over a concurrent control at the various mating intervals reflect clastogenic effects at specific germ-cell developmental stages, namely mature sperm, mid-early spermatids, late spermatids, spermatocytes, or differentiating spermatogonia (Table 4.5). This stage specificity is characteristic of individual chemicals, or chemical classes. In a large collaborative study, a standard protocol for dominant lethal tests and their statistical evaluation was validated by Ehling et al. (1978). Alternatively, the dominant lethal test can be performed by continuous treatment of the entire period of spermatogenesis followed by only one week of mating. This protocol saves on female animals; however, it only gives a yes-or-no answer and loses the information on stage sensitivity. The experimental protocols are described in the EU Directive Annex V.

Information on clastogenic effects of chemicals from dominant lethal tests leads to categorization in mutagenicity classes and, thereby, to regulation (EU category 2). The US EPA requires dominant lethal tests in the second stage of testing pesticides and other toxic substances.

Heritable Translocation Assay

The number of live progeny with reciprocal translocations is determined in the heritable translocation assay.

In contrast to the dominant lethal test, the heritable translocation assay detects structural chromosome aberrations that do not lead to loss of genetic material, e.g. reciprocal translocations (Figure 4.5) and, therefore, are compatible with fetal survival. Germ cells carrying reciprocal translocations can result in living offspring, which generally do not show any physical malformations but do display reduced fertility.

The analysis of progeny (F1-generation) of treated parental animals can be performed in two ways. The classical method described first by Hertwig in 1940 only analyses male

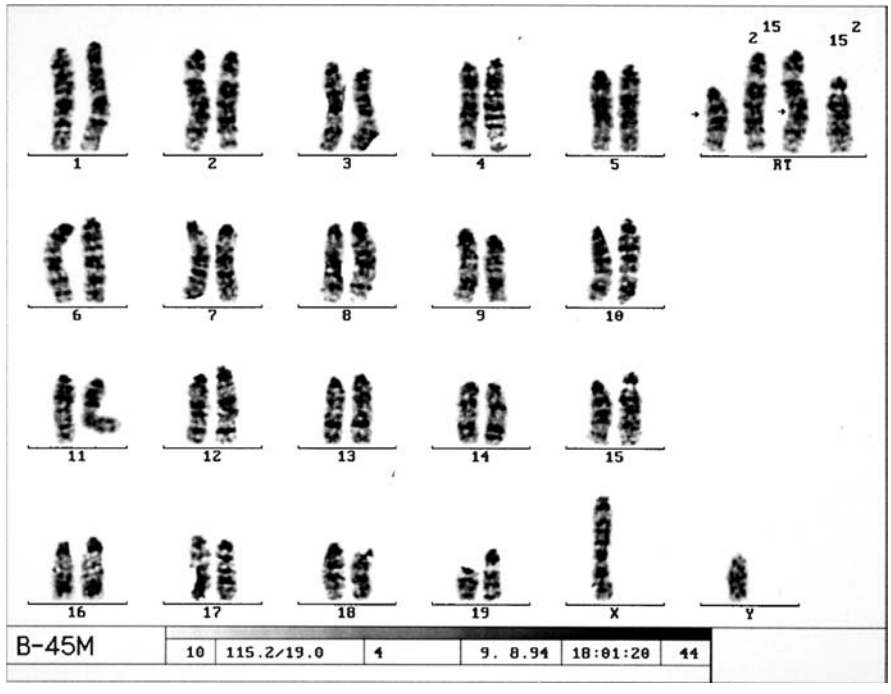


Figure 4.5 *Reciprocal translocation between mouse chromosomes 2 and 15. Banded karyotype analysed semiautomatically using the software IKARUS by MetaSystems.*

F1-progeny and uses reduced fertility as the selection criterion for the suspect reciprocal translocation carriers. During the meiotic segregation of the translocated chromosomes and their normal homologues three classes of gametes are formed, i.e. unbalanced gametes, balanced translocation-carrying gametes, and normal gametes at the proportions 2:1:1. After fertilization, the unbalanced gametes lead to embryonic death at the time of implantation of the embryo *in utero*. The fertility of F1-translocation carriers is, therefore, reduced to about 50% (semi-sterility). Complete sterility is observed in XXY mice (human Klinefelter syndrome), and in F1-females with certain types of reciprocal translocations, namely the c/t types with breaks near the centromere and telomere of the two chromosomes involved. The presence of a reciprocal translocation in a semi-sterile or sterile F1-animal is confirmed by cytogenetic analysis of first meiotic divisions in the form of translocation multivalents (Figure 4.4, right).

In a varation of the heritable translocation test female F1-progeny are also included in the analysis. Thereby, no additional females are needed for the fertility tests with F1-males and the number of analysed F1-progeny is almost doubled. The only drawback is that it takes another generation (F2-sons) to identify translocation-carrying F1-females. The cytogenetic verification of the presence of a translocation in oocytes is impossible due to the low numbers of available cells at MMI. Since all translocation carriers transmit the translocation to half of their offspring, suspect F1-translocation females are mated to collect 8–10 F2-sons for confirmation of the translocation in the F1-female. By this

protocol it is possible to also identify XO and XXY progeny, which result from nondisjunction of the sex chromosomes.

Another variant of the heritable translocation assay omits fertility testing. All F1-males are subjected to cytogenetic analysis without fertility prescreening. This protocol also detects carriers of Robertsonian translocations, which are lost from recognition in fertility testing because fertility impairment is only 20–30% in carriers of Robertsonian translocations.

The last step of the heritable translocation assay is the identification of the chromosomes and break sites within involved chromosomes. Giemsa-banded mitotic chromosomes from bone marrow or peripheral blood preparations are karyotyped with the aid of a computer software (Figure 4.5).

The stage specificity observed in the dominant lethal assay is important information for the planning of a heritable translocation test. Basically, both tests determine clastogenic effects in germ cells transmitted to the next generation. In the dominant lethal assay the transmitted clastogenic effect is embryonic death. In the heritable translocation assay the transmitted clastogenic effects are viable offspring, which carry reciprocal translocations and nondisjunction effects such as XO, XXY, and XYY, which actually pose a genetic hazard for future generations. Thus, the results of heritable translocation assays are important for the quantification of genetic risk.

The experimental protocol is described in the EU Directive Annex V. Information on clastogenic effects of chemicals from heritable translocation assays leads to categorization in mutagenicity classes and, thereby, to regulation (EU category 2). The US EPA requires heritable translocation assays in the third stage of testing pesticides and other toxic substances.

4.2A.5 Gene Mutations in Germ Cells

Gene mutations can be detected by examining 7 recessive loci that code for externally visible traits such as coat color and pattern, eye color, and ear size.

The specific locus test was developed in Oak Ridge (USA) with a specific mouse test stock (T-stock). T-stock mice are homozygous for 7 recessive alleles of genes that determine the coat color and pattern, eye color, and ear size (Figure 4.6). The homozygous T-stock animals have a white coat, red eyes, and small ears. Wild-type mice are homozygous for the dominant wild-type alleles of the 7 loci and have a brown coat, brown eyes, and normal size ears. Crosses of wild-type and T-stock mice produce offspring that have the phenotype of the wild-type parent because they are heterozygous for the dominant wild-type alleles.

At Harwell, UK, the HT-stock was developed, which is homozygous for 6 recessive loci. Four of these loci code for coat color, one for the length of extremities, and one for hair structure. When studying radiation-induced mutations the T-stock loci showed a higher mutation frequency than did the HT-stock loci.

Commonly, male wild-type mice are treated with the test chemical and mated to test-stock females. A germ-cell mutation in one of the wild-type alleles to a

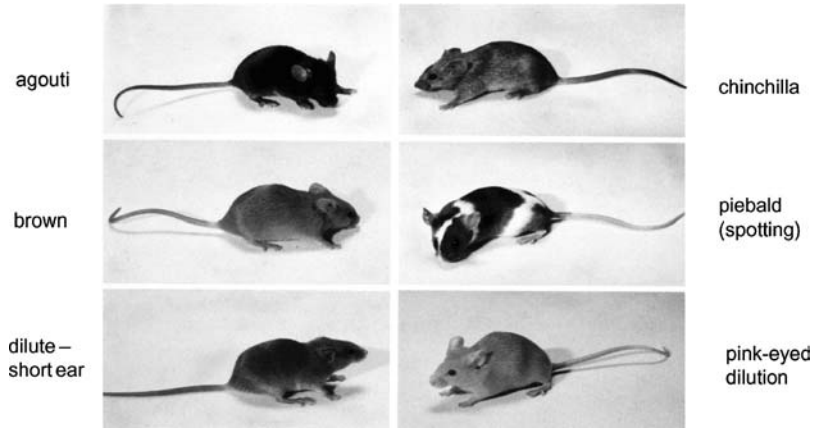


Figure 4.6 *Phenotypes of mutant offspring in the specific locus test. [Reprinted from Obe. Copyright 1984) with permission from Springer-Verlag.]*

recessive allele produces a phenotypically altered mutant offspring. The mutant is homozygous for the respective allele and displays the characteristics of the respective gene (Figure 4.6).

By taking advantage of the time intervals over which spermatogonia develop to mature sperm it is possible to determine at which stage in spermatogenesis mutations are chemically induced. Thus, T-stock females can be mated with chemically treated wild-type males at weekly intervals similar to the mating scheme of the dominant lethal assay (Figure 4.3). It was shown that most of the mutations induced in post-spermatogonial germ cells are inter- or intragenic deletions, indicating that these are clastogenic events. After 42 days of mating any mutations can be attributed to effects on spermatogonia and most of these are gene mutations.

Induction of mutations in spermatogonial stem cells is of special importance for the assessment of genetic risk. In contrast to mutations in post-spermatogonial stages, mutations in stem cells are reproduced and transmitted during the entire process of spermatogenesis. Thus, mutations induced in stem cells pose a permanent genetic risk throughout the reproductive life of the treated male.

The specific locus test cannot be applied routinely for every chemical because it requires large numbers of animals to detect an increased mutation rate over the spontaneous mutation rate of about one in 20 000 progeny. However, the specific-locus test permits the quantitative determination of the genetic risk posed by germ cell mutagens (Ehling, 1991).

The quantification of the genetic risk from a single exposure can be determined based on the mutation rate in spermatogonial stem cells. The quantification of the genetic risk from chronic exposures requires additional consideration of the mutation rates in post-spermatogonial germ-cell stages.

To enlarge the number of analysed loci, the ‘multiple-endpoint approach’ was developed (Ehling, 1991). In addition to the 7 recessive loci, this approach analyses 30 dominant genes that code for the formation of cataracts, 23 genes that code for charge

alterations in certain enzymes, and 12 genes that code for changes in enzyme activity. In radiation experiments, it could be shown that the mutation frequencies for the 7 recessive loci and for the enzyme activity genes were similar while for the cataract genes and the genes for changes in enzyme charge they were one order of magnitude lower. This difference in sensitivity of the various loci, especially the difference between recessive and dominant loci, underlines the necessity to increase the number of tested loci in order to obtain mutation rates at a representative sample of the entire genome for extrapolation of the results from animals to humans.

The quantification of genetic risk can be performed either indirectly or directly. Calculation of the 'doubling dose', namely the dose that induces as many mutations per generation as occur spontaneously, is an indirect method to determine the extent of the exposure-derived increase of the mutation rate. In a more direct method it is possible to determine the dominant mutations in a specific organ system, such as the eye lens, and to calculate the anticipated numbers of new mutations in the first generation following exposure to the mutagen.

Information on the induction of gene mutations by a chemical using the different variations of the specific-locus test lead to categorization into mutagenicity classes and, thereby, to regulation (EU category 2). The US EPA requires specific-locus test data in the third stage of testing pesticides and other toxic substances in order to quantify the genetic risk for humans.

4.2A.6 Summary

- The classical *in vivo* mutagenicity tests, i.e. the cytogenetic translocation test, coat-color test, dominant lethal test, heritable translocation test, and specific-locus test, were developed during the 1950s and 1960s to determine the genetic effects of ionizing radiation.
- Subsequently these methods were applied to test the mutagenicity of chemicals. The original intent of mutagenicity testing was to protect human progeny from adverse effects of mutational events in germ cells.
- With the realization that mutational events may lead to carcinogenesis, the primary interest in mutagenicity testing shifted to detecting the ability of chemicals to induce cancer. The mutagenicity data were used as a step in the strategy aimed at preventing chemically induced cancers. Tests performed *in vitro* were developed for the primary purpose of detecting potential genotoxic carcinogens as quickly and as reliably as possible.
- Mutagenicity tests were adopted worldwide because a good correlation between mutagenicity and carcinogenicity was found for certain classes of chemicals. For almost two decades, the original purpose of mutagenicity testing, namely to prevent genetic burden to future generations, was largely neglected.
- The emphasis on mutagenicity testing is again undergoing reconsideration. Despite a wealth of *in vitro* mutagenicity tests with microorganisms and mammalian cells in culture, the data do not reliably predict the actual carcinogenic or mutagenic hazard of chemicals for humans because of the complexity of metabolic processes in mammalian organisms, the multiplicity of stages in cancer development, and the biological

dynamics of germ-cell development. Currently, all internationally accepted criteria for the categorizing carcinogens and germ-cell mutagens require data derived from *in vivo* testing.

- Commercially available transgenic animal models such as the 'Muta^{TM}Mouse' (HazeltonLaboratories Corp.) or the 'Big Blue®' mouse (Stratagene) have introduced yet a new type of mutagenicity testing in which mutations resulting from *in vivo* exposure are detected using simple, rapid molecular biological methods *in vitro.* New molecular methodologies to detect mutations at restriction enzyme-cutting sites or within minisatellite regions of the DNA may also provide useful tools to complement or eventually replace the morphological mutagenicity tests (Singer et al., 2006).

References

Ehling, U.H., Machemer, L., Buselmaier, W., Dycka, J., Frohberg, H., Kratochvilova, J., Lang, R., Lorke, D., Müller, D., Peh, J., Röhrborn, G., Roll, R., Schulze-Schenking, M., and Wiemann, H. (1978) Standard Protocol for the Dominant Lethal Test on Male Mice. *Arch. Toxicol.*, **39**, 173–185.

Ehling, U.H. (1991) Genetic Risk Assessment. *Annu. Rev. Genet.*, **25**, 255–280.

EU Directive: Annex V to Directive 67/548/EEC on the Classification, Packaging and Labelling of Dangerous Substances: Testing Methods. Official J. Eur. Commun., No. L225, http://ecb.jrc.it/testing methods.

Hayashi, M., MacGregor, J.T., Gatehouse, D., Adler, I.-D., Blakeley, D.H., Dertinger, S.D., Krishna, G., Morita, T., Russo, A., and Sutou, S. (2000) In vivo Rodent Erythrocyte Micronucleus Assay. *Environ. Mol. Mutagen.*, **35**, 234–252.

IKARUS. Metasystems GmbH, D-68804 Altlussheim, Germany.

Obe, G. (1984) Mutations in Man. Springer-Verlag, Heidelberg.

Savage, J.R.K. (1976) Classification and Relationships of Induced Chromosomal Structural Changes, *J. Med. Genet.*, **13**, 103–122.

Singer, T.M., Lambert, I.B., Williams, A., Douglas, G.R., and Yauk, C.L. (2006) Detection of Induced Male Germline Mutation: Correlation and Comparisons between Traditional Germline Mutation Assays, Transgenic Rodent Assays and Expanded Simple Tandem Repeat Instability Assays. *Mutat. Res.*, **598**, 164–193.

For Further Reading

Allen, W.A., Bridges, B.A., Lyon, M.F., Moses, M.J., and Russell, L.B. (1990) Biology of Mammalian Germ Cell Mutagenesis. Banbury Report 34, Cold Spring Harbor Laboratory Press, New York, USA.

Fahrig, R. (1993) Mutationsforschung und Genetische Toxikologie. Wissenschaftliche Buchgesellschaft, Darmstadt.

Hollaender, A. and de Serres, F.J. (1971) Chemical Mutagens. Principles and Methods of Their Detection, Vol. 1, Plenum Press, New York, USA.

Hsu, T.C. (1982) Cytogenetic Assays of Environmental Mutagens. Allanheld, Osmun Publishers, Totowa, USA.

Kilbey, B.J., Legator, M., Nichols, W., and Ramel, C. (1984) Handbook of Mutagenicity Test Procedures. Elsevier Science Publishers, Amsterdam, Holland.

Venitt, S. and Parry, J.M. (1984) Mutagenicity Testing - a Practical Approach. IRL Press, Oxford, UK.

4.2B *In vitro* Tests for Genotoxicity

Ulrich Andrae and Günter Speit

4.2B.1 Introduction

Genotoxicity is a generic term that relates to the ability of a chemical or physical agent to induce mutations (gene mutations, structural chromosome mutations, genome mutations) or so-called indicator effects that are mechanistically associated with the formation of mutations (e.g., induction of DNA modifications, DNA repair, or recombination). In humans, germ-cell mutations increase the risk for offspring with heritable disease, and somatic mutations may lead to the development of cancer. Genetic damage may be the result of a direct reaction of a chemical or its metabolite with DNA, resulting in adduct formation, or an attack on DNA by reactive oxygen species. Moreover, genetic damage may also be the consequence of an interference with critical proteins involved in the maintenance or replication of DNA or the accurate transmission of genetic information, such as the proteins of the spindle apparatus. Therefore, genotoxicity should not be considered a synonym for DNA reactivity.

Genotoxicity tests of chemicals are conducted in order to obtain direct or indirect information on a compound's capability to induce mutations. The demonstration of genotoxicity may provide an alert that the compound might also be carcinogenic. Genotoxicity tests *in vitro* are also frequently employed to clarify the molecular mechanisms underlying the genotoxic and carcinogenic effects of chemicals.

Xenobiotic Metabolism is a Challenge when using in vitro *Test Systems for Genotoxicity*

A fundamental problem of all *in vitro* test systems is that uptake, distribution, and excretion cannot be simulated. The certainty with which the results of *in vitro* tests of genotoxicity can be extrapolated to people is limited by the inability to reproduce in a test tube the complex pharmacokinetics that regulate the disposition of chemicals in the body. In animals and man, the type and the extent of the effects of chemicals depend on various toxicokinetic parameters, i.e. uptake of the chemical into the organism, distribution into individual organs, metabolism, and excretion.

For the most part chemicals exert their toxicity in an organism only after metabolic activation. However, differences between species, individual organs, and the cell types within an organ may result in a diversity of metabolic pathways for any given chemical. Clearly, it would not be possible to use a single cell type for use in a genotoxicity test that would provide mutagenesis data applicable to all other cells. Since bacteria and the majority of the mammalian cells utilized in genotoxicity studies lack certain xenobiotic-metabolizing enzymes, most test systems depend on the addition of subcellular fractions of cells that contain the appropriate enzymes, or cells that contain the required enzyme activities. Ultimately, the choice of the test system and the suitable metabolizing system

depends on the nature of the chemical to be studied. Critical to the decision is the structure of the chemical, which may give a clue to the possible involvement of specific metabolic pathways. Information on organ-specific toxic effects of a chemical should also be taken into consideration when choosing the test system to be employed.

Genetic Endpoints of in vitro Test Systems for Genotoxicity

The genotoxic effects exerted by chemicals include alterations of the genetic material that are transmitted to the progeny of the treated cells. These heritable changes are termed **mutations**. However, genotoxic effects also include effects that can be detected in the treated cell itself, e.g. modifications of DNA structure, or certain responses of the cell occurring as a specific consequence of the damage to its genetic material. Thus, mutagenic effects represent a specific subgroup of genotoxic effects. The terms **genotoxic** and **mutagenic** or **genotoxicity test** and **mutagenicity test** should not be considered synonymous.

4.2B.2 Xenobiotic Metabolism *in vitro*

Primary Mammalian Cells

Primary mammalian cells, i.e. cells isolated from an organ and maintained in cell culture for a limited period of time, generally reflect the xenobiotic metabolism of the respective organ quite well, at least during the first hours after their isolation.

A disadvantage of these cells is that they generally exhibit a very low proliferative activity and, therefore, have a limited usefulness for the detection of heritable changes, i.e. mutations. However, these cells can be used in test systems for the detection of DNA strand breaks or DNA repair as indicators of a potentially mutagenic effect of the chemical. The most widely used of these assays is the detection of DNA repair synthesis in primary cultures of rat hepatocytes.

Cell Lines

Cell lines, i.e. descendants of primary cells that have gained the capacity to grow indefinitely, generally show a markedly reduced capacity for organ-specific xenobiotic metabolism.

The decrease in metabolic competence is a consequence of the close association between immortalization and dedifferentiation, i.e. the loss of organ-specific functions. Particularly affected are certain CYPs and sulfotransferases. An extreme example is the frequently employed V79 hamster cell line, which does not express these enzyme activities at all. A promising tool to overcome the problem that metabolic competence and proliferative activity are mutually exclusive is provided by recombinant cell lines genetically engineered to express genes of well defined xenobiotic-metabolizing enzymes.

Bacteria

Bacteria generally exhibit no activities of xenobiotic-metabolizing CYP enzymes.

Neither have epoxide hydrolase, UDP glucuronosyl transferase, or xenobiotic-metabolizing sulfotransferase been found. By contrast, several other enzymes that can play important roles in the metabolism of xenobiotics, such as nitroreductases, amidases, acetyltransferases, and glutathione transferases, have been detected in bacteria (Table 4.6).

Subcellular Organ Fractions

In most routine test systems employing cell lines or bacteria, subcellular fractions from rat liver are added to the cells as an exogenous metabolizing system in order to compensate for the lack of metabolic competence in the cells.

Cell fractions prepared from the livers of rats or mice treated with a mixture of polychlorinated biphenyls, such as Aroclor 1254, or phenobarbital plus β-naphthoflavone, which induce a broad spectrum of various CYPs, are usually employed. For preparation the liver is homogenized and the homogenate centrifuged at 9000 g in order to remove larger cell fragments, i.e., nuclei and mitochondria. The resultant supernatant, the so-called ''S9 fraction'' (**9**000 g **s**upernatant), comprises the membranes of the disrupted endoplasmic reticulum, which are the so-called 'microsomes,' and the cytosol of the cells. The microsomes contain the various CYP forms as well as other enzyme activities, such as cytochrome P450 reductase, UDP glucuronosyltransferase and epoxide hydrolase. The cytosolic portion of the S9 fraction contains, among other enzymes, the major part of the activities of various transferases, such as acetyl-, sulfo-, and glutathione transferases and cytosolic epoxide hydrolase. The cofactor NADPH, which is required for activity of the monooxygenases, is produced by enzymes present in the S9, i.e. glucose-6-phosphate dehydrogenase or isocitrate dehydrogenase, from NADP and glucose-6-phosphate or isocitrate, which are added to the S9 fraction. The S9 fraction, NADP, and isocitrate or glucose-6-phosphate, respectively, make up the so-called 'S9 mix'.

Table 4.6 *Presence of xenobiotic-metabolizing enzymes in various cell types employed in genotoxicity testing* in vitro.

	Cell type		
Enzyme activity	Hepatocytes	V79	*Salmonella*
Cytochrome P450-dependent monooxygenase (CYP)	++		
UDP glucuronosyltransferase	++		
Glutathione *S*-transferase	++	++	+
N-Acetyltransferase	++		+
Cytosolic sulfotransferase	++		
Epoxide hydrolase	++		

The activities of the conjugating enzymes in the S9 mix and the resultant incubation mixture are very low due to the dilution of the cofactors such as acetyl-coenzyme A, 3′-phosphoadenosine 5′-phosphosulfate, and UDP-glucuronic acid, which are required for enzyme activity. As a consequence, the proportion between activating and inactivating reactions as they occur in intact liver cells is generally shifted towards an over-representation of certain activating reactions in many test systems employing S9 mix. Furthermore, specific activating reactions catalysed by transferases in the S9 mix may not take place for the same reason.

A further problem associated with the use of S9 mix for metabolic activation results from the fact that metabolic activation occurs outside the target cells. Very-short-lived electrophilic metabolites or reactive species, which are unable to penetrate the cell membrane due to their electric charge, may not be able to interact with the genetic material of the cells. Their genotoxicity may therefore go undiscovered. Frequently, the S9 mix also exerts toxic effects on the cells, and even genotoxic effects have been observed. They probably originate from the formation of reactive lipid peroxides in microsomal membranes. Therefore, the treatment period with test compounds in the presence of S9 mix is usually limited to 3–6 h.

Co-cultivation Techniques

An alternative to the metabolic activation of test compounds by exogenous cell fractions is provided by co-cultivation techniques. Proliferating cells that allow the detection of mutations, such as V79 cells, are cultured together with metabolically competent cells, such as primary rat hepatocytes. A physical contact between the two cell types is usually required for an efficient transfer of the reactive metabolites formed. This approach is limited when the reactive metabolites cannot leave the activating cell because they are too reactive or electrically vcharged or when they are efficiently inactivated enzymatically in the activating cell.

4.2B.3 Test Systems Employing Bacteria

Tests for the Induction of Gene Mutations

Bacterial Gene Mutation Test

This assay, also colloquially referred to as the 'Ames test,' is the most frequently employed *in vitro* mutagenicity test. The indicator organisms for the detection of induced mutations are various strains of *Salmonella typhimurium* and *Escherichia coli* which, as a consequence of mutations in genes coding for enzymes for the biosynthesis of an amino acid (histidine or tryptophan, respectively), have lost the ability to grow on agar lacking this amino acid. Mutagenic chemicals induce reverse mutations in the genome of these cells, i.e. the mutated cells (revertants) regain the capability to synthesize histidine or tryptophan and to grow and form visible colonies in the absence of the respective amino acid in the agar.

Table 4.7 *Properties of the most frequently used S. typhimurium and E. coli strains.*

Tester strain	Repair defect	Plasmid	Base pair at primary reversion site	Type of mutation detectable
S. typhimurium				
TA1535	*uvr*B		GC	Base substitution
TA100	*uvr*B	pKM101	GC	Base substitution
TA1538	*uvr*B		GC	Frame shift
TA98	*uvr*B	pKM101	GC	Frame shift
TA1537	*uvr*B		GC	Frame shift
TA97	*uvr*B	pKM101	GC	Frame shift
TA102		pKM101, pAQ1	AT	Base substitution, frame shift
E. coli				
WP2 *uvr*A	*uvr*A		AT	Base substitution
WP2 *uvr*A (pKM101)	*uvr*A	pKM101	AT	Base substitution

The alterations of the nucleotide sequence of the DNA that resulted in amino acid dependence are precisely defined for the individual tester strains. Thus, the ability of a compound to give rise to reversions in specific tester strains enables conclusions to be drawn regarding the types of mutations induced. Whereas certain mutations are specifically reverted by base-substitution mutations, others are only reverted by frame-shift mutations (Table 4.7). The majority of the *Salmonella* strains have guanine–cytosine (GC) base pairs at the primary reversion site. In contrast, the strain TA102 and the *E. coli* WP2 strains have an adenine–thymine (AT) base pair at the primary reversion site, which renders them more sensitive to oxidizing and cross-linking mutagens. The target gene for reversion is carried on the bacterial chromosome with the exception of strain TA102, where it is located on a multicopy plasmid (pAQ1).

In addition to the mutations resulting in amino acid dependence, most tester strains carry additional mutations that were specifically introduced in order to increase the mutabilty of the cells by chemicals. One of these mutations (*rfa*) causes the partial loss of the lipopolysaccharide layer on the outer of the two cell membranes of *S. typhimurium.* As a consequence, the uptake of large hydrophobic chemicals into the bacteria is facilitated. Further mutations (*uvr*A or *uvr*B) result in a defective repair of DNA damage, thereby enhancing the mutagenicity of many chemicals. Moreover, several of the tester strains carry the plasmid pKM101, which allows the replicative bypass of DNA lesions and, thereby, increases the yield of mutations.

Table 4.7 gives an overview of the most frequently employed *S. typhimurium* and *E. coli* tester strains and the types of mutations that can be detected.

For routine testing of chemicals, the use of a test battery comprising the base substitution strains TA1535 and TA100, the frame-shift strains TA1537 (or TA97) and TA98, and the strain TA102 has gained general acceptance. Instead of TA102, the *E. coli* strain WP2 *uvr*A or WP2 *uvr*A (pKM101) may be employed.

Most commonly, the bacterial gene-mutation test is performed as so-called **plate incorporation assay.** Bacteria, the S9 mix, and the dissolved test material are mixed with soft agar and plated. In another common variant of the test, the **preincubation assay**,

bacteria, S9 mix, and the test material are first preincubated together for several minutes and subsequently mixed with the soft agar and plated. This variation has been shown to be more sensitive than the plate incorporation variant in several cases and is frequently used for the repeat test after a first plate incorporation test. Plating is done on agar that contains only traces of histidine or tryptophan (minimal agar). These traces enable the bacteria to pass through a few rounds of cell replication. During this limited-growth period induced DNA lesions become fixed as mutations. In contrast to the nonmutated cells, the revertants are capable of continuing growth on the minimal agar. Within 2–3 days they form macroscopically visible colonies that clearly stand out against the even, thin lawn of the histidine- or tryptophan-requiring bacteria. An increase in the number of revertant colonies over the number of spontaneously growing colonies following exposure of the cells to a test compound is taken as an indication of a mutagenic effect of the latter.

The widespread use of the bacterial mutagenicity assay, which has resulted in the assembly of a huge database, especially for the *Salmonella* tester strains, is largely due to the simplicity and speed with which mutagenicity can be detected. The system is very sensitive for the detection of mutations because of the tailor-made properties of the cells and the use of very large numbers of cells (about 10^8 per treatment), which allows the detection of small increases in mutant frequency by the test compound. A disadvantage of the assay is the very limited capacity of the bacteria to metabolize xenobiotics, which makes the use of exogenous S9 mix essential. Moreover, the structure of the bacterial chromosome is different from the much more complex structure of the chromatin in mammalian cells. As a consequence, some types of mutagens, e.g. compounds eliciting chromosomal mutations in mammalian cells as a consequence of an interaction with DNA-associated proteins, cannot be detected in bacterial test systems.

DNA Repair Assays

Bacterial DNA-repair assays are used to identify the DNA-damaging activity of chemicals by comparing their effects on two strains of bacteria differing only in their capacity for the repair of DNA lesions.

Usually, the test organisms employed are pairs of specific *Escherichia coli* or *Bacillus subtilis* strains. The numerous variants of the bacterial repair assays make use of two basic strategies for the detection of genotoxic effects:

- The demonstration of growth-inhibition zones of different size following exposure of the two strains to the test compound.
- The demonstration of different sensitivity to lethality of the test substance in the two strains.

The growth-inhibitory effects of compounds are frequently analysed using the so-called filter disk or **agar diffusion method**. The lethality is determined by measuring the number of cells capable of forming colonies following treatment.

The DNA-repair tests, or more precisely, the tests for differential killing have been less frequently used in recent years because they yield many ‘false positive’ results, probably

because several of the strains employed respond differentially not only to genotoxic, but also to toxic, effects of chemicals. Bacterial DNA-repair tests must not be confused with DNA-repair tests using mammalian cells, which are based on a completely different approach.

4.2B.4 Test Systems Employing Mammalian Cells

Mammalian Cell Gene-Mutation Tests

In gene-mutation assays, the mutagenicity of a test compound in cells becomes apparent from altered cellular properties, which are expressed after treatment and which are transmitted to the progeny of the cells. In contrast to the bacterial gene-mutation test, where the altered cell function is due to a reverse mutation, the common gene-mutation assays employing mammalian cells detect forward mutations, i.e. mutations that are associated with a loss of function. Gene-mutation assays allow the direct identification of heritable changes of the genetic material of the cell.

The functionally altered cells are identified by selection procedures that allow the detection and quantification of few mutants within a large number of nonmutated cells. In general, selection is based on the mutation-induced resistance to toxic substances that will kill the nonmutated cells. The mutants survive and form colonies because the resistance is passed on to the daughter cells. The number of resistant cell colonies related to the number of cells that survived the treatment is referred to as the mutant frequency.

For routine testing, primarily Chinese hamster cell lines, e.g. V79 or CHO cells, have gained acceptance for the detection of an induced resistance to guanine analogues and ouabain, and the mouse lymphoma cell line L5178Y for the detection of an induced resistance to thymidine analogues (Table 4.8). These cells exhibit a stable karyotype, proliferate very rapidly, and have a high potential for growing as colonies.

The cells are exposed to the test compound in the absence and the presence of a metabolic activation system. Subsequently, they are further cultivated in normal growth medium in order to allow a fixation of induced DNA damage as mutations by DNA replication. In mutation systems such as the HPRT test or the $TK^{+/-}$ test, where the induced resistance is due to the loss of an enzymatic activity, this post-treatment so-called

Table 4.8 *Genetic endpoints of the most common gene-mutation assays employing mammalian cells.*

Target gene	Genetics	Selection agent
Hypoxanthine guanine phosphoribosyltransferase (*hprt*)	Recessive, X chromosomal	6-Thioguanine
Thymidine kinase (*tk*)	Recessive, autosomal, hemizygous in test cells	Trifluorothymidine
Ouabain resistance locus/Na^+/K^{+-}ATPase gene	Dominant	Ouabain

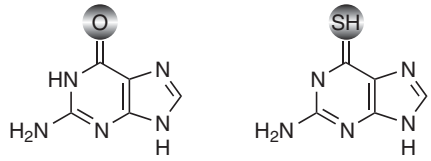


Figure 4.7 *Structures of guanine and 6-thioguanine (TG).*

'expression time' also provides the time necessary for the dilution of the enzyme proteins by cell division and their removal by proteolytic degradation. Following the expression time, the cells are cultured for several days in the presence of the selecting agent and then the number of the mutant colonies is counted. Simultaneously, the proportion of the cells surviving the exposure to the chemical is determined by growing the cells in the absence of the selecting agent.

HPRT Gene-Mutation Test

The HPRT gene-mutation test detects the heritable loss of the activity of the enzyme hypoxanthine guanine phosphoribosyltransferase (HPRT). It primarily identifies mutations within the *hprt* gene, e.g. base substitutions, frame shifts, or small deletions. Large deletions, which extend into neighboring essential genes and which show up microscopically as chromosome aberrations, are frequently lethal and cannot be detected by the HPRT test.

The HPRT converts free purine bases into the corresponding nucleoside monophosphates and, thus, salvages them for the synthesis of nucleic acids. The selection of the mutants is based on the different toxicities of the synthetic purine base 6-thioguanine (TG, Figure 4.7) to mutated and nonmutated cells. In cells with functional HPRT, TG is converted into nucleotides that are strongly cytotoxic and cause cell death. The loss of HPRT activity makes the cells resistant to TG.

The *hprt* gene is localized on the X chromosome. Cells from male animals, e.g. V79 cells, contain only one X chromosome, i.e. they are hemizygous for the *hprt* gene (Table 4.8). In cells from female animals, e.g. CHO cells, there are two X chromosomes, but one of them is genetically inactivated. Thus, these cells are also functionally hemizygous for the *hprt* gene. Therefore, in both cell types a single mutation can result in the loss of the HPRT activity and the formation of the TG-resistant phenotype.

The hamster cell lines V79 and CHO are almost exclusively employed for routine testing.

$TK^{+/-}$ Gene-Mutation Test

The $TK^{+/-}$ gene-mutation test detects the heritable loss of the activity of the enzyme thymidine kinase (TK) in cell lines heterozygous for the thymidine kinase gene ($tk^{+/-}$). Cells that have been found to be suitable for performing this test

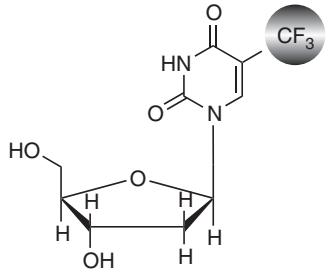
are the human lymphoblastoid cell line TK6 and the mouse lymphoma cell line L5178Y. For routine testing, the mouse lymphoma cells are usually employed, and this form of the test is also referred to as 'mouse lymphoma assay' (MLA). It detects both mutations limited to the *tk* gene and large mutations that can be identified microscopically as chromosome aberrations.

For the selection of the mutated cells, use is made of the different toxicities of the thymidine analogue trifluorothymidine (TFT, Figure 4.8) to the mutants and the wild-type cells. Following phosphorylation by thymidine kinase in the nonmutated cells, TFT is incorporated into DNA and methylated, which results in cytotoxicity and cell killing during subsequent rounds of DNA replication.

In contrast to the *hprt* gene, the *tk* gene is located on an autosome and is, therefore, present on two alleles in each cell. As a consequence, the TFT-resistant phenotype is recessive, i.e. it is expressed only when both alleles are inactivated by mutations. For mutagenicity testing, specific heterozygous cell lines are employed in which only one of two *tk* alleles is functionally active. In these cells, a single mutation in the tk^+ allele can be sufficient for the production of the $TK^{-/-}$ phenotype.

Mutants produced by a mutation within the tk^+ allele, which is not essential for growth, form rapidly growing, large colonies. This type of mutation apparently corresponds to the mutations detectable with the HPRT system. However, by means of the L5178Y $TK^{+/-}$ system large deletions spanning several thousand bases, which are associated with chromosomal mutations, can also be detected as 'gene mutations.' These mutations become apparent as slowly growing, small colonies of TFT-resistant cells.

The responsiveness of the L5178Y $TK^{+/-}$ system results in a markedly different specificity when compared with the HPRT system. In principle, one has to assume that all compounds or treatment conditions that result in the formation of chromosomal aberrations also lead to the appearance of small colonies. Electrophilic properties or reactivity of the compound or its metabolites with DNA are not essential. The differentiation of large and small colonies is, therefore, of special importance for the evaluation of positive results from mutation experiments using the L5178Y $TK^{+/-}$ system.



trifluorothymidine

Figure 4.8 *Structure of trifluorothymidine (TFT).*

Na^+/K^+-ATPase Gene-Mutation Test

The Na^+/K^+-ATPase gene-mutation test detects mutants resistant to the cardiac glycoside ouabain (G Strophanthin). The biological basis of this test system is not completely clear.

Ouabain binds to the Na^+/K^+ pump, which results in its inhibition and cell death due to a disturbance of energy homeostasis. It has been suggested that certain mutations in the structural gene of the Na^+/K^+-ATPase yield altered enzyme proteins that no longer contain a binding site for ouabain and that are, therefore, not inhibited by the drug. These mutants are able to survive in the presence of ouabain and can be selected. It is generally assumed that only base substitutions can be detected using this gene locus. On the other hand, the development of a resistance to ouabain as a consequence of an amplification of the ATPase gene has also been observed.

The test is usually performed with Chinese hamster cell lines, and a larger database of results obtained with the 'Na^+/K^+−ATPase' gene-mutation test is available only for V79 cells. In routine testing for mutagenicity, the system has not gained acceptance because the spectrum of detectable mutations appears to be very limited.

Tests for Chromosome Aberrations

The term 'chromosome aberrations' includes both alterations of chromosome structure that are microscopically visible in mitotic cells (**structural aberrations**) and alterations of chromosome number (**numerical aberrations**).

The *in vitro* chromosome-aberration test generally detects structural aberrations only. The first mitoses (metaphases) occurring after treatment of the cells with the test substance are scored, i.e. the analysed cells have not yet divided. Thus, the aberration assay does not detect mutations in the narrower sense, but potential precursors of heritable chromosome aberrations, i.e. chromosome mutations.

Therefore, the terms **chromosome-aberration test** and **chromosome-mutation test** are frequently used as synonyms. Chemicals that induce structural chromosome aberrations are also called **clastogens.**

For the testing of a substance with regard to potential clastogenic properties, cell cultures are incubated with the test compound in the absence and the presence of an exogenous metabolic activation system for a few hours. After a period equivalent to about 1.5 normal cell-cycle lengths the cells are fixed and prepared for chromosome analysis. The analysis of chromosomal aberrations is performed on metaphase cells, as only in this phase of the cell cycle are the chromosomes present in very compact form and amenable to an analysis of their structure by light microscopy. In order to increase the number of metaphase cells available for evaluation, spindle poisons such as colcemide or related compounds are added to the cultures before fixation.

A prerequisite for the induction of chromosomal aberrations by a compound is its ability to cause DNA double-strand breaks. These breaks can be induced directly by an attack of the compound on DNA, or they are formed during the processing or repair of DNA damage, usually during the S phase of the cell cycle.

In principle, chromosomal aberrations induced by chemicals can be detected in every proliferating mammalian cell. Usually various Chinese hamster cell lines, which are especially suitable for cytogenetic studies due to their comparatively low chromosome number $(2n = 22)$ and the characteristic structure of the individual chromosomes, are employed.

Even though the formation of DNA double-strand breaks is an essential step in the formation of chromosome aberrations, the clastogenicity of a compound is not necessarily due to the reactivity of the compound or its metabolites with DNA. Chromosomes are complex, dynamic structures made up of DNA and proteins, and their structural integrity depends on a large number of factors, such as the maintenance of a physiological intracellular milieu and the activity of several enzymes. Changes of chromosome structure can be induced by disturbances of DNA replication and repair, effects on topoisomerases, depletion of cellular energy, interference with cell membrane function, or triggering of apoptosis. Therefore, the evaluation of results from *in vitro* chromosome aberration tests requires specific care and experience.

Micronucleus Test

Micronuclei are small extranuclear chromatin-containing bodies. They consist of membrane-surrounded chromosomal fragments or complete chromosomes that have not been integrated into a nucleus of one of the daughter cells during cell division. Micronuclei are considerably smaller than the nucleus. They can be easily identified by light microscopy after staining.

Compounds can induce micronuclei by causing chromosome breaks or disturbances of the mitotic apparatus in proliferating cells (Figure 4.9). During mitosis, chromosomes assemble at the equatorial plate of the cell, and the chromosomes become separated and distributed to the cell poles. Subsequently, cell division occurs and new cell nuclei are formed. The separation of the two sister chromatids is brought about by the spindle apparatus of the cells. The spindle fibers attach to specific proteins of the centromeric region, the kinetochores, and pull the chromatids towards the cell poles. Chromosome breaks result in the formation of so-called acentric fragments, i.e. chromosome parts without the centromeric region. These fragments remain at the equatorial plate and form the **micronucleus** in one of the daughter cells after cell division. Micronuclei containing whole chromosomes are formed when the treatment of the cells affects the function of the spindle apparatus.

Since the formation of micronuclei consisting of acentric fragments and of whole chromosomes is based on completely different mechanisms, differentiation of the two types of micronuclei can yield valuable information on the mechanism of action of a chemical. The differentiation is achieved by examining the micronuclei for the presence of centromere-specific chromosome regions by either antikinetochore antibodies or fluorescence *in situ* hybridization (FISH) with pancentromeric DNA probes.

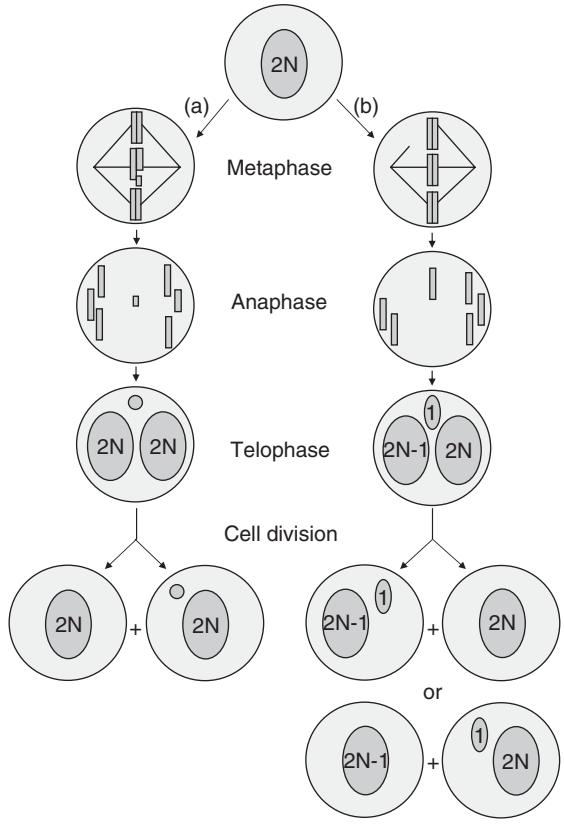


Figure 4.9 *Miconucleus test: schematic of the formation of cells with micronuclei consisting of: (a) An acentric chromosome fragment. (b) A whole chromosome.*

Thus, the micronucleus test allows the identification of **clastogens**, which cause chromosome breaks, and of **spindle poisons**, which disturb chromosome distribution. The micronucleus test is particularly valuable because no other assays for the reliable detection of aneuploidies are now available.

In vitro micronucleus tests are easy to perform, readily evaluated, and can be conducted with any mammalian cell line and certain primary cells, such as mitogen-stimulated freshly isolated hepatocytes.

Indicator Tests

The term 'indicator tests' stands for a number of different genotoxicity assays that do not detect the induction of mutations in the progeny of the treated cells. Rather, various effects mechanistically associated with the formation of mutations are identified directly in the treated cells.

Covalent Binding of Chemicals to DNA

The capability to react with DNA and to form DNA adducts is a characteristic property of many genotoxic mutagens and carcinogens or their metabolites.

In order to detect DNA binding, the cells can be treated with the radioactively labelled test chemical and subsequently analysed for the presence of radioactivity covalently bound to their DNA. However, the practical use of this approach is limited by the problem that the chemical in question has to be available in radioactively labelled form. Moreover, laborious control experiments are generally necessary to exclude experimental artifacts, such as:

- The exchange of tritium atoms between the test chemical and the DNA when [^{3}H]-labelled compounds are used;
- An apparent radioactive labelling of the DNA as a consequence of its contamination by radioactively labelled proteins;
- Enzymatic incorporation of radioactive decomposition products of the test chemical into DNA.

The additional analyses required to exclude these potential artifacts are so cumbersome that studies on the covalent binding are unsuited for routine testing.

Immunological methods and the ^{32}P post-labelling technique allow very sensitive detection of DNA adducts without being dependent on radioactively labelled test compounds, but these approaches are also associated with substantial experimental effort. They are, however, very valuable tools for investigations on the mechanism of action of genotoxic chemicals, and particularly useful in their *in vivo* version for the detection of organ-specific genotoxicity.

Induction of DNA Strand Breaks

Assays for the detection of chemically induced DNA strand breaks are easy to perform and highly sensitive. As a consequence of the extreme length of the DNA strands in chromosomes, even very few breaks suffice to cause a drastic reduction of the molecular mass, which can be easily detected.

Following treatment of the cells with the test chemical, double-stranded DNA is generally exposed to alkaline solutions in order to separate the DNA strands and to uncover single strand breaks. Subsequently, characteristic changes of the physicochemical properties of the DNA strands, which occur as a consequence of the strand breaks, are measured.

The induction of DNA strand breaks can occur directly, e.g. by radical attack of reactive species at the sugar phosphate backbone of the DNA, or by the action of endonucleases during the repair of DNA damage. In addition, there are DNA lesions that are not associated with strand breaks in the cell, but which are converted into breaks under the alkaline conditions employed for strand-break analysis. These so-called 'alkali-labile sites' include abasic sites formed by the enzymatic or spontaneous cleavage of a base from the deoxyribose, and phosphotriester, i.e. products of the alkylation of phosphoric acid groups of the DNA.

DNA strand breaks can be also induced by agents affecting the function of specific proteins involved in the maintenance of DNA structure. Such compounds include inhibitors of DNA polymerases and DNA repair enzymes or intercalating compounds that interfere with the activity of DNA topoisomerases. Strand breaks can also occur as a consequence of unspecific cytotoxic effects, such as membrane damage, shifts of the intra- and extracellular distribution of ions, lack of energy, and inhibition of protein synthesis. Thus, for the interpretation of observations from strand-break experiments it is mandatory to discriminate between unspecific or indirect actions of chemicals and specific effects originating from the reactivity of the test compound or its metabolites with the DNA. This differentiation is often very difficult and presents a problem that limits the assessment of strand-break measurements. Therefore, the measurement of DNA strand breaks is primarily used for the clarification of the mechanisms of genotoxic chemicals rather than for routine testing.

Alkaline elution

DNA can be deposited on membrane filters and subsequently eluted by an alkaline buffer. The rate of elution through the membrane pores depends on the length of the DNA molecules. Shortened strands of DNA resulting from strand breaks will be eluted prior to intact DNA. By comparing the elution rates of the DNA from untreated and from treated cells, it can be determined whether the test compound induced DNA strand breaks.

Following treatment of cells with the test compound they are collected on a membrane filter, lysed, and digested proteolytically. The DNA on the filter is washed with an alkaline buffer and individual fractions of the eluate are collected. Subsequently, the amount of DNA contained in each fraction is determined. The elution rate, i.e. the amount of DNA eluted per unit of time, is proportional to the number of DNA strand breaks per molecule. Modifications of the experimental protocol are available which preferentially detect DNA–DNA and DNA–protein cross-links as well as DNA double-strand breaks.

The alkaline-elution methodology may be used with every cell type. When proliferating cells, such as cell lines, are employed, the DNA can be labelled with radioactive thymidine prior to the treatment of the cells. The eluted DNA can subsequently be determined conveniently by counting the acid-insoluble radioactivity appearing in the eluate. When nonproliferating cells, such as hepatocytes, are employed, the DNA is quantified fluorimetrically.

Comet assay (alkaline single-cell gel electrophoresis)

The electrophoretic mobility of DNA molecules depends on their size and their conformation. By analysing the movement of the DNA of individual cells the presence of DNA strand breaks can be detected.

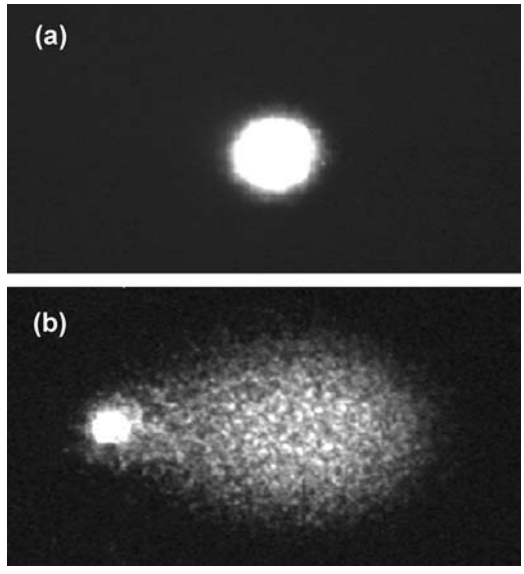


Figure 4.10 *Detection of DNA strand breaks by the Comet assay (alkaline single-cell gel electrophoresis). (a) Control cell. (b) Cell after treatment with a strand-break-inducing agent.*

The cells are exposed to a test compound, embedded in agarose on a microscope slide, and lysed with a detergent. Subsequently, they are treated with an alkaline buffer and then subjected to electrophoresis following which the DNA of the individual cells can be visualized by staining with a fluorescent dye such as ethidium bromide. Undamaged DNA is present in the nucleus in high molecular, extremely condensed form. Therefore, it does not migrate towards the anode under the test conditions applied. Strand breaks and alkali-labile sites in the DNA cause a partial relaxation of the compact DNA structure and a partial unwinding of the DNA strands. This enables the movement of DNA fragments and loops out of the nucleus towards the anode, which can be observed under the fluorescence microscope by the appearance of a 'comet tail' (Figure 4.10). Data evaluation can be performed in different ways: Usually, computer-aided image-analysis systems are used to measure DNA migration. If image analysis is not available, cells can also be classified into different damage classes visually according to tail size in order to calculate an arbitrary damage score.

Modifications of the technique employing post-treatment of the cells on the slides with lesion-specific endonucleases for the production of strand breaks allow the detection of specific types of DNA lesions, such as oxidative base damage.

Advantages of the single-cell gel electrophoresis technique, which can be applied on every cell type, are the high sensitivity of strand-break detection and the possibility to analyse individual cells. The *in vivo* version of this test has gained great practical importance in genotoxicity testing with regard to the detection of organ-specific genotoxic events.

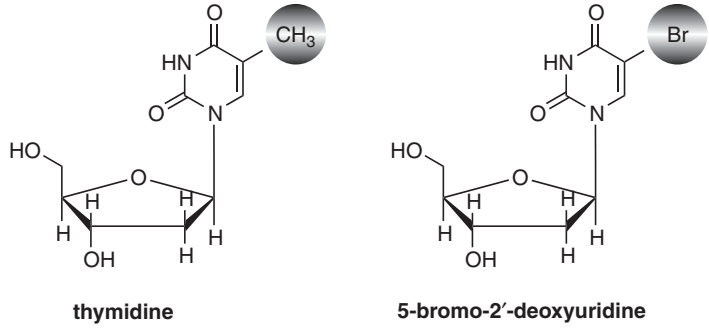


Figure 4.11 *Structures of thymidine and bromodeoxyuridine.*

Sister Chromatid Exchange Test (SCE test)

Sister chromatid exchanges (SCEs) represent the exchange of corresponding stretches of DNA molecules between the two chromatids of a chromosome during DNA synthesis. The exchange reaction is based on breaks of the DNA molecules of both chromatids at genetically identical sites with subsequent rejoining of both DNA molecules at the breakage sites during replication. Thus, the exchange is 'reciprocal' and does not usually lead to alterations in the DNA sequence.

For the detection of SCEs it must be possible to differentiate between the DNA molecules of the two sister chromatids. To achieve this differentiation, cells are grown for two cell cycles in the presence of the synthetic nucleoside 5-bromo-2′-deoxyuridine (BrdUrd). BrdUrd resembles thymidine (Figure 4.11), and during DNA synthesis it is incorporated into the newly synthesized DNA strand instead of thymidine. After two rounds of DNA synthesis the chromosomes consist of two chromatids that are asymmetrically substituted with BrdUrd (Figure 4.12). During metaphase, when the chromosomes are present in extremely condensed form, the two chromatids stain differentially and can therefore be discriminated. SCEs become apparent by color switches between the two chromatids. Quantitative evaluation of the experiments is performed by counting the number of color switches per metaphase.

SCE tests have been preferentially performed using Chinese hamster cell lines, such as V79 or CHO. In general, the cells are first incubated for 3–6 h with the test compound in the absence or presence of an exogenous metabolizing system. Subsequently, the cells are grown in the presence of BrdUrd for the duration of two cell cycles. Alternatively, for tests in the absence of S9 mix, treatment with the test compound and BrdUrd can be performed simultaneously for two cell cycles. In order to increase the number of metaphases available for analysis, the spindle poison colcemide, which arrests cells in metaphase, is added two hours before metaphase preparation.

The molecular mechanisms resulting in SCE formation are largely unclear. In general, SCEs appear to be the consequence of a perturbation of DNA synthesis. Such perturbations can be caused by the presence of unrepaired DNA lesions or by an inhibition of DNA repair. However, numerous other mechanisms that result in an inhibition of DNA

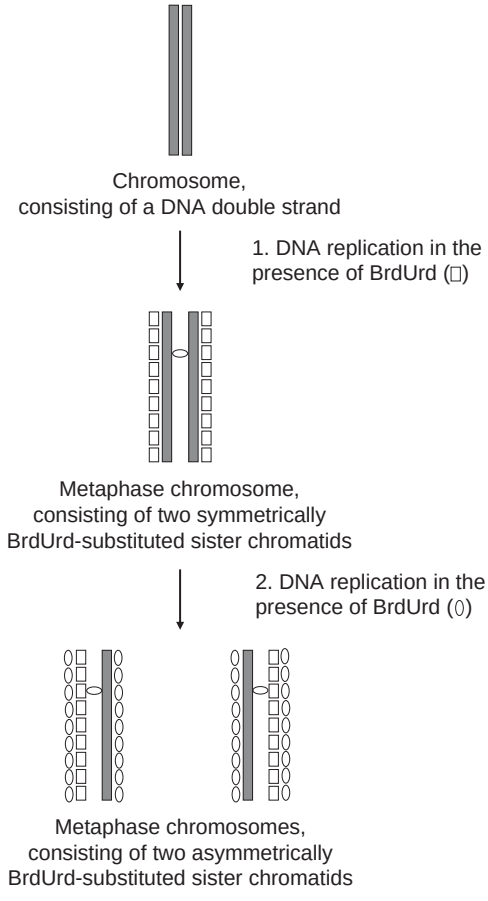


Figure 4.12 *Sister chromatid exchange (SCE) test: schematic of the formation of sister chromatids with different degrees of BrdUrd substitution.*

replication also induce SCEs. Since an exactly 'reciprocal' exchange between sister chromatids is not associated with alterations of the nucleotide sequence of the DNA, an SCE *per se* is not a mutagenic event. The exchange process between the two double-stranded DNA molecules is, however, a very complex, potentially error-prone biochemical process. A compound that is able to initiate this process is, therefore generally regarded as at least potentially mutagenic. This notion is supported by the observation that many chemicals that induce gene mutations or chromosomal aberrations also increase the frequency of SCEs.

The SCE test has been quite popular in genotoxicity testing of chemicals *in vitro* due to its simplicity and speed. It is not used any more in routine testing as the micronucleus test is also easy to perform and to evaluate and because the genetic endpoints detected with the micronucleus test, i.e. clastogenicity and aneugenicity, are better defined and of higher biological relevance.

DNA-Repair Assays

DNA-repair assays analyse the capability of a chemical to induce DNA repair in cultured mammalian cells. DNA repair is quantified by measuring DNA-repair synthesis. The capability of a chemical to induce DNA-repair synthesis is a specific indicator of the ability of the chemical or its metabolites to induce chemical alteration of the DNA molecule.

DNA-repair synthesis is a step in the process called 'excision repair' which comprises both the base-excision repair and the nucleotide-excision repair pathways, enables cells to detect DNA lesions and to excise them from the damaged strand. Excision creates a gap of 1 to about 30 nucleotides in length, which is subsequently filled by repair synthesis using the undamaged strand as template. This gap-filling step is termed 'repair synthesis' or 'repair replication.'

Repair synthesis can be detected and quantified by measuring the incorporation of radioactively labelled deoxyribonucleosides, such as [^{3}H]thymidine or [^{3}H]deoxycytidine, into the DNA. These nucleosides are also incorporated, in much larger amounts, by 'normal,' replicative DNA synthesis. Thus, the measurement of DNA-repair synthesis requires a clear-cut discrimination between both kinds of DNA synthesis in the cells.

Autoradiographic method (UDS test)

The autoradiographic determination of repair synthesis makes use of the fact that replicative DNA synthesis is restricted to the S phase of the cell cycle, whereas repair synthesis is independent of S phase and also occurs in other cell-cycle phases. It is, therefore, termed unscheduled DNA synthesis (UDS).

When cells are incubated with [^{3}H]thymidine and subsequently analysed autoradiographically, the nuclei of cells that were in S phase during the incubation appear heavily labelled due to the large amounts of incorporated radioactivity. Nuclei of cells that were not in S phase are substantially labelled only if the test compound induced UDS (Figure 4.13). DNA-repair synthesis can then be quantified by counting the number of silver grains or the area of the grains above these nuclei.

The UDS test has been widely used in routine genotoxicity testing. In general, primary cultures of rat hepatocytes are employed because of their outstanding metabolic competence. The cells are cultured with the test substance and [^{3}H]thymidine for 18–20 h. Unfortunately, the analysis of the autoradiographs is bedevilled by the presence of numerous silver grains above both the cell nuclei and the cytoplasm even in untreated cells. Whereas the labelling of the cytoplasm is due to the incorporation of radioactivity into mitochondrial DNA, the grains above the nuclei appear to originate from 'background' DNA repair going on even in the absence of a genotoxic test compound. Autoradiographs are evaluated by counting the number or the area of silver grains above the nuclei and correcting this value for the number of grains counted over a cytoplasmic area of the same size. The resultant number of 'net grains per nucleus' is an indirect measure for the incorporation of [^{3}H]thymidine into nuclei of non-S-phase cells, i.e. for UDS.

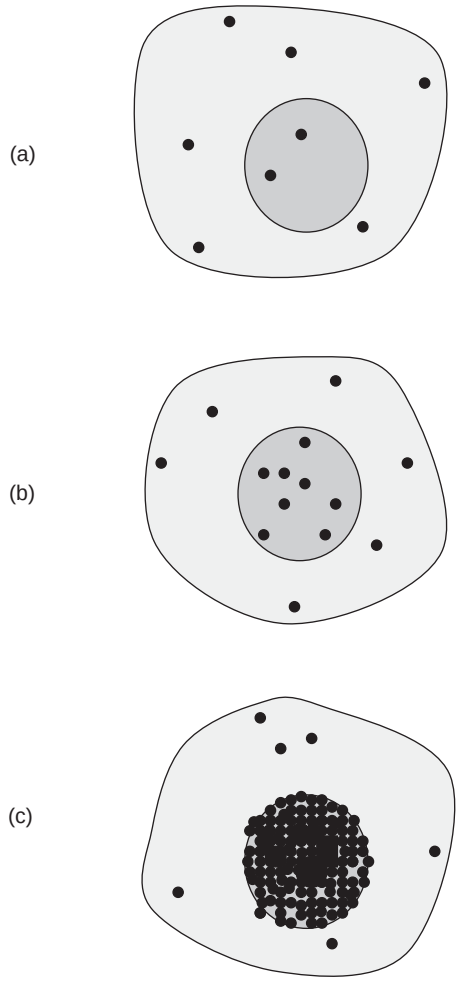


Figure 4.13 *Autoradiographic determination of DNA repair synthesis (UDS test): schematic of cell labelling: (a) Control cell. (b) Cell after treatment with a DNA-repair synthesis-inducing agent. (c) S-Phase cell.*

Measurement of [^{3}H]thymidine incorporation into DNA in the presence of an inhibitor of replicative DNA synthesis It has been frequently attempted to quantify DNA-repair synthesis by measuring the incorporation of [^{3}H]thymidine into DNA in the presence of an inhibitor of replicative DNA synthesis, generally *N*-hydroxyurea (HU). HU blocks the enzyme ribonucleotide reductase, which converts ribonucleoside diphosphates into deoxyribonucleoside diphosphates, which are precursors of DNA synthesis. Inhibition of this enzyme results in a strong inhibition of replicative DNA synthesis, whereas repair synthesis is not substantially affected. The technique is based on the assumption that the induction of an increased [^{3}H]thymidine incorporation by a compound in the presence of HU is indicative of an induction of repair.

However, this approach does not allow an unequivocal detection of DNA repair, as HU never inhibits replicative DNA synthesis completely and a change in the residual, HU-resistant [^{3}H]thymidine incorporation by the test compound can both mimic and obscure repair synthesis. For instance, mitogenic effects of the test compound or a reduction of the endogenous thymidine pool can result in an increased [^{3}H]thymidine incorporation. These problems can occur with every method relying on the use of inhibitors of replicative DNA synthesis, and are hardly suited for reliable detection of DNA-repair synthesis.

The measurement of the induction of DNA-repair synthesis enables the detection of the genotoxic properties of many different types of DNA-damaging chemicals. Since the biological consequences of the corresponding DNA lesions can be very different, no quantitative conclusions regarding the mutagenicity of chemicals can be drawn from the results of repair tests. However, an increased repair synthesis clearly indicates a chemical reactivity of the test compound or its metabolite(s) with the DNA. Since toxic effects not associated with genotoxicity do not result in an induction of repair, false positive effects cannot occur. Owing to this specificity the measurement of DNA-repair synthesis is suitable both for routine testing and for mechanistic studies on the effects of genotoxic compounds in mammalian cells.

4.2B.5 Cell-Transformation Assays

In certain mammalian cells, carcinogenic chemicals induce heritable alterations in cell morphology or growth behavior, which cause these cells to resemble tumor cells. The alterations occurring as a consequence of this 'transformation' include the loss of growth control, which can become manifest in various steps, such as the loss of contact inhibition, the acquired ability to grow in soft agar, or the capability for unlimited cell division (immortalization). Thus, cell transformation appears to be a complex process having many aspects in common with the process of cancer development.

The close relation of the alterations observed in transformed cells to the process of cancer development is shown by the observation that the implantation of transformed cells into experimental animals frequently results in tumor formation. Cell transformation resulting in cells capable of forming tumors is called 'malignant transformation.'

The molecular mechanisms responsible for the individual steps of cell transformation are largely unknown. It appears that test systems for cell transformation can not only detect mutagens, but also 'nongenotoxic carcinogens' such as certain types of tumor promoters.

Cell-transformation assays are not counted among the classical genotoxicity tests and have not gained broad acceptance in routine testing, primarily because of the uncertainties regarding the molecular mechanisms of cell transformation, the tremendous experimental effort required, and their sensitivity against modifications of experimental parameters. Large databases are available for systems employing mouse fibroblast cell lines and Syrian hamster embryo cells.

Cell-Transformation Assay with Mouse Fibroblast Cell Lines

Certain mouse fibroblast cell lines, such as C3H10T1/2, Balb/c3T3, or M2 cells, are capable of growing indefinitely, which is a consequence of partial transformation of the cells. They are still subject, however, to growth inhibition elicited by contacts with neighboring cells. Treatment of the cells with a transforming chemical can result in the loss of contact inhibition in individual cell clones, which becomes apparent by the development of so-called 'foci.' They consist of piled up cells, and after staining they can be easily differentiated from the underlying cell lawn. The frequency of these foci is taken as a measure of the transforming effect of the chemical. The evaluation of the culture dishes is not problematic, but the method makes high demands of the performing laboratory as the experiments usually take 1–2 months.

Cell-Transformation Assay with Syrian Hamster Embryo Cells

In comparison to the mouse fibroblast system, this transformation assay provides the advantage that the studies are performed on normal, nontransformed cells. The starting cells are obtained from whole Syrian hamster embryos (SHE) and, thus, are very heterogeneous. Usually, early subcultures of several embryos are pooled and frozen, which provides the advantage that a larger number of experiments can be performed with the same starting material, the suitability of which has been checked in a pilot experiment with a standard carcinogen.

The most frequently used variant of the SHE cell assay involves seeding of the cells at low density, exposing the cells to the test substance, and microscopic examination of the culture dishes after about one week. Transformed colonies are identified by their unordered (criss-cross) growth and their ratio between nucleus and cytoplasm, which is different from that of normal cells. Spontaneous transformations are extremely rare, but the frequency of induced transformations is also very low (mostly fewer than 5 transformed colonies per 1000 colonies), and clear concentration dependencies are seldom obtained. Usually, at least 1000 colonies per treatment group have to be analysed, which makes the evaluation very labourious.

A favorable property of SHE cells is their relatively broad expression of xenobiotic-metabolizing enzymes, at least during the first passages. Especially interesting is their high content of prostaglandin H synthase, which made these cells a popular tool for the study of the peroxidative activation of xenobiotics.

4.2B.6 Summary

The broad spectrum of genotoxic effects chemicals can exert in cells is reflected by the great variety of *in vitro* test systems available for their detection and quantification. *In vitro* tests for genotoxicity can be performed with various test organisms, which largely differ in their capacity for the metabolism of xenobiotics and which allow the detection of a broad spectrum of different genetic endpoints for the identification of genotoxic events. Most of these endpoints can in turn be investigated using different experimental approaches. As many compounds exert highly specific genotoxic effects on organisms, their impact on cells can be only detected by use of

test systems capable of indicating these effects. A sweeping characterization of a compound as 'genotoxic' or 'nongenotoxic' on the basis of an isolated result from a single test system and/or without indicating the test systems applied in the investigations would be meaningless. With the knowledge of the mechanisms underlying the various genotoxic effects, apparently contradictory results on the activity of compounds in different *in vitro* test systems can be often explained as the logical consequence of the specific properties of the compound. Thus, *in vitro* studies on the genotoxicity of chemicals are not only suited for use as fast and simple methods for the detection of genotoxic properties. They are also ideal tools for the clarification of the action mechanisms of chemicals.

4.2.C Strategies for the Evaluation of Genotoxicity

Stephan Madle, Peter Kasper, Ulrike Pabel and Günter Speit

4.2C.1 Introduction

During the 1970s the development of the so-called 'Ames test' initiated a new understanding of the relevance of mutagens for humans. Since then many different genotoxicity test systems have been developed and genotoxicity testing has become a routine part of toxicology.

Genotoxicity testing provides a basis for the assessment of deleterious effects of chemical substances, which may lead to germ cell mutations or cancer. Although systematic mutagenicity testing goes back to the mid-20th century, a major breakthrough occurred in the 1970s, when the so-called Ames test was established, enabling a quite simple and practical method to be developed for the identification of the mutagenic potential of chemical substances. The Ames test was adopted rapidly in many laboratories and resulted in new and important findings such as:

1. 'Carcinogens are mutagens.' Earlier mutagenicity testing was primarily performed with respect to mutations in germ cells, which would be heritable to subsequent generations. As it became evident that there was a close correlation between mutagenicity and carcinogenicity, the main purpose of mutagenicity testing changed to screening for carcinogens.
2. Mutagens are found in many parts of our environment (environmental mutagens). Substances with mutagenic potential were detected in nearly all areas of life.
3. Many mutagens need metabolic activation. With the use of 'S-9 mix' (rat liver homogenate) it became evident that the majority of mutagens expressed their mutagenic potential only after metabolic activation (pro-mutagens, also called indirect mutagens).

Many *in vitro* and *in vivo* test systems were established, in which various genetic endpoints could be investigated in different cell types. Simple methods were developed for the integration of S-9 mix into mammalian cell-culture assays. Later on, freshly isolated rat hepatocytes were added as a tool for metabolic activation.

In the early 1980s new legislation required systematic testing of chemical substances for mutagenic effects, and the Organisation for Economy Cooperation and Development (OECD) developed a number of test guidelines in order to ensure minimum standards for test performances. Triggered by the increasing number of test systems, new concepts were established. Testing methods were divided into two categories: mutagenicity tests, for investigating genetic effects heritable to daughter cells, and indicator tests, for investigating other genotoxic effects that are related to mutations.

Endpoints for indicator tests are of a different nature. 'Early' endpoints, such as unspecific DNA adducts, may or may not lead to mutations. Other endpoints, such as unscheduled DNA synthesis, are more reliable indicators of mutagenicity. The term 'mutagenicity testing' has been replaced by the term 'genotoxicity testing,' where genotoxicity is understood as an umbrella term including mutagenicity and other genotoxic effects. Nowadays the term 'genetic toxicology' indicates that genotoxicity testing has become a routine part of toxicology.

4.2C.2 Basics of Genotoxicity Testing

Based on the biology of mutations, genotoxicity testing shows a unique combination of characteristics: various genetic endpoints, show a preference for *in vitro* systems, complex modes of action with or without thresholds, and high relevance to carcinogenicity.

The interpretation of the results of genotoxicity testing and the strategies employed pose challenges not recognized in the investigation of other toxicological endpoints. These include:

- Different modes of action for mutagenesis.
  There are 3 types of mutations: gene mutations, structural chromosome mutations, and genome mutations, which include aneuploidies, i.e. deviations from the normal chromosome number. The different types of mutations are induced by different modes of action, although in practice there is large overlap of mutagens that induce gene and chromosome mutations. Therefore, a combination of test systems is needed for adequate assessment of genotoxicity.
- A broad spectrum of cell types, a variety of genetic endpoints, and systems for metabolic activation of pro-mutagens are used in genotoxicity testing.
  The spectrum comprises different levels: tests with bacteria, lower eukaryotes, and mammalian cells; animal tests with somatic and germ cells; and human studies. Obviously, some test systems are remote from the human situation. Therefore, it is particularly important not to extrapolate in an oversimplified manner from single test results to their relevance for man. Mutations are cell-bound events.
  Mutations are cellular events. Their detection does not require pathological alterations in organs or whole animals. Therefore, mutagenicity can easily be investigated in cell cultures *in vitro*, and mutagenicity testing was a vanguard for replacement of animal studies in toxicology.

- Background for genotoxic effects.
  Most genotoxic effects may occur spontaneously in the absence of external exposure to genotoxins. Some endpoints have high spontaneous frequencies. In mammalian cell cultures background levels vary considerably between and within the various genetic endpoints. For example:
  - Approximately 6000 DNA–protein cross-links are normally observed per cell.
  - Structural chromosome mutations and micronuclei are seen on the order of 1 in 100 cells.
  - In the mouse lymphoma $TK^{+/-}$ assay, approximately 10 in a million cells can be shown to have mutations.

- Genotoxic effects are not specific for a given inducing agent.
  Genetic effects induced by a specific substance cannot be distinguished from the same type of effects that occur spontaneously (with the exception of substance-specific DNA adducts).
- Genotoxic effects may be induced by secondary modes of action.
  When a substance induces a genotoxic effect, it is important to determine whether the effect is induced by direct interaction with the genomic material or by a secondary, indirect mode of action, e.g. chromosomal mutations may be caused by extreme cytotoxicity *in vitro*, or in rodents hypothermia can cause induction of micronuclei in bone marrow cells.
- The detection of mutated cells is limited to proliferating cell populations.
  Routine testing for induction of mutations can only be done when cell populations with high cell-proliferation rates (with the exception of gene-mutation tests in transgenic mice) are used. Otherwise, indicator tests have to be used.
- Somatic cell mutagens are germ cell mutagens.
  General mechanisms of mutagenicity are the same in somatic cells and germ cells. No mutagen is known which induces effects solely in germ cells. Therefore, *in vivo* testing is preferentially performed by the use of somatic cell tests, which are less expensive and require fewer animals.
- Germ cell mutagenicity vs. carcinogenesis - systemic vs. local genotoxicity.
  The most often used *in vivo* test systems are bone marrow assays. Negative results will generally lead to the conclusion that the investigated endpoint, e.g., micronuclei, is not induced in germ cells. Both bone marrow and germ cells are distant from the site of the body where the test substance is administered, and the availability of a substance to these cells is limited by its systemic availability. Directly exposed tissues such as those of the gastrointestinal tract after oral administration will be exposed to far higher concentrations, especially if the substance is poorly absorbed. Therefore, local genotoxic effects at the site of contact may not be detected by routine systemic genotoxicity tests such as the bone marrow micronucleus test. Organ-specific generation of DNA-reactive metabolites may be another reason for the potential failure of routine genotoxicity tests to detect relevant *in vivo* mutagens. This is of particular importance in the assessment of carcinogens when the organ in which tumors are found is not one of those examined during *in vivo* genotoxicity testing.
- Genotoxicity and thresholds.
  It was a generally accepted paradigm in the past that genotoxins have no thresholds, i.e. there is no dose of a genotoxin without genotoxic activity. However, various types of

genotoxic effects were identified to have thresholds, below which a hazard does not exist. Chemical substances that induce effects by indirect mechanisms such as nucleotide pool imbalances, inhibition of topoisomerases, damage to spindle proteins resulting in aneuploidy, or saturation of protective enzymatic activities can have thresholds for genotoxic effects. Therefore, the investigation of the mode of action is of critical importance for hazard assessment of genotoxins. The new paradigm is that genotoxins have no thresholds, unless a plausible threshold based on mode of action is demonstrated.

4.2C.3 Current Approaches for Assessing Genotoxicity

High Sensitivity of Basic Tests as a General Principle of Precaution

Thorough assessment of genotoxic effects includes the investigation of three types of mutation: gene mutations, structural chromosome mutations, and aneuploidies.

From a biological point of view there are two main requirements for appropriate mutagenicity testing: (1) consideration of all three relevant types of mutations, i.e. gene mutations, structural chromosome mutations, and aneuploidies; (2) employment of adequate metabolisation systems. Moreover, for the purpose of initial testing the use of highly sensitive *in vitro* systems is preferred in order to ensure that genotoxic potentials are detected with a high probability. The advantage of this approach is that negative results will give good evidence that the test substance does not induce germ cell mutagenesis or carcinogenesis with a genotoxic mode of action. The disadvantage is that some false-positive findings will be obtained and relatively high expense is needed to clarify whether or not the detected *in vitro* mutagenicity is relevant for the *in vivo* situation. Therefore, an important aspect of modern genotoxicity-testing strategies is to discriminate between relevant and irrelevant *in vitro* mutagens.

The following discussion describes general approaches, which are normally used for genotoxicity testing of chemical substances, and is outlined in Chart 4.I, Overview Chart 4.II Bacterial mutagens, and Chart 4.III *in vitro* mammalian cell mutagens.

Performance of Basic Testing (See Chart 4.I)

Basic testing for genotoxicity generally includes a bacterial gene mutation test and a mammalian cell test for structural chromosome mutations. All available data on the specific test substance are considered for adequate selection of test designs.

Bacterial Gene Mutation Test Nearly always the first step is to perform a bacterial gene mutation test with *Salmonella typhimurium* strains (sometimes supplemented with *E. coli* strains), irrespective of the specific properties of the test substance. Bacterial tests are practical and the results can be compared with those in the huge database available, both in the literature and at most testing facilities.

In the case of a negative result from the bacterial mutation test, all available data on the test substance should be checked to decide whether a deviation from the generally used approach

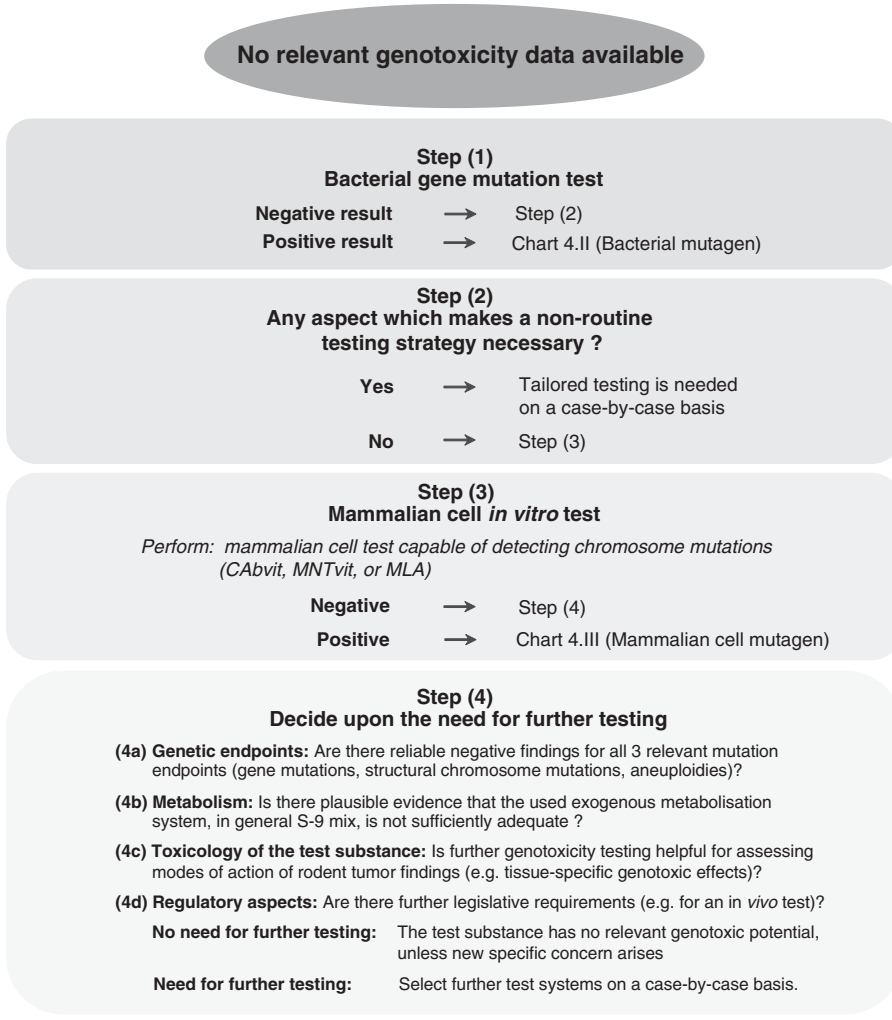


Chart 4.I. *General approach used for basic genotoxicity testing.*

is necessary. For example, physicochemical properties will influence the choice of further test systems, e.g. substances, that display very poor water solubility may require a nonroutine approach. If data on metabolisation of the test substance are available, it should be checked whether alternatives to the use of standard S-9 mix would be more appropriate. In general, over-representation of activating enzymes in S-9 mix renders this approach suitable as it further increases sensitivity. However, other approaches such as freshly isolated rat hepatocytes may sometimes be more appropriate and can be used as an alternative (or additional) metabolic activating system. Furthermore, information on chemicals with similar structures or on structural alerts may indicate a specific concern that should be clarified.

Mammalian Cell Test The second genotoxicity test is usually performed with mammalian cell cultures for the detection of structural chromosome mutations. For a long time the *in vitro* chromosomal mutation test has been used in these studies. However, the *in vitro* micronucleus test is often deemed advantageous, because it enables the detection not only of structural chromosome mutations but also of aneuploidies. The induction of chromosome mutations may also be investigated in the mouse lymphoma $TK^{+/-}$ assay by assessing the induction of slowly growing mutant colonies. These small colonies are usually caused by gross chromosomal mutations as compared with large mutant colonies in this assay, which are induced mainly by gene mutations.

In general, negative results from bacterial and mammalian cell tests can provide sufficient evidence that the test substance is devoid of relevant mutagenic potential. However, the need for further testing should be considered if (a) not all of the three relevant types of mutations are investigated, (b) modification of the metabolic activating system is indicated, (c) further genotoxicity testing would be helpful for assessing the mode of action of carcinogenesis by this chemical, or (d) regulatory requirements mandate further testing.

Positive results in basic testing generally indicate the need for follow-up tests.

Follow-up Testing of Bacterial Mutagens (See Chart 4.II)

In general, bacterial mutagens are further tested by *in vitro* genotoxicity tests with mammalian cells.

A comparison of bacterial mutagenicity tests with mammalian cell mutagenicity tests reveals that bacterial tests show a lower rate of positive findings. Bacterial mutagenicity tests are nearly always followed up by *in vitro* tests with mammalian cells; however, different approaches are used. Often the mouse lymphoma assay (step 1, option A) is performed for practical reasons. A negative result would give good evidence that the test substance does not induce gene or chromosome mutations in mammalian cells, whereas in the case of a negative result from a hypoxanthine phosphoribosyltransferase (HPRT) test (step 1, option B) further testing with respect to chromosome mutations is needed (step 2, option B). Furthermore, a negative result from a chromosomal aberration test or a micronucleus test (step 2, option C) leads to the need for further testing to investigate whether the substance is a specific inducer of gene mutations.

One major reason for compounds being found positive in bacteria but negative in mammalian cells is given by a bacterial-specific metabolisation of the compounds, as, for example, is known for the chemical class of nitroaromatic compounds. Such information would provide further assurance that the findings in bacteria are not relevant for mammalian cells.

Assessment of Findings from Mammalian Cell Tests

Mammalian cell genotoxicity tests show a very high rate of positive results. It is important to assess whether or not positive findings are limited to artificial conditions or extreme cytotoxicity, which are not relevant for the *in vivo* situation.

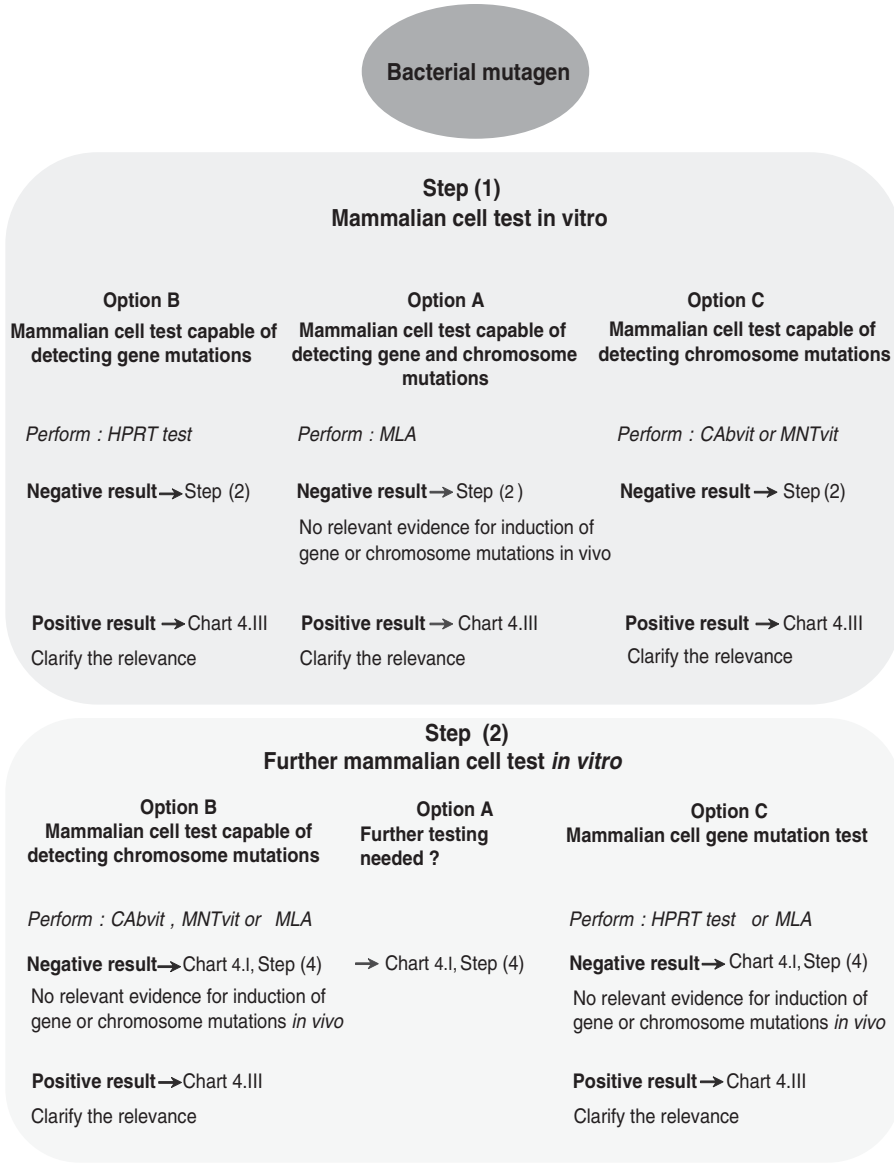


Chart 4.II. *Follow-up testing for bacterial mutagens.*

In general, findings from mammalian cell mutagenicity tests can clearly be assessed as 'negative' or 'positive.' Evaluation of reliability and plausibility of the findings can adequately be performed on the basis of validated guidelines and the familiarity with the systems. When contradictory findings are available for a substance obtained with the same test system, critical evaluation of the methodologies, plausibility of the findings, and consideration of good laboratory practice will be decisive criteria.

Critical Aspects: Artificial in vitro Conditions and Cytotoxicity Mammalian cell mutagenicity tests show a very high rate of positives, e.g. 30% of the new substances tested in Germany are shown to be positive in the *in vitro* chromosomal aberration test. In the evaluation of positive findings it is essential to discriminate between an assessment as 'positive' and the relevance of a finding. It is well known that induction of chromosomal aberrations (clastogenicity) may be bound to artificial *in vitro* conditions. Therefore, it is important to control culture conditions. Variations in culture conditions such as extreme shifts in osmolality or pH can be remedied. Precipitation of chemicals poses a problem for test substances with low water solubility. It is essential to evaluate the chemical at the highest concentration that it is possible to achieve. However, precipitation of the chemical due to water insolubility may lead to artificial positive results. Nevertheless, it is generally advised to select the analysed concentrations in such a way that the top dose leads to precipitation. Positive clastogenicity findings that are clearly bound to precipitating 'concentrations' are evaluated as of low relevance for the *in vivo* situation.

In practice, cytotoxicity is the most important confounding factor in tests with mammalian cell cultures. In the selection of the highest concentrations a balance should be achieved. The highest concentrations should lead to a sufficient level of cytotoxicity (in accordance with the guidelines) but should not cause extreme cytotoxicity, because extreme cytotoxicity can lead to secondary (indirect) modes of clastogenic action such as release of endonucleases and induction of apoptosis. Secondary effects may be restricted to very high *in vitro* concentrations, which are not relevant for the *in vivo* situation, because they are thresholded. For the so-called 'high-toxicity clastogens' (HTCs), induction of chromosomal mutations is limited to concentrations that also lead to high toxicity. However, clastogens can only be identified as HTCs when good data on cytotoxicity are available, e.g. on the time– and dose–effect relationship of cell growth.

Follow-up Testing of Mammalian Cell Mutagens (See Chart 4.III)

The relevance of mammalian cell mutagens *in vitro* is normally clarified by *in vivo* tests. The main criterion for the selection of an appropriate *in vivo* system is given by the adequate exposure of test cells.

In vitro/in vivo Contrast Given the awareness of limitations in mammalian cell mutagenicity tests, care must be taken when extrapolating positive results to *in vivo* situations. *In vitro* mutagens without major limitations in reliability and relevance need clarification, in general, by *in vivo* tests. According to existing experience the vast majority of *in vitro* mutagens lead to negative results in routine *in vivo* systems. For example, in Germany, only about 3% of chemicals identified as mutagens *in vitro* were shown to be mutagens *in vivo* using the micronucleus test. Understanding and interpreting of this *in vitro/in vivo* contrast is the most important challenge in evaluating genotoxicity data. *In vitro* systems are optimised with respect to their sensitivity and may produce results which are not relevant for the *in vivo* situation. *In vivo* tests are often performed without adequate information on the availability of the test substance to target cells. Thus, it would be imprudent to generally dismiss positive findings *in vitro* because of negative findings *in vivo*.

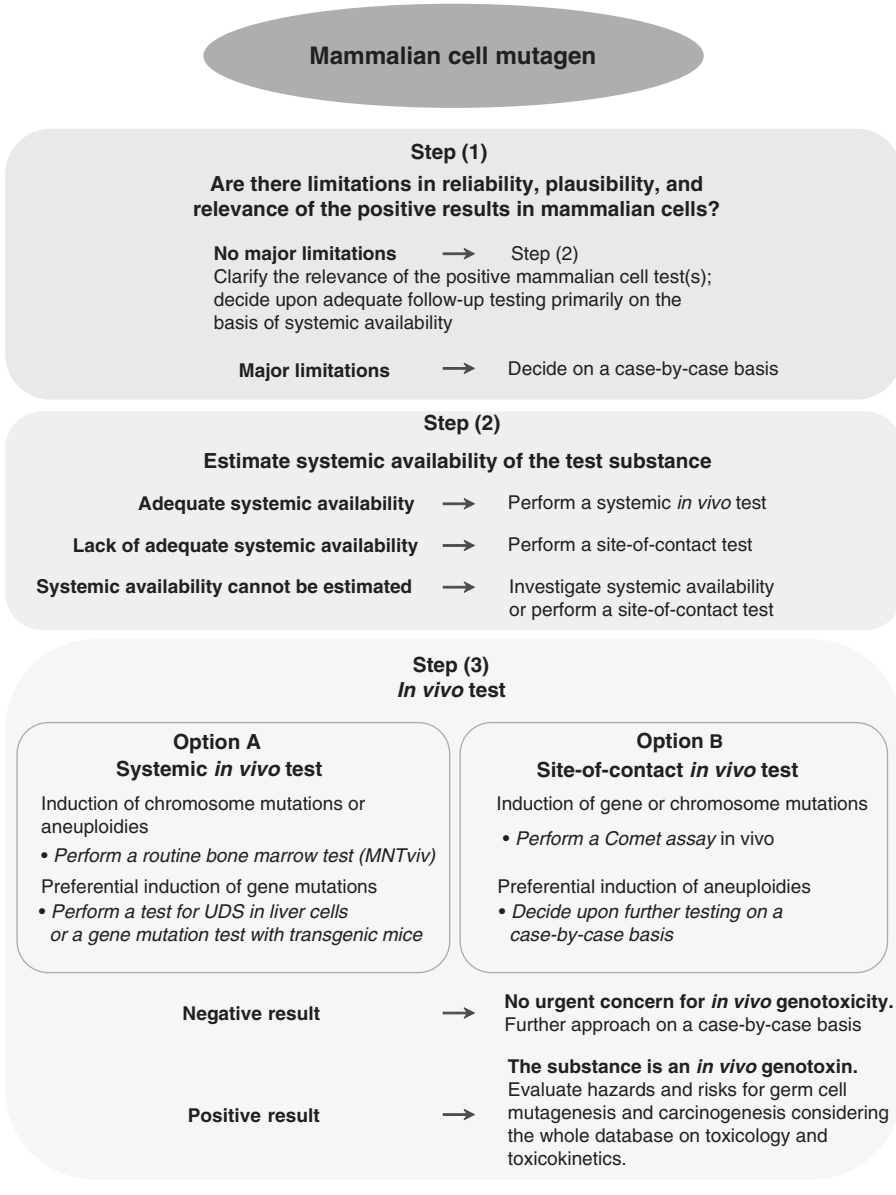


Chart 4.III. *Follow-up testing for mammalian cell mutagens.*

Systemic vs. Site of Contact Genotoxicity The main criterion for selection of an adequate *in vivo* system is the estimation of the systemic availability of the test substance. Routine systemic test systems like the bone marrow micronucleus test should not be performed when it cannot be assumed that the test substance, or a relevant metabolite, will reach the bone marrow in 'adequate' concentrations. However, data are often lacking upon which to judge the 'adequacy' of systemic availability.

If *in vivo* testing is performed in the absence of adequate toxicokinetic data, a local 'site-of-contact' genotoxicity test is the method of choice, which considers directly exposed tissues, i.e. the gastrointestinal tract after oral administration or the respiratory tract after inhalation. Maximum information can be obtained when further tissues are investigated within the same test, e.g. liver as a 'standard' tissue and other tissues that may be of interest with respect to carcinogenicity. The *in vivo* comet assay, which utilizes single-cell gel electrophoresis, is often viewed as the most suitable method for genotoxicity testing at the site of contact. The comet assay detects primary lesions of gene and chromosome mutations, is sufficiently validated, and enables investigation of nearly any tissue of interest. Furthermore, the comet assay can be combined with other genotoxicity tests that measure different genetic endpoints, such as the bone marrow micronucleus test, if a proper dosing schedule, which permits adequate time between treatment and sampling, is used. An example of appropriate timing might be treatment at 48 h, 24 h, and 3–6 h before sampling.

For *in vitro* clastogens and *in vitro* aneugens with adequate systemic availability, follow-up testing will normally include a bone marrow micronucleus test. Aneugens with low systemic availability must be evaluated on a case-by-case basis, considering the threshold mode of action for the induction of aneuploidy. In rare cases *in vitro* mutagens preferentially induce gene mutations. In those cases further investigations using a gene mutation test in transgenic mice or a test of unscheduled DNA synthesis in liver would be more appropriate.

Follow-up testing for *in vitro* mutagens is still a matter of controversial debate. There are doubts on the usefulness of *in vivo* indicator tests such as the comet assay, because they detect 'early' DNA effects which may or may not result in mutations. Therefore, the results should always be evaluated in the context of all genotoxicity data.

Assessment of Complex Genotoxicity Data bases

Complex sets of genotoxicity data often include contradictory findings. Careful evaluation of the limitations of the various data is needed, based on a weight-of-evidence approach.

When evaluating the genotoxicity of chemical substances, in practice, a relatively high number of test results may be available. Often findings are contradictory, and the significance of positive results is not always assessed using current strategies. Decisions must then be made by weighing the data on a case-by-case basis, assessing the reliability and the relevance of the various findings, and finally, by applying expert judgment.

Summary of Critical Aspects for the Evaluation of Genotoxicity Findings

- Genetic endpoint: All 3 types of mutations need to be considered. Indicator tests are helpful, particularly because they allow an investigation of genotoxic potentials in many different tissues. The significance of positive indicator tests *in vivo* depends on the *in vitro* test results for related mutation endpoints.
- Methodology: Investigations with major deficiencies in methodology should be disregarded. Preference should be given to those investigations which are in agreement with the guidelines and with good laboratory practice.

- Artificial findings and secondary modes of action: Positive results should be reviewed to determine whether they may be the result of extreme culture conditions or are caused by other secondary modes of action.
- Toxicokinetics: Special pathways in the metabolisation of substances may need to be considered by non routine metabolisation systems. In the interpretation of *in vivo* findings different conclusions may be drawn for germ cell mutagenicity and carcinogenicity, taking systemic and local availabilities of the substances into consideration.

Handling of in vivo *Genotoxins*

***In vivo* genotoxins are evaluated with respect to their modes of action in order to assess the relevance for humans. In particular, an understanding of the dose–effect relationship is of high practical importance.**

In vivo genotoxins should be critically evaluated with respect to their relevance for man. Positive results, which are limited to extreme exposure conditions, should be carefully considered because they may be induced by secondary modes of action. In the micronucleus test, for instance, genotoxic effects can be induced by physiological perturbations (such as bleeding or chemically induced hemolysis) and changes in body temperature. If possible, an understanding of the general mode of action for *in vivo* genotoxicity should be achieved to enable plausible extrapolations to the human situation - from animal species to man (toxicokinetics) and from high to low doses (dose–effect relationship). The basic assumption is that genotoxins have a nonthreshold mode of action, i.e. there is no dose without genotoxic effect. However, thresholds may arise from the type of reaction with the genetic material or by toxicokinetic properties (such as overloading of enzymatic capacities).

If *in vivo* genotoxicity data allow for a derivation of lowest observed effective doses (LOEDs), a comparison with tumorigenic doses may help us to understand whether genotoxicity has a relevant impact on the complex process of carcinogenesis. However, this comparison is often complicated when different types of exposure are used. For example, oral gavage exposure may lead to far higher systemic concentrations than feeding exposure, where the chemical is mixed in food. When a threshold mode of action is plausible, safe levels may be derived.

In the regulation of chemical substances there is, in general, no need for the performance of germ cell tests, because reliable demonstration of *in vivo* genotoxicity in somatic cells will provide a basis for strict regulation, in risk assessment as well as in classification. According to the Global Harmonization System (GHS; see table 1.2), which is in the process of worldwide implementation, *in vivo* genotoxins are classified as mutagens of category 2. If further evidence indicates that they reach the germ cells, they are classified as mutagens of category 1:

- Category 2 mutagen according to GHS: Chemicals that cause concern for man owing to the possibility that they may induce heritable mutations in the germ cells of humans.

- Category 1 mutagen according to GHS: Chemicals known to induce heritable mutations or to be regarded as if they induce heritable mutations in the germ cells of humans.

4.2C.4 Summary

Genotoxicity testing aims to detect substances that can induce germ cell mutations or cancer. Substances with genotoxic potential occur in various areas of our life. Many different test systems are available for their identification. In practice, basic testing for genotoxicity is normally done by a combination of two *in vitro* tests: a bacterial gene mutation test and a mammalian cell test. Thorough assessment of genotoxicity considers all three relevant types of mutations, i.e. gene mutation, structural chromosome mutation, and aneuploidy, and all the specific properties of the test substance.

Positive findings from *in vitro* tests are assessed with respect to their relevance for the *in vivo* situation. Genotoxic effects that are clearly limited to extreme or artificial test conditions do not necessarily need clarification by further testing. In general, the relevance of positive findings from mammalian cell tests is investigated by *in vivo* tests. However, negative *in vivo* findings cannot simply override positive results observed *in vitro*. A major aspect which needs consideration is the availability of the test substance to the specific test cells of the *in vivo* system used. If no adequate information is available, investigations considering site-of-contact tissues are preferred.

In the practice of genetic toxicology, for many substances complex data-sets are available which include contradictory findings. The data should then be evaluated by weighing relevant criteria such as genetic endpoints, methodology, mode of action, and toxicokinetic properties of the chemical.

In vivo genotoxins are assessed for their risks with respect to germ cell mutagenesis and carcinogenesis, and they are classified as mutagens. Understanding of the dose–effect relationship is of critical importance for the estimation of risks for humans after exposure to *in vivo* genotoxins.

For Further Reading

Dearfield, K. L. and Moore, M. M. (2005) Use of Genetic Toxicology Information for Risk Assessment. *Environ. Mol. Mutagen.*, **46**, 236–247.

Kirkland, D. J., Aardema, M., Henderson, L., and Müller, L. (2005) Evaluation of the Ability of a Battery of Three *in vitro* Genotoxicity Tests to Discriminate Rodent Carcinogens and Non-carcinogens. I. Sensitivity, Specificity and Relative Predictivity. *Mutat. Res.*, **584**, 1–256.

Speit, G., Autrup, H., Crebelli, R., Henderson, L., Kirsch-Volders, M., Madle, S., Parry, J. M., Sarrif, A. M., and Vrijhof, H. (2000) Thresholds in Genetic Toxicology. Concluding Remarks. *Mutat. Res.*, **464**, 149–153.

Stefan Pfuhler, A., Albertini, S., Fautz, R., Herbold, B., Madle, S., Utesch, D., and Poth, A. (2007) Genetic Toxicity Assessment: Employing the Best Science for Human Safety Evaluation. Part IV: Recommendation of a Working Group of the Gesellschaft fuer Umwelt-Mutationsforschung (GUM) for a Simple and Straightforward Approach to Genotoxicity Testing. *Toxicol. Sci.*, **97**, 237–240.

4.3 Biomonitoring

Michael G. Bird

4.3.1 Introduction

Biomonitoring is the measurement of chemicals, or their metabolites, in the blood, urine, or body tissues. Biomarkers are indicators for biological change such as binding with a protein to form a marker of exposure. A change of a cellular function such as an enzyme activity is a biomarker of effect. Biomarkers of 'susceptibility' identify (sensitive) subpopulations. Biomonitoring in its broadest terms can therefore also mean measurement of one or more of these types of biomarkers as well as the actual chemical itself.

The terms *biomonitoring* and *biomarkers* often are used interchangeably, but in fact biomarkers are indicators of where the chemical or metabolite causes some sort of biological change. This can be either by binding with a protein to form a marker of exposure or by changing a cellular function such as an enzyme activity, which would then be considered a biomarker of effect. There is also another class of biomarkers called biomarkers of 'susceptibility', which help to identify sensitive subpopulations. Biomonitoring in its broadest terms can therefore also mean measurement of one or more of these types of biomarkers as well as the actual chemical itself.

There are many other uses for biomonitoring data and they all center on its application to address one or more specific questions, which can include the identification of a specific exposure source such as lead in paint, or a potential health risk (e.g. by measuring a marker of genetic change). In fact there is a continuum from the initial exposure source to a possible health risk.

Biomonitoring can be used to answer specific questions as illustrated by Figure 4.14. The 'wheel' illustrated there has the main elements from the continuum of exposure, health risk, and associated susceptibility, but within each are important sub-questions or 'spokes', which may well determine the type of data needed, the population to be sampled, and the time period of study needed.

The World Health Organization (WHO, 2001) has defined 3 groups of biomarkers or biological indicators: those for identifying exposures, effects, or susceptibility.

Biomarker of Exposure

This provides a direct measure of body burden and is applicable for volatile and nonvolatile compounds. It can be the parent compound, metabolite, or protein-and-DNA adduct that is measured in a compartment in an organism. A biomarker of exposure integrates all routes of exposure and uptake to give a single measurement, which can provide unequivocal evidence of exposure; it can be used to indicate extent or magnitude of exposure, but may not be sensitive or specific enough for use in some investigations

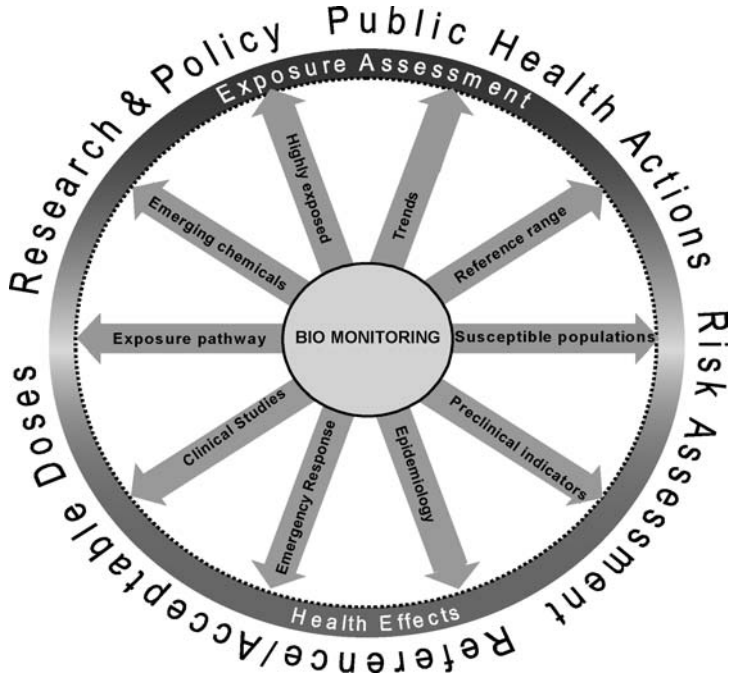


Figure 4.14 *Some specific applications of biomonitoring.*

and is often difficult to relate to risk of adverse outcome. The chemical's bound complex with various proteins can be used as a biomarker of exposure but here, similar to a nonbound chemical, the half-life of the protein is important. For instance, a compound bound to hemoglobin will reflect previous exposure of up to 180 days before the red blood cell is cleared from the circulation; for serum albumin up to 20 days; and for histones years in nondividing cells.

In using any biomarker of exposure, it's important to know the normal (background) level of the biomarker in order to determine if and to what extent a deviation from this background range has occurred as a result of exposure or of the disease process. There is also the need to know the background variability. A number of factors may influence an individual person's level, such as age, sex, and lifestyle, including smoking, alcohol, etc.

Biomarker of Effect

A measurable biochemical, physiological, behavioral, or other alteration in an organism that, depending on the magnitude, can be recognized as associated with an established or possible health impairment or disease. For instance, small and reversible departures from normal values - for example in packed-cell volume - may not signal an adverse health effect. Other biomarkers may indicate an early stage in the disease progression; examples of such early biomarkers of effect are somatic mutations, changes in tumor-suppressor genes, and cytogenetic changes (aberrations, micronuclei, and aneuploidy), whereas later indicators are of altered structure and function, and such biomarkers include changes in mutational spectra.

Biomarker of Susceptibility

An indicator of an inherent or acquired ability of an organism to respond to the challenge of exposure to a specific chemical substance. These biomarkers provide an indication of the extent to which an individual may be prone to progress from exposure to developing an adverse health effect. An example is glutathione *S*-transferase M, which is a Phase II conjugation enzyme.

Some biomarkers, such as DNA adducts, cross the above definitions and may be indicators of exposure and possibly of effect, while others are specific to exposure (hemoglobin adducts) or to effects (chromosomal aberrations). Ideally, biomonitoring links these markers of exposure, effect, and susceptibility to understand the implications to worker or public health of exposure to chemicals.

BIOMARKERS

'BIOMARKER OF DOSE/EXPOSURE'
- Determines exposures to toxic agents
- Identifies the presence of or past exposure to infectious organisms

'BIOMARKER OF EFFECT'
- Determines the effects of exposures to toxic agents

'BIOMARKER OF SUSCEPTIBILITY'
- Assesses susceptibility to toxic agents or infectious organisms

4.3.2 History

The earliest applications of biomonitoring were for the human monitoring of occupational exposures such as the monitoring of lead in the blood of workers, benzene in exhaled breath, or cadmium in urine. It continues to be an important adjunct to airborne measurements. In a number of cases biological exposure indices such as the Biological Exposure Index (BEI) of the American Conference of Governmental Industrial Hygienists (ACGIH) have been developed as guidance for the control of workplace exposures. Although these indices have been set to control exposure and to protect the worker population from harmful effects in the workplace, they may not directly be applicable to the general population, where long-term and continuous exposure are of concern.

Biomonitoring has become more widely used in the public sector. This has drawn increasing public attention, particularly as some of the biomonitoring programs, such as that of the Centers for Disease Control (CDC), have reported levels of many chemicals present in adults and children. Despite CDC's cautionary statement issued with these results, that 'the measurement of an environmental chemical in a person's blood or urine does not by itself mean the chemical causes disease,' without guidance as to interpretation, public opinion tends to default to 'presence equals harm.' This, as a minimum, can cause concern and may also result in unduly restrictive uses and/or product removal from the market.

4.3.3 Biomonitoring Programs

There are a number of important biomonitoring efforts ongoing, both in the US and also internationally. In some cases these are directed to measuring background levels in the general population or to assess trends and the efficacy of any public health control measures that may have been implemented. One such case is the removal of lead from gasoline in the US, which resulted in blood lead levels in children being reduced from 15 μg/dL in 1978 to 2 μg/dL in 1999. Other studies are focused on specific populations to assess possible health risks associated with the exposure measurements. An example is the measurement of biomarkers of methyl parathion to elucidate outbreaks of poisoning from the use of this insecticide.

In the US, programs include the US EPA sponsored National Human Exposure Assessment Survey (NHEXAS) in the 1990s and Centers for Disease Control (CDC) National Environmental Public Health Tracking Program. The CDC program reports every two years on the concentrations of chemicals and metabolites in blood and urine of a representative sample of the US population; the 2005 publication tracked 148 chemicals and the number will be increased to nearly 300 for the fourth CDC Report on Human Exposure to Environmental Chemicals due in 2007. A previous CDC program - the CDC National Health and Nutrition Examination Survey (NHANES) - included the tracking of lead and was responsible for the US regulation to reduce lead in gasoline, resulting in the reduction of blood lead levels in children as noted above.

In Europe, the SCALE initiative (Scientific evidence, focused in Children, meant to raise Awareness, improve the situation by Legal instruments and ensure a continual Evaluation of progress mode) (European commision, 2004).[1] There are also a number of national programs, which include environmental and occupational biomonitoring, the monitoring of breast milk and cord blood.

The World Health Organization has conducted three international surveys of chlorinated organic compounds in human breast milk, and a fourth survey will examine the presence of persistent organic pollutants.

Challenges

The ability to generate new biomonitoring data often exceeds the ability to evaluate whether and how a chemical measured in an individual or population may cause a health risk or to evaluate its sources and pathways for exposure (NRC, 2006). Of the 148 chemicals measured by CDC in its third exposure report of 2005, only 25 have established EPA reference values to help to assess health risk. Analytical methods have developed faster than our ability to determine a relationship of many internal dose biomarkers with adverse health effects, i.e. risk cannot be estimated directly from biomarkers for many chemicals, and detection of a compound does not necessarily indicate the presence of a significant risk.

However, biomonitoring data can be generated for many important purposes even if the link between a health effect and the presence of the chemical in the tissues has not been demonstrated. For example, these data can be used to detect trends, i.e. changes in the concentration of a chemical over time, within a particular population, or across a geographic region; this may or may not necessitate implementation of public health

[1] http://www.eu-humanbiomonitoring.org/

control measures. In most cases there have been few historical data available. One of the values of the above CDC program is to provide a base line so that future biomonitoring can be compared with this.

To assess trends in the concentration of a chemical over time, it is important to know information on the persistence and half life of the chemical in the environment, but for this application of assessing trends it is not necessary to know the rate of absorption or distribution and excretion within the body. Different supplementary types of information are needed, depending on the biomonitoring question being addressed.

Other applications do require toxicological, pharmacokinetic and metabolism information for interpretation of the biomonitoring data. These applications include identification of potential sources of exposure, use in epidemiological studies to identify highly exposed subpopulations and potential health effects, and in clinical evaluations of individual exposure and health risk. All applications require a rigorous scientific approach to data collection, analysis, and interpretation.

Biomonitoring applications, whether they be assessing a health risk or monitoring a trend, or identifying an exposure source, require a rigorous scientific approach to data collection, analysis, and interpretation.

4.3.4 Study Design

The design must be appropriate for the application being addressed. It needs to consider not only the specific population to be sampled, but also how directly the chosen biomarkers lie on the exposure/health risk continuum and therefore their ***biological relevance***. For example, the reporter gene for hypoxanthine guanine ribosyltransferase *(HPRT)* is often used as a biomarker of a mutational event but in fact it has no actual role in a disease process, whereas chromosomal aberration does have such an association with disease. These biomarkers reflect predisease processes - the biomarker may be a surrogate for an early pathogenic process or may be directly part of the process; it is important to understand the biological relevance of the biomarker(s).

The ***temporal relationship*** between the biomarker and the exposure and/or disease should also be known. Again *HPRT* determined from blood collected at the time of exposure will not reflect the consequences of that exposure since a 3-week period is needed for the mutation to be expressed. As mentioned previously, the binding of a chemical to hemoglobin to form a hemoglobin adduct can only reflect cumulative exposure to that chemical for the 180-day life span of the red blood cell, and albumin adducts have a shorter half-life of 20–25 days. Hence the pharmacokinetics of the marker provides an indication as to whether the exposure is recent or cumulative. All this bears on the frequency of sampling and from what media - blood, urine, expired air, etc.

The ***sensitivity*** of a biomarker is important. It should be a consistent and quantitative reflection of the range of exposure, including low doses, which are usually more typical of community exposures versus those in the occupational setting. However, in certain cases the analysis of the biomarker needs to be sufficiently sensitive to measure the potential contribution from an endogenous source as well as external sources. An example here would be ethylene, which is formed endogenously as a result of normal

metabolism but which also can be an external source of exposure as it is present in cigarette smoke and other combustion products.

Along with sensitivity is the need for ***specificity***: the biomarker should unequivocally reflect the chemical being assessed. Several chemicals may share some of the same metabolites and therefore use of a metabolite may in these cases be less specific than measuring the parent compound itself. External lifestyle and dietary factors can also contribute to these metabolic products. The dietary content of phenol, for example, present in curry, was found to significantly affect phenol measurements and potentially compromise urinary phenol measurements as a biomarker of benzene exposure.

Other considerations in selecting the biomarker(s) for the study and which impact the study design are ***reliability*** - the extent to which an experiment or measuring procedure yields the same results (tendency towards consistency) on repeated trials, and ***validity*** - the (relative) lack of systematic measurement error when comparing the actual observation with a standard - a reference method that represents the 'truth'.

Validation of the biomarker needs to be conducted both at the population level and also at the laboratory level. For the population, factors considered such as intra-individual variation over time - both without altering exposure and when exposure is removed or changed; inter-individual variation and response to a given exposure; and feasibility, i.e. amount and availability of tissue, cost, and time required for each assay. For the laboratory, factors are the dose–response curve, detection limit or low-dose sensitivity, exposure specificity, reliability of the assay (run-to-run, day-to-day, and inter-laboratory), and optimal conditions for sample collection, processing, and storage.

Population sampling requires adequate numbers of both sexes of all ages and also 'the possibly at risk' subgroups that might need to be included based on proximity to an exposure source, etc. The results of this sample will be extrapolated statistically to the larger population group as a whole. It is essential that appropriate statistical principles be followed when sampling populations for biological monitoring, and that the pitfalls of 'convenience sampling' through use of readily available, localized groups be avoided.

Ethics

Ethical considerations in the use of the biomonitoring data involving collection of human samples in any surveillance activity have to be carefully considered *a priori* and center around the issue of confidentiality and communication. Usually, for biomonitoring of workers carried out and in accordance with regulatory requirements, ethics and communication issues have been addressed through accepted procedures, including the need for approval by institutional review boards (IRBs). Biomonitoring of broader populations such as community studies are not so well defined and these public health surveillance activities need ethical review by the local and national authorities regardless of whether this is a research initiative or routine surveillance program.

Interpretation of Data

Once reliable biomonitoring measurements have been made, there are two types of approaches that are used in the interpretation. The qualitative statistical approach; the other is a risk-based approach.

Establishing a reference range is an example of the statistical approach in which the level of chemical is compared with the percentile of the normal range for the occupational or general population. Exceeding this range may indicate a health risk. Establishing a reference range over time and then comparing subgroups against this has been a main focus of the NHANES and CDC studies mentioned earlier. For the workplace, there have been actual reference values established for a number of chemicals; examples are the Biological Exposure Indices (BEI) of the American Conference of Industrial Hygienists, and the Deutsche Forschungsgemeinschaft (DFG) Biological Tolerance Values for Occupational Exposures (BAT). But while these values provide a useful frame of reference and can form the basis for pharmacokinetic modeling to derive population values, the adjusted BEI or BAT values per se do not appear appropriate as guidance values for the population except that they can be used as the starting point for further discussion.

The above, while identifying the potential of a health risk, do not provide any evaluation of this risk. For this, a risk-based approach, ideally using an existing risk assessment where the hazard has been identified and dose responses have been established, can enable an assessment of the new biomonitoring data by direct comparison. Where such human response data are not available, animal toxicology studies have been combined with human exposure data to estimate risks to the general population. Sometimes the only exposure information available is the biomarker/biomonitoring data, which can then be equated to the animal response through appropriate physiologically based pharmacokinetic modeling.

Communication of findings

Biomarkers may indicate undesirable exposures or potential adverse health effects and the need to intervene; however, this assessment should only be reached after proper consideration of all associated data and through a risk-assessment process as previously described. Mere presence does not necessarily mean harm. Before commencing a study, communication strategies need to be in place to preserve anonymity and minimize the chance for misinterpretation or misuse of the data, yet ensure an appropriate response is taken if a health risk is found to be present.

4.3.5 Case Study Examples

The following are examples of agents included in biomonitoring programs such as the CDC's. The first two, lead and mercury, are examples of where there is substantial epidemiological information as to their health effects and the associated dose–response, whereas the perfluorinated compound (PFOS) has few human health effect data - though animal toxicity data are accumulating.

Lead

Blood lead levels in excess of 10 μg/L are associated with neurological effects in children, which include behavioral changes, loss of concentration, and neurological damage. As noted, the withdrawal of lead in gasoline has led to a substantial fall in blood lead levels in children, but other sources of lead such as lead in paint - likewise

discontinued - still provide evidence of excess of blood lead levels through children's ingestion of paint flakes (a condition known as pica).

Mercury

The neurotoxicity/neurobehavioral effects of mercury are well documented in epidemiology studies. There are two biomarkers that have been used effectively as a determinant of internal dose: mercury in hair and mercury in blood, including cord blood. Both reflect total mercury - organic and inorganic and elemental mercury. It is normally dimethylmercury from eating fish that makes up the major proportion; inorganic mercury is converted into the organic form by bacterial action in aquatic sediments and enters the food chain through uptake by plankton, etc. Organic mercury in cord blood has been correlated with postnatal neurobehavioral effects; blood levels of 5.8 ug/L in children have been associated with losses in IQ.

Potassium Perfluorooctanesulfonate (PFOS)

This results from the degradation of intermediates used to produce paper, textiles, and food packaging. Recent studies show PFOS in environmental and biological matrices globally. It is persistent and not known to degrade. Median human serum level has been shown to be 35.8 ng/ml, with an estimated half-life of 8.7 years.

At low doses (< 10 mg/kg/day) PFOS causes liver damage in rodents and primates. It is also associated with carcinogenic and developmental toxicity. Pharmacokinetic modeling in the rat has been used in conjunction with the available human biomonitoring data for risk estimation and showed a Margin of Exposure (MOE) in excess of 1000-fold.

Methyleugenol

This final example, which occurs naturally in spices and foods, is not associated with any human health effects, but does produce cancer in rodents. Ingestion of ginger snap cookies containing approximately 216 µg methyleugenol resulted in peak serum concentrations of 25–100 pg/g. In the metabolism of methyleugenol, it appears that the 1-hydroxylation pathway is important at high doses (> 10mg/kg bw) and may be implicated in the cancer findings in rodents, but at low doses, such as encountered by the general population, the concentration of the 1-hydroxy metabolite based on the urinary excretion biomonitoring data has been found to be very low.

4.3.6 Summary

Biomonitoring is a powerful and versatile surveillance tool with demonstrated benefit in identifying health risks and unsuspected exposure sources. However, the ability to properly interpret new biomonitoring data, which are generated using increasingly sensitive analytical methods, is often outpaced by the technology. This challenge can only be met by integrating valid biomonitoring data, together with accompanying relevant information, into an appropriate risk-assessment process.

References

European Commission (2004). Baseline Report on 'Biomonitoring for Children' in the Framework of the European Environment and Health Strategy (COM(2003)338 final). Prepared by the Technical Working Group on Integrated Monitoring. [Online 9 January 2004] http://www.brussels-conference.org/Download/baseline_report/BR_Biomonitoring_final.pdf

WHO (World Health Organization) (2001). Biomarkers in Risk assessment: Validity and Validation. Environmental Health Criteria 222. Geneva, World Health Organization.

National Research Council of the National Academies (2006). Human Biomonitoring for Environmental Chemicals. The National Academies Press, Washington, DC.

Further Reading

Albertini, R. J. (1998). The Use and Interpretation of Biomarkers of Environmental Genotoxicity in Humans. *Biotherapy*, **11**, 155–167.

Albertini, R., Bird, M.G., Doerrer, N.G., Needham, L., Robison, S., Sheldon, L., and Zenick, H. (2006). The Use of Biomonitoring Data in Exposure and Human Health Risk Assessment. *Environ. Health Perspect*, **114**, 1755–1762.

Groopman, J. D. and Kensler, T.W. (1999). The Light at the End of the Tunnel for Chemical-specific Biomarkers: Daylight or Headlight? *Carcinogenesis*, **20**, 1–11.

4.4 Epidemiology

Kurt Ulm

4.4.1 Introduction

The goal of epidemiology is to study the distribution and determinants of health-related status or events in human populations, and the application of these studies to control of health problems.

One of the first epidemiological studies was undertaken in the middle of the 19th century, when in 1854 John Snow in London linked the mortality due to cholera with the quality of the drinking water.

A more systematic development of epidemiological methods started in the middle of the 20th century when the research focused on chronic diseases such as cancer and coronary heart diseases. In 1949 the Framingham study began to identify risk factors for cardiovascular events. In the field of cancer, the first studies were designed to explore the relationship between smoking and lung cancer.

4.4.2 Measures to Describe the Risk

The risk of a certain disease can be described for a specific point in time (prevalence) or for a time period (incidence)

The proportion in the population to be studied can be defined by demographic factors such as age, gender, or location. There are the two ways to describe the risk of disease. The **prevalence** rate is the number of individuals who have a specific disease at a particular time divided by the population at risk at that point in time [Equation (4.1)]. The **incidence** describes the rate at which new cases occur. The incidence rate is given as the number of new cases in a defined period of time divided by the population at risk of experiencing the event during this period, expressed as person-time or person-years at risk [Equation (4.2)].

$$\text{prevalence} = \frac{\text{number of persons with specific disease at a given point in time}}{\text{persons under risk at this time}} \quad (4.1)$$

$$\text{incidence} = \frac{\text{number of new cases within a defined period}}{\text{person-years at risk within this period}} \quad (4.2)$$

A high prevalence can be associated with a low incidence (e.g., diabetes) or a low prevalence with a high incidence (e.g., infectious disease). The link is the duration of the disease.

To give an example of prevalence a study investigating the association between phenacetin and renal disease will be considered. 7311 women were asked about the use of phenacetin and 623 women responded positive. 39 out of the 623 women had a renal disease, leading to a prevalence of 6.3% (= 39/623 · 100%). Among the remaining women not taking the drug a random sample of 621 women was selected. Out of this group 19 had a renal disease, denoting a prevalence of 3.1% (= 19/621 · 100%) among the nonusers.

To calculate the incidence the number of new diagnosed cases as well as the period of time is important. In a dynamic population, like the work force in a company, an individual observation period of each worker is necessary. The start of the observation is either a fixed point in time or the time at which the individual entered the company, whichever occurred later.

The end of the observation is defined by one of three events:

- The person is diagnosed with the disease;
- The end of observation is reached;
- The person drops out.

To give an example of *incidence* a study from Denmark can be used in which the cancer incidence of workers exposed to 2,4-D, 2,4,5-T, and dioxin, i.e. 2,3,7,8-TCDD, during the production of herbicides was estimated. In a cohort of 3390 workers 159 were diagnosed with cancer between 1964 and 1981. The person-years at risk of the 3390 workers totaled 49 879 years, or an average of 14.7 years per worker. The incidence rate per year is calculated as 159/49 879 = 0.0032 per person-year or 3.2 cases per 1000

person-years. It is important to note that 1000 person-years may add up when 100 persons are observed over 10 years or 1000 persons over 1 year.

The mortality rate is calculated in the same way as the incidence. Within the extended follow-up of a British-doctor study, which ran from 1950 until 1978, about 195 doctors died of lung cancer with 109 386 person-years at risk, leading to a lung cancer mortality rate of 1.78 per 1000 person years ($= 195/109\,386 \cdot 1000$).

In order to investigate the association between a certain factor and the disease one has to compare either the prevalence or the incidence between exposed and unexposed or high- and low-exposed persons. The difference between the two groups can be described by the ratio of either prevalence, or incidence, or mortality, and is denoted as **relative risk** (*RR*) [Equation (4.3)].

In the British-doctor study the mortality rate among the smokers was 1.78 per 1000 person-years. Among the nonsmokers the corresponding mortality rate was 0.086 per 1000 person-years. The relative risk for smokers dying of lung cancer compared with nonsmokers is therefore $RR = 1.78/0.086 = 20.7$. The interpretation of the relative risk is that the mortality rate for smokers dying of lung cancer is 20.7 times higher than for nonsmokers.

Besides the relative risk there are some other measures to describe the difference between exposed and the unexposed persons (e.g., Rothman and Greenland, 1998). Two of them are the **excess of risk** (*ER*), which is the difference between both rates [Equation (4.4)], and the **attributable risk** (*AR*), denoting the proportion of events among the exposed persons that are due to exposure [Equation (4.5)].

I_1: incidence rate in the exposed group;

I_0: incidence rate in the un-exposed group

$$\textbf{relative risk}(\textbf{RR}) = I_1/I_0 \tag{4.3}$$

$$\textbf{excess risk}(\textbf{ER}) = I_1 - I_0 \tag{4.4}$$

$$\textbf{attributable risk}(\textbf{AR}) = \frac{I_1 - I_0}{I_1} = \frac{RR - 1}{RR} \tag{4.5}$$

In the British-doctor study the excess risk gives a value of 1.694 per 1000 person-years. The attributable risk is $AR = 95.2\%(= 1.78 - 0.086/1.78)$. The interpretation is that 95.2% of all lung cancer death among the smokers is due to smoking ($195 \cdot 95.2\% \approx 185$ deaths).

The comparison of the rates for the exposed and unexposed persons gives an average value for the relative risk. More information can be drawn if the exposed group is divided with respect to the exposure and a dose–response relationship can be established.

In the British-doctor study the smokers are classified into certain categories with respect to the cigarettes smoked per day. In Table 4.9 the relative risk for certain

Table 4.9 *Results of the British-doctor study (Doll and Peto, 1978).*

Cigarettes/day	Lung cancer	Person-years	Mortality rate ·1000	*RR*
0	6	69 905	0.086	1
1–9	7	14 877	0.471	5.48
10–19	34	35 037	0.970	11.31
20–29	88	42 862	2.053	23.92
≥ 30	66	16 610	3.973	46.29

categories compared with nonsmokers is given. This analysis shows an increase in the risk depending on the cigarette consumption.

4.4.3 Standardization

The exposed and unexposed groups may also be different in other factors, like age, gender, smoking, etc. In this situation an increase in the relative risk cannot immediately be attributed to the exposure. One has to adjust for the differences in the other factors. Two approaches are possible, called direct and indirect standardization.

Direct Standardization

The procedure will be explained in relation to age.
The data are divided into several age groups ($i = 1, \ldots, I$). The incidence or mortality rates are calculated within each age group separately for the exposed and unexposed persons. Afterwards the rates are summed using weights w_i for the different groups. Mostly the weights w_i are determined by the proportion of person-years of that particular age group among the total person-years.

Importantly the same weights w_i are used for the summation of the rates for the exposed and the unexposed group, which gives the standardized rates.

The standardized incidences rate is given by Equation (4.6), where $I_i =$ incidence rate in age groups ($i = 1, \ldots, I$) and $w_i =$ weight of age group $i(\sum w_i = 1)$.

$$SI = \sum w_i I_i \tag{4.6}$$

The ratio of the standardized incidence rates between the exposed and unexposed persons is denoted as relative risk adjusted for age.

Example:

Again the British-doctor study will be used. The smokers with 30 cigarettes or more per day are compared with the nonsmokers. The unadjusted relative risk is 46.29 (see Table 4.9). If the data are separated into three age groups, the following result is obtained (see Table 4.10).

Table 4.10 *Results of the British-doctor study (Doll and Peto, 1976).*

Age group	Nonsmoker LC*	Smoker (≥ 30 cigarettes per day) $\sum t^*$	LC*	$\sum t^*$	w_i^*
<50	0	33.679	3	5.881	0.46
50–64	3	27.380	34	8.682	0.42
≥ 65	3	8.846	29	2.047	0.12
Σ	6	69.905	66	16.610	1.00

* LC = lung cancer.
$\sum t$ = sum of person-years at risk.
w_i = weight of age group i.

The standardized mortality rate (SM_s) among the smokers is given as:
$SM_s = 3/5881 \cdot 0.46 + 34/8682 \cdot 0.42 + 29/2047 \cdot 012 = 3.649$ per 1000 person-years.

The SM_s among the nonsmokers gives a value of $SM_s = 0.088$ per 1000 person years. Therefore, the relative risk adjusted for age is 41.29 (= 3.649/0.088), which is somewhat smaller than the unadjusted relative risk of 46.29.

Another way to adjust for certain factors is to use a regression model, like the logistic regression or the Cox model (see Rothman and Greenland, 1998), taking into account the various factors.

Indirect Standardization

The incidence or mortality rate of a certain group can be compared with a standard population, e.g. the population of a certain area or country. This approach is very common in occupational epidemiology in order to describe the risk of a certain group of workers in relation to the total population. The corresponding rates for the population are usually available. The rates for the population are given separately for age groups divided into 5-year intervals for gender and calendar time. Based on these rates the so-called expected number of events can be calculated using the observed person-years at risk. The ratio of the number of observed to expected events is denoted as standardized incidence ratio (SIR) or standardized mortality ratio (SMR) adjusted for age, gender, and calendar time.

Example:

The mortality rate of workers exposed to silica dust, who received worker's compensation from the stone and quarry industry for contracting silicosis on the job between 1988 and 2000, has been investigated (Ulm et al., 2004). A cohort of 440 workers was enrolled and followed up until the end of 2001. Within this period 144 workers died. Based on the mortality rates of Germany, 74.35 deaths could be expected, leading to an SMR of 1.94 (= 144/74.35). Therefore, nearly twice as many deaths were observed compared with the mortality rates for Germany. The calculation can be performed for overall mortality as well as for specific causes of death. It is important to note that the specific cause of death will be obtained from death certificates.

4.4.4 Types of Epidemiological Studies

In general there are two types of epidemiological studies: observational and experimental studies

The observational studies are further divided into cohort and case-control studies (Figure 4.15). Within a **cohort study**, also called a prospective study, a defined group of persons are followed up over a certain period of time. In the British-doctor study all physicians willing to participate were enrolled in this study. The whole cohort can further be divided into different sub-cohorts, e.g. with respect to the level of exposure, age, gender, etc.

Incidence, Mortality Rates, and Risk can be Determined in a Cohort Study

The span of time between the onset of exposure and the diagnosis of a disease is called the latency period. For mesothelioma, the latency period is a least 35 years. For a design of a cohort study to investigate the health effect of a certain factor one has to keep in mind that the follow-up period has to be long enough to avoid false negative results. The only way to shorten the follow-up period is to transfer the onset of the observation into the past. This approach is called an historical cohort study and is commonly used in occupational epidemiology.

In a **case-control study**, also called retrospective study, persons with a defined disease (cases) are compared with disease-free persons (controls). To investigate the association between smoking and lung cancer, patients with and without lung cancer are asked about their smoking habits. This type of study is faster and more efficient than cohort studies if the disease is rare or has a long latency period. The disadvantage of this type of study is that no incidence rate can be estimated and, therefore, no relative risk can be calculated. Nevertheless, case-control studies permit an approximation of the relative risk.

Assume we have a cohort study with P_1 exposed persons followed up over an average of t years. If a certain number (a) out of the P_1 persons develop the disease, the incidence

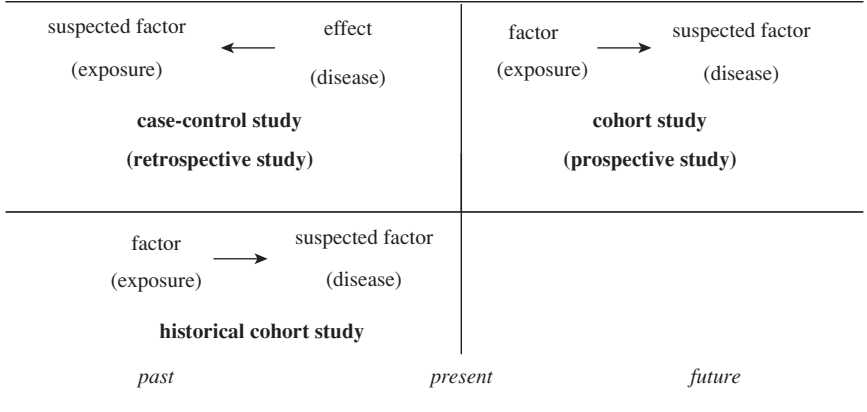


Figure 4.15 *Design of a cohort study and a case-control study.*

is given as by Equation (4.7). The incidence among the nonexposed group is expressed as Equation (4.8).

$$I_1 = a/(P_1 \cdot t) \tag{4.7}$$

$$I_0 = b/(P_0 \cdot t) \tag{4.8}$$

The number of new cases in both cohorts are denoted by a and b; the persons without the disease in both cohorts are, respectively, called c and d, where Equation (4.9a) and (4.9b) hold. If the disease is rare the size of each cohort (P_0 and P_1) can be replaced by the number of persons without the disease (c and d) and both incidence rates can be approximated by Equations (4.10) and (4.11).

$$P_1 = a + c \tag{4.9a}$$

$$P_0 = b + d \tag{4.9b}$$

$$I_1 = a/(P_1 \cdot t) \approx a/(c \cdot t) \tag{4.10}$$

$$I_0 = b/(P_0 \cdot t) \approx b/(d \cdot t) \tag{4.11}$$

The relative risk $RR = I_1/I_0$ can be estimated by Equation (4.12):

$$RR = I_1/I_2 \approx \frac{a/(c \cdot t)}{b/(d \cdot t)} = \frac{a/c}{b/d} = \frac{a/b}{c/d} \tag{4.12}$$

The numerator (a/b) is the ratio of the exposed and unexposed cases, called the odds of the cases [odds (K^+) with K^+ denoting the disease]. The denominator (c/d) gives the odds among the nondiseased persons, the controls [odds (K^+)]. The ratio of both odds is called odds-ratio (OR) and is an estimate of the relative risk. The great advantage of the odds-ratio is that it can be calculated even with a random sample of the controls. If among both cohorts (exposed and unexposed) the same proportion (p) of the non-diseased persons (controls) are sampled, e.g. $p = 10\%$, then the odds (K^-): are estimated by Equation (4.13). For estimating the relative risk it is necessary to get full information from all cases and only from a proportion of the controls. However, it is important to note that the proportion p has to be the same in both cohorts. Random sampling is essential, otherwise one gets a bias, which can lead to an over- or under-estimation of the relative risk.

$$\text{odds}(K^-) = (p \cdot c)/(p \cdot d) = c/d \tag{4.13}$$

A case-control study can also be performed within a cohort study. The cases are compared with a sample of controls. This approach is called a nested case-control study.

Example:

In order to investigate the lung cancer risk among plutonium workers, a cohort of about 22,000 workers from a plant in the US was identified (Brown et al., 2004). The internal lung dose was estimated for all lung cancer cases ($n = 180$) and a randomly selected group of workers without lung cancer ($n = 718$); 25 cases and 103 controls had no exposure, the remaining 165 cases and 615 controls were exposed. These figures give an OR of $(165/25)/(615/103) = 1.11$. The advantage of this approach is that for estimating the lung dose a relatively small sample is adequate.

In addition to observational studies, it is possible to perform an experimental study. For example, one can evaluate the effects of a certain intervention, which is similar to a controlled clinical trial. Subjects are randomly assigned to the intervention or the control group. An example is the Women's Health Initiative study (WHI, 2002). Within this study the risks and benefits associated with the use of hormones in healthy postmenopausal women were investigated. About 18 000 women were enrolled. About half of them received estrogen plus progestin, whereas the other half received placebo. After a follow-up of 5 years it was shown that treatment resulted in a decrease in fractures to the hip and other fractures ($RR = 0.76$ for hormone users). However, cardiovascular diseases and cancer showed increased incidence rates ($RR = 1.29$ and 1.26, respectively) in hormone users.

The intervention can also be a certain diet or exercise program. An example is the MRFIT study, in which several preventive measurements for reducing the risk of myocardial infarction in high-risk patients were investigated (Kolata, 1982).

Besides the cohort and case-control studies there are two other types of studies: cross-sectional and ecologic studies. Within a cross-sectional study the exposure and the disease are ascertained at the same point in time. With these data one can estimate the prevalence. However, it remains unclear whether the exposure has an influence on the risk. Within an ecologic study groups of persons rather than individuals are considered. In this type of study the correlation between a certain factor and a specific disease can be investigated.

4.4.5 Statistics

To ensure the significance of the values determined for incidence, mortality rate, prevalence, relative risk, or odds-ratio, statistical tests and the calculation of confidence limits are necessary.

The interpretation of a *P%* confidence interval is as follows: with a probability of *P%* (mostly *P%* = 95%) the 'true' value of the parameter of interest is within the interval. Based on statistical considerations it follows that the logarithm of the incidence (*I*) is approximately normally distributed. If the observed number of events is denoted by *d* and the person-years risk is denoted by *person-year*, the incidence *I* is estimated by Equation (4.14). The 95%-confidence interval (*95%-CI*) for the 'true' incidence is given by Equation (4.15).

$$I = d/\textit{person-year} \tag{4.14}$$

$$95\%\ CI = I \cdot \exp(\pm 1.96/\sqrt{d}) \tag{4.15}$$

Example:

TCDD study from Denmark:

$d = 159$; person-year = 49 879 years; $I = 3.2$ cases per 1000 person-years

95%-CI: 2.7–3.7 cases per 1000 person-years

Relative Risk (*RR*):
The incidence rates in the exposed and unexposed groups are denoted by I_1 and I_0, the number of observed events by d_1 and d_0. The 95%-*CI* around the estimate of *RR* $(= I_1/I_0)$ is given by Equation (4.16).

$$95\%\text{-}CI = RR \cdot \exp\left\{\pm 1.96\sqrt{1/d_0 + 1/d_1}\right\} \tag{4.16}$$

In addition to the calculation of the 95%-*CI* a statistical test can be performed. The hypothesis of interest is whether *RR* is equal to 1 or different from 1 ($H_0 : RR = 1$ vs. $H_1 : RR \neq 1$).

The test statistics can be derived as Equation (4.17):

$$T = \ln RR \Big/ \sqrt{1/d_1 + 1/d_0} \tag{4.17}$$

If T is between $\pm$ 1.96 the null-hypothesis ($H_0 : RR = 1$) cannot be rejected at a significance level of $\alpha = 5\%$. In this situation the 95%-*CI* covers the value of 1, the true value of *RR* under H_0. If T is larger than 1.96 or below -1.96, H_0 is rejected and the value of 1 is outside the 95%-*CI*.

Odds-ratio (*OR*):
The calculation of the exact confidence interval for the *OR* is difficult. Therefore, some approximations are available. One formula is called a test-based approach and uses the conventional χ^2-statistics for investigating the association between a factor and a disease. The 95%-*CI* is calculated as shown in Equation (4.18):

$$95\%\text{-}CI = OR^{1\pm 1.96/\sqrt{\chi^2}} \tag{4.18}$$

In the above example about the association between the exposure to plutonium and the lung-cancer risk, an *OR* of 1.11 had been observed. The corresponding χ^2-tests for proving the hypothesis $OR = 1$ gives a value of $\chi^2 = 0.17$. This leads to a 95%-*CI* of 0.69–1.77. Therefore H_0 cannot be rejected and no statistically significant association between plutonium and lung cancer can be observed for this example.

Standard mortality ratio (SMR):
The approximation of the 95% *CI* is based on the formula for the *RR*, and is shown in Equation (4.19). The number of events among the unexposed ($= d_0$) is much larger than the number of events among the exposed ($= d_1$) and $1/d_0$ is approximately 0.

$$95\%\text{-}CI = SMR \cdot \exp\left\{\pm 1.96/\sqrt{d_1}\right\} \tag{4.19}$$

There are also other formulas available (e.g., Ulm, 1990).

4.4.6 Causality, Bias, Confounding, Chance

In epidemiological studies only statistical associations can be observed. In order to establish causality certain criteria have to be fulfilled (Hill, 1965). Before proving these criteria three other aspects have to be considered: bias, confounding, and chance.

A series of criteria that should be met when attempting to establish a 'cause and effect' relationship between exposure to a certain compound and subsequent disease was postulated by Hill (1965). Prior to utilizing these criteria it is important to validate the data to ensure that they cannot be impugned because of bias, confounding factors, or chance.

Bias is any systematic error introduced into the process of collecting and evaluating data, which leads to failure to detect the truth. 'Selection bias' and 'Information bias' can be important sources of error. Selection bias in case-control studies can arise from the improper selection of controls or from a participation rate of study subjects. Information bias can result from investigation of cases and controls using different methodologies. The same holds for cohort studies in the event of misdiagnosis of disease or incorrect pronouncement of the cause of death. There is no way to correct for these errors.

Confounding represents a source of bias that occurs when an extraneous variable interferes with the determination of the effect of the variable under study. The extraneous variable may occur in both the control and experimental group and unless accounted for may lead to erroneous interpretation of the data. For example, a study is conducted to determine whether a black strain of mice metabolizes a drug more rapidly than a white strain of mice. The results do not indicate a difference. It is later determined that both strains were fed a diet containing a pesticide that induced the metabolism of the chemical in both strains. Repetition of the experiment using a noninducing diet shows that the black mice metabolized the drug faster than the white. Thus, the results were confounded by the unknown administration of the pesticide to both species.

Chance may play an important role in the outcome of an epidemiological study in the absence of bias. For example, if a variety of endpoints are investigated within a cohort study and several of them are statistically significant one is faced with a multiple-comparison problem. If several tests are performed the type-I error is inflated. The same problem occurs in case-control studies if more than one factor is considered. From the statistical point of view there are two ways to overcome this problem. One way is to select one primary endpoint, i.e. the disease or the factor of greatest interest. This analysis is called confirmatory. All other diseases or factors are considered as secondary endpoints. The outcomes of these diseases or factors can be used to generate new hypotheses. The other approach is to adjust the type-I error by altering the number of tests $(= n)$ performed. The type-I error for the single test is reduced to $\alpha^* = \alpha/n$, with α the overall type-I error, e.g. $\alpha = 5\%$. The p-value of a particular test has to be below α^* in order to be statistically significant at the α-level.

Criteria for Causality

If an association between a factor and a disease is statistically significant, and bias, confounding, or chance can be ruled out, criteria for establishing causality would include:

- *Temporality:* Taking appropriate latency periods into account, does exposure to a chemical precede the disease?
- *Consistency:* Consistency refers to the repeated observation of an association between exposure and outcome in different populations under different circumstances. This criterion can often help to distinguish between real associations and chance.

- *Dose–response relationships*: Strong support for a causal relationship can come from observing a biological gradient, i.e. a monotonic, increasing dose–response curve.
- *Strength:* A strong association is more likely to be causal than a weak association. However, the strength of the association may be modified by the level of exposure.
- *Experimental evidence:* This criterion refers to the situation in which an intervention is performed to reduce the exposure and a decrease in the risk is observed.
- *Plausibility:* There should be a biologically plausible explanation connecting the exposure and the disease. Plausibility may depend on the extent to which the mechanism of action of the agent is understood.

4.4.7 Summary

The goal of epidemiology is to identify risk factors for certain diseases. Several types of epidemiological studies are available. The most well known are the cohort study and the case-control study. The association between a factor and a disease is described by the relative risk or the odds-ratio. If a statistically significant association can be observed, causality has to be proven. A variety of criteria are mentioned. Not all of them have to be fulfilled in order to establish causality, but, before doing so, the influence of bias, confounding, or chance has to be ruled out.

References

Brown, S. C., Schonbeck, M. F., McClure, D., Barón, A. E., Navidi, W. C., Byers, T., and Ruttenber, A. J. Lung Cancer and Internal Lung Doses among Plutonium Workers at the Rocky Flates Plant: A Case-Control Study. *Am. J. Epidemiol.*, **2004**, *160*, 163–172.

Hill, A. B. The Environment and Disease: Association or Causation? *Proc. R. Soc. Med.*, **1965**, *58*, 295–300.

Kolata, G. Heart Study Produces a Surprise Result. *Science*, **1982**, *218*, 31–32.

Rothman, K. J. and Greenland, S. *Modern Epidemiology.* Little Brown and Company, Boston, **1998.**

Ulm, K. A Simple Method to Calculate the Confidence Interval of a Standardized Mortality Ratio (SMR). *Am. J. Epidemiol.*, **1990**, *131*, 373–375.

Ulm, K., Gerein, P., Eigenthaler, J., Schmidt, S. and Ehnes, H. Silica, Silicosis and Lung-cancer: Results from a Cohort Study in the Stone and Quarry Industry. *Int. Arch. Occup. Environ. Health*, **2004**, *77*, 313–318.

WHI: Women's Health Initiative Investigators: Risks and Benefits of Estrogen plus Progestin in Healthy Post-menopausal Women: Principal Results from the Women's Health Initiative Randomized Controlled Trial. *JAMA*, **2002**, *288*, 2819–2825.

For Further Reading

Dawber, T. R., Moore, F. E., and Mann, G. V. Coronary Heart Disease in the Framingham Study. *Am. J. Public Health*, **1957**, *47*, 4–24.

Doll, R. and Hill, A. B. A Study for the Aetiology of Carcinoma of the lung. *Br. Med. J.*, **1952**, *2*, 1271–1286.

Doll, R. and Peto, R. Cigarette Smoking and Bronchial Carcinoma: Dose and Time Relationships among Regular Smokers and Lifelong Non Smokers. *J. Epidem. Community Health*, **1978**, *32*, 303–313.

Dubach, U. C., Levy, P. S. and Mueller, A. Relationships between Regular Intake and Urorenal Disorders in a Working Female Population of Switzerland. I. Initial Results (1968). *Am. J. Epidemiol*., **1971**, *93*, 425.

Last, J. M. (Ed.): *A Dictionary of Epidemiology*, Oxford University Press, New York, **1988**.

Lynge, E. A Follow-up Study of Cancer Incidence among Workers in Manufacture of Phenoxy Herbicides in Denmark. *Br. J. Cancer*, **1985**, *52*, 259–270.

Snow, J. *On the Mode of Communication of Cholera*, 2nd ed., John Churchill, London, **1860**. (Facsimile of 1936 reprinted edition by Hafner, New York, **1965**.)

Wynder, E. L. and Graham, E. A. Tobacco Smoking as a Possible Aetiological Factor in Bronchogenic Carcinoma. A Study of Six Hundred and Eighty-four Proved Cases. *JAMA*, **1950**, *143*, 329–336.

4.5 Omics in Toxicology

Laura Suter-Dick and Thomas Singer

4.5.1 Introduction

Holistic approaches or 'omics' technologies are changing the way toxicology studies are performed and safety data are generated and interpreted. When speaking about 'Omics' technologies, we usually refer to genomics, proteomics, and metabonomics technologies. These new approaches provide the means to predict toxicity and to increase the understanding of the molecular events underlying a given toxicity. Moreover, toxicogenomics is helping scientists to integrate toxicology into the earlier discovery phases by including sensitive parameters that should help toxicologists recognize liabilities at lower doses (pharmacological rather than toxicological doses) or after short exposure times (acute rather than chronic exposures).

4.5.2 Concept of Toxicogenomics

***Toxicogenomics* emerged in the late 1990s as a contraction of *toxicology* and *genomics*. The development of the gene-expression microarrays, also known as chips, enabled simultaneous analysis of the transcription levels of thousands of genes to assess global gene expression (called *genomics or transcriptomics*). The application of these novel technologies in toxicology gave rise to toxicogenomics as a discipline. At present, toxicogenomics is understood as the use of genomics (or transcriptomics) for the study of toxic effects, while proteomics and metabolomics are often included into the broader definition of the toxicogenomics discipline.**

Originally, toxicogenomics referred to the application of genomics technologies to study adverse effects of chemicals. The advances in protein-sequence knowledge, in 2D-electrophoresis, and more recently on protein chips, as well as peptide identification made *proteomics* analysis possible. Thus, large numbers of proteins in a given biological sample can now be determined simultaneously. In the field of endogenous metabolites, *metabolomics* using H-NMR and/or LC-MS/GC-MS made it also possible to determine all the metabolites present in a biofluid (generally urine or serum) simultaneously. Thus

the holistic, global approach comprises the analysis of gene expression (*genomics or transcriptomics*), of proteins (*proteomics*) and of metabolites (*metabolomics*). The parameters most commonly assessed by each of these technologies are summarized in Table 4.11.

There are other *omics* that can be used and have been reported in literature, such as lipomics, cellomics, etc., but their use is not yet widespread in toxicology.

The concept of using gene- and protein-expression measurements, as well as metabolite composition, for mechanistic and predictive toxicology is not completely new. However, it is the amount of information that can be gathered with relatively small effort using new tools that has transformed molecular toxicology. It was less than 50 years ago that the structure and function of DNA was deduced and by the end of 2002 the genomes of 800 organisms, including the human, have been fully sequenced. Alongside the major genome-sequencing projects, protein databases now include thousands of identified proteins with their sequences and the structure of a fair number of endogenous metabolites is also available in the public domain or in proprietary databases. Several publications have corroborated the use of toxicogenomics for the prediction of toxic liabilities, for the investigation of toxic mechanisms, and for the identification of biomarkers of toxicity.

4.5.3 Technology Platforms

This part of the chapter introduces the technological tools more widely used in the toxicogenomics field and the general principles they are based on. For the interested reader, additional specific literature is recommended in the reference section.

Genomics

Nucleic acids have the characteristic of having a four-nucleotide code with very well defined complementarity, which is used during the biological processes of replication and transcription. Microarray technologies have made use of this biochemical property and use highly specific sequences (generally called *probes*) fixed to a solid support for the quantification of the genes of interest (generally called *targets*). For all represented genes, this technology quantifies the amount of transcript present and is therefore called transcriptomics or genomics. These two terms are commonly used as synonyms.

Methods For each target gene, one or several specific probes complementary to a stretch of its sequence are represented on the microarray. Most test systems (platforms) use 20- to 70-bases-long probes. The high spatial density of the microarrays allow tens of thousands of probes, covering virtually all known transcripts, to be represented in one single microarray and thus measured simultaneously. Hence, the whole *transcriptome* (transcriptome is the sum of all transcribed genes in a given tissue at a given time) for a given tissue can be assessed with one microarray experiment. For this, all transcripts (messenger RNA, mRNA) from a given tissue or cell culture are extracted, usually amplified, labeled, and hybridized onto the microarray. Hybridized probes are measured (usually fluorimetrically) generating an output where the intensity of the signal detected

Table 4.11 *Parameters most commonly assessed by ''Omics'' technologies.*

''Omics''	Description	Biologic material	Analytical method	Parameter	Example	No. of analytes
Genomics/ Transcriptomics	Gene-expression analysis	Tissue, cell cultures	Microarray qPCR	mRNA (transcripts)	Induction of Heat Shock Proteins after cell damage	>10 000
Proteomics	Protein-expression analysis	Tissue, cell culture, serum, urine	2D-PAGE, MS, MALDI, SELDI, DIGE	Protein products, including post-translational modifications	Increased expression of P450 after treatment with inducers. Adduct formation	>1000
Metabonomics/ Metabolomics	Metabolic profiles of endogenous metabolites	Urine, serum	NMR, LC-MS GC-MS	Endogenous metabolites, small molecules	Reduction of intermediate products of the Krebs cycle due to mitochondrial damage	>500

for each probe is proportional to the amount of transcript (mRNA target) in the original sample. Some microarray platforms use two-color dyes for the simultaneous determination of the expression level in two samples, one being typically a baseline (control) and the other being the treated sample. In these cases, both the control and treated sample are hybridized together onto the same microarray exactly under the same hybridization conditions, thus minimizing the technical variability. However, the disadvantage of obtaining relative rather than absolute gene-expression values is that they always refer to the sample that was arbitrarily selected as baseline. In all cases, microarrays provide a (semi)-quantitative measurement of absolute or relative expression level for each gene represented on the microarray. Owing to the differences in the genetic code between the species and of the exact complementarity required between a probe on a microarray and the target in the tissue of interest, microarrays are highly species-specific. There are many commercial providers of microarrays including (but not limited to) Affymetrix, Applied Biosystems, Agilent, Illumina, and Codelink. Also, several laboratories in academia and industry use spotting devices to spot oligonucleotides and produce their own microarrays. In addition to the microarray platforms that rely on hybridization for the detection of mRNA-species, qRT-PCR (quantitative reverse transcriptase polymerase chain reaction) uses specific amplification of a transcript of interest as a means to determine the expression level of a gene. The amplified product is highly specific due to the use of specifically designed primers and can be quantified after each PCR-cycle by means of a fluorescent dye. Thus the reaction can be followed in real time (qPCR is often also called real-time PCR). In general, qPCR is considered to be more sensitive, specific, and accurate than most hybridization-based microarray platforms. Strictly speaking, qPCR is not an 'omic' application, since it can only be applied to a limited number of genes, rather than being a global, genome-wide approach. However, automation on 384-well plates and the use of commercially available microfluidics cards (e.g., from Applied Biosystems) allows us to measure up to 50 or 100 genes simultaneously using qPCR. Moreover, qPCR is currently the technology of choice for most scientists to corroborate results obtained using microarrays and plays therefore a major role in toxicogenomics.

Proteomics

Proteomics is the study of the proteome, defined as the sum of all proteins expressed in a given tissue at a given time. The proteome is more complex than the transcriptome, as it includes the study of post-translational modifications such as phosphorylation. Most proteomics platforms use the strategy of separation, quantification, and identification of the protein products to determine the proteome in a given sample.

Proteins are more complex chemically than nucleic acids. First, it needs to be taken into consideration that there is no natural biochemical complementarity as with the genetic code, which precludes a simple hybridization process. Moreover, most proteins undergo extensive modifications after translation, such as cleavage of signalling peptides, phosphorylation/dephosphorylation, or glycosylation. In addition, treatment with xenobiotics can also lead to drug–protein covalent binding, a protein modification of relevance

in toxicology. These post-translational modifications have a major impact on the function of the proteins. For example, the activity of many proteins is regulated through phosphorylation/dephosphorylation, which can cause dramatic biological changes, whereas the total protein content remains unchanged. In addition, proteins are present in serum and tissues with a very wide dynamic range: some proteins are highly expressed, while others are up to 10 orders of magnitude less abundant. For example, the serum concentration of albumin, the most abundant protein in serum, is about 10 billion times greater than that of interleukin-6. This complex situation puts the technologies around proteomics to a challenge, since such a dynamic range is extremely difficult to achieve.

Methods Separation can be achieved by fractionation of cellular components (sub-cellular fractionation) and further by two-dimensional gel electrophoresis (2D-PAGE) or chromatographic methods. Generally, 2D-PAGE is considered the workhorse of proteomics, although it is very time-consuming and lacks sensitivity. Low-abundance proteins and highly hydrophobic proteins such as cytochromes P450 can usually not be assessed with this method. As a further development of 2D-PAGE, and similarly to the two-color dyes in microarrays, 2D-DIGE includes a fluorescence-labeling step of the proteins before the electrophoresis and the simultaneous separation (by 2D-PAGE) of two samples labeled each with a different color. The advantage of 2D-DIGE is that it allows relative quantification while controlling for technical variability, increasing thus the dynamic range and sensitivity of traditional 2D-PAGE.

Retentate chromatography–mass spectrometry (RC-MS), also known as SELDI (surface-enhanced laser desorption/ionization) is an alternative to 2D-PAGE-based methods. SELDI uses surfaces with different physicochemical properties to separate the proteins in a complex mixture. The adsorptive chromatographic support, placed on thin metal chips, acts as a bait to adsorb proteins in the sample. This is based on the fact that different chemical surfaces present affinities to groups of proteins with specific characteristics. A mass-spectrum profile is then created by desorbing the proteins with the laser from the MALDI (matrix-assisted laser desorption ionization) instrument. This technology is relatively easy to use, requires a small sample volume, and allows a higher throughput than 2D-PAGE. However, SELDI has several technical drawbacks and has been reported to suffer from large experimental variability. Identification of the proteins after 2D-PAGE is usually performed by digesting the protein spots of interest with trypsin and by determining the masses of the tryptic fragments with mass spectroscopic (MS) approaches. The mass spectrograms are like fingerprints that can be compared to public databases (DB) for identification purposes, thus this process is called peptide fingerprinting. After SELDI analysis, the identification of the identified possible 'markers' is rather difficult. Some researchers try to use this platform to identify 'fingerprints' without identifying the actual proteins, but this is not ideal. If identification is required, the peaks of interest must be isolated to allow subsequent peptide fingerprinting or tandem-MS to be performed.

A different approach to proteomics is the use of protein arrays. The most frequently used protein array is the antibody array. These arrays rely on the spotting of specific antibodies for the proteins of interest, and are conceptually similar to the DNA-microarrays. The antibodies fixed onto a solid support can capture through affinity binding the proteins present in the complex mixture hybridized onto the chip. The main

limiting factor of antibody arrays is the need for specific, good-quality antibodies for a large number of proteins of interest.

Metabolomics/Metabonomics

***Metabolomics* studies the collection of all endogenous metabolites present in body fluid or tissue. While transcriptomic and proteomic analyses provide partial information regarding the biological processes occurring in a cell, metabolic profiling gives an overall picture of the cellular physiology. Moreover, changes in the metabolic profile reflect the reaction of the system to external stimuli, such as feeding or exposure to chemicals. These changes in the metabolic profile can be investigated in a cell, tissue, or organ.**

The metabolome is the final downstream product of the genome and is defined as the total quantitative collection of low-molecular-weight compounds (metabolites) present in a cell or organism that participate in metabolic reactions required for growth, maintenance, and normal function. Therefore, *metabolomics* refers to the study of the collection of all metabolites such as in a biological organism, which are the end products of the activity of expressed genes and proteins and their interactions. Some investigators define the *metabonome* as the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification. Others consider the metabonome as the sum of the metabolomes of each individual cell, organ, and tissue of a complex organism. In any case, *metabonomics* measurements give an overall picture of the physiology of an individual and its reaction to a given treatment, which can be measured in body fluids such as serum and urine. However, there is still some disagreement over the exact difference between metabolomics and metabonomics and in practice both terms are often used interchangeably.

The applications of metabolomics include the characterization and prediction of drug toxicity diagnosis and monitoring of clinical disease, the evaluation of therapeutic intervention, and the understanding of the effects of genetic modifications. These are achieved by measuring the metabolite profiles, generally in biofluids (urine, serum, plasma), although metabolomics measurements can also be applied to tissues. The study of the metabolome faces similar challenges to the study of the proteome, since it cannot rely on simple chemical hybridization to separate, quantify, and identify metabolites of interest. Metabolites are even more chemically diverse than proteins and they are also present in a wide dynamic range of concentrations, with differences in concentration of up to nine orders of magnitude. It is a goal of researchers in this field to develop technologies that enable large-scale high-throughput screening, without losing sensitivity. However, there is currently no technology that fulfils all these requirements and investigators are forced to evaluate the advantages and disadvantages of each platform for each experiment.

Methods Metabolomics uses technologies such as H-NMR (proton nuclear magnetic resonance) or a chromatographic separation [liquid (LC) or gas chromatography (GC)] coupled to a mass spectroscopic (MS) determination. NMR spectroscopy is the only

detection technique that does not rely on separation of the analytes and is not destructive, e.g. it allows the sample to be recovered for further analyses. It requires relatively little sample processing and the measurement times are short, allowing for relatively high throughput. All kinds of small molecules can be measured simultaneously in a complex mixture. However, it is relatively insensitive compared with mass spectrometry-based techniques and the processing of the NMR-spectra requires very specialized technical expertise and is relatively time-consuming. Additionally, the identification of metabolites of interest poses a challenge. The other technology commonly used for metabolic profiling is LC-MS or GC-MS. Complex samples are separated by liquid or gas chromatography and the masses of the components of each fraction are determined by MS. This method is more sensitive than NMR, but requires additional sample processing and the sample cannot be recovered for subsequent analyses.

4.5.4 Bioinformatics and Biostatistics

The technologies listed above allow the study of the whole transcriptome, proteome, and metabonome. Therefore, omics experiments produce large amounts of data. These enormous data sets consist generally of few replicates (3 to 10 biological replicates at most) and tens of thousands of parameters (genes, proteins, and metabolites). This requires new concepts for data processing and data analysis.

From a statistical point of view, these technologies pose an enormous challenge. The huge amount of data does not usually fulfil prerequisites of commonly used parametric statistics such as normal distribution and independence of variables. Variables are highly correlated, since groups of genes/proteins might be co-regulated and different NMR signals or MS peaks might represent one and the same metabolite/protein. Several strategies are being used to analyse this kind of highly multivariate data. Commonly used univariate statistical tests include analysis of variance, pair-wise comparisons (i.e. t-test), and determination of false discovery rates, usually performed using specialized software. Tools for multivariate analysis include cluster analysis, principal-component analysis, and self-organizing maps. These unsupervised methods are used to determine if gene-expression patterns allow the discrimination of natural subpopulations that might bear a biological meaning, such as treated/untreated or healthy/diseased. In addition to the unsupervised analyses performed for each experiment, the use of toxicogenomics also covers the field of predictive toxicology, relying on mathematical predictive models that generate so-called *signatures* or *fingerprints* that enable the classification of unknown compounds. For this purpose, in genomics and metabolomics, scientists have produced databases with model compounds known to cause a certain type of toxicity, to which the compounds under investigation are compared. For this type of application, complex supervised multivariate analysis methods (e.g. support vector machines, discriminant analysis, partial least-squares) are required. These supervised methods necessitate a so-called training set (results obtained from the model compounds) that is used for the generation of the predictive model. Once the model has been generated and cross-validated to ensure its appropriate performance, test compounds (new compounds under investigation) can be classified based on the gene-expression results or the metabolite

profiles. The efficiency of the models depends highly on the size of the database and the appropriate training set.

An additional point that needs to be addressed when evaluating complex data sets is their appropriate standardization. This applies to experimental conditions and bioinformatics and biostatistical tools alike. Although many significant results have been derived from 'omics' studies, the lack of standards limits the exchange and independent verification of the data across groups. Therefore, there is a need to develop standards that are acceptable within the scientific community. Again, most progress in this area can be reported from the genomics environment, where the *Minimum Information About a Microarray Experiment* (MIAME) has been defined and is widely accepted. Similar standardization efforts are under way in the metabolomic and proteomic fields.

4.5.5 Applications of Toxicogenomics

The main goals from the use of toxicogenomics are *predictive* and *mechanistic investigations*. The ideal system would be either an *in vitro* system, or a short term animal study. As with any *in vitro* assay, it is difficult to select a biologically suitable cell system that provides information relevant to the whole organism. Consequently, *in vivo* investigations are preferred for predictive toxicogenomic and metabolomic studies.

It is the task of predictive toxicogenomics that gene, protein and metabolite profiles allow to segregate compounds with a toxic liability from others. Mechanistic investigations aim at unveiling the molecular processes underlying an observed or suspected toxicity. Thus, the goal of mechanistic investigations is to understand the reasons for a given toxicity, identifying the affected cellular pathways. Both predictive toxicogenomics based on signatures and mechanistic approaches based on specific pathways can lead to the discovery of novel and specific *safety biomarkers*, an additional application of these technologies (Figure 4.16).

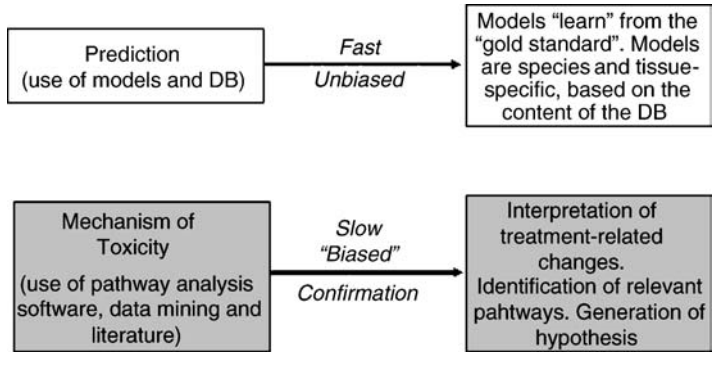


Figure 4.16 *Applications of Toxicogenomics for predictive and mechanistic investigations.*

Predictive Toxicogenomics

Predictive toxicogenomic studies usually compare the gene-expression patterns elicited by chemicals with unknown toxic potential to the profiles of model compounds with known toxicity with the goal of predicting organ toxicity.

The predictive approach is expected to yield faster, more accurate, and possibly more sensitive results than conventional toxicology assessment in the animal. Although *in vitro* test systems sound ideal to fulfil these purposes, advances in the field of *in vitro* toxicogenomics are still limited. This is due to various reasons. As with any *in vitro* assay, it is difficult to select a biologically suitable cell system that provides information relevant to the whole organism. Despite this, use of primary hepatocytes (mainly rodent, but also man), hepatic cell lines, and kidney cells have shown that prediction of toxicity using gene expression *in vitro* is feasible.

A substantial amount of gene-expression and metabolite-profiling data has been generated in animal models (mostly rat) with known toxicants, mainly hepatotoxicants and nephrotoxicants; but also with genotoxic, cardiotoxic, and testicular-toxic compounds. The results show that gene-expression and metabolite-composition analysis can provide information to allow classification of compounds according to their mechanism of toxicity as well as identifying cellular pathways related to the toxic event. In any case, for the use of toxicogenomics as a predictive tool, prior knowledge of gene-expression patterns related to toxicity is absolutely necessary. Consequently, this approach highly depends on the availability of a reference gene-expression database and robust software with appropriate algorithms for the comparison of complex fingerprints. Several commercial providers and industry have developed such databases. In addition, there are a number of consortia interested in generating similar datasets and some institutional efforts to create publicly available data.

Figure 4.17 represents an illustrative example of how predictive toxicogenomics can be used in practice. In this case, the hepatotoxicity database content was used to build a model with gene-expression data using Support Vector Machines (SVM), a supervised analysis tool. The generated model allows vehicle-treated animals (controls) and animals without histopathological findings (low-dose and nonresponders) to be discriminated from animals that had been exposed to a hepatotoxicant and had displayed liver necrosis (Figure 4.17A). Further, the use of the database in conjunction with a model including several classes of toxicity allows us also to distinguish several classes of hepatotoxicity, among others, direct acting (necrosis). In the selected example, the low-dose-treated animals were exposed to a subtoxic dose and could not be segregated from the controls, whereas the animals showing liver damage were assigned correctly to the direct-acting category. Nonresponder animals that did not display any toxicity as assessed by conventional endpoints were slightly different from the controls, suggesting a higher sensitivity of gene-expression analysis when compared with other toxicology endpoints (Figure 4.17B). In this validation study, phalloidin-treated animals could be assigned to the direct-acting category of compounds and liver necrosis was predicted based on gene-expression profiles. Furthermore, a clear distinction of doses, time, and extent of liver injury can be obtained using this type of analysis.

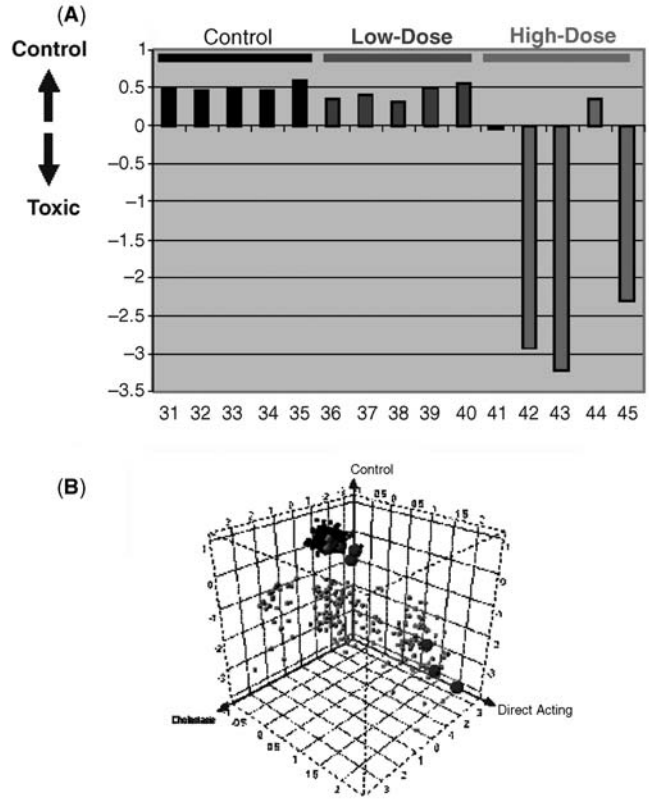


Figure 4.17 *Representation of classification of samples using models based on Support Vector Machines (SVM), as described by Steiner et al. (2004). Panel* ***A*** *represents the classification of animals treated with vehicle (control) or 2 doses of phalloidin. The three animals displaying liver necrosis (#42, #43, and #45) are classified as 'toxic' (negative values), while control, low-dose, and nonresponders are classified as controls (positive values). Panel* ***B*** *displays the classification of all animals in the database; black spheres represent vehicle-treated animals, and red spheres represent animals treated with hepatotoxicants. The blue spheres represent the animals treated with phalloidin. Note that the responder animals (dark blue) are correctly classified as direct acting (causing liver necrosis), while the low-dose animals are classified as controls (pale blue spheres). The two nonresponder animals (dark blue) show a slight difference in gene expression to the control group.*

Mechanistic Approach

Mechanistic analyses usually focus on genes, proteins, or metabolites of mechanistic interest and investigate the cellular and molecular mechanisms related to the exposure to the compound. This type of analysis concentrates either on changes or on pathways that are expected to be affected by the test compound.

The information obtained from mechanistic analyses is extremely useful to deepen the knowledge on the toxicity under investigation. However, it relies on the availability of accurate functional information on the affected genes, proteins, or metabolites. Moreover, it allows scientists to introduce knowledge-driven bias into the data, overemphasizing particular findings and disregarding perhaps important information. Hence it is vital that the generated mechanistic hypothesis be corroborated by appropriately designed follow-up experiments. These subsequent experiments aim to clarify two main sources of inaccuracy. On the one hand it needs to be elucidated whether the parameters of interest could be technical artefacts (false signals). On the other hand it should be assessed whether a change in the expression level of a gene or protein bears biological relevance and is actually causally associated with the toxic event under scrutiny. For the former, a technological validation needs to be performed, generally measuring the analyte(s) of interest with an alternative platform (such as qPCR or Western blot). This is particularly important for 'omics' applications, since, when analysing thousands of parameters, a small percentage of statistically expected false positives translate into a considerable number of data points that will wrongly be identified as significantly changed. For biological validation, the investigation of the hypothesized mechanism in another biological system, *in vivo* or *in vitro*, or the use of molecular biology tools such as gene knock-ins (transfection) or knock-outs (RNAi), will help the investigator realize if the pathway of interest is related to the studied phenomenon or not.

Mechanistic explanations based on 'omics' data have been reported with metabolomics, proteomics, and transcritpomics. For example, metabolomics analyses show that changes in intermediates of the citrate cycle are usually related to mitochondrial toxicity. In an investigative study, the investigators observed depletion in tricarboxylic acid cycle intermediates, and the appearance of medium-chain dicarboxylic acids. These findings led to the mechanistic hypothesis of defective metabolism of fatty acids in the mitochondria, which was confirmed by subsequent *in vitro* experiments. In addition, genomic data have often been used for the identification of a given molecular mechanism of toxicity. Fibrates, for example, are hypolipidemic drugs that rely on the activation of the nuclear receptor PPARα. Gene-expression profiling of livers of rats exposed to fenofibrate show the up-regulation of the fatty acid metabolic pathway, confirming a known pharmacological effect (Figure 4.18).

A recent publication showed the use of genomics data to study the differences between rodents and primates after exposure to several fibrates. These drugs cause hepatic peroxisome proliferation, hypertrophy, hyperplasia, and eventually hepatocarcinogenesis in rodents. However, primates are relatively refractory to these effects although the mechanisms for the species differences are not clearly understood. The authors used hepatic gene-expression profiles in the monkey and the results obtained led them to the following conclusions: On the one hand, genes related to the pharmacological effect of the compounds were strongly up-regulated, but to a lesser extent to that reported for rodents. On the other hand, a number of key regulatory genes known to be involved in proliferation, DNA-damage, oxidative stress, proliferation, and apoptosis were either unchanged or regulated in the opposite way to that in rodents. Thus, they concluded that the molecular mechanisms triggered in the liver of primates are qualitatively and quantitatively different from those in the rodent, which might be the reason for the species-specificity.

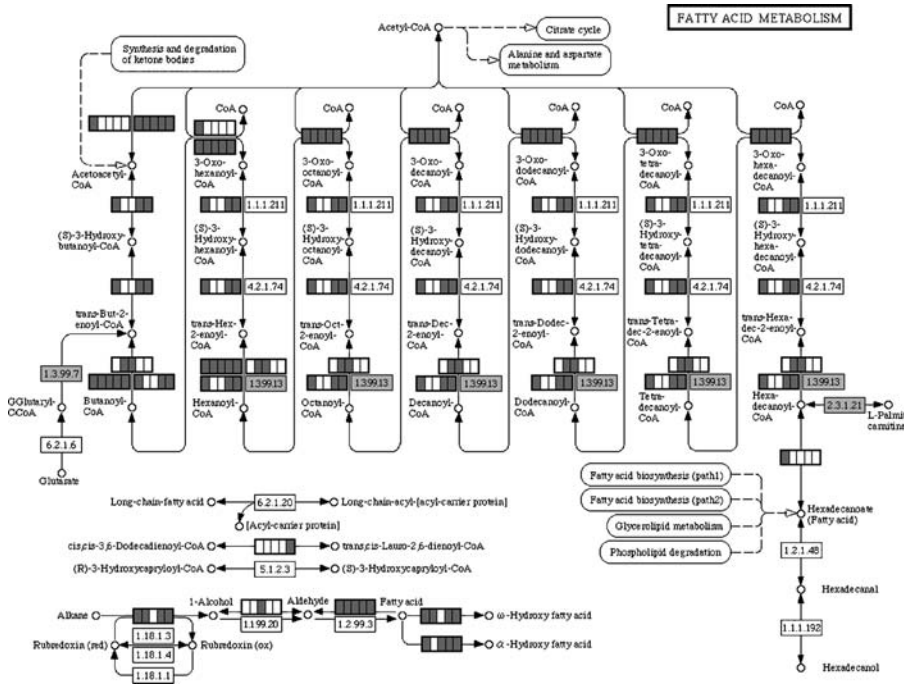


Figure 4.18 *[Reprinted from Cariello et al. Copyright (2005), with permission from Oxford University Press. Pathway mapping: fatty acid metabolism genes (KEGG pathway) up-regulated by ciprofibrate at different doses/times. A box representing a treatment condition is colored if any probeset reporting on the E.C. number was disregulated; the box is white if no disregulation occurred. Red indicates up-regulation.*

4.5.6 Summary

Toxicogenomics, as a new field in toxicology, needs to face many challenges and will doubtless undergo many changes before becoming a fully established discipline in toxicology. It has made major progress in terms of technology, data analysis, use, and general acceptance within the scientific community. However, there are currently still many ongoing discussions. First, scientists and managers are debating how to best use these technologies to get the most scientific and economic value from these experimentally complex and economically expensive approaches. Second, technology providers are investigating further in order to provide more robust, more high-throughput, and possibly less costly platforms to perform ever increasing amounts of 'omic' measurements. Last, but certainly not least, several discussion groups are debating the regulatory impact of toxicogenomics. It is clear that all these discussions will lead to new knowledge, new applications, and possibly new regulatory guidelines in the short to medium term. In addition, the results obtained from these different platforms are expected to lead to the identification of novel safety biomarkers, which in due time will need to be validated and might have a strong impact in how toxicology is performed. Hence the promise of toxicogenomics is well on the way to implementation and we should expect additional exciting news in the near future.

References

N. F., Cariello, E. H., Romach, H. M., Colton, H., Ni, L., Yoon, J. G., Falls, W., Casey, D., Creech, S. P., Anderson, G. R., Benavides, D. J., Hoivik, R., Brown, and R. T., Miller: Gene Expression Profiling of the PPAR-alpha Agonist Ciprofibrate in the Cynomolgus Monkey Liver. *Toxicol. Sci*., 2005, **88**, 250–264.

G., Steiner, L., Suter, F., Boess, R., Gasser, M. C., de Vera, S., Albertini & S., Ruepp: Discriminating different classes of toxicants by transcript profiling. *Environ. Health Perspect*., 2004, 112, 1236–1248.

For Further Reading

W. B. Dunn, N. J. Bailey, and H. E. Johnson: Measuring the Metabolome: Current Analytical Technologies. *Analyst* (*Cambridge, UK*) 2005, **130**, 606–625.

L. Eriksson, H. Antti, J. Gottfries, E. Holmes, E. Johansson, F. Lindgren, I. Long, T. Lundstedt, J. Trygg, and S. Wold: Using Chemometrics for Navigating in the Large Data Sets of Genomics, Proteomics, and Metabonomics (gpm). *Anal. Bioanal. Chem*., 2004, **380**, 419–429.

A. Luhe, L. Suter, S. Ruepp, T. Singer, T. Weiser, and S. Albertini: Toxicogenomics in the Pharmaceutical Industry: Hollow Promises or Real Benefit? *Mutat. Res*., 2005, **575**, 102–15.

R. J. Mortishire-Smith, G. L. Skiles, J. W. Lawrence, S. Spence, A. W. Nicholls, B. A. Johnson, and J. K. Nicholson: Use of Metabonomics to Identify Impaired Fatty Acid Metabolism as the Mechanism of a Drug-Induced Toxicity. *Chem. Res. Toxicol*., 2004, **17**, 165–173.

J. K. Nicholson, J. Connelly, J. C. Lindon, and E. Holmes: Metabonomics: a Platform for Studying Drug Toxicity and Gene Function. *Nat. Rev. Drug Discov*., 2002, **1**, 153–161.

B. A. Wetmore, and B. A. Merrick: Toxicoproteomics: Proteomics Applied to Toxicology and Pathology. *Toxicol. Pathol*., 2004, **32**, 619–642.

4.6 Introduction to the Statistical Analysis of Experimental Data

György Csanády

4.6.1 Introduction

Measurements carried out under the very same experimental conditions result in observations which are not exactly identical due to random variability. This variability might easily mislead the data interpretation guided solely by common sense. Applied statistical methods are useful to describe and analyse the random variability and thus can help researchers to reach a scientifically and statistically valid conclusion about the collected data.

Statistical procedures should be applied to the planning phase of the experiments, to summarise and analyse the data, and to draw inferences from the data.

This chapter is intended to assist scientists working in experimental settings by discussing:

- Summarizing data by using **descriptive statistics**;
- How uncertainty can be characterized by **error propagation** if variability of collected data is propagating in subsequent calculations;

- How **probability distributions** can be used to describe variability;
- How data collected from a sample of a population can be used to generate a **statistical inference** about the population using **estimation procedures** and **hypothesis tests**;
- The use of **regression analysis** to describe a linear relationship between two variables;
- Determination of the median lethal dose using **probit analysis**;
- Experimental design;
- Use of computers in statistical calculations.

In any toxicological experiment there are several variables and factors such as species, strain, sex and age of the animals, season, diet, dose, time, experience of the experimenter, etc., which may affect the outcome. Interpretation of experimental results is facilitated when all but one of the factors is held constant and the effect of a single **independent variable** is studied. In toxicology the dose is among the most frequently employed independent variable. Thus, in an experimental situation the effect of varying the dose leads to observations on the **dependent variable** collected in the form of data. (A single piece of information would be termed a datum.) Data can be either **qualitative**, e.g. tumor or inflammation present, or **quantitative**, e.g. body weight, concentration of a metabolite in blood.

If the second experiment is carried out under the same experimental conditions the outcome will not be exactly identical to that of the first. The difference between the two outcomes is governed in a well conducted experiment only by chance. Variability is the reason why in toxicological studies it is essential to make comparisons between 'treated' and 'control' groups, for which identical untreated animals are used. Some of the variability can be attributed to known factors. However, the major share of the variability is considered unpredictable and is called **random variability** composed of several factors, such as biological variability, sampling variation, and experimental error. This variability is the object of statistics. Statistical procedures can help to describe and characterise this variability in order to facilitate an accurate interpretation of experimental results.

The entirety of all possible outcomes of a given experiment conducted under identical conditions is called the **population**. The population - a theoretical concept - might refer to an infinite number of experiments, which could never be carried out experimentally. The outcome of a set of measurements results in data, which from a statistical point of view correspond to a sample independently drawn by chance (Figure 4.19). It is assumed that the sample is large enough to be representative of the population. After collection of the data, statistical procedures are applied with the aim of drawing an inference, i.e. a tentative conclusion, about the population itself. It is somewhat surprising that limited data can lead to any conclusion about a large, often infinite, population. The reason for this peculiarity lies in the fact that underlying assumptions govern the behavior of the population. Such assumptions must be known in order to perform a valid statistical analysis. For many statistical procedures it is assumed that the variability present in the observed data follows **Gaussian** or **normal distribution** (see below). However, biological data show often another distribution called **lognormal**.

Example

In a carcinogenicity study Sprague-Dawley rats are treated with a chemical. It is impossible to study all existing Sprague-Dawley rats that make up the theoretical rat

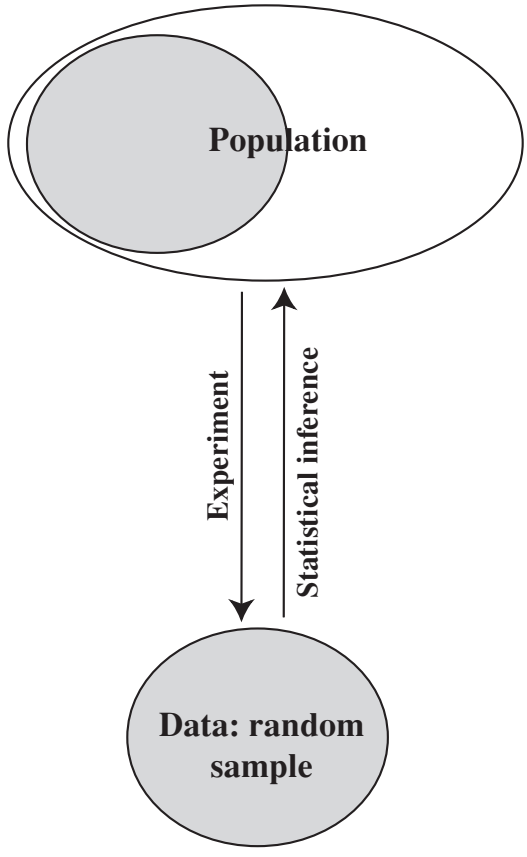


Figure 4.19 *Relationship between population and sample. The sample, consisting of a set of observations embodied in the data obtained in an experiment, is considered to be a random realization of the statistical population. The sample is analysed with the aim of drawing statistical inferences about the population itself.*

population. Instead, a given number, usually 50 rats, which comprise the sample are treated per dose group for about two years and the incidence of tumors is observed. The underlying assumption is that the finite sample investigated represents the infinite population (Figure 4.19). We are not interested in the fate of any particular rat but in the carcinogenicity of the tested chemical in the rat population. For this purpose the tumor incidence obtained in both a treated and a control group are compared.

Comment

Bias is a term used to describe **systematic errors** resulting in inaccuracies which lead to misinterpretation of the results. For example, use of a nonrandom sample of the population can lead to a constant deviation, which cannot be revealed by statistical analysis per se. **Systematic errors** can only be identified by using different experimental methods or by participating in **interlaboratory tests. Randomization** might help to minimise bias (see Section 4.6.8, Experimental Designs).

4.6.2 Descriptive Statistics

Large data sets or even a single set of similar numbers cannot be easily comprehended. Therefore, they should be organised and summarised in such a way that the main characteristics of the data are preserved. Descriptive statistics provide methods to present and summarise measured data by calculating specific sample characteristics determined solely by the type of the data (Table 4.12).

Data collected in an experiment are either **qualitative** (**categorical**) or **quantitative** (**numerical**). Many toxicological observations are qualitative and can be ordered into dichotomous categories, e.g. mortality data, fetal abnormalities, histopathology data, and clinical signs. There are also classifications requiring more than two categories, e.g. blood groups. These characteristics are given in the **nominal scale**, which does not allow the establishment of any hierarchical order. Other classifications, e.g. severity of a lesion (+, ++, +++) are established by a hierarchical order according to a predefined **ordinal scale**. The magnitude of the effect determines the ranking, although differences between ranks might not have an arithmetical meaning. Other experiments result in **quantitative data**, which can be either discrete or continuous. **Discrete data** can represent numerical values expressed as an integer, since they represent counts of events such as radioactivity measurements, some haematology data (cell counts), number of tumors, or size of litter. **Continuous data** are characterized by continuous quantities. The readings may have any values between certain limits. Such data include, for example, body and organ weights, food and water consumption, and concentrations of toxic agents in the blood.

For continuous variables full precision is retained during data analysis, except for final rounding. The number of significant digits presented should correspond to the precision of the measurement. For example, a single value of 9 usually implies that the true number is between 8.5 and 9.5. Therefore, the precision of this value is 11% if it was measured to the

Table 4.12 *The type of data collected determines the type of graphical representation as well as the sample characteristics.*

	Qualitative		Quantitative	
	Nominal scale	Ordinal scale	Discrete	Continuous
Examples	Gender, blood groups	Severity of lesions (+, ++, and +++)	Some haematology data, SCE, size of litter	Body and organ weight, concentration
Graphical representation	Bar/pie diagram	Bar/pie diagram	Histogram, box-and-whisker plot	Scatter plot, box-and-whisker plot
Sample characteristics	Frequency, mode	Frequency, mode, median	Mean/median, standard deviation/centiles	Mean/median, standard deviation/centiles

Mode: most frequent value. Median: the mid-point in an ordered series of numbers. SCE: sister chromatid exchange.

Table 4.13 *Number of significant digits of a datum in relation to its accuracy.*

Datum	Number of significant digits	Limit	Accuracy (%)
1	1	0.5–1.5	100 (100*1/1)
9	1	8.5–9.5	11 (100*1/9)
1.1	2	1.05–1.15	9.1 (100*0.1/1.1)
9.9	2	9.85–9.95	1.1 (100*0.1/9.9)
1.11	3	1.105–1.115	0.90 (100*0.01/1.11)
9.99	3	9.985–9.995	0.10 (0100*.01/9.99)

nearest digit (Table 4.13). To present results from most toxicological studies two significant digits are sufficient since the accuracy of the measured data is generally not below 1%. Statistical descriptors, such as mean values, can be given by one more significant figure.

Raw data are often represented in figures and tables. Information from a large body of data can be condensed by calculating their characteristic measures. Data can be summarized by reporting a measure for the **central tendency**, for the **variability**, together with the number of observations. There are several measures for the central tendency:

- The **arithmetic mean** or average, equals the sum of observations (X_i) divided by the number of observations (N) [Equation (4.20)]:

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{4.20}$$

- The **geometric mean** is the Nth root of the product of all observations [Equation (4.21)]:

$$\text{Geometric mean} = \sqrt[N]{\prod_{i=1}^{N} X_i)} \tag{4.21}$$

- The **median** divides the ordered data set into two equal groups.
- The **mode** is the most common value.

The most useful measure for the central tendency depends on the type of data. The mode describes nominal scale data, giving the most prevalent sample characteristics. For ordinal scale data the median is often considered to be useful. Mean values are appropriate for quantitative data. Outliers within the data confound the mean whereas the median and mode are less sensitive. Arithmetic mean, median, and mode are very similar (theoretically indistinguishable) if the data were collected from a symmetrical distribution. The variability or spread observed in the data can also be described by different measures termed the **range**, the **centiles**, the **standard deviation**, and the **coefficient of variation**.

The **range** is defined by the difference between the largest and the smallest value. It is very sensitive to outliers since it takes account of only the two most extreme values and does not reflect the majority of data. The **centile** or **percentile** gives the value below which a given percentage of the values lie. The data must be arranged in an order, e.g. $x_1 \leq x_2 \ldots . \leq x_n$, before centiles can be calculated. The **median** is defined as the 50th

centile because 50% of the values are below and 50% are above this value. The difference between the 25th and 75th centiles (**quartiles**) gives the **interquartile range**, which is also a measure for the spread in the data.

The **standard deviation** [Equation (4.22)]

$$\left(\mathrm{SD} = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N-1}} \right) \tag{4.22}$$

is related to the average distance of the data from their mean value. The SD is especially useful to help characterize the variation in the data when the underlying distribution is known. For example, the majority (68%) of the data lie within one standard deviation of the mean ($\bar{X} \pm \mathrm{SD}$) if the data follow normal distribution (see below). The share of data increases to 95% if the range is extended to about twice of the standard deviation ($\bar{X} \pm 1.96$ SD). The precise doubled range of standard deviation around the mean covers exactly 95.5% of the data. However, if the distribution is unknown then at least 75% of the data will be in the range of $\bar{X} \pm 2$ SD. The **coefficient of variation** [Equation (4.23)] is a measure of the relative standard deviation and is widely used to compare the variability among different samples and for estimating error propagation (see below). For example, the CV of chemical analysis is usually below 0.05 whereas that of biological experiments may often exceed 0.4.

$$\mathrm{CV} = \mathrm{SD}/\bar{X} \tag{4.23}$$

The **box-and-whisker plot** can be helpful to demonstrate the distribution of a large data set by showing the median and the quartiles of data in a box plot together with the extreme values indicated by the whiskers (Figure 4.20). The median is centred between the upper and lower quartiles if the data come from a symmetrical distribution. Asymmetric portions between the median and the quartiles as well as between the quartiles and the whiskers are indicative of a skewed distribution.

The majority of the toxicological data follow a **unimodal** distribution, having a single peak with tails on either side. The most common unimodal distribution is symmetrical, with equal tails on either side, and can be depicted by a symmetrical box-and-whisker plot corresponding to a bell-shaped curve (see below). A given sample from such a distribution is summarised by its mean and standard deviation.

A **skewed** or asymmetric distribution is characterized by asymmetric tails. Such a distribution is often encountered if the data are positive and the twofold standard deviation is larger than the corresponding mean value. In such a case the median or the geometric mean and the range between the 5th and 95th centiles can be used as measures for central tendency and variation.

It is also possible that data arise from distributions having more than one peak. For example a **bimodal** distribution might be observed if the data show a corresponding cluster, e.g. they include different sexes or age groups. In such cases, appropriate categorisation of the data could result in unimodal distributions.

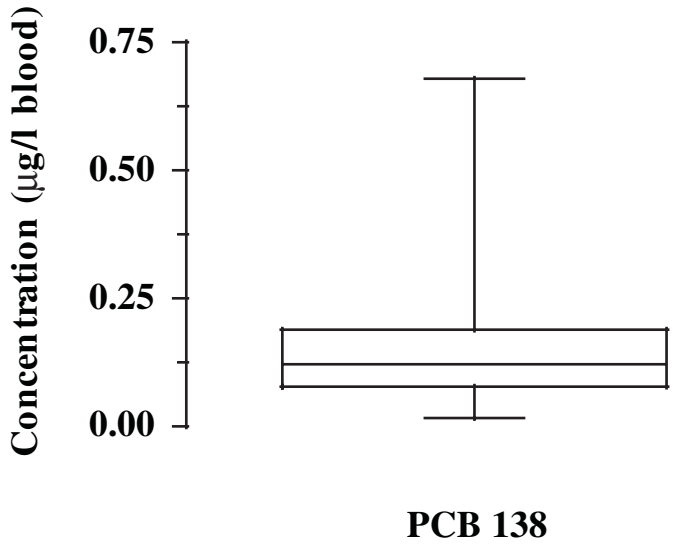


Figure 4.20 *Box-and-whisker plot of a polychlorinated biphenyl (PCB138) concentration measured in blood of children visiting a contaminated school (age: 6–16 y; single determinations in N = 218 individuals).*

Example

The polychlorinated biphenyl concentrations plotted in Figure 4.20 show a skewed distribution since the median is closer to the lower than to the upper quartile. Furthermore, the range between the lower quartile and lowest values is several times smaller than the range of the upper quartile and the largest data point. Consequently, the data are summarised using the median for central tendency, i.e. 0.122 µmol/l. The variability is given by 5th and 95th centiles, corresponding to the values of 0.056 and 0.357 µmol/l.

Comment

Before applying any statistical methods to the raw data they should be carefully checked for plausibility and consistency since any error in the data, e.g. due to the measurement itself or data transfer, may have serious implications for the outcome of most statistical procedures.

4.6.3 Error Propagation

Continuous raw data measured in an experiment are characterised by their mean and standard deviation. In many cases such mean values are used in further calculations. Since the mean values incorporated in the calculation are affected with uncertainty represented by their standard deviations the result will also exhibit a definite uncertainty.

To calculate this particular uncertainty in the result the law of propagation of error is used. If means of two independent variables (X, Y) are added or subtracted, Equation (4.24), then the squares of the resulting standard deviation equals the sum of the squares of the two standard deviations, Equation (4.25). If the two means are multiplied or divided, Equation (4.26), then the squares of coefficient of variations are added, Equation (4.27). Thereafter, the standard deviation of Z can be calculated using the value of CV_z^2. For more complicated cases formulas can be derived or found in the literature.

$$Z = \bar{X} + \bar{Y}, \quad \text{or } Z = \bar{X} - \bar{Y} \tag{4.24}$$

$$SD_Z^2 = SD_X^2 + SD_Y^2 \tag{4.25}$$

$$Z = \bar{X} * \bar{Y}, \quad \text{or } Z = \bar{X}/\bar{Y} \tag{4.26}$$

$$CV_Z^2 = CV_X^2 + CV_Y^2 \tag{4.27}$$

Comment

To use the above formulas it is assumed that the standard deviations are small with respect to their means ($< 20\%$) and the two data sets are uncorrelated. By subtracting two numbers of the same magnitude the result could be statistically insignificant (see below) due to an increase in the standard deviation of the result.

4.6.4 Probability Distribution

A probability distribution is a mathematical function describing the probability associated with all possible values of the random variable within the population.

Large data sets with more than 50 observations ($N > 50$), can also be plotted by means of a **frequency histogram** depicting the frequency of observations in ordered classes. To obtain an acceptable visualisation, the x-axis is divided into about 7–20 **classes** covering the full range of the data. The number of classes can be approximated as the square root of the number of observations if $N < 500$. Observations falling within a given class are grouped together. In the **relative frequency histogram** (Figure 4.21A) a bar represents the relative frequency of the observations falling within a given class, i.e. the number of observations within the interval divided by the total number of observations. Therefore, the total area of all bars equals one.

The **cumulative frequency histogram** is obtained by adding the bars of the frequency histogram together cumulatively. Each class contains the number of observations that fall within the class plus all that fall below it. The **relative cumulative frequency histogram** is obtained if the class values are divided by the total number of observations (Figure 4.21B). Therefore, this histogram approaches the value of one at its right boundary. The histogram would become less dispersed, i.e. the edges would disappear, and a continuous smooth function would appear as the number of data points approach infinity and the class width nears zero. These smooth functions depicted by the solid lines are referred to as the **probability density function** (A) and **probability distribution function** (B). The latter is also called **cumulative distribution function**.

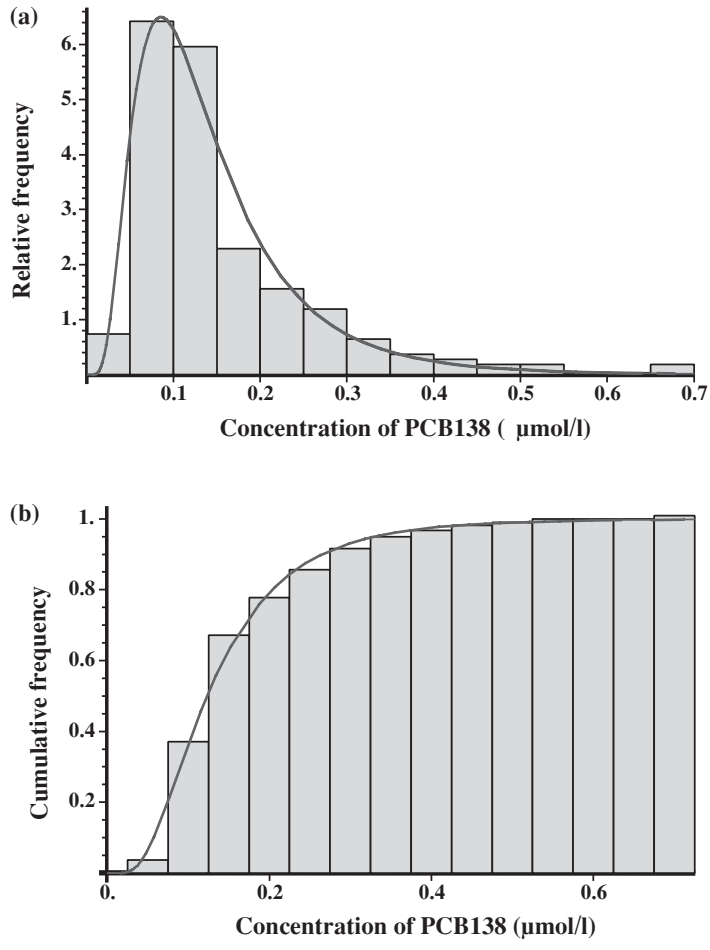


Figure 4.21 *Relative frequency histogram (A) and cumulative relative frequency histogram (B) of the concentration of a polychlorinated biphenyl (PCB138) measured in blood of children attending a contaminated school (age: 6–16 y; single determinations in N = 218 individuals). The solid grey lines correspond to the probability density function (A) and cumulative distribution function (B).*

In general, the probability density function is specified by a mathematical equation, and the cumulative distribution function can be calculated by integrating the density function. The area under a probability density function equals one as does the area under the relative frequency histogram.

Example

The relative frequency histogram of the PCB138 concentration (Figure 4.21A) shows that the data are not symmetrically distributed. Moving from the mode, which displays a maximum of about 0.09 µmol/l, to the left side, values rapidly decrease to zero whereas

on the right side the decrease takes place more slowly and values up to 0.7 µmol/l are reached. The majority of the data points (70%) are located on the right side of the mode. Such a distribution is called 'skewed right', displaying a longer tail on the right side.

The **normal** or **Gaussian distribution** plays a central role in statistics. The Gaussian distribution is often assumed to describe the random variation in the data especially when the variability resulting from several random sources acts additively. Such probability density functions [Equation (4.28)]

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{4.28}$$

are shown in Figure 4.22. The two parameters identified by the Greek letters μ and σ^2 describe the location of the curve on the x axis (**central tendency**, **expected value**) and the spread of the curve (**variance**), respectively. There are two curves with the same expected value of 10 (Figure 4.22A) located at the same place on the x-axis. One of these two curves has a fourfold larger variance, resulting in a flattened, wide shape. All possible normal distributions can be transformed [Equation (4.29)] to a single function located at zero ($z = 0$) and having a unit variance ($\sigma = 1$). This function is called the **standard normal distribution** [Equation (4.30)] and is depicted in Figure 4.22B.

$$z = \frac{x - \mu}{\sigma} \tag{4.29}$$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{4.30}$$

The area under the distribution curve between any two abscissae gives the probability that the x value falls between the specified limits. Tabulated values of the area under the standard normal distribution can be found in many statistical tables.

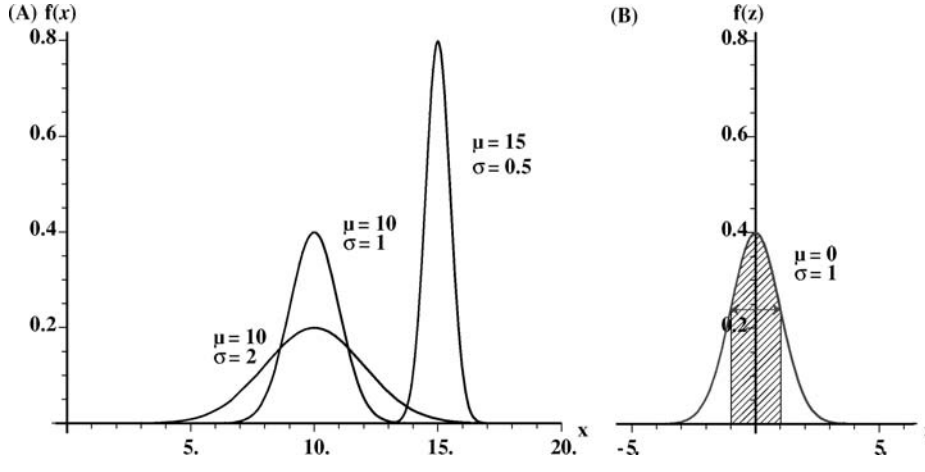


Figure 4.22 *Probability density functions of normal distributions (A) with different parameters (μ, σ). The standard normal distribution is depicted in Figure 4.22B. The shaded area has a value of 0.68, representing the probability that x falls within one standard deviation of the mean.*

Comment

The expected value of a distribution, referred to as $E(x)$, indicates its average or central value and is calculated as a probability-weighted average value [Equation (4.31)]:

$$E(x) = \int_{-\infty}^{\infty} x * f(x)dx \tag{4.31}$$

Applying this definition to the Gaussian distribution results in $E(x) = \mu$. Please note the analogy between the arithmetic mean and the expected value. It can be shown that the arithmetic mean of a sample approximates the population mean. Therefore, the sample mean value is an estimator of the unknown population mean, if the sample is large enough, sufficiently representative and unbiased.

The variance of a distribution, given as $V(x)$, indicates the spread of the curve. It is calculated as the expected value of the square of the deviation of X from its own expected value [Equation (4.32)]:

$$V(x) = E\{[X - E(X)]^2\} = \int_{-\infty}^{\infty} [x - E(X)]^2 * f(x)dx \tag{4.32}$$

The probability density function is the weighting function similar to the expected value. Applying this definition to the Gaussian distribution results in $V(x) = \sigma^2$. Owing to the analogy between the standard deviation and σ, it can be shown that the square of the standard deviation of a sample approximates the population variance.

Many measurements show a 'skewed right' distribution as exemplified for the **lognormal distribution** in Figure 4.21. Such a distribution becomes apparent if the variation results from multiplicative effects. Biological mechanisms are known for inducing lognormal distributions. Many biological variables such as alveolar ventilation or blood flow through tissue, concentrations of several exo- and endogenous substances in biological material, and latency periods of diseases follow such a distribution. Skewed data such as the blood concentrations of polychlorinated biphenyls can often be transformed into normal distribution just by using the logarithm of the data or by plotting them on a logarithmic scale. Consequently, lognormal distributions can be described by the log-transformed variable, resulting in a normal distribution, which is characterised by its expected value and variation expressed on a log scale. The skewed distribution (Figure 4.21A) becomes a symmetrical bell-shaped curve (Figure 4.23A) following the logarithmic transformation by stretching out the lower end and compressing the upper end of the original distribution. Statistical analysis can be carried out on the log-transformed data before converting the results back into the original scale.

Example

On a natural logarithmic scale, the mean value and standard deviation of the log PCB138 data amount to -2.072 and 0.621, respectively. By taking the exponent of these, values of 0.126 and 1.86 are obtained. Thereafter the range covering 95.5% of the data is calculated on the log scale, corresponding to -3.31 $(-2.072 - 2 * 0.621)$ and -0.83 $(-2.072 + 2 * 0.621)$. These values represent the PCB138 concentrations of 0.036

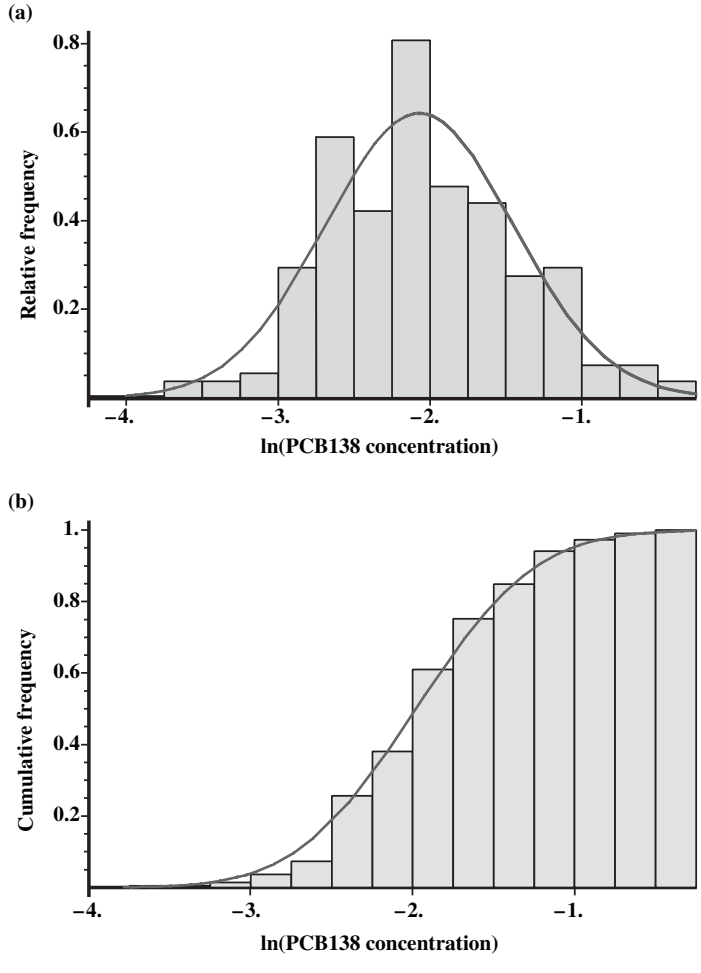


Figure 4.23 *The relative frequency histogram (A) and cumulative relative frequency histogram (B) obtained after a logarithmic transformation of the measured concentrations of a polychlorinated biphenyl PCB138 (see Figure 4.21). The grey lines correspond to the probability density function (A) and cumulative distribution function (B).*

[$0.126/(1.86^2)$ or $e^{-3.31}$] and 0.436 [$0.126*(1.86^2)$ or $e^{-0.83}$]. The interval is calculated in a multiplicative manner ($\bar{X}_{\log}/SD^2_{\log}$ and $\bar{X}^*_{\log}\,SD^2_{\log}$) in contrast to the additive manner demonstrated for the normal distribution ($\bar{X} - 2^*\,SD$ and $\bar{X} + 2^*\,SD$). The index log indicates that both the mean and the standard deviation were obtained using the log-transformed data that followed the 'back-transformation' [Equation (4.33)]. Note that the mean of the log-transformed data ($\bar{X}_{\log}$) corresponds to the geometric mean on the original scale.

$$\bar{X}_{\log} = \mathrm{Exp}\left(\frac{\sum \ln(X_i)}{N}\right) \tag{4.33}$$

A visual approach to evaluate whether a given data set follows a normal distribution is based on plotting the data on normal probability paper. This graph paper is the equivalent of 'logarithmic paper' on which exponential functions are depicted by straight lines.

On the **normal probability paper**, a normal distributions yields a straight line since the ordinate is scaled in percentages according to the inverse normal distribution. It covers the ranges 1–99% or 0.1–99.9%, but never includes the adjacent values 0 or 100%. Normally distributed data will appear close to a straight line in the probability plot where the percentiles of 16, 50, and 84% correspond to the abscissae of $\bar{X} - \text{SD}$, $\bar{X}$, and $\bar{X} + \text{SD}$. A departure of the data points from the theoretical line in the middle region indicates a deviation from normality, especially when the departure follows a given pattern, e.g. a hump, or a divergence. In contrast, a deviation of the data from the line towards its ends is less important since it can result from outliers or sampling effects. The PCB138 data demonstrate a significant deviation from the theoretical line in the probability plot (Figure 4.24A). However, after plotting the data on a logarithmically scaled *x*-axis they show a nearly normal distribution, which is fitted reasonably well by a straight line (Figure 4.24B). Probability paper can be used to test various distributions such as **lognormal**-, **exponential**- or **Weibull-distribution**. They are available commercially or can be downloaded (www.weibull.com/GPaper).

There are statistical tests to investigate formally whether a random sample stems from the normal distribution. These include the **Shapiro–Wilk test** if $N < 50$, the **D'Agostino–Pearson normality test** if $N =< 50$, and the **Kolmogorov–Smirnov test**. The first two tests are considered to be more robust and have more power than the Kolmogorov–Smirnov test, although any of these tests might fail if the sample size is too small ($N < 10$).

In statistical theory, the most important distribution is the normal distribution. There are other relevant distributions that can be applied to different fields of toxicological research. For example, the **Binomial distribution** can be useful to analyse results obtained by flow cytometry whereas the **Poisson distribution** can be applied to mutagenicity assays. Time-to-tumor data and survival data can be interpreted using the **Weibull distribution**.

For statistical analysis the **t-distribution**, which compares means, **F-distribution**, which compares variances, and **Chi** (χ^2) **squared distribution**, which compares distributions, are of importance. The areas under their probability density functions can be found in statistical books or tables.

Comment

If the random variable is discrete then the probability distribution is also discrete. For example, the lung cancer deaths in a sample of a given population depict a discrete distribution.

4.6.5 Inferential Statistics

The major challenge for many toxicologists is to draw reliable conclusions on the basis of measured or published data. Statistical inference can provide a major tool to support the assessment process, which relies on objective procedures.

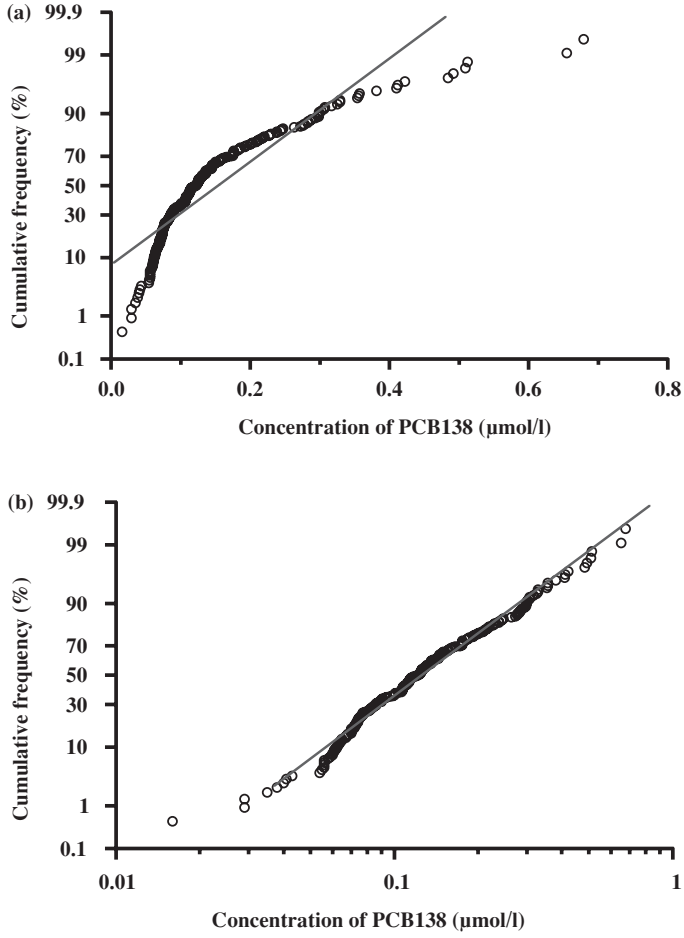


Figure 4.24 *Cumulative percentages of the PCB138 data, taken from Figure 4.21, depicted on normal (A) and lognormal (B) probability paper.*

Inferential statistics generalise about the population on the basis of a collected sample. There are two main methods used for this purpose: estimation and hypothesis testing. Estimation procedures are used to estimate a given parameter of a distribution such as point estimates for expected value or variance, based on the sample and to construct its confidence interval. Hypothesis testing is used when a comparison has to be made.

Estimation

The aim of estimation is to calculate an unknown population parameter such as the expected value of the underlying distribution, using the mean value

calculated from collected samples (a point estimate). Thereafter, around this single value a confidence interval is constructed enclosing the value of the population parameter with a predefined confidence level.

The mean value of a given sample (X) represents a single value known as a **point estimate** for the expected value of the normally distributed population (μ) since by increasing the size of the sample the sample mean approximates the expected value. By drawing repeated samples from the same population the calculated means of the samples differ from each other and show a scatter around the true population mean. Its true value remains unknown unless an exhaustive sampling is carried out, which is impossible in most cases. Therefore, it is necessary to provide the point estimate with a measure of confidence. This measure can be derived since the distribution of the sample mean is normal if the population from which the data were taken underlies a normal distribution. The variation of the random variable corresponding to the sample mean is characterised by the **standard error of the mean** (SE), which describes the expected extent of the variation among the means of different samples drawn from the same population. Its value is calculated by dividing σ by $\sqrt{N}$, where σ^2 is the variance of the population from which the sample including N measured data was drawn. The standard error of the mean is a measure of the uncertainty of a single sample mean related to the population mean. For example, the mean calculated from 68% of the samples will lie within the interval of $\bar{X} - \text{SE}$ and $\bar{X} + \text{SE}$ where $\bar{X} - \text{SE}$ is called the lower limit (L_L) and $\bar{X} + \text{SE}$ the upper limit (L_U). By increasing the probability from 68% to 95% the interval becomes less precise and almost two-times ($\bar{X} \pm 1.96\text{SE}$) wider. Such a symmetric range defined by its two limits (L_L and L_U) is called the **confidence interval** (CI) in which the population mean can be found with a specified level of confidence. The corresponding probability is called the **confidence level** ($1 - \alpha$) and is associated with the complementary probability called the **significance level** (α). The significance level is assigned before the experiments are carried out, and has the usual values of 0.05 or 0.01. The multiplier of the standard error can be obtained by looking up in a statistical table the $100(1 - \alpha/2)$ quantile of the standard normal distribution if σ is known. In many cases the squared standard deviation of the sample is used as a surrogate for the unknown variance of the population and SE is obtained as the ratio of SD and $\sqrt{N}$. For such samples, the critical values for calculating the confidence interval are taken from the **t-distribution** [$100(1 - \alpha/2)$ quantile] at the specified level of significance α. The t-distribution is also a symmetrical distribution, having one single parameter called degree of freedom. For estimating the confidence level for the mean by taking a sample with a size of N, the degree of freedom equals $N - 1$. The t-distribution approaches the normal distribution if the sample size becomes large enough. The deviation from the normal distribution becomes smaller than 2.5% if the sample size is larger than 50 ($N > 50$).

Example

Ten five-week-old mice were purchased with the following body weights: 22.3, 21.8, 22.7, 21.6, 19.8, 20.4, 22.6, 21.9, 20.8, 22.4 g. The mean value and the standard deviations of this sample are $\bar{X} = 21.6$ and $\text{SD} = 0.988$, respectively. The 95% confidence interval of the mean can be calculated to be 21.6 ± 0.706 ($21.6 \pm 2.26^{*}\ 0.988/\sqrt{10}$) by selecting the

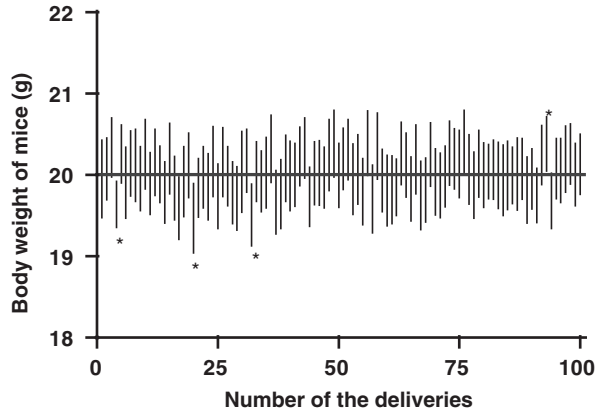


Figure 4.25 *The confidence intervals of the body weight calculated for 100 different deliveries containing 100 mice each of similar age by selecting a 0.95 confidence level (grey vertical lines). The solid black line shows the theoretical population mean of 20 g. In fact, there are 4 deliveries (marked with an asterisk) for which the calculated confidence interval does not include the population mean. The confidence intervals calculated for the major share of deliveries (96 %) enclose the population mean.*

critical *t*-value of 2.26 that can be taken from a statistical table given for the *t*-distribution for the degree of freedom = 9 and $\alpha = 0.05$ (two-sided test) or $\alpha = 0.025$ (one-sided test). Please note that the critical *t*-value is only slightly larger than the corresponding value obtained from the standard normal distribution (1.96). This deviation becomes more obvious if the degree of freedom is smaller.

The probabilistic concept of confidence interval is demonstrated in Figure 4.25, indicating that the confidence interval calculated from a single sample at a significance level of 0.05 is likely to contain the true population mean in 95% of cases.

The mean ($\bar{X}$) of a normally distributed sample represents only a single value estimate or a point estimate for μ whereas the confidence interval provides an interval estimate by quantifying the degree of confidence that the population mean lies within the calculated confidence interval. Therefore, a confidence interval is considered to be superior compared with a single point estimate.

Comments

About the distribution of the sample mean, interesting findings can be derived from the central limit theorem: If the sample size is large enough ($N > 50$) then the mean of repeatedly obtained samples (sampling distribution) is nearly normally distributed and independent of the distribution of the sample.

Confidence intervals can be constructed for other population parameters, e.g., the median, the variance, or slopes obtained by linear regression. The confidence interval is a concept similar to the reference interval (95%) used in medicine, usually defining the difference between the upper (97.5 centile) and the lower limit (2.5 centile) of a test value when measured in a population of apparently healthy individuals. The International

Federation of Clinical Chemistry and Laboratory Medicine recommends reference interval measurements in a least 120 individuals.

Upper limits (L_U) rather than estimates of expected values or confidence intervals are used in quantitative cancer-risk assessment.

Hypothesis Testing (Significance Test)

Knowledge is often generated by comparing treated and control groups. To draw an objective decision is possible only with the aid of statistical hypothesis testing. For this purpose, the probability is calculated that the difference observed between the two groups can be explained by chance alone. A distinct characteristic of statistical decision making is that uncertainty cannot be eliminated since decisions must be taken on the basis of limited samples.

Many toxicological studies are carried out to test potential adverse effects of a given chemical by formulating a specific research hypothesis, e.g. styrene is a carcinogen. For this purpose the outcome of a treatment-related effect, i.e. tumor incidence, is measured and a comparison between two groups called 'treated' and 'control' is made. The only way to confirm a research hypothesis is to reject its negated form. Therefore, the negated research hypothesis is formulated as the statistical **null hypothesis** (H_0) stating in general that the effect variable between the two groups does not differ, or in other words that the treatment has no effect. In contrast, the **alternative hypothesis** (H_1) describes the situation if the null hypothesis is invalid corresponding to the research hypothesis.

One of the most commonly used tests is the **t-test** which can be used in two different situations. Using this test either a known population mean can be compared with a sample mean or means of two different samples can be compared, provided that the sample(s) are taken from normally distributed populations. The use of *t*-test will be demonstrated on the following example:

Example A sample of ten five-week-old mice was obtained. The mean value and the standard deviation of the body weight are $\bar{X} = 21.63$ and SD = 0.988, respectively. The historical records of the breeding laboratory indicate that the mean body weight of such mice in the past has been 22 g. Is the weight of the animals in the delivered sample different from that of the historical controls?

The null hypothesis is that the ten mice have the same mean body weight as the historical population (H_0: $\bar{X} = \mu$). The alternative hypothesis is that they are different (H_1: $\bar{X} > \mu$ and $\bar{X} < \mu$).

By performing the t-test a probability (p) is calculated under the condition that the null hypothesis is true. In the above case the test statistics are given by the difference between the mean weights of the 10 mice (sample mean) and the weights of the historical population (μ) in relation to the standard error [Equation (4.34)]:

$$\left(t = \frac{|\text{sample mean} - \mu|}{\text{standard error of observed mean}} \right) \tag{4.34}$$

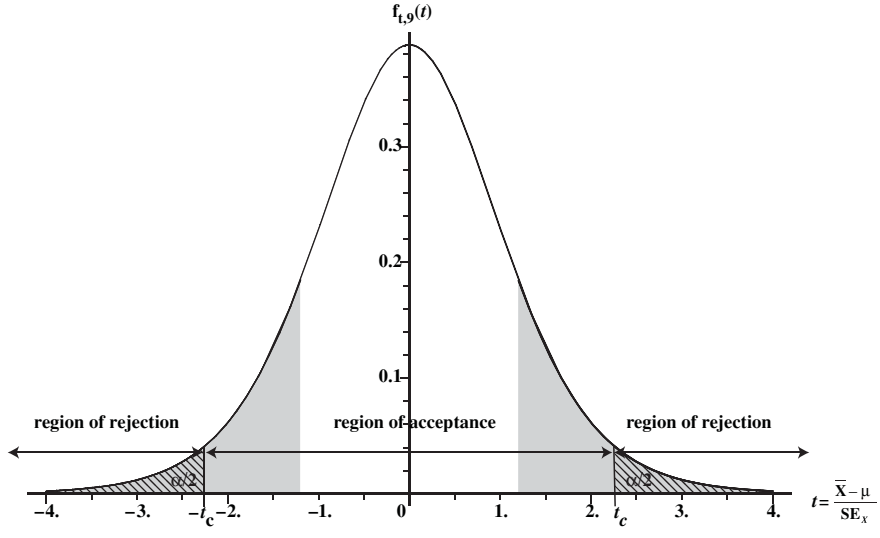


Figure 4.26 *The regions of acceptance and rejection for the hypothesis test H_0: $\bar{X} = \mu$ with H_1: $\bar{X} > \mu$ and $\bar{X} < \mu$ by selecting a significance level of $\alpha = 0.05$ as demonstrated using the mouse data set. The area below the curve colored in grey is 0.27. The area below the shaded curve is 0.05. The symbols $-t_c$ and t_c give the upper and lower limits of the region of rejection (see text for details).*

For this example the value of 1.2 $\frac{|21.63-22|}{(0.988/\sqrt{10})}$ is obtained, indicating that the population mean is 1.2 standard errors below or above the sample mean. The outcome of the *t*-test is the probability *p*, which corresponds to the grey area in Figure 4.26 and represents the area under the *t*-distribution over the limits between −infinity and −1.2 and between 1.2 and +infinity. The value of *p* gives the likelihood that a sample having at least such a large or even larger difference between measured and population mean values can be obtained by chance. In this example, *p* has a value of 0.27. This rather large value indicates that at least one in four cases a sample can be obtained in which the sample mean shows at least such a large departure from the historical control value. Because such a sample can be obtained frequently, the null hypothesis seems to be very likely. Therefore the null hypothesis cannot be rejected and the conclusion is that the sample mean is not significantly different from that of the historical population.

In contrast, a very small *p*-value, e.g. 0.001, would indicate that the sample collected is highly improbable if indeed H_0 is true. In other words it would be almost impossible (once in a thousand cases) to observe the data just by chance if the null hypothesis were true. Therefore, the null hypothesis should be rejected in favor of the most likely alternative hypothesis.

It is obvious how to deal with *p*-values above 0.2 and below 0.001. The grey area between these probabilities (0.001 and 0.2) is covered solely by convention. For this purpose an arbitrary threshold value or cutoff point of 0.05 (α) is defined, although another cutoff value (e.g., 0.01) might be considered as well. It is important that cutoff values are defined in advance before data collection takes place. When the *p*-value is below the cutoff point ($p \leq \alpha$) the sample is called significantly different from the

population tested, otherwise it is called not significantly different. Nowadays, statistical programs give the actual *p*-value, which should be included in the publication. For this purpose two significant digits are often sufficient for citing *p*-values. Alternatively, critical values for the test statistics (t_c) can be obtained from statistical tables for a given cutoff point. Its value can be used to construct the upper and lower limits of the region of rejection (Figure 4.26). If the absolute value of the test statistics exceeds the critical value (t_c) then the null hypothesis is rejected. For the example given above, statistical tables give a value of $t_c = 2.26$ by selecting $\alpha = 0.05$ and the degree of freedom (9). Consequently, the null hypothesis cannot be rejected, indicating no statistically significant difference between the sample mean and the population mean value ($\mu = 22$).

There is a close link between estimation and hypothesis testing since both methods are based on similar theoretical grounds. Therefore, the hypothesis test given above can be converted to calculate the confidence interval at the defined significance level (α).

The region of acceptance is essentially the same as the corresponding confidence interval. Consequently, the null hypothesis cannot be rejected (i.e., it is accepted) at the defined significance level if the confidence interval of a sample mean includes the hypothetical mean value (μ) given in the null hypothesis. For the body weight example given above, the 95% confidence interval of the mean is calculated to be 21.6 ± 0.706 enclosing the hypothetical mean value of 22. Consequently, no statistically significant difference can be reported between the sample mean (21.6) and the hypothetical mean value ($\mu = 22$).

Confidence intervals are considered to be superior compared with a single *p*-value since they present a numerical estimate of the uncertainty. In contrast, the *p*-value does not indicate the size of any difference. It solely gives the probability that the observed difference had arisen by chance alone. The most suitable statistical test depends on the nature of the hypothesis and on the kind of variable measured. The result of any hypothesis test is given by a *p*-value, which can be interpreted as discussed above. An overview on the different tests is given in Table 4.14.

Comment

The research hypothesis should be clearly defined in advance to allow researchers to be able to unequivocally formulate the statistical null hypothesis. The null hypothesis can be formulated by equality stating that there are no differences between the two groups. In this case there are alternative hypotheses stating that deviations may occur in both directions (shaded areas in Figure 4.26). For such a bidirectional hypothesis test, two-tailed or two sided *p*-values are used. The formulation of the null hypothesis by an inequality leads to a unidirectional test, which is performed on a single tail of the distribution. As a consequence the resulting significance is greater in the one-tailed test than in the two-tailed, i.e. a one-tailed test is more likely to detect an effect.

To make a comparison between treated and control groups they should be similar with respect to known sources of variability, e.g. test conditions, season, source of chemicals, etc. The statistical test can either result in a rejection of the null hypothesis or it can fail to reject.

Table 4.14 *Overview of statistical methods based on the data type and on the distributional assumption.*

Aim of the method	Continuous data from normal distribution	Ranks, scores, or data from nonnormal distribution	Dichotomous or binominal data
Comparison of the mean of one group to a reference value	*t*-Test	Wilcoxon test Kolmogorov–Smirnov test	χ^2-test
Comparison of the means of two unpaired groups	Unpaired *t*-test	Mann–Whitney test Kolmogorov–Smirnov test	Fisher exact probability test
Comparison of the means of two paired groups	Paired *t*-test*	Wilcoxon matched pairs signed rank sum test	McNemar sign test
Comparison of the means of three or more groups	Analysis of variance (ANOVA)	Friedman test Kruskal–Wallis test	χ^2-test
Linear association between two variables	Pearson correlation	Spearman rank correlation Kendall's coefficient of rank correlation	Contingency coefficient
Linear relationship between two variables	Simple linear regression	Grizzle–Starmer–Koch linear model	Simple logistic regression
Linear relationship among several variables	Multiple linear regression	Grizzle–Starmer–Koch linear model	Multiple logistic regression
Comparison of the distribution of data with a theoretical distribution	Shapiro–Wilk test D'Agostino–Pearson test Kolmogorov–Smirnov test	χ^2-test Kolmogorov–Smirnov test	χ^2-test Kolmogorov–Smirnov test

*The distribution of differences should be normal.

Failing to reject the null hypothesis is not the same as verifying that it is correct, since statistically nonsignificant results can be due to several circumstances: too small difference in the effect variable measured between the groups, too large variation in the samples, or limited **power** resulting from a small sample size. The sample size necessary to reach a more definitive answer can be calculated in advance by selecting the power of the test (see below).

Parametric statistical tests are based on the assumption of normality and homogeneity of variance. It is further assumed that the sample is randomly drawn from the population and the sample size is large enough to be representative. Some of the tests, such as the *t*-test, are considered to be robust with regard to these assumptions. Statistical tests developed for continuous data cannot be used with ordinal and nominal variables. **Nonparametric tests** should be applied if no assumption can be made about the underlying population, if it is known to be strongly skewed or if measurements were done on a nominal scale. Nonparametric methods may lack power compared with those that are parametric, especially if the sample size is small.

For comparison of multiple group means the **analysis of variance** (**ANOVA**) should be used. For this purpose the variation between the groups is compared with that within the groups. A multiple pairwise comparison based on the *t*-test is not correct since the cumulative probability of the Type I error increases rapidly, resulting in multiples of the preselected significance level (α). Typically, the Type I error is $n^* \alpha$, if n pairwise comparisons are to be made. Consequently, the chance of a false positive result will increase.

The hypothesis test may result in a false decision since the ultimate truth about the hypothesis is unknown. Furthermore, there is still a nonzero chance that the limited random samples taken might lead to a wrong decision.

Two types of errors are associated with hypothesis testing (Table 4.15). **Type I error** (**α-error**) occurs when one would claim a significant difference and reject the null hypothesis although in reality it is really true, i.e. the finding is false positive. The **type II error** (**β-error**) addresses the possibility that although the null hypothesis is false, one would fail to reject it and therefore generate a false negative finding. In other words, the β-error gives the probability of not detecting an existing difference. The size of an α-error has to be determined in advance, e.g. 0.05, whereas that of β-error depends on the α-error, the difference in the effect studied, its variation, and the sample size. The value of β is used to calculate the **power** of a test ($1 - \beta$), which is the probability of detecting an existing difference and rejecting a false null hypothesis. Therefore, the power is considered as a measure of the quality of the statistical test. A higher power is more likely to result in a test that detects statistically significant results.

Table 4.15 *An overview on the possible test outcomes related to a hypothesis test.*

	Test outcome	
	H_0 is rejected	Fail to reject H_0
H_0 is true	False positive (α-error)	Correct decision ($1 - \alpha$)
H_0 is false	Correct decision ($1 - \beta$)	False negative (β-error)

In each study enough experiments should be included to ensure that the outcome is not only scientifically but also statistically significant. The necessary sample size can be estimated in advance by **power analysis**. Qualitatively, a two-sided test, a more stringent significant level, and a larger power require a larger sample size. Larger sample sizes are also needed to get reliable results when the average effect difference is small and the measurement is variable. For two-sided problems, comparing a given difference (d) between two groups, a simple formula [Equation (4.35)] can be derived by assuming the difference (d) to follow a normal distribution, the same variation in both groups and by fixing $\alpha = 0.05$ and $\beta = 0.2$ (usually $\beta = 4^* \alpha$).

$$N = 1 + 15.7^* \left(\frac{\mathrm{SD}}{d}\right)^2 \tag{4.35}$$

For more general cases, the probable number of samples can be calculated either by using large and expensive full-featured statistical programs or small free programs (e.g., G*POWER - http://www.psycho.uni-duesseldorf.de/aap/projects/gpower or DSTPLAN - http://odin.mdacc.tmc.edu/anonftp) both running under Mac OS or Windows. Such estimates help to design studies with optimal sample sizes. If fewer experiments yield sufficient data, excessive experiments waste resources. If too few experiments are carried out the resulting data may not be adequate to reach statistically significant conclusions. In a study with human individuals or animals the size of the study should comply with obligatory ethical considerations as well.

Comment

The magnitude of the acceptable Type I and Type II errors depends on the aim of the research. For example, in a carcinogenicity study the α-error leading to the claim that a chemical is a carcinogenic when it is not is not the primary concern. In contrast, the reduction of the β-error, which argues that the chemical is not carcinogenic when it is, is of utmost importance.

Example

How many Sprague-Dawley rats are necessary in an experiment in which one wishes to measure at least a 15% exposure-related change in hepatic cytochrome P450 2E1 (CYP 2E1) content? The hepatic CYP 2E1 content in control rats is expected to be 9.7 ± 0.9 nmol/mg protein. The treatment might induce or reduce the CYP 2E1 content. Therefore, the data will be evaluated using a two-sided statistical test. By assuming normal distribution and equal variances in both groups and by fixing $\alpha = 0.05$ and $\beta = 0.2$, the number of animal in the control group as well as in the exposed group should equal $7\left(1 + 15.7^* \left(\frac{0.9}{0.15^* 9.7}\right)^2\right)$ [from Equation (4.35)].

After carrying out the experiments in 14 rats, the statistical test should have a power of 0.8 to detect a 15% difference between the group means (9.7 in the control group, either less than 8.2 or greater than 11.2 in the treatment group) if the standard deviation in both groups is 0.9.

4.6.6 Regression Analysis

The ultimate aim of many investigations is to derive and characterise dependencies between measured variables. To establish functional relationships among

variables regression analysis is a widely used statistical technique. The following investigates the simplest linear relationship between two continuous variables.

Propylene oxide (PO) concentrations measured in blood (C_{blood}; **effect variable** or **dependent variable**) of rats exposed to constant atmospheric concentrations (C_{air}; **independent variable** or **predictor variable**) of PO are expected on a theoretical basis to follow a linear relationship, at least over a certain range of exposure concentrations. Furthermore, no PO can be found in the blood of control, nonexposed rats since the compound is not endogenously formed. Therefore, the theoretical relationship can be described [Equation (4.36)] in the form of a straight line which passes through the origin

$$C_{blood} = m^* C_{air} \tag{4.36}$$

and has a slope equal to m. The 'best' value of this slope is determined using the data set shown in Figure 4.27A. The criterion of 'best' depends on the statistical approach used. For example, the **least-squares method** minimizes the summed deviations between measured (C_{blood}) and predicted blood concentrations ($m^* C_{air}$). Thus, the regression line, depicted by the straight line in Figure 4.27A, is fitted to the data by minimizing the vertical departure of the regression line Equation (4.37) from the measured data points. The smaller the residuals become the closer the line fits the data. The value of the sum of squares [Equation (4.38)]

$$\varepsilon_i = C_{blood} - m^* C_{air} \tag{4.37}$$

$$\left(\sum_{i=1}^{N} \varepsilon_i^2 = \sum_{i=1}^{N} (C_{blood,i} - m^* C_{air,i}) \right) \tag{4.38}$$

can be considered as a measure of the quality of the fitted line. Graphically, the sum of the squared residuals can be plotted versus the slope, depicting a function with a distinct minimum at about 0.058 (Figure 4.27B). Alternatively, statistical programs can be used to perform the necessary calculations resulting in an estimate for the slope and its standard error [Equation (4.39)].

$$C_{blood} = (0.058 \pm 0.0011)^* C_{air} \tag{4.39}$$

Comment

In order to carry out a valid regression analysis it is expected that the observations are independent and the dependent variable is normally distributed with a constant variability (homoscedasticity). The **weighted least-squares method** can be considered if these assumptions do not hold. Furthermore, it is assumed that the independent variable is error free. If the latter assumption does not apply, special regression techniques such as the **Bartlett regression** should be used. Finally, a linear relationship is expected between the two variables. The fulfilment of these assumptions can be revealed by a graphical analysis of residuals. The normal probability plot can be used to demonstrate whether the residuals are normally distributed. Departure from linearity, outliers, independence, and constant variance become evident by plotting the residuals versus predicted variable or independent variable. These plots should be evenly scattered and should not show any

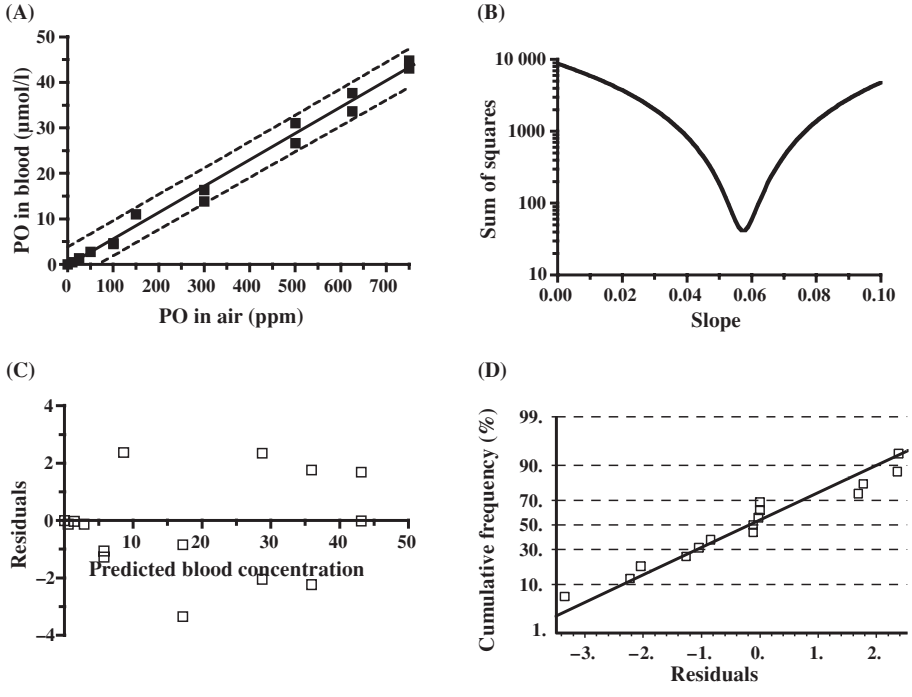


Figure 4.27 *Blood concentrations (C_{blood}) of propylene oxide (PO) measured in rats exposed to constant atmospheric concentrations (C_{air}) of PO (A). Symbols represent measured values. The solid line was obtained by linear regression [Equation (4.40)]*

$$C_{\text{blood}} = 0.058^{*}\, C_{\text{air}} \tag{4.40}$$

The dashed lines represent the prediction bands calculated at a confidence level of 0.95. Plot of the sum of the squared residuals versus the slope (B) indicates a minimum at the slope value of 0.058. Plot of residuals versus the predicted value of the blood concentrations (C) depicts no specific pattern. Normal probability plot of the residuals (D) demonstrates that residuals point closely to a straight line.

specific pattern such as lines having a nonzero slope, arch, or shape of a funnel. Regression analysis can also be conducted for observational studies when the assumptions given above are fulfilled. However, the interpretation of the regression parameters. i.e. slope or intercept, is often more complex.

The justification for the error-free independent variable is given by the fact that the variability of the air measurements can be neglected when compared with the variability of the blood determinations since the biological variability is expected to be several times larger than the variability resulting from the analytical procedure. Therefore, C_{air} might be considered practically error free when compared with C_{blood}. The necessary assumptions seem to be fulfilled, since the scatter of residuals versus predicted variables appears to be random (Figure 4.27C) and also the residuals point close to a straight line in the normal probability plot (Figure 4.27D).

The standard error can be used to calculate the confidence interval of the slope (0.058 ± 0.0022) at a confidence level of 0.95 under the assumption that the blood concentration follows normal distribution. In addition, a hypothesis test stating that the slope equals zero (H_0: $m = 0$ and H_1: $m \neq 0$) can be performed if necessary. For the example given above the slope differs significantly from zero and the null hypothesis should be rejected at $p < 0.0001$.

The regression line given in Figure 4.27A represents predicted values describing the average PO blood levels for a given exposure concentration. The expected confidence interval for the mean of many blood concentrations measured at an identical exposure concentration with a specified confidence level can be also calculated. By repeating the calculation over the entire range of exposure concentrations, upper and lower **confidence bands** for the regression line are obtained. Confidence bands contain the true regression line at a given confidence level.

A confidence interval can also be constructed describing the uncertainties for the blood concentration of a single new observation at a given exposure concentration. Repeating the calculation for a range of exposure concentrations with a confidence level of 0.95, **prediction bands** are obtained in which a measured blood concentration will be with 95% probability (dashed lines in Figure 4.27A). For example, it is expected with 95% confidence that rats exposed to PO concentrations of 400 ppm will exhibit blood concentrations between 19 and 26.9 μmol/l. The prediction band accounting for a single observation is expected to be broader than the confidence band related to an average response.

Comment

The regression line through the origin should be used when the regression line is expected on a theoretical basis to pass through the origin. Equations obtained by regression are often used to predict the value of the dependent variable. The predicted data may be invalid if extrapolation is carried out, i.e. the value of the independent variable is beyond its measured range. If the scatter plot of the data indicates a nonlinear relationship then nonlinear regression should be used. Computer programs provide effective algorithms to deal with nonlinear regression. A linearisation of the relationship by transforming the dependent and independent variables is, in general, not advised. The simple linear regression analysis can be extended to more than one independent variable (multiple or multivariate regression).

4.6.7 Probit Analysis

Probit analysis is used to determine the quantiles of dose–response curves such as the LD_{50}, by linearising the cumulative normal distribution.

Dose–response curves observed in biological assays are depicted by *S*-shaped sigmoid curves similar to a cumulative distribution function (see Figure 4.23B). In fact, the sigmoid curve resembles the cumulative normal distribution if the dose is given on a logarithmic scale and the tolerance levels (threshold below which the dose is tolerated or there are no response results) in the population are normally distributed. Sigmoid curves

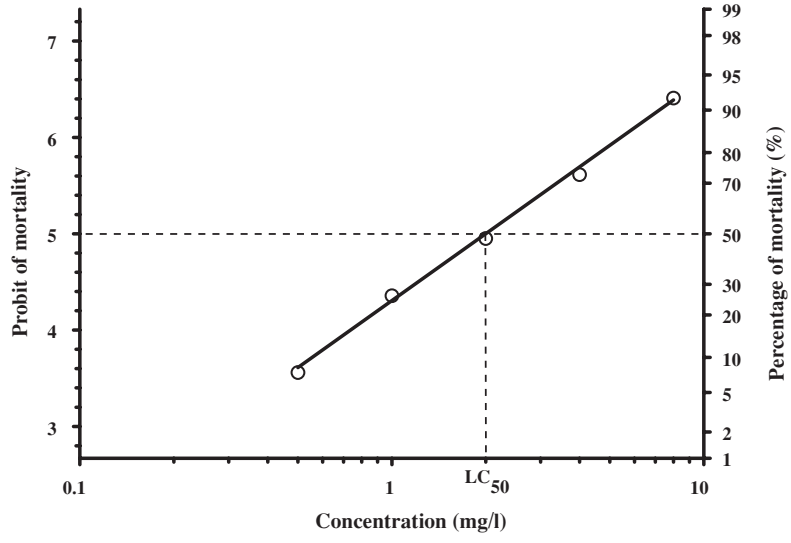


Figure 4.28 *Toxicity of an insecticide to the codling moth (Cydia pomonella). Percentage of mortality and probit of mortality as a function of the insecticide concentration. The median lethal concentration estimated graphically corresponds to about 2 mg/l. Symbols: simulated data. Solid line: obtained by visual fit.*

such as are obtained in an acute toxicity experiment are typically investigated by **probit analysis** in order to calculate the median lethal concentration (LC_{50}) or median lethal dose (LD_{50}). For this purpose, the *S*-shaped curve is transformed to a straight line by changing the scales of the dose and the observed frequencies, e.g. portion of animals affected, using the logarithmic and **probit-transformation**, respectively. This transformation is similar to that shown in Figure 4.24B. However, the ordinate is scaled linearly using the so-called **probit**, which is a normally distributed random variable with a mean of 5 and a variance of 1. The ordinate values are positive and linearly scaled in the range between 2 and 8. For example, if 50% of the investigated population are responding, the corresponding probit value on the ordinate is 5. The frequencies of 16% and 84% correspond to the probit values of 4 and 6 (Figure 4.28). To transform the experimentally observed incidences to probit values, suitable tables can be used. Thereafter, the resulting linear relationship is analysed and the median effective dose is calculated by appropriate regression techniques. For a rough estimation of the median lethal dose, a graphical method using **log-probit paper** or **lognormal probability paper** can be considered (Figure 4.28).

4.6.8 Experimental Designs

Statistical design of experiments ensures that the necessary amount of the right type of data will be generated in order to reach valid and objective conclusions with an acceptable degree of precision.

Experiments are performed to confirm or refute hypotheses. Carefully planned experiments are required to enable the researcher to reach an unequivocal conclusion. The ultimate aim of good experimental design is to rule out any alternative causes or confounders that might explain the results, and to minimise bias. For this purpose the goal of the experiment should be clearly defined by a research hypothesis before carrying out any experimental work. The further planning of the experimental procedure has to be based on three major principles: **replication**, **randomisation**, and **stratification (blocking)**.

Appropriate **replication** of the experiment is required to obtain a statistically significant result. The number of necessary repetitions required to ensure significant results can be estimated using the principles of **power analysis**. In general, **replication** reduces variability in the result, which in turn leads to a narrow confidence level and to statistical significance.

Randomization is regularly used in experiments to minimize an eventual bias by creating homogenous groups. A 'completely randomized design' is achieved if subjects are assigned to treatment or control groups completely at random. If the experimental subjects are not comparable with respect to age, sex, treatment schedule, etc., simple randomization will not account for heterogeneity.

Stratification is a technique used to divide heterogeneous experimental subjects into homogeneous blocks (strata) according to some of their properties such as gender. The resulting **blocking** balances the treatments across the variability represented by the blocks. In a carcinogenicity study each of the treatment and control groups might contain two different blocks assigned to the male and female animals.

Using these principles a detailed experimental plan can be established by listing the number of experimental subjects, how they will be selected and assigned to groups, what specific treatments will be applied, in which order will these treatments be carried out, which observations will be taken, and which statistical procedures will be used to evaluate the results.

Different types of experimental designs can be realised according to the research objective: **comparative design** [to compare the effect(s) of one independent variable] or **factorial design** [to compare the effect(s) of two or more independent variables or to identify important factors effecting the outcome]. A detailed description of such designs can be found in books on advanced statistics The quality of an experimental design is measured in terms of cost related to experimental efforts and in the accuracy of the conclusions that can be drawn. Finally, it should be taken into account that the most carefully planned experiments will lead only to trivial results if a reasonable research hypothesis is lacking.

Comment

To confirm a hypothesis, it is often required to demonstrate that the investigated phenomenon occurs exclusively under the conditions of a certain treatment. Therefore, it is a good practice to include both a positive and a negative control.

4.6.9 Statistical Software

Nowadays, statistical analysis is carried out using computers with the aid of a given statistical software package.

There are large differences among the packages with respect to price, user-friendliness, graphical design, accuracy, and complexity (for a review see http://www.statsci.org/statcomp.html). The major generalists are **SPSS** from SPSS Inc., USA (Windows), **SAS** from SAS Institute Inc., USA (Windows and UNIX), **Statistica** from StatSoft Inc., USA (Windows), **STATA** form StataCorp LP, USA (Macintosh OS X, Windows, and Unix) and **JMP** from SAS Institute Inc. USA (Macintosh OS 9, Windows, and Linux). Also, software designed for general mathematical or technical computing, e.g. **Mathematica** from Wolfram Research Inc., USA (Linux, Macintosh OS X, Windows, and UNIX) or **Matlab** from The Math Works Inc., USA (Linux, Macintosh OS X, Windows, and UNIX) can be considered. There are many other products, like Prism from Graphpad Software Inc., USA (Macintosh OS X, Windows), which cost less yet but offer limited scope and functionality (http://www.uni-koeln.de/themen/Statistik/softliste.html and http://www.statsci.org/statcomp.html). The public-domain software '**R Foundation for Statistical Computing**' is developed within the framework of the Free Software Foundation and can be used on Linux and Macintosh OS X as well as on Windows computers (http://www.r-project.org). The choice of the software is often determined by existing installations, convenience, or monetary constraints and by the complexity of the data analysis. In any event, a certain amount of time must always be invested in order to use a given software package efficiently.

The major disadvantage of most statistical software is that it allows the user to perform any test, even one incompatible with the collected data. This might easily lead to a wrong conclusion. Therefore, at least a basic understanding of the statistical tests is indispensable before using any statistical software package. It must be emphasized that data collected in a well done study can be analysed in many ways, leading to scientifically and statistically significant results, but even the most sophisticated statistical method cannot compensate for shortcomings in the experiments or study design.

4.6.10 Summary

Repeated measurements result in data that are not identical because of random variability. This variability may lead to erroneous interpretation. Statistical methods are indispensable for analysis and description of the random variability and to enable scientifically reliable conclusions to be drawn.

Statistics should be an integral part in designing an experiment, in particular to obtain a sufficient number of data for the required degree of precision and to minimise bias. Considering the corresponding sample characteristics, data should be presented by means of descriptive statistical methods. Inferential statistics are used to confirm or refute scientific hypotheses.

The present chapter gives an introduction into basic principles of statistical reasoning, including error propagation, linear regression, and probit analysis.

References

www.weibull.com/GPaper
G*POWER - http://www.psycho.uni-duesseldorf.de/aap/projects/gpower
DSTPLAN - http://odin.mdacc.tmc.edu/anonftp

http://www.statsci.org/statcomp.html
http://www.uni-koeln.de/themen/Statistik/softliste.html
http://www.r-project.org

For Further Reading

Online references

The Engineering Statistics Handbook (NIST/SEMATECH, 2004; http://www.itl.nist.gov/div898/handbook/index2.htm) provides an overview and in-depth details on several statistical methods. The main emphasis lies in the exploratory data analysis. There are many technical but no biological examples.

The online edition of Statistics at Square One (T. D. V. Swinscow) is a popular introduction to medical statistics: (http://bmj.bmjjournals.com/collections/statsbk)

Printed references

The reader might refer to books intended for an introductory treatise with practical examples (e.g., Weisbrot, 1985; Miller and Miller, 1988; Gad, 1999).

For a more exhaustive coverage it is recommended to consult in-depth statistical books either with less mathematical exposition (e.g. Altman, 1996) or with thorough details (e.g., Sokal and Rohlf, 1981).

There are books covering one or more aspects of statistical analysis, e.g. with major emphasize on experimental design (e.g., Box et al., 1978; Montgomery, 2005), on regression techniques (e.g., Kleinbaum and Kupper, 1978) on analysis of quantal response data (Morgan, 1992) and how to report statistical results properly (Lang and Secic, 1997).

Computational references

Straightforward introductions to the public domain software R, which provides a powerful environment for the statistical analysis on multiple computing platforms:

David Rossiter (2003–2006) Introduction to the R. http://www.itc.nl/personal/rossiter/teach/R/RIntro_ITC.pdf

Vincent Zoonekynd (2005) Statistics with R. http://zoonek2.free.fr/UNIX/48_R/all.html

Publications

Altman, D. (1996) Practical Statistics for Medical Research. Chapman & Hall, London, Weinheim, New York, Tokyo, Melbourne, Madras.

Box, G., Hunter, W. and Hunter, J. (1978) Statistics for Experimenters. John Wiley & Sons, New York, Chichester, Brisbane, Toronto, Singapore.

Gad, S. (1999) Statistics and Experimental design for Toxicologists. CRC Press, Boca Raton, London, New York, Washington, D.C.

Kleinbaum, D. and Kupper, L. (1978) Applied Regression Analysis and Other Multivariable Methods. Duxbury Press, Boston, Massachusetts.

Lang, T. and Secic, M. (1997) How to Report Statistics in Medicine. American College of Physicians, Philadelphia.

Miller, J. and Miller, J. (1988) Statistics for Analytical Chemistry. Horwood Limited, Chichester, Ellis.

Montgomery, D. (2005) Design and Analysis of Experiments. John Wiley and Sons, New York.

Morgan, B. (1992) The Analysis of Quantal Response Data. Chapman and Hall, London.

Sokal, R. and Rohlf, F. (1981) Biometry. W.H. Freeman and Company, San Francisco.

Weisbrot, I. (1985) Statistics for the Clinical Laboratory. Lippincott Company, Philadelphia.

5

Risk Assessment

5.1 Mathematical Models for Risk Extrapolation

Demonstration of the use of mathematical models to extrapolate the risk of low-dose occupational or environmental exposure starting with high-dose exposure data in experimental animals.

Jurgen Timm

5.1.1 Introduction

Limits of Observation

Experience has demonstrated that there are two approaches to determining the risk of exposure to carcinogens: the use of experimental animal data and the use of data derived from epidemiological studies.

Methods for the determination of carcinogenicity in animals have been standardized, and provided that significant sources of interference have been eliminated, one can assess the relationship between dose (d) and the probability (P) of tumorogenesis. Since many chemicals must undergo metabolic activation, the application of these data to humans is complicated by the need to thoroughly understand the metabolism and toxicokinetics of the chemicals in both animals and humans.

Epidemiological studies seeking to detect cancer due to exposure to a chemical are complicated because the human population is exposed to a wide range of chemicals, some of which may be carcinogenic. The additional risk caused by the chemical under study (Δd) can be estimated by comparing the exposed group with an appropriate control

Toxicology and Risk Assessment: A Comprehensive Introduction Edited by H. Greim and R. Snyder

group. Although the absolute risk (P) due to Δd cannot be known, the data will provide an estimate of relative risk (R).

The relationship between P and R may be calculated by using Equation (5.1). Thus, if one assumes that the control population is not exposed to the chemical, i.e., $d = d_0$, there is no risk, i.e. P_0. The risk posed by exposure to the chemical is then $P(d)$ and is based on $\Delta d + d_0$.

$$P(d) = R(\Delta d)P_0 = R(\Delta d)P(d_0). \tag{5.1}$$

Limits of Evidence

This method has proved to be successful in estimating cancer risk in the occupational setting or in cigarette smokers where exposures tend to be relatively high. For example, it has been suggested that these exposures may account for 10–20% of the overall cancer risk or cancer in 5–10 out of 100 000 people per year (Cederlöf et al., 1978).

When studying potential environmental carcinogens, however, technical considerations prevent the accurate determination of the risk below 10–20%, i.e. when $R = 1.1$ or 1.2. Thus, neither animal experiments nor epidemiology will, in general, provide direct risk determinations for environmental contaminants. Therefore, it is necessary to extrapolate the dose–response relationship from high-exposure concentrations derived from animal research or from data derived from epidemiological studies performed on workers to workers exposed at low concentrations.

The application of mathematical modelling to the solution of the problem requires clearly defining the relationship between the response at high-dose exposure leading to a well defined incidence of cancer and the incidence at very-low-dose exposure. The mathematics should not be excessively challenging and should provide a reasonable measure of upper bound risks.

Linear risk evaluation involves developing a mathematical model which fits the measured data. It should be as simple as possible or as complex as is necessary.

Figure 5.1 shows a dose–response curve which is extrapolated beyond the data range but is intended to show responses over the entire dose range. The Figure refers to the absolute risk but the results can be expressed as the relative risk using Equation (5.1).

Extrapolation with Straight Lines

The simplest curves used for estimating the risk of environmental contaminants are straight lines. Linear extrapolation from empirical data (dotted line g_1) tends to underestimate risks at the lower dose-range because the dose–response curve is actually sigmoidal in shape.

The straight-line extrapolation (straight-line g_2) involves extending the line through the origin of the coordinates. The WHO (1987) based its evaluations on this method. It defined a concept termed the 'unit risk' as 'the additional lifetime cancer risk in a

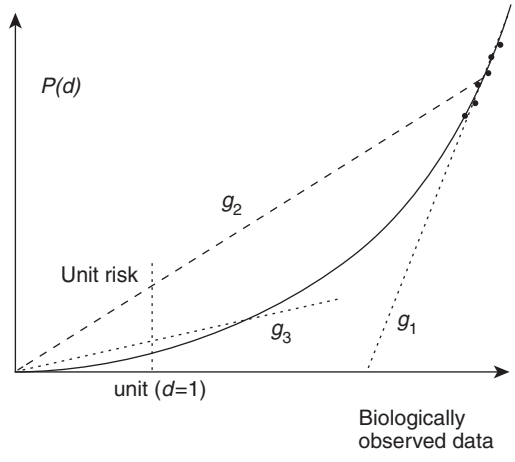


Figure 5.1

hypothetical population in which all individuals are exposed continuously from birth throughout their lifetimes to a concentration of 1 microgram/cubic meter of the agent in the air they breath.' Carcinogenic-potency comparisons can then suggest that a given chemical presents a risk of 1 in 10 000, 1 in 100 000, 1 in 1 000 000, etc, as measures of relative carcinogenic potency.

The straight lines through the origin are marked only by their slopes, i.e. by the risk 'a' at the unity point ($d = 1$, unit) of the scale. Conversely a dose d_b associated with a specific risk P_b (1%, 5% 10%, for example) is known as **benchmark dose.** One can derive the linear approximation P_l of the real risk $P(d)$ and the respective estimation d_{bl} of the benchmark dose d_b following Equation (5.2). P_l overestimates the true risk (g_3) between point zero and the intercept (point of departure, POD) of true and linear response in the case of sigmoidal dose–response curves following the principle of precaution. Consequently d_{bl} will underestimate the real benchmark dose within this interval.

$$P_l(d) = ad > P(d) \quad \text{and} \quad d_{bl} = P_b/a < d_b \tag{5.2}$$

Consistency with Empirical Facts

There is a risk of over simplification of the model to the extent that it does not correspond with the empiric data anymore. Therefore this model has to be cross checked with each single substance.

As far as the data in the field of measurement do differ in a statistically significant way from the linear extrapolation through the origin, a more complex reconsideration is necessary, which will analyse the curve form along the range of measurement in a more detailed way and perhaps incorporate further details concerning sub-populations and the biological basis of the toxic process. A relevant refinement of method was taken as a basis for the evaluation by the US Environmental Protection Agency (see EPA 1984, 1995; and following sections).

Straight Lines as Upper Boundary Estimates

The general idea of linear evaluation (see straight line g_3) near the origin need not be abandoned. It still makes sense to have a simple and understandable method of evaluation of upper boundaries of risks near the origin, as the method of unit risks or benchmark doses and a straight line through the origin are able to provide.

5.1.2 Methods of Extrapolation

For a rough classification one can establish two groups: One group will use empiric data only, while the other one will at first fit the suitable functions to the data. We are here introducing three typical variations:

Extrapolation from Single Data Points

In connection with the carcinogenic risk there are two proposals to mention:

WHO Method The most simple method has, as mentioned above, been used by the World Health Organization (WHO) when drawing up the *Air Quality Guidelines for Europe*. Here the individual records, which could be derived from different studies, are evaluated, and unreliable results are eliminated methodically. Using the average of the remaining data points one draws a straight line to the origin.

ICRP Method The International Commission for Ray Protection (ICRP) uses a slightly different method, incorporating only the two data points with the lowest dose of all data. Then, by using a linear regression analysis the straight line through the origin will be estimated, which gives the best fit to those two data points. For more details about assessment of low radioactive contamination see Rechard (1999)

Extrapolation with Function Adjustment (EPA Method)

A basically different approach is to fit a function (graphically a curve) to the data first of all, and by that to win a linear approximation of that function near the origin.

The EPA has done a consistent evaluation of quite many carcinogens with this method, obtaining a method for the tangent to the curve at the origin to be specified after 'fitting' a function. Preferably one uses the multistage model of Armitage and Doll (1961), but differing models may yield a better fit to data and knowledge. The method is feasible in general, as all functions in question can be approached by a so-called Taylor expansion [Equation (5.3)] near the origin, of which the first-order terms result in the tangent.

$$\mathrm{f}(d) = a_1 d + a_2 d^2 + a_3 d^3 + \ldots \tag{5.3}$$

However, this approach has two disadvantages. First of all it is quite sensitive to mistakes in single records, and secondly it underestimates the risk in the case of expected functions with left-curvature, this being the reason that the EPA uses straight lines connecting the origin with an estimated low-dose point (point of departure). This method tends to underestimated risks in the (rare) case of right-curvature, which may be reflected by taking the upper confidence limit instead of working with the estimated value directly.

Evaluation of Extrapolation Methods

The three extrapolation methods have different advantages and disadvantages.

As the mathematical effort can be kept small, the ICRP and WHO methods have the great advantage of being easily understood by nonexperts. In coherence with environment carcinogens the ICRP method has the disadvantage of being based on boundary data points of generally extremely poor precision compared with the remaining data set which is neglected, and therefore is prey to high random fluctuations.

This aspect is solved in a much more stable way by the WHO method, especially when pre-results have been averaged. However, it is questionable whether this method uses the available information sufficiently, and whether it is not in contradiction to empirical results. It should be checked to ascertain whether the approach is sufficiently correlating with the empirical results for each mathematical model used.

With the EPA method one uses the existing information completely, but one is bound by the specific choice of allowed functions.

If the data are quite precise on a function of high curvature, the confidence limit will become very narrow, and the linear evaluation will lead to too low results.

5.1.3 Consideration of the Time Aspect

Within the scope of evaluation of the degree of danger to a human population posed by specific environmental carcinogens, at first one is interested in the additional risk (excess of risk) only, which is caused by a relevant carcinogen, especially the risk of dying by certain kinds of cancer after life-long constant contamination. The quotient of contaminated and noncontaminated risk (relative risk, *RR*), however, is the standard result of epidemiological studies. Both risk estimates may be derived from basic contamination d_o, risk P_o, and the risk P_d with contamination $d = d_0 + \Delta d$ as shown in Equation (5.4):

$$ER = P_d - P_o; \qquad RR = P_d/P_0 \tag{5.4}$$

These risks can be approximated by using Unit Risks or Benchmark Doses, as described in Section 5.1.2, above. This approach is effective for toxicological situations with acute contamination, but it represents the real relation only incompletely in the case of chronic diseases, as it neither considers the variation in the amount of contamination over time, nor the quite different cancer risks in various age groups, and the natural mortality of the respective populations.

Mathematical Description of Risk and Survival

A more precise mathematical description may be obtained by understanding the risk as a function of time; that is, to introduce changes in the course of time into the concept.

The following survey shows some definitions and formulas of great importance for this concept. More details can be found in the relevant literature:

$L(t)$ basic **survival function**, the probability to survive at least until time t (in an uncontaminated population);

$s(t)$ **hazard function**, the border probability of dying by a relevant cancer at time t, if one has been living up to then in an uncontaminated population (with background dose d_0);

$d(t)$ biological **effective dose** (internal dose of ultimate carcinogen) at a time t;

$s(d,t)$ hazard function of contaminated population under contamination function $d = d(t)$;

$L^*(t)$ survival function in the contaminated population.

Simplification of Environment Carcinogenesis

The general mortality of the population is hardly changed by typical environmental carcinogens, as they contaminate at low dose levels and result in rare cases of cancer only. Therefore one can approximate the model by assuming that the overall survival function is hardly affected by such contamination [Equation (5.5)]:

$$L^*(t) = L(t) \tag{5.5}$$

Integrated over the entire life span the risks in question are given by Equation (5.6):

$$P_0 = \int_0^\infty s(t)L(t)\mathrm{d}t \tag{5.6}$$

for the uncontaminated (general) risk, and Equation (5.7):

$$P_d = \int_0^\infty s(d,t)L(t)\mathrm{d}t \tag{5.7}$$

for the risk of the contaminated population.

Usually one assumes that the hazard function may be split into one general hazard $s(t)$, depending only on time t (resp. age), and one factor, depending on the dose. In order to simplify later formulas this factor will be presented in the specific form $1 + \mathrm{f}(d)$ [Equation (5.8)]:

$$s(d,t) = s(t)[1 + \mathrm{f}(d(t))] \tag{5.8}$$

This is being cited as the proportional hazard assumption. Note that in combination with Equation (5.6) the specific formulation in Equation (5.8) implies that $\mathrm{f}(d_0) = 0$ and $P_0 > 0$ if $P_\mathrm{d} > 0$ for any dose d.

Using this formula the excess of risk may be written as Equation (5.9):

$$ER = P_d - P_0 = \int_0^\infty \mathrm{f}(d(t)) \cdot s(t) \cdot L(t)\ \mathrm{d}t \tag{5.9}$$

Practical Specification of Functions

From Integral to Sum In reality of course the precise functions are unknown, but one can use $s(t)$ and $L(t)$ approximations from mortality statistics life tables using at least

5-year intervals. Doing so one gets an approximation of the integral by the following sum [Equation (5.10)]:

$$ER = \sum_{i=1}^{N} \mathrm{f}(d_i) s_i L_i \quad \text{with} \quad i = 1, 2, \ldots, N \tag{5.10}$$

where i indicates the respective 5-year-period and only $\mathrm{f}(d_i)$ is still unknown.

Actual or Cumulated Dose? From results of experimental cancer research one can conclude that the biological effective dose, in a very complex way, relies on the accumulated dose. In any case, it is not only caused by the actual concentration of the contamination **(actual-dose principle)**, but earlier exposures may lead to cancer incidence later on. Often even the **cumulated overall dose** *(cumulated-dose principle)* can be the causing factor.

In the case of prognoses about the danger for the population caused by an exposure, to be on the safe side, one uses the cumulated overall dose as dose parameter. Therefore for a constantly contaminated population with actual dose δ per year the result is given by Equation (5.11)

$$d_i = \delta \cdot (i - 0.5) \cdot 5 \tag{5.11}$$

as the medium-relevant overall dose during time interval i [between 5 times $(i-1)$ and 5 times i years]. This leads to Equation (5.12).

$$ER = \sum_{i=1}^{N} \mathrm{f}[\delta \cdot (i - 0.5) \cdot 5] \cdot s_i \cdot L_i \tag{5.12}$$

5.1.4 Methods for Determination of Benchmark Doses and Unit Risks

As far as linear approximations of low-dose effects are in question, Unit Risks provide sufficient information to allow calculation of any dose effect for these extrapolations [see Equation (5.2)]. Benchmark doses d_b may be calculated as the inverse unit risk multiplied by the respective benchmark risk P_b as $d_\mathrm{b} = P_\mathrm{b}/a$ [using Equation (5.2)]. Benchmark doses are commonly preferred for interpretation and communication of results as they are defined by (meaningful) standardized risks instead of (arbitrary) dose units. From the mathematical point of view, however, the slope (unit risk) of the linearized function is the key parameter.

The dose–effect function f, introduced in Section 5.1.3., naturally has to be specified for the specific calculation of unit risks. Such specifications are usually based on epidemiological studies, for example working-place studies. They report specific *cumulated* overall doses $d_1, d_2, \ldots, d_\mathrm{k}$, respectively, with constant *actual* doses $\delta_1, \delta_2, \ldots, \delta_\mathrm{k}$, and associated cancer rates $P_1, P_2, \ldots, P_\mathrm{k}$ or relative risk estimates $RR_1, RR_2, \ldots, RR_\mathrm{k}$ compared with a control population with cancer rate P_0. Arguing that cancer mortality is mainly concentrated in the last time periods one may approximate the ER by replacing the time-dependent doses in Equation (5.12) by a constant dose near to the cumulated dose for the last interval. The failure will be great in the first time intervals but the hazard there is very small, such that the impact of this replacement may be tolerable. Using this approximation and cumulative doses corresponding to Equation (5.12) one

may derive relevant excess of risks for each sub-population j as shown in Equation (5.13).

$$ER_j = \mathrm{f}(d_j) \cdot \sum_{i=1}^{N} s_i L_i = \mathrm{f}(d_j) P_0 \tag{5.13}$$

Note that under these assumptions the relative risk is given by Equation (5.14).

$$RR_j = 1 + \mathrm{f}(d_j)\ j = 1, 2, \ldots, k \tag{5.14}$$

The relative and absolute risk for the function f can be estimated by Equations (5.13) and (5.14) using the different dose groups and their relative risks by interpolation and extrapolation. In most cases linear approximations are calculated for excess of risks, referring to small excess of doses Δd [see Equation (5.1)]. We have discussed problems arising from such extrapolations, and possible conventions in Section 5.1.2.

In many cases, however, the extrapolation may not be extended as far, as a comparison of the actual doses in the environment and at work would lead us to expect, as even short-term contaminated workers have been examined and found to have an overall dose which is comparable with life-long environmental contamination.

In the case of insufficient data, or in the case of data not differing significantly from a linear function (straight line through the origin), the existing data may be summarized according to the WHO approach for extrapolation (see Section 5.1.2) in an 'average' data point. It consists of the average relevant dose d_m (sum of excessive doses Δd_m and basic contamination d_0) and an average absolute (P_m) or relative risk (RR_m) depending on the data basis. The linear function of Equation (5.2) is commonly postulated for the relation between excess of risk and excess of dose. Applying Equations (5.2), (5.13), and (5.14) to this situation one receives the following terms [Equations (5.15)–(5.17)] for the unit risk a, the benchmark dose d_b (for excess of probability P_b), and the dose–response function:

$$a = (RR_m - 1) \cdot P_0 / \Delta d_m \tag{5.15}$$

$$d_b = P_b / a = P_b \cdot \Delta d_m / [(RR_m - 1) \cdot P_0] \tag{5.16}$$

$$ER(\Delta d) = a \cdot \Delta d = \Delta d \cdot (RR_m - 1) \cdot P_0 / \Delta d_m \tag{5.17}$$

If more information is available more sophisticated approaches are possible. The general method, however, follows the same lines. Instead of a point estimate $(\Delta d_\mathrm{m}, ER_\mathrm{m})$ a greater interval of the dose–response relation will be accessible by theory or by empirical data. As this curve becomes more and more uncertain as the origin is approached, one tries to find a low dose point sufficiently small and with a sufficiently certain response value. From this point (point of departure, POD) one draws a straight line to the origin taking the principle of precaution into account, i.e. construct lines 'on the safe side' overestimating possible risks. The widespread praxis is to use an appropriate benchmark dose d_b as point of departure. If this benchmark dose is estimated as excess of dose for the excess of risk P_b the linear extrapolation may be written as Equation (5.18).

$$ER(\Delta d) = a \cdot \Delta d = \Delta d \cdot P_b / d_b \tag{5.18}$$

5.1.5 Dose-conversion

Before practical use of extrapolation of the dose–effect relationship one has to consider further aspects which have not yet been mentioned. That is the conversion of the contamination situation in studies, for example at work, into life-long contamination in the environment.

The relevant dose, Δd, in Equation (5.17) has not been specified so far. For environmental toxicity in particular it should be calculated from small but nearly life-long constant doses δ (for example, in units per year). The WHO simplified this calculation by following these considerations: cancer mainly shows up in the high age intervals. Thus the relevant cumulated excessive dose is almost the life-long overall dose Δd_{ges} for 70-year-olds ($N = 18$ age groups). Taking this approximation, the effect will be overestimated and the result stays on the safe side [Equation (5.19)].

$$ER(\delta) = a \cdot \Delta d_{\text{ges}} = a \cdot 70 \cdot \delta \tag{5.19}$$

Another problem arises if data from working studies are used for the estimation of environmental risks. For this one may introduce conversion factors $f_1 = 8/24$ for the difference between the duration of the working day and the calendar day, and $f_2 = 240/365$ for the different numbers of working days, and the days of the calendar year. If T denominates the time of contamination in the working-place study, and T_{exp} is the expected lifetime, then the conversion of constant working-place dose δ into constant lifetime dose δ' will be as given by Equation (5.20).

$$\delta' = \delta \cdot f_1 \cdot f_2 \cdot (T/T_{\text{exp}}) \tag{5.20}$$

The situation is far more complex if an extrapolation from animal experiments has to be calculated. See, for example, Davidson et al. (1986).

5.1.6 Models of Carcinogenesis

Principally, one can extend these derivations to any kind of dose–effect relationship. Empirical data may be fitted to the theoretically expected curve, a benchmark dose and a linear approximation (characterized by unit risks) near the origin may be derived using standard analytical techniques, but this will lead to a considerable dependence from the choice of the respective formula being used. This is not satisfactory and leads to the demand that this approach must not be arbitrarily set up, but should be based on a simple and fundamentally plausible model.

Such a model has been introduced for carcinogenic processes in the early works of Armitage and Doll (1961), assuming that the cells have to pass a number of stages before they mutate to malignant cancer cells. The transition between certain stages can be triggered by small doses of harmful chemical substances. That implies, for mathematical models, that the transition probability between the stages are functions of the dose of the ultimate carcinogen. The reaction between carcinogen and biological compartments causing these transitions of cells from stage to stage are described as **hits.** Figure 5.2 shows a graphical representation of the model. Several hits may be necessary in order to

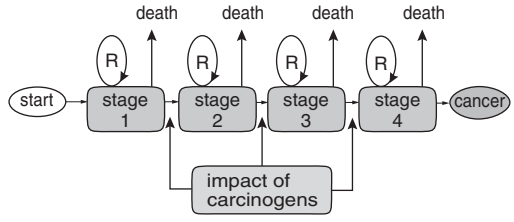


Figure 5.2 *Multihit–multistage model (R replication).*

induce a single stage transition and the probability of this determines the probability of stage transition. For details see Wahrendorf (1990) or Portier et al. (2000).

Depending on how many stages and hits one considers, one arrives at the following formulas for the expected risk, at dose d and time t, which contain each other as special cases [Equations (5.21)–(5.24)].

$$P(t,d) = 1 - \exp[-\exp(\mathrm{a} + \log d + \log t)] \tag{5.21}$$

one hit–one stage

$$P(t,d) = 1 - \exp[-\exp(\mathrm{a} + \mathrm{b}\ \log d + \log t)] \tag{5.22}$$

multihit–one stage

$$P(t,d) = 1 - \exp[-\exp(\mathrm{a} + \log d + \mathrm{k}\ \log t)] \tag{5.23}$$

one hit–multistage

$$P(t,d) = 1 - \exp[-\exp(\mathrm{a} + \mathrm{b}\log d + \mathrm{k}\ \log t)] \tag{5.24}$$

multihit–multistage
(Weibull distribution)

The most general formula of this sequence (5.24) with b hits and k stages is suitable for description of many relevant experiments in cancer research. The Druckrey law [see Druckrey and Kupfermüller (1948) and figure 5.3] could be used as an early example to verify the consequence of Equation (5.24) that the time points T where 50% of the population would have developed a tumour and follows Equation (5.25).

$$d^{\mathrm{b}} \cdot T^{\mathrm{k}} = \text{constant} \tag{5.25}$$

Dose Model

In order to check Equation (5.24), graphically one can use a double-logarithmic scale for age-specific tumour incidence and age, where Equation (5.24) predicts a bundle of parallel straight lines for various doses. Figure 5.4 shows this in a typical experimental result, where a great number of mice have been tested (Lee and O'Neil, 1971) with four different doses of benzopyrene. In human populations one can find similar curves, as Armitage and Doll (1961) have pointed out.

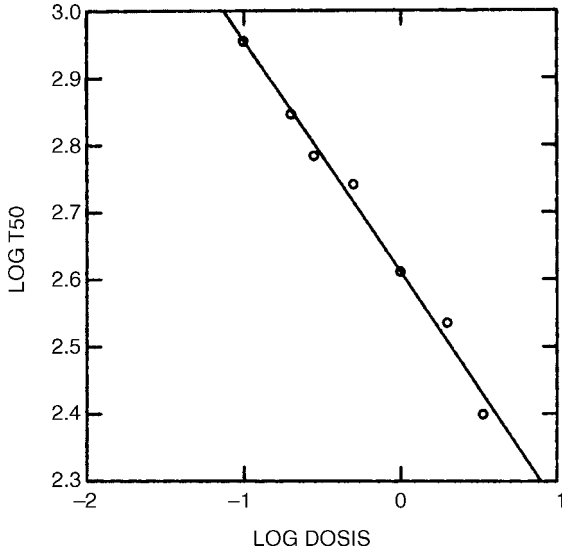


Figure 5.3 *An illustration of Druckrey's Law with an example taken from data published in H. Duckrey et al., 1962, showing tumor-induction times (time of 50%) tumor-bearing animals) versus dosis for 4-(dimethylamino)azobenzol and rats. The datapoints lie nearly on a straight line in a double-logarithmic scale. The publication gives numerous other examples for this relation.*

What does a transition to smaller doses at fixed observation times imply for this model? One easily recognizes that all models are approximated by the following formula [Equation (5.26)] near to zero (for fixed *t):*

$$P(d) = \mathrm{a} \cdot d^{\mathrm{b}} \tag{5.26}$$

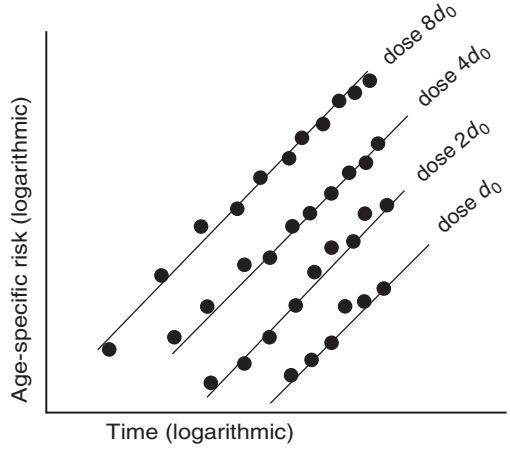


Figure 5.4 *Effect of benzopyrene on mice skin.*

As $a > 0$ and b is supposed to be an integer, the respective curves take a sub linear course, i.e. they stay below an extrapolated straight line through the origin. From that point of view even the simple estimation of the WHO method gives a conservative, i.e. upper-boundary, risk evaluation for this model group if extrapolation intervals are enlarged. More detailed linear upper boundaries could be derived, for example with the EPA method, without too large an effort, when the range of precaution is specified.

Age-specific Risk-time Models

If one is interested in simple models for risk evaluation or if more sophisticated information about the underlying biological process is not available, the multhit–multistage model family is the first choice for carcinogens. If more information is available it may serve as a starting point for more complicated model considerations, leading to in greater biologically based cancer-risk models. For an introduction in this wide field of research see Cogliano et al. (1999).

5.1.7 Assumptions and Limits of Extrapolation in Mathematical Models

Finally, underlying assumptions and limits of these methods should be discussed. We will limit ourselves here to the four most important requirements.

Linearity

Linearity surely is the most often mentioned and discussed assumption. We have mentioned above that a nonlinear dose–effect function can be approximated by a straight line at the origin for all differentiable functions, and it is possible to define a straight line as the upper limit in the field of precaution, if one knows the curve well enough. Thus linearity is not as important as is normally assumed as long as small doses are being studied.

It is of greater interest how the real dose–effect curve deviates from the estimated straight line near the origin. A sublinear curve form may be derived from the model mentioned above, which makes the construction of the straight line for precaution purposes much easier. If the dose–response curve has right-curvature (for example, effects proportional to the square root of the effective dose) it is no longer possible to find a linear function as the upper limit by a straight line through some benchmark dose and the origin. However, one can still find a straight line serving as the upper limit of the effects by using (a) the gradient at the origin or (b) a line through upper boundaries of an appropriate benchmark dose.

Most such deviant results can be explained by saturation effects in the pharmacokinetic process or by exceeding the 50% effect, where most of the dose–effect curves flatten to a sigmoid curve. Under such circumstances the question should be, whether the extrapolation interval has been stretched too far from zero and how it could be shortened.

Additivity

The additivity of risks within Equation (5.27) will be needed, when separately evaluated risks are summarized to give an overall risk in a population.

$$P(d_1, d_2 \ldots, d_n) = P(d_1) + P(d_2) + \ldots + P(d_n) \tag{5.27}$$

At the extremely low dose level for environment toxic substances this practically leads to an independent effect of small single contaminations, a situation which surely exists near the origin in a mathematical sense. The question is, whether the field of relevant doses as a whole is close enough to zero, to ensure this condition works. This seems to be the case with those carcinogens that have been evaluated up to now. Of course this does not mean that the interaction with massive other contamination such as smoking, stress, or certain working factors can be neglected. As far as there other sources are concerned, complementary reconsiderations are necessary.

Proportionality

The conclusion from the model presentation in Equation (5.4) is that the relative risks are constant for various doses over all observation times, inrespective of age groups. Often the so-called Cox model is used in epidemiology, which requires these assumptions to estimate the relative risks (Cox, 1972). When transferring results, for example from work studies to environment situations, one has to consider that for work contamination these assumptions are normally not true. Typically, one rather finds a temporal course of the relative risk between case and control group computed on the basis of the same age structure, as shown in Figure 5.5. One has to pay attention to a sufficiently extended observation time for the work study, so that the relative risk, which is connected to the contamination at work, can be correctly evaluated.

Accumulation of the Dose

The usual transformation of job-related data to life-long contamination problems, as being typical for environmental risks, assumes that the effect relates only to the overall dose and not to the time pattern of the contamination. This condition has to be seen critically, out of theoretical reasons, and will probably be quite rare. Time-dependent risk evaluation for cohent studies with working exposure show an increase of risk during exposure, which may be associated with the cumulative exposure. In general, further

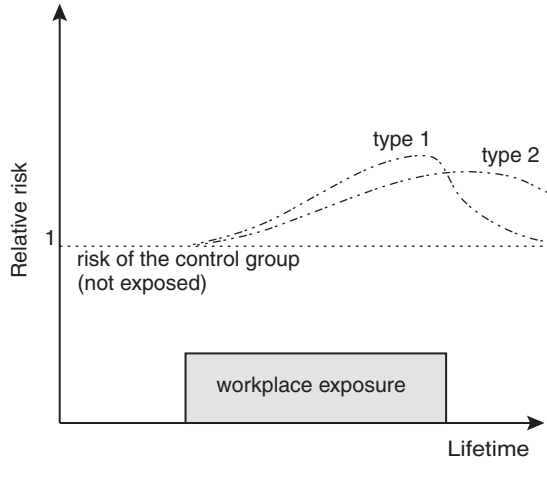


Figure 5.5 *Studies for evaluation of the relative risk at work.*

evaluation of these cohents does not present a stable level of risk after exposure but a decrease, which cannot be explained by the constant cumulated dose alone. The decrease may be steep (type 1 in the following figure) resulting in a risk at control-group level, or slowly flattening out and remaining above that level (type 2 in the following figure).

The problem of induction time, which for environmental cancer surely may last for decades, shows that the overall dose cannot be decisive alone. This problem deserves greater attention, considering the risk evaluation. Doing so, the questions of toxicokinetics and metabolism as well as repair mechanisms have to be considered. The results received so far for many of the carcinogens are too incomplete to solve this problem, such that there is need for a fundamentally new concept. First results indicate, however, that the evaluation on this basis of the total dose leads to upper limits, which should be corrected at least by using the concept of induction times.

5.1.8 Summary

The risk resulting from exposition to toxic substances, especially carcinogenic materials, can be evaluated by epidemiological studies or animal experiments. The human risk can be derived directly in epidemiological studies. However, in many cases adequate information is missing about the amount and duration of exposure for cancer cases as well as for controls. These data are known exactly in animal experiments, but for assessing the cancer risk for the human being an extrapolation from animal research to human being has to be calculated. For this, extensive knowledge of the biological effects and toxicokinetics of the material in animals and human beings is necessary. Epidemiological studies as well as the knowledge resulting from animal research mainly cover the risks resulting from comparably high exposure conditions. Therefore one has to extrapolate to the considerably lower concentrations found at work or in the environment by mathematical models.

With details about the risk at a certain exposure (unit risk) or the dose needed to induce a specific effect (benchmark dose), one may compare the carcinogenic power of various harmful substances, independently to the contamination situation. Carcinogens with a high unit risk are more dangerous (even and especially) at a very low contamination compared with those of a very low unit risk and a very flat course in the environmental interval of the dose–effect curve. For this reason the unit risk concept was originally developed.

If one sees the concept a little more in general, as a concept of estimating an upper boundary of the risk in the interval of environmental burden with contamination near the origin, one can use the linear function (dose multiplied by unit risk) to evaluate the specific situation between benchmark dose and the origin. This, for instance, is of greater interest for the establishment of priorities for protection and redevelopment than is a rather abstract comparison of substances.

Therefore, and because of its simple mathematical structure, the concept of straight lines through the origin is well suited as a base for further reconsiderations of risk management, but the limits and assumptions behind this method have to be taken into account.

For a specific evaluation of the risk different methods have been developed. A decision between them is difficult, as they have different advantages and disadvantages.

The EPA method, for example, is much more complex, and allows a consideration of important details of the empirical findings, which with the WHO method will inevitably not be considered. The method gives better results, which are also relative close to the real danger, if it is properly validated according to the proposal set out in the last two paragraphs of Section 5.1.2, above.

The uncertainty of specifying this linear approximation is rather significant when the data base is poor. As the EPA estimation depends on the confidence interval of the ascent, it tends to overestimate the risk considerably. If in this case the deviation is nonsignificant from the straight line through the origin, the WHO estimation should be preferred, because of its simple arrangement and its mostly smaller overestimation of the risk.

In case only one record of a summarized evaluation of contamination and relative risk is available, which regretfully happens to be the case with many epidemiological studies, the direct linear method has to be used.

In either case a biologically based cancer risk assessment should be the goal of further research and results within the frame of this research program will provide a better fit of mathematical models (some of them newly derived) and practical use of more sophisticated procedures based on such models.

References

V. J. Cogliano, E. G. Luebeck, and G. A. Zapponi (ed.), Perspectives on Biologically Based Cancer Risk Assessment, Nato. Challenges of a Modern Society, vol. 23, **1999**.

P. Armitage, and R. Doll, A Two-Stage Theory of Carcinogenesis in the Relation to Age Distribution of Human Cancer, *Br. J. Cancer*, **1961**, *11*, 161–169.

C. J. Portier, C. D. Sherman, and A. Kopp-Schneider, Multistage, Stochastic Models of the Cancer Process: A General Theory for Calculating Tumor Incidence, *Stochastic Environmental Research and Risk Assessment* **2000**, *14*, 173–179.

R. Cederlöf, R. Doll, B. Fowler, L. Friberg, N. Nelson, and V. Vouk (ed.), Risk Assessment Methodology and Epidemiological Evidence, *Environ. Health Prospect*, **1978**, *22*, 1–12.

D. R. Cox, Regression Models and Life Tables, *J. R. Statist. Soc. B*, **1972**, *34*, 187–220.

H. Druckrey and K. Kupfermüller, Quantitative Analyse der Krebsentstehung, *Z. Naturforsch Teit B*, **1948**, *3*, 254–266.

H. Druckrey, H. Schmähl, W. Dischler, and A. Schild bach, Quantitative Analyse der experimenteller Krebserzeugung, *Nauturwissenschafler*, **1962**, *48*, 217–228.

EPA (Environmental Protection Agency, USA), The Carcinogen Assessment Group's Final Risk Assessment on Arsenic, EPA-600/8-83-02IF, **1984**.

EPA (Environmental Protection Agency, USA), The Use of the Benchmark Dose Approach in Health Risk Assessment, Office of Research and Development, Washington DC, US EPA 630/R-94/007, **1995**.

P. Lee and J. O'Neill, The Effect of Time and Dose Applied on Tumor Incidence Rate in Benzopyrene Skin Painting Experiments, *Br. J. Cancer*, **1971**, *25*, 759–770.

J. W. F. Davidson, J. C. Parker, and R. P. Beliles, Biological Basis for Extrapolation Across Mammalian Species, *Regul. Toxicol. Pharmacol.*, **1986**, *6*, 211–237.

R. P. Rechard, Historical Relationship between Performance Assessment for Radioactive Waste Disposal and Other Types of Risk Assessment, *Risk Anal.*, **1999**, *19*, 763–807.

J. Wahrendorf, Quantitative Risiko-Abschätzung für ausgewählte Umweltkanzerogene, Umweltbundesamt-Berichte, **1990**, I.

WHO, Air Quality Guidelines for Europe, WHO Regional Publications. European Series, **1987**, 23.

Further Reading

H. Greim, Toxicologische Grundlagen der Risiko-Abschätzungen, VDI-Bericht, Krebserzeugende Stoffe in der Umwelt, 1991.
J. Timm, Mathematische Grundlagen, Modelle und Grenzen zur Risiko-Abschätzung, VDI-Berichte: Krebserzeugende Stoffe in der Umwelt, **1994**.

5.2 Regulations Regarding Chemicals and Radionuclides in the Environment, Workplace, Consumer Products, Foods, and Pharmaceuticals

Dennis J. Paustenbach and Pearl Moy

5.2.1 Introduction

The United States was among the first countries to establish a myriad of legislative and regulatory initiatives intended to control the release of chemicals to the ambient and workplace environment. Beginning in about 1970, the first of many of the modern era initiatives was promulgated. Advancements in science and technology played a major role in necessitating and evaluating these regulations. From that point forward, literally thousands of regulations were introduced in order to reduce exposure of humans and wildlife to various industrial chemicals. The significant changes in the regulatory environment in the United States regarding industrial chemicals are illustrated in Figure 5.6. This chapter presents a brief description of the various major initiatives that dictate the manufacture, use, and disposal of various chemicals in that country. Table 5.1 summarizes these major initiatives.

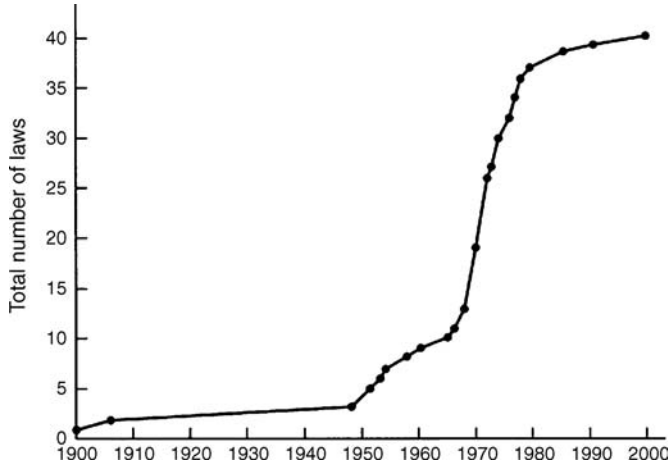


Figure 5.6 *Major legislative initiatives in the United States which dealt with the manufacture, use, transportation, sale, or disposal of hazardous materials. [Reprinted from Paustenbach. Copyright (2002), with permission from John Wiley & Sons].*

Table 5.1 *Summary of major US Regulations on occupational and environmental health.*

Law	Date	Agency	Scope
Occupational Health			
Occupational Safety and Health Act	1970	OSHA	Develop worker safety guildelines, Permissible Exposure Limits (PELs)
Food, Drugs, Cosmetics			
Food, Drug, and Cosmetic Act	1938	FDA	Protect foods from harmful additives and pesticides; regulate drugs and medical devices, including veterinary drugs
Food Quality Protection Act	1996	EPA	Lowered allowable level of residual pesticides on crops
Environment			
National Environmental Policy Act	1969	EPA	General environmental protection, requires environmental impact statements (EIS)
Clean Air Act	1970	EPA	Addresses public exposure to airborne contaminants; develop National Ambient Air Quality Standards (NAAQS)
Clean Water Act	1977	EPA	Addresses water pollution; restricts pollution discharge into waters and streams
Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA)	1980	EPA	Clean-up of abandoned contaminated sites; establishes financial responsibility for clean-up
Toxic Substances Control Act	1976	EPA	Develop tracking system for hazardous chemicals
Resource Conservation and Recovery Act	1976	EPA	Regulate chemicals 'cradle-to-grave' from manufacture, transportation, treatment, storage and disposal; also regulates nonhazardous waste
Federal Insecticide Fungicide Rodenticide Act	1972	EPA	Regulate use of pesticides through licensing and registration
Consumer Products			
Consumer Product Safety Act	1972	CPSC	Regulate and ensure safety of consumer products
Radionuclides			
Atomic Energy Act	1946	AEC, NRC (1975)	Assure safe handling and management of radioactive materials and facilities

5.2.2 Occupational Health Regulations

The development of early health guidelines on toxic substances in the workplace was a fragmented process involving private organizations. Early guidelines, for example, were written by a combination of the National Research Council (NRC), an independent council of scientists, the American National Standards Institute (ANSI), a conglomerate

of government, industry, and private individuals that develops voluntary standards, the National Safety Council (NSC), and the American Conference of Governmental Industrial Hygienists (ACGIH). Established in 1941, the ACGIH comprises hygienists and toxicologists employed in government or academia who had the objective of protecting workers.

The US Occupational Safety and Health Administration (OSHA), together with the National Institute for Occupational Safety and Health (NIOSH), are responsible for developing and enforcing regulations on workplace exposures. Occupational Exposure Limits, or OELs, are set by these two organizations. OELs can be Recommended Exposure Limits (RELs) prepared by NIOSH, which are based solely on health considerations and are used to prepare proposed regulations, and Permissible Exposure Limits (PELs), which are promulgated by OSHA, which are usually greater than the RELs and incorporate technical and economic considerations.

Occupational Safety and Health Administration (OSHA)

Established in 1970, OSHA is responsible for developing and enforcing workplace safety and health regulations. The agency's mission is to enssure the safety and health of America's workers by setting and enforcing standards, providing training, outreach, and education, establishing partnerships, and encouraging continual improvement in workplace safety and health.

On December 29, 1970, President Richard M. Nixon signed the Occupational Safety and Health Act. This legislation created both the Occupational Safety and Health Administration (OSHA) and the National Institute for Occupational Safety and Health (NIOSH). It was the first federal act which mandated that every state and nearly every business enterprise meet certain minimal standards with respect to safety and health. Prior to this time, regulations regarding the healthfulness of the workplace were promulgated by each state or regional governmental organization. OSHA officers had the responsibility for conducting inspections of the workplace and they could issue fines, as well as recommend criminal penalties for company officials who failed to comply with the regulations.

Occupational Safety and Health Act Congress passed the Occupational and Safety Health (OSH) Act to ensure worker and workplace safety. Their goal was to make sure employers provide their workers with a place of employment free from recognized safety and health hazards, such as exposure to toxic chemicals, excessive noise levels, nonionizing radiation, mechanical dangers, safety hazards (falls, electrocution, slips, and ergonomic stressors), heat or cold stress, or insanitary conditions.

The goal of the Act was to assure safe and healthful working conditions for working men and women by:

- AUTHORIZING enforcement of the standards developed under the Act;
- Assisting and encouraging the States in their efforts to ensure safe and healthful working conditions;

- providing for research, information, education, and training in the field of occupational safety and health.

In order to establish standards for workplace health and safety, the Act also created the National Institute for Occupational Safety and Health (NIOSH) as the research institution for the OSHA. OSHA is a division of the US Department of Labor that oversees the administration of the Act and enforces the various standards in all 50 states.

National Institute for Occupational Safety and Health (NIOSH)

Formed through the Occupational Safety and Health Act of 1970, the National Institute for Occupational Safety and Health (NIOSH) is the federal agency responsible for conducting research and making recommendations for the prevention of work-related injury and illness.

NIOSH is an agency established to help ensure safe and healthful working conditions for working men and women by providing research, information, education, and training in the field of occupational safety and health. NIOSH is part of the Centers for Disease Control (CDC) in the Department of Health and Human Services (DHHS). It is generally considered the agency which focuses on conducting research which will inform OSHA regarding the kind of standards that need to be promulgated in order to protect workers. It was intentionally placed in a separate division of the government to OSHA (which is in the Department of Labor) in an attempt to ensure objectivity in its recommendations.

Information pertaining to the responsibilities of NIOSH are found in Section 22 of the Occupational Safety and Health Act of 1970 (29 CFR § 671). The Institute is authorized to:

- Develop recommendations for occupational safety and health standards;
- Perform all functions of the Secretary of Health and Human Services under Sections 20 and 21 of the Act:
  - Conduct Research on Worker Safety and Health;
  - Conduct Training and Employee Education;
- Develop information on safe levels of exposure to toxic materials and harmful physical agents and substances;
- Conduct research on new safety and health problems;
- Conduct on-site investigations (Health Hazard Evaluations) to determine the toxicity of materials used in workplaces (42 CFR Parts 85 and 85a); and
- Fund research by other agencies or private organizations through grants, contracts, and other arrangements.

The Federal Mine Safety and Health Amendments Act of 1977 delegated additional authority to NIOSH for coal mine health research. The mine health and safety law authorized NIOSH to:

- Develop recommendations for mine health standards for the Mine Safety and Health Administration (MSHA);
- Administer a medical surveillance program for miners, including chest X-rays to detect pneumoconiosis (black lung disease) in coal miners;

- conduct on-site investigations in mines similar to those authorized for general industry under the OSH Act; and
- test and certify personal protective equipment and hazard-measurement instruments.

NIOSH provides national and world leadership to prevent work-related illness, injury, disability, and death by gathering information, conducting scientific research, and translating the knowledge gained into products and services. In order to help promote a safe and healthful workplace, NIOSH has issued literally hundreds of guidance documents for the public as well as for occupational health professionals. For most of the past 30 years, these documents have been relied upon by agencies and professionals around the world in their attempts to identify the appropriate workplace regulations or standards

5.2.3 Food and Drug Regulation

Amid public outcry due to health hazards from insanitary food processing and misused food additives, the Pure Food and Drugs Act was adopted in 1906. The act, administered by the Food and Drug Administration (FDA), has undergone a number of amendments, but remains the primary regulation for food and drug safety in the US today.

Food and Drug Administration (FDA)

The FDA is responsible for protecting the public health by ensuring the safety, efficacy, and security of human and veterinary drugs, biological products, medical devices, the food supply, cosmetics, and products that emit radiation.

Food and drugs are regulated by the Food and Drug Administration (FDA). The history of the US Food and Drug Administration dates back to 1862 when President Lincoln appointed a chemist, Charles M. Wetherill, to serve in the new Department of Agriculture (USDA). This was the beginning of the Bureau of Chemistry, the predecessor of the Food and Drug Administration. In 1927, The Bureau of Chemistry was reorganized into two separate entities. Regulatory functions were located in the Food, Drug, and Insecticide Administration, and nonregulatory research was located in the Bureau of Chemistry and Soils. In 1930, the name of the Food, Drug, and Insecticide Administration was shortened to Food and Drug Administration (FDA).

The United States Food and Drug Administration (FDA) regulates all food and food-related products, except commercially processed egg products, and meat and poultry products, which are regulated by the United States Department of Agriculture's Food Safety and Inspection Service (FSIS). Fruits, vegetables, and other plants are regulated by the Department's Animal and Plant Health Inspection Service (APHIS) to prevent intrusion of plant diseases and pests into the United States. The Agricultural Marketing Service (AMS) of the USDA is responsible for the voluntary grading of fruits and vegetables.

The FDA is responsible for advancing the public health by helping:

- facilitate innovations that make medicines and foods more effective, safer, and more affordable; and
- inform the public with accurate, science-based information on medicines and foods.

Food, Drug, and Cosmetic Act On 30 June 1906, President Theodore Roosevelt signed the Food and Drugs Act, also known simply as the Wiley Act. This legislation, which the Bureau of Chemistry was charged to administer, prohibited the interstate transport of unlawful food and drugs under penalty of seizure of the questionable products and/or prosecution of the responsible parties. The basis of the law rested on the regulation of product labeling rather than pre-market approval. Drugs or pharmaceuticals, defined in accordance with the standards of strength, quality, and purity in the *United States Pharmacopoeia* and the *National Formulary,* could not be sold in any other condition unless the specific variations from the applicable standards were plainly stated on the label. The law prohibited the addition of any ingredients that would substitute for the food, conceal damage, pose a health hazard, or constitute a filthy or decomposed substance.

The Food, Drug, and Cosmetic Act was signed on June 25, 1938 by President Franklin Delano Roosevelt. This new law brought cosmetics and medical devices under control, and it required that drugs be labeled with adequate directions for safe use. Moreover, it mandated pre-market approval of all new drugs, such that a manufacturer would have to prove to FDA that a drug were safe before it could be sold. It prohibited false therapeutic claims for drugs, although a separate law granted the Federal Trade Commission (FTC) jurisdiction over drug advertising. The act also corrected abuses in food packaging and quality, and it mandated legally enforceable food standards. Tolerances, or the safe concentration of certain chemical contaminants, for certain crops, foods, or drugs were addressed. The law formally authorized factory inspections, and it added injunctions to the enforcement tools at the agency's disposal.

Cosmetics and medical devices, which the Post Office Department and the Federal Trade Commission had limited oversight on prior to 1938, also came under FDA authority after 1938. While pre-market approval did not apply to devices, in every other sense the new law equated them to drugs for regulatory purposes. As the FDA had to deal with both increasing fraudulent medical devices and the expansion of medical technology in and after World War II, Congress passed the 1962 drug amendments for medical devices.

The 1938 act required colors to be certified as harmless and suitable by the FDA for their use in cosmetics. Further amendments to the Federal FD&C Act to included a series of laws addressing food additives in 1958 and color additives in 1960. These laws gave the FDA tighter control over the growing list of chemicals entering the food supply and gave manufacturers the responsibility to establish the safety of their products. The laws included a provision, known as the 'Delaney Clause,' which established that no food or color additive could be deemed safe–or given FDA approval–if found to cause cancer in humans or animals.

In 1962, the Delaney Clause was modified to permit the FDA to approve the use of carcinogenic compounds in food-producing animals if certain conditions were met. The modification to the Delaney Clause is known as the 'Diethylstilbestrol (DES) Proviso,' named for a hormone approved in 1954 to promote growth in cattle and sheep. DES had also been approved much earlier in humans. It was thought to prevent miscarriages in women, but the hormone later was linked to vaginal cancers in the daughters of women who were treated with the drug during pregnancy. Under the DES Proviso, the FDA could approve a carcinogen for food animal use if the concentration of any residue remaining in

the edible tissues was so low that it presented an insignificant risk of cancer to consumers.

The 1960 color amendments strengthened the safety requirement for color additives, necessitating additional testing for many existing cosmetics to meet the new safety standard. The FDA attempted to interpret the new law as applying to every ingredient of color-imparting products, such as lipstick and rouge, but the courts rebuffed this proposal.

Another agency responsibility, veterinary medicine, had been stipulated since the 1906 act; foods included animal feed, and drugs included veterinary pharmaceuticals. Likewise, animal drugs were included in the provisions for new drugs under the 1938 law and the 1962 drug amendments. However, the Food Additives Amendment of 1958 had an impact too, since drugs used in animal feed were also considered additives - and thus subject to the provisions of the food additive petition process.

5.2.4 Environmental Regulation

Environmental regulations in the US began with the adoption of the National Environmental Policy Act (NEPA) in 1969. The act led to the creation of a Council on Environmental Quality (CEQ) within the Executive office of the President, which was responsible for developing and coordinating the nation's environmental programs and policies. The nation's primary environmental regulatory agency is the Environmental Protection Agency (EPA), which was established in 1970.

Environmental Protection Agency (EPA)

> *The U.S. Environmental Protection Agency was established on December 2, 1970. EPA's mission is to protect human health and to safeguard the natural environment - air, water, and land. The agency leads the nation's environmental science, research, education, and assessment efforts.*

In July of 1970, the White House and Congress worked together to establish the EPA in response to the growing public demand for cleaner water, air, and land. Prior to the establishment of the EPA, the federal government was not structured to make a coordinated effort to minimize pollutants that harm human health and the environment. In 2006, the US EPA employs nearly 20 000 people across the country, including the headquarters offices in Washington, DC, 10 regional offices, and more than a dozen labs. The EPA works to:

- *Develop and enforce regulations.* EPA works to develop and enforce regulations that implement environmental laws enacted by Congress. EPA is responsible for researching and setting national standards for a variety of environmental programs, and delegates to States and Tribes the responsibility for issuing permits and for monitoring and enforcing compliance. Where national standards are not met, EPA can issue sanctions and take other steps to assist the States and Tribes in reaching the desired levels of environmental quality.

- *Offer financial assistance:* In recent years, between 40% and 50% of EPA's enacted budgets have provided direct support through grants to state environmental programs. EPA grants to States, nonprofits, and educational institutions support high-quality research that will improve the scientific basis for decisions on national environmental issues and help EPA achieve its goals.
  - EPA provides research grants and graduate fellowships.
  - The Agency supports environmental education projects that enhance the public's awareness, knowledge, and skills to make informed decisions that affect environmental quality.
  - The Agency also offers information for state and local governments and small businesses on financing environmental services and projects.
  - EPA also provides other financial assistance through state environmental programs.
- *Perform environmental research:* At laboratories located throughout the nation, the Agency works to assess environmental conditions and to identify, understand, and solve current and future environmental problems; integrate the work of scientific partners such as nations, private sector organizations, academia and other agencies; and provide leadership in addressing emerging environmental issues and in advancing the science and technology of risk assessment and risk management.
- *Sponsor voluntary partnerships and programs:* The Agency works through its headquarters and regional offices with over 10 000 industries, businesses, nonprofit organizations, and state and local governments, on over 40 voluntary pollution-prevention programs and energy-conservation efforts. Partners set voluntary pollution-management goals; examples include conserving water and energy, minimizing greenhouse gases, slashing toxic emissions, re-using solid waste, controlling indoor air pollution, and getting a handle on pesticide risks. In return, EPA provides incentives like vital public recognition and access to emerging information.
- *Further environmental education:* EPA advances educational efforts to develop an environmentally conscious and responsible public, and to inspire personal responsibility in caring for the environment.
- *Publish information:* Through written materials and its Web site, EPA informs the public about its activities.

Clean Air Act (CAA) Adopted in 1970, the Clean Air Act is the comprehensive Federal law that regulates air emissions from area, stationary, and mobile sources. This law authorizes the US Environmental Protection Agency to establish National Ambient Air Quality Standards (NAAQS) to protect public health and the environment.

The goal of the Act was to set and achieve NAAQS in every state by 1975. The setting of maximum pollutant standards was coupled with directing the states to develop state implementation plans (SIPs) applicable to appropriate industrial sources in the state. The Act was amended in 1977 primarily to set new goals (dates) for achieving attainment of NAAQS since many areas of the country had failed to meet the deadlines. Significant amendments were made to the act in 1990 to meet unaddressed or insufficiently addressed problems such as acid rain, ground-level ozone, stratospheric ozone depletion, and air toxics.

Clean Water Act (CWA) Growing public awareness and concern for controlling water pollution led to enactment of the Federal Water Pollution Control Act Amendments of 1972. As amended in 1977, this law became commonly known as the Clean Water Act. The Act established the basic structure for regulating discharges of pollutants into the waters of the United States. It gave EPA the authority to implement pollution control programs such as setting wastewater standards for industry. The CWA also continued requirements to set water quality standards for all contaminants in surface waters. The Act made it unlawful for any person to discharge any pollutant from a point source into navigable waters, unless a permit was obtained under its provisions. It also funded the construction of sewage-treatment plants under the construction grants program and recognized the need for planning to address the critical problems posed by nonpoint-source pollution. Subsequent enactments modified some of the earlier Clean Water Act provisions. Revisions in 1981 streamlined the municipal construction grants process, improving the capabilities of treatment plants built under the program. Changes in 1987 phased out the construction grants program, replacing it with the State Water Pollution Control Revolving Fund, more commonly known as the Clean Water State Revolving Fund. This new funding strategy addressed water quality needs by building on EPA–State partnerships.

Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) The Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), commonly known as Superfund, was enacted by Congress on December 11, 1980. This law created a tax on the chemical and petroleum industries and provided broad Federal authority to respond directly to releases or threatened releases of hazardous substances that may endanger public health or the environment. During the initial five years after passage (about 1986), $1.6 billion was collected and the tax went to a trust fund for cleaning up abandoned or uncontrolled hazardous waste sites. The Superfund or CERCLA:

- Established prohibitions and requirements concerning closed and abandoned hazardous waste sites;
- Provided for liability of persons responsible for releases of hazardous waste at these sites; and
- Established a trust fund to provide for cleanup when no responsible party could be identified.

The law authorizes two kinds of response actions:

- Short-term removals, where actions may be taken to address releases or threatened releases requiring prompt response.
- Long-term remedial response actions, that permanently and significantly reduce the dangers associated with releases or threats of releases of hazardous substances that are serious, but not immediately life threatening. These actions can be conducted only at sites listed on EPA's National Priorities List (NPL).

CERCLA also enabled the revision of the National Contingency Plan (NCP). The NCP provided the guidelines and procedures needed to respond to releases and threatened releases of hazardous substances, pollutants, or contaminants. The NCP also established

the NPL. CERCLA was amended by the Superfund Amendments and Reauthorization Act (SARA) on October 17, 1986.

The Superfund Amendments and Reauthorization Act (SARA) amended the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) on October 17, 1986. SARA reflected EPA's experience in administering the complex Superfund program during its first six years and made several important changes and additions to the program.

The reauthorization of Superfund (SARA):

- stressed the importance of permanent remedies and innovative treatment technologies in cleaning up hazardous waste sites;
- required Superfund actions to consider the standards and requirements found in other State and Federal environmental laws and regulations;
- provided new enforcement authorities and settlement tools;
- increased State involvement in every phase of the Superfund program;
- increased the focus on human health problems posed by hazardous waste sites;
- encouraged greater citizen participation in making decisions on how sites should be cleaned up; and
- increased the size of the trust fund to $8.5 billion.

SARA also required EPA to revise the Hazard-Ranking System (HRS) to ensure that it accurately assessed the relative degree of risk to human health and the environment posed by uncontrolled hazardous waste sites that may be placed on the National Priorities List (NPL). Superfund has not been reauthorized since 1986 even though it has been brought before the US Congress on a number of occasions.

Toxic Substances Control Act (TSCA) The Toxic Substances Control Act (TSCA) (pronounced 'tosca') was enacted by Congress in 1976 to give EPA the ability to track the 75 000 industrial chemicals currently produced or imported into the United States. EPA repeatedly screens these chemicals and can require reporting or testing of those that may pose an environmental or human-health hazard. EPA can ban the manufacture and import of those chemicals that pose an unreasonable risk.

Also, EPA has mechanisms in place to track the thousands of new chemicals that industry develops each year with either unknown or dangerous characteristics. EPA then can control these chemicals as necessary to protect human health and the environment.

Resource Conservation and Recovery Act (RCRA) The Resource Conservation and Recovery Act or RCRA (pronounced 'rick-rah'), was adopted in 1976 and gave EPA the authority to control hazardous waste from the 'cradle-to-grave.' The intent was to regulate all chemicals from the time of manufacture, transportation, treatment, storage, and disposal of hazardous waste. RCRA also set forth a framework for the management of nonhazardous wastes.

The 1986 amendments to RCRA enabled EPA to address environmental problems that could result from underground tanks storing petroleum and other hazardous substances. RCRA focuses only on active and future facilities and does not address abandoned or historical sites (see CERCLA).

HSWA (pronounced 'hiss-wa') - The Federal Hazardous and Solid Waste Amendments are the 1984 amendments to RCRA that required the phasing out of land disposal of

hazardous waste. Some of the other mandates of this strict law include increased enforcement authority for EPA, more stringent hazardous waste management standards, and a comprehensive underground storage tank program.

Federal Insecticide, Fungicide Rodenticide Act (FIFRA) The primary focus of FIFRA was to provide federal control of pesticide distribution, sale, and use. EPA was given authority under FIFRA not only to study the consequences of pesticide usage but also to require users (farmers, utility companies, and others) to register when purchasing pesticides.

Through later amendments to the law, users also must take exams for certification as applicators of pesticides. All pesticides used in the US must be registered (licensed) by EPA. Registration ensures that pesticides will be properly labeled and that if in accordance with specifications will not cause unreasonable harm to the environment.

Some key elements of FIFRA include:

- IS a product-licensing statute; pesticide products must obtain an EPA registration before manufacture, transport, and sale;
- Registration based on a risk/benefit standard;
- strong authority to require data - authority to issue Data Call-ins;
- Ability to regulate pesticide use through labeling, packaging, composition, and disposal;
- Emergency exemption authority - permits approval of unregistered uses of registered products on a time-limited basis;
- Ability to suspend or cancel a product's registration: appeals process, adjudicatory functions, etc.

Food Quality Protection Act (FQPA) With the enactment of the Food Quality Protection Act (FQPA) of 1996, Congress presented EPA with the enormous challenge of implementing the most comprehensive and historic overhaul of the Nation's pesticide and food safety laws in decades. The FQPA amended the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) and the Federal Food, Drug, and Cosmetic Act (FFDCA) by fundamentally changing the way EPA regulates pesticides.

Some of the major requirements include stricter safety standards, especially for infants and children, and a complete reassessment of all existing pesticide tolerances.

In 1996, Congress unanimously passed landmark pesticide food safety legislation supported by the Administration and a broad coalition of environmental, public health, agricultural, and industry groups. President Clinton promptly signed the bill on August 3, 1996, and the Food Quality Protection Act of 1996 became law. One of the key provisions of this Act was that the residual concentrations of a large fraction of pesticides on crops was to be substantially lowered; often by at least 10- to 100-fold.

EPA regulates pesticides under two major federal statutes. Under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), EPA registers pesticides for use in the United States and prescribes labeling and other regulatory requirements to prevent unreasonable adverse effects on human health or the environment. For the Federal Food, Drug, and Cosmetic Act (FFDCA), EPA establishes tolerances (maximum legally permissible levels) for pesticide residues in food. The FQPA streamlined these two regulations into one law that:

- Mandates a single, health-based standard for all pesticides in all foods;
- Provides special protections for infants and children;
- Expedites approval of safer pesticides;
- creates incentives for the development and maintenance of effective crop protection tools for American farmers; and
- Requires periodic re-evaluation of pesticide registrations and tolerances

5.2.5 Consumer Product Regulation

Consumer Product Safety Commission (CPSC)

The US Consumer Product Safety Commission, created in 1972 under the Consumer Product Safety Act, is charged with protecting the public from unreasonable risks of serious injury or death from more than 15 000 types of consumer products under the agency's jurisdiction.

Deaths, injuries, and property damage from consumer product incidents cost the nation more than $700 billion annually. The goal of the CPSC is to protect consumers and families from products that pose a fire, electrical, chemical, or mechanical hazard or can injure children. The CPSC's work to ensure the safety of consumer products - such as toys, cribs, power tools, cigarette lighters, and household chemicals - contributed significantly to the 30% decline in the rate of deaths and injuries associated with consumer products over the past 30 years.

The CPSC is involved in:

- Developing voluntary standards with industry;
- Issuing and enforcing mandatory standards or banning consumer products if no feasible standard would adequately protect the public;
- obtaining the recall of products or arranging for their repair;
- conducting research on potential product hazards;
- Informing and educating consumers through the media, State and local governments, private organizations, and by responding to consumer inquiries. These include local and national media coverage, publication of numerous booklets and product alerts, a web site, a telephone Hotline, the National Injury Information Clearinghouse, CPSC's Public Information Center, and responses to Freedom of Information Act (FOIA) requests.

Consumer Product Safety Act The CPSA, enacted in 1972, is CPSC's umbrella statute. It established the agency, defined its basic authority, and provided that when the CPSC finds an unreasonable risk of injury associated with a consumer product it can develop a standard to reduce or eliminate the risk. The CPSA also provides the authority to ban a product if there is no feasible standard, and it gives CPSC authority to pursue recalls for products that present a substantial product hazard. (Generally excluded from CPSC's jurisdiction are food, drugs, cosmetics, medical devices, tobacco products, firearms and ammunition, motor vehicles, pesticides, aircraft, boats and fixed-site amusement rides.)

5.2.6 Radionuclides Regulation

Nuclear Regulatory Commission

Created in 1974, the NRC's primary mission is to protect the public health and safety, and the environment, from the effects of radiation from nuclear reactors, materials, and waste facilities.

The NRC regulates nuclear materials and facilities to promote the common defense and security. The commission carries out its mission by conducting the following activities:

- Commission Direction-Setting and Policymaking - policy formulation, rulemaking, and adjudication oversight activities performed by NRC's five-member Commission;
- Radiation Protection - provide information about radiation and ensure protection of the public and radiation workers;
- Regulation - rulemaking, oversight, licensing and certification;
- Emergency Preparedness and Response - integration of the NRC emergency and preparedness programs and response to a wide spectrum of radiological emergencies;
- Nuclear Security and Safeguards - regulating licensees' accounting systems for special nuclear and source materials and security programs and contingency plans for dealing with threats, thefts, and sabotage relating to special nuclear material, high-level radioactive wastes, nuclear facilities, and other radioactive materials and activities that the NRC regulates;
- Public Affairs - interactions with the media and the public;
- Congressional Affairs - interactions with Congress;
- State and Tribal Programs - cooperative activities and interactions with Federal, State, and local governments, interstate organizations, and Indian Tribes;
- International Programs - cooperative activities with other governments and the international nuclear regulatory community and licensing for nuclear imports and exports

The NRC is headed by a five-member Commission. The President designates one member to serve as Chairman and official spokesperson. The Commission as a whole formulates policies and regulations governing nuclear reactor and materials safety, issues orders to licensees, and adjudicates legal matters brought before it. The Executive Director for Operations (EDO) carries out the policies and decisions of the Commission and directs the activities of the program offices.

The offices reporting to the EDO ensure that the commercial use of nuclear materials in the United States is safely conducted. As part of the regulatory process, the four regional offices conduct inspection, enforcement, and emergency response programs for licensees within their borders.

Atomic Energy Act The Atomic Energy Act of 1946, which was legislated when tensions with the Soviet Union were developing into the cold war, acknowledged the potential peaceful benefits of atomic power. The act established the five-member Atomic Energy Commission (AEC) to manage the nation's atomic energy programs.

In 1954, Congress passed new legislation that for the first time permitted the wide use of atomic energy for peaceful purposes. The 1954 Atomic Energy Act redefined the atomic energy program by ending the government monopoly on technical data and making the growth of a private commercial nuclear industry an urgent national goal. It instructed the agency to prepare regulations that would protect public health and safety from radiation hazards. Thus, the 1954 act assigned the AEC three major roles: to continue its weapons program, to promote the private use of atomic energy for peaceful applications, and to protect public health and safety from the hazards of commercial nuclear power.

In 1974, Congress divided the AEC into the Energy Research and Development Administration and the Nuclear Regulatory Commission. The Energy Reorganization Act, coupled with the 1954 Atomic Energy Act, constituted the statutory basis for the NRC. The new agency inherited a mixed legacy from its predecessor, marked both by 20 years of conscientious regulation and by unresolved safety questions, substantial anti-nuclear activism, and growing public doubts about nuclear power.

By 1974, the AEC's regulatory programs had come under such strong attack that Congress decided to abolish the agency. Supporters and critics of nuclear power agreed that the promotional and regulatory duties of the AEC should be assigned to different agencies. The Energy Reorganization Act of 1974 created the Nuclear Regulatory Commission; it began operations on January 19, 1975.

The NRC (like the AEC before it) focused its attention on several broad issues that were essential to protecting public health and safety. In many ways, the NRC carried on the legacy inherited from the AEC. It performed the same licensing and rulemaking functions that the regulatory staff had discharged for two decades. It also assumed some new administrative and regulatory duties. The NRC, unlike the AEC's regulatory staff, was the final arbiter of regulatory issues; its judgment on safety questions was less susceptible to being overridden by developmental priorities.

The AEC and the NRC published standards that were intended to provide an ample margin of safety from radiation that was generated by the activities of its licensees. The radiation standards embodied available scientific information and the judgment of leading authorities in the field.

Although reactor safety issues received the majority of public notice, the NRC also devotes substantial resources to a variety of complex questions in the area of nuclear materials safety and safeguards, such as the protection of nuclear materials from theft or diversion and the safety of depositories for the disposition of high-level and low-level radioactive waste.

5.2.7 Governmental Agencies on Human Health

Agency for Toxic Substances and Disease Registry (ATSDR)

In 1980, Congress created the Agency for Toxic Substances and Disease Registry (ATSDR) to implement the health-related sections of laws that protect the public from hazardous wastes and environmental spills of hazardous substances.

As mentioned previously, The Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (CERCLA) provided the Congressional mandate to

remove or clean up abandoned and inactive hazardous waste sites and to provide federal assistance in toxic emergencies. As the lead Agency within the Public Health Service for implementing the health-related provisions of CERCLA, ATSDR is charged under the Superfund Act to assess the presence and nature of health hazards at specific Superfund sites, to help prevent or reduce further exposure and the illnesses that result from such exposures, and to expand the knowledge base about health effects from exposure to hazardous substances.

In 1984, amendments to the Resource Conservation and Recovery Act of 1976 (RCRA), which provides for the management of legitimate hazardous waste storage or destruction facilities, authorized ATSDR to conduct public health assessments at these sites, when requested by the Environmental Protection Agency (EPA), States, or individuals. ATSDR was also authorized to assist EPA in determining which substances should be regulated and the levels at which substances may pose a threat to human health.

With the passage of the Superfund Amendments and Reauthorization Act of 1986 (SARA), ATSDR received additional responsibilities in environmental public health. This act broadened ATSDR's responsibilities in the areas of public health assessments, establishment and maintenance of toxicologic databases, information dissemination, and medical education.

Centers for Disease Control (CDC)

The Centers for Disease Control and Prevention (CDC), established in 1946, is one of the 13 major operating components of the Department of Health and Human Services (HHS), which is the principal agency in the United States government for protecting the health and safety of all Americans and for providing essential human services.

Since the CDC was founded in 1946 to help control malaria, CDC has remained at the forefront of public health efforts to prevent and control infectious and chronic diseases, injuries, workplace hazards, disabilities, and environmental health threats. Today, CDC is globally recognized for conducting research and investigations and for its action-oriented approach. CDC applies research and findings to improve people's daily lives and responds to health emergencies - something that distinguishes CDC from its peer agencies. The CDC seeks to accomplish its mission by working with partners throughout the nation and the world to:

- Monitor health;
- Detect and investigate health problems;
- Conduct research to enhance prevention;
- Develop and advocate sound public health policies;
- Implement prevention strategies;
- Promote healthy behaviors;
- Foster safe and healthful environments;
- Provide leadership and training.

Those functions are the backbone of CDC's mission. Each of CDC's component organizations undertakes these activities in conducting its specific programs. The steps

needed to accomplish this mission are also based on scientific excellence, requiring well trained public health practitioners and leaders dedicated to high standards of quality and ethical practice.

5.2.8 Regulations in the European Union (EU) and United Kingdom (UK)

This section provides a brief summary of some major chemical regulations in the EU and UK. Of particular interest is the implementation of Registration, Evaluation, Authorization and Restriction of Chemicals, or REACH, in 2007. This new regulation significantly changed the use of chemicals in the EU. Prior to the adoption of REACH, European countries regulated chemicals by classifying them as 'new' or 'existing' chemicals. Existing chemicals were defined as those listed in the European Inventory of Existing Commercial Chemical Substances (EINECS) which listed over 100 000 that were on the European market between January 1971 and September 1981. All other chemicals were categorized as 'new'.

The REACH legislation removes the distinction between 'existing' and 'new' chemicals, and requires that all chemicals produced in quantities greater than 1 ton undergo an evaluation and registration process. Chemicals must be also be approved by the Chemicals Agency based in Helsinki, Finland. Table 5.2 summarizes some of the major health and environmental regulations in the EU and UK.

5.2.9 International Chemical Regulations

The Intergovernmental Forum on Chemical Safety (IFCS) and the Interorganization Program for the Sound Management of Chemicals (IOMC) are responsible for coordinating international chemical safety programs in cooperation with the UN Environment Program (UNEP), International Labor Organization (ILO), UN Food and Agriculture Organization (FAO), World Health Organization (WHO), UN Industrial Development Organization (UNIDO), UN Institute for Training and Research (UNITAR), and Organization for Economic Co-operation and Development (OECD). The IFCS was established to promote chemical safety and management of chemicals as set out in the 1992 Rio Earth Summit. The major initiatives resulting from the Earth Summit were:

- The expansion and acceleration of international chemical risk assessments;
- Harmonization of chemical classification and labeling;
- Information exchange on toxic chemicals and chemical risks;
- Establishment of risk-reduction programs;
- strengthening of national capabilities and capacities for management of chemicals; and
- Sttopping illegal international traffic in toxic and dangerous products.

The following is a brief summary of some of the major international environmental conventions held by the UN.

Montreal Protocol on Substances that Deplete the Ozone Layer

In 1987, the Montreal Protocol introduced measures to restrict the production and use of chemicals which damage the ozone layer.

Table 5.2 *Summary of some environmental regulations in the European Union and UK.*

Law	Effective Date	Scope
European Union Directives and Regulations		
Directive 67/548	1967	Established laws, regulations and administrative provisions relating to the classification, packaging, and labeling of dangerous substances
New Substances Directive 92/32	1979 (revised 1992)	Mandated that a new chemical can only be placed on the market if the manufacturer or importer submits a notification to a competent authority (the competent authority is a body or bodies nominated by Member States)
Existing Substances Regulation No.: 793/93/EEC	1993	Introduced a scheme for assessing the risks to human health of 'existing' chemicals (those that are listed in the European Inventory of Existing Commercial Chemical Substances [EINECS]) that were on the market from Jan. 1971 to Sept. 1981
Directive No.: 19994R1488EC	1994	Sets forth principles for the assessment of risks to man and the environment of existing substances in accordance with Council Regulation (EEC) No 793/93
Directive No.: 1998L0008EC	1998	Concerns the placement of biocidal products on the market.
Directive No.: 2003R0304EC	2003	Concerns the export and import of dangerous chemicals
Registration, Evaluation, Authorization and Restriction of Chemicals (REACH), Regulation No.: 1907/2006	2007	Removes distinction between 'existing' and 'new' substances. Requires registration of all chemical substances produced or imported above a total quantity of 1 tonne per year. Chemical producers are also required to evaluate chemicals for health and environmental impacts. REACH also creates a new Chemicals Agency, which is responsible for authorizing chemical usage.
UK Legislation		
Environmental Protection Act (EPA) Sections 140 and 142	1990	Section 142 grants regulations to obtain relevant information from manufactures, importers, or suppliers about specific chemicals for the purpose of assessing their potential for environmental and human health risks. Section 140 prohibits or restricts importation into the UK of substances that may cause harm to the environment or human health.

Basel Convention on the Control of Transboundary Movements of Hazardous Wastes and Their Disposal

The Basel Convention was adopted in 1989 to regulate the movement of and develop criteria for the safe management of hazardous wastes.

Rotterdam Convention on the Prior Informed Consent Procedure for Certain Hazardous Chemicals and Pesticides in International Trade

Adopted in 1998, the Rotterdam Convention aimed to address problems associated with chemical usage (pesticides and industrial chemicals) in developing countries that lacked adequate legislation and infrastructure to ensure safe use of chemicals. The convention promoted the exchange of information to affected parties, including required notification of bans and labeling.

Stockholm Convention on Persistent Organic Pollutants (POPs)

The Stockholm Convention, adopted in 2001, is a global treaty that aims to protect human health and the environment from persistent organic pollutants (POPs). POPs are chemicals that persist in the environment, become widely distributed, accumulate in the fatty tissue of organisms, and are toxic to humans and wildlife. POPs include pesticides such as mirex, dieldrin, DDT, and hexachlorobenzene, which are no longer used in the EU due to human health and the environmental concerns. Other POPs are dioxins, furans, and PCBs.

5.2.10 Summary

The regulation of chemicals and chemical exposures has increased significantly over the years. This chapter presented a brief overview of major environmental, human health, and occupational safety regulations in the United States and European Union. The development of new technologies and scientific knowledge plays an important role in how these regulations are created and evaluated. By sharing knowledge and experiences about the implementation of such regulations, governments can anticipate that their overall effectiveness in protecting human health and the environment will continue to improve in the years to come.

References

Paustenbach, D.J. (2002). Chapter 1. Primer on Human and Environmental Risk Assessment. In: Paustenbach D.J. (ed.) *Human and Ecological Risk Assessment Theory and Practice* (pp. 1–83), John Wiley & Sons, New York, NY.

For Further Reading

Rothstein, M.A. (1998). Occupational Safety and Health Law, 4th edn, West Group, St. Paul, MN.
Ashford, N.A. (2000). Chapter 9. Government Regulation of Occupational Health and Safety. In: Levy, B.S. and Wegman, D.H. (eds.) *Occupational Health. Recognizing and Preventing*

Work-Related Disease and Injury, 4th edn. (pp. 177–199), Lippincott Williams & Wilkins, Philadelphia, PA.

Mazuzan, G.T. and Walker, S.J. (1997). Controlling the Atom: the Beginnings of Nuclear Regulation, 1946–1962, US Nuclear Regulatory Commission, Washington, DC.

Walker, S.J. (1992). Containing the Atom: Nuclear Regulation in a Changing Environment, 1963–1971, University of California Press, Berkeley, CA.

Mackenthun, K.M. and Bregman, J.L. (eds.) (1992). Environmental Regulations Handbook, Lewis Publishers, Boca Raton, FL.

Bingham, E. and Grimsley, L.F. (2001). Chapter 8. Regulations and Guidelines in the Workplace. In: Bingham, E., Cohrrsen, B., and Powell, C.H. (eds.) Patty's Toxicology, Volume 1. 5th edn. (pp. 321–352), John Wiley & Sons, New York, NY.

Moeller, D.W. (1997). Chapter 13. Environmental Law, and Chapter 14: Standards. In: Moeller, D.W. Environmental Health, Revised Edn. (pp. 270–294, 295–318), Harvard University Press, Cambridge, MA.

Food Safety Council (June 1982). A Proposed Food Safety Evaluation Process. Final Report of Board of Trustees, Food Safety Council, Washington, DC.

OSHA (n.d.) OSHA Mission. Retrieved January 5, 2007 from http://www.osha.gov/oshinfo/mission.html

CDC, NIOSH (n.d.) About NIOSH. Retrieved January 5, 2007 from http://www.cdc.gov/niosh/about.html

USEPA (n.d.) Major Environmental Laws. Retrieved January 5, 2007 from http://www.epa.gov/epahome/laws.htm

ATSDR (n.d.) ATSDR Congressional Background and Mandates. Retrieved January 5, 2007 from http://www.atsdr.cdc.gov/congress.html

CPSC (n.d.) CPSC Overview. Retrieved January 5, 2007 from http://www.cpsc.gov/about/about.html

CDC (n.d.) Vision, Mission, Core Values, Pledge. Retrieved January 5, 2007 from http://www.cdc.gov/about/mission.htm

FDA. August 2005. Milestones in US Food and Drug Law History. Retrieved January 5, 2007 from http://www.fda.gov/opacom/backgrounders/miles.html

European Chemicals Bureau (n.d.) Legislation. Retrieved February 5, 2007 from http://ecb.jrc.it/legislation/

DEFRA (September 2002). The Government's Chemicals Strategy: Sustainable Production and Use of Chemicals - A Strategic Approach. Retrieved February 5, 2007 from http://www.defra.gov.uk/environment/chemicals/strategy/04.htm#3

UNEP (September 2003). The Hazardous Chemicals and Wastes Conventions. Retrieved February 5, 2007 from http://www.pops.int/documents/background/hcwc.pdf

6

Toxicity of Selected Chemicals

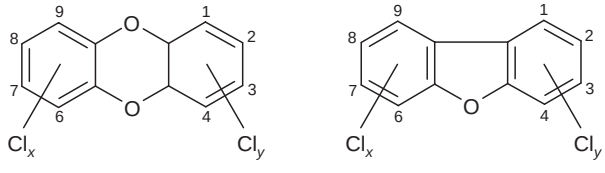
6.1 Persistent Polyhalogenated Aromatic Hydrocarbons

Kristian W. Fried and Karl K. Rozman

6.1.1 Introduction

Polyhalogenated aromatic hydrocarbons (PHAHs) are highly persistent organic compounds, which share general dispositional and toxicological properties due to structural similarities. The most prominent members of this class of toxicants are polychlorinated dibenzo-*p*-dioxins (PCDDs) and dibenzofurans (PCDFs), polychlorinated biphenyls (PCBs), as well as some pesticides, such as p,p'-dichlorodiphenyltrichloroethane (DDT) and hexachlorobenzene (HCB). These compounds are used as prototypes, representing the whole class of PHAHs. Brominated and mixed halogen congeners have similar physical and biological properties, resulting in comparable disposition and effects. PHAHs are ubiquitous environmental pollutants of mainly anthropogenic origin. Many are persistent and detrimental to the environment, as well as to humans, and have been listed with other persistent organic pollutants (POPs) as the ''Dirty Dozen'' (Table 6.1). Those listed have been banned or are being phased out in accordance with the Stockholm Convention of 2001, which was signed by 122 countries including the US. It became effective on May 17th, 2004, when it was ratified by 50 countries.

6.1.2 PCDDs and PCDFs



Toxicology and Risk Assessment: A Comprehensive Introduction Edited by H. Greim and R. Snyder

Table 6.1 *The Dirty Dozen, as identified by the Stockholm Convention. Almost half of the compounds are PHAHs (bold).*

Compound	Use	Structure
Aldrin	Insecticide	Chlorinated 1,4:5,8-dimethanonaphthalene derivative
Chlordane	Insecticide	Chlorinated 4,7-methanoindane derivative
DDT	**Insecticide**	**Chlorinated diphenylethane derivative**
Dieldrin	Insecticide	Chlorinated 1,4:5,8-dimethanonaphthalene derivative
Endrin	Insecticide, rodenticide	Chlorinated 1,4:5,8-dimethanonaphthalene derivative
HCB	**Fungicide, industrial compound, and by-product**	**Perchlorinated benzene**
Heptachlor	Insecticide	Chlorinated 4,7-methanoindene derivative
Mirex	Insecticide	Perchlorinated cyclobuta[*cd*]pentalene
PCBs	**Industrial compounds**	**Chlorinated biphenyls**
PCDDs	**Combustion/industrial by-products**	**Chlorinated dibenzo-*p*-dioxins**
PCDFs	**Combustion/industrial by-products**	**Chlorinated dibenzofurans**
Toxaphene	Insecticide	Chlorinated norbornane derivatives

PCDD/Fs are undesired by-products arising mainly from anthropogenic activity. Different chlorination patterns, combined with the consideration of symmetry, result in 75 possible dioxin and 205 furan congeners, the most toxic ones being halogenated in the 2,3,7,8 positions. PCDD/Fs are strictly planar molecules.

In 1872, Merz and Weith reported the first documented synthesis of PCDDs. The high biological potency of this class of compounds, however, was not noted until 1957, when Sandermann conducted research on the fungicide pentachlorophenol (PCP). He discovered PCDDs as by-products of a process for manufacturing plywood, using PCP as a wood preservative. The perchlorinated congener (octachlorodibenzo-*p*-dioxin, OCDD) showed no activity towards termites and mold, whereas tetrachlorodiphenylene dioxide (later referred to as 2,3,7,8-tetrachlorodibenzo-*p*-dioxin, TCDD) was highly effective. Researchers involved in this discovery soon developed symptoms of exposure, most prominently chloracne, which led to the end of Sandermann's research (Sandermann, 1984). Kimmig and Schulz linked these symptoms to identical clinical signs in workers at plants manufacturing trichlorophenoxyacetic acid (2,4,5-T). It was found that dioxins were generated as by-products in the industrial synthesis of this chlorinated pesticide as well as other chloro-organic syntheses, such as the production of trichlorobenzene and polyhalogenated biphenyls. PCDDs were found as contaminants of 2,4,5-T in the defoliant Agent Orange, used in Operation Ranch Hand during the Vietnam War. The concentration of TCDD ranged from 0.1–47 μg TCDD/g Agent Orange. It was public awareness and health problems in veterans that started research on dioxins at around

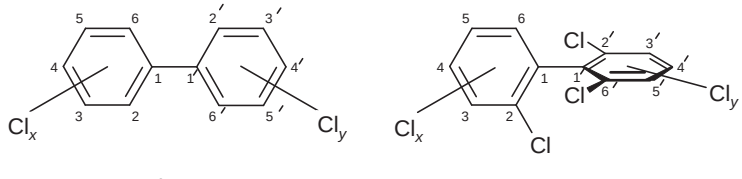
1970. Reports of industrial accidents with the release of PCDDs date back as early as 1949, but a major accident in a 2,4,5-T-producing plant in Seveso, Italy, in 1976 became the most infamous episode. Estimates of released TCDD, the most potent dioxin congener, range from 300 g–34 kg for this one incident. The most widely accepted estimate is ca. 1.3 kg, which contaminated an area of approximately one square mile, affecting about 37 000 people.

The main sources of dioxins and furans in the environment are thermal processes (production of metals and mineral products), combustion (waste incineration, heat and power generation, open-burning processes), and the chloro-organic industry. Each production condition generates a characteristic congener cluster.

Environmental samples can be linked to emission sources. PCDD/Fs are not commercially synthesized except in small amounts for scientific research. The thermal formation of PCDDs is a combination of *de novo* synthesis and generation from precursors in a temperature window of 250–350 °C, as found in smelters and outdated waste-combustion plants. PCDD/Fs are also generated during natural processes including forest fires, biodegradation and biosynthesis, and geothermal activities. Furthermore, congener clusters have been found in historical and prehistorical sediments worldwide (Australia, Germany, China, Japan, USA). The composition of those mixtures is not related to any pattern generated by anthropogenic processes, indicating a truly natural formation. The United Nations Environmental Programme (UNEP) published a comprehensive inventory of contributions to dioxin release, allowing countries to estimate total emissions according to domestic economy.

The environmental fate of TCDD is determined by its persistence. Studies after the accidental release in Seveso showed a half-life in soil of about 9–12 months. Studies at test sites revealed a climate-dependency of the half-life and reported values of 6.3 and 11 months in Florida and Utah, respectively. The number of bacteria capable of anaerobic, reductive dehalogenation in soil/sediment is limited and still under investigation for purposes of developing large-scale decontamination methods. The degradation of TCDD in the gas phase is rapid due to its susceptibility to photolysis by UV-light ($\lambda = 290$). Its theoretical half-life under photolytic conditions in the vapor phase is 1 h; under conditions of radical reactions with OH, on the other hand, its half-life in air is about 8.3 days. The persistence of TCDD in surface water depends on the extinction coefficient of the respective body of water and on seasonal changes in UV-radiation. It varies between 21 h (summer) and 118 h (winter).

6.1.3 Polychlorinated Biphenyls (PCBs)



PCBs are industrial chemicals, which are no longer used in developed countries. This substance class consists of 209 congeners, the toxicological potencies of which differ by the degree as well as by the pattern of chlorination. There are coplanar and nonplanar PCBs with differential toxicity profiles.

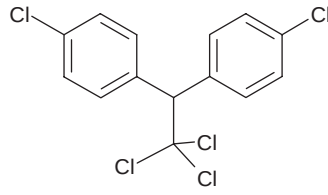
PCBs possess favorable technical properties, which have led to their widespread application in industry. These properties include a very high dielectric coefficient and high boiling points. PCBs are virtually noncombustible, heat resistant, chemically stable, and show very low acute toxicity. Their viscosity varies with the degree of chlorination. Therefore, PCBs have found use in electric transformers, capacitors, and heat exchangers, as lubricants, flame retardants, plasticizers, and additives in print inks and lacquers. An estimated cumulative total of 1.2 million metric tons were produced worldwide, half of which was manufactured by the USA. Production in America and new use was prohibited in 1978 under the Toxic Substances Control Act (TSCA).

PCBs are cost-effectively produced by chlorination of the biphenyl parent compound. Technical mixtures were marketed according to chlorine content (30–60 wt %), often indicated by their commercial trade name (e.g., Aroclor 1242 with 42 wt % chlorine). IUPAC Nomenclature of individual congeners is based on ascending numeric order of chlorination: PCB congeners were assigned numbers from 1–209 (Ballschmiter *et al.*, 1992). According to the Stockholm Convention, the use of equipment containing PCBs must be phased out by 2028.

The biological effects of PCBs greatly depend on their pattern of chlorination and, thus, their conformation. *Meta*-, *para*-, and mono-*ortho*-substituted congeners are coplanar and show a toxicological profile similar to PCDDs, but at much lower potency. The rings rotate freely around the central C—C bond. Bi-*ortho*-chlorinated PCBs show a rotation barrier of about 80 kJ/mole, still allowing racemization at room temperature. Higher chlorination in *ortho* positions, however, greatly increases the rotation barrier, since only cisoid transition states are possible. The energy required for C—C rotation is ca. 180 kJ/mole for tri- and 246 kJ/mole for tetra-*ortho*-chlorinated congeners, preventing these compounds from racemization, even at higher temperatures. Therefore, a planar transition state cannot be attained by these compounds under physiological or environmental conditions. It has been reported that some (+) and (−) enantiomers differ in toxic potency. Kinetic studies have shown slight differences in disposition, providing a possible explanation for different potencies. Tri- and tetra-*ortho* congeners elicit a toxicity profile different from those of planar PHAHs, by being also neurotoxic possibly by altering calcium signaling.

The first reported mass-poisoning with PCBs occurred 1968 on Kyushu Island, Japan. PCBs were used in the heat-exchanging system of a rice oil-manufacturing plant. Owing to leaky pipework, PCBs entered the final product and were consumed by approximately 1800 people. It was later shown that repeated heating of PCBs used in the heat exchanger caused the formation of traces of PCDFs in the product. Therefore, the symptoms observed in the affected population, such as chloracne and skin pigmentation, are considered a combination of PCB and PCDF toxicity.

6.1.4 DDT



The insecticidal effects of DDT were first discovered by Paul H. Müller in 1939. Its application greatly reduced outbreaks of vector-borne diseases, such as typhus and malaria. In 1948, Müller received the Nobel Price in Medicine for this contribution to preventive medicine. DDT was banned in the US in 1972 because of its persistence in the environment.

As a member of the Dirty Dozen, DDT was recently banned by the Stockholm Convention. However, exemptions were granted to about 25 countries for vector control, until cost-effective, environmentally safe and locally available alternatives have been developed. The cumulative worldwide production of DDT is estimated to exceed 2 million tons. US production peaked in the early 1960s (1962, 85 000 metric tons), with major applications in agriculture (cotton, peanut, soybean production). During the same period, Rachel Carson's book *Silent Spring* was published and fueled a growing environmental awareness in the US. Carson linked excessive use of pesticides to a reduction in bird populations, leading to the silence in the trees, reflected by the title. DDT and its metabolites were later found to be endocrine disruptors. In 1980, the wastewater pond of a pesticide-manufacturing facility containing DDT residues overflowed into Lake Apopka, FL, causing reproductive and developmental effects in alligators.

By 1968, several US states had banned the use of DDT, followed by a federal ban for most uses in 1972 and for all uses in 1989. However, US production remained at levels of one ton/day throughout the 1990s. Today, there is neither production, nor import, nor export of DDT in and out of the US.

Historically, the role of DDT in preventive medicine exceeded its economic importance in agriculture. DDT proved to be very efficient in controlling malaria, even in chronically plagued regions. This disease, transmitted by mosquitoes, was almost driven to extinction in Ceylon (Sri Lanka) by area and local spraying of DDT. After cessation of application in 1963, infection rates soon returned to previous levels (Tab. 2), indicating the effectiveness of DDT.

Table 6.2 *Effect of DDT usage on malaria infections in Ceylon (Sri Lanka).*

Year	1948	1963	1964	1969
Number of malaria cases	2 800 000 (before DDT usage)	17 (cessation of spraying)	150	2 500 000

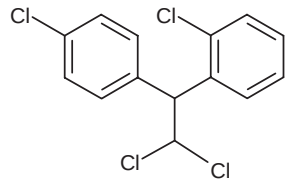
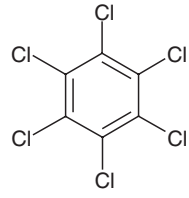


Figure 6.1 *Mitotane, a DDT analogue as drug in the treatment of adrenocortical carcinoma.*

DDT is a contact poison, irritant and repellant for insects. The combination of these properties increases its efficiency in controlling mosquitoes and lice by deterring them from entering sprayed areas as well as by killing them. However, gradually resistance to DDT started to develop, e.g. in *Anopheles*, requiring increasing doses to contain mosquito populations.

The human toxicity of DDT is extremely low, only one fatal poisoning having ever been reported. Its structural analogue mitotane (Figure 6.1) is used as a chemotherapeutic agent in the treatment of adrenocortical carcinoma. The mechanism of action is unclear, but mitotane appears to act as an endocrine disruptor, inhibiting certain functions in the adrenal cortex, and, thus, steroidogenesis. It causes adrenal atrophy, leading to starvation of the tumor and subsequent tumor destruction.

6.1.5 HCB



HCB is a potent fungicide, an industrial intermediate, and is generated as a by-product in chloro-organic syntheses (carbon tetrachloride, chlorophenols, vinyl chloride).

HCB was commercially synthesized by direct chlorination of benzene in the presence of a Friedel–Crafts catalyst, but the vast majority in the atmosphere was generated as by-products in the chloro-organic industry. Its primary use was the treatment of agricultural seed grain, such as wheat and sorghum, as well as onions, but it was also temporarily used as a flame retardant in plastics and in the production of pyrotechnics. Between 1955–1959, several poisoning episodes were reported with HCB in Turkey, when grain treated with HCB was used to bake bread during a period of starvation. An estimated 4000 people were affected, showing the Black Sore syndrome which exhibited symptoms like *porphyria cutanea tarda*. The fungicidal properties of HCB led to

widespread use in the 1950s and '60s. Its use as a fungicide was phased out in the USA starting in the mid-1960s and its registration was voluntarily canceled in 1984. Today, there is no US production, import, or export of HCB as a pesticide. However, it can be manufactured and used for chemical intermediates. Estimates of HCB generated as impurity in the US range from 68–690 tons/year in the mid-1990s.

In the US, most of the PHAHs covered by the Stockholm Convention are categorized as hazardous air pollutants under the Clean Air Act (CAA) and/or as priority toxic pollutants under the Clean Water Act (CWA).

6.1.6 Properties

PHAHs share common characteristics, such as high lipophilicity, low vapor pressure, a high melting point, and slow biodegradation, leading to biomagnification and environmental persistence.

The hydrophobicity of a compound increases with the degree of halogenation. This effect is amplified in PHAHs that lack further functional groups. Besides creating steric hindrance for enzymatic reactions, a high degree of halogenation also causes increased molecular weight, with elevated melting points and very low vapor pressure (Table 6.3). These properties are the predominant reason for the persistence of PHAHs in the environment.

Owing to (at least partial) planarity as a result of attached aromatic rings, they easily intercalate in microlayered minerals, such as clay, leading to geoaccumulation. Their Henry constants indicate low volatility from aqueous solution into the gas phase, quite independent of environmental conditions. Therefore, their vertical mobility in soil and sediment is very limited. Horizontal mobility occurs by erosion only. Consequently, PHAHs show high compartmental persistence in the environment. Ecotoxicological

Table 6.3 *Physical properties of selected PHAHs at 20–25 °C.*

	TCDD	Aroclor 1242	DDT	HCB
Molecular weight	321.97	266.5	354.49	284.78
Melting point (°C)	305	N/A	109	230
Boiling point (°C)	N/A	325–366	decomposes	322 (sublimates)
Water solubility	7.9 ng/l	0.10–0.34 mg/l	25 µg/l	6 µg/l
$pK_{O/W}$	6.79	5.6	6.91	3.59–6.08
Vapor pressure (mm Hg)	1.5×10^{-9}	4.06×10^{-4}	1.6×10^{-7}	1.089×10^{-5}
Henry constant (atm m^3/mol)	1.62×10^{-5}	5.2×10^{-4}	8.3×10^{-6}	6.8×10^{-4} to 1.3×10^{-3}

evaluations need to consider the resulting low bioavailability for these compounds from adsorbed states. The high $pK_{O/W}$ values of PHAHs indicate their high lipophilicity, causing bioaccumulation (i.e., uptake from medium and food) particularly in aquatic animals, and consequently, biomagnification (i.e., uptake from food only) throughout the food chain.

These compounds are slowly degraded by chemical and biochemical processes. However, they are subject to photolysis. Dehalogenation of higher chlorinated PCDD/Fs and PCBs can lead to the formation of more toxic congeners, increasing the total toxicity of a mixture. In aqueous media, OCDD is preferentially dechlorinated in the *peri* positions (C1, -4, -6, -9), yielding the more potent heptachloro congener. It has also been shown that UV-irradiation of PHAHs can result in the formation of other PHAH classes that were previously not present. Irradiation of water samples containing PCP yielded the formation of PCDDs and PCDFs in similar ratios. Therefore, degradation of PHAHs can, depending on substrates and conditions, lead to an increase or decrease in total toxicity.

6.1.7 Toxicity

The toxicity profile of PCDDs, PCDFs, and co-planar PCBs is qualitatively very similar (henceforth referred to as dioxin toxicity). However, amongst congeners, potency can differ by orders of magnitude. This relative potency is expressed in terms of Toxic Equivalency Factors (TEFs), normalized to TCDD, which is the most potent member of the family.

The most toxic congeners of PCDDs, PCDFs, and PCBs are chlorinated in the 2,3,7,8- or 3,3′,4,4′-positions, respectively. This is probably due to two main factors: First, this chlorination pattern is required for the compound to interact with the target site. Secondly, it renders these compounds largely resistant to metabolism. Therefore, they persist long enough to reach the target site at concentrations high enough to elicit an effect. The potency of these PHAHs is ranked in terms of TEFs (Table 6.4). The respective numbers were determined based on a combination of acute effects *in vitro* and *in vivo*. Evaluations of data by different groups of individuals led to slightly different TEF values between agencies. This may result in diverging risk assessments of mixtures.

It has been unequivocally demonstrated that high-dose acute toxicity effects as well as carcinogenicity effects of PCDDs, PCDFs, and co-planar PCBs are additive. More recently, additivity of medium-dose effects has also been demonstrated for reproductive endpoints. A mixture of 2,3,7,8-TCDD, 1,2,3,7,8-PeCDD, 1,2,3,4,7,8-HxCDD, 2,3,4,7,8-PeCDF, and 3,3′,4,4′,5-PeCB (PCB 126) inhibited ovulation in an entirely additive manner. Additivity allows the calculation of the total toxicity of mixtures based on the concentration of its individual components and the respective TEFs [Equation (6.1)]. The sum of these products is expressed as Toxicity Equivalent (TEQ):

Table 6.4 *TEF values of PCDDs, PCDFs, and PCBs according to NATO and WHO determinations. (Pe = penta, Hx = hexa, Hp = hepta, O = octa).*

Congener	NATO	WHO
PCDDs		
2,3,7,8-TCDD	1	1
1,2,3,7,8-PeCDD	**0.5**	**1**
1,2,3,4,7,8-HxCDD	0.1	0.1
1,2,3,7,8,9-HxCDD	0.1	0.1
1,2,3,6,7,8-HxCDD	0.1	0.1
1,2,3,4,6,7,8-HpCDD	0.01	0.01
1,2,3,4,6,7,8,9-OCDD	**0.001**	**0.0001**
PCDFs		
2,3,7,8-TCDF	0.1	0.1
1,2,3,7,8-PeCDF	0.05	0.05
2,3,4,7,8-PeCDF	0.5	0.5
1,2,3,4,7,8-HxCDF	0.1	0.1
1,2,3,7,8,9-HxCDF	0.1	0.1
1,2,3,6,7,8-HxCDF	0.1	0.1
2,3,4,6,7,8-HxCDF	0.1	0.1
1,2,3,4,6,7,8-HpCDF	0.01	0.01
1,2,3,4,7,8,9-HpCDF	0.01	0.01
1,2,3,4,6,7,8,9-OCDF	**0.001**	**0.0001**
PCBs		
IUPAC #	Structure	
77	3,3′,4,4′-TCB	0.0001
81	3,4,4′,5-TCB	0.0001
105	2,3,3′,4,4′-PeCB	0.0001
114	2,3,4,4′,5-PeCB	0.0005
118	2,3′,4,4′,5-PeCB	0.0001
123	2′,3,4,4′,5-PeCB	0.0001
126	3,3′,4,4′,5-PeCB	0.1
156	2,3,3′,4,4′,5-HxCB	0.0005
157	2,3,3′,4,4′,5′-HxCB	0.0005
167	2,3′,4,4′,5,5′-HxCB	0.00001
169	3,3′,4,4′,5,5′-HxCB	0.01
189	2,3,3′,4,4′,5,5′-HpCB	0.0001

Formula to calculate the TEQ of a mixture by applying the TEF concept.

$$\mathrm{TEQ} = \sum_{k=1}^{n} (c_k \cdot \mathrm{TEF}_k) \tag{6.1}$$

TEQ = toxicity equivalent
c = concentration
TEF = toxic equivalency factor
k, n = specific congeners

Table 6.5 *LD_{50} values for TCDD in selected mammals.*

Species	LD_{50} of TCDD (μg/kg)	Total body fat [% of body weight]
Guinea pig (Dunkin-Hartley)	0.7–2.0	4.5
Rat (Sprague Dawley)	45	8.9
Rhesus monkey	50	10.3
Mouse (C57 BL/6J)	100	7.9
Rabbit (New Zealand White)	115	10.1
Dog (Beagle)	1000	13.8
Hamster (golden Syrian)	5051	17.3

The TEF of the most potent congener (TCDD) is defined as 1. TEQ represents the theoretical amount of TCDD eliciting a response identical to that of a given mixture. Therefore, TEQ is also referred to as TCDD-equivalent.

Dioxin toxicity shows marked species differences, with the guinea pig being the most sensitive and the hamster being the most resistant genetically unmanipulated animal model.

The effects of PCDDs and PCDFs as well as PCBs are thought to occur by the same or similar mechanism. Therefore, most studies have focused on the most potent member of these substance classes, namely on TCDD. It serves as the standard model compound for toxicological studies. The LD_{50} of TCDD varies greatly among species (Table 6.5). The most sensitive mammal is the guinea pig with an LD_{50} between 0.7–2.0 μg/kg. It is well over 1000-times more sensitive than the hamster ($LD_{50} = 5051$ μg/kg), the most resistant species. LD_{50} values also vary greatly among strains, as most prominently revealed by two rat models: The LD_{50} in the sensitive Long Evans rat is 10 μg/kg, differing by a factor of 1000 from the resistant Han/Wistar rat ($LD_{50} > 9600$ μg/kg). To a lesser degree, this difference has also been observed in mice. C57 BL mice were found to be about 30-times more sensitive than DBA mice.

The explanation for some intra-species/inter-strain differences has turned out to be quite complex. Nevertheless, a striking general correlation ($R^2 = 0.834, P < 0.0001$) has been identified between total body fat content (TBF) of a species and the acute toxicity of TCDD [Equation (6.2)].

Formula to calculate the approximate LD_{50} of TCDD in mammals.

$$LD_{50} = 6.03 \cdot 10^{-4} (TBF)^{5.30} \tag{6.2}$$

TBF = total body fat (% of body weight)

A correlation between the toxicity of highly lipophilic compounds and TBF appears intuitively correct considering the disposition of such compounds. The larger the

compartment that serves as sink or pool (peripheral compartment) for the compound is, the lower is its concentration in the circulation, and, thus, at the target site (assuming the target site is not part of the peripheral compartment).

The above-mentioned empirical formula can be used to calculate a hypothetical acutely lethal dose for humans. The calculated LD_{50} of 6230 μg/kg TCDD for the 'Reference Western Man' (70 kg body weight, 21% TBF) supports the common notion that adult humans are not as sensitive to acute dioxin toxicity as are some animal models. Furthermore, these calculations allow an estimate of LD_{50} values across ages, using the respective body fat contents. This leads to the tentative conclusion that newborns (13.6% TBF) are about 10-times more sensitive to the acute toxicity of TCDD than are adult humans.

The effects of PCDD/Fs and PCBs are quite diverse and well characterized in experimental animals. They include a wasting syndrome and carcinogenicity (lungs, liver) at high doses, liver injury, immunosuppression, reproductive effects, and lowered serum insulin-like growth factor-1 (IGF-1) levels at medium doses, and effects on thyroid hormones, thymic atrophy, and enzyme inductions at low doses.

In humans, chloracne is the most sensitive symptom of elevated TCDD body burdens. It occurs approximately two weeks after exposure, manifested by swollen follicles. Three to five weeks post exposure, they turn into comedones. In the rabbit, the initial symptom of chloracne is dermatitis. This occurs two to four weeks after administration of TCDD, followed by swollen follicles and cysts several days later. In susceptible humans (young girls), symptoms may occur at TCDD concentrations of 800 ppt based on serum lipid content. In adolescents, differential diagnosis of acne-like skin conditions is notoriously difficult. Most individuals do not show signs below 11 000 ppt. The highest recorded level of TCDD in humans was 144 000 ppt blood fat in a 30-year-old woman, corresponding to a dose of 25 μg/kg TCDD. Symptoms included severe chloracne, nausea, vomiting, and gastrointestinal pain as well as cessation of menstruation. The hypothalamic-pituitary axis seemed to be unaffected.

Epidemiological studies could not link TCDD-exposure clearly to increased cancer mortality. Some studies associated occupational PCB exposure with hepatic, biliary, and intestinal cancers as well as with skin melanomas. Except for liver and biliary tract cancer, confirmatory evidence in animal models is lacking.

Investigations after accidental exposure suggest increased risks of digestive tract and respiratory tract cancer in TCDD-exposed smokers. This association supports the notion that TCDD is a carcinogen even though it is not a mutagen. Workers exposed to high concentrations of TCDD over prolonged periods of time showed significantly elevated to cancer incidence $\geq$20 years after initial exposure. However, when considering the total cohort, no increased mortality rate was found in comparison with the general population. Another cohort of 5100 workers, however, showed a decreased mortality rate from

strokes and gastrointestinal diseases. In summary, epidemiological data suggest that TCDD might have a weak tumorigenic effect at higher doses, particularly since, in the rat model, high doses of dioxins cause tumors of the lungs and liver. A lack of evidence for genotoxicity *in vitro* and *in vivo* suggests that TCDD, like PCBs and several other chlorinated hydrocarbons, exerts its carcinogenic effect through an epigenetic mechanism (e.g. promotion).

In many if not most species, TCDD has immunosuppressive effects at doses much lower than those causing acute toxicity.

Dose rates of 1 μg/kg/wk TCDD caused an increased susceptibility of mice to *Salmonella* infections. Immunosuppression has been observed in various lymphoid organs such as the thymus, spleen, and lymph nodes in different species exposed to a wide range of doses. Rats showed a decrease in cell-mediated immunity after a dose of 40 μg/kg TCDD, whereas 10 μg/kg TCDD caused an increase. The most prominent effect is thymic atrophy with impaired differentiation of T-lymphocytes. Antibody-mediated humoral immunity is also reduced. There are contradictory reports regarding effects on the immune system in humans. Epidemiological studies in Seveso cohorts could not establish an association.

In animals, dioxins have been shown to affect development and reproduction, and to be feto- and embryo-toxic as well as teratogenic. In humans, cessation of menstruation has been reported.

Studies in animal models demonstrated effects of TCDD on fertility, pre- and post-parturitional growth, and development. At higher doses, these effects included miscarriages and stillbirths. Morphological aberrations in the ovaries and uterus of the rat were seen after 1 μg/kg/d over a period of 13 weeks. The ED_{50} for the inhibition of ovulation in the immature rat model was found to be 3–10 μg/kg TCDD. A no-observed-adverse-effect level (NOAEL) of 0.03 μg/kg/d TCDD was reported for rats when administered from days 6–15 of gestation. In mice, teratogenic effects like cleft palate, hydronephrosis, and thymic hyperplasia were observed at doses below maternally toxic doses. The same effects have been reported in rats and hamsters, however, at doses that caused significant toxicity in the dam. A delayed onset of puberty in rats was found for the offspring when pregnant dams were dosed with 0.8 μg TEQ/kg on gestation day 15. Studies in primates also showed reduced fertility, lowered birth weights, and increased prenatal mortality, often accompanied by considerable toxicity in the pregnant female. Epidemiological studies on Seveso cohorts showed no increase in miscarriages or teratogenic effects but an increase in the number of female offspring, which was assumed to be linked to the paternal side. This could not be confirmed in cohorts of Operation Ranch Hand, which leaves a considerable uncertainty regarding reliability of the Seveso finding.

Other alleged effects associated with human exposure include hyperkeratosis, hyperpigmentation, hirsutism, liver damage, elevated blood fat content and cholesterol levels, intestinal effects such as diarrhea, cardiovascular effects, headache, peripheral neuropathy, reduced sensory performance, loss of libido, and psychiatric changes. Some of

these effects are also age-related and, therefore, it is difficult to assess if dioxins are indeed contributory or not. This demonstrates once more that the lack of information on dose or dose rates and exposure time in epidemiological studies often leads to equivocal conclusions.

TCDD has been shown to reduce hormone levels, such as thyroid hormones, insulin, and IGF-1, causing endocrine disruption and also dysregulation of intermediary metabolism.

It has been determined that TCDD causes a decrease in serum levels of total thyroxine (TT4) in rats within four days after a dose of 1 μg/kg. The concentration of total triiodothyronine (TT3) was, however, unaffected. Both hormones were lowered in a dose-dependent manner in C57 BL [lowest observable-effect lined (LOEL) 0.1 μg/kg] and DBA mice (LOEL 100 μg/kg). Sprague Dawley rats undergo a transient hypoinsulinemia at doses of 25 μg/kg together with insulin-hypersensitivity. It has been reported recently that rats maintained at steady state after a 3.2 μg/kg TCDD loading dose rate show a decrease in IGF-1 signaling within 8 days. Similar doses were shown to decrease ovulation but to prolong the life of experimental animals as well as to reduce cancer rates below those in controls. Figure 6.2 illustrates that these effects could be mediated by decreased IGF-1 signaling.

DDT and selected metabolites are endocrine disruptors, causing hormonal imbalances in wildlife.

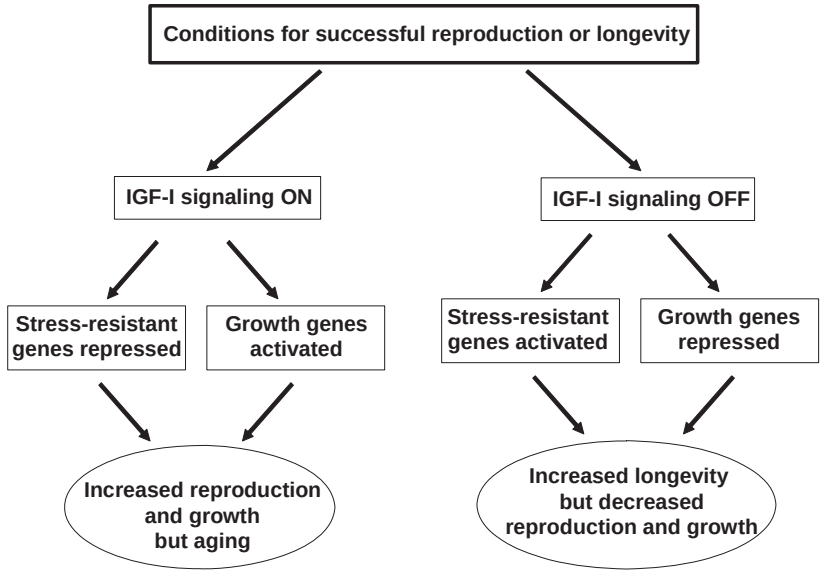


Figure 6.2 *Arking's theory of IGF-1 signaling and associated effects. Reprinted from Geyer et al. Copyright (1990/2002), with permission from Elsevier.*

Rachel Carson's *Silent Spring* triggered a large number of studies of the ecotoxic effects of DDT and other pesticides. It has been reported that the metabolites *p*,*p*′-dichlorodiphenyldichloroethylene (DDE) and *p*,*p*′-dichlorodiphenyldichloroethane (DDD) caused eggshell thinning in birds of prey, resulting in the breaking of eggs under the weight of the breeding parent. This led to decreased reproduction, e.g. in bald eagles, which recovered after the ban on DDT. In 1980, the wastewater pond of a pesticide-manufacturing facility containing DDT residues overflowed into Lake Apopka. In females alligators from that lake, plasma levels of 17 β-estradiol were two times higher than in controls, while males had significantly reduced testosterone levels. As a consequence of this endocrine disruption, testes of male alligators showed irregular structures, and phalli were abnormally small. Epidemiological studies on populations accidentally exposed to HCB showed an association with *porphyria cutanea tarda*, a heme imbalance.

The Black Sore syndrome, as observed in Turkey in the 1950s, included dermal sensitivity to sun, blistering, scarring, changed pigmentation, and hirsutism. As an inducer of P450 activity, HCB exposure also led to hepatomegaly. It is a hepatic and thyroid carcinogen in the hamster, the occurrence of which was found to be dose-dependent. In mice and rats, HCB has been shown to be a potent teratogen, causing renal and palate defects. At lower doses, HCB causes irritability and tremors.

6.1.8 Mechanisms of Action

The Aryl Hydrocarbon Receptor (AhR)

Much research has been conducted to elucidate the mechanism of dioxin toxicity, most of it related to the AhR. However, it has also been argued that a single mechanism is unlikely to explain a complex toxicity profile such as that displayed by dioxins.

Cytochrome P450s play a major role in phase I metabolism. Compounds can influence the transcription of these proteins via promoters. As first described by Poland in 1976, TCDD induces CYP 1A1 activity by interacting with its gene through a receptor-mediated mechanism. The cytoplasmic AhR and its nuclear partner ARNT (AhR nuclear translocator) play key roles in this signal transduction (Figure 6.3). Although the AhR is not TCDD-specific, TCDD is the most potent AhR agonist known. Therefore, TCDD is used as a model compound for AhR-binding kinetics and the study of downstream effects. The best known AhR signaling pathway is initiated by an agonist binding to the receptor. In this process, proteins associated with AhR are released. The AhR/agonist complex translocates into the nucleus and binds to the ARNT. This heterotrimer attaches to the dioxin response element (DRE), a core heptanucleotide sequence in the DNA, and acts as transcription factor. Mediated through translation, the AhR-signaling cascade elicits a biological response, such as CYP 1A1 induction *in vitro* and *in vivo*.

Although CYP 1A1 metabolizes many exogenous compounds, an endogenous substrate for the AhR has yet to be identified. In the rat, the highest tissue-concentrations of AhR were found in the thymus, lung, liver, and kidney. Some researchers attribute

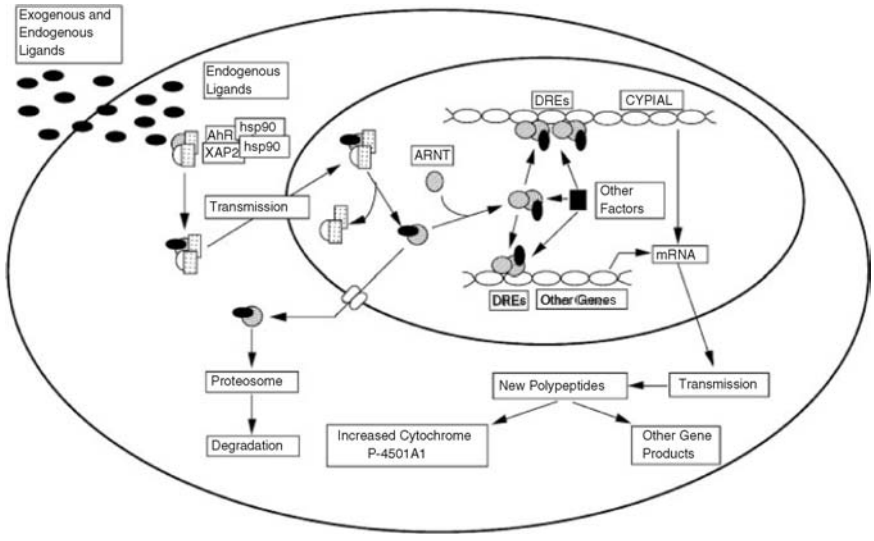


Figure 6.3 *AhR signaling pathway. Reprinted from Denison and Nagy. Copyright (2003), with permission from Annual Reviews.*

most or all toxic effects of PCDDs to AhR interaction. However, it has also been argued that the AhR might not be the universal key to acute toxicity, as the ED_{50} for CYP1A1 induction in rats is about two orders of magnitude lower (Sprague Dawley 0.2–0.3 μg/kg TCDD) than the LD_{50} (43 μg/kg TCDD). Furthermore, calculations of protein content in the liver yield a saturation of the AhR at 1.27 ng TCDD/g tissue, whereas sublethal doses (2.5 μg/rat) cause 100-times higher liver concentrations (117 ng/g). It has been argued that a saturated receptor cannot cause a variable such as a dose-response to occur at much higher doses. The study of sensitive and resistant rat strains also yields controversial results. The specific binding affinity of the AhR to TCDD does not differ between sensitive (Long Evans, 20 fmoles/mg cytosolic protein) and resistant (Han/Wistar, 23 fmoles/mg cytosolic protein) rat strains. In addition, different effects are produced at different doses of TCDD in sensitive and resistant mice (Table 6.6). While C57 BL mice show a decrease in thyroid hormones at 0.1 μg/kg TCDD, the same effect is only elicited in DBA mice by almost 1000-times higher doses. The ED_{50}s for CYP1A1 induction, however, differ by a factor of only 15.

Studies in AhR knock-out mice showed at least 10-times higher resistance to TCDD as compared with the wild type. However, TCDD-treated knock-out mice displayed scattered necrosis of the liver and lymphocytic infiltration of the lungs, which were effects also seen in TCDD-treated wild-type mice. Although less severe, the very presence of these effects suggests an AhR-independent component in the manifestation of toxicity.

Further studies with AhR knock-outs as well as wild-type animals will have to be conducted to address contradictions and to fully reveal if and to what extent this receptor modulates or mediates dioxin toxicity.

Table 6.6 *Differences in doses and effects between mice strains sensitive (C57 BL) and resistant (DBA) to TCDD toxicity.*

Mouse strain	ED_{50} CYP1A1 induction (liver) (μg/kg)	LD_{50} (μg/kg)	ED_{50} reduced serum glucose, activity of PEPCK and glucose-6-phosphatase (μg/kg)	Effects reduced T3, T4 serum levels (μg/kg)
C57BL	1.1	100	100	0.1
DBA	16	>3000	1000	97.5
Difference between strains	15-fold	>30-fold	10-fold	975-fold

Phosphoenolpyruvate Carboxykinase (PEPCK)

The hallmark of dioxin toxicity is the wasting syndrome in animal models: Reduced feed intake, combined with decreased body weight, and derailment of intermediary metabolism. This eventually leads to a lethal hypoglycemia, the apparent cause of acute death by TCDD.

Overall energy metabolism is directly related to the respiratory quotient. In TCDD-treated animals, this quotient is reduced as compared with pair-fed controls when the wasting syndrome has progressed beyond the stage of glycogen depletion. This has led to the suggestion that intermediary metabolism is impaired in animals treated with TCDD. In further studies, gluconeogenesis was confirmed as the target of that alteration. It has been shown in rats that a key enzyme of hepatic gluconeogenesis, PEPCK, is inhibited by TCDD in a dose-dependent manner, leading to a decreased activity to the extent of 44% of pair-fed controls. This reduction of liver PEPCK activity was found in the exact same dose range as acute toxicity occurs. In addition, hepatic tryptophan 2,3-dioxygenase was also dose-dependently reduced in rats and, consequently, plasma tryptophan was dose-dependently increased, which is compatible with the feed intake reduction observed. Subsequent studies in mouse strains with different sensitivity to TCDD (C57BL/6J and DBA/2J) revealed a plateau in the decreasing activity of PEPCK in liver, coinciding with the onset of acute toxicity in the same dose range. At lethal doses, PEPCK activity was reduced by 80% in liver of mice. However, hepatic tryptophan 2,3-dioxygenase activity and plasma tryptophan were unchanged in mice, which is compatible with a lack of reduced feed intake and a lack of reduced body weight in mice. Still, lethal hypoglycemia ensued, apparently, by the more pronounced inhibition of PEPCK in mice than in rats.

Studies on glucose homeostasis in the most TCDD-resistant and the most TCDD-susceptible species, the hamster and the guinea pig, respectively, did not support the notion of hepatic PEPCK being the mediator of lethal TCDD toxicity in all species.

A decreased PEPCK activity but unaffected liver glycogen levels were observed in hamsters at doses that did not induce body weight loss. Additionally, in hamsters a dose-dependent decrease in serum free fatty acids was also detected. While guinea pigs showed a dose-dependent decrease in liver glycogen levels they revealed a dose-dependent increase in serum free fatty acids. No inhibition of liver PEPCK activity was seen in liver cytosol of guinea pigs. It is noteworthy that in guinea pigs most of the hepatic PEPCK is located in mitochondria, whereas in both rats and mice PEPCK is mainly found in cytosol.

This leads to the conclusion that lethality in rats, mice, hamsters, and guinea pigs after the administration of TCDD is due to severe impairment of intermediary metabolism. However, a uniform mechanism of action for all species is unlikely to be discovered because of very different regulation of intermediary metabolism between hibernators (hamsters), herbivores (guinea pigs), and omnivores (rats, mice).

CNS Stimulation

DDT is a CNS stimulant by interfering with sodium/potassium conductance in neurons.

DDT is not toxic to humans up to an estimated acute dose of 10 mg/kg. Chlorinated insecticides such as DDT are CNS stimulants. In mammals, high doses cause neurotoxicity by interfering with sodium/potassium conductance in cells. This results in repetitive neural firing, causing tremors and seizures that can be treated with anticonvulsants. Although DDT is a liver carcinogen in mice, no evidence was ever presented for increased cancer rates in occupationally exposed workers.

Uroporphyrinogen Decarboxylase (UPD)

HCB is a porphyrinogen by inhibiting UDP in liver leading to the accumulation of porphyrins.

The toxicity of HCB is thought to be related to inhibition of UPD, a key enzyme in heme synthesis. This inhibition has been reported to precede porphyrinogenic symptoms. Specifically, HCB blocks the conversion of the first cyclic tetrapyrrole, uroporphyrinogen III, leading to increased re-oxidation and thus to the accumulation of uroporphyrins in liver. Since a direct inhibition of UPD could not be demonstrated *in vitro*, an indirect effect is assumed. A role for metabolites is unlikely considering the very low rate of metabolism of HCB. Lipid peroxidation as a result of CYP induction has been suggested as a possible mechanism of action. Lipid peroxidation and mitochondrial dysfunction are known to alter potassium membrane permeability, which was shown to increase heptic porphyrin levels. Furthermore, iron exacerbated the effects of HCB. Therefore, altered potassium and/or iron homeostasis is thought to be the mechanism of toxicity of HCB.

6.1.9 Metabolism

Biotransformation is the slowest and hence rate-limiting step in PCDD-elimination. Therefore, nonbiliary intestinal elimination by desquamating enterocytes and by redistribution of PHAHs into fecal fat become the main route of excretion.

PHAHs can be ring-hydroxylated, which requires vicinal unsubstituted aryl positions. Lacking these, biotransformation occurs on a very slow time-scale, leading to long half-lives of the different congeners. Therefore, PCDD/Fs chlorinated in the 2,3,7,8 positions are poor substrates for both oxidation and reductive dehalogenation. The same holds true for PCBs with *meta* and *para* substituents. One of the few low-yield metabolites of TCDD is 2,3,7-trichloro-8,9-dihydroxydibenzo-*p*-dioxin. These enzymatic reactions can also involve a chlorine-switch from a lateral position towards the central ring, as shown by identification of the minor TCDD metabolite 2-OH-1,3,7,8-TCDD. Other metabolic pathways are oxygen-bridge cleavage in PCDD/Fs, leading to tetrachlorodihydroxydiphenyl ether or hydroxylated PCBs, respectively. Perchlorinated PHAHs have to undergo reductive dehalogenation in order for further metabolism to occur. Epoxide intermediates as well as hydroxylated metabolites, once formed in low yield, are readily biotransformed by phase II metabolism. The derivatized PCB glutathione adduct 2,2′,5,5′-tetrachloro-4,4-bis(methylsulfonyl) biphenyl was reported to accumulate in lungs after *i.p.* administration to rats. Further methylsulfone derivatives of PCBs were identified in the liver and adipose tissue after chronic exposure, suggested to be due to protein binding. However, metabolites are generally less toxic than the parent compounds and much more rapidly excreted, due to increased hydrophilicity and additional phase II metabolism.

Overall, biotransformation plays a minor role in the disposition of PHAHs.

6.1.10 Enzyme Induction

PHAHs are potent enzyme inducers *in vitro* and *in vivo*. Enzyme induction is one of the most sensitive effects of TCDD and related compounds.

It has been shown *in vitro* that prior exposition of hepatocytes to TCDD increases the rate of its metabolism 3.2-fold for subsequent exposures. However, an auto-induction of TCDD-metabolism *in vivo* could not be confirmed. The previously described AhR signaling pathway is the best known route of enzyme induction by co-planar PHAHs. It predominantly leads to the induction of CYP1A1 and 1A2 by a factor of 50–100 compared with controls. The activity of NAD(P)H-quinone reductase was also induced by TCDD, however, to a far lesser degree. It has been also reported that PCBs induce activity of the phenobarbital group (CYP2B, 2C, 3A) as well as CYP2A1, epoxide hydrolase, DT-diaphorase [NAD(P)H-quinone oxidoreductsse], and aldehyde dehydrogenase activity. PCB atropisomers display differential induction patterns of enzymes by (+) and (−) enantiomers, including patterns observed by the phenobarbital-type inducers. PHAHs also induce phase II enzymes, but fewer studies have been conducted in this field. It was demonstrated that TCDD induces UDP-GT (uridine

diphosphate-glucuronosyltransferase) about 25-fold, whereas activities of glutathione *S*-transferase (GST), sulfotransferase, and *N*-acetyltransferase remained essentially unaffected.

6.1.11 Kinetics

After an initial distribution phase, PHAHs are deposited mainly in adipose tissue and also in liver. Owing to their persistence they have very limited elimination, and thus long elimination half-lives. Because of these very long half-lives the rate-limiting steps in the toxicity of this class of compounds are driven by their kinetics.

The bioavailability of TCDD after a single *per os* administration ranges from about 50% in the guinea pig to 70–85% in the rat and is also dependent on the choice of vehicle. Human data are available only on one self-administered dose. It was determined as >86% derived from fecal excretion of [1,6-^{3}H]TCDD. Feces are the main elimination pathway also in the rat. Urinary excretion plays a minor role at only 5–13% of dose. Following absorption, the initial distribution of TCDD depends on physiological parameters such as perfusion rate and relative size of tissues. Final distribution of TCDD follows the affinity of the compound to liver (5% of dose/g tissue) and white adipose tissue (1% of dose/g tissue) within the first 24 h. The half-life of TCDD in rats is 11 days for the parent compound in serum and 21 days in the liver. In guinea pigs, TCDD has a half-life of approximately 30 days, whereas the mean half-life of TCDD in humans is 7.8 years. The higher the degree of chlorination, the longer is the biological half-life of a PHAH as compared with its congeners.

In humans, DDT has an elimination half-life of about 5 years; its metabolite DDE is even more persistent. DDT is excreted mainly via urine. The fractions excreted via biliary and lactational pathways increase with increasing doses of DDT.

HCB shows a half-life of 24 days in rats, 32 days in rabbits, and about one year or more in humans. Much less of HCB distributes into liver than of TCDD because HCB, unlike TCDD, does not have high affinity to CYP1A2, which is the main and most abundant binding protein of PHAHs in the liver.

The kinetics of persistent xenobiotics in rats and humans can be calculated by an empirical formula [Equation (6.3)] which also allows the calculation of elimination half-lives of very persistent compounds, such as OCDD.

Formula for the extrapolation of elimination half-lives of persistent xenobiotics from rats to humans.

$$t_{1/2(\text{human})} = 17.78 \cdot \left(t_{1/2(\text{rat})}\right)^{1.34} \tag{6.3}$$

$t_{1/2} =$ elimination half-life

The difference in half-lives between rat and human not only determines species-dependent persistence of these chemicals in the body, but also steady-state levels during chronic exposure [Equation (6.4)].

Formula for the calculation of steady-state concentrations during chronic exposure.

$$c_{ss} = \frac{1.44 \cdot t_{1/2} \cdot f \cdot dose\ rate}{V_d \cdot \tau} \tag{6.4}$$

$c_{ss} =$ steady-state concentration
$f =$ fraction absorbed
$V_d =$ volume of distribution
$\tau =$ dosing interval

Small species-to-species variations in the volume of distribution and in the fraction absorbed have only minor impact on this equation. Differences in half-lives, however, are orders of magnitude larger and directly proportional to the steady-state levels achieved. The half-life in rats is approximately 140-times shorter than in humans. It takes 6.64 elimination half-lives to reach 99% of steady-state during chronic exposure. Therefore, rats reach it considerably faster than do humans [Equations (6.5)–(6.7)]. Furthermore, at identical daily dose rates, humans will eventually reach a steady-state about 140-times higher than rats, which represents the safety factor needed when extrapolating data obtained in rats to humans.

Calculation of eliminated fractions.

$$C_p = C_0 \cdot e^{-kt} \tag{6.5}$$

$$C_p = 0.5 \cdot C_0 \longrightarrow t_{1/2} = \frac{\ln 2}{k} \tag{6.6}$$

$$C_p = 0.01 \cdot C_0 \longrightarrow t_{1/99} = \frac{\ln 100}{k} = 6.64 \cdot t_{1/2} \tag{6.7}$$

$C_p =$ concentration in plasma
$C_0 =$ concentration in plasma at $t = 0$
$k =$ elimination rate constant
$t =$ time
$t_{1/2} =$ elimination half-life
$t_{1/99} =$ time to reach 99% elimination

The long period of time to steady-state also affects bioconcentration factors (BCFs). This is the reason why exposure studies with PHAHs in aquatic species must be extended for at least 6.64 half-lives. In the 1970s, a lack of appropriate study designs led to the assumption that highly chlorinated PCDDs would not bioconcentrate. This was clearly refuted both theoretically and experimentally.

Lipophilicity of PHAHs leads to high concentrations in breast milk. With the onset of lactation, lipophilic compounds are redistributed into this newly formed lipophilic compartment. This results in high concentrations, especially for the first pregnancy, when body burden of PHAHs after decades of accumulation is available for redistribution into milk fat. In later pregnancies, levels in breast milk are significantly lower, since the body burden has been reduced by the previous lactational elimination. Because of more frequent lactations, PHAH levels in cow's milk are considerably lower than in breast milk.

6.1.12 Summary

The production, emission, and use of persistent PHAHs peaked in the middle of the last century with an increase in synthetic chemistry and an upswing in agriculture and technology. Today, most of these materials are being phased out or have already been banned. Unintentional emissions have been reduced and clean-up of heavily contaminated sites is underway in developed countries.

These facts could lead to the wrong assumption that PHAHs belong to our past and merit little further attention. However, these compounds are highly persistent and are ubiquitously present in the environment and the human food chain both at the present time and in the foreseeable future. The effects of PHAHs on humans, animals, and the environment have been studied for decades. Yet, investigations of the toxicity of these compounds, applying classical and modern technologies, still yield new findings, providing new and improved bases for risk assessment and for the improvement of directives in regulatory agencies worldwide. PHAHs also serve as model compounds to study mechanism(s) of toxicity, which might facilitate the development of new drugs.

References

M.S. Denison and S. R. Nagy, Activation of the Aryl Hydrocarbon Receptor by Structually Diverse Exogenous and Endogenous chemicals. *Annu. Rev. Pharmacol. Toxicol.*, **2003**, *43*, 309–334.

H. J. Geyer, I. Scheuntert, K. Rapp, A. Kettrup, F. Korte, H. Greim, and K. Rozman. Correlation between Acute Toxicity of 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) and Total Body Fat Content in Mammals. *Toxicology*, **1990**, *65*, 97–107.

H. Geyer, K.-W. Schramm, E. A. Feicht, A. Behechti, C. Steinberg, R. Brüggemann, H. Poiger, B. Henkelmann, and A. Kettrup. Half-lives of Tetra-, Penta-, Hexa-, Hepta-, and Octachlorodibenzo-p-dioxins in Rats, Monkeys, and Humans - A Critical Review. *Chemosphere*, **2002**, *48*, 631–644.

K. Ballschmiter, R. Bacher, A. Mennel, R. Fischer, U. Riehle, and M. Swerev, *The Determination of Chlorinated Biphenyls, Chlorinated Dibenzodioxins, and Chlorinated Dibenzofurans by GC-MS. J. High Resolut. Chromatogr.,* **1992**, *15*, 260–270.

W. Sandermann, *Dioxin. Die Entstehungsgeschichte des 2,3,7,8-Tetrachlorodibenzo-p-dioxins (TCDD, Dioxin, Sevesogift). Natwiss. Rundsch.,* **1984**, *37*, 173–178.

For Further Reading

Toxicological Profile for DDT, DDE, and DDD (*Update*), US Department of Health & Human Services, Agency for Toxic Substances and Disease Registry, Atlanta, Georgia, **2002**.

Toxicological Profile for Hexachlorobenzene, US Department of Health & Human Services, Agency for Toxic Substances and Disease Registry, Atlanta, Georgia, **1996**.

Toxicological Profile for Polybrominated Biphenyls, US Department of Health & Human Services, Agency for Toxic Substances and Disease Registry, Atlanta, Georgia, **1995**.

Toxicological Profile for Polychlorinated Biphenyls (*Update*), US Department of Health & Human Services, Agency for Toxic Substances and Disease Registry, Atlanta, Georgia, **1997**.

L. W. Robertson and L. G. Hansen, *PCBs - Recent Advances in Environmental Toxicology and Health Effects*, The University Press of Kentucky, Lexington, Kentucky, **2001**.
K. K. Rozman, *A Critical View on the Mechanism(s) of Toxicity of 2,3,7,8-Tetrachlorodibenzo-p-dioxin. Implications for Human Safety Assessment. Dermatosen*, **1989**, *37*, 81–92.
K. K. Rozman, W. L. Roth, and J. Doull, *Influences of Dynamics, Kinetics, and Exposure on Toxicity in the Lung*, in *Toxicology of the Lung*, 4th edn, (Ed. D. E. Gardner), CRC Press, Boca Raton, FL, **2006**, pp. 195–229.
UNEP, *Standardized Toolkit for Identification and Quantification of Dioxin and Furan Releases*. **2005**. http://www.pops.int/documents/guidance/toolkit/en/Toolkit_2005_En.pdf

6.2 Metals

Karl-Heinz Summer, Stefan Halbach, Herrmann Kappus and Helmut Greim

6.2.1 Introduction

Metals, which occur naturally in the earth's crust, can be introduced into the environment via geological, biological, and anthropogenic pathways. Environmental sources are soil erosion, volcanic activity, and forest fires, and in our food derived from dietary plants and animals. The most important anthropogenic sources are the burning of fossil fuels, mining of minerals, smelting and refining of ores (Table 6.7), industrial utilization of metals, metal processing, and the disposal of local and industrial waste. In industrialized areas anthropogenic emission of metals into the atmosphere is usually greater than that of natural emissions.

Table 6.7 *World production (mine output) of metals*[a].

Metal	Production (tons/yr)
Crude steel	900×10^6
Aluminum	20.5×10^6
Copper	13.7×10^6
Chromium	12.6×10^6
Zinc	9.1×10^6
Lead	3.1×10^6
Nickel	1.3×10^6
Cadmium	19.9×10^3
Silver	18.3×10^3
Platinum metals[b]	13.2×10^3
Gold	2.6×10^3
Mercury	1.8×10^3

[a]From statistical data 2000–2002.
[b]Platinum and Palladium.

6.2.2 Toxicity

Metals have greatly varying effects that are frequently organ-specific.

Metals are presumably among the oldest poisons known to humans. Lead poisoning was known more than 2000 years ago. In view of their biological activity, metals are differentiated as essential or nonessential. The mammalian organism has a requirement for essential metals, e.g. iron, copper, chromium, molybdenum, magnesium, and many others. These are termed essential metals. Nonessential metals have no physiological function and are, therefore, occasionally called toxic metals. However, overdose of any metal can lead to toxicity. Generally, the absorption of metals is controlled by feedback mechanisms, which prevent excessive accumulation in the organism, despite higher than normal levels of exposure. However, several metals show a distinct tendency to accumulate. Thus, the biological half-life for mercury in humans is about 70 days, that for cadmium about 10–20 years. A metal may exhibit tissue-specific half-lives. The half-life for lead in soft tissue is several weeks, but more than 20 years in bone.

Owing to their different molecular interactions in cells, tissues, and organs, metals cause diverse effects in humans, and pathophysiological effects can become apparent locally on the skin, in the lung, and in the GI tract or systemically. Moreover, metals can be allergenic, mutagenic, teratogenic, and carcinogenic.

Metal atoms can form ionic, covalent, and coordinate bonds. Ligands containing oxygen, nitrogen, or sulfur are preferentially bonded. Consequently, many important biological compounds, such as proteins and nucleic acids, are targets for an interaction with metals. If the ligands are organic molecules with more than one group capable of coordination, and can form stable complexes, there complexes are called chelates.

Several chelating agents with varying specificity for metals are used in the treatment of metal poisonings (Table 6.8). Examples include: dimercaprol (BAL), the calcium disodium salt of ethylenediaminetetraacetic acid (EDTA), the hydroxamic acid derivative deferoxamine, 2,3-dimercaptopropanesulfonate (DMPS), and β,β-dimethylcysteine (D-penicillamine). The following characteristics are fundamental prerequisites for the therapeutic use of chelating agents as antidotes:

1. High affinity for toxic metals and low affinity for essential metals;
2. Low toxicity of chelating agent and chelate;
3. Insignificant metabolism of chelating agent and chelate;

Table 6.8 *Chelating agents as antidotes in metal poisoning.*

Chelator	Metal
Dimercaprol (BAL)	As, inorganic Hg, Pb (with EDTA)
EDTA[a]	Pb
Desferoxamine	Fe
DMPS[b]	Me-Hg, inorganic Hg, Pb, Cd, Cu, Ni
D-Penicillamine[c]	Cu, (Pb)

[a] Calcium disodium ethylenediaminetetraacetate.
[b] 2,3-Dimercaptopropanesulfonate.
[c] β,β-Dimethylcysteine.

4. Substantial excretion of chelate in urine or bile;
5. Stability of chelate at physiological pH and in acidic urine.

In general, chelating ligands do not react exclusively with the metal to be mobilized. Unintended chelation of essential metals, such as Ca, Zn, or Cu, depletes needed stores of the metal and can be life-threatening. In this context it must be emphasized that the diagnostic evaluation of the body burden of Cd, Hg, or Pb by the use of complexing agents such as DMPS is not justifiable. The body burdens of Pb and Hg can be extrapolated from measurements of the blood level of these metals.

Toxic effects of metals are elicited by interactions with receptors capable of binding the metal. The absorption and distribution of the metal may be dependent upon the chemical form in which it is administered. Thus, the metal may be poorly soluble, lipid-soluble, or easily absorbable as organic metal compounds. The toxicity of a metal, however, is directly influenced by the degree to which it is bonded to a specific cellular ligand. For all metals local irritation at the site of exposure is an important acute effect.

The chronic effects of metals are usually tissue-specific and are influenced by specific metal-binding proteins. Such metal-binding proteins have no enzymatic activity, are temporary depots and transport forms for essential metals, and exhibit detoxification functions by virtue of their metal-binding activity:

1. Calcium-binding calmodulin;
2. The iron-binding ferritin;
3. The glycoprotein transferrin;
4. The transport form for Cu, ceruloplasmin;
5. Metallothionein (Table 6.9).

Under physiological conditions metallothionein binds up to 7 Zn atoms or 12 Cu atoms per molecule. However, it can also bind up to 7 atoms of Cd, Hg, Co, and Ni, and up to 18 atoms of Ag. It is a ubiquitous protein with a molecular weight of 6500 and contains approximately 30% cysteine. All cysteine-sulfur atoms are involved in metal binding in the form of two clusters. Owing to its high affinity for metals and its inducible synthesis

Table 6.9 *Specific metal-binding proteins.*

Protein	Molecular weight	Occurrence	Metal	Function
Calmodulin	14 000	Ubiquitous	Ca	Activates several enzymes (second messenger)
Ferritin	470 000	Liver, spleen and bone	Fe	Storage protein
Transferrin	90 000	Plasma, extracellular space	Fe	Transport protein
Ceruloplasmin	132 000	Plasma	Cu	Transport protein, (Fe-Oxidase)
Metallothionein	6 500	Ubiquitous	Ag, Hg, Cu, Cd, Zn	Storage protein, detoxification

by metals, metallothionein is important in the metabolism and detoxification of metals, as well as the development of resistance to metals. Accordingly, the kidneys of people with excessive exposure to Cd, such as smokers, show increased Cd and metallothionein levels.

6.2.3 Carcinogenicity

Many metals are genotoxic *in vitro*. They interact with enzymes involved in DNA repair or cell regulation or they react with nucleophilic centers of nucleic acids. Several are carcinogenic in animals and humans. The carcinogenic potency increases with increasing electronegativity and decreasing solubility.

There is no general mechanism that explains the carcinogenic effect of the different metals. The carcinogenic potency of several metals and metal salts increases with increasing electronegativity and decreasing solubility. The electronegativity of most carcinogenic metals is in the range of 1.2 to 1.9, and the scarcely soluble oxides and sulfides of Ni and Cr are more potent carcinogens than are the soluble salts. Many metals are known to form complexes with nucleic acids or to inhibit basic processes of metabolic regulation and nucleic acid metabolism. For example, inactivation of sulfhydryl groups or displacement of essential metal ions such as Cu^{2+}, Mg^{2+}, Mn^{2+}, or Zn^{2+} can impact on DNA replication, transcription, or repair. Accordingly, many metals or metallic compounds show genotoxic effects *in vitro*. Based on epidemiological evidence, Ni, Cr, and As are classified as human carcinogens, and there are indications that Be, Cd, and Pb are also carcinogenic to man. In animal experiments, Co, Fe, Mn, Pt, Ti, and Zn have been shown to be carcinogenic at very high doses.

Mechanisms of Metal Carcinogenesis

Various molecular mechanisms contribute to metal carcinogenesis. Many metals are directly genotoxic and mutagenic. For most metals, however, indirect mechanisms, rather than direct interaction with DNA, are involved in the carcinogenic process. Generally, it is the metal ion that interacts with the critical cellular targets.

Among the metals, cadmium-induced carcinogenesis appears to be the best understood. The molecular and cellular effects of cadmium carcinogenesis are shown in Figure 6.4. At very low concentrations cadmium inhibits DNA repair by removal of thymine dimers from DNA. Other repair processes like nucleotide excision repair and mismatch repair are also inhibited. The inhibition most likely results from a replacement of zinc ions in these zinc-containing enzymes by cadmium ions. It has been shown that some cadmium-induced tumors can be prevented by dietary zinc supplementation.

Cd ions also induce oxidative stress. In contrast to iron, Cd ions do not produce oxygen radicals by reacting with hydrogen peroxide. Instead, oxidative stress is considered to

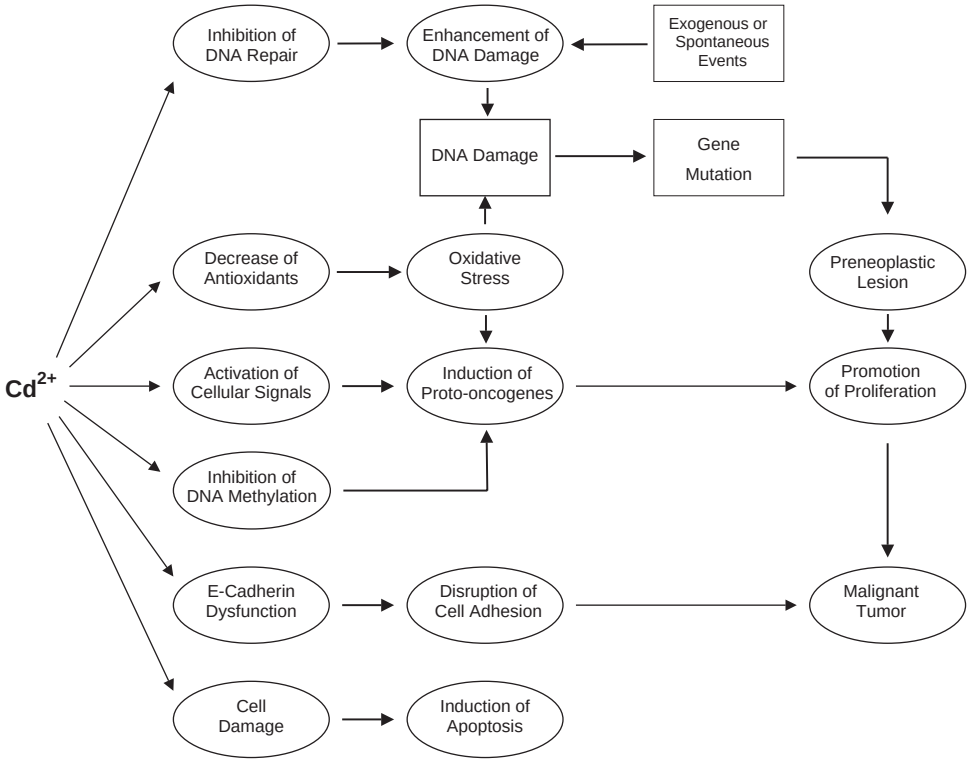


Figure 6.4 *Mechanisms involved in the carcinogenic process induced by cadmium. Reprinted from Waisberg et al. Copyright (2003), by permission from Elsevier.*

result from the well known decreased activities of enzymes like superoxide dismutase and catalase, and reduced concentrations of the antioxidant glutathione coupled with reduced glutathione synthesis. The consequence is an increase of the steady-state concentrations of oxygen radicals and an increase in oxidative DNA damage.

Several mechanisms have been described which contribute to Cd carcinogenesis:

1. Concentrations of Cd below those that would be toxic to the cell induce elevations in reactive oxygen species. Oxygen radicals have signalling properties. By this mechanism Cd can influence a number of genes involved in cell regulation. For example, Cd induces proto-oncogenes and tumor-suppressor genes, as well as genes that express metallothionein, heat-shock protein, and glutathione-related proteins.
2. Cd inhibits DNA methylation at very low concentrations. Cd may activate proto-oncogenes and inhibit tumor-supressor genes. DNA methylation is involved in several stages of cell regulation and hypomethylation plays a key role in carcinogenesis.
3. Cd interacts with E-cadherin-dependent junctions between cells probably by displacement of calcium. The consequence is disruption of cell adhesion, which also triggers tumor development.

4. Although Cd induces apoptosis, probably due to induction of several kinases, only a limited number of cells die. The result is selection of mutated cells, which are resistant to apoptosis. Clonal expansion of these cells leads to tumor development.

The carcinogenic effects of Cd appear to result from indirect effects on DNA that influence DNA repair, cellular growth regulation, and cell adhesion. It is likely that a combination of some or most of these mechanisms leads to tumor development. However, the activity of these mechanisms in different cells may vary. Thus, tumors may not develop in various organs despite the presence of Cd at similar concentrations.

The ionic form of lead (Pb^{2+}) is also carcinogenic. Pb inhibits DNA repair, induces oxidative stress, and replaces Zn ions in critical targets similar to Cd. In contrast to Cd, lead induces oxidative stress directly by interacting with hydrogen peroxide to form hydroxyl radicals. Decreases in antioxidants have not been demonstrated. Pb induces critical genes related to carcinogenesis. Protein kinases appear to play a major role and replacement of Ca by lead has been suggested. The other mechanisms described for Cd either have not been examined or are not involved in lead-induced carcinogenesis.

Although it is genotoxic and induces cell transformation, inorganic mercury has not been shown to be carcinogenic in animals or man. The underlying genotoxic mechanisms are less clear but formation of oxygen radicals has been shown. Furthermore, mercury ions interfere with sulfhydryl groups of the spindle fibers formed during cell mitosis. Whether this may lead to tumor formation is still under investigation.

6.2.4 Toxicity of Individual Metals

The metals that are described in greater detail, cadmium, lead, and mercury, were selected because of their particular relevance to the environment or to the workplace, as well as their toxicity. The consequences of deficiency of essential metals and the radioactive metals uranium, plutonium, and strontium are not discussed here.

Cadmium

Cadmium accumulates in the kidneys, the target organs for Cd toxicity.

Occurrence and Environmental Exposure Exposure to Cd occurs as a result of atmospheric emission during Cd production and processing, from combustion of fossil energy sources, from waste and sludge, from phosphate fertilizers, and deposition of waste and slag at disposal sites. The deposition of atmospheric Cd results in binding to and accumulation in soil substances such as humic acids in the upper soil layers. In comparison to metals such as Pb, Cd is readily absorbed by plants. Meat, fish, and fruits generally contain up to 50 μg Cd/kg fresh weight, whereas vegetables, potatoes, and grain products may contain up to 150 μg Cd/kg fresh weight. Higher concentrations are found in the kidneys of animals slaughtered for food, in wild mushrooms, and in seafood such

Table 6.10 *Typical cadmium content in foods.*

Cd (μg/kg fresh weight)	Food
≥200	Some mushrooms, molluscs, sea food (up to 10,000 μg/kg)
≤200	Liver and kidneys (up to 1000 μg/kg), various mushrooms
≤40	Grain and cereal products, potatoes, rice, vegetables
≤20	Bread, tomatoes, garden fruits, fresh water fish
≤5	Meat, wine, beer, fruit juices
≤1	Tap water, milk, dairy products

as mussels and oysters (Table 6.10). The accumulation of Cd in a particular food is related to the binding of the metal to specific proteins.

Cd concentrations have increased considerably in the human renal cortex since the beginning of the 20th century and today amount to approximately 30 mg Cd/kg fresh weight, reaching a maximum at 50 to 60 years of age. The sources of human exposure to cadmium are diet and smoking. Dietary-intake values calculated on the basis of recent food surveys, e.g. in Germany and the UK, are in the range of the Tolerable Weekly Intake (TWI) of 7 μg/kg bwt (body-weight) set by the WHO, although higher dietary exposures can occur. Smoking is a major contributor to total intake, since absorption of cadmium is higher upon inhalation (40–50%) than *per os* (2–8%). Smoking a pack of cigarettes per day results in an additional daily intake of about 2 μg cadmium, and thus exceeds the TWI from diet. This illustrates that lifestyle contributes significantly to total exposure and can result in a doubling of the Cd level in the renal cortex.

Toxicokinetics Depending upon particle size, approximately 10–50% of particle-bound Cd is absorbed from the air. About 5–7% of dietary Cd is absorbed but this may increase to as high as 10% in the face of relative Ca deficiency. Cd blood levels in adults with a normal body burden are generally less than 1 μg/l. In the liver Cd is bound to metallothionein. Cd–metallothionein is filtered in the kidney through the glomerular membrane and is reabsorbed from the proximal tubule cells. The absorption of metallothionein-bound Cd from the blood into the kidney is rapid and practically complete. In the case of chronic intake approximately 50% of the metal's body burden is found in the kidney, 15% in the liver, and 20% in muscle. Very low levels of Cd are observed in newborns because the placenta is an effective barrier to the metal

Chronic Toxicity The long-term inhalation of Cd-containing dust can cause inflammation, atrophy in the upper respiratory tract, and pulmonary emphysema. The renal cortex is the target tissue of chronic Cd toxicity regardless of the route of exposure and Cd nephropathy is the leading symptom of chronic Cd poisoning. Cd impairs the reabsorption function of the proximal tubule cells and, at higher exposure, also damages the glomeruli. This results in increased excretion of amino acids, of low-molecular weight proteins, e.g. β_2-microglobulin, and of Cd. The tubular dysfunction continues after Cd exposure has ended. The mechanism of Cd nephropathy is closely associated with Cd transport and binding of the metal to metallothionein. The protein moiety of the Cd–metallothionein taken up in the tubular cells is degraded with a half-life of a few days. The released Cd stimulates resynthesis of the protein. Released Cd interacts with

Zn. *In vitro* data suggest that Cd and Zn share the same transporter in the kidney. Cd is toxic to the transporter but pre-exposure to Zn may be protective.

Kidney dysfunction caused by Cd adversely influences the metabolism of bone minerals and results in an amplified excretion of calcium and phosphorus. Ca metabolism is impaired, so that Cd on the one hand binds to the Ca-binding protein in the small intestine and on the other hand inhibits the enzyme cholecalciferol hydroxylase, which is responsible for vitamin D_3 formation. The 'Itai-Itai' disease, a massive osteoporosis/osteomalacia with kidney dysfunction occurring in heavily Cd-contaminated regions of Japan after World War II, is now thought to result from both an increased Cd intake and a dietary deficiency of Ca and vitamin D.

Hypertension was observed in an animal study at Cd exposures without concomitant renal toxicity. The question of the association in humans of chronic exposure to low Cd concentrations and hypertension is currently not clear. Several studies on workers exposed to Cd and inhabitants of regions in Japan contaminated with Cd were unable to confirm such an association. However, among Japanese farmers who already exhibited proteinuria caused by Cd, the mortality from cardio- and cerebrovascular diseases was increased.

Critical Burden, Threshold Values The critical Cd concentration in the renal cortex is 200 μg/g. This value is reached in persons who inhale 13 μg Cd/m^3 daily at the workplace for 25 years. The same concentration is reached in a 50-year old adult, weighing 70 kg, after a daily ingestion of approximately 200 μg Cd. In 'uncontaminated' areas, the average intake for a 70 kg nonsmoker is in the range of 10–60 μg Cd/day. In smokers who regularly ingest food containing Cd, the potential for developing Cd-related health effects can no longer be ignored. The US EPA (Environmental Protection Agency) classifies cadmium as a probable human carcinogen (Group B1) and has established a reference dose (RfD) of 5×10^{-4} mg/kg/day in water and 1×10^{-3} mg/kg/day in food.

Lead

Lead toxicity is manifested as both hematological and neurological impairments and carcinogenicity.

Occurrence and Environmental Exposure Nonessential Pb is the most widely spread of all the metals in the environment and is found in all inorganic and organic systems. It is mainly used in the manufacture of batteries, pipes, cables, and paints. Tetramethyl- and tetraethyllead were used as gasoline additives. Human exposure to lead remains a serious public health problem. Today, the major route of exposure for the general population is food and drinking water. A definite decrease in the atmospheric lead concentration has occurred since it was banned for use in gasoline beginning January 1, 1996. Atmospheric Pb is deposited in the upper 1–3 inches of soil where the extent of accumulation depends on the humus content and the pH of the soil. The uptake of Pb by plants from the soil is relatively low except in broad-leaved plants such as lettuce, which can contain considerable contamination that may not completely be removed by washing. In foods of animal origin, meat, milk, and eggs

have relatively little lead contamination; liver and kidneys, however, are more heavily contaminated.

Toxicokinetics Whereas 30–80% of aerosol lead is absorbed through the lung, only about 8–15% of ingested lead is absorbed in the GI tract in adults. In children, intestinal Pb absorption may reach 50%. Pb appears in the bile but is partly reabsorbed via the enterohepatic circulation. In the large intestine Pb is transformed into the insoluble lead sulfide and is excreted with the feces.

More than 95% of the Pb circulating in blood is bound to erythrocytes. Systemically, Pb is initially found in the kidneys and liver. Later it is redistributed into calcium-rich tissues, such as bones and teeth, where it forms depots of insoluble Pb phosphate salts. The half-life of Pb in bone is more than 20 years. In soft tissues, such as kidney and brain, there is a labile Pb depot with a half-life of approximately 20 days. The Pb content of the urine correlates with the lead blood level. The normal values in nonexposed persons are about 5–20 µg Pb/l urine. Pb can penetrate the placenta; the Pb content in umbilical-cord blood correlates with the lead blood level of the maternal organism.

Lead Blood Level The blood Pb level provides a suitable measure of ambient Pb exposure. Upon a change in exposure it takes about two months until the blood level attains a new equilibrium. The WHO proposed a provisional TWI for adults of 25 µg/kg body weight, corresponding to about 215 µg lead per person per day. For occupational exposures to Pb, the employer must use engineering controls and work practices to achieve an occupational exposure of 50 $\mu g/m^3$ (0.006 ppm) or lower, based on an 8-hour time-weighted average. Although it has been suggested that a blood lead level of 30–40 µg/dl is acceptable, it would not protect against *in utero* effects or reductions in nerve conduction velocity, which have been associated with lower blood lead concentrations.

Average blood lead levels in the United States have fallen since the 1970s. In children the average level has decreased from 15 µg/dl in 1976–1980 to 2.7 µg/dl in 1991–1994. However, 4.4% demonstrated elevated blood lead levels in excess of 10 µg/dl. On a body-weight basis the uptake of lead in infants exceeds that in adults. For infants in general, and for especially for nonbreast-fed babies, the Pb content of drinking water has a particular significance. The suggested action level in drinking water is 15 µg/l.

Acute Toxicity Acute intoxication with Pb is very rare because of its relatively poor absorption and the large storage capacity of bones and erythrocytes for Pb. After ingestion of high doses of inorganic Pb compounds symptoms include: vomiting, abdominal colic, subnormal temperature, and hypotension. In addition, damage to the liver, kidneys, and the CNS, as well as development of hemolytic anemia and hemoglobinuria, may be observed.

Chronic Toxicity

Erythrocytes, the bone marrow, and the nervous system are primarily affected in chronic Pb poisoning.

Hematological effects Lead-induced anemia is characterized by a shortened erythrocyte life span and an impairment of heme synthesis. The effects of lead on heme synthesis are due to a stimulation of δ-aminolevulinic acid synthetase and a depression of δ-aminolevulinic acid dehydratase, coproporphyrinogen decarboxylase, and ferrochelatase. As a consequence of the effect on δ-aminolevulinic acid synthetase there is a marked increase of δ-aminolevulinic acid (ALA) in blood and urine. The effect on ferrochelatase, which catalyzes the incorporation of Fe into protoporphyrin, results in depressed heme formation. The changes in enzyme activities as well as the urinary excretion of ALA correlate well with lead blood levels and are, therefore, used as sensitive biochemical markers for current Pb exposure.

Neurological effects

Chronic Pb exposure leads to significant impairment of both the peripheral and central nervous systems. The central nervous system is particularly susceptible to Pb-induced damage both *in utero* and in early life.

The developmental processes leading to the adult CNS continue after birth and may not reach a mature state for several years. Early in life the blood–brain barrier does not exclude the entry of lead into the brain and as a result Pb can reach critical centers and alter the normal process of brain maturation at exceedingly low blood lead levels, e.g. at blood lead levels as low as 10 μg/dl. These effects may be quite subtle and not necessarily detectable using common clinical examinations. However, decrements in intelligence, attention deficit disorder, hearing impairment, and a variety of related persistent consequences such as poor scholastic performance, absenteeism, deficits in reading and vocabulary, and physiological shortcomings including hampered motor skills, slow reaction time, and poor hand–eye coordination have been reported in children and appear to persist long after childhood Pb exposure. In contrast, high blood lead levels can produce encephalopathy, which may include hyperirritability, ataxia, convulsions, stupor, and coma or death. These effects may develop at blood lead levels of of 70–80 μg/dl or higher. At doses between these extremes there is an increasing likelihood of neurological and behavioral damage.

Peripheral Pb neuropathy is based on dysfunctions of the motor nerve fibers and leads to paralysis of the upper extremities at blood levels of 500–700 μg Pb/l. Smooth muscle can also be paralysed, causing painful colic with persistent constipation in the large intestine. Central nervous system disturbances after chronic Pb exposure (Pb encephalopathy) are manifested by fatigue, headache, dizziness, and tremor. The pathogenesis of the neuronal changes caused by Pb is not well understood. Possibly the metal influences the excitatory transmission at the synapse by interfering with neurotransmitter function or disrupting calcium metabolism.

Organic Pb Compounds

Tetramethyl- and tetraethyllead, once used as antiknock agents in motor fuels, were the most significant sources of human Pb exposure.

Tetramethyl- and tetraethyllead are volatile and lipophilic derivatives of Pb, which can be taken up via the lung or upon direct contact via skin absorption. Tetra-alkyllead compounds are oxidatively metabolized to the corresponding trialkyllead compounds. The trialkyllead ion is the more stable active form with a half-life of several days. Both substances are far more neurotoxic than is inorganic Pb. Symptoms such as psychomotor excitation, convulsions, delirium, and later paralysis occur upon acute exposure to high concentrations. Mild neurological symptoms can occur as a consequence of chronic workplace exposure or as a consequence of sniffing leaded gasoline. Exposure to organic Pb compounds cannot be established by the lead content of the blood. Poisoning by organo-Pb compounds is not treatable by means of chelators.

Carcinogenicity More recent epidemiological studies investigating workers exposed to lead indicate slightly increased relative risks of tumors of the stomach and lungs. Tumors of kidneys, bladder, peripheral and central nervous system, prostate, liver, hypophysis, and thyroid, which were observed in lead acetate-treated rats, have not been verified.

Mercury

Occurrence and Environmental exposure

Mercury is transformed into dimethylmercury by aquatic microorganisms and, thus, accumulates in the food chain. Dimethylmercury from fish and elemental mercury from amalgam represent the major sources of human exposure (Table 6.11).

The toxic properties of mercury (Hg) were first observed by the Romans who used it for the refining of gold. It is the only metal that is liquid at ambient temperature and, therefore, has a considerable vapor pressure, which gives rise to gaseous Hg vapor, which plays an essential role in global Hg cycling. Natural emissions from volcanoes and evaporation from soil and water surfaces contribute about 1000 tons of Hg per year to the total atmospheric pool of approximately 5200 tons. Anthropogenic emissions (from combustion of fossil fuels) add about 2600 tons/yr. Emitted Hg resides in the atmosphere for approximately one year. Steady-state levels are maintained by its removal as a result of oxidation and rain, which promote terrestrial and oceanic deposition. In aquatic sediments, inorganic Hg is transformed by methanogenic bacteria into dimethylmercury, which enters the food chain mainly through fish.

The global primary production of mercury is estimated at 2000 tons per year while recycling contributes an additional 800 tons. Concerns about the damage to the environment have led to the reduction of the consumption rate. Use and application of

Table 6.11 *Mercury intake of adults without occupational exposure (μg/day).*

Source	General population	Fisher families[a]
Diet	5[a]–20[b]	up to 350
Amalgam	2.5–17.5	

[a] Dimethyl Hg.
[b] Total Hg.

mercury is mainly in the chemical and electrical industries where it is used for diverse purposes such as chloralkali electrolysis, synthesis of aldehydes and fungicides, and manufacture of switches and fluorescent light bulbs.

The biological effects of mercury are associated with its three principal chemical forms:

1. Elemental Hg vapour;
2. Inorganic Hg salts (Hg_2^{2+}, Hg^{2+});
3. Organic mercury compounds.

Occupational exposure occurs through mainly Hg vapor, whereas the general population is exposed to dietary dimethyl-Hg in fish and to Hg vapor released from dental amalgam.

Toxicokinetics

The toxicokinetics of Hg vapor, its inorganic salts, and of organically bound mercury are determined by variations in lipophilicity and permeation properties of these compounds.

Mercury vapor Approximately 80% of an inhaled dose of Hg vapor reaches the blood after alveolar absorption. Between one-half and two-thirds of the absorbed amount are associated with the red blood cells. The alveolar absorption is favored by the high lipophilicity of the vapor. In contrast, less than 1% is absorbed upon oral ingestion of the liquid metal. Because of its lipophilicity, the atomic vapor is rapidly distributed to body fluids, cells, and organs (Figure 6.5). In specific cells, e.g. erythrocytes and liver cells, the activity of catalase induces the oxidation to Hg^{2+} (Figure 6.6), which is a prerequisite for the binding of mercury to intracellular ligands. Thus, the distribution of the vapor after inhalation is similar to that after ingestion: more than half of the body burden is found in the kidney, and the second largest portion in the liver. After exposure to the vapor, mercury is eliminated as Hg^{2+} with a half-life of approximately 60 days; approximately

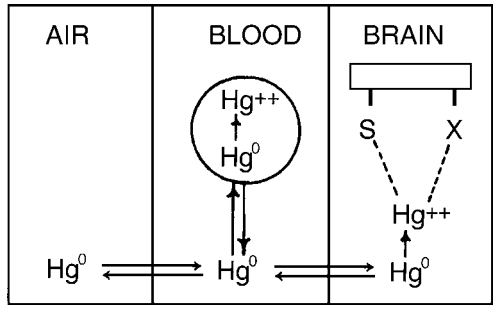


Figure 6.5 *Schematic diagram of the distribution and oxidation of elemental mercury in the organism. [Reprinted from Clarkson et al. Copyright (1980), by permission from Ann Arbor Science.]*

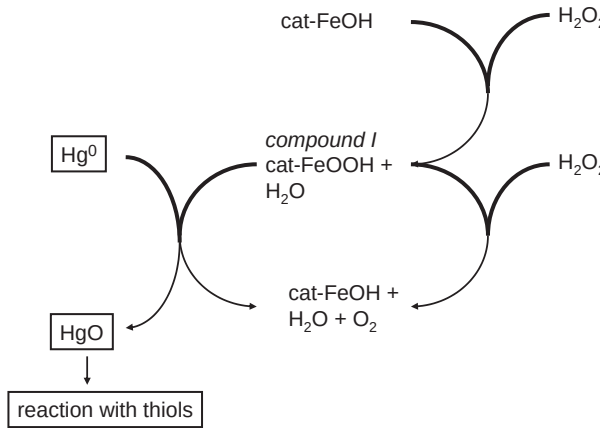


Figure 6.6 *Mechanism of the biological oxidation of elemental mercury by catalase and H_2O_2. Cat-FeOH: catalase; Cat-FeOOH: catalase–H_2O_2 intermediate (compound I).*

40% is excreted with the feces and 60% in the urine. The elimination half-time for the brain seems to be in the range of years, possibly because of the formation of biologically inactive complexes of mercury with selenium. Brain and kidney are the critical target organs for Hg^0 and Hg^{2+}.

Inorganic mercury The toxicokinetics of ionic mercury differ from those of the elemental form particularly with respect to uptake. About 5–10% of Hg^{2+} is absorbed after oral ingestion or after dermal exposure. The elimination of Hg^{2+} proceeds with a half-life of between 40 and 60 days.

Organic mercury Among these compounds, the short-chain alkyl mercurials, especially dimethylmercury, have the greatest importance for toxicology. Approximately 80–90% can be absorbed in the lungs and the gastrointestinal tract. About 90% of the dimethyl-Hg absorbed into blood is bound to the erythrocytes. Major accumulation occurs in the liver, while the brain and kidneys retain relatively small amounts. The main route of excretion is the bile (90%), followed by the urine (10%). In bile, dimethylmercury is excreted as complexes with glutathione and cysteine. Approximately half of the dimethylmercury in bile is demethylated to inorganic Hg and directly excreted, whereas the intact dimethylmercury is reabsorbed to a great extent and again enters the enterohepatic circulation. The elimination half-life of dimethylmercury is 70 days.

Acute Toxicity

The acute toxicity of mercury vapor and of ionic mercury is characterized by effects on the lung, the kidney, and the mucosal membranes of the gastrointestinal tract.

Acute toxicity of Hg can result from the ingestion of soluble Hg salts or as a result of the inhalation of Hg vapor. Mercuric bichloride has been described as ‘corrosive

sublimate.' Ingestion of solutions at 105 or greater causes serious lesions of the mucosal membranes with subsequent necrotic inflammation of the intestinal mucosa, abdominal cramps, and bloody diarrhea. In cases of attempted suicide intestinal membranes show evidence of ulceration with necrosis. Loss of blood can be sufficient to induce shock and circulatory collapse. If the patient survives the initial insult, damage to the intestinal mucosa can augment the absorption of the salts, which attack the proximal tubules of the kidney, leading to complete kidney failure. In contrast, mercurous chloride is much less toxic, in part because it is much less soluble. Inhalation of elemental mercury vapor may be accompanied with acute pneumonia, respiratory distress, and bronchitis-like symptoms. Nowadays, acute intoxications with mercury compounds are rare.

Chronic Toxicity

Mercury compounds with short alkyl chains cause irreversible lesions to the peripheral and central nervous system. The critical organs for exposure to mercury vapor are brain and kidney.

Elemental mercury is quite lipophilic and can enter cells by diffusion, and therein form oxidized ionic forms, which can exert their toxic effects within the cell. The brain and kidney are particularly sensitive to extended exposure to elevated concentrations. Typical signs of poisoning are tremor of the hands, changes in personality and behavior, and an increased irritability and impairment of memory. Mercury-induced nephrotoxicity originates from the formation of antibodies against the glomerular basal membrane, which causes a nephrotic syndrome and proteinuria. In addition, damage may also occur to tubular cells, resulting in the urinary excretion of enzymes and a decreased reabsorption of zinc and copper. The accumulation of inorganic mercury in kidney tissue induces an elevated urinary excretion of the metal. The ensuing correlation between mercury levels in ambient air and in urine is useful in monitoring the occupational exposure to mercury vapor. The Hg concentration in urine also reflects the absorbed dose upon long-term exposure to dusts of inorganic Hg, which is frequently associated with coexposure to the vapor. In such cases, disturbances of renal function prevail over neurological symptoms with increasing exposure to the inorganic form.

The subacute and chronic toxicity of dimethylmercury is characterized by a syndrome of marked disturbances of the extremities, uncoordinated gait, and impairment of speech, hearing, and vision. Manifestation of the symptoms is preceded by a latency period extending from several weeks to months. Detailed investigations of the epidemic outbreaks of dimethylmercury poisoning in Minamata (Japan) and in Iraq revealed that the neurological disturbances become apparent at body burdens between 25 and 40 mg, while above 200 mg the rate of mortality rises.

Fetotoxicity Dimethylmercury is a very potent neurotoxic agent with serious effects on the neuronal development of the fetus. The neonatal brain seems to be three to four times more sensitive than is the adult organ. The mercury concentration in cord blood is 1.2- to 1.9-fold higher than in maternal blood, even at moderate levels of fish consumption by the mother. In addition, the newborn is postnatally exposed to the mercury in breast milk, which has a concentration between 1- and 3-times of that in maternal blood at

background exposure. Owing to the marked neurotoxicity in early childhood, the WHO recommends that pregnant women should reduce the possibility of long-term exposure to elevated levels of dietary dimethylmercury.

Monitoring, Reference, and Threshold Limit Values The general population has a mercury burden below 1–2 μg/g of kidney, 0.25 to 0.75 μg/g of liver, and less than 0.25 μg/g for most other tissues. Appropriate indicator media for the exposure to inorganic mercury and Hg vapor are urine and blood, while the exposure to dimethylmercury is well correlated to the concentration in blood and in scalp hair. Values below 5 μg/l in blood and below 10 μg/l in urine are considered normal. The occupational exposure to Hg vapor indicates lowest observed adverse effect level (LOAEL) values for tremor between 50 and 100 μg/l in urine. A LOAEL of 300 μg/l in whole blood for paresthesia could be derived from the dimethylmercury-induced epidemics in Japan and Iraq; this corresponds to a daily intake of 300 μg dimethyl-Hg by an adult person. Assuming an uncertainty factor of 10, WHO deduced an acceptable daily intake (ADI) of 40 μg total mercury of which not more than 30 μg should be dimethyl-Hg (Table 6.11). Recommended exposure limits for workplaces vary between 0.05 and 0.1 mg/m^3 in air, 15 and 25 μg/l in whole blood, and 35 and 100 μg/l in urine. A environmental concentration of 1 μg/m^3 in the inhaled air is considered acceptable.

Aluminum

Occurrence and Exposure Human Al intake via the oral route may range up to 90 mg/day. Al-containing medicines may increase exposure. Data currently available indicate that aluminum plays only a minor role in biological processes. It cannot be stated with certainty whether it is an essential element because of the omnipresence and abundance of aluminum. It is practically impossible to rear and keep animals from exposure to Al despite an aluminum-free environment and an aluminum-free diet.

Toxicokinetics In the acidic medium of the stomach, and a short distance into the duodenum, Al is present as a cation. In the alkaline milieu of the small intestine Al precipitates as polymeric aluminum hydroxide. Some Al can remain in solution as an organic complex. At higher pH levels it becomes partly ionized again. Thus, although Al in an aluminum hydroxide antacid is subject to a series of conversions in the gastrointestinal tract, it is poorly absorbed.

Al is transported through the intestinal wall together with other metals using divalent-cation transporters. In blood Al can be carried by transferrin. Passive absorption seems to begin when the Al intake is about 125 mg per day. In total about 300 mg aluminum can be detected in the body of a healthy adult. The daily balance is approximately 3–30 mg aluminum and is in equilibrium in healthy persons. Fecal excretion removes about 95% of orally ingested Al. A small portion of the absorbed Al is returned to the intestine in the bile. Most of the absorbed Al is eliminated via the kidney.

Aluminum is enriched in the lysosomes of the proximal tubules in the form of $AlPO_4$ and enters the tubular fluid as insoluble particles of 50-Å diameter. Even patients ingesting large amounts of aluminum in medicines rarely excrete more than 1000 μg aluminum per day in the urine. Transiently increased levels of aluminum in blood and plasma are, without further information, not reliable indicators of deposition of the metal

in tissues. In persons with renal damage, aluminum excretion is impaired and increased aluminum concentrations in the blood can result in an enrichment of aluminum in various tissues.

Marked retention of aluminum is seen in almost all patients who are dependent on dialysis, especially in those who must ingest large amounts of antacid medicines containing aluminum to prevent hyperphosphatemia. As mentioned above, in patients with relatively high blood plasma aluminum levels, encephalopathy with pathognomonic enrichment of aluminum in the brain accompanied occasionally by changes in the skeletal system and other disorders have been observed.

Mode of Action A number of biochemical reactions of aluminum in the mammalian organism are known. It binds to protein, especially transferrin and to nonhistone proteins and DNA in cell nuclei. Vitamin D and parathyroid hormone seem to affect the homoeostasis of aluminum. In general, aluminum has an inhibiting effect on phosphate-transferring systems, probably because it combines with calmodulin, displacing calcium.

Chronic Toxicity Shortly before and during the Second World War the frequency of disorders in which aluminum was ascribed a causal role increased for workers in factories in which fine aluminum powder of respirable size was produced. Employees developed pulmonary fibrosis and this aluminosis has since become known as an occupational disorder. Since then, genuine aluminosis seems to have become extremely rare.

Since the beginning of the 1970s, when dialysis was introduced for patients suffering from renal insufficiency, various diseases have become frequent, namely encephalopathy, osteomalacia, and microcytic hypochromic anemia associated with other pathological conditions. In these diseases the metal was found in blood, bone, and brain tissue in remarkably high concentrations. The etiological involvement of aluminum is virtually certain because a reduction in the aluminum intake leads to a visible improvement in the symptoms.

It has been suggested that aluminum is involved in the development of *morbus Alzheimer* and in one form of amyotrophic lateral sclerosis since increased aluminum levels were found in the brain and spinal cord of patients. In these two diseases specific pathological processes may cause secondary enrichment of aluminum in the tissues.

At the workplace, only particles with an aerodynamic diameter less than 5 micrometer are considered to induce lung fibrosis, because they are small enough to enter the alveoli. Various national occupational safety regulations are based on this figure, and are considered to prevent fibrosis of the lungs (aluminosis), metal-fume fever, and critical increases in the level of serum aluminum.

Antimony

Occurrence and Exposure Antimony is a common trace element that normally occurs with sulfur and arsenic. It is used in alloys for batteries and motor bearings, and in the electronics industry for the manufacture of thermocouples. The tartrate is used for treatment of tropical parasites such as trypanosomes. Antimony oxides are released into

the environment from smelters, coal-fired power plants, and also from volcanoes. In the smelter industry workplace concentrations have reached up to 10 mg/m^3.

Toxicokinetics and Mode of Action The toxicity of antimony compounds depends on their oxidation state and solubility. Oral application of metallic antimony has the same systemic effects as have antimony trioxide and antimony potassium tartrate, which indicates that even metallic antimony is bioavailable.

At physiological conditions trivalent antimony seems to be the most stabile form. It reacts with thiol groups, especially with vicinal dithiols, and thereby acts as an enzyme inhibitor. Reaction of trivalent antimony with glutathione leads to reduced glutathione concentrations in the liver, and thereby to cytotoxicity and reduced glutathione conjugation of foreign compounds. There is an indication that antimony, similar to arsenic, inhibits the repair of DNA damage.

Like arsenic, antimony also seems to be methylated in the human organism. In workers employed in producing batteries, significantly more trimethylantimony dichloride was found in their urine when compared with controls. Methylstibane ($MeSbH_2$) and dimethylstibane (Me_2SbH) has been demonstrated in the urine of test subjects.

Chronic Toxicity Long-term inhalation of antimony trioxide or antimony ore dust (46% antimony, mainly as antimony trisulfide) caused lung tumors in female rats. There is an indication of an increase in lung cancer mortality in workers exposed to the one, metallic antimony, and antimony triodide. The mechanisms of the genotoxic and carcinogenic potential of antimony and its inorganic compounds are not well understood.

Arsenic and Inorganic Arsenic Compounds

Occurrence and Exposure Arsenicals have been widely used as insecticides, herbicides, fungicides, and rodenticides. Since it is known to promote the growth of animals arsenic has been used as an nutritional additive to enhance growth of pigs and poultry.

Toxicokinetics and Mode of Action Arsenic and inorganic arsenic compounds are absorbed via inhalation and ingestion and distributed rapidly to all organs. Accumulation is observed particularly in the liver, kidneys, and lungs. The metabolism of the inorganic arsenic compounds is independent of the route of absorption. Reduction and oxidation reactions lead to mutual transformation, with the preferred reduction of arsenate to arsenite. Only arsenite can be methylated to form first monomethyl (MMA) and then dimethyl arsenite (DMA). DMA, the main metabolite of the inorganic arsenic compounds, is excreted with the urine. At least for the genotoxic effects of inorganic arsenic compounds there is increasing evidence that DNA-damaging metabolites are also formed as a result of the methylation.

As inorganic arsenic compounds, arsenic(III) compounds react readily with SH groups of proteins, which is regarded as a mechanism for the toxicity. However, the mechanism of arsenic toxicity is more complex. Arsenic(III) compounds inactivate, possibly as a result of the reaction with SH groups, the phosphatases responsible for the inactivation of the *jun-N*-terminal kinases (JNKs). This results in an increase in the activity of the JNKs

and of proto-oncogenes of the *c-jun* family. This activation of JNKs may contribute to the carcinogenicity of arsenic(III) compounds.

A suggested mechanism for the genotoxicity of the substance is the generation of reactive oxygen species. DNA damage caused by arsenic(III) compounds is the result of the calcium-mediated formation of peroxynitrite, hypochlorous acid, and hydroxyl radicals. DMA forms a radical, possibly intracellularly, and is transformed in the presence of molecular oxygen into a peroxyl radical, which could be responsible for the observed oxidative DNA damage. Other studies show inorganic arsenic compounds to inhibit DNA-repair processes and by that regulation of repair.

Furthermore, interference of inorganic arsenic compounds with DNA methylation may result in changes in gene expression, as determined for the tumor suppressor gene *p53* and the oncogene *c-myc*. Moreover, inorganic arsenic compounds make certain cells more sensitive to mitogenic stimulus and increase cell proliferation and inhibit cell differentiation.

The disrupted regulation of certain interleukines is held responsible for impairment of the immune system in patients exposed to arsenic.

Acute Toxicity In case studies of the acute toxicity of the arsenic compounds, the first symptoms occur 30 to 60 minutes after ingestion and lead to acute paralytical and acute gastrointestinal syndromes. The acute paralytical syndrome is characterized by cardiovascular collapse, central nervous weakness, and death within hours.

Characteristic of the gastrointestinal syndrome is a metallic or garlic-like taste, a dry mouth, burning lips, difficulties swallowing, headaches, dizziness, and vomiting. This can then lead to multi-organ failure.

Chronic Toxicity Long-term exposure to arsenic in humans causes fever, sleep disorders, weight loss, swelling of the liver, dark discoloration of the skin, sensory and motor neuropathy, and encephalopathy with symptoms such as headaches, poor concentration, deficits in the learning of new information, difficulties in remembering, mental confusion, anxiety, and depression. Long-term exposure can also cause peripheral vascular effects such as acrocyanosis, Raynaud's disease, and tissue necrosis on the extremities (blackfoot disease). Cardiovascular diseases and the occurrence of *diabetes mellitus* are also described.

Long-term exposure to dust containing arsenic causes irritation of the conjunctiva and mucous membranes of the nose and throat. The pustular or follicular skin reactions observed after contact with inorganic arsenic compounds are usually attributed to irritative effects and not to sensitizing effects of the arsenic compounds. Possible contact allergy after exposure to inorganic arsenic compounds cannot be deduced with certainty as no clear separation from unspecific or toxic effects is possible.

Arsenic and inorganic arsenic compounds, with the exception of arsine, are carcinogenic in humans. After inhalation of the substance, carcinogenic effects are observed in the lungs and, after ingestion, in the bladder, kidneys, skin, and lungs. Increased incidence of respiratory cancer has been observed in workers in the smelting industry, in gold mines, and among employees in the manufacture and packaging of pesticides or who have sprayed pesticides. Inorganic arsenic compounds are clearly genotoxic in somatic cells *in vitro* and in rodents and humans.

Beryllium and its Inorganic Compounds

Occurrence and Exposure Beryllium is used to produce fatigue-resistant alloys, in neutron generators and nuclear reactors. Hazardous exposures occur in various fabrication industries by inhalation and skin contact. Environmental concentrations result from the burning of coal and oil, but are of minor concern.

Toxicokinetics Beryllium and its inorganic compounds are mainly absorbed by inhalation. The particle size and water solubility are decisive for the absorption. Nonionized soluble beryllium compounds are rapidly removed from the lungs, while the ionized forms remain in the lungs. Beryllium is transferred from the lungs into the blood in two phases. The rapid elimination of the substance via mucociliary clearance during the first two weeks is followed by a slower elimination in which beryllium is absorbed by alveolar macrophages and transported to the tracheobroncheolar lymph nodes. In man beryllium remains in the lungs for several years and also serves as evidence of previous exposure in cases of chronic beryllium disease. Beryllium and its inorganic compounds are usually not metabolized and are eliminated mainly in the urine and also with feces.

Acute Toxicity Acute beryllium disease occurs after inhalation exposure to high beryllium concentrations ($>100\,\mu g/m^3$) and is characterized by symptoms of acute pneumonitis, mostly by the direct toxicity of water-soluble beryllium compounds.

Chronic Toxicity Unlike acute beryllium disease, chronic beryllium disease (berylliosis) can also be triggered by insoluble beryllium compounds. Chronic effects are characterized by the formation of granulomas and functional restrictions of the lungs, such as reduction of the vital and total capacity and a reduced diffusion capacity. Repeated inhalation exposure to beryllium may also lead to cardiovascular, renal, hepatic, and hematological effects, and to weight loss.

As a result of direct irritation, water-soluble beryllium compounds may cause poorly healing dermatitis. If undissolved particles enter the skin as a result of wounds or impaired skin barriers, ulcers or necrosis may develop. Beryllium and its inorganic compounds may be the cause of allergic contact eczema or granulomatous skin reactions of immunological origin.

Soluble beryllium compounds induce sister chromatid exchange, chromosomal aberrations, and gene mutations in mammalian cells. Genotoxic effects are probably the result of the induction of DNA–protein complexes and the influence on DNA polymerases. Also, an indirect mechanism, the hypermethylation of promotor sequences, which leads to the inactivation of the corresponding genes, has been suggested.

Beryllium and its inorganic compounds are carcinogenic in humans. Increased mortality from lung cancer was determined in different retrospective studies of workers in beryllium-production plants. Death from lung cancer was disproportionately frequent in persons who had previously suffered from acute beryllium disease.

The lung damage caused by beryllium and its inorganic compounds is mainly provoked by an inflammatory process. Long-term effects probably depend more on the chemical and crystallographic properties and the surface properties of the particles. In chronic beryllium disease, in which the insoluble compounds are as effective or even

more effective than the soluble ones, there is an allergic reaction of the delayed type, in which beryllium appears to function as a hapten and stimulates the local proliferation and accumulation of beryllium-specific T-cells in the lungs.

Chromium and its Compounds

Occurrence and Exposure Chromium(VI) compounds are used for the manufacture of chromium-containing chemicals, metal treatment, magnetic tapes, montan wax, vitamin K, as mordants in wool dying, catalysts, and other minor uses. Chromium(VI) released into the environment is rapidly reduced to the 100–1000-fold less toxic chromium(III). It is present in human food, but more than half of the dietary chromium derives from food processing.

Toxicokinetics and Mode of Action The toxicity of chromium and its compounds is associated mainly with chromium in oxidation state VI. Chromium(VI) compounds are readily absorbed from the gastrointestinal tract, the skin, and by inhalation, whereas chromium(III) is not. Owing to their poor absorption, chromium(III) salts have none of the toxic effects demonstrated for the oxidation state VI compounds; only dermatitis has been described.

Acute Toxicity Chromic acid and its salts are corrosive or irritating, and chromates sensitize the skin. In animal studies, there is no difference in the acute toxicity, skin irritation and tolerability of intratracheally administered Cr_2O_3 (oxidation state III) and CrO_2 (oxidation state IV). The systemic effects of chromium(VI) compounds absorbed orally or percutaneously can include nausea, vomiting, diarrhea, spasms, and hemorrhagic nephritis.

Chronic Toxicity Chromium(II) compounds are not very stable. Their toxicology, like that of chromium(V) compounds, has not been described. Adverse effects of these compounds on health are still unknown; neither effects in humans nor results of animal experiments have been published. Systemic effects of metallic chromium in humans have not been described.

High concentrations (2%) of chromium(III) and chromium(VI) salts denature proteins at low pH (<5.4) and precipitate nucleic acids. Chromium catalyses various enzyme reactions. It is unknown to what extent such mechanisms are of significance in cancer development. Mutagenicity of inorganic chromium(VI) compounds is apparent if the bioavailability of the chromium(VI) is adequate.

Chromium salts are skin sensitizers in humans.

An increased incidence of malignant lung tumors has been seen in workers in industries producing chromate and chromium(VI)-based pigments. However, epidemiological studies have been unable to provide an answer to the question of whether or not chromic acid and the readily water-soluble chromates and dichromates have carcinogenic activity. In experimental animals the carcinogenic activity of sodium dichromate was only demonstrable after administration of toxic doses. Even under such

extreme conditions the carcinogenic activity of the substances was weak. Most of the lung tumors were benign and small and in no case were they the cause of death. Therefore, it cannot be stated whether or not the carcinogenic effect in an inhalation study is only demonstrable in the toxic range or close to it, as is the case in the instillation test.

Poor water solubility of some chromium(VI) compounds may contribute to their carcinogenic properties. Studies have been performed which featured intramuscular or intrapleural chromium salt injections or intrabronchial implantation of chromium-containing pellets. The deposited materials dissolve at a slow rate thereby permitting the accumulation of persistent high local chromium concentrations. It can be assumed that other readily water-soluble chromium(VI) compounds such as chromic acid and readily water-soluble chromates and dichromates are similarly carcinogenic. Currently available data indicate that a cancer risk exists only after inhalation of these substances.

Chromyl dichloride, a liquid with significant vapor pressure, decomposes in water to yield chromic acid and HCl. Since it yields chromic acid on contact with water, and because its genotoxicity in the Ames test is comparable to that of other chromium(VI) compounds, chromyl dichloride may also be carcinogenic.

Chromium(VI) in Cement Contact dermatitis is one of the most frequently reported disorders among construction workers who regularly work with cement. Chromium dermatitis can also be induced by other causes, particularly from chromium compounds in leather gloves. Chromium, especially chromium(VI), is a well known skin sensitizer. Water-soluble chromium(VI) compounds have a higher skin-penetration rate than do the less water-soluble chromium(III) compounds. The high alkalinity of the cement–water suspension results in irritation, thereby enhancing absorption of chromium(VI). In the skin, chromium(VI) is transformed into chromium(III), which readily binds to proteins and forms the hapten-carrier complex that acts as an antigen. As a protection measure, reduction of chromium(VI) to the less bioavailable chromium(III) is achieved by adding 0.35% ferrous [ion(II)] sulfate to the cement, which reduces the concentration of chromium(VI) to less than 2 ppm. The bioavailability of chromium(III) is further reduced by formation of insoluble precipitates of chromium hydroxide in the alkaline cement-water mixture. Reduction of chromium(VI) in cement to less than 2 ppm of chromium(VI) compounds has significantly reduced the prevalence of allergic cement eczema in workers.

Cobalt and Cobalt Compounds

Occurrence and Environmental Exposure Cobalt is a relatively rare element. It is used to produce parts of jet engines, gas turbines, and other devices that operate at high temperature, magnetic steels, and aluminum alloys. Human dietary cobalt intake is about 20–40 microgram per day; airborne exposure contributes a smaller portion.

Toxicokinetics Cobalt, being a constituent of vitamin B12, is an essential element. Absorption occurs in the small intestine and is coupled with iron uptake. Renal excretion of inhaled cobalt is biphasic. Most is eliminated with a half-life of 10 days, the remainder with a half-life of 90 days.

Acute Toxicity Cardiotoxic effects and polycythemia have been observed in persons who had ingested high doses of cobalt and in experimental animals after injection of the substance. Fatal congestive heart failures occurred among consumers of up to 12 liters of beer to which cobalt sulfate at 1 mg/l had been added as a foam stabilizer. Cardiomyopathy after occupational exposure has also been described occasionally.

Chronic Toxicity Long-term inhalation of cobalt dust can lead to diffuse inflammatory reactions of the bronchial mucosa and then to the development of chronic respiratory disease. It is conceivable that immunological processes are involved because sensitization to cobalt after inhalation and after dermal exposure has been described.

Cobalt and cobalt compounds are genotoxic *in vitro* and in the bone marrow of mice and hamsters. The genotoxic effect of cobalt ions probably takes place via the production of radical oxygen species, as 5 of 9 of the mutations found in tumor tissue within a carcinogenicity study with cobalt sulfate in mice were G–T transversions in codon 12 of the K-ras oncogene, which is seen as indication of indirect DNA damage by oxidative stress. This hypothesis is supported by direct proof that radical oxygen species are produced by the interaction between cobalt, tungsten carbide, and oxygen in hard metal dusts. Cobalt ions also inhibit repair of DNA damage. Empirical evidence in humans also provides indication for disturbed DNA repair due to cobalt. Both an increased number of DNA single-strand breaks and a reduced repair capacity for oxidative DNA damage in lymphocytes were found in 8 metal workers, who were exposed to $>4\ \mu g/m^3$ cobalt at the work site.

In inhalation tests with rats and mice the readily soluble cobalt sulfate induced lung tumors as well as pheochromocytomas. In rats intratracheal application of cobalt(II) oxide also induced lung tumors. Mechanistic studies indicate that the cobalt ion is the effective agent, whereas the solubility of metallic cobalt and the different cobalt salts determines their bioavailability and by that their potency.

Reliable epidemiological data demonstrating an increased incidence of tumors after occupational exposure to cobalt or its compounds are not available. All the data indicate that cobalt and cobalt compounds can be carcinogenic if the cobalt ion is bioavailable.

Copper and its Inorganic Compounds

Occurrence and Exposure Copper is an essential trace element, a constituent of many proteins and of more than 20 enzymes with important functions in cellular respiration, cellular energy metabolism, connective tissue biosynthesis, and iron metabolism. It also plays an important role in the regulation of gene transcription. The daily intake is between 1 and 4 mg, mainly via food but in smaller amounts also via inhalation and skin contact.

As a transition metal, copper is able to accept or donate one electron and thereby initiate redox reactions resulting in the formation of oxygen radicals. Copper ions are thus important catalytic co-factors for enzymatic redox reactions. Examples of copper-binding enzymes are copper–zinc superoxide dismutase, cytochrome c oxidase, dopamine β-hydroxylase, and ceruloplasmin (ferroxidase). Copper–zinc superoxide dismutase plays an essential role in the cellular defence against reactive oxygen species, such as the superoxide radical, which are normally formed during the cellular metabolism.

Toxicokinetics and Mode of Action The toxicity of copper can be explained by its direct effects on the structure and function of proteins, membranes, and DNA as well as by the reactivity of the oxygen radicals formed in the copper redox cycle. Therefore, it is assumed that reactive copper can lead to oxidative cell damage, such as lipid peroxidation, thiol oxidation, and DNA damage.

Acute Toxicity In general, single oral doses of copper salts by humans primarily causes reversible gastrointestinal complaints accompanied by nausea and vomiting. The no-observed adverse effect level (NOAEL) for nausea and vomiting is a copper concentration of 2–4 mg/l water.

Chronic Toxicity Gastrointestinal complaints have been associated with repeated oral ingestion of tap water containing copper concentrations of several milligrams per liter.

Newborns and infants react with particular sensitivity to high copper concentrations in drinking water. In neonates who are not breast-fed [or breast-fed only for a short time] high copper concentrations in drinking water are capable of producing liver damage with hepatomegaly, jaundice, ascites, hepatic coma, and hepatic failure.

Over recent years, investigations with the Long-Evans cinnamon (LEC) rat have yielded information on the mechanism of chronic copper intoxication. Long-term accumulation of copper in the liver of LEC rats involves the uptake of copper-laden metallothionein by lysosomes, where it is incompletely degraded and polymerizes to form an insoluble material containing oxidized metallothionein and reactive copper. This copper, together with iron, seems to damage the lysosomes via oxidation and results in hepatocyte necrosis. Subsequent to phagocytosis by Kupffer cells, accumulated reactive copper may amplify liver damage either directly or indirectly through stimulation of the Kupffer cells.

The mechanism underlying the carcinogenicity of copper in the LEC rat is attributed to the activation of the serum response factor (SRF) initiated by copper. The SRF is an important regulator for inducing the proto-oncogene c-fos. In the LEC rat liver, the concentrations of etheno-DNA adducts correlate with the copper concentration. Since oxygen radicals trigger lipid peroxidation, and since the etheno-DNA adducts are formed from lipid peroxidation products, it is assumed that formation of oxygen radicals is involved in cancer-inducing effects of copper in the LEC rat.

Soluble copper salts are not mutagenic in bacteria. Genotoxicity studies on mammalian cells *in vitro* revealed positive as well as negative results. Soluble copper salts induced chromosome aberrations and micronuclei in the bone marrow of some mice after intraperitoneal injection. Altogether, these studies indicate a clastogenic activity of the copper compounds. The carcinogenic potential of copper and its inorganic compounds in humans and animals cannot be evaluated on the basis of the available studies.

Only a few clinical observations in humans are available in which skin-sensitizing effects of copper or its inorganic compounds are confirmed. In the majority of cases, the reactions observed are probably to be interpreted as nonspecific reactions to unsuitable test preparations.

From animal experiments there is evidence that copper salts impair reproductive performance and that they are embryo- or fetotoxic, or teratogenic at higher doses.

Manganese

Manganese is an essential trace element. It is a component of various metalloproteins and a normal constituent of almost all tissues. Manganese is a co-factor essential for the activity of a variety of enzymes, e.g. pyruvate carboxylase, arginase, phosphatase, superoxide dismutase, glutamine synthetase, and manganese-dependent ATPase.

The most important target organs after prolonged uptake of manganese by man and animals are the central nervous system and the lungs. In workers exposed to manganese long-term exposure resulted in chronic manganism, a parkinsonism-like disorder, and in pneumonia. Exposure to lower concentrations leads to subclinical neurological symptoms.

The effects of manganese in its various target tissues are probably a result of the fact that manganese is a reducing agent or an oxidizing agent, dependent on its oxidation state. Manganese(II) can capture superoxide radicals ($O_2^{\bullet -}$) and so is a powerful antioxidant. In the higher oxidation states (III, IV) the manganese ion is an effective oxidizing agent. It catalyses the oxidation of biologically important substances such as catecholamines, unsaturated fatty acids, and glutathione.

The neurotoxicity of manganese is probably, at least in part, a result of the affinity of the metal to tissues containing high levels of neuromelanin, for example, the substantia nigra of the brain.

The lowest average manganese concentration which has been shown to cause slight neurotoxic symptoms was about 0.25 mg/m^3.

Nickel

Occurrence and Exposure Nickel is abundant in the earth's crust and widely distributed in the environment. Adverse effects have been observed at the workplace, especially in nickel refineries, where high concentrations of nickel sulfides and oxides occurred.

Toxicokinetics Nickel ions from readily soluble nickel salts are slowly taken up via ion channels in plasma membranes. The less soluble metallic, sulfidic, and oxidic forms of nickel are taken up by phagocytosis. Less soluble nickel compounds including elemental nickel lead to the release of nickel ions after having been taken up in mammalian cells. The better bioavailability of poorly soluble nickel compounds explains their more severe chronic toxic effects when compared with the readily soluble salts. First, an intracellular accumulation of nickel particles, localized mainly at the nuclear membrane, occurs. After they have dissolved, there is an accumulation in the nucleus in the form of nickel bound to nucleic acids and protein.

Acute Toxicity Nickel, its alloys, and nickel compounds can cause contact dermatitis, contact urticaria, allergic rhinitis, and allergic asthma.

Chronic Toxicity The lungs and nose are the target organs of the carcinogenicity of nickel and nickel compounds, especially in the form of inhalable dusts or aerosols.

In general, the release of nickel ions is considered responsible for the genotoxic, carcinogenic, and sensitizing effects of all forms of nickel.

Soluble nickel salts are nonmutagenic in almost all bacterial mutagenicity tests and only weakly mutagenic in tests with mammalian cells. Nickel ions cause chromosome

aberrations, sister chromatid exchanges, DNA breaks and DNA–protein cross-links in mammalian cells only in higher concentrations (mmol/l range). Three mechanisms are discussed for the carcinogenic effect of nickel:

(1) Enhanced formation of reactive oxygen species catalysed by nickel: In the presence of hydrogen peroxide, nickel(II) ions produce oxidative DNA damage to isolated DNA and isolated chromatin, these effects being reduced by antioxidants. However, oxidative DNA damage was observed in cell cultures only in cytotoxic nickel chloride concentrations ($\geq$0.75 mM).
(2) Epigenetic mechanisms inducing increased cell proliferation: Nickel chloride caused increased methylation of cytosine bases in tumor suppressor genes, resulting in increased cell proliferation. These nickel-induced changes in gene expression can be interpreted as a deregulation of normal growth control.
(3) Inhibition of the repair of DNA damage. This mechanism leads to an increase of the effect of other DNA-damaging agents. The inhibition of DNA-repair processes by nickel ion was identified as the mechanism of action on which this enhancement is based. Both DNA-excision repair and the repair of O^6-alkylguanine and of oxidative DNA damage are inhibited in mammalian cells by subtoxic concentrations of nickel. However, a co-mutagenic effect of nickel ions also agrees with the results of epidemiology, since exposure to readily soluble nickel compounds led to lung tumors only at relatively high exposure levels, and it increased the tumor incidence after simultaneous exposure to less readily soluble nickel compounds and tobacco smoke.
(4) A nonsubstance-specific mechanism of carcinogenicity based on a particle overload. According to this, the poorly soluble nickel oxide leads to chronic activation of macrophages, resulting in chronic inflammation and being carcinogenic only secondarily. While nickel oxide induced tumors in rats even at 1.25 mg/m^3, a carcinogenic effect was observed only at 250 mg/m^3 for titanium dioxide (pigment grade), at 18 mg/m^3 for talcum, and at 45 mg/m^3 for antimony trioxide.

Palladium

Palladium is preferentially used as an exhaust catalyst to oxidize combustion by-products.

Palladium and palladium compounds are most readily absorbed via the lungs, to a lesser extent also via the gastrointestinal tract or through the skin. After repeated inhalation or oral exposure to different palladium compounds or complexes, impaired general condition, reduced body weights, neurotoxic effects such as convulsions, sluggishness, ataxia, or tiptoe gait were found. The palladium-induced cardiovascular effects observed in animal experiments have been related to its inhibition of creatine phosphokinase activity. The damage to liver and kidneys is explained by the binding of palladium to sulfhydryl groups of functional proteins, which inhibits energy production, important metabolic processes, and also DNA, RNA, and protein biosynthesis.

Palladium chloride has sensitizing effects in animal experiments. A number of palladium chloro-compounds are irritant to the skin and corrosive to the eyes of rabbits. In many *in vitro* studies, a cytostatic effect of palladium complexes was observed.

Palladium compounds were not mutagenic or clastogenic *in vitro* or *in vivo*. The only available carcinogenicity study with palladium chloride in mice cannot be evaluated due to serious shortcomings.

Tantalum

Tantalum and tantalum oxides are practically nontoxic. In experimental animals, the acute toxicity of tantalum oxide is higher than that of the metal. In guinea pigs exposed long-term to tantalum dioxide, transient bronchitis, interstitial pneumonitis, and the associated hyperemia were observed.

Tellurium and its Inorganic Compounds

Occurrence and Exposure Tellurium and its compounds are used for thermoelectric and electronic applications, and in ferrous and nonferrous alloys. Concentrations in ambient air, water, and food are low and mostly result from the combustion of coal. Food is the main route of the general population's exposure.

Toxicokinetics Tellurium and its inorganic compounds are able to enter the organism via ingestion or inhalation in the form of inhalable dusts and aerosols. They are metabolized slowly and eliminated via the urine, feces, sweat, and expired air in the form of dimethyl telluride. According to the tellurium compound involved, intestinal absorption is between 10 and 25%.

Acute Toxicity On inhalation, hydrogen telluride produces mild nasal irritation, dyspnea, heart conditions, and general discomfort; tellurium hexafluoride leads to pronounced irritation of the respiratory tract and pulmonary oedema resulting in death. In humans, after the intake of tellurium, characteristic toxic symptoms occur such as a garlic-like odor to the breath, perspiration, and excrement, as well as nausea, vomiting, weight loss, greyish-black discoloration of the mucosa, and alopecia, which may persist over a long period.

Chronic Toxicity In experiments with rats, after the repeated intake of tellurium compounds, neuropathy with segmental demyelinization of the sciatic nerve and subsequent paralysis of the hind legs were observed in addition to damage to the liver and kidneys. Subsequently, after continuing treatment with tellurium, remyelination and a reversal of the paralysis occurred.

The neuropathy appears to be the result of reduced cholesterol synthesis, with subsequent destabilization and reduced remyelination. The disturbance of cholesterol synthesis is triggered by the inhibition of squalene epoxidase, a microsomal monooxygenase, the cause of which is thought to be a reaction between tellurium and the SH groups of the enzyme. The methylated tellurium compounds, dimethyltellurium dichloride and dimethyl telluride, inhibit the activity of squalene epoxidase to a greater extent than does tellurite; this is explained by their greater binding capacity to the enzyme.

Tellurium compounds cause direct damage to the DNA with mutagenic effects in bacteria, and are clastogenic in mammalian cells. In long-term drinking water studies, which were, however, not valid, no evidence of carcinogenic effects of sodium tellurite or potassium tellurate in rats and mice were found.

When administered in maternally nontoxic doses to pregnant rats and rabbits, tellurium and inorganic tellurium compounds produced hydrocephali.

Tin and its Inorganic Compounds

Occurrence and Exposure Metallic tin is processed to yield tin foil or solder and is used in the manufacture of numerous alloys such as bronze. Inorganic compounds of both bivalent and tetravalent tin are also processed industrially in large quantities. They are used in the tinning of sheet iron to produce tin plate, as catalysts, as pigments in the ceramics industry, and as starting materials for the production of organic tin compounds.

For the toxicological profile of metallic tin, practically no data are available. The toxicity of the inorganic tin compounds is determined by the bivalent or tetravalent tin cation, not by the associated anion. Therefore, the systemic toxicity of tin salts is largely a function of their solubility. In the inorganic tin compounds sodium pentafluorostannite and sodium chlorostannite, the tin is not present as a free cation but is bound in a complex; therefore, the toxicological profile of these substances is not comparable with that of the tin salts. Most toxicological studies have been carried out with tin(II) chloride, which is readily soluble in water.

Toxicokinetics Inorganic tin compounds are poorly absorbed from the gastrointestinal tract. The highest accumulation is found in the bones. The tin levels in other organs are also increased. Tin is excreted mainly with the urine; small amounts are found in the bile.

Acute Toxicity After ingestion of tin(II) chloride, the first effects are generally nausea and vomiting. In the only study with volunteers, tin doses of 0.7 mg/kg body weight had no effects. An effective dose was not determined. Ingestion of tin salts, for example with canned food, causes acute but reversible gastroenteritis in man from tin doses of about 50 mg. The gastroenteritis is ascribed to the irritant effects of these substances on the gastrointestinal mucosa. Tin salts in solution or formulations in petrolatum cause skin irritation in humans and experimental animals. Irritant effects on the eyes have not been studied. The sensitizing potential appears to be low. For both metallic tin and tin(II) chloride, positive reactions in patch tests have been described in only a few case reports.

Chronic Toxicity Repeated ingestion of tin doses of 0.7 mg/kg body weight/day in the form of tin(II) chloride is not toxic for humans. Oral administration of inorganic tin compounds to rats causes not only local effects on the gastrointestinal mucosa but also delayed growth, anemia, and histological changes in the liver. Effects on the homoeostasis of essential heavy metals and changes in biochemical blood parameters have been described. The data for doses without effects are not consistent.

The anemia observed in experimental animals given repeated doses of inorganic tin compounds is probably caused by several mechanisms. On the one hand, tin probably reduces the absorption of iron from the gastrointestinal tract, and on the other, tin induces heme oxygenase, a key enzyme in hemoglobin degradation.

In both prokaryotes and eukaryotes, inorganic tin compounds cause DNA damage, which could be the result of formation of reactive oxygen species. The available *in vivo*

studies provide no evidence of a genotoxic potential. Long-term studies in rats and mice do not suggest that the water-soluble tin(II) chloride has carcinogenic potential. Carcinogenicity studies have not been carried out with other inorganic tin compounds. In rats, the water-soluble compound tin(II) fluoride does not have toxic effects on reproduction.

There are practically no data for the threshold of effects or effect profile of repeatedly inhaled organic tin compounds. The values for the NOEL and LOEL determined in the various studies with experimental animals differ widely. Some of these differences may be ascribed to observations of the different endpoints. The highest NOEL of 60 mg/kg body weight/day (expressed as Sn) was found in a 13-week feeding study. In contrast, administration in the drinking water of concentrations yielding estimated tin doses of as little as 0.4 mg/kg body weight/day caused adverse effects. Thus, a NOEL value cannot be derived.

Organotin Compounds

Occurrence and Exposure Trialkyl- and triaryltin compounds have been used as biocides for many years. Special attention has been paid to the pollution of the marine environment, estuaries, and freshwater from their use in paint on boat bottoms (antifouling paint). After the discovery of their endocrine effects on marine snails, the use of these substances on small boats has been restricted in many countries, and there is now an international agreement to stop their use on large ships. The main nonbiocidal applications of organotin compounds are as stabilizers in polymers, such as PVC, or catalysts in the production of polymers. The final materials often appear in consumer products. There have recently been several reports on the appearance of organotins, such as tributyltin, in food, clothes, shoes, and other items in close contact with humans because of their use as bactericides, fungicides, and slimicides.

Tri-n-butyltin compounds Tributyltins are, in addition to antifouling paint, used as wood preservatives. The mono- and dialkyltins are used as stabilizers in PVC, as catalysts in the production of silicones and esters, as well as for electro-deposition.

Acute Toxicity Since they are cytotoxic, tributyltin compounds have a strong irritant effect on skin and mucous membranes, and inhaled particles produce severe local irritation.

Chronic Toxicity The lymphatic organs are the main targets of the systemic toxicity (thymus atrophy), resulting from repeated oral administration (10 days to 2 years) to experimental animals. In addition, suppression of thymus-dependent immune responses and of nonspecific resistance has been demonstrated in rats. The discussed mechanism involves dibutyltin metabolites of the tributyltin compounds. Higher doses ($>$80 ppm) damage the liver and bile ducts as well as producing microcytic anemia and endocrine disorders. Unlike trimethyltin and triethyltin compounds, tributyltin oxide (TBTO) is not neurotoxic.

Mutagenicity tests of various kinds have generally produced negative results. The highest dose of TBTO (50 ppm TBTO in the diet) used in a carcinogenicity study in rats induced an increase in benign tumors of the endocrine organs, probably by a

nongenotoxic mechanism, e.g. suppression of immune response or disturbance of hormonal regulation. In studies with mice, rats, and rabbits, TBTO has no embryotoxic effects except at concentrations which are also toxic to the mother.

Octyltin Compounds

Mixtures of about 80% dioctyltin- and 20% monooctyltin-compounds are used as stabilizers in the production of PVC. Dioctyltin dichloride (DOTC) causes thymus atrophy in rats after dietary application (decreased relative thymus weights at 0.7 mg/kg bw and associated histopathological findings at 6.5 mg/kg bw/day). Dioctyltin ethylhexylmercaptoacetate shows the same effects, i.e. decreased relative thymus weights at 1–1.5 mg/kg bw/day, reduced number of lymphocytes, and reduction in thymus size. DOTC does not only induce thymus atrophy at repeated doses of 1 mg/kg but also displays an immunosuppressive effect. *In vitro* the effects are similar in lymphocytes isolated from rodents and humans so that no qualitative species differences can be assumed. Whether there are interspecies differences in the toxic potencies of the organotin compounds is not known.

Repeated-dose studies with monooctyltin compounds revealed a lower toxic potency than that of the dioctyltins. The systemic NOEL in a 90-day study was 15 mg/kg bwt. Similarly, the effects on development and reproduction are less pronounced. In an F0- and F1-generation study female rats have received one dose of 300 ppm monooctyl in feed, starting 10 weeks prior to mating and continuing to the 21st day of lactation of the pups. The compound did not induce effects at 300 ppm (NOEL about 22.5 mg/kg day). This supports the information that monooctyltin compounds are less toxic that DOTC.

Vanadium

Vanadium is present in environmental dust as the sulfide or the oxidized form. Increased air concentrations in populated areas (450–1300 ng/m^3 are primarily from fly ash formed during burning of fossil fuels.

Vanadium inhibits Na^+-K^+-adenosine triphosphatase as well as the synthesis of cholesterol, L-ascorbic acid, and fatty acids. About 40–60% of ingested vanadium is excreted in the urine within 1–3 days of exposure. A small amount (10–12%) is excreted in the feces. In animal studies it could be demonstrated that within 14 days after intraperitoneal, intratracheal, or intragastric administration, vanadium had accumulated in bones, liver, lungs, kidneys, and placenta. Inhaled V_2O_5 or VO_2C1 are absorbed practically completely from the lungs whereas only a few percent of an intragastric dose is absorbed from the gastrointestinal tract.

Inhaled vanadium, particularly in the form of vanadium pentoxide, causes severe irritation of the mucous membranes of the conjunctiva and respiratory passages. Typical findings after short-term or long-term exposure include conjunctivitis, rhinitis, bronchitis, and asthmatoid symptoms. Skin contact can result in allergic dermatosis. In experimental animals V_2O_5 has been shown to cause anemia, weight loss, and liver and kidney damage. These effects are expected to occur at concentrations exceeding 0.1 mg/m^3.

In tests with a mutant *Bacillus subtilis*, DNA-damaging effects could be demonstrated with various vanadium compounds. In cultures of human fibroblasts, vanadate caused an increase in DNA synthesis and cell growth.

Zinc

Occurrence and Exposure Zinc is an essential element for all living organisms. The amount required by man (10–15 mg/day) is generally present in the diet. Especially in children, zinc deficiency can cause damage to the skin and zinc-containing enzymes such as alcohol dehydrogenase, superoxide dismutase, RNA and DNA polymerase, and pyruvate carboxylase. Overexposure by food or air generally poses no risk to the population because absorption is regulated by the zinc status.

Toxicokinetics Absorption in the intestine and homoeostasis of zinc are regulated by metallothionein. These processes are as complex as that of iron. In zinc deficiency more zinc is absorbed; less is absorbed when the zinc status is more than adequate.

Acute and Chronic Toxicity Whereas metallic zinc is practically nontoxic, inhalation of zinc oxide during smelting or welding can cause acute intoxication (metal-fume fever). This disorder develops only after inhalation of zinc oxide, not after oral, parenteral, or dermal application of the substance. The symptoms of metal-fume fever first appear between four and twelve hours after exposure as a sweetish metallic taste and a feeling of irritation in the throat. Unspecific influenza-like symptoms (general feeling of being unwell, coughing, weakness, etc.) can follow. Some of the affected persons develop none of these signs but instead show sudden chills and fever, a feeling of tightness in the chest, and mostly mild respiratory symptoms (dyspnea). During the further course of the disease, various organ systems can be affected. In addition to rapidly increasing temperature, the symptoms may include frontal headaches, visual disorders, pain in the abdomen, muscles, and joints, nausea, vomiting, and dysuria. Some exposed persons develop asthmatiform reactions. Challenge tests have yielded evidence of a hyper-reactive bronchial system.

The inhalation of zinc oxide fume causes inflammatory changes in the lungs of man and animals; ZnO concentrations of 2.5 mg/m^3 and more cause changes in inflammation markers in the bronchoalveolar lavage fluid (BAL), in lung function, and in histological parameters.

It has been suggested that the pathogenesis of metal-fume fever involves the increased formation of oxygen radicals in human neutrophils in the presence of Zn^{2+} and ZnO. The results of animal studies and the known effects in man suggest that the systemic effects of zinc oxide fume are consequences of the massive inflammatory changes in the lungs with release of mediators and cytokines. There is no agreement as to which mediators play the main role in the pathogenesis of metal-fume fever. Release of the tumor necrosis factor (TNF) and interleukin-6 (IL-6) seems to be involved.

Zirconium and its Compounds

Zirconium is used in the ceramics and electronics industry.

When insoluble zirconium compounds are inhaled they can cause mild bronchial asthma, and granulomatous and fibrotic changes in the lungs. Insoluble zirconium compounds are not readily absorbed from the gastrointestinal tract. The amount absorbed is much less than 1% of the dose.

Soluble zirconium compounds are more toxic than the water-insoluble compounds because they form aggressive aerosols and so can cause tissue damage on exposed sites

as well as systemic reactions. Rats, guinea pigs, and hamsters exposed for over 200 days to a zirconium concentration of 4.7 mg/m^3 (administered as zirconium lactate) showed signs of a diffuse interstitial pneumonitis with slight fibrogenic effects.

In man, hypersensitivity reactions of delayed type, mostly granulomatous changes, have been described on the skin and, more rarely, in the lungs of persons exposed to either water-soluble or water-insoluble zirconium compounds. The allergenic effects have also been demonstrated in the mouse, rabbit, and guinea pig.

One *in vitro* study and an inadequately documented study with the mouse suggest that zirconium has weak genotoxic potential. The animal studies reported to date provide no sufficient evidence of carcinogenicity.

6.2.5 Summary

Metals are ubiquitous in the environment. In contrast to organic chemical compounds, they cannot be degraded, which means that their anthropogenic use leads to redistribution and accumulation in ecological systems. The general population is mainly affected by metals in the diet, while at the workplace the exposure via inhalation prevails. The toxic action of metals originates from their reaction with atoms, such as oxygen, nitrogen or sulfur, which are present in all molecules of biological importance. This can also be the cause of the allergenic, genotoxic, and carcinogenic properties of some metals. Acute exposure may induce local irritations of skin and mucosa. Chronic toxicity of metals is reflected by specific effects related to organs and tissues and is frequently mediated by metal-binding proteins. Another general mechanism of action consists in the replacement of essential metal atoms bound to functionally important sites by toxic metals, which disrupts metal-dependent biological functions.

References

Clarkson, T.W., Halbach, S., Magos, L., and Sugata, Y. (1980) Molecular Basis of Environmental Toxicity (ed. Bhatnagar, R.S.), Ann Arbor Science

Waisberg, M., Joseph, P., Hale, B., and Beyersmann, D. (2003) Molecular and Cellular Mechanisms of Cadmium Carcinogenesis. *J. Toxicol.*, **192**, 95–117.

For Further Reading

Clarkson, T.W., Hursh, J.B., Sager, P.R., and Syversen, T.L.M. (1988) Mercury. In: Biological Monitoring of Toxic Metals (eds Clarkson, T.W. et al.), pp. 199–246, Plenum Press, New York.

Handbook on the Toxicology of Metals (1986) (eds Friberg, L., Nordberg, G.F., and Vouk, V.), 2nd edition, Elsevier, Amsterdam.

World Health Organization (1990) Methylmercury. Environmental Health Criteria, No. 101. WHO, Geneva.

World Health Organization (1991) Inorganic Mercury. Environmental Health Criteria, No. 118. WHO, Geneva.

The MAK-Collection for Occupational Health and Safety. (2006) Wiley-VCH. www.mrw.interscience.wiley.com/makbat (see list of substances)

UNEP: Global Assessment of Mercury. (2002) United Nations Environment Programme, Geneva.

6.3 Toxicology of Fibers and Particles

Paul J.A. Borm

6.3.1 Introduction

Historically, workers in the mining, sanding, wood, construction and agricultural industries were exposed to high concentrations of particles. Massive exposure to inhalable and respirable particles over a period of years at more than 5 mg/m^3 is known to cause both chronic obstructive pulmonary diseases (COPD), including bronchitis, emphysema, fibrosis, and pneumoconiosis, and lung cancers. In coal mines, dust and asbestos induce pneumoconiosis. Reducing particle exposure decreases, but does not eliminate, these disorders. Particle-induced lung diseases recognized more recently include allergic and nonallergic asthma caused by a wide variety of occupational allergens. During the past decade a series of epidemiological studies has shown that ambient exposure to particles which display a mass mean diameter of 10 μm (PM_{10}) is related to both respiratory and cardiovascular mortality and morbidity. Observations made on the workplace environment and experimental studies have led to the appreciation that particle components such as transition metals and ultrafine particles play a role in the induction of pulmonary oxidative stress and inflammation. These processes are considered crucial to many of the acute effects of ambient particle exposure. Furthermore, it is currently thought that ambient particulate matter (PM) with a mean diameter of 2.5 or 10 μm ($PM_{2.5}$, PM_{10}) can be carcinogenic, but it is not clear which components or characteristics of PM are responsible.

6.3.2 Particle Toxicology: Basic Concepts

To understand particle effects, one can apply the 'D-concept', i.e. the Dose, Dimension, Deposition, Durability, and Defense (Table 6.12). This paradigm builds a conceptual understanding of the adverse effects of particles and fibers.

Table 6.12 *Processes and particle parameters that play a major role in determining the toxic response upon particle inhalation.*

Dose	Cumulative dose for chronic effects; can be based on particle or fiber mass, fiber or particle number, or particle surface dose. Bulk composition is not equal to surface.
Dimension	Size (diameter, length)
Deposition	Dependent on dimension but also on airway properties (hot spots)
Durability	Biopersistence dependent on defense as well as particle properties (dissolution)
Defense	Mucociliary clearance, macrophage clearance, inflammatory cells. If macrophage clearance is saturated, overload occurs; dose increases exponentially with time.

Particles are a special case in toxicology. They are widespread in many different occupational and environmental situations, as well as in consumer products such as cosmetics, paints, etc. An overview of particles, their most important applications and/or exposure situations, and effects is given in Table 6.13. There are several characteristics that distinguish the toxicological properties of particles and soluble chemicals. First, if a particle is insoluble only its surface will interact with the biological environment. For example, a fly-ash particle with a hazardous bulk composition of heavy metals and quartz may be totally inert, since it is covered by an inert amorphous glass-layer. This cover does

Table 6.13 *Various sources of particles, exemplary exposure situations, and the current knowledge on their toxicological properties and effects.*

Particle type	Source	Example(s)	Who is exposed?	Effects
Crystalline silica	Quarrying	Coal mine dust, quartz flours, Kieselguhr	Miners	Silicosis, lung cancer
Coal mine dust	Coal mining		Miners	Pneumoconiosis
Asbestos	Mining, insulation	Crocidolite, chrysotile	Insulators, shipyard workers	Asbestosis, mesothelioma, lung cancer
Organic dusts	Agriculture	Grain, cotton, flour	Bakers, cotton workers	Asthma, COPD, allergic alveolitis
PM_{10}, $PM_{2.5}$, TSP[a]	Traffic, industry	Fly-ashes, diesel, sea salt, road dust	Everyone	Increase mortality from cardiovascular and respiratory causes
Man-made mineral fibers (MMVF)	Industry	Rockwool, ceramics	Occupational	
Organic synthetic fibers	Industry	Aramid, Polyethylene	Occupational	
Nanoparticles (NP) (<100 nm)				
Combustion-derived NP	Combustion	Diesel soot	Everyone	Pulmonary and CV effects in humans
Bulk manufactured NP	Combustion	Carbon black	Occupational	Fibrosis and lung cancer in rats
	Synthesis	Amorphous silica, TiO2	Occupational	
	Geological/ synthesis	alumina	Occupational	
Engineered NP	Nanotubes Q-dots	Combustion Sytnthesis	Occupational Patients	Lung damage rats Hepatic damage

[a]Definitions used in environmental particle control. TSP, Total suspened particles; PM_{10}, Particulate matter with diameter smaller than 10 μm; PM2.5, particulate matter with diameter smaller than 2.5 μm.

not allow leaching or contact of the potential toxic constituents with the biological environment. Second, particles do not distribute homogeneously throughout tissue compartments or even cells, in contrast to lipophilic agents such as solvents or drugs. Therefore, it is difficult to define the effective dose within cells. A single 5-micron particle in a cell may exert different actions in the same cell, caused by the fact that one side of the particle interacts with the cell membrane and the other part with mitochondria or the nucleus. Third, particles may carry other agents such as absorbed polycyclic aromatic hydrocarbons (PAHs), gases such as SO_2, bacteria, or proteins into the lung, thereby changing deposition and bioavailability of the absorbed contaminants.

Particle Deposition

Particle size largely determines where particles are deposited in the respiratory tract. The major fraction of the fine (<2.5 μm) and ultrafine (<100 nm) particles is deposited on the fragile epithelial structures of the gas-exchange region. Larger particles are deposited in the nasopharyngeal region (>30 μm) and tracheobronchial region (10 μm <diameter <30 μm). Particle deposition is further determined by shape and density.

The cumulative dose at a specific pulmonary site determines the adverse effects of particles. The deposited dose is dependent on the inhaled concentration and the dimensions of the particle. Smaller particles increase the probability of particle deposition in the respiratory tract. As illustrated in Figure 6.7 the particle size largely determines where particles will be deposited in the respiratory tract. The terminology described in Figure 6.7 includes the inhalable and respirable fractions, which are

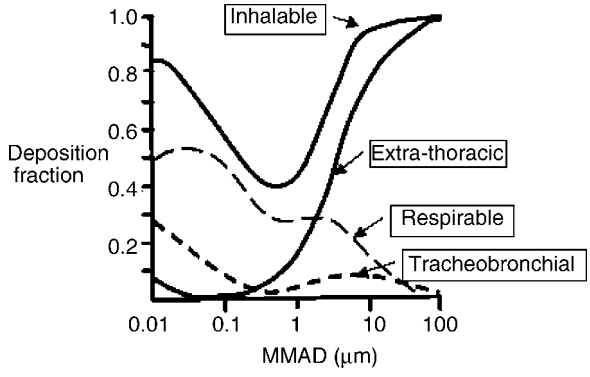


Figure 6.7 *The relationship between particle size (MMAD = mass median aerodynamic diameter) and the efficiency of deposition in various anatomical compartments of the lungs. The fractions depicted in the deposition curves (e.g. respirable fraction) are conventions used for the design of sampling devices and the setting of occupational standard limits for different particles.*

important definitions in the standard setting for occupational and environmental exposure limits.

The processes driving deposition are impaction of heavier particles, sedimentation in alveoli, interception of fibers, and, rarely, electrostatic interactions with airway wall constituents. Small ultrafine particles have little mass and behave by diffusion. This causes most of them to be deposited in the nasal compartment and the alveoli.

Shape and Density

The size for spherical particles is simply defined by the diameter. For nonspherical particles, the term 'diameter' does not appear to be strictly applicable because it does not properly describe the geometry of a flake of material or a fiber. Also, particles of identical shape can be composed of different chemical compounds and, therefore, have different densities. To provide a simple means of categorizing particles and fibers of different shape and density the term 'aerodynamic diameter' has been introduced, which is defined as the diameter of a spherical particle having a density of 1 gm/cm^3 that has the same inertial properties. Figure 6.8 explains the concept of aerodynamic diameter. Particle density affects the motion of a particle through a fluid and is taken into account in Equation (6.8). The Stokes diameter for a particle is the diameter of a sphere that has the same density and settling velocity. It is based on the aerodynamic drag force caused by the difference in velocity between the particle and the surrounding fluid. For smooth,

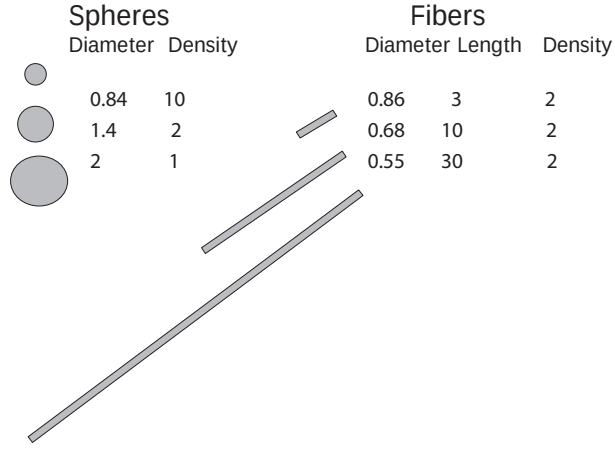


Figure 6.8 *Illustration of the concept of aerodynamic diameter by a set of spherical and fiber-shaped particles all with the same aerodynamic diameter of 2* μm*. A small particle with a high density behaves virtually as a bigger particle. On the other hand fiber length does not really affect aerodynamic diameter. This latter explains why long fibers pose a risk to the lower airways. Diameter and Length dimensions are in* μm*.*

spherical particles, the Stokes diameter is identical to the physical or actual diameter. The aerodynamic diameter for all particles greater than 0.5 μm can be approximated by using the Equation (6.8):

$$d_{pa} = d_{ps}\sqrt{P_p} \tag{6.8}$$

d_{pa} = Aerodynamic particle diameter, μm
d_{ps} = Stokes diameter, μm
ρ_p = Particle density, gm/cm^3

Most dosimetry models and calculations assume a uniform deposition at the bronchial or alveolar surfaces and, therefore, assume similar target dose for all epithelial cells within the respiratory tract. However, analysis of particles in human lung tissue, as well as mathematical modeling, shows that local particle deposition occurs at the bronchial airway bifurcations independently of the particle size. Moreover, major differences in particle deposition can be found between lung lobes. In particular, cells located in the vicinity of the dividing spur may receive local doses that are a few hundred times higher than the average dose for the total airway. It is no coincidence that it is at these so-called *hotspots* located at the bronchoalveolar bifurcations that tumors related to exposure to inhaled asbestos or other particles are usually seen.

Particle Clearance and Lung Overload

The lung has potent defense systems to remove particles. These include mucociliary clearance in the upper airways and macrophage clearance in the lower airways and the alveoli (Figure 6.9). Particle transport by macrophages from the alveolar region towards the larynx is rather slow in humans, even under normal conditions, and eliminates only about a third of the deposited particles in the peripheral lung. This implies that the other two-thirds remain in the lungs without significant clearance unless the particles are biodegraded or cleared by other mechanisms. In the course of inhalation studies high concentrations of particles, e.g. 1–3 mg, or 0.1 mm^3 per g of lung in rats, may exceed the macrophage clearance capacity and overburden the lung, a phenomenon known as 'lung overload.'

Lung overload results when deposition exceeds elimination. It is the consequence of exposure that results in a lung burden of particles that is greater than the steady-state burden predicted from the deposition rates and clearance kinetics of particles inhaled during exposure. The overload concept has important implications for hazard assessment as well as for setting occupational standard values for particles when based on the outcomes of animal studies.

The cause of particle overload is impaired macrophage clearance function. Volumetric overloading of macrophages starts at 6% of normal alveolar macrophage (AM) volume and this was originally used to develop to overload concept as illustrated in Figure 6.10. Nowadays the surface dose is suggested to be a better indicator of biological responses occurring at overload, and especially in explaining the effects of nanoparticles.

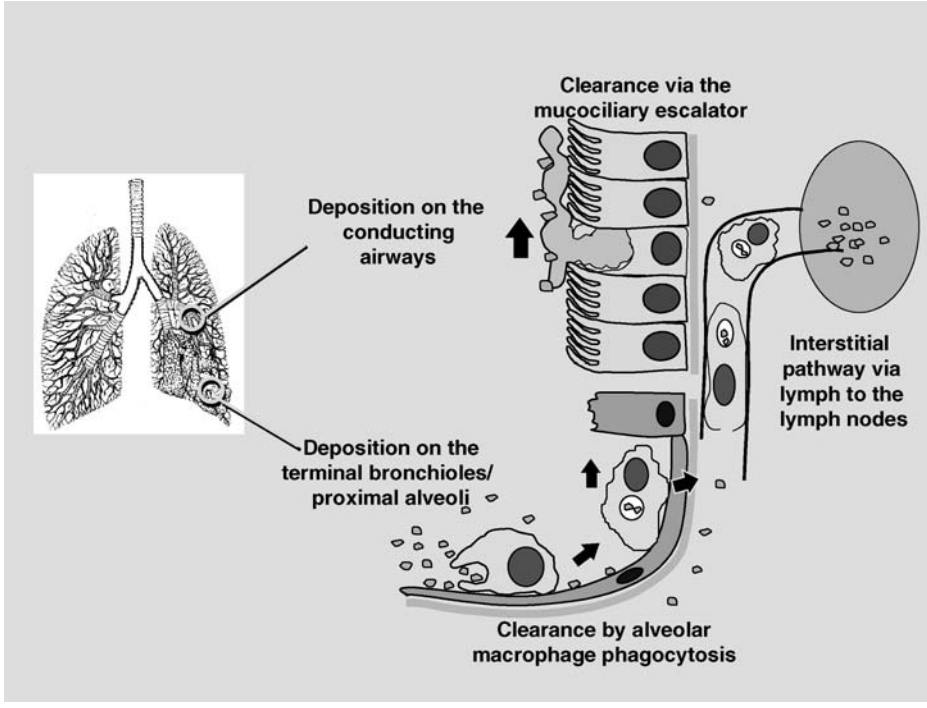


Figure 6.9 *Clearance pathways as host defenses in the respiratory tract are associated with the anatomical compartments in the lung. Larger particles (>10 μm) predominantly deposit in the tracheobroncial area, where the airways are covered with ciliated epithelial cells, and particles are transported upwards through the mucociliary escalator. Smaller particles that penetrate into the bronchoalveolar area can only be removed by phagocytosis from alveolar macrophages. This clearance is much slower than in the upper respiratory tract.*

At particle overload, macrophage clearance function is impaired and at this point particle accumulation starts and inflammatory cell influx increases sharply. It needs to be emphasized that this concept specifically applies to low-toxicity poorly soluble particles (PSP). Other more toxic particles such as crystalline silica, man-made fibers, and toxic metal particles affect AM-mediated clearance as well, but at much lower lung burdens since they can actively damage AM. Nanoparticles have also been shown to impair phagocytosis to a much greater extent than have larger particles when evaluated on an equal-mass basis (see Section 6.3.4, below). Thus, every impairment of AM-mediated particle clearance should not be viewed as a particle overload.

Particle Translocation

Particle translocation has only recently been recognized as a mechanism to explain possible systemic effects of inhaled ambient particles.

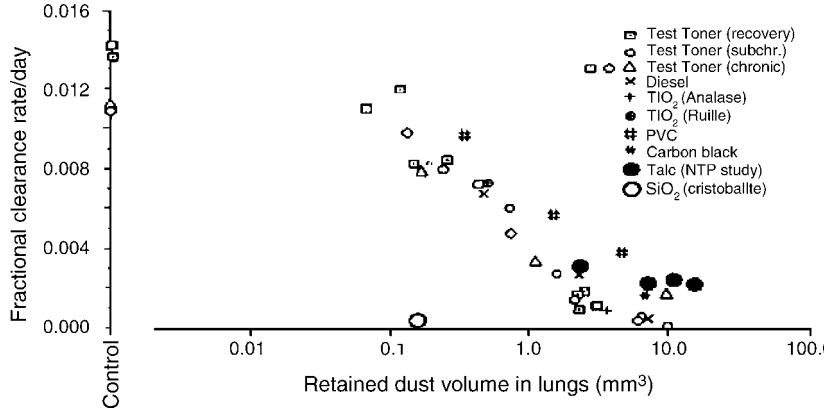


Figure 6.10 *The clearance rate of deposited particles in the lung is dependent on the total volume of retained particles in the lung. The graph is adapted from Morrow (1988) and shows the relation between clearance rate of various so-called poorly-soluble, low-toxicity particles (PSP) in the lung and particle volume (in mm^3) per gram of lung. Regardless of particle type, it is shown that beyond 0.1 mm^3 saturation of clearance occurs. This limit is the basis for many exposure limits of particles.*

Basically, inhaled particles, can translocate via two different routes:

- Passage through the pulmonary epithelium and endothelium to enter the general circulation. This may occur through endocytosis, transcytosis, or transport by immune cells.
- Passage to the central nervous system, more specifically the olfactory bulb or the cranial bulb, through the olfactory epithelium in the nose, or the trigeminal nerves in the upper airways.

Both routes seem to be limited to particle sizes under 100 or even 50 nm. Particle properties such as surface charge seem to be crucial in translocation through the lung barriers, while for transport along olfactory nerves size seems to be the main determinant. More details are discussed in paragraph 4 of Section 6.3.4 on nanoparticles.

Inflammation and Oxidative Stress

One of the crucial properties of particles is their ability to generate oxidants such as reactive oxygen species (ROS) and reactive nitrogen species (RNS). This factor is generally thought to be involved in many pathological outcomes of particle exposure, including fibrosis, and proliferative effects of particles.

One of the unifying paradigms in particle toxicology is the linkage between the generation of reactive oxygen species (ROS) and reactive nitrogen species (RNS) and the subsequent inflammatory response. The ability of particles to induce different endpoints

both in lung and systemic organs seems to be associated with their ability to generate oxidants that overwhelm the endogenous anti-oxidant defense mechanisms, and this is, therefore, called oxidative stress.

Particle exposure can induce oxidative stress by two mechanisms: The particles may have acellular oxidant-generating properties themselves, or they stimulate cellular oxidant generation. Furthermore, these actions may be subdivided into primary, particle-driven, and secondary, inflammation-driven, formation of oxidants. The inflammation-driven formation of oxidants is described here, while particle-driven formation is discussed under the subhead Surface Reactivity (Section 6.3.3, below).

Within the lung, various cell types, including vascular endothelial cells and lung epithelial cells, can endogenously generate ROS and RNS. For example, epithelial lung cells generate intracellular ROS upon exposure to fly-ash, Diesel exhaust particles (DEP), particulate matter (PM), and quartz particles, processes in which mitochondrial respiration and activation of NAD(P)H-like enzyme systems are involved. However, the pool of **inflammatory phagocytes** constitutes the most significant and important cellular ROS/RNS-generating system in the lung. During *in vivo* particle exposure, increased levels of ROS/RNS can be found in pulmonary tissue, and are associated with the influx of inflammatory phagocytes. Consequently, environmental particles and fibers such as asbestos, crystalline silica, heavy-metal-containing dusts, oil fly-ash, coal fly-ash, and ambient PM induce ROS production by neutrophils and macrophages *in vitro*. Several particle characteristics result in the activation of the phagocytic oxidative burst. For mineral dusts such as crystalline silica, it has been shown that ROS release from inflammatory cells is related to the physical dimensions and the surface-based radical-generating properties of the particles. In chemically complex particles such as PM or fly-ash (Table 6.13), the presence of metals is related to ROS release. Finally, organic substances adsorbed on the particle surface of PM have been related to the oxidative burst in neutrophils.

Genotoxic RNS such as nitric oxide (NO) and peroxynitrite (ONOO) can be generated in phagocytes by virtue of their inducible Nitric Oxide Synthase (iNOS) activity. Although neutrophils are considered as the most potent phagocytes with respect to particle-related ROS generation, the major source of RNS in the lung is the alveolar macrophage.

Apart from producing ROS, inflammatory phagocytes produce growth factors such as cytokines and chemokines that regulate cell migration, tissue remodeling, cell proliferation, and repair of damage. These mechanisms are induced by quartz, carbon black, and particulate matter.

Particles activate nuclear factor-kB (NF-kB) and other signaling pathways and cause increased release of various pro-inflammatory mediators such as interleukin-8 (IL-8), tumor necrosis factor-α (TNF-α), and interleukin-6 (IL-6), and anti-inflammatory mediators such as interleukin-10 (IL-10), and transforming growth factor-α (TGF-α). Such acute pro-inflammatory effects explain the exacerbations in chronic obstructive pulmonary disease (COPD) and asthma as well as cardiovascular events following air pollution episodes. Chronic inflammation helps us to understand, detect, and treat

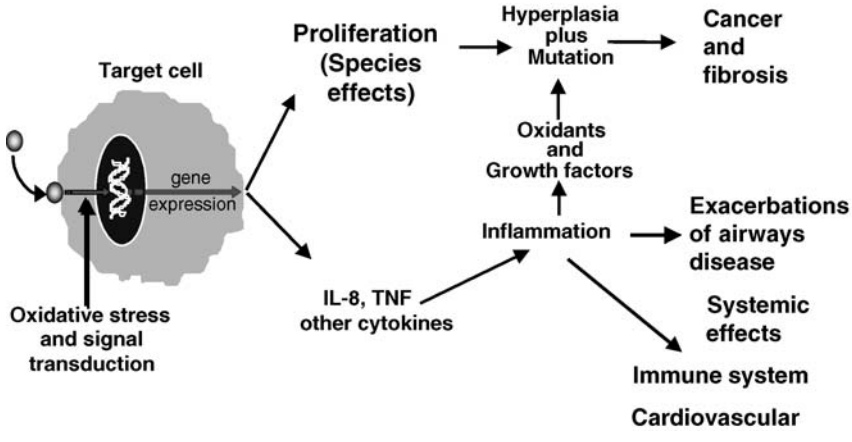


Figure 6.11 *The pathways and mediators leading to the acute and chronic pathology induced by particle or fiber inhalation. Particles induce oxidative stress in target cells, which is amplified by an inflammatory response. Both oxidants and pro-inflammatory mediators may contribute to acute effects of particles (lower part of diagram) as well as the chronic effects.*

long-term sequelae of particle exposure including fibrosis, emphysema, and lung cancer. TNF-α release from phagocytes is related to the risk of developing pneumoconiosis. Polymorphisms in the TNF-α gene have been associated with elevated risk of silicosis. Treatment with TNF-soluble receptor is now being applied as a treatment for pulmonary fibrosis. The mechanism as discussed in this paragraph is illustrated in Figure 6.11.

6.3.3 Particle Properties

The relationship between the types of particles to which workers are exposed and the lung diseases that they manifest is complex. For example, pneumoconiosis can be caused by different types of particles. Defining the dose of a particle requires information on the durability of the particle and its surface area. The locus of action, which is usually the site of deposition, is determined by factors that include the dimensions and shape of the particle and its inherent surface activity.

Durability

If a particle is not soluble or not degradable in the lung it has a high durability and there will be rapid local accumulation upon sustained exposure. Since durability, or biopersistence, is a major determinant of fiber pathogenicity, it is now incorporated into test protocols to characterize and classify new fibers.

Durability or biopersistence is a major determinant of fiber pathogenicity. In Table 6.14 a comparison of different indices of biopersistence or dissolution are given for various fibers. This shows that carcinogenic fibers have a high durability as illustrated

Table 6.14 *Comparison of different indices for biodurability of carcinogenic and noncarcinogenic fibers.*

Fiber	Carcinogen	$t_{1/2}$ (days)[a]	K_i[b]	K_d[c] (ng/cm^2.h)	KNB[d]
RCF1, 2 and 3	+	>60	>90	<6	<2
E-glass	+				
Crocidolite	+				
MMVF-21	−	30–60	<5	6–30	>18
MMVF-11	−	15–30	25	30–300	>18
Soluble fibers	−	<15	>30	>300	>18

[a] $t1/2$ is the half-life (days) after inhalation or intratracheal instillation.
[b] K_i is an index calculated based on KNB + BaO + B_2O_3 − $2Al_2O_3$.
[c] K_d is the measured dissolution rate in ng/cm^2.h in an acellular system.
[d] KNB is an index calculated as the sum of percentage composition of Na_2O + CaO + K_2O + MgO.

by their long half-life (>60 days), and a slow dissolution rate (<6 ng/cm^2 h) which is reflected in the chemical fiber indices K_i and KNB. Since *in vivo* and *in vitro* biopersistence assays are not always congruent, a tiered approach combining both simple and rapid bench-screening tests for *in vitro* dissolution and breakdown after intratracheal instillation in animal models is recommended. In addition, based on chemical composition a prediction can be made of dissolution rates of specific fiber classes (e.g. glass-fibers) and correlations between dissolution rates and *in vivo* carcinogenicity allow the screening out of potentially harmful fibers.

Dimensions: Size and Shape

Fiber dimensions can have profound effects on the ability of the body to defend against particles, and, thereby, on cumulative dose. Long (>20 µm) fibers are not taken up by alveolar macrophages, and, therefore, have longer half-lives in the lung than do shorter fibers of the same material, and, consequently, result in a higher cumulative dose at similar inhaled fiber number or mass. Inhaled fibers longer than 20 µm also show interception at the bronchiolar bifurcations due to their length and may achieve high local doses by this process.

For long fibers, which are not easily cleared by the immune system, durability (biopersistence) is the main determinant of whether or not the fiber is carcinogenic in animal studies. Fiber biopersistence *in vivo* and fiber dissolution *in vitro* are now used as screening methods for new synthetic fibers to select out potential durable and pathogenic products (see Subsection Durability, above, and Table 6.14). The intent is to develop new fibers which are rapidly broken down and dissolve under biological conditions. It is recognized that after phagocytosis by AM short fibers are trapped in an acidic (pH 4.5) environment. Fibers which are highly stable under normal conditions become biodegradable upon deposition in the lung.

Surface Area

The surface area of spherical particles has an impact on particle toxicity.

A larger surface area adsorbs more material, which is carried by inhaled particles to their deposition sites.

Smaller particles of a large surface area dissolve faster than do larger particles. Acute (inflammation) and chronic (lung cancer) responses to particle treatment are better correlated to surface area dose than to mass dose or volumetric dose.

Surface area is a particle property that has to be considered apart from surface reactivity. Equations (6.9) and (6.10) for calculating the volume and surface area of spheres are provided below, where r is the diameter of the sphere:

$$\text{Surface area of a sphere} = 4\pi.r^2 \tag{6.9}$$

$$\text{Volume of a sphere} = 4\pi.r^3/3 \tag{6.10}$$

The surface of the airborne particles adsorbs materials such as PAHs from the air and/or gases (see Subsection Adsorbed and Soluble Components, below. The adsorbed material could be toxic even if the particles are not. Alternatively, the adsorbed material could interfere with the normal chemical activity of the particles.

The larger the surface area, the more contaminating material reaches targets in the lung. On the other hand, stronger binding of adsorbed material on larger surfaces, mediated by van der Waals forces, diminishes the release of adsorbed components from the particle surface. An example is provided by PAHs adsorbed on carbon blacks that only become biovailable from highly contaminated versions of small-surface carbon black particles.

The time for dissolution is approximately proportional to the square of the particle size, which means much more rapid dissolution for smaller particles. This can be understood by looking at the surface/volume ratio of a sphere, which is inversely proportional to the diameter. For a particle of 6 micron the *S*/*V* ratio is 0.5 while for a 100 nm particle this is 30. The concentration at the site of deposition will be determined by fluid flow in the vicinity of the particle and by chemical reactions at the site. The faster the dissolution or reaction, the shorter the time before the material has been totally dissolved or removed.

Empirical observations show that both acute (inflammation) and chronic (lung cancer) responses to particle treatment are better correlated to surface area dose than to mass dose or volumetric dose. This concept has been mainly developed with PSP and, considering its relevance for dose–response curves, it is expected to impact on standard-setting, compliance measurements, and hazard identification. [For more detailed discussions on this issue see Oberdörster (2001), Borm et al. (2004), and Tran et al. (2000)].

Surface Reactivity

Differences in particle toxicity can be related to their surface reactivity and ability to induce oxidative stress either directly or by inducing inflammation (see Subsection Inflammation and Oxidative Stress, above).

As already indicated above, one of the crucial concepts in particle toxicity is their ability to induce oxidants. Various physicochemical properties of particles that play a role in oxidant generation have been characterized. Particles such as crystalline silica generate

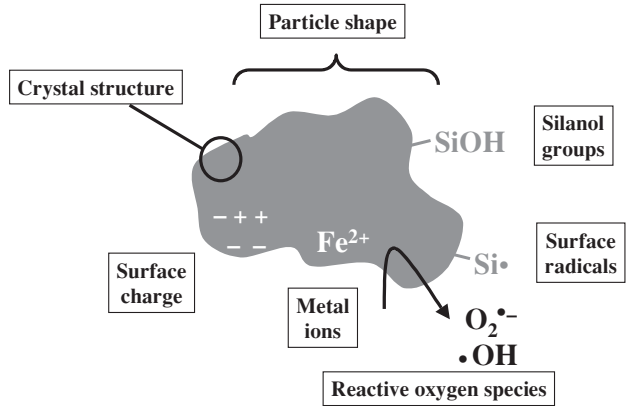


Figure 6.12 *Diagram illustrating the different properties that play a role in the reactivity of a particle, using a crystalline silica particle as an example. Reprint from Schins & Borm.*

oxidants as a function of their reactive particle surfaces (Figure 6.12). Selective blockade of the active quartz particle surface with organic polymers or metals (Al) is associated with a reduction of its ROS-generating capacities and its toxicity. On the other hand, freshly ground dusts, in which the reactive particle-surface is increased, not only contain more surface-bound reactive groups but are also more potent in generating free radicals in aqueous suspensions. Among the various existing types of particles one discriminates between those with inherent toxic activity such as hard-metal dusts, welding fumes, or quartz dust and those that, due to their material properties, have a much lower inherent particle reactivity. The latter are commonly referred to as poorly soluble particles (PSP) of low acute toxicity, or granular biodurable particles (GBP) without known specific toxicity.

Adsorbed and Soluble Components

The ultimate oxidant-generating capacity of particles is determined by their ability to adsorb various chemicals, including metals that also may enhance ROS-generation. The adsorbed material could be toxic even if the particles are not. Alternatively, the adsorbed material could interfere with the normal chemical activity of the particles.

Absorbed transition metals are of particular importance because they are involved in the generation of ROS *via* Haber–Weiss reactions. Indeed, in transition metal-containing particles like coal fly-ashes and PM, the generation of ROS can be easily modified by the use of metal-chelators or by removing the soluble particle fraction. Additionally, organic constituents associated with the particle surface may also contribute to oxidant formation. Semiquinone radicals produced during their metabolic activation can undergo redox-cycling, leading to the formation of reactive oxygen species. PAHs are present in many particle samples as a result of incomplete combustion. Particles may carry, hold, and release PAHs into cellular compartments that normally would not be reached. Also the kinetics of particle-associated PAHs and soluble PAHs have been demonstrated to differ

substantially and in some conditions particle-associated PAHs can be considered a slow-release deposit. On the other hand larger surfaces usually exhibit stronger binding of components through van der Waals forces, which diminishes the ability of absorbed components to be released from particle surface. An example of this is provided by PAHs absorbed on carbon blacks, which only become biovailable from low-surface, highly contaminated versions of carbon black.

6.3.4 Nanoparticles: A Special Case?

Nanotechnologies are expected to bring a fundamental change in manufacturing, mainly by producing new materials with highly added value. Anticipated applications and sectors are electronics, the automotive sector, food packaging, drug delivery, and imaging. Engineered nanoparticles (diameter<100 nm) are an important tool to realize these new applications and products.

The reason why nanoparticles (NP) are attractive for many purposes is based on their unique features, such as their high surface-to-mass ratio, their unique (quantum) properties, and their ability to adsorb and carry other compounds. Up to now many of these special purpose-engineered NP such as carbon nanotubes, quantum dots, dendrimers, and polymer-carriers are produced in small quantities, but production and applications are anticipated to increase steeply. Although chemically equivalent to larger particles of the same material, NP usually have a surface that is usually more reactive. Studies with fine and ultrafine TiO_2 particles have highlighted the fact that a material of low toxicity in the form of fine particles could be toxic in the form of ultrafine particles. Later studies have demonstrated that pulmonary inflammation, usually measured as the number of neutrophilic granulocytes (PMN) in bronchoalveloar lavage (BAL), is related to the surface area of particles although at similar surface areas some nanoparticles seem to be more inflammatory than others. Although the **deposition** of NP in the lungs follows largely the same distribution as that of fine particles, the underlying mechanisms are different. Nanoparticles have a size dimension that makes them less subject to gravity and turbidometric forces and, therefore, their deposition occurs mostly by diffusion. In addition, their size causes them to interact with other potential targets (brain, heart, endothelial tissue) more than do conventional fine particles. Secondly, NP may circumvent endogenous **defense**, since their distribution and cellular uptake is dependent on size and surface chemistry. In addition, NP particles can be **translocated** to body compartments away from their deposition sites in the respiratory system. NP may cross the epithelial and endothelial layers, especially in inflammatory conditions, and gain access to cells in the epithelium, the interstitium, and the vascular wall. Rapid translocation toward the liver of more than 50% of 26 nm gold NP occurred within 24 h in a rat model after inhalation or instillation. In contrast to such information from animal studies a rapid but no more than 3–5% uptake of radiolabeled carbonaceous NP into the bloodstream within minutes of exposure and subsequent uptake in the liver has been found in humans. NP can also be transported along the olfactory nerves into the brain after crossing the olfactory epithelium or via uptake along trigeminal nerves in the upper respiratory tract. It remains to be determined whether this uptake can lead to functional changes in the central nervous system. Experimental animal studies suggest that uptake

leads to activation of pro-inflammatory mediators such as TNF and cyclooxygenase-2 (COX-2) and degeneration of **dopaminergic** neurons. No human data are available at the moment, although circumstantial evidence for a relation between ambient particles in the brain and Alzheimer-like pathology has been observed in brain sections. Interestingly, the uptake and potential effects of NP in the brain have not been reported for fine particles and this effect seems to be limited to NP.

Nanoparticles partly use similar pathways and mechanisms as do fine particles to induce adverse effects. Quantitatively, however, nanoparticles can induce more inflammation at considerably lower gravimetric lung burdens than their larger analogues. Other mechanisms such as translocation to the brain, binding of subcelluler proteins, and effect on mitochondrial function are substantially different from that of their fine analogues.

Similar to fine-sized particles, cells in contact with NP such as macrophages and neutrophilic granulocytes, are activated and produce reactive oxygen species. Within hours, cytokines and chemokines are synthesized and secreted into the affected area. These mediators interact with specific receptors on the surfaces of many cell types and result in activation of local cells as well as those in the blood and other tissues. As a result, cells are attracted from the bloodstream and enter the fluid-filled interstitial spaces, where they can attack the foreign material. Consequently, particle-induced cell activation in the airways results in inflammation. Epithelial and nerve cells may also contribute to airway inflammation by producing pharmacologically active compounds such as capsacein. In this *neurogenic inflammation*, stimulation of sensory nerve endings releases neurotransmitters that may affect many types of white blood cells in the lung, as well as epithelial and smooth muscle cells.

For hazard characterization and classification of newly engineered nanomaterials several crucial questions need to be answered:

- Which effects are specific for nanomaterials, and which effects are merely stronger?
- Can we extrapolate available data and concepts generated with combustion-derived particles to newly engineered materials?
- Are current testing procedures specific enough to detect effects of nanomaterials?
- Is our current regulatory system ready to handle and communicate risks of nanomaterials?

6.3.5 Special Particle Effects

Carcinogenicity

Our current knowledge of particle-induced lung tumors in experimental animals can be summarized by stating that all inhaled particles, fibrous and nonfibrous, are likely to induce lung tumors in rats, provided that these particles are inhaled chronically or instilled intratracheally at sufficiently high doses, are respirable to the rat, and are highly durable.

The retained lung burden leading to lung tumors in the rat can differ for different particles and, apart from dose, greatly depends on particle properties such as surface area and chemistry, cytotoxicity, and size/dimensions. The gravimetric dose needed for the onset of particle overload and risk for subsequent neoplastic events is 1 mg/g lung tissue or 200–300 cm^2 surface burden of PSP. The surface dose where lung tumors start to develop after both inhalation and instillation of PSP lies between 0.2 and 0.3 m^2 per lung, which conforms to 5–15 mg of PSP of average surface area (20–40 m^2/g). It is now generally accepted that the continued presence of nontoxic particle material in the lungs, upon impairment of AM clearance, leads to a chronic inflammatory response, fibrosis, and tumor genesis in the rat. As discussed earlier, the overall pattern is one of chronic inflammation, which occurs upon saturation of lung clearance by overloading of macrophages. At this point, particle accumulation starts and inflammatory cell influx increases sharply.

The influx of neutrophils and associated DNA damage and proliferation are responsible for the mutagenicity, and the lung tumors after chronic particle exposure to PSP are due to their mutagenesis. However, several studies have generated data that deviate from this paradigm. A number of studies question the validity of the inflammation paradigm. In a rat study, mutations in the HPRT-gene in lung epithelial cells upon exposure to crystalline and amorphous silica were not increased in rats exposed to amorphous silica compared with controls, although the inflammation was similar to that induced by crystalline silica. Similarly, depletion of circulating neutrophils in rats by injection of anti-neutrophil serum before short-term inhalation of quartz particles (3 days, 100 mg/m^3) did not affect acute lung damage by quartz. It remains to be investigated whether the findings in these models using toxic quartz can be reproduced with PS.

A comparison of lung tumors at similar gravimetric dose (30 mg) of several NP (Table 6.15) shows that the amount of lung tumors induced by three different insoluble ultrafines is related to their surface area. DEP, with the lowest surface area (34 m^2/g), induced 19 tumors in 46 animals, while the carbon black (CB) (300 m^2/g) induced tumors in 32 out of 48 animals. Importantly, the high-surface amorphous silica induced 3 lung neoplasms, which might be due to its high solubility, i.e. a low durability *in vivo*. However, one has to caution against oversimplification and not rely solely on surface area, since apart from their small size, NP for commercial applications often have

Table 6.15 *Prevalence of lung tumors in rats 129 weeks after intratracheal administration of a 30-mg dose of different poorly soluble, low-toxicity particles.*[a]

Particle/treatment	Size (nm)	Surface Area (m^2/g)	Tumor incidence	
Control			0/91(%)	
TiO_2	30	50	29/47 (61.7%)	
Carbon Black	14	300	32/48 (68.1%)	
Diesel	ND	34	19/46 (41.3%)	1/46
Aerosil	14	200	3/38 (7.9%)	0/39

[a]Female Wistar rats were treated with multiple intratracheal injections (6 mg) at weekly intervals to reach the final cumulative dose (30 mg). Control animals remained untreated and received no injections. Tumors were evaluated by histopathological scoring of two sections per lung lobe.

chemically different surfaces. Evidence for the relevance of the chemical surface in PSP has also come from studies with surface-modified TiO_2. Surface modifications resulting in an enhanced hydrophobicity of TiO_2 have generally been found to lead to an amelioration of the inflammatory response, although initial studies showed an increased toxicity of surface-modified TiO_2. No data are available that connect this ameliorated response to chronic outcomes such as carcinogenicity.

Although some of the particles listed in Table 6.13 have been characterized as confirmed human carcinogens (Group I by IARC), PSP such as coal mine dust, pigmentary TiO_2, and CB were not associated with an increase in lung cancer in exposed workers. Thus, hazard assessment using rat studies raises the question of whether particles that induce tumors in this bioassay should be labeled as possible or even probable human carcinogens.

Cardiovascular Effects

Epidemiological studies indicate that the major cause of increased deaths upon ambient particle increase is among patients with cardiovascular diseases. In addition, experimental animal studies with combustion NP do show that high exposures to Diesel soot NP or other surrogate NP causes observable cardiovascular effects.

Combustion and model NPs can gain access to the blood following inhalation or instillation and can enhance experimental thrombosis. Diesel particles instilled into hamster lungs also enhance thrombosis but it is not clear whether this was an effect of pulmonary inflammation or of particles translocated to the blood. High exposures of combustion-derived NP (CDNP) by inhalation caused altered heart rate in hypertensive rats and dogs and this is interpreted as a direct effect of CDNP on the pacemaker activity of the heart. Inflammation in distal sites has also been associated with increased progression of atheromatous plaques in rabbits, ApoE-/- mice, and humans. In summary, the cardiovascular effects of inhaled particles (PM) are explained by theories that involve either impairment of the blood flow to or in the heart, or interference with autonomic innervation (Figure 6.13).

Inflammation can be seen as the main driver of the cardiovascular effects. There is evidence of systemic inflammation following increases in PM, as shown by elevated C-reactive protein, blood leukocytes, platelets, fibrinogen, and increased plasma viscosity.

Atherosclerosis is the underlying cause of acute coronary syndrome, the main cause of cardiovascular morbidity and mortality. Atherogenesis is an inflammatory process, initiated via endothelial injury, which produces systemic markers of inflammation that are risk factors for myocardial and cerebral infarction. Repeated exposure to PM_{10} may, by increasing systemic inflammation, exacerbate the vascular inflammation of atherosclerosis and promote plaque development or rupture of blood vessels. The inflammatory

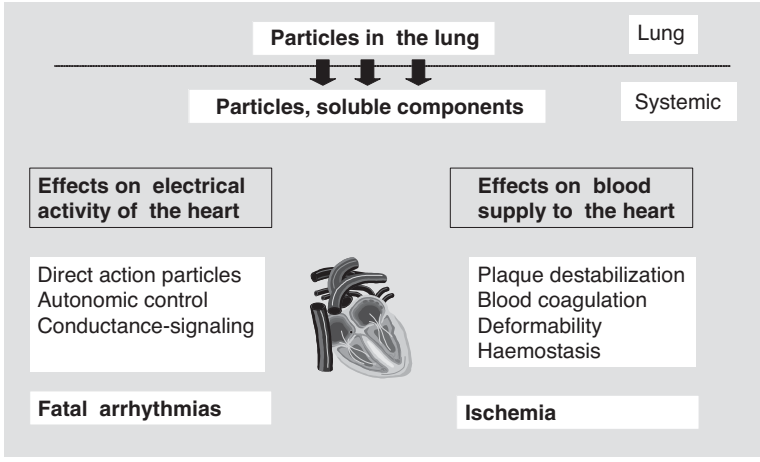


Figure 6.13 *Graphical illustration of the main mechanisms that may be involved in the cardiovascular effects of inhaled ambient particulate matter. It is suggested that particles cause an exacerbated inflammation in the lung, which may lead to translocation of particles, soluble components, and/or inflammatory mediators. These may affect electrical innervation of the heart and/or affect the blood flow to the heart, leading to arrhythmias or ischemic heart damage.*

response to particles or the particles themselves may also impact on the neural regulation of the heart, leading to death from fatal dysrhythmia. In support of this hypothesis, studies in humans and animals have shown changes in the heart-rate and heart rate variability in response to particle exposures.

6.3.6 Summary

Massive exposure to inhalable and respirable particles over a period of years at more than 5 mg/m^3 is known to cause both chronic obstructive pulmonary diseases including bronchitis, emphysema, fibrosis, pneumoconiosis, and lung cancers. Epidemiological studies have shown that ambient exposure to particles of 10 μm (PM_{10}) is related to both respiratory and cardiovascular mortality and morbidity. It is currently thought that ambient particulate matter with a mean diameter of 2.5 or 10 μm ($PM_{2.5}$, PM_{10}) can be carcinogenic, but it is not clear which components or characteristics of PM are responsible.

To understand particle effects, one can apply the 'D-concept' (Dose, Dimension, Deposition, Durability, and Defense). This paradigm builds a conceptual understanding of the adverse effects of particles and fibers. Fiber dimensions determine the ability of the body to defend against particles and the persistence. Long (>20 μm) fibers are not taken up by alveolar macrophages, and, therefore, have longer half-lives in the lung than shorter fibers and result in a higher cumulative dose at similar inhaled fiber number or mass. Inhaled fibers longer than 20 μm also show interception at the bronchiolar bifurcations due to their length and may achieve high local doses by this process. If a

particle is not soluble or not degradable in the lung it has a high durability and there will be rapid local accumulation upon sustained exposure. Particle size largely determines where particles are deposited in the respiratory tract. Particle deposition is further determined by shape and density. Small ultrafine particles have little mass and behave by diffusion. Most of them are deposited in the nasal compartment and the alveoli. Particle translocation has only recently been recognized as a mechanism to explain possible systemic effects of inhaled ambient particles.

One of the crucial properties of particles is their ability to generate reactive oxygen species (ROS) and reactive nitrogen species (RNS). This factor is involved in many pathological outcomes of particle exposure, including fibrosis, proliferative effects, and carcinogenicity. Apart from producing ROS, inflammatory phagocytes produce growth factors such as cytokines and chemokines that regulate cell migration, tissue remodeling, cell proliferation, and repair of damage.

Differences in particle toxicity can be related to their surface reactivity, their ability to adsorb various chemicals including metals that enhance ROS-generation and inflammation.

Most toxic mechanisms of nanoparticles are probably qualitatively not different from cell–particle interactions for fibers and fine particulates. Quantitatively, however, nanoparticles can induce more inflammation at considerably lower gravimetric lung burdens than can their larger analogues. Other mechanisms such as translocation to the brain or the vascular system are substantially different from their fine analogues.

References

Borm, P.J.A., Schins, R.P.F., and Albrecht, C.A. Inhaled Particles and Lung Cancer. Part B: Paradigms and Risk Assessment. *Int. J. Cancer*, 2004, **110**(1), 3–14.

Oberdorster, G. Plumonary Effects of Inhaled Ultrafine Particles. Int. Arch. Occup. Environ. Health, 2001, **74**(1), 1–8.

Morrow, P.E. Possible Mechanisms to Explain Dust Overloading of the Lungs. *Fundam. Appl. Toxicol.*, 1988, **10**, 369–384.

Schins, R.P.F. and Borm, P.J.A. Common Cellular and Molecular Mechanisms in the Development and Progression of Pneumoconiosis. Toxiological Characterization of Engineered Nanoparticles, Chapter 18 in Occupational Lung Disease: An International Perspective, 2nd edition (ed. Banks, D.) Hodder Arnold, in press.

Tran, C.L., Buchanan, D., Cullen, R.T., Searl, A., Jones, A.D., and Donaldson, K. Inhalation of Poorly Soluble Particles. II. Influence Of Particle Surface Area on Inflammation and Clearance. *Inhal Toxicol*., 2000, **12**, 1113–1126. **Surface area and not mass drives the inflammatory response to particles**.

For Further Reading

Borm, P.J.A. and Kreyling, W. Toxicological Hazards of Inhaled Nanoparticles – Potential Implications for Drug Delivery. *J. Nanosci. Nanotechnol*., 2004, **4**, 521–531.

Dockery, D.W., Pope, C.A. 3rd, Xu, X., *et al*. An Association between Air Pollution and Mortality in Six U.S. Cities. *New Engl. J. Med.*, 1993, **329**(24), 1753–1759. **First epidemiological study highlighting the potential effect of ambient particles on mortality.**

Donaldson, K., Tran, L., Jimenez, L.A., Duffin, R., Newby, D.E., Mills, N., MacNee, W., and Stone, V. Combustion-derived Nanoparticles: A Review of Their Toxicology following Inhalation Exposure. *Part. Fiber Toxicol.*, 2005, **2**, 10.

Greim, H., Borm, P., Schins, R., Donaldson, K., Driscoll, K., Hartwig, A., Kuempel, E., Oberdorster, G., and Speit, G. Toxicity of Fibers and Particulates – Report of a Workshop held in Munich, Germany, October 26–27 2000. *Inhal. Toxicol.*, 2001, **13**, 737–754.

Knaapen, A., Borm, P.J.A., Albrecht, C., and Schins, R.P.F. Inhaled Particles and Lung Cancer. Part A: Mechanisms. *Int. J. Cancer*, 2004, **109**(6), 799–809. **Review**.

Mills, N.L., Amin, N., Robinson, S.D., Anand, A., Davies, J., Patel, D., de la Fuente, J.M., Cassee, F.R., Boon, N.A., Macnee, W., Millar, A.M., Donaldson, K., and Newby, D.E. Do Inhaled Carbon Nanoparticles Translocate Directly into the Circulation in Humans? *Am. J. Respir. Crit. Care Med.*, 2006, **173**(4), 426–431.

Nemmar, A., Hoylaerts, M.F., Hoet, P.H., Dinsdale, D., Smith, T., Xu, H., Vermylen, J., and Nemery, B. Ultrafine Particles Affect Experimental Thrombosis in an in vivo Hamster Model. *Am. J. Respir. Crit. Care Med*., 2002, **166**(7), 998–1004.

Nel, A., Xia, T., Madler, L., and Li, N. Toxic Potential of Materials at the Nanolevel. *Science*, 2006, **311**(5761), 622–627. (2006). **Short review with the conceptual thinking of toxicologists on nanoparticle adverse effects.**

Oberdorster, G., Sharp, Z., Atudorei, V., Elder, A., Gelein, R., Kreyling, W., and Cox, C. Translocation of Inhaled Ultrafine Particles to the Brain. *Inhal. Toxicol.*, 2004, **16**(6–7), 437–445. **Key-publication for the future and has set the stage for further research on effects of particles in the brain.**

Oberdorster, G., Oberdorster, E., and Oberdorster, J. Nanotoxicology: an Emerging Discipline Evolving from Studies of Ultrafine Particles. *Environ. Health Perspect*., 2005, **113**(7), 823–839. **Comprehensive review on mechanisms and actions of inhaled nanoparticles by the founder of this field.**

Pope, C.A., 3rd, Burnett, R.T., Thurston, G.D., Thun, M.J., Calle, E.E., Krewski, D., and Godleski, J.J. Cardiovascular Mortality and Long-term Exposure to Particulate Air Pollution: Epidemiological Evidence of General Pathophysiological Pathways of Disease. *Circulation*, 2004; **109**(1), 71–77.

6.4 Xenoestrogens and Xenoantiandrogens

Gisela H. Degen and J. William Owens

6.4.1 Introduction

The central role of estrogens/androgens in developmental processes, reproduction, and in hormonal carcinogenesis (Section 2.7.2 and Chapter 2.8), has made the synthesis of reproductive steroids and their mode of action among the best studied and most well known in pharmacology. As a result, numerous pharmacological agents are available for daily use in humans and animals [e.g., estrogens: ethinylestradiol (EE); antiestrogens: tamoxifen, raloxifene, and faslodex; androgens: testosterone propionate, methyl testosterone, and trenbolone; antiandrogens: cyproterone acetate, flutamide, nilutamide, and bicalutamide; aromatase inhibitors: anastrozole and letrozole; 5α-reductase inhibitors: finasteride and dutasteride; and several more general steroidogenesis inhibitors: ketoconazole and spironolactone]. Further, a battery of *in vitro* and *in vivo* assays has been developed to guide pharmaceutical development, assessing the activity and potency of such compounds.

During the last decade *e*ndocrine *a*ctive *c*ompounds (EACs), i.e. compounds capable of interfering with the endocrine system, have become a widely discussed issue and topic of active research in toxicology and related disciplines. The focus of these concerns is that the agents may be widely used commercial chemicals and that exposure to such agents may result in detrimental effects in humans and animals. Chemicals which act as estrogen-mimics or function as antiandrogens are known as *xenoestrogens* and *xenoantiandrogens.* They are the best studied group of EACs and illustrate the issue of endocrine-related modes of action and potential toxicities of commercial chemicals. An endocrine activity is not necessarily associated with the induction of adverse effects, EACs that induce negative health effects (via their endocrine activity) are considered as 'endocrine disruptors.' The following definitions have been crafted by regulatory consensus to define the endocrine issue.

An **endocrine disruptor** is an exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse health effects in an intact organism or its progeny, or (sub)populations.

A **potential endocrine disruptor** is an exogenous substance or mixture that possesses properties that might lead to endocrine disruption in an intact organism or its progeny, or (sub)populations.

These definitions [adopted by the International Programme on Chemical Safety (IPCS) and the Organization for Economic Co-operation and Development (OECD)] are based on those drafted at the Weybridge Conference in 1996.

The essential point for toxicologists is that endocrine disruption is inherently described as a mode of action. These endocrine modes of action are among several that may potentially lead to adverse effects on reproduction, growth, and development. An underappreciated implication is to distinguish assays identifying modes of action (typically referred to as screens) from the classical assays used by toxicologists to characterize adverse effects, i.e. reproductive hazards and dose responses, such as multigenerational assays. These distinctions and their implications are emphasized in the following statement.

'Endocrine disruption is not considered a toxicological endpoint *per se* but a functional change that may lead to adverse outcomes.' [Damstra, *et al.*, IPCS, 2002]

6.4.2 Toxicities

Most EACs studied so far in laboratory experimental animals affect the target tissues of sex hormones; other organs may also be affected, such as the thyroid, immune, and neuroendocrine system, for instance by chemicals covered in Chapter 6.1. The effects vary from subtle changes in the physiology to permanently altered sexual differentiation.

The spectrum of changes is well known from reprotoxicity studies with potent hormones developed as drugs [Chapter 2.8].

Exposure to endocrine active compounds (EACs) is considered to be particularly **critical during prenatal** and **postnatal phases** and may later result in a permanent change of function or sensitivity to hormonal signals.

Exposure to EACs in adulthood may be compensated for by homeostatic mechanisms or result in toxicities that are reversible when exposure ends.

An instructive case is that of **diethylstilbestrol** (DES), an estrogen whose potency is equivalent to that of EE and that readily passes the placenta. In the 1950s and 1960s, DES was prescribed to a very large number of pregnant women at massive doses. These doses ranged from 5 mg per day up to 125 mg per day or approximately 2 mg/kg-bw/d, which compares with a normal contraceptive dose of 1 μg EE/kg-bw/d for birth control.

DES use was found to be associated with an increased risk of breast cancer in the 'DES-mothers' who received high doses. In their male offspring exposed to DES *in utero,* urogenital tract abnormalities (cryptorchidism, hypospadias, epididymal cysts, and hypoplastic testes) were found more frequently than in nonexposed males. In grown-up 'DES-sons,' average sperm counts were only slightly lower than in controls. Although sperm quality was clearly lower, this has been attributed to a higher incidence of hypoplastic testes in this group. However, there was no indication for reduced fertility when these persons were studied again at the age of 38–41. Earlier concerns on a possibly increased incidence of testicular tumors (untreated cryptorchidism is a known risk factor) have not been confirmed in more recent epidemiological studies.

DES has increased the incidence of a very rare tumor in young women exposed *in utero* during critical phases of development. The risk of this cancer (clear cell adenocarcinoma of the vagina) was estimated to be in the order of 1 per 1000 in the 'DES-daughters.' Other, nonmalignant reproductive tract abnormalities were observed much more frequently in this group.

Recent studies compared breast cancer incidence in women with or without prenatal DES exposure (where cohorts have been followed since the 1970s): The results indicate that women with prenatal exposure to DES have an increased risk of breast cancer after the age of 40 years. Apparently, *in utero* exposure to DES significantly increased postmenopausal, but not premenopausal, breast cancer incidence (Palmer *et al*., 2006).

In animal studies, the susceptibility of fetal developmental processes to this very potent synthetic hormone are further reinforced. Prenatally treated female mice showed a dose-dependent decrease in fertility, detectable already with daily maternal DES-doses of 0.01 μg/kg body weight. In male mice treated *in utero*, reduced fertility was found only with higher maternal DES-doses (10 μg/kg body weight).

Data for DES are consistent with other reprotoxicity studies that find stronger effects of estrogenic compounds on females than on males.

More recent studies investigated multi-generational effects of DES: Whilst reduced fertility observed in the DES F1 female mice was not transmitted to their descendants, an increased tumor susceptibility may be passed on to the next generation (Newbold, 2002). It is important to point out that increased occurrence of reproductive-tract tumors has not been reported in the offspring (grandchildren) of the prenatally DES-exposed human cohort.

Work with DES in experimental animals has greatly improved our knowledge of the spectrum of effects to be expected in males and females from perinatal EAC exposure, and even predicted some outcomes observed later in clinical studies on DES-exposed humans. DES studies provide also important data on **dose–effect relationships** in rodents and in humans. Epidemiological data gave some insight, because of marked differences in DES-dosing schedules used at various clinical centers: In cohorts of males prenatally exposed to comparatively low maternal doses (estimated mean total of 1.4 g at the Mayo Clinic) there were no indications for adverse consequences such as those described in the high-dose DES cohorts (estimated mean total dose of 11.6 g at the University of Chicago). The human data (reviewed by Golden *et al*., 1998) are consistent with the existence of maternal dose levels below which adverse noncancer effects may not occur. The extensive rodent DES reproductive/developmental toxicity data are also consistent with this view: DES effects on fertility and the occurrence of genital-tract abnormalities are consistent with the observations of dose levels below which adverse noncancer effects may not occur, i.e. dose–response thresholds.

Many well conducted studies on the reproductive/developmental toxicity of various xenoestrogens confirm the view that **adverse effects are related to dose and hormonal potency of the compound in question.**

Adverse effects observed in **wildlife** have been proposed as *sentinels* of human exposure to EACs. However, the impact of certain EACs on the developmental/reproductive health of humans and various animals can differ considerably due to differences in exposure scenarios, toxicokinetics, and/or toxicodynamics in a given species or class of animal. Caution must be taken in extrapolations, and also, in light of the diversity of wildlife (mammals, birds, reptiles, amphibians, fish, and invertebrates), in comparing their particular endocrine systems and/or habitats. Adverse effects observed in certain species can be linked to persistent chemicals acting as EACs, mostly in areas with high levels of pollution. In other cases such an association is weak and/or potential confounders cannot be excluded (Further Reading). The complex issue of ecotoxicology (see Chapter 2.9) is beyond the scope of this chapter; yet, some examples of toxicities attributed to environmental EAC exposure will be mentioned in the context of compounds and their mode of action.

6.4.3 Modes of Action and Testing

The fundamental hormonal signaling system is based on a feedback system involving the hypothalamic/pituitary/gonad axis, and this is schematically depicted in **Figure 6.14.** As the steroid hormones act via an intracellular, nuclear receptor (see Chapter 2.5), the

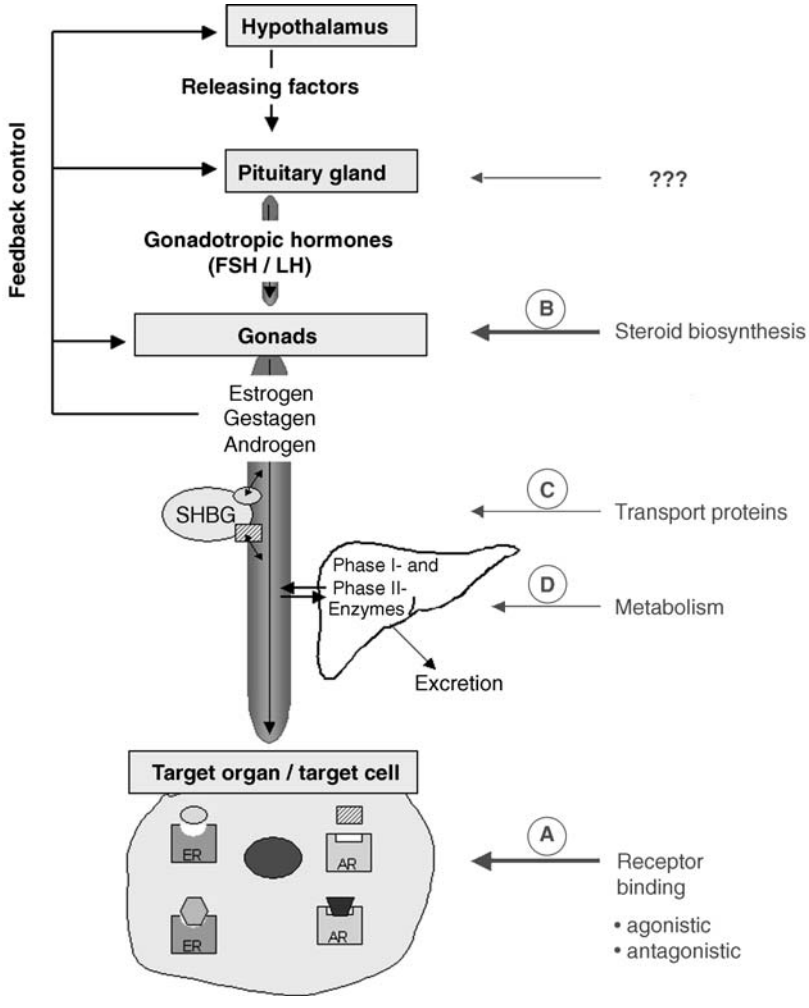


Figure 6.14 *Scheme of hormone signaling components. Chemicals can interfere by different mechanisms (A to D) with the sex hormone system (Section 6.4.3).*

feedback system can be modulated in two basic ways: 1) agonists or antagonists of the respective estrogen and androgen receptors, and 2) interference with steroid biosynthesis and metabolism, particularly important terminal steps such as the conversion of testosterone into estrogen by aromatase and the conversion of testosterone into the more potent dihydrotestosterone by 5α-reductase.

The set of *in vitro* and *in vivo* tools developed by research pharmacologists are now being applied to address the questions about weakly potent commercial chemicals. This section is devoted to briefly illustrating several of these tools and their potential application.

Screening for Modes of Actions in vitro

Data from *in silico* (structure–activity relationships) and *in vitro* studies can provide a very rapid orientation on whether compounds may interact with elements of the endocrine system and the (relative) hormonal potency of chemicals, e.g. an estrogenic activity in relation to estradiol, and do so while avoiding the use of laboratory animals. However, there are also limitations: i) *in silico* systems do not incorporate and cell-based *in vitro* systems often lack enzymes required for the conversion of pro-hormones into biologically active metabolites (example methoxychlor in Section 6.4.4, below); ii) neither do they reflect other important aspects of toxicokinetics (bioavailability, tissue distribution) of a given compound that also determine their potency *in vivo*; and iii) cell-culture systems may simulate the responses of one but not those of another target organ. Therefore, *in silico* systems and *in vitro* assays are used for rapid prioritization, thereby avoiding using significant numbers of animals. However, it is inappropriate to use *in silico* methods to identify substances as potential endocrine disruptors, and, *in vitro* systems should be used with caution.

Direct Receptor Binding A significant number of xenobiotics bind to estrogen receptors (ER, more precisely the subtypes ERα and ERβ). A smaller number interact with androgen receptors (AR), mostly acting as antagonists. While endogenous steroid hormones act in nM or sub-nM concentrations, the binding affinity of environmental chemicals is often several orders of magnitude lower than that of the endogenous ligands (Section 6.4.4, below). Thus, efficient receptor occupancy by xenobiotics requires μM to mM intracellular concentrations of the free ligand. Accordingly, higher xenobiotic doses are needed to elicit hormonal effects *in vitro* or in bioassays *in vivo*.

Advantages and limitations of existing test methods are briefly discussed below. Compounds and metabolites that bind directly to sex hormone receptors (Figure 6.14, case A) are rather easy to detect and this may explain the higher number of suspect compounds by comparison to EACs operating by indirect modes of actions.

Steroid Biosynthesis Chemicals may affect the biosynthesis of sex hormones (Figure 6.14, case B) and/or its regulation by the hypothalamic–pituitary–gonadal axis. With several enzymes catalysing the many steps in steroid biosynthesis from cholesterol, the challenge is to pinpoint the site of interference. The recently developed adenocarcinoma cell line H295R possesses an intact steroid-synthesis pathway and, therefore, can identify modulation of the entire steroidogenesis pathway from cholesterol recruitment by the STAR protein through the terminal conversion of androgens into estrogens. This is accomplished by analytical determination of the concentration of the final and intermediate products in response to doses of the putative disruptor.

Steroid and Thyroid Hormone Metabolism The transport and degradation of steroid and thyroid hormones (Figure 6.14C and D) may be altered by xenobiotics, e.g. by inducers or inhibitors of hepatic enzymes. A large number of drugs and natural food ingredients are known to modulate expression or activity of phase I/II enzymes. These indirect and difficult-to-demonstrate mechanisms reinforce the need for rather stringent definitions of (potential) endocrine disruptors. Those now widely adopted (Box 1,

Definition Section 6.4.1, above), and the following statement acknowledge the higher rank of *in vivo* studies for a risk characterization of EACs. Hepatocyte cell lines employing toxicogenomic and proteomic techniques are under development to identify the genes and metabolic pathways that can be induced by both natural and anthropogenic compounds. The capability of these enzymes to metabolize steroids and thyroid hormones can then be assessed.

Screening for Modes of Actions in vivo

The current regulatory strategy to identify potential EACs is to adapt and utilize existing bioassays as they are rapid, use a minimum number of animals, have proven to be highly specific for drugs, and often correlate to assays for adverse reproductive effects. The advantages are that they incorporate Absorption, Distribution, Metabolism, Excretion (ADME), but due to their short periods of administration, the maximum tolerated dose (MTD) in subchronic and chronic studies can be exceeded.

Estrogens and Antiestrogens The uterotrophic bioassay has been employed for over 70 years to identify estrogens and was adapted in the 1970s to identify antiestrogens. Uterine growth is a direct target for estrogens, and the uteri of ER knock-out mice do not respond to estrogens. The bioassay has been validated by the OECD using several weak, commercial estrogenic compounds using small group sizes ($n = 6$). The blotted uterine weight is equally sensitive to histological markers such as epithelial cell height of the uterus, and recent toxicogenomic studies have confirmed the estrogenic nature of the weight increase. For estrogens, with the short estrous cycle of rodents, after 3 consecutive days of administration, the uteri from the putative test substance groups are compared with the uteri from vehicle controls on day 4. For antiestrogens, a stimulating dose of EE and the putative antiestrogen are coadministered to assess the inhibition of uterine growth.

Androgens and Antiandrogens The Hershberger bioassay has also been employed for over 70 years to identify androgens and was adapted in the late 1960s to identify anitandrogens. Male accessory reproductive tissues and muscles such as the levator ani are direct targets for androgens, and the tissues of AR knock-out mice as well as humans with genetic defects in the AR (androgen-insensitivity syndrome) have little or no response to androgens. The bioassay has been validated by the OECD for regulatory use using several weak, commercial antiandrogen compounds; again, using small group sizes ($n = 6$). Androgens are administered to castrated males after a period of tissue regression, while antiandrogens are coadministered with a stimulating dose of testosterone propionate.

Aromatase The uterotrophic bioassay has been modified by the pharmaceutical industry for over 20 years to identify aromatase inhibitors. The immature rodent is administered a potent, aromatizable androgen, and upon conversion into estrogen rapid uterine growth is observed. To study aromatase inhibitors, the putative inhibitor is coadministered with the androgen, and, if effective, then the increase in uterine weight is inhibited.

5α-Reductase The antiandrogen protocol of the Hershberger bioassay has shown to present a feasible *in vivo* model for 5α-reductase inhibition, using finasteride as a model compound during the OECD validation program. Testosterone propionate was coadministered, and decreases in tissue weight gains were elicited by the finasteride.

General Steroidogenesis A protocol based upon the 15-day intact male assay has been developed which presents an *in vivo* model for steroidogenic inhibitors as well as other modes of action. The premise of the assay is to combine a battery of tissues and circulating levels of several serum hormones to produce a profile characteristic of endocrine modes of action. The assay requires a larger group size ($n = 15$) to provide sufficient power for the hormone analyses, a detailed protocol to minimize hormonal variation, and a laboratory with expertise in sampling animals and running a battery of hormonal assays. The assay has been demonstrated against pharmaceutical compounds, and awaits formal validation with weakly potent compounds. Classical assays like 28- or 90-day repeat-dose studies in rodents can also contribute to the evaluation of such effects; however, as normally only weights and histology of hormone-producing glands and their target tissues are investigated, little information on the exact mode of action is observed.

Thyroid Toxicity Thyroid toxicity is an example where classical assays are already in place and sufficient to evaluate this endocrine mode of action, such as thyroid weight changes and histopathological evaluations after 28- or 90-day repeat-dose studies in rodents. These can be supplemented with hormone assays where animal handling, serum sampling, and hormonal analyses are carefully done to minimize T3 and T4 variations. Alternatively, the 15-day intact male assay can be used. The rat is a sensitive model due to the lack of a specific, high-affinity serum carrier for thyroid hormones, and should be sensitive for vertebrates in general. However, the precise mode of action is not identified, *e.g.* symporter inhibition, peroxidase inhibition, or increased metabolic removal of circulating hormone.

Steroid and Thyroid Hormone Metabolism The liver is a primary target for the induction of metabolic pathways that increase the excretion of steroid and thyroid hormones. Classical observations of increased liver weight and histopathology are now being supplemented with identification of individual messenger RNA and protein quantities/activities for specific metabolic enzymes, while toxicogenomic and proteomic techniques are under development to elucidate complete pathways of metabolism. Metabolic screens of serum, bile, and feces also provide a capability to assess increased hormonal turnover when the potential for increased hormone metabolism has been identified.

Wildlife Mode-of-action assays have been also developed in certain wildlife models. The egg lipoprotein, vitellogenin (VTG), is being validated in fish as a marker for estrogens, antiestrogens, and aromatase inhibitors. VTG production in the liver and its circulating levels are under estrogen control. The external features of sexually dimorphic species such as the nuptial fat pad and sexual tubercles in the fathead minnow are under the control of androgens. They are now explored as suitable parameters as well as androgen specific proteins, such as the production of spigin protein in the kidney of the three-spined stickleback.

A Weight-of-Evidence Approach and Future Improvements

Data for the interaction of a chemical with components of the endocrine system and any adverse health outcome will come from a variety of sources, and, thus, evaluation will require a weight-of-evidence approach. The weight-of-evidence approach has several components, many of which were noted by Hill (1965), such as plausibility, reproducibility, coherence, consistency, and relevance. A set of decision rules is also needed, such as the toxicological realism of *in vivo* screens should outweigh or supersede *in vitro* findings. These *in vivo* screens then may provide a basis for estimating whether i) exposures may be sufficient to pose a threat to wildlife or humans or ii) large margins of safety may exist between doses that could lead to endocrine activity in screens and actual exposures. In this latter case, in order to use a screen in this context, the comparative data between the screen lowest-observed-effect level (LOEL) and those observed in definitive studies will be needed. Thus, the entire knowledge set of hazard identification, characterization of endocrine modes of action responses, and exposure should be considered together to identify those compounds requiring characterization for adverse effects.

In regards to identifying specific modes of action, the promise of the application of toxicogenomic and metabolic techniques is worthy of note. All of the above screens have limitations and a certain low rate of false positives has been observed in the uterotrophic and Hershberger validation programs. The inherent advantage to toxicogenomics is the specificity of a gene profile where the annotation of individual genes supports a specific mode of action. Work has already been conducted by several laboratories showing that a specific profile of uterine and ovarian genes can be identified that correlates with uterotrophic action. Work has also been recently published to elucidate similar profiles for androgens and antiandrogens.

Definitive tests are exemplified by the multi-generation reproductive and development assays which identify adverse effects in both adult sexes and adverse effects in offspring elicited in the absence of maternal toxicity. These definitive tests are the basis for both classification and labeling as well as the no-observed-adverse-effect levels (NOAELs) used to establish regulatory allowable daily intakes (ADIs) or margins of exposure or safety (MOS). A number of endpoints are associated with endocrine action, such as premature vaginal opening in the case of estrogens; and delayed preputial separation, the appearance of specific patterns of male reproductive tract malformations, and nipple/areola retention in the case of antiandrogens. However, such endpoints have not traditionally been commonly employed, so these data may be absent from many studies. Even with these endpoints, definitive assays for adverse effects are not specific for modes of action. Thus, clearly defined criteria to accept a causal linkage between positive findings in mechanistic assays and the appearance of particular endpoints in definitive assays is needed. Two criteria would be 1) a similarity of the dose–response characteristics [similar LOELs and lowest-observed-adverse-effect levels (LOAELs)] and 2) that the putative endocrine-related effect observed in the definitive study was indeed the clear primary effect in the definitive assay. Again, a weight-of-evidence approach will be needed combining both screening and definitive data to arrive at the conclusion that a compound fulfils the Weybridge definition of true endocrine disruptor - that adverse effects are the result of an endocrine mode of action as illustrated in the following box.

In assessing whether a compound is indeed an endocrine disruptor, it is important to distinguish between the responses observed in *in vitro* and *in vivo* screens and the observation in a definitive assay of an *adverse* effect. Screening assays such as the uterotrophic bioassay generate alerts to *potential endocrine disruptor*. These and other data can be utilized to prioritize substances for definitive tests. These assays, such as one- or two-generation rodent assays or life-cycle assays in nonmammalian species, determine the toxicological consequences and characterize the hazard. Together, these data indicate whether an adverse effect has occurred that is due to an endocrine mode of action, and, thus, fulfils the Weybridge definition of an *endocrine disruptor*.

6.4.4 Compound Assessment

EACs may be grouped according to their origin (e.g., natural and anthropogenic) and their mode of action (e.g., estrogenic or antiandrogenic). Table 6.16 lists naturally

Table 6.16 *Environmental chemicals of natural origin and of anthropogenic origin; agents and metabolites with hormonal activities other than estrogenicity [E] are marked as anti-/androgenic [AA; A], antiestrogenic [AE], gestagenic [G], or glucocorticoid-like [GC], and + positive, − negative, or +/− equivocal.*

Substances by class	*In vitro* screen[a]	*In vivo* screen[b]	Definitive data
Mycoestrogens natural chemicals			
Zearalenone [E]	+	+	+ Domestic animals
(α-, β-)Zearalenol [E]	+	+	+ Domestic animals
Phytohormones natural chemicals			
Coumestrol [E]	+	+	+ Domestic animals/ reproductive studies [c]
Daidzein [E]	+	+	+ Short-term developmental study [c]
Equol [E]	+	+	+ Domestic animals
Genistein [E]	+	+	+ Domestic animals/ reproductive studies
Glycitein [E]	+	+	No study available
Glycyrrhetinic acid [GC]	NA	NA	Reproductive studies not available; + human case reports [GC]
Resveratrol [E]	+	mixed	No study available
β-Sitosterol [E]	+	+	Studies inadequate for determining estrogenic and androgenic activity
Current and banned pesticides anthropogenic			
Aldrin [E]	−/+	−	− Reproductive study[d]
Atrazine [E, AE, AA]	−	−	− Reproductive study [e]
Chlordecone / Kepone	−	−	Studies inadequate for determining estrogenic and androgenic activity
o,p′-DDT, *o,p′*-DDE [E]	+	+	+ Short-term developmental study
p,p′-DDT, *p,p′*-DDE [AA]	+	+	+ Short-term developmental study
Dieldrin (major metabolite of aldrin) [E]	−/+	−	− Multi-generation repro study[d]
Endosulfan [E]	−	−	− Reproductive study [d]
Fenarimol [AA]	+	+	+ Reproductive study
Heptachlor	−	−	− Multi-generation repro study[d]

Table 6.16 *Continued.*

β- and γ-Hexachlorohexane (Lindane)	−	−	− Multi-generation repro study[d]
HPTE [E, AA]	+	+	+ Metabolite - see methoxychlor
Linuron [AA]	+	+	+ Reproductive study
Methoxychlor [E, AA]	+	+	+ Reproductive study
Mirex	−	−	Studies inadequate for determining estrogenic and androgenic activity
Procymidone [AA]	+	+	+ Multi-generation repro study
Toxaphene [E]	−/+	−	Studies inadequate for determining estrogenic and androgenic activity
Vinclozolin (M1 and M2) [AA]	+	+	+ Multi-generation repro study
Natural Steroids in Meat and Food Products			
Equilin, Equilenin [E]	+	+	No study available
Estradiol [E]	+	+	+ Multi-generation repro study
Estrone, Estriol [E]	+	+	+ Reproductive study
Progesterone [G]	+	+	+ Human volunteers and + Reproductive studies
Testosterone [A]	+	+	+ Reproductive study
PBTs (some banned, some under consideration for ban) anthropogenic			
PBDEs	−	−	− Reproductive studies[f]
PCBs	−	−	− Reproductive studies[f]
Hydroxy-PCBs [E]	+[g]	mixed	− Reproductive studies[f]
PCDDs and PCDFs [AE]	−	−	+ Reproductive study, cancer study
TCDD [AE]	−	−	+ Reproductive study, cancer study
Tributyltin [A; GC]	−	−	+ Environmental populations/ +/− Multi-generation repro study[h]
Commercial substances anthropogenic			
Amsonic acid [E]	+	+	No study available
BP A [E]	+	+	− Multi-generation repro study[d]
Butyl benzyl phthalate [AA]	−	−	+ Multi-generation repro study[i]
Di-n-butyl phthalate [AA]	−	−	+ Reproductive study[i]
DEHP [AA]	−	−	+ Multi-generation repro study[i]
NP [E]	+	+	+ Multi-generation repro study
OP [E]	+	+	− Multi-generation repro study
Organosiloxanes (D4, D5) [E]	+	+	− Multi-generation repro study
Pharmaceuticals anthropogenic			
Anastrozole [AE]	+	+	+ Multiple studies
Cyproterone acetate [AA, G]	+	+	+ Multiple studies
Ethinylestradiol [E]	+	+	+ Multiple studies
Finasteride [AA]	+	+	+ Multiple studies
Flutamide [AA]	+	+	+ Multiple studies
Tamoxifen [AE]	+	+	+ Multiple studies
Trenbolone [A]	+	+	+ Multiple studies

NA – not available.
[a]*in vitro* assay: Estrogen- or androgen-receptor binding or reporter gene cell-based assays.
[b]*in vivo* assay: Uterotrophic or Hershberger assays.
[c]Observation of estrogenic-related (e.g., accelerated vaginal opening in reproductive study or for antiestrogenic, decrease in mammary cancer) or androgenic-related (e.g., phallic clitoris and masculinization in female offspring and, for antiandrogenic, malformations after in utero exposure during gestational days 14–18) findings.
[d]No observation of estrogenic or androgenic-related findings. Findings indicate action as dopamine antagonist.
[f]Evidence suggests some congeners may displace T3 and T4 from carrier proteins, resulting in accelerated metabolism leading to thyroid toxicity.
[g]Only a subset of the family of congeners.
[h]Some changes in anogenital distance, suggesting possible androgenic activity, but other endpoints not affected.
[i]Evidence suggests no direct antiandrogenic action at the AR receptor, but action by interference with testosterone synthesis.

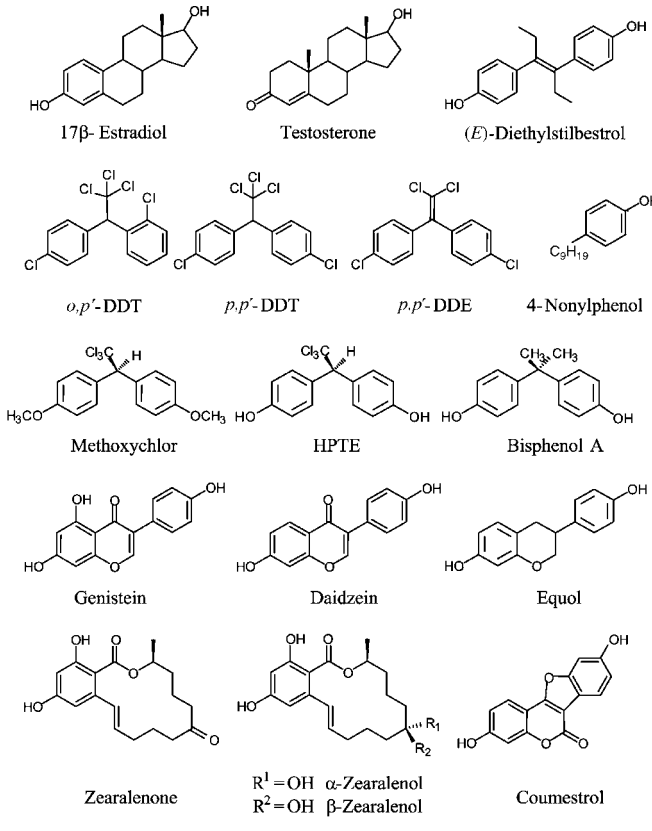


Figure 6.15 *Synthetic and natural compounds that bind directly to sex-hormone receptors and act as agonists or antagonists. See text for details (Section 6.4.4; abbreviations are given in a box at the end of this chapter).*

occurring compounds (phytochemicals, mycoestrogens, dietary steroids), and anthropogenic chemicals (pesticides, other chemicals and drugs) where the mode of hormonal activity (estrogenic, antiandrogenic, or other) is suspected or where evidence contradicts some original supposition that the compounds were involved in direct endocrine modes of action. Examples of EACs acting via *direct* binding to sex-hormone receptors are depicted in Figure 6.15. Compounds shown in Figure 6.16 apparently exert a hormone-mimetic or hormone-blocking activity by *indirect* modes of action.

The term *xenohormone* is often used for synthetic environmental chemicals alone, but in a stricter sense all agents listed here are *xeno*biotics. They are found in rather **different amounts** in foods, water, or elsewhere in the environment of humans and animals. Moreover, there are considerable differences in hormonal **potency** among EACs. This pertains to substances of both natural and anthropogenic origin.

Estrogens

The basic structural requirements for binding to the estrogen receptor are an aryl ring with a hydroxy group and a bulky, hydrophobic group in the *para* position, and no

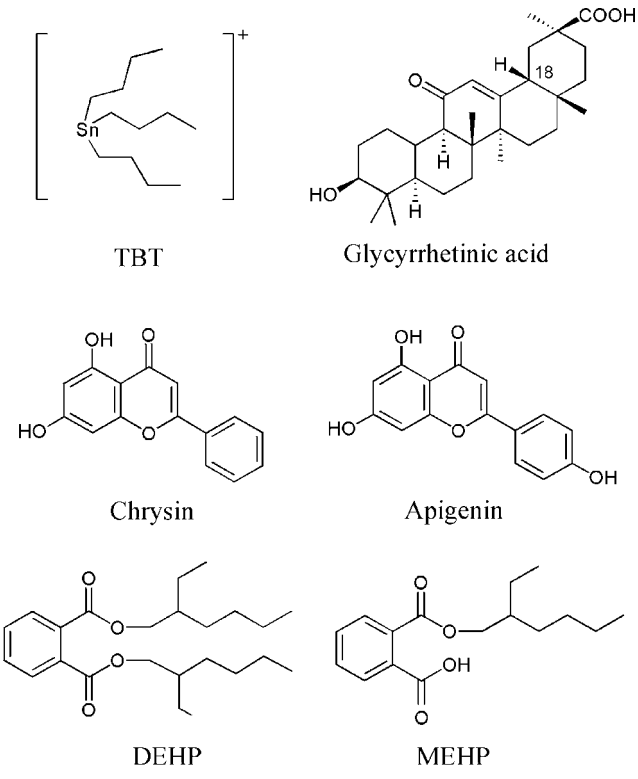


Figure 6.16 *Structures of compounds known to exert indirect effects on the endocrine system. See text for details (Section 6.4.4; abbreviations are given in a box at the end of this chapter).*

hindering groups at the *meta* and *ortho* positions. These structural features are common to many natural and anthropogenic compounds. This section highlights the diversity of these compounds and how a tiered assessment process can and has been utilized. The compounds have been chosen to represent a spectrum of potencies compared with endogenous estrogens. Their impact on human health may range from '*insignificant*' to '*likely to be of concern*.' Their impact on wildlife can be rather different due to differences in exposure scenarios and/or susceptibility.

Phytoestrogens are plant constituents belonging to several chemical classes: flavonoids, isoflavones, lignans, stilbenes, and steroids. The **phytoestrogens** bind to ERα and ERβ, elicit estrogenic responses in cell-based transcriptional assays, and are clearly positive in uterotrophic bioassays. Coumestrol is the most potent compound (100–1000-fold less active than estradiol), followed by the soy isoflavones genistein and daidzein (≥1000-fold less active than estradiol) and its metabolite equol (Figure 6.15). Other isoflavones (formononetin, biochanin A, glycitein) and lignans (enterolactone and enterodiol derived from colonic metabolism of matairesinol and secolairiciresinol) are less active. Certain phytoestrogens (coumestrol, genistein) are better ligands for one estrogen receptor subtype than another, and this raises the possibility of selective ER modulation and differential effects in target tissues with varying ERα/β contents.

Most phytohormones can be classified as weak estrogens. Yet, when consumed in sufficient amounts they can affect the endocrine system of animals and humans, with favorable as well as unfavorable consequences.

In the 1940s and '50s, reproductive disorders (including spontaneous abortion and sterility) in sheep had been linked to high isoflavone exposure consumed from forage, and rodent studies revealed developmental toxicity for several phytoestrogens (further reading). A reprotoxic and carcinogenic potential has been demonstrated for **genistein** in rats and mice with neonatal exposure to rather high genistein doses. Whilst the hazard profile of genistein resembles that of more potent estrogens, the isoflavone doses that cause adverse effects in rodents are several orders of magnitude higher than those of DES, and higher than those usually ingested with food.

Genistein and **daidzein** (Figure 6.15) are common dietary ingredients. The highest quantities are found in soybeans, soy milk, and other soy protein products (tofu, miso). Human exposure to isoflavones varies with dietary habits: It is relatively low in Westerners on an omnivorous diet (about 0.5 mg/day), and higher in vegetarians and Asians who consume a traditional soy-rich diet (5–50 mg/day). Babies fed with soy-based infant formula have the highest intake, in the range of 4.5–8 mg/kg bw, more than the calculated daily isoflavone intake for adults, which is between 0.1 and 1 mg/kg body weight. Human infants at 4 months fed on soy milk formula have blood levels of genistein and daidzein that are $\geq$5-fold higher than those of adults eating a soy-rich diet.

Whether phytoestrogen exposure results in relevant biological activity in humans, either beneficial or detrimental, is still a contentious issue.

Mycoestrogens The resorcylic acid lactone zearalenone (ZEA, Figure 6.15), and its metabolites α-ZOL and β-ZOL are potent estrogen-mimics, in both *in vitro* assays and *in vivo* in the uterotrophic bioassay. ZEA is formed by *Fusarium* species contaminating cereal crops. Pigs are quite sensitive to ZEA in their diets, showing adverse effects on reproduction with a no-observed-adverse-effect level (NOAEL) of 40 μg/kg bw/day estimated for pubertal pigs, the most sensitive species tested. Human exposure from food consumption is estimated to be 0.03–0.1 μg/kg bw/day, with children having the highest consumption based on body weight. Under unfavorable conditions, ZEA intakes may be up to 20 μg/day, or, for a 20 kg child, about 1 μg/kg bw/day. While still below the experimentally determined NOAEL in prepubertal pigs, this is sufficient reason to establish limit values for ZEA in foods.

Methoxychlor (MXC, Figure 6.15) is considered to be a proestrogen, since demethylation converts the parent to a more active metabolite, 2,2-bis(*p*-hydroxyphenyl)-1,1,1-trichloroethane (HPTE). **HPTE** binds to the estrogen receptor (subtype ERα) and exerts agonist activity in the uterotrophic bioassay. In developmental and reproductive assays, MXC results in effects on the female rodent that are consistent with estrogenic effects (i.e., accelerated vaginal patency). Thus, MXC can be defined as endocrine disruptor.

Table 6.17 *Binding of DDT isomers/metabolites to sex hormone receptors*[a]

Isomer/Metabolite	AR (μM)[b]	ER μM[b]
p,p′-DDT	75	≥1000
o,p′-DDT	95	5
p,p′-DDE	5	≥1000
Hydroxyflutamide[c]	0.5	≥1000
17β-Estradiol	0.5	0.002

[a]Data from Kelce and Wilson (2001).
[b] Concentration of compound which displaces 50% of the natural ligand at the androgen receptor (AR) or the estrogen receptor (ER).
[c] Hydroxyflutamide is a potent synthetic antiandrogen.

HPTE acts also as *antagonist* for the androgen receptor *in vivo*, to decrease steroidogenic enzyme activity (Figure 6.14, B) and to inhibit ($IC_{50} \sim 1$ μM) testosterone biosynthesis in Leydig-cell cultures, apparently as a result of an ER-mediated decreased expression of cholesterol-cleaving enzyme that initiates steroidogenesis (i.e., CYP450*scc*). While MXC is an interesting case from a mechanistic point of view, decreasing use, the lack of persistence, and the available human exposure data (<2 μg per day) speak for limited relevance.

Technical-grade **DDT**, a mixture of *p,p′*-dichlorodiphenyltrichlorethane and about 20% *o,p′*-isomer, is still used in certain areas against insects that transmit malaria. In most countries, DDT has been banned for decades because of the persistence and bioaccumulation of the parent compound and its metabolites (*o,p′*- and *p,p′*-DDE) leading to adverse impact on several species of wildlife. In experimental animals, *o,p′*-isomers of DDT and DDE act clearly estrogenic, showing an affinity for the ER receptor, an estrogenic activity in the uterotrophic bioassay, induction of VTG in fish, and have been observed to accelerate vaginal opening in young rodents. In addition, the *p,p′*-isomers exert antiandrogenic effects (see below and Table 6.17).

The examples of MXC and DDT underline the role of **metabolism**. Moreover, they illustrate that categories such as 'environmental estrogen' or 'antiandrogen' are sometimes simplifications:

> Compounds or metabolites may show **more than one type of hormonal activity**, e.g. the pesticide **methoxychlor** and its metabolite **HPTE**. Also the phytochemicals **genistein** and **8-prenylnaringenin** display both estrogenic and antiandrogenic activity in receptor- and cell-based *in vitro* assays. Yet, most EACs show predominantly one type of hormonal activity.

Bisphenol A (BPA) and **nonylphenol (NP)** are high-production volume chemicals. Both bind the ER with low affinity, are positive in *in vitro* transcriptional assays, induce VTG in fish, and are weakly positive in the uterotrophic bioassay (>50 and >200 mg/kg-bw/d po, respectively) Toxicokinetic studies show that both compounds undergo efficient conjugation by hepatic phase-II enzymes, resulting in metabolites that are rapidly

excreted and are not endocrine active. Definitive multi-generation studies have been conducted in rodents with NP, the related octylphenol (OP), and BPA. The observed NOAEL and LOAEL values for NP in three multi-generational studies are 10 and 50 mg/kg-bw/d, and both kidney lesions and an acceleration in vaginal opening occur at 50 mg/kg-bw/d. The latter is consistent with an estrogenic mechanism of action. For OP and BPA, systemic toxicity, but no evidence of estrogenic activity, was observed in the multi-generational studies. The systemic toxicity (body-weight decreases) NOAELs were 50 mg/kg-bw/d for both compounds.

BPA and 4-NP (Figure 6.15) are found as contaminants in foods and beverages at low ppb levels; human daily intake estimates arrive at about 10 μg for adults and lower amounts for infants. Are such exposures of toxicological relevance for humans? Exposures to BPA (<0.5 μg/kg bw in adults, 1.6 μg/kg bw in infants) and to NP (<0.13 μg/kg bw in adults, 0.3 μg/kg bw in infants) indicate that it is not plausible that internal xenoestrogen concentrations can reach a relevant level for receptor occupancy. Using the Margin of Safety (MOS) concept, the MOS for dietary ingestion of NP would be >40 000 for infants and 80 000 for adults, and the MOS values for BPA are comparable. These very high MOSs argue against a risk for human health. This conclusion is supported by another consideration: Comparing human exposure to BPA and NP to that for phytoestrogens and isoflavones, it appears that the estrogenic load from phytoestrogens is clearly larger, and even under an assumption of additivity, the contribution of NP and BPA is minute.

Antiandrogens

Jost used microsurgical techniques on fetal rabbits in the 1940s to show that the fetal testes are essential for the development of the male reproductive tract. Then in the 1960s, work by Neumann and co-workers with cyproterone acetate showed this compound to be an antiandrogen that could chemically block male reproductive tract development. Thus, it should be appreciated that a positive androgen signal is critical for the development and differentiation of the male reproductive tract in mammals, and that antiandrogens do represent the toxicological threat of irreversible damage. In contrast to estrogens, few nonsteroidal agonists for the androgen receptor are known, while a number of antagonists of nonsteroid structure have been identified. Again, the need for tiered assessment systems and for consideration of metabolism are illustrated with the examples.

The fungicide **vinclozolin** [3-(3,5-dichlorophenyl)-5-methyl-5-vinyloxazolidine-2,4-dione] is hydrolysed to M1 and M2 (3′,5′-dichloro-2-hydroxy-2-methylbut-3-enanilide). Metabolite M2 binds to the AR with an affinity (K_d 9.7 μM) similar that of to *p,p′*-DDE. In *in vitro* transcriptional assays, M2 displays antagonist characteristics, and vinclozolin demonstrates antiandrogenic activity in the Hershberger bioassay. After *in utero* exposure during the period of male reproductive tract differentiation and development, male rats exposed to vinclozolin at doses of 3–100 mg/kg bw/day, display alterations and frank malformations. Interestingly, dose–response curves for different effects of vinclozolin (i.e. shortened anogenital distance at birth, retained nipples, hypospadias, undescended testes, and no sex accessory glands) vary in shape, and also the effective dose values for these endpoints. Similar findings as in Long Evans rats are reported for Wistar rats, with

NOAELs of 1 and 3 mg/kg bw/day, based on the most sensitive parameter. Human exposure to vinclozolin is several orders of magnitude lower than the rodent NOAELs.

Linuron [3-(3,4-dichlorophenyl)-1-methoxy-1-methylurea] and ***p,p′*-DDE** are a herbicide and pesticide metabolite, respectively, that have been shown to act as antiandrogens. Both bind to the AR, both act as antiandrogens in transcriptional assays, and both are antiandrogens in the Hershberger assay at doses $\geq$100 mg/kg-bw/d. Linuron has been demonstrated to result in a spectrum of antiandrogenic effects in male rats after *in utero* exposure at 50 mg/kg-bw/d. In a multi-generational study, dietary levels of 625 ppm linuron (54 mg/kg-bw/d in the females) resulted in testicular and epididymal abnormalities, changes in hormonal profiles, and changes in reproductive organ weights in F1 males. The NOAEL in this study was 8.3 mg/kg-bw/d. However, in a one-year chronic study with dogs, hematological changes were observed with LOELs of 4.17 and 3.49 mg/kg-bw/day for males and females, and the NOEL doses were 0.79 and 0.77 mg/kg-bw/day, respectively. A 100-fold uncertainty (or 'safety') factor was used for these data, resulting in an allowable dose of 7.7 μg/kg-bw/d. Thus, the reproductive effects of linuron were not considered to be the most sensitive, primary effect.

Indirect Interference with the Hormone Systems

Whilst most of the compounds act by direct receptor binding, some may interfere with the sex hormone system by indirect effects.

Phthalate esters are high-production-volume chemicals, used mostly as plasticizers in flexible PVC products, but also for many other applications. Di(2-ethylhexyl) phthalate (**DEHP**, Figure 6.16), the most important member of this class of compounds, has been the subject of several reviews. A comprehensive review compiled by the National Toxicology Program Center for the Evaluation of Risks to Human Reproduction (NTP-CERHR-DEHP-05 at http://cerhr.niehs.nih.gov) covers all aspects of interest, i.e. exposure, kinetics, effects, and mode of action.

Owing to their use in plastics and common consumer products, there is ubiquitous human exposure to phthalates. Concern has focused on early exposure, since DEHP and other phthalate esters have been shown to disrupt male reproductive tract organogenesis when administered to rodents during gestation and/or early postnatal life. This was observed at relatively high dose levels for di-*n*-butyl phthalate, DEHP, butyl benzyl phthalate, and to a lesser extent diisononyl phthalate, whereas dimethyl phthalate, diethyl phthalate, and dioctyl terephthalate were inactive. Effects seen in rodents include reduced anogenital distances, reduced testis weight, and increased rates in other endpoints, e.g. cryptorchidism, hypospadias, epididymal agenesis: The higher the dose, the more profound are the changes and the more parameters are affected. The spectrum of effects points to an antiandrogenic mechanism of phthalate esters. They affect male sexual differentiation without interacting directly with the AR or the ER, but by interfering with testosterone synthesis in the fetal testis in a complex manner (further reading).

Conversion of DEHP into the active metabolite monoethylhexyl hydrogen phthalate (**MEHP**) (Figure 6.16) is an important aspect in its mode of action, as indicated by these observations: *Per os* application is more effective in eliciting effects on testis weight in

rats than is parenteral (*i.v.*) exposure, and there is more conversion of DEHP into MEHP by intestinal lipases with the oral route. Intestinal lipases appear to be at significantly greater levels in rodents than in primates, and marmosets showed no testis effects under conditions that produce testicular toxicity in juvenile rats.

Relevant studies involving exposure of rats to DEHP during gestation have been conducted, two of which are multi-generational studies with several dose levels. These data now provide a good basis for assessing the effect level in experimental animals. The lowest LOAEL identified for effects on testicular tract development was 14–23 mg/kg bw/day, with an NOAEL of 5–8 mg/kg bw/day based on an overall assessment of responses across generations.

Taking human exposure data and other relevant information into account, the CERHR report arrives at a differentiated view with regard to concerns about DEHP exposure in different groups of the human population. Exposure to phthalates is widespread and occurs with foods, drugs, medical devices (tubings, catheters), personal care products and indoor air. The estimated median DEHP intake of the general population (the largest source of which is dietary) varies by age group, between 8 μg/kg bw/day for adults up to 26 μg/kg bw/day for toddlers. Infants undergoing multiple medical procedures have parenteral exposures orders of magnitude higher than that: Recent biomonitoring studies indicate exposures up to about 6 mg/kg bw/day. Since such exposures approach toxic doses in rodents, there is 'serious concern' that multiple medical treatment of children may adversely affect male reproductive tract development. Although it is recognized that benefits of medical treatments can outweigh any risks, minimizing exposure to DEHP is now a goal. On the other hand, there is 'minimal concern' that current exposure (1–30 μg/kg bw/day) adversely affects adult human reproduction, since adult rodents require 1–2 orders of magnitude higher doses than juvenile rodents to produce any effects. For infants less than one year old, there is 'some concern' that exposure can adversely impact reproductive development, if the level of DEHP exposure is at the high end of the estimated range, because of their greater susceptibility and uncertainties regarding exposure; further concern is directed to phthalates shown to act through the same mode of action as DEHP.

Tributyltin (TBT, Figure 6.16), an organotin compound, was used in antifouling paints. Owing to its persistence, it has been banned for pleasure and smaller commercial vessels, and is now scheduled for phase-out for use on large, ocean-going vessels. TBT was found to have devastated oyster colonies along the French coast of the Bay of Biscay in the 1970s. Subsequently, an impact on many marine molluscs, particularly snails, has been demonstrated. The phenotypic condition in snails is the masculinization of female snails, a condition termed 'imposex'. As TBT is persistent in marine sediments and tends to bioaccumulate in aquatic organisms, the environmental effects outlive the antifouling use. The apparent mode of action is the inhibition of the aromatase enzyme. The enzyme (CYP450*arom*) converts androgens into estrogens, i.e. androstenedione into estrone or testosterone into estradiol (Figure 6.17). Selective aromatase inhibitors are capable of inducing imposex in neogastropods, but not as severely as TBT, which may suggest the involvement of additional mechanisms besides aromatase inhibition.

TBT also shows a complex toxicity profile in rodents, with reproductive and developmental toxicity observed at doses around 1 mg/kg-bw/d. Yet, the critical endpoint for risk assessment is considered to be immunotoxicity, with an NOAEL as low as 0.025 mg/kg

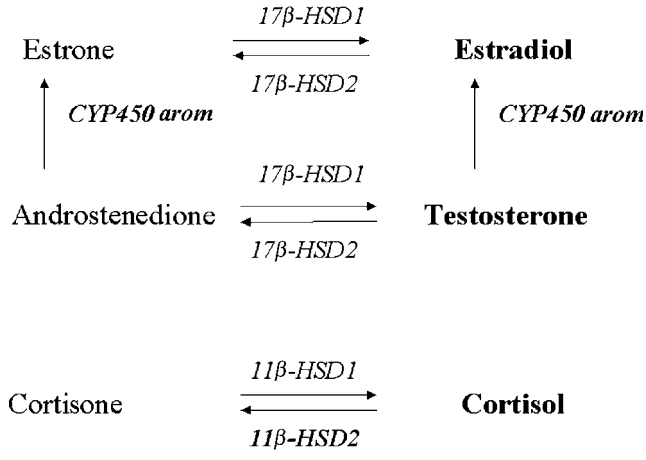


Figure 6.17 *Endocrine modulation via inhibition of steroid metabolizing enzymes: Some compounds acting as inhibitors of aromatase or hydroxysteroid dehydrogenases are discussed in the text (Section 6.4.4). Bold text indicates enzymes that are more active than the other(s) in a particular modulation, and those steriods that are more biologically active than their immediate precursors.*

bw/day identified in chronic feeding studies for TBT oxide. TBT can inhibit 11β-hydroxysteroid dehydrogenase type 2 (11β-HSD2) which converts cortisol into the inactive 11-ketosteroid, thereby raising glucocorticoid levels (Figure 6.17). Both organotins and elevated glucocorticoids cause low birth weight and thymus involution in animal models, reinforcing the association.

Health risks associated with consumer exposure to organotins have been assessed recently by the European Food Standards Agency (EFSA) (http://www.efsa.eu.int). Based on immunotoxicity in rodents, a group tolerable daily intake (TDI) of 0.25 μg/kg-bw/d was established for TBT and related organotin compounds, by applying a safety factor of 100 and considering the immunotoxic effects of these compounds as additive. Intake calculations based on high seafood consumption (as in Norway) showed that the combined organotin intake represents up to 70% of the group TDI. The Panel noted that the consumption of fish, mussels, and other marine animals from highly contaminated areas, such as the vicinity of harbors and heavily used shipping routes, may lead to organotin intake that exceeds the group TDI. While it is unlikely that dietary exposures would put humans at high risk, the example of TBT illustrates that consumers will benefit from restrictions taken to protect marine species.

Limited Evidence for Endocrine Disruption A number of compounds in Table 6.16 have been suggested to be endocrine disruptors based upon limited evidence or uncertain interpretations of data. Examples include lindane, mirex, heptachlor, and toxaphene. To date, these compounds do not have the structural alerts for well known modes of action such as estrogen- and androgen-receptor binding and have not been positive in well performed *in vitro* and *in vivo* screens. This points to the need for validated regulatory assays, a tiered framework for a battery of assays, and the need for a weight-of-evidence

approach to arrive at conclusions that substances are potential endocrine disruptors. Subsequently, there is the need for a clearly defined set of endpoints in definitive assays that would lead to a clear conclusion that adverse effects have occurred through an endocrine mode of action.

Other compounds present additional difficulties. The polychlorinated and polybrominated biphenyls, the polybrominated diphenyl ethers, and the polychlorinated dibenzodioxins and furans are complex chemical isomer families. Several isomers are potent binders of the Ah receptor, which appears to cross-talk with the ER, so that the compounds may have antiestrogenic properties. Other compounds may displace thyroid hormones from circulating serum carrier proteins, thereby accelerating metabolism. In these cases, the action is indirect, and the mode of action and relationship to adverse effects are difficult to establish with sufficient toxicological certainty to establish a weight of evidence. As a result, weight-of-the-evidence reviews by the International Program for Chemical Safety have concluded that the data are insufficient to say these compounds fulfil the Weybridge definitions for being endocrine disruptors.

6.4.5 Summary

Chemicals of different origin which act to mimic estrogens or act as antiandrogens may be detrimental to reproduction and development in humans and animals. The evidence for a causal link between exposures to such agents and adverse health effects in humans is limited to prenatal exposures to pharmacological doses of the potent drug diethylstilbestrol. Endocrine disruption in wildlife observed in areas with high levels of pollution has been linked to persistent and bioaccumulating chemicals such as DDT and TBT.

The question as to what extent xenoestrogens and xenoantiandrogens can indeed exert adverse effects on humans remains somewhat controversial. Some uncertainties remain regarding the role of combination effects, the existence of practical thresholds, and the complex regulation of the endocrine system. However, it is recognized now that i) risk assessments for synthetic chemicals with hormonal activity should take into account also naturally occurring compounds which may act as endocrine modulators; ii) adverse effects of xenoestrogens and xenoantiandrogens are related to time of exposure, dose, and potency of the compound in question, regardless of its mode of action (direct receptor binding or indirect effects on the hormone system) and origin.

A toxicological evaluation of the hazards, mode of actions, and risks from endocrine-active compounds is feasible within the existing regulatory framework. As illustrated by several examples (Section 6.4.4), this requires data from animal studies on dose–effect relationships for relevant endocrine endpoints and reliable information on human exposure, as well as data on the kinetics to aid in route-to-route and species extrapolation. The data base for judgments on the biological relevance of human exposure to endocrine-active chemicals with foods and other consumer products has improved: Margins of safety assessed case by case (e.g. bisphenol A, zearalenone) or when appropriate by a group approach (e.g. phthalate esters) cover a wide range, thereby allowing a focus on ‘suspects’ that deserve further investigation.

The more difficult challenge may be conducting a weight of evidence review that ascertains that a substance can indeed act in the intact animal by an endocrine mode of

action (potential EDC), that the endocrine-related effects are observed in a definitive tier test, and finally that the endocrine-related effects are the primary effects that are causally related to the mode of action (a true EDC).

References

Damstra, T., Barlow, S., Bergman, A., Kavlock, R., and Van Der Kraak, G. (eds.) (2002) Global Assessment of the State-of-the-Science of Endocrine Disruptors. International Programme on Chemical Safety; WHO/IPCS/EDC/02.2; http://ehp.niehs.gov/who/

Golden, R.J., Nollet, K.L., Titus-Ernsthoff, L., Kaufman, R.H., Mittendorf, R., Stillman, R., and Reese, E. A. (1998) Environmental Endocrine Modulators and Human Health: an Assessment of the Biological Evidence. *Crit. Rev. Toxicol*., **28**, 109–227.

Hill, A.B. (1965). The Environment and Disease: Association or Causation? *Proc. R. Soc. Med*., **58**, 295–300.

Kelce, W.R., and Wilson, E. (2001) Antiandrogenic Effects of Environmental Endocrine Disruptors. In: Metzler, M. (ed.) Endocrine Disruptors (Part I). Springer-Verlag, Berlin, Heidelberg, pp. 40–61.

Newbold, R. (2002) Effects of Perinatal Estrogen Exposure on Fertility and Cancer in Mice. In: Metzler, M. (ed.) Endocrine Disruptors (Part II). Springer-Verlag, Berlin, Heidelberg, pp. 171–186.

Palmer, J.R., Wise, L.A., Hatch, E.E., Troisi, R., Titus-Ernsthoff, L., Strohsnitter, W.C., Kaufman, R., Adam, E., Noller, K., Herbst, A., Robboy, S., Hartge, P., and Huwer, R.N. (2006) Prenatal Diethylstilbestrol Exposure and Risk of Breast Cancer. *Cancer Epidemiol. Biomarkers Prev*., **15**, 1509–1514.

For Further Reading

Bolt, H.M., Janning, P., Michna, H., and Degen, G.H. (2001) Comparative Assessment of Endocrine Modulators with Oestrogenic Activity. I. Definition of a Hygiene-based Margin of Safety (HBMOS) for Xenooestrogens against the Background of European Developments. *Arch. Toxicol.*, **74**, 649–662.

Committee on Toxicity of Chemicals in Food/Consumer Products and the Environment. COT Report – Phytoestrogens and Health. London: Food Standards Agency (2003); available at http://www.food.gov.uk/multimedia/webpage/phytoreportworddocs

David, R.M. (2006) Proposed Mode of Action for in utero Effects of Some Phthalate Esters on the Developing Male Reproductive Tract. *Toxicol. Pathol*., **34**, 209–219.

Degen, G.H. (2004) Endokrine Disruptoren in Lebensmitteln. *Bundesgesundheitsblatt*, **47**, 848–856.

European Commission, Scientific Committee on Food. Opinion of the Scientific Committee on Food on Bisphenol A. SCF/CS/PM3936 Final 3 May 2002; EC, Brussel, Belgium: http://europa.eu.int/comm/food/fS/sc/scf/index_en.html

European Food Standards Agency (2004) Opinion of the Scientific Panel on Contaminants in the Food Chain on a request from the Commission related to Zearalenone as undesirable substance in animal feed. *EFSA J.*, **89**, 1–35, available at http://www.efsa.eu.int/

Kennel, P.F., Pallen, C.T., and Bars, R.G. (2004) Evaluation of the Rodent Hershberger Assay Using Three Reference Endocrine Disruptors (Androgens and Antiandrogens). *Reprod. Toxicol*., **18**, 63–73.

Mueller, S.O. (2004) Xenoestrogens: Mechanisms of Action and Detection Methods. *Anal. Bioanal. Chem*., **378**, 582–587.

National Toxicology Program (NTP), Center for the Evaluation of Risks to Human Reproduction (CERHR) (2005). NTP-CERHR Expert Panel Update on the Reproductive and Developmental Toxicity of di(2-ethylhexyl) phthalate; available at http://cerhr.niehs.nih.gov/reports/index.html

Owens, J.W. and Koeter, H.B.W.M. (2003) The OECD Program to Validate the Rat Uterotrophic Bioassay: an Overview. *Environ. Health Perspect*., **111**, 1527–1529.

Sharpe, R.M. and Irvine, D.S. (2004) How Strong is the Evidence of a Link between Environmental Chemicals and Adverse Effects on Human Reproductive Health? *Br. Med. J*., **328**, 447–451.

Tinwell, H. and Ashby, J. (2004) Sensitivity of the Immature Rat Uterotrophic Assay to Mixtures of Estrogens. *Environ. Health Perspect*., **112**, 575–582.

Vos, J.G., Dybing, E., Greim, H.A., Ladefoged, O., Lambre, C., Tarazona, J.V., Brandt, I., and Vethaak, A.D. (2000) Health Effects of Endocrine-disrupting Chemicals on Wildlife, with Special Reference to the European Situation. *Crit. Rev. Toxicol*., **30**, 71–133.

Abbreviations

ADI, allowable daily intake; ADME, Absorption, Distribution, Metabolism, Excretion; AR, androgen receptor; BPA, bisphenol A; DDT, 2,2-bis(chlorophenyl)-1,1,1-trichloroethane; DEHP, di(2-ethylhexyl) phthalate; DES, diethylstilbestrol; EAC, endocrine-active compound; EE, ethinylestradiol; ER, estrogen receptor; HBMOS, Hygiene-based margin of safety; HPTE, 2,2-Bis(p-hydroxyphenyl)-1,1,1-trichloroethane; LOAEL, lowest observed (adverse) effect level; MEHP, monoethylhexyl hydrogen phthalate; MOS, margin of safety; MTD, maximum tolerated dose; MXC, methoxychlor; NO(A)EL, no observed (adverse) effect level; NP, 4-nonylphenol (branched); NTP, National Toxicology Program; OECD, Organisation for Economic Co-operation and Development; OP, 4-*t*-octylphenol; PBDE, polybrominated diphenyl ether; PCB, polychlorinated biphenyls; SHBG, steroid hormone-binding globulin; POP, persistent organic pollutants; TBT, tributyltin; TDI, tolerable daily intake; VTG, vitellogenin; ZEA, zearalenone.

6.5 Toxicology of Solvents

Wolfgang Dekant and Marion W. Anders

6.5.1 Introduction

Central nervous system depression, one form of which is general anesthesia, is an effect shared by many solvents after acute high-dose exposure. Chronic, repeated exposure to solvents may induce a wide variety of toxic effects characteristic for each chemical and not necessarily shared by other solvents. While the acute effects of solvent vapors depend on the accumulation of the parent compound in lipid-rich tissues of the central nervous system, the toxicities seen after chronic exposure are frequently the result of biotransformation into reactive or toxic metabolites.

Organic solvents form a very important group of industrial chemicals. They are used in a wide range of occupational settings. The general public may also be exposed to solvents in consumer products or solvent residues present in areas of environmental pollution. Fugitive vapors readily pollute ambient air in the general environment because they are volatile and are readily released into the atmosphere.

Commonly used solvents may be single compounds such as dichloromethane or toluene. They may also be simple mixtures of individual compounds containing a few chemicals or may be highly complex mixtures, such as those derived from petroleum, which may contain hundreds of compounds. Among the most frequent industrial uses for solvents are the degreasing of metal products, dry cleaning of textiles, and chemical syntheses. Furthermore, gasoline (petrol), which is perhaps the highest volume mixture of chemicals produced by industry, is often misused as a solvent. The major requirements for solvent applications are a high capacity for dissolving lipid-soluble materials and sufficient volatility to permit simple removal of the solvent. The selection of a solvent for a specific purpose is influenced by safety considerations such as flammability or the likelihood of formation of explosive mixtures with air. Toxicological considerations and the kinetics of degradation of the solvent in the environment are key considerations in selecting an appropriate solvent, assuring worker safety, and selecting the best process for dealing with residual waste solvents.

Inhalation is the most important pathway of human exposure to solvents because of their high volatility. However, skin contact with highly lipophilic solvents can result in transdermal uptake and rapid distribution of the chemical in the organism. In general, for all lipophilic solvents, four different types of toxic effects can be distinguished:

- After acute exposure to very high concentrations by inhalation, pulmonary uptake of solvents and distribution in the organism is very rapid, resulting in high solvent concentrations in the central nervous system, which can result in narcosis. Indeed, some previously widely used solvents such as chloroform and diethyl ether have been used as general anesthetics and the currently used anesthetic chlorofluoroalkanes or fluorinated ethers show typical solvent effects. The typical signs of solvent-induced narcosis are disorientation, euphoria, giddiness, paralysis, convulsions, and death from respiratory or cardiovascular arrest. After removal from exposure and elimination of the solvent, the effects, short of lethality, are reversible.
- Occupational exposures above currently acceptable exposure standards may cause nonspecific reactions in the central nervous system, which may be detected in behavioral toxicity tests. These include paresthesia, visual or auditory deficits, loss of memory, irritability, compulsive behavior, lack of coordination, and fatigue. Dose–response and causal relationships for these effects have been difficult to study due to the absence of appropriate animal models
- Chronic exposure to certain solvents may reveal specific toxicity in various organs or organ systems. The extent of toxicity and target organs affected differ between individual solvents and usually depend upon biotransformation into toxic metabolites. The solvent itself or its metabolites may concentrate in the target organs. For some solvents, biotransformation-dependent toxicities may also be observed after single high-dose exposures. The extent of these toxicities also depends on the kinetics of bioactivation and the chemical reactivity of the metabolites.

- After dermal contact, many solvents cause a rapid defatting of skin, resulting in irritation, cellular hyperplasia, swelling, and other skin damage. Exposure of the skin to solvents can cause an increase of the penetration rate of other compounds. Direct eye contact to some dilute or concentrated solvents may cause severe eye damage.

In the occupational setting workers may be exposed to solvent mixtures composed of many individual components at widely differing concentrations. The toxicological effects of any chemicals in the mixture may be exaggerated or diminished because of the presence of other compounds. Toxicological interactions may result from the simultaneous attack of two or more chemicals on a receptor. However, in many cases the toxicological effect is caused by metabolites rather than the original component of the mixture. Changes in metabolism due to enzyme induction or inhibition of biotransformation of one of the toxicologically relevant components in the mixture may lead to changes in observed toxic effects. However, because there may be many chemicals in a mixture the potential interactions are complex and our understanding of the basic mechanisms involved is still limited.

The toxic effects of exposure to many industrial solvents has been recognized for more than a century. Consequently, the toxicology of many solvents has been more thoroughly studied than that of chemicals used for other purposes. In support of safety concerns aimed at both the workforce and the public at large exposure limits have been defined. In the United States permissible occupational exposure limits (PELs) are established by the Occupational Safety and Health Administration of the Department of Labor (OSHA), and Threshold Limit Values (TLVs) by the nongovernmental American Conference of Governmental Industrial Hygienists (ACGIH). These controls are aimed at healthy workers and relate to exposure five days per week, 8 hours per day. They also establish Short-Term Exposure Limits (STEL) in which exposure should not exceed 15 minutes. For the general public Acute Exposure Guideline Levels (AEGLs) for acute exposures ranging from 10 minutes to 8 hours are determined by the AEGL Committee, which includes international representation and is managed by the United States Environmental Protection Agency (USEPA). The products of the Committee's recommendations are reviewed by the National Academies' Committee on AEGL and ultimately published as National Academy documents. Long-term exposure limits for the general public have been recommended by the USEPA under its Integrated Risk Information System (IRIS). For solvents, USEPA has established a program for determining Reference Concentrations (RfCs), i.e. air concentrations of noncarcinogenic chemicals, which are estimates (with uncertainty spanning an order of magnitude or greater) of daily inhalation exposures that are unlikely to have deleterious effects during a human lifetime. [There is a similar program for nonvolatile chemicals for which Reference Doses (RfD) are determined.] Toxicological Profiles, which contain summaries of the toxicology of chemicals and recommended exposure standards, are issued by the Agency for Toxic Substances and Disease Registry (ATSDR). Although these may be the standard setting examples of highest profile, there are many other agencies, both in the United Sates and other countries, where exposure standards for solvents continue to be established.

6.5.2 Toxicology of Selected Solvents

Halogenated aliphatics were widely used as solvents. The toxic responses observed include narcosis after single-inhalation exposure to high concentrations. Major target organs after repeated exposures are the liver and the kidney. The use of many of these compounds has been reduced or discontinued due to possible toxic effects or potential carcinogenicity.

Halogenated Aliphatic Hydrocarbons

The introduction of halogen atoms into both saturated and unsaturated aliphatic hydrocarbons increases their chemical stability and improves their solvent capacities for fat soluble materials. Chloro- and fluoroalkanes and alkenes are used extensively as solvents because of their excellent solvent properties, their low flammability, and the absence of formation of explosive mixtures with air. These compounds display high volatility, which often leads to their release during use. Because they are persistent in the environment, human exposure is likely and has prompted intensive study of their toxicology.

Trichloro- and tetrachloroethylene Owing to their low acute toxicity and lack of flammability, both tri- and tetrachloroethylene were widely used solvents. The main use of trichloroethylene was degreasing and tetrachloroethylene is still used for dry cleaning of textiles. Owing to their stability in the atmosphere and in ground water, and the availability of highly sensitive and simple analytical methods for determining their concentration, a wide distribution of these compounds in the environment has been demonstrated.

Toxicity After a single exposure to high doses of both tri- and tetrachloroethylene, narcosis is the typical effect. Other organ systems are only rarely affected in humans and the cause of death usually is respiratory depression. Trichloroethylene was widely used as a general anesthetic in the United Kingdom into the 1950s. Although trichloroethylene itself was benign, the decomposition of trichloroethylene to highly toxic dichloroacetylene in closed-circuit anesthesia equipment led to its discontinuation. Skin contact to both tri- and tetrachloroethylene results in defatting of the skin, often without further symptoms.

Repeated exposure to trichloroethylene under occupational settings with higher air concentrations may induce unspecific behavioral changes. Long-term administration of trichloroethylene or tetrachloroethylene to rodents has caused an increased incidence of liver tumors in mice and a very small increase in the incidence of kidney tumors in male rats only. Liver tumor formation by trichloroethylene and tetrachloroethylene is due to peroxisome proliferation (induced by the stable metabolite trichloroacetic acid), a phenomenon which is not considered relevant for human risk assessment. Some epidemiology studies indicate that, after prolonged exposure to very high trichloroethylene concentrations, an increased incidence of renal tumors in humans may be present; however, larger-cohort studies with lower exposures do not show increased risks for this

tumor type. Neither trichloroethylene nor tetrachloroethylene are genotoxic to mammalian cells nor are mutagenic in bacteria.

Toxicokinetics and biotransformation After inhalation exposure, uptake of tri- and tetrachloroethylene is rapid and both compounds may accumulate in adipose tissue after repeated exposure to high concentrations. Elimination of both compounds may occur by exhalation or by biotransformation to polar metabolites, which are almost exclusively excreted in urine. The biotransformation of both compounds in both experimental animals and in humans is highly dose-dependent. A comparison of the metabolism of trichloroethylene with tetrachloroethylene reveals that trichloroethylene is extensively metabolized, even after high doses have been administered, and much of the dose can be recovered as metabolites in the urine. In contrast, tetrachloroethylene is metabolized at low rates and very little of the dose can be accounted for as urinary metabolites.

The major pathway for biotransformation of both halogenoolefins is oxidation by cytochrome P-450 (CYP). CYP 2E1 is a major contributor to trichloroethylene oxidation in rodents and humans leading to the oxidation of trichloroethylene to chloral, and of tetrachloroethylene to trichloroacetic acid, in several steps. The major excreted metabolites of trichloroethylene are trichloroethanol and trichloroacetic acid. They are formed from chloral hydrate by reductive oxidation (Figure 6.18). Trichloroethanol is rapidly eliminated from the organism whereas elimination of trichloroacetic acid is slow due to binding to plasma proteins.

Both tri- and tetrachloroethylene are also transformed to a very small extent into glutathione conjugates that are nephrotoxic, and the formation of these conjugates has been implicated as a mechanism to explain the renal tumors seen after very-high-dose exposures to tri- and tetrachloroethylene in male rats. However, trichloroethanol, formed as a CYP-dependent metabolite from trichloroethylene, may also cause kidney damage after long-term exposure and could thus be responsible for tumor induction in this organ.

Chlorinated Methanes Chloromethanes are still used as solvents due to their high capacity for dissolving lipoid materials, their volatility, and economic considerations. The toxicological profiles of the individual chloromethanes are different and depend on the extent of biotransformation and the reactivity of the intermediates formed.

Tetrachloromethane (carbon tetrachloride) Tetrachloromethane was widely used as an inexpensive solvent in the past but has been discontinued due to its toxicity. Tetrachloromethane is also formed as a by-product in drinking-water chlorination; concentrations as high as 1 ppm have been reported in chlorinated drinking water.

Toxicity Intoxication with tetrachloromethane causes characteristic liver damage; narcosis is seen early only after high exposure or may not be present at all. The most prominent indication of liver damage after tetrachloromethane exposure is a massive increase in serum enzymes indicative of liver damage which may result in lethality. After repeated inhalation, the liver is the most prominent target organ; in sensitive species, liver damage is observed after long-term inhalation of concentrations of tetrachloromethane as low as 10 ppm. Long-term administration by inhalation causes liver tumors in rats; these tumors are formed as a consequence of the massive liver damage.

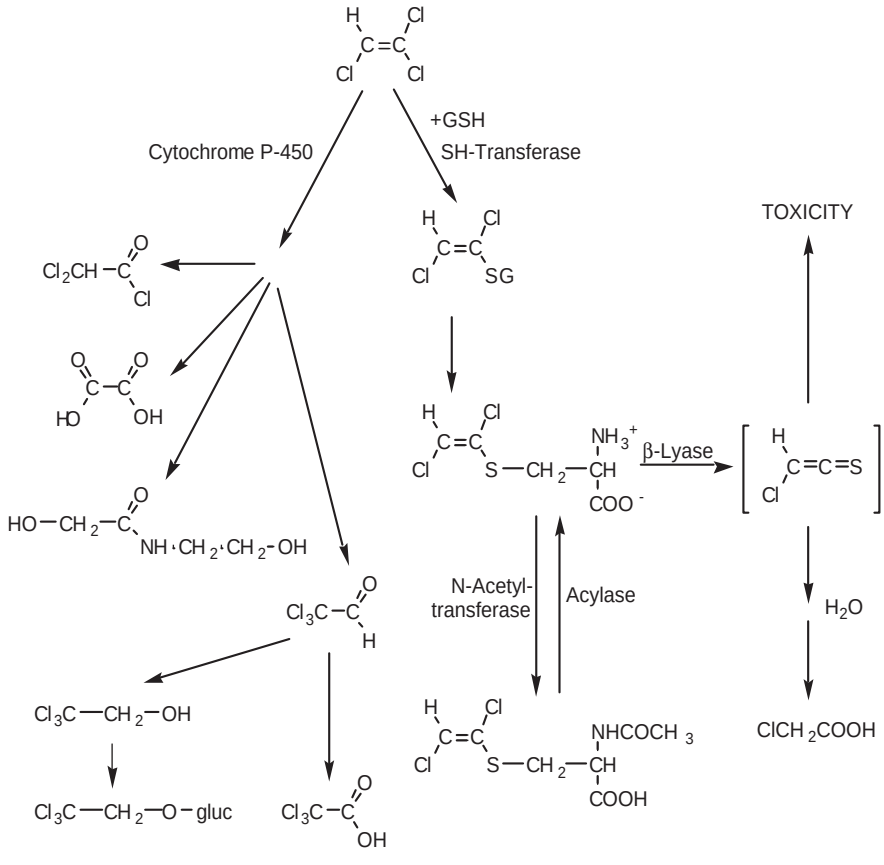


Figure 6.18 *Biotransformation of trichloroethylene by CYP and glutathione S-transferase in rodents and humans.*

Besides liver damage, exposure to very high doses of tetrachloromethane may induce toxic effects in the proximal tubules of the kidney, and renal failure has been described as a cause of death after some tetrachloromethane intoxications.

Toxicokinetics and biotransformation Owing to its lipophilicity, tetrachloromethane is rapidly absorbed after inhalation or ingestion; little is eliminated by exhalation of the parent compound. The toxicity of tetrachloromethane is due to a metabolic reduction catalysed by CYP. This reaction gives the trichloromethyl radical, which abstracts hydrogen atoms from polyunsaturated fatty acids of lipid membranes and, thereby, initiates lipid peroxidation (Figure 6.19). The trichloromethyl radical may also react with oxygen to give trichloromethylperoxyl radical, which may also cause lipid peroxidation and 'oxidative stress.' End products of the biotransformation of tetrachloromethane in rodents are carbon dioxide, hexachloroethane (which is formed by dimerization of the trichloromethyl radical) and chloroform. Abuse of alcohol increases the liver toxicity of

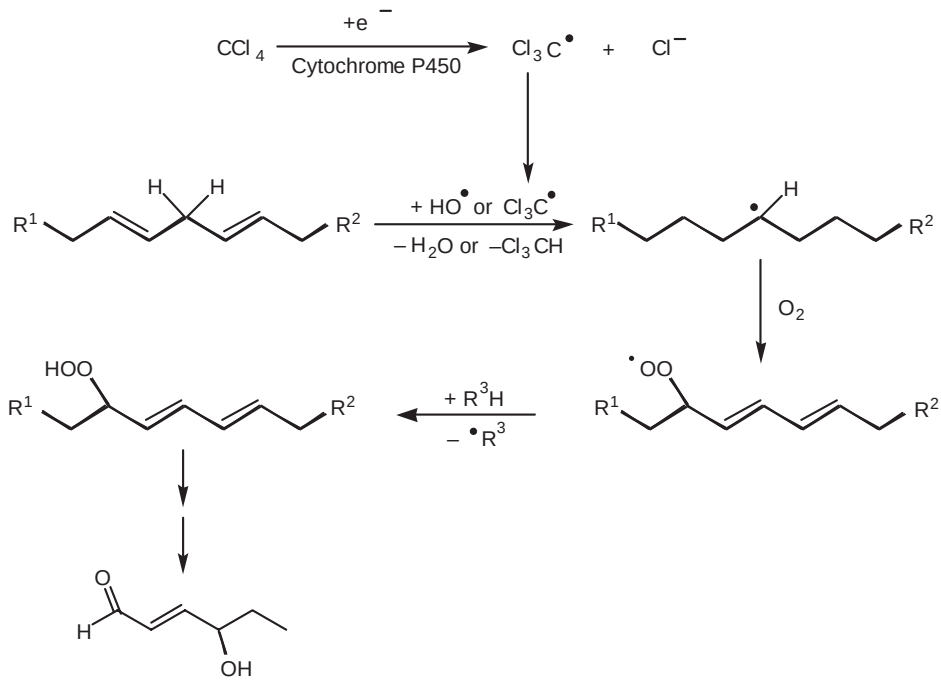


Figure 6.19 *Initiation of lipid peroxidation by reductive biotransformation of tetrachloromethane and interactions of the radicals formed with fatty acid in lipid membranes.*

tetrachloromethane due to induction of CYP 2E1, which catalyses formation of the trichloromethyl radical from tetrachloromethane.

Trichloromethane (chloroform) Trichloromethane had a long history of use as an inhalation anesthetic and was also used as a technical solvent. The use of chloroform for general anesthesia has caused a number of complications such as liver damage and death and was, therefore, discontinued early in the 20th century. At present, chloroform is a relevant by-product formed by drinking-water chlorination.

Toxicity Narcosis may be induced after single exposure to concentrations of chloroform above 1000 ppm for a few minutes. Higher concentrations may cause death by respiratory arrest. After long-term oral treatment, chloroform caused liver and kidney toxicity in rats and mice and increased the incidence of liver and kidney tumors.

Tumor induction in the liver was seen only after dosing regimens that caused massive liver toxicity. Doses that did not cause toxicity did not lead to the development of liver tumors. Therefore, liver tumor induction by chloroform is not relevant to the assessment of potential human tumor risks arising from the consumption of chlorinated drinking water because of much lower chloroform doses, which do not cause liver toxicity.

Toxicokinetics and biotransformation Owing to its lipophilicity, chloroform is rapidly absorbed both after inhalation and ingestion and is also rapidly metabolized.

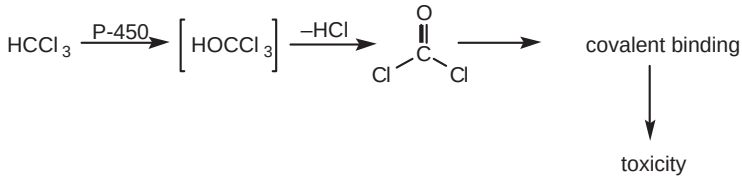


Figure 6.20 *Biotransformation of chloroform by CYP to phosgene.*

Chloroform-induced liver and kidney damage depends on biotransformation of chloroform to phosgene, a reactive metabolite formed by CYP (Figure 6.20) which may bind to cellular macromolecules and, thus, initiate cytotoxicity. Carbon dioxide, formed by hydrolysis of phosgene, is the major metabolite exhaled after chloroform exposure.

Dichloromethane Owing to its favorable properties, such as high volatility, low acute toxicity and high fat solubility, dichloromethane has been frequently used as a solvent for such uses as extracting caffeine from coffee and in typewriter correction fluid.

Toxicity In contrast to tetrachloromethane or chloroform, dichloromethane displays a low potential for toxicity after a single exposure. High concentrations cause narcosis and formation of carboxyhemoglobin as the major toxic effects. Liver damage or toxicity to other target organs is usually not present. Long-term inhalation of dichloromethane in mice caused liver and lung tumors; these tumors were not seen in rats and hamsters after identical exposure conditions or after administration of the same calculated doses to mice in drinking water. Dichloromethane is a bacterial mutagen, but does not show genotoxic effects in mammalian cells.

Toxicokinetics and biotransformation Dichloromethane is rapidly absorbed both after inhalation and oral administration. Parts of the absorbed dose may be eliminated by exhalation and the rapid clearance by exhalation and biotransformation prevents accumulation of dichloromethane in lipid-rich tissues. Biotransformation of dichloromethane occurs by two different pathways; stable products excreted are carbon monoxide and carbon dioxide. Carbon monoxide is formed by a CYP-catalysed oxidation of dichloromethane with formyl chloride as an intermediate (Figure 6.21).

Carbon monoxide formed from dichloromethane binds to hemoglobin; however, owing to the rapid clearance of dichloromethane by exhalation, the concentrations of carboxyhemoglobin in blood usually remain below 50% of total hemoglobin. After inhalation exposure to high dichloromethane concentrations, the CYP-catalysed formation of carbon monoxide becomes saturated and dichloromethane is partly metabolized to carbon dioxide in a glutathione-dependent reaction. Catalysis by glutathione S-transferases forms *S*-(chloromethyl)glutathione from dichloromethane. The glutathione conjugate is rapidly hydrolysed to give formaldehyde, and carbon dioxide is then produced by the sequential oxidation of formaldehyde and formic acid. The glutathione-dependent biotransformation reaction is thought to be the bioactivation pathway responsible for tumor induction by dichloromethane in mice and the electrophilic *S*-(chloromethyl)glutathione is thought to be the ultimate toxic metabolite.

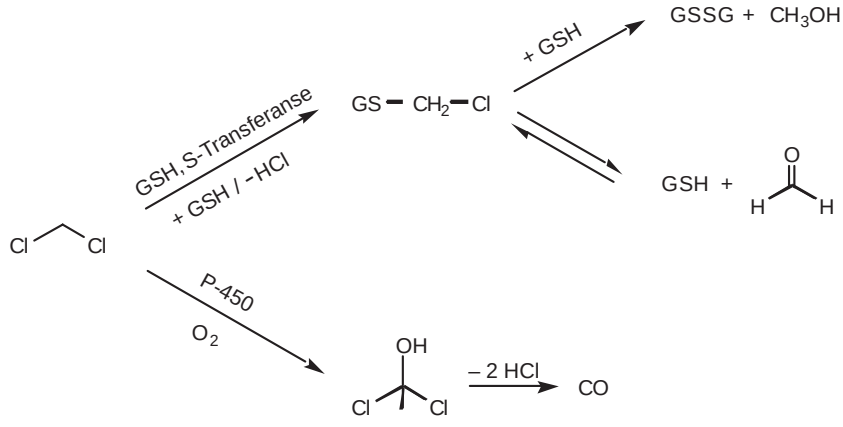


Figure 6.21 *Biotransformation of dichloromethane by CYP and by glutathione conjugation. The glutathione-conjugation pathway forms an electrophilic sulfur conjugate.*

Hydrocarbons

Aliphatic hydrocarbons, except for n-hexane, have a low potential for toxicity. Benzene has a specific toxicity profile with hematologic effects and induces leukemia in humans after occupational exposures.

Owing to their ready availability from crude oil fractions by repeated distillation, both aliphatic and aromatic hydrocarbons are inexpensive and widely used as solvents with a high capacity for the solution of lipids. Furthermore they display low chemical reactivity. Some of the compounds used for technical purposes are highly purified, whereas others are used as a mixture of isomers (e.g., hexane fraction) isolated from crude oil.

n-Hexane Owing to its favorable solvent properties, n-hexane was widely used as a solvent in paints and for purposes of extraction.

Toxicity Short-term exposure to n-hexane in concentrations up to 1000 ppm does not result in overt toxic effects; narcosis is seen only after inhalation exposure to very high concentrations ($LC_{50} = 96000$ ppm, 1 h, rat). n-Hexane is also irritating to the skin upon dermal exposure.

Repeated exposure to n-hexane under occupational conditions in the past has caused polyneuropathy in workers. The neurotoxicity of n-hexane in humans is described as a progressive motor or sensorimotor neuropathy. Symptoms after long-term occupational exposure are numbness and paresthesis in the extremities, most pronounced in the feet or hands. After cessation of exposure, symptoms may regress and minor cases may completely recover. In rats, exposure to n-hexane produces degeneration of the axon and the associated myelin sheath, leading to swelling of spinal tracts, accumulation of axonal neurofilaments, and inactivation of nerve fibers. In addition, testicular lesions are observed in rats, although they are not well documented in humans, after long-term

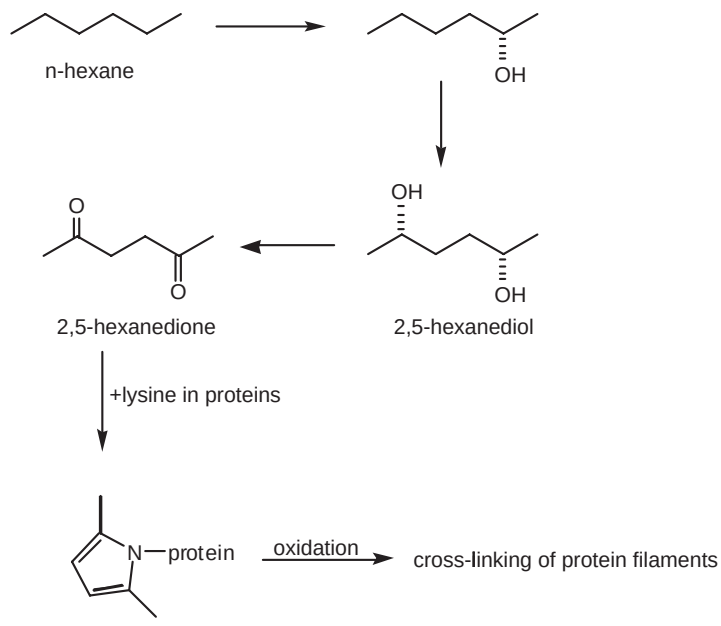


Figure 6.22 *Biotransformation of n-hexane by CYP, formation of lysine adducts, and autoxidation of protein pyrroles to form protein–protein cross-links.*

n-hexane exposure. The n-hexane metabolites 2-hexanone and 2,5-hexanedione cause nervous-system toxicity identical to that of n-hexane.

Toxicokinetics and biotransformation n-Hexane is rapidly absorbed in laboratory animals and humans and has an affinity for lipid-rich tissues. The neurotoxicity of n-hexane is dependent on biotransformation and n-hexane is metabolized to several isomers of hexanol and hexanediol (Figure 6.22). The pathway important for toxicity is the formation of 2,5-hexanediol with further oxidation to 2,5-hexanedione. 2,5-Hexanedione may react with lysine amine moieties to form pyrroles, which are easily oxidized, thus resulting in cross-linking of neurofilaments. These cross-links may cause difficulties in protein transport through narrow regions of the axon and an accumulation of proteins at the site of constriction, resulting in axonopathy.

The toxicity of n-hexane may be modified in the presence of other hydrocarbons due to competition for CYP. For example, concurrent exposure to toluene significantly reduces the neurotoxicity of n-hexane.

Benzene Several billion gallons of benzene are produced in the United States annually. Benzene has been used for many years as a solvent for paints, rubber, resins, and dyes. The use of benzene as a solvent is no longer justified. At present, benzene is present in unleaded gasoline in concentrations up to 2%. Human exposure to benzene, other than occupational exposures, may occur in the environment due to release of benzene present in gasoline. Indoor benzene exposure is often the result of the escape of benzene vapors from automobile gas tanks in attached garages. For smokers, smoking is the most important source of benzene exposure. Some benzene is also produced by the frying of food.

Toxicity Because of its high lipid solubility, acute exposures to benzene may cause narcosis with headache, dizziness, nausea, and vomiting. Exposure to high concentrations may also cause convulsions, ventricular irregularities, and respiratory failure, which may be the cause of death.

Chronic exposure to benzene has long been known to result in bone marrow depression characterized by decreases in circulating erythrocytes, leucocytes, and/or thrombocytes, otherwise termed anemia, leucopenia, and thrombocytopenia. Coincidental decreases in all three cell types is termed pancytopenia and usually is the result of pathologically altered bone marrow incapable of providing viable blood cells, i.e. a state of aplastic anemia. During chronic exposure to benzene the marrow often reaches a state of myelodysplasia, which most frequently proceeds to acute myeloid leukemia (AML).

Study of the mechanism of benzene-induced bone marrow damage has been hampered by the lack of animal models that mimic the human disease. Whereas bone marrow depression induced by exposure to benzene can be observed in many animal species, no animal model for benzene-induced leukemia is available. Benzene is not a classical mutagen but several of its metabolites are genotoxic. Nevertheless, benzene-induced leukemia appears to show some similarity to the leukemias that result several years after treatment of cancer patients with alkylating agents to achieve a remission of a variety of tumors. There is a latency period, myelodysplasia occurs, characteristic chromosomal disturbances are observed, and eventually AML develops.

Toxicokinetics and biotransformation Owing to the high vapor pressure of benzene, inhalation is the most important pathway of human exposure. The biotransformation of benzene to polar metabolites is dose-dependent; after low doses approximately 50% of the received dose is transformed into metabolites. The remainder is exhaled as unchanged parent compound. The major pathway for benzene biotransformation is formation of CYP 2E1-mediated benzene oxide, which is in equilibrium with the corresponding oxepin, and which rearranges to form phenol (Figure 6.23). 'Phenol sulfate' (phenyl hydrogen sulfate) is the major benzene metabolite excreted with urine in humans. Besides phenol, a number of other metabolites are formed from benzene oxide. The epoxide may be converted into a mercapturic acid via conjugation with glutathione, ultimately yielding *N*-acetyl-*S*-phenyl-L-cysteine. Benzene oxide is a substrate for epoxide hydrolase to yield benzene dihydrodiol, (1,2-cyclohexa-3,5-dienedid) which is, in turn, converted into catechol via dihydrodiol dehydrogenase. Hydroquinone and catechol are formed by further oxidation of phenol, and 1,2,4-trihydroxybenzene is thought to be a product of the hydroxylation of hydroquinone. The ring may be opened, presumably at the epoxide stage, by an unknown mechanism, to form mucondialdehyde, which is further converted into muconic acid.

Some of the benzene metabolites, e.g. *p*- and *o*-benzoquinone, the quinone forms of hydroquinone or catechol, or mucondealdehyde may yield a variety of free radicals, which may covalently bind to proteins. Since bone marrow has only a limited capacity for the biotransformation of benzene, it is assumed that benzene metabolites formed in the liver are distributed with blood to bone marrow and may undergo further metabolism by peroxidases to form phenoxyl radicals and semiquinone radical anions as ultimate toxic metabolites. The interactions of these reactive products with cellular macromolecules may be the basis of the bone marrow insult. It will be important to determine the

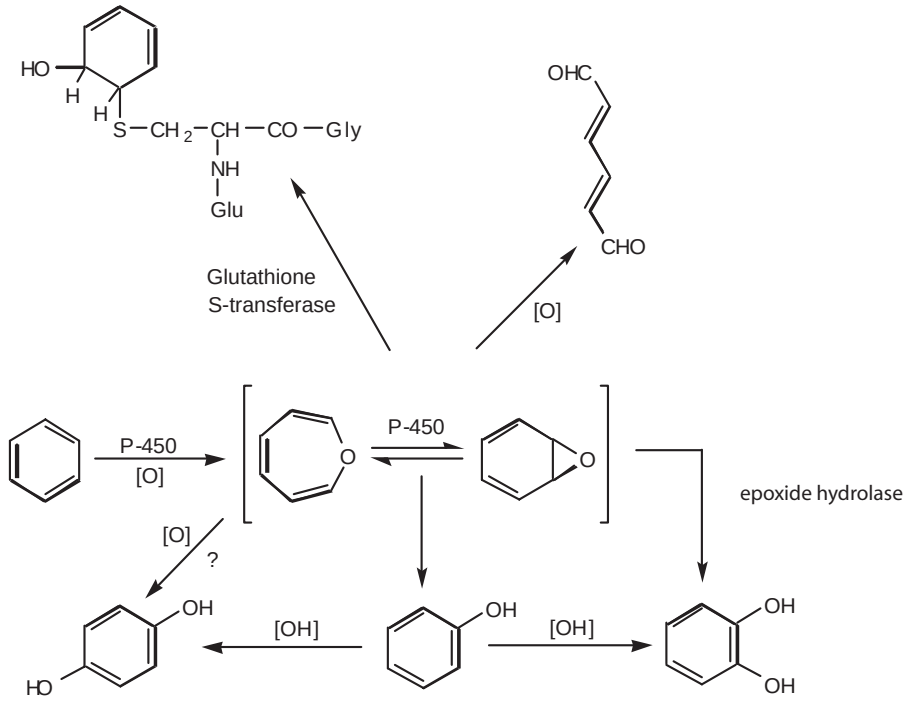


Figure 6.23 *Biotransformation of benzene by CYP, peroxidases, and epoxide hydrolase.*

significance of covalent binding of reactive metabolites of benzene to either proteins or nucleic acids.

Toluene Toluene is widely used as a solvent when a lower volatility is required, and toluene also serves as a starting material for chemical synthesis.

Toxicity After single high-dose inhalation exposures, toluene induces the typical signs of central nervous depression such as headaches, dizziness, and nausea. When toluene is abused to produce a euphoric effect (for example by the sniffing of glue containing toluene) symptoms such as euphoria, mild tremors, abnormality in walking, and behavioral changes have been described. Repeated inhalation exposures to high concentrations of toluene cause the unspecific symptoms also seen after single exposures, in addition to the other general solvent toxicity described above. The accompanying changes in the electroencephalogram are indicative of damage in the cerebellum and/or cortical spinal tract. Frequent inhalation of toluene-containing mixtures to induce euphoria has been reported to cause nephrotoxicity. Skin contact with toluene causes defatting of the skin, and aspiration of liquid toluene is associated with marked lung inflammation with tissue necrosis as a consequence.

Toxicokinetics and biotransformation Usually, the major pathway of human exposure to toluene is inhalation and approximately 75% of the inhaled toluene is retained. Retained toluene is metabolized by CYP to give benzyl alcohol, which is further oxidized to

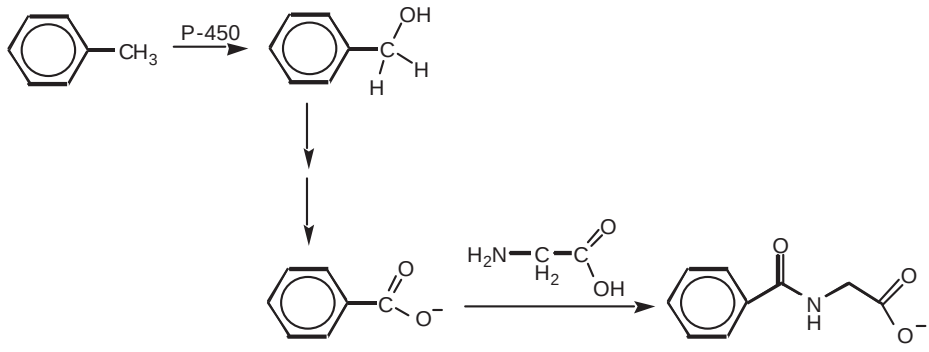


Figure 6.24 *Biotransformation of toluene by CYP and further oxidation of benzyl alcohol to benzoic acid followed by conjugation with glycine to give hippuric acid.*

benzoic acid; benzoic acid may further be conjugated with glycine to give hippuric acid (Figure 6.24). Oxidation of the aromatic ring is a minor metabolic pathway for toluene. The absence of ring oxidations may explain the absence of a carcinogenic effect or leukemogenic effect of toluene in the available studies.

Mixtures of Aliphatic Hydrocarbons Gasoline and kerosene (paraffin oil) are mixtures of saturated and unsaturated aliphatic hydrocarbons with a different content of aromatics. Gasoline used in passenger cars, which sometimes is also applied as a cleaning solvent, mainly contains isomers of hexane, heptane, and octane in addition to other components such as ethanol or methyl *tert*-butyl ether in concentrations of up to 10%.

Toxicity Despite the very wide availability of gasoline, intoxications are not frequent, due to the low toxicity of gasoline. Inhalation exposures to concentrations above 2000 ppm of gasoline cause typical symptoms of narcosis. Oral intoxication by gasoline occurs mainly in children and may result in a complex disease. Irritation of the mucosa of the stomach usually results in vomiting during which some gasoline may enter the lung and cause massive inflammation followed by tissue necrosis. In laboratory rodents, unleaded gasoline is nephrotoxic and causes renal tumors in male rats after long-term exposure. The mechanism of this rodent-specific tumor induction has been elaborated and represents one of the first examples where tumor induction seen in animals by a chemical was not considered relevant to human risk assessment.

Toxicokinetics and biotransformation Mechanistic studies have been performed with 2,2,4-trimethylpentane ('isooctane'), which is present in gasoline as a major component. Trimethylpentane is oxidized by CYP to give 2,2,4-trimethylpentan-2-ol, which binds to the circulating protein α_{2u}-globulin (alpha-2-urinary-globulin). α_{2u}-Globulin is a major circulating and urinary protein in the male rat and its synthesis is controlled by androgens. α_{2u}-Globulin undergoes glomerular filtration in the kidney, is partially reabsorbed in the proximal tubule, and is degraded in the lysosomes of the proximal tubular epithelial cells. α_{2u}-Globulin undergoes filtration and partial reabsorption. However, the modified protein cannot be efficiently degraded and thus accumulates in the lysosomes. As a consequence of this accumulation, lysosomes are destroyed and it is

the release of lysosomal enzymes which causes cell death. To repair the damage in the kidney, regenerative hyperplasia of the tubular epithelial cells occurs. The massive cell proliferation and inflammatory responses induced as a consequence of tissue necrosis then contribute to tumor formation (Figure 6.25).

Since humans, female rats, and both genders of mice do not synthesize α_{2u}-globulin, this mechanism does not result in tumor formation in these species and tumors in the kidney of male rats caused by this mechanism are not relevant for human risk assessment as concluded by the USEPA.

Alcohols and Ethers

The most important aliphatic alcohol used for solvent purposes, methanol, induces metabolic acidosis and, sometimes irreversible, eye damage. Formic acid is a toxic metabolite of methanol due to its slow catabolism. In contrast, the intermediary acetic acid formed from ethanol is rapidly utilized and does not accumulate. The widely used methyl *tert*-butyl ether is of low toxicity.

Methanol Methanol is a widely used solvent and has been considered as an additive to gasoline for high-performance engines. The most frequent cause of human intoxication, however, is the unintended addition of methanol to alcoholic beverages or confusion with ethanol.

Toxicity In humans, ingestion of methanol causes time-dependent characteristic symptoms of intoxication. Shortly after ingestion of toxic amounts of methanol, euphoria similar to that seen with ethanol may occur. The status of euphoria lasts for only a few hours. Then, usually after an interval of 12 to 24 hours without signs of toxicity, methanol induces metabolic acidosis with the characteristic effects of increased respiration, increased blood pressure accompanied by headache, disorientation, and vomiting. These effects may persist for up to four days depending on the dose of methanol. A second major consequence of methanol intoxication, which often develops approximately 20 hours after exposure, consists of eye damage which may range from photophobia to blurred vision to blindness when sufficiently high doses of methanol are consumed. After low doses, the eye damage with blurred vision may persist for several days, but permanent impairment of vision is not uncommon. The targets of methanol within the eye are the retina and the optic nerve. Blindness and death as an outcome of methanol intoxication have been reported after doses as low as 0.1 ml/kg body weight.

Toxicokinetics and biotransformation After oral administration, methanol is slowly, but completely, absorbed from the gastrointestinal tract and distributed to body water. Most of the methanol consumed is metabolized by alcohol dehydrogenase (Figure 6.26) to give formaldehyde, which is rapidly metabolized by formaldehyde dehydrogenase to formic acid. In humans and nonhuman primates, the further oxidation to carbon dioxide is slow. In addition, urinary excretion of formic acid in primates is also slow. Thus, formic acid may accumulate to cause metabolic acidosis and the other toxic effects seen in methanol poisoning. Indeed, formic acid accumulation in the eye seems to be responsible for the

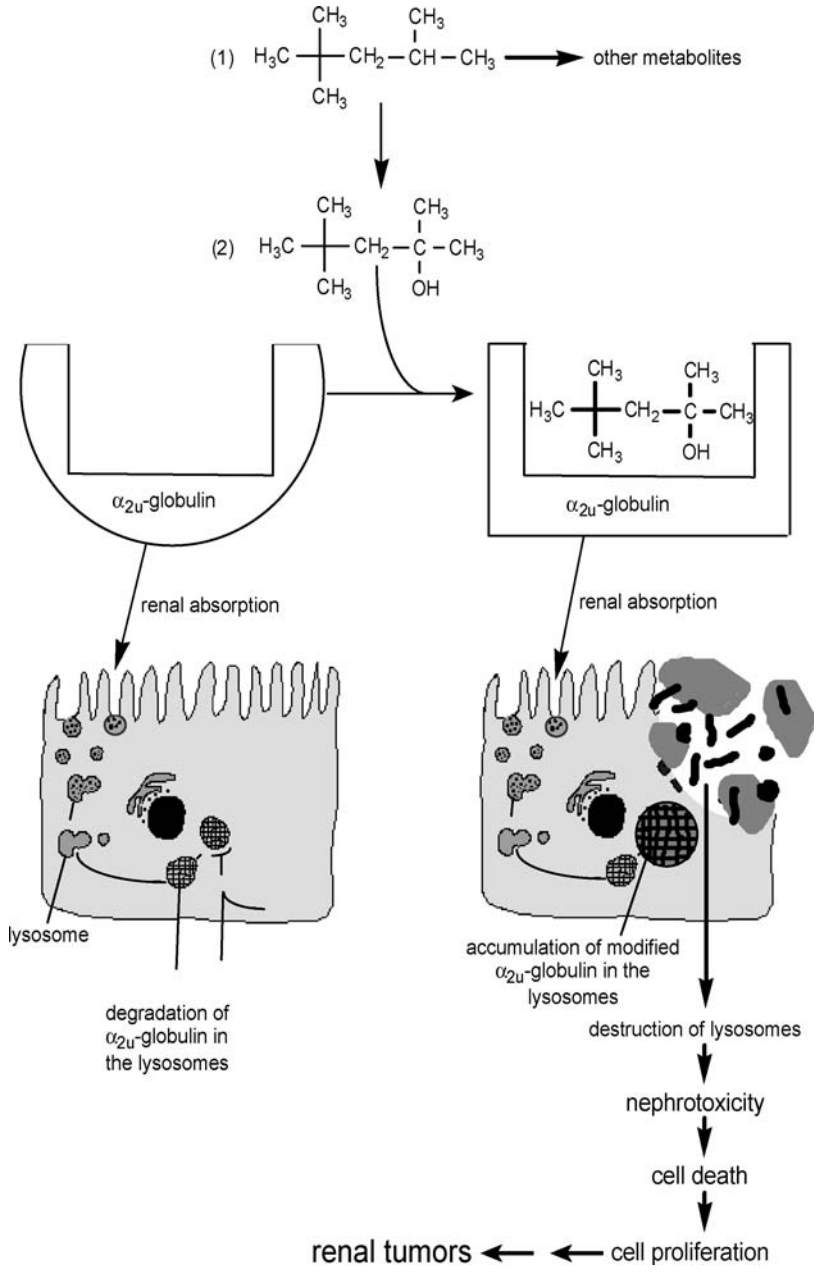


Figure 6.25 *Non-covalent binding of the 2,2,4-trimethylpentane metabolite 2,4,4-trimethylpentan-2-ol to α_{2u}-globulin impairs its lysosomal degradation, resulting in accumulation in the kidney tubule cells. Destruction of the lysosome and release of lysosomal-degrading enzyme into the cytosol is the initiating event leading to nephrotoxicity and formation of renal tumors.*

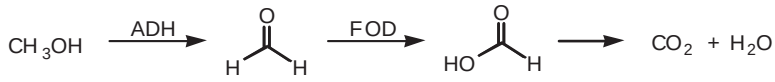


Figure 6.26 *Scheme for the biotransformation of methanol by alcohol dehydrogenase (ADH) and formaldehyde dehydrogenase (FOD) to formic acid and further catabolism of formic acid to carbon dioxide.*

eye damage. In rodents, catabolism of formic acid to carbon dioxide is more rapid when compared with primates, and rodents are generally much less sensitive to methanol toxicity. Blocking of methanol oxidation to formic acid by ethanol, a competitive inhibitor of alcohol dehydrogenase with higher affinity than methanol, is an efficient therapy of methanol intoxication.

Methyl tert-Butyl Ether (MTBE) Methyl *tert*-butyl ether (MTBE) is used as an additive to gasoline to increase oxygen content. Increased oxygen content is required in the United States to improve combustion efficiency and reduce emissions from motor vehicles. In addition, blending gasoline with MTBE and other ethers increases the octane number of gasoline for high-performance engines.

Toxicity Acute exposure to high concentrations of MTBE causes the typical symptoms of solvent-induced CNS depression. In rodents, long-term administration of very high doses of MTBE causes tumors at several target sites (kidney, testes and lymphohematopoetic system, liver). Owing to the lack of genotoxicity of MTBE and species-specific mechanisms elaborated for MTBE-induced tumors, the tumors observed in animals after MTBE exposure do not raise concern for tumorigenicity of MTBE in humans.

Toxicokinetics and biotransformation Absorption of MTBE from the gastrointestinal tract and the respiratory tract is rapid. Absorbed MTBE is uniformly distributed throughout the body. A higher concentration in fat tissue is reached after steady state due to the high fat solubility of MTBE. Elimination of MTBE metabolites occurs rapidly by exhalation and by excretion of metabolites with urine.

The biotransformation of MTBE also occurs by CYP-dependent metabolism (Figure 6.27), the intermediates *tert*-butyl alcohol (TBA) and formaldehyde are rapidly further metabolized. Formaldehyde is oxidized to formic acid and CO_2 and may enter the physiological one-carbon pool. Even at high MTBE doses, an increase in intracellular formaldehyde concentrations did not occur, due to rapid and efficient detoxication of formaldehyde. Further biotransformation of TBA produced from MTBE is dose-dependent; small amounts of TBA are excreted by exhalation and in urine.

The two other metabolites present in the urine of animals and humans exposed to MTBE were 2-methyl-1,2-propanediol and 2-hydroxyisobutyrate. The pathway for the formation of these metabolites involves oxidation of TBA to give 2-methyl-1,2-propanediol.

6.5.3 Summary

Many solvents, upon acute exposure at high doses, cause anesthesia. After repeated exposure, solvents may induce a wide variety of toxic effects, many of them specific for a defined chemical. While the acute effects of anesthesia are dependent on the accumulation of the parent compound in lipid-rich tissue of the central nervous system, the

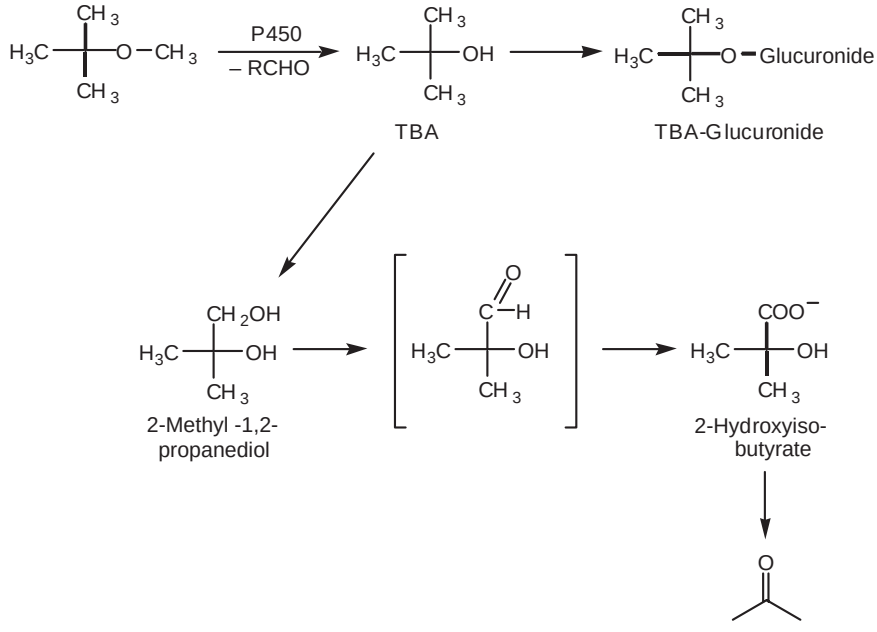


Figure 6.27 *Biotransformation of MTBE in rodents and humans.*

toxicities seen after long-term exposures are based on the biotransformation to reactive or toxic metabolites. Since the liver and the kidneys are the major sites of metabolism these organs are the preferential sites of solvent toxicity. Repeated skin contact leads to dermatitis due to the removal of the protecting fatty layers on the skin surface.

Further Reading

Andersen, M. E. (2003). Toxicokinetic Modeling and its Applications in Chemical Risk Assessment. *Toxicol. Lett.*, **138**, 9–27.

Barceloux, D. G., Bond, G. R., Krenzelok, E. P., Cooper, H., and Vale, J. A. (2002). American Academy of Clinical Toxicology Practice Guidelines on the Treatment of Methanol Poisoning. *J. Toxicol., Clin. Toxicol.*, **40**, 415–446.

Caro, A. A. and Cederbaum, A. I. (2004). Oxidative Stress, Toxicology, and Pharmacology of CYP2E1. *Annu. Rev. Pharmacol. Toxicol.*, **44**, 27–42.

Feron, V. J., Cassee, F. R., and Groten, J. P. (1998). Toxicology of Chemical Mixtures: International Perspective. *Environ. Health Perspect.*, **106 Suppl. 6**, 1281–1289.

Lucchini, R., Albini, E., Placidi, D., and Alessio, L. (2000). Mechanism of Neurobehavioral Alteration. *Toxicol. Lett.*, **112–113**, 35–9.

Needleman, H. L. (1995). Behavioral Toxicology. *Environ. Health Perspect.*, **103 Suppl. 6**, 77–79.

Sterzl-Eckert, H. and Greim, H. (1996). Occupational Exposure. *Food Chem. Toxicol.*, **34**, 1177–1178.

Wolf, D. C. and Butterworth, B. E. (1997). Risk Assessment of Inhaled Chloroform Based on its Mode of Action. *Toxicol. Pathol.*, **25**, 49–52.

6.6 Noxious Gases

Ladislaus Szinicz

6.6.1 Introduction

In order to warrant optimal gas exchange only a very thin membrane, consisting of alveolar epithelia, capillary endothelia, and the basal membrane in between, separates respiratory air from the blood in the lung. As gaseous compounds reach the alveolar space they pass practically unhindered into the blood and are rapidly distributed in the organism. The total exchange surface in the lung approximates the size of a tennis-court. Thus, even relatively polar noxious gases such as hydrogen cyanide and hydrogen sulfide, which hardly penetrate the skin, rapidly enter the circulation.

According to their mechanism of action, airborne toxicants may be classified as pulmonary irritants which act locally or systemic poisons. In general, pulmonary irritants cause irritation of the airways and damage leading to alveolitis and Acute Respiratory Distress Syndrome (ARDS), while systemic poisons cause asphyxia, hypoxia, and central respiratory arrest.

6.6.2 Airborne Systemic Poisons

Carbon Monoxide

Carbon monoxide (CO), a colorless, odorless gas, which exhibits about 200–300 times higher affinity than oxygen (O_2) for the divalent iron atom of hemoglobin (Hb). Relatively low concentrations of CO in air lead to substantial binding and formation of carboxyhemoglobin (COHb), thereby reducing the availability of hemoglobin for O_2 transport to the tissues. Exposure to 50 ppm CO in air results in a COHb concentration in the blood of about 5%. These values have been observed in smokers and are harmless for healthy persons. When COHb reaches a concentration of 10% the result is fatigue, weakness and/or headache. Persons with preexisting cardiovascular diseases are much more prone to CO toxicity.

Sources

Incomplete combustion in inadequately ventilated areas frequently leads to severe poisoning with carbon monoxide.

Combustion of carbon-containing material always produces some CO. The yield increases with decreasing availability of O_2. Severe CO poisoning in people is, therefore, almost completely limited to exposure in confined places. Faulty heating equipment and the use of charcoal grills, or open kerosene (paraffin oil) or gas heaters in closed spaces are frequent sources of CO exposure and poisoning. The most frequent cause of CO poisoning is automobile exhaust. While exposure in the car or in closed garages is often life threatening, exposure at poorly ventilated high-traffic places such as underground passages may cause some mild symptoms but is, in general, not a source of exposure leading to serious poisoning.

Methylene dichloride, an ingredient of paint and varnish removers, is metabolized to CO in the organism. Its use in poorly ventilated places for longer periods may lead to CO poisoning.

In the body CO is endogenously formed from Hb degradation, leading to physiological concentrations in the blood of nonsmokers of 0.4–0.7%. Smoking of one pack of cigarettes per day results in 4–6% COHb and of 2–3 packs in 7–9%. Natural sources of CO in the atmosphere are the oxidation of methane, forest fires, and marine microorganisms. The air usually contains less then 0.001% CO. The maximum allowable working place concentration in Germany (MAK) is 30 ppm. In the US the OSHA PEL is 50 ppm; the ACGIH TLV is 25 ppm.

Mechanism of Action The greater affinity of CO than oxygen for hemoglobin can result in substantial occupation of oxygen-binding sites by CO at relatively low CO concentrations, resulting in hypoxia.

The binding of CO to hemoglobin is reversible: CO + Hb <=> COHb.

The physiological impact of CO on the ability of hemoglobin to bind or release oxygen depends upon the source of hemoglobin and biological system under observation. CO is a competitive antagonist of O_2 at the divalent iron atom of Hb, exhibiting 200–300-times greater affinity. The effect may relate to the oxygen requirements of specific organs. Significant differences can be encountered between normal adults and the fetus *in utero*.

CO toxicity results from the reduction in Hb available for O_2 transport. Elevated levels of COHb can lead to tissue hypoxia. Organs and tissues with high metabolic activity such as the brain (nervous tissue) and heart (cardiac muscle) are most vulnerable. Thus, for the same proportion of binding to Hb, when compared with O_2, a 200–300-times lower concentration of CO in the air can suffice to produce toxicity. At 21% O_2 in the air, arterial blood is completely saturated. In contrast, only 0.1% (1000 ppm) CO in the air is sufficient to produce 50% saturation.

COHb does not bind and transport O_2. Moreover, in presence of COHb the ability of Hb to release O_2 in the periphery decreases, causing a further aggravation of hypoxia.

It has been suggested that the increased release of nitric oxide, destructive enzymes, and excitatory amino acids from leukocytes, platelets, and damaged vascular endothelia are contributory factors. CO is also bound to the divalent iron atom of myoglobin and other cytochromes, but the relevance for poisoning is not clear because of low affinity of myoglobin for CO.

CO can be of special concern in pregnant women. The fetus has a 10–15% higher burden of COHb, because of the higher affinity of CO for fetal hemoglobin. In addition, fetal blood exhibits lower oxygen tension, compared with that of the mother. Fetal Hb also exhibits a decreased O_2-release capacity in the periphery, which further aggravates the hypoxia. Thus, newborns show a higher sensitivity to CO than do adults.

Pharmacokinetics

Inhaled carbon monoxide readily enters the circulation and binds to hemoglobin.

The velocity of CO uptake through the lung is determined by the diffusion of the gas through the alveolar membrane. The polarity of CO and subsequent binding to Hb are relatively slow and the time required to reach maximum binding at a concentration of 1000 cm^3/m^3 is 8–10 h. The time may be reduced by a few hours by increasing the minute volume of respiration. About 85% of CO will be bound to Hb, the remainder being distributed to myoglobin and other proteins and hemoproteins. The elimination of CO, which depends upon the concentration of O_2 in air and the minute volume of ventilation, is slow. The half-life of COHb when breathing ambient air, i.e. air containing about 20% oxygen, is 3–4 h. However, at 100% O_2 $t_{1/2}$ is 30–40 min, and at hyperbaric conditions (100% O_2, 2.5 atm.), it is 15–20 min.

Therapeutic Principles

Free airways and plenty of oxygen are decisive for effective recovery.

Vomiting and aspiration are frequent in CO poisoning. Clear airways and removal of the victim from the contaminated area are necessary for recovery. Oxygen accelerates the elimination of COHb. Oxygen can be supplied either as 100% O_2 or better still as 95% O_2/5% CO_2 to stimulate the respiratory center and enhance respiratory minute volume. Hyperbaric oxygen (HBO) is the most effective therapy (up to 3 ATA), but it is dependent on the availability of an HBO facility.

Effects

Symptoms of mild to moderate carbon monoxide poisoning are not very specific, leading frequently to misdiagnosis.

Table 6.18 shows a list of symptoms observed with increasing COHb concentrations. Signs of mild to moderate poisoning might also be caused by other diseases such as

Table 6.18 *Symptoms of CO poisoning in relation to the COHb concentration in the blood.*

Symptoms	Blood COHb concentration (%)
Decreased psychomotor performance, decreased exercise tolerance	5
Headache, dizziness, malaise	10–20
Nausea, vomiting, throbbing headache, visual disturbances	20–30
Shortness of breath, headache, palpitation, tachycardia, syncope	30–40
Confusion, collapse	40–50
Coma, respiratory failure, cardiovascular depression, seizures	50–70
Respiratory failure, death	70–80

pneumonia, cardiac infarction, viral infections, cholecystitis, or epilepsy, which may as a result of CO poisoning, be misdiagnosed. In patients with preexisting cardiovascular diseases and limited tissue oxygenation even a small increase in COHb, e.g. to 5%, might induce signs of angina pectoris.

Chronic and Delayed Effects

Chronic and delayed effects, usually focusing on neuropsychiatric complaints, are frequent after CO poisoning.

Apathy, disorientation, amnesia, hypokinesia, mutism, irritability, urinary and fecal incontinence, gait disturbance, and increased muscle tone are among the most common delayed effects of CO poisoning and may not become apparent for 2 to 40 days after exposure.

Cyanide

Sources

Industry, combustion, pesticides, therapeutics, and plants are sources of cyanide exposures leading to poisoning.

Hydrocyanic acid (HCN) and the cyanide group (CN^-)-containing compounds (cyanides) are commercially used in electroplating, ore extraction, metal processing, and the synthesis of plastics, pesticides, and drugs, among others. Combustion of organic compounds containing carbon and nitrogen (e.g., polyurethane or polyacrylonitrile) may lead to the release of cyanides.

Hydrolytic enzymes from plants and human gut β-glucosidase can hydrolyse cyanogenic glycosides from plants such as bitter almond, apple seeds, plum, peach, and apricot kernels, and cassava beans and roots. HCN in the organism originates from the food and from vitamin B_{12} metabolism. In the blood, cyanide is concentrated in erythrocytes. Normal plasma levels are 4 ng/cm^3 in nonsmokers and 6 ng/cm^3 in smokers. These values may be expressed as 15 and 40 ng/cm^3, respectively, in whole blood. Sodium nitroprusside is an intravenous antihypertensive medication used in emergency medicine, which is metabolized to cyanide. The maximum allowable working place concentration in Germany is 1.9 ppm. In the US the OSHA PEL relates to skin exposure and is 11 mg/m^3 (10 ppm); both ACGIH and NIOSH recommend a STEL of 4.7 ppmμ.

Mechanism of Action

The high affinity of the cyanide ion for the trivalent iron in cytochromes is the cause of cytochrome a,a_3 inhibition and reduction of mitochondrial oxidative metabolism.

The high affinity of CN^- for oxidized iron in cytochromes of the respiratory chain blocks the last step of oxidative phosphorylation and consequently causes cellular ATP depletion. The binding to cytochrome iron is reversible.

Toxicokinetics

Rapid uptake of hydrocyanic acid occurs after inhalation whereas absorption through the skin is toxicologically not relevant.

Gaseous HCN rapidly enters the blood stream and is distributed throughout the body. Mucous membranes are also rapidly penetrated, whereas penetration via the skin is slow. After oral uptake of cyanide salts, the acid is liberated by hydrochloric acid in the gastric juice and rapidly absorbed. Cyanogenic glucosides such as amygdalin, prunasin, dhurrin, and linamarin yield HCN under the influence of β-glucosidase or linase in the gut or by hydrolysis and HCN is then absorbed.

HCN is detoxified predominantly by the enzyme rhodanese to thiocyanate in the presence of thiosulfate [Equation (6.11)]:

$$CN^- + Na_2S_2O_3 \xrightarrow{\text{rhodanese}} SCN^- + Na_2SO_3 \qquad (6.11)$$

In humans the enzyme metabolizes about 1 mg cyanide/kg body weight. Cyanide is also eliminated by binding to hydroxocobalamine with formation of cyanocobalamine (vitamin B_{12}). Small amounts are also exhaled and excreted by sweat causing a bitter almond odor.

Intoxications

Loss of consciousness with seizures, metabolic acidosis, and respiratory failure in the absence of cyanosis indicate acute cyanide poisoning. Chronic cyanide poisoning is an unusual event.

Breathing about 20 ppm of HCN for a few hours may result in headache. Death may follow exposure to 100 ppm after one hour and at 300 ppm after about 10 min. The oral lethal dose of KCN and NaCN in humans is between 1 and 5 mg/kg. The symptoms are the result of a deficiency of cellular energy. Tissues with high energy demand show highest sensitivity. CNS symptoms occur first, but are not very specific such as dizziness, nausea, dyspnea, tachypnea, restlessness, and anxiety. Severe poisoning is associated with seizures, loss of consciousness, respiratory failure, and dilated fixed pupils. The cardiovascular system is much less sensitive. In very severely poisoned individuals cyanosis might be absent initially despite respiratory arrest because of good saturation of the blood with oxygen through tachypnea and decreased intracellular oxygen consumption.

Continuous exposure to cyanide can cause headache, dizziness, nausea, vomiting, and bitter almond taste. Vitamin B_{12} and folate metabolism, as well as thyroid function, may be mildly disturbed. In Nigeria, a tropical ataxic neuropathy with enhanced thiocyanate levels was linked to chronic cyanide exposure from cassava consumption.

Therapeutic Principles

The aim of antidote use is rapid elimination of cyanide ions from the cells by binding in the blood and transformation into less toxic metabolites to be excreted in the urine.

Cyanide exhibits a high affinity for oxidized iron (Fe^{+++}) in cytochromes. Conversion of some hemoglobin into methemoglobin draws cyanide from tissue cytochromes to erythrocytes, enabling recovery of oxidative metabolism in the tissues.

Transformation of 30% Hb into methemoglobin results in enough capacity to bind about 8 times the LD_{50} of cyanide ion. This can be achieved by an i.v. injection of **4-(dimethylamino)phenol** (4-DMAP), 3 mg/kg.

In some countries, **sodium nitrite** or **amyl nitrite** are more commonly used for methemoglobin formation in cyanide poisoning. The dose at which either is used is limited, because of strong vasodilation and reduction in blood pressure. Up to 12 mg/kg $NaNO_2$, can be given i.v., sufficient for formation of about 15% methemoglobin. Amyl nitrite, a very volatile liquid, is in some textbooks recommended to be used by vapor inhalation in cyanide poisoning. However, inhalation of amyl nitrite is practically ineffective, because of rapid exhalation of the vapor.

The use of these methemoglobin-forming drugs should also be combined with sodium thiosulfate. A complimentary mechanism for detoxifying cyanide is through the enzyme rhodanese, a normal body constituent which mediates the reaction between cyanide and thiosulfate to yield the detoxication product thiocyanate. For maximum effectiveness $Na_2S_2O_3$ (100 mg/kg, i.v.) must be administered. In contrast to CN, thiocyanate is much less toxic and more rapidly excreted in the urine.

Hydroxocobalamine, a vitamin B12 congener, detoxifies cyanide by a reaction which yields cyanocobalamine, which is eliminated from the body by excretion in the urine. Cyanide exposure resulting from breathing smoke released by a fire may be accompanied by CO inhalation, which can impair tissue oxygenation. Unlike methemoglobin formation by nitrite, the hydroxocobalamine method of trapping cyanide does not interfere with tissue oxygenation. The International Programme on Chemical Safety/Commission of the European Communities suggests that large doses of hydroxocobalamine are required. It was estimated that about 1406 mg are required to inactivate 1 mmole of cyanide, i.e. 65 mg of KCN. In most countries commercial preparations of the antidote can be obtained only in ampoules containing 1–2 mg, although a 4 gram preparation is available in France. A solution of hydroxocobalamine in 5% dextrose, which may or may not contain thiosulfate, can then be administered intravenously. Some organ toxicity or allergic response to treatment may be encountered.

Hydrogen Sulfide

Sources

Hydrogen sulfide occurs in industry, agriculture, in sewage, and in the environment.

Industrial sources of hydrogen sulfide (H_2S), which yields a smell commonly described as that of 'rotten eggs,' include facilities for the production of viscose and rayon, paper, mineral oil, vulcanized rubber, and illuminating gas, and from heavy water production in nuclear plants. Waste from the production of cellulose, sugar, glue, and from tanneries contains the gas. Other sources include places where fermentation and the rotting of fish, manure, and raw sewage occurs. Natural sources are volcanoes and sulfur springs. Small amounts of H_2S are also formed in the gut of living organisms.

Mechanism of Action

Hydrogen sulfide, much like cyanide, is thought to inhibit cytochrome a,a_3 and, thereby, oxidative metabolism.

H_2S exhibits a somewhat higher affinity for oxidized iron (Fe^{+++}) in cytochrome oxidase than does cyanide. Consequently, oxidative metabolism is inhibited and metabolic acidosis may occur.

Pharmacokinetics

Rapid pulmonary but negligible dermal absorption, and rapid binding and metabolism in the tissues characterize hydrogen sulfide.

H_2S rapidly penetrates from the air into the blood and is distributed to the tissues. Dermal uptake is negligible. The gas is rapidly eliminated by oxidation to sulfate, methylation and synthesis of SH-containing proteins. The probability of accumulation is low and very little is eliminated unchanged.

Effects

Sudden collapse and a smell of 'rotten eggs' is indicative of poisoning. Chronic poisoning and delayed effects are uncommon.

Table 6.19 shows that perception of the smell occurs much earlier than irritation of eyes or respiratory epithelia. At concentrations causing mucous membrane irritation, olfactory perception rapidly becomes paralysed.

Table 6.19 *Effects of hydrogen sulfide relative to the concentration in air.*

Concentration (ppm)	Effect
0.02–0.1	Odor threshold
10–50	Irritation threshold (eyes, respiratory tract)
100–200	Loss of smell after 3–5 min
>250	Pulmonary edema after several hours of exposure
>500	Systemic poisoning with unconsciousness and respiratory paralysis after 0.5–1 hour exposure
>700	Respiratory failure and death after minutes

Therapeutic Principles

Rapid rescue with 100% oxygen respiration for the patient and airway protection for the rescuer are important measures.

Although H_2S binds to the oxidized iron of cytochrome oxidase, similar to cyanide, methemoglobin formation is an ineffective tool in mobilizing the poison from the tissue because of the rapid decay of the sulfmethemoglobin complex and the rapid elimination of sulfide. Furthermore, the H_2S–cytochrome oxidase complex seems to be very unstable. Patients, in general, recover very rapidly after termination of exposure.

6.6.3 Respiratory Tract Irritants

The primary localization of airway damage by irritant gases is largely determined by their water solubility.

Properties

An important protective factor which may determine the extent of damage by irritant gases is the initial warning provided by irritation and pain. Detection of painful stimuli motivates people to move away from the area of exposure. Mucous membranes of the eye, nose, throat, and larynx are very sensitive. In deeper lung areas perception of irritation and pain is absent. Therefore, water-soluble irritant gases have a higher warning capacity and cause a faster active termination of exposure. The sensory perception of exposure to lipid-soluble gases is much lower, and may be negligible, especially at low concentrations capable of causing damage.

The solubility of a gas in water largely determines depth of the lung at which irritant gases exert their toxicity. Thus, highly water-soluble gases tend to be taken up into the mucous layers higher in the lung than are less water-soluble gases. Gases that impact on the deeper lung areas are those that are relatively nonpolar. In the absence of upper airway irritation these gases may reach deep into the lung, undetected, in sufficient concentration to produce severe damage such as lung edema.

Ammonia, hydrogen chloride, hydrogen fluoride, formaldehyde, and acrolein are examples of water-soluble gases. Moderate water solubility characterizes sulfur dioxide, chlorine, and bromine, whereas oxygen, ozone, nitrogen oxides, and phosgene are poorly soluble in water.

Exposure to gases with high water solubility causes immediately irritation of eyes, nose, throat, and larynx and at higher concentrations erosions of the above mucous membranes, and laryngospasm. Characteristic of damage by gases with moderate water solubility are symptoms of bronchial damage such as coughing, expectoration, bronchospasm, and subsequent bronchitis and bronchopneumonia. Gases displaying poor water solubility exert their effects predominantly in the alveolus where they target alveolar epithelia and capillary endothelia. As a consequence of the damage, an inflammatory reaction is initiated with swelling of the alveolar membrane, leading to inhibition of gas exchange. Alveolar exudation is a result of damage to alveolar epithelia and the basal

membrane. Nevertheless, prolonged exposure to water-soluble irritant gases at high concentration may also cause deep lung damage. Thus, in accidental exposure to ammonia, a highly water-soluble gas, victims who are not able to escape the room where the gas was emitted may develop lung edema.

Therapeutic Principles

Remove victims from contaminated area and provide immediate support of airway and breathing. After exposure to gases of high and moderate water solubility, mucous membranes of eyes, nasopharynx and, if necessary, exposed skin should be decontaminated with a physiological saline solution or tap water.

Sympathomimetics may be used to treat cases of larynx edema and bronchospasm. Tracheostomy might become necessary.

Pulmonary edema is a consequence of alveolar damage which often results in mortality. Several hours may pass until alveolar membrane damage results in exudation of fluid into the alveolus, resulting in detectable disturbances of gas exchange. The development of edema can be visualized before symptoms occur by chest X-ray. Clinical symptoms such as coughing, exudation, cyanosis, and dyspnea can appear up to 24 h after exposure to phosgene. Among suggested treatments, which may or may not be effective, are oxygenation, bronchial lavage, and administration of high doses of glucocorticoids.

Chronic Effects

Highly water-soluble respiratory irritants may cause necrotic damage to the mucous membranes of the eyes and upper respiratory tract, and lead to scar formation. These lesions are slow to heal. Chronic bronchitis can follow extensive acute damage, leading to pulmonary fibrosis. Extensive necrosis of small bronchi and alveoli can lead to lung fibrosis. Increased lung vascular resistance can lead to increased work load for the right heart.

6.6.4 Irritant Gases

Hydrogen Chloride

Hydrogen chloride gas and its aqueous solution, hydrochloric acid, are widely used in industry and in most chemical, pharmaceutical, and medical laboratories.

Hydrogen chloride (HCl) is a highly water-soluble, colorless gas, with a pungent odor. It is heavier than air (density 1.27; air 1.0). HCl is widely used in industry, e.g. pharmaceutical manufacturing, synthesis of organic compounds, and vinyl chloride production. Combustion of poly(vinyl chloride), chlorinated acrylics, and flame-retardant materials produces hydrogen chloride.

The maximum allowable working place concentration in Germany is 5 ppm. In the US the PEL value is 5 ppm. The ACGIH suggests a ceiling value of 2 ppm.

Effects HCl gas and hydrochloric acid are very strong irritants causing coagulation necroses at all surfaces they come into contact with.

Hydrogen Fluoride

Hydrogen fluoride and its aqueous solution are among the most corrosive compounds known.

HF is a colorless, highly corrosive gas, which is heavier then air (density 1.27; air 1.0). Commercial use includes mineral oil processing, aluminum manufacture, separation of uranium isotopes, glass and enamel etching, and production of fluorinated resins and paint.

The maximum allowable working place concentration in Germany is 3 ppm. In the US the OSHA PEL is 3 ppm; the TLV is 0.5 ppm with a ceiling value of 2 ppm.

Effects HF causes enzyme inhibitions and cellular damage leading to necrosis. In contrast to other halogens HF binds calcium to form insoluble calcium fluoride. The deprivation of calcium exacerbates tissue damage, which can be ameliorated by calcium replacement.

Formaldehyde

Formaldehyde is a gas with pungent odor, broad commercial use, and is a suspected carcinogen.

Formaldehyde (HCHO) is a reactive, colorless, combustible gas with a pungent odor. (Synonyms: formalin, methyl aldehyde, methylene oxide). Commercial use of HCHO includes the production of phenol- and urea-based polymers. It is used for medical purposes as a disinfectant, antiseptic, deodorant, tissue fixative, and embalming agent. Because of suspected carcinogenicity, permissible exposure in Germany is limited to 0.5 ppm by technical regulations for hazardous materials. In the US the PEL is 0.75 ppm; ACGIH recommends only a ceiling value of 0.3 ppm.

Effects HCHO is a mucous membrane and skin irritant. It is metabolized to formic acid in the body. Carcinogenic effects were observed in the nose of mice and rats exposed to 5 ppm for their lifetime. In all cases local necroses were observed prior to tumor appearance in affected areas. Cytotoxicity stimulates cell proliferation and results in tumor promotion. In the absence of cytotoxicity no carcinogenicity was observed. Therefore, doses at which no cytotoxicity were observed are considered not to be carcinogenic.

Sulfur Dioxide

Sulfur dioxide is an important environmental toxicant, being a major cause of acid rain originating mainly from combustion of fossil fuels.

Sulfur dioxide (SO_2) a highly water-soluble, colorless, and noncombustible gas with a pungent odor that is 2.3 times heavier than air. Commercial applications include preservation of fruit and vegetables via fumigation, disinfection of breweries, and

bleaching of textiles, straw, wicker-work, gelatin, glue, and beet-sugar. Other sources of SO_2 include smelting, fossil fuel combustion, paper manufacturing, and fabrication of rubber.

The maximum allowable working place concentration in Germany is 5 ppm. In the US the PEL value is 5 ppm; The TLV value is 2 ppm with a STEL value of 5 ppm.

Effects In aqueous media such as mucus, SO_2 is found as sulfurous acid, which is capable of damaging the epithelia of the airways and inhibiting mucocilliary transport. The odor threshold is about 1 ppm. Airway irritation and, eventually, nosebleeds occur at about 10 ppm, and eye irritation at 20 ppm or higher. The maximum tolerable concentration is reached at 50–100 ppm. Concentrations higher than 400 ppm represent an immediate danger to life. Patients having hyperreactive airways indicative of bronchial asthma are particularly sensitive to SO_2. It is assumed that the severity of pulmonary diseases is increased by sulfur dioxide, leading to increased mortality from cardiovascular and pulmonary causes.

Chlorine

Chlorine is, after fluorine, the second most reactive and toxic gas.

Chlorine is a gas with a pungent odor. It is 2.49 times heavier than air, it is yellow-green in color and is very reactive. Commercial use includes the production of vinyl chloride and poly(vinyl chloride), solvents, and bleaching materials. Chlorine is widely used in water purification and paper bleaching. It forms explosive mixtures with hydrogen. During World War I chlorine was the first gas used as a chemical warfare agent on a large scale and its release resulted in many casualties.

The maximum allowable working place concentration in Germany is 0.5 ppm of air. In the US the PEL value is 1 ppm; the TLV value is 0.5 ppm with a STEL value 1 ppm.

Effects The odor threshold is about 0.2 ppm. Mucous membrane irritation occurs at about 1 ppm. At 30 ppm dyspnea, chest pain, nausea, vomiting, and coughing are observed. Inhalation of 40 ppm can cause pulmonary edema. Lethality occurs after 30 min at 400 ppm. In aqueous media (mucus) chlorine forms hydrochloric (HCl) and hypochlorous (HOCl) acid. The latter decomposes to chloric acid ($HClO_3$) and oxygen free radicals. Effects on biological tissues include necrosis and chlorination, and oxidation of biological materials. HOCl formed by leukocytes is an important mediator of inflammation.

Isocyanates

The highly reactive NCO-group of isocyanates is responsible for their broad use as precursors in the chemical industry.

Isocyanates are widely used for the production of plastics, polyurethane foam, lacquers, adhesives, and fibers. The most important representatives of this class are toluene 2,4-diisocyanate (TDI), 4,4′-methylenebis(phenyl is ocyanate) (MDI), hexa-methylene diisocyanate (HDI), and 1,5-naphthalene diisocyanate (NDI).

The maximum allowable working place concentration in Germany for TDI, MDI and HDI is 0.01 ppm. In the US the TLV for HDI is 0.005 ppm, and the PEL for MDI is set at a ceiling of 0.02 ppm.

Effects In December 1994, a severe accident in which approximately 27 tons of methyl isocyanate was released at a pesticide plant in Bhopal, India, resulted in more than 3300 deaths. Isocyanates cause irritation of mucous membranes in the eye, nose, and throat. At higher doses pulmonary symptoms (cough, dyspnea, choking sensation), gastrointestinal symptoms (nausea, vomiting, abdominal pain), skin inflammation, and neurologic symptoms may develop. In the occupational setting, sensitization frequently results in bronchial asthma. Although less frequent, skin reactions may also be seen.

Whereas the diisocyanates are strong sensitizers the monoisocyanates are not.

Nitrogen Oxides

Among the various nitrogen oxides, the toxicologically most important is nitrogen dioxide.

The nitrogen oxides include nitrous oxide (N_2O) 'laughing gas,' nitric oxide (NO), nitrogen dioxide (NO_2), and the commercial product nitrogen tetroxide (N_2O_4), the latter usually made up of a mixture of NO_2 and N_2O_4. NO is rapidly oxidized in air to NO_2. N_2O and NO nitrogen tetroxide are colorless gases. NO_2 is slightly yellow to reddish-brown. The commercial nitrogen tetroxide/nitrogen dioxide mixture is 1.61 times heavier than air, and, while not combustible, can promote the combustion of carbon, phosphorus, and sulfur. Water solubility is low, the odor of NO_2 faint and chlorine-like.

The maximum allowable working place concentration of NO_2 in Germany is 5 ppm, the TLV 3 ppm with a STEL of 5 ppm.

Occupational exposure can be observed in manufacture of dyes, fertilizers, or lacquers. Welding, glass blowing, and food bleaching are also known sources. Silo filler's disease is a consequence of NO, NO_2, and CO_2 inhalation in silos, originating from decomposing plants in the presence of high nitrate concentration. NO_x in the atmosphere originates from aerial nitrogen oxidation and combustion of fossil fuels (power plants, home heating, and automobiles).

Effects NO is not an irritant, but is a physiologically important mediator. It causes vasodilatation and methemoglobinemia. It represents only a small fraction of NOX (NO_x) fumes and its contribution to toxicity is negligible. The effect of nitrogen dioxide is typical of poorly water-soluble irritant gases. Initial irritation of eyes, nose, and throat is mild, but after a latent period of 3–30 h lung edema accompanied by fever, dyspnea, coughing, hemoptysis, wheezing, rales, and cyanosis may develop, leading to respiratory failure. 50% of patients surviving lung edema develop bronchiolitis obliterans 2–3 weeks later.

In the aqueous medium of the mucus, nitric acid formed from nitrogen oxides may damage the epithelium. The reaction of nitrogen dioxide with olefinic binding sites in fatty acids can result in radical formation and subsequent lipid peroxidation of capillary endothelia; 100 ppm causes airway damage after 30–60 min, and 200 ppm may be fatal.

Phosgene

Phosgene was responsible for 80% of fatalities caused by chemical agents in World War I.

Phosgene ($Cl_2C{=}O$) is a colorless gas, heavier than air, with a musty hay-like odor. It is used for production of isocyanates, dyes, insecticides, and pharmaceuticals.

Thermal decomposition of chlorinated hydrocarbons can lead to phosgene release and potential poisoning.

The maximum allowable working place concentration in Germany is 0.1 ppm. In the US both the PEL and the TLV values are 0.1 ppm.

Effects Phosgene has poor warning properties. Only mild irritation of the eyes and throat occur, even at concentrations causing pulmonary edema. The molecular mechanism of damage is unknown.

Ozone

In the stratosphere ozone provides protection from excessive ultraviolet radiation. In the troposphere this irritant oxidizing gas is an important pollutant formed in light by the interaction of oxygen with nitrogen oxides and volatile organic compounds (VOCs).

Ozone (O_3) is a bluish, explosive gas with low water solubility. The odor is pleasant at low concentrations (up to 2 ppm) and pungent at higher concentrations.

Occupational sources of exposure are electric arc welding, mercury vapor lamps, photocopy machines, water-purification processes, bleaching, and the synthesis of some organic compounds.

In the stratosphere the ozone concentration is about 10 ppm and in the troposphere 1 ppm. Ozone is the most important pollutant in photochemical smog, where it comprises about 90% of the oxidants.

The maximum allowable working place concentration in Germany is 0.1 ppm of air. In the US the PEL value is 0.1 ppm. The TLV values are based on work effort and range from 0.5 ppm for heavy work to 0.2 ppm for light work.

Effects The damage by ozone is based on direct oxidation of various molecules in the tissues (-SH, $-NH_2$, -OH, phenol groups, etc.), formation of free radicals, diene conjugation, and lipid peroxidation. Increased release of cytokines in asthmatics exposed to ozone appears to exacerbate the disease.

Because of a somewhat higher water solubility, compared with phosgene and nitrogen oxides, the irritation of eyes and upper airways is more pronounced. High concentrations of ozone can cause lung edema. At low concentrations ozone causes initially increased reactivity of the bronchi, which disappears after prolonged exposure, as the airways adapt to oxidative stress.

Smoke Inhalation

Up to 80% of fire-related deaths are due to poisoning and not to thermal injuries.

Table 6.20 *Toxic gases in fires.*

Gas	Source	Effect[a]
Carbon monoxide	Organic materials at oxygen deficiency	S
Cyanide	Polyacrylates, polyurethanes, nylon, wool, silk	S
Hydrogen sulfide	Rubber	S, I
Sulfur dioxide	Rubber, sulfur-containing materials	I
Hydrogen chloride	Poly(vinyl chloride)	I
Phosgene	Chlorinated hydrocarbons	I
Isocyanate	Polyurethanes	I
Nitrogen oxides	Cellulose nitrate, wool, silk	I
Acrolein	Polyolefins and cellulose pyrolysis	I

[a]S = Systemic effects, I = Irritation (local effects).

Smoke is the volatilized product of combustion consisting of various gases and particulate matter. The gaseous phase can contain irritant and systemically poisonous gases. Particulate matter can transport toxic materials, with smaller particles reaching the lower respiratory tract. Hot smoke can cause burns of the upper airways, but the damage usually does not reach areas deeper than the larynx. Damage to deep lung areas develops from exposure to toxic materials.

Fires in closed or poorly ventilated places bear an especially high risk of poisoning. In most cases carbon monoxide is the main toxicant, but depending on the combusted material a large number of other systemically or locally poisonous gases can be formed (Table 6.20).

Effects Release of toxic gases in fires can lead to systemic poisoning (mainly asphyxiation) and to lung irritation (Table 6.20). While systemic gases have to be taken up and distributed in the circulation to the tissues, locally acting gases primarily cause irritation and damage to respiratory airway epithelia.

6.6.5 Summary

Acute poisoning by airborne toxicants is based either on systemic effects or on local damage of airway epithelia. Systemic poisoning is caused by gases such as carbon monoxide, hydrogen cyanide, or hydrogen sulfide. These compounds interfere with oxygen availability or its utilization in the tissues, thereby reducing oxidative metabolism in the mitochondria and the availability of energy for metabolic processes. High-energy-demanding tissues such as the CNS are most sensitive, leading to respiratory and circulatory failure in severe poisoning.

Locally acting irritant gases affect the epithelia of the respiratory tract, leading to irritation at lower doses and necrosis at higher concentrations. Extensive damage may lead to fibrosis of the lung. Gases displaying high water-solubility such as ammonia and formaldehyde primarily affect the upper respiratory airways such as the nasopharynx, the larynx, and the eyes. This area has a very dense sensory innervation, resulting in early warning in the form of irritation and pain. In contrast, smaller amounts of gases with low water-solubility, such as phosgene and nitrogen oxides, are dissolved in the mucus and

reach the epithelia of the upper respiratory tract and eyes. The irritation and the warning effect is not prominent and higher concentrations reach and damage the lower parts of the lung, i.e. small bronchi and alveoli. Therefore, lung edema is more characteristic of poisoning by gases with low water-solubility.

For Further Reading

Hardman, J.G., Limbird, L.E., Molinoff, P.B., Ruddon, R.W., and Goodman Gilman, A. (1996) Goodman and Gilman's The Pharmacological Basis of Therapeutics. McGraw-Hill, New York.

Ellenhorn, M.J., Schonwald, S., Ordog, G., and Wasserberger, J. (1997) Ellenhorn's Medical Toxicology: Diagnosis and Treatment of Human Poisoning. Williams & Wilkins, Baltimore.

6.7 Animal and Plant Toxins

Thomas Zilker

6.7.1 Introduction

There is an enormous variety of naturally occurring toxins. The chemical structure of these toxins ranges from simple organic compounds to very complex proteins. Animal toxins consist mostly of a mixture of polypeptides and different digestive enzymes. In plants alkaloids and glycosides are responsible for the toxic effects.

The morbidity of intoxications due to natural toxins differs from country to country and continent to continent according to the habitat of the toxic plants and animals. Nevertheless, as a consequence of travelling and the habit of keeping exotic animals in aquaria and terrariums, intoxications with animal toxins can happen around the world via animals which are not indigenous to that particular place. Moreover, snakes, scorpions, or spiders can transmigrate from continent to continent through modern transportation. Mushroom and plant intoxications occur mostly by oral intake; dermal toxic reactions are possible following contact with some plants.

A description of all known toxins can fill whole books. In this chapter only those intoxications are described which can lead to severe illness or even death.

6.7.2 Animal Toxins

Intoxications caused by animal toxins happen either by bites, stings, through dermal transmission, or by oral intake of sea-borne food, which is contaminated with a toxin. Bites happen through snakes and spiders. Stings happen through hymenoptera, scorpions, and fish.

Snakes

Snake bites from poisonous animals are the most common cause of severe morbidity and mortality among poisonous animals. In some areas of the world snake-bites are among the top ten fatal diseases. The Elapidae and Viperidae families are those that cause the most morbidity and mortality.

There are four families of snakes, which contain either exclusively venomous or some toxic species. The families with only venomous species include the Elapidae, the Viperidae, and the Atractaspididae. The last family is found only in Africa and the Middle East. Envenomations with the toxins of these snakes are usually mild. The family Colubridae includes mostly nonvenomous snakes. A few species of this family have fangs with venom glands, others have a toxic saliva which leads to an inoculation of the venom after a bite with ordinary teeth.

The Viperidae can be divided in two subfamilies, Viperinae and Crotalinae. The species *Vipera berus* (common in Europe) belongs to the Viperinae; the Crotalinae, which are also called pit vipers, includes the North American rattlesnakes. The Elapidae consist, among others, of the cobras and the kraits in Asia, the mambas in Africa, the tiger snake and taipan in Australia, and the coral snakes in the USA. Though the poisons of many snakes in the Elapidae family are neurotoxic, there are few species in this family that produce only local tissue damage (African tree cobra). Most venoms of the Viperinae are more or less toxic, producing mild to extreme local damage, but there are some that are neurotoxic (e.g., Middle East horned viper).

The four families of snakes have all different sorts of fangs. The Elapidae have hollow front fangs, which are short because they are permanently erect whereas the Viperinae and Crotalinae have fangs which fold back on hinges and can therefore be long. The Atractaspididacl have a small maxilla but still bear a big hollow front fang, which is erect and sticks out by the side of the mouth, so-called side-striking fangs.

The poisonous Colubridae have grooved fangs at the back of their mouth (rear fang snakes), which are permanently erect, or else have no fangs and instead have poisonous saliva.

Venom Variability There is significant intraspecies and even intraindividual venom variability in snake venom. This depends on the geographic range and even the season. The different groups of toxins and their various effects will be described.

Neurotoxins

There are two different kinds of neurotoxins in snake venom: presynaptically or postsynaptically acting venoms. The former are phospholipases, the latter small polypeptides.

The *presynaptic neurotoxins* belong to the phospholipase A_2 family of enzymes. Their target is the terminal axon of the neuromuscular junction. The early effect is a release of the neurotransmitter, the later effect a destruction of the axonal structure leading to a disruption of the synaptic vesicles with a complete cessation of transmitter release.

Signs and symptoms in a bitten patient are a flaccid paralysis starting with the cranial nerves, ptosis coming first, followed by ophthalomoplegia, dysarthria, loss of airway protection, paralysis of the respiratory muscles and the diaphragm, and finally limb paralysis with loss of deep tendon reflexes.

Postsynaptic neurotoxins are polypeptides with about 60–70 amino acids. Their relatively small-sized proteins contain four (neurotoxin I) or five (neurotoxin II) disulfide bonds. These bonds fold the peptide to three loops. One of the loops (B) probably attaches to the acetylcholine-binding site of the acetylcholine receptor. The receptor is a ligand-gated channel protein, allowing ions to pass through when activated. The postsynaptic snake neurotoxins attach to the two subunits (a glycoprotein) of the receptor and block the formation of the ion-channel.

The onset of symptoms commences one hour after the bite. Whereas the effect produced by the postsynaptic toxins can be blocked by a specific antivenom, the damage done to the presynaptic structures cannot be influenced. If the presynaptic effect is dominant it usually takes weeks to resolve. This means that long-time respiratory therapy may be necessary even when antivenom treatment was applied.

Among the Elapidae, mambas have a special venom containing two neurotoxic groups: dendrotoxins and fasciculins. Their action is different from the action of the already mentioned snake neurotoxins. *Dendrotoxins* affect the potassium channels in the terminal axon membrane, leading to a continuous release of acetylcholine and an over-stimulation of the muscle endplate. *Fasciculins* act as cholinesterase inhibitors in the junctional cleft, reducing the degradation of acetylcholine. Together the two actions lead to a surplus of acetylcholine similar to that in organophosphate-poisoning. The over-stimulation of the endplate leads to fasciculation and myocloni of the peripheral muscles with the consequence of the victim being unable to use the peripheral musculature, leading to respiratory paralysis.

The action of these two groups of toxins may be very fast, leading to symptoms within less than one hour. The fasciculins have a similar structure to the postsynaptic neurotoxins of other Elapids. They have an amino-acid chain consisting of 57–60 amino acids cross-linked by three disulfide bonds. Dendrotoxins have a molecular weight of around 7000. They have considerable sequence homology with X-bungarotoxin venom of the krait, which has a presynaptic action and is also a sodium-channel blocker.

Myotoxins

Snake venoms that cause damage to the musculature are phospholipase A_2 enzymes, which destroy skeletal muscles.

The myotoxins destroy muscle cells sparing the basement membrane. This allows regeneration of the muscle cells in due time (one month). As muscles are destroyed a release of myoglobin occurs; potassium and, as a diagnostic indicator for the damage, creatine kinase (CK) are released. Myoglobin damages the tubulus cells of the kidney, leading to secondary kidney failure which can be treated by hemodialysis. It is questionable whether antivenom can do any good if myolysis has already ensued.

Hematotoxins

The mechanism by which a mixture of enzymes in snake venom damages blood homeostasis is two-fold: On a cellular level the functioning and the amount of platelets are affected, and on a humoral level the blood coagulation is deranged. The major venom constituents are procoagulants, anticoagulants, platelet-aggregation inhibitors or promoters, as well as hemorrhagins that cause local damage to vessel walls.

A mixture of enzymes interacts via hemostasis. The strange effect of these components is that hypercoagulopathy with thrombosis and bleeding can happen simultaneously. The cause of the tendency to bleed lies in the procoagulant activation of Factor V, Factor X, Factor IX, and prothrombin (Factor II), thus using up fibrinogen.

In addition to the mechanism of using up fibrinogen, there is a toxin with a direct fibrinolytic action promoting fibrinogen degradation. Therefore, in many snake poisoning cases, especially by Crotalinae, a coagulopathy with no measurable fibrinogen is present. As the platelets are affected by platelet-aggregation inducers, results of laboratory testing [low platelets, no fibrinogen, high partial thromboplastin time (PTT) and thrombin time (PT)] may point to a disseminated intravascular clotting (DIC) which in fact is not happening.

In nearly all snake venoms are hemorrhagins, which are zinc metalloproteinases that cause local damage to vessel walls, leading to local bleeding and ecchymosis. Bleeding is enhanced by substances with anticoagulant action such as Protein C, thrombin inhibitor, and phospholipase A_2. The chemistry of all these diverse toxins ranges from small to large molecules that resemble the prothrombinase complex. Although after a snake bite, especially from Viperidae, severe coagulopathy is notable, spontaneous bleeding is rare. Difficult-to-control hemorrhage may happen after medical interventions such as puncture from an intravenous line, surgical debridement, or surgical cuts for treatment of swelling and compartmental syndrome. In these circumstances the substitution of clotting factors alone cannot stop the bleeding so that use of an antivenom is essential.

There is one venomous snake that has a particular nephrotoxin, the Russell's viper, which leads to acute renal failure that is not secondary to shock or myolysis.

Local damage

Most snake bites - with the exception of some Elapidae (e.g., Krait) - create more or less local damage (blisters, necrosis) around the bite due to the venoms' content of enzymes, kinins, leukotrienes, histamine, phospholipases, collagenases, and metallic ions.

Generally, within minutes after the bite, the region around the fang marks starts to swell and becomes painful. Within hours ecchymosis, blisters, and tissue necrosis develop and the swelling spreads over the whole limb, sometimes involving adjacent areas of the trunk. This can lead to compartmental syndrome, in that the swelling gets so bad that the circulation in the affected limb is impaired. Without surgical intervention this

could lead to the loss of the limb. The toxic components act at varying times. The venoms may remain local and be fixed to tissue but can reach the circulation as well and thereby cause systemic toxicity.

Systemic symptoms Usually the venom is injected subcutaneously during the snake bite. It spreads through lymphatic and superficial venous vessels and reaches the circulatory system, slowly inducing systemic signs. The rare intravascular envenomation produces systemic symptoms within minutes. Direct intravenous venom injection seldom happens but may account for the majority of fatalities by snakes, which are thought not to be toxic. In some cases the venom of the snake can lead to overt anaphylaxis with a fatal or near fatal reaction. Less severe systemic signs are weakness, nausea, malaise, and anxiety. More severe signs manifest as abdominal pain, vomiting, diarrhea, dyspnea, profuse sweating, salivation, metallic taste, and confusion. The most severe symptom is circulatory collapse with hypotension and tachycardia.

Management of Snake bites

First Aid

The 'traditional' first aid methods like incision or sucking the wound are ineffective. Instead, it is effective to immobilise the bitten limb and the patient to reduce the distribution of the venom.

The only approved first aid for a snake bite is to immobilise the bitten limb and keep the victim still. For bites of many elapids the Australian pressure immobilisation method is effective. It is able to retard venom transport in the lympathics by immobilising the limb on a splint. The limb is wrapped with an elastic bandage just hard enough to stop the transport of lymph and compress the superficial veins. Though this method has proven to be effective for the extremely toxic Australian elapids, immobilisation of the bitten limb on a splint makes sense for all snake bites to reduce the transport of the venom through the muscle pump.

In-hospital treatment

Hospital treatment requires careful medical history of the event, description of symptoms and swelling to judge the progression, neurological and hematological investigations, antivenom therapy and, if necessary, anticoagulant therapy and surgical intervention.

First of all a medical history has to be taken, not forgetting to ask for the current tetanus immunisation status and for known allergies. A careful examination – with photographs taken of the extremity that was bitten – has to be done. The diameter of the limb has to be measured to be able to compare with later measurements to thus judge the progression. The extent of the edema and skin discoloration can be marked by a pen to help recognise the development of the envenomation. A physical examination records the vital signs, and cardiorespiratory and neurological status. The disposition for

bleeding can be clinically seen by searching for blood in urine, blood in stool, and gingival petechiae.

The baseline laboratory tests cover a complete blood count including the platelets, electrolytes, urine analysis, blood urea nitrogen (BUN), glucose, international normalised ratio (INR), PTT, and fibrinogen. If normal, these tests have to be repeated within 6 hours. If there are any signs of circulatory depression, intravenous(IV) fluid replacement and, when this is not sufficient, vasopressors, have to be administered.

Antivenom Therapy The treatment with antivenom depends to some extent on the toxicity of the snake involved. Whereas for the European viper antivenom is only indicated if there are systemic signs and/or if the edema is progressing fast, an antivenom treatment for crotaline or elapid envenomations is more often necessary.

Besides antivenoms derived from horses, ovine-derived fragment antigen binding (Fab) fragments have been developed, which have fewer side effects than horse serum. 'CroFab' is available for the most common American crotaline and 'ViperaTab' for the most common European vipers. (see http://www.toxinfo.org).

Treatment of coagulopathy Crotalinae, far more than Viperinae or Elapidae, envenomation leads to severe coagulopathy. Coagulation can only be restored if sufficient antivenom is applied, as all blood products are neutralized by circulating snake venom. Bleeding calls for a great amount of antivenom, before fibrinogen and clotting factors are given. The same holds true for platelet transfusion, as platelets are destroyed by the venom as long as not enough antivenom is present.

Surgical intervention Surgical intervention should be reduced to a minimum as most snake bites, even with enormous swellings, can achieve full recovery. There are two reasons for surgical treatment. First, wound debridement around the bite is indicated if necrosis of the tissue has developed and hemorrhagic blebs and blisters have become large enough for resection. This usually occurs 3–6 days after envenomation. Second, a compartmental syndrome has to be decompressed by fasciotomy. Fasciotomy has to be instituted only if a high tissue pressure is found (by direct measurement), if there is no more blood flow through the limb, and if the neuromuscular transmission is interrupted.

Scorpions

Scorpion stings exhibit a high morbidity and mortality in India, the Middle East, North Africa, Brazil, Mexico, southern states of the USA, and in Central and South Africa. Scorpion venoms are both cardiotoxic and neurotoxic as a consequence of increasing neuronal sodium influx by interfering with the sodium channel. The initial symptoms are similar to poisoning by organophosphates.

Scorpions are the next most poisonous animals after snakes. Scorpion stings are common in various regions but not in Europe. In Mexico it has been estimated that 1000 people die as a result of 200 000 scorpion stings per year. In Saudi Arabia the fatality rate for scorpion stings is about 1.5%.

High mortality rates of up to 25% have been reported from India in patients who received only supportive care. The most toxic species are scorpions of the Buthus, Leiurus,

Androctonus, Tityus, Centruroides, and Parabuthus genera, of the Buthidae family. As a rule the thinner the claws and bigger the stingers are, the more toxic is the scorpion.

Venoms of the Scorpions Scorpion venoms are neurotoxic by influencing the vegetative and the central nervous system (CNS). Similar to snake venom, the constituents are a mixture of mucopolysaccharides, hyaluronidases, serotonin, histamine, protease inhibitors, histamine releasers, and amino acids. The local necrotising reactions are far less than in snake bites with the exception of stings of *Hemiscorpius lepturus*. Scorpion venoms increase neuronal sodium influx by interfering with the sodium-channel. This leads to an increased conductance at the voltage-gated calcium-channel in presynaptic nerve fibres with an increased release of neurotransmitters including acetylcholine. Similar to poisoning by organophosphates, first a heavy sympathetic stimulation with the release of epinephrine and norepinephrine ensues. This is followed by a parasympathetic reaction with low cardiac output. Salivation, abdominal pain, nausea, and vomiting are common signs of this vagal reaction. Pancreatitis can occur.

The nicotinergic neuromuscular transmission is manifested as tongue and muscle fasciculations and involuntary movements of skeletal muscles. The catecholamine over-stimulation leads to hypertension creating lung congestion or even pulmonary edema. Myocardial ischemia and even infarction can occur. Central nervous system effects may be due to hypertensive encephalopathy as well as to a direct central mechanism of toxicity. Agitation, hyperthermia, seizures, and even coma may be the signs and symptoms of this evolvement. Whereas most scorpion species exhibit both cardiotoxicity and neurotoxicity there are some scorpions of the Centruroides genus that are merely neurotoxic.

Clinical Presentation of Scorpion Stings Scorpions sting humans most often in the extremities after hiding in clothing. Children are more often stung than adults and show more serious reactions than adults, especially if they are under the age of 10. The severity grade of scorpion envenomation depends on the scorpion's species, age, and size. Very often local reactions are the only symptoms.

Paresthesia and pain at the sting site are usual. In more severe cases these symptoms spread out over the limb involved and can even reach the perioral area. Systemic signs and symptoms comprise stimulation, followed by depression of the CNS, hypertension, lung edema, dysrhythmia, and rarely shock. Cranial nerve palsy and neuromuscular dysfunction like involuntary movement, restlessness, and shaking can occur. Parasympathetic stimulation leads to gastrointestinal over-stimulation and may cause priapismus. In the circulatory system the sympathomimetic effects prevail. This is reflected by tachycardia and hypertension; bradycardia and hypotension can happen but are rare. In cases with fatal outcome, cardiac and respiratory failures are the predominant symptoms. A dilated cardiomyopathy with sequelae is found in a few cases. Scorpion stings may produce some laboratory abnormalities.

Hyperglycemia, leucocytosis, and C-reactive protein (CRP) elevation are commonly seen. The latter are due to an inflammatory reaction caused by interleukin release. There maybe a transient elevation of creatine kinase muscle brain (CKMB) and Troponin T due to myocardial ischemia which can be diagnosed by ischemic signs in the electrocardiogram (ECG). It was possible to demonstrate cardiac dysfunction in echocardiography, showing a decreased ejection function.

Therapy of Scorpion Stings A light constricting bandage proximal to the wound to prevent lymphatic spread, and application of ice at the sting site, slowing venom absorption and relieving pain, is generally recommended. In cases of systemic signs and symptoms an antivenom should be used. Nevertheless, the use of antivenom is still controversial as supportive care has improved over the years and the death rate has come down in the last 40 years.

Spiders

All spider venoms are complex mixtures consisting of toxic peptides and proteins, but most spiders are unable to penetrate human skin in their bites, which therefore do not lead to envenomation. For humans the most toxic is the Australian funnel-web spider.

Taxonomy divides spiders into the orders Mygalomorphae and Araneomorphae. Within the Mygalomorphae is the Australian funnel-web spider (*Atrax robustus*) the most toxic spider for humans. All the other spiders that are toxic for humans belong to the Araneomorphae, namely *Latrodectus* spp. (widow spiders), *Phoneutria* spp. (banana spiders), and *Loxosceles* spp. (recluse spiders).

Spider venoms are, as snake and scorpion venoms, complex mixtures consisting of toxic peptides and proteins. There is a great divergence from one spider venom to the other. Some venoms are very potent neurotoxins, others create an endogenous catecholamine storm, some contain necrotoxins which can destroy large skin areas of the human body.

Australian Funnel-web Spiders The *Atrax robustus*, the Sydney funnel-web spider, has caused the most severe bites and fatalities of all spider bites.

The mouse spiders (*Missulena*) have venom similar to the above-mentioned spider but only a few severe human envenomings occur. The venom of these spiders has a neurotoxic component which is potentially lethal for humans but does little harm to other mammals. The venom of the male spiders is more toxic than that of the females. The neurotoxic component is robustoxin. It is a protein of molecular weight 4854 Da with 42 amino acid residues and four disulfide bridges.

Clinical signs and symptoms Since the spider is large, and, after biting, hangs on to the victim, the bite is usually recognised. Bite marks are present and the bite is painful.

Similar to some scorpion stings perioral paresthesias and tongue fasciculations develop. This is followed by a sympathomimetic phase, as the catecholamines are forced out of their repositories, including tachycardia, hypertension, piloerection, cardiac arrhythmias, and pulmonary edema, which is partly due to left ventricular congestion and partly of neurogenic nature. As in scorpion stings muscarine-cholinergic over-stimulation can ensue with hypersalivation, hyperlacrimation, increased sweating – first around the bite, later generalised – with nausea, vomiting, and abdominal pain. The nicotinic-cholinergic hyperactivity leads to muscle fasciculation and involuntary movements. As soon as catecholamine depletion becomes manifest, the second life-threatening phase is reached in which excitatory effects come to an end, secretion stops, and hypertension resolves only to progress into terminal circulatory failure, apnea, and cardiac arrest.

Treatment of spider bites by Australian funnel-web spiders There are bites by the Australian funnel-web spiders which cannot be survived without the use of intravenous antivenom treatment. On the Australian market an antivenom is available against the male *A. robustus* venom. It is a rabbit IgG with the name FWSAV.

Latrodectus Spiders

***Latrodectus* spiders are found around the world, usually between 50° N and 45° S. They are very rarely found in central Europe. 2500 Cases of latrodectus bites are reported to US poison-centres per year. In Australia 5000–10 000 bites occur annually. The venoms contain high-molecular-weight proteins that cause release of transmitters in various synapses.**

There are about 50 species of the genus *Latrodectus* of which about 9 are of medical importance. In North America: *L. mactans*, *L. hesperus*, *L. bishopi*, and *L. variolus*; in Australia: *L. hasseltii* (red back spider); in New Zealand: *L. katipo*; in South Africa: *L. indistinctus*; around the world: *L. geometricus* (brown widow) and *L. tredecimguttatus*. The male spiders have a small biting apparatus and therefore cannot envenomate humans, whereas the female can. The *Latrodectus* venoms contain high-molecular-weight proteins. Some of the proteins have been cloned and sequenced. Alpha-LTX is produced by all *Latrodectus* species, it has a molecular weight of 150 000 Da, and seems to be the constituent of the venom that is responsible for human envenomation.

Alpha-LTX causes massive release of transmitters in various synapses and is effective not only at cholinergic but also at noradrenergic and glutaminergic sites due to a depletion of the synaptic vesicles. The main action of *Lactrodectus* venom is mediated by the binding of alpha-LTX to the presynaptic CIRL (Ca^{2+}-independent receptor). CIRL is believed to be coupled to phospholipase C interfering with the phosphoinositide metabolism, which is responsible for secretion.

There are several CIRLs 1–3, with different affinities to alpha-LTX. CIRL 1 and CIRL 3 are high-affinity receptors in neuronal tissue and in many other organs such as the kidney, spleen, ovary, heart, lung and brain. CIRL belongs to the family of 7-transmembrane domain G-protein-coupled receptors. The nervous system is the first target for alpha-LTX but other tissues can be involved due to the presence of CIRLs.

Signs and symptoms of latrodectism

Within 60 minutes after a bite local pain, local sweating, and piloerection may occur. Only a quarter of the bitten victims experience systemic toxicity such as muscle pain, which in severe cases may become generalised and in very severe cases require treatment with analgesics.

A pain similar to a pinprick is what the victim feels if bitten by a *Latrodectus* species. A pair of red spots with a halo can be seen at the site. Very often the bite remains unnoticed. The local reaction at the bite site is uneventful until 60 minutes after the bite. Within that time a pathognomonic triad appears: Local pain, local sweating, and piloerection. It may take hours before it is fully developed.

Muscle pain starts around the bite and spreads over the extremity involved. Lymph tenderness of the regional nodes evolves. In severe cases the pain becomes generalised. The pain is felt in the neck, chest, abdomen, lower back, and thighs. The pain can become very severe and mimic that of acute peritonitis. Severe diaphoresis is present. Dysphoria, restlessness, horror, and fear of death torment the patient. The victims show - if it comes to a fully developed latrodectism - a fascia latrodectismica which means trismus, flushing, grimacing, and blepharoconjunctivitis. Symptoms peak within 12 hours after the bite and resolve within 24 hours. Symptoms may last for a week, although this is rare, and muscle weakness may be present over months. Laboratory tests exhibit leucocytosis, hyperglycemia, and elevation of CK and lactic dehydrogenase (LDH), which are of course nonspecific.

Management of Latrodectism The treatment of a patient who was bitten by a spider of the *Latrodectus* genus depends on the severity of symptoms. It is most important to treat the patients with analgesics to relieve pain. Since muscle cramping plays a prominent role in lactrodectism the use of benzodiazepines has been recommended. The indication for the use of antivenom is given if hypertension or pain is not manageable by the above-mentioned treatment.

Loxosceles Spiders

***Loxosceles* venom is one of the most potent natural toxins containing necrotising and hemolysing substances. Most bites occur in South and Central America.**

Loxosceles spiders are found throughout the world. *L. reclusa* the brown recluse spider, lives in the southern states of the US. *L. intermedia* and *L. laeta* are found in South Africa. *L. rufescens* resides in the Mediterranean regions of Europe and Africa. Very few bites happen in Europe, some in North America, and a lot in South and Central America.

So far, 11 major components of the venom have been identified. The most important subcomponents are hyaluronidase, deoxyribonuclease, ribonuclease, alkaline phophatase, lipase and, last but not least, sphingomyelinase-D. Hyaluronidase works as a spreading factor that allows the venom to penetrate tissue. Sphingomyelinase-D (MW: 32 000) is the necrotising and hemolysing substance. Sphingomyelinase-D stimulates platelets to release serotonin and interacts with the membrane of cells, which contain sphingomyelin. This leads to a release of hemoglobin, choline, and *N*-acylsphingosine phosphate from red blood cells. The resulting hemoglobinemia can cause kidney failure. The other substances trigger a cascade of reactions by releasing inflammatory mediators such as thromboxanes, leukotrienes, and prostaglandins and by the activation of neutrophil granulocytes. The sequelae of this release is vessel thrombosis, tissue ischemia, and skin necrosis. Complete hemolysis is induced by the activation of a metalloproteinase, leading to a cleavage of glycophorins from the erythrocyte surface, mediating complement activation. The hemolytic process is maintained by transfer of the toxins from one erythrocyte to another, attacking even newly synthesized erythrocytes. Cytokinin activation by the toxic proteins of the spider venom can evoke an endotoxin shock resembling a septic shock.

A typical histopathology develops around the lesion of the bite. It starts as perivasculitis with a polymorphonuclear inflammation, followed by local hemorrhage and edema

with epidermal necrosis and ulcerations. A panniculitis with subcutaneous fat necrosis is the next step in this development together with arterial wall necrosis. The ulcerations become eschar-covered. Slow resolution of the lesion leads to scarring of the dermis and subcutis. Skin transplantation is sometimes needed after several weeks.

Symptoms of Loxosceles spider bites *Loxosceles* bites can be classified in 3 severity grades:

1. A bite with little venom injected leads to a small erythematous papule associated with a localised urticarial reaction.
2. A cytotoxic reaction from a painless bite that evolves to blistering, local bleeding, and ulcerations after 2–8 h. Within 3 days a violaceous necrosis develops surrounded by ischemic blanching of the skin with an outer erythema (blue, white, red: French flag). The central blister becomes necrotic within 3–4 days with eschar formation between 5–7 days. After several weeks the ulceration starts to heal by secondary intention.
3. A systemic manifestation with fever, chills, general edema, vomiting, arthralgia, petechial bleeding, rhabdomyolysis, disseminated intravascular clotting (DIC), and renal failure.

Treatment of Loxoceles bites Treatment is according to severity grading. Grade 1 doesn't need treatment, besides tetanus vaccination – if necessary – and observation for several days. Grade 2 needs general wound care with immobilisation of the extremity that is involved. Stimulation of natural healing by wound debridement and delayed primary wound closure is a useful surgical measure. Antibiotics should be used to treat cutaneous or systemic infections. For systemic signs and symptoms (grade 3) there is an antivenin available but no clinical studies about its side effects or effectiveness are available.

6.7.3 Plant Toxins

Introduction

Plant chemistry is most variable and one plant can contain multiple compounds that act and interact together. This makes a grouping of the constituents of plants necessary to get an idea of the underlying chemistry and action.

The occurrence of plant poisoning varies from country to country. In Europe there are many cases involving children who eat berries or other parts of a plant in summer, leading to diarrhea or vomiting; whereas in Sri Lanka there are deadly plant poisonings from yellow oleander occurring when people commit suicide. A classification of plant toxins in 5 groups allows differentiation among the most toxic substances.

The major plant toxins are alkaloids, glycosides, phenols and phenylpropanoids, lectins, and terpenes.

Classification of Plant Toxins

Alkaloids Alkaloids are bases containing at least one nitrogen and their structure is heterocyclic. The best known alkaloids belong to the *tropane* group consisting of

atropine, hyoscyamine, scopolamine, apoatropine, and mandragonine (Belladonna alkaloids). Most illegal drugs stem from alkaloids like codeine, morphine, cocaine, psilocybin and LSD. Some natural stimulants like nicotine and caffeine (theine) are alkaloids as well. The most severe intoxications with alkaloids happen with aconitine, colchicine, and strychnine.

Glycosides Glycosides consist of a sugar or sugar derivative (glycone) and a nonsugar moiety called an aglycone. The aglycone allows further classification into *saponin glycosides*, *cyanogenic glycosides*, *anthraquinone glycosides, atractylosides*, and *salicines.* The best known glycosides are the cardioactive *saponins* like digoxin and digitoxin. Intoxication with naturally occurring plant toxins of this group are poisonings with oleander (*Nerium oleander*), lily of the valley (*Convallaria majalis*) and squill (*Urginea maritima*). Not all *saponin* glycosides are cardiotoxic. There are, for example, saponin glycosides in the holly berry (*Ilex*), which induces gastrointestinal symptoms, and glycyrrhizin in the licorice (*Glycyrrhiza glabra*) that inhibits 11β-hydroxysteroid dehydrogenase, producing pseudohyperaldosteronism. *Cyanogenic glycosides* are amygdalin, cyacosin, linamarin, and sambunigrin that can under certain circumstances lead to acute or, more likely, to chronic cyanide poisoning. *Anthraquinone glycosides* are laxatives. *Atractyloside* is hepatotoxic; *salicin* is hydrolysed to salicyl alcohol.

Phenols and Phenylpropanoids Phenylpropanoids have a phenyl ring with a propane side chain. There is no nitrogen in the molecule. This group includes the *coumarins* with a lactone side chain, the *flavonoids* built around a flavan 2,3-dihydro-2-phenylbenzopyran nucleus, the *lignans* with linked double phenylpropanoids like podophyllin, the *lignins*, which are polymers of lignans, and *tannins* with phenyl hydroxyl groups that can condense with proteins.

Lectins Lectins are glycoproteins with different carbohydrate ligands like galactosamines. The most toxic lectins are the *toxalbumins* like ricin and abrin. Ricin has an A- and a B-chain with, respectively, 267 and 262 amino acids linked by a disulfide bridge. The sugars are lactose, galactose, and *N*-acetylgalactosamine.

Terpenes Terpenes are plant constituents of the essential oils. The basic structure of terpenes is the hydrocarbon isoprene (CH_2=CH-(CH_3(-CH=CH_2). The number of isoprene units is used for the classification into monoterpene (2 isoprenes), sesquiterpene (3 isoprenes), diterpene (4 isoprenes), sesterterpene (5 isoprenes), triterpene (6 isoprenes), and tetraterpene (8 isoprenes). There may be many functional groups attached, like alcohols, phenols, ketones, and esters (terpenoids). Toxic terpenes and terpenoids are, for example, ginkgolides, kava lactones, thujone, anisatin, ptaquiloside, and gossypol.

Alkaloids

Belladonna Alkaloids

Atropine, the best known representative of these alkaloids, is a competitive muscarinic receptor antagonist. It competes with acetylcholine and other muscarinic agonists for a common binding site on the muscarinic receptor.

Belladonna alkaloids show a strong anticholinergic effect by competitive inhibition at the muscarinic receptor. The receptor is activated by acetylcholine or blocked by atropine. There are 5 different muscarinic receptors located in different places of the body: M_1–M_5. All known muscarinic receptors belong to the family of G protein-coupled receptors. M_1 receptors are found in the cerebral cortex, the hippocampus, and in ganglia, in the GI-tract, and in parietal cells of the stomach. M_2 receptors are present in the conducting tissue of the heart and in presynaptic terminals. M_3 receptors exist in exocrine glands like lacrimal and sweat glands and in smooth muscles including the endothelium of vessels. They are also seen in lymphocytes. M_4 receptors dwell in the cerebral cortex, amygdala, and hippocampus. Outside the brain they are in keratinocytes of the skin. M_5 receptors can be detected in keratinocytes and lymphocytes.

Clinical signs of poisoning by belladonna alkaloids The anticholinergic syndrome caused by belladonna alkaloids includes peripheral and central effects. Peripheral muscarinic receptors differ in their sensitivity to antagonists. First, excretory glands are antagonised, decreasing salivary and bronchial secretion and sweating. This is followed by mydriasis, loss of accommodation, and tachycardia. Finally, urinary sphincter, intestinal, and gastric motion are influenced, leading to atony of these organs. Atropine causes peripheral anticholinergic signs at lower doses; at higher doses central anticholinergic effects are added. They start with stupor, followed by delirium with restlessness and hallucinations. These hallucinations are usually visual, less often auditory. Seizures are common signs of anticholinergic poisoning. The inhibition of sweating, which reduces the ability to dissipate heat, leads to hyperthermia. Hyperthermia together with increased muscle activity can create rhabdomyolysis with kidney failure.

A very high dose of atropine can result in immediate respiratory failure with a deep coma and shock; death may ensue within a short time.

The cardiac effects of belladonna alkaloids may consist of abnormal conduction, such as bundle branch block and atrioventricular dissociation besides sinus tachyarrhythmia. Rarely, ventricular fibrillation can be induced. Treatment includes sedation with benzodiazepines and the use of physostigmine. In severe poisoning deep sedation has to be induced, which makes mechanical ventilation mandatory.

Aconitines

Aconitines include the three (19 diterpenoid-ester) alkaloids aconitine, mesaconitine, and hypoaconitine. These alkaloids influence the sodium channel in the conductive tissue of the heart as well as in central and peripheral nerves. The sodium influx through these channels is increased, increasing inotropy and at the same time delaying repolarisation, promoting premature excitation.

Signs of cardiotoxicity are bradycardia and later on ventricular dysrhythmia. Symptoms can occur 5 minutes to hours after ingestion. Early signs of peripheral neurotoxicity are paresthesias around the mouth and of the oral mucosa, followed in severe cases by progressive skeletal muscle weakness. Vegetative symptoms consist of nausea, vomiting, diarrhea, and hypersalivation. CNS involvement is manifested by seizures. There is no real antidote for aconitine. Atropine may be useful to treat bradycardia and

hypersalivation. For ventricular dysrhythmias sodium-channel blockade – by lidocaine, flecainide or, probably safest, amiodarone – is recommended.

Strychnine Strychnine alkaloid is found in the seed of the strychnine tree (*Strychnos nux-vomica*). It was isolated in 1818 and is a colourless crystalline powder with an extremely bitter taste.

In the past, strychnine was used in many medical preparations as a digestive stimulant, as an analeptic, and as an antidote for barbiturate overdose. It caused many fatal poisonings in the 1920s and 1930s and has become a very rare cause of intoxication in the present western world. It is still a common poisoning agent in Asia often occurring from unintentional overdose of a Chinese herbal medicine.

The lethal dose of strychnine is around 50–100 mg but much higher doses have been survived with proper treatment.

Strychnine is well absorbed. Protein binding is minimal. It is quickly distributed to peripheral tissue and has a volume of distribution of about 13 l/kg. Strychnine is metabolised by hepatic P450 microsomes. The main metabolite is strychnine *N*-oxide. The half-life amounts to 10–16 h.

Mechanism of toxicity of strychnine

Strychnine competitively inhibits the binding of glycine to the α-subunit of a glycinergic chloride channel. Although glycine receptors are found in the whole nervous system, the predominant target of strychnine is the spinal cord.

The inhibition of the glycinergic receptor by strychnine interrupts the control of the neurons in the ventral horn of the spinal cord by muscle tone. The result is an increase in impulse transmission to the muscles manifested by generalised muscular contractions.

This is a similar action as seen in tetanus where a blockade of glycine release is due to the toxin of *Clostridium tetani* (tetanospasmin).

Clinical signs and symptoms of strychnine poisoning After ingestion of strychnine, symptoms commence within one hour. The characteristic signs of poisoning are generalised involuntary muscle contractions. Each period of musculature contraction lasts for 1–2 min. Owing to different strengths of various muscle groups flexion and extension are asymmetric, leading to opisthotonos, flexion of the upper limbs, and extension of the lower limbs. Other symptoms are hyper-reflexia with convulsions, nystagmus, and hyperthermia up to 43 °C as consequence of muscular activity. Loss of consciousness can follow due to severe metabolic acidosis as lactate is produced in surplus by muscle contraction.

Late fatality is a consequence of multi-organ failure induced by hyperpyrexia plus hypotension.

Treatment of strychnine poisoning The primary aim of treatment is to stop the muscular hyper-reactivity, thereby preventing metabolic acidosis and respiratory impairment, by administration of benzodiazepines. Hyperthermia requires aggressive cooling with ice-water, cooling blankets, and fanning. In severe cases hemodialysis is the treatment of choice as it resolves metabolic acidosis and allows rigorous cooling.

Colchicine The name colchicine can be traced back to the king of Colchis, a kingdom in Asia Minor. Colchicine is found mainly in two plants, *Colchicum autumnale* (autumn

crocus) and *Gloriosa superba* (glory lily). Colchicine is readily absorbed in the jejunum and ileum. Owing to a strong first-pass metabolism the bioavailability is about 25%.

Plasma elimination half-lives follow a two-compartment model with a distribution half-life of between 90–100 min and a delayed terminal elimination half-life of 2–30 h. Colchicine can be found in leukocytes or urine even ten days post exposure. Protein binding is circa 50%. Colchicine is found in blood cells in 5–10-times higher concentration than in plasma. Peak plasma concentration occurs between 1–3 h after ingestion. Toxic effects are only found if plasma concentrations are above 3 µg/l.

Pathophysiology of colchicine poisoning

Colchicine inhibits microtubulin formation by binding to tubulin, resulting in its conformational change, which affects microtubule growth. The impaired formation of the microtubule spindle in metaphase leads to cellular dysfunction and cell death. Colchicine competitively inhibits the γ-aminobutyric acid A ($GABA_A$) receptors in the brain and impairs leukocyte mobility and adhesiveness.

The best known pathomechanism is the inhibition of microtubule formation, which is responsible for cellular mitosis. Microtubules also take part in the maintenance of cellular structures and transport within cells. Colchicine binds to subunits of the tubules and alters tubulin secondary structure. These conformational changes in the tubules prevent binding of the next tubule subunit and thereby interrupt microtubule growth. These alterations result in a lack of conformation of the microtubule spindles in the metaphase of mitosis, leading to cellular dysfunction and cell death.

The effects of colchicines are dose dependent and reversible. High doses disintegrate microtubules, low concentrations prevent new formation of microtubules. Colchicine also inhibits intracellular granule transport, which is mediated by microtubules. As shown from animal experiments colchicine might also be able to inhibit DNA synthesis.

Colchicine can furthermore influence the $GABA_A$ receptors in the brain by competitive inhibition, which might be responsible for some of the neurological symptoms seen in colchicine poisoning.

As colchicine accumulates in leukocytes it exhibits an inhibitory effect on leukocyte mobility and adhesiveness. It reduces expression of adhesion molecules on endothelial cells and inhibits cytokine production in polymorphonuclear leukocytes. The last mentioned mechanisms are at therapeutic doses responsible for the anti-inflammatory action of colchicine in the treatment of gout or familial Mediterranean fever (FMF).

Symptoms of colchicine poisoning There are three phases of colchicine poisoning. Within hours following ingestion there is gastrointestinal disturbance, within a week organ dysfunction follows, with possible death from cardiac or lung failure or septicemia due to bone-marrow suppression. Myopathy, neuropathy, and a combined myoneuropathy can result in respiratory insufficiency. During recovery sequelae like alopecia and myoneuropathy are seen after a week to several months.

Treatment of colchicine poisoning Besides cardiovascular monitoring, treatment with catecholamines is indicated if lowered blood pressure is not responsive to fluids. In such a

case multi-organ failure is imminent and early endotracheal intubation and artificial respiration becomes necessary. In case of a very likely infection a calculated treatment with antibiotics is warranted. Hemodialysis is needed if kidney failure is occurring.

Glycosides

Cardioactive Steroids The best known cardiac glycosides are digoxin and digitoxin because they were used for centuries to treat cardiac insufficiency. Digitoxin stems from the purple foxglove (*Digitalis purpurea*) whereas digoxin and digitoxin are found in the Grecian and woolly foxglove (*D. lanata*). Similar digitaloids are contained in oleander (*Nerium oleander*) as oleandrin, in the yellow oleander (*Thevetia peruviana*) as thevetin A plus B, in the lily of the valley (*Convallaria majalis*) as convallatoxin, convalloside, and convallotoxol, in squill (*Urginea maritima*) as scillaren and scillarenin, and finally in Christmas rose (*Helleborus niger*) as helleborin. The most toxic of them is the yellow oleander: 4 seeds can kill an adult. The next most toxic one is the oleander with high morbidity in children; again the seeds have the highest concentration of glycosides.

Leaves and flowers of the lily of the valley have a low glycoside concentration. The red fruits taste very bitter and the seeds with the highest toxin concentrations are too hard to be bitten open. Poisoning with foxglove is very rare probably due to the bitter taste of these plants. Even after the ingestion of 6 leaves in a suicidal attempt only a mild intoxication was seen. There are no known severe cases of Christmas rose poisonings.

Pathophysiology of cardiac glycosides

Toxicity of cardiac glycosides is due to the same mechanism as their therapeutic action. They bind to specific subunits of the sodium–potassium ATPase in all sorts of muscles, resulting there in an inhibition of the Na^+ –K^+ pump. Potassium is no longer pumped into the muscle cells; in severe poisoning hyperkaliemia can be seen. Since sodium and calcium are exchanged, more calcium is taken up intracellularly, resulting in increased contractility. Excessive increase of intracellular calcium concentration results in elevation of the resting potential, producing myocardial sensitisation and predisposes to dysrhythmia. Depolarisation is delayed.

Within the CNS cardiac glycosides interact with the vegetative nervous system. At therapeutic levels the vagal tone is increased, at toxic levels the sympathetic system prevails, which can cause an increase in automaticity and life-threatening dysrhythmias.

Symptoms of cardiac glycoside poisoning The noncardiac manifestation of cardiac glycoside poisoning includes gastrointestinal, neurological, and visual symptoms. The cardiac toxicity is manifested by conduction abnormalities and dysrhythmia. The disturbed conduction includes bradycardia, sinoatrial block, and I–III degrees of

atrioventricular block. Patients usually die in ventricular fibrillation not responsive to defibrillation.

Laboratory findings in cardiac glycoside poisoning Serum digitoxin and digoxin glycoside concentration can be measured by different immunoassays. If cardiac glycosides from plants are ingested the immunoassay gives positive results due to a cross-reactivity within both immunoassays. No nontoxic or toxic levels are defined for these glycosides. Any measurable value after a plant ingestion means that glycosides were within the plant. Therapeutic levels of digoxin or digitoxin found after such plant ingestion indicate a severe intoxication by these naturally occurring cardiac glycosides.

Treatment of cardiac glycoside poisoning For gastrointestinal decontamination a single dose of 10 grams of activated charcoal can be administered within 2 h of ingestion. Gastric lavage is not advised as it can trigger dysrhythmias.

Since digoxin-specific fragment antigen binding (Fab) fragments have become available, all previous treatments such as the administration of potassium, phenytoin, lidocaine, or an external pacemaker are no longer necessary.

For I–II degree AV (atrioventricular)-block atropine can still be used. Whereas for digitoxin or digoxin poisoning the dosage of digitalis antidote is well defined if the amount taken in or the serum levels are known, the dosing for plant digitaloid intoxications has to be pragmatic.

Cyanogenic Glycosides Cyanogenic glycosides are found in about 2500 plants. The best known is amygdalin in bitter almonds (*Prunus* spp) but also in seed kernels of apricots, peaches, pears, apples, and plums. Others are prunasin, linamarin, dhurin/dhurrin, sambunigrin and cyacosin. The chemical structure of these compounds consists of mono- or disaccharides with a benzaldehyde cyanohydrin attached to it. On digestion the saccharides are split off by β-glucosidase and the benzaldehyde is freed. Cyanide is produced by further hydrolysis.

Eating the fruits or leaves can lead to cyanide poisoning. Treatment is very seldom needed as the natural detoxification is faster than the liberation of cyanide. Treatment of cyanide poisoning with sodium thiosulfate or hydroxocobalamin is theoretically possible. (Treatment of cyanide poisoning see Chapter 6.6). A chronic sort of poisoning by cyanogenic glycosides is quite common in East Africa from the vegetable cassava (*Manihot esculenta*), which contains linamarin. The disease caused by the ingestion of cassava is called konzo. The manihot is usually dried by oven or by sun exposure. Mass poisoning happened when, due to food shortages, there was insufficient drying of this traditional food. Symptoms of the disorder are a symmetric spastic paraparesis and hypothyroidism with the development of a goitre.

Phenols and Phenylpropanoids

From this group of plant toxins only podophyllin can be discussed here. Podophyllin or podophyllotoxin is a phenylpropanoid, the main toxic compound in the rhizome and roots of may apple (*Podophyllum peltatum*) and wild mandrake (*Podophyllum emodi*). Podophyllin in purified form is still used as ointment for the treatment of anogenital warts. It is also used to treat oral hairy leukoplakia.

Pathophysiology of Podophyllin Poisoning

Podophyllin, similar to colchicines, binds reversibly to tubulin at the same site as does colchicine. This leads to mitotic arrest. Since the mitotic spindles cannot be formed without microtubules the metaphase comes to a halt and the chromosomes become clumped. Podophyllotoxin inhibits nucleoside transport into the cell nucleus. Disruption of the microtubules hampers all sorts of intracellular transport mechanisms.

Signs and Symptoms Intoxications are reported after ingestion of wild mandrake root, herbal remedies containing podophyllin, and from topical application of podophyllin. The symptoms start within several hours with nausea, vomiting, abdominal pain, and diarrhea.

Compared with colchicine poisoning podophyllin exhibits more CNS symptomatology and less delayed hematologic toxicity.

Patients rapidly develop confusion and even coma. Auditory and visual hallucinations are found early in the course of the poisoning. Paresthesia, loss of tendon reflexes, and pyramidal tract signs can be seen. The cranial nerves are included in the neuropathology with diplopia, nystagmus, dysconjugate gaze, and facial nerve paralysis.

As in colchicine poisoning a peripheral sensorimotoric axonopathy with slow, or no, recovery can develop. After an initial leukocytosis a pancytopenia with its nadir around day 7 can develop. As a complication infection can develop. Treatment is the same as described for colchicine poisoning.

Lectins

Lectins are glycoproteins. The most important are the toxalbumins such as ricin and abrin. They are so toxic that they even might be used as biological weapons. Abrin is found in prayer beans (*Abrus precatorius*), ricin in castor beans (*Ricinus communis*). Further plants containing toxalbumins are the black locust tree (*Robinia pseudoacacia*) and the European mistletoe (*Viscum album*).

Abrin and Ricin

Pathophysiology of toxalbumin poisoning Ricin and abrin are the two main toxalbumins. They consist of an A and B chain bound by a disulfide linkage. The molecular weights are about 65 000 Da. Cleaving the disulfide bond eliminates toxicity. The B chain binds to a galactose-containing receptor in the cell wall and both chains are internalised by the cells thereafter.

Toxalbumins disrupt DNA synthesis, and thereby protein synthesis, at the 60 S ribosomal subunit. An agglutinin in the *R. communis* plant can stimulate hemagglutination and intravascular hemolysis by binding to glycoproteins on red blood cell surfaces. The hemolysis can lead to kidney failure. Abrin and ricin can cause multi-organ failure. They can damage the endothelium of the alveolar space, leading to pulmonary edema. Liver failure can result from direct hepatotoxicity. Direct toxin contact with gastrointestinal mucosa results in ulceration, fluid loss, and bleeding. Damage to all parenchymatous organs was seen in for instance, the pancreas, spleen, liver, and kidney.

Symptoms of toxalbumin poisoning Oral ingestion of abrin or ricin creates a biphasic toxidrome. In the first stage – usually 2 hours after ingestion – patients have nausea, vomiting, diarrhea, and blood in feces. Even corrosive burns can be found in the throat and stomach. Thereafter more systemic symptoms develop like circulatory problems with signs of hypovolemic shock. Multi-organ involvement follows – partially through a direct toxic effect on the different organs, partially due to the shock itself. Liver, kidney, heart, and lungs are involved. Cardiac dysrhythmia and epileptic convulsions can be seen. The second phase develops over 3–4 days.

Treatment of toxalbumin poisoning Early administration of activated charcoal may be helpful. When diarrhea has started fluid resuscitation has to be commenced with infusion of electrolytes and fluid. Daily, or even twice daily, laboratory testing for liver enzymes, BUN, creatinine, clotting parameters, and blood count is necessary. If, after volume replacement, the blood pressure stays low, catecholamine infusion has to be given for circulatory support. Pulmonary edema and prolonged shock call for intubation and ventilation. Massive hemolysis requires transfusion and alkalinisation to keep up urinary output. If these measures are not successful, hemodialysis may be required.

Terpenes and Terpenoids

This group of plant toxins, though structurally similar, have quite different effects. Ginkgolides from *Ginkgo biloba* have an antiplatelet aggregation effect. Kava lactones found in kava kava (*Piper methysticum*) cause CNS effects and may be hepatotoxic. Ptaquilosides found in bracken fern (*Pteridium aquilinum*) lead to hemorrhage following a severe thrombocytopenia.

Kava lactones, thujone, and anisatin (*Illicium* spp.) are terpenes or terpenoids that have an effect at $GABA_A$ and $GABA_B$ receptors. They are all associated with seizures. Thujone – the main ingredient in absinthe, an extract from wormwood (*Artemisia absinthium*) – creates CNS excitation followed by depression and seizure. Thujone has a similar structure to tetrahydrocannabinol (THC), which might explain its psychoactive effects. Chronic intake of thujone produced a clinical picture of absinthism with hallucinations, cognitive impairment, and personality change. Thujone has an α and a β isomer. The toxic one is the α-stereoisomer that antagonises the $GABA_A$ receptor at the picrotoxin site of the chloride-channel, leading to neuroexcitation and seizure.

Treatment of Terpene and Terpenoid Poisoning The treatment is symptomatic. In case of hemorrhage, blood transfusion or platelet transfusion has to be performed. For hepatic failure in kava kava poisoning liver transplantation has to be undertaken in some cases. The acute CNS effects such as excitation, hallucination, and seizures can be treated with benzodiazepines dosed as necessary.

6.7.4 Summary

Animal Toxins

Snake bites, spider bites, and scorpion stings can cause severe morbidity and even mortality in humans. Four families of snakes are responsible: Viperidae, Elapidae,

Colubridae, and Atractaspididae. The most toxic snakes belong to the Viperidae, especially the sub-family of Crotalinae (rattlesnakes) and the Elapidae.

All these snake families produce neurotoxins, myotoxins, hematotoxins, and necrotoxins. Most elapid bites cause neurotoxicity, most crotaline bites result in coagulopathy. Most snake toxins can produce mild to severe tissue damage and systemic symptoms. The neurotoxins act presynaptically and postsynaptically. The myotoxins destroy muscle cells sparing the basement membrane. The hematotoxins induce coagulopathy, thrombocytopenia and hemolysis. Systemic symptoms are gastrointestinal and cardiovascular. Antivenom therapy is available but well tolerated antivenoms are not on the market for all poisonous snakes.

Scorpion venoms are neurotoxic inducing little local tissue damage. Their toxins stimulate acetylcholine release, leading to cholinergic crisis.

The most toxic spiders are the Australian funnel web spiders, the widow spiders and the recluse spiders. The funnel web spider's venom leads to a catecholamine storm, the widow spiders cause a massive release of different neurotransmitters like norepinephrine, acetylcholine, and glutamine. These neurotransmitters induce generalized pain, sweating, piloerection, dysphoria, and restlessness. Recluse spider bites are necrotizing and lead to extended skin necroses. They can create disseminated intravascular clotting and hemolysis. Antivenoms are available for some scorpions and spiders but their effectiveness is not well studied.

Plant Toxins

Plant toxins can be classified either by their chemistry or their action. The toxic components belong to 5 different chemical groups: Alkaloids, glycosides, phenols, lectins and terpenes. Their action can include peripheral and central nervous system, cardiovascular or gastrointestinal toxicity. The most important toxic alkaloids are atropine, aconitine, colchicine and strychnine.

Atropine shows a strong anticholinergic effect by competitive inhibition at most muscarinic receptors.

Aconitine enhances the influx of sodium at the sodium channel in peripheral, central and conductive nerve tissue.

Strychnine is a competitive inhibitor of glycine.

Colchicine interrupts the formation of microtubulin, blocking mitosis in the metaphase.

The relevant toxic glycosides are digitalis or digitaloids, and cyanogenic glycosides. Digitalis and digitaloids are cardiotoxic by binding to the sodium–potassium ATPase, especially at the myocardium.

The cyanogenic glycosides can interact with cytochrome a_3 in the respiratory chain, blocking oxygen consumption. Chronic poisoning leads to a peripheral neuropathy.

Podophyllin is a toxic phenylpropanoid leading to mitotic arrest.

The most toxic lectins are ricin and abrin. They disrupt protein synthesis at the 60 S ribosomal subunit.

Terpenes interfere with the $GABA_A$ receptor, inducing excitation and CNS depression.

Reference

Website on antivenoms for viperid envenomation: http://www.toxinfo.org

For Further Reading

Abroug, F., Elatrous, S., Nouria, S., Hagviga, H., Touzi, N., and Bouchoucha, S., Serotherapy in Scorpion Envenomation. A Randomised Controlled Trial. *Lancet*, 1999, **354**, 906–909.

Brent, J., Wallace, K.L., Burkhart, K.K., Phillips, S.D., and Donovan, J.W. (eds), Critical Care Toxicology, Philadelphia, Elsevier Mosby, 2000.

Chang, L.W. and Dyer, R.S. (eds), Handbook of Neurotoxicology, New York, Marcel Dekker Inc., 1995.

Escouba, P., Diochot, S., and Corzo, G., Structure and Pharmacology of Spider Venom Neurotoxins. *Biochimie*, 2000, **82**, 897–907.

Flomenbaum, N.E., Goldfrank, L.R., Hofmann, R.S., Howland, M.A., Lewin, N.A. and Nelson, L.S. (eds), Goldfrank's Toxicologic Emergencies, New York, McGraw-Hill, 2006.

Ford, M.D., Delaney, K.A., Ling, L.J., and Erickson, T., (eds), Clinical Toxicology, Philadelphia, WB Saunders Company, 2001.

Frohne, D. and Pfländer, H.J., Giftpflanzen, Stuttgart, Wissenschaftliche Verlagsgesellschaft, 2004.

Gold, B.S. and Barish, R.A., Venomous Snakebites: Current Concepts in Diagnosis, Treatment and Management. *Emerg. Med. Clin. North Am.*, 1992, **10**, 249–267.

Junghans, T. and Bodio, M., Notfall-Handbuch Gifttiere, Stuttgart, New York, Georg Thieme Verlag, 1996.

Jyaniwura, T.T., Snake Venom Constituents: Biochemistry and Toxicology, Part I + II. *Vet. Hum. Toxicol*. 1991, 468–480.

von Mühlendahl, K.E., Oberdisse, U., Bunjes, R., and Brockstedt, M., (eds), Vergiftungen im Kindesalter, Stuttgart, New York, Georg Thieme Verlag, 2003.

Winkel, K.D., Hawdon, G.M., and Levick, N., Pressure Immobilization for Neurotoxic Snake Bites. *Ann. Emerg. Med.*, 1999, **34**, 294–295.

Index

Page numbers in bold indicate tables; page numbers in italics indicate figures.

Toxicology and Risk Assessment: A Comprehensive Introduction Edited by H. Greim and R. Snyder