General Distribution

ENVIRONMENT MONOGRAPHS

This series is designed to make available to a wide readership selected technical reports prepared by the OECD Environment Committee and Directorate. Following a recommendation by the Environment Committee, this report has been derestricted by the Secretary-General on his own responsibility. Additional copies of Monographs on a limited basis can be forwarded on request.

This monograph is also available in French.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT





28992

## FOREWORD BY THE SECRETARIAT

The OECD is a leading international organisation in the promotion of internationally acceptable methods for the testing of chemicals for regulatory purposes. A large section of the OECD Guidelines for Testing concerns toxicological tests and, apart from the bacterial and cell-culture tests for determination of mutagenic potential, these are for the most part traditional *in vivo* procedures using mammalian laboratory animals.

The use of laboratory animals for evaluating toxic effects of chemicals raises ethical concerns. The OECD therefore has the obligation to revise its Guidelines for Testing of Chemicals any time a possibility arises to reduce the numbers and the suffering of animals used in the tests and, when the time has come, to replace whole-animal tests by *in vivo* procedures.

*In vitro* tests for toxicological endpoints are being developed in many laboratories around the world. Before promising new methods can gain wide acceptance as alternatives for traditional in *in vivo* methods, they need to be validated. Validation is a crucial step that brings a method forward from its development stage in a single laboratory to a stage where it can be considered for generalized use in regulatory schemes of testing. There is much scientific debate about the validation proces.

Discussions held by OECD policy bodies led to the request for a general document on criteria for validation of *in vitro* methods. The Secretariat invited Prof. J. Frazier, Associate Director of the Center for Alternatives to Animal Testing at Johns Hopkins University, to prepare this report. The report is inteded to contribute to the discussion of validation processes. It sets out a broad comprehensive model of validationl It is made public on the responsibility of the Secretary-General of the OECD.

The views presented in this report are those of the author and they do not represent a consensus view of the OECD and its Member countries.

# SCIENTIFIC CRITERIA FOR VALIDATION OF *IN VITRO* TOXICITY TESTS

John M. Frazier, Ph.D.

The Johns Hopkins University
School of Hygiene and Public Health
615 N. Wolfe Street
Baltimore, MD 21205

Document Prepared

for

The Organisation for Economic Co-Operation and Development

1990

ENVIRONMENT DIRECTORATE

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

CONTENTS

# I. INTRODUCTION

## I.A. Uses and Needs for Toxicity Testing

Simply stated, all chemicals are toxic. What distinguishes one chemical from the next is the dosage which produces adverse biological responses. Thus, the difference in the acute toxicity of mercuric chloride and sodium chloride, as measured by the LD50, is the quantity which must be ingested to produce lethality. Based on the wide difference in human sensitivity to the acute toxic effects of these two chemicals, mercuric chloride is considered a highly toxic chemical and its availability is carefully regulated while sodium chloride is found on most dinner tables and the package in which it is purchased is not even labelled as hazardous. The societal judgement on the toxicity of these two chemicals, as is true in all cases, is based on a combination of the intrinsic toxicity of chemicals and the expected conditions of use. Therefore, in spite of its potential toxicity, sodium chloride is considered non-toxic.

Any new chemical which appears in the market or finds its way into drinking water, the food supply or the atmosphere must be classified as to its potential level of toxicity and appropriate controls placed on its production, distribution and disposal. This is the objective of regulatory toxicology and is essential to protecting the health and welfare of man and his environment.

In most cases, substances which must undergo safety evaluation fall into four categories: (1) old chemicals which have never received adequate toxicological evaluation, (2) old chemicals for new uses (3) new chemicals, and (4) new formulations of old and/or new chemicals. The first category consists of the many existing chemicals which have never undergone adequate toxicological safety evaluation. These chemicals represent a large backlog of toxicological testing which must be addressed.

The second category of substances requiring safety evaluation is old chemicals which are being introduced into the market for new uses. A chemical which may produce no toxicity when applied to the skin in an ointment may exhibit dangerous levels of toxicity when ingested in food products. The question of whether the safety evaluation for the new use will require additional toxicity testing will depend to a certain extent on how extensive the initial testing of the chemical was when it first came onto the market and what epidemiological information is available concerning its toxicity related to its initial application.

Thirdly, any totally new chemical entity must undergo safety evaluation. The actual extent of the toxicological testing of this new chemical will depend to a certain extent on its planned use. For example, a medicinal product which may be expected to be taken by human beings for a significant portion of their lifetime, or a food additive which will be consumed by large populations, will require a total toxicological evaluation prior to marketing. On the other hand, a new solubilizing agent used in the production of an industrial product which is restricted in use such that only a small occupational group will be exposed under highly controlled conditions will, in all likelihood, receive less extensive toxicological testing.

Finally, toxicity safety evaluations are conducted on new formulations of chemicals, whether it is mixtures of old chemicals, new chemicals or a combination of the two. The basis for this evaluation is to determine whether interactions between chemicals which individually are considered safe for the proposed use will result in an enhanced toxicity of the mixture which would be considered unacceptable. The scientific bases for this concern are the well documented toxicological problems with drug interactions. Since modern industrial technology can produce literally thousands of new chemicals and formulations each year and there is an extensive

accumulation of older chemicals, numbering in the tens of thousands, which have not received complete toxicological evaluation, the requirement for large scale toxicological testing programs is critical and will continue into the foreseeable future.

There are several distinct categories into which toxicological evaluations required by industry and regulatory agencies fall. One major need is for toxicity screening purposes. The situation where this activity is particularly important is during early stages of new product development in industry and for evaluation of large collections of materials to set priorities for additional toxicity evaluations by regulatory agencies. For screening purposes, what is required are rapid, inexpensive methodologies to provide information for decision making. In the case of product development, industrial management must make decisions which have significant economic implications as to which chemicals should be developed to the market level. As with any decision process, the greater the ability to eliminate uncertainties, the more effective the decision process becomes. One major uncertainty is whether a new product will exhibit unsatisfactory toxicological characteristics which will limit or even prevent its introduction as a consumer product. Toxicity screening methods, which provide reliable predictions of final product toxicity, are extremely important. At this stage, the decision maker is willing to trade off absolute accuracy for speed and cost since complete toxicological evaluations will be undertaken before the product actually goes to market. The greater the reliability of the screening test, the greater will be its influence in decision making and product development.

A second use for toxicity screening tests is to evaluate large collections of samples to set priorities for additional, in-depth toxicological evaluations. This situation arises most often in regulatory agencies which must evaluate water, wastewater, air, food and soil samples for potential toxicity. As a consequence of the large number of samples to be evaluated, it is highly desirable to screen the samples and pick out the individual cases which are most likely to present toxicological hazards. Screening of large collections of samples can be important to industry when toxicity is the desired pharmacological activity, such as for antineoplastic drugs, antibiotics or pesticides. Again, in these cases speed and low cost are essential characteristics of the testing methodology.

The second major use for toxicity testing is for regulatory safety evaluations. In the USA, toxicity testing is required to meet the regulatory requirements of the Food and Drug Administration, the Environmental Protection Agency, the Consumer Product Safety Commission, the Department of Transportation and the Occupational Safety and Health Administration. The specific testing requirements vary depending on the product and its intended use. For regulatory safety evaluations, absolute accuracy and reliability are essential. The chance for an inaccurate classification or undetected toxicological risk must be minimal. As a consequence, a toxicity test which is more than adequate for screening purposes may be totally unacceptable for regulatory decision making.

Finally, even if there were no regulatory requirements for toxicity testing, the ethical issue of allowing a dangerous product on the market which could produce morbidity or mortality in human populations or degrade environmental quality would require toxicological testing with the best technology available to prevent such a situation from occurring. Reports of successful liability litigation in the public media merely reflect society's demand that corporations be ethically responsible for the safety of their products. As long as the judicial system allows liability litigation to be brought against corporations, industry will be required to conduct state of the art safety evaluations - and the state of the art is defined by what will be accepted in the legal process as "good faith" testing.

These three areas — screening, regulatory decision making, and liability protection — encompass the basic purposes for which toxicity testing is conducted.

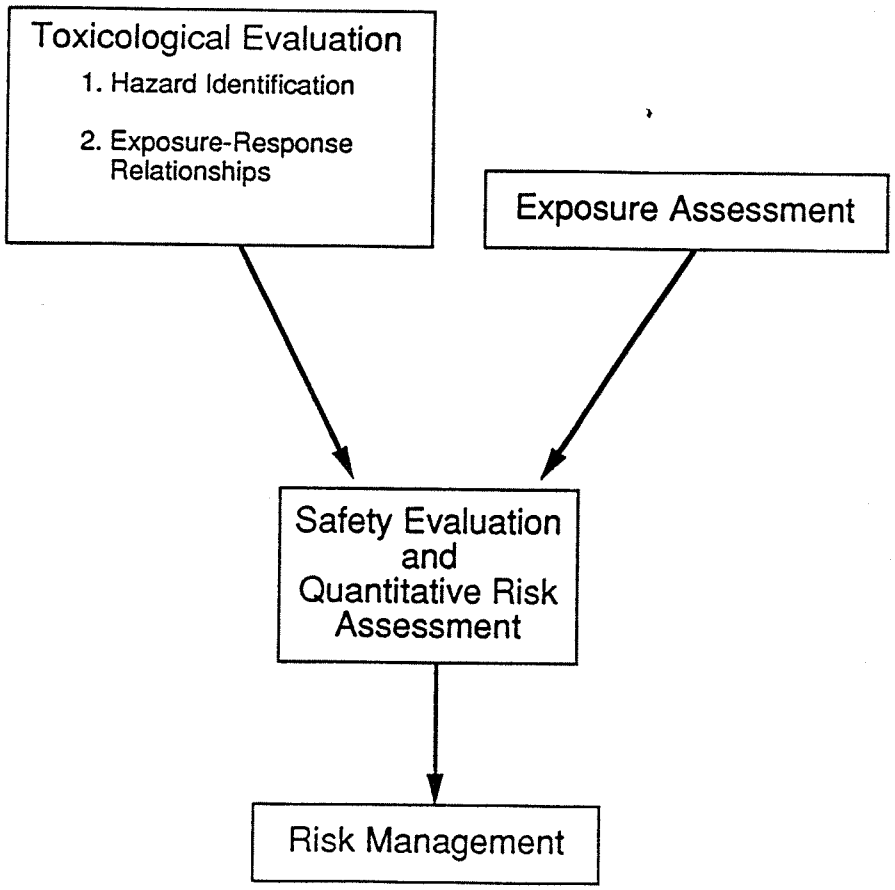


**Figure I-1: The toxicological risk assessment process**

Data provided by either animal toxicological studies or epidemiological studies are used to identify whether a potential toxicological hazard exists regarding human exposure to a particular chemical (hazard identification). If such data suggests that a potential problem exists, then estimation of exposure provides information about expected levels of exposure of specific populations to the chemical under evaluation. The exposure-response relationship provides a risk estimate for adverse effects expected at that level of exposure. If human epidemiological exposure-response data exists which is relevant to the chemical being evaluated, then that information is taken into consideration. In the absence of sufficient experimental toxicity data for the reliable estimate of risk for a particular chemical, then structure-activity relationships are used to infer potential levels of toxicity based on previously studied chemicals of similar structure. The final step in the process is to incorporate a safety factor into the estimate of the safe exposure levels which will produce no adverse effects. The magnitude of the safety factor will depend on the degree of reliability of the database available for risk assessment; a larger safety factor is used where less data is available. Once the risk has been characterized, the data are turned over to risk managers who must decide how the hazard should be regulated and what control measures must be taken.

## I.B. Safety Evaluation and Toxicity Testing

The procedures used to conduct chemical safety evaluations have developed over the years and continue to evolve with our increasing understanding of the science of toxicology (Hayes, 1989). Three major activities of the safety evaluation process are hazard identification, estimation of exposure and determination of the exposure-response relationship (Figure I-1). Hazard identification is the process by which potential toxicological concerns resulting from the manufacture, use and disposal of chemicals are identified for consideration. Whether or not these potential problems will be manifested in the real world will depend on other factors taken into consideration by the risk assessment process. The exposure estimate gives an indication of how many people will be exposed to what concentrations for how long and under what conditions (route of exposure). For a chemical to pose a risk of concern, there must exist the likelihood of human exposure to the agent in quantities sufficient to produce adverse biological effects: a highly toxic substance which is totally confined during an intermediate chemical reaction in an industrial process may pose less risk than a substance of much lower toxicity to which a large portion of the population may be exposed on a regular basis.

A second component of safety evaluation, the exposure-response relationship, is a quantitative relationship which describes the likelihood of developing a particular adverse response as a function of exposure to the chemical. Presumably there is a specific relationship for each possible adverse outcome (*i.e.*, cancer, kidney damage, reproductive effects). Exposure-response relationships depend on many factors which modulate the intrinsic toxicity of a chemical, such as route of exposure (ingestion, inhalation, dermal contact) and host characteristics (age, nutritional status and genetic background, among others). These two fundamental components of the safety evaluation process, exposure and the exposure-response relationship, when combined, provide a quantitative estimate of the potential toxicological risk to man of a given chemical under the proposed conditions of use.

Within the context of safety evaluation, the main objectives of toxicity testing are: (1) to determine which potential adverse effects are of concern for a given chemical (hazard identification) and (2) to provide adequate data to estimate the quantitative exposure-response relationship in man and other organisms. Traditionally, these goals were attained through the use of whole-animal studies. In fact, in the 1920's, at the time when regulatory toxicity testing began, the use of whole animals for toxicity testing was a logical choice and a significant advance since few options existed to predict product safety. Early toxicity testing focused mainly on evaluating acute lethality. The classical LD50 test, originally developed to test the potency of digitalis and other biological materials used for medical purposes, provided a statistically defined estimate of the dose of a chemical which produced 50 percent mortality in a population of animals. Later, in the 1940's, ocular irritation testing, which had been performed for many years, was standardized by the Draize protocol (Draize, *et al.*, 1944). The result was to define the procedures by which test chemicals are placed in the eyes of rabbits, and to develop a numerical scale by which specific ocular effects were evaluated. Over the years, other *in vivo* procedures were developed to evaluate various aspects of toxicological concern. Some of these tests are listed in Table I-1.

Testing procedures employed in each of the testing categories listed in Table I-1 are defined by standard guidelines which prescribe how the tests are conducted (OECD, 1981). In acute toxicity testing (the LD50 test falls into this category) test animals, usually rats or mice, are individually exposed to single high doses of the test chemical and lethality is observed. Acute toxicity tests not only provide information concerning dosages expected to be lethal, which is important in cases of accidental exposures, but also provide information concerning symptomology related to toxicity. This information gives clues to mechanisms of action and target organs. Identification of target organ toxicity is a major objective of subchronic toxicity testing where groups of animals (in these studies both rodents and dogs are usually involved) are exposed to

**TABLE I-1: Traditional whole animal toxicity tests utilized in the toxicological risk assessment process.**

Acute toxicity tests (single exposure — observation period up to 14 days)
Subchronic toxicity tests (repeated exposure — observation period up to 90 days)
Chronic toxicity tests (repeated exposure — observation period up to 2 years)
Reproductive toxicity test
Developmental toxicity (teratogenicity) tests
Ocular/Skin irritation tests
Hypersensitivity tests
Phototoxicity tests
Toxicokinetic studies
Behavioral tests

---

dosages of the test chemical for extended periods of time — up to 90 days. The dosages used are less than the acute toxicity dosage in order for the animals to survive the entire test period. However, dosages and routes of exposure (in the diet, via inhalation or on the skin) are selected to provide information relevant to the expected use of the chemical. At the end of the study, the animals are humanely killed and many tissues are examined for pathological changes. In addition to tissues, blood samples are carefully evaluated for indices of toxicity. These studies are designed to identify which, if any, tissues are affected by longer term exposure to the test chemicals.

Chronic toxicity studies are usually employed only if there is concern about the potential carcinogenicity of the chemical or if it is expected that people will be exposed to the chemical for a significant portion of their lifetime. Two species, usually mice and dogs, are tested and again, full pathology studies are included to evaluate whether adverse responses have occurred in any tissues. Reproductive and developmental toxicity testing evaluates whether chemicals will affect reproductive success or induce teratogenic effects. These studies use both rodents and dogs.

Ocular and skin irritation testing based on the Draize eye and skin test uses rabbits to determine whether chemicals will damage these tissues. The eye test involves administering a fixed dose of a chemical (0.1 ml for a liquid or 0.1 g for a solid) to one eye of a rabbit; the other eye serves as a control. The reaction in the test eye is evaluated (scored) at 24, 48 and 72 hours as to damage to various ocular tissues (see Appendix). For skin testing, rabbits are usually the test animal of choice. The back of the rabbit is shaved and the test chemical applied directly to the skin. The site of application is occluded (wrapped in gauze) and the effect on the skin evaluated at 24, 48, and 72 hours and later if effects are persistent. In both eye and skin testing, the degree of damage is then classified by various regulatory schemes as non-irritating, irritating or corrosive. Hypersensitivity testing is an extension of skin testing to evaluate chemicals which may not directly damage the skin, but which may elicit an immunological response similar to that produced by poison ivy. Phototoxicity testing is used to determine whether sunlight will activate the test chemical and, thus, produce skin irritation or hypersensitivity where the parent compound prior to exposure to sunlight may be inactive. Phototoxicity tests are particularly important for materials applied to exposed portions of the body, such as suntan oils.

Toxicokinetic studies are usually undertaken only when questions or inconsistent results arise. Many times, species differences in toxicity are observed, for example, between rats and mice. In these cases the differences can usually be accounted for on the basis of differences in absorption, metabolism or excretion of the chemical, *i.e.* toxicokinetics. Finally, behavioral tests are employed

to evaluate potential neurological effects of chemicals, either directly or through neurological damage during fetal development. In the latter case, pregnant rats are exposed to test chemicals and the behavioral patterns of the neonate are followed through early stages of development.

As one can see from this brief description of *in vivo* chemical testing procedures, complete toxicological evaluation of new chemicals is extremely time consuming and expensive. It has been estimated that complete testing of one chemical including a two year feeding study would cost between $500,000 and $1,500,000, involve several thousand animals and take 2-3 years to complete. Obviously, most chemicals to which people are exposed have never been completely tested.

Historically, the database to be utilized in the safety evaluation process was developed using whole-animal testing, human epidemiological studies and, in some cases, accidental human exposure data. However, as a result of recent biotechnological advances in the areas of cell culture and bioanalytical methodologies, new possibilities for *in vitro* studies and their application to toxicity testing have been created. In the light of these developments, the traditional approach to toxicity testing should be reevaluated. Before looking at recent developments in several specific areas of *in vitro* toxicity testing, it is useful to discuss the theoretical basis of toxicology to place the general scientific strategy for developing *in vitro* testing methodologies in perspective.

An important component of the toxicological safety evaluation process provided by toxicity testing is the exposure-response relationship (Figure I-2). This relationship gives a quantitative estimate of the percentage of a group of animals which will exhibit a specific adverse biological response at a given level of exposure. The exposure which produces a response in 50 percent of the population (ED50) is traditionally used by toxicologists to indicate the level of toxicity. If a chemical substance is highly toxic with respect to a particular effect (such as liver damage), then the exposure-response relationship is located with its 50 percent response level at a low exposure, to the left of the graph in Figure I-2A (Chemical I). A relatively non-toxic agent will be located to the right (Chemical II). In the case of oral ingestion of mercuric chloride or sodium chloride, mercuric chloride would behave more like Chemical I whereas sodium chloride would exhibit a response more like Chemical II. The position of the ED50 alone is often not adequate to describe completely the potential toxicity of a chemical. Consider two compounds with widely different ED50s, but whose exposure-response relationship crosses over at lower concentrations such as in Figure I-2B. The substance which is more toxic when comparing ED50s becomes the lesser toxic substance at lower exposure levels. Thus, both the location (ED50) and shape (slope) of the exposure-response relationship are important for ranking the potential toxicity of new chemicals. Since the quantitative parameters which describe the location and shape of this relationship are used to define the toxicological classifications of chemicals, it becomes important to understand the basic toxicological mechanisms which influence these parameters.

To appreciate fully the exposure-response relationship one must explore the fundamental nature of the toxicological process (Figure I-3); its two major components are toxicokinetics and toxicodynamics. Toxicokinetics describes the movement of the chemical in the biological system, its absorption, distribution to tissues, metabolism (chemical conversion to derivative forms, usually by enzymes), storage and excretion. On the other hand, toxicodynamics refers to the alterations in the biological system which are a consequence of the presence of the chemical in the system. At the molecular level these alterations are biochemical, such as the inhibition of enzymes involved in normal cellular functions, while at higher levels of biological organization these alterations are manifested as tissue pathology or clinical toxicity. If human exposure occurs, the toxicokinetic properties of the agent will determine whether the agent or one of its metabolites will ultimately reach a sensitive cellular/molecular target and initiate a biological response. If the reactive form of the chemical reaches the potential molecular target (*e.g.*, a biological macromolecule with which

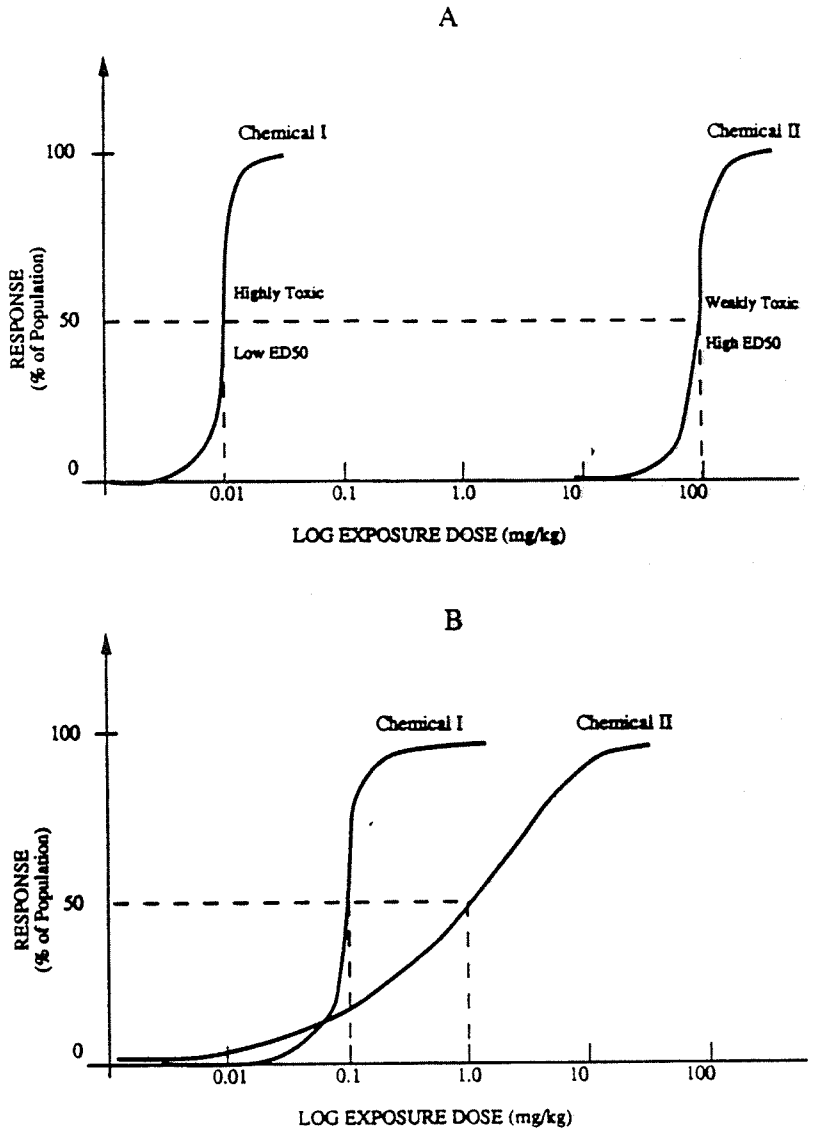


**Figure I-2: The exposure-response relationship**

The exposure-response relationship is a quantitative relationship between the level of exposure and the expected percentage of the population exposed which will have a particular toxic response at that exposure level. If the response measured is lethality, then the exposure which produces lethality in 50% of the animals exposed is designated as the LD50, for lethal dose to 50%. On the other hand, if the toxicological endpoint is something other than lethality (for example, liver damage), then the term ED50 is used for the effective dose which produces a response in 50% of the population. (A) illustrates the situation when two chemicals have similar shapes of the exposure-response relationship but different ED50s. (B) illustrates the situation when two chemicals have both different shapes and different ED50s. In this latter case, at low doses chemical II is more toxic than chemical I while at a high doses the reverse is true. If large populations of people are expected to be exposed to a given chemical, then it becomes critical to determine the exposure which will affect an extremely small percentage of the population. In this situation, the critical region of the exposure-response relationship for that chemical is the tail-end of the curve at low doses.

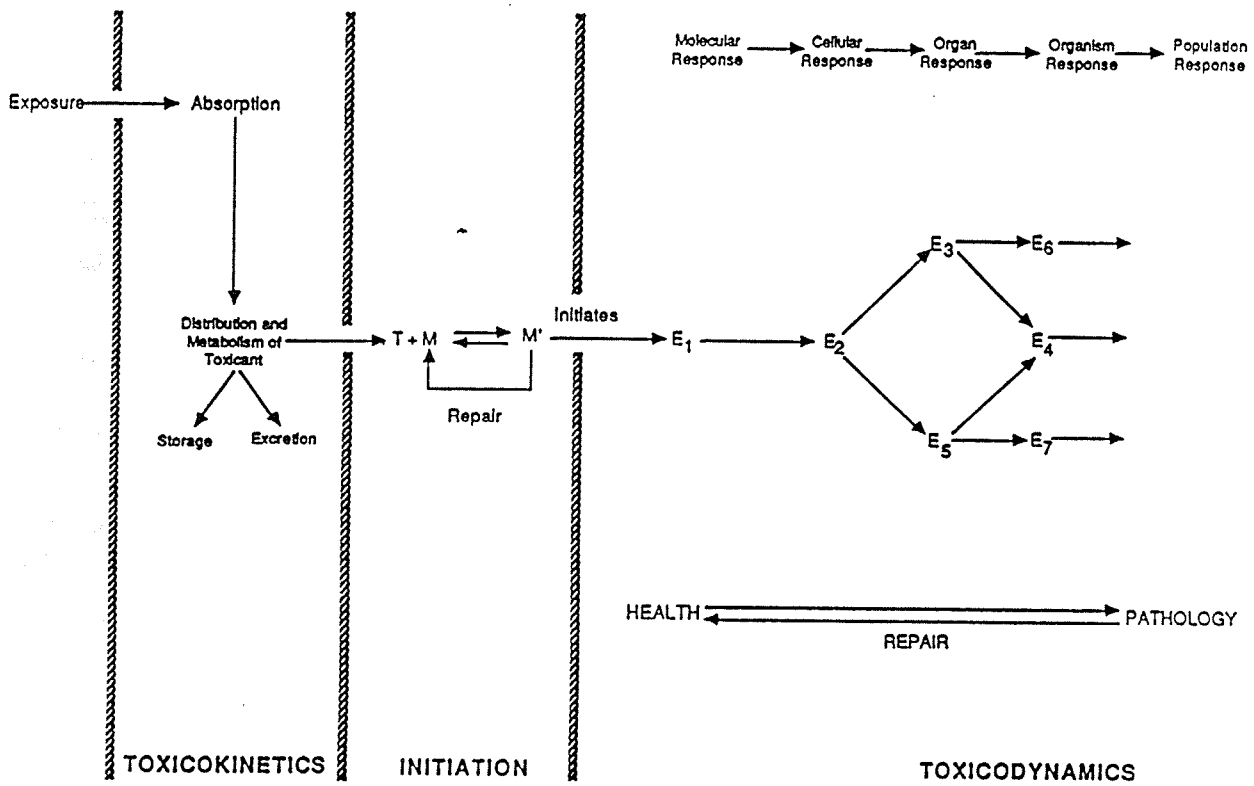


**Figure I-3: The toxicological process**

The processes which are involved in the ultimate expression of an adverse effect in an organism are illustrated in this schematic diagram. Following exposure, a chemical may be absorbed into the organism and various biochemical and physiological mechanisms control the distribution, metabolism, storage and excretion of the chemical. Together these processes control the toxicokinetics of the chemical in the organism. The specific combination of these competing processes determines the amount of the active form of the chemical (T) which reaches a molecular target (M — a macromolecular component of the cell) in a sensitive tissue. The reaction of the chemical (or its active metabolite) with the molecular target results in a molecular alteration (M*) which initiates the cascade of events (E) resulting in the expression of the toxicological effect. Initially, these events are expressed at the molecular level but with time the sequence of events successively involves higher levels of biological organization — cellular, organ, organism and ultimately the population of organisms. This cascade of events is referred to as toxicodynamics. Note that repair processes, either at the molecular level, possibly involving the molecular target itself, or at higher levels of biological organizations, tend to reverse the toxic effects. The balance between the rate of damage and the rate of repair is an important factor in determining the observable toxicity of a chemical.

the agent interacts), the question becomes "to what degree will this reaction induce the cascade of biochemical/physiological events which results in an adverse effect in man (*i.e.*, initiate the toxicodynamic response)?".

An important biological factor which modulates the relationship between the initiation of a toxic effect and its ultimate expression as some form of adverse pathology is the ability of the system to repair toxicant-induced damage at both the molecular/cellular level and at the tissue level. At the cellular level, repair can occur by mechanisms which recognize alterations in macromolecules and either correct the damage directly (such as DNA repair enzymes) or eliminate the altered macromolecule by cellular degradative pathways. At the tissue level, repair usually involves proliferation of surviving cells to replace necrotic or functionally inhibited cells. The ability of the biological system to effect repair at both of these levels will significantly influence the quantitative exposure-response curve.

The integration of all of these processes — toxicokinetics, initiation, the biological response cascade and repair — determines the quantitative exposure-response relationship. The complexities of these processes are immense and rule out the possibility of predicting this relationship from theoretical principles. Current predictive toxicology utilizes the database provided by whole-animal experimentation to estimate this relationship for man and other organisms. The general success of the *in vivo* approach over the years is clearly positive on balance and human health and welfare have been protected in the main. When considering the introduction of new testing methodologies, one must ask whether alternative testing methodologies can provide an adequate database to predict human toxicological responses, as well as provide accurate quantitative predictions of exposure-response relationships.

### I.C. *In Vivo* Toxicity Testing

Before delving into the *in vitro* world, the strengths and weaknesses of the *in vivo* (whole-animal) approach to toxicity testing should be outlined. As discussed above, the objective of toxicity testing is to generate a toxicological database, *i.e.*, a toxicological profile, for each chemical which can be utilized for safety evaluation. *In vivo* testing has certain advantages which have made these methods attractive over the years. First and foremost, a whole-animal model for toxicity testing provides an integrated biological system which serves as a surrogate for the complexities of the human and other animal systems. This variety of interacting biochemical and physiological systems offers a wide net in which to catch potential toxicological responses. Complex, integrated biological functions such as behavior and immunological responses are present which can be used to simultaneously evaluate many potential adverse effects. Thus, the whole-animal serves as a broad spectrum sentinel. Secondly, the physiological and anatomical arrangement of mammalian test organisms provides a useful tool for investigating essential questions of whether systemic toxicity is affected by routes of exposure (dietary, inhalation, dermal) and the exposure levels by each route which are relevant to potential toxicity. A most important value of whole-animal studies is in the area of chronic toxicity. Will repeated exposure over long periods (90 days in subchronic tests) to low doses of the chemical produce unique forms of toxicity not seen in acute (less than 14 day) high dose studies? Finally, whole-animal studies can be designed to determine whether particular toxic effects are reversible or not. Chemicals which produce irreversible, permanent damage are of much greater concern than chemicals which cause easily repaired damage to the biological system. Thus, whole-animal toxicity testing has several strengths which support its continued use. However, there are also significant problems in using whole animals which must be recognized.

First, use of living animals, particularly in toxicity testing, raises significant public concerns with respect to animal welfare. Although the issue of animal welfare and toxicity testing is not

new, public awareness has been heightened in the 1980's. Second, whole-animal testing is expensive in terms of the time and cost involved in generating the complete database needed for risk assessment. As discussed previously, whole animal testing can cost more than $1,000,000 and last for several years if carcinogenicity testing is required. For economic reasons, most chemicals are not thoroughly tested. A third major problem area involves the extrapolation of the available animal toxicological data to man and other organisms.

Two major extrapolations are required for safety evaluations based on traditional testing: inter-species extrapolation and high dose-low dose extrapolation. Inter-species extrapolation addresses the issues of how to make data obtained from animal models relevant to man. Many of the inter-species extrapolation questions are of a toxicokinetic nature. Is absorption of the chemical from the gastrointestinal tract or skin similar in animals and man? Does man metabolize the chemical in the same way as the animal model? Are excretion processes similar? Techniques have been developed to assist in answering these questions, such as physiologically based pharmacokinetic modelling: (NRC, 1987). This technique attempts to utilize knowledge of anatomic, physiological and biochemical processes to predict kinetics of chemicals in man based on animal data. Other inter-species extrapolation problems relate to differences in the toxicodynamic phase of the toxicological process. Does man have the same molecular targets as the animal model? Are human repair mechanisms more efficient/less efficient? Lack of basic toxicological knowledge relating to these questions introduces uncertainties in the ability to accurately quantify human risk resulting from exposure to toxic chemicals.

A second major extrapolation problem is what is referred to as high dose-low dose extrapolation. If a new chemical produces neurotoxicity in a rat when exposed to high doses of the chemical by gastric intubation (placing the test chemical in the stomach of the animal) under experimental conditions, will this same chemical produce toxicity in man when exposure occurs at a fraction of the dose over long time periods as a result of ingesting contaminated food? To make such judgments, extensive information concerning the dependence of absorption, distribution, metabolism, excretion, target organ sensitivity, mechanism of action and repair on level of exposure are needed. Again, techniques are available to make these extrapolations, but the uncertainty of such predictions is great. Furthermore, if any of the essential data is missing, confidence in predicting human responses deteriorates significantly. These limitations imposed by the necessity to extrapolate data from animal models to man, from high dose to low dose, from one route of exposure to another, are offset by the long historical precedents established by *in vivo* toxicity testing which allows non-quantifiable factors, such as human judgement and experience, to play a strong role in the safety evaluation process. Be that as it may, in practice the ultimate solution to the uncertainties inherent in the various extrapolations is to include a safety factor in determining the final estimates for safe human exposure.

## I.D. *In Vitro* Toxicity Testing

During the 1980's, *in vitro* toxicity testing systems have received increasing attention (Goldberg and Frazier, 1989). This interest relates to several advantages which *in vitro* testing systems offer. First, *in vitro* testing has the potential to be more rigorously standardized than *in vivo* testing. This has important advantages since reliable, quality-controlled data can be generated. This is usually not fully possible in whole animal testing due to the prohibitive cost of including positive and negative controls in each test protocol. Inclusion of such controls in *in vitro* testing makes it possible to standardize each test independently. Secondly, *in vitro* systems are in general faster and less expensive, thus offering an economic advantage which is important because it will allow testing of a larger number of chemicals for the same cost.

As discussed above, there are serious problems associated with extrapolation of *in vivo* animal data to man and other organisms. The question of species differences can be eliminated by using cells derived from the species in question in *in vitro* toxicity testing systems. In the case of human safety evaluation, human cells can be used directly in *in vitro* test systems. Another benefit of *in vitro* systems is experimental control of the cellular dose of a chemical, thus it is possible to define precisely the critical concentrations of toxicants. *In vivo*, both in animals and man, it is often difficult to determine the precise chemical dose to the tissues of concern and when this becomes important, extensive toxicokinetic analyses must be undertaken to obtain the required information. *In vitro*, samples of the medium and cells can be taken directly for analysis of the test chemical and its metabolites. Furthermore, samples can be taken at various time intervals to follow the time course of events and thus resolve mechanisms involved in pathological responses. An additional advantage of *in vitro* systems is that small quantities of test chemicals are needed. This allows tests to be conducted on novel compounds produced by the research chemist which are usually available in limited quantities. Finally, *in vitro* toxicity testing offers the advantage of reducing the use of live animals in toxicity evaluation studies which is important from a societal point of view. These points are summarized in Table I-2.

---

**TABLE I-2: Advantages of *in vitro* systems**

Utilize human cells.
Allow for control of environmental conditions.
Eliminate interactive systemic effects.
Utilize large number of test organisms (cells) per dose level.
Reduce variability between experiments.
Allow for simultaneous and/or repeated sampling over time.
Make feasible complex interactive experiments.
Are often quicker and cheaper.
Require smaller quantities of test chemicals.
Produce small quantities of toxic waste.
Reduce animal usage.

---

Considering these advantages, what is delaying the replacement of *in vivo* methods by *in vitro* toxicity testing systems? The major limitation is that *in vitro* toxicity testing methodology is not yet fully accepted by either the scientific or regulatory communities. Other than in the area of mutagenicity, teratogenicity and screening for carcinogenicity, *in vitro* toxicity testing is a new concept. Prior to 1980 there were few peer-reviewed *in vitro* toxicity testing systems for toxicological safety evaluation. Today there is significant effort in this area (Goldberg, 1983-1989; Brown, 1983; Purchase and Conning, 1986; Balls and King, 1988). However, the historical database needed to define fully the limitations of these systems does not yet exist. Major problems which must be addressed include questions concerning what set of *in vitro* test systems will provide a sufficient database to allow reliable risk assessment. To simulate the complexity of the responses of the whole animal will require a battery of *in vitro* tests. Will it be necessary to have one *in vitro* test for every potential target cell type in the body? How will *in vitro* systems evaluate toxicological responses which involve interactions between different systems, such as immunological processes or regulation of blood pressure? How can *in vitro* systems evaluate chronic toxicity, a biological change which develops over long periods of time, or recovery from toxic insults? Finally, how will chemical dosages determined for cells *in vitro* be extrapolated to

human exposure via the diet, the skin or by inhalation? These are only some of the toxicological questions which must be answered before *in vitro* toxicity testing can replace *in vivo* methods.

In principle, *in vitro* toxicity testing systems can be introduced into the safety evaluation process within the context of: (1) screening, (2) adjuncts and (3) replacements. For screening purposes rapid, inexpensive *in vitro* testing methods can be used to provide preliminary toxicity evaluations. Due to the nature of the screening process *i.e.,* definitive toxicity testing will be conducted at a later time, the criteria for a good screening test are less rigorous than for other testing purposes. As adjuncts, *in vitro* toxicity testing data are used as an integral part of safety evaluation. The data provided by *in vitro* testing systems supplement that provided by existing *in vivo* methodologies. In this context, *in vitro* data can be used in a tier testing strategy which can reduce the use of live animals in the overall safety evaluation. Finally, *in vitro* toxicity testing protocols may someday totally replace some, if not all, traditional *in vivo* testing protocols, thus eliminating the use of live animals. In spite of the emphasis in recent years on the development of *in vitro* toxicity testing batteries as replacements for existing *in vivo* testing, the current state of *in vitro* testing is not adequate to eliminate *in vivo* testing. Because of this emphasis on replacement, the full value of *in vitro* testing systems for screening and adjunct testing activities has not been fully realized.

Significant scientific and technical problems must be overcome in the development of *in vitro* toxicity testing, however these methodologies are viewed as a future direction for the science of toxicity testing. Two major factors contribute to this perception: technological developments and rapid expansion of basic, mechanistic toxicology. Over the past 20 years, biotechnological advancements have provided the research toxicologist with new tools. Developments in cell culture allow for the growing of cells with differentiated properties in defined conditions which can be used for toxicity testing. New bioanalytical tools, such as high performance liquid chromatography and monoclonal antibody based assays, provide ways to make measurements which were not possible only a few years ago. The combination of new cell culture techniques and new, more sensitive analytical techniques to evaluate toxicologically important endpoints stimulates new approaches to *in vitro* testing. Other factors which promote *in vitro* toxicity testing are: (1) better understanding of the initiation of toxicological processes and (2) the mechanisms of expression of toxic effects. Such knowledge permits the interpretation of *in vitro* testing results. Clearly, *in vitro* methodologies are not yet ready to replace *in vivo* toxicity testing, but the future applications are coming into focus.

What is the current status of *in vitro* initiatives in toxicity testing? Some important categories of research include: (1) cytotoxicity, (2) irritation and inflammation, (3) genotoxicity, (4) teratogenicity, (5) target organ toxicity, (6) toxicokinetics and (7) structure-activity relationships. This is not an inclusive listing, but it covers the most general categories in which *in vitro* approaches are actively being investigated.

1. Cytotoxicity

All chemicals are toxic, what is of concern is the dose level which will produce toxic effects. What defines a dangerous chemical is its ability to produce toxicity under expected conditions of exposure, whether in the food we eat, the drugs we take, the exposure we receive during occupational, recreational or household activities, or through contact with products used for personal hygiene or cosmetic purposes. *In vitro* cytotoxicity assays are designed to evaluate the intrinsic ability of a chemical to kill cells.

Many *in vitro* cytotoxicity assays have been developed over the years. Such assays have a long historical background. Some were developed for special purposes, such as screening potential antineoplastic drugs for their ability to kill cancer cells, while others were developed for more

general purposes. Cytotoxicity can be evaluated with any cell type that can be cultured *in vitro* and methods for evaluating whether or not cells are dead have multiplied rapidly in recent years. Thus, anyone interested in evaluating the cytotoxic potential of a chemical will have a multiplicity of test systems from which to choose. Two cytotoxicity test systems which have received particular attention in the toxicity testing arena are the cytotoxicity assay developed by FRAME (The Foundation for the Replacement of Animals in Medical Experimentation — an animal welfare organization centered in England; Clothier, *et al.*, 1988) and the neutral red uptake assay developed by the research group at Rockefeller University which was sponsored by Revlon (Borenfreund and Puerner, 1985).

To give some idea of how these tests are conducted, the FRAME assay consists of growing cells in plastic dishes: various concentrations of test chemicals are added to the culture medium in different culture dishes and exposed cells as well as control cells (*i.e.* not exposed to the test chemical) are grown for 24 h. The test chemical is washed out and an analytical reagent is added to the cultures which reacts with proteins in the cells to give a colored product which can be measured. Control culture dishes containing cells which have not been affected by the test chemical will have a dark blue color due to the presence of many healthy cells. Culture dishes exposed to increasing concentrations of test chemicals will have decreasing levels of color due to the progressively increasing killing of cells. A concentration (exposure)-response curve can be generated, similar to those described for animal testing (Figure I-2), and the concentration that produces 50% inhibition of protein content (IC50) is determined. This value can be compared to the IC50 for known chemical toxins to obtain an evaluation of the relative cytotoxicity of a new chemical. Variations on this assay can include the addition of hepatic enzyme preparations (S9) which normally metabolize chemicals in the liver to determine whether *in vivo* metabolism to reactive chemical intermediates is necessary for a chemical to produce cytotoxicity. An advantage of a test such as this is that it can be automated so that many chemicals can be rapidly tested at relatively low costs. In general, these same advantages accrue to many *in vitro* test systems.

The neutral red cytotoxicity assay is slightly different. The basic principle of this test is that neutral red is a vital dye, *i.e.* living cells will take up the dye and store it in lysosomes. Again cells are grown in plastic dishes for 24 h in the presence or absence of the test chemical. At the end of the exposure period, the test chemical is washed out and fresh media containing the neutral red dye is added. The cells are incubated for an additional 3 h. Unabsorbed dye is washed out and the remaining dye which has been taken up by the cells is measured. The amount of dye retained by the cells is a measure of the number of living cells. The quantitation of this assay is similar to the FRAME assay and the IC50 of the test chemical is compared to the IC50 of known toxic chemicals to obtain a relative ranking of the cytotoxicity.

There are many such assays currently available ranging from rather simple tests, as exemplified by the two assays described above, to very sophisticated assays requiring expensive analytical instrumentation. In all cases, the data obtained provides basic information concerning the intrinsic cellular toxicity of chemicals and gives analogous data to that provided by acute *in vivo* toxicity testing. However, *in vitro* cytotoxicity tests do not replace the *in vivo* acute toxicity test since in the intact animal many possible toxic responses are screened simultaneously, while the *in vitro* test will detect universal toxins (*i.e.* toxins which will kill all cell types through a common mechanism, such as cyanide which blocks essential cellular energy metabolism in all cells) but may miss tissue specific effects or effects which involve specific cell-cell interactions. In spite of the limitations of *in vitro* cytotoxicity tests, they do provide essential toxicological information on the intrinsic cellular toxicity of pure chemicals, mixtures and formulations.

2. Irritation and Inflammation

Research in this area addresses problems relating to eye and skin irritation. The *in vivo* Draize ocular irritation tests have been used for many years and there is a general perception that alternative approaches to ocular irritation testing are needed. Significant research effort has focused on developing alternative *in vitro* methods for both the Draize eye and skin test. In a recent review of the ocular irritation problem, The Johns Hopkins Center for Alternatives to Animal Testing identified over 30 *in vitro* tests which can potentially be utilized for ocular irritation testing (Frazier, *et al.*, 1987). The various tests identified were classified on the basis of the toxicological endpoints which they evaluated, *e.g.* cytotoxicity, impairment of cell function or release of inflammatory mediators. The status of alternatives to ocular irritation testing is summarized in more detail in Section XII. The inability of single *in vitro* tests to predict ocular irritation has been demonstrated in a recent publication (Kennah, *et al.*, 1989).

3. Genotoxicity

This category includes research approaches to mutagenicity and carcinogenicity testing. Many approaches, beginning with the Ames bacterial assay for genetic damage by chemicals, have been proposed and/or are under development. This area has probably been the most adequately funded and active area of *in vitro* approaches to toxicity testing as a consequence of the fact that a whole animal test for carcinogenicity is extremely expensive and time-consuming. Therefore, there is significant economic pressure to develop alternatives to *in vivo* carcinogenicity testing. Currently, *in vitro* tests are widely used as screens for potential genotoxicity, but do not replace the lifetime bioassay in rodents as the definitive test for regulatory purposes.

4. Teratogenicity Testing

Developmental abnormalities resulting from chemical agents are another very important toxicological concern. Birth defects caused by thalidomide are a constant reminder of the toxicological risks associated with beneficial chemicals. Thalidomide is an important example in the history of toxicity testing because it demonstrated limitations of whole-animal testing as it was conducted at the time. Based on the lessons learned from thalidomide, *in vivo* testing strategies were modified to prevent the recurrence of such problems. This evolutionary aspect to toxicity testing strategies should be kept in mind when considering the future development of *in vitro* toxicity testing: greater knowledge of the basic mechanisms of toxicity will lead to better designs of *in vitro* tests to evaluate these processes.

The key to teratogenicity testing is to establish the relationship between *in vitro* indices of toxicological response and the complex process of differential toxicity in the developing organism (Kimmel *et al.*, 1982). Proposed test systems range from lower animals, such as *hydra* and *drosophila*, to cultures of rodent limb bud cells and whole embryo cultures. Although these test systems show promise, significant problems still exist in predicting human teratogenicity.

5. Target Organ Toxicity

In contrast to general cytotoxicity, this category of *in vitro* testing includes all aspects of organ-specific toxicity. In subchronic and chronic *in vivo* testing, animals are continuously treated with a test chemical and at the termination of the study all organs are examined for pathological changes. Studies designed in this manner allow for the identification of specific effects in specific tissues. *In vitro* research efforts in this area focus on cell cultures from specific organs to evaluate

selective toxicity that could potentially occur in those particular organs. In addition, *in vitro* research is conducted in order to elucidate mechanisms of toxic action of test chemicals in target organs. Extensive progress has been made in the use of *in vitro* tests (at least as screens) for heart, kidney, liver, lung and nervous system toxicity (Davenport, *et al.*, 1989).

As an illustration of how *in vitro* target organ toxicity testing is advancing, one area of rapid development is hepatotoxicity testing. The methods used for *in vitro* hepatotoxicity testing are derived from experimental techniques which were developed for liver research — isolated liver cells, liver slices, and isolated, perfused whole livers. The nature of the testing using these *in vitro* systems ranges from identifying chemicals which specifically produce toxicity in the liver to determining the metabolic kinetics and biliary excretion of chemicals. Test systems based on isolated rat hepatocytes (Rauckman and Padilla, 1987) can also evaluate cellular markers for potential toxicity such as peroxisome proliferation, which is thought to be related to epigenetic mechanisms of cancer formation, or unscheduled DNA repair, which is an index of DNA damage.

An important advantage of *in vitro* techniques in hepatotoxicity testing is that human liver cells and liver slices can be used for toxicity testing (Frazier, *et al.*, 1989). Advances in the biotechnology of human hepatocyte culture, in the cryopreservation of human liver tissue and genetic engineering techniques to provide continuous supplies of normal, differentiated human hepatocytes, are helping to solve the problems of interspecies extrapolation of testing data as well as to reduce the need for animals to provide cells and tissues for toxicity testing systems.

The approaches described to address the issues of hepatotoxicity evaluation are parallelled by developments in other target organs. As more knowledge is obtained about mechanisms of chemical toxic action in each organ system, new *in vitro* methods can be developed to test for these effects.

## 6. Toxicokinetics Testing

Understanding the absorption, distribution, metabolism and excretion of toxicants is essential for risk assessment. Many studies have investigated the relationship between *in vitro* and *in vivo* toxicokinetics. However, much remains to be accomplished before it will be possible to accurately predict *in vivo* toxicokinetics from *in vitro* observations. In this area, mathematical models, exemplified by the physiologically based pharmacokinetic models (NRC, 1987) are proving their importance. Once these techniques have been fully established, it should be possible to relate *in vitro* dose-response relationships to the effects which could be expected from human exposure. Because of the importance of toxicokinetics in extrapolating from *in vitro* to *in vivo*, this area must receive high priority in future research.

## 7. Structure-Activity Relationships

The objective of the structure-activity approach is to be able to predict toxicological effects of chemicals based on an analysis of their molecular structure (ECETOC, 1986). Current efforts toward this goal attempt to correlate general toxicological responses (lethality, ocular irritation, mutagenicity, cytotoxicity) with chemical parameters computed from the molecular structure of the test chemical. Currently these methods are highly empirical, and as such the reliability of prediction is difficult to estimate. Structure-activity relationships are currently used for screening purposes. When available toxicological data are limited, structure-activity relationships become important for regulatory purposes. In the future when specific mechanistic effects can be related to chemical structure, these techniques will prove to be valuable *in vitro* methods since they are based on computer analysis of the physical-chemical structure of molecules.

These seven areas or general categories of *in vitro* toxicity testing are characterized by active research efforts in many laboratories around the world. It should be re-emphasized that there is no one *in vitro* test which is going to answer all toxicological questions. As implied by the broad range of *in vivo* toxicity testing, there are many different areas of toxicological concern, each involving very different biochemical/mechanistic processes. At this time and probably in the future, no single *in vitro* test is adequate to evaluate this broad spectrum of toxicological endpoints. In the best case it will take a battery of several *in vitro* tests to obtain the necessary information to evaluate specific human risks resulting from exposure to toxic chemicals.

It should be obvious from these examples that *in vitro* toxicity testing is a new and developing field of toxicology. Its ultimate success will depend on scientific breakthroughs in various areas of toxicology and cell culture. A major goal of *in vitro* toxicity testing is to use human cells for all testing systems to eliminate species extrapolation questions. To attain this goal, several technological obstacles must be surmounted. First, not all human cells can be adequately cultured and those that can suffer from the phenomenon of dedifferentiation, *i.e.*, the cells take on the characteristics of more primitive cells than the normal cell found *in situ*. Significant research activities are in progress to maintain differentiated cells in culture, but more research is needed in this area. Secondly, the supply of normal human cells for toxicological testing activities is limited. Biotechnology and genetic engineering must be applied to solve this supply problem and make human cells a commonly available resource. Advances in the science of toxicology are needed to provide insights into biological indices of toxicity which can be evaluated *in vitro*. The greater our understanding of mechanistic toxicology, the more rapidly *in vitro* toxicity testing will develop. Finally, the problems of *in vitro* to *in vivo* extrapolation must be overcome. *A priori*, the extrapolation from tissue culture to man can be expected to be a difficult problem. However, with a focused and well supported research effort, the problems can be defined and extrapolation procedures established.

## II. OVERVIEW OF VALIDATION PROCESS

Validation is the process by which the credibility of a candidate test is established for a specific purpose. There are clearly two parts to this activity, establishing both reliability and relevance. Reliability means that when the test is conducted using the appropriate test protocol any technically competent laboratory can reproducibly obtain accurate results. This requirement applies to different laboratories as well as the same laboratory at different times. Relevance implies that the data have meaning for some scientific decision making process. Obviously a test which produces highly reliable, consistent data is useless if it is unclear how to interpret the data for safety evaluation purposes. Conversely, a test which is highly relevant, *i.e.* evaluates mechanistically defined endpoints, is useless if different laboratories get different results or the same laboratory gets different results at different times. Hence, the validation process must evaluate both of these aspects of a new test in the context of the objective for which that test is designed.

The overall process of validation of a new test methodology consists of several stages which are outlined in Figure II-1. Before committing to a validation program, it is essential to define the objective of the validation exercise. In this context, there are two types of objectives which have been pursued. The first objective has been to validate a replacement for an existing *in vivo* test procedure. Most interest has been placed on replacements for the LD50 acute toxicity protocol and the Draize eye test. This approach uses the *in vivo* test to define the objective for the validation process. As has been discussed in previous sections, for this objective to be satisfied, the replacement strategy must encompass a battery of tests to ensure success, and therefore will only be successful if groups of tests are validated together. The second objective is to validate an individual *in vitro* toxicity test for some specific toxicological objective, *i.e.*, to validate a test

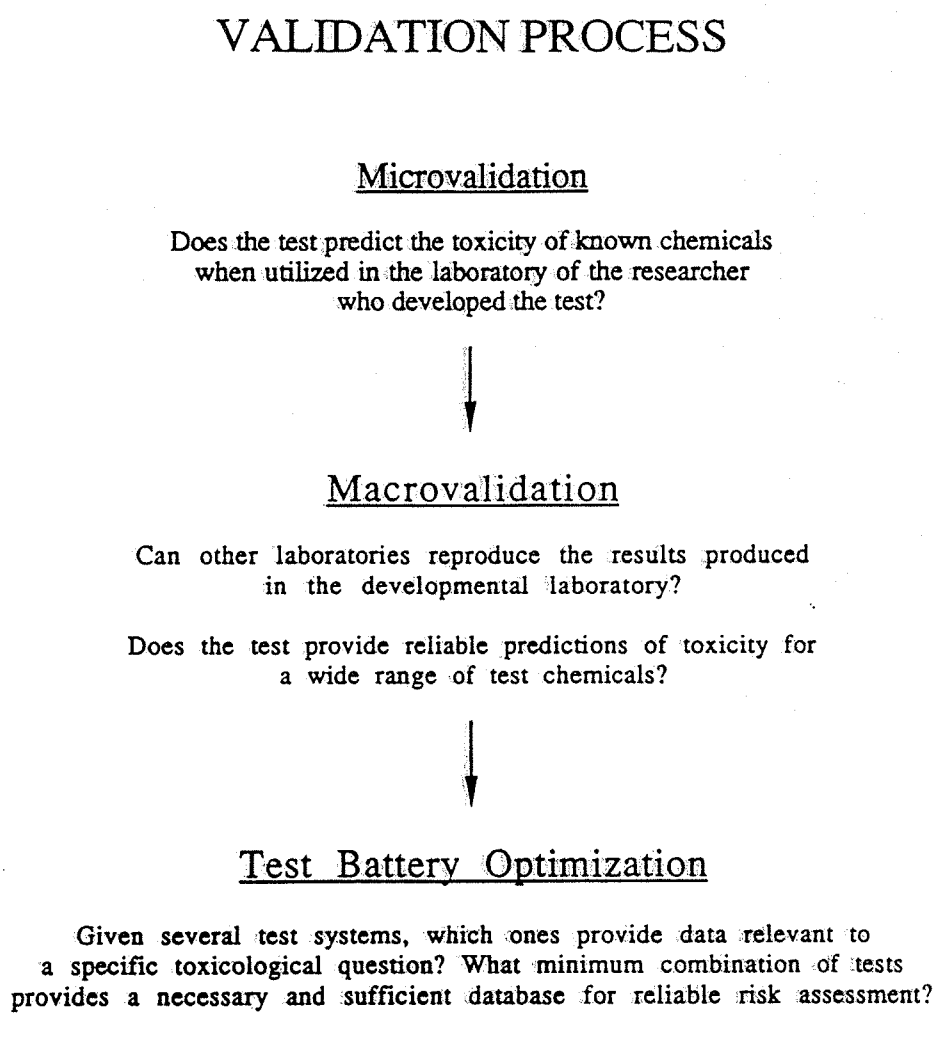


**Figure II-1: Validation process**

The validation of new *in vitro* tests requires a series of steps which provide the necessary data to evaluate the potential of a new test. In order to carry out this process, it is necessary to have a set of test chemicals and a relevant toxicological database for these chemicals to evaluate whether the predictability of the *in vitro* test is reliable as measured by sensitivity (how often the test predicts known toxicants as positive) and specificity (how often the test predicts that known non-toxicants are negative). Based on these data, the role of a particular new test in the overall test battery can be established.

which will reliably predict the hepatotoxicity of a series of test chemicals. When the objective of the validation program is defined in this manner, the strategy and design of the validation process will differ from the first case. For example, a single test system can be validated in this context.

Once the objective of the validation activity is defined, several preliminary steps must be taken. If the objective is to replace an *in vivo* test, then it is essential that a set of potential candidate tests be selected. If, on the other hand, the objective of the validation exercise is to validate a specific test for a specific toxicological purpose, then the candidate test becomes the focus of the activity. Having defined the candidate test or tests, the next step is to select specific chemicals which will be utilized in the validation scheme to evaluate the test system(s) in order to define the range of validity of the test and its performance characteristics. In addition, mechanisms must be set up to conduct the validation activity as a blind study, *i.e.* the tester should not know the identity of the chemicals being tested. Once these preliminary activities have been completed, the actual validation process begins.

Intralaboratory Assessment is the initial phase of the process. This activity is usually conducted by the laboratory which develops a specific alternative test and is designed to standardize the test protocol and establish the feasibility of using the proposed test for its intendedpurpose. Having successfully completed Intralaboratory Assessment, the Interlaboratory Assessment consists of transferring the test to several laboratories to establish its reproducibility. In addition, a large number of chemicals must be tested in one or more laboratories to accurately define the rates of false positives and false negatives. This process is referred to as Test Database Development. With these data in hand, it is possible to make scientific judgments as to the value of specific alternative tests for various aspects of safety evaluation. If the objective is to replace a given *in vivo* testing procedure, then the data generated by carrying a collection of alternative tests through the validation procedure can be used to select the optimum test battery to minimize errors.

In the following section, each of the steps involved in the validation process is described in more detail. The objective of these discussions is to identify the most important issues which must be considered when conducting alternative test validation.

## III. SELECTION OF TESTS FOR VALIDATION

### III.A. Intended Use of Test and Selection Criteria

The first step in any validation exercise is to fully define the purpose of the activity. If the objective is to validate a specific test for a specific purpose, then test selection becomes trivial. If, on the other hand, the objective of the validation project is to establish a test battery as a replacement for an existing *in vivo* toxicity testing procedure, then test selection becomes a critical activity. This section will deal with some of the factors which must be considered in test selection to ensure success of the validation program.

As previously discussed in Section I.D., the types of *in vitro* toxicity testing activities fall into three categories — screening, adjuncts or replacements. In the first area, the test is used in a manner where definitive information is not required and an understanding of the mechanistic basis of the test is not required. Thus, if the purpose of the validation exercise is to evaluate a single test or a collection of tests as screens for some toxicological purpose, then a wide range of tests may be selected. However, if the objective of the validation project is to select a single test or a battery of tests as either adjuncts to or replacements for existing *in vivo* testing protocols for regulatory decision making, then selection will be restricted to tests for which a mechanistic

rationale exists to interpret the data. Several other factors are important when tests are selected for validation of screening tests *versus* adjuncts and replacements. As each test selection factor is discussed below, this distinction will be highlighted.

## III.B. Scientific Considerations

When selecting potential *in vitro* toxicity testing systems to be evaluated in a validation study, it is important to understand the fundamental nature of the predictive relationship of the proposed test. There are two basic philosophical approaches to developing and using experimental data from model systems to predict human toxicity. The correlative approach is basically to develop empirical mathematical relationships between variables in model systems and expected toxicological endpoints in man. The mechanistic approach is to use model systems to determine how the test chemical produces its toxicity, measuring the dose-response relationship for this effect and extrapolating the data from the model system to man to provide estimates of human risk under the expected conditions of exposure.

In the correlative approach, any biological or chemical variable whether it is measured experimentally or calculated theoretically, can be used to generate an empirical correlation. The mechanistic relationship between the predictor variable and the toxicological endpoint of concern is unimportant. A mathematical relationship is empirically generated by a set of learner compounds and the confidence intervals for this relationship are determined. A value judgement on the predictive power of the correlation is based on a measure of goodness of fit between the predictor variable and the toxicological outcome to be predicted. A good correlation is taken to mean that this relationship is a good predictor. The major drawback to this approach is that chemicals which act by different mechanisms are likely to produce different correlations between the predictor variable and the toxicological outcome. Experimentally this will become apparent when a large number of chemicals are studied and the graphical data segregate into distinct curves (Figure III-1). The inherent limitations of this approach become obvious when the toxicity of a totally unknown chemical is to be predicted. For this unknown chemical a particular value of the observed/calculated predictor variable is obtained. The question becomes which mechanism applies to this particular chemical, Mechanism 1 or Mechanism 2. The estimates of the toxicity of the unknown chemical will differ greatly depending on which correlation is used as the estimator. A further complication is that the unknown chemical may even produce its toxicity through a third, completely unknown mechanism and thus neither estimation will provide accurate predictions. This basic uncertainty in whether or not an unknown chemical falls within the domain of validity of a known estimator relationship results in the inability to determine *a priori* the reliability of predictions using a correlative method. Furthermore, this approach does not have the ability to identify when such an occurrence has happened. For these reasons, correlative toxicity is not a conservative basis for chemical safety evaluation.

This is not to mean that correlative toxicological data are of no value. They are extremely useful under two conditions: (1) if no other toxicological data are available, and (2) where benefits are perceived to outweigh the risks, such as screening methods in early stages of product development. The higher risk of incorrect predictions means that the incidence of false negatives and false positives can be high. A false negative prediction will have negative economic implications to the industry since a toxic chemical will be carried through extensive research and development only to be found to be non-viable as a marketable product. False positives will have negative impacts on society since a potentially beneficial product will not be developed. These are the risks that a decision maker must consider when using correlative toxicity evaluations.

The mechanistic approach is reductionist in nature in the sense that it is believed that the pathological effects of a chemical can be identified in the model system, the fundamental nature of the process determined, the dose-response for the reactions involved evaluated, the modulating

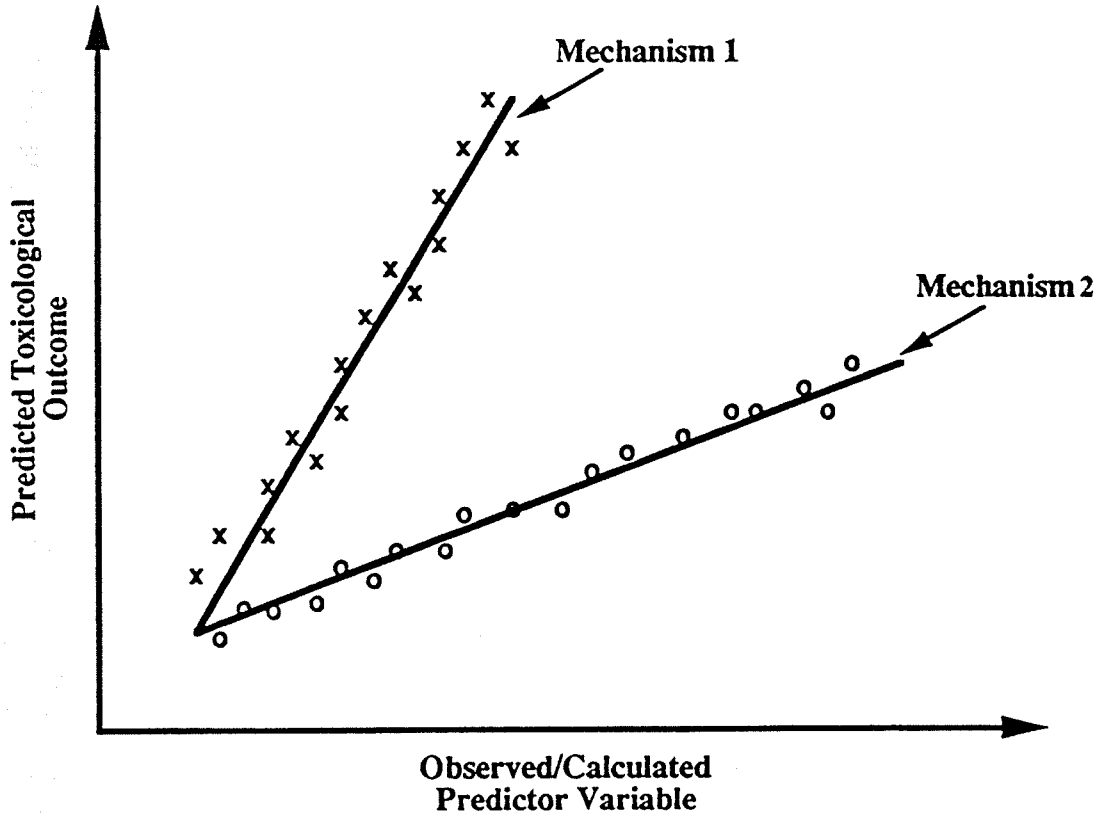


**Figure III-1: Correlative toxicity test**

A correlative toxicity test is established by mathematically determining the relationship between the predictor variable, which may be measured in a test system or calculated from chemical structure, and the toxicological endpoint in the organism of concern. A set of learner chemicals, *i.e.* chemicals with known toxicity, are used to establish this relationship. In the hypothetical example indicated in the figure, the chemicals tested fall into two classes which act through different mechanisms, thus giving two different relationships. Unless other factors can be used to determine into which of the two classes an unknown chemical falls, the ability to predict toxicity is confounded since two possible outcomes exist for each value of the predictor variable. The possibility that such circumstances can occur makes it difficult to determine the reliability of predictions using correlative testing approaches.

factors such as metabolism and repair taken into account and the toxicokinetics which relate exposure to target organ dose predicted. Given this information and making corrections for differences in these factors between the model system and man, the mechanistic approach contends that it is possible to estimate the risk for each possible toxicological endpoint. An advantage of this approach is that a scientifically based rationale exists for interpretation of test data, thus providing greater assurance of reliable predictions. The disadvantage is that in order to be quantitative this approach requires extensive toxicological and kinetic data input. Mechanistic *in vitro* tests imply that a rational relationship exists between the endpoint measurements in the test system and human pathology *in vivo*.

It is clear from this discussion that correlative tests can be used satisfactorily for screening purposes where the interpretability of the data is not critical for successful testing. Mechanistic tests can certainly be used for screening and thus enhance the value of data generated from screening activities. For adjuncts and replacements, mechanistic tests become essential. The greater the understanding of the relationship between the results of a mechanistically based test and the development of human pathology the more powerful the testing results become. Thus, when selecting tests for any validation study, consideration of the scientific basis of the test is important in relation to the objectives of the validation project.

### III.C. Economic Considerations

The cost of an *in vitro* toxicity testing methodology, both in monetary value and time, is an important consideration in the ultimate implementation of these methodologies in the safety assessment process. The criteria on which selection decisions are based again differ between screening tests and regulatory tests. For screening tests, which involve processing large numbers of samples, fast and inexpensive tests are essential. Therefore, when selecting tests for inclusion in a validation study for screening tests, it will be important to consider cost of: (1) test system — cells, media, culture vessels, (2) support systems — incubators, water baths, (3) endpoint analysis — reagents, instrumentation, and (4) technician time — for running the test, for maintaining the biological component of the test system and for calculating and reporting results. Since the time required for most *in vitro* test methods is usually under 48 hours, there is little basis for discriminating between methods on test time alone. However, time required to obtain and prepare the biological component of the test system for the actual test can differ significantly between test systems and may influence selection of one test over another.

For regulatory tests, the requirements for speed and cost are somewhat different. Clearly, rapid and inexpensive tests are the optimum, but fewer materials are carried through regulatory testing and at this stage, investment of more time and money has a bigger payoff — a product which is accepted by regulatory agencies. Therefore, more expensive tests with more predictive power can be considered for inclusion in validation of tests as adjuncts and/or replacements. By considering screening and regulatory testing separately, it becomes apparent that sophisticated *in vitro* testing methodologies based on high technology instrumentation can have an important role in *in vitro* toxicity testing, particularly in the regulatory arena.

### III.D. Logistical Considerations

To a certain degree, logistical considerations overlap with economic considerations since logistical requirements of the test system imply costs. Obtaining and preparing the biological component of the test system, whether it is a cell line supplied by various cell collections or primary cells, explants or tissue slices which are prepared directly from animals, can be important considerations if their availability is limiting. Furthermore, special requirements of the cell culture

system — serum, hormones growth factors — may be available in limited quantities. These factors must be carefully evaluated when testing systems are selected for validation studies.

These various factors must be considered in test system selection, yet because of their disparity, it is difficult to develop a formula which combines all of these factors to give a single measure of potential test value. At this time, all of these factors must be considered individually within the context of the objectives of the validation study.

## IV. CHEMICALS SELECTION FOR VALIDATION STUDIES

In setting up any validation program to evaluate *in vitro* toxicity testing methodologies for a specified toxicological objective, it is necessary to select sets of test chemicals to be utilized in the evaluation program. The basic scientific criteria for selection of test chemicals has not been fully established. The discussion below highlights some of the scientific considerations which are important for this selection process. In specific cases, overriding considerations of a practical or political nature may play a role in chemical selection. In these cases, the scientific consequences of basing decisions on these external factors must be carefully evaluated and documented.

Three sets of test chemicals must be selected for the overall validation process. The first set, referred to as the calibration set, is a small number of chemicals (n=10-40; see Section VII.B) which is used both in the intralaboratory assessment and as quality control standards for the first phase of reproducibility testing in the interlaboratory assessment. The second set of test chemicals (n=10-20), referred to as the intralaboratory evaluation set, is to be used in a blind study to provide the first full evaluation of the test potential. The third set of test chemicals is a large set of chemicals (n=50-100) which is used to fully define the performance characteristics of a test system. The basic criteria for selection of test chemicals for all three sets of test chemicals are similar, however the smaller number of test chemicals in the first two sets will influence how rigorously these criteria are applied in these cases.

There are basically six factors which must be taken into consideration for selection of test chemicals:

(A) the objective of the validation exercise,
(B) the available toxicological database,
(C) sufficient numbers of toxic and non-toxic test chemicals to obtain accurate estimates of false positives and false negatives,
(D) sufficient numbers of chemically related test chemicals to establish structure-activity relationships,
(E) a selection of formulations to investigate interactive toxicology, and
(F) additional test chemicals which are non-toxic in the target organ of concern but are toxic in other tissues (selective toxicity).

Not all of these factors have equal importance, and their ultimate influence on the final selection of chemicals will vary from one validation project to the next. Below is a discussion of each of these factors.

### IV.A. Objectives of the Validation Exercise

The set of chemicals selected for validation of *in vitro* tests for ocular irritation will differ significantly from those selected for validating test systems for evaluating hepatotoxicity. In the former case, chemicals selected will likely include alcohols, surfactants, detergents, preservatives,

propellants and colorants, *i.e.*, chemicals to which the eye is likely to be exposed. On the other hand, chemicals which would be selected for hepatotoxicity test systems would include many systemic drugs, chlorinated hydrocarbons, pesticides and well known hepatotoxins such as galactosamine and halothane. Thus, the set of chemicals selected for a particular validation study will depend to a large extent on the target organ toxicity under consideration.

### IV.B. Available Toxicological Database

Chemicals selected for inclusion in any validation study must be well documented with respect to expected *in vivo* toxicity. This issue is discussed in detail in Section V below. If a particular test chemical is to be included in the validation study, then, whenever possible, the database for that chemical must be compatible with the other chemicals selected. Complications in interpreting validation results will arise if toxicity data are not consistent, *e.g.*, toxicity data for all chemicals should be obtained in the same species, all data should relate to the same toxicity index (serum enzyme data or hepatic pathology data but not a mixture of the two endpoint data sets). It may be necessary to independently develop a comprehensive and compatible database for chemicals selected for the validation study before the exercise can be undertaken.

### IV.C. Distribution of Toxic and Non Toxic Test Chemicals

As will be discussed in more detail below (Section IX) an extremely important objective of the validation exercise is to determine the false positive and false negative rates for each candidate test. Much discussion has centered on the concept of prevalence — the fraction of the total number of test chemicals which are true toxicants as defined by the particular pathology under consideration. In actual fact, this factor is not particularly important in chemical selection for validation. It does become important in defining the predictive power of a test under practical testing circumstances — see Section IX. In the case of test chemical selection for validation, the more important consideration is to include a sufficiently large selection of test chemicals representing each toxicity classification in order to obtain statistically accurate estimates of the false positive and false negative rates. These rates can be used later to compute the predictive power of the test in the context of practical toxicity testing.

Another factor affecting the distribution of test chemicals selected for method validation is the need to include sufficient numbers of test chemicals which elicit toxic responses through known mechanisms. Subgrouping test chemicals by mechanism of action can be used to identify specific test systems which are predictive for particular mechanisms. Thus, a test may be highly predictive for chemicals which produce their effects by a certain mechanism and, yet, fail to identify toxic chemicals which act through other mechanisms. The apparent sensitivity of the test (its ability to give a positive test outcome for a toxic chemical) will then depend on the fraction of all toxic chemicals selected for the validation study which act through that particular mechanism of action. Thus, a test system which exhibits a relatively low overall sensitivity, yet gives a high sensitivity for a subgroup of chemicals which act through a common mechanism is a potentially useful test because of its ability to segregate toxic chemicals on a mechanistic basis. The opportunity to identify such test systems in a validation study will depend on the distribution of chemicals selected.

### IV.D. Structure-Activity Relationships

In the later phase of validation, where large numbers of chemicals are being evaluated, it is scientifically justifiable to include series of test chemicals with defined structural relationships in

order to elucidate structural factors which contribute to the degree of toxicity as well as tissue specificity of toxicity. Although this consideration is not essential to the valid design of a validation study, in the long run attention to structure-activity relationships will provide valuable data for developing predictive toxicity methods. Within this context, it is important to include pairs of chemicals which are structurally related, yet exhibit significant differences in toxicity. Such pairs of chemicals can be used to determine the ability of test systems to discriminate between chemically similar yet toxicologically different products.

### IV.E. Formulations

Since many commercial products are in fact complex formulations of various substances, there is significant interest in determining the ability of *in vitro* test systems to correctly predict the toxicity of these complex mixtures and products. However, at the initial phases of new test validation the tendency to use formulations as test substances should be resisted until the basic performance of the test is determined. Testing of formulations should only be introduced into the validation process after purified components are tested individually. In this manner, the ability of the test system to evaluate complex formulations can be properly determined. The nature of formulations will be highly dependent on the particular product lines of specific industries. Thus, the validation phase dealing with formulations should be tailored to each special interest group.

### IV.F. Selective Toxins

When validating alternative testing systems for specific target organ toxicity evaluations — neurotoxicity, hepatotoxicity, nephrotoxicity, etc. — it is important to determine whether the new methodologies can discriminate against toxins which have selective toxicity for other target organs *e.g.*, a primary neurotoxin should give a negative response in a test for hepatotoxicity. Thus, when selecting test chemicals to carry out an evaluation of target organ predictability, it is important to include chemicals which are active toxicants in the target organ of concern, chemicals which are non-toxic in general and chemicals which are active toxicants in organs other than the target organ of concern. Chemicals which fall into the latter category should be selected on the basis that they exhibit their selective toxicity as a consequence of tissue specific sensitivity, not as a result of unique systemic toxicokinetics. *In vitro* toxicity test systems should not be expected to replicate *in vivo* kinetic phenomena. Using a collection of such test chemicals, it will be possible to determine the ability of the test system to identify correctly target organ specific toxins.

These six factors should be taken into consideration when selecting test chemicals for validation studies. The overall objective of any specific validation exercise will determine which of these factors are most important.

## V. REFERENCE CLASSIFICATION OF CHEMICALS FOR VALIDATION

### V.A. Primary Standards — Human Databases

The term reference classification refers to the accurate classification of the set of chemicals selected for the validation process as to their *in vivo* toxicity. Using carcinogens as an example, this means to correctly identify whether or not a specific chemical is a carcinogen. Thus, each chemical used in the validation of an alternative test for carcinogenicity must be so classified — carcinogen or non-carcinogen. (Note, this is a quantal classification, either a chemical is or is not a carcinogen.) In the case of toxicity, the question arises as to whether the database selected for

evaluating alternative testing methodologies will be the human toxicity database, which is most relevant to the main objective of safety evaluation — predicting human toxicity of chemicals — or the more readily available animal database.

There are several important points which must be taken into consideration in selecting the database used to define the reference classification. First and foremost, does a quantitative human toxicity database exist? If such a human database does exist, then this would create the ideal situation. Since a human toxicity database is directly relevant to validating alternative methods, they will be referred to as primary databases. Unfortunately, this information is, in general, not readily available. For carcinogenicity, such data is sparse. Epidemiological studies have been able to relate exposures and carcinogenicity, (*e.g.*, vinyl chloride and angiosarcomas, asbestos and mesotheliomas, and chromium and skin cancer) but uncertainties exist for other suspected carcinogens and weak carcinogens which do not produce unique cancers are difficult to detect. Thus, most validation approaches for alternative carcinogenicity testing methods have relied on the rodent carcinogenicity database which is referred to as a secondary database here. These databases have their own set of confounding factors, such as species specific responses, *e.g.*, a rat carcinogen is not necessarily a mouse carcinogen and vice versa. Thus, whichever database is chosen to define the carcinogenicity of a test chemical in the validation study, difficult problems of interpretation result. Similar problems exist for other categories of toxicity.

### V.B. Secondary Standards — Animal Databases

The predictive power of an alternative test can be directly compared to a secondary animal database. This, in effect, establishes the validity of an alternative test in terms of its ability to reproduce the results of an *in vivo* animal test. Thus, an alternative test which predicted the exact same results as an *in vivo* test, *i.e.* 100% correlation, would be validated as being "as good as the *in vivo* test". Whether or not this result would imply 100% correlation with human toxicology would depend on how well the *in vivo* animal test predicts human responses for the set of test chemicals selected for validation. In most cases of toxicity responses, it can be shown that *in vivo* animal testing is not 100% predictive of the human situation when dealing with the universal set of all chemicals. The exact predictability of the human toxicological response by *in vivo* animal testing has not been determined for most categories of *in vivo* toxicity. However, for a subclass of the universe of all chemicals, sufficient data for both *in vivo* animal testing and human toxicological responses is available to confirm the accuracy of the *in vivo* animal toxicity test predictions. Thus, when comparing alternative testing methods to *in vivo* animal tests, whenever possible the chemicals selected should be restricted to those for which sufficient data are available to substantiate that the animal test data are accurate predictors of human response.

## VI. TECHNICAL PROBLEMS ASSOCIATED WITH VALIDATION STUDIES

### VI.A. Confounding Factors

*In vitro* cell culture systems are a unique combination of a biological component (cells or tissues), fluid phase (culture media) and solid support (glass, plastic and/or artificial membranes). This combination of components is designed to provide a system in which the biological component can be maintained in a defined environment which supports the growth and maintenance of particular cells in a well characterized, differentiated state. The physical, chemical and biological factors necessary to attain this condition are well developed for many cell lines, but are less well understood for many primary cultures of human, primate and rodent cells. Significant research efforts are devoted to the problem of maintaining differentiated, primary cultures of