13th Meeting of the Scientific Group on Methodologies for the Safety Evaluation of Chemicals (SGOMSEC): Validation and Acute Toxicity Testing

Rodger Curren, Leon Bruner, Alan Goldberg, and Erik Walum

1Institute for In Vitro Sciences, Gaithersburg, Maryland; 2The Procter & Gamble Company, Staines, Middlesex, United Kingdom; 3Center for Alternatives to Animal Testing, Johns Hopkins University, Baltimore, Maryland; 4Department of Cell Biology, Pharmacia & Upjohn, Stockholm, Sweden

Scientific principles demand that before newly developed alternative methods for safety testing are fully embraced by the industrial or regulatory community, they reliably and reproducibly predict the designated toxic end point. The process used to determine reliability and reproducibility is termed validation, and it generally culminates with a highly controlled, blinded study using multiple chemicals and laboratories. It is imperative that the validation study is designed to confirm the previously established reproducibility and predictive power of the assay. Much has been learned recently about the practical aspects of validation through investigation of alternative methods for acute toxicity testing, i.e., those methods that assess acute systemic toxicity, skin irritation, and eye irritation. Although considerable progress has been made—many alternative tests are now commonly used in various industrial settings—there are few tests that have successfully passed a complete validation. Some of the barriers to successful validation have been a) lack of high-quality, reproducible animal data; b) insufficient knowledge of the fundamental biologic processes involved in acute toxicity; and c) the development of truly robust in vitro assays that can accurately respond to materials with a wide range of chemical and physical characteristics. It is recommended that to progress in the areas of eye and skin irritation we need to expand our knowledge of toxic markers in humans and the biochemical basis of irritation; progress in the area of acute systemic toxicity will require the development of in vitro models to determine gastrointestinal uptake, blood-brain barrier passage, and biotransformation.

Key words: alternative methods, toxicology, in vitro methods, reduction, refinement, replacement, testing, ocular irritation, dermal irritation, acute toxicity, LD50

Introduction

In applying scientific principles to the problem of determining acceptability of new, alternative methods for safety testing, there is a need for a standard validation process. This validation process is described and referenced in this document. Scientific methodologies have been validated throughout the history of science by less structured but scientifically rigorous approaches. This validation has been accomplished during the development of individual methodologies, publication of results in peer-reviewed journals, repetition of studies by independent laboratories, and acceptance by widespread use of the method. This less structured, but rigorous and scientifically acceptable approach will and should continue. However, it has become clear that more rapid progress will result if a more structured, formal approach to validation is adopted. This would include having a set of explicit criteria to use in the evaluation of any new method. As methods are reviewed for regulatory use, supportive data for compliance with how each method meets well-established criteria will provide a more rapid and effective review.

Use of Alternative Methods in Acute Toxicity Testing

Considerable research has been directed toward the development and validation of alternative methods for acute toxicity testing. Acute toxicity tests assess acute oral toxicity, skin irritation, and eye irritation. These tests are important for assessing chemical hazards due to a single short-term exposure to a chemical. Historically, these assessments have been conducted in whole animals with the assumption that the results can be extrapolated to humans. For several important reasons discussed throughout this document, it would be useful to replace these animal-based tests with nonanimal alternative methods. A scheme for the inclusion of alternative methods in a typical safety assessment process is shown in Figure 1. The first step in such processes usually includes a review of the historical data available on the chemical and an assessment of its physical-chemical properties. In many cases, quantitative structure–activity relationship (QSAR) models may be useful for such assessments. In some cases it is possible to characterize the chemical hazard based on these data alone. If more data are needed, the next step in the process is to conduct testing in some type of in vitro biological system. The new data are integrated with any other available information, and again, it is often possible to characterize the chemical hazard based on the total data package. If it is still not possible to complete the hazard assessment, the next step in the process is to determine if human data are needed. If so, it is necessary to consider whether the available information is sufficient to allow controlled human exposure. If the data are sufficient, and if the testing is deemed ethical after independent ethical review, controlled studies may be undertaken. If the
data are insufficient, testing in appropriate animal models should be conducted. When all data are available, the hazard characterization may be completed, and human exposure may be allowed depending on the results of the safety assessment.

**Evolution of Alternative Methods forAcute Toxicity Testing**

Experience in developing and optimizing in vitro assays for acute toxic end points has revealed a general pathway through which most tests pass on their way to general acceptance by the scientific community. The pathway is generally characterized as test development → optimization/prevalidation → validation → acceptance → routine use. These stages can be defined as follows:

- **Test development:** Identification of a toxicological need, developing a model, defining end point measurements, and characterizing output.
- **Optimization/prevalidation:** Further definition of operating parameters, standardization of protocol and standard operating procedures (SOPs), defining appropriate controls, obtaining evidence of transferability and reproducibility, and defining a prediction model.
- **Validation:** Generally a formalized method of independently evaluating the reproducibility and relevance of a method. Often conducted with coded materials and in multiple laboratories.
- **Acceptance:** May indicate formal regulatory acceptance or acceptance by a set of users after appropriate peer review of the validation data.
- **Routine use:** Standard use of the defined assay for its specific purpose as defined within its prediction model.

This progression can occur on many different levels, for example, within a single laboratory, within a company, or at a more general national scale where it may lead to full regulatory acceptance. However, validation of the test is necessary at any of these levels if other researchers are expected to accept the usefulness of the method and use it on a routine basis.

**Lessons Learned**

Much attention has been given to the development and validation of alternative methods for acute toxicity testing in recent years. There are three reasons for this. The first is to improve the quality of the data obtained from acute toxicity tests. This is necessary for toxicologists to better predict and manage the risks associated with use of chemicals under defined conditions. Second, it is hoped that it is possible to develop methods that are less expensive and less time consuming. This will certainly be true for complex long-term studies such as 2-year bioassays. However, it is likely that alternative methods for other types of studies, such as some acute toxicity tests, will be more costly and may take more time. Finally, acute toxicity tests may be highly stressful to test animals. This has led to criticism of the procedures on ethical grounds. The development of nonanimal methods would allow toxicologists to obtain necessary information without the need for animals in the safety assessment process.

Numerous methods have been proposed as viable alternatives, and several large validation studies have been conducted to assess the ultimate utility of these procedures. Although these studies have not yet led to both the validation and acceptance of an alternative method, much has been learned about the test methods and also about the procedures that should be used to assess their acceptability. It is, therefore, useful to provide a brief overview of the most important lessons learned from these efforts.

**Lessons Learned about Alternative Method Data Development**

The development, optimization, prevalidation, and validation of an alternative method for toxicity testing is a long and complicated process because many of the mechanisms leading to toxic reactions are not well defined. This lack of understanding makes it difficult to develop tests that provide relevant information needed for making correct decisions in the safety assessment process.

The basic research needed to develop mechanistic understanding takes significant time. Often this time exceeds the patience of groups seeking the end of animal testing through political means, and the extended effort consumes resources of commercial developers who attempt to market new test methodologies. Pressures brought by groups desiring rapid progress are not unlike those observed in many other fields, i.e., at times the pressures facilitate the ability to focus resources on the problem. However, such pressures must not lead to shortcuts that not only could damage the scientific credibility of new methods, but could even seriously endanger human health and safety.
Lessons Learned about Evaluation of Data from Validation Studies

Validation must be a disciplined process. Validation studies must be designed so that they test clearly stated hypotheses and allow judgment of alternative method performance relative to predefined success criteria. This is important whether validation studies are conducted in a coordinated fashion or conducted independently in several unrelated laboratories. Once completed, the results from a validation study should be subjected to appropriate peer review. Developers of methods should present their work in forums where the results and conclusions can be openly debated. Constructive criticism should be given and accepted in the spirit of improving the scientific quality of new methods. For example, the careful review of instances where a method has failed may lead to greater understanding of basic mechanisms and improved alternative methods.

Lessons Learned about the Technical Aspects of Validation

Much has been learned about the development and validation of alternative methods from a technical point of view. First, it is important that alternative methods be fully optimized before they are assessed in validation studies. Scientists leading validation studies must assure that clearly defined protocols and SOPs are available, that prediction models have been defined, and that the method is relevant for its intended use before a validation study is started. Second, attempts to define criteria of success have reminded us that the data obtained from in vivo tests may be less precise and accurate than desired. Scientists responsible for developing alternative methods must find ways to overcome the limitations imposed by the technical inadequacies of current tests. The use of human data in some limited, ethically acceptable cases may address this issue to some extent, but will not solve the overall problem. Finally, there is a need to increase our understanding of toxicity mechanisms and then develop new methods based on this understanding. A sound mechanistic basis increases confidence in predictions obtained from such tests and ultimately leads to the generation of better data for risk assessments.

Lessons Learned about the Adequacy of Data Sources for Validation Studies

Human toxicity data can be found in handbooks, the scientific literature, and in the databanks of poison information centers throughout the world; however, these data are not standardized or normalized. Nonetheless, there is adequate information on the acute systemic human toxicity of a number of compounds to allow alternative method development, prevalidation, and validation to be conducted with human data as the standard.

Current Situation

Validation

The purpose of the validation process is to provide independent confirmation that an alternative method provides correct information needed for making risk assessments or for other safety-related decisions. Accordingly validation must be considered a confirmatory process, and not a process involving test method development or test method optimization. The following paragraphs provide a brief review of the steps that should be considered in the validation process. The discussion is centered around the flowchart depicted in (Figure 2) and is based on information contained in reviews of validation prepared by international bodies such as the Organisation for Economic Co-operation and Development (OECD) (1) and from the scientific literature (2-4).

Definitions. Validation has been defined as the establishment of the reliability and relevance of an alternative method for a specific purpose (2-4). Reliability refers to the demonstration of reproducibility of data obtained from a test method in the laboratory and the reproducibility of the predictions of toxic hazard following application of a clearly stated prediction model to the alternative method data (5). Relevance refers to the establishment of scientific meaningfulness and usefulness of results from an alternative method for a particular purpose (2-4). The establishment of usefulness and meaningfulness is important because hazard predictions obtained from scientifically credible alternative methods have a higher probability of being correct.

Validation Process. CONFIRMING THAT AN ALTERNATIVE METHOD IS DEVELOPED SUFFICIENTLY TO ALLOW AN ASSESSMENT OF RELIABILITY AND RELEVANCE. The validation of an alternative method cannot be undertaken until it has been sufficiently developed in the pre-validation stages. Factors that should be considered during prevalidation have been reviewed by Curren et al. (6). The most important are a) there should be strong evidence that the alternative method is relevant for the intended purpose; b) there must be clearly written protocols and SOPs that describe how to conduct a test method in an appropriately equipped laboratory; c) there should be evidence that data obtained from the method can be generated reproducibly across several laboratories; and d) there must be a prediction model available that allows correct interpretation of results from the alternative method.

Importance of the Prediction Model. Experience gained during the conduct of recently completed validation studies has highlighted the importance of the concept of the prediction model. In order for an alternative method to be useful for making safety assessments, it must be possible to translate its results into correct predictions of in vivo toxicity. This is usually done by applying algorithms to the alternative method data that converts them into toxicity predictions. These algorithms are usually developed from experimental data generated during the prevalidation stages of development. Since such algorithms constitute models that allow the prediction of toxicity, they have been called prediction models (5). If an alternative method does not have an adequate prediction model, there would be no way to use it in the safety assessment process. It is therefore essential that validation programs test the utility of the prediction models. In fact, if there is no prediction model, there can be no validation study (7).

Measuring the Reliability of an Alternative Method in a Validation Study. Factors that must be considered in the design and conduct of a validation study are listed in Figure 2. The validation study must be designed based on a knowledge of the data to be provided by the study. The participating laboratories must be identified and recruited. A reference set of test substances must be assembled, and the quality of the in vivo data must be assessed. Once these steps have been completed, each test substance must be evaluated in the alternative methods being evaluated in the program. Finally, when the data from the laboratory work are available, the prediction model must be used to predict the in vivo toxicity of each test substance. If the toxicity predictions are similar to the actual toxicity of the test substances, and if the same results were obtained in each of the participating laboratories, it would provide evidence that the method is reliable. If, however, the toxicity is not predicted correctly or if the results are not similar across the participating laboratories, it would not be possible to

Environmental Health Perspectives • Vol 106, Supplement 2 • April 1998 421
consider the method reliable. If the alternative method is not reliable, the validation process should end so that further method development may be undertaken (Figure 2). If the optimization process is successful, the method may be evaluated in a subsequent study. Alternatively, the alternative method may be abandoned if additional work is unlikely to be fruitful. If the alternative method is reliable, the next step in the validation process is to assess its relevance (Figure 2).

**Figure 2.** The validation process. The flow chart depicts the series of steps that may be used as a guide to design and conduct a validation program. The steps proceeding down the left side of the chart represent the validation process. The steps proceeding up the right side of the chart depict the steps associated with improving the performance of the alternative method and defining another prediction model prior to inclusion of the method in a subsequent validation study. Any new method, whether it is based on a fundamental understanding of toxic mechanisms or based on empirical correlations may be assessed for validity using this approach. From Bruner et al. (5).

Assessing the relevance of an alternative method. As noted above, an alternative method may be considered relevant when the predictions of toxicity obtained are meaningful and useful for a specific purpose. The establishment of relevance is a judgmental process that requires the evaluation of all the available supporting data and scientific evidence supporting the use of an alternative method. It may also involve evaluation of key performance benchmarks that can provide a useful context for interpreting the results obtained from a validation study.

Establishing performance benchmarks based on performance characteristics of the alternative method and the tests it will replace. One of the approaches that can be used for assessing relevance is to estimate the theoretical best performance that may be expected from an alternative method. Ideally, there should be a good correlation between the in vivo and alternative method data, and a narrow 95% prediction interval associated with each prediction of toxicity. However, technical limitations in the in vivo tests and alternative methods most commonly prevent this ideal from being reached. Computer simulations based on an understanding of both the alternative method and the in vivo test to be replaced may be used to provide guidance on the performance levels that may be expected in a validation study. Practical examples demonstrating how this may be done have been described elsewhere (5).

Another important benchmark to consider is a comparison between the performance of the alternative method with the performance of the method that will be replaced. The predictive capacity of the alternative method should be equivalent or greater than the method that will be replaced. Again, computer simulations may provide data-based guidance useful for judging the acceptability of predictions from an alternative method. A practical example showing how computer simulations may be conducted has been presented elsewhere (5).

Other factors to consider in assessment of relevance. In addition to the performance benchmarks given above, it is important to consider other factors supporting the relevance of an alternative method. One of the most important is a consideration of the mechanistic basis of an alternative method. This is important because a strong mechanistic basis often supports the rationale for an alternative method, and helps increase confidence that the predictions from a new test will be correct. It is also important to define the known limitations in the use of an alternative method. For example, a new procedure may be valid for only a small number of substances relative to the universe of materials that must be tested. If the method is limited in its application, it may not be relevant for general use in the safety assessment process. The technical limitations of an alternative method must also be known. An assay that can handle all types of test substances may ultimately be more relevant for general use than one restricted to only one type (e.g., water-soluble test materials). Finally, the scientific literature concerning the performance of the alternative should also be taken into account.

Once all of this information has been assembled and evaluated, the overall relevance of the method for its defined purpose must be assessed. If the conclusion is
that the alternative method is not relevant, the test cannot be considered valid and it is necessary to consider whether there is value in optimizing the assay, developing a new prediction model, and assessing it in a subsequent validation study (Figure 2) (1). Conversely, if the data support its relevance, then it would suggest the alternative method can be used in the safety assessment process and should be considered for official acceptance by regulatory authorities (Figure 2) (1).

**Ocular Irritation**

**Status.** In recent years many methods have been proposed as alternatives to the animal ocular irritation test. To date, none of these have been shown to be complete replacements for the existing animal test. The failure to find a complete replacement is likely due to several reasons:

- The eye is a complex organ made up of multiple tissues, each of which responds somewhat differently to injury. Current animal tests for ocular irritation use a complex scoring system involving a combination of data from three important ocular tissues.
- *In vitro* tests have generally been designed to model only one, or just a few ocular tissues, not the whole eye. This can be very helpful in obtaining mechanistic information, but it is a disadvantage if the *in vitro* tests are required to accurately reproduce Draize scores that reflect total eye injury.
- Validation of *in vitro* ocular irritation assays will be difficult because the animal test itself is not very reproducible and because compounds listed in the historical databases may no longer be available.
- Many *in vitro* tests, by their physical nature, may be unable to cope with materials that are solids, water insoluble, have extremes of pH, or cause damage by direct physical means (e.g., by abrading the cornea).

However, successes have been obtained either by using natural or reconstructed tissue equivalents that allow direct application of test material to target cells [e.g., tissue equivalent assay (8) or the bovine corneal opacity and permeability assay (9)], or when use of the test system is restricted to specific conditions such as using the fluorescein leakage assay for very mild surfactant-containing materials.

**Strategies for Use.** *In vitro* ocular irritation assays can be most successfully utilized by referring to Figure 1. There may be sufficient information available in step 1 that allows a decision to be made, e.g., extreme pH coupled with high acid or alkali reserve may lead to a classification of severe irritant. If a decision cannot be made at this point, an appropriate ocular alternative method is conducted (step 3), if one is available. It is extremely important that only appropriate methodologies are used, i.e., those that have been well evaluated (validated) for the specific type of test material and degree of irritant response that could be expected, and that have well-defined prediction models.

The resulting information can then be used in conjunction with information gathered in step 1 to attempt to make a hazard evaluation. It is possible that this information is sufficient; indeed some safety decisions are made at this stage today. However, it may be that an appropriate *in vitro* test does not yet exist for the type of test material being examined—or that the information obtained from integrating the available data is still not sufficient for a decision. Animal tests (or human tests, if appropriate) may then need to be conducted with the knowledge that only a very limited study may be necessary to supply the missing information or to add to the confidence in the previous data. For example, study of a single rabbit may be sufficient to add the confidence necessary to label a material as nonirritating.

**Barriers to Progress.** Several barriers to progress still exist. The major ones are:

- Lack of high quality *in vivo* data (animals or human) with which to calibrate and subsequently validate *in vitro* assays and their associated prediction models.
- Insufficient knowledge of the processes involved in human eye injury. This prevents improvement of current tests so that they more accurately predict human hazard.
- Robust *in vitro* models that can accommodate test materials with wide-ranging physical characteristics and potential toxicity.
- An adequate understanding of recovery from eye injury, a process not examined in any of the existing tests. Knowledge of this phenomenon is very important to risk assessment.

**Recommendations**

- Expand knowledge of toxic markers in humans, especially those that are predictive of permanent injury, and develop *in vitro* tests in which the same toxic markers can be observed.
- Continue research into characterizing models where there is direct application of test material to the target cells.
- Conduct parallel *in vitro* testing whenever an *in vivo* animal eye test is conducted. Create a system that will allow these paired data and the identity of the tested chemical to be easily available to researchers in the area of alternative models.

**Dermal Toxicity**

Dermal toxicity expresses itself either as irritation or allergic contact sensitization. Irritation (acute, primary, delayed, or traumatic) can occur rapidly or over years and can exist as mild erythema through various degrees of severity resulting in disruption of the integrity of skin, i.e., corrosion. Allergic contact dermatitis is a cutaneous T-lymphocyte-mediated reaction to exogenous chemicals.

Phototoxicity is a specialized case of either irritancy or allergy but requires light activation of the toxic chemical.

Numerous factors can modify the biological response of the skin to exogenous chemicals. Among the most important are age, gender, race, and preexisting and previous skin diseases.

**Current Systems for Evaluation.**

Historical animal and human data, QSAR, and physical/chemical assays can provide excellent data for initial evaluation and an appropriate road map to identify necessary studies. In some cases, if sufficient data exists, the material can be evaluated without further testing. In other cases it may be possible to directly use these materials in human testing.

A number of *in vitro* biological methods are currently being used and/or evaluated for their usefulness. These include cell culture, reconstructed tissue equivalent (RTE), and skin explants.

For irritation a number of biological end points are used as a measure of cellular toxicity. Specific biochemical markers interleukin-1 (IL1), arachidonic acid, and the prostaglandins provide information on this inflammatory process. Histopathology and noninvasive techniques (e.g., transepidermal water loss, CO2 transport) are being evaluated mainly in RTE and in skin explants.

A potential strategy for irritation can be evaluated by using the general scheme in Figure 2. Step 1 includes physical and chemical properties, literature review, QSAR, historical data, and physical/chemical assays.

Once the data are integrated and evaluated, then *in vitro* methods can be used (step 4). These could include cytotoxicity (e.g.,
MTT reduction end point), biochemical measures, and histopathology. After further integration and evaluation of all data, it will be possible, in many cases, to go directly to human testing (step 9).

Corrosive testing, the extreme case of irritation, may be evaluated completely in step 1 as described above. In other cases histopathology, biochemical, or Corrositex® testing may be required. If a known or suspected corrosive material is being evaluated for degree of severity, animal or human testing is inappropriate.

For allergic contact dermatitis (ACD) an understanding of the biological basis will provide appropriate mechanistic tests (10). Specific approaches are currently under evaluation and one anticipates rapid understanding and significant methodology development. The biology of ACD suggests that specific cytokines (and/or their messenger RNA), adhesion molecules, and histochemistry will provide appropriate batteries of tests.

The biological systems most likely to be used include keratinocyte cell cultures, dendrite cell model (Langerhan cells and/or blood monocytes), and RTE. The results of ongoing QSAR studies are most encouraging. In fact, one might anticipate that QSAR and a single or limited battery of in vitro tests will provide sufficient data to identify and manage the risk of skin contact allergens.

A potential strategy for ACD using the general scheme in Figure 1 is step 1: QSAR, historical data, and literature; step 4: cell or RTE with measurements of an adhesion molecule, interleukin-8 (IL8), and histochemistry.

It is anticipated that in many cases results from this approach will provide the appropriate information to describe and manage the risk of these chemicals.

Next Steps and Current Needs. There is a need to develop: a) more completely defined biological systems necessary to study dermal toxicity; b) better and more accessible preparations of Langerhan cells and/or model dendritic cell systems (e.g., blood monocytes); c) an expanded knowledge base for adhesion molecules so that relationships between exogenous chemicals and changes in adhesion molecules are understood; d) an understanding of the relationships of cytokines and other mediators of the inflammatory process with exogenous irritant chemicals; and e) an understanding of the interaction and time relationships of the numerous cytokine and adhesion molecules.

Acute Oral Toxicity
Data on acute oral toxicity are produced for reasons that include obtaining information on the biological activity of a chemical and gaining insight into its mechanism of action. Usually long-term studies begin with a dose-finding exercise under acute conditions. Furthermore, the information on acute systemic toxicity generated by the test is used in hazard identification and risk management for the production, handling, transportation, and use of chemicals. The LD₅₀ value (precise or approximate) is currently the basis for toxicologic classification of chemicals and is thus required by the regulatory authorities in different situations.

Status. Today all tests for acute systemic toxicity rely on the use of laboratory animals. The original LD₅₀ test (using a large number of animals in each dose group, two species, and both sexes) is outdated; there are no scientific or legal reasons for performing the test. Nevertheless, accepted new methods that fulfill both the reduction and refinement criteria for alternatives are based on animal studies. Furthermore, the older, more animal intensive alternatives are used more frequently than the newer, more humane methods.

The general opinion is that nonanimal methods, addressing the replacement criteria, are not likely to come into use in acute toxicity testing in the near future.

The fact that correlations can be shown between the blood concentration of a chemical that is lethal to man and the concentration that kills a defined fraction of the population of cells in culture indicates that in vitro alternatives are possible to develop. These postulated nonanimal methods will inevitably have to be more complicated than just simple cytotoxicity measurements. It is already clear that the prediction of human toxicity by cytotoxicity measurements is greatly improved by the introduction of biomarkers information in the prediction model. Such information can be obtained from in vivo or in vitro studies or physiology-based computer models of kinetics. However, if one considers the possibility of replacing some isolated aspects of acute systemic testing, e.g., dose-finding or quantitative determinations (LD₅₀ value finding), cytotoxicity determinations represent at least partial alternatives. There is obviously a role for in vitro assays in acute toxicity testing as outlined in Figure 1.

Strategies for Use. Acute systemic toxicity should be seen in the context of the stepwise procedure presented in Figure 2.

The collection (step 1) and integration (step 2) of information on physical and chemical properties of a compound, literature reviews, and an analysis of the structure–activity relationships, when possible, are always desirable before any testing of toxicity begins. However, this information will not be sufficient to allow a definite judgment on the systemic toxicity of the compound (step 3); biological tests will still be necessary. It is therefore recommended that acute systemic toxicity testing (step 4) begin with the determination of general cytotoxicity in an in vitro system. Data on gastrointestinal uptake, on the penetration of the blood–brain barrier, and on biotransformation may be obtained using in vitro models. Furthermore, this information can be reinforced and supplemented (distribution and elimination data) by the use of computer-based biokinetic modeling. When integrated (step 5) with steps 1 and 2 the in vitro results may provide sufficient information for an evaluation of hazard (step 6). If not, the animal studies (step 8) can be performed with a wealth of background information and thus a minimum use of laboratory animals. When the hazard characterization is completed, it may be appropriate to perform human studies of acute systemic toxicity (step 7) as in the case of drug development, where confirmation of the lack of toxicity in the therapeutic dose range is required.

Barriers to Progress. To successfully perform acute systemic toxicity testing according to the strategy outlined above, the following problems have to be addressed:

• There is insufficient information on structure–activity relationships with respect to acute systemic toxicity. This may be explained by the fact that a large number of mechanisms participate in the expression of this type of toxicity (see organ toxicity section). Substantial additional investigation of the cause of chemically induced lethality is needed.

• The in vitro models used to determine gastrointestinal uptake, blood–brain barrier passage, and biotransformation are currently and frequently used in pharmacology and toxicity studies. However, these methods have not been formally validated. This is needed to identify the most reliable ones.

• Projects are ongoing to establish procedures for the integration of in vitro data into hazard characterization. However, more efforts must be directed to promote cooperative programs in the
development of alternatives to acute toxicity testing.

- In spite of a number of in vivo alternatives for oral acute toxicity testing, the most frequently used are the more traditional, animal-intensive methods. Therefore, active measures must be taken to phase out older methods and to promote more humane alternatives.

Conclusions Regarding Acute Toxicity Testing in General

- A great deal of effort has been directed toward development and validation of alternative methods for acute toxicity testing.
- Validation and acceptance have occurred on many different levels, e.g., within single laboratories and within individual companies. These methods are routinely used in the risk assessment process.
- Some alternative tests for reduction and refinement have been validated and accepted by international authorities (OECD). Examples of new methods include the fixed-dose procedure and limit test for acute oral toxicity testing, and the refinement of guidelines for eye irritation testing.
- Some alternative methods have been validated for limited use in the risk assessment process. Some of these methods are used routinely.
- No single battery of alternative assays has been validated as a complete replacement for an acute toxicity test.
- This substantial progress in the successful use of QSAR, cell and molecular biology, computer modeling, and statistics leads us to conclude that use of nonanimal alternative methods is a strategy that improves acute toxicity testing and the safety evaluation of chemicals.
- Recently completed validation studies have led to substantial refinements in the validation process. These improvements will allow faster, more efficient, and objective assessment of alternative methods in the future.

Recommendations

- Continued support of the development and validation of alternative methods is essential and is expected to lead to continued improvement in the safety evaluation of chemicals.
- Research should be directed toward a better understanding of the relevant biology (especially human) of the toxic events and mechanisms.
- Development of better methods for synthesizing the information obtained from the battery of alternative tests to improve the interpretation of the integrated results is needed.
- Users of alternative methods should adopt the general scheme for toxicity testing (Figure 1).
- Validated and accepted alternative methods should be immediately utilized. Implementation of methods providing equivalent information should be prioritized according to the degree that they use fewer animals or cause less stress.
- Given the expected rapid progress in the development of new alternative methods, it is recommended that scientists, regulators, potential users, and policy-makers establish and use continuing education and information distribution programs to stay current with progress.

References and Notes

1. OECD. Final Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods. 1966. Paris: Organisation for Economic Co-operation and Development, 1996.
2. Frazier JM. OECD Environment Monograph 36: Scientific Criteria for Validation of In Vitro Toxicology Tests. Paris: Organisation for Economic Co-operation and Development, 1990.
3. Balls M, Blaauboer B, Brusick D, Frazier J, Lamb D, Pemberton M, Reinhardt C, Roberfroid M, Rosenkranz H, Schmid B et al. Report and recommendations of the CAAT/ERGATT Workshop on the Validation of Toxicity Test Procedures. ATLA 18:313–337 (1990).
4. Balls M, Blaauboer B, Fentem JH, Bruner L, Combes RD, Ekwall B, Fielder RJ, Guillouzo A, Lewis RW, Lovell DP et al. Practical aspects of the validation of toxicity test procedures. The report and recommendations of ECVAM Workshop 5. ATLA 23:129–147 (1995).
5. Bruner LH, Carr GJ, Chamberlain M, Curren RD. Validation of alternative methods for toxicity testing. Toxicol in Vitro 10:479–501 (1996).
6. Curren RD, Southee JA, Spielmann H, Liebsch M, Fentem, Balls M. The role of prevalidation in the development, validation and acceptance of alternative methods. ATLA 23: 211–217 (1995).
7. Bruner LH, Carr GJ, Chamberlain M, Curren RD. No prediction model, no validation study ATLA 24:139–142 (1996).
8. Osborne R, Perkins MA, Roberts DA. Development and intralaboratory evaluation of an in vitro human cell-based test to aid ocular irritancy assessment. Fundam Appl Toxicol 28:139–153 (1995).
9. Gautheron P, Dukic M, Alix D, Sina JF. Bovine cornea opacity and permeability test: an in vitro assay of ocular irritancy. Fundam Appl Toxicol 18:442–449 (1992).
10. Karol MH. Target organs and systems: methodologies to assess immune system function. Environ Health Perspect 106(Suppl 2):533–540 (1998).