Alternatives to Toxicity Testing in Animals: Challenges and Opportunities

George P. Daston and Pauline McNamee

George Daston (left) directs research on mechanistic and reproductive toxicology at Procter & Gamble. He has conducted research on in vitro *toxicological methods for over 20 years and currently leads Procter & Gamble's team on alternative toxicological approaches. He is editor-in-chief of* Birth Defects Research B: Developmental and Reproductive Toxicology.

Pauline McNamee (right) is a biochemist by training who joined the Procter & Gamble Company in 1987 as a product safety toxicologist. She is currently a principal scientist in the company's Central Product Safety Division where one of her focus areas is the development of alternative methods for the evaluation of eye irritation. In this context, she chairs the European cosmetics industry trade association's Eye Irritation Task Force that co-ordinates their research program in this end point.

We have learned over time that the development of successful alternative methods in toxicology testing requires the successful integration of three elements: First, there must be a solid foundation of understanding the basic biology and toxicology of the tissues and organs being studied. Second, *in vitro* platforms must be available that can be modified to make them amenable for toxicity testing. Third, one needs to convince the scientific community, which is skeptical by nature and training (and rightfully so), that the alternative methods fulfill their intended purpose and have been rigorously validated. *In vitro* mutagenicity screening methods have been used for many years and are a good illustration of these three points. Initially, the basic biology that needed to be understood was that DNA is the molecular basis for heredity and that mutations are, in fact, manifestations of damage to the DNA. Furthermore, several types of mutations (e.g., point mutations, insertions, deletions) require the development of different *in vitro* models. *In vitro* platforms, the second element, were adapted from extensive research into the molecular biology of prokaryotes, and later, eukaryotic cells. The third element involved years of assay standardization, replication of results in multiple laboratories, and comparisons with *in vivo* results.

For more complex end points, the development of alternatives has been a daunting task. Even supposedly simple targets for replacement, such as the Draize test for eye irritation, have proved difficult to model *in vitro* and progress through successful external validation despite major efforts by the European Centre for the Validation of Alternatives (ECVAM), industry trade associations, individual companies, and academia.

An extensive list of *in vitro* models that have been proposed as alternatives to the Draize test has been published

The development of successful alternatives requires the successful integration of [fundamental] . . . understanding of . . . basic biology and toxicology . . . , in vitro *platforms . . . amenable for toxicity testing [and the ability] . . . to convince the scientific community . . . that the alternative method fulfills its intended purpose and has been rigorously validated.*

(Bruner et al. 1991). Such alternative assays can be categorized as target organ/tissue assays [e.g., the bovine corneal opacity and permeability (BCOP) test, isolated rabbit eye (IRE) test, chicken enucleated eye test (CEET)]; organotypic models [e.g., the hen's egg test–chorioallantoic membrane (HET-CAM) assay, chorioallantoic membrane vascular assay (CAMVA), tissue equivalent assay]; cytotoxicity assays (e.g., neutral red assays, red blood cell lysis assay, fluorescein leakage assay); and chemical reaction assays (e.g., the Irritection Assay System). Although some of the many alternative assays developed have received limited attention, substantial effort has been invested in evaluating a significant number of the assays. Six major validation or evaluation studies were conducted between 1991 and 1997 in different locations: in Europe, the European Commission/British Home Office study (Balls et al. 1995), a European Cosmetic, Toiletry, and Perfumery Association (COLIPA) study (Brantom et al. 1997), and a Bundesgesundheitsamt/German Department of Research and Technology (BGA/BMBF) study (Spielmann et al. 1993, 1996); in the United States, the Cosmetics, Toiletries and Fragrance Association (CTFA) study (Gettings et al. 1991, 1994, 1996) and Interagency Regulatory Alternatives Group (IRAG) study (Bradlaw et al. 1997); in Japan, the Japanese Ministry of Health and Welfare/Japanese Cosmetic Industry Association (MHW/JCIA) study (Ohno et al. 1994). Unfortunately, none of the methods included in these validation/evaluation studies met all the formal validation requirements of the regulatory authorities for replacing the current animal test accepted by the Organisation for Economic Co-operation and Development (OECD) for acute eye irritation/corrosion (OECD 2002). It is reasonable to conclude, on the basis of several reviews that have been conducted on this topic, including a COLIPA workshop on mechanisms of eye irritation held in 1997 (Bruner et al. 1998) and an ECVAM workshop titled "Eye Irritation Testing: The Way Forward" held in 1998 (Balls et al. 1999), that the reasons for this lack of success are multiple and include a lack of understanding of the underlying physiological mechanisms of eye irritation, the variability of the *in vivo* Draize test data, and the ability of the Draize test to reliably predict the human response. In essence, none of the three elements of successful alternatives development was met during this early phase of the development of *in vitro* assays for eye irritation, mainly because their importance was not known at the time.

Although not formally validated by external scientific organizations (e.g., ECVAM) for the overall evaluation of eye irritation, the usefulness of some of these *in vitro* methods is well established for specific and limited purposes within some regulatory agencies and within industry. For example, the isolated eye tests IRE and CEET as well as the BCOP and HET-CAM tests are accepted by some European regulatory authorities on a case-by-case basis for the identification of severe eye irritants for the purposes of classification and labeling within the European Union on chemicals and products.

The development of alternative methods addresses the eventual replacement of animals in the evaluation of eye irritation. Reduction and refinement approaches such as the OECD tiered testing strategy now included as part of the OECD guideline for acute eye irritation/corrosion (OECD 2002) for hazard identification and regulatory classification of new chemicals are being used but do not eliminate the need for an *in vivo* test when the result of the *in vitro* test is negative.

Reduction and refinement methods/approaches for the evaluation of eye irritation are available today, but a validated replacement method(s) has not yet been achieved. There remains a clearly identified need to define alternative methods that reliably predict the human eye response to chemical exposure and that replace the *in vivo* test. Therefore, a fundamental understanding of what is needed to fill the knowledge gaps is essential to continued progress.

As a result of the reviews mentioned above, which have been conducted to define the future direction of the development and validation of eye irritation alternative methods, the key focus that emerged for future research is the need for mechanistic understanding of eye injury resulting from chemical exposure. Therefore, for *in vitro* replacement eye irritation tests to be reliable and predictive of the human response, they must be based upon mechanistically relevant biological events. Mechanistically based *in vitro* tests for ocular irritation likely will depend on *a*) well-characterized ocular cellular models, *b*) assays that measure biochemical end points of cellular injury, and *c*) a database of human responses. All of these cover a wide range of chemical classes and varying degrees of eye irritation.

Over the years we have gained a better understanding of the pathological events at the tissue and cellular levels that lead to corneal damage of varying degrees and the ability of the eye to recover from the initial injury (Maurer et al. 2002). This has led to the design and conduct of research programs that address development of alternative methods based on mechanistically relevant biological events. An example of one such program is being conducted by COLIPA, whose Steering Committee for Alternatives to Animal Testing has developed a collaborative research program with academia. The COLIPA research program is directed toward understanding the mechanism of eye injury and identification of new *in vitro* end points predictive of the *in vivo* response to chemical injury. There are three integrated parts of the research program: *a*) investigation of whether the kinetics and patterns of change in physiological function and signals of injury released from the cornea *in vitro* can predict a chemical's potential to damage the eye, with a focus on recovery, *b*) development of human corneal cell cultures and three-dimensional constructs for the study of chemically induced injury and recovery, and *c*) a genomics project. The outcome of the research program is that investigators better understand the cellular and molecular mechanisms of chemically induced eye irritation.

Although these developments satisfy the first and second elements—understanding of the basic biology and toxicology, and platforms amenable for toxicity testing—the third element, validation and regulatory acceptance, has been more difficult. There are several reasons for this, including the fact that the process for validating alternative assays was still being developed. But perhaps the most important reason was that the *in vivo* data set—results from the Draize test—against which the *in vitro* data were being compared was of variable quality. Weil and Scala (1971) determined that the numerical scores for the Draize test could not be reproduced in different laboratories. The Draize results continued to be used for regulatory decisions but with the understanding that the scores for individual components of the test were of dubious utility (Marzulli and Ruggles 1973). Unfortunately, these same data are the single largest source of *in vivo* information against which to compare the performance of alternative models even today. The low-volume eye test developed by Procter & Gamble in the late 1970s is more reproducible and more relevant to human responses (Freeberg et al. 1986), but fewer chemicals have been evaluated in this assay.

Although the development of *in vitro* methods is occurring more systematically than ever before, it continues to be a slow and uncertain procedure to model the complex biological processes that underlie toxicological assessments for end points such as subchronic and chronic toxicity, reproductive toxicity, and carcinogenicity. In the remainder of this article we evaluate the trends in each of the three elements needed for successful alternatives and make predictions as to what lies in store for *in vitro* methods development.

State of the Science: The First Element

Traditional toxicity tests are apical in nature: they evaluate the end result of exposure to a toxicant but provide little or no information about how that result occurred. For example, chronic bioassays provide information about the potential of the test agent to produce tumors, and in which tissues, but do not shed light on the mechanism by which the tumors arise. Apical tests have been used to predict human toxicity potential because it is inferred that all possible mechanisms of toxicity are represented in the animal model, including

those that are unknown, and that an effect on any of these mechanisms leads to a manifestation of toxicity. The tests are also useful in that the end points evaluated correspond to the biological processes in humans that one wishes to protect (e.g., fertility, normal organ function, etc.).

In vitro models, on the other hand, are the brainchildren of reductionist thinking. They are simple systems intended to facilitate the testing of hypotheses without the complexities and interrelationships that are inherent to intact organisms and that can hinder interpretation. The stereotypical *in vitro* model focuses on the mechanistic level of understanding.

The apical nature of the *in vivo* safety assays makes them ill-suited for identifying relevant mechanisms of action to be modeled. Therefore, the mechanistic basis for the *in vitro* assays has been developed through basic research, either to characterize the mechanism of action of a specific toxicant, or to understand the basic biology of a system. Considerable progress has occurred on both fronts, and our understanding of biological responses at a fundamental level is likely to increase exponentially with the advent of the tools of functional genomics.

The advent of genomics tools such as microarrays and related technologies makes it possible in a single experiment to evaluate all the changes in gene expression that occur in a cell, tissue, or organ as a result of an environmental perturbation. It appears that changes in gene expression occur after virtually any toxic insult (Nuwaysir et al. 1999) and it is possible that these changes are integral to the toxic response. If so, and if these changes are sufficiently specific, then it may be possible to use changes in gene expression as the basis for alternative screens.

We already know that gene expression is integral to the biological and toxicological responses to one group of chemicals—the steroid hormones and agents that activate or inhibit steroid hormone receptors. It has been established that the signal transduction pathway for steroid hormones involves interaction of the hormone–receptor complex with sites on DNA to promote or suppress the expression of specific genes. It is these changes in gene expression and the subsequent changes in the protein complement of the affected cells that constitute the cellular response to steroid hormone receptor agonists or antagonists.

The biological response of estrogen-sensitive tissues has been examined using microarrays. The time course for gene expression in the mouse uterus (Fertuck et al. 2003) and dose response (Naciff et al. 2003) for gene expression in the rat uterus and ovaries after treatment with an exogenous estrogen have been determined. These studies reveal that the uterotrophic response involves the coordinated action of genes that control cell proliferation, differentiation, tissue remodeling, angiogenesis, and apoptosis, among others. Although much of this could have been inferred by observations at the histological level, the identification of specific genes involved in the process could not.

The advances in our understanding of this and other biological responses at such a fundamental level of biological organization are enormous and surpass by orders of magnitude the pace at which information on gene expression was being added to the literature using the gene-by-gene technology that was state of the art only 5 years ago. Functional genomics is allowing scientists to formulate hypotheses not only about the role of single genes in biological responses (which, except in rare instances, are unlikely to be acting alone) but also about the role of entire suites of related genes whose functions are coordinated. At this point, hypothesis generation may be the most productive use of microarray technology.

The potential explosion of information about gene expression will be beneficial to the development of *in vitro* alternatives in two ways: First, it will support the selection of model systems that are mechanistically relevant. Second, it will provide end points for assessment (i.e., the expression of specific sets of genes) that are the same as, and can be measured in, the *in vivo* system being modeled. This would allow for the optimization of existing *in vitro* methods and/or the development of new methods. The relevance of an *in vitro* assay is often questioned because the nature and range of response of the system is unlikely to resemble fully that of the *in vivo* system. For example, a culture of uterine epithelial cells might be expected to proliferate and/or show changes in morphology in response to an estrogen but would not respond in ways that are so obvious in the intact uterus, such as thinning of the uterine wall or imbibition of fluid. However, if changes in gene expression (or at least of the subset expressed by the epithelium) are comparable with those *in vivo*, given a comparable stimulus, then the likelihood increases that the response is relevant to circumstances *in vivo*.

Despite the possible benefits from the information explosion, we should not fool ourselves into believing that the acceleration in hypothesis generation from genomics experiments will lead to accelerated hypothesis testing and *in vitro* methods development. The hypotheses are likely to be more complicated and difficult to test, commensurate with the increased complexity of the information feeding the hypotheses. However, advances in statistical analysis and bioinformatics now provide us with new methods of compression, analysis, and interpretion of complex data, so we have good reason to be optimistic that we are on a path that will provide the deep biological understanding needed for the development of useful *in vitro* methods.

Another scientific advance with considerable relevance for alternatives is the elucidation of fundamental biological processes, especially in the context of embryonic development in nonmammalian species, particularly *Caenorhabditis elegans* (a free-living nematode), *Drosophila melanogaster* (fruit fly), *Danio rerio* (zebrafish), and *Xenopus laevis* (African clawed frog).

Drosophila has been an especially useful model for genetic experiments for almost a century because of its small size and short life cycle as well as the ease with which it can be maintained and handled in the lab. It also has become an important model for developmental biology. Saturation mutagenesis research, which began in the 1970s, to investigate mutations in developmental control genes, resulted in the identification of virtually all the susceptible genes that are important developmentally (Nusslein-Volhard and Weischaus 1980). Detailed analysis of these genes has shown that most are involved in signal transduction and/or the regulation of gene expression. Furthermore, the sequence and function of these genes have been highly conserved across phylogenetic groups. Not only does this underscore the importance of these genes for regulation of cell function, but it also provides a basis for the hypothesis that lower organisms can be used for toxicity screening purposes, particularly if these screens evaluate the function (and perturbation of function) of the conserved genes.

Perhaps the most widely known example of the conservation of these genes is that of the *Hox* gene complex. These genes were first identified in *Drosophila* as the molecular basis for homeotic transformations, mutations in which a body part acquires the characteristics of a different body part. Antennapedia is one such mutation and is characterized by the development of legs where the antennae should be. Ultimately, a set of eight of these *Hox* genes was identified in *Drosophila*, and a homologous but expanded set of 13 *Hox* genes also was identified in mammals. These gene clusters were duplicated twice during early chordate evolution such that there are four paralogous groups. Not only is the sequence of the genes highly conserved but also the sites of expression along the anterior–posterior axis of the embryo between *Drosophila* and mammals.

Several other genes and gene clusters are highly conserved in sequence and function and are responsible for signal transduction. Of particular significance for alternatives test development is the existence of a finite number of signal transduction pathways: less than 20 have been identified (Gerhart 1999). Below are the intercellular signaling pathways listed according to developmental/physiological function [National Research Council (NRC) 2000].

- Early development and tissue growth/renewal: wingless-Int, tumor growth factor-β, hedgehog, receptor tyrosine kinases, notch-delta, cytokine receptor (STAT)
- Differentiation: interleukin-1/toll NF-κB, nuclear hormone receptor, apoptosis, receptor phosphotyrosine phosphatase
- After differentiation: receptor guanylate cyclase, nitric oxide receptor, G-protein–coupled receptor, integrin, cadherin, gap junction, ligand-gated cation channel.

Although it is possible that a few more may be found, it is likely that most of the pathways are already known. These pathways tend to be used repeatedly, not only in embryonic development but also in differentiated cells as a part of physiological function and tissue remodeling and renewal. To develop alternative methods, it may be possible to exploit the small number of pathways, as they may be a common step in the cascade of events that constitute the mechanisms of action for a disparate and large number of toxicants. The National Research Council Committee on Developmental Toxicology (NRC 2000) has suggested that model organisms such as those listed above for which the outcome of a perturbation in a specific signaling pathway is easily measured could be used as preliminary screens for toxicity. Much work is needed to determine whether this concept is pragmatically feasible, but the idea has a solid biological foundation.

Practical *in Vitro* Platforms: The Second Element

The second element necessary for successful alternatives consists of platforms or models that use the burgeoning information base in basic biology. These platforms must be selected or constructed so that critical aspects of a mechanism of toxicity are expressed and the outcome of perturbing those critical factors manifests as something that can be easily and reproducibly measured.

Many successful assay systems in the existing *in vitro* toxicology armory are intact structures or organs, or primary cultures. Examples of the former are the organotypic *in vitro* preparations of bovine, rabbit, or chicken eyes (obtained as a by-product of the slaughter of these animals for food) used as eye irritation screens, or rodent whole-embryo culture used to screen for teratogens. Examples of the latter are micromass cultures of embryonic rodent limb or brain to screen for teratogens or Syrian hamster embryo cells used to screen for carcinogens. Most of these models were selected because of *a*) the reasonable expectation that they would respond to toxicants in a manner similar to the *in vivo* structure from which they were derived and *b*) the inference that they contain the critical factors that mediate toxicity by most or all mechanisms that affect that structure.

The performance of these models supports the contention that they can serve as alternatives to *in vivo* screening. Although none has been validated to the point that it can completely replace *in vivo* testing, the results published to date are encouraging for their use in specific applications/situations, for example, use of BCOP, IRE, and CEET to identify severe eye irritants. One real benefit of these systems, particularly of the organotypic *in vitro* preparations, is that the manifestation of toxicity can be extrapolated immediately to the manifestation *in vivo*; for example, corneal damage in the enucleated eye corresponds directly to potential corneal damage *in vivo* (although it must be recognized that these assays do not address the key parameter of recovery), or a neural tube defect in whole-embryo culture is expected to predict the potential for a limb defect *in vivo*. Such coordinate responses eliminate the uncertainty from the interpretation of the *in vitro* results.

The disadvantage of these models is obvious: They require the continued use of animals as the source of organs, tissues, or cells. Although the models are a step in the right direction of refinement and reduction, they do not meet the ultimate goal of replacement.

Established cell cultures have occasionally made good models for *in vivo* alternatives, but these tend to be for acute end points such as cutaneous or ocular toxicity in which the mechanisms for the toxicity are limited and for which the end point measured is a sensitive evaluation of cellular function. Cell culture systems are becoming increasingly refined; three-dimensional cultures grown on a structural protein matrix tend to preserve the differentiated characteristics of epithelial cells. Many of these cultures have a medium-air interface that improves the quality of the culture and also facilitates treatment with test materials not compatible with the culture media. An example is a three-dimensional culture used to evaluate eye irritation (Osborne et al. 1995). It is also possible to immortalize cells while maintaining their differentiated characteristics, which has led to the development of human corneal equivalents (Griffith et al. 1999) that may be useful for eye irritation screening and form a basis for ongoing and future research programs for *in vitro* methods development.

In addition to providing the tools for immortalizing cells, molecular biology provides other techniques that have been applied to the screening of large numbers of chemicals for biological activity. In the pharmaceutical industry, it is now common practice to screen large libraries of compounds for their abilities to interact with a specific protein target (receptor, enzyme, etc.). This is accomplished either by making large quantities of recombinant receptor and conducting binding assays or by transfecting the receptor along with a reporter gene, which indicates that the receptor has been activated (or inhibited) into a cellular system. These high-throughput screening systems may be applied to toxicity screening but have the disadvantage of possibly screening for only one mechanism at a time. Therefore, until we have a more comprehensive understanding of toxic mechanisms, the concern remains that we have not adequately screened for toxicity. Still, for some applications such as screening compounds for their ability to act as an estrogen or androgen, these high-throughput methods may be useful.

It is also now possible to use gene expression as an end point for toxicity. As noted in the preceding section, gene expression patterns are likely to be mechanism specific; therefore, it is possible theoretically to

conduct screening systems by identifying transcript profiles that are diagnostic of specific toxicities. The literature increasingly describes transcript profiles that are specific for various mechanisms of action. The next step will be to determine whether comparable profiles can be elicited from *in vitro* models. This approach continues to have the limitation that not all mechanisms may be represented in the model, but unlike the high-throughput reporter gene assays described previously, expression of the cell's genome is almost certain to provide more information than a reporter gene assay about more mechanisms.

In the preceding section we described advances in our understanding of signal transduction and the idea that nonmammalian systems could be used as models to evaluate the effect of test agents on key signaling pathways. Because of saturation mutagenesis experiments, *Drosophila* and zebrafish mutants now exist that could be adapted for this purpose.

Of course, many obstacles must be overcome before cell-based systems can be relied on to predict systemic or chronic toxicity. These obstacles include the lack of adequate modeling of the complicated pharmacokinetics that occurs in the intact animal and usually incomplete or qualitatively different metabolism of the test agent. One of the most intractable problems is that *in vivo*, the upper limit on dosing is established by the inability of the animal to tolerate a higher dose; *in vitro*, the only limit tends to be the solubility of the test material, often leading to positive results with no relevance for predicting *in vivo* response. Some attempts have been made to solve these problems (e.g., comparing the concentration that produces a specific response with that which causes cytotoxicity), but these approaches do not account adequately for the complexity of the *in vivo* situation.

Validation and Regulatory Acceptance: The Third Element

The third element in the development of alternative toxicity assays is their acceptance by skeptical scientific and regulatory communities. The skepticism of both is warranted. On the scientific side we know the difficulties in developing predictive models. On the regulatory side there is concern that the goal of regulation, that is, the protection of public health, will be compromised if the alternative assays are not as reliable as the existing *in vivo* approaches.

It became clear during the early days of alternative methods development that a process was needed to assure all stakeholders that proposed new methods were adequate to serve in the stead of traditional methods. In the United States, ICCVAM (the Interagency Coordinating Committee on Validation of Alternative Methods) has developed a rigorous, objective, and peer-reviewed process to determine whether proposed new assays are suitable alternatives to existing ones. The federal agencies that regulate chemical safety are members of ICCVAM, and the review process is administered through the National Toxicology Program's Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM).

Dr. Ken Olden, director of the National Institute of Environmental Health Sciences and the National Toxicology Program, during his tenure has provided critical support to NICEATM and its director, Dr. Bill Stokes. Under Dr. Stokes's leadership, with full support from Dr. Olden, NICEATM has developed a process for reviewing potential alternative methods that is objective and consistent and involves the expertise of the external scientific community in such a way that maximize the chances for scientific acceptance of the outcome of ICCVAM reviews. Dr. Olden is also to be commended for his support for the National Center for Toxicogenomics (NCT), also established during his tenure. The NCT provides critical scientific support to studies on the effects of exogenous agents on gene expression, research that likely will serve as the foundation for the next generation of alternative assays.

The review process is a good one in that it does what is intended. However, all parties involved agree that the process is too long. Much of the time is consumed with assay development, standardization, and intra- and interlaboratory validation studies that provide the basis for the review. This typically takes many years. An example, the local lymph node assay (LLNA), is an alternative test for skin sensitization that was conceived in 1984, with the first paper on the assay published in 1986 (Kimber et al. 1986). Improvements on the assay continued over the next few years, until the assay was ready for interlaboratory validation studies in the United States and Europe in 1989. These validation studies required many years to complete, with final ICCVAM acceptance 10 years later and OECD guidelines published soon after (for a review of the assay, see Gerberick et al. 2000).

It is likely that subsequent validation and acceptance of alternatives will have a shorter timeline because assays such as LLNA have paved the way, but probably not by much. Development of assays is a complicated business, and even the process of transferring a protocol so that the results in that laboratory are qualitatively and quantitatively equivalent across laboratories does not always work. Validation of the uterotrophic assay for detecting estrogens has taken several years (Owens and Koeter 2003) and is still not complete at this writing, although it has existed in some form since the 1920s.

Similarly, in Europe, ECVAM was created by the European Parliament in October 1991 to address a requirement in the Protection of Laboratory Animals Directive (86/609/EEC) on the protection of animals used for experimental and other scientific purposes. This directive requires that the commission and the member states actively support the development, validation, and acceptance of methods that could reduce, refine, or replace the use of laboratory animals. As such, ECVAM's mission is to promote the scientific and regulatory acceptance of nonanimal tests that are important to biomedical sciences. This is to be accomplished through research, test development, and the validation and establishment of a specialized database service through European coordination of the independent evaluation of the relevance and reliability of tests for specific purposes, so that chemicals and products of various kinds, including medicines, vaccines, medical devices, cosmetics, household products, and agricultural products, can be manufactured, transported, and used more economically and more safely. This Directive should progressively reduce the current reliance on animal test procedures. Examples of recent *in vitro* method validations by ECVAM in the area of topical toxicity are 3T3 neutral red uptake phototoxicity test, EpiSkin skin corrosivity test, rat transcutaneous electrical resistance skin corrosivity test, and EpiDerm skin corrosivity test.

Another possible impediment to alternatives development and validation is that the traditional tests that are used as the gold standard against which to compare results are not always useful for that purpose. The problems with using the Draize eye irritation assay as the gold standard for *in vitro* eye irritation tests are discussed in the introductory remarks of this essay. The low-volume eye test mentioned above correlates reasonably well with the Draize results and is more reproducible and more relevant to human responses (Cormier et al. 1996), but the database for this test is smaller than that for the Draize test, and it is not universally accepted or widely approved by regulatory agencies.

The issue of benchmarks against which to compare will be even more complicated for more complex protocols. For example, *in vivo* developmental toxicity tests cover a large span of development and several manifestations of toxicity: structural malformation, growth retardation, *in utero* death at a minimum, and functional deficits for protocols with a postnatal leg. Any given *in vitro* alternative covers only a fraction of the developmental period. All tests developed to date cover only the embryonic period and probably predict only the potential to cause structural malformation. Therefore, a consensus must be developed as to which chemicals constitute positives (or negatives) for comparison of assay concordance with *in vivo* results. This is not as simple a task as it seems. Previous attempts to create such a list (Genschow et al. 2002; Smith et al. 1983) have met with criticism.

Perhaps the greatest challenge in determining the feasibility of using *in vitro* methods is the difficulty of comparing the results of reductionist, mechanism-based assays with those from apical *in vivo* tests. The mechanism-based assays are likely to be very reliable but because of their restricted nature will be able to predict only a fraction of the toxicity observed in apical tests. Establishing a battery of such tests may be highly predictive of toxicity potential, but the task of validating each particular test is likely to be daunting. The question will continually be asked, Is the failure of the *in vitro* test to detect an *in vivo* toxicant attributable to the fact that the toxicity is caused by another mechanism, or because the assay is inadequate? That question will be answered only through mechanistic research using the *in vivo* models. It is possible that, in the short run at least, the validation of alternatives could require more animals than are currently being used.

Although scientists have been developing alternative methods for more than two decades, recent legislation in Europe has added to the urgency of those efforts. The legislation calls for a ban of most animal testing for substances used in cosmetic products by 2009 and a ban on all animal testing by 2013. The European definition of cosmetic products is broad and

includes items such as dentifrice that are regulated in the United States as over-the-counter drugs.

The deadlines imposed by the European Parliament pose the greatest challenge, by far, not only to the research enterprise that has been dedicated to the "3Rs" (reduction, refinement, and replacement) of alternative methods/approaches but also to predictive toxicology in general. Whether the deadlines are achievable is a matter for debate: the European Union's own Scientific Committee for Cosmetics and Non-Food Products [now known as the European Union Scientific Committee on Consumer Products (2004)] has issued an opinion that it is not. Regardless of the prevailing scientific opinion, industry's only viable option is to continue its existing programs and collaborations in alternatives development at an even more accelerated pace.

Conclusions

Opportunities to develop alternative tests to predict toxicity have never been greater. The amount of information being generated on basic biology and how it can be perturbed by exogenous agents is increasing exponentially and is likely to continue as new tools such as genomics become more widely available and applied to toxicology. Similarly, the ability to develop better *in vitro* models is increasing. We have the chance not only to replace traditional tests but also to better predict and prevent adverse responses in humans. This would follow in the tradition of LLNA, which used a 3Rs method to help us better predict allergen potency by taking advantage of advances in biological understanding and statistical methodology.

It must be recognized, however, that investigators will need time to take advantage of these opportunities. The deadline for full replacement of animal testing for consumer products in Europe is so near that it will hinder the development of tests that use the new knowledge and new technology. Because several years are needed to validate and gain regulatory acceptance for alternative methods, the only methods with a chance of meeting the deadline are those that have already been developed and standardized to some extent, which may mean that they do not use the latest technology. It may be possible over the next several years to develop tests with more promise, but these tests will not be available by the deadline set by the European Union. Neither of these alternatives—defaulting to less than optimal tests, or short-circuiting the validation and peer-review process—fulfills the goal of protecting and improving public health.

The next 4 years will be interesting and difficult ones for those of us working on alternatives. We hope that all those interested in this area will work together to find solutions that are in the best interest of animal welfare and public health. These goals do not have to be, and indeed should not be, mutually exclusive.

Summary

The development and application of alternative methods in toxicology have been active areas of research for decades. The pace of alternatives development is determined by three elements. First, the basic biology of adverse responses to toxicants must be understood with sufficient mechanistic depth to support the selection of models and end points relevant to the process being studied. Second, *in vitro* methodology must be developed that is amenable for, or can be adapted to, toxicological applications. Third, the scientific basis and performance of assays in validation programs must be sufficiently robust to convince the scientific and regulatory communities that proposed alternative assays can replace traditional methods. Each of these three elements is rate limiting to the replacement of animal testing; however, new scientific advances coupled with streamlined review processes for alternative methods should accelerate the pace of new methods development. New, genomics-aided research on the molecular basis of toxic response will enhance our ability to select appropriate test systems and will expand (and possibly make more relevant) the end points that we measure in those systems. Adaptation of molecular biological approaches to create *in vitro* systems that are more relevant to humans—by incorporating human metabolizing systems, human receptors, and so forth—will improve the performance of the assays measuring those end points. Finally, objective and comprehensive review processes, such as the one administered at the National Institute of Environmental Health Sciences via ICCVAM/NICEATM (Interagency Coordinating Committee on Validation of Alternative Methods/National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods), provide alternative methods researchers with a venue for gaining scientific and regulatory acceptance of their methods. The pace of methods development will need to accelerate markedly during the current decade to meet the deadline imposed by the European Parliament that calls for a ban of most animal testing by 2009, and all animal testing by 2013, for any substance to be used in a cosmetic product. Although it is unlikely that science will be

able to meet the legislatively imposed deadlines for animal replacements, progress will be made toward that goal during the coming years.

doi:10.1289/ehp.7723 available via *http://dx.doi.org/*

Notes

Address correspondence to G. Daston, Miami Valley Laboratories, Procter & Gamble, PO Box 538707, Cincinnati, OH 45253 USA. Telephone: (513) 627-2886. Fax: (513) 627-0323. E-mail: daston.gp@pg.com

The authors have competing financial interests in that they are employed by Procter & Gamble, a company impacted by the necessity to conduct toxicological testing.

References

Balls M, Berg N, Bruner LH, Curren RD, de Silva O, Earl LK, et al. 1999. Eye irritation testing: the way forward. The report and recommendations of ECVAM workshop 34. Altern Lab Anim 27:53–77.

Balls M, Botham PA, Bruner LH, Spielmann H. 1995. The EC/HO international validation study on alternatives to the Draize eye irritation test. Toxicol in Vitro 9:871–929.

Bradlaw J, Gupta K, Green S, Hill R, Wilcox N. 1997. Practical application of non-whole animal alternatives: summary IRAG workshop on eye irritation. Food Chem Toxicol 35:175–178.

Brantom PG, Bruner LH, Chamberlain M, De Silva O, Dupuis J, Earl LK, et al. 1997. A summary report of the COLIPA international validation study on alternatives to the Draize rabbit eye irritation test. Toxicol in Vitro 11:141–179.

Bruner HL, de Silva O, Earl LK, Easty DL, Pape W, Spielmann H. 1998. Report on the COLIPA workshop on mechanisms of eye irritation. Altern Lab Anim 26:811–820.

Bruner LH, Shadduck J, Essex-Sorlie D. 1991. Alternative methods for assessing the effects of chemicals in the eye. In: Dermal and Ocular Toxicology: Fundamentals and Methods (Hobson DW, ed). Boca Raton, FL:CRC Press, 585–606.

Cormier EM, Parker RD, Henson C, Cruse LW, Merritt AK, Bruce RD, et al. 1996. Determination of the intra- and interlaboratory reproductibility low volume eye test and it's statistical relationship to the Draize eye test. Regul Toxicol Pharmacol 23:156–161.

European Union Scientific Committee on Consumer Products. 2004. Opinion concerning Report for Establishing the Timetable for Phasing Out Animal Testing for the Purpose of the Cosmetics Directive issued by ECVAM 30/04/2004. http://europa.eu.int/comm/health/ph_risk/committees/sccp/documents/out285_en.pdf [accessed 1 July 2004].

Fertuck KC, Eckel JE, Gennings C, Zacharewski TR. 2003. Identification of temporal patterns of gene expression in the uteri of immature, overiectomized mice following exposure to ethynylestradiol. Physiol Genom 15:127–141.

Freeberg FE, Nixon GA, Reer PJ, Weaver JE, Bruce RD, Griffith JF, et al. 1986. Human and rabbit eye responses to chemical insult. Fundam Appl Toxicol 7:626–634.

Genschow E, Spielmann H, Scholz G, Seiler A, Brown N, Piersma A, et al. 2002. The ECVAM international validation study of vitro embryotoxicity tests: results of the definitive phase and evaluation of prediction models. Altern Lab Anim 30:151–176.

Gerberick GF, Ryan CA, Kimber I, Dearman RJ, Lea LJ, Basketter DA. 2000. Local lymph node assay: validation assessment for regulatory purposes. Am J Contact Dermat 11(1)3–18.

Gerhart J. 1999. 1998 Warkany lecture: signaling pathways in development. Teratology 60:226–239.

Gettings SD, DiPasquale LC, Bagley DM, Casterton PL, Chudkowski M, Curren RD, et al. 1994. The CTFA evaluation of alternatives program: an evaluation of *in vitro* alternatives to the Draize primary eye irritation test (phase II) oil/water emulsions. Food Chem Toxicol 32:943–976.

Gettings SD, Lordo RA, Hintze KL, Bagley DM, Casterton PL, Chudkowski M, et al. 1996. The CTFA evaluation of alternatives program: an evaluation of *in vitro* alternatives to the Draize primary eye irritation test (phase III) surfactant-based formulations. Food Chem Toxicol 34:79–117.

Gettings SD, Teal JJ, Bagley DM, Demetrulias JL, DiPasquale LC, Hintze KL, et al. 1991. The CTFA evaluation of alternatives program: an evaluation of *in vitro* alternatives to the Draize primary eye irritation test (phase I) hydro-alcoholic formulations: Part 2. Data and biological significance. In Vitro Toxicol 4:247–288.

Griffith M, Osborne R, Munger R, Xiong X, Doillon CJ, Laycock NL, et al. 1999. Functional human corneal equivalents constructed from cell lines. Science 286:2169–2172.

Kimber I, Botham PA, Rattray NJ, Walsh ST. 1986. Contact-sensitizing and tolerogenic properties of 2,4-dinitrothiocyanobenzene. Int Arch Allergy Appl Immunol 81(3):258–264.

Marzulli FN, Ruggles DI. 1973. Rabbit eye irritation test: collaborative study. J Assoc Off Anal Chem 56:905–914.

Maurer JK, Parker RD, Jester JV. 2002. Extent of initial corneal injury as the mechanistic basis for ocular irritation: key findings and recommendations for the development of alternative assays. Regul Toxicol Pharmacol 36:106–117.

Naciff JM, Overmann GJ, Torontali SM, Carr GJ, Tiesman JP, Richardson BD, et al. 2003. Gene expression profile induced by 17-β-ethynyl estradiol in the pre-pubertal female reproductive system of the rat. Toxicol Sci 72:314–330.

NRC (National Research Council Committee on Developmental Toxicology). 2000. Scientific Frontiers in Development Toxicology and Risk Assessment. Washington, DC:National Academy Press.

Nusslein-Volhard C, Weischaus E. 1980. Mutations affecting segment number and polarity in *Drosophila*. Nature 287:795–801.

Nuwaysir EF, Bittner M, Trent J, Barrett JC, Afshari CA. 1999. Microarrays and toxicology: the advent of toxicogenomics. Mol Carcinog 24:153–159.

OECD 2002. OECD Guideline for the Testing of Chemicals, No. 405: Acute Eye Irritation/Corrosion. Paris:Organisation for Economic Co-operation and Development.

Ohno Y, Kaneko T, Kobayashi T, Inoue T, Kuroiwa Y, Yoshida T, et al. 1994. First phase validation of the *in vitro* eye irritation tests for cosmetic ingredients. In Vitro Toxicol 7:89–94.

Osborne R, Perkins MA, Roberts DA. 1995. Development and intralaboratory evaluation of an in vitro human cell-based test to aid ocular irritancy assessments. Fundam Appl Toxicol 28:139–153.

Owens W, Koeter HB. 2003. The OECD program to validate the rat uterotrophic bioassay: an overview. Environ Health Perspect 111:1527–1529.

Smith MK, Kimmel GL, Kochhar DM, Shepard TH, Spielberg SP, Wilson JG. 1983. A Selection of candidate compounds for *in vitro* teratogenesis test validation. Teratog Carcinog Mutagen 3:461–480.

Spielmann H, Kalweit S, Liebsch M, Wirnsberger T, Gerner I, Bertram-Neis E, et al. 1993. Validation study of alternatives to the Draize eye irritation test in Germany: cytotoxicity testing and HET-CAM test with 136 industrial chemicals. Toxicol in Vitro 7:505–510.

Spielmann H, Liebsch M, Kalweit S, Moldenhauer F, Wirnsberger T, Holzhutter H-G, et al. 1996. Results of a validation study in Germany on two *in vitro* alternatives to the Draize eye irritation test, the HET-CAM test and the 3T3 NRU cytotoxicity test. Altern Lab Anim 24:741–858.

Weil CS, Scala RA. 1971. Study of intra- and interlaboratory variability in the results of rabbit eye and skin irritation tests. Toxicol Appl Pharmacol 19:276–360.