Introduction
Consumers and workers are exposed to a wide range of naturally occurring and manmade chemicals and products in foods, medicines, consumer products, pesticides, water, air, and other sources. Safety testing is required by regulatory authorities to determine if chemicals and products are safe or if they may produce adverse health effects to people, animals, and the environment. Such testing determines the nature and severity of health hazards that might be produced by accidental or intentional exposures, and it is used as the basis for hazard labeling or to establish safe levels of exposure.

Safety assessments traditionally have used animal models, but in recent years a number of new in vitro and in vivo models have been developed and accepted that have significantly reduced, refined (less or no pain and distress), and replaced animal use (Stokes and Wind, 2010b). Regulatory acceptance of these new alternative methods was supported by scientific validation studies that characterized the usefulness and limitations of the new proposed methods for identifying specific hazards. Validation studies provided the information needed by regulatory authorities to determine that using the proposed test method would provide equivalent or improved protection of people, animals, and/or the environment (Stokes and Schechtman, 2007; Birnbaum and Stokes, 2010).

Several national validation centers and committees are charged with evaluating the scientific validity of new test methods and/or conducting test method validation studies. Much of the recent progress in the acceptance of alternative test methods has resulted from the efforts of these organizations. In the United States, these include the National Toxicology Program’s (NTP) Interagency Center for the Evaluation of Alternative Methods (NICEATM), and its Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM, 2008, 2010; Birnbaum and Stokes, 2010; Stokes and Wind, 2010c). Other organizations include the European Centre for the Validation of Alternative Methods (ECVAM), the Japanese Center for the Validation of Alternative Methods (JaCVAM), and the Korean Center for the Validation of Alternative Methods (KoCVAM). At least 43 alternative methods have now been adopted by national and international regulatory authorities (NICEATM, 2011).

More recently, new research initiatives have focused on a systems biology approach to understanding and detecting the molecular, genetic, structural, and cellular perturbations that may lead to adverse health outcomes. Referred to as the 21st Century Toxicology Toolbox, these include a wide range of applications, or tools, that incorporate toxicogenomics, metabolomics, and computational models. These tools are now being used to create complex biological activity profiles for specific chemicals, with an expectation that these profiles eventually will predict toxicity and safety without the use of animals. This paper discusses the emergence and application of new systems
biology approaches for toxicity testing, collectively referred to as the 21st Century Toxicology Toolbox, and discusses considerations for their scientific validation for use in regulatory decision-making.

2 A new paradigm for toxicity testing

In the past few years there has been a growing emphasis on developing and using new testing approaches that measure early molecular and cellular perturbations in cells and simple systems as predictors of adverse health effects rather than observing the actual adverse outcome in animal models. This concept was the basis for the 2004 NTP Roadmap, which envisioned moving toxicology from such an observational science to a predictive science (NTP, 2004). Similarly, the 2007 National Research Council (NRC) publication, Toxicity Testing in the 21st Century, A Vision and a Strategy, called for transforming toxicology “from a system based on whole-animal testing to one founded primarily on in vitro methods that evaluate changes in biologic processes using cells, cell lines, or cellular components, preferably of human origin.” (NRC, 2007a)

A key aspect of this new testing approach is the concept of chemical activation of toxicity pathways. Under normal conditions, biologic inputs to cells are translated into normal biologic function. However, exposure to toxicants can perturb these pathways. Low-level non-toxic exposures result in early cellular changes that are met by an adaptive response and subsequently return to normal biologic function. However, at higher exposure levels, these early cell changes may be severe enough in a sufficient number of pathways to result in cell injury, which in turn could result in tissue or organ damage and potentially lead to morbidity or even mortality. The NRC report calls for a stronger mechanism-based approach for making chemical safety decisions using a comprehensive battery of toxicity pathway-based in vitro tests that can then be complemented, if necessary, with targeted animal tests.

The NRC report recognizes that successfully implementing the vision will depend on a number of critical factors. Infrastructure changes will be necessary to support basic and applied research to develop the tests and pathway models. Validation of tests and test strategies will ultimately be needed to provide the necessary evidence to justify that a toxicity-pathway approach is adequately predictive of adverse health outcomes for use in decision-making. And, finally, it recognizes that much work over a sustained period of time will be needed to make this transformation succeed. A subsequent NRC report, Science and Decisions: Advancing Risk Assessment, predicted that realizing the vision of 21st century toxicological test methods is at least a decade away (NRC, 2009).

3 Tools in the 21st Century Toxicology Toolbox

Examples of tools in the 21st Century Toxicology Toolbox include high throughput screening (HTS) using cell-based assays, toxicogenomics, metabolomics, proteomics, biochemical activity profiles, and computational models. This toolbox is continually expanding and improving as a result of new scientific advances and innovative technologies.

HTS minimizes manual collection and processing of data by using computerized robots to conduct laboratory procedures, allowing analyses of large numbers of compounds in multiple in vitro assays within a short period of time (Stokes and Wind, 2010a). The National Institutes of Health (NIH) has created a laboratory to carry out HTS at the NIH Chemical Genomics Center (NCGC) (Michael et al., 2008). A consortium of U.S. agencies subsequently was created as a mechanism to achieve the vision of the NTP Roadmap and the 2007 NRC report by incorporating the combined resources of the consortium partners (Collins et al., 2008). This consortium includes the NTP, the NCGC, and the U.S. Environmental Protection Agency. In 2010, the U.S. Food and Drug Administration (FDA) joined the partnership (Tox21, 2011). The FDA brings to the partnership their experience in human diseases and in animal models of human disease, as well as toxicity pathway analysis and computational toxicology. This so-called Tox21 consortium is focused on research, development, validation, and translation of innovative chemical testing methods that characterize toxicity pathways.

High throughput testing is performed at the NIH NCGC (Michael et al., 2008), which uses 1536-well plates to test fifteen concentrations of each chemical (Collins et al., 2008). More than 100,000 concentration response profiles can be generated per week, which can then be evaluated to determine if there is a correlation between in vitro responses and known adverse health effects that have been observed in standardized animal-based toxicology studies or in people during pre-market clinical trials or post-marketing surveillance. Bioinformatics techniques will be used to identify complex relationships between different types of biological responses that may provide insights into critical toxicity pathways (Schmidt, 2009).

Tools in the 21st Century Toxicology Toolbox can be used in a systems biology approach to evaluate the complex interactions within biological systems at the molecular level. Perturbation of a system resulting from chemical exposures can be studied by monitoring molecular expression, integrating response data, and modeling system structure and function (Ideker et al., 2001) (Tab. 1). To fully elucidate a person’s likely response to a specific chemical exposure, it is necessary to include functional characterization and accurate quantification of all levels of alterations in gene products, mRNA, proteins, and metabolites associated with responses to the exposure (Zhang et al., 2010). These are measured using techniques such as transcriptomics, proteomics, and metabolomics, which provide information used to understand cellular processes as one “integrated system” rather than as a collection of different parts. Using the ‘omic’ technologies, chemical-induced perturbations can be detected at multiple levels of biological organization: expression patterns of genes, proteins, and metabolites in cells, tissues, and organisms. Such technologies are being investigated for their potential to improve the prediction...
of safety or potential hazards of chemicals to human health. The NRC published a report in 2007 that addressed the application of toxicogenomic technologies to predictive toxicology and risk assessment (NRC, 2007b). In this report, the NRC recommends that regulatory agencies incorporate toxicogenomic data into risk assessments using tools like gene expression profiling (transcriptomics), metabolomics, and proteomics.

Computational models provide additional non-animal methods that can be used to estimate the absorption, distribution, metabolism, and excretion of chemicals (ADME) (NRC, 2007a). These models seek to estimate the relationship between the dose, or amount of chemical exposure via oral, dermal, or inhalation routes, and the concentration of chemical that reaches individual cell types in various critical organs and tissues (Stokes and Wind, 2010a). These estimates will be essential for non-animal estimates of exposure levels that are safe and those that are likely to be associated with toxic effects. It is also important that data used to construct computational models is of high quality and derived from adequately designed studies.

NIH is supporting an increasing number of National Centers for Systems Biology. Among these is a 2011 award of a $13 million five-year grant for development of a computational model referred to as the “Virtual Physiological Rat” (http://the-scientist.com/2011/08/12/the-virtual-physiological-rat/). The project will integrate widespread data into a single model so that rat physiology as a whole can be better understood. Researchers will be able to predict how multiple systems interact in response to environmental and genetic causes of disease. The model is expected to enhance our understanding of the interaction of genes and environmental factors in determining phenotype.

### 4 Applying new science and technology to regulatory decision making

Linking in vitro perturbations of normal biological pathways to the prediction of adverse health effects in animals and people requires understanding the mechanisms and modes of action involved in chemically-induced phenotypic adverse effects in animals and humans (Stokes and Wind, 2010a). This issue was addressed at a recent NICEATM-ICCVAM International Workshop on Acute Chemical Safety Testing – Advancing In Vitro Approaches and Humane Endpoints for Systemic Toxicity Evaluations (NICEATM, 2008). A key workshop recommendation was the need to collect additional mechanistic and pathway perturbation data from current in vivo testing models that could be used to support the development of predictive mechanism-based in vitro alternative models. Workshop participants recommended collecting mechanistic information during in vivo studies using available technologies such as noninvasive telemetry systems for real-time monitoring of physiological parameters, automated systems for collecting behavioral information, and noninvasive analytical devices to analyze small volumes of blood and urine for early biomarkers of molecular and cellular damage. These mechanistic data could then be used to inform the development of in vitro test methods to detect chemically induced perturbations of specific toxicity pathways using tissue-specific cellular models. The workshop also recommended that such data could help identify predictive biomarkers that could serve as earlier, more humane endpoints where animal studies are still necessary. In addition, early in vivo biomarkers predictive of the eventual in vivo adverse effect could, in turn, serve as the basis for validation of in vitro test methods that measure perturbations of pathways that are proposed to be predictive of the in vivo adverse effect.

### 5 Validation and acceptance of test methods based on new science and technology

In the United States, Federal law requires that agencies ensure that new test methods proposed for regulatory safety assessment decisions must be determined to be valid for their proposed use prior to requiring, recommending, or encouraging the application of such test methods (USC, 2000). Governments have developed national and international harmonized principles for validation and regulatory acceptance to assess the validity and regulatory acceptability of new test methods (ICCVAM, 1997; OECD, 2005). Determination of validity involves assessing the accuracy and reliability of a test method for a specific proposed purpose (ICCVAM, 1997; OECD, 2005; Stokes and Schechtman, 2007). Regulatory acceptance involves reviewing the validation database to determine if the proposed use of the method for decision making will provide equivalent or improved protection compared to existing methods (USC, 2000). Reliability assessments determine whether

<table>
<thead>
<tr>
<th>Pathway Constituent</th>
<th>Tool for Measuring Perturbations</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>Genomics</td>
</tr>
<tr>
<td>RNA</td>
<td>Transcriptomics</td>
</tr>
<tr>
<td>Protein</td>
<td>Proteomics and Interactomics</td>
</tr>
<tr>
<td>Biochemicals (Metabolites)</td>
<td>Metabolomics</td>
</tr>
</tbody>
</table>

Tab. 1: Systems biology pathways and molecular tools for their measurement
reproducible results can be obtained in different laboratories when using the proposed standardized test method protocol.

A common approach to assessing test method accuracy is to directly compare measurements or predictions from the new method or strategy to high quality results for the same chemicals from an accepted reference test method (Fig. 1). When predicting human toxicity, it is always important to consider all available human data and information, in addition to traditional reference data (Stokes and Wind, 2010a). This can be especially helpful when human data differ from reference testing data, in which case human data or experience is usually given priority. Accuracy assessments provide performance parameters that characterize the extent that the new test method can correctly measure or predict the biological effect of interest. The resulting calculations are provided as accuracy, sensitivity, specificity, false positive rate, and false negative rate. The extent of erroneous predictions or measurements that will be acceptable to regulatory authorities varies depending on the intended purpose of the test method and the implications of accepting a false result.

When results from the proposed method are to be used for regulatory hazard classification and labeling, ideally there would be no false negatives and no false positives. Regulatory and public health authorities strive to avoid test methods that produce false negatives because these could lead to injuries and disease from exposures to hazards that are not appropriately labeled. Such hazard labeling is necessary to warn consumers and workers of the precautions necessary to avoid exposures that could lead to injuries, disease, or even death. While false positives may result in a hazard label when there is no significant hazard, such precautions do not adversely impact human health. Nonetheless, high false positive rates also are undesirable because requirements associated with certain hazard classifications can result in additional costs for packaging, shipping, and handling. Accordingly, decision criteria for new methods and testing strategies are normally selected to avoid false negatives while tolerating some false positives, if necessary.

### 6 Validation and intended purposes of test methods

National and international authorities have agreed on validation and regulatory acceptance criteria for new, revised, and alternative test methods (ICCVAM, 1997; OECD, 2005). These are general criteria that should be appropriately addressed when considering the validity of test methods. The published criteria emphasize that flexibility is essential in applying and interpreting the criteria and that the extent that the criteria should be addressed will depend on the intended purpose and nature of the proposed test method (ICCVAM, 1997; OECD, 2005). Validation strategies need to be flexible in order to accommodate a wide variety of purposes for a proposed test method. For example, test methods used for product development decisions, those used to inform prioritization decisions, and those used for mode/mechanism of action investigations might require less extensive validation than those methods proposed for international adoption for regulatory hazard and risk assessment decisions.

Industry has used biological activity profile data for many years in making decisions on whether chemicals should be advanced for further product development. Those substances that lack evidence of potential efficacy or for which there is evidence of potential toxicity often are given a lower priority for further investigation. Similarly, biological profile data have been proposed for prioritizing chemicals for further testing in standardized regulatory-approved test methods (Judson et al., 2009). For example, chemicals that appear likely to cause the toxicity of interest based on their biological activity profile might receive a higher priority for definitive testing than those with a profile that indicates a low or negligible likelihood. In some situations, the lack of certainty associated with the prediction of safety or hazard might result in a higher priority for definitive testing.

Mechanistic data also can be considered to inform weight of evidence decisions on chemical safety, hazard, and risks and, potentially, to reduce uncertainties in risk assessment. Howev-
er, the use of biological profile data to replace current validated regulatory test methods for regulatory hazard assessments and decisions will require validation to demonstrate that the proposed data and decision strategies can provide equivalent or improved protection of consumers and workers compared to existing risk assessment procedures (USC, 2000).

Flexibility in the validation of these new tools and strategies will be essential (ICCVAM, 2003). Validation study designs will vary depending on the intended purpose of new methods and strategies, the proposed applicability domain, and existing data that can also be used to support the validity of the proposed methods or strategy. Established validation criteria should be considered and used to ensure appropriate validation study designs. Consideration and use of appropriate validation principals early in the test method development process will help ensure efficient use of experimental resources and expedite acceptance of new tools and approaches.

### 7 Levels of validation

When considering validation strategies for methods in the 21st Century Toxicology Toolbox, three levels of validation must be considered (NRC, 2007c):

First, **technical validation** focuses on whether a new technology platform provides reproducible and reliable results. For example, testing of the same chemicals across a range of responses is repeated to determine if the technology platform provides consistent and reproducible answers. Technical validation occurs early in the test method development process.

**Biologic validation** evaluates whether the underlying biology is reflected in the outcomes obtained from the new technology platform. This determines the extent that the measured qualitative and quantitative response in the test system is indicative of the true biologic response and whether there are other factors causing unrelated positive, negative, or quantitatively-altered responses.

**Regulatory validation** often is considered following technical and biologic validation, and when test methods using the new technology platform are proposed as regulatory decision tools. Regulatory validation determines the extent that the test system generates information useful for regulatory decisions on safety or hazard, and the extent that use of a proposed standardized test method protocol produces similar results in different qualified laboratories.

### 8 Best validation practices

Early consideration of the potential application of new technologies for regulatory testing during research and development stages provides an important opportunity to incorporate efforts that will support the eventual validation of the test methods (Stokes and Wind, 2010a). Early standardization and use of harmonized technology platforms for approaches such as toxicogenomics and HTS will allow for data from different studies to be compared and combined for data analyses. This also will help minimize experimental variables, aid in achieving more reproducible results across labs, and contribute to achieving a high signal to noise ratio. For example, recommendations have been developed for the standardization and validation of toxicogenomic platforms that will be used to support test methods proposed for safety assessment decisions (Corvi et al., 2006).

Early consideration of validation principles will save time and resources by ensuring adequate validation study designs and will minimize the need to repeat testing. For example, the generation of high quality data during research and development may contribute to the validation database supporting the validity of proposed test methods and approaches (Stokes and Wind, 2010a). Several critical factors should be considered during research, development, translation, and validation stages for new technologies:

- A database of reference data and other chemical/physical information about the chemicals to be tested by the new method or approach should be compiled and fully referenced in advance of the validation study.
- A database of concurrent positive and negative control responses should be established to provide information to determine whether each independent test is functioning appropriately and to identify if there are changes in response over time. An appropriate negative control substance ensures that the test can appropriately identify an unknown substance that has no activity. An appropriate positive control substance ensures that the test method can appropriately identify a substance with activity.
- National validation authorities should be consulted during the test method development process in order to ensure regulatory applicability and to obtain guidance and advice on study design, protocol development, and selection of reference substances. Early and frequent consultation will save time and resources.

### 9 Conclusions

Advances in science and innovative technologies are providing new opportunities to develop improved safety testing methods and strategies that are also expected to reduce animal use. Consideration of validation principles and potential application to regulatory decision-making during early stages of research, development, and validation will help expedite more efficient scientific validation of these new methods and strategies. Validation databases will need to adequately characterize the usefulness and limitations of new proposed test methods and strategies, and support determinations of whether the new method or approach can provide equivalent or improved protection compared to existing safety assessment procedures. New methods and integrated strategies should be developed and validated in consultation with relevant stakeholders and national validation centers in order to ensure adequate and appropriate studies. Comprehensive and optimal validation study designs are expected to expedite the validation and regulatory
acceptance of new test methods and strategies that support improved safety assessments and contribute to reducing, refining, and replacing animal use for regulatory safety testing.

References


Correspondence to

William S. Stokes, D.V.M.
Director, NICEATM
National Institute of Environmental Health Sciences
P.O. Box 12233, Mail Stop K2-16
Research Triangle Park, NC 27709
USA
e-mail: stokes@niehs.nih.gov