FutureTox: Building the Road for 21st Century Toxicology and Risk Assessment Practices

J. Craig Rowlands,*,[1] Miriam Sander,† James S. Bus*,[2]; FutureTox Organizing Committee[3]

*The Dow Chemical Company, Midland, Michigan 48674; and †Page One Editorial, Boulder, Colorado 80304

[1]To whom correspondence should be addressed at The Dow Chemical Company, 1803 Building, Midland, MI 48674. Fax: (989) 638-9305.
E-mail: jcrowlands@dow.com.
[2]Present address: Center for Toxicology and Mechanistic Biology, Exponent, Midland, MI 48640.
[3]J. Craig Rowlands (The Dow Chemical Company), James S. Bus (The Dow Chemical Company), Kim Boekelheide (Brown University), Russell Thomas (Hamner Institutes), Martin Stephens, (Human Toxicology Project Consortium), George P. Daston (Procter & Gamble), Suzanne Compton Fitzpatrick (U.S. Food and Drug Administration [U.S. FDA]), Raymond Tice (National Institute of Environmental Health Sciences [NIEHS]), Robert J. Kavlock (U.S. Environmental Protection Agency [EPA]), and Laurie C. Haws (ToxStrategies).

This article reports on the outcome of FutureTox, a Society of Toxicology (SOT) Contemporary Concepts in Toxicology (CCT) workshop, whose goal was to address the challenges and opportunities associated with implementing 21st century technologies for toxicity testing, hazard identification, and risk assessment. One goal of the workshop was to facilitate an interactive multisector and discipline dialog. To this end, workshop invitees and participants included stakeholders from governmental and regulatory agencies, research institutes, academia, and the chemical and pharmaceutical industry in Europe and the United States. The workshop agenda was constructed to collectively review and discuss the state-of-the-science in these fields, better define the problems and challenges, outline their collective goals for the future, and identify areas of common agreement key to advancing these technologies into practice.

Key Words: in vitro **and alternatives; risk assessment; regulatory/policy.**

BACKGROUND

The urgent need to develop new tools and increased capacity to test chemicals and pharmaceuticals for potential adverse impact on human health and the environment has been recognized for quite some time. In response, a committee formed by the U.S. National Academy of Sciences (NAS) formally challenged the entire scientific-academic-industrial-regulatory community to develop streamlined, more efficient methods to assess risk from human exposure to known and novel chemicals and pharmaceuticals. This challenge was outlined in great detail in 2 heavily cited reports: *Toxicity Testing in the Twenty-first Century: A Vision and a Strategy* (National Research Council, 2007) and *Science and Decisions:*

Advancing Risk Assessment (National Research Council, 2009), both of which advocated for "science-informed decision making" in the field of human risk assessment.

The NAS reports also recognized and recommended that effective implementation of new technologies in risk assessment requires interdisciplinary and intersector dialog as well as a consistent and coherent strategy for validating and then integrating new technologies into the existing risk assessment framework.

This article reports on the outcome of FutureTox, a Society of Toxicology (SOT) Contemporary Concepts in Toxicology (CCT) workshop, whose goal was to develop consensus among stakeholders on how to implement 21st century technologies for toxicity testing, hazard identification, and risk assessment. To facilitate an interactive multisector and multidiscipline dialog, the workshop invitees and participants included stakeholders from governmental and regulatory agencies, research institutes, academia, and the chemical and pharmaceutical industry in Europe and the United States. Workshop participants reviewed and discussed the state-of-the-science in these fields, attempted to define collective goals for the future, and identify areas of common agreement key to advancing use of novel technologies. The meeting was attended by 206 individuals distributed across government (93), industry (78), academia (11), and NGOs (24). The FutureTox organizing Committee included J.C.R. and J.S.B. (cochairs, The Dow Chemical Company), Kim Boekelheide (Brown University), Russell Thomas (Hamner Institutes), Vicki L. Dellarco (U.S. Environmental Protection Agency [U.S. EPA]), Martin Stephens (Human Toxicology Project Consortium), George P. Daston (Procter & Gamble), Suzanne Compton Fitzpatrick (U.S. Food and Drug Administration [U.S. FDA]), Raymond Tice (National Institute of Environmental Health Sciences [NIEHS]), Robert

J. Kavlock (U.S. EPA), and Laurie C. Haws (ToxStrategies). Workshop support was provided by Clarissa Wilson (Society of Toxicology). FutureTox was held in Arlington, Virginia, United States, on October 18–19, 2012.

In order to accomplish the objectives set by the organizing committee, the workshop addressed the following topics: (1) How can predictive toxicology be integrated into existing safety assessment practices? (2) What is the best use of human and environmental exposure, toxicity, and dosimetry data in the context of risk assessment? and (3) How can emerging science impact and reshape current risk assessment practice? Throughout each workshop session, presenters discussed how to apply 21st century toxicology methods for improved and more efficient safety assessment, and how to promote an ordered and scientifically defensible transition to a new toxicity testing and risk assessment paradigm. Workshop participants also discussed how to translate *in vitro* chemical exposures to "real-world" systemic doses and how to eliminate use of arbitrary "uncertainty" factors. The CCT format enabled these discussions and helped participants begin to develop a vision of the path forward.

GETTING FROM HERE TO THERE: FACING MULTIPLE ROADS TO THE FUTURE

The workshop opened with a plenary lecture by Thomas Hartung (Johns Hopkins Bloomberg School of Public Health, Baltimore, MD). In "setting the stage" for the remaining workshop presenters, Hartung described recent explosive growth of the pharmaceutical and chemical industries; he estimated that approximately 300 chemicals and pesticides are currently tested per year in the United States alone at a cost of approximately $2.5 million per chemical and a total cost of $750 million per year. Moreover, new classes of compounds, such as nanoparticles, are being generated and tests for new endpoints, such as endocrine modulation, will be or are already being required to ensure product safety. One needs also to add to this workload the challenge of evaluating the potential health effects of exposures to chemical mixtures. The current rate of growth in the chemical industry alone indicates that the demand for health testing will increase at least 15-fold over the next 2 decades, clearly portending that new testing approaches will be required to avoid unnecessarily crippling chemical innovation with the costly and restrictive funnel of conventional toxicity testing practice. As a rough estimate, without any change in conventional testing requirements, the cost of testing is expected to increase to approximately $11 billion per year just for environmental chemicals alone (excluding pharmaceuticals) and will potentially require use of as many as 4 billion animals in the United States alone. In Europe, there could be additional demand to test as many as 144,000 substances at a cost of €9 billion, potentially using 54 million animals. The magnitude of the problem is obvious,

and the demand for accurate and efficient cost-effective chemical testing cannot be met using existing tools and resources.

Although Hartung's commentary focused on the pragmatic needs for increasing productivity of toxicity testing, a later presentation by Linda Birnbaum (Director, NIEHS) outlined new approaches and initiatives on the fundamental biological/toxicological initiatives already underway at the National Institutes of Health (NIH) that could help meet the challenges ahead. These research investments emphasize exploration of the use of alternative *in vitro* non–animal-based methods and their associated value as efficient tools for prioritization of chemical testing. In addition, an increased focus on more rapidly and clearly understanding fundamental mechanisms of toxicity, when coupled with an improved understanding of human exposure and dosimetry potential, was viewed as a key opportunity to better inform health risks of chemical mixtures, the cumulative effects of low-dose exposure, epigenetic gene-environment interactions, and implications for susceptible human populations.

William Slikker, Director of FDA National Center for Toxicological Research, described another federal initiative between FDA-NIH and Defense Advanced Research Projects Agency (DARPA) to develop engineering platforms more reflective of complex physiological processes that are well-established determinants of toxicity. He described several research initiatives such as "tissue on a chip" that combine cells from several different organs into one integrated platform replicating whole animal physiology and use of stem cells for evaluating developmental toxicity. Dr Slikker referred to the challenges of meeting these objectives as a "moon shot," reflecting both the undeniable complexity and necessity of robust intersector coordination and collaboration required to attain such a goal. However, he tempered this sobering observation by noting that it took less than a decade to accomplish that daunting task when both the resources and political and technical will were directed at the problem.

Workshop attendees also heard from a series of presenters about their direct experiences implementing predictive toxicology and their perspectives on the future of toxicity testing and risk assessment. Timothy Pastoor (Syngenta Crop Protection), cochair of ILSI-HESI Risk21 project, strongly asserted that the collective group of stakeholders associated with that project, with few exceptions, believed that the toxicology and risk communities have not succeeded in efficiently using resources available to identify hazards, predict risk, and ultimately more confidently protect human and environmental health against the potential adverse effects of exposure to drugs and chemicals. Consistent with this view, nearly every speaker at FutureTox concurred that the existing "toolbox" for chemical testing and risk assessment is inadequate, leading to an imperative to change the inventory of that toolbox. In fact, the response to this imperative is already well underway. Hartung and other FutureTox participants outlined ongoing major efforts to modernize parts of the toxicology toolbox and/or novel ways to

put those tools together to develop a new testing and new risk assessment framework. Hartung identified several major initiatives including NexGen, ToxCast, Risk21, the Transatlantic Thinktank for Toxicology, Expocast, the Critical Path Institute, and the FDA-DARPA-NIH initiative to develop complex 3D *in vitro* models for human organs. Importantly, consistent with the complexity and challenge of the tasks, each of these initiatives is a multistakeholder collaborative effort. As stated by George P. Daston (Proctor & Gamble) in a closing summary statement at the end of the workshop, "bigger is better"; more participants are better than fewer participants; more data are better than less data; and more chemicals, cell lines, assays, and models need to be examined using ever more demanding and complex computational resources. Given the magnitude of the challenge, the consensus among the workshop participants was that success will require that all stakeholders leverage all available resources to overcome technical challenges. At the same time, the stakeholders may need to adapt their vision of the future, participating in programs that fit in to a strategic roadmap built on consensus. Thus, regular opportunities for intersector and interdisciplinary dialog are and will continue to be critical.

PRIORITIZATION AND TIER-BASED RISK PROFILING

Workshop participants agreed that screening all chemicals for all possible toxic outcomes using conventional whole animal or even high-throughput testing is neither feasible nor necessary. Rather, it will likely prove far more productive to develop and optimize integrated tools that accurately prioritize chemicals into high-, moderate-, and low-risk bins. However,

the factors that go into the prioritization decision tree, their relative weight, and more importantly, the selection of tools in the prioritization tool box are a matter far from resolved. Nevertheless, a "tier-based" approach was uniformly endorsed at FutureTox as a reasonable and productive "roadmap" objective for advancing these new technologies into practice.

Tier-based prioritization requires rank ordering a large universe of chemicals from high concern to low concern using high-throughput techniques that are informative without being expensive or labor intensive. The relative risk for any compound increases due to high exposure, high toxicity, or both. Risk21 cochair Timothy Pastoor presented one of several graphic representations, termed the Matrix, which captures this concept (Fig. 1). The Matrix is a tool designed to assist in problem formulation. In other words, the Matrix helps to define "what you know and what you need to know." Dr Pastoor emphasized that the problem should not always be formulated as needing to precisely define the hazard; the more relevant way to formulate the problem is to ask "what is the context of human exposure and how does that compare with toxicity-associated doses in a relevant test system." At its heart, the Matrix framework rests on the assumption that the Margin of Exposure is central to interpreting integrated *in vivo* or *in vitro* outputs. Thus, in prioritization frameworks such as the Matrix, it may not always be necessary to identify apical hazard responses in order to assess the potential for human risk. Rather, one should identify the concentration that elicits an adverse response (see PoT, MoA, or AOP: Toxicity as Perturbation(s) of Biological System(s) section) and try to translate or extrapolate that context and dose to real-world human exposures through modeling. In this context, several workshop presenters emphasized that rapid advances in exposure science are providing more precise
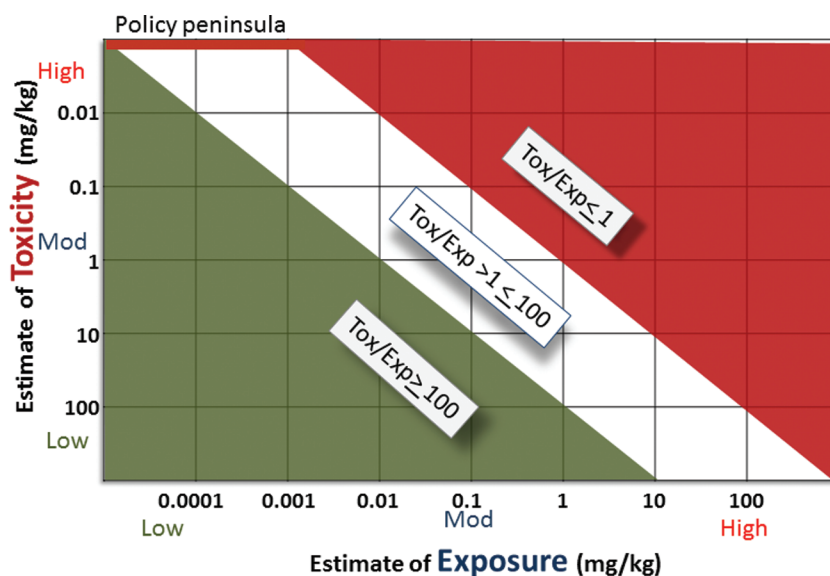
FIG. 1. The RISK21 matrix graphically evaluates risk as a function of relative exposure and relative toxicity, binning compounds into 3 classes (lower left space, low; upper right space, high; zone between high and low, intermediate). Allowable or "safe" exposure to compounds with predicted intermediate risk can be achieved through policy decisions and enforced regulatory guidance. Abbreviations: Exp, exposure; Tox, toxicity.

quantitation of human exposures and helping develop tools for systemic dosimetry in humans. Soon, it will be unacceptable to state "we know little about exposure" and use of arbitrary uncertainty factors will no longer be justifiable. Many presenters at the workshop discussed a tier-based approach to screening for hazard and/or risk, including Ila Cote (U.S. EPA), Alan Boobis, Richard Judson (U.S. EPA), Russell Thomas (Hamner Institute), Richard Becker (American Chemical Council, William Pennie (Pfizer), and J.C.R. (Dow Chemical Company). In addition, several specific strategies for "binning" compounds into high and low priority groups were also described (see further discussion below).

POT, MOA, OR AOP: TOXICITY AS PERTURBATION(S) OF BIOLOGICAL SYSTEM(S)

Although it has been proposed that one can catalog and define a discrete number of cellular pathways leading to toxicity, it is still a matter of controversy how to identify and even how to name such pathways, which have alternatively been called pathways of toxicity (PoT), adverse outcome pathways (AOP), or chemical modes of action (MoA). Thomas Hartung reminded the participants that the initial event leading to adversity is an interaction between a chemical or drug and a biological target and that this interaction initiates disruption of (or perturbs) normal cell homeostasis provoking an acute response (Fig. 2). A cell adjusts to continued impact and dose from the chemical or drug and a new cellular state is achieved, which he called "homeostasis under stress." However, many toxicological assays neglect these first 2 stages of the response and, instead, measure biological parameters unique to the third phase when cellular homeostasis is overwhelmed (ie, during acute intoxication) or when systemic function declines under chronic adversity, leading to hazard manifestation (ie, overt toxicity). Whether the pathways are called PoT, AOP, or MoA,

it is seen as valuable to classify the nature of the initial toxicant cell interaction and to develop predictive models that build on detailed understanding of the steps that lead, in a dose- and time-dependent manner, from the initial interaction to a specific toxicity. However, as noted in the previous section, identification of a specific toxicity may not always be necessary. In a framework based on toxicity pathways (ie, PoT, AOP, or MoA), one needs only to focus on critical early events and link them to one or more specific human exposures.

ATTAINING THE GOAL: EFFICIENT, ACCURATE HIGH-THROUGHPUT SCREENING TOOLS AS A NEW ROAD TO INFORMING HEALTH RISKS

Richard Judson (U.S. EPA) presented an overview of the U.S. EPA ToxCast program and the collaboration U.S. EPA, NIH, U.S. FDA Tox21 program. The goal of these programs is to screen approximately 10 000 compounds using an *in vitro* robotic platform. The first phase of ToxCast was recently completed, producing data on 309 chemicals using 600 assays, with the objective of coupling such test platform outputs with associated data reflecting 1100 *in vivo* endpoints. Ultimately EPA intends to use the data to formulate a complete set of toxicologic pathways and responses to be used in hazard prediction. However, the difficulties associated with developing predictive models of whole organism hazard outcomes from *in vitro* data were underscored by Russell Thomas (Hamner Institutes). Thomas described a systematical study that asked, "how well do the current set of ToxCast assays perform for predicting hazard outcome." In this study, the output of current ToxCast assays was compared with data from an extensive toxicology database of well-characterized chemicals. Somewhat surprisingly, Thomas reported that the *in vitro* ToxCast assays performed only slightly better than expected by chance in predicting hazard outcomes as a whole, and they
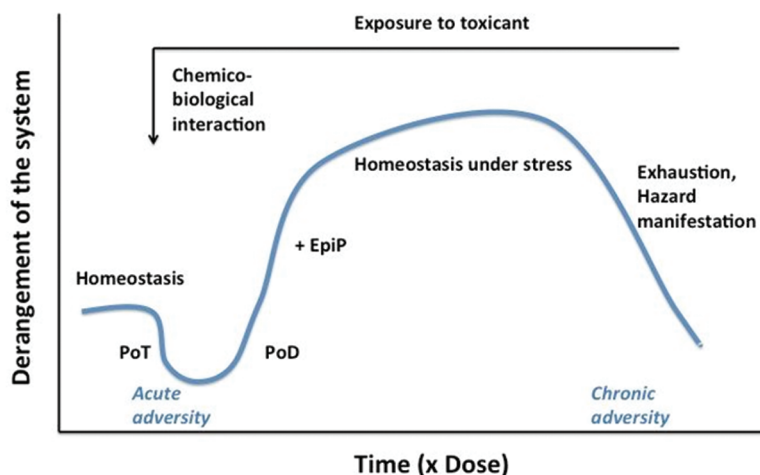
FIG. 2. A system adapts to moderate exogenous stress, such that a new homeostasis is established; repeated stress leads to chronic adversity, and ultimately to exhaustion. Abbreviations: EpiP, epiphenomena (Selye, 1956); POD, point of departure; PoT, pathway of toxicity.

did not perform significantly better than equivalent predictions based on quantitative structure-activity relationship (QSAR) modeling. Thomas concluded that the current suite of 600 *in vitro* ToxCast assays *per se* is poorly suited to the task of hazard prediction. William Pennie (Pfizer Global Research) described similar conclusions reached at Pfizer when describing an approach used in Pfizer for identifying compounds likely to cause low-dose toxicity relative to doses necessary for pharmacologic efficacy. Pennie underscored his view that the available toxicity evaluation tools, regardless of how they are coupled, are not currently capable *a priori* of predicting organ-specific endpoints. Despite this complication, Thomas emphasized that the relatively broad range of *in vitro* assays could still be highly informative for identifying potential human risk. In particular, better results are achieved when target selectivity and potency are considered. Thomas suggested that target-specific endpoints can and should be differentiated from nonspecific endpoints in test systems. Thomas concluded by describing a framework to use *in vitro* and genomic data in which margins of exposure between human doses encountered in the real world and concentrations eliciting responses in high-throughput test systems can be compared. This approach

can be used to determine whether real-world exposures are sufficiently differentiated to be considered safe (Fig. 3). He noted that additional human exposure data and modeling are critical important to the future success of the framework, as highlighted in the previous section and detailed below. There was general consensus by all presenters that conversions between reverse dosimetry and high-throughput exposure modeling must be developed as a parallel and ultimately converging road, if high-throughput testing approaches are ultimately to be employed with success in a new risk assessment paradigm.

THE THREE E'S: EXPOSURE, EXPOSURE, AND EXPOSURE

As noted above, early in the first workshop session Timothy Pastoor challenged the audience by asserting that it is no longer an option to attempt to assess risk without a reasonable estimate of exposure. John Wambaugh (U.S. EPA), Harvey Clewell (The Hamner Institutes), Amin Rostami-Hodjegan (University of Manchester, Simcyp Limited, United Kingdom), J.S.B. (The Dow Chemical Company), and Sean Hays (Summit
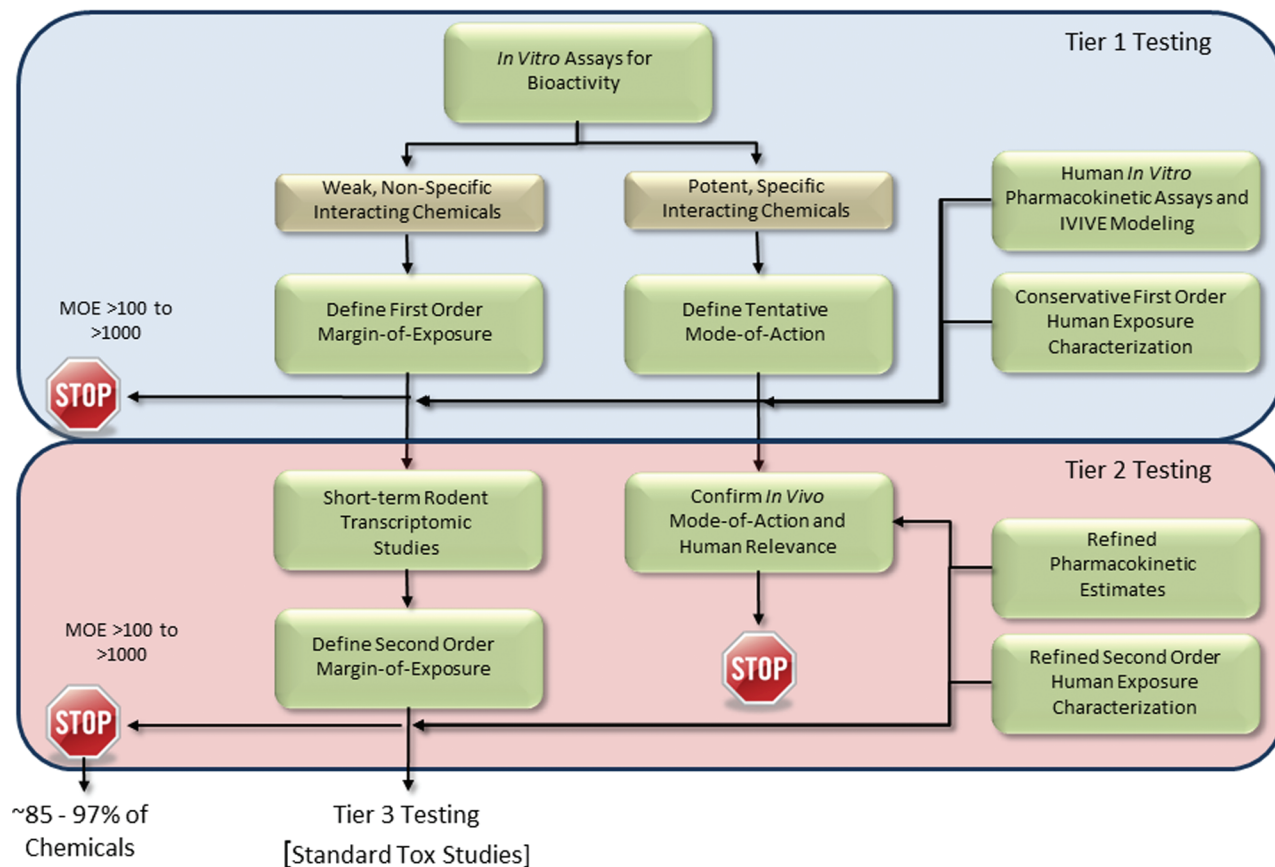

FIG. 3. Proposal for a data-driven framework for toxicity testing and risk assessment.

Toxicology) addressed this challenge in the third workshop session. John Wambaugh spoke to the general feasibility of high-throughput exposure analysis as an alternative and/or supplement to conventional exposure analysis. Wambaugh described the EPA ExpoCast project, the exposure counterpart to ToxCast, which is attempting to extend the effective realm for exposure modeling, from simple models based solely on chemical fate and transport such as USETox (Hauschild *et al.*, 2008) and The Risk Assessment IDentification And Ranking (RAIDAR) Model (Arnot *et al.*, 2010) to complex models suitable for evaluating exposures in indoor environments, including use of consumer products. Very revealingly, when ToxCast and ExpoCast data were compared by Wambaugh, the estimated human exposures to approximately 90% of the chemicals analyzed fell several orders of magnitude below the lowest concentration eliciting a response in the ToxCast assays (Fig. 4). Thus, even with a low degree of confidence in predicted exposure, the data suggest that ExpoCast and similar exposure estimates, when coupled with outputs from test systems such as ToxCast, may be sufficient to allow compounds to be binned according to relative risk for potential adverse human health effects. However, the exposure models used in ExpoCast rely on QSAR modeling for novel chemicals, which is a significant limitation of this approach. This problem, that QSAR-based models are restricted to a limited domain of applicability, was raised several times during FutureTox presentations and discussions. Because QSAR modeling is currently most successful for drugs

(ie, compounds that are nonlipophilic, water soluble, and non-volatile), there is great need to expand the fraction of chemical space to which QSAR-based models can be reliably applied. On the other hand, human biomonitoring data are currently limited to a relatively small number of chemicals. Nevertheless, expansion of the domain for QSAR-based models is seen as critical future need.

Harvey Clewell extended the discussion of human exposure modeling at the workshop, focusing on the multiple pharmacokinetic processes that can be incorporated into biokinetic models in order to calculate the human exposure that is equivalent to a chemical concentration from an *in vitro* assay. Amin Rostami-Hodjegan addressed the need to account for population variance in human exposure, arguing that it is possible to consolidate data from diverse sources and leverage physiologically based pharmacokinetic (PBPK) modeling toward this end (Jamei *et al.*, 2009; Rostami-Hodjegan, 2012). Sean Hays suggested further value of constructing human exposure paradigms reflecting internal systemic dose estimates, ie, toxicant concentration in human blood or urine samples, primarily because the systemic doses reflect aggregate exposure over all possible sources and thus require fewer assumptions and associated uncertainty factors. However, Hays also pointed out that spot urine collection, the method most commonly associated with human biomonitoring programs, must be interpreted in light of the fact that intraindividual measurements of a specific chemical can vary over 3 orders of magnitude in a single day
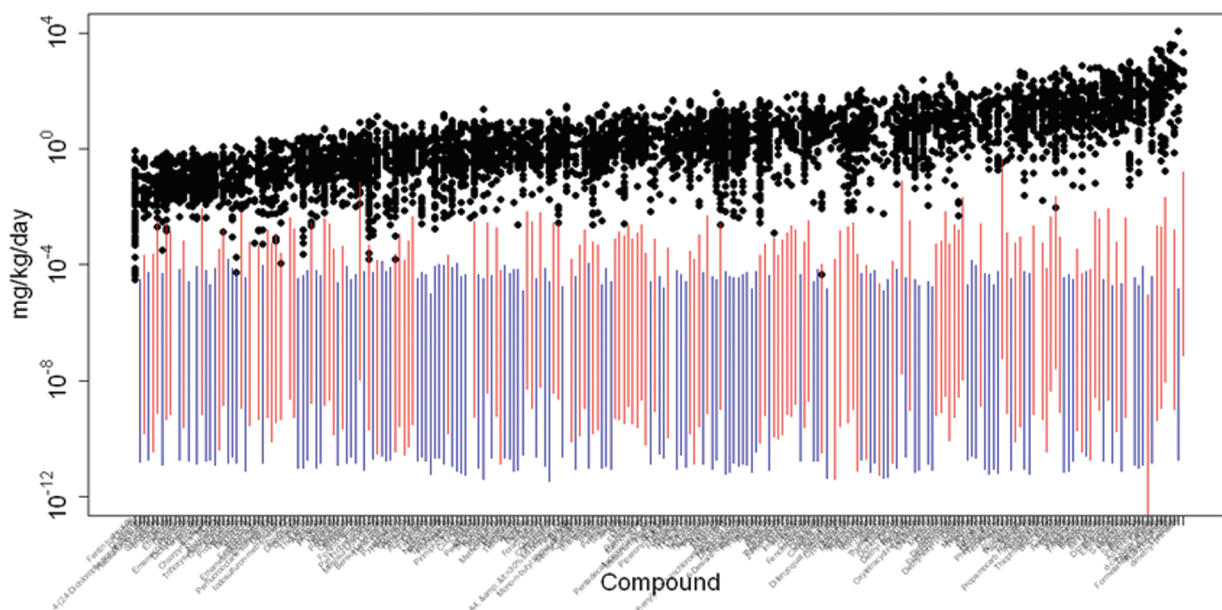
FIG. 4. Exposure-based data from ExpoCast (statistical and graphical extrapolation) and toxicity data from ToxCast are used in an integrated manner to prioritize chemicals. ToxCast data (oral equivalent of AC_{50}); ExpoCast, near field (indoor/consumer) use, and far field (industrial/agricultural) use bars.

for toxicokinetically short half-life substances, whereas the daily or weekly average concentrations of a chemical in urine of a single individual or in groups of individuals vary over 2 or 1 orders of magnitude, respectively (Li *et al.*, 2010).

J.S.B. argued for the importance and value of measuring blood and/or target tissue concentrations of toxicants and/or key metabolites in laboratory animals during conventional dietary, drinking water, or inhalation toxicity studies, with a particular focus on identifying systemic blood and/or tissue concentrations at both the no and lowest observed effect level doses. He observed that rapid advances in analytical sciences have substantially increased the feasibility of such approaches (Saghir *et al.*, 2006). J.S.B. termed this approach "dosimetric anchoring," ie, exploiting comparisons of systemic blood/tissue concentrations eliciting both no-effect and effect level responses *in vivo* with concentrations inducing responses in *in vitro* high-throughput test systems as a valuable aid in interpreting the potential health impacts of the outputs *in vitro* tests. Toxicity responses observed in *in vitro* tests at concentrations substantially higher than blood concentrations measured at NOEL and/or LOEL doses in *in vivo* animal toxicity studies are unlikely to represent a human risk if real-world human exposures also result in blood concentrations well differentiated from those at animal NOEL/LOEL doses. J.S.B. noted that dosimetric anchoring is an exposure dose parallel to phenotypic anchoring, in which apical toxicity and genomic and/or other biomarker responses are correlated and have been productively used to inform health implications of those responses. Dosimetric anchoring data also will be useful for guiding and/or interpreting future toxicity tests based on *in vitro* assays and for developing validated QIVIVE models. The dosimetric anchoring concept is especially attractive because there is an extensive database of exposure-outcome data in rodent short- and long-term toxicity testing assays; therefore, there is a valuable opportunity to leverage this legacy knowledge toward improving future risk assessment frameworks.

A HOLY GRAIL: REPLACE UNCERTAINTY FACTORS WITH SCIENCE-INFORMED RISK ASSESSMENT

In the risk assessment framework of the future, toxicologists, risk assessors, and risk managers aim to either reduce or entirely remove the uncertainty factors associated with cross-species extrapolation, *in vitro* to *in vivo* extrapolation, and human population variation. Advances in QIVIVE modeling, as discussed above, are facilitating progress toward this goal. In addition, direct *in vitro* approaches have been developed to study comparative toxicity in human and other animal model test systems. J.C.R. (The Dow Chemical Company) pointed out that there are no roadblocks for applying 21st century methods today for reducing uncertainty in risk assessments. He demonstrated this using transcriptional profiling of toxicant-exposed primary cells to directly evaluate uncertainty factors for quantitative

and qualitative differences in toxicant responses across species. Analysis of the transcriptomic profiles of primary rat and human cell cultures exposed to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) allowed a direct comparison of the response to TCDD between the 2 species (Black *et al.*, 2012). After measuring dose-dependent transcriptional induction of key aryl hydrocarbon receptor–regulated genes in primary cells, benchmark dose analysis showed an average 18-fold difference in potency among orthologs that were differentially expressed in human and rat, with human hepatocytes less sensitive than rat hepatocytes. The data also demonstrate significant qualitative cross-species differences at the gene and pathway level, with many orthologous genes and pathways responding in a discordant manner in rats and humans. These data demonstrate that transcriptional profiling of toxicant-exposed primary cells is a useful approach to directly evaluate quantitative and qualitative differences in toxicant responses across species.

Approaches were presented for using 21st century methods to understand health impact of population sensitivity. Raymond Tice (NIEHS) described ongoing experiments at NIEHS in which toxicant response profiles are being studied in a genetically diverse population of mice (the Diversity Outbred Mouse; Chesler *et al.*, 2008). Gene toxicity studies revealed that resistant and sensitive strains can be differentiated, and that in sensitive strains, the number of micronuclei in bone marrow cells after exposure to benzene varies up to 100-fold. Such experimental information may be relevant to understanding human variation in the response to benzene. Ivan Rusyn (University of North Carolina) is exploiting repositories of immortalized human lymphoblast cell lines, generated by the International HAP-MAP Consortium and the 1000 Genomes Project, to study variation in susceptibility to toxicants in these human cell lines (O'Shea *et al.*, 2011). These cell lines represent the geographic and racial diversity of the entire global human population and allow assessment of population-wide variability in response to chemicals and the heritability of such responses. In one experiment, the response to 240 compounds was examined at 12 chemical concentrations (0.3nM–46mM) in 81 cell lines—including 27 genetically related trios in 2 assays: 1 assay measured ATP production and 1 assay measured Caspase activation, both of which are used as general measures of cytotoxicity (Lock *et al.*, 2012). In another experiment, 1086 cell lines were exposed to 179 compounds at 8 concentrations (0.33nM–92mM) and tested in 1 assay. Although more work is needed in this area, these types of approaches are very promising for understanding the distribution of sensitivities to specific chemicals and the concurrent influence of mode of action considerations on toxicity variability.

THE ELEPHANT AND THE GORILLA

Participants at FutureTox achieved consensus on many immediate goals and future needs including reducing use of "blind"

uncertainty factors related to cross-species extrapolation in human risk assessment, to improve links of biokinetic models with exposure refinements to support quantitative *in vitro* to *in vivo* extrapolation in humans, to continue dialog that encourages interdisciplinary and cross-sector collaborations and consensus building among all stakeholders, to engage the regulatory risk assessment community in development and implementation of new risk assessment frameworks reflecting advances in both biological and exposure sciences, to increase biomonitoring efforts and generate more "real-world" human exposure data, and to expand the QSAR and/or quantitative structure property/activity relationship (QSPAR) domain of applicability so that it is relevant to environmental chemicals as well as pharmaceuticals.

Workshop participants also identified several barriers to success, including the inherent difficulty in predicting organ- and tissue-specific responses solely from responses in either single or a battery of *in vitro* tests and in measuring organ- and tissue-specific dose and the need to rethink whether convergence of phenotypic anchoring is necessary to inform risk decisions, resistance to change from past risk assessment approaches emphasizing need to identify a hazard response, and inadequate resources to address the critical needs for transitioning to more productive testing and risk characterization paradigms. Workshop presenters mentioned one "elephant," and one "800-pound gorilla," standing in the way of future progress: the "elephant in the room" is "chemical mixtures"; the "800-pound gorilla" is "metabolism." Clearly, it is challenging to understand the complex influence of metabolizing systems on exposure outcomes *in vivo*. Thus, it is important to understand the metabolic capacity of *in vitro* models and their relationships to whole animal metabolism in order to accurately interpret results. Similarly, the dose-dependent roles of metabolism and detoxification will be difficult to assess. In addition, parent chemical metabolism and the role of specific metabolite(s) as drivers of MoA are frequent weaknesses in many risk assessments. Humans are not exposed simply to 1 compound by 1 route of exposure for a defined length of time, and the predominant practice of conventional toxicity testing does not account for this complexity. Insights into the health implications of mixtures will likely be aided by biomonitoring studies, but at present remains a difficult part of the overall exposure equation. Another "real-world" complication is the unaccounted for exposure to a diverse mixture of natural chemicals in the human diets and the typical human indoor environment. The need to improve exposure models in this regard is well recognized.

CONCLUSIONS

A common theme that emerged from presentations and the roundtable discussions at FutureTox is that a "paradigm shift" in toxicology will take place through incremental steps rather than by revolution. Specific uses for new technologies should be explored and targeted for decision making. Among these are prioritization of chemicals for further testing; read-across hazard profiles to estimate the hazards associated with a novel chemical, for which little data are available; improved understanding of chemical modes of action; reduced uncertainty associated with human variability and susceptibility; improved cross-species comparison and relevance; better understanding and more information on outcomes after low-dose exposure, leading to a more complete more relevant analysis of dose-response relationships in humans.

Historically, the link between policy-based default assumptions and science-informed risk has been weak. Workshop participants recognized that future risk assessment paradigms may be informed better by science by relying less on apical hazard identification through conventional toxicity testing and instead increasing the emphasis toward PoT/AOP/MoA pathways and refined assessments of real-world human exposure. Importantly, the dialog and presentations at this workshop revealed that rapid progress is being made in addressing many technical challenges of these new methods so that they can eventually be used in risk assessment. The focus has largely been methods for hazard identification, but it was clear to all that more emphasis is needed on exposure assessment and the dose-dependent roles of metabolism and detoxification. The advent of 21st century testing technologies has opened the door for toxicity testing refinements that could lead to science-informed decision making in the near future. Achieving that objective will require that regulatory agencies and other stakeholders work in concert to define a clear strategic roadmap toward science-informed risk assessment.

By bringing together toxicologists, exposure scientists, risk assessors, and regulators, FutureTox provided a unique environment for formulating a new vision of 21st century toxicity testing, exposure science, and risk assessment.

REFERENCES

Arnot, J. A., Mackay, D., Parkerton, T. F., Zaleski, R. T., and Warren, C. S. (2010). Multimedia modeling of human exposure to chemical substances: The roles of food web biomagnification and biotransformation. *Environ. Toxicol. Chem.* **29,** 45–55.

Basketter, D. A., Clewell, H., Kimber, I., Rossi, A., Blaauboer, B., Burrier, R., Daneshian, M., Eskes, C., Goldberg, A., Hasiwa, N., *et al.* (2012). A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing—t4 report*. *ALTEX* **29,** 3–91.

Blaauboer, B. J., Boekelheide, K., Clewell, H. J., Daneshian, M., Dingemans, M. M., Goldberg, A. M., Heneweer, M., Jaworska, J., Kramer, N. I., Leist, M., *et al.* (2012). The use of biomarkers of toxicity for integrating *in vitro* hazard estimates into risk assessment for humans. *ALTEX* **29**, 411–425.

Black, M. B., Budinsky, R. A., Dombkowski, A., Cukovic, D., LeCluyse, E. L., Ferguson, S. S., Thomas, R. S., and Rowlands, J. C. (2012). Cross-species comparisons of transcriptomic alterations in human and rat primary hepatocytes exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Toxicol. Sci.* **127**, 199–215.

Chesler, E. J., Miller, D. R., Branstetter, L. R., Galloway, L. D., Jackson, B. L., Philip, V. M., Voy, B. H., Culiat, C. T., Threadgill, D. W., Williams, R. W., *et al.* (2008). The collaborative cross at Oak Ridge National Laboratory: Developing a powerful resource for systems genetics. *Mamm. Genome* **19**, 382–389.

Hardy, B., Apic, G., Carthew, P., Clark, D., Cook, D., Dix, I., Escher, S., Hastings, J., Heard, D. J., Jeliazkova, N., *et al.* (2012). Food for thought. A toxicology ontology roadmap. *ALTEX* **29**, 129–137.

Hauschild, M. Z., Huijbregts, M., Jolliet, O., MacLeod, M., Margni, M., van de Meent, D., Rosenbaum, R. K., and McKone, T. E. (2008). Building a model based on scientific consensus for Life Cycle Impact Assessment of chemicals: The search for harmony and parsimony. *Environ. Sci. Technol.* **42**, 7032–7037.

Jamei, M., Dickinson, G. L., and Rostami-Hodjegan, A. (2009). A framework for assessing inter-individual variability in pharmacokinetics using virtual human populations and integrating general knowledge of physical chemistry, biology, anatomy, physiology and genetics: A tale of 'bottom-up' vs 'top-down' recognition of covariates. *Drug Metab. Pharmacokinet.* **24**, 53–75.

Judson, R. S., Kavlock, R. J., Setzer, R. W., Hubal, E. A., Martin, M. T., Knudsen, T. B., Houck, K. A., Thomas, R. S., Wetmore, B. A., and Dix, D. J. (2011). Estimating toxicity-related biological pathway altering doses for high-throughput chemical risk assessment. *Chem. Res. Toxicol.* **24**, 451–462.

Judson, R. S., Martin, M. T., Egeghy, P., Gangwal, S., Reif, D. M., Kothiya, P., Wolf, M., Cathey, T., Transue, T., Smith, D., *et al.* (2012). Aggregating data for computational toxicology applications: The U.S. Environmental Protection Agency (EPA) Aggregated Computational Toxicology Resource (ACToR) System. *Int. J. Mol. Sci.* **13**, 1805–1831.

Li, Z., Romanoff, L. C., Lewin, M. D., Porter, E. N., Trinidad, D. A., Needham, L. L., Patterson, D. G., Jr, and Sjödin, A. (2010). Variability of urinary concentrations of polycyclic aromatic hydrocarbon metabolite in general population and comparison of spot, first-morning, and 24-h void sampling. *J. Expo. Sci. Environ. Epidemiol.* **20**, 526–535.

Liebsch, M., Grune, B., Seiler, A., Butzke, D., Oelgeschläger, M., Pirow, R., Adler, S., Riebeling, C., and Luch, A. (2011). Alternatives to animal testing: Current status and future perspectives. *Arch. Toxicol.* **85**, 841–858.

Lock, E. F., Abdo, N., Huang, R., Xia, M., Kosyk, O., O'Shea, S. H., Zhou, Y. H., Sedykh, A., Tropsha, A., Austin, C. P., *et al.* (2012). Quantitative high-throughput screening for chemical toxicity in a population-based *in vitro* model. *Toxicol. Sci.* **126**, 578–588.

National Research Council (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. The National Academies Press, Washington, DC.

National Research Council (2009). *Science and Decisions: Advancing Risk Assessment*. The National Academies Press, Washington, DC.

O'Shea, S. H., Schwarz, J., Kosyk, O., Ross, P. K., Ha, M. J., Wright, F. A., and Rusyn, I. (2011). *In vitro* screening for population variability in chemical toxicity. *Toxicol. Sci.* **119**, 398–407.

Price, D. A., Blagg, J., Jones, L., Greene, N., and Wager, T. (2009). Physicochemical drug properties associated with *in vivo* toxicological outcomes: A review. *Expert Opin. Drug Metab. Toxicol.* **5**, 921–931.

Rostami-Hodjegan, A. (2012). Physiologically based pharmacokinetics joined with *in vitro-in vivo* extrapolation of ADME: A marriage under the arch of systems pharmacology. *Clin. Pharmacol. Ther.* **92**, 50–61.

Rotroff, D. M., Wetmore, B. A., Dix, D. J., Ferguson, S. S., Clewell, H. J., Houck, K. A., Lecluyse, E. L., Andersen, M. E., Judson, R. S., Smith, C. M., *et al.* (2010). Incorporating human dosimetry and exposure into high-throughput *in vitro* toxicity screening. *Toxicol. Sci.* **117**, 348–358.

Saghir, S. A., Mendrala, A. L., Bartels, M. J., Day, S. J., Hansen, S. C., Sushynski, J. M., and Bus, J. S. (2006). Strategies to assess systemic exposure of chemicals in subchronic/chronic diet and drinking water studies. *Toxicol. Appl. Pharmacol.* **211**, 245–260.

Selye, H. (1956). *The Stress of Life*. McGraw-Hill Book Company, New York, NY.

Thomas, R. S., Black, M. B., Li, L., Healy, E., Chu, T. M., Bao, W., Andersen, M. E., and Wolfinger, R. D. (2012). A comprehensive statistical analysis of predicting *in vivo* hazard using high-throughput *in vitro* screening. *Toxicol. Sci.* **128**, 398–417.

Wetmore, B. A., Wambaugh, J. F., Ferguson, S. S., Sochaski, M. A., Rotroff, D. M., Freeman, K., Clewell, H. J., 3rd, Dix, D. J., Andersen, M. E., Houck, K. A., *et al.* (2012). Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. *Toxicol. Sci.* **125**, 157–174.