FORUM

FutureTox II: In vitro Data and In Silico Models for Predictive Toxicology


aUnited States Environmental Protection Agency, Research Triangle Park, North Carolina 27711, bSanofi, Bridgewater, New Jersey 08807, cPage One Editorial Services, Boulder, Colorado 80304, dDow Chemical Company, Midland, Michigan 48674, eHealth and Environmental Sciences Institute, Washington, District of Columbia 20005, fUniversity of Washington, Seattle, Washington 98105, gUnited States Food and Drug Administration, Silver Spring, Maryland 20993, hSanofi, Bethesda, Maryland 20814, iNational Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, jUniversity of North Carolina, Chapel Hill, North Carolina 27599, kThe Hamner Institutes, Research Triangle Park, North Carolina 27709, and lEuropean Commission Joint Research Centre, I-21027 Ispra, Italy

1These authors contributed equally to this report. 2To whom correspondence should be addressed at Sanofi, 55C-420A, 55 Corporate Drive, Bridgewater, NJ 08807. E-mail: douglas.keller@sanofi.com.
3Present address: Hastings Toxicology Consulting LLC, Mount Airy, Maryland 21771.

Disclaimer: The information in these materials is not a formal dissemination of information by FDA and does not represent agency position or policy. The views expressed in this article are those of the authors and do not necessarily reflect the views of policies of the U.S. Environmental Protection Agency, the National Institute of Environmental Health Sciences, the National Institute of Health, or any other U.S. federal institute. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ABSTRACT

FutureTox II, a Society of Toxicology Contemporary Concepts in Toxicology workshop, was held in January, 2014. The meeting goals were to review and discuss the state of the science in toxicology in the context of implementing the NRC 21st century vision of predicting in vivo responses from in vitro and in silico data, and to define the goals for the future. Presentations and discussions were held on priority concerns such as predicting and modeling of metabolism, cell growth and differentiation, effects on sensitive subpopulations, and integrating data into risk assessment. Emerging trends in technologies such as stem cell-derived human cells, 3D organotypic culture models, mathematical modeling of cellular processes and morphogenesis, adverse outcome pathway development, and high-content imaging of in vivo systems were discussed. Although advances in moving towards an in vitro/in silico based risk assessment paradigm were apparent, knowledge gaps in these areas and limitations of technologies were identified. Specific recommendations were made for future directions and research needs in the areas of hepatotoxicity, cancer prediction, developmental toxicity, and regulatory toxicology.

Key words: in vitro; in silico; modeling; predictive toxicology; risk assessment

©The Author 2015. Published by Oxford University Press on behalf of the Society of Toxicology.
All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com
In 2007, the National Research Council (NRC) published Toxicity Testing in the 21st Century: A Vision and a Strategy. Hazard-based values for regulatory toxicology are traditionally estimated from laboratory animal studies. However, because of the large number of chemicals in commerce with little or no toxicity information, the NRC report highlighted the need for high-throughput screening (HTS) technologies to identify chemically induced biological activity in human cells and cell lines and to develop predictive models of in vivo biological response for targeted testing. In the 7 years since publication of the NRC report, the implementation of the 21st century vision and implications of this new strategy for basic science and regulatory decision-making have been extensively discussed and debated (Andersen and Krewski, 2010; Boekelheide and Campion, 2010; Bus and Becker, 2009; Chapin and Stedman, 2009; Kavlock et al., 2012; MacDonald and Robertson, 2009; Meek and Doull, 2009; NRC 2007a; Sturla et al. 2014; Walker and Bucher, 2009). The toxicology community has made significant progress developing assays and tools that will help achieve the predictive toxicology goals outlined by the NRC in 2007. This raises the important question of how in vitro data and in silico models can be used to understand and predict in vivo toxicity. This question was the central focus of a Society of Toxicology (SOT) Contemporary Concepts in Toxicology (CCT) Workshop held in January 2014 in Chapel Hill, North Carolina. FutureTox II, the second in the FutureTox series (Rowlands et al., 2014), was attended in person and via webcast by more than 400 scientists from governmental and regulatory agencies, research institutes, academia, and the chemical and pharmaceutical industries in Europe, Canada, and the United States. The agenda for FutureTox II can be found at https://www.toxicology.org/ai/meet/cct_futureToxII.aspx#program (accessed October 22, 2014). Workshop participants reviewed and discussed the state of the science in toxicology and human risk and exposure assessment and attempted to define collective goals for the future. This article reports on the key issues covered in FutureTox II with regard to state of the science and challenges to implementation of the new testing paradigm in understanding toxicological risks.

Many efforts have been initiated toward developing new assays for toxicity testing and models to integrate large datasets into an emerging risk assessment framework. A major challenge to progress is the complex nature of biological systems. Reducing the complexity of test systems to simpler cell and small model organisms (such as Caenorhabditis elegans and zebrafish) enables the application of higher throughput testing strategies, but in so doing may lose many of the systems-level characteristics that make a human toxicological response complex. Programs such as Tox21, ToxCast™, SEURAT-1 (and other EU-sponsored efforts), and NIH-NCATS-FDA-sponsored work on microphysiological systems have made significant contributions toward implementation of approaches that are scalable to the problem, but much remains to be accomplished. Mathematical modeling of complex integrated biological systems remains somewhat of a bottleneck. Biological models of metabolism, pharmacokinetics, and risk estimation have been prominent in toxicology for a few decades; however, new graphical and analytical tools and methods are needed in order to “decode the toxicological blueprint of active substances that interact with living systems” (Sturla et al. 2014). Ultimately, the goal is a detailed mechanistic, quantitative understanding of toxicological processes—from a specific molecular initiating event to a specific biological outcome (Ankley et al., 2010), in the context of an integrated biological system.

A number of factors drive society’s need for an accurate and efficient paradigm for predictive toxicology (Figure 1). Although some of these drivers are specific to a region of the world, others are relevant worldwide, and can only be addressed through international cooperation and harmonization. In the European Union (EU), Directive 2010/63/EU legislates an end to “the use of animals in toxicology testing and biomedical research as soon as scientifically feasible to do so.” Also in the EU, “REACH” is a directive that restricts animal testing only “as a last resort to satisfy registration information requirements,” whereas Regulation 1223/2009/EU legislates a comprehensive ban on marketing within the EU of any cosmetic product (or ingredient thereof) that has been tested on animals since March 2013. This means that, in the EU, effective in vitro / in silico tools for predictive toxicology are an urgent priority. Predictive toxicology will also play an important role in satisfying other societal and legislative demands: these include, the need to protect human health and the environment from (1) endocrine-disrupting chemicals, and (2) the effects of chemical mixtures. The US FDA endorses the effort “to reduce animal testing [and] to work towards replacement of animal testing” as a basis for regulatory action. This policy may reflect concern that predictive models based on animal testing fail to account for the rate of reported adverse events and lethal adverse events among pharmaceutical drug users in the USA. Notably, the frequency of reported adverse events increased steadily from 1998 to 2005 (Moore et al., 2007). This suggests that current test methods may neglect factors that contribute to “false negative” predictions, such as: polypharmacology; age-, gender-, or ethnicity-linked drug susceptibility; genetic diversity of human populations, and failure to independently verify study outcomes. Alternatively, some animal models may simply be inadequate for extrapolation to humans. Looking to overcome these problems, one must ask the question, “what types of models need to be developed, and how will human genetic diversity be incorporated into the models?”

FutureTox II explored a broad range of current toxicological research to address those and other questions. Priority concerns and emerging areas in the field of predictive toxicology are summarized in Box 1. Priority concerns include, for example, predicting, modeling, or experimentally evaluating the role of metabolism on toxicological outcomes; modeling chemical mixtures; understanding the controls of cell growth and differentiation; identifying and characterizing human subpopulations susceptible to specific adverse outcomes; and developing tools for integrating multiple types of data from diverse experimental systems into a unified risk assessment paradigm. Emerging trends in technologies include, for example, increased use of induced pluripotent stem cell (iPSC)-derived human cells; development of defined heterotypic cell and three-dimensional (3D) organotypic culture models; engineered micro-scale physiological systems; mathematical modeling of cellular processes and morphogenesis; Adverse Outcome Pathways (AOPs) as a regulatory tool; and high-content imaging of in vivo systems and small model organisms. In addressing these key issues, the question, “how do currently available in vitro / in silico methodologies compare to in vivo methods—are they as good or superior, and will their predictive accuracy eventually be high enough to obviate the need for in vivo models?” must be asked.

The charge presented by conference organizers to FutureTox II participants included the following 3 goals: (1) to address progress and advances toward a paradigm where improvements to predictivity and concordance are based on in vitro / in silico approaches; (2) to provide a forum to integrate new in vitro methodologies and computational (in silico) modeling approaches with advances in systems biology; and (3) to clarify
the usefulness and validity of new and emerging technologies and approaches, so that expectations can be managed in both the regulatory and regulated scientific communities. Suggestions in considering the state of the science since the NRC 2007 report and insight into future technologies led to recommendations for future activities to achieve the goals projected in that report. The following synopsizes these main themes, an analysis of the progress, significant
PREDICTIVE TOXICOLOGY

Converting Tools into Solutions

One of the largest predictive toxicology programs undertaken to date is the US-based Tox21 consortium of federal agencies (Collins et al., 2008) comprising US Environmental Protection Agency’s (EPA) National Center for Computational Toxicology, the National Institutes of Environmental Health Sciences (NIEHS)/National Toxicology Program (NTP), the National Center for Advancing Translational Sciences (NCATS), and the US Food and Drug Administration (FDA). The goals of Tox21 and EPA’s ToxCast™ are to identify targets and pathways linked to toxicity, develop relevant high-throughput assays and predictive signatures, and apply predictive models to prioritizing chemicals for testing chemicals of interest (Kavlock et al., 2012; Tice et al., 2013). Significant bioinformatics resources are devoted to storing, curating, analyzing, and modeling the millions of chemical assay data points and disseminating the output to the public through the Interactive Chemical Safety for Sustainability (iCSS) dashboard (http://actor.epa.gov/dashboard/; accessed September 18, 2014) and the Chemical Effects in Biological Systems (CEBS) database (http://www.niehs.nih.gov/research/resources/databases/cebs/; accessed September 18, 2014). Univariate models correlating in vitro bioactivity with in vivo toxicity utilize knowledge of target gene for various assays, gene family or pathways, and cellular state changes to produce multivariate models of toxicity (e.g., predictive assay-endpoint combinations). Extrapolating in vitro effects to in vivo prediction faces the general problem of false-positive (in vitro positive, in vivo negative) and false-negative (in vitro negative, in vivo positive) results that may arise for many reasons: pharmacokinetic issues that impact biotransformation and/or clearance in vivo; incomplete assay coverage of molecular pathways and biological processes; physical limitations of complex multi-cellular networks and interactions between diverse cell types; statistical power in analyzing diverse, multidimensional data sets; and the potential for in vivo adaptation through homeostatic mechanisms. In vitro to in vivo modeling does hold promise in predictive toxicology when in vitro assays can plausibly be used to link specific molecular endpoint perturbations to an adverse outcome pathway (AOP) in humans or ecological populations. Before this vision becomes a reality, AOPs themselves must be established with enough precision and detail to enable an understanding of false-positive and -negative predictions, noted above. The complexity of the ToxCast data set is exemplified by the bioactivity profiles for chemicals tested in 8 cell systems with 87 endpoints (Kleinstreuer et al., 2014). Although some known activities could be predicted by these assays, a large number of gaps exist in both the assay capabilities and the in vivo toxicity profiles of the chemicals. Additional analysis of ToxCast data by Wetmore et al. (2013) indicated that although adjustment for pharmacokinetic factors did not improve the ability of the in vitro assay to predict in vivo effects, there was some in vitro-in vivo correlation of effects indicating potential to identify key molecular and cellular effects from in vitro HTS profiling.

Knowing the potential targets and pathways of toxicity linked to AOPs can lead to development of relevant HTS assays in areas for improvement in the science and future needs in these areas.

Box 1

Key Workshop Discussions: Challenges and Potential Solutions

The needs of the global toxicology community can only be met through global collaborative efforts devoted to achieving international harmonization.

The EU cosmetics testing ban challenges the “status quo” for chemical testing, and creates incentive and motivation to move beyond animal testing.

The potential danger from human exposure to chemical mixtures is very great, but until the mechanisms/modes of action (MOA) of the components of a mixture are known, it will remain impossible to estimate the magnitude or nature of the danger. In silico prediction and modeling of the metabolic fate of a toxicant is complex and not currently accomplished with automated systems.

The mechanisms that regulate cell growth are incompletely understood; as this understanding improves, toxicological models will also improve.

An important future goal is to incorporate human genetic diversity into toxicological models in order to predict specific effects on specific subpopulations.

Big data sets and complex biological systems are presenting challenges in data analysis that are not being met with current analytical tools. New analytical tools are urgently needed.

SEURAT-1 and similar programs are working toward classifying compounds into groups of mechanistic analogs, based on both structure and MOA; this could point the way toward toxicological risk management using an in vitro/in silico approach.

Stem cell technologies are emerging and evolving rapidly. Once mature, these methods will provide a wide variety of genetically well-defined differentiated human cells, which will be used to study diseases and susceptibility factors for toxicity.

Adverse Outcome Pathways (AOPs) link key initiating events and apical effects. Challenges to using an AOP framework include: modeling responses to mixtures and multiple or mixed modes of action, quantitative prediction of outcomes, incorporating absorption, distribution, metabolism, and excretion (ADME) and compensatory responses.

Advanced methods of mathematical modeling now provide tools for simulating complex biological processes in silico, including cell migration, limb growth, and vasculogenesis.

Systems that model the complex heterotypic cellular interactions in the human liver and other tissues are yielding insight into mechanisms of toxicity, and dynamic pharmacokinetic models of these systems will improve predictions of fate and effect.

In vitro assays can plausibly be used to quantify specific molecular endpoints that are unequivocally linked to AOPs in humans. However, before this goal becomes a reality, AOPs must be elucidated with greater precision and detail, and the factors leading to false-positive and -negative predictions must be understood in greater detail.
and novel systems models to allow in silico — in vitro — in vivo prediction. An example of how this could work is the strategy for effectively prioritizing test chemicals for the US Endocrine Disruptor Screening Program (EDSP). Under this program, US EPA has a mandate to evaluate 100 chemicals per year for potential endocrine disruptor activity, an activity that has been estimated by the agency to cost up to $1 million per chemical. A set of 18 ToxCast assays are being used to prioritize chemicals, assign relative risk, and select a subset of chemicals for additional study of estrogenic responses (Rotroff et al., 2013). An analysis of the relative specificity or promiscuity of assays/targets/chemicals have revealed the general concept of an “assay burst” of promiscuity at high chemical concentrations implying a concentration transition at which cytotoxic responses begin to dominate. Bisphenol A and raloxifene hydrochloride were used to calibrate the system to a true estrogen agonist and antagonist, respectively, and other compounds were used to introduce assay interference to a greater or lesser extent. A combined metric of specificity and potency called the “Gene Score”, defined operationally as \(-\log(AC50) + Z\)-score\) (where AC50 is the concentration of a half-maximal response), flagged chemicals for further study based on a value >7 across a range of 4–10 (~1% of 1800 tested chemicals were flagged as potential endocrine disruptors for further study).

**Complex (Cell-Based) and Virtual (Computer-Based) Models of Human Tissues and Organs**

Newer tools and methods for analyzing toxic responses with human stem cell-derived differentiated (or differentiating) lineages and/or engineered human tissues have led to novel dynamic systems that incorporate multiple cell types and/or complex architectures/geometries. Importantly, mathematical models are being developed to predict and/or account for the dynamical behavior(s) of these systems. Although these systems are not yet “ready for prime-time,” this is undoubtedly one of the most exciting emerging research areas in 21st century toxicology.

Induced pluripotent stem cell (iPSC) technology (Hou et al., 2013; Rabinovich and Weissman, 2013) is an exciting emerging technology for several reasons. The iPSCs can be used to derive genetically well-defined differentiated human cells, and they can be manipulated genetically in vitro. They can be derived from patients with inherited disease syndromes to evaluate differential sensitivity to potential therapeutics or environmental agents and for in vitro assays emulating tissue-specific functions (i.e., cardiomyocyte contraction, airway expansion, and intestinal peristalsis). Several recent examples of iPSC models for cardiac safety have been published (Guo et al., 2011; Moretti et al., 2010); however, the challenges in applying iPSC technology to toxicology are not trivial: robust manufacturing methods are needed, media and reagents must be rigorously and precisely defined, and quality control standards must be established for iPSC-derived cell populations. A major challenge facing iPSC technology at present is the relatively immature phenotype attained by differentiated cells (Jonsson et al., 2012; Kang et al., 2012), a problem evident with (but not unique to) cardiomyocytes and hepatocytes. Until this critical issue is resolved, there will be limitations on the broad use and application of iPSC-based platforms. As the US FDA has proposed future drug screening based on in vitro studies with iPSC-cardiomyocytes, working toward full implementation of a revised testing guideline by July 2016 (Chi, 2013), resolution of the phenotype issue is a key point for the near-term. Human iPSC technology is poised to have a large impact on screening drugs in the pre-clinical or clinical stages for potential neurotoxic or hepatotoxic effects. The iPSC-derived models for many inherited human diseases are also being developed at a rapid rate, and these models could potentially support development of novel therapies to treat human diseases considered to be either terminal or extremely difficult to treat.

Human iPSCs can also be a source of material for “biologically driven” assembly of rudimentary tissues under chemically defined conditions in vitro, primarily using iPSC-derived progenitors as input. The process of biologically driven assembly occurs “spontaneously” in vitro in the presence of complex, poorly defined biological substrata, such as Matrigel or fibrin. This renders it a non-trivial process to systematically analyze and characterize the chemical, biological, and physio-spatial determinants that lead to a specific outcome in such a tissue assembly system in vitro. Advanced microtiter plate screening systems that reproducibly generate complex tissues from cellular assembly utilizing synthetic substrata provide a path forward (Hou et al., 2013). For example, vascularized cortical tissue can be generated in vitro from appropriately stimulated/maintained co-cultures of endothelial progenitor cells and pericytes (Figure 2). These artificial tissues demonstrate reciprocal signaling, and functional evidence of vascularization (i.e., transport of exogenous microspheres through the vasculature via spontaneous perfusion). In a preliminary study, tissues were exposed to a screening test set (29 neurotoxicants, 16 non-toxicants), and the system demonstrated >90% reproducibility for selected outcome measures. Such complex microsystems could be used to screen for developmental toxicity and neurotoxicity, and in the future to generate for exploitation complex cellular culture models for human neurogenic endpoints or neurological diseases.

Engineered human 3D organotypic culture models (OCM’s) and human “organ-on-chip” micro-scale physiological systems have the potential to change the landscape of toxicology research. For example, one of the most advanced OCM platforms for studying liver physiology incorporates hepatocytes, hepatic stellate cells, liver sinusoidal epithelial cells, and Kupffer cells—with or without a defined structure or extracellular matrix components or controlled directional flow of media. Early results emulate expected liver metabolic processing of toxicants and demonstrate known toxic responses over extended periods in culture. For example, phenobarbital-induced expression of CYP2b2 has been observed and maintained for as long as 4 weeks, and systems with directional flow are up to an order of magnitude more sensitive to exogenous toxicants (e.g., phenobarbital and rifampicin) than comparable static culture systems (Dash et al., 2013). Flow systems may be important to adequately reproduce liver physiology in a synthetic culture model and must be appropriately configured, because the perisinusoidal space (or Space of Disse) dampens shear forces at the hepatocyte level. Systems that model these complex heterotypic cellular interactions in the human liver can someday be used to better understand hepatic metabolism and hepatotoxicity in humans, bringing together pharmacodynamics with pharmacokinetics (PK) in models structured from the native cellular hierarchy: hepatocytes > Kupffer cells > stellate cells > endothelial cells. Advances in understanding the role of the adaptive immune system in idiosyncratic drug-induced liver injury (DILI) will be of high importance in this area and are consequently supported by the development of data and cell banks such as the DILI network (DILIN). Combining micropatterned OCMs with flow through dynamics
of immune system activation may unravel the mystery of the idiosyncratic DILI response.

In addition to liver models that are micro-patterned or flow through, OCMs are being developed for kidney, female reproductive tract, heart, lung, gastrointestinal tract, bone, and many other tissues and organ systems. A promising feature of these models is the synthetic reconstruction of stem cell-derived tissues into flexible, physiological constructs that recapitulate spatial dynamics in a 3D system. Additional linkage of so-called miniorganoids and bioreactors into functional human physiological systems makes the technology attractive for studying toxicity of metabolites and organ-level interactions. This "human organ-on-a-chip" technology is thoroughly reviewed in a series of papers in Stem Cell Research and Therapy (Sutherland et al., 2013).

Virtual tissue models (VTMs) are spatially dynamic computer models that capture complex tissue- or organ-level cellular interactions in silico such as amplification, compensation, movement, or rearrangement of cells, features that are disrupted a priori in cell-free or single cell-based systems. As such interactions should not be neglected outright in toxicity, even if they are difficult to measure, VTM are a useful tool for predicting toxicological impact at the tissue and organ level, when only molecular and cellular in vitro data are available. One new approach is the decision to use modular open source shareable code (CompuCell3D) and to make the code freely available on the internet (Swat et al., 2012). The code has so far been applied to heuristic tissue simulations such as vasculogenesis, age-related macular degeneration, eye development, radiation treatment of tumors, gastrulation, limb-bud development, liver toxicology, polycystic kidney disease, segmentation, and vascular-dependent tumor growth. There are caveats with the use of VTM. The method of inquiry is based on empirically derived inter-relationships of objects and events as they are understood and hence must be validated experimentally. As they are built from existing knowledge they can miss (or omit) key mechanisms and can only show sufficiency, not necessity (although VTM simulations may be useful to inform new experiments to fill critical gaps in modeling the system under investigation). Although this is only one type of mathematical modeling of biology being done in toxicology, it is an advanced example that shows the promise of the technology for mechanistic prediction. The ability to use mathematical models for hypothesis generation now, hazard prediction, and ultimately for risk estimation down the road has broad appeal.

High Content in vivo Models

In addition to in vitro and computational models, there are new in vivo models that may have the ability to identify complex toxicities, and model human variability in response. Genetically diverse mouse models, whose use could simulate the genetic heterogeneity of human populations, are an emerging area of...
interest. The Collaborative Cross (CC) and the Diversity Outbred (DO) Mouse Models are complementary resources that capture 90% of the heterogeneity in laboratory mice, and provide “controlled genetic complexity” distributed randomly throughout the entire mouse genome at a higher level than is thought to be present in the human genome (Collaborative Cross Consortium, 2012). The haplotype blocks in DO mice are estimated to be on average 20 Mb in size, allowing for relatively straightforward mapping of quantitative trait loci (QTL), when appropriate screens are performed on several CC strains. A recent study was conducted at NIEHS, identified and mapped QTLs for benzene resistance and susceptibility in DO mice (French et al., forthcoming). In future studies, the complete genome sequences of a large number of CC mouse strains will be determined and made publicly available. The resource will be enhanced further, as pluripotent embryonic cell lines corresponding to individual DO mouse strains are developed and curated for shared use among researchers in many disciplines, including toxicologists. Application of these cell lines for HTS is anticipated.

The zebrafish model provides an integrated biological system in which chemicals can rapidly be screened for complex neuro-modulatory effects on multiple behavioral endpoints. The system is capable of screening 2000 compounds × 20 000 zebrafish per day. The simplest assay described is a quantitative and highly sensitive measure of neuronal cell death, based on binding of fluorescence-labeled Annexin V to phosphatidyl serine in dead versus living neurons. In more complex assays, fish are exposed to a series of stimuli (i.e., light, sound, and electro-stimulation) in patterns that produce complex responses, such as fear, aggression, movement, habituation/sensitization and associative or non-associative learning. Peterson and colleagues developed methods to quantify hundreds of features of the responses to these stimuli, which collectively generate a “behavioral barcode” for a specific neuroactive compound (or groups of related compounds that have similar effects) (Kokel et al., 2012). This approach, when combined with polypharmacology models, is suitable for identifying neuroactive compounds, investigating the mechanism of action of neuroactive compounds, and for identifying compounds that suppress an aberrant behavioral pattern in a zebrafish mutant.

**INTEGRATING DIVERSE TYPES OF DATA: MOVING TOWARDS AN AOP FRAMEWORK**

Realization of the NRC vision is dependent on developing a comprehensive understanding of the pathways and networks underlying toxicity, such that in vitro studies and modeling can provide a prediction of in vivo toxicity. The framework being used to develop this pathway and network understanding is the AOP (Ankley et al., 2010). Examples of AOPs are now being published both in the ecological toxicology and human health areas, and much research is being put into this concept. (Vinken, 2013; Volz et al., 2011; Watanabe et al., 2011). For example, inhibition of aromatase activity in vitro is an example of a framework for developing and using AOPs in predictive toxicology, describing the process by which toxicants that inhibit aromatase in vitro are rationally linked to adverse impact on fecundity and health of populations of fish and other oviparous species (Kramer et al., 2011; Villeneuve et al., 2009). Development of an AOP is dependent on identification of a molecular initiating event (MIE), in this case the inhibition of aromatase by a compound. Aromatase inhibition in fish models leads to reduced estrogen levels, and in turn reduced vitellogenin, and impaired oocyte development in fish. The adverse outcome is decreased ovulation and spawning, resulting in a decline in the population (Figure 3). Molecular and biochemical experiments generate data that collectively contribute to the “weight of evidence” supporting the proposed AOP. Over time, the essentiality of key events, and the quality, biological plausibility, and consistency of the linkages between key events in an AOP must be examined and validated. Once an AOP has been established, the MIE can be used to develop screening assays for compounds that might affect the AOP and networks of interacting AOPs.

There are other areas where AOP research is yielding benefits or has the potential to be integral to success. Developmental toxicity is one, as there are dozens of known developmental pathways and processes (the 2007 NRC report suggests a minimum of 17). Toxicants may act on multiple pathways in the developing embryo converging onto final common pathways in teratogenesis, or toxicants may act on one MIE that cascades into multiple pathways to different outcomes that are dependent on the stages of gestation during which exposure occurred. Assessing developmental toxicity in the face of such complexity is traditionally dependent on whole animal testing studies, but may benefit from AOPs that address highly relevant developmental processes and toxicities. Embryonic vascular development, for example, has been considered as a general model for an AOP network linked to multiple MIEs affected by ToxCast chemicals and HTS data (Knudsen and Kleinstreuer, 2011). Cancer research could also benefit from a well-established AOP network/framework built on well-characterized MOAs. A comprehensive international effort is needed to coordinate the elucidation and development of cancer AOPs and the relationships to chemistry/SAR of carcinogens.

Three seminal publications (i.e., Ankley et al., 2010; Meek et al., 2014; NRC, 2007b) with a similar conceptual basis have paved the way for understanding how to use mechanistic information and in vitro/in silico tools in toxicological risk assessment. Building on these ideas, the AOP Development Program at the Organisation for Economic Co-operation and Development (OECD) has embarked on building an AOP Knowledge Base (www.aopwiki.org), which is destined to become an important resource for the toxicology community. Progress will be accelerated by a concerted effort at the outset toward effective problem formulation, with a strong later emphasis on computational approaches and modeling of systems described by AOPs and their networks. As in vitro/in silico toxicological methods are generating vast amounts of data, a future challenge for AOP-based approaches is to map “big-data” can be used to produce accurate mechanistic knowledge in a user-friendly manner.

The recent OECD guidance document states that, by definition, an AOP has a single MIE and a single adverse outcome, which are represented as the first and last nodes in the pathway. One or more key events are represented as intermediate nodes in the pathway, and the relationships between the key events are represented as edges. An adverse outcome can be the terminal node in multiple AOPs, and a MIE (and key events) can link to multiple AOPs, where each set of linkages is a fragment of a larger AOP network. This simplification of biological complexity is designed to achieve and manage AOP elucidation. Other challenges to using an AOP framework include: modeling responses to mixtures, modeling multiple or mixed modes of action; probabilistic or quantitative prediction of outcomes; incorporating absorption, distribution, metabolism, excretion (ADME), and compensatory responses; and the fact that refining
AOPs and the AOP network is a resource-intensive process. Nevertheless, the level of momentum toward developing and refining AOPs is relatively high; therefore, it would be advantageous to engage now in strategic prioritization, so that current efforts are directed toward areas where the need and potential benefits are high.

Beyond formal AOP elucidation, a major area of research in toxicology is the use of systems biology modeling to construct pathways and networks of toxicity (Sturla et al., 2014). In order for this to be effective the experimental support for the modeling of such pathways should be quantitative (although relatively few examples of fully elucidated pathways currently exist). Consider, for example, the challenges in modeling the distinct dynamics/kinetics of the response(s) to influenza infection in human cells, mouse, or macaque cells and their interactions to higher-order biological processes in the different animal model systems (McDermott et al., 2011). Transcriptome analysis revealed conserved pathways in the post-infection responses in human and non-human cells, rendering plausibility to a transcriptional regulatory model based on human cell culture data that could potentially be used to make predictions about specific biological responses in whole animals. On the other hand, effective analytical tools to make such predictions are not yet in hand, and automated analytical tools may not be up to the task, given the inherent complexity of the data sets and the potential for non-linear and/or threshold effects. Pattern and ontology-based approaches do not account for regulatory mechanisms or cross-talk between pathways and new approaches will be needed to quantitatively integrate ‘omics data into biological pathway models.

Of course not all pathway perturbations will be adverse. Adaptive responses may follow a particular MIE and may be difficult to associate with a point of departure in a pathway leading to an adverse outcome. Case studies of prototypical, data-rich compounds are considered appropriate as a means to improve understanding AOPs in risk assessment, an example being dimethylarsenic acid (Keller et al., 2012). A major bottleneck to using an AOP-based framework for science-based decision making is the lack of detailed dosimetry and exposure data on a time and dose continuum. Dose transitions remain poorly understood and biological context (cell, tissue, or organism) should be considered. In vitro studies model only a discrete set of steps in a complex cascade of events that lead to a phenotypic expression of toxicity. An extensive effort to characterize AOPs is needed before they can be used to build a new human risk assessment paradigm.

The difficulty of making correlations between in vitro and in vivo activity may be a result of the narrow chemical space of the training sets, the limits of the assay biology, or other factors. For example, computational models can be constructed to fit the HTS data from ToxCast or other screening programs, but this does not yet guarantee that the models will be able to predict outcomes for chemicals outside the dataset. Useful predictive power for in vitro assay data may be present, depending on details of how available chemicals, endpoints, and tests were selected.

**IMPACT ON RISK ASSESSMENT**

Those outside the regulatory environment may perceive reluctance to use new approaches, new types of data, and new technologies for human risk assessment. Most of these concerns are based on well-recognized limitations of in vitro systems such as, for example, limited metabolic capacity; inability to account for ADME and pharmacokinetics; uncertainty in cell-cell network interactions; lifestage considerations; reductionism in the biological complexity of systems being tested; and biological diversity in general (Chiu et al., 2013; Kavlock et al., 2012; Sturla et al., 2014). In order to build confidence in new methods their value needs to be proven, and this is best accomplished by applying them in a “fit-for-purpose” manner to address areas of uncertainty that are difficult to address using conventional toxicology methods. For example, the assessment of mixtures, read-across based on mechanism of action, and the role of human genetic variability are all areas for which new methods offer unique contributions toward reducing uncertainty. Acceptance of test methods by regulatory authorities is difficult to achieve and usually occurs gradually. Regulatory authorities such as FDA and EPA have legal responsibilities to protect public health and are reluctant to accept methodologies that have not been...
demonstrated to be consistent with this obligation; however, when sufficient information exists to demonstrate the reliability of new methods to protect public health, authorities have shown a willingness to adopt these for regulatory decision-making. The key is not necessarily formal validation of new methods, but demonstration that alternative methods are sufficiently robust (can be performed in various laboratories with reasonably consistent results), reliable (results obtained can be used to make regulatory decisions), and scientifically sound (the methodology is based on well-established understanding of the effect being assessed).

In the absence of detailed information on AOPs or mechanisms of toxicity there are still approaches available to classify chemicals based on known activities. These approaches use molecular toxicology data such as toxicogenomic information and physical/structural/chemical properties to predict adverse effects in vivo (Chiu et al., 2013). Initial attempts to determine benchmark concentrations on the basis of transcriptomics data have already been undertaken in vivo (Thomas et al., 2013). Their application to in vitro test systems can yield key information about principles of “point of departure” for low-dose effects during in vivo toxicity response systems. The TG-GATES (Toxicogenomics Genomics-Assisted Toxicity Evaluation System; http://toxico.nibio.go.jp/english/index; accessed July 7, 2014) demonstrates that gene expression signatures form self-organizing maps that can be used to cluster chemicals according to their biological impact, their structure, and other properties. Using a “chemical-biological” annotation tool, it is possible to use TG-GATES (or analogous) molecular toxicology databases to develop networks of chemicals and chemical pathways, in the same manner as one deduces networks of genes in a biological system. Such a chemical network could be useful for postulating chemical MOAs in a biological system. One caveat is that chemicals interact differently with biological systems in vitro and in vivo; therefore, pattern recognition in a molecular toxicology dataset (if based solely on in vitro data) may report (and/or predict) outcomes different than what would be observed in vivo. Recognizing this fact, it is advisable to consider and include a “degree of confidence” metric for predictions derived solely or primarily within molecular toxicology pattern space.

Risk Assessment by analogy is an approach where similar toxicological outcomes are predicted for relatively poorly characterized chemicals, based on molecular properties and/or a mode of action similar or identical to one or more well-characterized toxicants. Risk assessment by analogy is not at all akin to traditional computational commercially available QSAR; instead, it requires significant investment of time, resources, and human expertise to identify chemicals that are true analogs of one another. As a proof of principle, a decision tree was constructed to “predictive toxicology by analogy.”

Gene expression data can also be used to generate a connectivity map for toxicants, in a similar manner as described by Lamb and colleagues for pharmaceuticals (Lamb, 2007; Peck et al., 2006), and such data can be used to confirm predicted toxicity of chemical analogs. Non-specific effects at high toxicant concentration and pleiotropic effects (where a single toxicant impacts more than one biological pathway) are problem areas that must be recognized as such, when present.

Quantitative in vitro to in vivo extrapolation (QIVIVE) and reverse dosimetry can be described as the critical “end game” in the work flow of predictive toxicology. QIVIVE is essential in order to transition away from animal model-based toxicology to entirely in vitro/in silico-based toxicological science. It is not likely that in the near future a priori prediction of the rate at which a potential toxicant is metabolized in vivo will be possible; however, in a preliminary study of 12 ToxCast™ chemicals for which human in vivo PK data are available, 8 of 12 QIVIVE predictions of hepatocellular clearance resulted in a satisfactory estimate of in vivo steady-state concentration, if non-restrictive clearance was assumed (Wetmore et al., 2012). On the other hand, if in vivo clearance is neglected, approximately 2 orders of magnitude variation in biological potency could fail to be captured when predicted directly using HTS in vitro measurements of AC50. In another preliminary study, the OECD QSAR toolbox was used to predict metabolites of 12 chemicals. Although a toxicologically relevant metabolite was identified for 9 of 12 chemicals, other low-yield or non-toxic metabolites were also predicted, so that the usefulness of the predictions were questionable. New tools are being developed to improve QIVIVE. For example, a human enterocyte cell line that expresses the intestinal form of carboxylesterase toxicant metabolism is being modeled in the context of liver bioreactors, and flow-based multi-organ systems (i.e., human on a chip) are being developed as an extension of current liver bioreactor models. There is optimism that these complex in vitro systems, which are first steps toward mimicking in vivo human exposure to a toxicant and its metabolites.

Q SAR is often perceived to be a “black box” that generates prediction without measurement. To ensure success, a multi-step “data curation” workflow, involving careful inspection of the data to exclude duplicates, eliminate unreliable data sources, detect and verify activity cliffs, adjust dataset modelability, and identify and correct mislabeled compounds should be adopted. Dataset modelability, MODI, is a concept developed to quantify the distribution of structural similarity/dissimilarity within a dataset, such that a dataset with a higher mean similarity has higher modelability (Golbraikh et al., 2014). Initial attempts to create hybrid models, that incorporate biological and chemical descriptors, based on gene expression signatures and bioassay data for 127 drugs from the second Toxicogenomics Informatics Project of Japan (TG2P) of Japan showed that hybrid models perform less well than models built on biological descriptors alone; however, the performance of hybrid models was optimized simply by considering a compound’s similarity to its biological nearest neighbors in biological space and its similarity to its chemical nearest neighbors in chemical space (Low et al., 2011). This is referred to as chemical-biological read across. Additional models and concepts on QSAR are available at http://chembench.mml.unc.edu (accessed July 7, 2014). SEURAT-1 (Safety Evaluation Ultimately Replacing Animal Testing, http://www.seurat-1.eu/; accessed July 7, 2014), a large EU-based 5-year program launched January 2011, is designed to demonstrate through a “proof-of-concept” cases studies that the modern toxicological toolbox is up to the challenge. Based on the output of SEURAT and similar programs, it is anticipated that in the short/medium term it will be possible to form robust chemical categories and carry out read-across toxicological predictions using an integration of both chemical structure and MOA information. If successful, SEURAT-1 and similar programs could point the way towards practical toxicological risk management using an in vitro / in silico approach.
Three major opportunities exist for improving the current human risk assessment paradigm: (1) derivation of probabilistically based human-specific toxicity estimates to replace deterministic estimates based on rodent models; (2) genetically-diverse experimental animal systems to assess phenotypic variation for adverse outcomes; and (3) integrated quantitative analysis of human variability and susceptibility. Opportunities exist to incorporate modeling and more quantitative estimates of data values and variability into all areas of risk assessment, from in vitro to PK to animal data to human data (Zeise et al., 2013). Ultimately, probabilistic toxicology will be able to break down artificial dichotomies, moving the scientific community toward a broader conception of population health, where toxicological responses are modeled using a continuum from ‘no effect’ to ‘effect’ (i.e., non-toxic to toxic), based on probabilistic measures, such as the chemical potency confidence interval and the benchmark dose confidence interval in reference to a genetically diverse population with a characteristic distribution of susceptible subpopulations.

CONCLUSIONS AND FUTURE DIRECTIONS

FutureTox II identified a number of advances and challenges in moving toward an in vitro/in silico based human risk assessment paradigm. Although these challenges were recognized in the 2007 report from the NRC (2007a) and FutureTox I (Rowlands et al., 2104), FutureTox II built on recent research experience, identifying both the successes and the limitations of current models and methods of predictive toxicology as the 21st century vision is realized. The growing abundance of in vitro data are now addressing the dearth of mechanistic toxicity data for large chemical inventories and new in silico models are addressing potential solutions to specific challenges in predictive toxicology.

Despite the demonstrated progress and new directions, achieving the vision of the NRC 2007 report has a long way to go. Knowledge gaps in biological modeling, cell biology, systems biology, and the genetic basis of human susceptibility must be closed. Limitations of technologies such as iPSC culture, OCMs and imaging will have to be overcome. Each of these advances will need to be developed in a “fit-for-purpose” manner, to be used in the way that is most advantageous to the scientific and regulatory community. Some tools will be most useful for screening and determining which chemicals need more in-depth analysis, while other tools will play a role in estimating human risk. Given the nature and magnitude of the challenges, it is clear that while stem cells and OCMs are relevant and can accelerate progress toward an in vitro/in silico paradigm, animal studies are expected to remain a critical part of the toxicologist’s toolbox for the foreseeable future. Four topical breakout-group sessions considered future directions and research needs, summarized as follows.

With respect to hepatotoxicity, FutureTox II highlighted two priority research areas for in vitro data generation and in silico model development: prediction of dose-dependent toxicities in animal and humans, and predicting idiosyncratic DILI for which there are no good animal models currently. Continued improvements in maintenance of differentiated phenotype of hepatocytes and combination with other liver cell types and stem-cell based lineages are needed to advance the predictivity of current in vitro liver models. Flow systems may be important to reproduce liver physiology in various genetic backgrounds, as well as defined culture conditions to calibrate with ‘reference’ compounds for hepatotoxicity. Introduction of inflammation, fatty liver disease and viral hepatitis in silico models may better deliver useful information for predictivity of idiosyncratic DILI. Although instances where environmental chemicals cause dose-dependent or idiosyncratic liver injury in humans are rare outside an industrial setting, the priorities are similar due to the central role of hepatic metabolism in chemical-induced liver injury.

With respect to cancer prediction, FutureTox II highlighted the need for accelerative assembly of key events into animal models (short-term) and computational models (long-term) that integrate in vitro data with advanced literature mining techniques for AOP-based approaches to identify potential carcinogens. Elucidating assays and cellular-complex culture models that reflect key events in cancer AOPs and account for genetic, dietary, inflammatory heterogeneity in human populations is a research priority, building onto a MOA framework to help define the thresholds that drive key events in cancer AOPs. Computational models to help predict human cancers will need to be calibrated with rodent models to deal with the complexity of different cell types, the potentially large number of MIEs, and the role of ADME and kinetics for a more quantitative integration of pathway/AOPs into tiered testing strategies. Such strategies will use in vitro assays, short-term animal studies, human exposure models, and IVIVE to provide relevant information about dose-response profiles of biological targets of chemicals and their toxicological relevance in light of human exposure predictions.

With respect to developmental/reproductive toxicity, FutureTox II highlighted the need for advanced models to integrate in vitro data from models of morphogenesis and differentiation such as zebrafish embryos and stem-cell systems. Developmental toxicity is complex, with multiple endpoints being both possible and relevant. Perturbation to developmental pathways from exposure at earlier life-stages can broadly impact the health and well-being of an individual throughout their lifecourse. A major challenge for translating the in vitro data and in silico models is to determine what specific pathways, and threshold of effect at different stages of development, would give rise to an adverse phenotype and to pinpoint the lifestage processes at which an adverse outcome would likely manifest itself. Predicting toxicity to the developing organism would include considerations of male/female fertility, maternal-fetal interactions, adverse pregnancy outcomes, and children’s health through puberty and beyond. Understanding of mechanisms will be essential for successful prediction and requires relevant whole animal studies and future OCMs.

With respect to regulatory toxicology, FutureTox II highlighted the need for QSAR and targeted in vitro assays to generate reliable information for use in regulatory decisions. Assay selection and validation that is “fit-for-purpose” must be considered as adequate to the task for QSAR models. For example, validation of ToxCast/Tox21 results is needed to bolster confidence in taking the new testing paradigm to a decision-making level that will reduce uncertainties in risk assessment for specific endpoints. The question is what data are appropriate for what purpose in assessing environmental chemicals (prioritization and predicting the lowest observable effect level [LOAEL] or no observable adverse effect level [NOAEL]) and pharmaceuticals or medical devices? These tools, approaches, and technologies also have the potential to solve issues around modeling of mixtures, chemical metabolism, subpopulation susceptibility and integration of diverse data types. How they are used to solve human and environmental problems will also depend on
the ability of the toxicology community to continue innovative research, and to effectively communicate among themselves as well as to the regulatory community and the public. As part of this important consideration, a third conference in the FutureTox series, entitled “FutureTox III: Transforming 21st-Century Science into Risk Assessment and Regulatory Decision-Making”, has been formally approved by the Society of Toxicology and is being held in Arlington, Virginia on November 19–20, 2015.

FUNDING

U.S. Environmental Protection Agency grant 83559201.

Workshop Sponsors included the Society of Toxicology, American Academy of Clinical Toxicology, American Chemistry Council, American College of Toxicology, Consumer Specialty Products Association, Elsevier, U.S. Environmental Protection Agency, U.S. Food and Drug Administration, Grocery Manufacturers Association, The Hamner Institutes for Health Sciences, ILSI Health and Environmental Sciences Institute, Office of Environmental Health Hazard Assessment/CalEPA, Teratology Society, Society of Toxicologic Pathology, and the UNC Gillings School of Global Public Health.

ACKNOWLEDGMENTS

The FutureTox II Organizing Committee included the authors of this article and Ivan Rusyn (SOT Scientific Liaison Committee). Marcia Lawson and Clarissa Wilson participated in the Workshop as SOT Staff. Breakout group leaders were: Paul Watkins, Doug Keller (Liver Disease and Hepatotoxicity); Chris Corton, Maurice Whelan, Charles Wood, and Margaret Pratt (Cancer); Ed Carney and Tom Knudsen (Developmental/Reproductive Toxicity); Ken Hastings and Kevin Cross (Regulatory Toxicology).

REFERENCES


