OXFORD

SOT | Society of Toxicology
www.toxsci.oxfordjournals.org

TOXICOLOGICAL SCIENCES, 155(2), 2017, 326–336

doi: 10.1093/toxsci/kfw207
Advance Access Publication Date: December 19, 2016
Forum

FORUM

How Adverse Outcome Pathways Can Aid the Development and Use of Computational Prediction Models for Regulatory Toxicology

Clemens Wittwehr,[a,1] Hristo Aladjov,[b] Gerald Ankley,[c] Hugh J. Byrne,[d] Joop de Knecht,[e] Elmar Heinzle,[f] Günter Klambauer,[g] Brigitte Landesmann,[a] Mirjam Luijten,[e] Cameron MacKay,[h] Gavin Maxwell,[h] M. E. (Bette) Meek,[i] Alicia Paini,[a] Edward Perkins,[j] Tomasz Sobanski,[k] Dan Villeneuve,[c] Katrina M. Waters,[l] and Maurice Whelan[a]

[a]European Commission, Joint Research Centre, Ispra 21027, Italy; [b]Bulgarian Academy of Sciences, Sofia 1113, Bulgaria; [c]US Environmental Protection Agency, Duluth, Minnesota 55804; [d]FOCAS Research Institute, Dublin 8, Ireland; [e]National Institute for Public Health and the Environment (RIVM), Bilthoven, MA 3721, The Netherlands; [f]Universität des Saarlandes, 66123 Saarbrücken, Germany; [g]Johannes Kepler Universität, Linz 4040, Austria; [h]Unilever Safety and Environmenta Assurance Centre, Sharnbrook, MK44 1LQ, UK; [i]University of Ottawa, Ontario K1N 6N5, Canada, [j]US Army Engineer Research and Development Center Vicksburg, Mississippi 39180; [k]European Chemicals Agency, ECHA, 00121 Helsinki, Finland; and [l]Pacific Northwest National Laboratory, Richland, Washington 99352

[1]To whom correspondence should be addressed at European Commission, Joint Research Centre, Via E. Fermi 2749 – TP 126, Ispra 21027, Italy. E-mail: clemens.wittwehr@ec.europa.eu.

ABSTRACT

Efforts are underway to transform regulatory toxicology and chemical safety assessment from a largely empirical science based on direct observation of apical toxicity outcomes in whole organism toxicity tests to a predictive one in which outcomes and risk are inferred from accumulated mechanistic understanding. The adverse outcome pathway (AOP) framework provides a systematic approach for organizing knowledge that may support such inference. Likewise, computational models of biological systems at various scales provide another means and platform to integrate current biological understanding to facilitate inference and extrapolation. We argue that the systematic organization of knowledge into AOP frameworks can inform and help direct the design and development of computational prediction models that can further enhance the utility of mechanistic and *in silico* data for chemical safety assessment. This concept was explored as part of a workshop on *AOP-Informed Predictive Modeling Approaches for Regulatory Toxicology* held September 24–25, 2015. Examples of AOP-informed model development and its application to the assessment of chemicals for skin sensitization and multiple modes of endocrine disruption are provided. The role of problem formulation, not only as a critical phase of risk assessment, but also as guide for both AOP and complementary model development is described. Finally, a proposal for actively engaging the modeling community in AOP-informed computational model development is made. The contents

serve as a vision for how AOPs can be leveraged to facilitate development of computational prediction models needed to support the next generation of chemical safety assessment.

Key words: Adverse Outcome Pathways; AOP; quantitative AOP; computational prediction model.

In recent years, several conceptual papers advocating the use of innovative mechanistically based approaches for assessing the potential hazards and risks of chemicals for both human health and the environment have been published (Garcia-Reyero and Perkins, 2011; Krewski *et al.*, 2010). Ideally, these new approaches would involve combining high-throughput and high-content data from *in vitro* assays or small organisms with computational modeling to predict toxicological effects of concern. Many computational methods have already demonstrated their usefulness in toxicology. For example, machine learning and pattern recognition methods have been used for identification of quantitative structure–activity relationships (QSARs) and genomic biomarkers (Bercu *et al.*, 2010; Buick *et al.*, 2015; Garcia-Serna *et al.*, 2015; Guyton *et al.*, 2009; Maltarollo *et al.*, 2015; Neagu *et al.*, 2007; Thomas *et al.*, 2013). However, in order to be considered credible for use in a regulatory decision-making context, predictive models require a sound mechanistic foundation built from knowledge of toxicological processes.

Adverse outcome pathways (AOPs) are intended to outline and capture existing knowledge concerning the biologically plausible and empirically supported foundations for predicting apical toxicity from mechanistic data (OECD, 2013). In recent years, activities in the international community, coordinated through the OECD, have focused on providing a single, harmonized, point of access to summary descriptions of AOPs, including associated weight of evidence assemblies and evaluations via an AOP knowledge base (AOP-KB; https://aopkb.org; last accessed October 24, 2016). The AOP-KB and content therein is organized in a systematic, searchable, and transparent manner according to an established set of guidelines and principles (OECD, 2016; Villeneuve *et al.*, 2014) that facilitates evaluation of suitability for various regulatory applications. In particular, the AOP-Wiki module of the AOP-KB, which has been publically accessible since September 2014, is a valuable source of documented pathway information that incorporates assembly and evaluation of weight of evidence supporting causal relationships between various key events in the pathway, including quantitative understanding where it is available. As a consequence of this organizational structure, the growing knowledge base of AOP descriptions is well suited to aid the development of computational prediction models that can help bring about a new era in regulatory toxicology, using AOPs for developing computational prediction models.

Recognizing the potential value of integrating the AOP framework with development of computational prediction models, the European Commission's Joint Research Centre organized a workshop on the topic of "AOP-informed predictive modeling approaches for regulatory toxicology". September 24–25, 2015, 28 invited international experts in computational modeling, toxicology, and risk assessment from academia, government, and the private sector came together to explore the scientific opportunities and stimulate greater communication and collaboration between the AOP development and computational modeling communities. Modelers were introduced to the principles and practice of AOP description and invited to think about how this systematic organization of knowledge could aid model development. A number of case examples exploring the different ways models and AOPs could be integrated to address regulatory challenges were discussed. Issues related to regulatory uptake and application were considered. Finally, the group discussed ways to further engage the modeling community in the endeavor of AOP-informed predictive toxicology. The following highlights major points of discussion and conclusions from the workshop and sets the stage for further dialog regarding the role of computational modeling and AOPs in the emerging paradigm for regulatory toxicology.

HOW CAN AOPs INFORM COMPUTATIONAL MODEL DEVELOPMENT?

Adverse outcome pathways are a conceptual framework that portrays existing knowledge concerning the linkage between some molecular initiating event (MIE) and an adverse outcome (AO) that occurs at a level of biological organization considered relevant to regulatory decision-making (Ankley *et al.*, 2010). Individual AOPs are represented as sequences of measureable key event (KE) nodes that reflect a causal progression from an initial perturbation of normal biology, caused through direct interaction with a chemical, to a series of system failures at higher levels of biological organization (Figure 1). Key events are linked via Key Event Relationships (KERs) that define both the structural and functional relationship between a given pair of KEs and compile specific empirical evidence that supports the idea that if the upstream KE is altered to a sufficient degree, predictable changes (qualitative or quantitative) can be expected in the downstream event in the sequence. AOPs are described using modular assemblies of KE and KER descriptions. These modular descriptions, properly structured and connected in the AOP-KB, provide the foundation for construction and analysis of AOP networks (Figure 1) that can provide a more comprehensive, integrated, and biologically realistic synthesis of available knowledge concerning the ways chemicals can adversely impact organisms. Overall, AOPs and AOP networks provide structure for our knowledge of how an MIE (or MIEs) can lead to deviations from normal healthy function of a biological system (ie, adverse outcome[s]). From a modeling perspective, structuring of knowledge is extremely informative for model design and development. In particular, AOPs can help reduce an initially overwhelmingly complex biology to the essentials necessary for a predictive model, avoiding model overload. First, the summary of an AOP represented in the form of a box and arrow diagram that identifies the KEs and KERs (Figure 1) provides an overall conceptual model that bounds the modeling challenge within a specific biological domain. It defines, for example, the key chemical–biological interaction that triggers a toxicologically relevant biological perturbation by identifying the MIE. This can immediately inform the development of quantitative structure–activity relationship (QSAR) models and chemical categories useful for defining the chemical space for which the AOP is likely to have relevance. It then identifies the key biological pathways, functions and compartments (ie, cell types, tissues, organs) in which the biology to be modeled operates. Because each KE is defined at a particular level of biological organization, the AOP also provides a road map of the biological scales at which a single model, or series of models, must operate. In this way, the AOP suggests the heuristic domain and

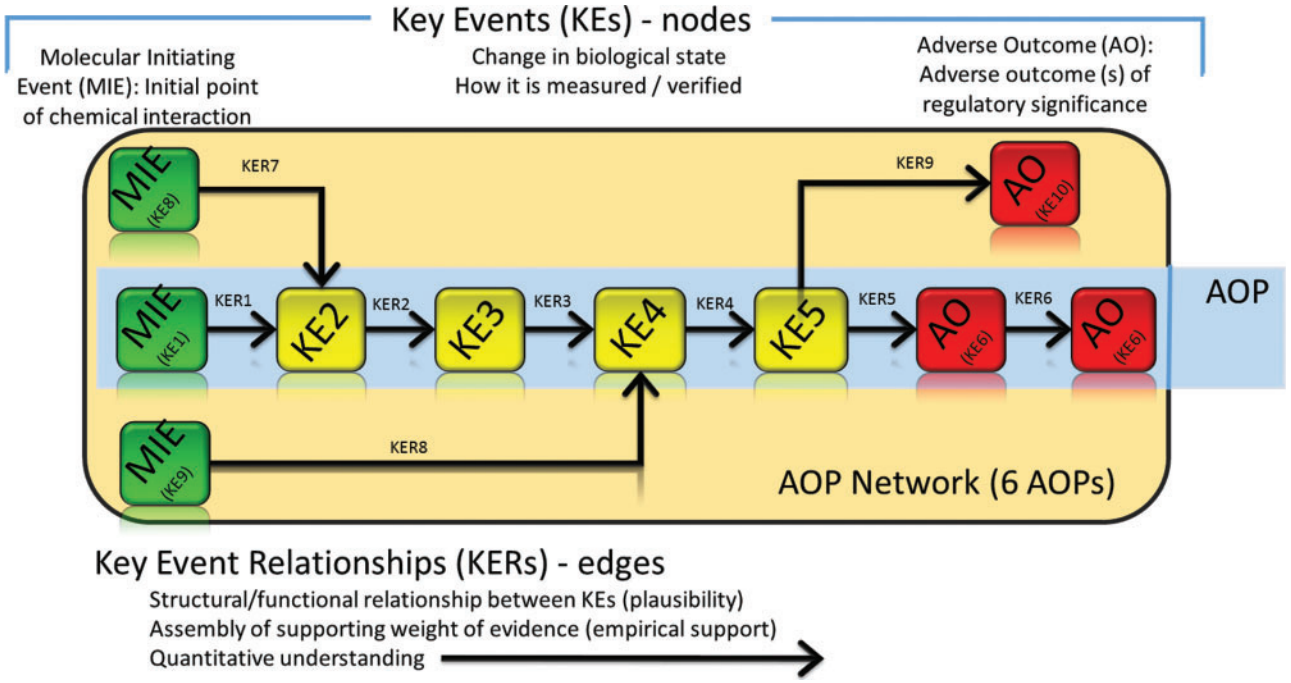

FIG. 1. Conceptual representation of the adverse outcome pathway (AOP) framework including modular representation as Key Events (KEs) and Key Event Relationships (KERs), 2 specialized types of KEs, molecular initiating events (MIEs) and adverse outcomes (AOs), that serve as upstream and downstream anchors in an AOP, and assembly of multiple AOPs sharing common KEs and/or KERs into an AOP network.

biological scope/space in which the prediction models should function. A second level of information in the AOP description that guides the development of prediction models is the description of the KEs (OECD, 2016). In many cases, those with the computational modeling expertise may not be familiar with the biology represented in the AOP. The biological description of the KEs presented in the AOP-WIKI module of the AOP-KB (https://aopwiki.org/; last accessed October 25, 2016; the first information field on each KE page) provides the entry point or gateway that can introduce the modeler to the biology encompassed by the AOP and its events. Whereas this description may not be sufficient to fully support model design and formulation, it can suggest the appropriate biological subject matter experts with which the modeler may want to consult and partner. Additionally, it may be helpful in identifying the type of modeling approaches, mathematical, formalisms, and parameters that could be employed. For example, a KE involving enzyme inhibition may require identification of the type of inhibition (eg, competitive, irreversible) and related kinetic constants. A KE involving cell proliferation or selective cell death may indicate the need to employ an agent-based modeling approach, whereas one involving an increased risk of disease may require a probabilistic approach.

Key Event descriptions associated with an AOP also contain useful information regarding how a KE is measured. This information helps define the kinds of data likely to be available, or which could be generated to inform model development. Identification of specific assays may, in some cases, provide useful information regarding data sources that the modeler(s) could utilize for model development purposes. For example, for KEs measured in ToxCast assays (Kavlock *et al.*, 2012), association of the KE with an assay identifies a database (Actor/iCSS dashboard; http://actor.epa.gov/dashboard/; last accessed October 24, 2016) of relevant data that may be useful for model development. Depending on the experimental method(s) used, additional information might be required to translate the raw output of the method to *in vivo* relevant data. Future modules of the AOP-KB (ie, Effectopedia) aim to provide standardized summaries of the data itself along with meta-information describing the test methods and transformation functions need to put those data into appropriate in vivo context. The identification of specific approaches used to measure a given KE can suggest the types of data that may serve as inputs to the model, and parameters that may be useful to simulate from an interpretive standpoint. For example, if the AOP involves enzyme inhibition that leads to a decrease in a circulating hormone followed by a loss of function in a particular cell type, one might want to design a model that can take a standardized measure of a chemical's potency to inhibit the enzyme and predict the dose-response and time-course behavior of the circulating hormone concentration, subject to feedback regulation and other modulating factors represented in the model (see Case Example 1). Finally, identification of the methods used to measure the KEs can provide insights into the time-scales over which the variables represented as KEs can be measured. This provides information regarding the level of temporal resolution that the models should be designed to predict.

Key Event Relationship descriptions (OECD, 2016) are similarly useful. The KER description gives a summary of the weight of evidence (WoE) that establishes the causal nature of the relationship between 2 measurable biological events (Becker *et al.*, 2015). The structure of the KER immediately defines key input and output parameters relevant for model simulation. Defining the biological plausibility of the relationship between the pair of KEs highlights the important biological context and the processes that need to be captured in the relationship model. Furthermore, the empirical evidence summarized in the KER description, provides references that can provide data for model parameterization, fitting, and/or testing.

The KER descriptions also include a narrative section on the quantitative understanding of the relationship between the KEs (OECD, 2016). At present, OECD guidance provides only a very

general indication of the types of information which should be captured. Modelers attending brought new insights regarding the types of information that AOP developers could capture in order to help facilitate model development. Specifically, it was recommended that AOP developers more explicitly identify sources of data on response–response relationships (ie, input–output relationships between neighboring KEs). Where possible, this should include the general form of the relationship (eg, whether there is a linear, sigmoidal, threshold or Bayesian relationship between changes in KEs upstream (KE_{up}) and those downstream (KE_{down})). It should also indicate the range of uncertainty in the relationship when it can be at least approximated. For example, based on a measure of KE_{up}, can KE_{down} be predicted within a factor of 2, or 1000? Alternatively, if the relationship is probabilistic, how much certainty is there that KE_{down} will be significantly perturbed given a defined magnitude of change in KE_{up}? Known sources of uncertainty should be defined: For example, is there uncertainty due to inter-individual variability, measurement error (informed by the description of approaches used to measure the KE), or modulating environmental factors such as temperature, pH, photoperiod, etc. This type of information can direct how models are structured and which key parameters need to be incorporated to accurately predict response–response behaviors.

Beyond the form and uncertainty of the response–response relationships, it would be useful to describe the approximate time-scale over which a change in KE_{up} impacts KE_{down}. Are these tightly coupled, whereby KE_{down} changes in a matter of seconds following an alteration in KE_{up}, or do they operate on much different time-scales, KE_{down} lagging behind the changes in KE_{up} by minutes, hours, or days? Ideally, data on the kinetics or dynamics of the relationship between KEs would be included. However, even a general sense of the temporal scale over which the relationship operates may be highly informative for model development.

Effectopedia, another module of the AOP-KB under development, is expected to make the AOP-KB even more useful for modeling purposes. Effectopedia is being designed to provide data structure for explicit representation of response–response relationships along with their corresponding uncertainty bounds. Implementation of this module should greatly streamline access to the data describing response–response relationships. It may even allow for implementation of predictive models directly in the KB.

Overall, information captured in the descriptions of AOPs and associated KEs and KERs provide a wealth of information that can aid predictive model development. The modular organization of content in the AOP-KB allows for implicit creation of AOP networks, in which certain KEs, KERs, or sequences of KEs and KERs, serve as points of convergence for multiple independent AOPs (Figure 1). Modeling these convergence points within an AOP network may yield greater return on the resources invested. It can also suggest how a series of models could most effectively be structured. For example, Ankley *et al.* (2010) described a series of AOPs converging at a common KE of impaired production of a key egg yolk precursor, vitellogenin (VTG), in fish. Whereas multiple models may be needed to link diverse upstream MIEs (eg, aromatase inhibition [see Case Example 1], estrogen receptor antagonism, androgen receptor agonism) to predicted reductions in VTG, the downstream KEs for multiple AOPs converge such that only a single model is needed to predict impacts reduced VTG on fish reproduction. Conversely, in the case of multiple AOPs related to thyroid axis disruption (Case Example 3), genes and proteins involved in thyroid hormone synthesis and homeostasis and their regulation is well conserved among vertebrates, such that a common model that captures multiple MIEs and upstream portions of the pathways could be developed. However, the downstream consequences of reduced serum thyroid hormone concentrations (a key point of convergence in that AOP network) diverge across species, suggesting independent models may be needed to translate well-conserved effects of chemical perturbation on thyroid hormone concentrations to divergent AOs across species. Examples like this illustrate how AOP networks can be an important source of information to both prioritize the AOP elements for which computational models should be developed and facilitate efficient structuring and coupling of models that can be used to simulate individual AOPs, or form more comprehensive predictive models of network behavior. Building AOPs in such a network environment is also consistent with the intrinsically interconnected nature of biological systems and promotes understanding of important modifying factors, resulting from intersection and interactions with other AOPs, that may be needed to build effective predictive models.

Development of predictive models is best achieved through multi-disciplinary teams often composed of modelers, subject-matter experts, and experimentalists. The contact information for developers, captured in the AOP-KB, provides a useful resource for modelers developing prediction models based on AOPs by facilitating collaboration and enabling insights that may not be represented in the existing AOP descriptions. This type of interchange is particularly important given that the AOP developers may not have the background needed to identify the types of data and information that would be informative for modeling. Thus, there is a strong role for the AOP-KB to play in not only developing AOP networks, but also professional networks and multidisciplinary teams that can implement a successful iterative approach to model development.

CASE EXAMPLES OF AOP-INFORMED COMPUTATIONAL PREDICTION MODELS

The utility of AOPs for development of computational prediction models for regulatory toxicology is more than just conceptual. There are already examples of how AOP knowledge has informed the development of predictive models suitable to a variety of applications, ranging from screening and prioritization of chemical testing to quantitative risk assessment. These examples demonstrate the diversity of model types and data that can be employed. They also highlight that modeling can be focused either on an entire AOP, or portions of an AOP, as appropriate to a given assessment.

Case Example 1: Aromatase Inhibition Leading to Reproductive Impairment in Fish

A well-established AOP (https://aopwiki.org/wiki/index.php/Aop:25; last accessed October 25, 2016) describes the linkage between inhibition of ovarian aromatase (CYP19; which converts testosterone to 17β-estradiol [E2]) in female fish and AOs at the individual (embryo production) or population level (total number of fish). Intermediate KEs in the AOP capture E2-induced synthesis of vitellogenin (VTG; egg yolk protein) in the liver, and subsequent incorporation of the protein into developing oocytes in the ovary. This AOP description includes a WoE evaluation, which rated the support for the AOP as "strong" from a technical perspective (Becker *et al.*, 2015). Given interest in aromatase inhibition as a mode of endocrine disruption, this AOP was used as a case study to illustrate how computational prediction models can be aligned with and complement an AOP.

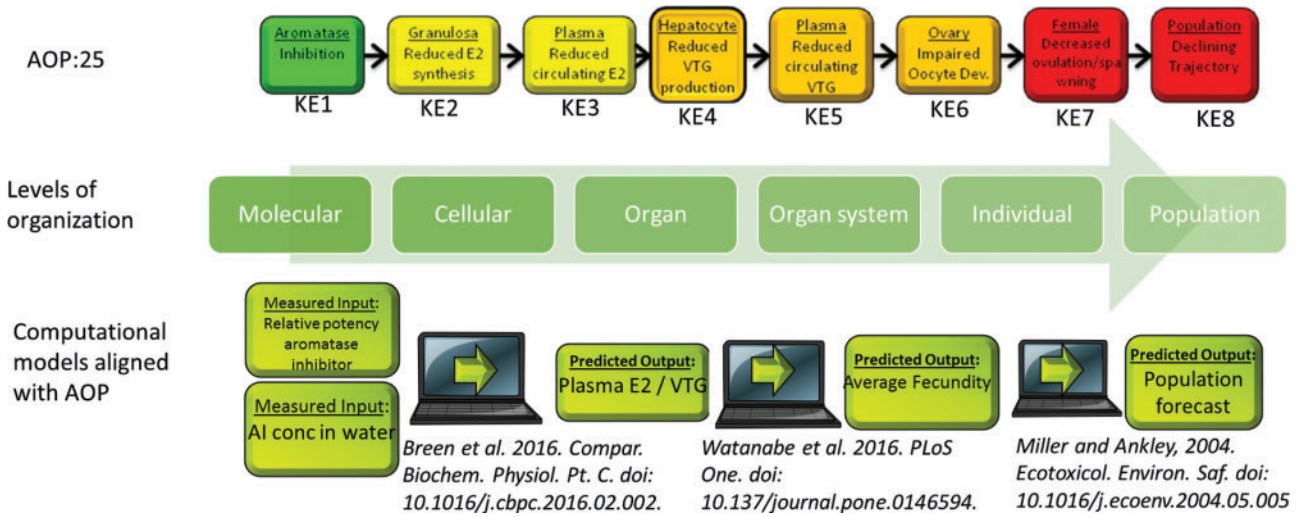

FIG. 2. Illustration of the alignment between multiple computational models and an adverse outcome pathway linking aromatase inhibition to reproductive dysfunction and declining population trajectory in fish. Model constructs allow for quantitative extrapolation across key events at multiple levels of biological organization ranging from molecular scale interactions to population-level effects.

The overall approach involved linking a series of models reflecting different portions of the AOP (Figure 2). The first portion of the AOP (first 5 KEs) is simulated by a biologically based dose–response (BBDR) model of hypothalamic–pituitary–gonadal (HPG) axis that quantitatively associates the degree of inhibition of CYP19 activity (KE1) with (reduced) plasma E2 concentration (KE3) and, ultimately, depressions in hepatic VTG production (KE4) and circulating VTG (KE5) (Breen *et al.*, 2013, 2016; Cheng *et al.*, 2016; Li *et al.*, 2011a). The HPG axis model accounts for feedback and adaptation/compensation in the system, in order to simulate complex response–response relationships among the KEs as a function of dose and time (Breen *et al.*, 2013, 2016). The next portion of the AOP (KEs 6–7) is captured by a statistical model that uses circulating VTG concentrations as an input parameter to simulate and predict oocyte growth dynamics in the context of egg production (Li *et al.*, 2011b; Watanabe *et al.*, 2016). Finally, egg production data are used as input parameters into the Leslie matrix of a density-dependent population model that can be used to forecast potential impacts on population trajectories (Miller and Ankley, 2004; Miller *et al.*, 2015).

This series of linked models can be used for quantitative risk assessment, with the initial input based upon degree of inhibition of CYP19 by a test chemical, a direct measurement of the MIE that can be made using either an *in vivo* or *in vitro* test system. This is a relevant application in the context of international endocrine disruptor screening and testing programs (OECD, 2010). With this basic model construct, it is also possible to use measures of intermediate KEs within the AOP to quantitatively estimate "downstream" responses, including the AOs. For example, Miller *et al.* (2015) recently used the approach to successfully make quantitative predictions of fecundity of the white sucker (*Catostomus commersonii*) at a site impacted by a pulp and paper mill where depressed sex steroid synthesis was measured in resident fish. From this, Miller *et al.* (2015) were able to make quantitative long-term predictions of status of the white sucker population under varying scenarios, in order to inform resource managers evaluating different risk mitigation/treatment options. Even if one feels the certainty in the models predictions is insufficient to directly support a risk assessment or regulatory decision, they provide hypotheses which can guide the design of studies that test both the quantitative veracity of an AOP and the presumption that the relationship of KEs to the adverse outcome are correct.

Case Example 2: Skin Sensitization in Mammals

Another well-characterized AOP involves a MIE of covalent modification of cellular proteins in the skin by electrophilic chemicals, which results in an AO of sensitization of the skin to allergens (https://aopwiki.org/wiki/index.php/Aop:40; last accessed October 25, 2016; OECD, 2012). Intermediate KEs in the AOP capture processes related to the induction of inflammatory cytokines by dendritic cells and keratinocytes, and activation and proliferation of T-cells that ultimately cause sensitization (Kimber *et al.*, 2012). The scientific support for this AOP is considered to be strong (Patlewicz *et al.*, 2015; Perkins *et al.*, 2015) so it also offers an excellent basis for development of quantitative models relating the MIE to the AO (Maxwell *et al.*, 2014). In particular, there is a strong regulatory and scientific interest in applying mechanistic understanding captured in the AOP to help reduce and replace the need for animal testing associated with the hazard characterization and risk assessment of skin sensitizing chemicals for use in cosmetic and other consumer care products (eg, soaps, lotions).

To date, over 20 *in vitro* test methods have been developed to either assess skin sensitization hazard potential or characterize sensitizer potency [reviewed in Reisinger *et al.* (2015)]. The skin sensitization AOP has enabled a clearer dialogue with regulatory authorities and risk assessors on the mechanistic relevance of each of these *in vitro* approaches either when applied in isolation or when these datasets are combined using integrated testing strategies (ITS)/data integration procedures (DIP). Twelve skin sensitization DIPs have been identified and discussed as case studies within the OECD "Skin Sensitization IATA guidance working group" (including Bauch *et al.*, 2012; Gomes *et al.*, 2012; Hirota *et al.*, 2015; MacKay *et al.*, 2013; Natsch *et al.*, 2015; Patlewicz *et al.*, 2014; Takenouchi *et al.*, 2015; van der Veen *et al.*, 2014). As part of the research to develop these DIPs, a variety of different statistical and modeling approaches have been applied to skin sensitization datasets and this analysis has refined our mechanistic understanding of the disease process and enabled these mechanistic insights to be applied to decision-making. Jaworska *et al.* (2013) represent a good example of how these mechanistic insights can be applied to structure the weighting

and integration of skin sensitization data to predict sensitizer potency information. The authors describe a quantitative modeling approach for prediction of skin sensitization based on a Bayesian network construct that utilizes cumulative evidence from multiple *in vitro* assays reflecting different KEs within the AOP. Pirone *et al.* (2014) describe an optimized, open-source version of the skin sensitization model that classifies chemicals as non-, weak-, moderate-, or strong-sensitizers. The Bayesian network approach appears to be particularly well suited to interfacing with the AOP framework, as it allows the integrated consideration of mechanistic data from multiple biological levels of organization.

A further example is the toxicokinetic/toxicodynamic (TK/TD) modeling approach taken by MacKay *et al.* (2013) who have extended an existing skin bioavailability model (Davies *et al.*, 2011) to enable the probability of allergy in a given human population to be predicted using a TD model of skin protein haptenation, DC antigen presentation and $CD8^+$ T cell activation. As these pathways represent only a subset of the skin sensitization AOP steps/key events, an analysis of model uncertainty has been undertaken to enable the impact of these assumptions on the model prediction to be quantified and explicitly represented. As this approach has been designed for application in skin sensitization risk assessment, the consequence of this level of mechanistic transparency is to empower the risk assessor to judge for themselves whether the level of uncertainty is appropriate for a given application of the model.

Case Example 3: Inhibition of Thyroid Hormone Synthesis/Degradation Leading to Varied Developmental Effects

Pathways associated with the hypothalamic–pituitary–thyroidal (HPT) axis control aspects of early development and metabolic homeostasis in all vertebrates. Enough is known concerning basic HPT functional biology, and the effects of chemical perturbation on this biology, that it is possible to depict the axis as an AOP network comprised of multiple MIEs and discrete early KEs that converge at a shared node/KE of (reduced) plasma thyroid hormone (TH) concentrations, which subsequently can be associated with various stage/species-specific AOs (Crofton, 2008; Perkins *et al.*, 2013). Prominent MIEs include perturbation of proteins involved in the production of TH (eg, decreased activity of thyroperoxidase, inhibition of the sodium iodide symporter [NIS]), as well as its enhanced degradation/elimination (eg, via induction of UDPG-transferases [UDPGT]). The types of AOs that can be linked to quantifiable depressions in plasma TH can be as varied as increased incidence of developmental neurological effects (eg, IQ decreases) in humans, to delayed amphibian metamorphosis that result in potential population-level impacts.

There have been a number of efforts to produce quantitative models focused on associations between TH and AOs. Some of this work has involved consideration of chemical dose in the context of response–response relationships among KEs. For example, Parham *et al.* (2012) and Wise *et al.* (2012) describe quantitative statistical models for the association between decreased TH and neurodevelopmental effects of PCBs of differing potency relative to their induction of UDPGTs, using both rodent and human data. Employing a more mechanistic approach, Fisher *et al.* (2013) developed a BBDR model that relates reductions in serum iodide concentrations (which could be caused by inhibition of NIS) to decreases in TH that could ultimately lead to developmental neurotoxicity. In more recent work, Lumen *et al.* (2015) conducted a quantitative sensitivity analysis for a BBDR model describing relationships between maternal TH synthesis and subsequent transfer of hormone to the developing fetus. These types of analyses can be guided by KEs described and depicted in HPT AOP networks (Perkins *et al.*, 2013).

This Case Example differs from the approach of the previous 2 in a couple regards. First, the theoretical modeling framework is considered as an interactive network in which discrete responses elicited via multiple MIEs converge on a shared node, decreased TH concentrations, which provide a basis for predicting any of a number of different AOs. Second, in this context, as opposed to Case Examples 1 and 2, only the initial portion of the AOPs leading to the convergent KE of reduced serum TH concentration is quantitatively modeled. However, this is completely adequate for some intended regulatory uses. The evidence for an association between a defined percent reduction in maternal serum TH concentration and adverse neurodevelopmental effects in humans is adequate for assessors to reasonably conclude that the effect is adverse (Gilbert, 2011). Indeed, the connection between alterations in serum TH and adverse effects at the individual or population level are supported by the remainder of the AOP description (eg, https://aopwiki.org/wiki/index.php/Aop:134; last accessed October 25, 2016). In this manner, the strength of the AOP downstream of reduced serum TH KE helped focus model development on upstream key events for which relevance to adversity was less widely accepted.

Case Example 4: Activation of Estrogen Receptor-α Leading to Diverse Adverse Outcomes

For some applications, quantitative modeling may only need to capture the MIE and/or very early KEs of an AOP. An example of this is illustrated by ongoing activities through the US Environmental Protection Agency's Endocrine Disruptor Screening Program (EDSP), the objective of which is to identify chemicals with potential to cause adverse effects through alteration of pathways associated with HPT and HPG function (US EPA, 2014). One MIE of concern is activation of the estrogen receptor-alpha (ERα). Estrogenic chemicals have been associated with a large number of different AOs involving reproduction and development in vertebrates (WHO/IPCS, 2002). Scientists involved with the EDSP effort recently described the development of a network model to predict the potential for chemicals to act as estrogens *in vivo* based on a chemical's ability to elicit responses in high-throughput (HTP) *in vitro* assays that capture multiple aspects of the MIE, including binding to ERα, receptor dimerization, chromatin binding, transcriptional activation, and ER-dependent cell proliferation (Browne *et al.*, 2015; Judson *et al.*, 2015). The quantitative model provides potency values for test chemicals relative to E2 (the ERα endogenous ligand), and was evaluated/validated by comparing model output to results from the uterotrophic assay, an *in vivo* pathway-based system considered to be a "gold standard" for identifying ERα agonists (Browne *et al.*, 2015; Kleinstreuer *et al.*, 2015).

Whereas this particular quantitative model only reflects early portions of AOPs relevant to interaction with the ERα, it nonetheless has substantial utility for addressing one of the challenges faced by EDSP. Specifically, it is being utilized by the USEPA to prioritize 10 000-plus chemicals for more resource-intensive *in vivo* testing necessary to assess potential risks, based on their predicted estrogenic potency. A recent "proof of concept" study conducted through the EDSP indicates that the quantitative model predictions, in conjunction with a rapid exposure assessment, provide a reasonable basis for test chemical prioritization based on agreement between the *in vitro*-based predictions and *in vivo* results available for a reference set of estrogenic compounds (US EPA, 2014). In this hazard-based scenario, the AOP provides a toxicological "anchor" for

selection/use of the HTP assays in the context of application of the computational model to hazard assessment.

Based on case examples already available, the potential utility of the integrated application of AOPs and computational models is apparent. Whereas an AOP description lays out the sign posts within a biological system that indicate progression toward an AO, computational models can quantitatively simulate the dynamics of the complex biology at multiple scales that dictate dose–response and time–course behaviors and define the conditions under which perturbation of early KEs in the pathway will ultimately lead to the AO, or not (eg, Case Example 1, Case Example 3). In particular, computational models are well equipped to represent multi-factorial cellular processes and mechanisms such as feedback, feed forward, and ultrasensitive motifs (Zhang *et al.*, 2014) that give rise to systems behaviors that are neither intuitive, nor easily represented in a text-based AOP description. In addition to representing and simulating the biological systems themselves, computational models can help integrate diverse data streams from a range of pathway-based assays to provide potency estimates that can be used to directly inform risk (eg, Case Example 2) or to prioritize compounds for more detailed testing (eg, Case Example 4). Ultimately, whereas both AOPs and computational models help organize and apply existing knowledge and data they do so in different ways. Consequently, through effective integration computational models can enhance the predictive utility of AOPs and vice versa.

ENSURING AOP INFORMED PREDICTIVE MODELS ARE FIT FOR REGULATORY PURPOSES

Understanding and acceptance by the regulatory community is essential to the application of AOPs and associated quantitative models in a range of envisaged applications, including integrated approaches to testing and assessment (IATA) and/or dose–response assessment for individual and/or groups of chemicals. Principles that contribute to regulatory uptake of evolving technologies and quantitative modeling, described previously (Meek and Lipscomb, 2015), are relevant in the context of AOPs and computational prediction models that align with them (IPCS, 2010; Meek *et al.*, 2013). In particular, consideration of the uncertainties associated with an AOP or model-based prediction is a critical in determining fit-for-purpose.

In principle, the degree of uncertainty that can be tolerated in AOP and/or associated quantitative models derives from the problem formulation phase of a risk assessment (NRC, 2009). During problem formulation, model developers, intended users, and decision makers collaborate closely to define the scope and regulatory purposes for which a prediction model is needed, the type of model best suited to meet the regulatory objectives, the data criteria, and the model's domain of applicability (US EPA, 2009). An individual AOP is not a complete, detailed, representation of complex biological processes. Rather it describes, in a simplified but structured way, existing knowledge concerning specific motifs of failure of a biological system, culminating in an adverse outcome (Villeneuve *et al.*, 2014). In this respect, an AOP itself can be considered a conceptual model, in that it represents a useful abstraction of the real world. As described above, these conceptual models (AOPs) can be gainfully employed to aid develop computational models predictive of toxicity. The desired level of detail in the resulting AOP or computational model(s), ie, the complexity or granularity required, depends, among other factors, on its regulatory application (ie, problem formulation). Different regulatory applications have different needs in terms of level of confidence in the assessment, which determines the level of uncertainty of the model that is acceptable. Model uncertainty is influenced by model complexity (US EPA, 2009). Model complexity, in turn, depends on the toxicological knowledge that is available; compared with information-rich AOPs, information-poor AOPs may give rise to computational models which are based on more assumptions and thus have a higher level of uncertainty. Although a higher level of uncertainty may sound undesirable, the development of a computational model focused on the regulatory objectives, rather than complete recapitulation of the system, is actually useful as it saves time and reduces data requirements. Consultation and engagement of both the development (ie, modelling) and application communities in model refinement (IPCS, 2010) is also advised as a basis to inform the extent of testing and sensitivity analysis of either models or AOPs require for specific application.

Because AOPs are, by definition, not chemically specific (Villeneuve *et al.*, 2014), predictive models derived from them will also be non-chemically specific by nature. This is advantageous in the sense that resources invested in developing AOP-based prediction models should have general applicability for a range of chemicals exhibiting similar bioactivity. However, it is also recognized that chemical-specific exposure considerations which drive many of the uncertainties in a given risk assessment are not explicitly accounted for in AOP-based models. Therefore, in practice, the output of AOP-informed models needs to be evaluated together with the output of physiologically-based toxicokinetic (PBTK) models and models that estimate exposure, preferably in a quantitative fashion. Recognizing this need, the Aggregate Exposure Pathway (AEP) framework was recently proposed as a companion to the AOP framework (Teeguarden *et al.*, 2016). The interface of AEPs with AOPs and between toxicokinetic and toxicodynamic models occurs at the MIE, where a chemical reaching its target site of action triggers an AOP. In applying AOP and AEP-based prediction models for regulatory purposes, the relevant uncertainties in each need to be considered.

ENGAGING THE MODELING COMMUNITY—A PROPOSAL

The discussion and examples in the preceding sections have demonstrated that AOPs and context-specific problem formulation can help focus and inform the development of computational toxicity prediction models. However, the question remains, "how can we engage the broader computational modeling community in utilizing AOP knowledge and related data to develop tractable and effective predictive models that transform chemical risk assessment and regulatory decision-making in accordance with the vision for toxicity testing in the 21st century?" The OECD and other organizations have provided AOP training to a variety of stake-holders in the toxicology research and regulatory communities. However, to date, those outreach efforts have not specifically targeted major biological modeling audiences. Increased efforts in that regard would be useful. Likewise, longer term there is a need to reshape academic program introduce more quantitative modeling into toxicology curricula and to stimulate stronger interaction and collaboration between quantitatively trained individuals and applied toxicologists.

In the near term, crowd sourcing-based challenges offer another possibility that has been effective for engaging the modeling community in specific problems. In general, such

challenges have defined an intellectual problem for modelers to address and provided a forum for them to share their proposed solution(s) along with a peer-reviewed publication, notoriety, and maybe a cash prize for winners. An early example related to modeling biology is the Dialogue on Reverse Engineering Assessment and Methods, or DREAM conference, which initially focused on computational methods to infer cellular networks from high-throughput data (Stolovitzky *et al.*, 2007). In the past 10 years, the DREAM challenges have continued to grow in both scope and complexity, not only building communities of researchers focused on a common goal, but also producing robust performance evaluation criteria to assess and rank solutions (Meyer *et al.*, 2011). Similar crowd-sourcing challenges are being applied to a wide range of biological problems including: drug development (Meyer *et al.*, 2012), cross-species translation (Biehl *et al.*, 2015; Rhrissorrakrai *et al.*, 2015), population-wide chemicals sensitivity in the Tox21 1000 Genomes project (Eduati *et al.*, 2015), and chemical toxicity predictions in the Tox21 Data Challenge (Abdelaziz Sayed *et al.*, 2016; Capuzzi *et al.*, 2016; Mayr *et al.*, 2016; Huang *et al.*, 2016; Koutsoukas *et al.*, 2016; Stefaniak, 2015). These challenges do not just solve a problem, they also bring scientific communities together in a repeatable model that could provide a series of benefits to the AOP community.

In order to stimulate development of quantitative prediction models aligned with AOPs, and attract greater involvement of the broader modeling community, we recommend initiating a program of AOP challenges to develop real-world applications based on established AOP practices (OECD, 2013; https://aopwiki.org/; last accessed October 25, 2016), such as the recent announcement of the AOP Wiki Data Challenge (http://www.piscltd.org.uk/aop-prize/; last accessed October 25, 2016). Challenge formats could include: building the most predictive model for existing AOPs using a specific challenge data set; quantitatively defining a new AOP for an outcome of interest (eg, replacing animal testing for a particular purpose); integration of multiple AOPs or an AOP network into one consistent decision-making workflow (eg, to enable risk assessment of a given class of compounds). Challenges could be constrained to specific goals using a defined data set such as response-response data, *in vitro* assays, or 'omics data. This would help focus efforts and reduce potential advantages of groups with access to better data sets. Depending on the goal, online resources could be drawn upon through the course of the challenge, such as biological data for benchmark chemicals, contact details for experts in the field and/or webinars to help the challenge community use the AOP knowledgebase. Each challenge should be time-bound and would ideally have been demonstrated to be theoretically feasible in advance. There could also be very closely defined criteria for 'winning' the challenge (eg, level of predictive accuracy, improvement over existing approaches, novel insights of interest), a test set of anonymized data that could be used at the conclusion of the challenge to demonstrate a "win" and "winners" would be offered the opportunity to present their entry at a high impact conference relevant to the challenge problem and to publish in a special issue of a high impact journal.

Adverse outcome pathway (AOP) challenge competitions have the potential to achieve both short-term and long-term benefits for the AOP community. In the short-term, it should accelerate development of AOP-based predictive models that could potentially be used in real-world applications, generate a wealth of user input to broaden the content of the AOP knowledgebase, refine our quantitative understanding of the nominated pathways and provide guidance into how different approaches can be used to model AOPs for different applications. Longer-term benefits could be even more significant, the introduction of a generation of computer scientists and mathematical modelers into a new field where they find productive careers in predictive toxicology.

CONCLUSIONS

Predictive application of mechanistic data is envisioned as the future of regulatory toxicology and risk assessment. AOPs are, by design, intended to aid application of mechanistic data in regulatory decision-making. However, qualitative description of biologically plausible and empirically supported causal connections between biological changes measured at different levels of organization and outcomes of regulatory significance may not yield the quantitative precision and sophistication needed for some regulatory applications. Development of predictive computation models aligned with AOP knowledge and designed to meet the needs of context-specific problem formulations represents another critical piece of the emerging paradigm. The time is right for the toxicology community to engage and partner with the computational modeling expertise to advance the science of predictive toxicology.

ACKNOWLEDGMENTS

Mention of trade names or commercial products does not constitute endorsement or recommendation for use. The contents of this manuscript neither represent, nor necessarily reflect official US EPA or the US Army Corps Engineers policy. The opinions expressed and arguments employed herein are those of the authors and do not necessarily reflect the official views of the OECD or the governments of its member countries. Pacific Northwest National Laboratory is a multi-program laboratory operated by Battelle for the U.S. Department of Energy under Contract DE-AC05-76RL01830. The contents of this paper are the views of the authors and do not necessarily represent the views and policies of the European Commission. The workshop participants (full listing) include: Hristo Aladjov (OECD), Gerald T. Ankley (US EPA), David Asturiol Bofill (EC, JRC), Elisabet Berggren (EC, JRC), Hugh Byrne (Dublin Institute of Technology), Mark Cronin (Liverpool John Moores University), Joop de Knecht (OECD), Stephen Edwards (US EPA), Elena Fioravanzo (S-IN Soluzioni Informatiche), John-Paul Gosling (Leeds University), Elmar Heinzle (Saarland University), Günter Klambauer (Johannes Kepler Universität), Nicole Kleinstreuer (Integrated Laboratory Systems), Brigitte Landesmann (EC, JRC), Igor Linkov (US Army Corps of Engineers), Mirjam Luiten (RIVM), Cameron MacKay (Unilever), Gavin Maxwell (Unilever), Bette Meek (University of Ottawa), Sharon Munn (EC, JRC), Edward Perkins (US Army Corps of Engineers), Anna Price (EC, JRC), Tomasz Sobanski (ECHA), Daniel Villeneuve (US EPA), Katrina Waters (Pacific Northwest National Laboratory), Maurice Whelan (EC, JRC), Clemens Wittwehr (EC, JRC), Andrew Worth (EC, JRC).

FUNDING

The *AOP-informed Predictive Modeling Approaches for Regulatory Toxicology Workshop* (24-25 September 2015) was funded by

the European Commission's Joint Research Centre (JRC), Ispra, Italy.

REFERENCES

Abdelaziz Sayed, A., Spahn-Langguth, H., Werner-Schramm, K., and Tetko, I. V. (2016). Consensus approach for modeling HTS assays using in silico descriptors. *Front. Environ. Sci.* **4**, 2.

Ankley, G. T., Bennett, R. S., Erickson, R. J., Hoff, D. J., Hornung, M. W., Johnson, R. D., Mount, D. R., Nichols, J. W., Russom, C. L., Schmieder, P. K., *et al.* (2010). Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ. Toxicol. Chem.* **29**, 730–741.

Bauch, C., Kolle, S. N., Ramirez, T., Eltze, T., Fabian, E., Mehling, A., Teubner, W., van Ravenzwaay, B., and Landsiedel, R. (2012). Putting the parts together: Combining in vitro methods to test for skin sensitizing potentials. *Regul. Toxicol. Pharmacol.* **63**, 489–504.

Becker, R. A., Ankley, G. T., Edwards, S. W., Kennedy, S. W., Linkov, I., Meek, B., Sachana, M., Segner, H., Van Der Burg, B., Villeneuve, D. L., *et al.* (2015). Increasing scientific confidence in adverse outcome pathways: Application of tailored Bradford-Hill considerations for evaluating weight of evidence. *Regul. Toxicol. Pharmacol.* **72**, 514–537.

Bercu, J. P., Jolly, R. A., Flagella, K. M., Baker, T. K., Romero, P., and Stevens, J. L. (2010). Toxicogenomics and cancer risk assessment: A framework for key event analysis and dose-response assessment for nongenotoxic carcinogens. *Regul. Toxicol. Pharmacol.* **58**, 369–381.

Biehl, M., Sadowski, P., Bhanot, G., Bilal, E., Dayarian, A., Meyer, P., Norel, R., Rhrissorrakrai, K., Zeller, M. D., and Hormoz, S. (2015). Inter-species prediction of protein phosphorylation in the sbv IMPROVER species translation challenge. *Bioinformatics* **31**, 453–461.

Breen, M., Villeneuve, D. L., Ankley, G. T., Bencic, D., Breen, M. S., Watanabe, K. H., Lloyd, A. L., and Conolly, R. B. (2016). Computational model of the fathead minnow hypothalamic-pituitary-gonadal axis: Incorporating protein synthesis in improving predictability of responses to endocrine active chemicals. *Comp. Biochem. Physiol. C. Toxicol. Pharmacol.* **183–184**, 36–45.

Breen, M., Villeneuve, D. L., Ankley, G. T., Bencic, D. C., Breen, M. S., Watanabe, K. H., Lloyd, A. L., and Conolly, R. B. (2013). Developing predictive approaches to characterize adaptive responses of the reproductive endocrine axis to aromatase inhibition: II. Computational modeling. *Toxicol. Sci.* **133**, 234–247.

Browne, P., Judson, R. S., Casey, W. M., Kleinstreuer, N. C., and Thomas, R. S. (2015). Screening chemicals for estrogen receptor bioactivity using a computational model. *Environ. Sci. Technol.* **49**, 8804–8814.

Buick, J. K., Moffat, I., Williams, A., Swartz, C. D., Recio, L., Hyduke, D. R., Li, H. H., Fornace, A. J. J., Aubrecht, J., and Yauk, C. L. (2015). Integration of metabolic activation with a predictive toxicogenomics signature to classify genotoxic versus nongenotoxic chemicals in human TK6 cells. *Environ. Mol. Mutagen.* **56**, 520–534.

Capuzzi, S. J., Politi, R., Isayev, O., Farag, S., and Tropsha, A. (2016). QSAR modeling of Tox21 challenge stress response and nuclear receptor signaling toxicity assays. *Front. Environ. Sci.* **4**, 3.

Cheng, W., Zhang, Q., Schroeder, A., Villeneuve, D. L., Ankley, G., and Conolly, R. (2016) Computational modeling to evaluate alternative hypotheses for the linkage of aromatase inhibition to vitellogenin levels in fathead minnows. 55th Annual Meeting of Society of Toxicology, New Orleans, LA.

Crofton, K. M. (2008). Thyroid disrupting chemicals: Mechanisms and mixtures. *Int. J. Androl.* **31**, 209–223.

Davies, M., Pendlington, R. U., Page, L., Roper, C. S., Sanders, D. J., Bourner, C., Pease, C. K., and MacKay, C. (2011). Determining epidermal disposition kinetics for use in an integrated nonanimal approach to skin sensitization risk assessment. *Toxicol. Sci.* **119**, 308–318.

Eduati, F., Mangravite, L. M., Wang, T., Tang, H., Bare, J. C., Huang, R., Norman, T., Kellen, M., Menden, M. P., Yang, J., *et al.* (2015). Prediction of human population responses to toxic compounds by a collaborative competition. *Nat. Biotechnol.* **33**, 933–940.

Fisher, J. W., Li, S., Crofton, K., Zoeller, R. T., McLanahan, E. D., Lumen, A., and Gilbert, M. E. (2013). Evaluation of iodide deficiency in the lactating rat and pup using a biologically based dose-response model. *Toxicol. Sci.* **132**, 75–86.

Garcia-Reyero, N., and Perkins, E. J. (2011). Systems biology: Leading the revolution in ecotoxicology. *Environ. Toxicol. Chem.* **30**, 265–273.

Garcia-Serna, R., Vidal, D., Remez, N., and Mestres, J. (2015). Large-scale predictive drug safety: From structural alerts to biological mechanisms. *Chem. Res. Toxicol.* **28**, 1875–1887.

Gilbert, M. E. (2011). Impact of low-level thyroid hormone disruption induced by propylthiouracil on brain development and function. *Toxicol. Sci.* **124**, 432–445.

Gomes, C., Nocairi, H., Thomas, M., Collin, J. F., Ibanez, F., and Saporta, G. (2012). Stacking prediction for a binary outcome. *COMPSTAT, 20th International Conference on Computational Statistics, Limassol, Cyprus*, 271–282.

Guyton, K. Z., Kyle, A. D., Aubrecht, J., Cogliano, V. J., Eastmond, D. A., Jackson, M., Keshava, N., Sandy, M. S., Sonawane, B., Zhang, L., *et al.* (2009). Improving prediction of chemical carcinogenicity by considering multiple mechanisms and applying toxicogenomic approaches. *Mutat. Res.* **681**, 230–240.

Hirota, M., Fukui, S., Okamoto, K., Kurotani, S., Imai, N., Fujishiro, M., Kyotani, D., Kato, Y., Kasahara, T., Fujita, M., *et al.* (2015). Evaluation of combinations of in vitro sensitization test descriptors for the artificial neural network-based risk assessment model of skin sensitization. *J. Appl. Toxicol.* **35**, 1333–1347.

Huang, R., Xia, M., Nguyen, D. T., Zhao, T., Sakamuru, S., Zhao, J., Shahane, S. A., Rossoshek, A., and Simeonov, A. (2016). Tox21Challenge to build predictive models of nuclear receptor and stress response pathways as mediated by exposure to environmental chemicals and drugs. *Front. Environ. Sci.* **3**, 85.

IPCS. (2010). Characterization and Risk, application of physiologically based pharmacokinetic models in assessment. World Health Organisation, International Programme on Chemical Safety, Geneva, Switzerland. WHO/PCS/EDC/02.2. Available at: http://www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en/. Accessed October 25, 2016.

Jaworska, J., Dancik, Y., Kern, P., Gerberick, F., and Natsch, A. (2013). Bayesian integrated testing strategy to assess skin sensitization potency: from theory to practice. *J. Appl. Toxicol.* **33**, 1353–1364.

Judson, R. S., Magpantay, F. M., Chickarmane, V., Haskell, C., Tania, N., Taylor, J., Xia, M., Huang, R., Rotroff, D. M., Filer, D. L., *et al.* (2015). Integrated model of chemical perturbations of a biological pathway using 18 in vitro high-throughput

screening assays for the estrogen receptor. *Toxicol. Sci.* **148**, 137–154.

Kavlock, R., Chandler, K., Houck, K., Hunter, S., Judson, R., Kleinstreuer, N., Knudsen, T., Martin, M., Padilla, S., Reif, D., *et al.* (2012). Update on EPA's ToxCast program: Providing high throughput decision support tools for chemical risk management. *Chem. Res. Toxicol.* **25**, 1287–1302.

Kimber, I., Maxwell, G., Gilmour, N., Dearman, R. J., Friedmann, P. S., and Martin, S. F. (2012). Allergic contact dermatitis: A commentary on the relationship between T lymphocytes and skin sensitising potency. *Toxicology* **291**, 18–24.

Kleinstreuer, N. C., Ceger, P. C., Allen, D. G., Strickland, J., Chang, X., Hamm, J. T., and Casey, W. M. (2015). A curated database of rodent uterotrophic bioactivity. *Environ. Health Perspect.* **2**, 2.

Koutsoukas, A., Amand, J. S., Mishra, M., and Huan, J. (2016). Predictive toxicology: Modeling chemical induced toxicological response combining circular fingerprints with random forest and support vector machine. *Front. Environ. Sci.* **4**, 11.

Krewski, D., Acosta, D. J., Andersen, M., Anderson, H., Bailar, J. C., 3rd, Boekelheide, K., Brent, R., Charnley, G., Cheung, V. G., Green, S. J., *et al.* (2010). Toxicity testing in the 21st century: A vision and a strategy. *J. Toxicol. Environ. Health. B. Crit. Rev.* **13**, 51–138.

Li, Z., Kroll, K. J., Jensen, K. M., Villeneuve, D. L., Ankley, G. T., Brian, J. V., Sepulveda, M. S., Orlando, E. F., Lazorchak, J. M., Kostich, M., *et al.* (2011a). A computational model of the hypothalamic: pituitary: gonadal axis in female fathead minnows (*Pimephales promelas*) exposed to 17alpha-ethynylestradiol and 17beta-trenbolone. *BMC Syst. Biol.* **5**, 63.

Li, Z., Villeneuve, D. L., Jensen, K. M., Ankley, G. T., and Watanabe, K. H. (2011b). A computational model for asynchronous oocyte growth dynamics in a batch-spawning fish. *Can. J. Fish. Aquat. Sci.* **68**, 1528–1538.

Lumen, A., McNally, K., George, N., Fisher, J. W., and Loizou, G. D. (2015). Quantitative global sensitivity analysis of a biologically based dose-response pregnancy model for the thyroid endocrine system. *Front. Pharmacol.* **6**, 107.

MacKay, C., Davies, M., Summerfield, V., and Maxwell, G. (2013). From pathways to people: Applying the adverse outcome pathway (AOP) for skin sensitization to risk assessment. *Altex* **30**, 473–486.

Maltarollo, V. G., Gertrudes, J. C., Oliveira, P. R., and Honorio, K. M. (2015). Applying machine learning techniques for ADME-Tox prediction: A review. *Expert Opin. Drug Metab. Toxicol.* **11**, 259–271.

Maxwell, G., MacKay, C., Cubberley, R., Davies, M., Gellatly, N., Glavin, S., Gouin, T., Jacquoilleot, S., Moore, C., Pendlington, R., *et al.* (2014). Applying the skin sensitisation adverse outcome pathway (AOP) to quantitative risk assessment. *Toxicol. In Vitro* **28**, 8–12.

Mayr, A., Klambauer, G., Unterthiner, T., and Hochreiter, S. (2016). DeepTox: Toxicity prediction using deep learning. *Front. Environ. Sci.* **3**, 80.

Meek, M. E. B., Barton, H. a., Bessems, J. G., Lipscomb, J. C., and Krishnan, K. (2013). Case study illustrating the WHO IPCS guidance on characterization and application of physiologically based pharmacokinetic models in risk assessment. *Regul. Toxicol. Pharmacol.* **66**, 116–129.

Meek, M. E. B., and Lipscomb, J. C. (2015). Gaining acceptance for the use of in vitro toxicity assays and QIVIVE in regulatory risk assessment. *Toxicology* **332**, 112–123.

Meyer, P., Alexopoulos, L. G., Bonk, T., Califano, A., Cho, C. R., de la Fuente, A., de Graaf, D., Hartemink, A. J., Hoeng, J., Ivanov, N. V., *et al.* (2011). Verification of systems biology research in the age of collaborative competition. *Nat. Biotechnol.* **29**, 811–815.

Meyer, P., Hoeng, J., Rice, J. J., Norel, R., Sprengel, J., Stolle, K., Bonk, T., Corthesy, S., Royyuru, A., Peitsch, M. C., *et al.* (2012). Industrial methodology for process verification in research (IMPROVER): Toward systems biology verification. *Bioinformatics* **28**, 1193–1201.

Miller, D. H., and Ankley, G. T. (2004). Modeling impacts on populations: Fathead minnow (Pimephales promelas) exposure to the endocrine disruptor 17beta-trenbolone as a case study. *Ecotoxicol. Environ. Saf.* **59**, 1–9.

Miller, D. H., Tietge, J. E., McMaster, M. E., Munkittrick, K. R., Xia, X., Griesmer, D. A., and Ankley, G. T. (2015). Linking mechanistic toxicology to population models in forecasting recovery from chemical stress: A case study from Jackfish Bay, Ontario, Canada. *Environ. Toxicol. Chem.* **34**, 1623–1633.

Natsch, A., Emter, R., Gfeller, H., Haupt, T., and Ellis, G. (2015). Predicting skin sensitizer potency based on in vitro data from KeratinoSens and kinetic peptide binding: Global versus domain-based assessment. *Toxicol. Sci.* **143**, 319–332.

Neagu, D. C., Guo, G., Trundle, P. R., and Cronin, M. T. D. (2007). A comparative study of machine learning algorithms applied to predictive toxicology data mining. *Altern. Lab. Anim.* **35**, 25–32.

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

OECD. (2013). Guidance document on developing and assessing adverse outcome pathways. Environment, Health and Safety Publications, *Series on Testing and Assessment, No. 184*, Vol. ENV/JM/MONO(2013)6, p. 45. Organisation for Economic Cooperation and Development, Paris, France.

OECD. (2012). The Adverse Outcome Pathway for Skin Sensitisation initiated by covalent binding to proteins. Part 1: Scientific Evidence. Environment, Health and Safety Publications, *Series on Testing and Assessment, No. 168*, Vol. ENV/JM/MONO(2012)10/PART1. Organisation for Economic Cooperation and Development, Paris, France.

OECD. (2016). Users' Handbook supplement to the Guidance Document for developing and assessing Adverse Outcome Pathways. Environment, Health and Safety Publications, *Series on Testing and Assessment No. 233*, Organisation for Economic Cooperation and Development, Paris, France. Available at: http://aopkb.org/common/AOP_Handbook.pdf; Accessed October 25, 2016.

OECD. (2010). Workshop report on OECD countries activities regarding testing, assessment and management of endocrine disrupters. Environment, Health and Safety Publications, *Series on Testing and Assessment, No. 118*, Vol. ENV/JM/MONO(2010)2. Organisation for Economic Cooperation and Development, Paris, France.

Parham, F., Wise, A., Axelrad, D. A., Guyton, K. Z., Portier, C., Zeise, L., Thomas Zoeller, R., and Woodruff, T. J. (2012). Adverse effects in risk assessment: Modeling polychlorinated biphenyls and thyroid hormone disruption outcomes in animals and humans. *Environ. Res.* **116**, 74–84.

Patlewicz, G., Kuseva, C., Kesova, A., Popova, I., Zhechev, T., Pavlov, T., Roberts, D. W., and Mekenyan, O. (2014). Towards AOP application–implementation of an integrated approach to testing and assessment (IATA) into a pipeline tool for skin sensitization. *Regul. Toxicol. Pharmacol.* **69**, 529–545.

Patlewicz, G., Simon, T. W., Rowlands, J. C., Budinsky, R. A., and Becker, R. A. (2015). Proposing a scientific confidence framework to help support the application of adverse outcome

pathways for regulatory purposes. *Regul. Toxicol. Pharmacol.* **71**, 463–477.

Perkins, E. J., Ankley, G. T., Crofton, K. M., Garcia-Reyero, N., LaLone, C. A., Johnson, M. S., Tietge, J. E., and Villeneuve, D. L. (2013). Current perspectives on the use of alternative species in human health and ecological hazard assessments. *Environ. Health Perspect.* **121**, 1002–1010.

Perkins, E. J., Antczak, P., Burgoon, L., Falciani, F., Garcia-Reyero, N., Gutsell, S., Hodges, G., Kienzler, A., Knapen, D., McBride, M., *et al.* (2015). Adverse outcome pathways for regulatory applications: Examination of four case studies with different degrees of completeness and scientific confidence. *Toxicol. Sci.* **148**, 14–25.

Pirone, J. R., Smith, M., Kleinstreuer, N. C., Burns, T. A., Strickland, J., Dancik, Y., Morris, R., Rinckel, L. A., Casey, W., and Jaworska, J. S. (2014). Open source software implementation of an integrated testing strategy for skin sensitization potency based on a Bayesian network. *Altex* **31**, 336–340.

Reisinger, K., Hoffmann, S., Alepee, N., Ashikaga, T., Barroso, J., Elcombe, C., Gellatly, N., Galbiati, V., Gibbs, S., Groux, H., *et al.* (2015). Systematic evaluation of non-animal test methods for skin sensitisation safety assessment. *Toxicol. In Vitro* **29**, 259–270.

Rhrissorrakrai, K., Belcastro, V., Bilal, E., Norel, R., Poussin, C., Mathis, C., Dulize, R. H. J., Ivanov, N. V., Alexopoulos, L., Rice, J. J., *et al.* (2015). Understanding the limits of animal models as predictors of human biology: Lessons learned from the sbv IMPROVER Species Translation Challenge. *Bioinformatics* **31**, 471–483.

Stefaniak, F. (2015). Prediction of compounds activity in nuclear receptor signaling and stress pathway assays using machine learning algorithms and low dimensional molecular descriptors. *Front. Environ. Sci.* **3**, 77.

Stolovitzky, G., Monroe, D., and Califano, A. (2007). Dialogue on reverse-engineering assessment and methods: The DREAM of high-throughput pathway inference. *Ann. N. Y. Acad. Sci.* **1115**, 1–22.

Takenouchi, O., Fukui, S., Okamoto, K., Kurotani, S., Imai, N., Fujishiro, M., Kyotani, D., Kato, Y., Kasahara, T., Fujita, M., *et al.* (2015). Test battery with the human cell line activation test, direct peptide reactivity assay and DEREK based on a 139 chemical data set for predicting skin sensitizing potential and potency of chemicals. *J. Appl. Toxicol.* **35**, 1318–1332.

Teeguarden, J. G., Tan, Y. M., Edwards, S. W., Leonard, J. A., Anderson, K. A., Corley, R. A., Kile, M. L., Simonich, S. M., Stone, D., Tanguay, R. L., *et al.* (2016). Completing the link between exposure science and toxicology for improved environmental health decision making: The aggregate exposure pathway framework. *Environ. Sci. Technol.* **50,** 4579–4586.

Thomas, R., Thomas, R. S., Auerbach, S. S., and Portier, C. J. (2013). Biological networks for predicting chemical hepatocarcinogenicity using gene expression data from treated mice and relevance across human and rat species. *PLoS One* **8**, e63308.

US EPA. (2009). Guidance on the Development, Evaluation, and Application of Environmental Models. Office of the Science Advisor. Council for Regulatory Environmental Modeling U.S. Environmental Protection Agency, Washington, DC. EPA/100/K-09/003.

US EPA. (2014). Integrated Bioactivity and Exposure Ranking: A Computational Approach for the Prioritization and Screening of Chemicals in the Endocrine Disruptor Screening Program. U.S. Environmental Protection Agency Endocrine Disruptor Screening Program, Washington, DC. Available at: http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPP-2014-0614-0003; Accessed October 25, 2016.

van der Veen, J. W., Rorije, E., Emter, R., Natsch, A., van Loveren, H., and Ezendam, J. (2014). Evaluating the performance of integrated approaches for hazard identification of skin sensitizing chemicals. *Regul. Toxicol. Pharmacol.* **69**, 371–379.

Villeneuve, D. L., Crump, D., Garcia-Reyero, N., Hecker, M., Hutchinson, T. H., LaLone, C. A., Landesmann, B., Lettieri, T., Munn, S., Nepelska, M., *et al.* (2014). Adverse outcome pathway (AOP) development I: Strategies and principles. *Toxicol. Sci.* **142**, 312–320.

Watanabe, K. H., Mayo, M., Jensen, K. M., Villeneuve, D. L., Ankley, G. T., and Perkins, E. J. (2016). Predicting fecundity of fathead minnows (*Pimephales promelas*) exposed to endocrine-disrupting chemicals using a MATLAB(R)-based model of oocyte growth dynamics. *PLoS One* **11**, e0146594.

WHO/IPCS. (2002). Global Assessment of the State-of-the-science of Endocrine Disruptors. World Health Organization/International Programme on Chemical Safety. WHO/PCS/EDC/02.2, 180 pp. Available at: http://www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en; Accessed October 25, 2016.

Wise, A., Parham, F., Axelrad, D. A., Guyton, K. Z., Portier, C., Zeise, L., Zoeller, R. T., and Woodruff, T. J. (2012). Upstream adverse effects in risk assessment: A model of polychlorinated biphenyls, thyroid hormone disruption and neurological outcomes in humans. *Environ. Res.* **117**, 90–99.

Zhang, Q., Bhattacharya, S., Conolly, R. B., Clewell, H. J., Kaminski, N. E., and Andersen, M. E. (2014). Molecular signaling network motifs provide a mechanistic basis for cellular threshold responses. *Environ. Health Perspect.* **122**, 1261–1270.