2014

In Vitro and Modelling Approaches to Risk Assessment from the U.S. Environmental Protection Agency ToxCast Programme

Richard Judson
U.S. EPA, National Center for Computational Toxicology, Research Triangle Park, judson.richard@epa.gov

Keith Houck
U.S. EPA, National Center for Computational Toxicology, Research Triangle Park

Matt Martin
U.S. EPA, National Center for Computational Toxicology, Research Triangle Park

Thomas Knudson
U.S. EPA, National Center for Computational Toxicology, Research Triangle Park

Russell Thomas
U.S. EPA, National Center for Computational Toxicology, Research Triangle Park

See next page for additional authors

Authors
Richard Judson; Keith Houck; Matt Martin; Thomas Knudson; Russell Thomas; Nisha Sipes; U.S. EPA, National Center for Computational Toxicology, Research Triangle Park; John Wambaugh; and Kevin Crofton

BCPT Basic & Clinical Pharmacology & Toxicology

Basic & Clinical Pharmacology & Toxicology Doi: 10.1111/bcpt.12239

MiniReview

In Vitro and Modelling Approaches to Risk Assessment from the U.S. Environmental Protection Agency ToxCast Programme

Richard Judson, Keith Houck, Matt Martin, Thomas Knudsen, Russell S. Thomas, Nisha Sipes, Imran Shah, John Wambaugh and Kevin Crofton

U.S. EPA, National Center for Computational Toxicology, Research Triangle Park, NC, USA

(Received 27 January 2014; Accepted 24 March 2014)

Abstract: A significant challenge in toxicology is the 'too many chemicals' problem. Human beings and environmental species are exposed to tens of thousands of chemicals, only a small percentage of which have been tested thoroughly using standard *in vivo* test methods. This study reviews several approaches that are being developed to deal with this problem by the U.S. Environmental Protection Agency, under the umbrella of the ToxCast programme (http://epa.gov/ncct/toxcast/). The overall approach is broken into seven tasks: (i) identifying biological pathways that, when perturbed, can lead to toxicity; (ii) developing high-throughput *in vitro* assays to test chemical perturbations of these pathways; (iii) identifying the universe of chemicals with likely human or ecological exposure; (iv) testing as many of these chemicals as possible in the relevant *in vitro* assays; (v) developing hazard models that take the results of these tests and identify chemicals as being potential toxicants; (vi) generating toxicokinetics data on these chemicals to predict the doses at which these hazard pathways would be activated; and (vii) developing exposure models to identify chemicals for which these hazardous dose levels could be achieved. This overall strategy is described and briefly illustrated with recent examples from the ToxCast programme.

The last decade has seen a convergence of the need for alternatives to animal testing with the development and wide availability of large numbers of robust, high-throughput *in vitro* assays. Such alternatives are needed because of the large number of chemicals in commerce with inadequate toxicological information and concern about cost and animal welfare with large-scale animal testing [1]. Even where high-quality, guideline animal studies are available, there is a need for human-focused mechanistic information to help assess the human relevance of any adverse findings.

The U.S. Environmental Protection Agency (U.S. EPA) is contributing to the development and evaluation of high-throughput *in vitro* alternative toxicology methods through the ToxCast™ programme (ToxCast website: http://epa.gov/ncct/toxcast/). ToxCast was initiated in 2007 with the goal of testing a large and diverse set of environmentally relevant chemicals in a correspondingly broad range of *in vitro* high-throughput screening (HTS) and high-content screening (HCS) assays [2]. The initial aim was to provide a data set that could be used to evaluate the value of different assay technologies and computational/modelling approaches. To date, ToxCast has screened more than 1800 chemicals in as many as 700 assays. Several applications and modelling approaches have been developed and published, mainly focusing on the first 300 ToxCast chemicals, which are primarily data-rich pesticide active ingredients [3]. We are currently applying these approaches to the larger Phase II data set, which includes a more diverse set of chemicals. All of these data are publically available, so researchers outside of the U.S. EPA can perform independent evaluations and analyses through the ToxCast website and the ToxCast dashboard (http://actor.epa.gov/dashboard). ToxCast is one part of the U.S. multiagency toxicity testing programme called Tox21, which comprises researchers from the National Institutes of Health (NIH), the National Toxicology Programme (NTP), the Food and Drug Administration (FDA), and the U.S. EPA. Tox21 has many of the same aims as ToxCast, but covers a broader chemical space for a subset of HTS assay technologies [4].

A key element of the ToxCast/Tox21 paradigm is a focus on the molecular and cellular pathways that are targets of chemical interactions. By learning about these interactions and their consequences to tissue and organ function, we can manoeuvre tactically towards a more mechanistic understanding, as opposed to merely what diseases or adverse health effects are associated with potential chemical exposure. This deeper understanding leads to increasing confidence in identifying which populations might be at risk, significant susceptibility factors and key influences on the shape of the dose–response curve. The high-level plan to achieve this can be broken into seven major tasks: (i) identify biological pathways that, when sufficiently perturbed, can lead to specific manifestations of toxicity; (ii) develop high-throughput *in vitro* assays to test chemical perturbations of these pathways; (iii)

Author for correspondence: Richard Judson, U.S. EPA, National Center for Computational Toxicology, Research Triangle Park, NC 27711, USA (e-mail judson.richard@epa.gov).

identify the universe of chemicals with likely human or ecological exposure (for the moment ignoring compounds in development or not yet introduced into the environment); (iv) test as many of these chemicals as possible in biologically relevant *in vitro* assays; (v) build predictive models that take the results of these tests and identify chemicals as being potential toxicants; (vi) generate toxicokinetic data on these chemicals to predict the doses at which these hazard pathways would be activated; and (vii) develop exposure models to identify chemicals for which these hazardous dose levels could be achieved. These steps are summarized in fig. 1.

This MiniReview briefly summarizes the main tasks carried out in the ToxCast programme and lessons learned to date. The results are focused largely on the published work from Phase I of ToxCast, which tested about 300 chemicals, mostly pesticidal active ingredients. Phase II, which extends testing to as many as 1877 chemicals, is now close to completion, and results are beginning to be released [5,6].

Materials and Methods

Chemicals. Chemicals for testing in ToxCast and Tox21 were nominated by U.S. EPA, NIH, FDA, various stakeholder groups (industry, academia and non-governmental organizations), international governmental agencies and working groups of the Organisation for Economic Co-operation and Development. The ToxCast Phase I library contains 309 unique chemicals; the Phase II library contains 767 additional chemicals; and the E1K library (chemicals screened in endocrine-related assays) comprises another 880 chemicals (totalling 1000 samples, hence the designation E (for endocrine) 1K). A number of additional chemicals were added at several stages, so that the current data set contains some screening data on 1877 unique chemicals. The Tox21 list totals ~8200 chemicals, including pesticides, pharmaceuticals, food additives and food contact substances, cosmetics ingredients, personal care ingredients and industrial chemicals [7]. The complete list of chemicals with identity and structures can be found at the ToxCast website.

The testability and availability of chemicals are other important considerations when selecting chemicals for the ToxCast testing paradigm. In our current testing programme, all assays required that a chemical be soluble and stable in DMSO (assays are optimized to use DMSO-solubilized chemicals) and that it is not volatile (assays are run in unsealed micro-titre plates, so volatile chemicals would escape). We used calculated physico-chemical properties in an attempt to exclude chemicals that do not meet these criteria, but a small fraction of purchased chemicals fails at least one of these criteria based on follow-up analytical testing. When chemicals are tested, they are solubilized in DMSO and visually checked for the evidence of crystals or particles. After solubilization, all chemicals are diluted into aqueous buffer (to a final concentration of 100 μM) and analytically tested for presence and purity of the intended structure. This process is still ongoing but we see that approximately 5% of the compounds fail. Chemicals that fail this QC process are flagged, but the data are not removed. Failure could result from volatilization, degradation or poor initial sample quality. Finally, chemical availability is an important consideration. Many industrial chemicals are not available in small and pure quantities (we aim for 95% purity but will accept lower values). Some compounds (especially pharmaceutical ingredients) are excluded because they are very expensive or otherwise difficult to procure.

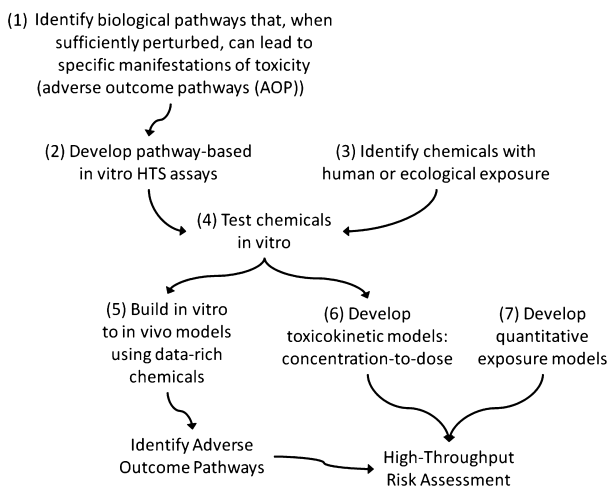

Fig. 1. Summary of the flow of major tasks in the ToxCast programme that lead from initial identification of relevant biological pathways, inclusion of exposure information, to cost-efficient preliminary risk assessments for thousands of chemicals.

The systematic cataloguing of chemicals, their uses and potential exposure routes were a major driver in the development of the ACToR database (Aggregated Computational Toxicology Resource, http://actor.epa.gov). Currently, ACToR has four main modules: (i) 'ACToR main'; (ii) ToxRefDB; (iii) ToxCastDB; and (iv) ExpoCastDB [7,8]. The main portion of ACToR ('ACToR main') compiles data from 1359 public databases with information on hazard, exposure, risk assessment, risk management, use, regulations and chemical properties (all statistics are for the database release actor_2012q1). There are 551,763 chemicals in the database, although data coverage varies from full data sets to very minimal information. The fact that many chemicals have almost no public information that can help with risk assessment is one of the main drivers of ToxCast and other screening programmes. ToxRefDB catalogues data from guideline *in vivo* toxicity studies and serves as the anchor for *in vitro*-to-*in vivo* modelling efforts described below [9–11]. ToxCastDB contains HTS data from ToxCast Phase I. ExpoCastDB provides data from biomonitoring studies [12]. The Chemical and Product Category (CPCat) database, which compiles data from ACToR on chemical uses, is available at http://actor.epa.gov/cpcat.

In vitro *assays and testing of chemicals. In vitro* assays were obtained from commercial testing laboratories, in-house laboratories at the U.S. EPA, collaborators at the NIH National Center for Advancing Translational Science (NCATS) and academic partners. In total, there are up to 700 assays or assay end-points being used as part of the ToxCast programme. These cover a large range of technologies, including cell-free biochemical assays, assays targeting nuclear and other receptors and other molecular targets, assays measuring downstream integrated cell processes and model organisms (especially zebrafish) [5,6,13–24]. The complete set of assay descriptions can be found at the ToxCast website. Where possible, the assays are annotated with target gene and other metadata. With the exception of a preliminary screen for a large number of cell-free biochemical assays (the Novascreen assays), all assays are run in concentration–response format. The Novascreen assays initially were run in duplicate at a single concentration (50 μM for most assays and 25 μM for cytochrome-P450 assays). If activity exceeded a pre-defined threshold, the chemical–assay pair was repeated in concentration–response mode. A customized data processing pipeline was developed to manage data analysis from each of the assays, beginning with mapping well-level data and continuing to chemical identification and concentration and to curve fitting and hit calling (the process of deciding whether there is a statistically significant concentration response). This process is documented in the reference [25]. The complete data set (Phase I, II,

E1K) contains 1,526,359 chemical–assay pairs and is available at the ToxCast website and dashboard.

Building of predictive models. For predictive model development, we combined ToxCast *in vitro* data with *in vivo* toxicity data from guideline studies compiled into ToxRefDB. Several preliminary first-generation (Phase I) models have been published, including ones for reproductive, developmental and chronic/cancer end-points [21,26–30]. These models used a combination of statistical and biologically based modelling approaches.

Reverse toxicokinetics. Recently, several publications have applied *in vitro*-to-*in vivo* extrapolation methods in pharmacokinetics to make first-order predictions of the scaling from ingested dose to a chemical's plasma concentration [31–34]. This approach requires that two experimental *in vitro* measurements be carried out: (i) clearance of the parent chemical in primary hepatocytes and (ii) the unbound fraction of the chemical present in the plasma. These measurements can be carried out using either human or rat hepatocytes and plasma. In a process called reverse toxicokinetics (RTK), a scaling factor is calculated for each chemical that coverts a steady-state blood concentration (Css) in micromoles to a daily dose in milligram per kilogram per day. The scaling factor can then be applied to the AC50 values (the concentration at which 50% of maximal activity is seen) in the *in vitro* assays to provide the daily dose that would be required to activate that pathway.

Biological coverage of toxicities. The ToxCast/Tox21 groups performed a broad survey of biological targets and pathways that, when perturbed by chemicals, could lead to adverse outcome [4]. This survey was focused on several modes of toxicity, namely endocrine disruption, cancer, prenatal developmental outcome and reproductive toxicity which includes fertility impairment. Many of the pathways tested involve pharmaceutical targets that could lead to adverse effects if improperly activated.

Simultaneously, the ToxCast/Tox21 groups surveyed the range of assay technologies that were available in high-throughput mode. This was motivated partly by the need to compare and contrast different technologies for their utility and value in screening large numbers of environmental chemicals. An early decision was made in ToxCast to initially build the assay portfolio with large batteries of commercially available HTS assays already used in the pharmaceutical industry [2]. As such, the programme could begin testing much sooner and over a broader assay space than would be possible if all assays were developed *de novo*. Subsequent to the initial selection of assays for Phase I of ToxCast, other assay sets have been added through collaborations with commercial suppliers, academic groups and in-house laboratories at the U.S. EPA. Although the majority of the assays that were run in Phase I of ToxCast yielded data that were useful for further analyses, some were purged from the data set (or discounted) because the data were noisy or otherwise unreliable. The complete data set (excluding the purged data) is available as a collection of files at the ToxCast website or through the ToxCast dashboard.

Identifying the chemical universe. A long-term goal of the ToxCast and Tox21 programmes is to test as many chemicals with potential for human exposure as possible. Classes of chemicals of interest include pesticidal active and inert ingredients; pharmaceuticals; industrial chemicals; cosmetic and personal care ingredients; food additives and contaminants and food contact substances; chemicals found in furniture, paints and other household products; and chemicals found in children's products, such as clothes, bedding and toys. Cataloguing chemicals, their uses and potential exposure routes were a major driver in the development of the ACToR database.

Using data from ACToR, we compiled a list of 9912 chemicals that were of high priority for screening, based on their use patterns [35]. These chemicals were largely limited to those subject to U.S. EPA regulatory authority. Of them, two-thirds had no publicly available toxicity/hazard data. A similar survey was carried out for the availability of exposure data (occurrence in surface or drinking water, air, food or soil; human biomonitoring data). A universe of 10,000 chemicals was surveyed, and only about 1% had any relevant exposure data [12]. Currently, we are carrying out a more complete survey of how chemicals are used, based on data from the CPCat database (http://actor.epa.gov/cpcat), to refine the universe of chemicals to test. The present survey goes beyond U.S. EPA-regulated chemicals to include a host of additional use classes listed in the previous paragraph. From this analysis, we identified 41,866 chemicals with at least one known use (e.g. industrial, consumer, etc.), of which 19,788 chemicals had at least one consumer use. Future testing is expected to include chemicals from this list.

Results

First-generation predictive models.

One key goal of the ToxCast programme is to discover correlations and linkages between molecular/cellular perturbations and apical toxicity end-points (adverse outcome). We approach this problem by building *in vitro*/*in vivo* computational (*in silico*) models. To do this, the ToxCast data are being combined with *in vivo* toxicity data from guideline studies in U.S. EPA's ToxRefDB. Several first-generation (Phase I) models have been published to date, including ones for reproductive, developmental and chronic/cancer end-points [11,21,27–30, 36,37]. These models currently are being tested and refined using the newest (Phase II) ToxCast data. An important point about all of these models is that the *in vitro* data are mostly from human cells, while the *in vivo* data are from rodents and rabbits.

Reproductive toxicity model. Initial models of reproductive toxicity were built using the data set compiled by Martin *et al.* [10]. This data set compiled information on 75 reproductive effects for 256 chemicals with data from both ToxCast and guideline multi-generation rat reproductive guideline studies performed as part of pesticidal active ingredient registrations. A total of 19 parental, offspring or reproductive end-points had a sufficiently high incidence after chemical exposure and were used as predictive end-points in the model. These included reproductive performance indices, male and female reproductive organ pathologies, offspring viability, growth and maturation, and parental systemic toxicities. Next, these end-points were combined with the ToxCast data to build a model of generalized reproductive toxicity. A reproductive toxicant was defined as a chemical with a reproductive adverse effect seen at <500 mg/kg/day. A total of 68 chemicals in the data set were considered to be reproductive toxicants. Using the *in vitro* assay data from ToxCast, a linear discriminant analysis (LDA) model was constructed that predicted the reproductive toxicity with a 74% balanced accuracy (BA = mean of sensitivity and specificity) based on cross-validation and a 76% BA using an external validation set. The *in vitro* assays used in the model

included activity in nuclear receptors [oestrogen receptor, androgen receptor, peroxisome-proliferator-activated receptor (PPAR)], cytochrome P450s, G-protein-coupled receptors and other cell signalling pathways. This model was also evaluated for its utility in prioritizing chemicals for further testing based on a scenario where many chemicals were tested *in vitro*, but where only a few could be tested *in vivo* because of cost and animal welfare considerations [29]. Two regulatory environments were evaluated in this study – one consistent with industrial chemicals where little data are required to be generated unless there is prior evidence of risk (screen in) and another where many studies are required for registration, but the U.S. EPA has the ability to waive (screen out) certain studies.

Developmental toxicity model. Models of prenatal developmental toxicity used data compiled from ToxRefDB on guideline rat and rabbit developmental toxicity studies [9]. A total of 383 rat and 368 rabbit studies were available, covering 387 chemicals, mostly pesticidal active ingredients. Of these chemicals, 283 were tested in both species, and, of those, 53 chemicals were specifically developmentally toxic (no overt maternal toxicity or maternal toxicity at doses higher than observed for the developmental defects). The primary expressions of developmental toxicity in pregnant rats were foetal weight reduction, skeletal variations and abnormalities, and foetal urogenital defects. Relative to rats, general pregnancy/foetal losses were more frequently observed in the rabbit as were structural malformations to the visceral body wall and CNS. Species-specific models were built on these data, linking *in vitro* ToxCast data to developmental defects (LDA with cross-validation) [28]. Specifically, 271 chemicals (187 unique) with ToxCast and ToxRefDB data were used, with 251 for the rat model (146 identified as developmental toxicants) and 234 for the rabbit model (106 identified as developmental toxicants). A developmental toxicant was defined as eliciting any significant end-point (i.e. foetal weight reduction, various malformations, prenatal loss) regardless of the maternal toxicity dose. The overall risk of a chemical causing development defects was linked to disruption of the following main targets and pathways: transforming growth factor beta (TGFβ), retinoic acid receptor (RAR) and G-protein-coupled receptors in rat; and interleukins and chemokines in rabbit. Species-specific models had a BA of about 70%. A key finding was that the molecular effects driving prenatal developmental toxicity showed strong species dependence in prediction models for pregnant rats and rabbits. Because the same set of *in vitro* assays was used for both species models, the differences are assumed to reflect model input parameters related to (i) the chemical space tested in each species; (ii) the apical end-points (*in vivo* outcome) recorded for each species, toxicokinetic differences between rats and rabbits, and/or toxicodynamic differences between the responses in pregnant dams and their conceptuses for either species.

Developmental vascular disruption model. Several of the molecular targets that were associated with developmental defects suggested a broad linkage between disruption of vascular development and the emergence of gross phenotypic developmental defects. This hypothesis led to the concept of 'putative vascular disrupting compounds' (pVDCs) [9,26,27,38]. An AOP linking multiple molecular initiating events to outcome was developed around the biomedical literature and Mouse Genome Informatics (MGI) database to provide a framework for identifying pVDCs based on ToxCast *in vitro* signatures. Particular targets included inflammatory chemokine signalling (CK), the vascular endothelial growth factor (VEGF) pathway and the plasminogen-activating system (uPAR). Consistent with the species dependence of prediction models built for prenatal developmental toxicity in pregnant rats and rabbits [28], we also observed species differences in models predicting pathway-level sensitivity to angiogenic signals, particularly those mediated by CK and uPAR pathways. This suggests a mechanistic link to species-dependent processes for inflammatory responses and extracellular (ECM) remodelling, respectively. The group of pVDCs with rat developmental toxicity correlated with down-regulation of pro-inflammatory CK assays, whereas pVDCs with rabbit activity often resulted in up-regulation of these signals. The rabbit pVDCs generally showed greater bioactivity across assays, which can be inferred to entail ECM degradation and release of angiogenic growth factors. The observed *in vivo* developmental toxicity also showed a distinct trend across species, with skeletal malformation in rats and prenatal death in rabbits being the most prevalent end-points for the pVDCs [27]. To further investigate this linkage, a cell/tissue-level dynamic signalling *in silico* model was developed [26] using the CompuCell3D (CC3D) software (http://www.compucell3d.org/). The *in silico* model was able to recapitulate self-directed assembly of endothelial cells into a completed vascular network utilizing signal-response pathways involving an exchange of CK, VEGF and uPAR among several cell types. By incorporating parameters from ToxCast HTS data into this 'virtual tissue model', the concentration-dependent disruption of angiogenesis was shown for 5HPP-33, an anti-angiogenic thalidomide analogue (fig. 2).

Cancer model. We have also published a first-generation prediction model linking *in vitro* effects and the likelihood that a chemical will be an *in vivo* carcinogen [36]. This model began with the hypothesis that chemicals perturbing cancer hallmark processes would increase the likelihood of those chemicals being carcinogens [39,40]. To test this hypothesis, univariate associations were calculated between each gene tested by the ToxCast assays and each cancer end-point (rat or mouse) in ToxRefDB. We found that the vast majority of cancer-linked genes (defined as having an odds-ratio >2, with confidence intervals not overlapping with zero after permutation testing) were either hallmark-associated or involved in xenobiotic metabolism. A scoring function was used that combined the cancer-associated gene hits for each chemical into an overall score. This was applied to an external test set of 33 chemicals that were not used in the model development process. The results were that the model

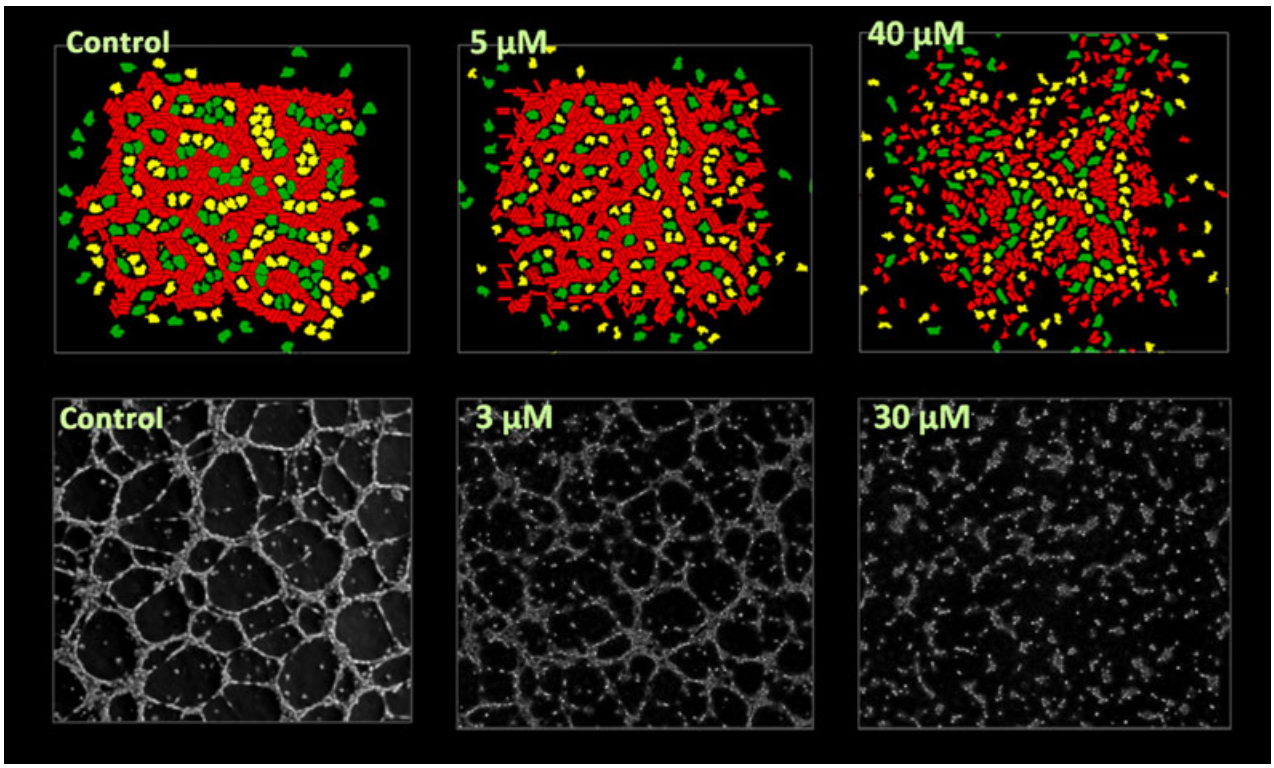

Fig. 2. Cellular systems-based simulation of endothelial cell organization into a complete endothelial network using CompuCell3D. Results comparing *in silico* disruption for 5HPP-33 based on ToxCast data (upper panel) *versus* a human angiogenesis (human umbilical vein endothelial cells – HUVEC) assay (lower panel). The middle and right panels show increasing lack of vascular organization with increasing dose. Reprinted from Kleinstreuer, Dix *et al.* 2013 [26].

distinguished between carcinogens and non-carcinogens with statistical significance ($p = 0.024$). Future work on all of the models will expand them to look in more detail at the molecular mechanisms linked to the adverse outcome and to forward validate using ToxCast Phase II data.

In vitro *pharmacokinetics – reverse toxicokinetics.*

To improve the accuracy of quantitative predictions of *in vivo* toxicity, it is necessary to have appropriate pharmacokinetic (PK) models [31,32,34,41]. New methods and models are being developed that increase throughput and efficiency and allow generation of simple PK data for hundreds of chemicals [41]. Measurements of clearance of the parent chemical in primary hepatocytes and the fraction unbound in the presence of plasma protein can be carried out using either human or rat hepatocytes and plasma. These experimental values are inputs to *in vitro*-to-*in vivo* extrapolation models, and the conversion of steady-state serum concentrations to administered dose using RTK. One use of this approach has been to compare doses at which each of the ToxCast assays would be predicted to be activated with the predicted doses from pesticide residues in the human diet [31,32,34,41]. This calculation showed that, for almost all pesticidal ingredients, no measured pathway would be activated, which is defined as a concentration reaching the AC50 for an assay, at the expected doses. One notable exception is triclosan, which has intentional exposures through its use in soaps, toothpaste and other consumer products. Note that the choice of the AC50 as a metric for the activating concentration is one of several points of departure that could be used. For the thirteen chemicals where human-relevant data (either PBPK models or actual serum measurements) is available, the steady-state serum predictions of the RTK method have been evaluated. For eight of the thirteen chemicals, the predictions were on the order of the data, while for the other cases, the predictions differed by orders of magnitude. The RTK predictions were conservative for all but perfluorinated compounds, which are hypothesized to be actively reabsorbed by the kidney after glomerular filtration [31].

High-throughput exposure modelling.

In order to perform risk assessments on the large numbers of chemicals to which human beings are exposed, we need methods for estimating exposures that are as high-throughput as HTS methods are for estimating hazard. To meet this need, we are developing high-throughput exposure models for this type of application under the U.S. EPA ExpoCast programme. Important aspects of these models are accurate statistical evaluation and estimation of uncertainties [42]. Exposure is a consequence of many factors, including human activity and the environment (both inside and outside of the home). To be high-throughput, exposure models must assume that chemicals can be described using only those parameters that can be

obtained easily. Depending on the specific model, these assumptions may be correct for some, none or all chemicals evaluated. The ExpoCast model evaluation framework treats ensembles of models as related assays for predicted chemical exposure. To evaluate the predictivity of a given model, evaluation sets must be obtained. Given that directly measuring human exposure is difficult, exposure has been inferred from biomonitoring data using simple pharmacokinetic models. This inference process is fraught with additional assumptions and is also a contributor to overall uncertainty in model predictions. With careful analysis, it has been shown that current models can be 95% confident that exposures lie within roughly eight orders of magnitude (e.g. 0.01 μg/kg/BW/day to 1 g/kg/BW/day) band, but that the upper 95% limit for roughly a third of 1936 chemicals evaluated to date was below the lowest oral equivalent dose identified by Wetmore and coworkers [31,32,34]. As both exposure prediction and biomonitoring exposure inference models improve, it is expected that uncertainty will continue to be reduced, enabling rapid risk-based prioritization of chemicals.

High-throughput risk assessment: combining the pieces.
A long-term goal of the ToxCast and related programmes is to develop methods for carrying out quantitative risk assessments in the absence of *in vivo* animal data. One key piece is an estimate of a no-effect level (NEL) for a chemical (e.g. a no-adverse-effect level) (NOAEL) or an NEL adjusted for uncertainty factors (e.g. a reference dose, RfD). The other is an estimate of the expected exposure level. Our approach to this latter quantity was outlined in the previous section. Here, we describe our initial approach to determining *in vitro* equivalents of an NEL or an RfD [33]. Firstly, because *in vitro* data are generated target by target or pathway by pathway, we derive a lowest concentration at which a particular pathway is altered, called the biological-pathway-altering concentration (BPAC). This quantity is divided by the scaling factor derived from the RTK modelling described above, to generate the biological-pathway-altering dose (BPAD). Both the BPAC and the scaling factor are subject to various types of uncertainty and population variability, which yield a distribution of the BPAD. We then take the lower end of the BPAD distribution (BPAD-L) as the pathway-wise equivalent of the RfD. This approach was demonstrated by calculating the BPAD-L for a set of conazole fungicides with the end-point being pregnane-X-receptor (PXR)-related liver hypertrophy. These values were all within a factor of 10 of the *in vivo* equivalent, and most were more conservative (BPAD-L < RfD) [33].

Discussion

We have briefly summarized the results derived from Phase I of the ToxCast project. The goals of Phase I largely have been met and include the following: (i) demonstrating that a large set of environmentally relevant chemicals can be screened in a diverse battery of *in vitro* assays; (ii) that predictive models of toxicity can be developed using these data; and (iii) that *in vitro* pharmacokinetic data can be integrated with the *in vitro* assay data, enabling us to make initial quantitative comparisons with *in vivo* rodent toxicity data.

That said, a number of challenges lie ahead. Some of these have been outlined by other researchers who performed independent analyses of the ToxCast data [43,44]. One challenge is presented by the broad diversity of chemicals, chemical–biological activities *in vitro* and chemical effects *in vivo*. At the very least, these pose a classic statistical power issue. For instance, if there are N different mechanisms by which a chemical can cause a specific phenotype, one needs a significant multiple of N examples for each such pathway-end-point pair in the data set to discover this linkage using purely statistical methods [37]. This argues for the need to increase the size of the data set (number of chemicals), which we are beginning to do with Phase II of ToxCast. This presents increasing challenges, however, because the number of chemicals with high-quality toxicity data (typically using guideline studies) has not increased as rapidly, simply because there is such a limited amount of these data available. A second challenge is how to link *in vitro* activity and *in vivo* effects. Purely, statistical methods that treat the data as just a set of numbers, in the same way that quantitative structure–activity relationship (QSAR) models are built, have inherent limitations [43,44]. One approach to solve this problem is to develop network models of adverse outcome pathways or modes of action [41,45–49]. These help form hypotheses for statistical testing, and also help minimize the multiple testing issues inherent in the purely statistical approach.

Another solution to both of these issues is to develop and employ more complex, cellular and multi-scale, biologically based models that incorporate information and knowledge of the structure of the biological pathways being altered that include pharmacokinetic information and that explicitly include multiple levels of biological organization [9,26,27,50–53]. This is the driver of our virtual tissue modelling research (http://www.epa.gov/ncct/v-Embryo/ http://www.epa.gov/ncct/virtual_liver/).

A complementary approach is to use *in vitro* data and model data in a tiered manner with risk-based metrics driving the elevation to higher tiers [33,41]. These approaches could, for instance, use point of departure metrics to guide further testing, and for many chemicals short circuit the need to focus on specific hazard identification. This is because expected exposures are likely much lower than the *in vitro*-derived point of departure. We suspect that this will leave a significantly smaller set of chemicals for which more complex, biologically based models, and follow-up testing, will be needed. There is also a potential (although this is unproven) that there are a relatively small set of critical pathways that will then be active in these remaining chemicals at relevant concentrations.

While these challenges are daunting, new modelling and *in vitro* methods and computational models provide avenues for improvement to existing methods. These efforts will improve risk predictions for human beings and ecosystem health for thousands of currently untested chemicals.

References

1 NRC. Toxicity Testing in the 21st Century: A Vision and a Strategy. National Academies Press, Washington D.C., 2007.

2 Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ. The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci 2007;**95**:5–12.

3 Kavlock R, Chandler K, Houck K, Hunter S, Judson R, Kleinstreuer N *et al.* Update on EPA's ToxCast program: providing high throughput decision support tools for chemical risk management. Chem Res Toxicol 2012;**25**:1287–302.

4 Tice RR, Austin CP, Kavlock RJ, Bucher JR. Improving the human hazard characterization of chemicals: a Tox21 update. Environ Health Perspect 2013;**121**:756–65.

5 Rotroff DM, Dix DJ, Houck KA, Kavlock RJ, Knudsen TB, Martin MT *et al.* Real-time growth kinetics measuring hormone mimicry for ToxCast chemicals in T-47D human ductal carcinoma cells. Chem Res Toxicol 2013;**26**:1097–107.

6 Sipes NS, Martin MT, Kothiya P, Reif DM, Judson RS, Richard AM *et al.* Profiling 976 ToxCast chemicals across 331 enzymatic and receptor signaling assays. Chem Res Toxicol 2013;**26**:878–95.

7 Judson RS, Martin MT, Egeghy P, Gangwal S, Reif DM, Kothiya P *et al.* Aggregating data for computational toxicology applications: the U.S. environmental protection agency (EPA) aggregated computational toxicology resource (ACToR) system. Int J Mol Sci 2012;**13**:1805–31.

8 Judson R, Richard A, Dix D, Houck K, Elloumi F, Martin M *et al.* ACToR–aggregated computational toxicology resource. Toxicol Appl Pharmacol 2008;**233**:7–13.

9 Knudsen TB, Martin MT, Kavlock RJ, Judson RJ, Dix DJ, Singh AV. Profiling the activity of environmental chemicals in prenatal developmental toxicity studies using the U.S. EPA's ToxRefDB. Reprod Toxicol 2009;**28**:209–19.

10 Martin MT, Judson RS, Reif DM, Kavlock RJ, Dix DJ. Profiling chemicals based on chronic toxicity results from the U.S. EPA ToxRef database. Environ Health Perspect 2009;**117**:392–9.

11 Martin MT, Mendez E, Corum DG, Judson RS, Kavlock RJ, Rotroff DM *et al.* Profiling the reproductive toxicity of chemicals from multigeneration studies in the toxicity reference database. Toxicol Sci 2009;**110**:181–90.

12 Egeghy PP, Judson R, Gangwal S, Mosher S, Smith D, Vail J *et al.* The exposure data landscape for manufactured chemicals. Sci Total Environ 2012;**414**:159–66.

13 Chandler KJ, Barrier M, Jeffay S, Nichols HP, Kleinstreuer NC, Singh AV *et al.* Evaluation of 309 environmental chemicals using a mouse embryonic stem cell adherent cell differentiation and cytotoxicity assay. PLoS ONE 2011;**6**:e18540.

14 Houck KA, Dix DJ, Judson RS, Kavlock RJ, Yang J, Berg EL. Profiling bioactivity of the ToxCast chemical library using BioMAP primary human cell systems. J Biomol Screen 2009;**14**:1054–66.

15 Huang R, Xia M, Cho MH, Sakamuru S, Shinn P, Houck KA *et al.* Chemical genomics profiling of environmental chemical modulation of human nuclear receptors. Environ Health Perspect 2011;**119**:1142–8.

16 Knight AW, Little S, Houck K, Dix D, Judson R, Richard A *et al.* Evaluation of high-throughput genotoxicity assays used in profiling the US EPA ToxCast chemicals. Regul Toxicol Pharmacol 2009;**55**:188–99.

17 Knudsen TB, Houck KA, Sipes NS, Singh AV, Judson RS, Martin MT *et al.* Activity profiles of 309 ToxCast chemicals evaluated across 292 biochemical targets. Toxicology 2011;**282**:1–15.

18 Martin MT, Dix DJ, Judson RS, Kavlock RJ, Reif DM, Richard AM *et al.* Impact of environmental chemicals on key transcription regulators and correlation to toxicity end points within EPA's ToxCast program. Chem Res Toxicol 2010;**23**:578–90.

19 Rotroff DM, Beam AL, Dix DJ, Farmer A, Freeman KM, Houck KA *et al.* Xenobiotic-metabolizing enzyme and transporter gene expression in primary cultures of human hepatocytes modulated by ToxCast chemicals. J Toxicol Environ Health B Crit Rev 2010;**13**:329–46.

20 Shah I, Houck K, Judson RS, Kavlock RJ, Martin MT, Reif DM *et al.* Using nuclear receptor activity to stratify hepatocarcinogens. PLoS ONE 2011;**6**:e14584.

21 Reif DM, Martin MT, Tan SW, Houck KA, Judson RS, Richard AM *et al.* Endocrine profiling and prioritization of environmental chemicals using ToxCast data. Environ Health Perspect 2010;**118**:1714–20.

22 Zaldívar Comenges JM, Wambaugh J, Judson R. Modelling *in vitro* cell-based assays experiments: cell population dynamics. In: Jordan F, Jorgensen SE, (eds). Models of the Ecological Hierarchy. Elsevier, Amsterdam, 2012; 92.

23 Cowden J, Padnos B, Hunter D, MacPhail R, Jensen K, Padilla S. Developmental exposure to valproate and ethanol alters locomotor activity and retino-tectal projection area in zebrafish embryos. Reprod Toxicol 2012;**33**:165–73.

24 Padilla S, Corum D, Padnos B, Hunter DL, Beam A, Houck KA *et al.* Zebrafish developmental screening of the ToxCast phase I chemical library. Reprod Toxicol 2012;**33**:174–87.

25 U.S. EPA. FIFRA SAP Meeting Held January 29–31, 2013 on the scientific issues associated with "prioritizing the universe of endocrine disruptor screening program (EDSP) chemicals using computational toxicology tools". http://ntp.niehs.nih.gov/NTP/About_NTP/SACATM/2013/September/SAPMtgRpt_Jan2013_508BE.pdf. 2013 (last accessed on 28 October 2013).

26 Kleinstreuer N, Dix D, Rountree M, Baker N, Sipes N, Reif D *et al.* A computational model predicting disruption of blood vessel development. PLoS Comput Biol 2013;**9**:e1002996.

27 Kleinstreuer NC, Judson RS, Reif DM, Sipes NS, Singh AV, Chandler KJ *et al.* Environmental impact on vascular development predicted by high-throughput screening. Environ Health Perspect 2011;**119**:1596–603.

28 Sipes NS, Martin MT, Reif DM, Kleinstreuer NC, Judson RS, Singh AV *et al.* Predictive models of prenatal developmental toxicity from ToxCast high-throughput screening data. Toxicol Sci 2011;**124**:109–27.

29 Martin MT, Knudsen TB, Judson RS, Kavlock RJ, Dix DJ. Economic benefits of using adaptive predictive models of reproductive toxicity in the context of a tiered testing program. Syst Biol Reprod Med 2012;**58**:3–9.

30 Martin MT, Knudsen TB, Reif DM, Houck KA, Judson RS, Kavlock RJ *et al.* Predictive model of rat reproductive toxicity from ToxCast high throughput screening. Biol Reprod 2011;**85**:327–39.

31 Wetmore BA, Wambaugh JF, Ferguson SS, Li L, Clewell HJ 3rd, Judson RS *et al.* Relative impact of incorporating pharmacokinetics on predicting *in vivo* hazard and mode of action from high-throughput *in vitro* toxicity assays. Toxicol Sci 2013;**132**:327–46.

32 Wetmore BA, Wambaugh JF, Ferguson SS, Sochaski MA, Rotroff DM, Freeman K *et al.* Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicol Sci 2012;**125**:157–74.

33 Judson RS, Kavlock RJ, Setzer RW, Cohen Hubal EA, Martin MT, Knudsen TB *et al.* Estimating toxicity-related biological pathway altering doses for high-throughput chemical risk assessment. Chem Res Toxicol 2011;**24**:451–62.

34 Rotroff DM, Wetmore BA, Dix DJ, Ferguson SS, Clewell HJ, Houck KA *et al.* Incorporating human dosimetry and exposure into high-throughput *in vitro* toxicity screening. Toxicol Sci 2010;**117**:348–58.

35 Judson R, Richard AM, Dix DJ, Houck K, Martin MT, Kavlock RJ *et al.* The toxicity data landscape for environmental chemicals. Environ Health Perspect 2009;**117**:685–95.

36 Kleinstreuer NC, Dix DJ, Houck KA, Kavlock RJ, Knudsen TB, Martin MT *et al. In vitro* perturbations of targets in cancer hallmark processes predict rodent chemical carcinogenesis. Toxicol Sci 2013;**131**:40–55.
37 Judson R, Elloumi F, Setzer RW, Li Z, Shah I. A comparison of machine learning algorithms for chemical toxicity classification using a simulated multi-scale data model. BMC Bioinformatics 2008;**9**:241.
38 Knudsen TB, Kleinstreuer NC. Disruption of embryonic vascular development in predictive toxicology. Birth Defects Res C Embryo Today 2011;**93**:312–23.
39 Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 2000;**100**:57–70.
40 Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;**144**:646–74.
41 Thomas RS, Philbert MA, Auerbach SS, Wetmore BA, Devito MJ, Cote I *et al.* Incorporating new technologies into toxicity testing and risk assessment: moving from 21st century vision to a data-driven framework. Toxicol Sci 2013;**136**:4–18.
42 Wambaugh JF, Setzer RW, Reif DM, Gangwal S, Mitchell-Blackwood J, Arnot JA *et al.* High-throughput models for exposure-based chemical prioritization in the ExpoCast project. Environ Sci Technol 2013;**47**:8479–88.
43 Thomas RS, Black MB, Li L, Healy E, Chu TM, Bao W *et al.* A comprehensive statistical analysis of predicting *in vivo* hazard using high-throughput *in vitro* screening. Toxicol Sci 2012;**128**:398–417.
44 Benigni R. Evaluation of the toxicity forecasting capability of EPA's ToxCast Phase I data: can ToxCast *in vitro* assays predict carcinogenicity? J Environ Sci Health C Environ Carcinog Ecotoxicol Rev 2013;**31**:201–12.
45 Sonich-Mullin C, Fielder R, Wiltse J, Baetcke K, Dempsey J, Fenner-Crisp P *et al.* IPCS conceptual framework for evaluating a mode of action for chemical carcinogenesis. Regul Toxicol Pharmacol 2001;**34**:146–52.
46 Meek ME, Bucher JR, Cohen SM, Dellarco V, Hill RN, Lehman-McKeeman LD *et al.* A framework for human relevance analysis of information on carcinogenic modes of action. Crit Rev Toxicol 2003;**33**:591–653.
47 Seed J, Carney EW, Corley RA, Crofton KM, DeSesso JM, Foster PM *et al.* Overview: using mode of action and life stage information to evaluate the human relevance of animal toxicity data. Crit Rev Toxicol 2005;**35**:664–72.
48 Boobis AR, Doe JE, Heinrich-Hirsch B, Meek ME, Munn S, Ruchirawat M *et al.* IPCS framework for analyzing the relevance of a noncancer mode of action for humans. Crit Rev Toxicol 2008;**38**:87–96.
49 Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD *et al.* Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 2010;**29**:730–41.
50 Kleinstreuer NC, Smith AM, West PR, Conard KR, Fontaine BR, Weir-Hauptman AM *et al.* Identifying developmental toxicity pathways for a subset of ToxCast chemicals using human embryonic stem cells and metabolomics. Toxicol Appl Pharmacol 2011;**257**:111–21.
51 Knudsen T, Martin M, Chandler K, Kleinstreuer N, Judson R, Sipes N. Predictive models and computational toxicology. Methods Mol Biol 2013;**947**:343–74.
52 Shah I, Wambaugh J. Virtual tissues in toxicology. J Toxicol Environ Health B Crit Rev 2010;**13**:314–28.
53 Wambaugh J, Shah I. Simulating microdosimetry in a virtual hepatic lobule. PLoS Comput Biol 2010;**6**:e1000756.