Navigating the Minefield of Computational Toxicology and Informatics: Looking Back and Charting a New Horizon

Grace Patlewicz*

Independent Researcher, Durham, NC, United States

INTRODUCTION

As we enter 2020, it is worth looking back at the development and progression of the computational toxicology discipline, how it has evolved and what some opportunities might be going forward. Computational toxicology stands poised to broadly and directly inform chemical safety assessment, and as such, the demands of computational toxicology are growing due to international regulatory needs. Critical to increasing scientific confidence in the use of computational toxicology approaches in applied toxicology decision-making will be: (1) transparency and reproducibility in the underlying data and data analysis approaches utilized; (2) accessibility of information to evaluate the fitness of the computational toxicology approach for a particular problem; and (3) sharing of ideas and approaches internationally. Herein the progress in applied computational toxicology is considered, with a call for additional research to continue this rapid advancement.

EARLY COMPUTATIONAL TOXICOLOGY: APPLICATION OF QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIPS [(Q)SARs]

A quarter of century ago, the field of computational toxicology might simply have been summarized as the intersection of three scientific domains: toxicology, chemistry, and statistics, packaged in predictive models such as SAR and QSAR models, collectively referred to as (Quantitative) Structure Activity Relationships [(Q)SARs]. (Q)SARs are theoretical models that can be used to predict in a quantitative (e.g., potency) or qualitative manner (e.g., active/inactive) the physicochemical, biological [e.g., an (eco)toxicological endpoint] and environmental fate properties of compounds from the knowledge of their chemical structure (Worth et al., 2005). A SAR is a (qualitative) association between a chemical substructure and the potential of a chemical containing the substructure to exhibit a certain biological effect. The classic example of a SAR was the supramolecule published by Ashby and Tennant (1988) which related chemical structure moieties to genotoxic carcinogenicity. Typical toxicity endpoints under study were those with a greater preponderance of data such as the Ames test for bacterial mutagenicity (see Benigni and Bossa, 2019 for a recent review), or the fathead minnow fish acute lethality test (Adhikari and Mishra, 2018). Physicochemical properties such water solubility, octanol-water partition coefficient (LogKow) were also modeled. The algorithms underpinning these predictive models (QSAR) tended not to be overly complex mainly because the data volume was usually limited, "big" data was confined to a few hundred data points and typically far less. The algorithms used to develop QSARs relied upon conventional statistical approaches such as linear regression or logistic regression,

in part because the data volume did not merit more complex models, and in part since the models being developed relied on a limited set of descriptors that could be readily computed and interpreted relative to the property being modeled. Indeed, many fish toxicity models and Ames models relied upon LogKow as the main determining factor. Computing descriptors for chemicals was mainly dependent on commercial software within specific QSAR modeling platforms, e.g., TSAR from Oxford Molecular, Biovia's QSAR Workbench (https://www.3dsbiovia.com/products/collaborative-science/biovia-qsar-workbench/). The types of QSAR models for toxicity were often "local," i.e., based on defined mechanisms or chemical classes. The exception tended to be for physicochemical parameters where models were categorized as "global," i.e., heterogenous training datasets comprising a diversity in chemical structure.

The underlying principle within this framework was that the toxicity (property) being predicted was a function of chemical structure. The notion of similarity where similar chemicals were expected to cause similar toxicities also formed the complementary basis around the concept of read-across (Patlewicz et al., 2018) as well as Thresholds for Toxicological Concern (TTC) approaches (Kroes et al., 2004).

The application of these models was also limited, usually in providing preliminary indications of activity rather than in lieu of additional empirical data. The toxicity would be characterized by a single summary endpoint e.g., point of departure such as a No Adverse Effect Level (Concentration) [NOAEL(C)] and usually a single value for a given substance. The concept of reproducibility of the test method was not a major consideration, since studies tended not be repeated due to cost, animal use, and time constraints.

EVOLVING REGULATORY LANDSCAPE FOR (Q)SARs IN APPLICATION

In the late 1990s, there started to be a need to make predictions for a broader coverage of chemicals beyond the smaller datasets that underpinned the mechanistic chemical class type QSAR models to date. The shift was partly driven by interest in improvements to quantitative descriptions of chemical structure for toxicity prediction, and the availability of computing power.

Decision contexts were also changing and provided an additional impetus for new model development. Two main drivers were influencing this change: the need for non-animal alternatives, largely prompted the EU Cosmetics Regulation (European Commission, 2009) and the EU Chemicals legislation known as REACH (European Commission, 2006). REACH in particular had a profound effect in the development, evaluation, and application of QSARs, primarily since the decision context was to use QSAR predictions as supporting information in the construct of an Integrated Approaches to Testing and Assessment (IATA) (Tollefsen et al., 2014) and/or in lieu of new experimental testing. In the run up to REACH coming into force, there was a concerted effort to characterize a framework to facilitate the use of (Q)SARs for regulatory purposes (Cronin et al., 2003a,b). This culminated in the formulation of the OECD QSAR principles for

validation namely: a defined endpoint, unambiguous algorithm, appropriate measures of predictivity (e.g., external validation), goodness of fit (e.g., cross-validation), an applicability domain and mechanistic interpretation if possible (OECD, 2004, 2007; Patlewicz et al., 2016). The QSAR Validation principles largely provided the impetus to develop new approaches to characterize the applicability domain of models (Netzeva et al., 2005; Nikolova-Jeliazkova and Jaworska, 2005) as well as consider integration of models e.g., consensus models (e.g., Votano et al., 2004). In the development of Frontiers in Toxicology: Computational Toxicology and Informatics, a focus on the validation principles and their applicability to (Q)SARs and beyond is a component of advancing the scientific confidence in using these approaches in applied decision-making.

BROADENING COMPUTATIONAL TOXICOLOGY TO A STRATEGIC *IN SILICO* AND *IN VITRO* APPROACH AS SUPPORTED BY INFORMATICS

At the same time, the NRC report was published (NRC, 2007) which outlined the change in how toxicity testing could be undertaken. Subsequent reports on computational methodology for exposure (NRC, 2012) and risk assessment (NRC, 2017) have broadened the call. The NRC reports, together with the synergism of increased computing resources, increased access to laboratory automation for toxicology, and development of methodologies that efficiently generated large volumes of data, generated a disruptive change in the field and an expansion of what computational toxicology represented. Instead of summarizing toxicity on the basis of traditional toxicity tests, a shift was proposed to predict genotoxic vs. non-genotoxic substances, and then to have *in vitro* bioactivity and predicted exposure define a bioactivity:exposure ratio, which would inform the need for models of greater biological complexity (Thomas et al., 2013, 2019). This shift is dependent on high throughput and high content screening methods (HTS/HCS), including high throughput transcriptomics (HTTr) and high throughput phenotypic profiling (HTPP) of cellular morphology (Harrill et al., 2019; Thomas et al., 2019; Nyffeler et al., 2020).

The data needed for a rapid, high-throughput safety assessment requires application of a range of computational approaches for data analysis, data storage, and *in silico* predictive modeling. These challenges are directly identified in the title of this journal as the necessary "informatics" component of realizing computational toxicology for safety assessment. How to meet these informatic challenges is the subject of ongoing research as the volume and variety of data require tools for large scale data processing, databasing and informatics for single-dimension and multi-dimensional datasets, visualization for heterogeneous information, demonstrating reproducibility, and quality control, and perhaps most challenging, for interpretation and communication in the appropriate format and context for chemical safety assessment. Many aspects of the vision articulated by the initial NRC report have been realized in preliminary form by the ToxCast (Kavlock et al., 2012) and Tox21 research

programs (Tice et al., 2013; Thomas et al., 2018) which have generated publicly available HTS data on thousands of chemicals. In addition to the data generated, data processing pipelines have been developed (Hsieh et al., 2015; Filer et al., 2017) and many different models continue to be derived using the data, including those designed to understand mode-of-action (e.g., Shah et al., 2011; Judson et al., 2015; Kleinstreuer et al., 2016; Saili et al., 2019;) as well as models that use HTS data as descriptors or training information for (Q)SARs (e.g., Liu et al., 2015; Mansouri et al., 2016). The informatic needs of data driven predictive modeling, and how to standardize and openly transmit these models, is a clear need in the field. Recent progress in advancing computational toxicology and the associated challenges were discussed in Ciallela and Zhu (2019). Noteworthy examples of recent data driven models include those for acute oral toxicity (Russo et al., 2019) and liver toxicity (Zhao et al., 2020).

The databasing and informatic challenges for computational toxicology also have a legacy component: to bolster scientific confidence and for fit-for-purpose evaluations, the use of *in vivo* animal study data and any available *in vivo* human data have been important in the early application of computational toxicology as replacements or alternatives to existing approaches (Kleinstreuer et al., 2016, uterotrophic database; Hoffmann et al., 2018, LLNA database; Watford et al., 2019b; ToxRefDBv2). To enable quantitative comparisons of dose in animals or humans, internal exposures as modeled using HT toxicokinetic data and modeling (Wetmore et al., 2012; Pearce et al., 2017; Wambaugh et al., 2018) have been developed to support *in vitro* to *in vivo* extrapolation (IVIVE). Examples of how IVIVE has enabled greater utilization of HTS data for safety assessment are discussed in more detail in Thomas et al. (2019), Paul Friedman et al. (2020).

CURRENT STATE OF COMPUTATIONAL TOXICOLOGY AND INFORMATICS

The Rising Need of Informatics and Data Engineering

The validation principles of QSARs are perhaps still relevant today but in providing a framework to make explicit the provenance of the data, how it has been processed, the assumptions made, and the transparency and reproducibility of any models derived (Patlewicz et al., 2015). The three pillars of statistics, toxicology, and chemistry have since extended, in part due to the demand to make rapid decisions (Judson et al., 2010), with greater transparency. Perhaps the term "Data Science" now better captures the skill sets and needs encompassed in computational toxicology. Thus, computational toxicology considers the disciplines of toxicology, chemistry, and statistics, but also a number of front-end data science techniques relying on programming skills to facilitate data acquisition, data processing, storage and retrieval, data manipulation and interpretation, and beyond traditional statistics, other machine learning and deep learning techniques. The wealth of open source tools has also facilitated the change in the skills and approaches now applied. Commercial bespoke tools are being somewhat superseded by open source libraries developed on top of programming languages such as R and python. A skill set that so far has not been a strong focus as yet is that of data engineer, the backend of data science: models developed need to be deployed, and for reproducible models, a different set of considerations regarding versioning of models, their underlying inputs, and algorithms e.g., docker containers.

Evaluating Fit-For-Purpose Utility

The variety and volume of data now being generated and analyzed has also raised challenging questions for the "legacy" or existing *in vivo* data available: the level of curation, study reproducibility, and how these data may be used to benchmark new approach methodologies are all of high interest (Pham et al., 2019) Using *in vivo* study data to benchmark the performance of or directly train new approach methodologies for human or ecological health assessment should include some evaluation of how variable the *in vivo* study data may have been. Fit-for-purpose evaluations require not only acquisition of and curation of reference data and meaningful assessments of variability and uncertainty, but also efforts to increase data interoperability (Watford et al., 2019a). That requires ontologies in data storage and domain knowledge, as well as standards that permit sharing and exchange of data and models. Another consideration is the fact that these new data stream technologies are evergreen and in constant state of evolution and improvement. Evaluation of the fitness of these information needs to be flexible to deal with the changes in the methods of a specific technology and increased understanding of method performance using a large number of substances (Judson et al., 2018; Ciallela and Zhu, 2019).

The concept of an "applicability domain" is central in an evaluation of fit-for-purpose use of computational toxicology approaches, taking on an extended meaning as we want to understand the relevance of the model and when it can be applied and the extent to which it can be used to forecast other substances and to what extent there is confidence for that to occur. The uncertainties associated with the prediction need to be clearly specified and linked back to the decision context and purpose intended. The appropriate measures of fit and predictivity remain important considerations. What are the steps and procedures that were applied during the model building phases, including selection of approach, cross validation, performance metrics, and hyperparameter optimization. Before final model evaluation and application to new data (prediction), consideration of how to deploy a model should include a plan for ensuring reproducibility.

And Now: A Call for Research and Action Using the QSAR Validation Principles as a Guide

Clearly the landscape of computational toxicology has significantly evolved in the last two decades and much progress has already been made. Notable examples include data challenges that have been organized by NCATS (e.g., Huang and Xia, 2017) and NICEATM (e.g. Kleinstreuer et al., 2018). Most progress has been realized for individual discrete substances

but major areas of effort that remain to be tackled include: (1) the challenge of big data itself e.g., how to fit for large datasets (Ciallela and Zhu, 2019) (2) difficult substances to test in existing HTS systems, e.g., volatiles, insoluble in solvents; (3) mixtures–to date progress has been made on developing individual models but less focus on ensemble models; (4) wide implementation of cloud resources for data accessibility and data processing; (5) metabolism and degradation aspects in inferring effects of parent chemicals; (6) the use of unbalanced datasets in model training and development; (7) predicting dose response in conjunction with effects rather than extracting a summary metric from a study (Moran et al., 2019); (8) mining and extraction of insights from unstructured literature data; (9) standardized application of epidemiology; and likely a myriad of other challenges yet to be identified.

In many respects the (Q)SAR validation principles from 2004 remain relevant. The defined endpoint of a model or new approach methodology, the purpose and the goal of the model, and its basis need to be specified, albeit characterized differently to meet the requirements of 2020 and beyond. "The unambiguous nature of the algorithm" reads now as a call for increased reproducibility of the methodology and the approach. Are the assumptions of the modeling approach and the underlying data clearly specified? What are the data processing steps taken? How has the data been generated, summarized, and stored? These and other considerations should feature prominently in *Computational Toxicology and Informatics*.

AUTHOR CONTRIBUTIONS

GP prepared and wrote this article.

ACKNOWLEDGMENTS

REFERENCES

Adhikari, C., and Mishra, B. K. (2018). Quantitative structure-activity relationships of aquatic narcosis: a review. *Curr. Comput. Aided Drug Des.* 14, 7–28. doi: 10.2174/1573409913666170711130304

Ashby, J., and Tennant, R. W. (1988). Chemical structure, *Salmonella* mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested in rodents by the U.S. NCI/NTP. *Mutat. Res.* 204, 17–115. doi: 10.1016/0165-1218(88)90114-0

Benigni, R., and Bossa, C. (2019). Data-based review of QSARs for predicting genotoxicity: the state of the art. *Mutagenesis* 34, 17–23. doi: 10.1093/mutage/gey028

Ciallela, H. L., and Zhu, H. (2019). Advancing computational toxicology in the Big data era by artificial intelligence: data-driven and mechanism-driven modelling for chemical toxicity. *Chem. Res. Toxicol.* 32, 536–547. doi: 10.1021/acs.chemrestox.8b00393

Cronin, M. T., Jaworska, J. S., Walker, J. D., Comber, M. H., Watts, C. D., and Worth, A. P. (2003a). Use of QSARs in international decision-making frameworks to predict health effects of chemical substances. *Environ. Health Perspect.* 111, 1391–1401. doi: 10.1289/ehp.5760

Cronin, M. T., Walker, J. D., Jaworska, J. S., Comber, M. H., Watts, C. D., and Worth, A. P. (2003b). Use of QSARs in international decision-making frameworks to predict ecologic effects and environmental fate of chemical substances. *Environ. Health Perspect.* 111, 1376–1390. doi: 10.1289/ehp.5759

European Commission (2006). Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. *Off. J. Eur. Union* L136:3. Available online at: http://data.europa.eu/eli/reg/2006/1907/2014-04-10

European Commission (2009). Regulation (EC) No 1223/2009 of the European Parliament and the Council of 30 November 2009 on cosmetic products. *Off. J. Eur. Union* L342:59. Available online at: http://data.europa.eu/eli/reg/2009/1223/oj

Filer, D. L., Kothiya, P., Setzer, R. W., Judson, R. S., and Martin, M. T. (2017). tcpl: the ToxCast pipeline for high-throughput screening data. *Bioinformatics* 33:618. doi: 10.1093/bioinformatics/btw680

Harrill, J., Shah, I., Setzer, R. W., Haggard, D., Auerbach, S., Judson, R., et al. (2019). Considerations for strategic use of high-throughput transcriptomics

chemical screening data in regulatory decisions. *Curr. Opin. Toxicol.* 15, 64–75. doi: 10.1016/j.cotox.2019.05.004

Hoffmann, S., Kleinstreuer, N., Alépée, N., Allen, D., Api, A. M., Ashikaga, T., et al. (2018). Non-animal methods to predict skin sensitization (I): the Cosmetics Europe database. *Crit. Rev. Toxicol.* 48, 344–358. doi: 10.1080/10408444.2018.1429385

Hsieh, J. H., Sedykh, A., Huang, R., Xia, M., and Tice, R. R. (2015). A data analysis pipeline accounting for artifacts in Tox21 quantitative high-throughput screening assays. *J. Biomol. Screen.* 20, 887–897. doi: 10.1177/1087057115581317

Huang, R., and Xia, M. (2017). Editorial: Tox21 Challenge to build predictive models of nuclear receptor and stress response pathways as mediated by exposure to environmental toxicants and drugs. *Front. Environ. Sci.* 5:3 doi: 10.3389/fenvs.2017.00003

Judson, R. S., Magpantay, F. M., Chickarmane, V., Haskell, C., Tania, N., Taylor, J., 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. doi: 10.1093/toxsci/kfv168

Judson, R. S., Martin, M. T., Reif, D. M., Houck, K. A., Knudsen, T. B., Rotroff, D. M., et al. (2010). Analysis of eight oil spill dispersants using rapid, *in vitro* tests for endocrine and other biological activity. *Environ. Sci. Technol.* 44, 5979–5985. doi: 10.1021/es102150z

Judson, R. S., Thomas, R. S., Baker, N., Simha, A., Howey, X. M., Marable, C., et al. (2018). Workflow for defining reference chemicals for assessing performance of *in vitro* assays. *ALTEX* 36, 261–276. doi: 10.14573/altex.1809281

Kavlock, R., Chandler, K., Houck, K., Hunter, S., Judson, R., Kleinstreuer, N., 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. doi: 10.1021/tx3000939

Kleinstreuer, N. C., Ceger, P. C., Allen, D. G., Strickland, J., Chang, X., Hamm, J. T., et al. (2016). A curated database of rodent uterotrophic bioactivity. *Environ. Health Perspect.* 124, 556–562. doi: 10.1289/ehp.1510183

Kleinstreuer, N. C., Karmaus, A., Mansouri, K., Allen, D. G., Fitzpatrick, J. M., et al. (2018). Predictive models for acute oral systemic toxicity: a workshop to bridge the gap from research to regulation. *Comput. Toxicol.* 8, 21–24. doi: 10.1016/j.comtox.2018.08.002

Kroes, R., Renwick, A. G., Cheeseman, M., Kleiner, J., Mangelsdorf, I., Piersma, A., et al. (2004). Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. *Food Chem. Toxicol.* 42, 65–83. doi: 10.1016/j.fct.2003.08.006

Liu, J., Mansouri, K., Judson, R. S., Martin, M. T., Hong, H., Chen, M., et al. (2015). Predicting hepatotoxicity using ToxCast *in vitro* bioactivity and chemical structure. *Chem. Res. Toxicol.* 28, 738–751. doi: 10.1021/tx500501h

Mansouri, K., Abdelaziz, A., Rybacka, A., Roncaglioni, A., Tropsha, A., Varnek, A., et al. (2016). CERAPP: collaborative estrogen receptor activity prediction project. *Environ. Health Perspect.* 124, 1023–1033. doi: 10.1289/ehp.1510267

Moran, K. R., Dunson, D., and Herring, A. H. (2019). Bayesian joint modeling of chemical structure and dose response curves. *arXiv* arXiv:1912.12228. Available online at: https://arxiv.org/abs/1912.12228

Netzeva, T. I., Worth, A., Aldenberg, T., Benigni, R., Cronin, M. T., et al. (2005). Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52. *Altern. Lab. Anim.* 33, 155–173. doi: 10.1177/026119290503300209

Nikolova-Jeliazkova, N., and Jaworska, J. (2005). An approach to determining applicability domains for QSAR group contribution models: an analysis of SRC KOWWIN. *Altern. Lab. Anim.* 33, 461–470. doi: 10.1177/026119290503300510

NRC (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy.* Washington DC: National Academy Press.

NRC (2012). *Exposure Science in the 21st Century: A Vision and A Strategy.* Washington DC: National Academy Press.

NRC (2017). *Using 21st Century Science to Improve Risk-Related Evaluations.* Washington DC: National Academy Press.

Nyffeler, J., Willis, C., Lougee, R., Richard, A., Paul-Friedman, K., and Harrill, J. A. (2020). Bioactivity screening of environmental chemicals using imaging-based high-throughput phenotypic profiling. *Toxicol. Appl. Pharmacol.* 389:114876. doi: 10.1016/j.taap.2019.114876

OECD (2004). *ENV/JM/MONO/(2004)24.* Available online at: http://www.oecd.org/env/ehs/risk-assessment/37849783.pdf (accessed June 14, 2020).

OECD (2007). *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships [(Q)SAR] Models.* Series on Testing and Assessment No. 69. OECD Environment Health and Safety Publications.

Patlewicz, G., Cronin, M. T. D., Helman, G., Lambert, J. C., Lizarraga, L., and Shah, I. (2018). Navigating through the minefield of read-across frameworks: a commentary perspective. *Comput. Toxicol.* 6, 39–54. doi: 10.1016/j.comtox.2018.04.002

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. doi: 10.1016/j.yrtph.2015.02.011

Patlewicz, G., Worth, A. P., and Ball, N. (2016). Validation of Computational Methods. *Adv. Exp. Med. Biol.* 856:165. doi: 10.1007/978-3-319-33826-2_6

Paul Friedman, K., Gagne, M., Loo, L. H., Karamertzanis, P., Netzeva, T., Sobanski, T., et al. (2020). Utility of *in vitro* bioactivity as a lower bound estimate of *in vivo* adverse effect levels and in risk-based prioritization. *Toxicol. Sci.* 173, 202–225. doi: 10.1093/toxsci/kfz201

Pearce, R. G., Setzer, R. W., Strope, C. L., Wambaugh, J. F., and Sipes, N. S. (2017). httk: R package for high-throughput toxicokinetics. *J. Stat. Softw.* 79, 1–26. doi: 10.18637/jss.v079.i04

Pham, L., Sheffield, T. Y., Pradeep, P., Brown, J., Haggard, D. E., Wambaugh, J., et al. (2019). Estimating uncertainty in the context of new approach methodologies for potential use in chemical safety evaluation. *Curr. Opin. Toxicol.* 15, 40–47. doi: 10.1016/j.cotox.2019.04.001

Russo, D. P., Strickland, J., Karmaus, A. L., Wang, W., Shende, S., et al. (2019). Non animal models for acute toxicity evaluations: applying data-driven profiling and read-across. *Environ. Health Perspect.* 127:47001. doi: 10.1289/EHP3614

Saili, K. S., Franzosa, J. A., Baker, N. C., Ellis-Hutchings, R. G., Settivari, R. S., Carney, E. W., et al. (2019). Systems modeling of developmental vascular toxicity. *Curr. Opin. Toxicol.* 15, 55–63. doi: 10.1016/j.cotox.2019.04.004

Shah, I., Houck, K., Judson, R. S., Kavlock, R. J., Martin, M. T., Reif, D. M., et al. (2011). Using nuclear receptor activity to stratify hepatocarcinogens. *PLoS ONE* 6:e14584. doi: 10.1371/journal.pone.0014584

Thomas, R. S., Bahadori, T., Buckley, T. J., Cowden, J., Deisenroth, C., Dionisio, K. L., et al. (2019). The next generation blueprint of computational toxicology at the U.S. Environmental Protection Agency. *Toxicol. Sci.* 169, 317–332. doi: 10.1093/toxsci/kfz058

Thomas, R. S., Paules, R. S., Simeonov, A., Fitzpatrick, S. C., Crofton, K. M., Casey, W. M., et al. (2018). The US Federal Tox21 Program: a strategic and operational plan for continued leadership. *ALTEX* 35, 163–168. doi: 10.14573/altex.1803011

Thomas, R. S., Philbert, M. A., Auerbach, S. S., Wetmore, B. A., Devito, M. J., Cote, I., et al. (2013). Incorporating new technologies into toxicity testing and risk assessment: moving from 21st century vision to a data-driven framework. *Toxicol. Sci.* 136, 4–18. doi: 10.1093/toxsci/kft178

Tice, R. R., Austin, C. P., Kavlock, R. J., and Bucher, J. R. (2013). Improving the human hazard characterization of chemicals: a Tox21 update. *Environ. Health Perspect.* 121, 756–765. doi: 10.1289/ehp.1205784

Tollefsen, K. E., Scholz, S., Cronin, M. T., Edwards, S. W., de Knecht, J., Crofton, K., et al. (2014). Applying Adverse Outcome Pathways (AOPs) to support Integrated Approaches to Testing and Assessment (IATA). *Regul. Toxicol. Pharmacol.* 70, 629–640. doi: 10.1016/j.yrtph.2014.09.009

Votano, J. R., Parham, M., Hall, L. H., Kier, L. B., Oloff, S., et al. (2004). Three new consensus QSAR models for the prediction of Ames genotoxicity. *Mutagenesis* 19, 365–377. doi: 10.1093/mutage/geh043

Wambaugh, J. F., Hughes, M. F., Ring, C. L., MacMillan, D. K., Ford, J., Fennell, T. R., et al. (2018). Evaluating *in vitro-in vivo* extrapolation of toxicokinetics. *Toxicol. Sci.* 163, 152–169. doi: 10.1093/toxsci/kfy020

Watford, S., Edwards, S., Angrish, M., Judson, R. S., and Paul Friedman, K. (2019a). Progress in data interoperability to support computational toxicology and chemical safety evaluation. *Toxicol. Appl. Pharmacol.* 380:114707. doi: 10.1016/j.taap.2019.114707

Watford, S., Ly Pham, L., Wignall, J., Shin, R., Martin, M. T., and Friedman, K. P. (2019b). ToxRefDB version 2.0: improved utility for predictive and retrospective toxicology analyses. *Reprod. Toxicol.* 89, 145–158. doi: 10.1016/j.reprotox.2019.07.012

Wetmore, B. A., Wambaugh, J. F., Ferguson, S. S., Sochaski, M. A., Rotroff, D. M., et al. (2012). Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. *Toxicol. Sci.* 125, 157–174. doi: 10.1093/toxsci/kfr254

Worth, A. P., Bassan, A., Gallegos, A., Netzetva, T. I., Patlewicz, G., Pavan, M., et al. (2005). The characterisation of (Q)uantitative Structure-Activity Relationships: preliminary guidance. EUR 21866EN. Available online at: https://publications.jrc.ec.europa.eu/repository/bitstream/JRC31241/QSAR%20characterisation_EUR%2021866%20EN.pdf (accessed June 14, 2020).

Zhao, L., Russo, D. P., Wang, W., Aleksunes, L. M., and Zhu, H. (2020). Mechanism-Driven Read-across of chemical hepatotoxicants based on chemical structures and biological data. *Toxicol. Sci.* 174, 178–188. doi: 10.1093/toxsci/kfaa005