



Supporting Read-Across Using Biological Data

Hao Zhu¹, Mounir Bouhifd², Elizabeth Donley³, Laura Egnash³, Nicole Kleinstreuer⁴, E. Dinant Kroese⁵, Zhichao Liu⁶, Thomas Luechtefeld², Jessica Palmer³, David Pamies², Jie Shen⁷, Volker Strauss⁸, Shengde Wu⁹ and Thomas Hartung^{2, 10}

¹Department of Chemistry, Rutgers University, Camden, NJ, USA; Center for Computational and Integrative Biology, Rutgers University, Camden, NJ, USA; ²Johns Hopkins Bloomberg School of Public Health, Center for Alternatives to Animal Testing (CAAT), Baltimore, MD, USA; ³Stemina Biomarker Discovery Inc., Madison, WI, USA; ⁴National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA; ⁵Risk Analysis for Products in Development, TNO Zeist, The Netherlands; ⁶US FDA, NCTR, Little Rock, Arkansas, USA; ⁷Research Institute for Fragrance Materials, Inc. Woodcliff Lake, NJ, USA; ⁸BASF Aktiengesellschaft, Experimental Toxicology and Ecology, Ludwigshafen, Germany; ⁹Procter & Gamble, Cincinnati, OH, USA; ¹⁰University of Konstanz, CAAT-Europe, Konstanz, Germany

Summary

Read-across, i.e., filling toxicological data gaps by relating to similar chemicals for which test data are available, is usually done based on chemical similarity. Besides structure and physico-chemical properties, biological similarity based on biological data adds extra strength to this process. In the simplest case, chemically similar substances also show similar test results in relevant *in vitro* assays. This is a well-established method for the read-across of, e.g., genotoxicity assays. Larger datasets of biological and toxicological properties of hundreds and thousands of substances are becoming available, enabling big data approaches in read-across studies. In the context of developing Good Read-Across Practice guidance, a number of case studies using various big data sources were evaluated to assess the contribution of biological data to enriching read-across. An example is given for the US EPA's ToxCast dataset which allows read-across for high quality uterotrophic assays for estrogenic endocrine disruption. Similarly, an example is given for REACH registration data that enhances read-across for acute toxicity studies. A different approach is taken using omics data to establish biological similarity: Examples are given for *in vitro* stem cell models and short-term *in vivo* repeated dose studies in rats used to support read-across and category formation. These preliminary biological data-driven read-across studies show the way towards the generation of new read-across approaches that can inform chemical safety assessment.

Keywords: read-across, biological similarity, safety assessment, big data

1 Introduction

Read-across has become a primary approach to filling data gaps for chemical safety assessments. Chemical similarity, based on structure, reactivity and physico-chemical property information, is the main approach followed for this purpose. Chemical toxicity, however, normally arises from complicated biological

mechanisms, so only using chemical similarity to justify the read-across will be error-prone, especially when chemically similar compounds show dissimilar toxicity phenomena. The availability of massive biological datasets for environmental compounds makes biological similarity approaches feasible within the scope of read-across. This paper, which accompanies the efforts toward Good Read-Across Practice (Ball et al.,

*A report of t4 – the transatlantic think tank for toxicology, a collaboration of the toxicologically oriented chairs in Baltimore, Konstanz and Utrecht sponsored by the Doerenkamp Zbinden Foundation. The views expressed in this article are those of the contributing authors and do not necessarily reflect those of their institution of employment.

Disclaimer: The information in these materials is not a formal dissemination of information by the U.S. Environmental Protection Agency or the National Institutes of Health, and does not represent the position or policy of any agency.





2016, this issue) scopes the opportunities for biological support to strengthen read-across based on a number of examples.

Three different types of biological similarity approaches have been applied: 1) When several bioassays represent key mechanisms of target toxicity endpoints, they may have straightforward predictive power and in some instances it has been possible to show this for a larger applicability domain by traditional validation. In some cases, however, biological similarity can be successfully applied only for specific parts of the chemical universe and only for one toxicity endpoint. This has been termed “local validity” (Patlewicz et al., 2014); 2) When biological similarity is based on a large number of bioassays, the read-across study may be successful for various types of toxicity endpoints. This kind of study is so far rarely pursued because of high costs to screen the same target compounds against many (often several hundred) different bioassays (Zhu et al., 2014). The recent public availability of large high-throughput screening (HTS) datasets (e.g., ToxCast and Tox21 data) has made this kind of study feasible. New statistical tools need to be applied for read-across studies using such complex biological data; 3) The concept of toxicity pathways (or the molecularly defined pathways of toxicity, PoT (Kleensang et al., 2014)) represents a new opportunity for risk assessment. When all targets within a potential pathway are included in the biological testing approach, mechanism-based read-across studies are feasible. In the current big data era, some “popular” compounds, such as well-known toxicants, have been extensively studied worldwide and a complex data landscape is available for them (e.g., omics data). To this end, more biological data is needed for target compounds, more toxicity mechanisms need to be clarified and novel computational tools (e.g., big data approaches) need to be developed. The recent efforts steered by OECD toward developing Adverse Outcome Pathways (AOP) have added much useful information and new tools to approach these challenges (Vinken et al., 2013a,b; Ankley et al., 2010). Biological similarity studies have been greatly enhanced by the rapidly increasing biological data (Zhu et al., 2014).

Omics analyses also allow assessing similarity; here *in vivo* models, such as short-term animal studies as well as stem cell-derived developmental and organ models lend themselves for comparison of signatures of toxicity. While test-across deviates from traditional methods only by acknowledging the small applicability domain of proven usefulness, the HTS and omics approaches are based on what is now called “big data”, i.e., curated large datasets for data-mining.

2 The state of the art of read-across using biological data

2.1 Empirical read-across studies using biological data

2.1.1 Moving from chemical structure information to biological data

Traditional read-across studies, mostly using quantitative structure-activity relationship (QSAR) approaches, are normally based on chemical structure information (Solimeo et al., 2012; Zhu et al., 2008, 2009; Schultz et al., 2003). Certain structural

fragments (e.g., structural alerts) (Klopman et al., 2004), physico-chemical properties (Klopman et al., 1999) or other molecular properties (e.g., molecular sizes) (Moss et al., 2002) are used to estimate the chemical toxicity potential. In contrast to these efforts, the early stage of using biological data in read-across normally uses limited biological data obtained from one or a few bioassays for small sets of compounds.

In the studies using only chemical information for large parts of the chemical universe, activity cliffs (i.e., small changes in structure inducing significant changes of toxicity) resulted in major prediction errors (Maggiora, 2006). For this reason, evaluation of the uncertainty of read-across, as well as of *in silico* methods (Hartung and Hoffmann, 2009) and *in vitro* assays, is crucial. Critically, the part of the chemical universe for which a given method is applicable needs to be defined. This means that reliable predictions can be made within a defined applicability domain (Hartung et al., 2004). Without knowing all parts of the chemical universe to which a method is applicable, it is often possible to demonstrate that a method works for a certain group of chemicals. The term “local validity” was introduced to describe this issue in read-across studies (Patlewicz et al., 2014). When applying biological data in read-across, it also should focus on the areas of local validity to carry out *in vitro* tests that represent key aspects of the pathophysiology, a concept earlier introduced as test-across (Hartung, 2007). Because of the recent emergence of the public toxicity data from the European Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) and the US Tox21 datasets, which notably resulted in a more than 1,700 overlap of chemicals (Luechtefeld et al., 2016a), this concept can now be empirically evaluated. These new efforts will help to move from a pragmatic use of weight-of-evidence to a quantitative biological data read-across with an associated measure of evaluation uncertainty (Linkov et al., 2015).

2.1.2 Case study: Developing bioassays for read-across evaluation of developmental toxicity

Substantial efforts have already been undertaken to develop alternative assays for the assessment of reproductive and developmental toxicity (Adler et al., 2011; Leist et al., 2014). Of these alternatives only a few have been formally validated for developmental toxicity, such as the whole embryo culture (WEC), the mouse embryonic stem cell test (mEST), and the mammalian micromass (MM) test (Pamies et al., 2011). While characterizing these assays, it was also recognized that none of them alone could cover the whole mammalian reproductive cycle due to its inherent complexity, covering male and female fertility, implantation, and embryonic development (Adler et al., 2011; Leist et al., 2014). Therefore, more recent studies have attempted to combine several *in vitro* assays into a test battery instead of applying individual assays. As part of a European FP6 project called ReProTect, a feasibility study was performed, in which ten compounds were studied in a battery of 14 assays (Schenk et al., 2010). This battery, which consisted of several assays detecting endocrine disruption (androgen receptor (AR) and estrogen receptor (ER) binding assays, and ARE and ERE promoter driven reporter assays), 3 tests detecting embryotoxicity (mEST, WEC, and ReProGlo), and several assays detecting adverse ef-

fects on male and female fertility (mouse follicle bioassay, bovine maturation and fertilization assays, mouse peri-implantation assay, and Ishikawa test, which measures mRNA levels of progesterone receptor). This battery was able to detect all reproductive toxicants for which the modes of action were actually represented in at least one of the assays. In a subsequent study of the European FP7 project ChemScreen, this battery approach was developed further (Piersma et al., 2013). The zebrafish embryo test (ZET) and the embryonic stem cell test (EST) were included as more apical assays to detect effects on the development of a whole egg from fertilization until the hatching stage 72 h later (Hermsen et al., 2011), and to detect effects on cellular differentiation of cardiomyocytes (Scholz et al., 1999), respectively. Again the ReProGlo assay, monitoring interference with the WNT pathway (Uibel et al., 2015), and assays for CYP17 and CYP19, to detect effects on steroidogenesis enzymes essential for reproductive hormone homeostasis (van Duursen et al., 2010; Hecker et al., 2011) were included. Finally, a panel of 24 high-throughput CALUX assays were added to measure changes in the activity of key transcription factors, varying from nuclear receptors (e.g., reproductive hormone receptors) to transcription factors involved in cellular signaling (Sonneveld et al., 2005; van der Burg et al., 2013). The approach also encompassed toxicokinetic modeling to reveal whether effective *in vitro* concentrations observed in the battery were in the range expected from the *in vivo* reproductive toxicity data, in line with suggestions by Daston et al. (2010). The ChemScreen battery approach (including the toxicokinetic model) successfully identified eleven out

of twelve compounds with varying mechanisms of action, while the missed compound, glufosinate ammonium, had a mechanism not covered by the battery (Piersma et al., 2013). This result encourages further optimization of this battery into one ultimately able to detect all reprotoxic compounds.

The use and interpretation of battery results depends strongly on the purpose of testing and the information that may already be available. In the absence of any *in vivo* test information relevant to potential reproductive toxicity and/or in the absence of any structural alerts pointing to such effects, the battery could be applied as a filter to optimize and/or reduce testing of potential reproductive toxicants in animal studies (Wu et al., 2013). The battery could also help prioritize chemicals for further investigation and/or to select candidate compounds (e.g., drug candidates) for further development (van der Burg et al., 2011). On the other hand, when there are clear indications for potential reproductive toxicity based on close structural similarity of a query chemical to a reproductive toxicant, the battery could be used to confirm any reproductive toxicity and to avoid any further *in vivo* studies.

Ideally, this battery should then also be capable of correctly distinguishing reproductive toxicants from non-reproductive toxicants, even when both have high structural similarity. This has been investigated by Kroese et al. (2015) via testing three groups of structurally-related chemicals differing in their reproductive toxicity: two valproic acid (VPA) analogs, i.e., 2-ethylhexanoic acid (EHA) and 2-methylhexanoic acid (MHA), two analogs of monoethylhexyl phthalate (MEHP), i.e., monobenzyl

Tab. 1: General information and properties of the analogs

Property	Flusilazole	Hexaconazole	Propiconazole	Triadimefon	Myclobutanil
Use	Fungicide/ antibacterial drug	Fungicide	Fungicide	Fungicide	Fungicide
Structural representation					
CASRN	85509-19-9	79983-71-4	60207-90-1	43121-43-3	88671-89-0
Molecular weight (g/mol)	315.3927	313.0749	341.0698	293.0931	288.1142
Physical state at 20°C & 101.3 kPa	Solid	Solid	Liquid	Solid	Solid
Water solubility (mg/l) at 20 or 25°C	43	1.29	100	71.5	142
Log P (octanol-water)	4.68	3.7	3.7	2.8	2.9



Tab. 2: Summary of test article ADME and toxicity data

	Flusilazole ^a	Hexaconazole ^b	Propiconazole ^c	Triadimefon ^d	Myclobutanil ^e
ADME properties					
<i>In vivo</i> half-life (plasma/serum)	NA	NA	24-31 h	~4 h	Biphasic rapid phase: 5.25 Slow phase: 25.7
Rate and extent of oral absorption	Rapid & extensive (up to 80%)	NA	> 80% in 48 h	28% in females, 67% in males as urinary excretion	Rapidly absorbed (> 89%)
Distribution	Widely distributed	Widely distributed; highest concentrations in liver, intestinal tract and adrenal cortex	Widely distributed; highest concentrations in the liver and kidney	Widely distributed in kidneys and liver	Widely distributed
<i>In vivo</i> metabolism	Extensive	Extensive	Extensive	Rapid & extensive	Rapid & extensive
Most active CYPs	NA	NA	NA	CYP2C and CYP3A	CYP2C and CYP3A
Excretion*	96 h	72 h	24 h	96 h	96 h
Primary route of excretion	urine	43% urine/ 53% feces (m) 66% urine/ 29% feces (f)	Even distribution in urine & feces	Feces (m) urine (f)	Even distribution in urine & feces
Toxicological data					
Acute toxicity, oral, LD ₅₀ (mg/kg bw)	674	2189	1517	363-1855	1600
Genotoxicity	Negative	Negative	Negative	Negative	Negative
Short-term toxicity studies					
Target	Liver and urinary bladder	Liver	Body weight, liver, erythrocytes	Liver	Liver
Oral NOAEL (mg/kg bw per day)	9	2.5	76	150	51.5
<i>In vivo</i> developmental toxicity studies					
Target/critical effect	Skeletal anomalies, malformations at higher doses	Fetal toxicity, skeletal variations	Skeletal variations	Skeletal variations	Fetal toxicity/ increased number of early resorptions and lower fetal weights
Developmental toxicity NOAEL** (mg/kg bw per day)	2	2.5	30	30	93.8
Maternal toxicity NOAEL (mg/kg bw per day)	10	25	90	10	93.8
devTOX^{QP} results					
dTP (μM)	17	22	26	35	51

Note: NA: Data not available. *In vivo* data summarized from rat studies.

*Excretion is greater than or equal to 90% of radiolabel.

**Developmental toxicity includes embryo/fetal toxicity and teratogenicity.

^aAdcock and Tasheva, 2009; ^bden Tonkelaar and van Kotten-Vermeulen, 1991; ^cDewhurst and Dellarco, 2006; ^dZarn et al., 2006;

^eYoshida and McGregor, 2015



phthalate (MBzP) and monomethyl phthalate (MMP), and three organotin analogs, i.e., tributyltin chloride, dibutyltin dichloride (TBTC), dibutyltin chloride dibutyltin dichloride (DBTC) and monobutyltin trichloride (MBTC). MHA and MMP were not considered reproductive toxicants, while MBTC was considered a weak reproductive toxicant. The battery correctly distinguished MMP and MBTC as non- or weak developmental toxicants. MHA was identified as a weak developmental toxicant, while it is negative in *in vivo* studies. However, available toxicokinetic data for MHA clearly show lower predicted fetal concentrations as compared to those of VPA and EHA, and at these lower concentrations the battery showed no developmental toxicity for MHA. This result clearly shows the relevance of toxicokinetic information for any assessment based on alternative *in vitro* models (Bosgra and Westerhout, 2015).

Such *in vitro* batteries can be combined with *in silico* tools, bioassays in lower species or even short-term *in vivo* tests. An important improvement lies in combining these batteries of tests with algorithms for optimizing the employment and interpretation of the different components. Such integrated testing strategies (ITS) (Hartung et al., 2013; Rovida et al., 2015) hold promise for the *in vitro* prediction of complex endpoints. However, they represent enormous challenges for validation of such a battery. Before such validation efforts are actually possible, test-across approaches for groups of chemicals similar to the examples given above already could be carried out, where non-tested compounds are profiled together with chemically similar tested ones.

Besides the above efforts, other complex bioassays representing broad biological processes possibly disturbed by a given chemical can serve to profile substances. The devTOX quickPredict (devTOXqP) assay was developed around the principal that toxicity is a function of exposure. The assay uses human embryonic or induced pluripotent stem (hPS) cells to predict a test article's developmental toxicity potential based on changes in the metabolites ornithine and cystine (Palmer et al., 2013). Changes in these metabolites are measured in response to treatment and then used in a ratio (o/c ratio) across an 8-point dose response curve. The developmental toxicity potential (dTP) is the interpolated exposure level (concentration) of a test article where the dose response curve crosses a defined developmental toxicity threshold (dTT). Exposure levels greater than this concentration are associated with developmental toxicity. The assay was 85% accurate in predicting the developmental toxicity potential of 80 compounds with a broad range of chemotypes (89% specificity, 82% sensitivity). The data generated with the devTOXqP assay presents an opportunity to include an *in vitro* human endpoint in read-across or weight of evidence approaches.

A set of five structurally similar triazole fungicides (flusilazole, hexaconazole, propiconazole, triadimefon and myclobutanil) were evaluated in the devTOXqP assay to demonstrate how the assay can be used to strengthen read-across and weight of evidences approaches (Tab. 1). Myclobutanil was used as an example compound with an uncertain developmental toxicity profile. Flusilazole was the most potent chemical *in vivo* and in

the devTOXqP assay, with a developmental toxicity potential at 17 μ M (Tab. 2). Myclobutanil had a developmental toxicity potential similar to its analogs and was classified as a developmental toxicant, which is consistent with published *in vivo* data (Tab. 2). Additionally, myclobutanil had the highest NOAEL for developmental toxicity *in vivo*, which is consistent with it being the least potent in the devTOXqP assay. Taken as part of a weight of evidence approach, this human data point would help to define potential for risk.

2.2 Using big data to establish chemical profile for biological similarity read-across studies

2.2.1 Available sources of biological data

The term “big data” describes a collection of datasets that are so large and complex that the data is too difficult to process by traditional data analysis tools. Modern toxicity research has moved into the big data era as massive biological data for compounds of interest (e.g., toxicants) became available (Zhu et al., 2014). There are two major sources of biological data: One biological data source is HTS of large libraries of compounds in toxicity studies. There has been a huge increase in the number of compounds and associated testing data in different *in vitro* screenings. Besides that, there are also efforts to curate historical *in vivo* toxicity data to share with the public. Table 3 shows some examples of these data collections distributed through various data sharing programs.

PubChem is a public repository for chemical structures and their biological properties (Wang et al., 2009, 2010). Bioactivity data in PubChem were contributed by hundreds of institutes, research laboratories, and specifically those screening centers under the NIH Molecular Libraries Program (MLP) (Austin et al., 2004). For example, the NIH Chemical Genomics Center (NCGC) was created in 2005 as a comprehensive screening center in the NIH MLP (Thomas et al., 2009). The mission of the NCGC is to apply the tools of small molecule screening and discovery to toxicology studies. Every year the NCGC generates millions of toxicity bioassay data points by testing thousands of diverse compounds and shares all the data with the research community via PubChem. The unique quantitative high throughput screening (qHTS) technique developed and optimized by the NCGC generates data in high quality and standardized form (Inglese et al., 2006). Another large reservoir of toxicity bioassay data in PubChem comes from the European Bioinformatics Institute (EBI) (ChEMBL¹). The EBI's goal is to provide freely available data and bioinformatics services to all parts of the scientific community. As part of this goal, the ChEMBL database was constructed for screening data of both chemical toxicity and absorption, distribution, metabolism and excretion (ADME) properties. ChEMBL version 11 (ChEMBL_11) was launched in 2011. It includes 3.3 million bioassay measurements covering 629,943 compounds (Gaulton et al., 2012). This was obtained from curating over 42,500 scientific publications.

The ToxCast program of the US EPA was initiated to identify alternatives to animal models (Reif et al., 2010; Dix et al.,

¹ www.ebi.ac.uk/chembl/index.php

**Tab. 3: Public databases of toxicity data**

Name	General information	Data description
PubChem	Over 50 million compounds, over 700,000 bioassays, over 13 billion data points	Toxicity, genomics and literature data
ChEMBL	Over 600,000 compounds, 3.3 million bioassay readout data	Literature toxicity data
ACToR	Toxicity results from 100 various data resources	Both <i>in vitro</i> and <i>in vivo</i> toxicity data
ToxNET	Over 50,000 environmental compounds from 16 different resources	Both <i>in vitro</i> and <i>in vivo</i> toxicity data
SEURAT	Over 5,500 cosmetic-type compounds in the current COSMOS database web portal	Animal toxicity data
REACH	816,048 studies for 9,800 substances and 3,600 study types	Data submitted in EU chemical legislation, made machine-readable by Luechtefeld et al., 2016a
CTD	Over 13,000 compounds, over 32,000 genes, over 6000 diseases	Compound, gene and disease relationships
CEBS	About 10,000 toxicity bioassays from various sources	Gene expression data
DrugMatrix	About 600 drug molecules and 10,000 genes	Gene expression data
Cmap	About 1,300 compounds and 7,000 genes	Gene expression data

2007). For this purpose, this program intentionally tested compounds with rich animal toxicity information, to generate a database containing both *in vitro* and *in vivo* toxicity data. Currently the ToxCast data, along with animal toxicity data, is shared via the Aggregated Computational Toxicology Resource (ACToR²) (Judson et al., 2008, 2012) portal. Similarly, but different from ACToR, ToxNET³ contains and allows navigation through 16 separate databases of much more diverse chemicals (Fonger et al., 2000). ToxNET was developed by the National Library of Medicines' (NLM) Division of Specialized Information Services (SIS). By grouping the databases together, ToxNET allows for all information to be accessed from one query form. Although the databases are separate, some are very similar and are grouped together in the example report.

In response to the shortage of alternative testing methods, the European Commission and the European Cosmetics Trade Association, Cosmetics Europe, launched the 5-year research initiative called Safety Evaluation Ultimately Replacing Animal Testing (SEURAT-1⁴) in 2011 (Vinken et al., 2012). It is called "SEURAT-1", indicating that more steps have to be taken before the ultimate goal of full animal replacement will be reached. Under the SEURAT-1 initiative, five research projects and one coordinating project were funded and extensive data curation/management work was involved (Kohonen et al., 2013). For example, one of these projects, the COSMOS project, was dedicated to the development of freely available tools and workflows to predict the safety of cosmetic ingredients in humans (Yang et al., 2013). The COSMOS database web portal⁵ contains over 5,500 unique cosmetic-relevant compounds with their respec-

tive *in vivo* toxicity data. A similar effort is the recent curation of REACH toxicity data (Hartung, 2010; Hengstler et al., 2006) from the publicly available registration summary data (Luechtefeld et al., 2016a-d, this issue).

Another rapidly growing area of interdisciplinary research generating big data is toxicogenomics (TGx), which aims to study the underlying molecular mechanisms of toxicity and address challenges that are difficult to overcome by conventional toxicology methods by integrating genomic technology with bioinformatics. Toxicogenomics is a field of toxicology that addresses information concerning gene expression changes, and in extension also protein and metabolite changes (Bouhifd et al., 2013; Ramirez et al., 2013), within a particular cell or tissue of an organism in response to chemicals. It has to be noted that transcriptomics is certainly most advanced among the omics technologies (van Vliet, 2011) with regard to standardization and quality assurance, but other omics technologies such as metabolomics are catching up (Bouhifd et al., 2015). Many modern *in vitro* toxicity studies now address relevant toxicity mechanisms, and these findings can be translated into biomarkers that could be applied to human exposure studies (McHale et al., 2010; Blaauboer et al., 2012). Several extensive publicly available TGx databases based on good experimental designs, such as the Japanese Toxicogenomics Project (TGP) (Uehara et al., 2010) and PredTox (Suter et al., 2011), provide enormous opportunities to evaluate and investigate a large set of TGx assays systematically, which provides a landscape of TGx and a more objective understanding of toxicity mechanisms. TGx investigations generate enormous amounts of "omics" data that are meant to predict toxicity or ge-

² <http://actor.epa.gov/actor>

³ <http://toxnet.nlm.nih.gov>

⁴ <http://www.seurat-1.eu>

⁵ <http://cosmosdb.cosmostox.eu>

netic susceptibility induced by chemicals. The Chemical Effects in Biological Systems (CEBS) database developed by the NIEHS is now the public repository for all NTP conventional toxicology and carcinogenicity data as well as NCGC HTS data (Waters et al., 2008) along with the Comparative Toxicogenomics Database (CTD) at North Carolina State University (Davis et al. 2014), which aims to promote comparative studies of genes and proteins across species (Mattingly et al., 2003, 2004, 2006a,b). CTD data is searchable through the ToxNET portal. Similar efforts in toxicogenomics data curation, but with a more specific research goal, are DrugMatrix (Ganter et al., 2005) and Cmap (Lamb et al., 2006).

2.2.2 Evaluating biological similarity based on big data

Although the read-across is based on the hypothesis that similar structures have similar toxicological profiles, the information derived from the chemical structure usually is limited. Therefore, information regarding the biological properties of the chemicals, both target and analogs, is the key support to read-across. Biological similarity refers to similar results from one or more assays for two chemicals. One of the approaches is to use the results from a large number of assays, usually high throughput assays, to profile the biological fingerprint of a chemical (Kim et al., 2016; Zhang et al., 2014; Sipes et al., 2013). If two chemicals have similar bioprofiles, they will be considered to be biologically similar (Low et al., 2013; Zhang et al., 2014).

However, it is not easy to apply this in any real case of read-across. First, it requires comprehensive information from toxicogenomics studies and/or high-throughput assays for both target and analog.

It is unusual that both target and analog have been tested in the same toxicogenomics studies and/or high-throughput assays. If we generate the data required as needed, the costs may not be much less than just testing the chemical of interest for any specific toxicity endpoint. Second, irrelevant information might be included when conducting read-across for a specific toxicity endpoint. For example, if using read-across to fill the data gap for skin sensitization, endocrine system related *in vitro* assays are not relevant to this endpoint. Therefore, one should be cautious of using a universal bioprofile to evaluate the biological similarity to support an endpoint specific read-across.

Another approach is based on the understanding of the mechanism of the specific toxicity, which is using one or a few closely related bioassays to compare the biological similarity. For example, one can use the Direct Peptide Reactivity Assay (DPRA) and KeratinoSens assays to profile the biological similarity for skin sensitization, or use the BlueScreen assay to profile the biological similarity for genotoxicity. Assays that map to the estrogen receptor pathway can be used to define biological similarity for potential endocrine disrupting compounds. It is worth mentioning that biological similarity should serve as a weight of evidence to evaluate the read-across, but structural similarity usually will be the first tier for similarity criteria.

In the current big data era, the bioassay response profile can be very large for some compounds (e.g., well-known toxicants) (Zhu et al., 2014). If all the public data for these compounds are used to create a profile, the initial profile can be large, complex and unorganized. For example, Figure 1 shows the PubChem re-

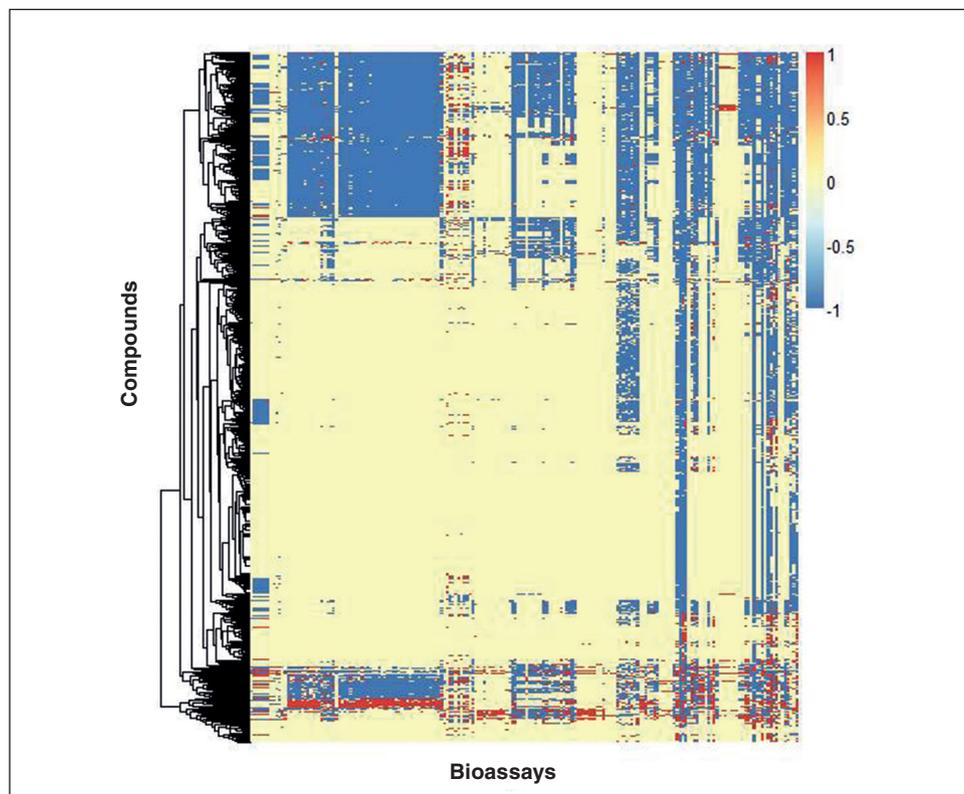
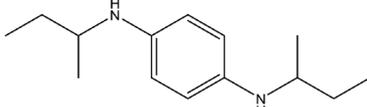
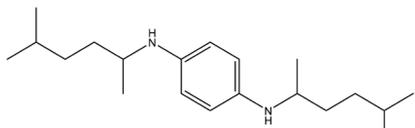
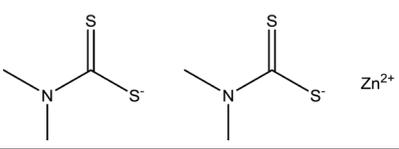
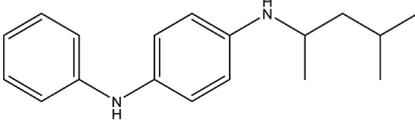
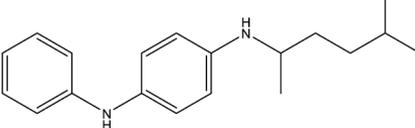
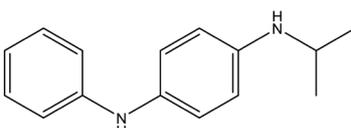
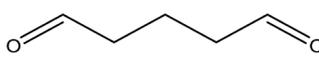
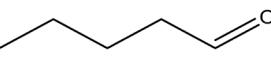
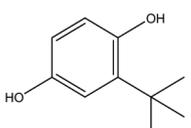


Fig. 1: The response space of 7,581 ToxCast and Tox21 compounds represented by the data obtained from 483 PubChem bioassays

The red dots represent active responses; the blue dots represent inactive responses, and the yellow dots represent no available testing data or inconclusive results.



Tab. 4: Three REACH compounds (the first compound) with their chemical nearest neighbor (the second compound) and biological nearest neighbor (the third compound)

Compounds		LD ₅₀ (mg/kg)	Bioprofiles*
1		181	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
		730	N/A**
		320	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2		949	1 1 1 -1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
		2,100	N/A
		520	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
3		206	-1 -1 -1 -1 1 -1 -1 1 1 1 1 -1 1 -1 -1 -1
		6,490	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
		1,041	-1 -1 1 -1 -1 1 -1 -1 1 1 1 1 -1 1 1 1 -1 -1

*The bioprofile consists of 18 PubChem assays (PubChem assay AIDs 427, 542, 544, 545, 546, 921, 963, 964, 966, 968, 973, 974, 993, 504832, 651802, 686979, 743041, 743086), which were selected for calculation since they contain the largest number of active responses per assay in REACH compounds. The red color indicates active response, blue color indicates inactive response and white color indicates no data available.

** N/A indicates there is no data available for this compound within these assays.



sponse space of 7,581 ToxCast and Tox21 compounds by using 483 PubChem assays (accessed December 2015). So the public resources shown in Table 3 contain lots of biological data that will be useful for read-across purposes.

It is understandable that most areas within the initial response map are either “no testing” or “inconclusive” because many bioassays have only been applied to a small portion of this large chemical set. Furthermore, the nature of HTS assays, many of which represent specific interactions, results in a biased distribution of responses for the target chemicals (many more “inactives” than “active” data entries). Since not all the bioassay data are relevant or useful for a particular type of toxicity, in the big data scenario, the most critical issue is to identify useful *in vitro* data. In principle, this could be done by a human expert using the knowledge of the design and quality of each particular bioassay (e.g., the “Confidence Score” assigned during manual curation to each assay in ChEMBL). However, in the big data era, the data-driven approaches should be developed preferably by fully automatic techniques. We recently developed an automatic bioassay system to evaluate and extract the relevant bioassay data based on the *in vitro-in vivo* relationship (Zhang et al., 2014; Wang et al., 2015; Kim et al., 2014). Using this approach, we analyzed the current REACH compounds with their rat oral acute toxicity. Table 4 shows three REACH compounds with their chemical nearest neighbors in the same set. It is obvious that these three pairs of chemicals’ nearest neighbors have quite different acute toxicity results. Any read-across approaches only based on molecular structures will not be able to differentiate them. Therefore, these activity cliffs in the REACH dataset will prompt prediction errors of any QSAR model.

We can integrate public bioassay results for these compounds as extra information for read-across purposes. By searching the PubChem portal using an in-house profiling tool, hundreds of PubChem assays containing experimental data for REACH compounds were automatically extracted (Luechtefeld et al., 2016a, this issue). These experimental biological data can be viewed as extra descriptors and a similarity search can be applied by using these data as bioprofiles. This way, the biological nearest neighbors for the three target compounds can be found within the REACH dataset, as shown as the third compound in each group of Table 4. It is clear that the biological nearest neighbors have much more similar acute toxicity results to the three target compounds than to the chemical nearest neighbors, indicating the value of using these biological data in the read-across procedure. This effort provides a potential solution to the pitfall of applying QSAR as read-across approaches with these activity cliff issues.

2.2.3 Case study: Using complex high-throughput biological data to support read-across – BioActivity-based read-across (BaBRA) using ToxCast Data

As highlighted above, the advent of high-throughput screening and research initiatives such as Tox21 and ToxCast provide data on a range of targets and pathways that may be linked to tox-

icity (Judson et al., 2010; Betts, 2013). The Tox21 program has screened over 8,000 chemicals in approximately 60 assays, and ToxCast testing includes a much broader range of assays, with around 800 targets, on a lower number of chemicals (~2,000). The ToxCast dataset⁶ in particular affords a unique opportunity to attempt BioActivity Based Read-Across (BaBRA), due to the wide coverage of biological space and range of assays from different cell types, species and technology platforms. A number of predictive models have identified critical pathways, such as embryonic vascular development, and characterized similar chemical activity against the identified molecular targets as a way to prioritize chemicals for their potential to cause toxicity, e.g., developmental defects (Kleinstreuer et al., 2011; Knudsen and Kleinstreuer, 2012). Supervised analyses such as support vector machines (SVM) have successfully predicted mechanisms such as phosphodiesterase inhibition and glucocorticoid receptor agonism for unknown chemicals based on similar protein expression profiles in primary human cells (Kleinstreuer et al., 2014). Others have described an approach that is pathway-agnostic and more closely resembles traditional structure-based read-across, with the addition of all available *in vitro* assay data as features to determine biological similarity (Low et al., 2013; Kim et al., 2016). Here we examined analogous approaches, both encompassing and pathway-specific, with a novel mathematical definition of similarity that used *in vitro* bioactivity data from ToxCast as well as structural features to classify different adverse effects *in vivo*.

Here, ToxCast *in vitro* assay data was used to perform BaBRA to predict *in vivo* endpoint information for one chemical by using data from the same *in vivo* endpoint from another chemical that had similar *in vitro* activity. This biological data-based similarity was also enriched with structural similarity (St.BaBRA) and used to make predictions for a chemical’s *in vivo* toxicity based on its nearest neighbors. The measure of chemical similarity was calculated using an unsupervised random forest approach to produce a proximity matrix. Briefly, a random forest is a collection of tree predictors such that each tree depends on the values of an independently sampled random vector from within the feature space (i.e., *in vitro* assay data and/or structural descriptors). When random forest is run in an unsupervised fashion to calculate the proximity matrix, the original data is considered as class 1 and a synthetic second class of the same size is created by sampling at random from the univariate distributions of the original data and labeled as class 2. In this way class 2 maintains the distributions of the variables but destroys the dependency structure in the original data. The N×N proximity matrix is formed by growing a large number of trees (here 10,000) based on the artificial two-class problem, and for each tree if chemicals x and y end in the same terminal node their proximity increases by one. Finally, the proximity scores are normalized by dividing by the number of trees. Figure 2 shows the proximity matrix for the ToxCast Phase I and II chemical library based on the entire set of *in vitro* assays.

Several proximity matrices were calculated to characterize similar bioactivity across the ToxCast chemical library, based on all the *in vitro* assays (Fig. 2), on a subset that were run in

⁶ <http://actor.epa.gov/actor/toxcastdbata.jsp>

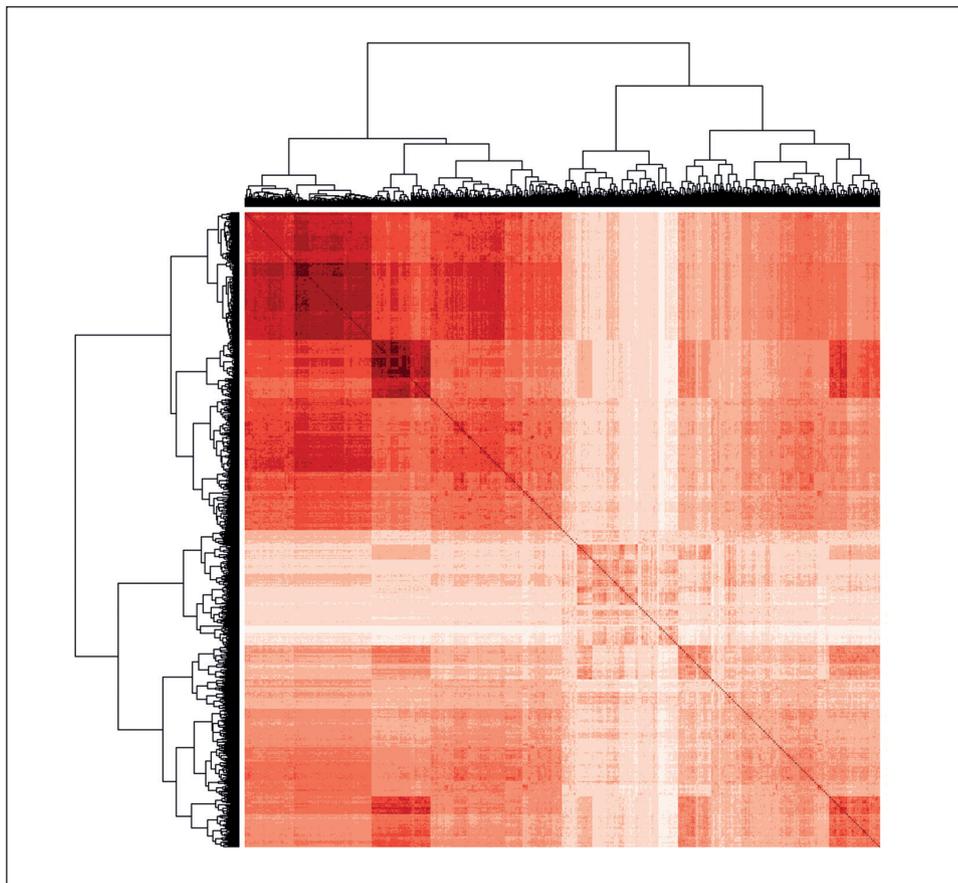


Fig. 2: Random Forest Proximity Matrix for 1,056 ToxCast Phase I/II chemicals

Chemicals are clustered based on their similarity across all 800 ToxCast *in vitro* assay targets. Chemical ordering is the same on each axis, and the unity correlation is shown along the diagonal. Darker red coloring indicates a higher degree of similarity.

concentration response (excluding those that were only run in a single concentration screen), or on assays that mapped to a particular biological pathway relevant to the endpoint of concern. These proximities, with and without enrichment for structural similarity, were used to make predictions for a particular chemical based on the *in vivo* outcomes observed in *k*-nearest neighbor (*k*NN) space. The following equations define the predicted activity for a chemical against a toxicity endpoint based solely on *in vitro* biological similarity,

$$A_{pred} = \frac{\sum_{i=1}^k P_i * A_i}{\sum_{i=1}^k P_i}$$

or with the inclusion of structural similarity,

$$A_{pred} = \frac{\sum_{i=1}^k P_i * A_i + \sum_{i=1}^k S_i * A_i}{\sum_{i=1}^k P_i + \sum_{i=1}^k S_i}$$

where A_{pred} is the predicted activity, *k* is the number of nearest neighbors, P_i is the proximity score based on the ToxCast data, and S_i is the structural similarity score based on the Tanimoto index (Abdo and Salim, 2009). The BaBRA and St.BaBRA predictions were produced based on a variety of proximity matrices, as mentioned previously, and compared to a range of

in vivo endpoints from the Toxicological Reference Database (ToxRefDB⁷) and from a database of guideline-like uterotrophic studies curated by the National Toxicology Program Inter-agency Center for Evaluation of Alternative Toxicological Methods (NICEATM⁸).

Multiple study types are represented in ToxRefDB, namely prenatal developmental, multigenerational reproductive, subchronic and chronic cancer studies, with corresponding lowest effect levels (LELs) on a per-chemical basis for a hierarchy of apical endpoints (e.g., skeletal malformations, litter size, liver tumors, etc.). Many of these endpoints are highly unbalanced, with either positive or negative results significantly overrepresented. To deal with the biased data, random sampling from the positive/negative space was used to create a balanced dataset for each ToxRefDB endpoint. Parameter sweeps were used to define the optimal values for the respective *k*NN space and apply a minimum threshold for the similarity scores. Endpoints from all four study types were predicted using the BaBRA and St.BaBRA approaches, and showed generally poor predictive performance. The best BaBRA model for a ToxRefDB endpoint used a proximity matrix based on the ToxCast assays that were run in concentration response to predict the existence of a LEL for reproductive impairment. This model was based on 256 compounds

⁷ <http://actor.epa.gov/toxrefdb>

⁸ <http://ntp.niehs.nih.gov/pubhealth/evalatm/tox21-support/endocrine-disruptors/edhts.html>



(even distribution between positives and negatives), an optimal k NN space of 6 nearest neighbors, and achieved sensitivity, specificity and balanced accuracy of 70% (p -value $< 1 \times 10^{-6}$). The addition of structural similarity did not improve the model. Interestingly, this level of predictivity may actually reflect the degree of variability in the *in vivo* endpoint being predicted and/or the presence of multiple mechanisms contributing to an observed endpoint.

To create BaBRA/St.BaBRA frameworks that are informed by biological relevance and anchored to highly curated data, we used only assays that mapped to the estrogen receptor (ER) pathway to create a proximity matrix, and attempted to predict the outcome in high quality uterotrophic studies that specifically measure an estrogenic response *in vivo*. The NICEATM uterotrophic database (Kleinstreuer et al., 2015) contains results from > 700 studies whose protocols were evaluated based on adherence to a set of minimum criteria from internationally harmonized test guidelines from the U.S. EPA and the Organization for Economic Cooperation and Development (OECD). The best model performance was achieved using a proximity matrix calculated with only the 18 ToxCast assays that map to the ER pathway, with an optimal k NN of 3 nearest neighbors, for compounds with uterotrophic studies that met all the minimum criteria to be considered guideline-like and had reproducible results across multiple labs. Both BaBRA and St.BaBRA approaches resulted in a sensitivity of 95%, a specificity of 98%, and a balanced accuracy of 97% (p -value $< 1 \times 10^{-15}$), with dibutyl phthalate as the only false positive and octylmethyltetrasiloxane (D4, a highly volatile compound) as the only false negative.

BaBRA and St.BaBRA are approaches that show great promise within certain applicability domains and well-curated datasets. However, broad *in vitro* activity patterns across a wide range of assays are difficult to correlate with apical *in vivo* toxicity endpoints, even when enriched with structural similarities. Feature selection and optimization methods should be explored to improve predictive accuracy and applicability. For example, identifying features that provide the best separation between positive and negative space for each ToxRefDB endpoint in combination with *in vivo* data curation (e.g., lowest adverse effect levels instead of LOELs, study quality evaluation) will improve the applicability of read-across. Further, biological pathway knowledge can be used to define the assay/proximity space that is relevant to the endpoint of interest (e.g., endocrine targets to predict reproductive impairment, or cancer hallmarks to predict carcinogenesis).

2.3 Using omics data to establish biological similarity for read-across

Grouping of chemicals based on structural relationships should also be complemented by omics data: The advantage can be illustrated by two examples: 1) 2-acetylaminofluorene (2-AAF) and 4-acetylaminofluorene (4-AAF) are structurally very similar. However, the toxicological profile of these two compounds differs significantly. 2-AAF is a strong liver enzyme inducer, leading in long-term studies to liver tumors, whereas 4-AAF only slightly induces liver enzymes and does not induce the formation of liver tumors. This is reflected in different metabolome

changes induced by these two compounds in rat plasma (van Ravenzwaay et al., 2012); 2) A lot of compounds with different structures (e.g., fibrates, phthalates, perfluorinated fatty acids) stimulate the peroxisome proliferator activated receptor alpha (PPAR α) leading to hepatomegaly and liver tumors in rodents (Youssef, 1998). These compounds can be grouped by typical metabolite changes in rat plasma into one class, and can be differentiated from other liver tumor-inducing compounds, e.g., liver enzyme inducers (van Ravenzwaay et al., 2010a,b).

A prerequisite for using omics data for read-across is a standardized technique and a database with reference compounds for application of grouping with data-poor chemicals. Regarding metabolomics, BASF and Metanomics have established a standardized technology (Looser et al., 2005) and built up such a database (MetaMap[®] Tox) with about 600 compounds administered to rats in repeated dose studies (van Ravenzwaay et al., 2015). The toxicological activity of data-poor chemicals in rats can be assessed by a standardized evaluation procedure with this database: 1) Profile strength: it is assessed if the number of metabolite changes in rat plasma is above a threshold representing a treatment-related effect (van Ravenzwaay et al., 2014); 2) Pattern ranking: metabolomics mode of action patterns are defined with reference compounds representing a unique set of metabolites changed in the same way. The fit of the metabolome of new compounds to these patterns is evaluated statistically and by toxicological expert judgement; 3) Treatment correlation: The measured metabolome of data-poor chemicals is compared with reference compounds in the database by correlation statistics; 4) Pathway analysis: eventually, endogenous metabolite changes can explain or monitor key events in the adverse outcome pathway (e.g., accumulation of tyrosine and 4-hydroxyphenylpyruvate of 4-hydroxyphenylpyruvate dioxygenase inhibitors, i.e., a herbicide compound class). As a result of the mentioned evaluation process, an assessment can be made regarding: 1) target organ; 2) systemic toxicity modes of action by comparison with reference compounds; 3) which pathways or which chemical groups of metabolites (e.g., aromatic amino acids, unsaturated long chain fatty acids, etc.) in the rat physiology are affected. For example, differentiating direct thyroid hormone synthesis inhibitors from compounds accelerating thyroid hormone clearance (Montoya et al., 2014). The assessment is restricted to the set of reference compounds in the database and the established metabolite patterns, defining which modes of action can be covered.

To increase confidence in the results, different levels of validation procedures have to be performed. Apart from the technical validation of the applied methods and the statistics, influencing factors and the variation of the biological system (here metabolomics in rat plasma) have to be assessed. Regarding the MetaMap[®] Tox database several aspects have been published, such as influence of rat strains (Strauss et al., 2009), influence of the diet (Mellert et al., 2011), reproducibility and robustness of the biological system (Kamp et al., 2012).

Concerning read-across for the absence of a toxic effect, a quantitative risk assessment for the respective endpoints is necessary. The no adverse effect level (NOAEL) can be determined with omics technologies as the absence of a consistent pattern of



change associated with an adverse effect (ECETOC, 2008, 2010, 2013). This is assessed for metabolomics by the fit of the metabolite profile of new compounds to the established adverse mode of action patterns in the MetaMap[®] Tox database (van Ravenzwaay et al., 2014). There are some publications that compare transcriptomic/metabolomic data and most sensitive traditional toxicity data regarding the benchmark dose or NOAEL dose, stating that the sensitivity of the omics technologies compared to “traditional” toxicity measurements (histopathology, clinical pathology) is in a comparable order of magnitude (van Ravenzwaay et al., 2014; Thomas et al., 2011, 2012).

3 Discussion

The above sections and examples together show that the concept of biological similarity enhances read-across: if the target of interest and the similar compounds have been tested in the same set of high-throughput assays, one can use a bioprofile (i.e., a collective set of results from different assays) to profile the target compounds against the tested compounds, and then compare the bioprofile between them. The key of this procedure is to prove the selected assays are related to the toxicological endpoint of interest, either from the understanding of the toxicological mechanism (e.g., as characterized by an AOP) or from the correlative data analysis (e.g., significant relationship between the bioprofile and the toxicological effect). If there is any data gap for generating the bioprofile, i.e., lacking information for certain *in vitro* assays, one might use QSAR models to predict the results of the *in vitro* assay. When applying QSAR modeling one should follow the respective OECD guidance for QSAR (Gramatica, 2007).

The increasing availability of biological data via the data sharing depositories will augment such support of read-across and grouping by big data. The curation of such datasets and the respective data-sharing by companies, organizations and individual researchers needs to be further encouraged and possibly furthered with some incentives. Alternatively, holistic profiling, typically by transcriptomics or metabolomics, of the biological effect of substances in complex systems representing many targets for perturbation can allow an individual assay to support similarity arguments.

4 Conclusions

Taken together, the approaches presented here cannot yet be considered as standardized tools for read-across. However, they promise already now on a case-by-case basis to support read-across considerations and should be considered when the respective test data are available or can be obtained with reasonable efforts. For the future, more accessible standardized testing environments might offer bioprofiling of substances and thereby open the doors to enhanced read-across of substances that have not been broadly studied in the scientific literature.

References

- Abdo, A. and Salim, N. (2009). Similarity-based virtual screening with a Bayesian inference network. *Chem Med Chem* 4, 210-218. <http://dx.doi.org/10.1002/cmdc.200800290>
- Adcock, C. and Tasheva, M. (2009). Flusilazole. In *Pesticide Residues in Food – 2007 Evaluations. Part II – Toxicological* (317-347). Geneva: World Health Organization.
- Adler, S., Basketter, D., Creton, S. et al. (2011). Alternative (non-animal) methods for cosmetics testing: Current status and future prospects – 2010. *Arch Toxicol* 85, 367-485. <http://dx.doi.org/10.1007/s00204-011-0693-2>
- Ankley, G. T., Bennett, R. S., Erickson, R. J. et al. (2010). Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ Toxicol Chem* 29, 730-741. <http://dx.doi.org/10.1002/etc.34>
- Austin, C. P., Brady, L. S., Insel, T. R. and Collins, F. S. (2004). NIH molecular libraries initiative. *Science* 306, 1138-1139. <http://dx.doi.org/10.1126/science.1105511>
- Ball, N., Cronin, M. T. D., She, J. et al. (2016). Toward good read-across practice (GRAP) guidance. *ALTEX* 33, 149-166. <http://dx.doi.org/10.14573/altex.1601251>
- Betts, K. S. (2013). Tox21 to date: Steps toward modernizing human hazard characterization. *Environ Health Perspect* 121, a228. <http://dx.doi.org/10.1289/ehp.121-a228>
- Blauboer, B. J., Boekelheide, K., Clewell, H. J. et al. (2012). The use of biomarkers of toxicity for integrating *in vitro* hazard estimates into risk assessment for humans. *ALTEX* 29, 411-425. <http://dx.doi.org/10.14573/altex.2012.4.411>
- Bosgra, S. and Westerhout, J. (2015). Interpreting *in vitro* developmental toxicity test battery results: The considerations of toxicokinetics. *Reprod Toxicol* 55, 73-80. <http://dx.doi.org/10.1016/j.reprotox.2014.11.001>
- Bouhifd, M., Hartung, T., Hogberg, H. T. et al. (2013). Review: Toxicometabolomics. *J Appl Toxicol* 33, 1365-1383. <http://dx.doi.org/10.1002/jat.2874>
- Bouhifd, M., Beger, R., Flynn, T. et al. (2015). Quality assurance of metabolomics. *ALTEX* 32, 319-326. <http://dx.doi.org/10.14573/altex.1509161>
- Daston, G. P., Chapin, R. E., Scialli, A. R. et al. (2010). A different approach to validating screening assays for developmental toxicity. *Birth Defects Res* 89, 526-530. <http://dx.doi.org/10.1002/bdrb.20276>
- Davis, A. P., Grondin, C. J., Lennon-Hopkins, K., et al. (2014). The Comparative Toxicogenomics Database's 10th year anniversary: update 2015. *Nucleic Acids Res* 43, D914-20. <http://dx.doi.org/10.1093/nar/gku935>
- den Tonkelaar, E. M. and van Koten-Vermeulen, J. E. M. (1991). Hexaconazole. In *Pesticide residues in food – 1990 evaluations. Toxicology*. World Health Organization, WHO/PCS/91.47, 1991, nos 802-816 on INCHEM. <http://www.inchem.org/documents/jmpr/jmpmono/v90pr09.htm> (accessed November 13, 2015).
- Dewhurst, I. and Dellarco, V. (2006). Propiconazole. In *Pesticide residues in food – 2004 evaluations. Part II – Toxicological* (281-323). Geneva: World Health Organization.

- Dix, D. J., Houck, K. A., Martin, M. T. et al. (2007). The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol Sci* 95, 5-12. <http://dx.doi.org/10.1093/toxsci/kfl103>
- ECETOC (2008). Workshop on the application of 'omics technologies in toxicology and ecotoxicology: Case studies and risk assessment. Workshop Report No. 11. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels.
- ECETOC (2010). 'Omics in (eco)toxicology: Case studies and risk assessment. Workshop Report No.19. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels.
- ECETOC (2013). 'Omics and risk assessment science. Workshop Report No. 25. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels.
- Fonger, G. C., Stroup, D., Thomas, P. L. and Wexler, P. (2000). TOXNET: A computerized collection of toxicological and environmental health information. *Toxicol Ind Health* 16, 4-6. <http://dx.doi.org/10.1177/074823370001600101>
- Ganter, B., Tugendreich, S., Pearson, C. I. et al. (2005). Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J Biotechnol* 119, 219-244. <http://dx.doi.org/10.1016/j.jbiotec.2005.03.022>
- Gaulton, A., Bellis, L. J., Bento, A. P. et al. (2012). ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res* 40, D1100-D1107. <http://dx.doi.org/10.1093/nar/gkr777>
- Gramatica, P. (2007). Principles of QSAR models validation: Internal and external. *QSAR Comb Sci* 26, 694-701. <http://dx.doi.org/10.1002/qsar.200610151>
- Hartung, T., Bremer, S., Casati, S. et al. (2004). Modular approach to the ECVAM principles on test validity. *Altern Lab Anim* 32, 467-472.
- Hartung, T. (2007). Food for thought ... on validation. *ALTEX* 24, 67-72. <http://www.altex.ch/en/All-issues/Issue.50.html?iid=86&aid=3>
- Hartung, T. and Hoffmann, S. (2009). Food for thought ... on in silico methods in toxicology. *ALTEX* 26, 155-166. http://www.altex.ch/resources/altex_2009_3_155_166_FFT_Hartung.pdf
- Hartung, T. (2010). Food for thought ... on alternative methods for chemical safety testing. *ALTEX* 27, 3-14. http://www.altex.ch/resources/altex_2010_1_3_14_FFT_Hartung.pdf
- Hartung, T., Luechtefeld, T., Maertens, A. and Kleinsang, A. (2013). Integrated testing strategies for safety assessments. *ALTEX* 30, 3-18. <http://dx.doi.org/10.14573/altex.2013.1.003>
- Hecker, M., Hollert, H., Cooper, R. et al. (2011). The OECD validation program of the H295R steroidogenesis assay: Phase 3 final inter-laboratory validation study. *Environ Sci Pollut Res Int* 18, 503-515. <http://dx.doi.org/10.1007/s11356-010-0396-x>
- Hengstler, J. G., Foth, H., Kahl, R. et al. (2006). The REACH concept and its impact on toxicological sciences. *Toxicol* 220, 232-239. <http://dx.doi.org/10.1016/j.tox.2005.12.005>
- Hermesen, S. A., van den Brandhof, E. J., van der Ven, L. T. and Piersma, A. H. (2011). Relative embryo-toxicity of two classes of chemicals in a modified zebrafish embryotoxicity test and comparison with their in vivo potencies. *Toxicol In Vitro* 25, 745-753. <http://dx.doi.org/10.1016/j.tiv.2011.01.005>
- Inglese, J., Auld, D. S., Jadhav, A. et al. (2006). Quantitative high-throughput screening: A titration-based approach that efficiently identifies biological activities in large chemical libraries. *Proc Natl Acad Sci U S A* 103, 11473-11478. <http://dx.doi.org/10.1073/pnas.0604348103>
- Judson, R. S., Richard, A., Dix, D. et al. (2008). ACToR – aggregated computational toxicology resource. *Toxicol Appl Pharmacol* 233, 7-13. <http://dx.doi.org/10.1016/j.taap.2007.12.037>
- Judson, R. S., Houck, K. A., Kavlock, R. J. et al. (2010). In vitro screening of environmental chemicals for targeted testing prioritization: The ToxCast project. *Environ Health Perspect* 118, 485-492. <http://dx.doi.org/10.1289/ehp.0901392>
- Judson, R. S., Martin, M. T., Egeghy, P. et al. (2012). Aggregating data for computational toxicology applications: The U.S. Environmental Protection Agency (EPA) aggregated computational toxicology resource (ACToR) system. *Int J Mol Sci* 13, 1805-1831. <http://dx.doi.org/10.3390/ijms13021805>
- Kamp, H., Strauss, V., Wiemer, J. et al. (2012). Reproducibility and robustness of metabolome analysis in rat plasma of 28-day repeated dose toxicity studies. *Toxicol Letters* 215, 143-149. <http://dx.doi.org/10.1016/j.toxlet.2012.09.015>
- Kim, M. T., Sedykh, A., Chakravarti, S. K. et al. (2014). Critical evaluation of human oral bioavailability for pharmaceutical drugs by using various cheminformatics approaches. *Pharmaceut Res* 31, 1002-1014. <http://dx.doi.org/10.1007/s11095-013-1222-1>
- Kim, M. T., Huang, R., Sedykh, A. et al. (2016). Mechanism profiling of hepatotoxicity caused by oxidative stress using the antioxidant response element reporter gene assay models and big data. *Environ Health Perspect*, in press.
- Kleinsang, A., Maertens, A., Rosenberg, M. et al. (2014). Pathways of toxicity. *ALTEX* 31, 53-61. <http://dx.doi.org/10.14573/altex.1309261>
- Kleinstreuer, N., Judson, R., Reif, D. et al. (2011). Environmental impact on vascular development predicted by high throughput screening (HTS). *Environ Health Perspect* 119, 1596-1603. <http://dx.doi.org/10.1289/ehp.1103412>
- Kleinstreuer, N., Houck, K., Yang, J. et al. (2014). Bioactivity profiling of a diverse 800 chemical library in human primary cell systems. *Nat Biotechnol* 32, 583-591. <http://dx.doi.org/10.1038/nbt.2914>
- Kleinstreuer, N. C., Ceger, P., Allen, D. et al. (2015). A curated database of rodent uterotrophic bioactivity. *Environ Health Perspect*, Epub ahead of print. <http://dx.doi.org/10.1289/ehp.1510183>
- Klopman, G., Saiakhov, R., Rosenkranz, H. S. and Hermens, J. L. M. (1999). Multiple computer-automated structure evaluation program study of aquatic toxicity 1: Guppy. *Environ Toxicol Chem* 18, 2497-2505. <http://dx.doi.org/10.1002/etc.5620181116>
- Klopman, G., Zhu, H., Fuller, M. A. and Saiakhov, R. D. (2004).



- Searching for an enhanced predictive tool for mutagenicity. *SAR QSAR Environ Res* 15, 251-263. <http://dx.doi.org/10.1080/10629360410001724897>
- Knudsen, T. B. and Kleinstreuer, N. C. (2012). Disruption of embryonic vascular development in predictive toxicology. *Birth Defects Res C Embryo Today* 93, 312-323. <http://dx.doi.org/10.1002/bdrc.20223>
- Kohonen, P., Benfenati, E., Bower, D. et al. (2013). The Tox-Bank data warehouse: Supporting the replacement of in vivo repeated dose systemic toxicity testing. *Molec Inform* 32, 47-63. <http://dx.doi.org/10.1002/minf.201200114>
- Kroese, E. D., Bosgra, S., Buist, H. E. et al. (2015). Evaluation of an alternative in vitro test battery for detecting reproductive toxicants in a grouping context. *Reprod Toxicol* 55, 11-19. <http://dx.doi.org/10.1016/j.reprotox.2014.10.003>
- Lamb, J., Crawford, E. D., Peck, D. et al. (2006). The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929-1935. <http://dx.doi.org/10.1126/science.1132939>
- Leist, M., Hasiwa, N., Rovida, C. et al. (2014). Consensus report on the future of animal-free systemic toxicity testing. *ALTEX* 31, 341-356. <http://dx.doi.org/10.14573/altex.1406091>
- Linkov, I., Massey, O., Keisler, J. et al. (2015). From “weight of evidence” to quantitative data integration using multicriteria decision analysis and Bayesian methods. *ALTEX* 32, 3-8. <http://dx.doi.org/10.14573/altex.1412231>
- Looser, R., Krotzky, A. J. and Trethewey, R. N. (2005). Metabolite profiling with GC-MS and LC-MS. In *Metabolome Analyses: Strategies for Systems Biology* (103-118). New York: Springer US. http://doi.org/10.1007/0-387-25240-1_7
- Low, Y., Sedykh, A., Fourches, D. et al. (2013). Integrative chemical-biological read-across approach for chemical hazard classification. *Chem Res Toxicol* 26, 1199-1208. <http://dx.doi.org/10.1021/tx400110f>
- Luechtefeld, T., Maertens, A., Russo, D. P. et al. (2016a). Global analysis of publicly available safety data for 9,801 substances registered under REACH from 2008-2014. *ALTEX* 33, 95-109. <http://dx.doi.org/10.14573/altex.1510052>
- Luechtefeld, T., Maertens, A., Russo, D. P. et al. (2016b). Analysis of public oral toxicity data from REACH registrations 2008-2014. *ALTEX* 33, 111-122. <http://dx.doi.org/10.14573/altex.1510054>
- Luechtefeld, T., Maertens, A., Russo, D. P. et al. (2016c). Analysis of Draize eye irritation testing and its prediction by mining publicly available 2008-2014 REACH data. *ALTEX* 33, 123-134. <http://dx.doi.org/10.14573/altex.1510053>
- Luechtefeld, T., Maertens, A., Russo, D. P. et al. (2016d). Analysis of publically available skin sensitization data from REACH registrations 2008-2014. *ALTEX* 33, 135-148. <http://dx.doi.org/10.14573/altex.1510055>
- Maggiore, G. M. (2006). On outliers and activity cliffs – why QSAR often disappoints. *J Chem Inf Model* 46, 1535. <http://dx.doi.org/10.1021/ci060117s>
- Mattingly, C. J., Colby, G. T., Forrest, J. N. and Boyer, J. L. (2003). The comparative toxicogenomics database (CTD). *Environ Health Perspect* 111, 793-795. <http://dx.doi.org/10.1289/ehp.6028>
- Mattingly, C. J., Colby, G. T., Rosenstein, M. C. et al. (2004). Promoting comparative molecular studies in environmental health research: An overview of the comparative toxicogenomics database (CTD). *Pharmacogenomics J* 4, 5-8. <http://dx.doi.org/10.1038/sj.tpj.6500225>
- Mattingly, C. J., Rosenstein, M. C., Colby, G. T. et al. (2006a). The comparative toxicogenomics database (CTD): A resource for comparative toxicological studies. *J Exp Zool A Comp Exp Biol* 305, 689-692. <http://dx.doi.org/10.1002/jez.a.307>
- Mattingly, C. J., Rosenstein, M. C., Davis, A. P. et al. (2006b). The comparative toxicogenomics database: A cross-species resource for building chemical-gene interaction networks. *Toxicol Sci* 92, 587-595. <http://dx.doi.org/10.1093/toxsci/kfl008>
- McHale, C. M., Zhang, L., Hubbard, A. E. and Smith, M. T. (2010). Toxicogenomic profiling of chemically exposed humans in risk assessment. *Mutat Res* 705, 172-183. <http://dx.doi.org/10.1016/j.mrrev.2010.04.001>
- Mellert, W., Kapp, M., Strauss, V. et al. (2011). Nutritional impact on the plasma metabolome of rats. *Toxicol Letters* 207, 173-181. <http://dx.doi.org/10.1016/j.toxlet.2011.08.013>
- Montoya, G. A., Strauss, V., Fabian, E. et al. (2014). Mechanistic analysis of metabolomics patterns in rat plasma during administration of direct thyroid hormone synthesis inhibitors or compounds increasing thyroid hormone clearance. *Toxicol Lett* 225, 240-251. <http://dx.doi.org/10.1016/j.toxlet.2013.12.010>
- Moss, G. P., Dearden, J. C., Patel, H. and Cronin, M. T. D. (2002). Quantitative structure-permeability relationships (QSPRs) for percutaneous absorption. *Toxicol In Vitro* 16, 299-317. [http://dx.doi.org/10.1016/S0887-2333\(02\)00003-6](http://dx.doi.org/10.1016/S0887-2333(02)00003-6)
- Palmer, J. A., Smith, A. M., Egnash, L. A. et al. (2013). Establishment and assessment of a new human embryonic stem cell-based biomarker assay for developmental toxicity screening. *Birth Defects Res B Dev Reprod Toxicol* 98, 343-363. <http://dx.doi.org/10.1002/bdrb.21078>
- Pamies, D., Martínez, C. E., Sogorb, M. A. and Vilanova, E. (2011). Mechanism-based models in reproductive and development toxicology. In R. C. Gupta (ed.), *Reproductive and Developmental Toxicology* (135-146). San Diego, CA: Academic Press-Elsevier. <http://dx.doi.org/10.1016/B978-0-12-382032-7.10011-6>
- Patlewicz, G., Ball, N., Becker, R. A. et al. (2014). Read-across approaches – misconceptions, promises and challenges ahead. *ALTEX* 31, 387-396. <http://dx.doi.org/10.14573/altex.1410071>
- Piersma, A. H., Bosgra, S., van Duursen, M. B. M. et al. (2013). Evaluation of an alternative in vitro test battery for detecting reproductive toxicants. *Reprod Toxicol* 38, 53-64. <http://dx.doi.org/10.1016/j.reprotox.2013.03.002>
- Ramirez, T., Daneshian, M., Kamp, H. et al. (2013). Metabolomics in toxicology and preclinical research. *ALTEX* 30, 209-225. <http://dx.doi.org/10.14573/altex.2013.2.209>
- Reif, D. M., Martin, M. T., Tan, S. W. et al. (2010). Endocrine

- profiling and prioritization of environmental chemicals using ToxCast data. *Environ Health Perspect* 118, 1714-1720. <http://dx.doi.org/10.1289/ehp.1002180>
- Rovida, C., Alépée, N., Api, A. M. et al. (2015). Integrated Testing Strategies (ITS) for safety assessment. *ALTEX* 32, 171-181. <http://dx.doi.org/10.14573/altex.1506201>
- Schenk, B., Weimer, M., Bremer, S. et al. (2010). The ReProTect feasibility study, a novel comprehensive in vitro approach to detect reproductive toxicants. *Reprod Toxicol* 30, 200-218. <http://dx.doi.org/10.1016/j.reprotox.2010.05.012>
- Scholz, G., Pohl, I., Genschow, E. et al. (1999). Embryotoxicity screening using embryonic stem cells in vitro: Correlation to in vivo teratogenicity. *Cells Tissues Organs* 165, 203-211. <http://dx.doi.org/10.1159/000016700>
- Schultz, T. W., Cronin, M. T. D. and Netzeva, T. I. (2003). The present status of QSAR in toxicology. *J Molec Struct – Theo Chem* 622, 23-38. [http://dx.doi.org/10.1016/S0166-1280\(02\)00615-2](http://dx.doi.org/10.1016/S0166-1280(02)00615-2)
- Sipes, N. S., Martin, M. T., Kothiyia, P. et al. (2013). Profiling 976 ToxCast chemicals across 331 enzymatic and receptor signaling assays. *Chem Res Toxicol* 26, 878-895. <http://dx.doi.org/10.1021/tx400021f>
- Solimeo, R., Zhang, J., Kim, M. et al. (2012). Predicting chemical ocular toxicity using a combinatorial QSAR approach. *Chem Res Toxicol* 25, 2763-2769. <http://dx.doi.org/10.1021/tx300393v>
- Sonneveld, E., Jansen, H. J., Riteco, J. A. C. et al. (2005). Development of androgen- and estrogen-responsive bioassays, members of a panel of human cell line-based highly selective steroid responsive bioassays. *Toxicol Sci* 83, 136-148. <http://dx.doi.org/10.1093/toxsci/kfi005>
- Strauss, V., Wiemer, J., Leibold, E. et al. (2009). Influence of strain and sex on the metabolic profile of rats in repeated dose toxicological studies. *Toxicol Lett* 191, 88-95. <http://dx.doi.org/10.1016/j.toxlet.2009.08.004>
- Suter, L., Schroeder, S., Meyer, K. et al. (2011). EU Framework 6 project: Predictive toxicology (PredTox) – overview and outcome. *Toxicol Appl Pharmacol* 252, 73-84. <http://dx.doi.org/10.1016/j.taap.2010.10.008>
- Thomas, C. J., Auld, D. S., Huang, R. et al. (2009). The pilot phase of the NIH Chemical Genomics Center. *Curr Top Med Chem* 9, 1181-1193. <http://dx.doi.org/10.2174/156802609789753644>
- Thomas, R. S., Clewell, H. J. III, Allen, B. C. et al. (2011). Application of transcriptional benchmark dose values in quantitative cancer and noncancer risk assessment. *Toxicol Sci* 120, 194-205. <http://dx.doi.org/10.1093/toxsci/kfq355>
- Thomas, R. S., Clewell, H. J. III, Allen, B. C. et al. (2012). Integrating pathway-based transcriptomic data into quantitative chemical risk assessment: A five chemical case study. *Mutat Res* 746, 135-143. <http://dx.doi.org/10.1016/j.mrgentox.2012.01.007>
- Uehara, T., Ono, A., Maruyama, T. et al. (2010). The Japanese toxicogenomics project: Application of toxicogenomics. *Molecular Nutrition Food Res* 54, 218-227. <http://dx.doi.org/10.1002/mnfr.200900169>
- Uibel, F. and Schwarz, M. (2015). Prediction of embryotoxic potential using the ReProGlo stem cell-based WNT reporter assay. *Reprod Toxicol* 55, 30-49. <http://dx.doi.org/10.1016/j.reprotox.2014.09.009>
- van Duursen, M. B., Nijmeijer, S. M., Ruchirawat, S. and van den Berg, M. (2010). Chemo-preventive actions by enterolactone and 13 VIOXX-related lactone derivatives in H295R human adrenocortical carcinoma cells. *Toxicol Lett* 192, 271-277. <http://dx.doi.org/10.1016/j.toxlet.2009.11.001>
- van der Burg, B., Kroese, E. D. and Piersma, A. H. (2011). Towards a pragmatic alternative testing strategy for the detection of reproductive toxicants. *Reprod Toxicol* 31, 558-561. <http://dx.doi.org/10.1016/j.reprotox.2011.02.012>
- van der Burg, B., van der Linden, S. C., Man, H. Y. et al. (2013). A panel of quantitative CALUX[®] reporter gene assays for reliable high throughput toxicity screening of chemicals and complex mixtures. In P. Steinberg (ed.), *High Throughput Screening Methods in Toxicity Testing* (519-532). New York: John Wiley and Sons, Inc. <http://dx.doi.org/10.1002/9781118538203.ch28>
- van Ravenzwaay, B., Coelho-Palermo Cunha, G., Fabian, E. et al. (2010a). The use of metabolomics in cancer research. In W. C. S. Cho (ed.), *An Omics Perspective of Cancer* (141-166). Berlin, Germany: Springer. http://dx.doi.org/10.1007/978-90-481-2675-0_8
- van Ravenzwaay, B., Coelho-Palermo Cunha, G., Strauss, V. et al. (2010b). The individual and combined metabolite profiles (metabolomics) of dibutylphthalate and di(2-ethylhexyl)phthalate following a 28-day dietary exposure in rats. *Toxicol Lett* 198, 159-170. <http://dx.doi.org/10.1016/j.toxlet.2010.06.009>
- van Ravenzwaay, B., Herold, M., Kamp, H. et al. (2012). Metabolomics: A tool for early detection of toxicological effects and an opportunity for biology based grouping of chemicals from QSAR to QBAR. *Mutat Res* 746, 144-150. <http://dx.doi.org/10.1016/j.mrgentox.2012.01.006>
- van Ravenzwaay, B., Montoya, G. A., Fabian E. et al. (2014). The sensitivity of metabolomics versus classical regulatory toxicology from a NOAEL perspective. *Toxicol Lett* 227, 20-28. <http://dx.doi.org/10.1016/j.toxlet.2014.03.004>
- van Ravenzwaay, B., Peter, E., Krennrich, G. et al. (2015). The development of a data base for metabolomics – looking back on 10 years of experience, in press.
- van Vliet, E. (2011). Current standing and future prospects for the technologies proposed to transform toxicity testing in the 21st century. *ALTEX* 28, 17-44. <http://dx.doi.org/10.14573/altex.2011.1.017>
- Vinken, M., Pauwels, M., Ates, G. et al. (2012). Screening of repeated dose toxicity data present in SCC(NF)P/SCCS safety evaluations of cosmetic ingredients. *Arch Toxicol* 86, 405-412. <http://dx.doi.org/10.1007/s00204-011-0769-z>
- Vinken, M., Landesmann, B., Goumenou, M. et al. (2013a). Development of an adverse outcome pathway from drug-mediated bile salt export pump inhibition to cholestatic liver injury. *Toxicol Sci* 136, 97-106. <http://dx.doi.org/10.1093/toxsci/kft177>



- Vinken, M. (2013b). The adverse outcome pathway concept: A pragmatic tool in toxicology. *Toxicol* 312, 158-165. <http://dx.doi.org/10.1016/j.tox.2013.08.011>
- Wang, W. Y., Kim, M. T., Sedykh, A and Zhu, H. (2015). Developing enhanced blood-brain barrier permeability models: Integrating external bio-assay data in QSAR modeling. *Pharmaceut Res* 32, 3055-3065. <http://dx.doi.org/10.1007/s11095-015-1687-1>
- Wang, Y., Xiao, J., Suzek, T. O. et al. (2009). PubChem: A public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res* 37, W623-W633. <http://dx.doi.org/10.1093/nar/gkp456>
- Wang, Y., Bolton, E., Dracheva, S. et al. (2010). An overview of the PubChem BioAssay resource. *Nucleic Acids Res* 38, D255-D266. <http://dx.doi.org/10.1093/nar/gkp965>
- Waters, M., Stasiewicz, S., Merrick, B. A. et al. (2008). CEBS – chemical effects in biological systems: A public data repository integrating study design and toxicity data with microarray and proteomics data. *Nucleic Acids Res* 36, D892-D900. <http://dx.doi.org/10.1093/nar/gkm755>
- Wu, S., Fisher, J., Nacliff, J. et al. (2013). Frame-work for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. *Chem Res Toxicol* 26, 1840-1861. <http://dx.doi.org/10.1021/tx400226u>
- Yang, C. H., Ambrosio, M., Arvidson, K. et al. (2013). Development of new COSMOS repeat dose and non-cancer threshold of toxicological concern (TTC) databases to support alternative testing methods for cosmetics related chemicals. *Toxicol Lett* 221, S80. <http://dx.doi.org/10.1016/j.toxlet.2013.05.082>
- Yoshida, M. and McGregor, D. (2015). Myclobutanil. In *Pesticide Residues in Food – 2014 Evaluations. Part II – Toxicological* (357-405). World Health Organization, 2015.
- Youssef, J. and Badr, M. (1998). Extraperoxisomal targets of peroxisome proliferators: Mitochondrial, microsomal, and cytosolic effects. Implications for health and disease. *Crit Rev Toxicol* 28, 1-33. <http://dx.doi.org/10.1080/10408449891344182>
- Zarn, J., Davies, L. and Boobis, A. (2006). Triadimenol and Triadimefon. In *Pesticide Residues in Food – 2004 Evaluations. Part II – Toxicological* (325-386). World Health Organization, WHO/PCS/06.1.
- Zhang, J., Hsieh, J. H. and Zhu, H. (2014). Profiling animal toxicants by automatically mining public bioassay data: A big data approach for computational toxicology. *PLoS One* 9, e99863. <http://dx.doi.org/10.1371/journal.pone.0099863>
- Zhu, H., Tropsha, A., Fourches, D. et al. (2008). Combinatorial QSAR modeling of chemical toxicants tested against *Tetrahymena pyriformis*. *J Chem Inf Model* 48, 766-784. <http://dx.doi.org/10.1021/ci700443v>
- Zhu, H., Martin, T. M., Ye, L. et al. (2009). Quantitative structure-activity relationship modeling of rat acute toxicity by oral exposure. *Chem Res Toxicol* 22, 1913-1921. <http://dx.doi.org/10.1021/tx900189p>
- Zhu, H., Zhang, J., Kim, M. T. et al. (2014). Big data in chemical toxicity research: The use of high-throughput screening assays to identify potential toxicants. *Chem Res Toxicol* 27, 1643-1651. <http://dx.doi.org/10.1021/tx500145h>

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

T.L. was supported by NIEHS training grant (T32 ES007141). Support from the EU Horizon 2020 project EUToxRisk is gratefully appreciated.

Correspondence to

Thomas Hartung, MD PhD
Center for Alternatives to Animal Testing
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe Street
Baltimore, MD 21205, USA
e-mail: thartun1@jhu.edu