



“It is the mark of an instructed mind to rest easy with the degree of precision which the nature of the subject permits, and not to seek an exactness where only an approximation of the truth is possible”

Aristotle (384-322 BC)

Food for Thought

Making Big Sense from Big Data in Toxicology by Read-Across

Thomas Hartung

Johns Hopkins Bloomberg School of Public Health, Center for Alternatives to Animal Testing (CAAT), Baltimore, USA and University of Konstanz, CAAT-Europe, Konstanz, Germany

Summary

Modern information technologies have made big data available in safety sciences, i.e., extremely large data sets that may be analyzed only computationally to reveal patterns, trends and associations. This happens by (1) compilation of large sets of existing data, e.g., as a result of the European REACH regulation, (2) the use of omics technologies and (3) systematic robotized testing in a high-throughput manner. All three approaches and some other high-content technologies leave us with big data – the challenge is now to make big sense of these data. Read-across, i.e., the local similarity-based intrapolation of properties, is gaining momentum with increasing data availability and consensus on how to process and report it. It is predominantly applied to *in vivo* test data as a gap-filling approach, but can similarly complement other incomplete datasets. Big data are first of all repositories for finding similar substances and ensure that the available data is fully exploited. High-content and high-throughput approaches similarly require focusing on clusters, in this case formed by underlying mechanisms such as pathways of toxicity. The closely connected properties, i.e., structural and biological similarity, create the confidence needed for predictions of toxic properties. Here, a new web-based tool under development called REACH-across, which aims to support and automate structure-based read-across, is presented among others.

Keywords: computational toxicology, *in silico*, data-mining, read-across, databases

1 Introduction

Read-across has been termed an “ugly duckling” (Teubner and Landsiedel, 2015). For many it still has the stigma of GOBSAT (“good old boys sitting around the table”), i.e., a very pragmatic discussion trying to take a shortcut and avoid testing by arguing that we know enough from similar substances. But REACH, the European chemicals legislation (Regulation (EC) No 1907/2006), has changed the game by first making read-across an official tool (see Box 1 detailing some of the official guidance of the European Chemicals Agency, ECHA) and providing

a Read-Across Assessment Framework (RAAF) published mid 2015. This has placed the bar high, too high many feel, as it definitely is not an easy thing to meet these criteria as the emerging Good Read-Across Practice guidance (Ball et al., 2016) impressively documents.

The article presented here makes the case that big data can nurture the ugly duckling to becoming a beautiful swan. And an essential contribution to the big data appears to come from REACH itself: The mineable REACH database also presented in this issue of ALTEX (Luechtefeld et al., 2016a-d) is already now the largest toxicological database, and has enormous growth

¹ http://echa.europa.eu/documents/10162/13628/raaf_en.pdf

Received March 9, 2016
<http://dx.doi.org/10.14573/altex.1603091>



This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is appropriately cited.



Box 1

Extract from European Chemicals Agency (2010)

Summary of "Guidance on requirements for substances in articles," 1-19.

Grouping of substances and read-across

Animal tests on a substance can be avoided if there is enough evidence on similar substances which the registrant can show should be "read across" to their own substance. Substances whose physicochemical, toxicological and ecotoxicological properties are likely to be similar or follow a regular pattern as a result of structural similarity, may be considered as a "group", or "category" of substances. Applying the group concept means that the physicochemical properties, human health effects and environmental effects or environmental fate may be predicted from data for one substance within the group by interpolation to other substances in the group (read-across approach). This avoids the need to test every substance in the group for every hazard endpoint. Preferably, a category should include all similar substances. REACH Annex XI, Section 1.5. sets out the requirements for the application of this strategy.

Recommendations

1. Results from the read-across approach should be adequate for the purposes of classification and labelling and/or risk assessment (see section R6.2.3 of *Guidance on Information Requirements and Chemical Safety Assessment*).
2. Substance identity must be specified and documented for all relevant members of the category, including purity/impurity profiles. The *Guidance for identification and naming of substances under REACH* should be used.
3. Where substances have been accepted as members of categories under other regulatory programs (for example OECD HPV categories), the registrant should refer to them in the dossier. The registrant must nevertheless include all

available information (including information which became available after assessment in the other regulatory programme) and reassess the validity of the category.

4. The read-across hypothesis used and its justification must be detailed in the dossier. An acceptable read-across justification is normally based on multiple lines of evidence. Different routes of exposure should also be taken into account. A consideration of information from studies on toxicokinetics may improve the robustness of the read-across hypothesis.
5. The documentation must detail which hazard end-points are covered by the read-across, and the source chemical which is used for the read-across must be identified. It is also important that the reliability indicator (Klimisch score*) reflects the *assumptions* of similarity. Thus, a score of 1 (reliable without restrictions) should normally not be used for results derived from read-across.
6. A comparison of experimental data for hazard endpoints for all category members (also presented in a tabular data matrix) is recommended, ideally highlighting trends within the category.

Further information can be found in the *Guidance on information requirements and chemical safety assessment in Chapter R.6: (Q)SARs and grouping of chemicals and in the Practical Guide 6: How to report read-across and categories*.

* Klimisch, H., Andreae, M. and Tillmann, U. (1997). A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regul Toxicol Pharmacol* 25, 1-5.

prospects considering the ongoing REACH registration process. It allows us to find the chemical structure and data availability of substances that are similar to queried substances in fractions of seconds – results that would take ages based on manual compilation and that could never be complete. It is in the best interest of all stakeholders that such a powerful tool becomes fully available and is maximally exploited for the purpose of REACH and beyond. A proactive data curation and sharing could boost this and it seems that the next version of IUCLID, the REACH registration software, to be published later in 2016, takes a further step in this direction, with more structured and standardized data input. At this moment, our database project and a tool to be presented here, which we call REACH-across and are currently developing, fills the gap. Quite ironic: filling the gap for a data-gap filling tool...

Chemical structure determines chemico-physical properties and reactivities, both key determinants of the interactions

with biological systems. Large parts of the chemical universe act quite promiscuously, i.e., at the concentration at which they start to cause biological effects, many aspects of physiology are simultaneously perturbed (Thomas et al., 2013). In these cases, it makes sense that very general properties play a role. Other substances exert their damaging effects via specific targets and pathways. Here, often more narrow structural requirements are needed to trigger the mechanism. However, many other aspects blur this interaction, such as the kinetics (adsorption, distribution, metabolism and excretion) of the substance, which again strongly depend on structural and reactivity features but not necessarily the same ones. This leads to many overlapping non-linear relationships, which are highly dynamic and interdependent. Some of them have thresholds or other sudden changes in properties, which lead to discontinuities, often called "activity cliffs," i.e., sudden changes in a property with minimal changes in structure.

Rarely will we be able to fully describe these connections. However, the probability is very high that properties are shared in a given local environment. This will not give us absolute certainty, but absolute certainty is an illusion anyway in the safety sciences considering the poor approximation of the human situation of our models and their inherent limitations, not least with regard to reproducibility (Hartung, 2013). *In silico* approaches based on information on similar substances, such as read-across and grouping / category approaches, thus often represent a reasonable tool to fill data gaps or prioritize testing and risk management needs. Their enormous saving potential was for example shown in the US High-Production Volume Chemical program (Bishop et al., 2012): The potential consumption of 3.5 million animals in new testing was brought down to approximately 127,000. These methods are not perfect, but more efficient and not necessarily less predictive than testing on animals, which are not little humans on four legs.

2 The availability of good big data

Making sense of big data starts with having good big data! Another discussion on the shortcomings of both *in vivo* and *in vitro* methods? The well-disposed reader of these Food for thought ... articles has noticed this fundamental argument recurring (Hartung, 2013). However, the big data can finally enable an objective assessment. How reproducible are our animal tests? In this issue of ALTEX, we show some examples of how this information can be extracted from the REACH registration data. An easy target is the Draize eye test. Its non-reproducibility already was shown by Weil and Scala in 1971! However, defenders of the test have always claimed that this was before the test was standardized in an OECD guideline and before Good Laboratory Practice. It is shocking how often some chemicals have been tested in the same test, for example, every chemical on average three times in rabbit eyes. Two chemicals were tested 90 times, sixty-nine chemicals were tested 45 times (Luechtefeld et al., 2016c). This demonstrates how important the mutual acceptance of data brokered by OECD between countries is: Before 1981 tests often had to be repeated for registrations in different countries. It also stresses how important it is that REACH requests all companies that are interested in one substance to work together and share their data. Before, they might have independently registered their chemicals, not knowing of each other's data. However, this waste of animals now shows for the first time how bad this test really is. A lottery: If a substance was identified as a severe irritant in the first test, there is a 20% probability that it is characterized a mild irritant in the repetition and 10% that it is deemed a non-irritant. The other way around, many irritants will go undetected. This result should speed up the replacement of the rabbit eye test. Some alternative methods have proven to be better than this reproducibility of the animal test. And this is not the only example: When different acceptable animal tests correspond only around 80% of the time, as shown for both acute oral toxicity (Luechtefeld et al., 2016b) and skin sensitization (Luechtefeld et al., 2016d), this measure

is easily achieved by some of the validated and partially accepted *in vitro* methods as stand-alone tests, not even requiring complicated and costly integrated testing strategies (Hartung et al., 2013; Rovida et al., 2015) currently discussed.

The number and size of databases and available tools is steadily increasing (Nigsch et al., 2009; Rusyn and Daston, 2010; Raunio, 2011; Greene and Pennie, 2015). The prerequisite for making sense of these big data is that we deal with good big data. A major concern for all chemical databases is that chemical structures may be entered incorrectly, which has been found in 0.1 to 3.4% of the cases (Young and Martin, 2008; Fourches et al., 2010). The needs for quality assurance differ among the different sources: If we look at existing (especially animal study) data, these tests are typically not performed and reported in a standardized way, documentation quality differs, and, as safety data have in the past typically been considered proprietary, are difficult to access, compile and analyze in an aggregated manner. REACH, the European chemicals regulation, was a game-changer here, as it made the publication of at least some summarizing information mandatory. REACH forces the creation of consortia, i.e., SIEF (Data Sharing & Substance Information Exchange Forum), mandatory for submitting information jointly: "One substance, one dossier". This is, together with the fact that immediate regulatory consequences can result, the best way to assure the quality of submissions. All interested parties have a say in how a substance is registered and classified, and access to all available data is much easier than for any registrant in isolation. This provides us with big data in two dimensions: broad and deep. We can get information on many substances and with many replicates and variants of assessment. This basis allows us to map the chemical universe and assess the quality or comparative performance of the different tools that are employed.

Very differently, omics technologies produce a holistic description of a single substance's effect in one biological test system. However, this requires the use of highly standardized test systems as well as a high level of standardization of the omics technology and its processing / analysis. The high number of assessed variables creates problems from the start, further amplified by the noise associated with most of these technologies. Measuring lots of things on a noisy background is doomed to disaster from the start. Thus, the requirements of the test systems are even higher, i.e., a bad test does not become better by adding sophisticated omics technology. On the contrary, we will spot the shortcomings (the noise) of the test more easily.

High-throughput testing (HTS) is now quite common in drug screening, sometimes involving hundreds of thousands or millions of substances. However, it is typically carried out only at a single concentration and very little quality assurance is applied as the interest is in the hits, i.e., the substances showing the property of interest, and the things that are missed are of little concern. The use of HTS for safety concerns must be quite different: It must not miss substances of concern and one concentration typically is not sufficient to assure this. Thus there is a far higher need to quality assure, for example, the identity and purity of the substances studied. This approach has been impressively steered by the ToxCast and Tox21 programs of US



agencies. They have stimulated many attempts to mine these data for predictive toxicity^{2,3} (more general: Sun et al., 2012; Rusyn et al., 2012).

Omics technologies and the integrated analysis of multi-omics are in the focus of the Human Toxome Project (Bouhifd et al., 2015). The lessons learned here and in the use of HTS data shall be the subject of a future article in this series. Both approaches enhance read-across, offering opportunities for biological support data (Zhu et al., 2016), and thus need to be mentioned here. They have in common that they make most sense when data are reflecting a mechanism of toxic action. In comparison, read-across is first of all agnostic of mechanism, but by embracing the mechanistic similarity of molecules we can dramatically enhance the predictive value of read-across. Very similar to structure, the multidimensional biological assessments of substances create fingerprints allowing us to find similarities, e.g., by machine learning algorithms. This allows reading across a shared mechanism – and, for example, when some substances have been annotated to a PoT/AOP, these other chemicals can be linked to the same pathway.

3 Trash in, trash out

This is a simple golden rule, beautifully condensed in line one of Figure 1. If using inappropriate (bad) computation, it only gets worse (line two). But we can really mess up badly, when we complicate this with omics-type big data with lots of noise and far too many variables measured (line three).

The author has been quite skeptical as to computational approaches in toxicology for many years (Hartung and Hoffmann, 2009). An important reason was the overselling of one type of computational approach for the purpose of REACH, i.e., quantitative structure-activity relationships (QSAR). Oversimplifying, the aim of (Q)SAR is to find a formula to predict the properties of the chemical universe by some descriptors or rules from chemical structure. There is no doubt that there are some relationships between structure and properties of a chemical. However, the intentional overselling of the prospects of these methods for REACH was appalling.

Table 1 shows as key example of the expectations raised for these approaches in the impact analysis of the European Commission (Pedersen et al., 2003).

The proposed up to 92% use of computational methods drastically reduced the expected impact of the REACH legislation, which was in its making at this time. Notably, this estimate has never been revoked but fiercely defended. Already a simple plausibility check showed that this could not be possible. At the time we commissioned a simple study, which asked how high the percentage of substances under REACH is that qualify for computational approaches because they represent a defined structure amenable to computational toxicology. The result for 200 high- and 200 low-production volume chemicals was very

$$f(\text{trash}) = \text{trash} \quad \text{in vivo / vitro}$$

$$f(\text{trash}) = \text{trash}^2 \quad \text{in vivo / vitro}$$

$$f(\text{trash}) = \text{trash}^3 \quad \text{in vivo / vitro plus omics}$$

Fig. 1: Illustration of the principle of trash in - trash out and its corollaries

“Trash in - trash out” nicely condenses the need for good data to create models. In toxicology, this refers both to *in vivo* and *in vitro* data. The problem increases with inadequate modeling and becomes even worse, when high-content endpoints such as omics technologies are combined with the experimental system.

Tab. 1: Optimal use of (Q)SARs, grouping and read-across techniques according to Pedersen et al. (2003)

Endpoint	Acceptance
Skin irritation/corrosion	80%
Eye irritation	40%
Skin sensitization	60%
<i>In vivo</i> mutagenicity screening tests	91%
Acute toxicity	92%
Short-term repeated dose	92%
Sub-chronic toxicity	40%
Long-term repeated toxicity	40%
Development toxicity screening	86%
Development toxicity study	25%
Two-generation reproduction toxicity	10%
Short-term Daphnia toxicity	83%
Growth inhibition study on algae	85%
Short-term fish toxicity	85%
Long-term Daphnia toxicity	45%
Long-term fish toxicity	45%
Ready biodegradability	82%
Hydrolysis	45%
Adsorption/desorption	80%
Accumulation in aquatic species	80%

Modified from (Pederson et al., 2003), Table 4. Available at: <http://home.kpn.nl/reach/downloads/reachttestingneedsfinal.pdf>

² <http://www.pubfacts.com/search/ToxCast>

³ <http://ntp.niehs.nih.gov/results/hts/index.html#Publications>

clear: only about half of the substances qualify. However, this study was considered too politically sensitive to be published. We later included it in our 2009 re-analysis of the impact of REACH (Hartung and Rovida, 2009a):

“Moreover, it should be further highlighted that (Q)SAR calculations are based on specific organic chemicals with a well defined structure. For this reason, any inorganic compounds, organometallic compounds, mixtures, and UVCBs are excluded by default. Barrat et al. (2007) tested 400 chemicals randomly selected from ESIS HPV and LPV. The output was that (Q)SAR could be taken into consideration for only half of them, just by excluding inorganic and ionic compounds, complex mixtures, and those chemicals with no unique chemical structure.”

Our analysis at the time made very clear, that there was a desperate need for alternative methods to satisfy the information requirements, but that these could not be expected from (Q)SAR (Hartung and Rovida, 2009a): *“On average, the ‘good’ (Q)SAR approach implies a reduction of 4.6%, the ‘fair’ approach a reduction of 1.9%, and the ‘poor’ approach a reduction of 0.1%. If this is true, the (Q)SAR benefit is practically negligible.”*

This was very much challenged at the time as reported (Gilbert, 2009, 2010), but we pointed out that all previous analysis was based on a single database from 1991, not adjusted to the continued growth of the industry and the EU – from 12 member states only in the database to now 28 plus some associated countries applying REACH (Hartung and Rovida, 2009b). The prospects for using cellular assays are very dim (Rovida, 2010). Already our analysis of the 2010 registrations for the key hazard reproductive toxicity confirmed this (Rovida et al., 2011): of the 400 dossiers of REACH phase 1 analyzed, 40% used existing data, 27% read-across, 15% waiving, 5% provided no information, 0.5% suggested alternative methods and 10% new guideline studies. The use of (Q)SAR, though suggested up to 86% for developmental toxicity screening studies, was in fact negligible. Noteworthy, the proposed studies and studies carried out for REACH since 2007 only for reproductive and developmental toxicity and only for the first deadline of REACH would use 1,600,000 animals at € 210 million and exceed testing capacities in Europe.

Now, seven years later, we might ask how the 2009 predictions stand. Two of the three deadlines have passed. The EC had predicted 2,704 and 5,165 substances to be registered, respectively, i.e., a total of 7,869 substances. We said it could be 12,007; Katy Taylor from BUAV counted 13,328 by mid last year⁴ ... place your bets for 2018.

This means that we need alternative approaches to satisfy the legislative information needs more than ever (Schaafsma et al. 2009). As we pointed out (Hartung and Rovida, 2009b), another previously not anticipated problem is that the execution of the accepted testing proposals from the 2010 and 2013 deadlines coincides with the testing needs of the 2018 deadline, for which not testing proposals but complete dossiers must be registered. This means, our warning that we will face tremendous shortages of testing capacities just before the 2018 registration deadline

stands. It is foreseeable that many, especially small companies, will have problems to produce the required data and will have to rely on read-across approaches to meet the deadline.

4 The power of big data

Big data adds quantity and quality to the analysis of the subject at stake: Quantity is obvious, but because of the interdependence of data, there are network effects, i.e., the combined analysis allows seeing connections, diluting mistakes and consolidating the overall interpretation. And the latter simply works the bigger, the better. If I want to know something about one node in the network, it is informed by all others – the more, the closer the other node; the impact on the prediction also will be stronger the more pronounced the effect (potency) and the better the information (repetitions, quality assurance, etc.), but all nodes can contribute. In a very pragmatic way, we base the prediction on local similarity. In addition, we might use regional similarity, including somewhat less similar chemicals to assess whether this is a rather homogenous part of the chemical universe: Is the property of interest generally present or absent or are there activity cliffs, i.e., sudden changes of properties with small changes in structure (Fig. 2)? This might allow assigning some uncertainty measure to the prediction. The theoretical example of Figure 2 shows an untested substance (indicated by the

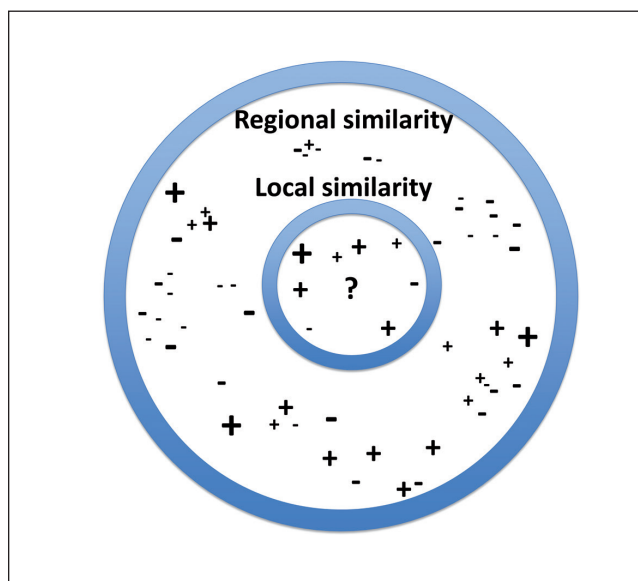


Fig. 2: Illustration of the concept of local and regional similarity when basing read-across on larger datasets

The graph assumes a similarity map, where more similar chemicals are placed closer to each other. In the center, indicated by “?”, is the not tested substance of interest. The different symbol sizes indicate differences in information value, such as certainty and potency.

⁴ bit.ly/1TR7nrC



question mark) for which most close neighbors are positive, but the region is full of both positive and negative substances, suggesting some uncertainty. The difference in size of the positive and negative symbols indicate that the information value might be different, e.g., because of available repeat tests with either homogenous or controversial results, quality scores or potency information.

This issue of ALTEX is to a large extent devoted to the REACH dataset made available by automated data download. Its organization in a computable format and making it machine-readable by natural language processing is a game changer. The full potential of the new database lies in computational toxicology. Similar chemicals have similar toxicity. The simplest form is read-across – concluding from a group of chemicals on the properties of non-tested ones. Arguably, it is also the most robust one as it is only based on local validity of the prediction, no attempt to generate a chemical “world formula.” We can now tell for any substance whether there are neighbors in the database and how they behave. Besides enabling predictions based on similarity of chemicals, this dataset allows three important things:

1. To assess how frequent certain hazards are (which is critical for the design of testing strategies (Hoffmann and Hartung, 2005)).
2. To assess objectively the quality of the traditional animal tests as many repeat tests are in the database.
3. To monitor the REACH registration process with its economical impact but also to identify possible mistakes in the registrations.

What we found is astonishing: for instance, far fewer chemicals are labeled with the different hazards than expected. Only one of four chemicals produced any effect in rabbit eyes, for example. But it is shocking how often some chemicals have been tested in the same test (see above). However, this waste of animals now shows, for the first time with so much evidence, how bad some animal tests are. Some are pretty much a lottery! The analysis of repeat tests as done in the articles presented in this issue of ALTEX for acute oral, eye irritation and skin sensitization (Luechtefeld et al., 2016b-d), show that some accepted protocols for animal tests are only 70-80% reproducible or concordant. Read-across actually might overcome even the problems of this lottery – if some neighbors were wrongly classified, the majority can still show the true result. Perhaps we can say: More trash data in, less trash out?

Other approaches resulting in big data in toxicology are the high-content (omics) and high-throughput (HTS) methods. We have discussed them already quite extensively in this series of articles (Hartung and McBride, 2011; Hartung et al., 2012). The Human Toxome Project (Bouhifd et al., 2015) is a key activity for making sense of big data as it aims to condense the multi-om-

ics data to pathways of toxicity (PoT) (Kleensang et al., 2014). In the end, it is not about recommending omics measurements as the tool for measurement but for the identification of the biomarkers (Blauboer et al., 2012). The expectations are high that these approaches will strengthen each other. The accompanying paper on biological support to read-across (Zhu et al., 2016) makes this case and gives examples. Omics and HTS fingerprints allow establishing biological similarity; pathways, such as the increasingly popular Adverse Outcome Pathways, make sense of these similarities, showing how joined mechanism is the basis of similar behavior. The fact that the REACH database already now has an overlap of 1,700 substances with Tox21 (Luechtefeld et al., 2016a) allows exploring this systematically.

5 The REACH database

The chemical universe can now be mined thanks to the new database and complemented with the developed guidance to help make sense of it (Ball et al., 2016, this issue). This database was first presented at AAAS, the annual general meeting of the American Association for the Advancement of Science, in Washington on February 12, 2016 and events in Brussels on February 26, 2016 and in Washington on March 1, 2016. The first four articles analyzing the database and the two read-across guidance documents were published online to coincide with their presentation at AAAS. The motivation for adopting such an approach was simple: First we just wanted to understand how the chemical universe looks with all the animal test data currently submitted to the public database of ECHA, the European Chemicals Agency in Helsinki. Due to the requirements of the REACH Directive, about 800,000 studies on 10,000 chemicals had already been registered in the ECHA database by December 2014. This is the largest toxicological database in the world.

It took our PhD student Tom Luechtefeld one year and extensive programming to train a computer to make sense of all the free text information stored in the public database of ECHA. Until then, this information was not all computer-readable. Previously, to state that a given chemical is an eye irritant, the information on file might read: “category 1”, “corrosive”, “cat. I” or “highly irritating.” Even for data for which there is a pull-down menu in IUCLID, registrants would often click “other” and enter free text.

The database was welcomed with great enthusiasm^{5,6,7,8} (Rabesandratana, 2016). We hope that some concerns by ECHA about making the database fully public⁹ can be overcome and most recent discussions in this direction are very promising. It would be a scandal if we had to test on animals although sufficient information is publicly available! This is

⁵ <http://www.ft.com/intl/cms/s/0/ca195b3c-da86-11e5-98fd-06d75973fe09.html>

⁶ <https://chemicalwatch.com/45087/us-team-maps-chemical-landscape-using-echa-data>

⁷ <http://www.rsc.org/chemistryworld/2016/02/chemicals-database-safety-animal-testing>

⁸ <https://www.washingtonpost.com/news/speaking-of-science/wp/2016/03/01/how-this-map-of-chemicals-could-help-reduce-animal-testing/>

⁹ <http://www.scientificamerican.com/article/legal-tussle-delays-launch-of-huge-toxicity-database/> and <http://www.nature.com/news/legal-tussle-delays-launch-of-huge-toxicity-database-1.19365>

not in the spirit of REACH or animal welfare legislation in Europe. And more than 20,000 chemicals are to be registered by 2018. Article 14 of the German Constitution uses a very nice phrasing: “Property entails obligations” – in this sense the ownership of these data should entail the obligation of the owners to make them fully available in the interest of chemical safety and animal welfare.

For a while around 2009, the author used a slide “*Status quo of in silico for REACH & TSCA*,” which stated among others “we cannot wait until REACH delivers the data to model.” Obviously, we did have to wait, but we might at least come in just in time for the last deadline in 2018. The good thing is that this deadline concerns the biggest number (20-40,000 depending on the forecast) of chemicals and there are only a limited number of hazards to be evaluated. The only animal tests required for all chemicals produced or marketed in the EU below 10 tons per year, i.e., the majority of substances, are acute oral toxicity and skin sensitization; and those were quite promising in our very preliminary computational predictions (Luechtefeld et al., 2016b,d).

6 Good Read-Across Practice

Read-across plays an important role in hazard assessment of chemicals under numerous internal and external regulatory programs. In many regulatory environments and for many toxicological endpoints, it is the only currently available non-animal alternative method. Use of read-across offers the potential for significant savings in terms of animal testing, product development time, and costs. However, acceptance of read-across by regulators has been slow and unpredictable (Ball et al., 2016).

Over the past five years industry has worked to address the challenges that read-across acceptance presents. There have been several projects and working groups set up to identify opportunities for making read-across more robust, less uncertain, and more available to a broader array of stakeholders. Several guidance documents (OECD, ECHA, ECETOC, etc.) aim to instruct a practitioner on the key considerations in preparation of a read-across justification. Until now this guidance has been very generic and, as a result, there is little understanding of what constitutes a universally acceptable best practice for read-across.

Upon request of its chemical/consumer product sponsor companies, namely BASF, Dow, ExxonMobil, Procter & Gamble and Shell, CAAT started an initiative to facilitate read-across use. A white paper was developed to scope the program (Pattlewicz et al., 2014). A team of thirty experts then developed guidance on how to do read-across properly. Until the beginning of 2016, the working title of the manuscript was actually “Good Read-Across Practice (GRAP) Guidance version 1.0”. Over the last few months, however, the discussion was very strongly impacted by the fact that ECHA published its Read-

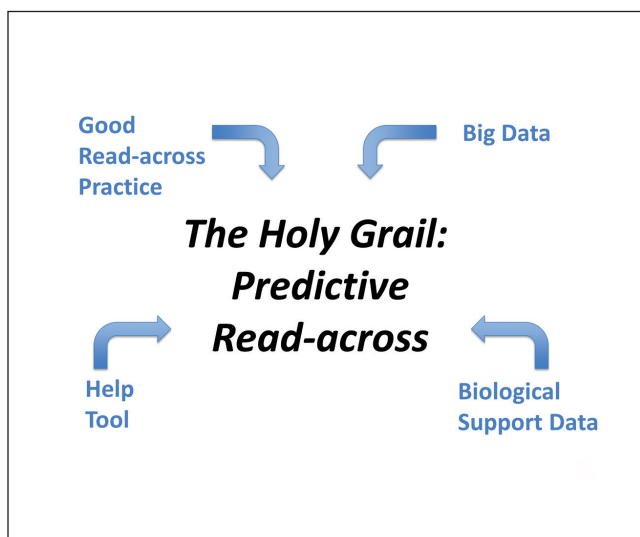


Fig. 3: Enabling good read-across

Different measures to support good read-across are shown, i.e., access to the necessary data, guidance on how to do it and how to support it with biological information and tools to carry it out more easily.

Across Assessment Framework – the RAAF¹⁰. This is a thorough, comprehensive and logical framework for evaluating the scientific robustness of a read-across assessment. ECHA has set a high bar for the information required to support a read-across hypothesis, and the RAAF contains sufficient detail to illustrate ECHA’s expectations for documentation and supporting evidence required in a read-across justification. The detailed RAAF provides industry with the opportunity to formulate a tool to “walk” a practitioner through the complicated process of constructing a robust read-across justification. This convinced the steering group to choose the more humble title “Toward Good Read-Across Practice (GRAP) Guidance,” seeing the need to adapt the guidance to respond directly to the RAAF, which could not be achieved in the timeframe before the planned stakeholder meetings.

Together with Nick Ball (Dow) and Sharon Stuard (P&G) we developed the idea to now develop a tool to guide read-across directly responding to the requirements of the RAAF. The respective proposal is under review by CEFIC-LRI. The use of this type of tool has the potential to help increase the transparency and acceptability of read-across by regulators. With the 2018 REACH registration deadline fast approaching, such a tool would be particularly useful for small company registrants less experienced at read-across. In addition, the tool may be useful for cleaning up past registrations, addressing future developments and REACH-like regulations elsewhere – all of which are of critical importance to the chemical industry. We need to give the companies not only access to the data, but guide them on how to do read-across with high certainty (Fig. 3).

¹⁰ http://echa.europa.eu/documents/10162/13628/raaf_en.pdf

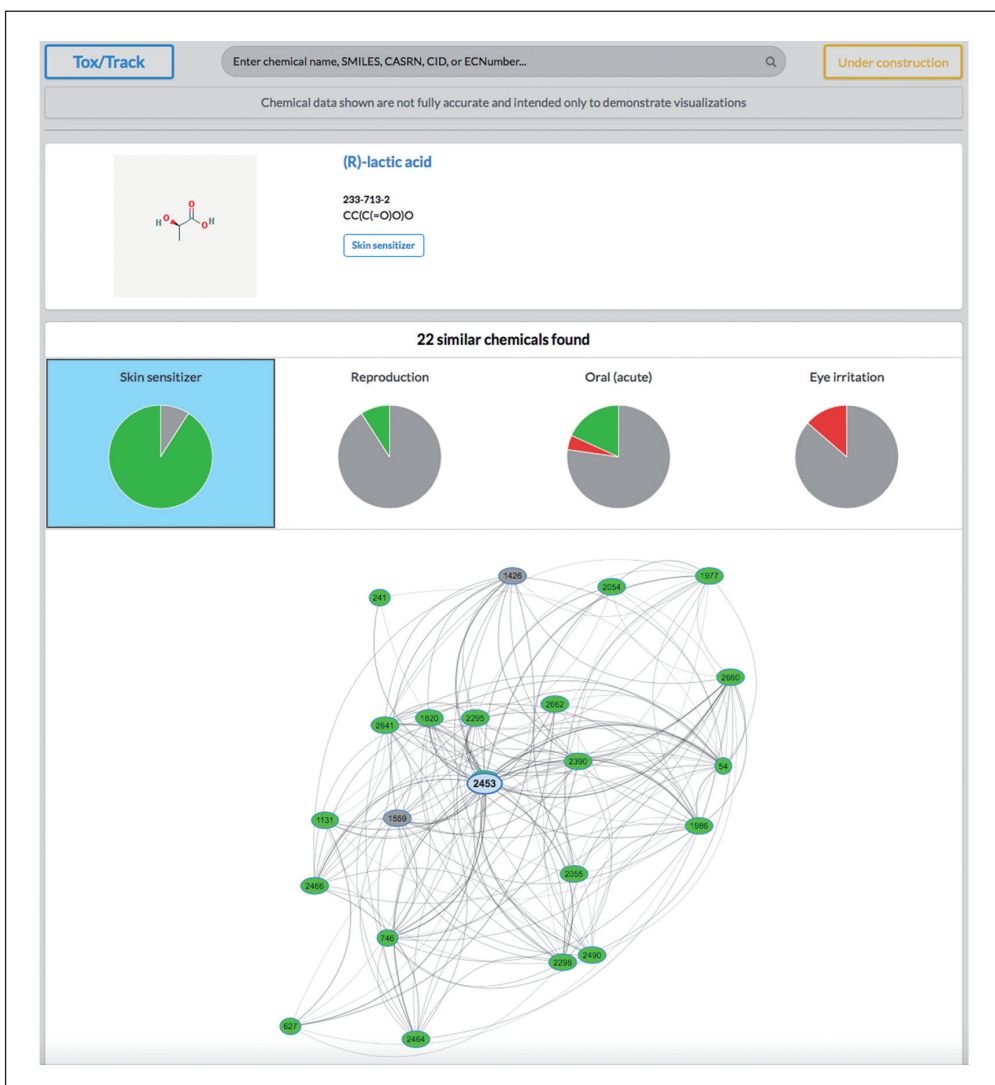


Fig. 4: Screenshot of the beta version of REACH-across™ – example lactic acid

Lactic acid was chosen as an example of a harmless substance. The light color of the field “skin sensitizer” below its smile code indicates that the substance has been tested and found negative. All substances of the database that have at least a Tanimoto similarity of 0.7 are shown as neighbors. The pie-charts indicate the properties of the 22 neighbors, in this case no positives (red) for skin sensitization or reproductive toxicity, but some for oral acute toxicity and eye irritation. The tool is under development with Insilica LLC and will be made available on the website <http://toxtrack.com>.

7 REACH-across – a new tool in the making

A number of tools have been developed to support read-across and are summarized in the two GRAP papers (Ball et al., 2016; Zhu et al., 2016). However, they all require expert knowledge and training. We therefore envisaged as part of Tom Luechtfeld’s PhD a web-based tool to do the read-across more or less automatically (REACH-across™). Nobody should test on animals just because it is too complicated to first try the existing data. However, the necessary programming of a user-interface is costly and traditional research funding does not typically cover this. Thus, a spin-off company tentatively called ToxTrack is currently being considered and is looking for investors. With the prospect of considerable cost and animal saving for REACH, this promises to be an economically interesting perspective.

The progress of the tool can be monitored on <http://www.tox-track.com>. Figures 4 and 5 show screen shots of the beta version. Figure 4 shows lactic acid, a harmless chemical, and no alerts are seen here for skin sensitization; however, it is an acid,

thus not surprisingly, some eye irritants are among its neighbors as summarized in the pie-chart in the upper right.

Figure 5 shows m-phenyldiamine, a well-known skin sensitizer and indeed many skin sensitizers are similar here. What is quite impressive is the number of similar chemicals found, in both cases here with a Tanimoto similarity of 0.7, which is often taken as a cut-off for read-across.

The REACH-across™ tool is under development. Currently, the input of untested structures is being implemented. Alternative similarity measures (other than the Tanimoto index) and prediction models as well as filters (e.g., considering only certain Klimisch quality scores) are foreseen. The tool will benefit from the inclusion of additional databases and an optional inclusion of proprietary data by the user. By creating an IUCLID-compatible reporting, the use for REACH registrations will be further simplified.

The large number of chemicals also allows evaluating this read-across systematically, e.g., by cross-validation, i.e., an approach where in a repeated permutation a subset of chemicals is

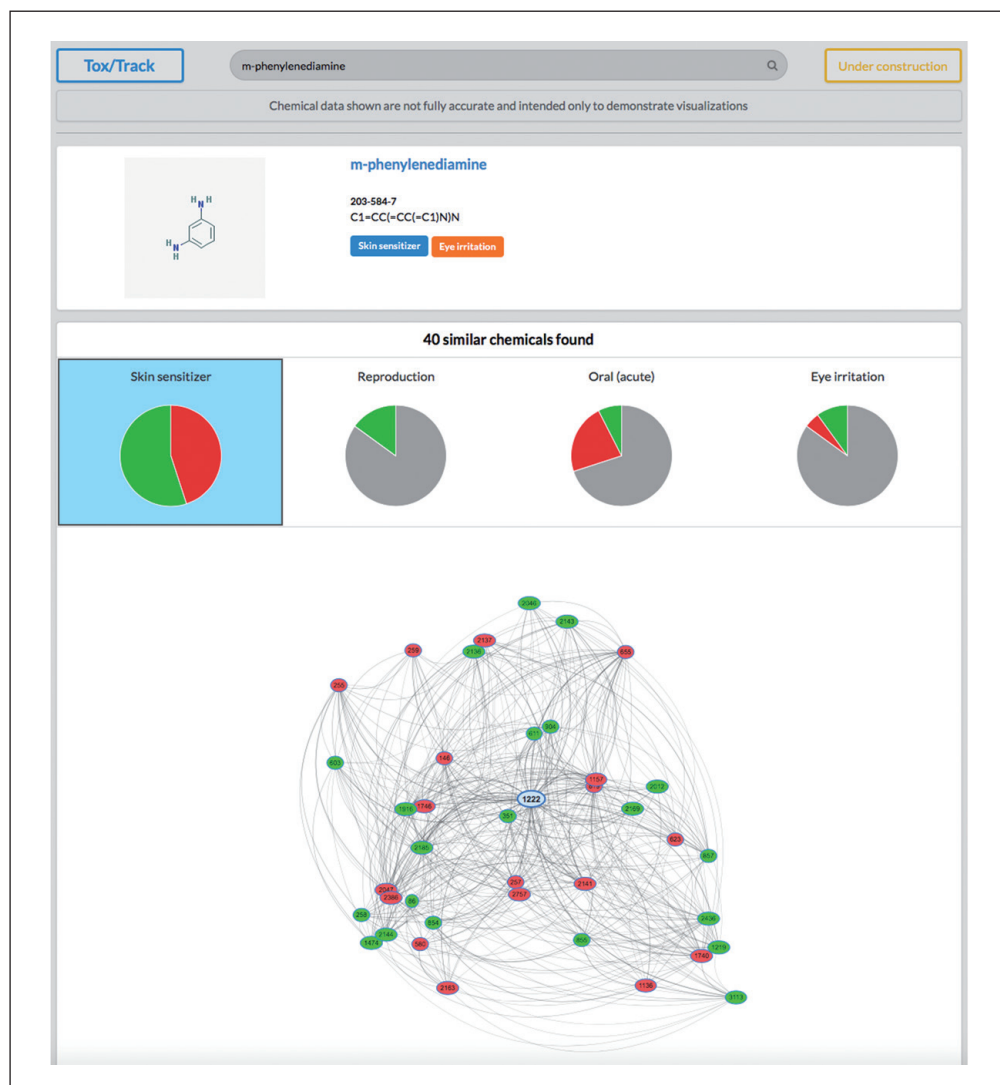


Fig. 5: Screenshot of the beta version of REACH-across – example m-phenylenediamine

m-phenylenediamine has been chosen as an example of a skin sensitizer. The dark blue color of the field “skin sensitizer” below its smile code indicates that the substance has been tested and found positive. All substances of the database that have at least a Tanimoto similarity of 0.7 are shown as neighbors. The pie-charts indicate the properties of these 40 neighbors, in this case positives (red) for skin sensitization, oral acute toxicity and eye irritation, but none for reproductive toxicity. The tool is under development with Insilca LLC and will be made available on the website <http://toxtrack.com>.

left out of the training set and predicted as if they were untested. Leaving certain subsets completely out of the optimization process or employing an external validation dataset as a challenge are also possible and discussions in this direction have started with NICEATM at the US National Toxicology Program. Rigorous validation of such a tool is of critical importance (Tropsha et al., 2003; Worth et al., 2004; Hartung and Hoffmann, 2009; Tropsha, 2010). Overfitting of any computer algorithms is a key problem (Hawkings, 2004): Computational scientists can fine-tune their calculations to fit any dataset – but in the end only this one dataset will be predicted, as general connections are no longer mirrored. Formal validation could bring the status of the REACH-across™ tool toward the one of a validated (Q)SAR, with enormous implications for the acceptability of its predictions in REACH.

And there are many more possible applications in addition to using it for REACH: Not only are there emerging similar programs in the US and Asia, such as the US Toxic Substances Control Act (TSCA) reauthorization as well as Korea and Tai-

wan REACH or similar plans in China. One area is Green Toxicology (Maertens et al., 2014). Imagine that a company wants to go for greener chemistry and get rid of the toxicants to protect workers and consumers. How would it find out where to start? Such a web tool could help identify what smells like a problem. Similarly, chemists could place their structures into the similarity map even before synthesizing them. If they get flags for toxicity, they might choose a different molecule and not find out about toxic liabilities only at the end of a costly product development (Voutchkova et al., 2010). The tool shall therefore enable direct comparison of substances to identify alternative chemistry for substitutions or the prioritization of substance lists, such as product ingredients or substances used in the supply chain.

The REACH-across™ tool is only one of many possible uses of the REACH database. With the 2018 deadline only 2 years away, the number of substances in it will at least triple and other databases also will add to it. It will be of utmost importance that a resource of this value is fully available for research and safety



assessments. The efforts of ECHA to make this happen are recognized and appreciated, as it is a fact that they are acting in a red-tape area. We should not exclude that it might ultimately take policy-makers to amend the legislation to make the free use of the data happen. This allows us to learn from the past to avoid the animal testing in the future.

References

- Ball, N., Cronin, M. T. D., Shen, J. et al. (2016). Toward good read-across practice (GRAP) guidance. *ALTEX* 33, 149-166. <http://dx.doi.org/10.14573/altex.1601251>
- Barrat, M., Hoffmann, S. and Hartung, T. (2007). The suitability of 400 REACH-relevant existing chemicals for evaluation by in silico methods. Unpublished results.
- Blauboer, B. J., Boelheide, 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>
- Bishop, P. L., Manuppello, J. R., Willett, C. E. and Sandler, J. T. (2012). Animal use and lessons learned in the U.S. High Production Volume Chemicals Challenge Program. *Environ Health Perspect* 120, 1631-1639. <http://dx.doi.org/10.1289/ehp.1104666>
- Bouhifd, M., Andersen, M. E., Baghdikian, C. et al. (2015). The human toxome project. *ALTEX* 32, 112-124. <http://dx.doi.org/10.14573/altex.1502091>
- ECHA (2010). Practical guide 10: *How to avoid unnecessary testing on animals*. Available at: http://echa.europa.eu/documents/10162/13655/pg_avoid_animal_testing_en.pdf
- Fourches, D., Muratov, E. and Tropsha, A. (2010). Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J Chem Inf Model* 50, 1189-1204. <http://dx.doi.org/10.1021/ci100176x>
- Gilbert, N. (2009). Chemical-safety costs uncertain. *Nature* 460, 1065-1065. <http://dx.doi.org/10.1038/4601065a>
- Gilbert, N. (2010). Crucial data on REACH not disclosed. *Nature* 464, 1116-1117. <http://dx.doi.org/10.1038/4641116a>
- Greene, N. and Pennie, W. (2015). Computational toxicology, friend or foe? *Toxicol Res* 4, 1159-1172. <http://dx.doi.org/10.1039/C5TX00055F>
- Hartung, T. and Hoffmann, S. (2009). Food for thought ... on in silico methods in toxicology. *ALTEX* 26, 155-166. <http://www.altex.ch/Current-issue.50.html?iid=107&aid=1>
- Hartung, T. and Rovida, C. (2009a). Chemical regulators have overreached. *Nature* 460, 1080-1081. <http://dx.doi.org/10.1038/4601080a>
- Hartung, T. and Rovida, C. (2009b). That which must not, can not be... A reply to the EChA and EDF responses to the REACH analysis of animal use and costs. *ALTEX* 26, 307-311. http://www.altex.ch/resources/altex_2009_4_307_311_Hartung.pdf
- Hartung, T. and McBride, M. (2011). Food for thought ... on mapping the human toxome. *ALTEX* 28, 83-93. <http://dx.doi.org/10.14573/altex.2011.2.083>
- Hartung, T., van Vliet, E., Jaworska, J. et al. (2012). Food for thought ... systems toxicology. *ALTEX* 29, 119-128. <http://dx.doi.org/10.14573/altex.2012.2.119>
- Hartung, T. (2013). Look Back in anger – what clinical studies tell us about preclinical work. *ALTEX* 30, 275-291. <http://dx.doi.org/10.14573/altex.2013.3.275>
- Hartung, T., Luechtefeld, T., Maertens, A. and Kleensang, A. (2013). Integrated testing strategies for safety assessments. *ALTEX* 30, 3-18. <http://dx.doi.org/10.14573/altex.2013.1.003>
- Hawkins, D. M. (2004). The problem of overfitting. *J Chem Inf Comput Sci* 44, 1-12. <http://dx.doi.org/10.1021/ci0342472>
- Hoffmann, S. and Hartung, T. (2005). Diagnosis: Toxic! – Trying to apply approaches of clinical diagnostics and prevalence in toxicology considerations. *Toxicol Sci* 85, 422-428.
- Hoffmann, S., Kinsner-Ovaskainen, A., Prieto, P. et al. (2010). Acute oral toxicity: Variability, reliability, relevance and interspecies comparison of rodent LD₅₀ data from literature surveyed for the ACuteTox project. *Regul Toxicol Pharmacol* 58, 395-407. <http://dx.doi.org/10.1016/j.yrtph.2010.08.004>
- Kleensang, A., Maertens, A., Rosenberg, M. et al. (2014). Pathways of toxicity. *ALTEX* 31, 53-61. <http://dx.doi.org/10.14573/altex.1309261>
- Luechtefeld, T., Maertens, A., Russo, D. 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. 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. 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. et al. (2016d). Analysis of publicly available skin sensitization data from REACH registrations 2008-2014. *ALTEX* 33, 135-148. <http://dx.doi.org/10.14573/altex.1510055>
- Maertens, A., Anastas, N., Spencer, P. J. et al. (2014). Green Toxicology. *ALTEX* 31, 243-249. <http://dx.doi.org/10.14573/altex.1406181>
- Nigsch, F., Macaluso, N. M., Mitchell, J. B. and Zmuidinavicius, D. (2009). Computational toxicology: An overview of the sources of data and of modelling methods. *Expert Opin Drug Metabol Toxicol* 5, 1-14. <http://dx.doi.org/10.1517/17425250802660467>
- 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>
- Pedersen, F., de Bruijn, J., Munn, S. and van Leeuwen, K. (2003). Assessment of additional testing needs under REACH – effects of (Q)SARs, risk based testing and voluntary industry initiatives. *JRC Report EUR 20863*. <http://home.kpn.nl/reach/downloads/reachttestingneedsfinal.pdf> (accessed 07.03.2016).
- Rabesandratana, T. (2016). A crystal ball for chemical

- safety. *Science* 351, 651-651. <http://dx.doi.org/10.1126/science.351.6274.651>
- Raunio, H. (2011). In silico toxicology – non-testing methods. *Frontiers Pharmacol* 2, 33. <http://dx.doi.org/10.3389/fphar.2011.00033>
- Rovida, C. (2010). Food for thought ... why no new in vitro tests will be done for REACH by registrants. *ALTEX* 27, 175-183. http://www.altex.ch/resources/altex_2010_3_175_183_FFT_Rovida.pdf
- Rovida, C., Longo, F. and Rabbit, R. R. (2011). How are reproductive toxicity and developmental toxicity addressed in REACH dossiers? *ALTEX* 28, 273-294. <http://dx.doi.org/10.14573/altex.2011.4.273>
- 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>
- Rusyn, I. and Daston, G. P. (2010). Computational toxicology: Realizing the promise of the toxicity testing in the 21st century. *Environ Health Perspect* 118, 1047-1050. <http://dx.doi.org/10.1289/ehp.1001925>
- Rusyn, I., Sedykh, A., Low, Y. et al. (2012). Predictive modeling of chemical hazard by integrating numerical descriptors of chemical structures and short-term toxicity assay data. *Toxicol Sci* 127, 1-9. <http://dx.doi.org/10.1093/toxsci/kfs095>
- Schaafsma, G., Kroese, E. D., Tielemans, E. L. J. P. et al. (2009). REACH, non-testing approaches and the urgent need for a change in mind set. *Regul Toxicol Pharmacol* 53, 70-80. <http://dx.doi.org/10.1016/j.yrtph.2008.11.003>
- Sun, H., Xia, M., Austin, C. P. and Huang, R. (2012). Paradigm shift in toxicity testing and modeling. *AAPS J* 14, 473-480. <http://dx.doi.org/10.1208/s12248-012-9358-1>
- Teubner, W. and Landsiedel, R. (2015). Read-across for hazard assessment: The ugly duckling is growing up. *Altern Lab Anim* 43, P67-71.
- Thomas, R. S., Philbert, M. A., Auerbach, S. S. 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. <http://dx.doi.org/10.1093/toxsci/kft178>
- Tropsha, A., Gramatica, P. and Gombar, V. K. (2003). The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR Models. *QSAR Combinat Sci* 22, 69-77. <http://dx.doi.org/10.1002/qsar.200390007>
- Tropsha, A. (2010). Best practices for QSAR model development, validation, and exploitation. *Molecular Informatics* 29, 476-488. <http://dx.doi.org/10.1002/minf.201000061>
- Voutchkova, A. M., Ferris, L. A., Zimmerman, J. B. and Anatas, P. T. (2010). Toward molecular design for hazard reduction – fundamental relationships between chemical properties and toxicity. *Tetrahedron* 66, 1031-1039. <http://dx.doi.org/10.1016/j.tet.2009.11.002>
- Weil, C. S. and Scala, R. A. (1971). Study of intra- and interlaboratory variability in the results of rabbit eye and skin irritation tests. *Toxicol Appl Pharmacol* 19, 276-360. [http://dx.doi.org/10.1016/0041-008X\(71\)90112-8](http://dx.doi.org/10.1016/0041-008X(71)90112-8)
- Worth, A. P., Hartung, T. and Van Leeuwen, C. J. (2004). The role of the European centre for the validation of alternative methods (ECVAM) in the validation of (Q)SARs. *SAR QSAR Environ Res* 2004, 15, 345-358. <http://dx.doi.org/10.1080/10629360412331297362>
- Young, D. and Martin, T. (2008). Are the chemical structures in your QSAR correct? *QSAR Comb Sci* 27, 1337-1345. <http://dx.doi.org/10.1002/qsar.200810084>
- Zhu, H., Bouhifd, M., Kleinstreuer, N. et al. (2016). Supporting read-across using biological data. *ALTEX* 33, 167-182. <http://dx.doi.org/10.14573/altex.1601252>

Conflict of interest

The author is currently exploring with partners opportunities to incorporate ToxTrack LLC in order to develop and commercialize the REACH-across™ tool described here.

Correspondence to

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