Toxicology Research

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/toxicology

Computational Toxicology, Friend or Foe?

Nigel Greene[1] and William Pennie[2]

[1]Worldwide Medicinal Chemistry, Pfizer Inc. Groton, CT 06340, USA.

[2]Drug Safety Research and Evaluation, Takeda Pharmaceuticals International Inc., Cambridge, MA 02139, USA.

Abstract

There is increasing public pressure to reduce animal testing and yet maintain public safety from exposure to chemicals either in the environment we live in, the food that we eat or the drugs that we take to treat illnesses. Computational approaches offer the attraction of being both fast and cheap to run being able to process thousands of chemical structures in a few minutes. As a result these approaches have seen an increase in interest and effort over the last decade most notably in the pharmaceutical industry where costs for new drug development is soaring and the failure rate for safety reasons is high. Many applications and approaches have been published covering a wide variety of different human and environmental health issues. As with all new technology, there is a tendency for these approaches to be hyped up and claims of reliability and performance may be exaggerated. So just how good are these computational methods? This review is intended to provide an overview of the state of the art in computational toxicology and to illustrate where some of the limitations of these approaches exist so that these valuable tools are applied and interpreted correctly.

Introduction

The field of computational toxicology has seen an increase in both interest and effort in recent years. This has been as a consequence of greater accessibility to toxicological databases, a drive to reduce animal testing wherever possible and the considerable practical/ economic pressures in

industries and agencies charged to test the safety of novel molecules in a more rapid and cost effective way. The specific expectations and requirements of computational toxicology tools and approaches are likely to be distinct across different applications.

In the early pharmaceutical discovery process, for example, computational toxicology offers an opportunity to quickly identify problematic chemical space, potentially contributing to reducing the well-recognized high cost of discovering and developing new drugs [1]. Early guidance from computational predictions can be used to prioritize testing and to support the selection of emerging drug candidates with the best profile with regards to toxicity hazard. Such toxicity hazard could be driven by chemical property space, general promiscuity to secondary pharmacology targets or specific interactions with off-target molecules, all molecular features which are amenable to computational predictive models. As the selected candidate molecules progress through discovery and into preclinical development they will be subjected to more exhaustive safety characterization regimen of *in vitro* and *in vivo* assessments and therefore computational toxicology applications of this type have most utility with a high positive predictive power but the user may tolerate some false negatives (as false negative compounds will hopefully be caught by "wet-lab" testing).

The situation is different when prioritization the testing of chemicals with occupational or environmental exposure for potential toxicity. Comprehensive toxicity testing is both time consuming and expensive and as a result only a small fraction of synthetic compounds with the potential to cause toxicity have been evaluated. For example, in the USA the Environmental Protection Agency (EPA) maintains a database of the health effects of environmental contaminants; the Integrated Risk Information System (IRIS, see http://www.epa.gov/iris/). The EPA has also a stated commitment to reduce the backlog of untested compounds in the Toxic

Substances Control Act (TSCA) inventory (for more details see: http://www.epa.gov/oppt/existingchemicals/pubs/tscainventory/) and for this application, and others, is investing in computational toxicology to provide more efficient ways to assess chemicals for human health effects [2]. Where the goal is to quickly prioritize compounds for more exhaustive toxicological assessment, computational toxicology could be used to help prioritize compounds which should be tested first. In this situation the ideal application will have a low false negative rate. However, current safety testing protocols are very slow, and the effective backlog is decades long. Under these circumstances if one develops a computational prioritization approach developed today with a low false positive rate, would focus on those chemicals of highest priority allowing time for a new a potentially more sensitive prioritization approach to be developed.

There are also distinct considerations when computational toxicology is being introduced into a regulatory framework. A notable example in this regard is effort made by the multinational International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), recognizing that *in silico* methods can essentially rule out mutagenic potential of pharmaceutical impurities. ICH has developed a particular guideline, M7, to harmonize the application of Quantitative Structure Activity Relationship modeling to predict the outcome of a bacterial mutagenicity assay to support hazard assessment; (http://www.ich.org/products/guidelines/multidisciplinary/article/multidisciplinary-guidelines.html). In certain cases the implemented M7 guideline can avoid having to test impurities or degradants for their ability to cause bacterial mutation. Analysis of data across multiple pharmaceutical companies has confirmed that structural assessment, supported by expert knowledge databases is sufficient to conclude that an impurity is non-mutagenic, with a

Negative Predictive Value of approaching 99% when the output of such expert systems is vetted by a human expert [3]. The ICH M7 takes a balanced approach to the application of computational toxicology in this space, favoring a low false negative rate through the application of two complementary methods (structure based and statistical) and the vetting of the predictions by a subject matter expert.

In this review we have attempt to summarize the current state of computational toxicology, including

- representative sources of public domain and commercial toxicology data, examples of the application of computational methods in hazard identification (genotoxicity, carcinogenicity, reproductive and developmental toxicity, skin and respiratory sensitization, hepatotoxicity, mitochondrial dysfunction),
- a consideration of computational modeling in predicting compound absorption, distribution and clearance
- a recognition of the limitations of computational toxicology
- some examples of consortia activity for data sharing and methodology development in the field of computational toxicology.

Data Sources for In Silico Modeling

A fundamental requirement for developing any computational approach is access to data on which to build the model. Ideally, this data set should contain an adequate number of compounds that represent the universe of possible scenarios to be modeled. Predictive models tend to work best when the effect being predicted is elicited through a single, common

mechanism. For toxicology endpoints, this typically means that the biological response or toxicity is caused by a discrete molecular interaction, for example, inhibition of a protein or enzyme such as the voltage-gated sodium channel, $Na_v1.5$, and the relative strength of the molecular effect of each chemical in the data set is known. In practice, this is often unobtainable as depending on the measurement, the molecular interactions of even closely related chemical structures maybe different. Similarly, the observed phenotypes in toxicity studies may well result from multiple, different mechanisms and so grouping chemicals by their molecular mechanisms of action is often challenging.

Public Sources for Toxicology Data

Under these circumstances, it is often necessary to have large numbers of chemicals in the hope that this will allow the modeling algorithm to "learn" the important features needed to elicit a response from having sufficient examples of similar structures. Unfortunately, there is no single source for toxicology data. However, a number of public data sources exist that can be used to retrieve and build data sets for modeling purposes. These include ToxNet (http://toxnet.nlm.nih.gov/index.html) maintained by the US National Library of Medicine and as the name suggests, this is a number of data sources assembled under a single user interface. It is searchable by chemical structure using the ChemIDPlus function which facilitates the searching of all sources using either exact, substructure or structural similarity. Other useful resources for accessing toxicology data include the US Environmental Protection Agency's National Center for Computational Toxicology ToxCast Initiative (http://epa.gov/ncct/index.html) which includes links to multiple sources for chemical structures, in vitro assay profiles and in vivo toxicology studies as well as the ACToR database. The US Centers for Disease Control and NIOSH maintain a potentially useful resource in the Registry of Toxic Effects of Chemical Substances

(RTECS) (http://www.cdc.gov/niosh/rtecs/default.html) however access to this data source requires the user to have a license for the database although it can be searched via the TOXNET interface.

More focused data sets can be found in the public domain and one good example of this is the Open TG-GATEs database (http://toxico.nibio.go.jp/english/index.html) [4] that contains both preclinical and transcriptional profile data on a data set of chemicals, primarily drugs, that have been linked to causing liver and/or kidney injury in humans. On a similar vein, the US Food and Drug Administration are now making the Summary Basis of Approval documents freely available via their website (http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm). These documents contain summaries of the clinical and preclinical data supplied by the company sponsor at the time of filing the New Drug Application (NDA), but require extensive data extraction and reformatting in order to be amenable to modeling techniques. Similarly, the FDA makes their Adverse Event Reporting System database (FAERS) ,that contains adverse event and medication error reports submitted to FDA, freely available to the public via their website but this requires the user to be familiar with the creation of relational databases in order to search the data effectively, (http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/). However, care should be taken as there is no certainty that the adverse event was actually due to the product it is being associated with and similarly not all adverse events are reported to the FDA. Factors such as publicity about an event and the length of time a product is on the market can influence whether an adverse event is reported. It should not be used to calculate the frequency of a particular event.

Sources such as PubChem (https://pubchem.ncbi.nlm.nih.gov/) and ChemBL (https://www.ebi.ac.uk/chembl/) [5] can be useful for compiling data sets against discrete molecular targets such as the hERG channel linked with causing QTc prolongation or the serotonin receptor type 2b (5HT2b) where agonism of this receptor has been linked to causing cardiac valvulopathy [6].

Commercial Toxicology Data Sources

Besides the public domain sources, there are commercial databases available that can be used for model building. These commercial sources offer the attraction of having some added degree of curation on the data contained in the database. Data quality is an important consideration for any model development as the introduction of noise through the incorrect annotation of activity can lead to a decrease in model performance. Some examples of commercial sources include the Vitic database from Lhasa Limited [7] that originated from a consortia effort managed by the Health and Environmental Sciences Institute (HESI), part of the International Life Science Institute (ILSI). Vitic is searchable by chemical structure as well as by toxicological effects that include genotoxicity, carcinogenicity, hepatotoxicity, HERG, and sensitization. It has also been used as the main repository for several cross-industry collaborations that share data on the toxicity of various chemicals most prominent of which is the Innovative Medicines Initiative (IMI) eTox project.

Other example commercial data sources include but are not limited to ToxWiz from Cambridge Cell Networks (http://camcellnet.com/products/toxwiz/) that contains a database of relationships between chemicals, proteins, gene and pathologies. In addition to the database, ToxWiz also incorporates predictive models as part of the package. PharmaPendium from Elsevier is another example that offers access to both preclinical and clinical safety data by incorporating Food and

Drug Administration (FDA) and European Medicines Agency (EMA) approval summary documents on new drug applications as well as the FAERS database (https://www.pharmapendium.com).

It should be noted that, irrespective of the origin of a particular data source, both in vivo and in vitro data are subject to numerous sources of errors and noise. These imperfections in the data used to train and evaluate models are often one of the reasons for the lack of predictive power observed with computational approaches.

Approaches for Predicting Toxicity

Hazard Identification

Most *in silico* methods for toxicity prediction have focused on hazard identification, for example does a compound have features or properties that have been associated with liver injury. However the majority of these computational approaches do not tell you the dose at which these effects are likely to happen. Models exist for a variety of human health endpoints but depending on the endpoint being predicted the accuracy of these can vary dramatically. Here we discuss briefly the models and approaches available for predicting genotoxicity, carcinogenicity, skin sensitization, reproductive and developmental toxicity and hepatotoxicity.

Genotoxicity

The *in silico* prediction of genotoxicity has been a major research focus since the publication of structural alerts for DNA reactivity from Ashby and Tennant [8] over 3 decades ago. Access to large public domain data sets [9-12] have helped to stimulate progress and have

resulted in a fair degree of success in the prediction of genotoxicity and in particular, the prediction of the Ames salmonella assay for mutagenicity [13, 14].

There are many commercial systems available and all have their strengths and weaknesses. Commercially available software packages such as Derek for Windows (DfW) [15] now called Derek Nexus, MC4PC [16], and Leadscope Model Applier (LSMA) [17] are now commonly used within the pharmaceutical industry for the prediction of genotoxicity and other toxicological endpoints. Other freely available systems like Toxtree [18] are also being evaluated for its usefulness. Their comparative performances have been extensively reviewed and published [19, 20] but it is clear that no single system performs significantly better than any of the others.

The comparison of systems is heavily dependent on the source of the data being used to evaluate a system's performance, for example, a public domain data or a set of proprietary pharmaceutical-like compounds. The overall concordance of these tools range between 70% and 85% and it is worth noting that these values are close to the inter- and intra-laboratory reproducibility of the Ames assay, reported as 87% [21]. However, a system's sensitivity i.e. its ability to accurately predict an Ames positive compound, can vary much more dramatically from up to 85% for public domain sets to just 17% for some proprietary data sets [19].

This variability in performance most likely results from the fact that most software applications are trained using only the public data sets which tend to be industrial and environmental chemicals. Most pharmaceutical-like compounds, i.e. the active pharmaceutical ingredients in drug products, tend not to contain the classical DNA-reactive functional groups that are a common cause of genotoxicity. It is possible that these pharmaceutical compounds

undergo rare or unusual metabolic activation and hence are not obviously reactive in of themselves, or they elicit a positive response in the Ames assay through non-reactive mechanisms such as intercalation [22]. It should also be noted that the ratio of positive to negative compounds are significantly lower with the pharmaceutical-like data sets where typically only 6-10% are shown to be mutagenic in the Ames assay compared to 40-60% in the case of some public domain sets and so maintaining an appropriate balance between correct positives and false positives becomes a key challenge for any computational tool.

It is worth noting that although models exist for the prediction of chromosomal aberrations such as clastogenicity and anugenicity, these systems are generally less accurate and are not commonly used in industry settings.

As mentioned in the introduction, computational prediction of mutagenicity has now matured to the extent that it is been incorporated within a formal regulatory guidance document for the first time in the ICH M7 guidance on the "Assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk" (http://www.ich.org/products/guidelines/multidisciplinary/article/multidisciplinary-guidelines.html).

Carcinogenicity

Various methods for the structure-based prediction of carcinogenicity, as with genotoxicity, have been in developed over the last several decades including some commercial applications such as Derek, Case Ultra, Leadscope Model Applier, ToxTree and OncoLogic. However, the degree of success of these methods have been more limited in the prediction of carcinogens mainly due to the fact that less data is available in the public domain and the

complexity of the endpoint itself. Carcinogenicity can occur through both genotoxic and non-genotoxic mechanisms. Most structure-based approaches are able to more accurately predict the former rather than the latter due to the successes of being able to predict DNA-reactive compounds. However, some systems such Derek contain structural alerts specifically targeting certain classes of non-genotoxic carcinogens. Other predictive packages such as Case Ultra however, do not always differentiate these two classes in their predictions.

There have been some comparisons of the performance of computational models for carcinogenicity most notable of which were the two prospective exercises conducted by the National Toxicology Program (NTP) in the mid to late 1990's. In these exercises the NTP invited interested parties and developers of models to predict and publish them on a set of chemicals that were scheduled to be tested by them in the gold standard two year rodent bioassay. Once the tests had been completed the in vivo results were compared to the predictions. The first set of 65 chemicals were reasonably well predicted with computational models achieving between 50-65% accuracy. However, the second set consisting of only 30 chemicals were not predicted so well by the in silico systems and tended to over predict non-carcinogens as carcinogenic [23].

The evaluation of computational models depends heavily on both the size and composition of the data sets being used for the comparison. Unfortunately, the cost and time required to conduct carcinogenicity tests inevitably means that the size of any evaluation set is small and so these exercises can be somewhat misleading. Despite this limitation, efforts to predict carcinogenicity through structure-based approaches continue to be developed with some recent examples from Fjodorova *et al.* [24]and Kar *et al.* [25].

Reproductive and Developmental Toxicity

Developmental and reproductive toxicity (DART) occurs through many different mechanisms and involves a number of different target sites, making it very difficult to predict this end point [26]. In silico prediction of reproductive and developmental toxicity have been limited by the quantity and quality of data available in the public domain for model development. Most of the published QSAR development has been done through collaborative projects with the computational toxicology group within the US Food and Drug Administration using data collected from preclinical and clinical data submitted by pharmaceutical companies.

Matthews *et al*. [27] claim to be able to predict male and female reproductive toxicity, fetal dysmorphogenesis, functional toxicity, mortality, growth, and newborn behavioral toxicity with high specificity (i.e. the number of correctly predicted negatives) and positive predictive value (i.e. the number of correct positive predictions when compared to the total number of positive predictions) of greater than 80%. However the sensitivity (i.e. the number of correctly identified positive compounds) was often less than 50%. These results were obtained through a 10-fold cross-validation exercise where 10% of the data set is withheld for testing and a model built on the remaining 90%. Unlike the NTP carcinogenicity exercises, to date there has been no published prospective tests of performance of these models and so the accuracy against a set of novel compounds cannot be ascertained. These published models are available in commercial packages such as Case Ultra and Leadscope Model Applier. In addition, Derek Nexus also contains some structural alerts for DART effects although these alerts and their respective performance have not been formally published [28].

Wu *et al*. [26] recently published an empirically-based decision tree for determining whether or not a chemical has receptor-binding properties and structural features that are

consistent with chemical structures known to have toxicity for DART end points. As with the above models and structural alerts, the performance of this decision tree has not been independently assessed and so its performance for truly novel chemical series that have not been previously tested may well be limited.

Skin Sensitization

As with other key toxicity endpoints, structural alerts for skin sensitization have been developed in response to changes in the regulatory environment, most notably through changes in legislation in Europe banning the testing of cosmetic ingredients in animal models. Skin sensitization is primarily driven through modification of proteins usually through covalent binding forming haptens that then trigger the immune system and generate an inflammatory response [29]. This requirement for chemical reactivity makes the prediction of skin sensitizers more feasible and, as with mutagenicity, there has been substantial progress in this area. Structural alerts for skin sensitization have been implemented in commercial systems such as Derek Nexus and are also implemented in ToxTree. Similarly, numerous QSAR methods have been published looking at the prediction of allergic contact dermatitis (ACD) based on specific chemical classes or on non-congeneric data sets. The relative performances of these approaches have not been reviewed so it is difficult to make comparisons on the relative merits of one approach over another.

The manifestation of ACD is moderated by the ability of the chemical to penetrate viable epidermis and so while a chemical may possess the intrinsic capability to cause skin sensitization the inability of the chemical to penetrate the skin may result in the apparent lack of toxicity.

Respiratory Sensitization

There are currently no validated or widely accepted models for identifying and characterizing the potential of a chemical to induce respiratory sensitization yet this may lead to severely incapacitated human health [30]. There is a great deal of uncertainty about the immunological mechanisms through which respiratory sensitization may be acquired. Despite the lack of a universally accepted test method, REACH regulations and others still require the assessment of respiratory sensitization as part of a risk assessment. The REACH guidance describes an integrated evaluation strategy that includes a consideration of well-established structural alerts and existing data (whether it be derived from read-across, (Quantitative) Structure Activity Relationships ((Q)SAR), *in vivo* studies etc.).

Efforts to model respiratory sensitization *in silico* have been variable and to some extent mirror those for skin sensitization itself. Structural alerts have been developed notably by Aguis et al. [31, 32] and more recently by Enoch et al. [33]. Typical alerting groups have been encoded into the Derek Nexus knowledge based expert system developed by LHASA Ltd. Other efforts have been focused on trying to establish statistical QSAR models; examples include those first derived by the developers of MCASE [34], Jarvis *et al.* [35] and more recently by Warne *et al.* [36] who investigated the use of pattern recognition methods to discriminate between skin and respiratory sensitizers.

As with other toxicological endpoints, there has been no published comparison of these methods for prediction and so it is difficult to draw conclusions on the relative merits and accuracy of the models.

Hepatotoxicity

Drug-induced liver injury (DILI) is a major issue of concern in the pharmaceutical industry and has led to the withdrawal of a significant number of marketed drugs [37, 38]. Adverse effects can range from hepatic enzyme elevations to liver failure [39, 40] and are often difficult to predict in the preclinical stages. For the pharmaceutical industry hepatotoxicity discovered late in development or after the launch of the drug, leading to its withdrawal, has huge financial implications [41]. As a result of this interest, numerous in silico approaches for predicting hepatotoxicity have been developed. These range from structural alerts associated with causing DILI [42, 43] to QSAR methods [44]. Most of these methods claim to have a sensitivity and specificity between 65-70% depending on the method and test set. However, no independent evaluation has been published so true head to head performances are difficult to ascertain.

Clearance and metabolism of xenobiotics (foreign compounds, including drugs) into hydrophilic metabolites to facilitate their excretion is one of the liver's main physiological roles. As such, a postulated mechanism of DILI involves the generation of reactive metabolites that covalently bind to proteins and subsequently cause cellular damage or stimulate the immune system. This mechanism may account for some of the success of structure-based methods for predicting DILI. However, as with DART and carcinogenicity, the mechanisms involved in DILI are diverse and often complex making the accurate prediction of this using QSAR approaches challenging. Similarly, factors and properties that influence exposure at the site of injury will undoubtedly have a significant impact in the manifestation of DILI even if the chemical possess the intrinsic capability to cause the injury.

Mitochondrial Dysfunction

Cells derive their energy from adenosine triphosphate (ATP), a ubiquitous chemical that can be synthesized in the cytoplasm through glycolysis but is predominantly generated in the mitochondria through oxidative phosphorylation. Oxidative phosphorylation is a process by which the bond energy in nicotinamide adenine dinucleotide is extracted to create a proton-based electrochemical gradient that drives the phosphorylation of adenosine diphosphate (ADP) by ATP synthase. The disruption of this critical mitochondrial function is proposed to have severe implications for organ health, but such a causal hypothesis is challenging to demonstrate, notably because organ toxicity is a complex and multifactorial process [45]. Mitochondrial uncoupling mechanisms are common toxic pathways and a correlation of outcome with p*K*a and clogP for phenolic mitochondrial uncouplers has been demonstrated. In addition, for the benzoic acid class of NSAID, a correlation of the HOMO LUMO gap with cytotoxicity was also noted, probably reflective of the ease of oxidation of the diphenylamine template of many compounds in this class. There is no doubt that there is a link between the physicochemical properties of a compound and the risk of *in vivo* toxicological outcomes.

Off-target pharmacology

Pharmacological interactions with specific proteins been linked to causing a variety of adverse effects in humans. For example, the human ether-a-go-go-Related Gene (hERG) channel is involved in the repolarizing I_{Kr} current in the cardiac action potential and so inhibition of this important ion channel can result in a prolonged QT interval which has been linked to causing potentially fatal cardiac arrhythmias. The structural properties associated with hERG inhibition have been extensively studies and often include a basic center and one or more lipophilic chains.

However, there are numerous other examples of pharmacological interactions that have been associated to adverse events where the structural requirements are not clearly understood. These include but are not limited to the agonism of 5HT2b associated with cardiac valvulopathy [46], inhibition of VEGF signaling pathways has been associated with causing hypertension etc. For a more extensive review of pharmacology and the associated effects of a variety of proteins please refer to http://www.iuphar-db.org/index.jsp.

In addition to discrete interactions, it has been shown that promiscuity across multiple pharmacological targets at a concentration of 10μM can lead to an increased likelihood of observing toxicity in vivo at low exposures [47, 48]. In addition, interactions with multiple ion channels and other CNS protein targets can lead to an increased risk of seizures in preclinical studies. Increased target promiscuity has been associated with higher lipophilicity (LogP) and low polar surface area as well as pKa [47]. Acidic molecules tend to interact with different classes of receptors such as cyclooxygenases and the nuclear hormone receptors such the PPARs whereas basic molecules tend to interact with the aminergic G-protein coupled receptors.

Endoplasmic Recticulum (ER) Stress

The endoplasmic reticulum (ER) is a cytoplasmic organelle involved in protein folding, maturation and secretion, cholesterol and lipid biosynthesis, as well as gluconeogenesis, and it is the main calcium storage compartment in the cell. Disturbances in redox regulation, calcium regulation, glucose deprivation, and viral infection or the over-expression of proteins can lead to endoplasmic reticulum stress (ER stress), a state in which the folding of proteins slows, leading to an increase in unfolded proteins. This stress is emerging as a potential cause of damage in hypoxia/ischemia, insulin resistance, and other disorders has been implicated in many diverse diseases and has also been linked to pharmacologically-induced toxicity [49].

Key structural properties that have been shown to influence the induction of ER stress are the high receptor promiscuity, high lipophilicity, low polar surface area and low passive permeability which can be estimated computationally (see sections herein).

Reactive metabolites

Many idiosyncratic adverse effects of drugs have been associated with the formation of reactive metabolites. The precise mechanisms of these adverse effects remain unclear; however, it is believed that the majority of these reactions are caused by immunogenic conjugates formed from the reaction of an electrophilic reactive metabolite of a chemical with cellular proteins resulting in direct cellular dysfunction or an immune response via the formation of a hapten. Structural alerts relating to the potential for generating reactive metabolites have been published [50] and could be used in the identification of potential hazards associated with a chemical.

Phospholipidosis

Phospholipidosis is a condition primarily characterized by excessive accumulation of phospholipids in different cell types, giving the affected cells a finely foamy appearance. This phospholipid storage disorder is characterized by the excessive accumulation of phospholipids and the inducing drug in the lysosomes of the affected tissues. The hallmark feature of phospholipidosis is the formation of the characteristic lamellar bodies in cells, which can be detected by electron microscopy. In case of alveolar phospholipidosis, foamy macrophages accumulate within the alveolar spaces of the lung. Whether this effect is considered a toxic response or an adaptive response remains an unanswered question but in silico approaches to predict the likelihood of observing this biological response have been developed and extensively published [51]. Most models use a measure of a chemicals amphiphilicity, pKa and lipophilicity

(LogP) where amphiphilic, cationic (basic) molecules are more likely to induce phospholipidosis.

Nuclear hormone receptor activity and steroidogenesis

Chemicals that disrupt the endocrine system have been linked to a wide variety of human health effects depending on the pathway that is disrupted. It has been hypothesized that exposure to xenoestrogens and xenoandrogens can lead to an increased prevalence of breast cancer, prostate cancer and testicular cancer. In addition to cancer, infertility and loss of sperm count are likely associated with the exposure to chemicals the disrupt the endocrine system. Most of the effects of xenoestrogens and xenoandrogens are mediated *via* estrogen receptors (ER) and androgen receptors (AR). Numerous compounds, including environmental chemicals such as DDT and phthalates, have been classified as ER and/or AR modulators, acting either as direct agonists or antagonists or altering receptor expression.

Various in silico approaches can be used to predict the endocrine effects of chemicals from (Q)SAR methods to pharmacophore and docking [52]. These methods have been used to prioritize chemicals for further experimental testing and confirmation.

Moving towards Risk Assessment

When assessing safety risk it is necessary to not only identify the potential hazards presented by a chemical but also to combine them with the level of exposure where these occur. Exposure to a chemical is usually described in terms of an amount or mass of chemical per unit of body weight per day. In pharmaceutical development, the exposure of an individual to chemical is often known but in the case of environmental contamination the exposure is often difficult to assess. However, to fully understand and therefore model the relationship between the molecular

properties of a chemical and its ability to cause toxicity it is necessary to know the exposure of the chemical at the circulating plasma, organ, tissue or cellular level. Recent efforts in the field of computational toxicology have begun to focus on this aspect of toxicity prediction. Here we describe some of the molecular properties that can influence the exposure at which toxicity is observed and efforts to use these to predict toxicity.

Absorption

Toxicology is founded on the basic principle of the dose determines the poison. Everything is potentially toxic and it is the administered dose that defines whether something will elicit observable toxic effect. Therefore the expression of human and mammalian toxicity for the most part is predicated on the absorption of the substance into the circulating bloodstream of the organism. Oral bioavailability (F) is a product of fraction absorbed (Fa), fraction escaping gut-wall elimination (Fg), and fraction escaping hepatic elimination (Fh). In 1997, Lipinski proposed the 'rule of 5' as set of guiding principles for designing oral drugs that had good absorption from those that were more likely to be poorly absorbed [53]. This rule of 5 comprised of 4 rules of MW<500; cLogP<5 number of hydrogen bond donor atoms HBD <5; and the number of hydrogen bond acceptor atoms HBA <10. Chemicals are less likely to have good oral absorption if they violate 2 or more of these rules. However, there are numerous examples of drugs that adhere to these criteria yet have bioavailability of <30% of the administered dose, for example, acyclovir has a fraction absorbed below 30%. Similarly, there are examples of drugs that do not fulfill these criteria yet are readily absorbed, for example, cyclosporine has a bioavailability of up to 60% and thus has reasonable absorption. Since Lipinski's rules are based on the distribution of these properties across several thousand drugs outliers are expected. Most

drugs rely on passive transport across membrane barriers but active transport mechanisms, both uptake and efflux, also exist and these might explain some of these surprises in bioavailability.

Trend analysis clearly indicated molecular weight (MW), ionization state, lipophilicity, polar descriptors, and free rotatable bonds (RB) influence bioavailability [54]. These trends were due to a combination of effects of the properties on Fa and first-pass elimination (Fg and Fh). Higher MW significantly impacted Fa, while Fg and Fh decreased with increasing lipophilicity. Parabolic trends were observed for bioavailability with polar descriptors. Interestingly, RB has a negative effect on all three parameters, leading to its pronounced effect on bioavailability. In conclusion, physicochemical properties influence bioavailability with typically opposing effects on Fa and first-pass elimination. This analysis may provide a rational judgment on the physicochemical space to optimize oral bioavailability.

Dermal absorption

Dermal or topical absorption predictive models have been in existence since the early 1990's when Potts & Guy [55] published a simple model that showed a relationship between the molecular volume or molecular weight and the lipophilicity of a chemical and its ability to permeate the skin. Although many other models have been proposed and published most rely on these key properties to determine the skin permeation rate.

Ocular and Respiratory exposure

There are few published reports on the physicochemical properties of compounds that influence the likelihood of either ocular or respiratory exposure. However, it can be speculated that the following may play a role in the accidental exposure to chemicals via the ocular or respiratory routes:

1) Highly volatile chemicals or those with low melting and boiling points will have an increased risk of exposure to fumes.

2) Liquids or solutions of a chemical may result in exposure through splashing.

3) Chemicals that can exist as fine particles or dust that can be inhaled or get into the eye.

Once the chemical has come in contact with the surface of the eye or the lungs then the properties that influence oral or dermal absorption will most likely have a similar influence on the absorption through these alternative routes.

Distribution

One important measure of compound distribution that has been demonstrated to have a link to toxicity in mammals is the concept of Volume of Distribution, Vd. This is defined as the theoretical volume that the total amount of administered drug would have to occupy (if it were uniformly distributed), to provide the same concentration as it currently is in blood plasma. Higher values of Vd shows that the drug is more diluted than it should be in the bloodstream implying that more of the chemical is distributed into the tissues. Drugs with high lipophilicity (non-polar), not ionized at physiological pH or have low plasma protein binding have higher volumes of distribution than drugs which are more polar, more highly ionized or exhibit high plasma protein binding. Vd directly influences the half-life of a compound whereby large Vd leads to a longer half-life, i.e. prolongs the duration of exposure.

Several publications exist that look at the structure-based prediction of volume of distribution which could be used in the assessment of hazard and risk given the above relationship of Vd in determining the observed LOAEL in a rodent study. Lombardo *et al*. [56] looked at human Vd_{ss} values and determined that the physicochemical properties of LogD and pKa along with a

measurement of the plasma protein binding in human plasma would be enough to make a reasonable prediction of the human Vd_{ss}. Gombar *et al*. [57] have also published on using QSAR models to predict the VD_{ss} and clearance values (CL) using only structural descriptors. Other computational models of clearance such as the work by Hsiao *et al.* [58] have similarly shown a strong correlation between clearance and lipophilicity or LogP and polar surface area descriptors.

Plasma Protein Binding

In general, molecules within in vivo systems are either bound to proteins and lipids in plasma (more commonly referred to as plasma protein binding (PPB)), or to proteins and lipids in tissues, or are free (that is, unbound) and diffuse among the aqueous environment of the blood and tissues [59]. Among other factors, PPB strongly influences volume of distribution and half-life of chemicals [60]. In most cases it is the unbound fraction of molecules that interact with protein receptors to produce a pharmacological effect on the system. These pharmacological interacts can be considered to be either therapeutic or toxicological effects and are often context dependent. For example, a molecule that produces a pharmacological effect that results in a drop of blood pressure can be deemed to have a therapeutic effect in patients suffering from hypertension but if administered to patients with hypotension this can be seen as a potential adverse effect.

Chemicals that act via a pharmacological interaction with a protein receptor, such as the estrogen receptor, that are also highly bound to plasma proteins will generally require higher doses to achieve the required free concentrations to elicit an equivalent response to a chemical that has a lower PPB level provided the rate and fraction absorbed for both are equivalent. Physicochemical properties such as lipophilicity (LogP) and pKa can have a strong influence on

the degree of PPB observed for a given chemical. In general, molecules with high lipophilicity will have a lower fraction unbound and acidic molecules will similarly have a greater degree of PPB than basic compounds [61].

Clearance (or Metabolism and Excretion)

Clearance (CL) describes is a proportionality factor that relates the rate of elimination of chemical to its concentration in plasma. For first-order elimination, CL has a constant value and is measured by the plasma volume completely cleared of the chemical per unit time (e.g. mL/min). In nonlinear elimination, CL depends on plasma concentration. Total clearance describes the elimination of a chemical from the body without identifying the mechanisms involved in the process but most chemicals are eliminated primarily via the liver and/or kidney.

Clearance is one of the most important of all pharmacokinetic parameters. It is affected significantly by the binding of chemicals to serum proteins and only the free (unbound) fraction of a compound is able to be cleared. The unbound clearance CLu is the clearance with reference to unbound clearance in plasma and is independent of the plasma protein binding so only depends on chemical structure and properties. In studies published the rate of clearance is heavily dependent on the lipophilicity of molecule at pH 7.4 as expressed by the term LogD7.4 which is ultimately related to the LogP and pKa of a compound [62].

Physicicochemical Properties Associated with Toxicity

Recent studies have looked at the relationship between physicochemical properties and a chemical's ability to cause to *in vivo* toxicity at low plasma exposures. For example, Hughes *et al.* [47] reported that compounds with an cLogP value greater than 3 and topological polar surface area (TPSA) of less than 75Å^2 were 6 times more likely to show toxicity at a measured

C_{max} less than 10μM than those that had cLogP<3 and TPSA>75 Å^2. Following on from this work, Price *et al*. [63] suggest in a later review that compounds with an increasingly basic center were more likely to exhibit off-target pharmacology which can similarly lead to an increased likelihood of observing toxicity at low exposures. They also noted that compounds that were both cationic and amphiphilic, i.e. contained both hydrophilic and lipophilic elements, were more likely to cause phospholipidosis in vivo. Manallack *et al* [64] have also raised the importance of pKa in the drug discovery process as this influences aqueous solubility and absorption as well as other important factors such as plasma protein binding. For some properties, such as lipophilicity, the nature of the *in vivo* finding is difficult to predict; however, there is now a known quantifiable increase in the chances of a significant finding.

They go on to highlight the association between a chemicals ability to absorb light with wavelength > 290 nM and its ability to cause phototoxicity. However while it is true that compounds with known phototoxicity have a UV absorption and a large extinction coefficient but there are similarly many examples exist where chemicals can absorb UV light and do not exhibit toxicity in this excited state.

In summary, changes in key physicochemical properties such as pka, lipophilicity and polar surface area can lead to dramatic effects on the toxicity of a chemical, either through influencing the ADME properties such as clearance of the compound or its ability to interact with a biological system in the form of pharmacological interactions or non-specific protein binding events or even both of these.

Repeat Dose or NOAEL Prediction

Repeat dose toxicology study results for a compound are often summarized as either the no observable effect level (NOEL); the no observable adverse effect level (NOAEL); or the lowest observed adverse effect level (LOAEL). Often these levels are expressed in terms of the administered dose in milligrams per kilogram of bodyweight. However, more recent work looking this type of study data has used the plasma exposure concentrations defined by the maximum concentration in micromolar units (μM) of compound in the circulating plasma, C_{max}. Other plasma concentration values such as the area under the curve (AUC) which measures concentrations over a 24 hour time period measured in μM times hour (μM.h) units ; or the average concentration observed over a 24 hour period (C_{av}) in μM which is simply the AUC value divided by 24 to define the toxic concentration of a chemical. Simply plotting the administered dose against the observed plasma concentration across sets of chemicals show a complex relationship between these two measures with the plasma exposure of compounds administered at identical doses varying by over five log units meaning the that relative rates of absorption, distribution, metabolism and excretion of a chemical plays a significant role in assessing risk.

Subsequent to the publications on the role of physicochemical properties in toxicity, researchers have shown that lower cytotoxic concentrations [65, 66] and and Wang and Greene increased off-target pharmacological activity [47, 48] can also lead to an increased likelihood of observing toxicity at a C_{max} <10μM.

Sutherland *et al*. [67] looked at the lowest observed adverse effect level (LOAEL) as determined by the identification of the lowest dose that causes adverse histological changes or death in rats and assigned the corresponding compound concentration in plasma (i.e., C_{max} value)

associated with this dose. Small values or concentrations denote compounds with unfavorable toxicology outcomes. The authors examined the ability of surrogate properties to predict changes in the LOAEL within a given chemical series. While the quantitative agreement between surrogates and the observed LOAEL is modest, several surrogates provided useful qualitative information: large increases in Volume of Distribution (Vd) or plasma clearance in rats corrected for fraction of compound unbound, (CLu) from low-dose rodent PK studies tend to result in significantly lower LOAEL values; the converse is observed for large increases in in vitro rat primary hepatocyte (RPH) cytolethality or AUC from low oral dose rodent PK studies. Among computed molecular properties, a large increase in molecular weight or heavy atom count tends to decrease LOAEL, whereas increasing hydrogen bond donors tends to increase LOAEL.

Limitations of Computational Models for Toxicology

The use of any *in silico* system or model is limited by both the accuracy of the predictions and the confidence in those predictions but this accuracy and confidence is context dependent. For example, the Ames test is considered to be the most accurate surrogate assay for genotoxic carcinogenicity [68], yet there is little confidence that non-genotoxic carcinogens will display activity in this *in vitro* assay. In the case of *in silico* models, accuracy can be generally considered a property of the system, and confidence, also known as trust or reliability, assigned to individual predictions. The accuracy of *in silico* toxicity predictions is typically measured through internal and external validation of the model using data sets of known experimental activity. Internal validation is used during development to show that statistically-derived models are robust, but provide little information about their ability to predict the activity of compounds outside of the training set [69, 70]. External or prospective validation is the gold standard

method for evaluating model performance, but results have proved to be very data set and therefore context dependent.

Computational models in biology and toxicology predominantly rely on the assumption that similar molecules will have similar biological effects. However, the definition of what is similar may be very different depending on the biological effect being measured. When using in silico methods or read-across approaches to infer toxicological activity there are two main limitations that need careful consideration in the assessment:

1. Measures of chemical similarity and their appropriate application to the effect being predicted;
2. The reported applicability domain of a prediction and hence the reliability of the prediction being made.

These aspects of computational models have been discussed in more detail in reviews by Patlewicz *et al*. [71] and Modi *et al*. [72] but the issues are briefly summarized in the following sections.

Limitations of Chemical Similarity and Read-Across

Chemical similarity is often used in read-across and QSAR models to identify structures with known activities that could be used to infer the activity of a molecule with unknown activity. However, this presents the dilemma of how to define what is similar and what is not. This problem of defining chemical similarity has been debated for decades and no one method has been agreed upon as being optimal as this often use case dependent. For example, in genetic toxicology when a chemical bears the same structural alert as the experimentally Ames negative comparison compound (and no other known structural alert) in the same position and

environment and possesses a similar molecular weight, then the compound is often considered to be negative in the Ames test.

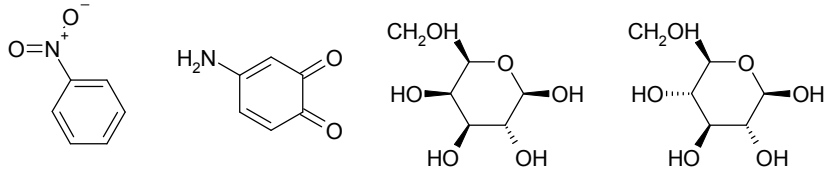
In a read-across assessment, a chemically-defined category of known adverse activity can be represented by a series of compounds with common structural features, showing similar trends in their physicochemical properties. The presence of a common biological or chemical behavior can be generally associated with a common underlying mechanism of action (e.g. alkylating compounds).

This categorical approach provides the basis of identifying trends in properties across the category of compounds resulting in the possibility to extend the use of measured data to similar untested chemicals. These estimates of biological activity may be deemed adequate for regulatory purposes (e.g. classification and labeling and/or impurity hazard assessment for classification with respect to toxicity potential) without further testing. A description of chemical category function has been given by Enoch *et al*., [73]. However the standardization of this approach for defining structural similarity to a chemically-defined class of known biological actives is much more difficult where the mechanisms of action are both diverse and complex.

In QSAR approaches, the definition of structural similarity is crucial to the final result of an in silico prediction [74]. The typical starting point for any computational approach for assessing chemical similarity is to obtain a quantitative description of the molecular structure or fingerprint. Comparisons between structures are then performed using one of a variety of indices that have been developed for example, Euclidean distance measures or maximum common substructures.

However, similarity is a multi-dimensional concept and the similarity between two compounds can be difficult to determine and even more challenging to create a set of guidelines for. For instance, compounds **(1)** and **(2)** in Figure 1taken from Naven *et al.* [74] have the same molecular formulae ($C_6H_5NO_2$) but will unlikely to be considered similar as they have different atom connectivity, different electron delocalization properties or aromatic behavior, physicochemical properties and most importantly, probably dissimilar biological properties. Conversely, glucose **(3)** and galactose **(4)** also have the same molecular formulae ($C_6H_{12}O_6$) and visually appear almost structurally identical but from a pharmacological perspective, these compounds have very different properties. Many methods to measure the structural similarity between two compounds have been developed but the more relevant question to ask would be is structural similarity an important factor for the toxicological endpoint being studied.

Figure 1: Selected Examples of Similar Compounds



1 nitrobenzene

2 4-amino-2-benzo-quinone

3 glucose

4 galactose

For toxicological endpoints like mutagenicity or the uncoupling of oxidative phosphorylation that are dependent on the presence or absence of structural alerts, the less-applicable the concept of similarity becomes. This is because minor modifications to the structural alert can significantly influence toxicological activity, yet major modifications to the periphery of the chemical structure may have little impact on activity *so long as* the structural

alert remains intact. When assessing the relevance of a prediction, it is not enough to ask how similar the query compound is to other inactive compounds, but to identify the features of structurally-alerting, active compounds that would attenuate the activity and to assess if these features can be adequately extrapolated to the compound being studied.

Limitations of Defining Applicability Domains

OECD guidelines (http://www.oecd.org/chemicalsafety/risk-assessment/validationofqsarmodels.htm) currently recommend that QSAR models should define the domain within which the predictions of a model can be deemed reliable. Many methods exist for defining the applicability domain (AD) of a QSAR model and have been extensively reviewed [75-77]. The AD of a model can be broadly described using two different yet non-exclusive descriptions:

(i) the region of chemical or response space relating to the model training set; and

(ii) the region of chemical or response space where a model makes an acceptable prediction error.

In the first definition (i), the underlying assumption is that those predictions that are based on interpolation from data in the training set are generally more reliable than those based on extrapolation. The second definition (ii) is based on the assumption that by assessing where compounds are predicted well we can gain valuable information, whereas inevitably a subset of the training set will be incorrectly classified and so similarity to these compounds will have no guarantee that predictions are reliable. In addition, this definition does not automatically assume

that predictions for compounds that are considered dissimilar to the training set are unreliable [75].

Defining the applicability domain of any model is difficult and presents challenges to the end user as to whether a prediction is reliable or not. In addition, although the scope of structural alerts can be used to define their AD, this provides little information to a user when alerts are not matched to the compound in question. Expert systems that rely on structural alerts do not have a model training set per se, as the alerts are often based on disparate data sources such as toxicity data, information pertaining to the biological mechanism and knowledge of chemistry and reactivity, which are synthesized into the development of an alert *in cerebro*. Furthermore, not all data is publically-available, thus current approaches cannot reflect this expert knowledge and often require a complete model training set.

Most of the methods for defining applicability domains have been trained to reduce the error in continuous output QSARs where the assay data provides homogeneous responses, for example LogP values or an experimentally derived IC_{50} for protein inhibition. It should be noted that there is a distinct gap on the *applicability* of ADs to categorical models that are based on assays which generate a more diverse range of outputs such as carcinogenicity or reproductive effects. There are a few exceptions but generally it was shown that there was only value in using an AD to qualify confidence in a positive response, rather than a prediction for absence of activity. [76, 78, 79]

Looking Ahead; Consortia Efforts on Database Development and Toxicity Prediction

It can be argued that the challenge of building predictive toxicology models that predictive across a broad chemical space is too large for single organizations to effectively address alone. Indeed, there are several active consortia efforts which seek to engage industry, academia, private companies and regulators in combined efforts to share data and build tools. Some of these initiatives come with the support of significant government funding. This is the case for the European eTOX project which aims to leverage historic toxicology study data held by participating pharmaceutical companies for new model development. At the time of publication, eTOX, which is sponsored by the European Innovative Medicines Initiative (IMI), involves active participation of 11 academic groups, *6 in silico* model technology companies and 13 pharmaceutical companies (see http://www.etoxproject.eu). The overall framework of the initiative centers on a software platform (eTOXsys) which contains both the underlying data from the participating organizations and access to computational predictive models under development. To date the initiative has reported success in assembling >5000 toxicology study reports from members (described in a common ontology of toxicology terms) and 74 individual *in silico* models for hazard identification or prediction of compound disposition in the body [80].

In the USA perhaps the most prominent consortia effort with somewhat similar goals is the "Toxicity Testing in the 21st Century" initiative, commonly referred to as "Tox21". This multidisciplinary project spans several government research partners (including the EPA, National Institutes of Health, National Center for Advancing Translational Sciences and the Food and Drug Administration) and is exploring alternative approaches to *in vivo* toxicity testing with a particular emphasis on understanding the critical molecular pathways in cells and tissues

which could be diagnostic of toxicology mechanisms. In the initial phases of this program a well characterized screening library of >10,000 discrete molecules was established as the substrate to be systematically profiled through approximately 50 high throughput screening assays selected for their potential relevance to toxicity mechanisms [81]. As the project has advanced additional high throughput assays, including secondary assays to refine mechanistic understanding have been added and the partnerships extended. The effort has linked with a separate project at EPA called "ToxCast" (http://www.epa.gov/ncct/toxcast/) with broadly similar goals around combining in vitro data and *in silico* modeling to predict potential toxicity. In the case of the ToxCast initiative, the source of the library compounds has been across a diverse range of industrial partners (including pharmaceuticals, consumer products, food additives etc) and its testing paradigm includes broad assays, or panels of assays developed externally. As a consequence the ToxCast compound collection has been used as test set to examine areas as diverse as endocrine disruption [82], genotoxicity [83] and to classify toxic and therapeutic mechanisms [84].

While initiatives such as eTOX, Tox21 and ToxCast attempt to develop a very broad framework for toxicity prediction there have also been more targeted efforts which focus on individual toxicities which have been problematic in existing testing schemes. The area of drug induced liver injury (DILI), for example, is a longstanding problem in pharmaceutical discovery and development (for a recent review see Leise *et al* [85]). In Europe IMI has funded an integrated assessment platform for DILI named Mechanism Based Integrated System for the Prediction of DILI or MIP-DILI (http://www.imi.europa.eu/content/mip-dili). This 5 year project was launched in 2012 and involves the participation of almost 30 organizations from industry, technology companies and academia. The efforts of MIP-DILI span the assembly of an

annotated compound training set, the testing of *in vitro* systems for utility, novel in vivo approaches (including reporter or humanized transgenic animals), and a bioinformatics hub to assemble the data and facilitate the building and testing of predictive models. In the USA, a jointly led initiative by The Hamner Institutes for Health Sciences and the University of North Carolina is working to develop software predictive algorithms to predict DILI and the initiative is partnering with over a dozen global pharmaceutical companies to partner on the production of an integrated software platform DILIsym (http://dilisym.com/) which is envisaged as a platform for both industry and regulators to evaluate potential DILI concerns of new chemical entities [86]. The model looks at essential cellular processes and hepatotoxicity-specific cellular mechanisms (such as reactive metabolites, mitochondrial dysfunction and transporter inhibition) and the models are developed using supporting data from well-characterized DILI-producing compounds from member companies and public domain sources.

Clearly computational predictive modeling in toxicology assessment continues to advance as a discipline with the sharing of data and experience through efforts such as those described above. Inevitably the approach has significant limitations when being considered as a definitive toxicity assessment such as that which would be required in formal risk assessment strategies however. Nevertheless it appears clear that in the areas of hazard identification, hypothesis generation and in prioritizing compounds for more extensive toxicity evaluation that computational approaches have utility today with the promise of even greater impact through consortia effort and shared experience in the future.

Conclusions

The rise of computational toxicology as a scientific discipline has seen significant investment over recent years and has become a mainstream activity in many commercial and regulatory

organizations. However, the application of these tools to both hazard and risk assessment applications needs to be carefully thought through. The computational predictions are only as good as the data used to train the model and the inherent noise in these data sets is often overlooked when training or assessing performance. Similarly, the current lack of knowledge about mechanisms of toxicity along with the fact that multiple molecular initiating events can lead to the same observed phenotype make it difficult to select the best measures of chemical similarity to use in any given model. Clearly, the need for higher throughput and more cost effect approaches for safety assessments make computational approaches a useful tool in toxicology but if used inappropriately or without consideration for the limitations of the approaches then this may lead to poor or regrettable decisions being made.

References:

1. Paul, S.M., et al., *How to improve R&D productivity: the pharmaceutical industry's grand challenge.* Nat Rev Drug Discov, 2010. **9**(3): p. 203-14.
2. Judson, R.S., et al., *In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project.* Environ Health Perspect, 2010. **118**(4): p. 485-92.
3. Dobo, K.L., et al., *In silico methods combined with expert knowledge rule out mutagenic potential of pharmaceutical impurities: an industry survey.* Regul Toxicol Pharmacol, 2012. **62**(3): p. 449-55.
4. Uehara, T., et al., *The Japanese toxicogenomics project: Application of toxicogenomics.* Molecular Nutrition & Food Research, 2010. **54**(2): p. 218-227.
5. Gaulton, A., et al., *ChEMBL: a large-scale bioactivity database for drug discovery.* Nucleic Acids Research, 2012. **40**(D1): p. D1100-D1107.

6. Elangbam, C.S., *Drug-induced Valvulopathy: An Update.* Toxicologic Pathology, 2010. **38**(6): p. 837-848.

7. Judson, P.N., et al., *Towards the creation of an international toxicology information centre.* Toxicology, 2005. **213**(1-2): p. 117-28.

8. Ashby, J. and R.W. Tennant, *Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the U.S. NTP.* Mutat Res, 1991. **257**(3): p. 229-306.

9. Haworth, S., et al., *Salmonella mutagenicity test results for 250 chemicals.* Environ Mutagen, 1983. **5 Suppl 1**: p. 1-142.

10. Mortelmans, K., et al., *Salmonella mutagenicity tests: II. Results from the testing of 270 chemicals.* Environ Mutagen, 1986. **8 Suppl 7**: p. 1-119.

11. Zeiger, E., et al., *Salmonella mutagenicity tests: IV. Results from the testing of 300 chemicals.* Environ Mol Mutagen, 1988. **11 Suppl 12**: p. 1-157.

12. Zeiger, E., et al., *Salmonella mutagenicity tests: III. Results from the testing of 255 chemicals.* Environ Mutagen, 1987. **9 Suppl 9**: p. 1-109.

13. Lynch, A.M., et al., *New and emerging technologies for genetic toxicity testing.* Environ Mol Mutagen, 2011. **52**(3): p. 205-23.

14. Naven, R.T., S. Louise-May, and N. Greene, *The computational prediction of genotoxicity.* Expert Opin Drug Metab Toxicol, 2010. **6**(7): p. 797-807.

15. Marchant, C.A., K.A. Briggs, and A. Long, *In silico tools for sharing data and knowledge on toxicity and metabolism: derek for windows, meteor, and vitic.* Toxicol Mech Methods, 2008. **18**(2-3): p. 177-87.

16. Saiakhov, R.D. and G. Klopman, *Benchmark performance of MultiCASE Inc. software in Ames mutagenicity set.* J Chem Inf Model, 2010. **50**(9): p. 1521.

17. Valerio, L.G., Jr. and K.P. Cross, *Characterization and validation of an in silico toxicology model to predict the mutagenic potential of drug impurities.* Toxicol Appl Pharmacol, 2012. **260**(3): p. 209-21.

18. Benigni, R., et al., *Alternatives to the carcinogenicity bioassay: in silico methods, and the in vitro and in vivo mutagenicity assays.* Expert Opin Drug Metab Toxicol, 2010. **6**(7): p. 809-19.

19. Hillebrecht, A., et al., *Comparative evaluation of in silico systems for ames test mutagenicity prediction: scope and limitations.* Chem Res Toxicol, 2011. **24**(6): p. 843-54.

20. Sutter, A., et al., *Use of in silico systems and expert knowledge for structure-based assessment of potentially mutagenic impurities.* Regul Toxicol Pharmacol, 2013. **67**(1): p. 39-52.

21. Kamber, M., et al., *Comparison of the Ames II and traditional Ames test responses with respect to mutagenicity, strain specificities, need for metabolism and correlation with rodent carcinogenicity.* Mutagenesis, 2009. **24**(4): p. 359-66.

22. Snyder, R.D., *Assessment of atypical DNA intercalating agents in biological and in silico systems.* Mutat Res, 2007. **623**(1-2): p. 72-82.

23. Benigni, R. and A. Giuliani, *Putting the Predictive Toxicology Challenge into perspective: reflections on the results.* Bioinformatics, 2003. **19**(10): p. 1194-200.

24. Fjodorova, N., et al., *Quantitative and qualitative models for carcinogenicity prediction for non-congeneric chemicals using CP ANN method for regulatory uses.* Mol Divers, 2010. **14**(3): p. 581-94.

25. Kar, S., O. Deeb, and K. Roy, *Development of classification and regression based QSAR models to predict rodent carcinogenic potency using oral slope factor.* Ecotoxicol Environ Saf, 2012. **82**: p. 85-95.

26. Wu, S., et al., *Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants.* Chem Res Toxicol, 2013. **26**(12): p. 1840-61.

27. Matthews, E.J., et al., *A comprehensive model for reproductive and developmental toxicity hazard identification: II. Construction of QSAR models to predict activities of untested chemicals.* Regul Toxicol Pharmacol, 2007. **47**(2): p. 136-55.

28. Cronin, M.T.D. and A.P. Worth, *(Q)SARs for Predicting Effects Relating to Reproductive Toxicity.* QSAR & Combinatorial Science, 2008. **27**(1): p. 91-100.

29. Vocanson, M., J.-F. Nicolas, and D. Basketter, *In vitro approaches to the identification and characterization of skin sensitizers.* Expert Review of Dermatology, 2013. **8**(4): p. 395-405.

30. Mekenyan, O., et al., *A mechanistic approach to modeling respiratory sensitization.* Chem Res Toxicol, 2014. **27**(2): p. 219-39.

31. Agius, R.M., et al., *Occupational asthma and the chemical properties of low molecular weight organic substances.* Occup Med (Lond), 1994. **44**(1): p. 34-6.

32. Agius, R.M., et al., *Structure activity hypotheses in occupational asthma caused by low molecular weight substances.* Ann Occup Hyg, 1991. **35**(2): p. 129-37.

33. Enoch, S.J., et al., *Development of mechanism-based structural alerts for respiratory sensitization hazard identification.* Chem Res Toxicol, 2012. **25**(11): p. 2490-8.

34. Graham, C., H.S. Rosenkranz, and M.H. Karol, *Structure-activity model of chemicals that cause human respiratory sensitization.* Regul Toxicol Pharmacol, 1997. **26**(3): p. 296-306.

35. Jarvis, J., et al., *Relationship between chemical structure and the occupational asthma hazard of low molecular weight organic compounds.* Occup Environ Med, 2005. **62**(4): p. 243-50.

36. Warne, M.A., et al., *A QSAR investigation of dermal and respiratory chemical sensitizers based on computational chemistry properties.* SAR QSAR Environ Res, 2009. **20**(5-6): p. 429-51.

37. Holt, M.P. and C. Ju, *Mechanisms of drug-induced liver injury.* AAPS J, 2006. **8**(1): p. E48-54.

38. Kaplowitz, N., *Idiosyncratic drug hepatotoxicity.* Nat Rev Drug Discov, 2005. **4**(6): p. 489-99.

39. Williams, D.P., *Toxicophores: investigations in drug safety.* Toxicology, 2006. **226**(1): p. 1-11.

40. Zimmermann, H.J., *The Adverse Effects of Drugs and Other Chemicals on the Liver*, ed. H.J. Zimmerman, Ed. 1999, Philadelphia: Lippincott Williams & Wilkins.

41. Kola, I. and J. Landis, *Can the pharmaceutical industry reduce attrition rates?* Nat Rev Drug Discov, 2004. **3**(8): p. 711-5.

42. Greene, N., et al., *Developing structure-activity relationships for the prediction of hepatotoxicity.* Chem Res Toxicol, 2010. **23**(7): p. 1215-22.

43. Hewitt, M., et al., *Hepatotoxicity: a scheme for generating chemical categories for read-across, structural alerts and insights into mechanism(s) of action.* Crit Rev Toxicol, 2013. **43**(7): p. 537-58.

44. Chen, M., et al., *Quantitative structure-activity relationship models for predicting drug-induced liver injury based on FDA-approved drug labeling annotation and using a large collection of drugs.* Toxicol Sci, 2013. **136**(1): p. 242-9.

45. Naven, R.T., et al., *The development of structure-activity relationships for mitochondrial dysfunction: uncoupling of oxidative phosphorylation.* Toxicol Sci, 2013. **131**(1): p. 271-8.

46. Elangbam, C.S., et al., *5-hydroxytryptamine (5HT)-induced valvulopathy: compositional valvular alterations are associated with 5HT2B receptor and 5HT transporter transcript changes in Sprague-Dawley rats.* Exp Toxicol Pathol, 2008. **60**(4-5): p. 253-62.

47. Hughes, J.D., et al., *Physiochemical drug properties associated with in vivo toxicological outcomes.* Bioorg Med Chem Lett, 2008. **18**(17): p. 4872-5.

48. Wang, X. and N. Greene, *Comparing Measures of Promiscuity and Exploring Their Relationship to Toxicity.* Molecular Informatics, 2012. **31**(2): p. 145-159.

49. Koslov-Davino, E., X. Wang, and T. Schroeter, *Target promiscuity and physicochemical properties contribute to pharmacologically induced ER-stress.* Toxicol In Vitro, 2013. **27**(1): p. 204-10.

50. Stepan, A.F., et al., *Structural alert/reactive metabolite concept as applied in medicinal chemistry to mitigate the risk of idiosyncratic drug toxicity: a perspective based on the critical examination of trends in the top 200 drugs marketed in the United States.* Chem Res Toxicol, 2011. **24**(9): p. 1345-410.

51. Goracci, L., et al., *Modeling phospholipidosis induction: reliability and warnings.* J Chem Inf Model, 2013. **53**(6): p. 1436-46.

52. Vuorinen, A., A. Odermatt, and D. Schuster, *In silico methods in the discovery of endocrine disrupting chemicals.* J Steroid Biochem Mol Biol, 2013. **137**: p. 18-26.

53. Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J., *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings* Advanced Drug Delivery Reviews, 1997. **23**(1-3): p. 3-25.

54. Lipinski, C.A., *Drug-like properties and the causes of poor solubility and poor permeability.* J Pharmacol Toxicol Methods, 2000. **44**(1): p. 235-49.

55. Potts, R.O. and R.H. Guy, *Predicting skin permeability.* Pharm Res, 1992. **9**(5): p. 663-9.

56. Lombardo, F., et al., *Prediction of volume of distribution values in humans for neutral and basic drugs using physicochemical measurements and plasma protein binding data.* J Med Chem, 2002. **45**(13): p. 2867-76.

57. Gombar, V.K. and S.D. Hall, *Quantitative structure-activity relationship models of clinical pharmacokinetics: clearance and volume of distribution.* J Chem Inf Model, 2013. **53**(4): p. 948-57.

58. Hsiao, Y.W., U. Fagerholm, and U. Norinder, *In silico categorization of in vivo intrinsic clearance using machine learning.* Mol Pharm, 2013. **10**(4): p. 1318-21.

59. Smith, D.A., L. Di, and E.H. Kerns, *The effect of plasma protein binding on in vivo efficacy: misconceptions in drug discovery.* Nat Rev Drug Discov, 2010. **9**(12): p. 929-39.

60. Hollosy, F., et al., *Estimation of volume of distribution in humans from high throughput HPLC-based measurements of human serum albumin binding and immobilized artificial membrane partitioning.* J Med Chem, 2006. **49**(24): p. 6958-71.

61. Yang, Y., et al., *Beyond Size, Ionization State, and Lipophilicity: Influence of Molecular Topology on Absorption, Distribution, Metabolism, Excretion, and Toxicity for Druglike Compounds.* Journal of Medicinal Chemistry, 2012. **55**(8): p. 3667-3677.

62. Zhivkova, Z. and I. Doytchinova, *Quantitative structure--clearance relationships of acidic drugs.* Mol Pharm, 2013. **10**(10): p. 3758-68.

63. Price, D.A., et al., *Physicochemical drug properties associated with in vivo toxicological outcomes: a review.* Expert Opin Drug Metab Toxicol, 2009. **5**(8): p. 921-31.

64. Manallack, D.T., et al., *The significance of acid/base properties in drug discovery.* Chem Soc Rev, 2013. **42**(2): p. 485-96.

65. Benbow, J.W., et al., *Predicting safety toleration of pharmaceutical chemical leads: cytotoxicity correlations to exploratory toxicity studies.* Toxicol Lett, 2010. **197**(3): p. 175-82.

66. Greene, N., et al., *Using an in vitro cytotoxicity assay to aid in compound selection for in vivo safety studies.* Bioorg Med Chem Lett, 2010. **20**(17): p. 5308-12.

67. Sutherland, J.J., et al., *Relating molecular properties and in vitro assay results to in vivo drug disposition and toxicity outcomes.* J Med Chem, 2012. **55**(14): p. 6455-66.

68. Kirkland, D., et al., *Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I. Sensitivity, specificity and relative predictivity.* Mutat Res, 2005. **584**(1-2): p. 1-256.

69. Gramatica, P., *Principles of QSAR models validation: internal and external.* QSAR & Combinatorial Science, 2007. **26**(5): p. 694-701.

70. Tropsha, A., P. Gramatica, and V.K. Gombar, *The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models.* QSAR & Combinatorial Science, 2003. **22**(1): p. 69-77.

71. Patlewicz, G., et al., *Use of category approaches, read-across and (Q)SAR: general considerations.* Regul Toxicol Pharmacol, 2013. **67**(1): p. 1-12.

72. Modi, S., et al., *The value of in silico chemistry in the safety assessment of chemicals in the consumer goods and pharmaceutical industries.* Drug Discov Today, 2012. **17**(3-4): p. 135-42.

73. Enoch, S.J., M.T. Cronin, and C.M. Ellison, *The use of a chemistry-based profiler for covalent DNA binding in the development of chemical categories for read-across for genotoxicity.* Altern Lab Anim, 2011. **39**(2): p. 131-45.

74. Naven, R.T., N. Greene, and R.V. Williams, *Latest advances in computational genotoxicity prediction.* Expert Opin Drug Metab Toxicol, 2012. **8**(12): p. 1579-87.

75. Dragos, H., M. Gilles, and V. Alexandre, *Predicting the predictability: a unified approach to the applicability domain problem of QSAR models.* J Chem Inf Model, 2009. **49**(7): p. 1762-76.

76. Ellison, C.M., et al., *Assessment of methods to define the applicability domain of structural alert models.* J Chem Inf Model, 2011. **51**(5): p. 975-85.

77. Hewitt, M. and C.M. Ellison, *Developing the applicability domain of in silico models: relevance, importance and methodology*, in *In Silico Toxicology: Principles and applications*, M.T.D. Cronin and J.C. Madden, Editors. 2010, Royal Society of Chemistry: Cambridge, UK. p. 301-333.

78. Jaworska, J., N. Nikolova-Jeliazkova, and T. Aldenberg, *QSAR applicabilty domain estimation by projection of the training set descriptor space: a review.* Altern Lab Anim, 2005. **33**(5): p. 445-59.

79. Kuhne, R., R.U. Ebert, and G. Schuurmann, *Chemical domain of QSAR models from atom-centered fragments.* J Chem Inf Model, 2009. **49**(12): p. 2660-9.

80. Cases, M., et al., *The eTOX data-sharing project to advance in silico drug-induced toxicity prediction.* Int J Mol Sci, 2014. **15**(11): p. 21136-54.

81. Betts, K.S., *Tox21 to date: steps toward modernizing human hazard characterization.* Environ Health Perspect, 2013. **121**(7): p. A228.

82. Filer, D., et al., *Test driving ToxCast: endocrine profiling for 1858 chemicals included in phase II.* Curr Opin Pharmacol, 2014. **19**: p. 145-52.

83. Kligerman, A.D., et al., *An evaluation of 25 selected ToxCast chemicals in medium-throughput assays to detect genotoxicity.* Environ Mol Mutagen, 2014.

84. Kleinstreuer, N.C., et al., *Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms.* Nat Biotechnol, 2014. **32**(6): p. 583-91.

85. Leise, M.D., J.J. Poterucha, and J.A. Talwalkar, *Drug-induced liver injury.* Mayo Clin Proc, 2014. **89**(1): p. 95-106.

86. Bhattacharya, S., et al., *Modeling drug- and chemical-induced hepatotoxicity with systems biology approaches.* Front Physiol, 2012. **3**: p. 462.