Computational toxicology—a tool for early safety evaluation

Cédric Merlot

Genyotex SA, Chemin des Aulx 14, 1228 Plan les Ouates, Geneva, Switzerland

Although inappropriate pharmacokinetic properties were a major cause of attrition in the 1990s, safety issues are recognized as today's single largest cause of drug candidate failure. It is expected that the right balance of *in vivo*, *in vitro* and computational toxicology predictions applied as early as possible in the discovery process will help to reduce the number of safety issues. This review focuses on recent developments in computational toxicology. Direct modeling of toxic endpoints has been deceiving and hampered the wide acceptance of computer predictions. The current trend is to make simpler predictions, closer to the mechanism of action, and to follow them up with *in vitro* or *in vivo* assays as appropriate.

Introduction

Although *in vivo* toxicology remains the gold standard for identifying the side effects induced by a drug, it is now considered that this approach alone cannot help to reduce the large attrition rate in late clinical development stages. Moreover, there is pressure to reduce the number of *in vivo* experiments and, therefore, an extensive development of new *in vivo* tests is not an option. Several initiatives (e.g. parts of the FDA's Critical Path Initiative or the EU's Seventh Research Framework Programme) aim to improve how the toxicity of new molecules can be evaluated all along the discovery and development pipelines. A new paradigm for toxicity testing – combining the strengths of the *in vivo*, *in vitro* and *in silico* worlds – is sought.

This review focuses on recent developments that try to increase the reliability of toxicity prediction algorithms that use the structure only. We also report on the application of these new approaches in lead optimization projects.

Computational methods to predict side effects of small molecules

Many computational approaches are available to predict the toxicity induced by a small molecule from its chemical drawing [1]. Traditionally, they have been tuned to predict global toxicity endpoints, such as carcinogenicity, or mutagenicity [2–5]. Although they are useful in a few cases, the broad application of such predictions has been hampered by their lack of accuracy [6]. It is generally deemed that this lack of accuracy is due to the complexity of the predicted endpoints, rather than to the poor performance of data analysis methods. The focus, therefore, is on modeling more simple endpoints, such as off-target activity, to increase accuracy and to combine the results with experiments from other fields (e.g. -omics) to try to make a link with potential modes of action.

Expert systems versus statistical modeling

Of the several approaches to predicting the effects of small molecules from the structure, most algorithms can be put into two classes: expert systems and statistical modeling.

On the one hand, expert systems, such as Oncologic [7] or Derek for Windows (http://www.lhasalimited.org), are a repository of expert knowledge. The computer is there to store, then use on demand a piece of knowledge that has been formalized and input by human experts. The power of the system is linked to the amount of expert time invested in feeding it and to the availability of reliable and high-quality datasets. Its expansion, therefore, is limited by the time it takes for humans to collect and digest lots of information. Although the information contained in these systems is considered reliable enough, it suffers from a lack of sensitivity [5]. The direct consequence is that many side effects

E-mail address: cedric.merlot@genkyotex.com.

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2009.09.010

are likely to be missed. Muster *et al.* [8] reported that medicinal chemists have to use this tool cautiously because the outcome of the program requires a deep understanding of the system to interpret results. Indeed, the use of toxicity prediction software without a critical review of such predictions can easily lead to misleading interpretations of the data, resulting in a negative impact on discovery projects [9].

On the other hand, statistical modeling software – such as Topkat (http://accelrys.com/products/discovery-studio/toxicology), PASS [10], TPS–SVM [11] and Multicase (http://www.multicase.com/) – aims to analyze existing data and automatically build models, with a reduced need for human intervention. Just as for expert systems, the first step consists of assembling a relevant training set of compounds with experimental biological data. The system will then perform a statistical analysis that shall be reviewed by a scientist. These systems require a lot of attention for the selection of modeling techniques and structural descriptors. They have several advantages over expert systems, however: a model can be optimized on internal data more quickly and objectively than through an expert analysis. It can also be combined with a quantitative structure–activity relationship (QSAR) when only a single chemical series is involved. Although statistical analysis highlights trends in diverse structures (i.e. all molecules containing a given fragment are flagged as potentially toxic), the QSAR handles the more subtle structural changes that, in a set of similar compounds, flag some compounds as toxic and others as less harmful. The combination of statistical analysis and QSAR, therefore, facilitates lead optimization and the removal of toxicophores [8].

The most common softwares were cited herein to illustrate the difference between expert systems and statistical modeling. A more exhaustive list of computer systems for toxicity prediction can be found in two reviews [8,12]. The details of the algorithms have been reviewed [2,3] and their relative performance evaluated [5,13].

Applicability domain for early safety evaluation

As stated above, evaluating the reliability of a prediction for a given compound or series of similar compounds is necessary to make informed decisions [14]. Since the inception of REACH (registration, evaluation and authorization of chemical substances), much effort has been devoted to defining the domain of applicability of computational models, a measure of the degree of reliability of a prediction. As part of the new procedures, the EU proposes a framework and list of requirements to develop and validate QSAR methods for the evaluation of physicochemical properties or toxic endpoints [15].

One reason for the failure of computational approaches is the difference in the chemical space covered by compounds used to train a model and those to which models are applied [16]. It seems obvious that computational models will work better when test compounds are close to compounds used to train the model, provided that an appropriate measure of distance, or measure of similarity of compounds, is defined. The major issue here is to define the measure of distance [17] and between which objects said distance must be measured. In some applications, structures are described as a set of features (a fingerprint): a large vector is filled with 1 if the corresponding feature is in the structure and with 0 otherwise. The distance is usually defined as a function of the number of common features normalized with the number of (shared) features, using, for example, the Tanimoto distance. These are referenced as similarity-based applicability domain. Another example is used in Topkat (http://accelrys.com/products/discovery-studio/toxicology), which assimilates the applicability domain to a box containing the training set. Structures are described by a vector of continuous parameters, and the distance is calculated as the distance between the position in the hyperspace of the structure and the space filled by the molecules in the training set [18,19]. More advanced definitions of an applicability model exist and involve the consensus prediction variance or the dissimilarity to outlier-free training sets [20].

The concept of applicability domain is useful and efficient for the prediction of physicochemical properties (e.g. lipophilicity and aqueous solubility) [16,19]. One reason for this is that these properties are often calculated by adding the values of local properties such as atom-based or small fragments, resulting in a small number of descriptors (typically, a few hundred) to describe all potential structures.

Defining applicability domains for a method using fragment-based descriptors is trickier. Fragments of different sizes are used in Oncologic, Derek, Multicase (from 2 to 10 atoms) and PASS (on average 6–7 atoms; a fragment is defined by all the neighbors of the neighbors of a central atom). The numbers of fragments required to describe all potential structures grows up to billions, and most test structures, therefore, will be considered as 'far' from the training set because the fragments it contains are not in the training set. If the test structure is not new, fragments will be found in the training set and the test structure might be considered close to the training set. Being at a small distance from the training set, however, does not mean that the prediction is correct. Off-target activity prediction is impacted by small changes (steric hindrance, conformational change, and so on) in the structure, which can lead to different toxicological responses [21,22].

Defining the domain of applicability, therefore, seems restricted to endpoints that can be predicted from a few set of descriptors (e.g. physicochemical properties) for a limited set of chemical series.

Dissecting complex endpoints for increased coverage

It seems that the complexity (i.e. the number of underlying mechanisms of action) of a toxicological endpoint is a factor of prime importance in the quality of predictions [6,23]. The complexity of a biological system to model increases with the number of histological observations (e.g. cirrhosis, steatosis and fibrosis with regards to liver toxicity); the number of proteins whose modulation can potentially induce these observations; and for each protein, the number of binding sites/modes available. In other words, the complexity is intrinsically related to the total number of mechanisms of action. Although there can be some exceptions, the more complex an endpoint, the more difficult it is to predict.

Estrogen receptor modulation is a simple endpoint deemed to be involved, for example, in carcinogenicity [24,25]. It can be predicted successfully [24,25]. The Ames test (salmonella mutagenicity assay) is a somewhat more mechanistically complex assay that measures genetic toxicity—one of the mechanisms by which compounds can induce cancer. The Ames test is, nevertheless, seen

as a simple assay because it involves a low number of underlying mechanisms, and many models with respectable accuracy exist (see below). By contrast, carcinogenicity, liver toxicity and developmental toxicity are complex endpoints that can be caused by a wide variety of events, and the prediction of these endpoints is usually far from acceptable [6,26].

Eventually, the more complex an endpoint, the more data will be needed to build an accurate model applicable for a diversity of structures. The Ames test can help us to estimate the number of data points needed. Because it is a standardized assay, it is expected that the data will be reproducible in any laboratory. The inter-laboratory reproducibility is estimated to be 85% [27]. This provides us with a large, publicly available set of comparable data. It is possible to assemble a dataset containing approximately 4000 diverse structures (drugs and chemicals) annotated for Ames [28], half of which are positive. In our own experience, this is close to the ideal set for building a predictive model. It is then easy to build a model capable of predicting results of the predominant mechanism (i.e. chemical alkylation, although the Ames test was not designed to detect only alkylating agents), and the best predictive models reach 85% accuracy. This is good in that this is also the estimated reproducibility of experimental data. This example shows that a dataset of 4000 annotated compounds can provide models for the major mechanism(s) of action; however, it will fail to uncover minor mechanisms.

Provided one can extrapolate these numbers, predicting an endpoint with many underlying mechanisms (such as carcinogenicity) requires datasets of hundreds of thousands of annotations for diverse structures. In comparison with what is required, current datasets of a few hundred structures (http://potency.berkeley.edu/chemicalsummary.html) seem too small, and the coverage (i.e. the sensitivity) of carcinogenicity models based on these datasets is questionable [6].

Dissecting complex models into several mechanistically understood events has, therefore, been a focus for several groups. Simon-Hettich *et al.* [6] demonstrated that this approach could lead to more predictive methods. Several groups have also applied this approach, which will be detailed below. They all attempt to predict complex endpoints using surrogate, less complex endpoints, which are simpler to evaluate or to predict.

Predicted safety pharmacological profiling

The prediction of biological activity using structure–activity relationships has been a major focus of cheminformatics. The literature abounds with techniques and methods to predict the biological activity from the chemical structure alone [29] or from biological profiles either complete [30] or focused on a given class of target (e.g. kinases [31,11,32] or nuclear receptors [33]). Any of the methods that have been developed over the years (fragment-based analysis, QSAR, 3D pharmacophores, docking, and so on) are applicable virtually [1], although one must take care because they were developed to increase affinity on a selected target and not to find all potential liabilities of a structure [34]. However, the preference seems to be for fragment-based analysis (e.g. Multicase and PASS [10,35,36]) and QSAR [37,38] (e.g. Topkat), probably because they are fast to implement and fast to run but still provide results comparable to those from more time-consuming 3D approaches [39].

The most time-consuming step is the collection of data. All the computational models employed in the field of toxicity prediction have in common that they are based on the structure of existing, annotated ligands. Typically, scientists gather data from different sources: literature, patents and, when available, in-house data. It is a common practice in the pharmaceutical industry to prepare a few datasets around a target of interest. Computational toxicology platforms, however, have to support many more targets, from 70 at Novartis [40] to 350 in our own experience [41]. Whereas this is still a fraction of what would be required to predict complete activity profiles, it represents a huge amount of work that has, for a long time, prevented the development of systems.

Thanks to several initiatives aiming to collect information from legacy, unstructured systems (e.g. NTP reports and literature) and bring them into structured, open-access databases [42–47], more and more data are becoming available for the construction of new models [1]. Today, many databases are freely available [1,48], providing easy access to public information and fostering the development of new analysis methods and models. They enable all research groups to better understand the relations between pharmacological profiles and toxicity [49,50] and to develop innovative analysis tools with a broad coverage.

From predicted safety pharmacological profiling to toxicity prediction

Compound profiles are nice to have as tools; they answer questions such as 'Is my compound devoid of off-target activity?' Drugs, however, are seldom completely clean. Although on several occasions, their lack of selectivity has improved clinical efficacy [51], most of the time it gives rise to unwanted side effects. The step after predicting a profile comprising off-target modulation, genotoxicity and other relevant tests, therefore, is to understand the pharmacological consequences implied by the compound's profile, in the context of a therapeutic indication, to progress promising compounds and deprioritize molecules with potential side effects that might require additional tests in the clinical development phase.

Matthews *et al.* have applied pharmacological profiling to the prediction of carcinogenicity, reproductive and developmental toxicity. They have decomposed the process in two phases. First, surrogate endpoints (*in vitro* and *in vivo* assays) are chosen to predict the toxicological endpoint [52]. Second, computational models are built to predict these surrogate endpoints [23,53].

In the first study [52], experimental results for gene mutation tests (e.g. *Salmonella* assay), clastogenicity tests (e.g. *in vitro* chromosome aberrations and *in vivo* micronucleus), DNA damage tests, cell transformation tests and reprotox tests were evaluated to find a correlation with carcinogenicity in rodents. Among the 63 *in vitro* tests evaluated, 14 were found to be correlated to carcinogenicity (e.g. gene mutation in *Salmonella* and *in vivo* micronucleus), whereas others were poorly correlated (e.g. mouse lymphoma gene mutation [MLA] and *in vitro* chromosome aberration [CA]). In the second step [53], surrogate endpoints were predicted and used when no experimental data were available. The results showed that the correlations between carcinogenicity and the 14 *in silico* predicted surrogate endpoints were as good as the correlations between carcinogenicity and the experimental results. This demonstrates that computational approaches could be used to

complement experimental data in the pharmacological profiling approach.

Several combinations of surrogate assays have been tested to predict carcinogenicity. Matthews *et al.* reported results comparable to those published by Kirkland *et al.* [54]: combining an experimental *Salmonella* mutation assay with *in vitro* mammalian assays achieved high sensitivity (95%) but low specificity (5%) [52]. This was expected; they report in the same paper that several of these *in vitro* mammalian assays (MLA and CA) are poorly correlated with carcinogenicity. In the companion paper [53], they report high positive predictive values (76.9%) with high specificity (76.4%) for another combination of surrogate assays. Interestingly, complementing experimental results with predicted values did not decrease the overall performance of predictions significantly [53].

Fliri *et al.* [55–57] have extended the pioneering studies of Weinstein [58]. They based their study on a matrix of biological activities (target modulation) for 1567 drugs (marketed, withdrawn or failed) against 92 targets. The compounds were clustered according to their pharmacological profile (termed 'biological spectra analysis'). Biospectra not only aim to predict the affinity of new compounds [57] but also expect to help to identify how a small molecule perturbs a host system and, as such, to predict *in vivo* effects [59].

Finally, Bender *et al.* [40] report a method for predicting adverse reactions across hundreds of categories from the structure only. The authors built computational models for predicting the biological activity of novel structures on 70 targets using data collected in the Wombat database [60], as well as in-house information. Besides, they build computational models for predicting adverse effects using the World Drug Index (Thomson Scientific). The two sets of models are then matched with the aim of creating a link between adverse effects and biological profiles.

Although comparable in essence to the chemical diversity approach described above, the latter two approaches are new by their coverage (number of targets and adverse effects).

Applications

Extensive animal toxicity studies will usually not start before the preclinical candidate stage, and human toxicity studies will start even later. When one of these studies reveals significant toxicity and stops a project, a significant amount of time has already been spent optimizing the potency and the pharmacokinetic profile of the compound, and huge amounts of money have been invested in clinical trials. Eventually, all the money and the time invested are completely lost.

The role of computational toxicology is not to eliminate attrition; rather, it aims to shift attrition earlier in the discovery process to fail early and fail cheap [61]. The current trend, therefore, is to 'be aware' of potential toxicological issues to prioritize chemical series at the hit discovery stage or to try to fix the proven liabilities during lead optimization, just as potency and pharmacokinetics properties are being optimized. By delivering better candidates, it is hoped that discovery projects will have an increased probability of success in clinical development. The following sections focus on several applications of the computational methods described above to support a project in its early stages (hit discovery and lead optimization).

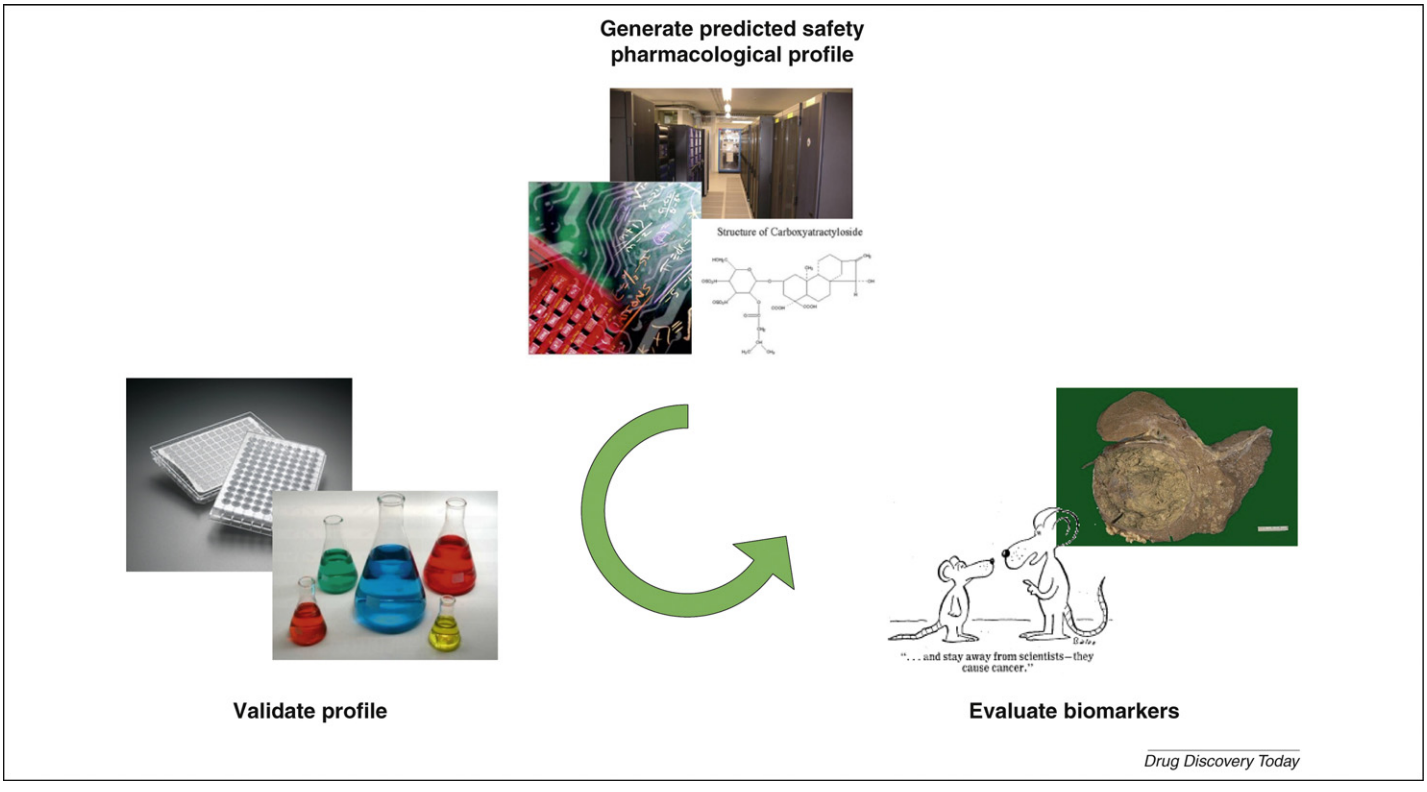

FIGURE 1

Assay prioritization. The predicted safety pharmacological profile is generated from the structure. Predictions have to be confirmed *in vitro* on isolated targets or functional assays. Biomarkers are then evaluated during early pharmacology studies to monitor *in vivo* the confirmed *in vitro* activities.

Predictions to ask the right question

By definition, little is known about new chemical entities and, in particular, about their toxicity. A straight comparison with other known chemicals is hardly possible because the structural changes that are directed at improving on-target potency or pharmacokinetics can also substantially modify their off-target profile [21].

Ideally, one would test new compounds as early as possible in every available *in vitro* profiling assay to identify potential off-target activities. Because of cost and practicality aspects, however, this is not possible, and only a subset of available assays is selected for a few compounds per project.

Computer predictions do not suffer much from cost and speed limitations. As soon as a model has been built, any number of structures can be tested, in a very short time frame. They suffer, however, from a large number of false positives: they tend to predict too many side effects for a given structure. Looking at this deficiency with a positive eye, computer predictions can be used as a pre-screen and followed up by *in vitro* or *in vivo* testing. In other words, they can help to prioritize *in vitro* assays [34,62,63] (Figure 1).

Pelletier *et al.* [64] reported on the use of computer models to evaluate the need for phospholipidosis testing. This work describes a process that aims to spend as little time as possible on compounds that will not be easy to develop (the fail early, fail cheap strategy) [65]. The authors have tuned their predictive model to increase the negative predictive value of the model (i.e. the proportion of compounds that are predicted negative that are correctly predicted). Therefore, when a structure is predicted to be negative, it is likely that it will be devoid of the liability and can be progressed quickly in the project without testing. The likelihood that this compound will fail in clinical development because it induces phospholipidosis is low. Structures that are predicted to be positive can potentially induce phospholipidosis (although there will be a lot of false positives), and phospholipidosis induction is tested first *in vitro* and then, if necessary, *in vivo* to validate or reject the prediction.

Supporting animal testing

Although the testing of classical endpoints can eliminate major problems, minor events can remain undetected, simply because the right biomarker was not evaluated *in vivo*. One benefit of employing *in silico* predictions is to highlight which biomarkers should be evaluated and in which populations, in addition to the classical endpoints required by regulatory authorities.

As for *in vitro* safety pharmacological profiling, it is not practical to evaluate every biomarker, and only the major potential issues are studied (e.g. genotoxicity and hERG blockade). In addition, an unwanted event might not happen in a population selected randomly. Individuals might respond differently when they are healthy or diseased, or aged, and so on. It is, therefore, not just necessary to evaluate all biomarkers; they should also be evaluated in relevant populations.

One use of computer predictions is to provide a predicted profile that is then validated *in vitro* (Figure 1). Once the actual profile is known, it suggests a list of predicted adverse events with their mechanisms of action. This list helps to design new experiments (Figure 1). In particular, they help to select appropriate biomarkers (through systems biology) and to select a population in which an unwanted side effect is likely to happen. Put in another way, they give experimentalists some guidance on how to further investigate and validate predicted effects. These results are then available to support informed decision in the context of the project.

From observations to new compounds

Conversely, adverse events are sometimes observed in pharmacology studies at the discovery stage. They are seldom reported to the

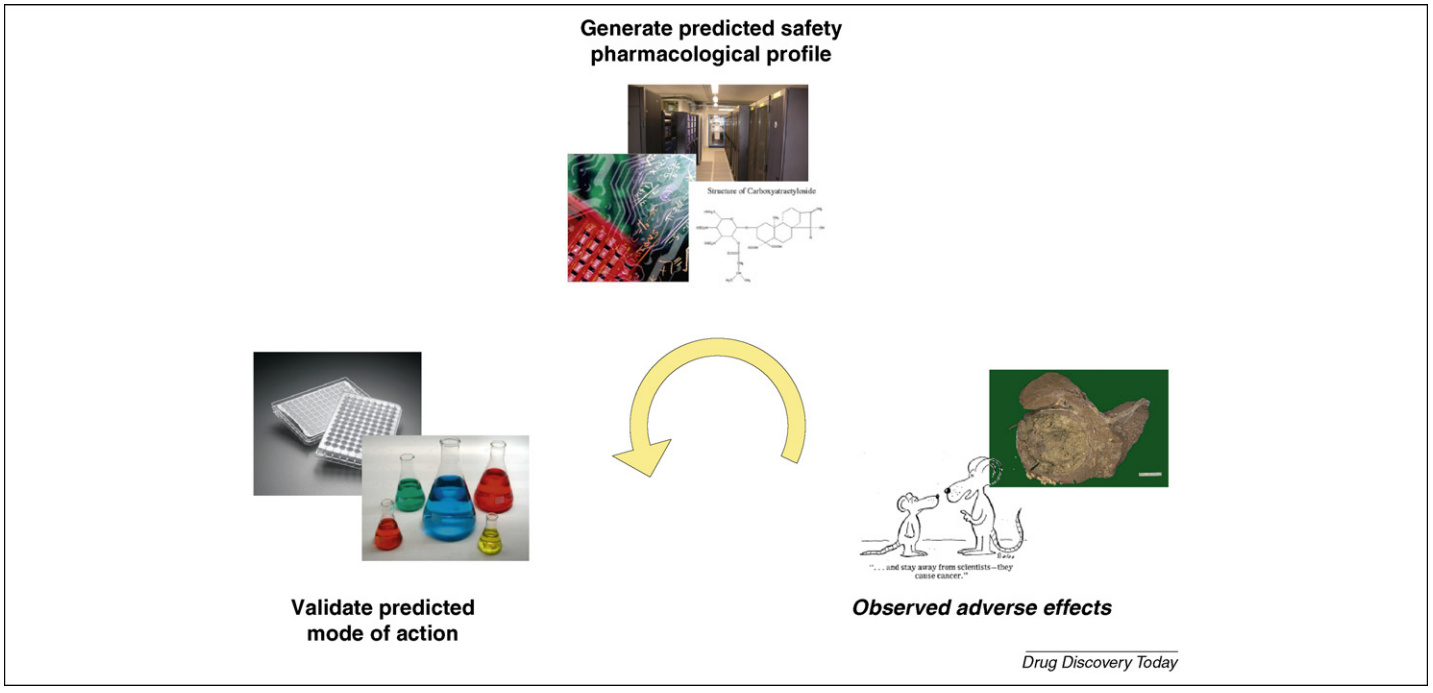

FIGURE 2

Explain *in vivo* observed toxicity. After an adverse effect has been observed *in vivo*, a predicted safety pharmacological profile is generated; this profile is then validated *in vitro* on isolated targets or in functional assays to suggest potential modes of action.

project team because they are thought to be too general, and most of the time they can result from many mechanisms of action. Even if they are reported, they do not alone help to improve compounds because they do not provide any indication of which part to modify in the structure.

Computational methods can help to identify the mode of action, as well as the chemical determinants that are associated with the observed effect (Figure 2). After an adverse effect has been observed *in vivo*, a predicted safety pharmacological profile is generated; this profile is then validated *in vitro* on isolated targets or in functional assays. The role of the predicted safety pharmacological profiling (PSPP) is to identify relevant assays in which the compound has to be tested. Once the predicted mode of action is validated, PSPP proposes structural modifications to chemists to help to remove the liability (see case study, below).

Case study

Despite the fact that Genkyotex's computational toxicology platform is still under construction, we routinely apply predicted safety pharmacological profiling in our lead optimization programs. The following example is about the identification, *in vivo* confirmation and removal of a potential liability. This case study illustrates that using predicted safety pharmacological profiles neither kills valuable leads nor slows down lead optimization. Indeed, it helped us to identify potential liabilities early and fix them with a limited impact on the budget.

During lead optimization, the pharmacological profile was predicted for several compounds in a chemical series. One of the compounds was predicted active on three targets. The list of pharmacological effects associated with these three targets was assembled from the literature, with the constraint that the effects should be easily observable. A list of four effects was presented to the project's pharmacologist. Among the side effects, one had actually been observed *in vivo* during the pharmacokinetics study. A biomarker was developed to further investigate the side effect observed, and it was monitored in a two-week *in vivo* experiment. This biomarker was measured at the end of the experiment, and changes seemed to be significant with a p-value of 5.10^{-4} ($n = 14$). The prediction also provided us with the chemical moieties and structural requirements for the inhibition of the target we predicted associated to this effect. These features were removed from new compounds, either by removing the chemical moieties or by introducing steric hindrance. The biomarker is still used to monitor the side effect. Interestingly, the new compounds do not induce a change in the biomarker level of treated animals versus control animals, which demonstrates that we removed the liability by altering the structure.

Without the predictions, the side effect would probably have remained hidden until clinical development. Depending on the intended indication, this could have forced us to stop the project, which is not the best scenario for a start-up company. Predictions helped us to identify the liability early in the process, and in this particular case, the liability has probably been removed. This example shows that, contrary to a common thought, predictions do not kill valuable leads; instead, they help turn them into better candidates.

Outlook

The relevance of *in silico* and *in vitro* profiling combined with toxicogenomics studies to predict toxic events is increasing [66,67]. It can be more predictive than results from animal studies when there are significant genetic differences between human and rodent [68]. For example, human and rodent bradykinin receptors B1 and B2 are inhibited by different types of molecules [69,70]. If a molecule inhibits human bradykinin receptors, toxicology in rodents would be of limited interest because it does not correlate with human data. Instead *in silico* and *in vitro* pharmacological profiling would suggest testing on human recombinant protein and, therefore, point out the need for monitoring nociception, inflammation, vasodilation and cough in human clinical studies [68]. Put in another way, *in vivo* results obtained from animal studies might not be of relevance for humans, but *in silico* and *in vitro* profiling can help to identify and then correct those issues [1].

Exploiting today's structure–activity relationships could prevent scientists from repeating errors of the past. One should remember, of course, that no statistical or data mining method is able to extract relationships from information that was not provided to it [37]. However, a lot of information is available today. Systematic analysis of this data and interpretation of predictions to raise the right questions can participate in the delivery of safer drugs, more quickly, and at a lower cost.

References

1 Nigsch, F. *et al.* (2009) Computational toxicology: an overview of the sources of data and of modelling methods. *Exp. Opin. Drug Metab. Toxicol.* 5, 1–14

2 Benigni, R. and Bossa, C. (2006) Structural alerts of mutagens and carcinogens. *Curr. Comput. Aid. Drug Discov.* 2, 169–176

3 Greene, N. (2002) Computer systems for the prediction of toxicity: an update. *Adv. Drug Deliv. Rev.* 54, 417–431

4 Langham, J.J. and Jain, A.N. (2008) Accurate and interpretable computational modeling of chemical mutagenicity. *J. Chem. Inf. Model.* 48, 1833–1839

5 Valerio, L.G., Jr *et al.* (2007) Prediction of rodent carcinogenic potential of naturally occurring chemicals in the human diet using high-throughput QSAR predictive modeling. *Toxicol. Appl. Pharmacol.* 222, 1–16

6 Simon-Hettich, B. *et al.* (2006) Use of computer-assisted prediction of toxic effects of chemical substances. *Toxicology* 224, 156–162

7 Woo, Y.T. *et al.* (1995) Development of structure–activity relationship rules for predicting carcinogenic potential of chemicals. *Toxicol. Lett.* 79, 219–228

8 Muster, W. *et al.* (2008) Computational toxicology in drug development. *Drug Discov. Today* 13, 303–310

9 Muller, L. *et al.* (2007) Strategies for using computational toxicology methods in pharmaceutical R&D. In *Computational Toxicology: Risk Assessment for Pharmaceutical and Environmental Chemicals* (Ekins, S., ed.), pp. 545–580, Wiley-Interscience

10 Anzali, S. *et al.* (2001) Discriminating between drugs and non-drugs by prediction of activity spectra for substances (PASS). *J. Med. Chem.* 44, 2432–2437

11 Kawai, K. *et al.* (2008) Predictive activity profiling of drugs by topological-fragment-spectra-based support vector machines. *J. Chem. Inf. Model.* 48, 1152–1160

12 Mohan, C.G. *et al.* (2007) Computer-assisted methods in chemical toxicity prediction. *Mini Rev. Med. Chem.* 7, 499–507

13 Snyder, R.D. (2004) Assessment of the sensitivity of the computational programs DEREK, TOPKAT, MCASE in the prediction of the genotoxicity of pharmaceutical molecules. *Environ. Mol. Mutagen.* 43, 143–158

14 Gombar, V.K. (2008) Quantitative Structure Activity Relationships (QSAR). *ADMET, SMI*, The Hatton, at etc. venues, London
15 Worth A.P. *et al.* (2007) The characterisation of (quantitative) structure–activity relationships: preliminary guidance (http://ecb.jrc.ec.europa.eu/qsar/background/)
16 Tetko, I.V. *et al.* (2006) Can we estimate the accuracy of ADME-Tox predictions? *Drug Discov. Today* 11, 700–707
17 Tetko, I.V. *et al.* (2008) Critical assessment of QSAR models of environmental toxicity against *Tetrahymena pyriformis*: focusing on applicability domain and overfitting by variable selection. *J. Chem. Inf. Model.* 48, 1733–1746
18 US Patent (2009) US6,6,036,049
19 Gombar, V.K. and Enslein, K. (1996) Assessment of n-octanol/water partition coefficient: when is the assessment reliable? *J. Chem. Inf. Comput. Sci.* 36, 1127–1134
20 Dragos, H. *et al.* (2009) Predicting the predictability: a unified approach to the applicability domain problem of QSAR models. *J. Chem. Inf. Model.* 49, 1762–1776
21 Scheiber, J. (2007) *Virtual Design of Compound Libraries: Consideration of the Success and Challenges of Pharmacophore-Derived Compound Libraries and Scaffold Hopping.* The Design and Synthesis of Quality Compound Libraries
22 Kubinyi, H. (2002) Chemical similarity and biological activities. *J. Braz. Chem. Soc.* 13, 717–726
23 Matthews, E.J. *et al.* (2007) A comprehensive model for reproductive and developmental toxicity hazard identification: II. Construction of QSAR models to predict activities of untested chemicals. *Regul. Toxicol. Pharmacol.* 47, 136–155
24 Wang, C.Y. *et al.* (2006) Identification of previously unrecognized antiestrogenic chemicals using a novel virtual screening approach. *Chem. Res. Toxicol.* 19, 1595–1601
25 Mukherjee, S. and Saha, A. (2006) Molecular modeling studies of estrogen receptor modulators. *Curr. Comput. Aid. Drug Des.* 2, 229–253
26 Guengerich, F.P. and Macdonald, J.S. (2007) Applying mechanisms of chemical toxicity to predict drug safety. *Chem. Res. Toxicol.* 20, 344–369
27 Piegorsrch, W.W. and Zeiger, E. (1991) Measuring intra-assay agreement for the Ames Salmonella assay. In *Statistical Methods in Toxicology* (Hotham, L., ed.), pp. 35–41, Springer-Verlag, Heidelberg
28 Kazius, J. *et al.* (2005) Derivation and validation of toxicophores for mutagenicity prediction. *J. Med. Chem.* 48, 312–320
29 Agrafiotis, D.K. *et al.* (2007) Recent advances in chemoinformatics. *J. Chem. Inf. Model.* 47, 1279–1293
30 Nigsch, F. and Mitchell, J.B.O. (2008) Toxicological relationships between proteins obtained from protein target predictions of large toxicity databases. *Toxicol. Appl. Pharmacol.* 231, 225–234
31 Sheridan, R.P. *et al.* (2009) QSAR models for predicting the similarity in binding profiles for pairs of protein kinases and the variation of models between experimental data sets. *J. Chem. Inf. Model.* 49 (8.), 1974–1985
32 Sciabola, S. *et al.* (2008) Predicting kinase selectivity profiles using Free-Wilson QSAR analysis. *J. Chem. Inf. Model.* 48, 1851–1867
33 Ai, N. *et al.* (2009) Understanding nuclear receptors using computational methods. *Drug Discov. Today* 14, 486–494
34 Rabinowitz, J.R. *et al.* (2008) Computational molecular modeling for evaluating the toxicity of environmental chemicals: prioritizing bioassay requirements. *Environ. Health Perspect.* 116, 573–577
35 Poroikov, V. *et al.* (2001) Top 200 medicines: can new actions be discovered through computer aided prediction? *SAR QSAR Environ. Res.* 12, 327–344
36 Poroikov, V.V. (2003) PASS biological activity spectrum predictions in the enhanced open NCI database browser. *J. Chem. Inf. Comput. Sci.* 43, 228–236
37 Helma, C. and Kazius, J. (2006) Artificial intelligence and data mining for toxicity prediction. *Curr. Comput. Aid. Drug Discov.* 2, 123–133
38 Gallegos-Saliner, A. *et al.* (2008) Toxmatch—a chemical classification and activity prediction tool based on similarity measures. *Regul. Toxicol. Pharmacol.* 52, 77–84
39 Merlot, C. *et al.* (2002) Fragment analysis in small molecule discovery. *Curr. Opin. Drug Discov. Dev.* 5, 391–399
40 Bender, A. *et al.* (2007) Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. *ChemMedChem* 2, 861–873
41 Merlot, C. (2008) *Early Safety Evaluation—Looking for Safety at the Discovery Stage.* SMI ADMET
42 Richard, A.M. (2006) Future of toxicology—predictive toxicology: an expanded view of 'chemical toxicity'. *Chem. Res. Toxicol.* 19, 1257–1262
43 Yang, C. *et al.* (2006) Landscape of current toxicity databases and database standards. *Curr. Opin. Drug Discov. Dev.* 9, 124–133
44 Richard, A.M. *et al.* (2006) Chemical structure indexing of toxicity data on the internet: moving toward a flat world. *Curr. Opin. Drug Discov. Dev.* 9, 314–325
45 Judson, R. *et al.* (2008) ACToR—aggregated computational toxicology resource. *Toxicol. Appl. Pharmacol.* 233, 7–13
46 Judson, R. *et al.* (2009) The toxicity data landscape for environmental chemicals. *Environ. Health Perspect.* 117, 685–695
47 Benz, R.D. (2007) Toxicological and clinical computational analysis and the US FDA/CDER. *Expert Opin. Drug Metab. Toxicol.* 3, 109–124
48 Yang, C. *et al.* (2006) The art of mining the minefields of toxicity databases to link chemistry to biology. *Curr.Comput. Aid. Drug Des.* 2, 135–150
49 Martin, M.T. *et al.* (2009) Profiling chemicals based on chronic toxicity results from the U.S. EPA ToxRef Database. *Environ. Health Perspect.* 117, 392–399
50 Williams-Devane, C.R. *et al.* (2009) Toward a public toxicogenomics capability for supporting predictive toxicology: survey of current resources and chemical indexing of experiments in GEO and ArrayExpress. *Toxicol. Sci.* 109, 358–371
51 Mestres, J. *et al.* (2006) Ligand-based approach to *in silico* pharmacology: nuclear receptor profiling. *J. Chem. Inf. Model.* 46, 2725–2736
52 Matthews, E.J. *et al.* (2006) An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: I. Identification of carcinogens using surrogate endpoints. *Regul. Toxicol. Pharmacol.* 44, 83–96
53 Matthews, E.J. *et al.* (2006) An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: II. Identification of genotoxicants, reprotoxicants, and carcinogens using *in silico* methods. *Regul. Toxicol. Pharmacol.* 44, 97–110
54 Kirkland, D. *et al.* (2005) Evaluation of the ability of three *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens. I. Sensitivity, specificity, and relative predictivity. *Mutat. Res.* 584, 1–256
55 Fliri, A.F. *et al.* (2007) Analysis of system structure–function relationships. *ChemMedChem* 2, 1774–1782
56 Fliri, A.F. *et al.* (2005) Biospectra analysis: model proteome characterizations for linking molecular structure and biological response. *J. Med. Chem.* 48, 6918–6925
57 Fliri, A.F. *et al.* (2005) Biological spectra analysis: linking biological activity profiles to molecular structure. *Proc. Natl. Acad. Sci. U. S. A.* 102, 261–266
58 Weinstein, J.N. *et al.* (1997) An information-intensive approach to the molecular pharmacology of cancer. *Science* 275, 343–349
59 Froloff, N. (2005) Probing drug action using *in vitro* pharmacological profiles. *Trends Biotechnol.* 23, 488–491
60 Olah, M. *et al.* (2007) WOMBAT and WOMBAT-PK: bioactivity databases for lead and drug discovery. In *Chemical Biology: From Small Molecules to Systems Biology and Drug Design* (Schreiber, S.L., Kapoor, T.M., Wess, G., eds), pp. 760–786, Wiley-Interscience
61 Fielden, M.R. and Kolaja, K.L. (2008) The role of early *in vivo* toxicity testing in drug discovery toxicology. *Expert Opin. Drug Saf.* 7, 107–110
62 Jacobson-Kram, D. and Contrera, J.F. (2007) Genetic toxicity assessment: employing the best science for human safety evaluation part I: early screening for potential human mutagens. *Toxicol. Sci.* 96, 16–20
63 McGee, P. (2005) Modeling success with *in silico* tools. *Drug Discov. Dev.* 8, 24–28
64 Pelletier, D.J. *et al.* (2007) Evaluation of a published *in silico* model and construction of a novel Bayesian model for predicting phospholipidosis inducing potential. *J. Chem. Inf. Model.* 47, 1196–1205
65 Merlot, C. (2008) *In silico* methods for early toxicity assessment. *Curr. Opin. Drug Discov. Dev.* 11, 80–85
66 Fielden, M.R. *et al.* (2002) *In silico* approaches to mechanistic and predictive toxicology: an introduction to bioinformatics for toxicologists. *Crit. Rev. Toxicol.* 32, 67–112
67 Leighton, J.K. *et al.* (2006) Workgroup report: review of genomics data-based experience with mock submissions—view of the CDER Pharmacology Toxicology Non-clinical Pharmacogenomics Subcommitee. *Environ. Health Perspect.* 114, 573–578
68 Whitebread, S. *et al.* (2005) *In vitro* safety pharmacology profiling: an essential tool for successful drug development. *Drug Discov. Today* 10, 1421–1433
69 Mayne, J.T. *et al.* (2006) Informed toxicity assessment in drug discovery: systems-based toxicology. *Curr. Opin. Drug Discov. Dev.* 9, 75–83
70 Regoli, D. *et al.* (1997) B1 and B2 kinin receptors in various species. *Immunopharmacology* 36, 143–147