# The Future of Computational Models for Predicting Human Toxicities

*Sean Ekins[1,2,3] and Antony J. Williams[4]*

[1]Collaborations in Chemistry, Fuquay-Varina, NC, USA; [2]Collaborative Drug Discovery, Burlingame, CA, USA; [3]Department of Pharmaceutical Sciences, University of Maryland, MD, USA; [4]Royal Society of Chemistry, Wake Forest, NC, USA

**Summary**

*New regulations requiring toxicity data on chemicals and an increasing number of efforts to predict the likelihood of failure of molecules earlier in the drug discovery process are combining to increase the utilization of computational models to toxicity. The potential to predict human toxicity directly from a molecular structure is feasible. By using the experimental properties of known compounds as the basis of predictive models it is possible to develop structure activity relationships and resulting algorithms related to toxicity. Several examples have been published recently, including those for drug-induced liver injury (DILI), the pregnane X receptor, P450 3A4 time dependent inhibition, and transporters associated with toxicities. The versatility and potential of using such models in drug discovery may be illustrated by increasing the efficiency of molecular screening and decreasing the number of animal studies. With more computational power available on increasingly smaller devices, as well as many collaborative initiatives to make data and toxicology models available, this may enable the development of mobile apps for predicting human toxicities, further increasing their utilization.*

*Keywords: Bayesian models, databases, drug-induced liver injury, mobile apps, P450 3A4, Pregnane X receptor, REACH, ToxCast*

## 1 Introduction

In the last decade we have witnessed a perfect storm in terms of the impact on our ability to predict toxicity. Compared to a decade ago, compute power is far cheaper, resulting in faster predictions. Hardware continues to shrink at a rate that much of the available modeling software can now easily run on a laptop or notebook computer. The move to cloud-based servers means that compute power is available for intensive calculations with results served up easily via web-based interfaces. The number of chemical compounds now available via public databases is in the tens of millions. The data associated with the structures is provided in a manner that allows it to be downloaded and used to build computational models. There are also increasing amounts of data collated for toxicity endpoints, such as for drug-induced liver injury (DILI) (Cruz-Monteagudo et al., 2007; Ekins et al., 2010b; Fourches et al., 2010; Greene et al., 2010), the human Ether-à-go-go Related Gene (hERG) ion channel (Shamovsky et al., 2008; Hansen et al., 2009), Pregnane X receptor (PXR) (Pan et al., 2010) and for transporters involved in toxicities (Diao et al., 2009, 2010; Zheng et al., 2009). There are far more examples of computational models and software that are now freely accessible that can be used for toxicity prediction, and as the community becomes aware of these tools they will be used increasingly often. Computational toxicity prediction is certainly starting to reach more researchers and is becoming more widely accepted (Cronin and Livingstone, 2004; Helma, 2005; Ekins, 2007).

What is driving this change? Based on the discourse in so many venues (online, in journals, and in popular media), it is clear that the pharmaceutical industry is seen as having reached a productivity tipping point, and the cost of new drug development is extremely high (Munos, 2009). Initiatives such as REACH and other new legislations are likely to require even more testing of compounds for which "there is insufficient information on the hazards that they pose to human health and the environment" (http://ec.europa.eu/environment/chemicals/reach/reach_intro.htm). To respond to this need, initiatives such as ToxCast aim to "develop ways to predict potential toxicity and to develop a cost-effective approach for prioritizing the thousands of chemicals that need toxicity testing" (http://www.epa.gov/ncct/toxcast/). Ultimately, we need to understand the toxicity of far more compounds than is reasonable or ethical to screen *in vivo* in animals or even *in vitro*. There has to be a first tier prioritization to filter molecules of interest to pharmaceutical, consumer products, and environmental researchers. The predictive models that have been built in recent years, and those that are becoming available with efforts such as eTox and OpenTox, may be of some utility. So a perfect storm of many factors could position computational models well for predicting human toxicities for the future.

## 2 Simple methods

In pharmaceutical research there has been a proliferation of rules since the Rule of Five described orally active compounds in terms of a few simple molecular properties (Lipinski et al., 1997). Could very simplistic approaches like this be used to

predict toxicity? For example, Pfizer scientists studied the relationship between physicochemical properties and animal *in vivo* tolerance for 245 preclinical compounds. They determined that compounds with ClogP <3 and total polar surface area >75A$^2$ were preferable, with fewer toxicity findings. (Hughes et al., 2008). Many pharmaceutical companies have developed computational filters to remove reactive molecules from their screening datasets (Hann et al., 1999; Walters and Murcko, 2002; Pearce et al., 2006). Abbott developed an assay to detect thiol reactive molecules by NMR (ALARM NMR) (Huth et al., 2005, 2007), and this data was used to create a Bayesian classifier model to predict reactivity (Metz et al., 2007). It recently has been suggested that molecules failing such reactivity filters may correlate (Ekins and Freundlich, 2011) with the number of violations of the Rule of Five (Lipinski et al., 1997). There is certainly potential to develop more advanced rules that can be interpreted readily by chemists and biologists, and this may be possible by using larger datasets or those focused on individual toxicity types.

## 3 Examples of human toxicities modeled

The amount of biological data being created due to high-throughput screening requires databases for storage (see next section) and, in the process, also creates a wealth of information for developing computational models. For example, Pfizer used data for more than 100,000 compounds extracted from the literature and measured in their own laboratories against many different assays. They used these data to develop a Bayesian model for predicting cytotoxicity (Langdon et al., 2010) with training Receiver Operator Characteristic = 0.84. Other generic models of human toxicities have been published as well.

Drug-metabolism in the liver can convert some drugs into highly reactive intermediates that, in turn, can adversely affect the structure and functions of the liver. DILI is therefore one of the most important reasons for drug development failure at both pre-approval and post-approval stages. A list of approximately 300 drugs and chemicals with a classification scheme based on clinical data for hepatotoxicity has been assembled previously by Pfizer in order to evaluate an *in vitro* human hepatocyte imaging assay technology (HIAT), which had a concordance of 75% with clinical hepatotoxicity (Xu et al., 2008). A Bayesian classification model generated with this data was evaluated by leaving out either 10%, 30% or 50% of the data and rebuilding the model 100 times in order to generate the cross validated ROC (Ekins et al., 2010b). In each case the leave out 10%, 30% or 50% testing AUC value was comparable to the leave-one-out approach (0.86) and these values were very favorable, indicating good model robustness (Ekins et al., 2010b). The Bayesian model was tested with 237 new compounds with concordance ~60%, specificity 67% and sensitivity 56%, which were comparable with the internal validation statistics. A subset of 37 compounds, of most interest clinically, showed similar testing values with a concordance greater than 63% (Ekins et al., 2010b). This example represents the first large-scale testing of a machine learning model for DILI that uses a similarly sized training and test set. The overall concordance of the model is lower (~60-64% depending on test set size) than that observed previously for *in vitro* HIAT (75% (Xu et al., 2008)). However, the test-set statistics are similar to those reported elsewhere using structural alerts (Greene et al., 2010). This work suggests that currently available data on compounds can be used to predict, with reasonable accuracy, future compounds and their potential for DILI.

An *in silico-in vitro* approach was used to predict compounds likely to cause time-dependent inhibition (TDI) of P450 3A4 in human liver microsomes. The Bayesian classification approach (Xia et al., 2004; Bender, 2005), along with simple, interpretable molecular descriptors as well as FCFP_6 descriptors (Jones et al., 2007), was used to classify P450 3A4 TDI. The models used between 1853 and 2071 molecules and were tested with molecules excluded from the models (Zientek et al., 2010). All of the receiver operator characteristic curves show better than random ability to identify the TDI positive molecules and these models were integrated into the Pfizer testing paradigm (Zientek et al., 2010).

Drug transporters, another class of proteins that is starting to be well studied for its potential role in toxicity, can actively take up or efflux compounds, or plays a key physiological role in the transport of metabolites and endogenous compounds. Interference with transporters may therefore result in toxicity. For example, many classes of compounds may cause rhabdomyolysis or muscle weakness, and this may be a result of inhibition of the Organic Cation/Carnitine Transporter 2 (Diao et al., 2009, 2010). The human apical sodium-dependent bile acid transporter (ASBT) is an important mechanism for intestinal bile acid reabsorption and plays a critical role in bile acid and cholesterol homeostasis. Since ASBT is the main mechanism for intestinal bile acid re-absorption, ASBT impairment may be associated with colorectal cancer development. Separate 3D-QSAR and Bayesian models were developed using 38 ASBT inhibitors (Zheng et al., 2009). Validation analysis showed that both models exhibited good predictability in determining whether a drug is a potent or non-potent ASBT inhibitor. Many additional FDA-approved drugs from diverse classes, such as the dihydropyridine calcium channel blockers and HMG CoA-reductase inhibitors, were found to be ASBT inhibitors, and work is ongoing to associate these with any toxicity because of this activity.

## 4 More accessible data and software

The prediction of metabolites and sites of metabolism is important for commercial and environmental chemicals. It is a critical first step prior to running computational models for other endpoints, as it is likely that potential metabolites could be as important as the parent compound. There are several new methods for metabolite prediction that have focused on CYP3A4, including MLite for CYP3A4 (Oh et al., 2008), RegioSelectivity-Predictor (RS-Predictor) (Zaretzki et al., 2011)

and SMARTCyp, a fragment-based method only validated for CYP3A4 (Rydberg et al., 2010). Another method is MetaPred, based on SVM models for CYPs (Mishra et al., 2010). The MetaPrint2D (Boyer et al., 2007; Carlsson et al., 2010) method is a more generic method in that it does not focus on a single enzyme but is reaction-based. This tool has been developed for the prediction of metabolic sites from input chemical structures and is freely available online (http://www-metaprint2d.ch.cam.ac.uk/metaprint2d/). The method uses an internal database based on historical metabolite data derived from the 2008.1 version of the Accelrys Metabolite database for all transformations, or only those in human, dog, or rat. This method is fast (50 ms per compound), (Afzelius et al., 2007) and in the limited testing described to date by AstraZeneca performs well compared to other algorithms in the field such as SmartCyp (Rydberg et al., 2010) and MetaSite (Cruciani et al., 2005). A second tool developed by this group, MetaPrint2D-React, also lists potential metabolites at each location that is predicted as a site of metabolism. To date there are no publications that validate this tool, but it would be relatively straightforward for anyone interested to run a defined set of compounds of interest to see how the predictions compare to their own or published knowledge. This tool has some promise as an alternative or adjunct to the older rule-based methods for metabolite prediction (Jolivette and Ekins, 2007).

ADME properties have been modeled using an array of machine learning algorithms such as support vector machines (Kortagere et al., 2008), Bayesian modeling (Klon et al., 2006), Gaussian processes (Obrezanova et al., 2007), or others (Zhang et al., 2008). A major challenge remains the ability to share such models (Spjuth et al., 2010). Researchers have both proposed and provided a proof of concept using open descriptors and modeling tools to model very large ADME datasets at Pfizer (Gupta et al., 2010). Very large training sets of approximately 50,000 molecules and a test set of approximately 25,000 molecules were used with human liver microsomal metabolic stability data (Gupta et al., 2010). A C5.0 decision tree model demonstrated that the Chemistry Development Kit (Steinbeck et al., 2006) descriptors together with a set of SMARTS keys had good statistics (Kappa = 0.43, sensitivity = 0.57, specificity = 0.91, positive predicted value (PPV) = 0.64) equivalent to models built with commercial MOE2D software and the same set of SMARTS keys (Kappa = 0.43, sensitivity = 0.58, specificity = 0.91, PPV = 0.63). This observation also was confirmed upon extension of the dataset to ~193,000 molecules and generation of a continuous model using Cubist (http://www.rulequest.com/download.html). When the continuous predictions and actual values were binned to get a categorical score, an almost identical Kappa statistic (0.42) was observed (Gupta et al., 2010). The same group also evaluated other large datasets and found that open source tools and commercial descriptors were equivalent. Such studies raise the possibility that open source cheminformatics methods could be used to share toxicity models developed by different industrial or academic groups, thus surmounting the considerable commercial costs involved in software licensing.

The increased interest in groups sharing data underscores the need for better databases for housing the toxicity data. Williams (Williams et al., 2009; Williams and Ekins, 2011) analyzed the quality of public domain databases in relation to the curation of the ChemSpider database, and he identified common issues regarding the relationships between chemical structures and associated chemical names, generally drug names and associated synonyms. A recently published alert on data quality for internet-based chemistry resources used the NCGC (Williams and Ekins, 2011) "NPC browser" database to illustrate the need for government funding for curation of public chemistry databases. We also have proposed the construction of a validated ADME/Tox database (Ekins and Williams, 2010) and suggested an updated strategy for how the scientific community could build such a resource (Williams et al., 2011b). So, it will be important that any new toxicity databases ensure that molecule and data quality are high enough to be used for computational modeling efforts.

## 5 Making toxicity predictions mobile

With the shrinking size of computers and the increasing computational power of mobile devices such as smartphones and tablet computers, there is an opportunity to deliver toxicity models and other cheminformatics tools as mobile "Apps" (Williams et al., 2011a). The area of "green chemistry," which recently has recognized the importance of computational models in designing new compounds with reduced hazard (Voutchkova et al., 2010), has already seen the development of the first green chemistry app called Green Solvents (http://www.scimobileapps.com/index.php?title=Green_Solvents). While it is not a predictive tool, its development was rapid and it brought forth a green solvent list developed by the ACS Green Chemistry Institute (GCI), with many pharmaceutical partners, in the hope that chemists can use it as a look-up resource while they are in the lab. To date, the major cheminformatics tool vendors have not focused on mobile apps and certainly have not devoted research to new human toxicity models. As a result, there is a need for new technologies in this space, as the historical toxicity prediction tools are immature and companies have focused on developing similar technologies rather than innovating new approaches (Ekins et al., 2010a).

## 6 Collaboration

Scientific research creates and consumes tremendous volumes of data, but the data available on human toxicities makes up only a small fraction. Many industries have realized that they cannot do everything in house and have developed complex networks of collaborators or partners throughout R&D. Current areas of major interest are Open Innovation, Collaborative Innovation, Open Source and Open Data. Of relevance to toxicity prediction are the e-Tox (http://www.etoxproject.eu/), OpenTox (http://www.opentox.org/), OECD toolbox (http://www.oecd.

org/document/54/0,3746,en_2649_34379_42923638_1_1_1_1,00.html), OCHEM (http://ochem.eu/), and Open PHACTS (http://www.openphacts.org/) initiatives that promote a collaborative research and data sharing agenda. It is our belief that collaborative research and development is going to continue to feature even more prominently in the future of scientific research. It is essential, therefore, to manage information and computational resources across collaborations. From the computational toxicology perspective this means data and models that need to be stored and shared in order to prevent experiments from being repeated. We need intelligent information systems and an understanding of how to use them effectively to create and manage knowledge across these collaborations. A recent publication looked at this challenge for biomedical research (Ekins et al., 2011b), but many of the topics covered are equally applicable to consumer product and environmental chemical research. As research costs increase, all scientists will be driven to collaborate more and this will require novel databases and tools to selectively share data (Hohman et al., 2009; Ekins et al., 2011a) and perhaps create a new business model (Bunin and Ekins, 2011). The continued utilization of computational models for human toxicity will be critical to any researcher developing new molecules and to regulators and decision makers on all sides.

## 7 Conclusions

The efforts described in section 4 suggest that computational models for human toxicities may help to fill in data gaps that currently exist. Computational models can be used before screening compounds to prioritize those of most interest, as was recently demonstrated by scoring the ToxCast library with PXR docking models as a way to test a small fraction of the chemicals (Kortagere et al., 2010). Obviously in the case of ToxCast there are hundreds of endpoints that could be predicted to give a more complete picture of toxicity. But this dataset may be useful for validating computational models with the caveat that the chemistry space may be pretty narrow (although it will expand in phase II of this project). Computational toxicity models may also help us repurpose FDA approved drugs with a better side effect profile (Ekins and Williams, 2011; Ekins et al., 2011c). Many different industries now require some degree of toxicity prediction, and therefore the provision of computational tools, databases, and scientists with an understanding of how such methods can be used will be critical. It may, therefore, be equally important for those with a cross-industry perspective to collaborate and share toxicity data for the benefit of all. Ultimately, the focus on computational models for human toxicity should decrease the use of animal studies and perhaps ultimately replace them.

## References

Afzelius, L., Arnby, C. H., Broo, A., et al. (2007). State-of-the-art tools for computational site of metabolism predictions: comparative analysis, mechanistical insights, and future applications. *Drug Metab. Rev. 39*, 61-86.

Bender, A. (2005). *Studies on molecular similarity*. Cambridge: University of Cambridge.

Boyer, S., Arnby, C. H., Carlsson, L., et al. (2007). Reaction site mapping of xenobiotic biotransformations. *J. Chem. Inf. Model 47*, 583-590.

Bunin, B. A. and Ekins, S. (2011). Alternative business models for drug discovery. *Drug Disc. Today 16*, 643-645.

Carlsson, L., Spjuth, O., Adams, S., et al. (2010). Use of historic metabolic biotransformation data as a means of anticipating metabolic sites using MetaPrint2D and Bioclipse. *BMC Bioinformatics 11*, 362.

Cronin, M. T. D. and Livingstone, D. J. (2004). *Predicting chemical toxicity and fate*. Boca Raton, FL USA: CRC Press.

Cruciani, G., Carosati, E., De Boeck, B., et al. (2005). MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist. *J. Med. Chem. 48*, 6970-6979.

Cruz-Monteagudo, M.,Cordeiro, M. N., and Borges, F. (2007). Computational chemistry approach for the early detection of drug-induced idiosyncratic liver toxicity. *J. Comput. Chem. 29*, 533-549.

Diao, L.,Ekins, S., and Polli, J. E. (2009). Novel inhibitors of human organic cation/carnitine transporter (hOCTN2) via computational modeling and in vitro testing. *Pharm. Res. 26*, 1890-1900.

Diao, L., Ekins, S., and Polli, J. E. (2010). Quantitative structure activity relationship for inhibition of human organic cation/carnitine transporter. *Mol. Pharm. 7*, 2120-2130.

Ekins, S. (2007) *Computational Toxicology: risk assessment for pharmaceutical and environmental chemicals*. Hoboken, NJ, USA: John Wiley and Sons.

Ekins, S. and Williams, A. J. (2010). Precompetitive preclinical ADME/Tox data: Set it free on the web to facilitate computational model building to assist drug development. *Lab on a Chip 10*, 13-22.

Ekins, S., Gupta, R. R., Gifford, E., et al. (2010a). Chemical space: missing pieces in cheminformatics. *Pharm. Res. 27*, 2035-2039.

Ekins, S., Williams, A. J., and Xu, J. J. (2010b). A predictive ligand-based Bayesian model for human drug induced liver injury. *Drug Metab. Dispos. 38*, 2302-2308.

Ekins, S. and Freundlich, J. S. (2011). Validating new tuberculosis computational models with public whole cell screening aerobic activity datasets. *Pharm. Res. 28*, 1859-1869.

Ekins, S. and Williams, A. J. (2011). Finding promiscuous old drugs for new uses. *Pharm. Res. 28*, 1786-1791.

Ekins, S., Hohman, M., and Bunin, B. A. (2011a). Pioneering use of the cloud for development of the collaborative drug discovery (cdd) database. In S. Ekins, M. A. Z. Hupcey, and A. J. Williams (eds.), *Collaborative computational technologies for biomedical research*. Hoboken, NJ, USA: John Wiley and Sons.

Ekins, S., Hupcey, M. A. Z., and Williams, A. J. (2011b). *Col-*

*laborative computational technologies for biomedical research*. *Hoboken*, NJ, USA: John Wiley and Sons.

Ekins, S., Williams, A. J., Krasowski, M. D., et al. (2011c). In silico repositioning of approved drugs for rare and neglected diseases. *Drug Disc. Today 16*, 298-310.

Fourches, D., Barnes, J. C., Day, N. C., et al. (2010). Cheminformatics analysis of assertions mined from literature that describe drug-induced liver injury in different species. *Chem. Res. Toxicol. 23*, 171-183.

Greene, N., Fisk, L., Naven, R. T., et al. (2010). Developing structure-activity relationships for the prediction of hepatotoxicity. *Chem. Res. Toxicol. 23*, 1215-1222.

Gupta, R. R., Gifford, E. M., Liston, T., et al. (2010). Using open source computational tools for predicting human metabolic stability and additional ADME/TOX properties. *Drug Metab. Dispos. 38*, 2083-2090.

Hann, M., Hudson, B., Lewell, X., et al. (1999). Strategic pooling of compounds for high-throughput screening. *J. Chem. Inf. Comput. Sci. 39*, 897-902.

Hansen, K., Rathke, F., Schroeter, T., et al. (2009). Bias-correction of regression models: a case study on hERG inhibition. *J. Chem. Inf. Model 49*, 1486-1496.

Helma, C. (2005). *Predictive toxicology*. Boca Raton, FL, USA: Taylor and Francis.

Hohman, M., Gregory, K., Chibale, K., et al. (2009). Novel web-based tools combining chemistry informatics, biology and social networks for drug discovery. *Drug Disc. Today 14*, 261-270.

Hughes, J. D., Blagg, J., Price, D. A., et al. (2008). Physiochemical drug properties associated with in vivo toxicological outcomes. *Bioorg. Med. Chem. Lett. 18*, 4872-4875.

Huth, J. R., Mendoza, R., Olejniczak, E. T., et al. (2005). ALARM NMR: a rapid and robust experimental method to detect reactive false positives in biochemical screens. *J. Am. Chem. Soc. 127*, 217-224.

Huth, J. R., Song, D., Mendoza, R. R., et al. (2007). Toxicological evaluation of thiol-reactive compounds identified using a la assay to detect reactive molecules by nuclear magnetic resonance. *Chem. Res. Toxicol. 20*, 1752-1759.

Jolivette, L. J. and Ekins, S. (2007). Methods for predicting human drug metabolism. *Adv. Clin. Chem. 43*, 131-176.

Jones, D. R., Ekins, S., Li, L., et al. (2007). Computational approaches that predict metabolic intermediate complex formation with CYP3A4 (+b5). *Drug Metab. Dispos. 35*, 1466-1475.

Klon, A. E., Lowrie, J. F., and Diller, D. J. (2006). Improved naive Bayesian modeling of numerical data for absorption, distribution, metabolism and excretion (ADME) property prediction. *J. Chem. Inf. Model 46*, 1945-1956.

Kortagere, S., Chekmarev, D. S., Welsh, W. J., et al. (2008). New predictive models for blood brain barrier permeability of drug-like molecules. *Pharm. Res. 25*, 1836-1845.

Kortagere, S., Krasowski, M. D., Reschly, E. J., et al. (2010). Evaluation of computational docking to identify PXR agonists in the ToxCastTM database. *Env. Health Perspect. 118*, 1412-1417.

Langdon, S. R., Mulgrew, J., Paolini, G. V., et al. (2010). Predicting cytotoxicity from heterogeneous data sources with Bayesian learning. *J. Cheminform. 2*, 11.

Lipinski, C. A., Lombardo, F., Dominy, B. W., et al. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev. 23*, 3-25.

Metz, J. T., Huth, J. R., and Hajduk, P. J. (2007). Enhancement of chemical rules for predicting compound reactivity towards protein thiol groups. *J. Comput. Aided Mol. Des. 21*, 139-144.

Mishra, N. K., Agarwal, S., and Raghava, G. P. (2010). Prediction of cytochrome P450 isoform responsible for metabolizing a drug molecule. *BMC Pharmacol. 10*, 8.

Munos, B. (2009). Lessons from 60 years of pharmaceutical innovation. *Nat. Rev. Drug Discov. 8*, 959-968.

Obrezanova, O., Csanyi, G., Gola, J. M., et al. (2007). Gaussian processes: a method for automatic QSAR modeling of ADME properties. *J. Chem. Inf. Model 47*, 1847-1857.

Oh, W. S., Kim, D. N., Jung, J., et al. (2008). New combined model for the prediction of regioselectivity in cytochrome P450/3A4 mediated metabolism. *J. Chem. Inf. Model 48*, 591-601.

Pan, Y., Li, L., Kim, G., et al. (2010). Identification and validation of novel human pregnane X receptor activators amongst prescribed drugs via ligand-based virtual screening. *Drug Metab. Dispos. 39*, 337-344.

Pearce, B. C., Sofia, M. J., Good, A. C., et al. (2006). An empirical process for the design of high-throughput screening deck filters. *J. Chem. Inf. Model 46*, 1060-1068.

Rydberg, P., Gloriam, D. E., and Olsen, L. (2010). The SMARTCyp cytochrome P450 metabolism prediction server. *Bioinformatics 26*, 2988-2989.

Shamovsky, I., Connolly, S., David, L., et al. (2008). Overcoming undesirable HERG potency of chemokine receptor antagonists using baseline lipophilicity relationships. *J. Med. Chem. 51*, 1162-1178.

Spjuth, O., Willighagen, E. L., Guha, R., et al. (2010). Towards interoperable and reproducible QSAR analyses: exchange of datasets. *J. Cheminform. 2*, 5.

Steinbeck, C., Hoppe, C., Kuhn, S., et al. (2006). Recent developments of the chemistry development kit (CDK) – an open-source java library for chemo- and bioinformatics. *Curr. Pharm. Des. 12*, 2111-2120.

Voutchkova, A. M.,Osimitz, T. G., and Anastas, P. T. (2010). Toward a comprehensive molecular design framework for reduced hazard. *Chem. Rev. 110*, 5845-5882.

Walters, W. P. and Murcko, M. A. (2002). Prediction of 'drug-likeness'. *Adv. Drug Del. Rev. 54*, 255-271.

Williams, A. J., Tkachenko, V., Lipinski, C., et al. (2009). Free online resources enabling crowdsourced drug discovery. *Drug Discov. World 10*, 33-38.

Williams, A. J. and Ekins, S. (2011). A quality alert and call for improved curation of public chemistry databases. *Drug Disc. Today 16*, 747-750.

Williams, A. J., Ekins, S., Clark, A. M., et al. (2011a). Mobi-

lizing chemistry in the world of drug discovery. *Drug Disc. Today*, in press.

Williams, A. J., Ekins, S., Spjuth, O., and E. L. Willighagen (2011b). Accessing, using and creating chemical property databases for computational toxicology modeling. In B. Reisfeld and A. Mayeno (eds), *Methods in Molecular Biology: Computational Toxicology*. New York, NY, USA: Humana Press.

Xia, X. Y., Maliski, E. G., Gallant, P., et al. (2004). Classification of kinase inhibitors using a Bayesian model. *J. Med. Chem. 47*, 4463-4470.

Xu, J. J., Henstock, P. V., Dunn, M. C., et al. (2008). Cellular imaging predictions of clinical drug-induced liver injury. *Toxicol. Sci. 105*, 97-105.

Zaretzki, J., Bergeron, C., Rydberg, P., et al. (2011). RS-Predictor: A new tool for predicting sites of cytochrome P450-mediated metabolism applied to CYP 3A4. *J. Chem. Inf. Model. 51*, 1667-1689.

Zhang, L., Zhu, H., Oprea, T. I., et al. (2008). QSAR modeling of the blood-brain barrier permeability for diverse organic compounds. *Pharm. Res. 25*, 1902-1914.

Zheng, X., Ekins, S., Rauffman, J.-P., et al. (2009). Computational models for drug inhibition of the Human Apical Sodium-dependent Bile Acid Transporter. *Mol. Pharm. 6*, 1591-1603.

Zheng, X., Ekins, S., Raufman, J. P., et al. (2009). Computational models for drug inhibition of the human apical sodium-dependent bile acid transporter. *Mol. Pharm. 6*, 1591-1603.

Zientek, M., Stoner, C., Ayscue, R., et al. (2010). Integrated in silico-in vitro strategy for addressing cytochrome P450 3A4 time-dependent inhibition. *Chem. Res. Toxicol. 23*, 664-676.

## Acknowledgments

SE gratefully acknowledges the many collaborators involved in the work cited herein.

## Correspondence to

Sean Ekins, PhD
Collaborations in Chemistry
5616 Hilltop Needmore Road
Fuquay-Varina
NC 27526
USA