ADMET Property Prediction: The State of the Art and Current Challenges

Joelle Gola*, Olga Obrezanova, Ed Champness and Matthew Segall

Inpharmatica Ltd., 127 Cambridge Science Park, Milton Road, Cambridge CB4 0GD, UK, E-mail: j.gola@inpharmatica.co.uk

Keywords: ADMET, Computational chemistry, Drug design, Structure–activity relationships

Received: July 21, 2006; Accepted: August 21, 2006

DOI: 10.1002/qsar.200610093

Abstract

In this article, we review recent developments in the prediction of Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) properties by Quantitative Structure–Activity Relationships (QSAR). We consider advances in statistical modelling techniques, molecular descriptors and the sets of data used for model building and changes in the way in which predictive ADMET models are being applied in drug discovery. We also discuss the current challenges that remain to be addressed. While there has been progress in the adoption of non-linear modelling techniques such as Support Vector Machines (SVM) and Bayesian Neural Networks (BNNs), the full advantages of these 'machine learning' techniques cannot be realised without further developments in molecular descriptors and availability of large, high-quality datasets. The largest pharmaceutical companies have developed large in-house databases containing consistently measured compound properties. However, these data are not yet available in the public domain and many models are still based on small 'historical' datasets taken from the literature. Probably, the largest remaining challenge is the full integration of predictive ADMET modelling in the drug discovery process. Until *in silico* models are applied to make effective decisions in a multi-parameter optimisation process, the full value they could bring will not be realised.

1 Introduction

The importance of optimising Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) properties of potential drug molecules is now widely recognised [1]. A potent molecule is not sufficient to achieve an efficacious drug. For efficacy, a drug must reach the site of the target in the body at sufficient concentration with a specific time to achieve the required pharmacological effect. Furthermore, the drug must be safe at a therapeutic concentration, exhibiting minimal side effects. Therefore, it is the *balance* of potency, selectivity and ADMET properties that will ultimately determine the success of a potential drug molecule.

Historically, ADMET properties were often considered relatively late in the drug discovery process, in late lead optimisation or even preclinical development. The result of this was a high attrition rate in the later stages of R&D, where the costs increase dramatically. This in turn contributed to the ever-increasing average cost of developing a marketed drug, now estimated by some as US$ 0.8–1.7 billion [2]. The reason for the delay in consideration of ADMET properties was the high cost and low throughput of measurement, which relied predominantly on *in vivo* experiments.

Attitudes began to change in light of notable successes of drugs such as Fluconazole (1991) [3], which achieved an excellent balance of potency with ADMET properties to achieve market dominance in their therapeutic area. The chemistries from which these drugs were derived were optimised for improved ADMET properties early in the drug discovery process. This illustrated the advantage of early optimisation of ADMET properties and motivated the development of *in vitro* technologies that could be used to measure key properties at higher throughput and lower cost than conventional *in vivo* experiments [4]. The effects of these could be seen by the late 1990s, when the reported contribution of ADME to attrition in clinical development fell, although failures due to toxicity remained high [5].

A major hurdle to performing toxicity studies earlier in the discovery process is that the causes and consequences of toxicity are various and variable compared to ADME properties. Toxicity is frequently a multi-factorial event with a plethora of possible responses from lachrymation to cancer. The toxic response observed may be the end result of a whole series of chemical and biochemical events that can only occur in an intact animal and may be dose- and time-dependent. Performing whole animal studies is expensive and is not practicable for a large number of compounds at an early stage of investigation. Furthermore, these tests are often designed to look for specific effects that have already been reported for related compounds with known toxicophores.

Clearly, the earlier it is possible to consider the ADMET properties of molecules, the larger the potential impact on the productivity of drug discovery. Ideally, one would wish to examine as wide as possible a range of chemistries, to identify those that are most likely to have appropriate ADMET properties as well as potency. This is most efficient if the ADMET properties can be predicted from the chemical structure, so that large numbers of compounds could be considered at low cost prior to choosing a synthetic strategy. The advent of *in vitro* methods for measuring specific ADMET properties led to an increase in the availability of data on a wide range of compounds, making it possible to investigate the rules by which the chemical structure of a molecule determine its ADMET properties, *i.e.* to build predictive models of ADMET properties. Thus, the late 1990s and early 2000s saw a dramatic increase in the development of predictive, or *in silico*, models of ADMET properties [6–8].

In silico modelling of ADMET properties can be broadly divided into three categories: molecular modelling, physiologically based pharmacokinetic modelling and statistical modelling [9]. Molecular modelling approaches include quantum and classical mechanical methods, homology modelling and pharmacophore models and can be used where the underlying molecular mechanism of a property is understood. Physiologically based pharmacokinetic modelling integrates several factors responsible for ADME processes in one model and attempts to simulate the pharmacokinetics of a drug in the whole organism. Statistical modelling is applied when the molecular mechanism of an ADMET property is not clear or cannot be efficiently modelled at the molecular level and largely uses Quantitative Structure–Activity Relationship (QSAR) approaches. In this paper, we will concentrate on QSAR modelling techniques.

This article will review the progress made over the past 5 years in the development of predictive ADMET models and the current state-of-the-art. We will discuss trends in three fundamental aspects of predictive model development

(1) Data: The availability of sets of molecules for which ADMET properties have been determined.
(2) Descriptors: Methods for characterisation of molecule structures that capture the biological or chemical mechanisms determining a molecule's properties.
(3) Modelling techniques: Statistical techniques that can identify the key descriptors and their quantitative relationship with the property being modelled.

Predictive ADMET modelling has moved from being predominantly of 'academic' interest to a position of increased practical importance in pharmaceutical R&D. In particular, the application of predictive ADMET modelling is increasingly undertaken by non-computational scientists and the tools for deployment of models to the desktop have advanced. The increased availability of data, both *in silico* and *in vitro* on a wide range of properties has created an interest in 'multi-parameter optimisation' [1], whereby the explicit goal is to achieve a *balance* of multiple properties simultaneously. We will discuss this evolution of modelling tools from simple interfaces to run models and collect the resulting data, into powerful tools for analysis and decision-support.

The following sections discuss the trends in data, descriptors and modelling techniques for predictive ADMET models and the application of models in drug discovery. Finally, we will draw these threads together to reach conclusions and contemplate the potential future developments in this field.

2 Data

The development of accurate and applicable predictive models for ADMET properties is highly dependent upon the data on which the models are built. A dataset would be characterised as high quality if the measurements are reliable, consistent and reproducible with low experimental errors. Unfortunately, the availability of experimental data for ADMET properties is often limited in quantity and quality. This is particularly true of *in vivo* properties obtained directly from humans, where data is typically only available for compounds in clinical development. Published data from *in vitro* experiments also show a high degree of variability. Reported inhibition data for P450 isoform CYP2D6 for ketoconazole have a 17-fold variation [9]. Ideally, data used to build ADMET models must be carefully quality-controlled and should be generated under a single experimental protocol for any one model [9]. With the increased application of ADMET evaluation during the course of drug discovery, one would expect an abundance of experimental ADMET data available to model. However, the reality is often different with a paucity of high-quality ADMET data available for model building, except in the largest companies with comprehensive in-house databases. Where in-house databases are not available, modellers must resort to peer-reviewed literature, internet resources or commercially available databases. Compilations of data for human oral absorption and bioavailability, plasma protein binding and blood-brain permeation have been extensively published and used to formulate predictive models. These 'historical' datasets were put together from various sources [10] and while often carefully quality controlled, they are subject to variability due to experimental conditions and inter-laboratory errors. Models built on literature data may hold only limited value for true predictive use within pharmaceutical companies [11].

Another important dataset characteristic is its chemical diversity or coverage [11]. Global models cover as wide a range of chemical space as possible. However, as a drug discovery project progresses, the chemistry under consideration often focuses on a small number of chemical series in which the molecules are structurally similar. Global models may lack the resolution required to distinguish between molecules with subtle differences.

Predictive models based on proprietary datasets have emerged during the past 5 years. The need to know the ADMET properties of drug candidates has propelled the development of numerous high-throughput screening methods, which have resulted in a sufficient quantity of data for the analysis of the relationship between structural properties and ADMET properties. Furthermore, efficient data management has also significantly eased the access to data for model building. For instance, knowledge-based analyses of proprietary oral bioavailability data in rat, a measurement frequently used as a surrogate for human oral bioavailability, have been published; Veber *et al.* based their studies on a dataset of 1100 compounds from diverse GSK projects [12], 434 Pharmacia compounds were analysed by Lu *et al.* [13] and Martin published a study of 553 compounds from an Abbot in-house project [14]. In addition, O'Brien and de Groot published the results of *in silico* modelling of affinities for the human ether-a-go-go-related gene (hERG) ion channel and Cytochrome P450 2D6 (CYP2D6) drug metabolising enzyme of a very large dataset of 58963 and 2410 Pfizer compounds, respectively [15]. Seierstad and Agrafiotis recently reported a QSAR model of hERG binding using a diverse training set of 400 compounds, tested in a single assay under the same experimental conditions [16].

Questions have recently arisen regarding the meaning and relevance of experimental data used in model building for the desired ADMET endpoints [17]. For instance, the majority of *in silico* models for drug brain penetration use log (BB) as the index of Blood-Brain Barrier (BBB) permeability, where BB is equal to the brain:blood drug concentration ratio at some defined time point. There are log (BB) data for approximately 150 compounds in the public domain [18]. However, the value of the log (BB) parameter is now uncertain. The main concern expressed over these data is that the brain concentration is the sum of the bound and free drug concentration, and it has been suggested that future BBB permeation models be based not on log (BB), but on log PS; where log PS is the permeability-surface area product that represents a true measure of the rate of transfer of the compound from the blood to the brain. Two models predicting log PS values have been published, both based on very small datasets; 50 and 23 compounds, respectively [18].

Over the past few years, new priorities have been identified for modelling ADMET properties. For example, the cardiotoxicity of many chemical entities has been linked to their potent inhibition of the hERG channel; an effect that can lead to prolongation of the QT interval of the heart beat and in the worst case, death [19]. The link between hERG inhibition and QT prolongation has become another important component of preclinical safety evaluation and compound, testing as blockade of this channel is also used as a preliminary assessment of proarrythmic liability. Not surprisingly, the urgency of the matter has led to the development of a variety of *in vitro* and *in vivo* tests to assess the hERG inhibition of new chemical entities and inhibition data used to build a number of *in silico* models to predict this potential toxicity (Table 1). However, hERG affinity measurements are highly dependent on experimental measurement conditions such as gating voltage and temperature and, hence, prone to large inter-laboratory discrepancies. Variations in literature data for a single compound can be quite significant and, consequently, conclusions regarding hERG-drug binding interactions based on a model built with such data should be made with caution.

Whilst datasets may be available in-house or in the literature, to build predictive models for the major ADMET hurdles, the properties covered are far from comprehensive. A few years ago, it was expected that advances in technology and better understanding of the phenomena would lead to an increase in the range of properties being modelled [19]. However, this has yet to happen in practice.

For example, insufficient data are currently available to predict drug transport by proteins expressed in the major clearance organs, the liver and kidney. Furthermore, models to predict the major routes of elimination for a compound are not widespread and building successful *in silico* models for such properties will depend upon the implementation and conduct of suitable screens and subsequent dataset generation [19].

3 Descriptors

Using an appropriate set of descriptors, the most important features of the mechanism giving rise to a measured property are captured and a mathematical correlation with those observations can be derived. It is important that the predictions can be interpreted to provide guidance on the effects of chemical modifications on the predicted property. For this reason, wherever possible chemically interpretable descriptors are used, despite some potential penalty in model accuracy.

There are currently over 3000 molecular descriptors that can be used in *in silico* modelling [20]. The form of the descriptors employed by a model often depends on the endpoint to be predicted. Some descriptors contain information about the conformation of a compound, which is im-

Table 1. Recent publications of in silico modelling of hERG affinity.

Ref.	Dataset[a]	Descriptors	Modelling technique[b]	Statistic on test set
[40]	28/4	Homology model and 3-D (CoMSIA)	CoMSIA	R^2_{CV}[d] $=0.571$
[41]	31/6	Pharmacophore and 3-D (CoMFA)	PLS	$R^2=0.744$
[42]	15/22	3-D Catalyst	Catalyst	$R^2=0.83$
[43]	322/16	3-D pharmacophore (GRIND)	PLS	$R^2=0.94$
[43]	518/26	3-D pharmacophore (GRIND)	PLS	$R^2=0.90$
[44]	244/38+57	1-D, 2-D and 3-D	ANN	93% of negatives 71% of positives
[45]	332/83	2-D and 3-D	DT	85% of positives 71% of negatives
[16]	439/40	1-D and 2-D	ANN	$R^2=0.52$
[46]	55/13	2-D	PLS	$R^2=0.81$
[24]	71/19+20	2-D	SVR	$R^2=0.848$
[15]	46967/11996	1-D and 2-D	ANN/BS[c]	87% of positives 86% of negatives

[a] Number of compounds in training/number of compounds in test set.

[b] PLS, partial least square; ANN, neural network; DT, decision tree; SVR, system vector regression; BS, Bayesian statistics.

[c] Consensus modelling using models built on neural network and Bayesian statistics.

[d] R^2_{CV}, cross-correlation coefficient.

portant if predicting specific interactions with a receptor. Other descriptors do not contain any geometrical information and are non-specific in nature. These are often used for the prediction of physicochemical properties but can also provide more general descriptions of the underlying properties required for receptor binding.

Descriptors have been designed to take into account size, lipophilicity and electronic effects as well as the hydrogen-bond propensity of molecules. Descriptors can be defined by the dimensionality of the structural representation [21]. Two-Dimensional (2-D) descriptors are calculated from the molecular graph alone, and they do not use information related to the Three-Dimensional (3-D) conformation of model compounds. 2-D descriptors are typically simple descriptors that count atoms or functional groups that may be relevant to specific mechanisms, for example, hydrogen-bond donors and acceptors, acidic and basic functionalities. 3-D descriptors are more complex descriptors that try to explain the variance in biological activity by capturing the effects relevant to compound/protein interaction.

3-D descriptors are often more chemically intuitive than 2-D descriptors, as they capture, more efficiently, the specific interactions occurring between a receptor and a compound. A medicinal chemist can use this information to modify chemical structure to try and reach the desired affinity. For instance, pharmacophore models and 3-D models developed on CoMFA [22] and CoMSIA [23] to predict interaction between drugs and the hERG ion channel have highlighted important chemical features which can lead to high hERG affinity. Hence, the presence of a nitrogen atom positively charged at pH 7.4, surrounded by a bulky hydrophobic moiety, has been shown to increase hERG affinity [24]. However, despite providing important insights for the protein/compound interactions, 3-D models have had limited application due to the lack of a hERG crystal structure, effective techniques for the sampling of active conformations and the need for 3-D molecular alignment of diverse structures [21]. In addition, the speed of calculation for 3-D descriptors can limit their use in 'real time' compound assessment and design, and for large datasets 2-D QSAR models are often the preferred option [21].

The majority of *in silico* models for ADMET properties have been developed using descriptors that have been available for some time. Indeed, there have not been significant changes in descriptor calculations for the past 5 years. Some of the commonly used descriptors, such as solvation descriptors [25], the E-state descriptors [26] and BCUT [27] descriptors, were first implemented in the 1990s. Even VolSurf, a relatively new approach developed by Cruciani and applied to build *in silico* models for a large number of ADMET properties [28], was introduced some 6 years ago.

While the available descriptors capture general trends in properties, often large numbers of descriptors are necessary to obtain good correlations across diverse chemistry. This, in turn, necessitates the use of large datasets to train and test these models, and these are often unavailable (see Sec. 2, Data). Therefore, the development of new descriptors that relate more closely to the mechanisms of interaction leading to observed ADMET properties and correlate more strongly, is a challenge that remains to be addressed.

4 Modelling Techniques

What are the requirements for a good computational tool, which would be useful and effective in ADMET modelling? These are; a high prediction accuracy, ability to deal with multiple mechanisms of action, effective modelling of non-linear relationships, ability to handle multi-dimensional data and to ignore irrelevant descriptors, computational efficiency and robustness to overtraining. An ability to produce an easily interpretable model is a very desirable feature for a computational technique, but usually conflicts with some of the above requirements. One of the recent demands on a modelling technique is the possibility of estimating confidence in a prediction which can be used in decision making when selecting the best compounds.

Since the inception of the ADMET modelling field, Partial Least Squares (PLS), Multiple Linear Regression (MLR), Artificial Neural Networks (ANNs) and Decision Trees (DT) have been the most common methods for modelling of ADMET properties. They still remain the most popular even though they have a number of weaknesses. PLS and MLR are linear techniques which cannot adequately model non-linear relationships and multiple mechanisms of action. ANNs can handle non-linear problems, but are prone to overtraining, have problems with network optimisation and model selection and are not efficient in dealing with high-dimensional data without preselection of descriptors.

DTs, also called recursive partitioning, is a very popular method in QSAR modelling of ADMET properties. It handles datasets containing a high number of variables, has an embedded ability to select important descriptors and is suitable for multiple mechanisms of action. However, DTs often have quite low predictive ability, although this drawback may be overcome by the tree ensemble techniques such as Boosting [29, 30], Bagging (see papers cited in [29]) and Random Forest [29], which give higher prediction accuracy than a single tree. In recent years these powerful methods have been increasingly used for ADMET modelling.

A more recent entrant in the field is the Support Vector Machines (SVM) technique which is widely recognised for its remarkable generalisation ability. In the last 5 years, it has become very popular in ADMET modelling as a non-linear classification technique, but it is also applicable to regression problems. An extensive review of applications of this technique to ADMET modelling is given in [31].

Bayesian Neural Networks (BNNs) represent a special type of neural net which overcome the problems of conventional ANNs described above. BNNs are based on a probabilistic interpretation of network training. Network weights are found by Bayesian inference that gives an objective solution to the problem of conventional network optimisation. This approach also provides the model predictions as probability distributions and therefore permits evaluation of the confidence in prediction [32]. BNNs can be successfully used together with an Automatic Relevance Determination (ARD) procedure to select relevant descriptors and develop an optimal model [32, 33]. Another new promising method based on a Bayesian approach is the Gaussian Processes technique which has been recently applied to QSAR problems [34]. Gaussian processes have been shown to be equivalent to an ANN with a single hidden layer containing an infinite number of nodes [35].

An unsupervised neural network algorithm called Kohonen Self-organising Maps has been successfully used to build ADMET classification models [36]. This technique is particularly efficient for dealing with properties having multiple mechanisms of action and can be used as a descriptor selection technique.

Many of the modelling techniques described above, such as SVM and ANN, have trouble in handling high-dimensional data, *i.e.* the presence of many descriptors. Therefore, methods for variable selection have to be employed with the aim of selecting a subset of relevant descriptors for building a model. There are descriptor selection techniques that perform variable preselection prior to model building, for example filtering descriptors with low linear correlation with the target property or using a small number of principal components of the descriptor matrix to build a model [16]. Some variable selection methods are embedded within the modelling algorithm; examples of such algorithms are DTs and BNNs with ARD. However, the most commonly used descriptor selection techniques in ADMET prediction are those that are employed iteratively with the model building algorithm, such as forward selection [16], backward elimination, Genetic Algorithm (GA) [30, 37] and simulated annealing [16]. In the last few years, novel algorithms such as artificial ant colony [38] and particle swarms [31] have been applied to QSAR datasets and these methods could potentially be used to model ADMET properties. The iterative selection methods are more computationally demanding because the model has to be trained for each subset of descriptors considered.

The demand for fast model (re)building whenever new data becomes available combined with the use of a variety of modern modelling techniques gave rise to a trend to develop computational algorithms for automatic model generation [39]. The purpose of such algorithms is to save scientists' time, explore more modelling possibilities and make a process for QSAR model building accessible to non-experts. Automatic approaches to model building face some difficulties, principally a tendency to generate 'black box' models which are difficult to interpret, choosing the best model from a number of potentials and selection of compounds for training and test sets. In our experience, the latter represents the major difficulty and expert knowledge seems to be invaluable at this stage of model generation.

5 Application of ADMET Predictions

Whilst the development of descriptors and modelling techniques has increased our ability to develop models for individual ADMET properties, there has been a shift in focus with respect to the application of these models. A large proportion of drug discovery project teams now have access to ADMET models in some form, either through in-house computational teams or using third party software. However, with the availability of technology capable of generating numbers as fast as computers can process the mathematics, comes the problem of analysis.

Popular software packages have been available since many years' which enable the user to 'analyse' large sets of data. Multi-dimensional datasets can be displayed on 2-D or 3-D charts with additional dimensions made available using colours, shapes, sizes, *etc.* Although these make for excellent data representations in presentations and reports, there still remains the question as to whether these have truly allowed a rigorous analysis of the data. Ultimately, what the user is looking to achieve is the ability to make decisions on the basis of the properties they have predicted and measured. Even better, the user would hope to be able to gain some measure of confidence around any decision they make.

Traditionally, the major stumbling block for those advocating the use of *in silico* models has been the accuracy of the prediction. By definition, a model is only an estimation of an unknown value on the basis of previously collected knowledge, yet often it will be evaluated and criticised for not exactly predicting a value for a single molecule. Perhaps, what is often forgotten is that all the technology used in the drug development process, be it *in silico*, *in vitro* or *in vivo*, has some degree of uncertainty around it. Decisions have been made on the basis of *in vivo* and *in vitro* data for many years now, despite the known experimental error and in some cases poor correlation between the experimental system and human *in vivo* properties. With this in mind, is there any reason to believe *in silico* predictions cannot be used equally well for decision making? However, the problem still remains to be one of multi-dimensional optimisation in which there are varying levels of uncertainty around different data sources and differences in the relative importance of each property.

With the gradual acceptance of *in silico* technology, the requirement for an ability to analyse data using a new set of skills has arisen. The same people who might originally have been looking at analysing the results of a single *in vivo* study are now being asked to analyse and make decisions about datasets containing thousands of datapoints, all with varying levels of uncertainty and relevance.

In terms of a conceptual model for applying *in silico* technology, a very simple model exists which uses *in silico* technology alongside *in vitro* and *in vivo* techniques (Figure 1). The underlying premise is that at the start of a drug discovery project there is an understanding of the techniques available and the relevance of the data generated by each. There must also be known measures for the uncertainties around the data generated by different techniques. The first phase of the project can be purely theoretical, a process of molecule design using only *in silico* techniques. An appropriate subset of all the molecules initially considered can then be synthesised and evaluated against a number of appropriate screens, such as potency. Assuming there are 'hits' at this stage, a further subset of molecules could be progressed for further *in vitro* and then finally *in vivo* screening. At every stage *two* processes must take place. Firstly, a selection process involving multi-dimensional optimisation which takes into account uncertainties and relevance; this yields the subset of molecules for progression. One example of a method that could be employed to perform the multi-dimensional optimisation would be probabilistic scoring [1], however, what is important is that the same, rigorous and comprehensive method is used during each selection process. Secondly, and equally important, a feedback loop of all data is generated into the *in silico* process, either for model evolution or for model generation. The feedback loop guarantees that at any stage, if there is a need for further molecule design, then everything already known about those molecules already tested is taken into consideration. Clearly such a process is difficult to introduce once a project is in progress, yet this is almost always the way *in silico* technology gets introduced in drug discovery.

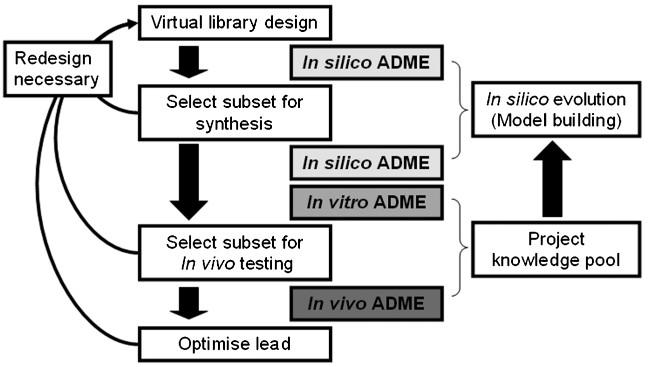
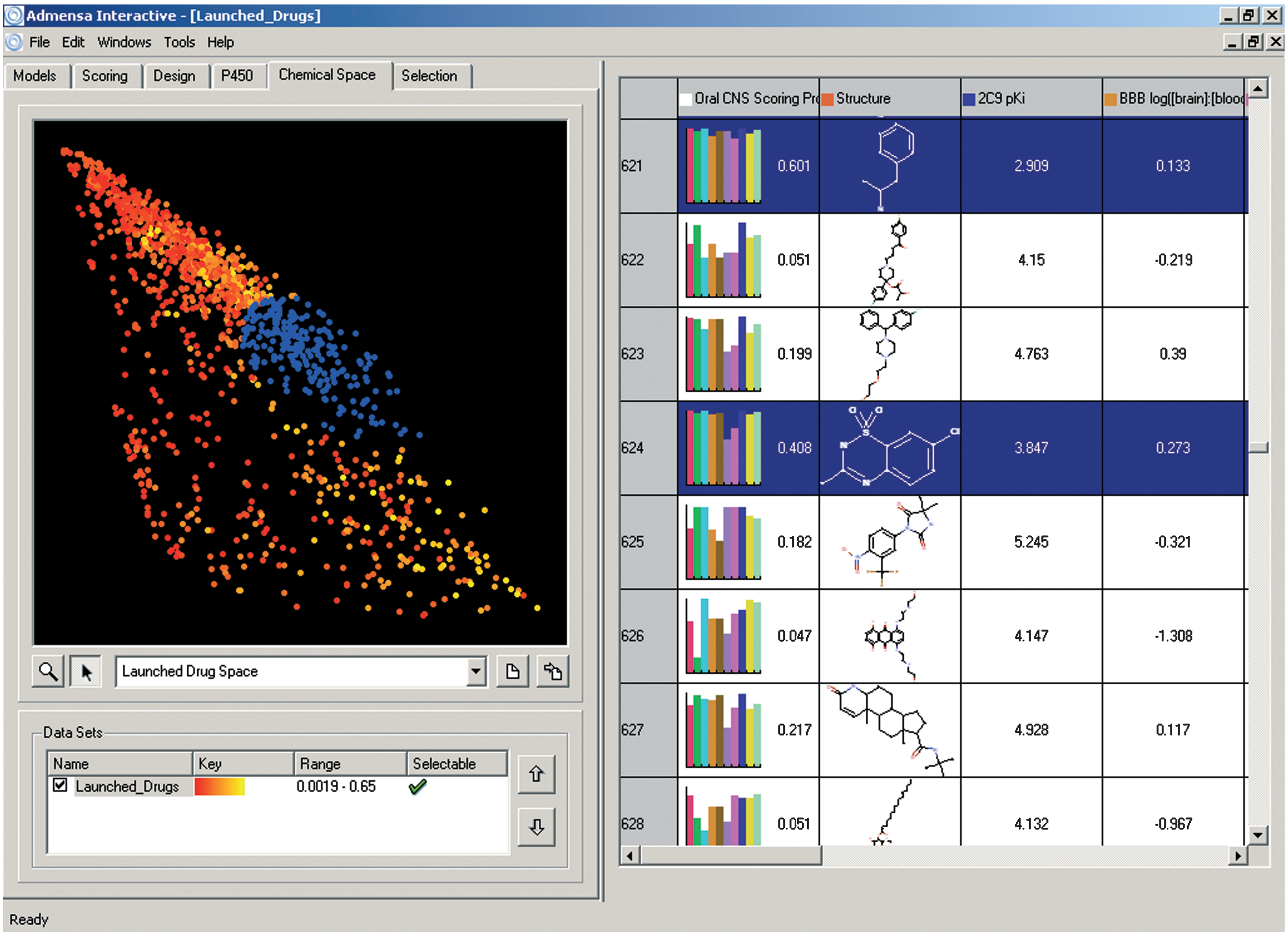


Figure 1. Conceptual model for application of *in silico* technology within the drug discovery process. See text for discussion.

As *in silico* technology creeps into the drug discovery process, it is often viewed as something 'nice to have', that can be used alongside the existing technologies and contribute to decision making when there are no other alternatives. It is clear to see that in order to implement a drug discovery process that fully integrates *in silico* technology, such as that described in the conceptual model, there can be no half-way house. Without implementing a process in which *in silico* technology is used from the very beginning, it will never be possible to realise all the benefits and insights that such technology is capable of delivering. Indeed, it could be argued that without the *in silico* technology contributing to the decision making process it is likely that maximum benefit is not being derived from the *in vitro* and *in vivo* technologies either. This, therefore, presents what is probably the greatest barrier to be overcome by *in silico* technology if it is to become established in the


Admensa Interactive - [Launched_Drugs]
File Edit Windows Tools Help
Models Scoring Design P450 Chemical Space Selection
Launched Drug Space
Data Sets
Ready

	Oral CNS Scoring Pr	Structure	2C9 pKi	BBB log([brain]:[blood
621	0.601		2.909	0.133
622	0.051		4.15	-0.219
623	0.199		4.763	0.39
624	0.408		3.847	0.273
625	0.182		5.245	-0.321
626	0.047		4.147	-1.308
627	0.217		4.928	0.117
628	0.051		4.132	-0.967

Name	Key	Range	Selectable
☑ Launched_Drugs		0.0019 - 0.65	✓

Figure 2. Admensa Interactive^{TM}, an example of an intuitive desktop interface providing *in silico* models within a decision support framework to support drug discovery projects.

drug discovery process; namely that the drug discovery process must evolve to get the maximum value from it. Unfortunately, the pharmaceutical industry has too often experienced false promises that a change to the process may result in saving time and money, whether through applying scientific, technological or organisational changes. It is likely that *in silico* technology may need to become accepted as an adjunct to the current process in the short-term. Only then can the underlying architecture of the whole process be gradually transformed. This process will only work if it can be done without demanding wholesale change at any one time.

This presents an interesting challenge for those advocates and developers of *in silico* technology. Many lessons have been learned from the difficulty of introducing *in silico* technology that attempts to replace an older technology, such as using a computer model instead of an *in vitro* screen. Despite potential time and cost benefits, there has always been a lengthy process of convincing users of the old technology to take the chance and adopt what's new. However, there are many opportunities for *in silico* technology to be developed in less contentious fields. An example of this is decision making – this is something that has always taken place but with no one 'correct' approach. A consistent and rigorous framework for decision support can be of benefit to professionals at all levels in the hierarchy and it is arguable that there is not yet any existing technology to be replaced. One example of such an approach is illustrated in Figure 2 where multi-dimensional optimisation can be performed in a probabilistic scoring framework and used alongside chemical space and compound selection algorithms to allow prioritisation. By providing the means to get the best out of technologies that are already in place, it is much easier to help the drug discovery process evolve.

6 Conclusions

This review has described recent developments in the field of ADMET property predictions by QSAR models. It is clear that there have been developments in the technology for building predictive models, particularly in the sophistication of statistical modelling techniques and availability of large datasets in 'big pharma'. However, it is equally apparent that in order to obtain the full value from these, further developments will be necessary in the descriptors used to characterise molecules as inputs to predictive models and in the availability of data in the public domain. Furthermore, the demand to generate and update models more quickly, whenever new experimental data is available, poses an additional challenge in automating all aspects of the model generation process.

Probably the biggest challenge facing the field is the integration of *in silico* ADMET predictions into the drug discovery process. For this to occur, approaches supporting decision-making based on the simultaneous optimisation of multiple parameters in the face of experimental and statistical uncertainty must be made available to the key decision-makers in drug discovery projects. *In silico* predictions, no matter how accurate, have no value unless they are used to make decisions regarding the direction of a project.

Acknowledgements

The authors would like to thank the members of the Admensa group of Inpharmatica Ltd. for contributions to the ideas discussed in this paper.

References

[1] M. D. Segall, A. P. Beresford, J. M. R. Gola, D. Hawksley, M. H. Tarbit, *Expert Opin. Drug Metab. Toxicol.* **2006**, *2*, 325–337.
[2] J. Gilbert, P. Henske, A. Singh, *In Vivo, The Business and Medicine Report* **2006**, *21*, 73–82.
[3] K. Richardson, K. Cooper, M. S. Marriott, M. H. Tarbit, P. F. Troke, P. J. Whittle, *Rev. Infect. Dis.* **1990**, *12(Suppl 3)*, S267–S271.
[4] M. H. Tarbit, J. Berman, *Curr. Opin. Chem. Biol.* **1998**, *2*, 411–416.
[5] I. Kola, J. Landis, *Nat. Rev. Drug Discov.* **2004**, *3*, 711–715.
[6] S. Ekins, J. Rose, *J. Mol. Graph. Model.* **2002**, *20*, 305–309.
[7] S. Ekins, C. L. Waller, P. W. Swaan, G. Cruciani, S. A. Wrighton, J. H. Wikel, *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 251–272.
[8] H. van de Waterbeemd, E. Gifford, *Nat. Rev. Drug Discov.* **2003**, *2*, 192–204.
[9] M. Segall, A. P. Beresford, Virtual ADME-Tox: The Promise of Technology in Preclinical Development, in: C. Sansom (Ed.), *Enabling Technologies: Delivering the Future for Pharmaceutical R&D*, PJP Publications Ltd., London **2002**, pp. 93–110.
[10] G. Colmenarejo, *Curr. Comput. Aided Drug Design* **2005**, *1*, 365–376.
[11] A. M. Davis, R. J. Riley, *Curr. Opin. Chem. Biol.* **2004**, *8*, 378–386.
[12] D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward, K. D. Kopple. *J. Med. Chem.* **2002**, *45*, 2615–2623.
[13] J. J. Lu, K. Crimin, J. T. Goodwin, P. Crivori, C. Orrenius, L. Xing, P. J. Tandler, T. J. Vidmar, B. M. Amore, A. G. Wilson, P. F. Stouten, P. S. Burton, *J. Med. Chem.* **2004**, *47*, 6104–6107.
[14] Y. C. Martin, *J. Med. Chem.* **2005**, *48*, 3164–3170.
[15] S. E. O'Brien, M. J. de Groot, *J. Med. Chem.* **2005**, *48*, 1287–1291.
[16] M. Seierstad, D. K. Agrafiotis, *Chem. Biol. Drug Des.* **2006**, *67*, 284–296.
[17] X. Liu, M. Tu, R. S. Kelly, C. Chen, B. J. Smith, *Drug Metab. Dispos.* **2004**, *32*, 132–139.
[18] D. E. Clark, Computational Prediction of Blood-Brain Barrier Permeation, in: A. M. Doherty (Ed.), *Annual Reports in Medicinal Chemistry*, Elsevier Inc., San Diego **2005**, pp. 403–415.
[19] A. P. Beresford, M. Segall, M. H. Tarbit, *Curr. Opin. Drug Discov. Devel.* **2004**, *7*, 36–42.
[20] R. Todeschini, M. Lasagni, *J. Chemometrics* **1994**, *8*, 263–272.
[21] R. Lewis, P. Ertl, E. Jacoby, M. Tintelnot-Blomley, P. Gedeck, R. Wolf, M. C. Peitsch, *Chimia* **2005**, *59*, 545–549.
[22] R. D. Crammer, D. E. Patterson, J. D. Bunce, *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
[23] G. Klebe, U. Abraham, T. Mietzner, *J. Med. Chem.* **1994**, *37*, 4130–4146.
[24] M. Song, M. Clark, *J. Chem. Inf. Model.* **2006**, *46*, 392–400.
[25] M. H. Abraham, *Chem. Soc. Rev.* **1993**, *22*, 73–83.
[26] L. H. Hall, L. B. Kier, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1039–1045.
[27] F. R. Burden, D. A. Winkler, *J. Med. Chem.* **1999**, *42*, 3183–3187.
[28] G. Cruciani, *Molecular Interaction Fields*, WILEY-VCH, Weinheim **2006**.
[29] V. Svetnik, T. Wang, C. Tong, A. Liaw, R. P. Sheridan, Q. Song, *J. Chem. Inf. Model.* **2005**, *45*, 786–799.
[30] J. K. Wegner, H. Frohlich, A. Zell, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 931–939.
[31] S. J. Barrett, W. B. Langdon, Advances in the Application of Machine Learning Techniques in Drug Discovery, Design and Development, in: A. Tiwari, J. Knowles, E. Avineri, K. Dahal, R. Roy (Eds.), *Applications of Soft Computing: Recent Trends (Advances in Soft Computing)*, Springer, Berlin/Heidelberg **2006**.
[32] P. Bruneau, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1605–1616.
[33] F. R. Burden, M. G. Ford, D. C. Whitley, D. A. Winkler, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1423–1430.
[34] F. R. Burden, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 830–835.
[35] R. Neal, *Bayesian Learning for Neural Networks*, Springer-Verlag, New York **1996**.
[36] Y. H. Wang, Y. Li, S. L. Yang, L. Yang, *J. Chem. Inf. Model.* **2005**, *45*, 750–757.
[37] A. Yasri, D. Hartsough, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1218–1227.
[38] S. Izrailev, D. K. Agrafiotis, *SAR QSAR Environ. Res.* **2002**, *13*, 417–423.
[39] J. Cartmell, S. Enoch, D. Krstajic, D. E. Leahy, *J. Comput. Aided Mol. Des.* **2005**, *19*, 821–833.
[40] R. A. Pearlstein, R. J. Vaz, J. Kang, X. L. Chen, M. Preobrazhenskaya, A. E. Shchekotikhin, A. M. Korolev, L. N. Lysenkova, O. V. Miroshnikova, J. Hendrix, D. Rampe, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1829–1835.
[41] A. Cavalli, E. Poluzzi, F. De Ponti, M. Recanatini, *J. Med. Chem.* **2002**, *45*, 3844–3853.
[42] S. Ekins, W. J. Crumb, R. D. Sarazan, J. H. Wikel, S. A. Wrighton, *J. Pharmacol. Exp. Ther.* **2001**, *301*, 427–434.

[43] G. Cianchetta, Y. Li, J. Kang, D. Rampe, A. Fravolini, G. Cruciani, R. J. Vaz, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3637 – 3642.

[44] O. Roche, G. Trube, J. Zuegge, P. Pflimlin, A. Alanine, G. Schneider, *Chembiochem* **2002**, *3*, 455 – 459.

[45] A. M. Aronov, B. B. Goldman, *Bioorg. Med. Chem.* **2004**, *12*, 2307 – 2315.

[46] G. M. Keseru, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2773 – 2775.