Journal of Computer-Aided Molecular Design, **16:** 381–401, 2002.
KLUWER/ESCOM

Towards a new age of virtual ADME/TOX and multidimensional drug discovery

Sean Ekins[1,*], Bruno Boulanger[2], Peter W. Swaan[3] and Maggie A.Z. Hupcey[1]
[1]*Concurrent Pharmaceuticals Inc, 502 West Office Center Drive, Fort Washington, PA 19034, USA;* [2]*Eli Lilly, 11 rue Granbompré, B1348 Mont-Saint-Guibert, Belgium;* [3]*Division of Pharmaceutics, Ohio State Biophysics Program, OSU Heart & Lung Institute Core Laboratory for Bioinformatics and Computational Biology, The Ohio State University, 500 West 12th Avenue, Columbus, OH 43210-1291, USA*

Summary

With the continual pressure to ensure follow-up molecules to billion dollar blockbuster drugs, there is a hurdle in profitability and growth for pharmaceutical companies in the next decades. With each success and failure we increasingly appreciate that a key to the success of synthesized molecules through the research and development process is the possession of drug-like properties. These properties include an adequate bioactivity as well as adequate solubility, an ability to cross critical membranes (intestinal and sometimes blood-brain barrier), reasonable metabolic stability and of course safety in humans. Dependent on the therapeutic area being investigated it might also be desirable to avoid certain enzymes or transporters to circumvent potential drug-drug interactions. It may also be important to limit the induction of these same proteins that can result in further toxicities. We have clearly moved the assessment of in vitro absorption, distribution, metabolism, excretion and toxicity (ADME/TOX) parameters much earlier in the discovery organization than a decade ago with the inclusion of higher throughput systems. We are also now faced with huge amounts of ADME/TOX data for each molecule that need interpretation and also provide a valuable resource for generating predictive computational models for future drug discovery. The present review aims to show what tools exist today for visualizing and modeling ADME/TOX data, what tools need to be developed, and how both the present and future tools are valuable for virtual filtering using ADME/TOX and bioactivity properties in parallel as a viable addition to present practices.

Introduction

'Virtual' and 'in silico' applied to ADME/TOX

It is important to first define and distinguish 'in silico' and 'virtual', which are often used interchangeably to describe the use of computational tools for drug discovery. 'Virtual' has become a frequently used word in the pharmaceutical industry over the last five years as the pharmaceutical industry has joined the e-business paradigm. There have been moves towards a 'virtual' company structure proposed as a way to obtain more resources internal and external to its business. The outsourcing of routine discovery research activities including combinatorial chemistry, high throughput screening, metabolism and toxicology studies, specialized legal and patent capabilities as well as clinical trials are common to all drug and small biotechnology companies to obtain these 'virtual' resources and expert capabilities as needed [1]. This type of 'virtual' strategy has the potential to increase the efficiency of drug commercialization (Lehman Brothers, London 1996*), however it should not be confused with 'in silico' drug discovery approaches encompassing the use of computational software to generate and evaluate molecules which are also 'virtual'.

In the future it is suggested the pharmaceutical industry will be forced to decide which leads to select and optimize with fewer resources due to the escalat-

*Author for correspondence: E-mail: ekinssean@yahoo.com, sekins@concurrentpharma.com

*This report was described in SCRIP World Pharmaceutical News, January 9, 1996

ing cost of drug discovery [2]. This is perhaps relevant now as the recognition in recent years of the need to understand attrition of molecules in early development has caused companies to focus on rationales to skew the odds in their favor so that the probability of success increases. If more molecules are successful earlier, they bolster the pipeline and impact the size of the company portfolio. Computational tools to aid us in this decision making process would ultimately make drug discovery more efficient.

Developing a drug is a complex and lengthy process and failure during the development can occur as a result of a combination of poor pharmacokinetics, animal toxicity, lack of efficacy and perhaps most importantly, adverse effects in humans [3]. These adverse reactions observed in humans are not only the most public and visible form of failure but also the most expensive when they occur in late stage clinical trials or after market launch. By this point, millions of dollars have been invested in the molecule. However, the majority of adverse effects are essentially related to measurable ADME/TOX properties that may be predictable. The future of drug discovery and development will rest on better ways to select and test molecules efficiently and earlier.

Already pharmaceutical companies have started using computational approaches for metabolism and toxicology. This began approximately a decade after the computational tools for assisting in target prediction and optimization came of age. In parallel there have been numerous studies that have generated rules and filters from large databases or internal company data for predicting solubility and absorption potential [4–6]. Similarly tools have been developed to discriminate drug-like and non-drug-like molecules [7–8], differentiate between leads and drugs [9], and describe common structural features of known drugs [10, 11]. The latest approaches in this area have most recently led to pharmacophore point filters [12].

These various applications of in silico approaches (Figure 1) are perhaps the correct definition of 'virtual drug discovery' as molecules do not need to physically exist in order to be tested. Virtual approaches therefore represent an effective and rapid means to filter and prioritize molecules based on desirable structural descriptors and predicted physicochemical properties such as solubility and their ability to dock into a model of the target proteins. Although it is too early to see an increase in the number of drugs approved to date using these computational techniques, the future impact of these computational chemistry approaches will require their multivariate use for predicting all properties necessary for success [13, 14] to avoid costly investment in problematic molecules [6]. We may reach a point where lead generation and lead optimization occur simultaneously through computational tools.

Converging in vitro and in silico tools

Combinatorial synthesis, high throughput screening (HTS) and computational approaches are widely thought to have emerged as a result of continual pressure on pharmaceutical companies to shift from the 'natural product' based discovery of previous decades and to accelerate the drug discovery process and decrease drug development costs. At the same time the importance of ADME/TOX has come to prominence. To many discovery chemists ADME/TOX is a rather mysterious field that for decades has been assessed by specialist colleagues in these disciplines. This is changing as a result of implementing ADME/TOX screening earlier in the discovery timeline. There have been numerous accessible reviews written recently which provide useful information on the state of ADME/TOX at present [14–17]. The past few years have seen researchers indicating that we can predict a molecule's probability of success in vivo from simple chemico-physical descriptors [4,6] that ultimately impact the absorption, metabolism and distribution. The development of in vitro assays in the past decade have also yielded data sets that have been used to build computational models to predict blood brain barrier penetration [18], solubility [4], absorption [19, 20], metabolism, drug-drug interactions, drug transport and toxicity [14].

With the relative explosion in generating this kind of in vitro data for larger numbers of molecules, the ADME/TOX scientists and chemists have at their disposal a seemingly large array of information requiring visualization and interpretation to guide the synthesis of further molecules. This has led to the recent urgency with which companies have been hiring computational chemists as well as making strategic alliances with companies either involved in visualization software generation (Spotfire, Leadscope, Toxscope), modeling software (Tripos, Accelrys etc), ADME/TOX models (Camitro, LION bioscience, Accelrys) or combinatorial and high throughput crystal structure generation for the enzymes involved in detoxification (Syrxx, Astex). The potential for learning and iterative molecule improvement using traditionally derived empirical data

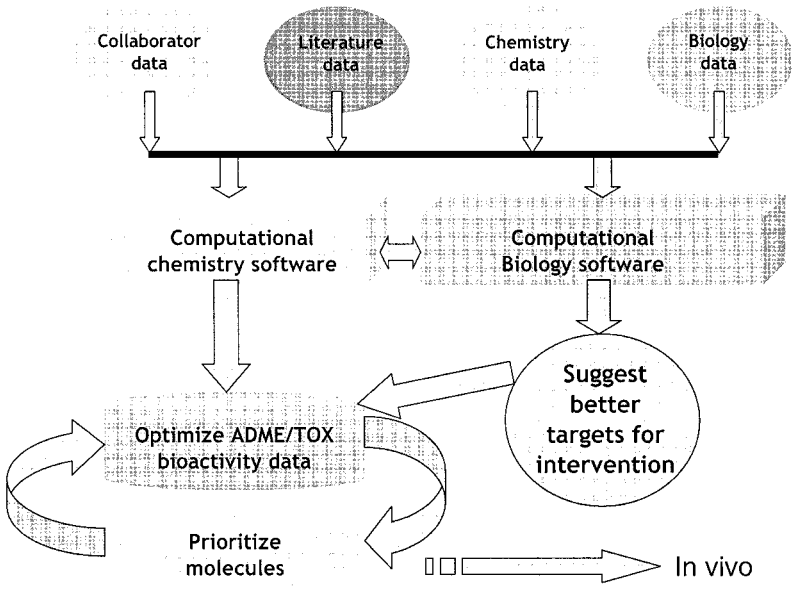


Figure 1. In silico centric drug

of course assumes that there is a valid in vitro-in vivo correlation. As most of the in vitro systems are quite artificial in terms of the experimental conditions employed, the use of this type of data for in silico model building involves quite a leap of faith, softened by the years of validation most in vitro models have undergone. However there are still significant caveats which may require assessment for each in vitro ADME/TOX system of which one should be aware. These models in total give us a basis then to go from in vivo to in vitro to in silico using the higher throughput ADME/TOX data to generate predictive computational models (See section 3).

Ultimately these in silico models will enable an exponential increase in the number of compounds that can be screened if we use large virtual libraries [21, 22] that can be scored for ADME/TOX and bioactivity parameters in parallel with the desired weighting factors required. Clearly there may be some overlaps between approaches that have been applied to assess drug likeness as previously described and how they may also be used for metabolism and toxicology end points. Less well studied to date appear to be the common functional 3D-configurations of molecules relating to ADME/TOX properties, which are ultimately more computationally intensive than the 1D and 2D descriptors that have been more widely used. In addition these 3D configurations add further complexity as they may need as relaxed tolerances as many of the drug targets [23] due either to the complexities or conformational flexibilities of the drug metabolizing enzymes [24] or the proteins that may be involved in a toxicological pathway. Fast pharmacophore identification tools [25] that can be used with thousands of molecules are already available and may be applicable to ADME/TOX data sets alongside other models in future (Figure 1).

ADME/TOX model building requirements and issues

One could classify the types of ADME/TOX in silico models that can be generated into primary and secondary groups (Table 1). This could be based on our present physicochemical understanding of an endpoint such as the apparent mechanism involved. Perhaps as important these models may be prioritized based on the critical importance or need for the models such as solubility, absorption, mutagenicity, bioavailability and metabolic stability which may be based on interrelated molecular properties and have a major influence on whether a molecule will be viable in vivo as a drug. These can be thought of as 'make or break' factors. The ability of a molecule to enter the central nervous system via the blood brain barrier may also be desirable or undesirable depending on the therapeutic target [18]. Further understanding the drug distribution by prediction of plasma protein binding

Table 1. Computational ADME/TOX models needed in drug discovery.

Primary Models
Solubility
Absorption (Caco-2)
Mutagenicity (AMES)
Bioavailability
Metabolic stability (multiple species)
Blood brain barrier penetration (presence or absence)
Cardiac toxicity (hERG)
Plasma protein binding
Secondary Models
Transport (uptake and efflux)
General toxicity
Hepatotoxicity (PXR, CAR)
Nephrotoxicity
Immunogenicity
Neurotoxicity (receptor binding)
Drug–drug Interactions (cytochrome P450)

is also important as this ultimately affects the circulating free concentration of the drug available to reach the target. Lipophillicity is thought to be a key determinant of this property [26,27]. The final primary model in this list is cardiac toxicity which is clearly an unwanted risk regardless of the ultimate therapeutic target in question. The recent withdrawal from the market of terfenadine, mibefradil, grepafloxacin and cisapride displaying this undesirable characteristic has highlighted the importance of evaluating cardiac toxicity earlier in the drug discovery process. For example, two quantitative structure activity relationships for potassium channel (hERG) inhibition have been described so far to date using drugs with inhibition constants [28,29].

Depending on the therapeutic target in question and the duration of the therapy other more secondary in silico models may be necessary (Table 1). In particular cytochrome P450 mediated drug-drug interaction models (see Section 3.3) could be applicable if the molecule was to be co-administered with a drug possessing a narrow therapeutic index [13]. Models for predicting active transport of drugs would enable selection of molecules that are poorly predicted by passive permeability in silico absorption models. Additional toxicity models would also be valuable to capture other potential adverse events (as described in Section 3.4).

The future may require integrated approaches that automatically update these types of computational ADME/TOX models (Table 1) as the data is generated from the various in vitro assays, in such a way that the transition from data generation to data storage to model is seamless (Table 2). The requirements for ADME/TOX model building can be thought to represent a pathway in which all ten steps have equal importance for the eventual success of a model in our opinion (Table 2). Obviously this process would be equally applicable to many ADME/TOX properties as well as bioactivity models which both undergo multiple sequential screening cycles at present [30]. At some point a limit may be reached in terms of the computational model size and predictivity depending on the structural diversity of the data set, its representation of chemistry space and that of the molecules it is predicting. Clearly the molecular descriptor-based tools are developing alongside pharmacophore based models for predicting drug-drug interactions [31–34], the latter will perhaps have a greater viability within a structural series while the former type of approach will be more amenable to rapid screening of more diverse chemotypes. Some researchers have suggested that modeling techniques themselves may provide a method to evaluate whether historical data is consistent with models derived from acceptable data [15] which in turn can be used in further model building. This may be a questionable approach as data is generated quite differently compared with historical data as newer assays have developed. These in silico models (and potentially others) may at present be dependent on the molecules in the model and ultimately limited. For example, models for hERG inhibition do not account for other cardiac ion channels that may be inhibited or the protein binding of the molecule which affects circulating concentrations of drugs and the clinical relevance of inhibition.

Multidimensional ADME/TOX data

With data arriving at chemists from multiple directions for many endpoints including both experimental and predicted results of physicochemical, pharmacological and toxicological significance there has to be ample technology to see through many dimensions. Software such as Spotfire, Leadscope, Diva and others to some degree fulfill this requirement enabling multiple dimensional data to be assessed on a per-

Table 2. Ten requirements for integrating in vitro data to result in in silico ADME/TOX models

1.	A consistent source of in vitro data for each system modeled
2.	A large, diverse data set of molecular structures tested with a range of activities
3.	Multiple integrated molecular descriptor generation tools
4.	Methods for removing correlating descriptors and missing values
5.	Multiple algorithms (neural networks, support vector machines, genetic algorithms, recursive partitioning, pharmacophore, linear regression etc) for deriving a structure activity relationship for continuous and categorical data
6.	Statistical methods for testing, validating and selecting models for use
7.	Providing the model/s in an accessible intuitive output
8.	Integration of models with other available tools
9.	Ensure the models are used at the appropriate point in drug discovery
10.	Continual update and refinement of models with new in vitro data

sonal computer with little more training than using a spreadsheet program. These tools readily include plug-ins to display molecular structure to aid in the visualization of potential structure activity/ADME/TOX relationships. Rather than focusing on the traditional X vs Y plot, n dimensions can now be displayed, although it may take some practice for the inexperienced viewer to visualize. For example, the data for a well studied class of molecules like the β-blockers, where the in vivo data has been reviewed [35] alongside predicted physicochemical data like $\log P$ and pKa data [36], can be visualized (Figure 2). In this case the x, y and z dimensions represent calculated $\log P$, in vivo clearance and human bioavailability, respectively. This type of multidimensional approach could be applied to a previous series of analyses of relationships between physicochemical properties, potency, clearance, volume of distribution and bioavailability for the same class of molecules [37]. A further example of data for visualization is that derived for the rhinovirus inhibitors in which the in vitro data for bioactivity, metabolic stability [38] and permeability [19] of a series of molecules has been published. In this case the x, y and z dimensions represent permeability (cm/s $\times 10^{-5}$), potency (mean IC_{50} μM) and metabolic stability (% disappearance), respectively (Figure 3).

Three-dimensional plots can also be used to visualize the relationship between in silico, in vitro and in vivo data. An important example where this can be applied is to the prediction of in vivo absorption based on either calculated polar surface area or in vitro derived Caco-2 permeability data (Figure 4). The x, y and z dimensions represent Caco-2 permeability, polar surface area and percent oral absorption, respectively for 49 molecules (data collated from [20, 39, 40]). It has been suggested that there may be a strong relationship between Caco-2 permeability, human absorption and the polar surface area with relatively small well behaved data sets [20, 39–45]. In the present case there would appear to be a grey area between polar surface area values of 80–130 Å^2 between which it is difficult to predict absorption in vivo. In addition, there is little apparent relationship between polar surface area alone and the Caco-2 data. This would appear to confirm suggestions by others that the relevance of using Caco-2 cells to predict human absorption could be questionable due to a non-linear relationship and interlaboratory differences in generating the Caco-2 data [46]. In particular these authors suggested Papp values below 5×10^{-6} cm/s could not be used to predict in vivo absorption. Some understanding of molecules that are actively transported and that have large polar surface areas is obviously important as has been suggested previously [43] with a cut off for this calculated value ≥ 140 Å^2. In the present example (Figure 4) this criteria for poor absorption would appear successful. However tackling the grey area defined by recent molecules with polar surface area values in the range 80–130 Å^2 and published with Caco-2 data [19, 48] is more difficult (Table 3). All of these molecules representing two different homologous series pass the rule of 5, and additionally when scored with the Cerius 2 absorption model [47] there seems to be no relationship with the Caco-2 data. This is probably due to the Cerius 2 model containing polar surface area and AlogP descriptors for calculating absorption. The 28 LY compounds (Table 3) are all passively ab-

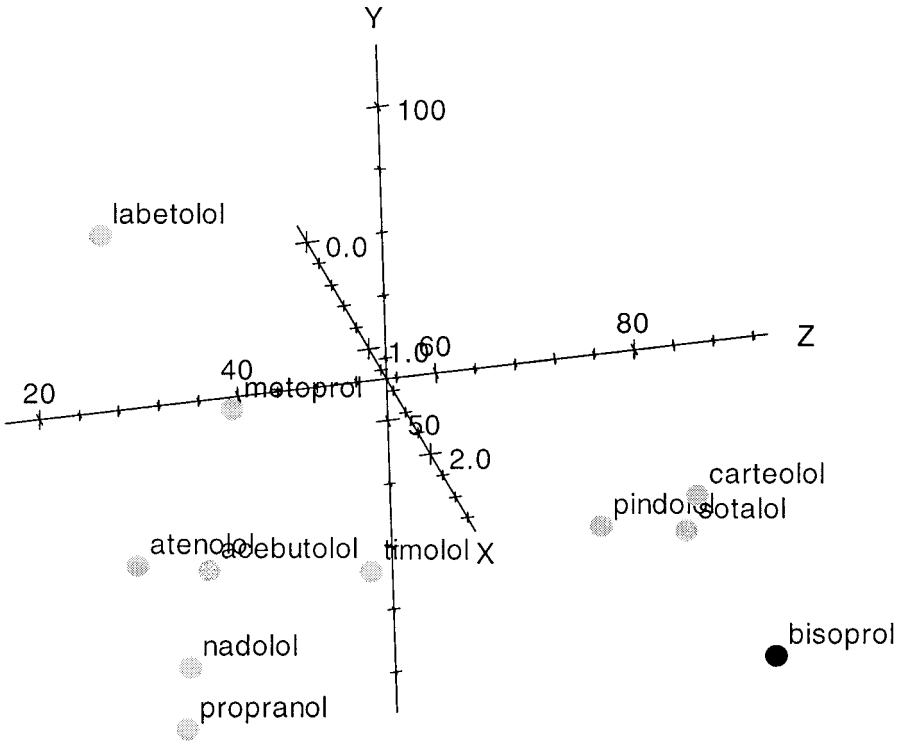


Figure 2. β-blocker in silico and in vivo data analysis. (x =calculated log$P$, y =Clearance total (L/h/70kg), z =oral bioavailability (%)). Best molecule highlighed in black. Output derived with Diva (Accelrys, San Diego, CA).

sorbed [19] whereas the remaining PNU molecules are effluxed to differing extents by an unidentified transporter [48]. Neither the rule of five or polar surface area alone offer any value in ranking the Caco-2 permeability of these molecules and therefore these factors must be considered in addition to other descriptors [49]. At the same time models may be required to decompose the differing mechanisms of drug transport and absorption [50] and improve upon the rule of five which has been shown to demonstrate a high level of false positives for predicting bioavailability [51]. By combining knowledge of important determinants such as polar surface area [39], hydrogen bonding [49, 52], lipophilicity [53], 3 dimensional structure [19, 50] and substructural fragments [51] with multiple algorithms for deriving relationships, we may be able to generate more predictive models based on a consensus of these different approaches [19].

It can be seen from these examples of visualizing data that the power of these tools will be more apparent when the data sets involved are much larger. Visualizing outputs of specific molecular descriptors may aid in understanding the (lack of) correlation with a particular dependent variables. Also these types of multidimensional visualization approaches may be useful for illustrating predicted (virtual) bioactivity and ADME/TOX parameters and then prioritizing which molecules to synthesize.

Statistical approaches: Multiple optimization

It is important at this stage to say that the in-silico predictive models for ADME/TOX are at the present time used in the drug development cycle almost the same way their in-vivo or in-vitro counterparts are used: sequentially and linearly, or one at a time, after the efficacy properties have been optimized. With this perspective, ADME/TOX predictive models are solely appreciated with respect to time and resource savings these models can contribute balanced with their perceived, but not always adequately assessed, lack of prediction. Most companies view in silico computational approaches as a technology upgrade in the way ADME/TOX properties can be predicted instead of as a change in the optimization paradigm or fundamental shift as is needed. In one perspective, in-vivo and in-vitro assays can be thought of as another type of predictive model with the same statistical properties

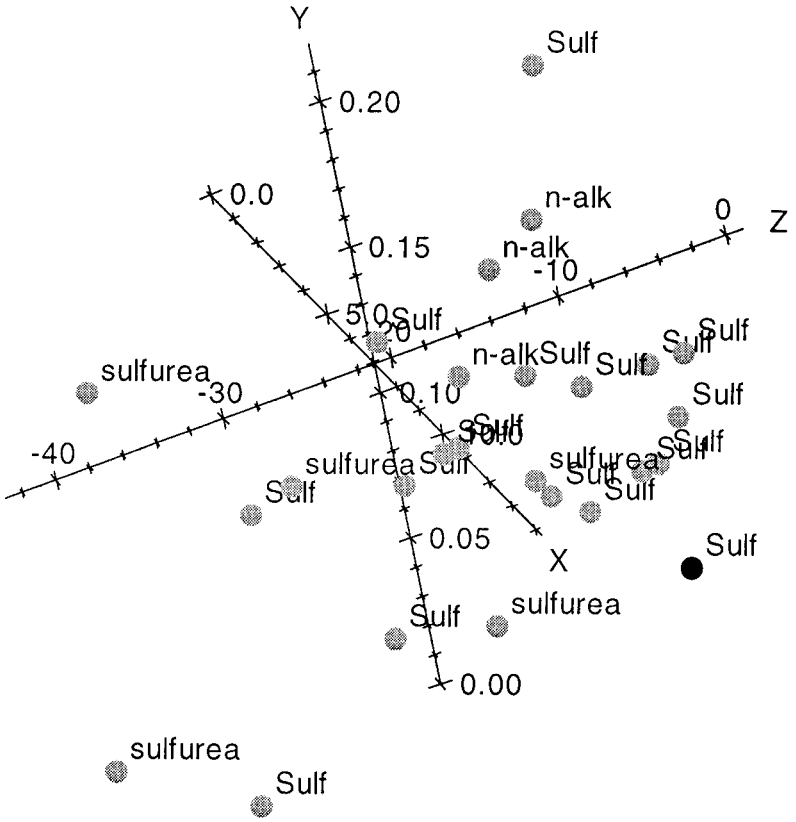


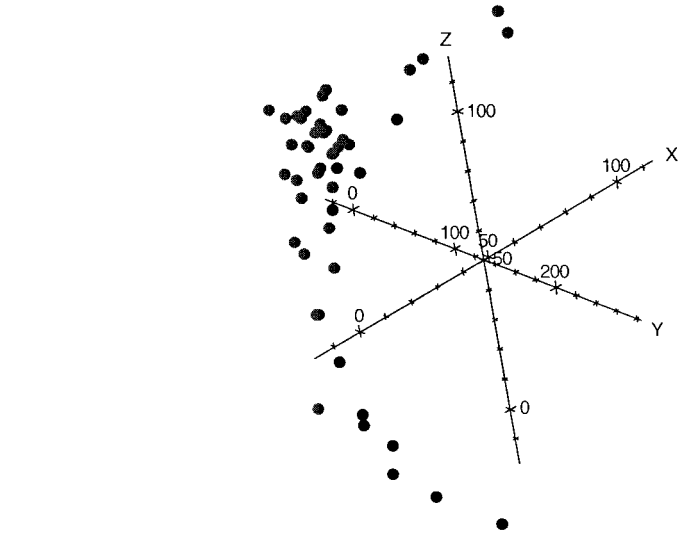
Figure 3. Rhinovirus inhibitor in vitro data analysis. (x =permeability (cm/s $\times 10^5$), y =potency (mean IC^{50} μM), z =metabolic stability (% disappearance, e.g. –20 =20% loss). Best molecule highlighed in black. Structural classes: Sulf = sulfonamide; n-alk = n-alkyl; sulfurea = sulfonylurea. Output derived with Diva (Accelrys, San Diego, CA).

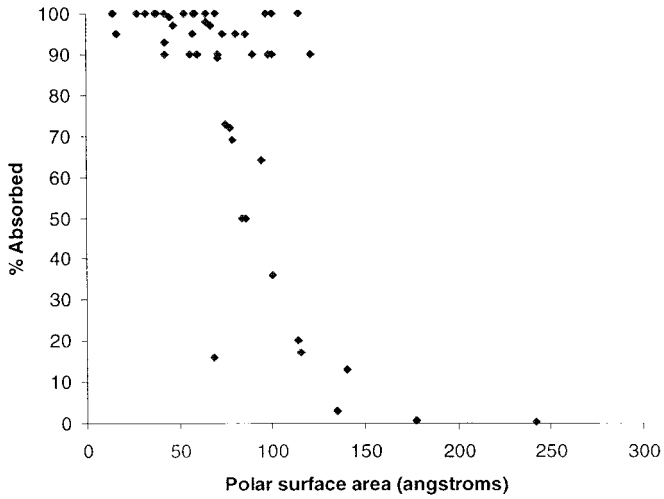
and error rate as in-silico ones. The difference among them is the way they are perceived by the different kind of scientists, whether biological, chemical or mathematical scientists.

The first challenge to implementing several predictive models in parallel with the optimization of efficacy/selectivity properties is the use of the techniques of Multi-Criteria Decision Methods (MCDM), [54] also called Multiobjective Optimization [55, 56]. The goal every drug hunter should dream of achieving is to have all properties or criteria being improved simultaneously. But the difficult challenges every drug hunter faces in daily practice are compromises that may need to be made between two or more properties. The dilemma is that modifying one part of a structure to improve property A by some positive percentage, but at the same time often alters the property B by some percentage. Is this a reasonable modification and direction to keep moving in? How does property A weight the global 'efficacy/druggability' ratio of the compound as compared to property B? Is there any better move that is allowed within the chemical space or platform envisaged? This quandary is difficult with only two properties, and becomes even more challenging with thirty or more properties.

Too often the complexity and the lack of understanding of the problem leads scientists to dichotomize on arbitrary criteria for the acceptability of the properties obtained by either in-vivo, in-vitro or in-silico models. A classical use of dichotomization rules is for example the 'Lipinski rule of 5': if a (predicted or measured) property Z for a compound is greater or less than a critical value Z^* then the compound is acceptable. If all properties envisaged are assessed in this way, then the compounds fulfilling those requirements can move to the next stage. Such a dichotomist approach might be useful in filtering a library of its worst compounds, but by no way can help in giving direction for optimization. In addition, this yes/no technique can also potentially jeopardize a research

A





B

Figure 4. (A) The 3-dimensional relationship between Caco-2 permeability (x), polar surface area (y) and percent oral absorption (z) for 49 molecules. Molecules highlighted in black have a polar surface area $> 100A^2$ (data from [19, 38, 39]. 3D output derived with Diva and PSA values generated with Cerius2 (Accelrys, San Diego, CA). (B) the 2 dimensional relationship between polar surface area and percent oral absorption is essentially similar to previous observations [42].

project if criteria or limits of acceptance are all too strictly adhered to such that very few compounds or too many are accepted. If too many compounds are retained, how and why will one strengthen criteria A preferably to criteria B? The same method that can be used for filtering or eliminating the bottom of the library can't necessarily be used for improving and maintaining the top of the library as the result is not symmetrical since the objective itself is not symmetrical. The recommended alternative is of course to work with the continuous values or all of the information available if not all variables measured/predicted are themselves continuous.

For example, Figure 5 represents graphically how the different MCDM rules operate on a simple theoretical problem where both properties A and B have to be maximized. In the four figures the set of observed compounds measured on two criteria is represented by

Table 3. Relationships between Caco-2 permeability data [19, 48] and different in silico predictions for absorption potential. All in silico predictions were generated with Cerius2 software (Accelrys, San Diego, CA).

	Papp cm/s $\times 10^{-6}$	Cerius2 absorption	PSA Å^2	Rule of 5 Violations
LY366853	0.05	good	86.99	0
LY368177	0.07	good	73.26	0
LY368288	0.18	moderate	133.9	0
LY366799	0.29	moderate	121.6	0
LY366349	0.33	moderate	130.5	0
LY354400	0.35	moderate	121.6	0
LY362683	0.37	moderate	121.6	0
LY362898	0.41	good	120.1	0
LY368228	0.43	moderate	121.6	0
LY153186	0.44	good	121.6	0
LY368227	0.44	moderate	121.6	0
LY341904	0.47	moderate	121.6	0
LY366856	0.47	moderate	121.6	0
LY355081	0.55	good	107.9	0
LY366572	0.58	moderate	121.6	0
LY366094	0.6	good	73.26	0
LY357822	0.61	good	107.9	0
LY357132	0.63	good	107.9	0
LY366659	0.63	moderate	121.6	0
LY362546	0.64	good	107.9	0
LY353462	0.67	good	124.9	0
LY341908	0.73	moderate	121.6	0
LY366347	0.73	good	107.9	0
LY359353	0.8	good	111.2	0
LY368766	0.86	good	107.9	0
LY354030	0.88	good	111.2	0
LY366092	0.93	good	107.9	0
LY122772	1.36	good	109.9	0
pnu200648	0.09	good	113.5	0
pnu200603	0.14	good	87.2	0
pnu200001	0.2	good	102.3	0
pnu200681	0.2	good	122.6	0
pnu200969	0.7	good	114.8	0
pnu200647	22.3	good	96.26	0

this shapeless convex set. In Figure 5A, the critical values fixed for both criteria A and B is defined in such a way that there are no compounds that fulfill the A&B rule. In Figure 5B however, the requirements on both criteria is too weak such that too many compounds are left. Here, compound X, is definitively less suitable than compound Z but is also retained despite this. In Figure 5C, the set of Pareto Optimal compounds is represented by the bold line in the upper right side of the observed space. It also show how the number of solution increase with the number of dimension or criteria retained. Figure 5D, illustrates how the number of solutions decreases when changing from one criteria (A or B) to two criteria together (A and B). As a rule of thumb, the Pareto Optimality Principle, the fundamental property in MCDM, should be applied. It states that a solution, a compound in our case, is Pareto Optimal if there is no other solution (compound) that is better for one criteria without being worse in the other criteria. Graphically on Figure 5C, a Pareto Op-

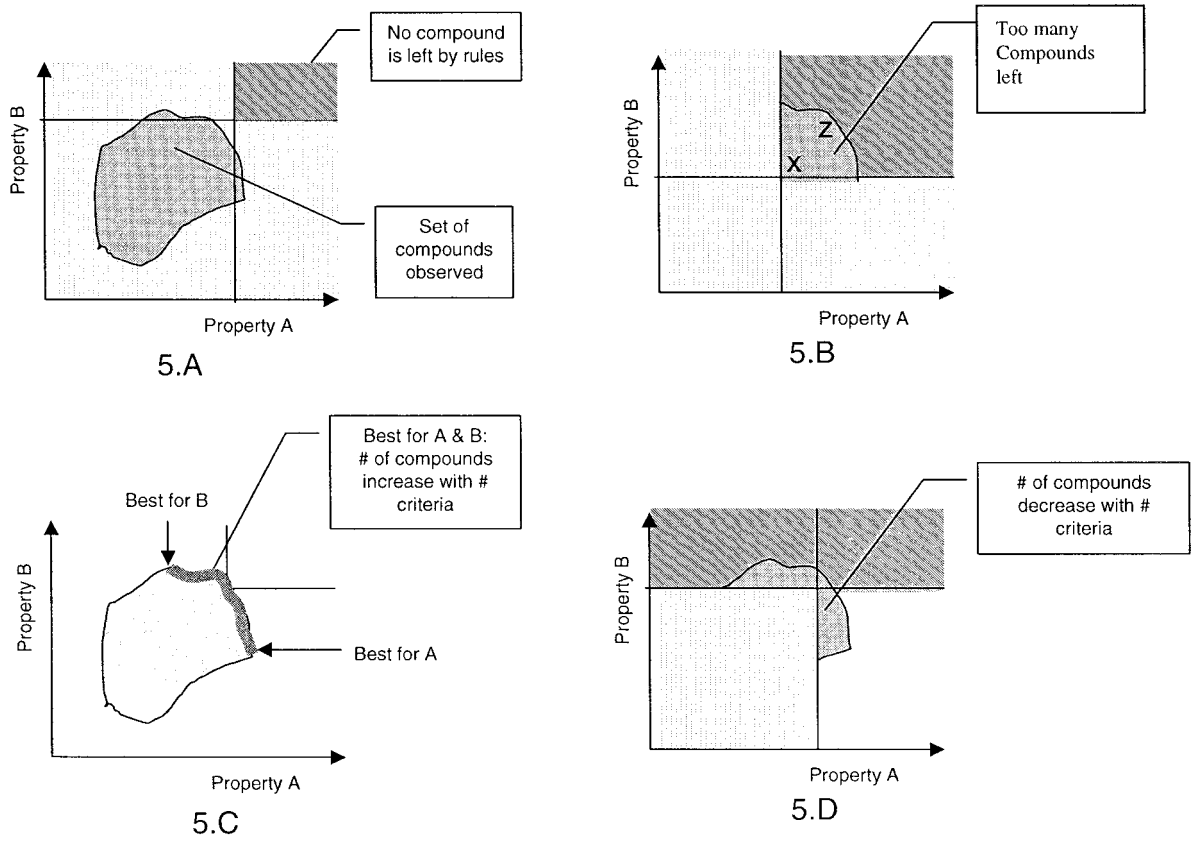


Figure 5. Graphical representation of Multi-Criteria Decision Methods applied to a theoretical problem where both properties A and B have to be maximized. In the four figures the set of observed compounds measured on two criteria is represented by the shapeless convex set.

timal compound is a compound with no other located in the upper right quadrant. This rule usually gives a more or less extensive set of solutions or compounds represented in Figure 5C by the bold line. The more dimensions or criteria are envisaged, the more extensive is the set of solutions, which is an acceptable and suitable behavior opposite to the one obtained by applying the dichotomization rule, as can be seen by comparing Figure 5C and 5D.

This subtle but unsuitable behavior of the dichotomization rule over time leads scientists to unconsciously select only a few properties to optimize, because the more properties that are maintained, the less likely it is they will find an acceptable compound. Applying modern MCDM techniques on all information available will, on the contrary, allow retention of more properties which the compounds should fulfill during the optimization process to become drugs. The success of the integration of the new early ADME/TOX models, whether in-vivo, in-vitro or in-silico, will only be achieved by implementing in the development and decision processes the modern Multi-Criteria Decision Methods.

The second major change in the discovery and optimization paradigm should be the implementation at the molecular level of the well-known and proven concepts of Design of Experiments (DOE) theory. When dealing early on with multiobjective optimization, a shift in the paradigm is required to explore efficiently the chemical space up to the optimal set of solutions. Medicinal chemists are frequently champions of proposing new innovative hypothesis for improving binding, metabolism or biological selectivity, but these sequential proposals are no more sufficient for building the best optimization strategy. The recommended technique in this review appears to be counter intuitive to most scientists and does not fit very well with the traditional hypothesis testing approach. Our experience however suggests that minimum training and exposure to concepts of Design of Experiments rapidly bridges the gap. Expanding this approach with many criteria will assume that a large number of parallel

and sometimes concurrent hypotheses will be generated and examined simultaneously. This is of course unrealistic. It however points to the fact that new multiobjective optimization is now more of an operational research and statistical issue than purely a chemical one. This will require multidisciplinary teams including in-silico scientists and specialists in optimization heuristics and optimization strategies applied to chemical sciences, if drug discovery is to capitalize on this approach.

An additional reason for developing and implementing powerful evolutionary algorithms is the fact that current combinatorial space is becoming huge. Even for pure in-silico computation, the chemical space is mostly too large for handling within a reasonable timeframe or allowable disk space. Several authors already proposed the implementation of different evolutionary strategies such as the GA (Genetic Algorithms), Simulated Annealing or different kinds of Fedorov-like search processes. All these techniques are suitable and constitute a significant improvement to the available intuitive techniques. Depending on the nature of the chemical space (whether combinatorial or not) to be optimized and the outcome desired (an optimal compound or a lead) some approaches are more efficient than others. Many efforts have been conducted to refine the efficiency of these algorithms but the present situation is that to efficiently explore a chemical space many chemical or structural features should be changed simultaneously instead of one at a time. The later is still the rule in most drug discovery teams and is known in theory as being the worst approach.

Another dimension ADME/TOX models should stimulate interest in developing evolutionary algorithms is the cost function to be optimized, an important aspect that has been too frequently neglected in research. The MCDM techniques, such as the Derringer's Desirability function can dramatically help define the cost function to minimize and could be applied easily to most optimization problems encountered such as the multiobjective optimization. The 'many–many' challenge (many structural features and many properties) has yet to be fully understood and developed but hopefully will be implemented one day in the discovery practice in order to give in silico ADME/TOX models maximal success.

In silico models for metabolic stability and drug-drug interactions

Most companies are assessing the metabolic stability and potential for drug-drug interactions of molecules as early as possible as this will be a critical piece of information that can point to the likely viability of a new molecule. However, potential for drug-drug interactions may be more appropriate as a secondary assay after metabolic stability particularly if the therapeutic agent fulfils an unmet medical need. Predicting drug-drug interactions may become important for differentiating a new molecule from one already on the market. Obviously if the parent molecule is extensively metabolized it is unlikely to be involved in drug-drug interactions, in which case assessment of the metabolites may be more important (Table 1). Metabolic stability and drug-drug interaction assays are automated such that the throughputs can be in excess of thousands of compounds per week with human or other species liver microsomes used. On the other hand some companies generate data with one or more of the commercially available recombinant CYPs and fluorescent probes (BD Gentest, Woburn, MA and Aurora, San Diego CA). In particular there is focus on CYP3A as it represents a large percentage of the human liver CYP enzyme constitution [57]. Some groups have screened for drug-drug interactions using a single concentration of inhibitor [58] in an attempt to 'bin' molecules. An alternative is to use human hepatocytes such that phase II metabolism is also captured, which is normally missed when using liver microsomes. Small data sets from these types of studies have been used to generate a structure-pharmacokinetic relationship using a decision tree and to model hydrogen abstraction using ab-initio calculation methodologies [59]. Large data sets will be computationally expensive with the latter approach and will require computational model building that include algorithms like neural networks or tree based approaches with a range of molecular descriptors described in Table 2 [60]. With in vitro microsomal or hepatocyte data, some studies have generated intrinsic clearance models using multiple linear regression, PLS, neural networks, Catalyst or Cerius2 with the genetic function approximation [61–63]. Others have developed neural networks with pharmacokinetic data such as in vivo half-life of drugs, a method which avoids the in vitro data and the in vitro-in vivo correlation altogether. Such a study using in vivo half-life data for antihistamines appeared to be predictive for a

small test set of similar molecules [64] though it remains to be seen whether this model could be applied to predicting a more structurally diverse data set.

The future of modeling drug safety

Modeling critical drug safety toxicological endpoints is a slow moving field that has not progressed as far as ADME in the past decades [65]. This is perhaps a result of toxicity modeling being dependent on information in the public domain and low throughput commercially available modeling software [17, 66–68]. Perhaps the most heavily modeled area from a toxicology perspective is genotoxicity in which a number of tools are available to recognize key molecular features involved in this process [69–72]. Recently, a comparative analysis of literature data for mutagens and non-mutagens has been enabled using a combination of statistical approaches, recursive partitioning and a number of descriptors for molecules tested in four Salmonella strains [73]. The authors concluded that compounds with a low number of hydrogen bond acceptors as well as numerous rings are more likely to be mutagenic whereas more flexible molecules are likely to be non-mutagenic. Obviously the models generated in these cases are highly dependent on the literature data available which may include molecules tested for non-pharmaceutical uses.

Computational prediction of acute toxicity has been well reviewed previously [74] with a heavy focus on molecular orbital based approaches. One area that appears to be developing within the field of drug safety is the search for non-mammalian approaches for prescreening chemical toxicity. This is not without controversy. On the one hand *Daphnia magna*, more widely used as a test organism for ecotoxicology screening, has recently been shown to be more specific an indicator of toxicity to the rat as evaluated with 54 compounds [75]. This type of in vitro model could have implications for generating very large data sets for acute toxicity. These data sets would be amenable for computational modeling using invertebrates without the use of as many expensive and ethically sensitive mammals as would normally be used. On the other hand, some studies have shown that mammals may be ultimately better predictors of human acute toxicity than aquatic invertebrates [76]. Therefore it remains to be seen whether ecotoxicological derived systems will provide a valid model for human toxicity and facilitate rapid computational toxicology modeling. In the meantime there has been an application of 3D-QSAR tools including CoMFA and Catalyst to ecotoxicological data derived for a series of chlorophenols [77]. Other aquatic organisms like *Tetrahymena pyriformis* have been used to generate large data sets for QSAR building [78] although the correlation with mammalian toxicity is less clear.

With increasing investigations of novel drug delivery routes there are perhaps unique opportunities for assessing potential toxicity at these sites. For example, with intraocular delivery being widely used for glaucoma and other treatments the ability to assess the potential for eye irritation would be of value. At present there are few such computational models capable of predicting this property for molecules of therapeutic interest. One model has used a membrane-interaction QSAR derived from solute-membrane interaction and solute aqueous solvation free energy descriptors and the genetic function approximation [79]. The result suggested increased aqueous solubility and membrane binding of the irritant molecules studied increased irritation as measured by the Draize assay. A second model studied a small set of cationic surfactants with $\log P$, log-critical micelle concentration and molecular volume as descriptors along with a neural network [80]. Presently there is no model which deals with irritation data from diverse drugs delivered via the intraocular route.

Clearly, there are many facets of toxicology and focusing on those areas with the greatest current mechanistic insight will increase the chance of enabling some success in the short term. Computational models for drug metabolizing enzyme induction are possible [81–82] and this has advanced as larger data sets of relevant inducers have been investigated [83]. With the new hypothesis of Pregnane-X-Receptor (PXR)-mediated CYP3A induction and the publication of a crystal structure for the ligand binding domain [84], there has been some interest in developing high throughput computational tools for evaluating binding [85]. However there are drawbacks with the application of these techniques due to varying interpretations of how PXR works [86]. As more data is also generated for the constitutive androstane receptor (CAR) involved [86] in the induction of CYP2B [87], CYP3A and possibly other genes, models will likely be generated for the ligands and inhibitors. As it stands now a simple alignment of 5 molecules suggested to be CAR activators [88] yields a pharmacophore (Figure 6A) with 2 hydrophobic features and one hydrogen bond acceptor in a similar arrangement to the CYP2B6 substrate pharmacophore [89]. This could indicate that

CAR is a less promiscuous receptor than PXR with ultimately less flexibility in the ligands involved in activation, as suggested by its high affinity for the rigid androstane metabolites [90] that are inhibitors of CAR (Figure 6B). Computational approaches will likely be applied to other important proteins with a role in the drug metabolizing enzyme and drug-transport induction processes and at the same time may shed some light on understanding their mechanisms and interactions.

In conjunction with computational models based on ligand structure, fast docking and accurate scoring algorithms [91–93] that could use the protein structure of targets of toxicological significance alongside 3D-QSAR approaches need more in depth work in the future. Modeling toxicity further requires large sets of quality experimental data and appreciation for the complexity of the cellular processes involved. In addition a better understanding of the potential of computational modeling by toxicologists, more powerful computational models that have a high level of predictivity, and a larger amount of risk taking outside of this heavily regulated science would also be desirable.

The importance of transporters as targets for improving bioavailability

Oral absorption remains the favored route of administration from a patient comfort and compliance point of view. As a result, the marketing push for the development of orally administrable drugs and/or dosage forms poses a significant attrition threat to pharmacologically potent drug candidates. Despite intensive research in this area, oral drug delivery remains the greatest challenge to the pharmaceutical scientist, not only for the delivery of macromolecules but also for small lead compounds with poor membrane permeability. The drug industry is acutely aware of this challenging conundrum and has shown an increasing interest in oral drug delivery approaches. Much progress has been made in the area of advanced drug delivery systems such as nanotechnology devices and biodegradable polymers; however, the drug industry's aversion to complicated, costly systems may ultimately limit the application of these approaches. Simplified systems, such as small molecules that can be synthesized on a large scale are desirable de facto, and exploring pathways and systems to accommodate this requirement has drawn intense attention.

A promising approach for increasing oral bioavailability is targeting to intestinal transporter systems. Transporters are polytopic membrane proteins that are indispensable to the cellular uptake and homeostasis of many essential nutrients. During the past decade it has become clear that a vast number of drugs share transport pathways with nutrients. Moreover, a critical role has been recognized for transport proteins in the absorption, excretion, and toxicity of drug molecules, as well as in their pharmacokinetic and pharmacodynamic (PK/PD) profiles. Because cellular transporter expression is often regulated by nuclear orphan receptors that simultaneously regulate the translation and expression of metabolic enzymes in the cell (e.g. P-glycoprotein and cytochrome P450 regulation by the pregnane X receptor), they indirectly control drug metabolism. Thus, transport proteins are involved in all facets of drug ADME and ADME/TOX, conferring an important field of study for pharmaceutical scientists involved in these areas. As a result, in-depth knowledge of membrane transport systems may be extremely useful in the design of new chemical entities (NCE). After all, it is now well-appreciated that the most critical parameter for a new drug to survive the drug development pipeline on its way to the market is its ADME/TOX profile.

Despite the involvement of solute transporters in fundamental cellular processes, most are poorly characterized at the molecular level. As a result, we are unable to accurately predict the interaction of drugs with this important class of membrane proteins a priori, and detection of drug-transporter interactions has been unacceptably serendipitous.

Transporter pathways can be exploited to increase oral drug absorption in two ways: 1) designing NCE that mimic the molecular structure of the endogenous substrate to the intestinal transporter system or 2) piggybacking compounds through the cell by conjugation to substrates for transport systems. Both approaches require detailed knowledge of the three-dimensional arrangement(s) between the substrate and transport protein. At the present time, very few transport proteins have been successfully crystallized and – as a result – structural details and affinity requirements of transport proteins are absent. The application potential of in silico techniques is finally gaining acceptance in the transporter field as more manuscripts appear that use a combined biology/molecular modeling approach to infer useful information on transporter structure-activity relationships. This represents an opportunity for optimizing molecules as substrates for the solute

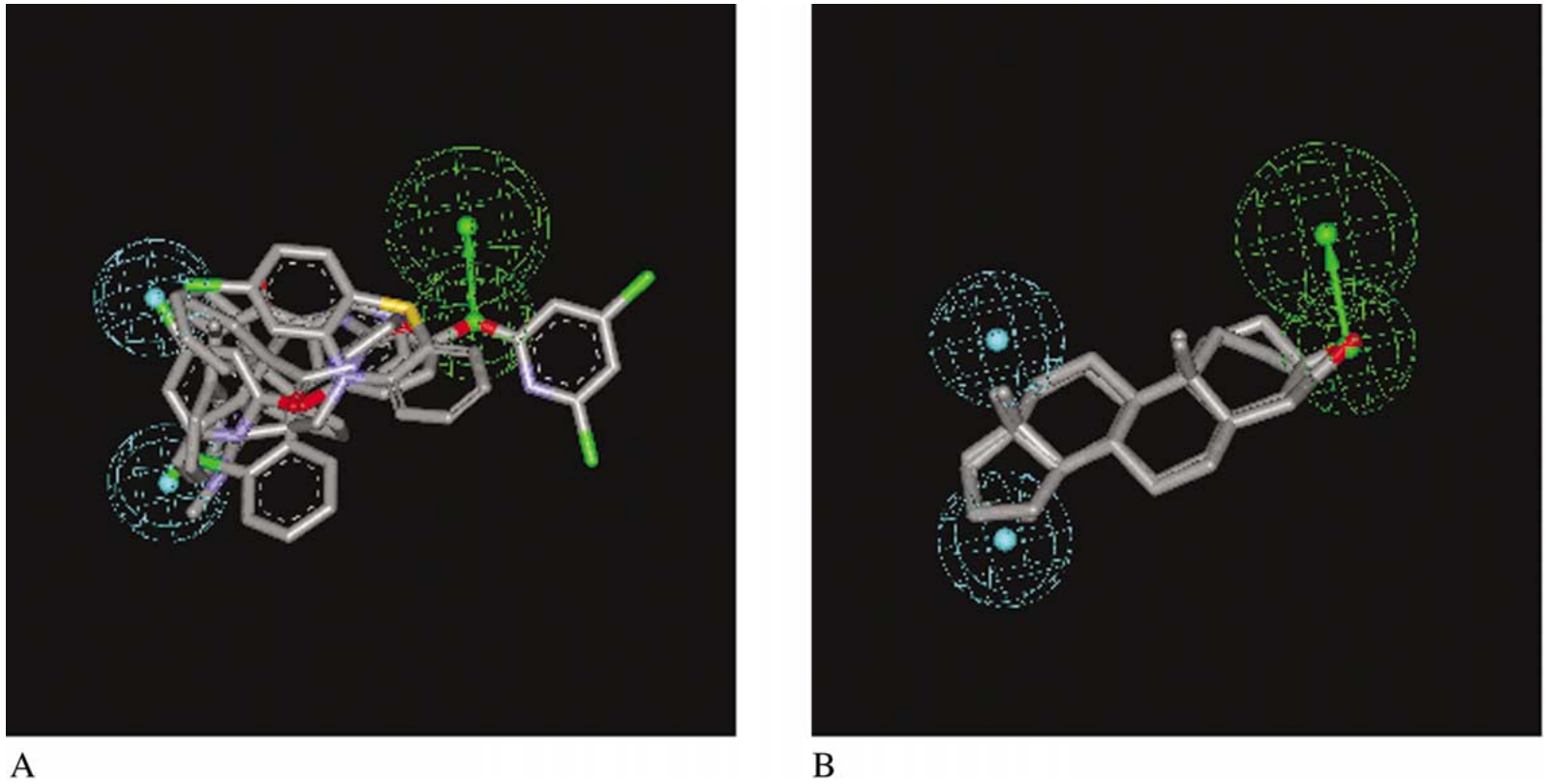


Figure 6. Common feature alignment (HIPHOP function in Catalyst, (Accelrys, San Diego, CA)) of ligands for the constitutive androstane receptor (CAR). (A) CAR activators chlorpromazine, clotrimazole, phenobarbital, metyrapone, 1,4-bis[2-(3,5-dichloropyridyloxy)]benzene and phenobarbital [87] used to generate an alignment in Catalyst to result in 2 hydrophobic features (cyan) and one hydrogen bond acceptor (green). (B) CAR inhibiting steroids 16, 5α-androsten-3α-ol and 5α-androstan-3α-ol fitted to the common features alignment generated with the structurally diverse CAR activators in Figure 6A.

transporters and providing a further screening system for drug discovery. Clearly the future growth in knowledge of transporter function will be lead by integrated in vitro and in silico approaches.

Structural models of transport proteins and methods to design substrates

The primary and, to a lesser extent, secondary structures of many transport systems are known. Due to the absence of suitable crystallization methods, however, only a few polytopic membrane proteins have yielded to X-ray crystallographic analyses [94–97] resulting in high-resolution three-dimensional structural information. Recently, combinatorial approaches have been applied to successfully crystallize membrane proteins in a high-throughput approach, resulting in the structure of a homolog of the multidrug resistant ATP binding cassette transporters to a resolution of 4.5 angstrom [98]. In this study, approximately 96,000 crystallization conditions using about 20 detergents were tested to yield crystals with good quality for x-ray structure determination. Unfortunately, the resulting protein was not functionally active, but the structural information corroborated several previously obtained experimental results. The combinatorial crystallization approach used by Chang and Roth [98] will likely stimulate other researchers to attempt crystallization of other important membrane proteins.

In the mean time, we base our views of solute transport on molecular models that provide working models of transport systems [99–101]. It is well recognized that any two proteins that show sequence homology (i.e., share sufficient primary structural similarity to have evolved from a common ancestor) will prove to exhibit strikingly similar 3-D structures [102]. Furthermore, the degree of tertiary structural similarity correlates well with the degree of primary structural similarity. Phylogenetic analyses allow application of modeling techniques to a large number of related proteins and additionally allow reliable extrapolation from one protein family member of known structure to others of unknown structure. Thus, once 3-D structural data are available for any one family member, these data can be applied to all other members within limits dictated by their degrees of sequence similarity.

To date, a few examples of in silico models with potential application to prodrug design exist. The Swaan group and the joint laboratories of Baringhaus and Kramer have reported on the three-dimensional structural requirements of the apical sodium-dependent bile acid transporter (ASBT) that will be of specific use in the development of novel substrates for this transport protein. Using a training set of 17 chemically diverse inhibitors of ASBT, Baringhaus and colleagues [103] developed an enantiospecific Catalyst pharmacophore that mapped the molecular features essential for ASBT affinity. The ASBT pharmacophore model is in good agreement with the 3D-QSAR model previously developed using a series of 30 ASBT inhibitors and substrates [104]. In this study, the electrostatic and steric fields around bile acids were mapped using comparative molecular field analysis (CoMFA) to identify regions of putative interaction with ASBT (Figure 7). This model enabled the in silico design of substrates for ASBT, especially for conjugation at the C-17 position. The two aforementioned models should facilitate the rational design of (pro)drugs for specific targeting to ASBT. There may also be the possibility of combining models from different groups although this may present challenges on its own.

One potential limitation of indirect structure-activity models is the inherent lack of information on substrate-transport protein interactions at the molecular level. In the absence of high-resolution molecular information on ASBT, we have recently constructed an in silico model for the ligand binding of the ASBT transport protein using knowledge-based homology modeling [105]. Using the transmembrane domains of bacteriorhodopsin as a scaffold, the extracellular loop regions were superposed and optimized with molecular dynamics simulations. By probing the protein surface with cholic acid putative binding domains were identified that provide rational leads for future site-directed mutagenesis studies. In addition to the previously described pharmacophore and 3D-QSAR models, the molecular representation of ASBT may provide a powerful tool in the design of novel substrates or inhibitors for this transporter. Recently, Zuniga and colleagues [106] presented a 3D model of the human facilitative glucose transporter, GLUT1. Their model, based on lac permease topology, explains most of the structural features earlier resolved by biophysical techniques and clearly illustrates the future direction and utilization of modeling transport proteins in silico.

Future directions in utilizing transporter systems

It is clear that over the past decade we have gained a great deal of knowledge around the functioning of transport proteins and how they can be manipulated as drug targets for maximizing absorption and bioactiv-

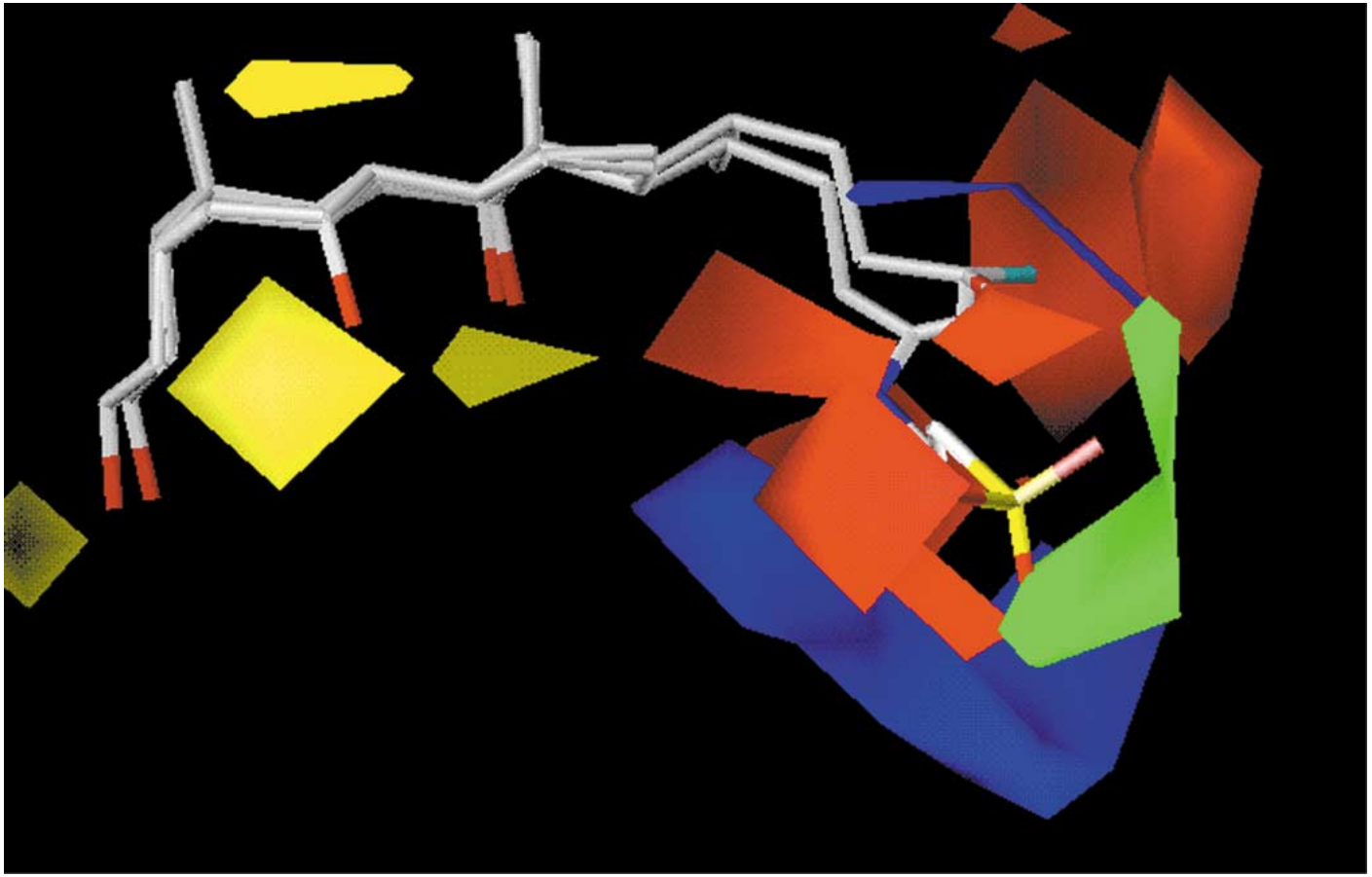


Figure 7. CoMFA (Comparative molecular field analysis) steric and electrostatic fields surrounding the natural bile acids cholic acid and taurocholic acid. Red and blue fields indicate areas where negative and positive charge are favored, respectively. Yellow and green fields represent areas were steric bulk or hydrophobic interactions are favorable or non favorable, respectively. CoMFA areas can serve the medicinal chemist to optimize a lead compound for interaction with a transporter system, ASBT in this case.

ity through prodrug approaches. Although we are still without a crystal structure for these proteins there has been an explosion in interest in structure-activity relationships. This has been seen largely as part of a wider growth in using computational approaches for understanding in vitro data and in particular in the fields of drug design, optimization and ADME properties [14]. The resolution of the three-dimensional structures of solute transporter protein models is presumably low, however, they can be justified by their ability to confirm biologically relevant phenomena. As with all models, the continued input of novel experimental data and revalidation of the model may eventually lead to highly predictive systems capable of in silico detection and design of novel solute transporter substrates. Furthermore, we can expect that a combination of indirect (3D-QSAR) and direct (homology models) techniques may lead to newer, higher-resolution screening systems. It is likely that the integration of bioinformatics, computational and in vitro models will drive our understanding of transport proteins to new levels and afford an opportunity for further possible therapeutic targets and drug design opportunities.

The development of cell culture systems that stably express specific transporters or sets of cooperative transporters/enzymes are now useful tools in screening potential transporter candidates; this, in turn, should facilitate rational design and rapid screening of numerous potential substrates for transporter systems. A worthwhile example in this case is the development of valacyclovir (Valtrex®), a prodrug of acyclovir known to be recognized in the GI tract by the small intestinal peptide carrier, PepT1 (SLC15A1) [107, 108]. Although not rationally designed to target an intestinal transporter, this approach increased the oral bioavailability of acyclovir from approximately 20% to 70%. Increased knowledge of the molecular and structural biology as well as substrate requirements for transport systems in combination with high-throughput screening techniques will allow a more rational development of substrates for intestinal solute transporters.

Learning from complex biological systems modeled in silico

In order to expand the impact of in silico approaches even further in the pharmaceutical industry, the next arena for computational approaches is that of biology. This field has already been initiated with the influx of data generated by molecular biology, creating the field of bioinformatics which includes under its umbrella applications in sequence comparison, multiple alignment, DNA arrays, gene prediction and proteomics. Clearly if we could simulate all biological processes and focus on a higher level of the cell [109] by expanding this growing list of computational approaches, then we will go beyond the reductionist genomic paradigm [110] to a higher level understanding of where and how pharmaceutical intervention could have the greatest impact. It appears biology is ripe for theoretical approaches after centuries of historical research focused on finding and naming discoveries. It has recently been proposed that the developing field of computational molecular biology requires biologists in the 21st century to know the basics of discrete mathematics and algorithms [111]. Clearly this will extend the alliance of biology with chemistry that occurred during the 20th century to the integration of many more fields and approaches simultaneously.

Due to the use of high throughput approaches in industry the quality and quantity of available experimental biology observations is no longer a road-block to understanding the whole cell. The challenge of requiring a data-driven approach for biology pressures computing science as it stands at present to develop advanced integrated tools that organize and structure existing data for a complex system. The metabolic pathways traditionally pasted to laboratory walls are likely to change from a metabolite centric viewpoint to one where the enzyme involved in catalysis, the protein itself, is central [112]. This has been the case for many years in drug metabolism where there has been intense focus on the enzyme specificity [57, 113, 114]. This type of approach will aid in changing the way we view and interpret complex biological pathways and their interconnections. As this type of knowledge increases in quantity, databases of such interactions and pathways with links to other relevant sources of information, adding to their overall utility and multidimensionality becomes essential. Some of these databases have already been developed as object-orientated tools with graphical and textual representations. For example, in some cases they may contain search links to literature, experimental data and sequence information [115]. Computational models could be derived by combining such databases of pathways to influence and feed further experiments [116]. This biological understanding can be summarized in the words of Dietrich Dorner; 'The more we know, the more clearly we realize what we don't know' [117]. The process of building in silico models for complex biological systems represent the need for an iterative model building

relationship that has been suggested in computational chemistry, whereby a model should be updated as new data or knowledge becomes available.

Computational models continue to be generated to explain complex cellular signaling pathways or metabolic cascades. One example is a recent study on the serotonergic modulation of sensory neurons in Aplysia. These simulations were performed using the program SNNAP (Simulator for Neural Networks and Action Potentials). The model was used to show that PKC had a role in serotonin-induced spike broadening while PKA was involved in increased excitability [118]. Another example is in drug metabolism where the modeling of a single enzyme's catalysis of one molecule to multiple metabolites [119] using commercially available software, Kinecyte was demonstrated. This study showed that toxicity of a substrate is likely to be minimized when there are multiple metabolites and this therefore may represent a biological advantage. To date the complex interconnections where one molecule may be metabolized by different enzymes with different kinetic constants and then the metabolite cascade may follow a different pathway has not been modeled.

This brings us into the realm of predicting drug safety. For example, the need for in silico approaches to predict action on the cardiac cell as described earlier is at present the most high profile due to the life threatening consequences of cardiac arrhythmia [120]. Certainly this is just one example of a complex cellular process that could be modeled more completely by understanding whole cell and whole organ behavior not only in terms of the modulated ion channels but by integrating the metabolic information, potential for drug-drug interactions and protein binding. The potential power of combining computational ADME/TOX, computational biology and computational chemistry (Figure 1) will lead to an ability to predict such characteristics for a molecule much earlier and preempt synthesis of poorly scored virtual molecules.

There is a unique opportunity that awaits the groups that go beyond the present sequential drug discovery paradigm using combinatorial chemistry and high throughput screening (HTS) to instead integrate simultaneous multiple approaches to provide predictive models which together can increase the efficiency and overall speed of drug discovery and development. This situation is analogous to the recent proposal for model-based discovery as applied to the postgenomic era which ties together visualization using graphical models and data into a systems based approach that capitalizes on multiple competing models to judge the relative potential for competing hypotheses to occur [121].

Conclusions

The goal of validated and generally applicable computational models is the avoidance of the need to synthesize and screen every molecule for a given program and the simultaneous optimization of bioactivity with ADME/TOX properties (Figure 1). By incorporating these types of filters into combinatorial library design and HTS screening, molecules with optimal ADME properties and structural diversity may be readily selected early in drug discovery [14, 16, 122]. Numerous reviews of in silico modeling for ADME/TOX [13, 14, 16, 45, 123] and the potential for impact [14, 16, 30] have appeared. The power of these tools will not be achieved by talking about them alone. Using these in silico tools and their adoption across many disciplines in the pharmaceutical industry alongside other advanced tools in development will be the true acid test for their impact on the discovery process.

Visualization and mining of data derived from an array of computational models presently available for ADME/TOX and bioactivity will provide the next paradigm for drug discovery and be amenable to pharmaceutical companies of all sizes. By its nature this approach will not demand a large library of synthesized molecules, but instead will require expertise in the computational tools derived from internal or external experimental data. Rather than using the traditional sequential process flow [124] apparent in most companies we should ensure that molecules are selected after concurrent assessment of predicted properties [5, 14, 16] (Figure 1). In silico tools for drug discovery have the potential to rapidly change the shape of the pharmaceutical industry and the economics of new drug discovery such that research costs will decrease. Such a new paradigm will require scientists to adopt and adapt to diverse types of computational models and visualization software and will require the development of new technologies that make this a reality and turn virtual data and molecules into drugs faster.

Acknowledgements

The authors gratefully acknowledge Dr Hung-Ir Li for assistance in initial designs for Figure 1, and Dr James Grace and our many colleagues for their discussions

References

1. Love, B., PSTT, 1 (1998) 89–90.
2. Cavalla, D., PSTT, 1 (1998) 91–92.
3. Kennedy, T., Drug Discovery Today, 2 (1997) 436–444.
4. Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J., Adv. Drug. Del. Rev. 23 (1997) 3–25.
5. Lipinski, C.A., J. Pharm. Toxicol. Methods, 44 (2000) 235–249.
6. Lipinski, C.A., Curr. Drug Discovery, April (2001) 17–19.
7. Gillet, V.J., Willett, P. and Bradshaw, J., J. Chem. Inf. Comp. Sci., 38 (1998) 165–179.
8. Sadowski, J. and Kubinyi, H., J. Med. Chem., 41 (1998) 3325–3329.
9. Oprea, T.I., Davis, A.M., Teague, S.J. and Leeson, P.D., J. Chem. Inf. Comput. Sci., 41 (2001) 1308–1315.
10. Bemis, G.W. and Murcko, M.A., J. Med. Chem., 39 (1996) 2887–2893.
11. Bemis, G.W. and Murcko, M.A., J. Med. Chem., 42 (1999) 5095–5099.
12. Muegge, I., Heald, S.L. and Brittelli, D., J. Med. Chem., 44 (2001) 1–6.
13. Ekins, S., Ring, B.J., Bravi, G., Wikel, J.H. and Wrighton, S.A., (2000b) Predicting drug-drug interactions in silico using pharmacophores: A paradigm for the next millenium, in Pharmacophore, perception, development, and use in drug design (Guner OF ed) 269–299, University International Line, San Diego.
14. Ekins, S., Waller, C.L., Swaan, P.W., Cruciani, G., Wrighton, S.A. and Wikel, J.H., J. Pharm. Tox. Methods, 44 (2000d) 251–272.
15. Barratt, M.D., Env. Health. Perspectives., 106 (1998) 459–465.
16. Ekins, S., Ring, B.J., Grace, J., McRobie-Belle, D.J. and Wrighton, S.A., J. Pharm. Tox. Methods, 44 (2000c) 313–324.
17. Barratt, M.D. and Rodford, R.A., Curr. Opin. Chem. Biol., 5 (2001) 383–388.
18. Lombardo, F., Blake, J.F. and Curatolo, W.J., J. Med. Chem. 39 (1996) 4750–4755.
19. Ekins, S., Durst, G.L., Stratford, R.E., Thorner, D.A., Lewis, R., Loncharich, R.J. and Wikel, J.H., J. Chem. Inf. Comput. Sci., 41 (2001b) 1578–1586.
20. Stenberg, P., Norinder, U., Luthman, K. and Artursson, P., J. Med. Chem., 44 (2001) 1927–1937.
21. Walters, W.P., Stahl, M.T. and Murcko, M.A., Drug Discovery Today, 3 (1998) 160–178.
22. Leach, A.R. and Hann, M.M., Drug Discovery Today, 5 (2000) 326–336.
23. Barnum, D., Greene, J., Smellie, A. and Sprague, P., J. Chem. Inf. Comput. Sci. 36 (1996) 563–571.
24. Ekins, S., Bravi, G., Wikel, J.H. and Wrighton, S.A., J. Pharmacol. Exp. Ther., 291 (1999d) 424–433.
25. Chen, X., Rusinko, I., Tropsha, A. and Young, S.S., J. Chem. Inf. Comput. Sci., 39 (1999) 887–896.
26. Saiakhov, R., Stefan, L.R. and Klopman, G., Persp. Drug. Disc. Design., 19 (2000) 133–135.
27. Colmenarejo, G., Alvarez-Pedraglio, A. and Lavandera, J.-L., J. Med. Chem., 44 (2001) 4370–4378.
28. Crumb, Jr. W.J., Ekins, S., Sarazan, D., Wikel, J.H., Wrighton, S.A., Carlson, C. and Beasley, C., J. Clin. Psychiatry., (2002) in press.
29. Ekins, S., Crumb, W.J., Sarazan, R.D., Wikel, J.H. and Wrighton, S.A., J. Pharmacol. Exp. Ther., (2002) in press.
30. Manly, C.J., Louise-May, S. and Hammer, J.D., Drug Discovery Today, 6 (2001) 1101–1110.
31. Ekins, S., Bravi, G., Binkley, S., Gillespie, J.S., Ring, B.J., Wikel, J.H. and Wrighton, S.A., J. Pharmacol. Exp. Ther., 290 (1999b) 429–438.
32. Ekins, S., Bravi, G., Binkley, S., Gillespie, J.S., Ring, B.J., Wikel, J.H. and Wrighton, S.A., Pharmacogenetics, 9 (1999a) 477–489.
33. Ekins, S., Bravi, G., Binkley, S., Gillespie, J.S., Ring, B.J., Wikel, J.H. and Wrighton, S.A., Three and four dimensional-quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2C9 inhibitors, 28 (2000a) 994–1002.
34. Ekins, S., de Groot, M. and Jones, J.P., Drug. Metab. Dispos., 29 (2001a) 936–944.
35. Bertz, R.J. and Granneman, R.G., Clin. Pharmacokinet., 32 (1997) 210–258.
36. Destroyer, A., Vander Heyden, Y., Carda-Broch, S., Garcia-Alvarez-Coque, M.C. and Massart, D.L. J. Chromatogr. A., 912 (2001) 211–221.
37. van de Waterbeemd, H., Smith, D.A. and Jones, B.C., J. Comput.-Aid. Mol. Des., 15 (2001) 273–286.
38. Stratford, R.E.J., Clay, M.P., Heinz, B.A., Kuhlfeld, M.T., Osborne, S.J., Phillips, D.L., Sweetan, S.A., Tebbe, M.J., Vasudevan, V., Zornes, L.L. and Lindstrom, T.D., J. Pharm. Sci., 88 (1999) 747–753.
39. Palm, K., Stenberg, P., Luthman, K. and Artursson, P., Pharm. Res., 14 (1997) 568–571.
40. Yazdanian, M., Glynn, S.L., Wright, J.L. and Hawi, A., Pharm. Res., 15 (1998) 1490–1494.
41. Palm, K., Luthman, K., Ungell, A.-L., Strandlund, G. and Artursson, P., J. Pharm. Sci., 85 (1996) 32–39.
42. Palm, K., Luthman, K., Ungell, A.-L., Strandlund, G., Beigi, F., Lundahl, P. and Artursson, P., J. Med. Chem., 41 (1998) 5382–5392.
43. Clark, D.E., J. Pharm. Sci., 88 (1999) 807–814.
44. Stenberg, P., Luthman, K., Ellens, H., Lee, C.P., Smith, P.L., Lago, A., Elliot, J.D. and Artursson, P., Pharm. Res., 16 (1999) 1520–1526.
45. Stenberg, P., Luthman, K. and Artursson, P., J. Control. Rel., 65 (2000) 231–243.
46. Ren, S. and Lien, E.J., Prog. Drug. Res., 54 (2000) 1–23.
47. Egan, W.J., Merz, K.M.J. and Baldwin, J.J., J. Med. Chem., 43 (2000) 3867–3877.
48. Schipper, N.G.M., Osterberg, T., Wrange, U., Westberg, C., Sokolowski, A., Rai, R., Young, W. and Sjostrom, B., Pharm. Res., 18 (2001) 1735–1741.
49. Oprea, T.I. and Gottfries, J., J. Mol. Graph. Model., 17 (1999) 261–274.
50. Ghuloum, A.M., Sage, C.R. and Jain, A.N., J. Med. Chem., 42 (1999) 1739–1748.
51. Andrews, C.W., Bennett, L. and Yu, L.X., Pharm. Res., 17 (2000) 639–644.
52. van de Waterbeemd, H., Camenenisch, G., Golkers, G. and Raevsky, O.A., QSAR, 15 (1996) 480–490.
53. Yoshida, F. and Topliss, J.G., J. Med. Chem., 43 (2000) 2575–2585.
54. Derringer, G. and Suich, R., J. Quality Technology, 12 (1980) 214–219.
55. Rassokin, D.N. and Agrafiotis, D.K., J. Mol. Graph. Model., 18 (2000) 368–382.
56. Agrafiotis, D.K., J. Chem. Inf. Comput. Sci., 41 (2002) 159–167.

57. Wrighton, S.A., Schuetz, E.G., Thummel, K.E., Shen, D.D., Korzekwa, K.R. and Watkins, P.B., Drug. Metab. Revs., 32 (2000) 339–361.
58. Gao, F., Johnson, D.L., Ekins, S., Janiszewski, J., K.G. K., Meyer, R.D. and West, M., (2001) Submitted.
59. Bursi, R., de Gooyer, M.E., Grootenhuis, A., Jacobs, P.L., van der Louw, J. and Leysen, D., J. Mol. Graph. Modelling., 19 (2001) 552–556.
60. Ekins, S. and Rose, J.P., J. Mol. Graph., 20 (2002) 305–309.
61. Lave, T., Dupin, S., Schmitt, C., Valles, B., Ubeaud, G., Chou, R.C., Jaeck, D. and Coassolo, P., Pharm. Res. 14 (1997) 152–155.
62. Schneider, G., Coassolo, P. and Lave, T., J. Med. Chem., 42 (1999) 5072–5076.
63. Ekins, S. and Obach, R.S., J. Pharmacol. Exp. Ther., 295 (2000) 463–473.
64. Quinones, C., Caceres, J., Stud, M. and Martinez, A., Quant Struct Act Rel., 19 (2000) 448–454.
65. Hansch, C., Drug Metabolism Reviews, 15 (1984) 1279–1294.
66. Cronin, M.T., Pharm. Pharmacol. Commun., 4 (1998) 157–163.
67. Richard, A.M., Toxicol. Lett., 102–103 (1998) 611–616.
68. Johnson, D.E. and Wolfgang, G.H.I., DDT, 5 (2000) 445–454.
69. Smith, C.J., Hansch, C. and Morton, M.J., Mutation Res., 379 (1997) 167–175.
70. Benigni, R., Passerini, L., Gallo, G., Giorgi, F. and Cotta-Ramusino, M., Env. Mol. Mutagenesis., 32 (1998) 75–83.
71. Mathews, E.J. and Contrera, J.F., Reg. Toxicol Pharmacol, 28 (1998) 242–264.
72. Tafozoli, M., Baeten, A., Geerlings, P. and Kirsch-Volders, M., Mutagenesis, 13 (1998) 115–126.
73. Llorens, O., Perez, J.J. and Villar, H.O., J. Med. Chem., 44 (2001) 2793–2804.
74. Stoffers, A.E.M.F., Boersma, M.G., Vaes, W.H.J., Vervoort, J., Tyrakowska, B., Hermens, J.L.M. and Rietjens, I.M.C.M., Toxicol in vitro, 15 (2001) 539–551.
75. Guilhermino, L., Diamantino, T., Silva, M.C. and Soares, A.M.V.M., Ecotoxicol Env. Safety, 46 (2000) 357–362.
76. Calleja, M.C., Persoone, G. and Geladi, P., Fd. Chem. Toxic., 32 (1994) 173–187.
77. Briens, F., Bureau, R. and Rault, S., Ecotxicol. Env. Safety., 43 (1999) 241–251.
78. Cronin, M.T.D. and Schultz, T.W., Chem. Res. Toxicol., 14 (2001) 1284–1295.
79. Kulkarni, A.S. and Hopfinger, A.J., Pharm. Res., 16 (1999) 1245–1253.
80. Patlewicz, G.Y., Rodford, R.A., Ellis, G. and Barratt, M.D., Toxicol in vitro, 14 (2000) 79–84.
81. Rekka, E.A. and Kourounakis, P.N., Eur. J. Drug Metab. Pharmacokinetics, 21 (1996) 7–11.
82. van der Burght, A.S.A.M., Clijsters, P.J., Horbach, G.J., Andersson, P.L., Tysklind, M. and van den Berg, M., Toxicol Appl. Pharmacol., 155 (1999) 13–23.
83. Schuetz, E.G., Beck, W.T. and Schuetz, J.D., Mol. Pharmacol., 49 (1996) 311–318.
84. Watkins, R.E., Wisely, G.B., Moore, L.B., Collins, J.L., Lambert, M.H., Williams, S.P., Willson, T.M., Kliewer, S.A. and Redinbo, M.R., Science, 292 (2001) 2329–2333.
85. Ekins, S. and Erickson, J.A., Drug. Metab. Dispos., 30 (2002) 96–99.
86. Ekins, S. and Schuetz, E., TIPS, (2002).
87. Honkakoski, P., Zelko, I., Sueyoshi, T. and Negishi, M., Mol. Cell. Biol., 18 (1998) 5652–5658.
88. Honkakoski, P., Jaaskelainen, I., Kortelahti, M. and Urtti, A., Pharm. Res. 18 (2001) 146–150.
89. Ekins, S., Bravi, G., Ring, B.J., Gillespie, T.A., Gillespie, J.S., VandenBranden, M., Wrighton, S.A. and Wikel, J.H., J. Pharmacol. Exp. Ther., 288 (1999c) 21–29.
90. Forman, B.M., Tzameli, I., Choi, H.-S., Chen, J., Simha, D., Seol, W., Evans, R.M. and Moore, D.D., Nature, 395 (1998) 612–614.
91. Bohm, H.-J. and Klebe, G., Angew. Chem. Int. Ed. Engl., 35 (1996) 2588–2614.
92. So, S.-S. and Karplus, M., J. Comput.-Aid. Mol. Des., 13 (1999) 243–258.
93. Shimada, J., Ishchenko, A.V. and Shakhnovich, E.I., Protein Sci., 9 (2000) 765–775.
94. Deisenhofer, J. and Michel, H., Annu. Rev. Biophys. Chem., 20 (1991a) 247–266.
95. Deisenhofer, J. and Michel, H., Annu. Rev. Cell. Biol., 7 (1991b) 1–23.
96. Unwin, N., Nature, 373 (1995) 37–43.
97. Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R. and Yoshikawa, S., Science, 272 (1996) 1136–1144.
98. Chang, G. and Roth, C.B. Science, 293 (2001) 1793–1800.
99. Goswitz, V.C. and Brooker, R.J., Protein. Sci., 4 (1995) 534–537.
100. Kaback, H.R., Voss, J. and Wu, J., Curr. Opin. Struct. Biol., 7 (1997) 537–542.
101. Varela, M.F. and Wilson, T.H., Biochim. Biophys. Acta, 1276 (1996) 21–34.
102. Doolittle, R.F., Johnson, M.S., Husain, I., Van Houten, B., Thomas, D.C. and Sancar, A., Nature, 323 (1986) 451–453.
103. Baringhaus, K.H., Matter, H., Stenglin, S. and Kramer, W., J. Lipid Res., 40 (1999) 2158–2168.
104. Swaan, P.W., Szoka, F.C.J. and Oie, S., J. Comput. Aid. Mol. Des., 11 (1997) 581–588.
105. Zhang, E.H., Phelps, M.A., Helsper, F. and Swaan, P.W., (2002) submitted.
106. Zuniga, F.A., Shi, G., Haller, J.F., Rubashkin, A., Flynn, D.R., Iserovich, P. and Fischbarg, J., J. Biol. Chem., 276 (2001) 44970–44975.
107. Balimane, P.V., Tamai, I., Guo, A., Nakanishi, T., Kitada, H., Leibach, F.H., Tsuji, A. and Sinko, P.J., Biochem. Biophys. Res. Commun. 250 (1998) 246–251.
108. Han, H., de Vrueh, R.L., Rhie, J.K., Covitz, K.M., Smith, P.L., Lee, C.P., Oh, D.M., Sadee, W. and Amidon, G.L., Pharm. Res., 15 (1998) 1154–1159.
109. Palsson, B., Nature Biotechnology, 18 (2000) 1147–1150.
110. Ekins, S. and McGowan, R.J., Phil. Sci., 9 (2001) 1–23.
111. Pevzner, P.A., (2000) Computational molecular biology. MIT Press, Cambridge, MA.
112. Gerrard, J.A., Sparrow, A.D. and Wells, J.A., Trends Biochem Sci., 26 (2001) 137–140.
113. Wrighton, S.A. and Stevens, J.C., Crit. Rev. Toxicol., 22 (1992) 1–21.
114. Rendic, S. and Di Carlo, F.J., Drug Metab. Rev., 29 (1997) 413–580.
115. Ellis, L.B.M., Speedie, S.M. and McLeish, R., Bioinformatics, 14 (1998) 803–806.
116. Bower, J.M. and Bolouri, H., Computational modeling of genetic and biochemical networks. (2001) The MIT press, Cambridge, MA.

117. Dorner, D., (1996) The logic of failure. Perseus Books, Reading, MA.
118. Baxter, D.A., Canavier, C.C., Clark, Jr. J.W. and Byrne, J.H., J. Neurophysiol., 82 (1999) 2914–2935.
119. Cook, D.L. and Atkins, W.M., Biochemistry, 36 (1997) 10802–10806.
120. Noble, D. and Colatsky, T.J., Emerg. Ther. Targets, 4 (2000) 39–49.
121. Gifford, D.K., Science, 293 (2001) 2049–2051.
122. Darvas, F., Dorman, G. and Papp, A., J. Chem. Inf. Comput. Sci., 40 (2000) 314–322.
123. Ekins, S. and Wrighton, S.A., J. Pharm. Tox. Methods, 44 (2001) 1–5.
124. Hall, A.H., Toxicol. Lett., 102–103 (1998) 623–626.