

## ADMET *IN SILICO* MODELLING: TOWARDS PREDICTION PARADISE?

Han van de Waterbeemd\* and Eric Gifford‡

Following studies in the late 1990s that indicated that poor pharmacokinetics and toxicity were important causes of costly late-stage failures in drug development, it has become widely appreciated that these areas should be considered as early as possible in the drug discovery process. However, in recent years, combinatorial chemistry and high-throughput screening have significantly increased the number of compounds for which early data on absorption, distribution, metabolism, excretion (ADME) and toxicity (T) are needed, which has in turn driven the development of a variety of medium and high-throughput *in vitro* ADMET screens. Here, we describe how *in silico* approaches will further increase our ability to predict and model the most relevant pharmacokinetic, metabolic and toxicity endpoints, thereby accelerating the drug discovery process.

### Why is *in silico* ADMET needed?

Traditionally, drugs were discovered by testing compounds synthesized in time-consuming multi-step processes against a battery of *in vivo* biological screens. Promising compounds were then further studied in development, where their pharmacokinetic properties, metabolism and potential toxicity were investigated. Adverse findings were often made at this stage<sup>1</sup> (FIG. 1), with the result that the project would be halted or re-started to find another clinical candidate — an unacceptable burden on the research and development budget of any pharmaceutical company.

Today, this paradigm has been re-worked in several ways. The testing of drug metabolism, pharmacokinetics and toxicity is today done much earlier; that is, before a decision is taken to evaluate a compound in the clinic. However, the rate at which biological screening data are obtained has dramatically increased, and (ultra)high-throughput screening (HTS) facilities are now common at large pharmaceutical companies and at specialized biotechs<sup>2</sup>. In response to these developments, a new approach to chemistry — combinatorial chemistry — has been adopted to feed these highly efficient hit-finding machines. Combinatorial chemistry makes it possible to synthesize large series of closely related libraries of chemicals using the same chemical reaction and appropriate reagents. Such

libraries are then run through the HTS to find hits around which further, more focused, series are designed and synthesized in a next round.

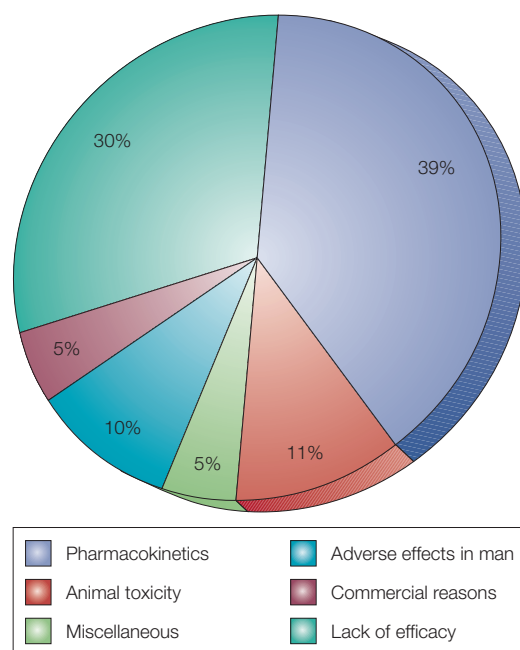
As the capacity for biological screening and chemical synthesis have dramatically increased, so have the demands for large quantities of early information on absorption, distribution, metabolism, excretion (ADME) and toxicity data (together called ADMET data). Various medium and high-throughput *in vitro* ADMET screens are therefore now in use. In addition, there is an increasing need for good tools for predicting these properties to serve two key aims — first, at the design stage of new compounds and compound libraries so as to reduce the risk of late-stage attrition; and second, to optimize the screening and testing by looking at only the most promising compounds.

**Drug-like properties.** Which properties make drugs different from other chemicals? A number of studies have been performed with the aim of answering this question (for examples, see REFS 3–6). A particularly influential example — the analysis of the World Drug Index (WDI)<sup>5</sup>, which lead to Lipinski's 'rule-of-five' — identifies several critical properties that should be considered for compounds with oral delivery in mind. These properties, which are usually viewed more as guidelines rather than absolute cutoffs, are molecular mass <500 daltons

\*Pfizer Global Research & Development, PDM, Sandwich, Kent CT13 9NJ, UK.

‡Pfizer Global Research & Development, Discovery Research Informatics, 2,800 Plymouth Road, Ann Arbor, Michigan 48105, USA.

Correspondence to H.v.d.W.  
e-mail: han\_waterbeemd@sandwich.pfizer.com  
doi:10.1038/nrd1032



**Figure 1 | An analysis of the main reasons for attrition in drug development<sup>1</sup>.** In this analysis, published five years ago, half of all failures were attributed to poor pharmacokinetics (39%) and animal toxicity (11%). Such analyses clearly indicated that these two areas should be focused on as early as possible in the drug-discovery process (although it should be noted that the interpretation of such data is often hampered by the fact that compounds may have more than one flaw and, as the project was halted, these might not always have been identified). An even better approach would be to use predictive tools in the design phase of the synthesis of compounds and compound libraries.

(Da), calculated octanol/water partition coefficient (CLOGP) <5, number of hydrogen-bond donors <5 and number of hydrogen-bond acceptors <10. In general, such studies, and others not cited here, point to the most important physicochemical and structural properties characteristic of a good drug in the context of our current knowledge. These properties are then typically used to construct predictive ADME models and form the basis for what has been called property-based design<sup>7</sup>. To a certain extent, similar molecules can be expected to have similar ADME properties<sup>8</sup>. This concept is the basis of software called SLIPPER-2001, in which physicochemical DESCRIPTORS and molecular similarity are used for the prediction of properties such as lipophilicity, solubility and fraction absorbed in humans<sup>9</sup>.

**How are ADMET data obtained?** The quest for early, fast and relevant ADMET data is tackled in three ways. First, a variety of *in vitro* assays have been further automated through the use of robotics and miniaturization. Second, *in silico* models are being used to assist in the selection of both appropriate assays, as well as in the selection of subsets of compounds to go through these screens. Third, predictive models have been developed that might ultimately become sophisticated enough to replace *in vitro* assays and/or *in vivo* experiments.

**When is ADMET data needed?** The need for ADMET information starts with the design of new compounds. This information can influence the decision to proceed with synthesis either via traditional medicinal chemistry or combinatorial chemistry strategies. Obviously, at this stage, computational approaches are the only option for getting this information, but it is also acceptable that the predictions are not perfect at this point. Once a series of molecules is focused around a lead and is further optimized towards a clinical candidate, more robust mechanistic models will be required.

**What ADME properties do we want to predict?** A deeper understanding of the relationships between important ADME parameters and molecular structure and properties has been used to develop *in silico* models that allow the early estimation of several ADME properties<sup>10–17</sup>. Among other important issues, we want to predict properties that provide information about dose size and dose frequency (BOX 1), such as oral absorption, bioavailability, brain penetration, clearance (for exposure) and volume of distribution (for frequency).

As a result of the availability of experimental data in the literature, considerable effort has gone into the development of models to predict physicochemical properties relevant to ADME, such as lipophilicity. However, despite its importance, the prediction of pharmacokinetic properties such as clearance, volume of distribution and half-life directly from molecular structure is making slower progress owing to a lack of published data. Similarly, the prediction of various aspects of metabolism and toxicity is also underdeveloped.

**What computational tools are used?** Here, there are two aspects to consider: data modelling and molecular modelling, which have different toolboxes. Molecular modelling includes approaches such as protein modelling<sup>18</sup>, which uses quantum mechanical methods to assess the potential for interaction between the small molecules under consideration and proteins known to be involved in ADME processes, such as cytochrome P450s. This requires three-dimensional structural information on the protein, which can be built by homology modelling of related structures if the human protein structure is not available. If no structural information on the protein is available, an alternative way of assessing the potential of a small molecule to interact with a particular protein is to use PHARMACOPHORE models, which are built from a superposition of known substrates of the protein.

For data modelling, quantitative structure–activity relationship (QSAR) approaches<sup>19</sup> are typically applied. QSAR and quantitative structure–property relationship (QSPR) studies have been performed since the 1960s with a variety of biological and physicochemical data. These studies use statistical tools to search for correlations between a given property and a set of molecular and structural descriptors of the molecules in question. Once such a QSAR model has been ‘TRAINED’ using a set of molecules for which experimental data on the property in question are available, it can be used to make

#### DESCRIPTOR

A structural or physicochemical property of a molecule or part of a molecule. Examples include log P, molecular mass and polar surface area.

#### PHARMACOPHORE

A pharmacophore is the ensemble of steric and electronic features that are necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response.

#### TRAINING

The building of a model using part of the data (that is, the training set), followed by validation of the model using the rest of the data (that is, the validation set). Finally, the model is tested using compounds (the test set) not used for training and validation.

## Box 1 | Pharmacokinetics

Pharmacokinetics is the study of the time course of a drug within the body and incorporates the processes of absorption, distribution, metabolism and excretion (ADME)<sup>76</sup>. Pharmacokinetic parameters are derived from the measurement of drug concentrations in blood or plasma. The key pharmacokinetic parameters and their importance for the dose regimen and dose size are shown in the figure<sup>80</sup>.

Most drugs are given orally for reasons of convenience and compliance. Typically, a drug dissolves in the gastro-intestinal tract, is absorbed through the gut wall and then passes the liver to get in to the blood circulation. The percentage of the dose reaching the circulation is called the bioavailability. From there, the drug will be distributed to various tissues and organs in the body. The extent of distribution will depend on the structural and physicochemical properties of the compound. Some drugs can enter the brain and central nervous system by crossing the blood–brain barrier. Finally, the drug will bind to its molecular target, for example, a receptor or ion channel, and exert its desired action.

- Volume of distribution (Vd) is a theoretical concept that connects the administered dose with the actual initial concentration (C<sub>0</sub>) present in the circulation:

$$Vd = \text{Dose}/C_0$$

Most drugs will bind to various tissues and in particular to proteins in the blood, such as albumin. As only the free (unbound) drug will bind to the molecular target, the concept of unbound volume of distribution is used:

$$Vdu = Vd/f_u, \text{ where } f_u \text{ is the fraction unbound.}$$

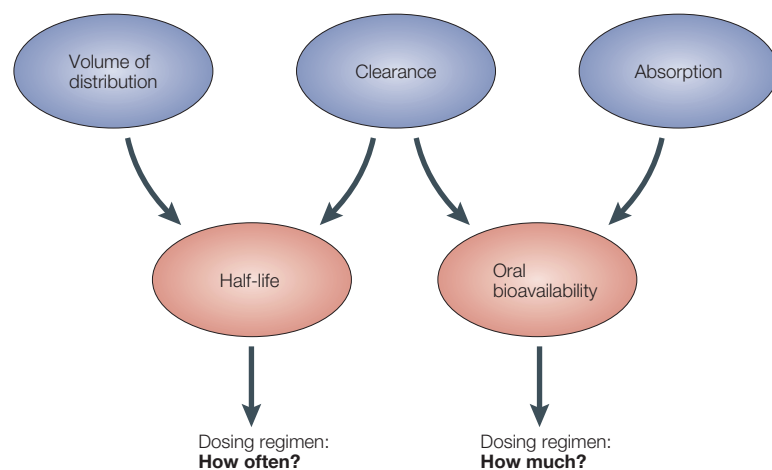
- Clearance (Cl) of the drug from the body mainly takes place via the liver (hepatic clearance or metabolism, and biliary excretion) and the kidney (renal excretion).

By plotting the plasma concentration against time, the area under the curve (AUC) relates to dose, bioavailability and clearance.

$$AUC = F \times \text{Dose}/Cl$$

- Half-life (t<sub>1/2</sub>) — the time taken for a drug concentration in the plasma to reduce by 50% — is a function of the clearance and volume of distribution, and determines how often a drug needs to be administered.

$$t_{1/2} = 0.693 Vd/Cl$$



predictions on molecules not in the training set, although, in general, reliable predictions are only possible for molecules similar to those in the training set.

A wide variety of descriptors for use in QSAR studies have been developed over the last 40 years<sup>20</sup> (for example, those available in the program Dragon). A subset of these descriptors is potentially useful for predicting ADME properties. Indeed, with the increased interest in the prediction of ADME properties, specifically tailored descriptors have already been reported, for example, those in the VolSurf program<sup>21</sup>. Some of the descriptors used are close to the chemist's

intuition, such as molecular size and hydrogen bonding. Other descriptors are merely topological or quantum chemical concepts, but can produce highly predictive models, although these might be 'black boxes' for most people.

Using appropriate descriptors, QSAR approaches — ranging from simple multiple linear regression to modern MULTIVARIATE ANALYSIS techniques, such as partial least squares (PLS) — are now being applied to the analysis of ADME data<sup>22</sup>. Data-mining and machine-learning methods originally developed and used in other fields are now also successfully being used for this purpose. Examples of such methods include NEURAL NETWORKS (NN), self-organizing maps (SOM; also called Kohonen networks), RECURSIVE PARTITIONING (RP) and support vector machines (SVM).

Good predictive models for ADMET parameters depend crucially on selecting the right mathematical approach, the right molecular descriptors for the particular ADMET endpoint, and a sufficiently large set of experimental data relating to this endpoint for the validation of the model (BOX 2). Insight is growing as to which of the available descriptors and QSAR tools are most appropriate, although there often seems to be different options with similar predictive power. In particular, more needs to be learnt about how the size of the training set influences the choice of the most capable model.

#### Prediction of physicochemical properties

The physicochemical properties of a drug have an important impact on its pharmacokinetic (BOX 1) and metabolic (BOX 3) fate in the body, and so a good understanding of these properties, coupled with their measurement and prediction, are crucial for a successful drug discovery programme.

**Lipophilicity.** Poor biopharmaceutical properties — in particular, poor aqueous solubility and slow dissolution rate — can lead to poor oral absorption and hence low oral bioavailability. In general, poor solubility is related to high lipophilicity, whereas hydrophilic compounds generally show poor permeability and hence low absorption. Therefore, the measurement of solubility and lipophilicity, as well as ionization constants affecting these two properties, has been automated and integrated in the high-throughput drug discovery paradigm.

The relationship between lipophilicity and pharmacokinetic properties has been discussed by various workers in the field<sup>23–25</sup>. Lipophilicity is the key physicochemical parameter linking membrane permeability — and hence drug absorption and distribution — with the route of clearance (metabolic or renal). Measuring the lipophilicity of a compound is readily amenable to automation. The gold standard for expressing lipophilicity is the partition coefficient P (or log P to have a more convenient scale) in an octanol/water system; alternatives include applications of immobilized artificial membranes (IAM), immobilized liposome chromatography (ILC) and liposome/water partitioning.

**MULTIVARIATE ANALYSIS**  
A subset of statistical techniques that can deal with larger sets of molecular descriptors that is aimed at finding relationships or patterns in data sets. Examples include multiple linear regression (MLR) and partial least squares (PLS).

## Box 2 | The need for good data

Clearly, larger databases of marketed drugs are required to establish more robust models to predict various ADME properties, including drug–drug interactions. Several published ADME data sets are available for data modelling<sup>13,36,47,105–107</sup>, but the quality of the data and the number of available training examples remain important issues. In the future, service providers such as Cerep, Novascreen and Cyprotex will be able to offer larger data sets with which to build more robust models.

In a recent symposium, the question of whether the Internet can help as a resource to collect relevant ADME data was addressed<sup>108</sup>. The current opinion is that there are some good and well-maintained websites available, but unfortunately also many of questionable use in research owing to a lack of control of data quality or reference to the original data.

There is continued interest in developing and improving log P calculation programs, and there are many such programs available. Most calculation approaches rely on fragment values, although simple methods based on molecular size and hydrogen-bonding indicators for functional groups to calculate log P values have also been shown to be extremely versatile<sup>22</sup>.

However, log P values can only be a first estimate of the lipophilicity of a compound in a biological environment. For partition processes in the body, the distribution coefficient D (log D) — for which an aqueous buffer at pH 7.4 (blood pH) or 6.5 (intestinal pH) is used in the experimental determination — often provides a more meaningful description of lipophilicity, especially for ionizable compounds. However, in our experience, programs that can reliably predict log D are scarce at present.

**Solubility.** The first step in the drug absorption process is the disintegration of the tablet or capsule, followed by the dissolution of the active drug. Obviously, low solubility is detrimental to good and complete oral absorption, and so the early measurement of this property is of great importance in drug discovery. Reflecting this need, rapid, robust methods reliant on turbidimetry and nephelometry have been developed to efficiently measure the solubility of large numbers of compounds<sup>6,26</sup>.

Ideally, only soluble compounds would be synthesized in a drug-discovery programme. Predictive solubility methods — for example, neural networks — might assist in this effort. However, at present, no approaches are robust enough to accurately predict low solubility. Many current predictive solubility programs<sup>27</sup> use training data from different laboratories with varying quality and different experimental conditions. Hopefully, by measuring many compounds under standardized conditions, current predictive models can be improved<sup>28</sup>.

**pKa.** As ionization can also affect the solubility, lipophilicity (log D), permeability and absorption of a compound, approaches have been developed for the rapid measurement of pKa values of sparingly soluble drug compounds. Using experimental data reported in the literature, several approaches have been used to develop pKa calculators. Programs include ACD/pKa (ACD), Pallas/pKa (CompuDrug) and SPARC<sup>29</sup>.

**Hydrogen bonding.** The hydrogen-bonding capacity of a drug solute is now recognized as an important determinant of permeability. In order to cross a membrane, a drug molecule needs to break hydrogen bonds with its aqueous environment. The more potential hydrogen bonds a molecule can make, the more energy this bond-breaking costs, and so high hydrogen-bonding potential is an unfavourable property that is often related to low permeability and absorption.

Initially, AlogP — the difference between octanol/water and alkane/water partitioning — was used as a measure for solute hydrogen-bonding, but this technique is limited by the poor solubility of many compounds in the alkane phase. A variety of computational approaches have addressed the problem of estimating hydrogen-bonding capacity, ranging from simple heteroatom (O and N) counts, the consideration of molecules in terms of the number of hydrogen-bond acceptors and donors, and more sophisticated measures that take into account such parameters as free-energy factors<sup>30</sup> and (dynamic) polar surface area (PSA)<sup>31</sup>. The latter are easily calculated, and it is now believed that a single minimum-energy conformation is sufficient to compute the PSA, instead of the more computationally demanding and time-consuming dynamic polar-surface-area calculation<sup>31</sup>. A fast fragment-based algorithm for PSA has been reported<sup>32</sup>, which allows PSA calculations to be implemented in virtual screening approaches.

**Permeability.** Efforts have been undertaken to predict the permeability of compounds through Caco-2 cells, which serve as a model for human intestinal absorption, in an approach called membrane-interaction quantitative structure–activity relationships (MI-QSAR)<sup>33</sup>. But one could ask the question, “Why model the model of human absorption?”. A more direct approach is to model processes that would address ‘pure’ measures of permeability. These include octanol/water partitioning, liposome partitioning, retention on immobilized artificial membranes (IAM), the parallel artificial membrane-permeability assay (PAMPA) and binding to liposomes measured by surface-plasmon-resonance (SPR) biosensors.

#### Prediction of ADME and related properties

**Absorption.** For a compound crossing a membrane by purely passive diffusion, a reasonable permeability estimate can be made using single molecular properties, such as log D or hydrogen-bonding capacity. However, besides the purely physicochemical component contributing to membrane transport, many compounds are affected by biological events, including the influence of transporters and metabolism (further discussed in later sections). Many drugs seem to be substrates for transporter proteins, which can either promote or hinder permeability. In particular, the combined role of **cytochrome P450 3A4** (CYP3A4) and **P-glycoprotein** (P-gp) in the gut as a barrier to drug absorption has been well studied<sup>34</sup>. Currently, no theoretical SAR basis exists to account for these effects.

#### NEURAL NETWORKS

Neural networks are computational models that are based on the principles of the functioning of the brain. They can be used to model nonlinear relationships between dependent (biological endpoint to be predicted) and independent (molecular and structural descriptors) variables. Examples include back-propagation and self-organising maps (SOM; also called Kohonen neural networks).

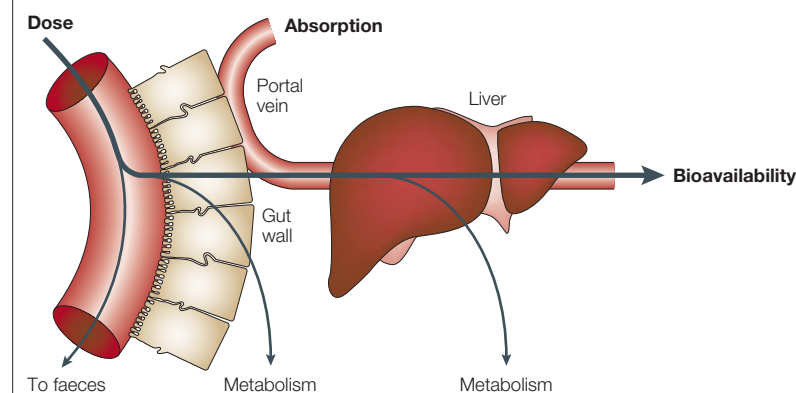
#### RECURSIVE PARTITIONING OR DECISION TREES

A supervised learning method producing a tree-structured series of rules to predict a particular property using a set of molecular descriptors as input.

## Box 3 | Metabolism

The body will eventually try to eliminate xenobiotics, including drugs. For many drugs, this first requires metabolism or biotransformation, which takes place partly in the gut wall during uptake, but primarily in the liver. The figure shows where metabolism occurs during the absorption process. The fraction of the initial dose appearing in the portal vein is the fraction absorbed, and the fraction reaching the blood circulation after the first-pass through the liver defines the bioavailability of the drug.

Traditionally, a distinction is made between phase I and phase II metabolism, although these do not necessarily occur sequentially. In phase I metabolism, a molecule is functionalized, for example, through oxidation, reduction or hydrolysis. The most important enzymes involved are the cytochrome P450s. In particular, CYP3A4, CYP2D6, CYP2C9 and CYP2C19 are important for the metabolism of drugs in humans. In phase II metabolism, the functionalized drug molecule is further transformed in so-called conjugation reactions. These include for example, glucuronidation and sulfation, as well as conjugation with glutathione. It should be noted that the metabolism in animals might be different from that in humans, and therefore the prediction of human pharmacokinetics and metabolism from animal data might not be straightforward.



*In vitro* methods, such as Caco-2 or Madin-Darby canine kidney (MDCK) monolayers, are widely used to make oral absorption estimates. These cells also express transporter proteins, but only express very low levels of metabolizing enzymes. Similarly, there is a continued interest in finding a relevant *in vitro* screen for estimating the permeability of drugs for diseases of the central nervous system (CNS). The bovine microvessel endothelial cell (BMEC) model has been explored as a possible *in vitro* model of the blood–brain barrier<sup>35</sup>.

Considerable effort has also gone into the development of *in silico* models for the prediction of oral absorption<sup>36–42</sup>. The simplest models are based on a single descriptor, such as log P or log D, or polar surface area, which is a descriptor of hydrogen-bonding potential<sup>31</sup>. Different multivariate approaches, such as multiple linear regression, partial least squares and artificial neural networks<sup>41</sup>, have been used to develop quantitative structure–human–intestinal–absorption relationships. In all approaches, hydrogen bonding is considered to be a property with an important effect on oral absorption.

Absorption-simulation programs, such as GastroPlus<sup>43</sup> and Idea<sup>44</sup>, might eventually become a valuable tool in lead optimization and compound selection. These programs, which have recently been compared<sup>45</sup>, are computer simulation models developed and validated to predict ADME outcomes, such as rate of absorption and extent of absorption, using a limited

number of *in vitro* data inputs. They are based on advanced compartmental absorption and transit (ACAT) models, in which physicochemical concepts, such as solubility and lipophilicity, are more readily incorporated than physiological aspects involving transporters and metabolism. In more recent versions, attempts are being made to model the influence of transporters, in addition to gut-wall metabolism, on gastrointestinal uptake. For example, the oral bioavailability of ganciclovir in dogs and humans was simulated using a physiologically based model that utilized many biopharmaceutically relevant parameters, such as the concentration of ganciclovir in the duodenum, jejunum, ileum and colon at a variety of dose levels and solubility values. The simulation results demonstrated that the low bioavailability of ganciclovir is limited by compound solubility rather than permeability due to partitioning, as previously speculated<sup>44</sup>.

**Bioavailability.** Recently, the first attempts to predict bioavailability directly from molecular structure have been published. However, this is not an easy task, as bioavailability depends on a superposition of two processes: absorption and liver first-pass metabolism. Absorption in turn depends on the solubility and permeability of the compound, as well as interactions with transporters and metabolizing enzymes in the gut wall. Important properties for determining permeability seem to be the size of the molecule, as well as its capacity to make hydrogen bonds, its overall lipophilicity and possibly its shape and flexibility. Molecular flexibility, for example, as evaluated by counting the number of rotatable bonds, has been identified as a factor influencing bioavailability in rats<sup>46</sup>.

Yoshida and Topliss<sup>47</sup> trained a QSAR model with log D at pH 7.4 and 6.5 as inputs for the physicochemical properties and the presence/absence of typical functional groups most likely to be involved in metabolic reactions as the structural input. This approach used ‘fuzzy adaptive least squares’, and drugs could be classified into one of four predefined bioavailability ranges. Using this approach, a new drug can be assigned to the correct class with an accuracy of 60%. An unpublished effort based on classification using the SIMCA approach and which seems to achieve similar success has also been reported<sup>12</sup>.

In another approach, regression and recursive partitioning have been used<sup>48</sup>. In this study, 591 compounds were included and a set of 85 structural descriptors was generated. The authors noted that the mean error in the experimental data used to generate the model is ~12%, with an increase in error for well-absorbed drugs. Therefore, the models should not be expected to generate predictions that are more accurate than the variability inherent in the biological measurements.

Genetic programming, which is a specific form of evolutionary programming, has recently been used for predicting bioavailability<sup>49</sup>. The results show a slight improvement compared with the Yoshida-Topliss approach, although a direct comparison is difficult owing to a different selection of the bioavailability ranges of the four classes.

A method for predicting bioavailability using adaptive fuzzy partitioning (AFP) has recently been presented at conferences<sup>50</sup>. The best molecular descriptors were selected with a genetic algorithm, and in the next step SOMs were used for the classification, which correctly classified the molecule in the right bioavailability class in 64% of cases.

The methods described above demonstrate that at least qualitative (binned) predictions of oral bioavailability seem tractable directly from molecular structure. Approaches using *in vitro* data are also under continual development. For example, a graphical approach for bioavailability prediction based on the combined measurement of Caco-2 flux and microsomal stability was recently presented<sup>51</sup> that uses a reference plot to make a prediction of bioavailability for a new compound. Typically, the prediction will classify a compound as 0–20%, 20–50% or 50–100% bioavailable. Extending this approach to include solubility, for example, might increase its predictive power.

**Blood–brain barrier penetration.** Drugs that act in the CNS need to cross the blood–brain barrier (BBB) to reach their molecular target. By contrast, for drugs with a peripheral target, little or no BBB penetration might be required in order to avoid CNS side effects. A key issue in the development of models to predict BBB penetration is the use of appropriate data to describe brain uptake of compounds. There is an ongoing discussion about the use of total-brain data versus extracellular fluid (ECF) or cerebro-spinal fluid (CSF) data or data generated by microdialysis<sup>52</sup>. Another point of debate relates to the time point of measurement, which is clearly crucial. Overall, data in the literature are rather limited in number, and are also generated from different experimental protocols. All of these factors limit the development of highly predictive models of BBB penetration.

Nevertheless, a variety of models for the prediction of uptake into the brain have been developed<sup>53–59</sup>. ‘Rule-of-five’-like recommendations regarding the molecular parameters that contribute to the ability of molecules to cross the BBB have been made to aid BBB-penetration predictions<sup>53</sup>; for example, molecules with a molecular mass of <450 Da or with PSA <100 Å<sup>2</sup> are more likely to penetrate the BBB. Most of the early predictive models are based on a multiple linear regression approach and many use physicochemical properties<sup>60</sup>. One example of such a model is based on the combination of only three descriptors, namely the calculated octanol/water partition coefficient, the number of hydrogen-bond acceptors in an aqueous medium and the polar surface area<sup>55</sup>. More recently, other multivariate techniques have been tried using new ADME-tailored properties, such as the Volsurf approach, in which a variety of three-dimensional molecular field descriptors are transformed into a new set of descriptors, which are inputs for the construction of a model using a discriminant partial least squares procedure<sup>56,57</sup>. As this method is based on computed properties only, it can be used as a tool in virtual screening.

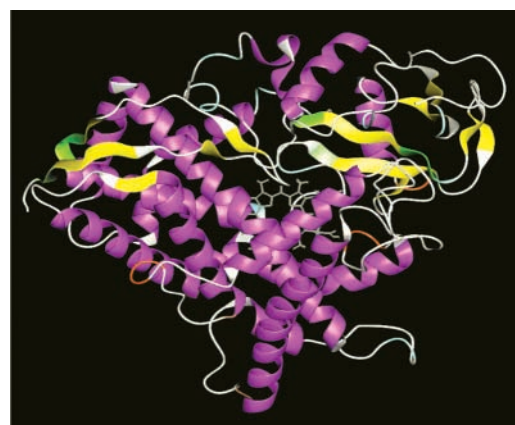


Figure 2 | **Model of the CYP2D6 metabolizing enzyme**<sup>87</sup>. This shows the secondary and tertiary structure of the enzyme.

**Transporters.** Transport proteins are found in most organs involved in the uptake and elimination of endogenous compounds and xenobiotics, including drugs<sup>61</sup>. As mentioned above, a better understanding of the role of transporters in oral absorption and uptake in the brain<sup>62</sup> and liver is of particular interest. Consequently, several *in vitro* systems, some with double-transfected transporters, are being developed and these might become a valuable tool in screening for optimal pharmacokinetic properties.

One of the best-studied transporters is P-gp, a member of the ATP-binding cassette (ABC) transporter family that was identified first as the transporter responsible for multiple-drug resistance (MDR) observed with anti-tumour agents. A better understanding of the relationships between the structure of P-gp binders (substrate or inhibitor) has been obtained using QSAR, as well as from pharmacophore and protein modelling. Such models for P-gp function have recently been reviewed<sup>63</sup>. Although this paper focused on MDR reversers, the QSAR studies discussed demonstrate the first steps towards a better understanding of P-gp SAR. A crude filter to discriminate between P-gp substrates and non-substrates — with an accuracy of 63% — has been suggested<sup>64</sup>. However, even for large virtual combinatorial libraries, this does not seem good enough, as it is too close to random.

A set of well-defined structural elements required for interaction with P-gp has been derived from the analysis of a set of known P-gp substrates<sup>65–67</sup>. The key recognition elements in this model are two or three electron-donor groups (hydrogen-bond acceptors) with a fixed spatial separation. However, this preliminary model does not take into account the directionality of the hydrogen bonds, or the conformational flexibility of certain compounds. Models are now sufficiently sophisticated to begin to rationalize earlier observations for well-known P-gp substrates in terms of molecular weight, lipophilicity, hydrogen bonding, presence of a basic nitrogen and so on. The program MolSurf has been used to generate descriptors to build a PLS model to predict P-gp-associated ATPase activity. This model identified the main contributing

descriptors for predicting ATPase activity as the size of the molecular surface, polarizability and hydrogen-bonding potential<sup>68</sup>.

Some initial attempts have also been made to undertake P-gp modelling<sup>69</sup>. Using the primary sequence of human P-gp and a low-resolution structure, a pseudo-receptor has been constructed and attempts have been made to model its interaction with MDR modulators. A P-gp pharmacophore model consisting of two hydrophobic points, three hydrogen-bond-acceptor points and one donor point was reported<sup>70</sup>. In another approach, a

THREE-DIMENSIONAL-QSAR P-gp model was generated using the Catalyst program<sup>71</sup>. This model allows qualitative rank-order and predicts IC<sub>50</sub> values for P-gp inhibitors.

Other transporters potentially involved in limiting the oral uptake of drugs include the MDR-associated proteins MRP1 and MRP2, and the recently discovered breast-cancer-resistance protein (BCRP). It is important to expand our knowledge about these transporters, with a view to developing better tools for the prediction of oral uptake. Other transporters could also have a key role in hepatic uptake<sup>72</sup> and in order to understand their

Table 1 | Sources for commercial ADMET software

Company/Institute	Software product	URL
<b>Biotechs</b>		
Aber Genomic Computing	Gmax-Bio	www.abergc.com
Accelrys (Pharmacopeia)	Cerius2, C2.ADME, Topkat	www.accelrys.com
Advanced Chemistry Development	ACD/logP, ACD/logD, ACD/pKa	www.acdlabs.com
Amedis Pharmaceuticals		www.amedis-pharma.com
Arqule		www.arqule.com/insilico/camitro.html
Biobyte	CLOGP, CQSAR	www.biobyte.com
Bioreason	LeadPharmer	www.bioreason.com
Chemical Computing Group	Molecular Operating Environment (MOE)	www.chemcomp.com
Compudrug	Pallas, MetabolExpert, HazardExpert	www.compudrug.com
Daylight	MedChem db, CLOGP	www.daylight.com
EduSoft	Molconn-Z	www.edusoft-lc.com
Entelos	Physiolab	www.entelos.com
Genomatica		www.genomatica.com
Iconix	DrugMatrix	www.iconixpharm.com
IDBS		www.id-bs.com
Incyte	DrugMatrix	www.incyte.com
LeadScope	LeadScope, ToxScope	www.leadscope.com
Lhasa	DEREK, Meteor	www.chem.leeds.ac.uk/luk
Lion Bioscience	iDEA	www.lionbioscience.com
Logichem	Oncologic	www.logichem.com
MDL Information Systems	MDL QSAR, MDL Discovery Predictive Science, Metabolite db, Toxicity db	www.mdl.com
Molecular Discovery	VolSurf	www.moldiscovery.com
Molecular Networks	KMAP, Petra	www.mol-net.de
Multicase	M-CASE, Meta	www.multicase.com
Omniviz	Omniviz Chemoinformatics	www.omniviz.com
Pharma Algorithms	Advanced QSAR Builder	www.ap-algorithms.com
pION	Absolv, Algorithm Builder	www.pion-inc.com
Schrödinger	QikProp	www.schrodinger.com
SciVision	ToxSys, QSARIS	www.scivision.com
SimCyp	SimCYP	www.simcyp.com
SimulationsPlus	GastroPlus, QMPRPlus	www.simulations-plus.com
Sirius Analytical Instruments	AbSolv	www.sirius-analytical.com
Spotfire	Spotfire	www.spotfire.com
Syracuse Research		www.syrres.com/default.htm
Tripos	VolSurf	www.tripos.com
Umetrics	SIMCA	www.umetrics.com
ZyxBio	OraSpotter	www.zyxbio.com

THREE-DIMENSIONAL-QSAR  
A technique that uses the three-dimensional molecular structures to derive a quantitative relationship between a biological property and properties derived from these three-dimensional structures, for example, related to their size and electrostatic fields.

Table 2 | Sources for commercial ADMET software

Company/Institute	Software product	URL
<b>Data providers</b>		
Cerep	Bioprint	www.cerep.fr
Cyprotex	Cloe PK	www.cyprotex.com
Novascreen	Profile	www.novascreen.com
TNO Pharma		www.pharma.tno.nl/

effects on pharmacokinetics it might be necessary to model them. For example, some modelling work on the peptide transporter (PepT1) and the apical sodium-dependent transporter (ASBT), which might be involved in active drug uptake, has been reported<sup>73</sup>.

**Dermal and ocular penetration.** Although much attention has been given to oral absorption models, some drugs are administered through alternative routes, such as the skin or eye. For many years, QSAR models have been developed to predict the optimal percutaneous penetration<sup>74</sup> (a recent example is given in REF. 75). These models resemble oral absorption and BBB models, and often employ very similar properties and descriptors. The existing transdermal models are typically a function of the octanol/water partition coefficient and terms that have been associated with aqueous solubility, including hydrogen-bonding parameters, molecular weight and molecular flexibility. Commercial models for the prediction of solute-permeation rates through the skin are available, for example, the QikProp and DermWin programs. However, it seems that there is little difference between the commercially available models and models published in the literature. Most, if not all, of the published skin-permeation models have been constructed from various compilations of published skin-permeation data sets.

**Plasma-protein binding.** It is generally assumed that only free drug can cross membranes and bind to the intended molecular target<sup>76</sup>, and it is therefore important to estimate the fraction of drug bound to plasma proteins. Drugs can bind to a variety of particles in the blood, including red blood cells, leukocytes and platelets, in addition to proteins such as albumin (particularly acidic drugs),  $\alpha_1$ -acid glycoproteins (basic drugs), lipoproteins (neutral and basic drugs), erythrocytes and  $\alpha, \beta, \gamma$ -globulins.

A tentative sigmoidal relationship is observed between plasma-protein binding (in percentage of drug bound or unbound) and log D at pH 7.4 (REF. 23). As there is a considerable scatter of the data around these sigmoidal trends, adding further descriptors might lead to better predictive models. One attempt in this direction is based on 107 descriptors and uses a technique called genetic function approximation (GFA)<sup>77</sup>. For a set of 80 compounds, a QSAR with 12 descriptors and a correlation coefficient  $r = 0.91$  between measured and predicted serum-albumin binding data was obtained<sup>77</sup>. Using the multiple computer-automated structure evaluation (M-CASE) program and protein-affinity data for

154 drugs, models were generated that correctly predicted the percentage of drug bound in plasma for ~80% of the test compounds with an average error of ~14% (REF. 78). A generic model to predict drug-association constants to human serum albumin (HSA) using a pharmacophoric-similarity concept and PLS was reported using a data set of 138 compounds<sup>79</sup>.

**Volume of distribution.** The volume of distribution, together with the clearance rate, determine the half-life of a drug and therefore its dose regimen, and so the early prediction of both properties would be of great benefit. When the logarithm of the volume of distribution is plotted against log D, a scatter plot is obtained and no correlation is observed. However, when these data are corrected for plasma-protein binding, the resulting plot of the logarithm of the unbound volume of distribution (log V<sub>du</sub>; BOX 1) against log D reveals a clear linear trend, with log V<sub>du</sub> increasing at higher lipophilicities<sup>23</sup>. This can be used as a simple guide in modifying and optimizing the V<sub>du</sub>.

Recently, an approach for predicting volume of distribution values has been presented that used experimental distribution coefficients at pH 7.4 in octanol/water, the ionization constant (pK<sub>a</sub>) of the compounds and measured plasma-protein-binding data<sup>80</sup>. In principle, this approach could be fully computational, as predictive models are available for log P and pK<sub>a</sub>, and models for plasma-protein binding are under development, as described above. Several groups are exploring such fully computational models.

**Clearance.** Clearance is an important pharmacokinetic parameter that defines, together with the volume of distribution, the half-life, and thus the frequency of dosing, of a drug. For a series of adenosine A<sub>1</sub> receptor agonists, not only their clearance, but also their volume of distribution and protein binding, could be predicted using the multivariate PLS technique<sup>81</sup>. As pointed out by the authors, further improvements might be obtained using nonlinear models, such as neural networks, although the application of neural networks is still a relatively new data-modelling method in the field of pharmacokinetics.

It was concluded from an exploratory study using neural networks in addition to multivariate techniques that human hepatic drug clearance was best predicted from human hepatocyte data, followed by rat hepatocyte data. In the studied data set, however, animal *in vivo* data did not significantly contribute to the predictions<sup>82</sup>. However, only a rather limited data set was used in this study, and generalizations from these results should be made with caution at this stage. This study also demonstrates that computational predictions can be successful if the models can use experimental data as part of their input. Obviously, however, these models can then only be used at stages in the drug discovery process where such experimental data are being generated.

**Half-life.** The half-life of a drug is a hybrid concept that involves clearance and volume of distribution, and it is arguably more appropriate to have reliable estimates of



Table 3 | Sources for commercial ADMET software

Company/Institute	Software product	URL
<b>Academics/consultants</b>		
M. G. Ford (Centre for Molecular Design, University of Portsmouth, UK)	Paragon	www.cmd.port.ac.uk/cmd/software.shtml
Hall Associates Consulting	Molconn-Z	
J. McFarland	Hybot, Slipper	reckon.dat@attglobal.net
V. Poroikov (Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, Moscow, Russia)	PASS	www.timtec.net/software/pass.htm
O. A. Raevsky (Department of Computer-Aided Molecular Design, Russian Academy of Sciences, Moscow, Russia)	Hybot, Slipper	www.ibmh.msk.su/molpro/
P. Sjöberg	MolSurf	qemist@swipnet.se
University of Washington	Drug Interaction Database	depts.washington.edu/didbase/

these two properties instead. Nevertheless, neural networks have been used to predict drug half-life values of antihistamines<sup>83</sup>. Unfortunately, the method relied on topographical coding of the molecule using an in-house program, and also involved a number of strongly inter-correlated calculated properties such as log P, pKa, molecular mass and molar refractivity.

#### *Physiologically based pharmacokinetic modelling.*

There are several approaches to pharmacokinetic modelling, including empirical, compartmental, clearance-based and physiological models. In the latter, full physiological models of blood flow to and from all organs and tissues in the body are considered. Such physiologically based pharmacokinetic (PBPK) models can be used to study concentration–time profiles in individual organs and in the plasma<sup>84,85</sup>. In the future, we expect that PBPK models will be linked to absorption modelling, as discussed above, and the first examples of this type of linkage are Cyprotex' Cloe PK and Simulations Plus' GastroPlus.

Pharmacokinetic/pharmacodynamic (PK/PD)-modelling links dose–concentration relationships (PK) and concentration–effect relationships (PD). This approach facilitates the description and prediction of the time course of drug effects resulting from a certain dose regimen. Further progress in understanding PK/PD relationships and the availability of specialized software not further discussed here, in combination with advanced PBPK modelling, will greatly enhance our capability to perform reliable PK predictions in humans. The linkage of absorption simulation and PBPK models will bring us closer to a full simulation of drug disposition, one that will hopefully be based on only a few properties that can be readily measured *in vitro* and/or computed.

**Metabolism.** Several aspects of metabolism are relevant to drug discovery, including the rate and extent of metabolism (turnover), the enzymes involved and the products formed, each of which can give rise to different concerns. The extent and rate of metabolism affect clearance, whereas the involvement of particular enzymes might

lead to issues related to the polymorphic nature of some of these enzymes and to drug–drug interactions.

QSAR and molecular modelling approaches for predicting metabolism could have an increasingly important role as a possible alternative to *in vitro* metabolism studies. *In silico* approaches to predicting metabolism can be divided into QSAR and three-dimensional-QSAR<sup>86</sup> studies, protein and pharmacophore models<sup>87,88</sup> and predictive databases. Some of the first-generation predictive-metabolism tools currently require considerable input from a computational chemist, whereas others can be used as rapid filters for the screening of virtual libraries, for example, to test for CYP3A4 liability<sup>89</sup>.

Perhaps the most intellectually satisfying molecular modelling studies are those based on the crystal structure of the metabolizing enzymes (FIG. 2). Historically, these structure-based models have relied on crystal-structure information from bacterial homologues<sup>87,88</sup>. However, the crystal structures of the more relevant mammalian cytochrome P450s have been announced (CYP3A4 and CYP2C9) and the structure of CYP2C8 is publicly available.

Early predictions of the vulnerability to metabolism of certain positions in the molecule might help to eliminate metabolic liabilities. An example of such an approach is the use of reaction energetics to develop a predictive CYP3A4 metabolism model<sup>90</sup>. Another program, called MetaSite (Cruciani, G.; abstract presented at Euro QSAR 2002), is based on a pharmacophore representation obtained from interaction fields for the protein structure and a pharmacophoric fingerprint for the potential substrate.

Several approaches that use databases to predict metabolism are available or under development<sup>91</sup>, including expert systems, such as MetabolExpert (Compudrug), META (MultiCASE) or Meteor (Lhasa), and the databases Metabolite (MDL) and Metabolism (Accelrys)<sup>92</sup>. Ultimately, such programs might be linked to computer-aided toxicity prediction on the basis of quantitative structure–toxicity relationships and expert systems for toxicity evaluation such as DEREK (Lhasa) and MultiCase.

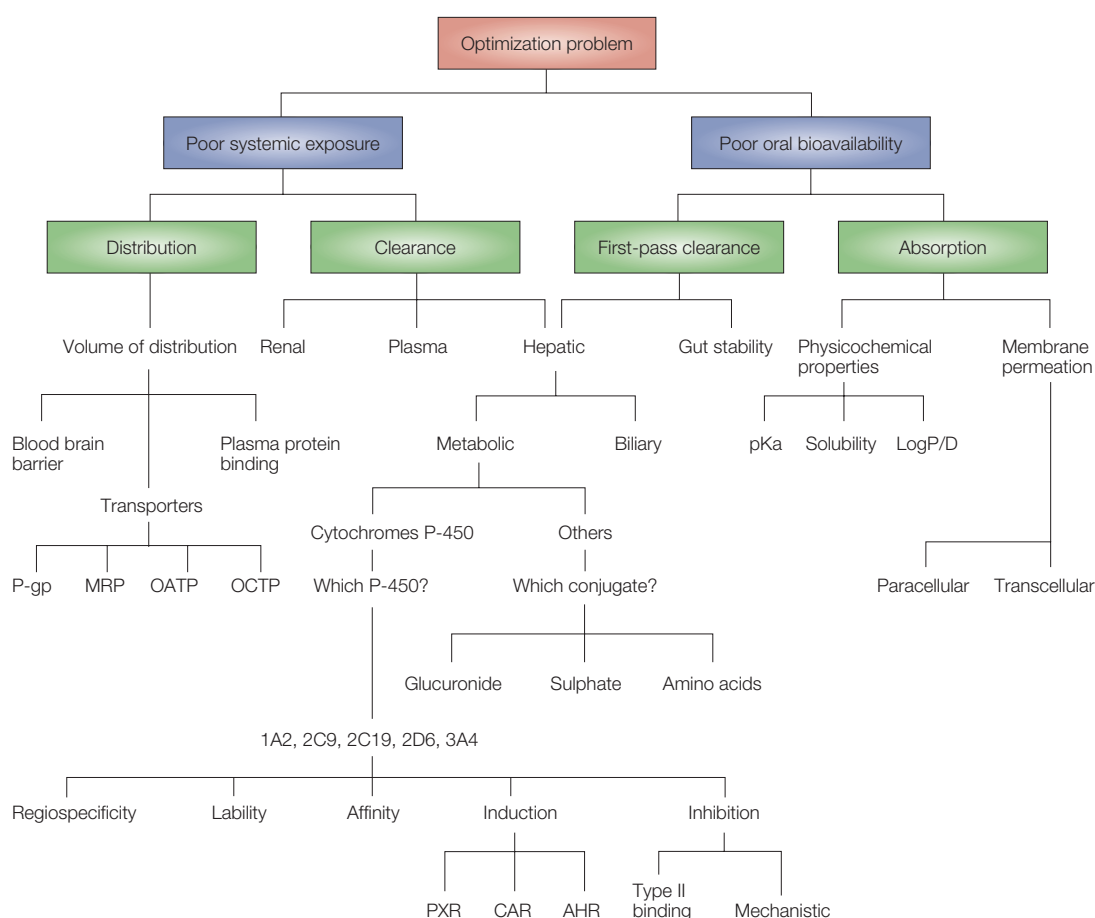


Figure 3 | **An analysis of the crucial ADME processes for which predictive models are available or are being developed**<sup>11</sup>. This figure does not suggest a logical flow in ADME studies, but rather tries to group the problem areas for which predictive models could be of help.

### ***In silico* prediction of toxicity issues**

Toxicity is responsible for many compounds failing to reach the market and for the withdrawal of a significant number of compounds from the market once they have been approved. It has been estimated that ~20–40% of drug failures in investigational drug development can be attributed to toxicity concerns (for example, as shown in FIG. 1)<sup>1</sup>.

The existing commercially available *in silico* tools for forecasting potential toxicity issues can be roughly classified into two groups. The first approach uses expert systems that derive models on the basis of abstracting and codifying knowledge from human experts and the scientific literature. The second approach relies primarily on the generation of descriptors of chemical structure and statistical analysis of the relationships between these descriptors and the toxicological end-point. Recent reviews have compared and contrasted the commercially available *in silico* toxicology software<sup>93–95</sup>. As has been previously mentioned, the most important part of any *in silico* approach is the quality of the underlying data used to develop the models (BOX 2). The limited availability of public-domain toxicology data has limited the number of toxicological end-points forecasted by the commercially available

systems. The primary emphasis of the current software packages is carcinogenicity<sup>93</sup> and mutagenicity, although some packages do also include models and/or knowledge bases for other end-points, such as teratogenicity, irritation, sensitization, immunotoxicology and neurotoxicity. There is currently an unmet need for *in silico* predictive toxicology software<sup>94,95</sup> for other end-points important in drug development, such as QT prolongation<sup>96,97</sup>, hepatotoxicity and phospholipidosis<sup>98</sup>.

**Drug–drug interactions.** Patients often receive several medications at the same time, and if the drugs involved compete for the same enzymes to be metabolized, or if the same transporters are involved in transporting the drugs across membranes, this can lead to undesired effects with possibly even fatal results. Therefore, during drug development, new chemical entities intended for use as a new drug are now often screened *in vitro* for potential drug–drug interactions.

The quantitative prediction of such interactions has been attempted in a system called Q-DIPS (quantitative drug interactions prediction system)<sup>99</sup>. Another approach is the SimCYP project at the University of Sheffield, which integrates human physiological, anatomical and genetic information with human *in vitro* data, and which can

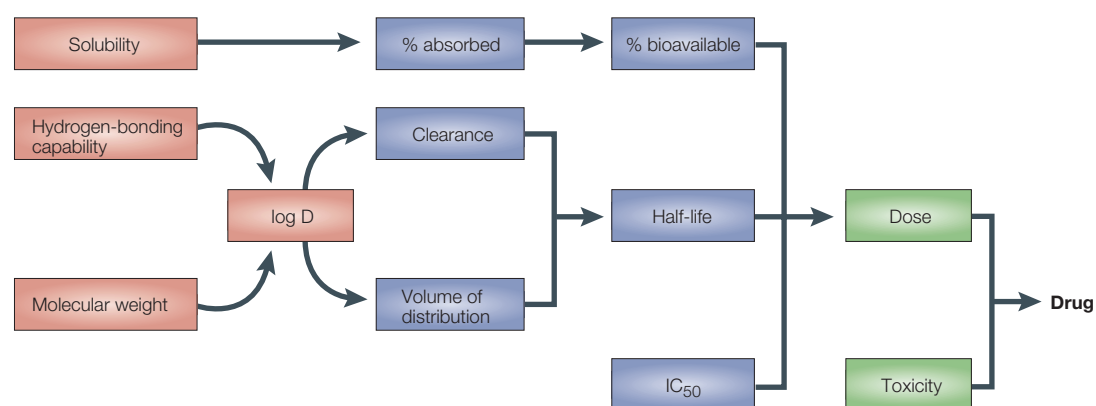


Figure 4 | **Towards prediction paradise.** As more and more robust models for the crucial endpoints in the drug discovery process become available, we will increasingly be in a position to map out the potential qualities of a new chemical purely from its molecular structure and appropriate descriptors using a suite of predictive models. These range from models for simple physicochemical properties, such as hydrogen bonding-capability, molecular mass, solubility and lipophilicity (log D), to models for ADME properties, such as percentage drug absorbed and bioavailable, clearance, volume of distribution and half-life, to complex endpoints, such as the binding ( $IC_{50}$ ) to the molecular target of the new drug, its required dose and toxicity potential.

simulate PK data for a population of individuals. Published data on drug–drug interactions are now available in a database at the University of Washington, which will facilitate future model development.

**Induction of drug metabolism.** A further concern in drug metabolism related to drug–drug interactions is the induction of drug metabolism caused by some drugs. Recently, two related nuclear hormone receptors — the pregnane X receptor (PXR) and the constitutive androstane receptor (CAR) — have emerged as transcriptional regulators of cytochrome P450 expression<sup>100</sup>. Drugs that bind to these receptors can induce the expression of cytochrome P450s, thereby accelerating the metabolism of other drugs that are substrates for these enzymes. In particular, PXR is a key regulator of the inducible expression of CYP3A, which metabolizes 50–60% of all prescription drugs, and therefore methods to identify compounds that will not activate PXR could be valuable in avoiding drug–drug interactions. At present, such approaches to predict the induction of drug metabolism are in an early phase of development.

### Outlook

**Commercial software.** Software vendors traditionally active in the field of molecular modelling and/or QSAR have also recently begun to create modules to assist high-throughput screening and combinatorial library design, as well as the estimation of ADME and toxicity properties<sup>101</sup>. The key players developing this software are listed in TABLES 1, 2 and 3. This review does not intend to describe nor evaluate each of the individual products. Products and vendors change rapidly and the interested reader is encouraged to obtain current product information from the vendor's websites, specialized conferences or recent reviews<sup>94</sup>.

**How far are we from prediction paradise?** Computational chemists are now using ADMET filters in the very early stages of drug discovery, for example, in library design and virtual screening. The first generation of predictive ADMET models are now commercially available and others have been published and can now be implemented<sup>2,102</sup>. In the short term, these tools should allow chemists and drug-metabolism scientists to concentrate on compounds with the highest chances of meeting the required pharmacokinetic and safety criteria, and should contribute to a reduction in late-stage compound attrition.

However, the models are clearly only as good as the data they are based on, and, unfortunately, in most cases, the data sets are rather limited. It is worth noting that even though the database for log P prediction has more than 10,000 compounds, the predictions derived from these data are far from perfect. This is because of the many innovative chemical groups that appear in modern drug-discovery programmes that are not part of the legacy database used to derive the model(s). So, the learning/modelling will need to be an ongoing, iterative process in which the models are continuously refined. Driven by the changes in the working paradigm in the pharmaceutical and

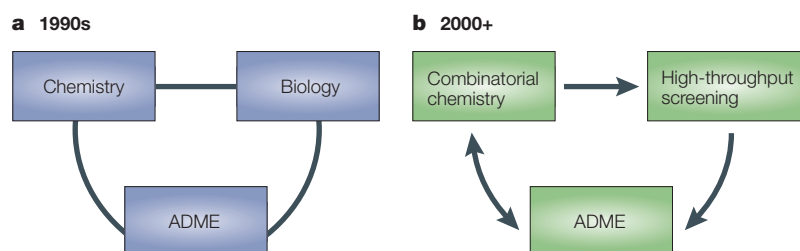


Figure 5 | **The evolution of drug discovery and the changing role of ADME studies.** The transition from **a** | the classical project-collaboration approach between chemistry, biology and drug metabolism (ADME) groups in the 1990s to **b** | a much more automated world at the start of this millennium in which combinatorial chemistry (CombiChem), high-throughput screening and ADME studies are linked together in a streamlined fashion. Note that these three activities can even be carried out by separate companies. Furthermore, the wide introduction of *in silico* and high-speed *in vitro* methods could redefine the traditional meaning of ADME to Automated Decision-Making Engine.

biotechnology industry, *in silico* approaches will inevitably find their place. As expressed recently, “the insilicoids are coming and will save the world”<sup>103</sup>.

During the next few years, the range of models will further expand to include, for example, metabolism by non-P450 enzymes, models for various transporters, predictors for volume of distribution, plasma protein binding, and so on (FIG. 3). The ability to continuously adapt and refine the existing models by building on larger and higher-quality data sets will be crucial to the success of the *in silico* approaches. FIGURE 4 outlines the key parameters in the prediction of a safe drug given in an acceptable dose, which it is ultimately hoped will be reliably obtainable from molecular structure and appropriate descriptors using a suite of predictive models.

Today, most models are rule-based and may use descriptors that are not easily understood by the chemist and not easy to translate into better molecular structures. Clearly, there is a need for a new generation of mechanism-based models that will provide the required understanding and which can be successfully used for prediction and simulation of ADMET properties. Such second-generation predictive models could be combinations of models (that is, meta-models) for the partial processes; for example, oral absorption

would be a combination of models for log D, solubility, permeability, CYP3A4 metabolism and efflux and influx mediated by transporters such as P-gp.

A recent report on *in silico* technology estimates that by 2006 ~10% of pharmaceutical R&D expenditure will be on computer simulation and modelling, a figure set to rise to 20% by 2016. It seems clear that as a whole, the pharmaceutical R&D landscape will further change<sup>104</sup>, and that computational ADMET will be part of this change.

In the next 10 years or so, building on what is already starting in several companies, the degree of automation in traditional drug metabolism departments will continue to increase, and fully automated medium- and high-throughput *in vitro* assays will be used alongside *in silico* modelling and data interpretation. Whereas ADME today stands for absorption, distribution, metabolism and excretion, in the future we might instead speak of the ‘automated decision-making engine’ (FIG. 5). There could well be two types of ADME technology in the future: the early discovery paradigm (based on *in vitro* and *in silico* approaches) and the regulatory one (close to today’s approach). In any case, there is much basic science that needs to be done first, making this an especially exciting time to be involved in ADMET prediction and drug discovery.

- Kennedy, T. Managing the drug discovery/development interface. *Drug Disc. Today* **2**, 436–444 (1997).
- Van de Waterbeemd, H. High-throughput and *in silico* techniques in drug metabolism and pharmacokinetics. *Curr. Opin. Drug Disc. Dev.* **5**, 33–43 (2002).
- Sadowski, J. & Kubinyi, H. A scoring scheme for discriminating between drugs and nondrugs. *J. Med. Chem.* **41**, 3325–3329 (1998).
- Anzali, S., Bamickel, G., Cezanne, B., Krug, M., Filimonov, D. & Poroikov, V. Discriminating between drugs and nondrugs by prediction of activity spectra for substances (PASS). *J. Med. Chem.* **44**, 2432–2437 (2001).
- Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug. Del. Revs.* **23**, 3–25 (1997).
- Paper introducing Lipinski’s rule-of-5.**
- Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **44**, 235–249 (2001).
- Van de Waterbeemd, H., Smith, D. A., Beaumont, K. & Walker, D. K. Property-based design: Optimisation of drug absorption and pharmacokinetics. *J. Med. Chem.* **44**, 1313–1333 (2001).
- Johnson, M. A. & Maggiora, G. M. *Concepts and Applications of Molecular Similarity Analysis* (Wiley Interscience, New York, 1990).
- Raevsky, O. A., Trepalin, S. V., Trepalina, H. P., Gerasimenko, V. A. & Raevskaja, O. E. SLIPPER-2001 – Software for predicting molecular properties on the basis of physicochemical descriptors and structural similarity. *J. Chem. Inf. Comput. Sci.* **42**, 540–549 (2002)
- Janssen, D. The power of prediction. *Drug Disc.* 38–40 (January 2002).
- Carlson, T. J. & Segall, M. D. Predictive, computational models of ADME properties. *Curr. Drug Disc.* 34–36 (March 2002)
- Podlogar, B. L., Muegge, I. & Brice, L. J. Computational methods to estimate drug development parameters. *Curr. Opin. Drug Disc. Dev.* **4**, 102–109 (2001).
- Ekins, S., Waller, C. L., Swaan, P. W., Cruciani, G., Wrighton, S. A. & Wikel, J. H. Progress in predicting human ADME parameters *in silico*. *J. Pharmacol. Toxicol. Methods* **44**, 251–272 (2001).
- Excellent summary of *in silico* ADME and tables with references to data sets for permeability/absorption, brain penetration, P-gp, clearance and bioavailability.**
- Ekins, S. & Wrighton, S. A. Application of *in silico* approaches to predicting drug–drug interactions. *J. Pharmacol. Toxicol. Methods* **45**, 65–69 (2001).
- Rodrigues, A. D., Winchell, G. A. & Dobrinska, M. R. Use of *in vitro* drug metabolism data to evaluate metabolic drug–drug interactions in man: the need for quantitative databases. *J. Clin. Pharmacol.* **41**, 368–373 (2001)
- Bachman, K. A. & Ghosh, R. The use of *in vitro* methods to predict *in vivo* pharmacokinetics and drug interactions. *Curr. Drug Metab.* **2**, 299–314 (2001).
- Beresford, A. P., Selick, H. E. & Tarbit, M. H. The emerging importance of predictive ADME simulation in drug discovery. *Drug Disc. Today* **7**, 109–116 (2002).
- Guner, O. (ed). *Pharmacophore Perception, Development and Use in Drug Design* (IUL Biotechnology Series, 2000)
- Van de Waterbeemd, H. & Rose, S. In *The Practice of Medicinal Chemistry* 2nd (ed Wermuth, L. G.) 1367–1385 (Academic Press, 2003).
- Todeschini, R. & Consonni, V. *Handbook of Molecular Descriptors* (Wiley–VCH, Weinheim, 2000).
- Cruciani, G., Crivori, P., Carrupt, P. A. & Testa, B. Molecular fields in quantitative structure–permeation relationships: The VolSurf approach. *Theochem.* **503**, 17–30 (2000)
- Buchwald, P. & Bodor, N. Computer-aided drug design: the role of quantitative structure–property, structure–activity and structure–metabolism relationships (QSPR, QSAR, QSMR). *Drug Future* **27**, 577–588 (2002).
- Van de Waterbeemd, H., Smith, D. A. & Jones, B. C. Lipophilicity in PK design: methyl, ethyl, futile. *J. Comput. Aid. Mol. Des.* **15**, 273–286 (2001).
- Van de Waterbeemd, H. In *Pharmacokinetic Challenges in Drug Discovery* (Eds) 213–234 (Ernst-Schering Research Foundation Workshop Series No. 37, Springer, 2001).
- Walther, B., Vis, P. & Taylor, A. In: *Lipophilicity in Drug Action and Toxicology* (eds Pliska, V., Testa, B., Van de Waterbeemd, H.) 253–261 (VCH, Weinheim, 1996).
- Bevan, C. D. & Lloyd, R. S. A high-throughput screening method for the determination of aqueous drug solubility using laser nephelometry in microtiter plates. *Anal. Chem.* **72**, 1781–1787 (2000).
- Jorgensen, W. L. & Duffy, E. M. Prediction of drug solubility from structure. *Adv. Drug Del. Rev.* **54**, 355–366 (2002).
- Bergstrom, C. A. S., Norinder, U., Luthman, K. & Artursson, P. Experimental and computational screening models for prediction of aqueous drug solubility. *Pharm. Res.* **19**, 182–188 (2002).
- Hilal, S. H., Karickhoff, S. W. & Carreira, L. A. A rigorous test for SPARC’s chemical reactivity models: estimation of more than 4300 ionization pKas. *Quant. Struct.-Act. Relat.* **14**, 348–355 (1995).
- Raevsky, O. A., Fetisov, V. I., Trepalina, E. P., McFarland, J. W. & Schaper, K.-J. Quantitative estimation of drug absorption in humans for passively transported compounds on the basis of their physico-chemical parameters. *Quant. Struct.-Act. Relat.* **19**, 366–374 (2000).
- Stenberg, P., Norinder, U., Luthman, K. & Artursson, P. Experimental and computational screening models for the prediction of intestinal drug absorption. *J. Med. Chem.* **44**, 1927–1937 (2001).
- Ertl, P., Rohde, B. & Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* **43**, 3714–3717 (2000).
- Kulkarni, A., Yi, H. & Hopfinger, A. J. Predicting Caco-2 cell permeation coefficients of organic molecules using membrane-interaction QSAR analysis. *J. Chem. Inf. Comput. Sci.* **42**, 331–342 (2002).
- Cummins, C. L., Jacobsen, W. & Benet, L. Z. Unmasking the dynamic interplay between intestinal P-glycoprotein and CYP3A4. *J. Pharmacol. Exp. Ther.* **300**, 1036–1045 (2002).
- Gumbleton, M. & Audus, K. L. Progress and limitations in the use of *in vitro* cell cultures to serve as a permeability screen for the blood-brain barrier. *J. Pharm. Sci.* **90**, 1681–1698 (2001).
- Zhao, Y. H., Le, J., Abraham, M. H., Hersey, A., Eddershaw, P. J., Luscombe, C. N., Boutina, D., Beck, G., Sherborne, B., Cooper, I. & Platts, J. A. Evaluation of human intestinal absorption data and subsequent derivation of a quantitative structure–activity relationship (QSAR) with the Abraham descriptors. *J. Pharm. Sci.* **90**, 749–784 (2001).
- Key reference for oral absorption data of 169 compounds.**
- Van de Waterbeemd, H. Quantitative structure–absorption relationships. In: *Pharmacokinetic Optimization in Drug Research: Biological, Physicochemical and Computational Strategies*, Testa, B., Van de Waterbeemd, H., Folkers, G. & Guy, R. (Eds), Verlag HCA, Zurich (2001), pp. 499–511.
- Norinder, U. & Osterberg, T. Theoretical calculation and prediction of drug transport processes using simple parameters and partial least squares projections to latent structures (PLS) statistics. The use of electrotopological state indices. *J. Pharm. Sci.* **90**, 1076–1085 (2001).
- Yu, L. X., Gatlin, L. & Amidon, G. L. Predicting oral drug absorption in humans. *Drugs Pharm. Sci.* **102** (Transport Processes in Pharmaceutical Systems), 377–409 (2000).

40. Ho, N. F. H., Raub, T. J., Burton, P. S., Barsuhn, C. L., Adson, A., Audus, K. L. & Borchardt, R. T. Quantitative approaches to delineate passive transport mechanisms in cell culture monolayers. *Drugs Pharm. Sci.* **102** (Transport Processes in Pharmaceutical Systems), 219–316 (2000).
41. Agatonovic-Kustrin, S., Beresford, R. & Yusuf, A. P. M. Theoretically-derived molecular descriptors important in human intestinal absorption. *J. Pharm. Biomed. Anal.* **25**, 227–237 (2001).
42. Fu, X. C., Liang, W. Q. & Yu, Q. S. Correlation of drug absorption with molecular charge distribution. *Pharmazie* **56**, 267–268 (2001).
43. Agoram, B., Wolosz, W. S. & Bolger, M. B. Predicting the impact of physiological and biochemical processes on oral drug bioavailability. *Adv. Drug Del. Rev.* **90**, S41–S67 (2001).
44. Norris, D. A., Leesman, G. D., Sinko, P. J. & Grass, G. M. Development of predictive pharmacokinetic simulation models for drug discovery. *J. Contr. Rel.* **65**, 55–62 (2000).
45. Parrott, N. & Lavé, T. Prediction of intestinal absorption: comparative assessment of GastroPlus and IDEA. *Eur. J. Pharm. Sci.* **17**, 51–61 (2002).
46. Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W. & Kopple, K. D. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **45**, 2615–2623 (2002).
47. Yoshida, F. & Topliss, J. G. QSAR model for drug human oral bioavailability. *J. Med. Chem.* **43**, 2575–2585 (2000).  
**Good reference for data on bioavailability of more than 200 compounds.**
48. Andrews, C. W., Bennett, L. & Yu, L. X. Predicting human oral bioavailability of a compound: development of a novel quantitative structure-bioavailability relationship. *Pharm. Res.* **17**, 639–644 (2000).
49. Bains, W., Gilbert, R., Sviridenko, L., Gascon, J.-M., Scoffin, R., Birchall, K., Harvey, I. & Caldwell, J. Evolutionary computational methods to predict oral bioavailability QSPRs. *Curr. Opin. Drug Disc. Dev.* **5**, 44–51 (2002).
50. Pintore, M., Van de Waterbeemd, H., Piclin, N. & Chrétien, J. R., Prediction of oral bioavailability by adaptive fuzzy partitioning. *Eur. J. Med. Chem.* (in the press).
51. Mandagere, A. K., Thompson, T. N. & Hwang, K. K. A graphical model for estimating oral bioavailability of drugs in humans and other species from their Caco-2 permeability and *in vitro* liver enzyme metabolic stability rates. *J. Med. Chem.* **45**, 304–311 (2002).
52. De Lange, E. C. M. & Danhof, M. Considerations in the use of cerebrospinal fluid pharmacokinetics to predict brain target concentrations in the clinical setting. *Clin. Pharmacokinet.* **41**, 691–703 (2002).
53. Van de Waterbeemd, H., Camenisch, G., Folkers, G., Chrétien, J. R. & Raevsky O. A. Estimation of blood-brain barrier crossing of drugs using molecular size and shape, and H-bonding descriptors. *J. Drug Target.* **6**, 151–165 (1998).  
**Discussion of the critical physicochemical properties required for brain penetration.**
54. Clark, D. E. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 2. Prediction of blood-brain barrier penetration. *J. Pharm. Sci.* **88**, 815–821 (1999).
55. Feher, M., Sourial, E. & Schmidt, J. M. A simple model for the prediction of blood-brain partitioning. *Int. J. Pharmaceut.* **201**, 239–247 (2000).
56. Crivori, P., Cruciani, G., Carrupt, P. A. & Testa, B. Predicting blood-brain barrier permeation from three-dimensional molecular structure. *J. Med. Chem.* **43**, 2204–2216 (2000).
57. Ooms, F., Weber, P., Carrupt, P. A. & Testa, B. A simple model to predict blood-brain barrier permeation from 3D molecular fields. *Biochim. Biophys. Acta* **1587**, 118–125 (2002).
58. Kaznessis, Y. N., Snow, M. E. & Blankley, C. J. Prediction of blood-brain partitioning using Monte Carlo simulations of molecules in water. *J. Comput. Aid. Mol. Des.* **15**, 697–708 (2001).
59. Rose, K., Hall, L. H. & Kier, L. B. Modeling blood-brain barrier partitioning using the electrotopological state. *J. Chem. Inf. Comput. Sci.* **42**, 651–666 (2002).
60. Abraham, M. H. & Platts, J. A. Physicochemical factors that influence brain uptake. *Blood-Brain Barrier Drug Delivery CNS*, 9–32 (2000).
61. Ayrton, A. & Morgan, P. The role of transport proteins in drug absorption, distribution and excretion. *Xenobiotica* **31**, 469–497 (2001).
62. Van Asperen, J., Mayer, U., Van Tellingen, O. & Beijnen, J. H. The functional role of P-glycoprotein in the blood-brain barrier. *J. Pharm. Sci.* **86**, 881–884 (1997).
63. Wiese, M. & Pajeva, I. K. Structure-activity relationships of multidrug resistance reversers. *Curr. Med. Chem.* **8**, 685–713 (2001).
64. Penzotti, J. E., Lamb, M. L., Evensen, E. & Grootenhuys, P. D. J. A computational ensemble pharmacophore model for identifying substrates of P-glycoprotein. *J. Med. Chem.* **45**, 1737–1740 (2002).
65. Seelig, A. A general pattern for substrate recognition by P-glycoprotein. *Eur. J. Biochem.* **251**, 252–261 (1998).
66. Seelig, A. & Landwojtowicz, E. Structure-activity relationship of P-glycoprotein substrates and modifiers. *Eur. J. Pharm. Sci.* **12**, 31–40 (2000).
67. Seelig, A., Blatter, X. L. & Wohnsland, F. Substrate recognition by p-glycoprotein and the multidrug resistance-associated protein MRP1: a comparison. *Int. J. Clin. Pharmacol. Ther.* **38**, 111–121 (2000).
68. Österberg, Th. & Norinder, U. Theoretical calculation and prediction of P-glycoprotein-interacting drugs using MolSurf parametrization and PLS statistics. *Eur. J. Pharm. Sci.* **10**, 295–303 (2000).
69. Pajeva, I. K. & Wiese, M. Human P-glycoprotein pseudoreceptor modeling: 3D-QSAR study on thioxanthene type multidrug resistance modulators. *Quant. Struct. Act. Relat.* **20**, 130–138 (2001).
70. Pajeva, I. K. & Wiese, M. Pharmacophore model of drugs involved in P-glycoprotein multidrug resistance: Explanation of structural variety (Hypothesis). *J. Med. Chem.* **45**, 5671–5686 (2002).
71. Ekins, S., Kim, R. B., Leake, B. F., Dantzig, A. H., Schuetz, E. G., Lan, L.-B., Yasuda, K., Shepard, R. L., Winter, M. A., Scheutz, J. D., Wikel, J. H. & Wrighton, S. A. Three-dimensional quantitative structure-activity relationships of inhibitors of P-glycoprotein. *Mol. Pharmacol.* **61**, 964–973 (2002).
72. Goh, L.-B., Spears, K. J., Yao, D., Ayrton, A., Morgan, P., Wolf, C. R. & Friedberg, T. Endogenous drug transporters in *in vitro* and *in vivo* models for the prediction of drug disposition in man. *Biochem. Pharmacol.* **64**, 1569–1578 (2002).
73. Zhang, E. Y., Phelps, M. A., Cheng C., Ekins, S. & Swaan, P. W. Modeling of active transport systems. *Adv. Drug Del. Revs.* **54**, 329–354 (2002).
74. Pugh, W. J., Degim, I. T. & Hadgraft, J. Epidermal permeability-penetrant structure relationships. 4. QSAR of permeant diffusion across human stratum corneum in terms of molecular weight, H-bonding and electronic charge. *Int. J. Pharm.* **197**, 203–211 (2000).
75. Ghafourian, T. & Foadadi, S. The effect of structural QSAR parameters on skin penetration. *Int. J. Pharm.* **217**, 1–11 (2001).
76. Smith, D. A., Van de Waterbeemd, H. & Walker, D. K. *Pharmacokinetics and Metabolism in Drug Design*, (Wiley-VCH, Weinheim, Germany, 2001).
77. Colmenarejo, G., Alvarez-Pedraglio, A. & Lavandera, J.-L. Chemoinformatic models to predict binding affinities to human serum albumin. *J. Med. Chem.* **44**, 4370–4378 (2001).
78. Saikhov, R. D., Stefan, L. R. & Klopman, G. Multiple computer-automated structure evaluation model of the plasma protein binding affinity of diverse drugs. *Perspec. Drug Disc. Des.* **19**, 133–155 (2000).
79. Kratochwil, N. A., Huber, W., Müller, F., Kansy, M. & Gerber, P. Predicting plasmas protein binding of drugs: A new approach. *Biochem. Pharmacol.* **64**, 1355–1374 (2002).
80. Lombardo, F., Obach, R. S., Shalavea, M. Y. & Gao, F. Prediction of volume of distribution values in humans for neutral and basic drugs using physicochemical measurements and plasma protein binding data. *J. Med. Chem.* **45**, 2867–2876 (2002).
81. Van de Graaf, P. H., Nilsson, J., Van Schaick, E. A. & Danhof, M. Multivariate quantitative structure-pharmacokinetic relationships (QSPKR) analysis of adenosine A<sub>1</sub> receptor agonists in rat. *J. Pharm. Sci.* **88**, 306–312 (1999).
82. Schneider, G., Coassolo, P. & Lavé, T. Combining *in vitro* and *in vivo* pharmacokinetic data for prediction of hepatic drug clearance in humans by artificial neural networks and multivariate statistical techniques. *J. Med. Chem.* **42**, 5072–5076 (1999).
83. Quiñones, C., Caceres, J., Stud, M. & Martinez, A. Prediction of drug half-life values of antihistamines based on the CODES/neural network model. *Quant. Struct.-Act. Relat.* **19**, 448–454 (2000).
84. Poulin, P. & Theil, F. P. A priori prediction of tissue: plasma partition coefficients of drugs to facilitate the use of physiologically-based pharmacokinetic models in drug discovery. *J. Pharm. Sci.* **89**, 16–35 (2000).
85. Poulin, P., Schoenlein, K. & Theil, F. P. Prediction of adipose tissue: plasma partition coefficients for structurally unrelated drugs. *J. Pharm. Sci.* **90**, 436–447 (2001).
86. Ekins, S., Bravi, G., Binkley, S., Gillespie, J. S., Ring, B. J., Wikel, J. H. & Wrighton, S. A. Three- and four-dimensional quantitative structure-activity relationship (3D/4D-QSAR) analyses of CYP2C9 inhibitors. *Drug Metab. Dispos.* **28**, 994–1002 (2000).
87. De Groot, M., Ackland, M. J., Home, V. A., Alex, A. & Jones, B. C. Novel approach to predicting P450-mediated drug metabolism: Development of a combined protein and pharmacophore model for CYP2D6. *J. Med. Chem.* **42**, 1515–1524 (1999).
88. Ekins, S., De Groot, M. & Jones, J. P. Pharmacophore and three-dimensional quantitative structure-activity relationship methods for modelling cytochrome P450 active sites. *Drug Metab. Dispos.* **29**, 936–944 (2001).
89. Zuegge, J., Fechner, U., Roche, O., Parratt, N. J., Engkvist, O. & Schneider, G. A fast virtual screening filter for cytochrome P450 3A4 inhibition liability of compound libraries. *Quant. Struct. Act. Relat.* **21**, 249–256 (2002).
90. Higgins, L., Korzekwa, K. R., Rao, S., Shou, M. & Jones, J. P. An assessment of the reaction energetics for cytochrome P450-mediated reactions. *Arch Biochem Biophys* **385**, 220–230 (2001).
91. Langowski, J. & Long, A. Computer systems for the prediction of xenobiotic metabolism. *Adv. Drug Delivery Rev.* **54**, 407–415 (2002).
92. Ehrhardt, P. W. (Ed). *Drug Metabolism Databases and High-Throughput Testing During Drug Design and Development* (IUPAC, Blackwell Science, Malden, Massachusetts, 1999).
93. Richard, A. M. & Benigni, R. AI and SAR approaches for predicting chemical carcinogenicity: survey and status report. *SAR and QSAR Environ. Res.* **13**, 1–19 (2002).
94. Greene, N., Computer systems for the prediction of toxicity: an update. *Adv. Drug Deliv. Rev.* **54**, 417–431 (2002).
95. Durham, S. K. & Pearl, G. M. Computational methods to predict drug safety liabilities. *Drug Disc. Dev.* **4**, 110–115 (2001).
96. Roche, O., Trube, G., Zuegge, J., Pflimlin, P., Alanine, A. & Schneider, G. A virtual screening method for prediction of the hERG potassium channel liability of compound libraries. *ChemBioChem* **3**, 455–459 (2002).
97. Cavalli, A., Poluzzi, E., De Ponti, F. & Recanatini, M. Toward a pharmacophore for drugs inducing long QT syndrome: Insights from a CoMFA study of HERG K<sup>+</sup> channel blockers. *J. Med. Chem.* **45**, 3844–3853 (2002).
98. Fischer, H., Kansy, M., Potthast, M. & Csato, M. Prediction of *in vitro* phospholipidosis of drugs by means of their amphiphilic properties. In: *Rational Approaches to Drug Design* (eds. Hölte, H.-D. & Sippl, W.) 286–289 (Prous Science, Barcelona, 2001).
99. Bonnabry, P., Sievering, J., Leemann, Th. & Dayer, P. Quantitative drug interactions prediction system (Q-DIPS). *Clin. Pharmacokinet.* **40**, 631–640 (2001).
100. Willson, T. M. & Kiewer, S. A. PXR, CAR and drug metabolism. *Nature Rev. Drug Disc.* **1**, 259–266 (2002).
101. Farr-Jones, S., Computational methods to predict ADME/tox properties for drug discovery, Decision Resources Inc. November 28 (2001).
102. Clark, D. E. & Grootenhuys, P. D. J. Progress in computational methods for the prediction of ADMET properties. *Curr. Opin. Drug Disc. Dev.* **5**, 382–390 (2002).
103. Smith, D. A. Hello Drug Discovery, I am from *Insilico*, take me to your president. *Drug Disc. Today* **7**, 1080–1081 (2002).
104. Anderson, R. J. 20/20 Vision: A brave new world of drug development. *Curr. Drug Disc.* 25–29 (August 2002).
105. Sietsema, W. K. The absolute oral bioavailability of selected drugs. *Int. J. Clin. Pharmacol. Ther. Toxicol.* **27**, 179–211 (1989).
106. Thummel, K. E. & Shen, D. D. In *Goodman & Gilman's The Pharmacological Basis of Therapeutics* (eds Hardman, J. G. & Limbird, L. E.) 1917–2023 (McGraw-Hill, New York, 2001).  
**Good reference for pharmacokinetic data of over 300 marketed drugs.**
107. Sakaeda, T., Okamura, N., Nagata, S., Yagami, T., Horinouchi, M., Okumura, K., Yamashita, F. & Hashida M. Molecular and pharmacokinetic properties of 222 commercially available oral drugs in humans. *Biol. Pharm. Bull.* **24**, 935–940 (2001).
108. Van de Waterbeemd, H. & De Groot, M., Can the Internet help to meet the challenges in ADME and e-ADME? *SAR QSAR Environ. Res.* **13**, 391–401 (2002).

Acknowledgement  
We would like to thank Don Walker for helpful comments in the preparation of this manuscript.

 Online links

**DATABASES**  
The following terms in this article are linked online to:  
LocusLink: <http://www.ncbi.nlm.nih.gov/LocusLink/>  
CYP2C9 | CYP2C19 | CYP2D6 | CYP3A4 | P-glycoprotein

**FURTHER INFORMATION**  
Glossary of Terms Used in Medicinal Chemistry:  
<http://www.chem.qmul.ac.uk/iupac/medchem/>  
The QSAR and Modelling Society: [www.qsar.org](http://www.qsar.org)  
UK QSAR and Chemoinformatics Discussion Group:  
[www.ukqsar.co.uk](http://www.ukqsar.co.uk)  
Access to this interactive links box is free online.