8 *In Silico* Models of Drug Metabolism and Drug Interactions

RICHARD J. DIMELOW,[1,2] PAUL D. METCALFE,[1] and SIMON THOMAS[1]

[1] Scientific Computing Group, Cyprotex Discovery Ltd, Macclesfield, Cheshire, UK
[2] Modelling and Simulation, DMPK, AstraZeneca Ltd, Macclesfield, Cheshire, UK

8.1 SUMMARY

Determination of the nature and the rates of any active processes that are involved in the absorption, distribution, metabolism, and elimination (ADME) of drug candidates is a primary goal of drug discovery. These processes include metabolism and active transport. Metabolism entails the (usually) irreversible enzyme-mediated conversion of the compound to a different form. Active transport entails the movement of an unchanged compound, mediated by a transporter protein, across a biological membrane (frequently the plasma membrane). Within drug discovery, *in vitro*, *in situ*, and *in vivo* methods for estimating human drug metabolism and transport are becoming increasingly supplemented by a wide range of *in silico* methods aimed at reducing costs, reducing animal usage, and increasing throughput of compound screening.

These active processes are potentially subject to drug–drug interactions, whereby one drug (or, in general, any xenobiotic) alters the rate of metabolism or transport of a second xenobiotic *in vivo*, leading to changes in plasma and target-site concentrations of the affected drugs. Several mechanisms for such interactions exist, including reversible or irreversible inhibition of transporters and enzymes and increases in their expression mediated by activation of nuclear receptor proteins. *In silico* methods are being developed to predict these effects, to reduce the cost and time burden of experimental screening and to mitigate the likelihood that undetected drug–drug interactions with undesirable consequences could compromise otherwise promising drug candidates.

Encyclopedia of Drug Metabolism and Interactions, 6-Volume Set, First Edition.
Edited by Alexander V. Lyubimov.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

Following a brief review of drug metabolism and drug–drug interactions (Section 8.2), we present a high level view of some of the areas in which *in silico* modeling can contribute to the quantitative or qualitative understanding and prediction of drug metabolism and drug–drug interactions (Sections 8.2.1 and 8.2.2). We then discuss some of the available *in silico* approaches to these problems (Section 8.3). We start our discussion with methods that just depend on descriptors calculated from compound structure (Section 8.3.1), go on to discuss more complex models that use compound structure more explicitly (Section 8.3.2), then discuss approaches that try to model in detail the interaction between compound and protein (Section 8.3.3). Following this, we discuss in some detail a number of specific application areas in each of which one, or more, of the previously described methods have been used (Sections 8.4.1–8.4.3). These application areas include prediction of the rate of metabolism of a compound; determination of the substrate and/or inhibitor specificity of enzyme and transporter isoforms; and activation of the pregnane X receptor (PXR), a nuclear receptor which is known to be involved in the xenobiotic-mediated induction of a number of drug-metabolizing enzymes and transporters.

The approaches discussed up to this point do not permit the understanding of the interactions between different active processes *in vivo*; we, therefore, then (Section 8.4.5) discuss physiologically based modeling, which, in principle, allows the creation of an integrated model of compound pharmacokinetics that allows the modeling and understanding of all relevant physiological processes involved in drug metabolism and interactions. We finish with a discussion of some practical aspects of applying *in silico* methods in drug discovery (Section 8.5) and our conclusions on the topic (Section 8.6). Having such a large and diverse subject matter to cover, we realize that we cannot be comprehensive and have therefore cited several relevant recent papers for readers to follow-up.

8.2 INTRODUCTION

Drugs and other xenobiotics are potentially subject to metabolic transformation by enzymes and enzyme systems located in the liver, intestine, and other organs. Metabolism of drugs usually leads to reduction, or total loss, of their therapeutic effect, in which case the rate of metabolism can be a prime determinant of their *in vivo* therapeutic utility. Metabolism can sometimes generate toxic metabolites [1], thus contributing to the toxicity of a drug, or to active metabolites, contributing to its therapeutic effect [2]. The latter effect can be deliberately harnessed in the production of prodrugs that have, for example, favorable absorption properties but require metabolism to generate the therapeutically active compound [3,4]. The bulk of xenobiotic metabolism is performed by the cytochrome P450 (CYP) superfamily of enzymes [5], which tend to transform compounds into more polar compounds. These metabolites can be either eliminated directly or become subject to further metabolism [6].

Until relatively recently, the role of active transport in the disposition of drugs had been poorly studied in comparison to the effort put into investigating and understanding drug transport. The effects of phosphoglycoprotein (PGP) (MDR1/ABCB1) and MRP1 (ABCC1) on drug disposition had originally been observed because of their roles in the development of multidrug resistance in cancer cells [7], and there was an appreciation

of their potential in inhibiting drug uptake from the gastrointestinal (GI) tract [8] and from the blood across the blood–brain barrier [9]. Some compounds must be transported into the interior of the hepatocyte to reach the metabolizing enzymes and some compounds can be transported out of the hepatocyte into the bile [10]; the rate of transport can have an effect on the overall rate of removal from the bloodstream [11]. Transporter proteins also affect the absorption of a compound from the gut—some help compound uptake and some pump compound back into the gut lumen [12]—and hence also have an effect on the overall extent of metabolism. Transporter polymorphism can have a significant effect on pharmacokinetics [13–15].

The activity of the metabolizing enzymes and transporter proteins can be induced by a range of drugs or environmental factors and can be inhibited by a range of drugs or foodstuffs [16]. Xenobiotic concentration in plasma or tissues can significantly be affected by interactions with other compounds and can vary by an order of magnitude or more [17]. This can lead to unwanted behavior: either a lack of efficacy at a standard dose caused by an increase in overall clearance or toxic effects caused by a decrease in overall clearance. For a compound with a narrow therapeutic window—when the effective dose is close to the toxic dose—this is particularly significant. Assessment of the potential of a clinical candidate to participate in drug–drug interactions is an integral part of review by regulatory bodies [17]. Compounds that are found to participate in drug–drug interactions must undergo *in vivo* tests to determine the extent of any *in vivo* drug–drug interactions [18].

Inhibition of CYP3A4 has been a particular concern in drug development, and pharmacokinetic changes following CYP3A4 inhibition have led to the withdrawal from market of a number of compounds. Terfenadine, cisapride, and astemizole are CYP3A4 substrates that, in excess, can cause cardiac arrhythmias. If their concentrations rise—due to the inhibition of their metabolism—adverse, potentially fatal, effects can ensue. These side effects on coadministration with CYP3A4 inhibitors led to their withdrawal. The significance of CYP3A4 inhibition relates to two factors. First is the preponderance of this isoform in the metabolism of many drugs and drug-like compounds. Second is that its substrates are often lipophilic (as discussed further below), which increases the likelihood of unwanted off-target interactions *in vivo*, such as those with the human ether-a-go-go-related gene (hERG) protein [19,20] that caused the arrhythmias noted above. The postmarketing discovery of enzyme inhibition has also led to the withdrawal of mibefradil, an inhibitor of CYP3A4 and PGP [21].

Induction of drug-metabolizing enzymes is less dangerous, insofar as the consequence is reduced concentrations of the substrate(s) of the induced enzyme(s), but this does raise the possibility of therapeutic failure of such substrates, with potentially serious consequences. A common mechanism for CYP induction is binding of ligand to nuclear receptors that regulate transcription, which alters the expression of the CYP genes [22]. This means that potential CYP inducers can be identified as potential binders to the relevant nuclear receptors. As examples, plasma levels of CYP3A4 substrates—for instance, oral contraceptives, triazolam, and midazolam—can be reduced by coadministration of CYP3A4 inducers [22–24], leading to a lack of efficacy at standard doses.

There are a number of transporter-based interactions documented in the literature: for instance, an increase in repaglinide plasma concentration following atorvastatin dosing due to an inhibition of OATP1B1 [25], an increase in plasma concentration of rosuvastatin following cyclosporin dosing due to the same effect [26], and an increase

in atorvastatin plasma concentration following rifampin dosing [27,28]. Transporter inhibition does not yet seem to have been as serious as metabolic inhibition, but the capacity for clinically significant changes in pharmacokinetics exists.

Just in this brief discussion of the biochemistry and physiology that affects xenobiotic metabolism, we find some questions that must be answered—qualitatively or quantitatively—during drug discovery, to produce high quality development candidates:

1. The overall rate of metabolism, which affects the tissue concentrations reached after administration, and consequent duration of action of the drug;
2. The identity of major metabolites along with knowledge of whether, or not, they are toxic and/or therapeutically active
3. The enzyme/enzymes that is/are responsible for the majority of metabolism and, where more than one enzyme is involved, the relative proportions of metabolism via each enzyme. This determines the potential for
   a. Differences in metabolism between different individuals of the target population
   b. Effects of polymorphisms in drug-metabolizing enzymes
   c. Inducibility of drug metabolism, and the identities of potential inducers
   d. Inhibitability of drug metabolism, and the identities of potential inhibitors
4. Potential effects on metabolism of other drugs that could be coadministered, whether by induction or inhibition of drug-metabolizing enzymes.

Furthermore, with the increasing recognition of the importance of active transport in affecting the disposition and elimination of drugs, there is a corresponding increase in emphasis being given to determining the analogous properties (excluding 2) relevant to drug transport in the liver, kidneys, intestines, and other tissues.

8.2.1 Why Is *In Silico* Modeling Necessary?

All the above properties can be determined, in principle, by appropriate *in vitro* assay. But a liability in any one property could cause costly clinical failure of a candidate, so there is a consequent drive to determine these properties as early as possible within drug discovery. This leads to increasing costs and compound requirements in lead generation and optimization, as additional assays are developed and used to inform the discovery process. This generates competition for resources in lead identification and lead optimization that requires careful prioritization for successful outcomes at affordable costs. Improvements in assay miniaturization and throughput [29] have helped offset these costs, but there is still a potentially crucial role for *in silico* modeling to inform prioritization of assays, and hence to ease the competition for resources throughout drug discovery. First, predictive models for *in vitro* assay endpoints can be used to reduce the number of assays to be performed and/or to increase the number of compounds—whether real or virtual—that can be screened. Second, models can be used to extrapolate from *in vitro* assay endpoints (or predictions thereof) to predict *in vivo* pharmacokinetics. A variety of *in silico* techniques are used across the entire range of ADME, toxicity, pharmacology, and physicochemical properties [30–40], of which applications relating specifically to drug metabolism and drug–drug interactions [17,41–47] are a subset.

8.2.2 Problems for *In Silico* Modeling

This leaves us with a number of problems that could be addressed by *in silico* modeling.

1. Predict the total rate of metabolism of a compound.
2. Predict whether a compound is metabolized by a given enzyme, or enzyme isoform, and the rate of such metabolism. One difficulty is that the enzymes, such as CYPs and uridine 5′-diphosphoglucuronic acid (UDPGA) glucuronosyl transferases (UGT), that control a significant fraction of drug metabolism are promiscuous, metabolizing a wide range of substrates. If interactions with a given metabolizing enzyme are predicted, predict the kinetics of transformation, and also the metabolite or metabolites formed.
3. Predict whether a compound is transported by a given transporter and the rate of such transport. Just like the metabolizing enzymes, the transporters are promiscuous proteins that interact with a variety of drugs, and any interaction with a transporter is a potential source for drug–drug interactions with compounds that affect that transporter. If interactions with a given transporter are predicted, this can give some idea of the population variability that may be expected.
4. Predict whether a compound inhibits metabolism or transport of itself or other compounds. Is the compound likely to prevent the metabolism or transport of other compounds? Does it affect important enzymes, risking interactions with many other drugs, or does it just affect a relatively insignificant enzyme?
5. Predict whether a compound induces metabolizing enzymes or transporters. Does the compound bind to a nuclear receptor, causing increased expression of CYP or transporter protein? Which enzymes does it induce; are they important, or relatively insignificant in the metabolism of the drug, or of coadministered drugs?
6. Predict or model, quantitatively, total *in vivo* clearance or concentration time profiles. What effects do metabolism and transport have on the compound's pharmacokinetics, and how do they interact with absorption and distribution. If one can model *in vivo* metabolism and transport to the extent of individual enzymes, one could begin to understand the compound's susceptibility to population variability and enzyme induction or inhibition. A model of this class is a useful tool: one can ask the questions "by how much do I have to inhibit metabolism to increase plasma concentration of my drug or a probe substrate by x-fold?" or "by how much do I have to induce metabolism to decrease plasma concentration of my drug or a probe substrate x-fold?," where the value of x can be dependent on the drug's therapeutic window. It is possible to begin to understand the true *in vivo* significance of an interaction discovered *in vitro*.
7. Predict or model, quantitatively, inhibition or induction effects due to the dosed compound or caused by a coadministered drug. With knowledge of effective and toxic concentrations, a model of this class directly addresses changes in the therapeutic window of the compound or of a coadministered drug. Such a model also aids in the design of *in vivo* clinical studies that examine an interaction in detail. One may, perhaps, attempt full prediction from *in vitro* data, or one may take the single-compound pharmacokinetics as given and just attempt to model the compound–compound interaction.

8.3 TOOLS

In this section, we discuss in some detail the different *in silico* approaches to drug metabolism and drug–drug interaction. Generally, we must consider compound interactions with varying promiscuous proteins, whether transporters, metabolizing enzymes, or nuclear receptors. Given this similarity among the different problems, we look at each modeling approach in turn and discuss some advantages and disadvantages of the methods.

8.3.1 Compound Descriptor Methods

A compound can be characterized by a set of molecular descriptors that are classed according to the dimension of the chemical formula used to derive them. One-dimensional descriptors are calculated from the empirical formula and include quantities such as the molecular weight and atom counts. Two-dimensional descriptors come from the planar projection of the chemical structure and include, for example, the number of hydrogen bond donors and acceptors. Finally, the three-dimensional structure of the compound gives rise to three-dimensional descriptors such as charge distribution parameters like the dipole and quadrupole moments [48]. These molecular descriptors form the basis of a quantitative structure–activity relation (QSAR), which relates these molecular descriptors with a specific activity, such as metabolic rate or protein binding affinity, for a training set of compounds [39,49,50]. A huge number of molecular descriptors can be calculated for each compound [48]. However, sensitivity analysis can identify those descriptors having greatest influence over the structure–activity relationship, and restricting the descriptors to this select group helps prevent over-fitting to the compound training set. The major advantage of using compound descriptor methods to predict enzyme metabolism/inhibition is that structural knowledge of the active site is not required so this method can be used to predict enzyme activity when the enzyme structure is unknown.

8.3.2 Compound Structure Methods

The following methods utilize the compound geometry: either the three-dimensional structure or its two-dimensional projection onto a plane. The geometry provides a structural understanding of binding into an active site and we therefore expect compound structure methods to be superior to methods based purely on compound descriptors (Section 8.3.1). Furthermore, as with compound descriptor methods, compound structure-based methods do not utilize the protein structure and can therefore be used when the protein structure is unknown [51–55].

8.3.2.1 Pharmacophore Methods. A pharmacophore is the collection of the most important structural features relevant for a given biological activity from a series of molecules with a similar mechanism of action [56]. Pharmacophore features are, for example, hydrogen bond donors, hydrogen bond acceptors, acidic centers (centers with a negative charge at neutral pH), hydrophobic regions, and aromatic rings [57]. The set of features with their corresponding positions within the molecule then define the pharmacophore. By comparing a set of known enzyme substrates to a set of known

nonsubstrates, the common set of pharmacophore features required for binding to the enzyme active site can be identified. This method can therefore classify other compounds as being substrates or nonsubstrates, depending on whether they contain the same pharmacophore features. However, it will not predict the strength of binding to the enzyme active site and so cannot differentiate between compounds of the same class.

8.3.2.2 Comparative Molecular Field Analysis. Comparative molecular field analysis (CoMFA) is a particular three-dimensional QSAR technique. It requires that a training set of compounds are all aligned, according to some rule, within a three-dimensional grid. The steric and electrostatic fields are calculated for each compound at each point on the grid, which are then correlated with the compound activity. CoMFA is a widely used method in drug discovery, but its success is sensitive to the molecular alignment step. By aligning the compounds, the assumption is made that all compounds have the same orientation in the active site. In reality, however, different compounds will bind to the active site in different orientations, perhaps even several orientations for a given compound [58].

8.3.2.3 Comparative Molecular Similarity Indices Analysis. Comparative molecular similarity indices analysis (CoMSIA) is another three-dimensional QSAR technique where molecules are aligned on a grid and the steric and electrostatic fields calculated [59]. However, CoMSIA is considered superior to CoMFA as the hydrogen bond donor, hydrogen bond acceptor, and hydrophobic fields are also taken into account. At each grid point, these interaction fields are converted to a similarity index [60], which is the property that is correlated with the compound activity. The steric and electrostatic fields, given by the Lennard–Jones and Coulombic potential, respectively, can be excessively large for grid points close to atoms of the compound. The Gaussian function of the similarity index can be designed to moderate this dominant behavior, ultimately resulting in greater predictive accuracy [61].

8.3.3 Protein Structure Methods

The following methods require the three-dimensional structure of the protein, which restricts these methods to systems where this is known, or at least predicted using protein homology models [62]. Since the proteins typically contain several thousands of atoms, these methods are computationally expensive and are not high throughput. However, this disadvantage is somewhat offset since these methods can reveal the exact nature of binding into the active site of an enzyme.

8.3.3.1 Molecular Dynamics. A molecular dynamics simulation provides an atomistic view of enzyme–substrate binding over time. It can capture the different orientations and conformations of the compound in the active site, and the role of solvent molecules if they are included in the simulation too [63]. The accuracy of the simulations depend on the classical force field used to model the interactions between compound and enzyme, which may require parameterizing to higher level quantum calculations. Add this to the already computationally intensive molecular dynamics simulation and it is clear that this method cannot currently be used as a general screening tool in drug discovery.

8.3.3.2 Docking. As the name suggests the compound is docked into the binding site of the enzyme. Each atom pair between the ligand and the enzyme is considered, the separation between the two atoms calculated, and the result converted into an atom-pair potential according to a scoring function. The total binding potential is the sum of the potential energies across all atom pairs. Where entropic and solvation effects have been incorporated into the scoring function, the free energy of binding is obtained [64,65]. However, knowledge-based atom-pair potentials are parameterized on selected sets of ligand–protein structures and may not predict well outside this domain [64].

8.4 APPLICATIONS

Here, we present a variety of examples where *in silico* methods have been used to characterize substrates of CYP, UDP-glucuronosyltransferase, and the PXR according to the different problems that are outlined in Section 8.2.2.

8.4.1 Prediction of the Rate of Metabolism

We start with prediction of overall rate of metabolism, not because it is one of the easiest to address by *in silico* modeling—far from it—but because it is a fundamental property that can significantly affect the likelihood of its progression in drug discovery. Consequently, the *in vitro* determination of metabolic rate (using isolated hepatocytes, or subcellular fractions such as microsomes or S9) is a routine screen utilized in lead optimization, or even lead identification, during drug discovery. Even modest drug discovery programs can result in large numbers of compounds being screened for this property [66]. Consequently, any contribution that *in silico* modeling can make to reduce the need for *in vitro* screening is potentially valuable.

The problem has been studied since the early days of QSAR modeling. For example, Hansch *et al*. [67] developed linear regression models for oxidative deamination of aliphatic amines and benzylamines by a variety of *in vitro* systems (purified amine oxidases, liver extracts, and beef liver mitochondria), for glucuronide formation of benzoic acids and aliphatic alcohols in the rabbit *in vivo* and for hippuric acid formation, also in the rabbit *in vivo*. The models they developed were for small series (7–30 compounds) of closely related, relatively simple compounds, having calculated $\log P$s in the range 0.7–3.5. They considered hydrophilic/hydrophobic nature, hydrogen bonding capability, polarization, dipole moment, molar volume, pK_a, and electronic and steric properties. In essence, they found that the logarithm of the rate of metabolism (MR) was a parabolic function of calculated $\log P$, of the general form:

$$\log(\mathrm{MR}) = -a \cdot \log P^2 + b \cdot \log P + c$$

with, for a small number of cases, a contributing term from the Taft intramolecular steric substituent coefficient, E_s. These relationships had correlation coefficients between 0.8 and 0.95 on the training set, thus explaining between 64% and 90% of the variance of the training sets. Now, these relationships are unlikely to generalize beyond the training sets (because of the relatively large ratio of fitted parameters to compounds), and are not necessarily applicable to larger, more complex drugs and drug leads, but they did provide valuable insight into important characteristics of drug metabolism.

First, the parabolic fits imply the existence of a $\log P$ value ($\log P_0$) for which metabolic rate is predicted to be maximal. The authors interpreted this as a balance between increasing lipophilicity enhancing the ability of compounds to diffuse through the lipoidal membranes (*in vitro* or *in vivo*) to reach the active site(s) of the metabolizing enzyme(s) and the tendency, when lipophilicity is particularly high, to strongly adhere to lipids or lipophilic proteins, so slowing this necessary diffusive step. They observed that, across the range of series and systems they investigated, the mean $\log P_0$ was ~2.0. Second, particular determinants of metabolic rate were observed for certain compound series. Thus, predictability of the *in vivo* rate of glucuronidation of primary aliphatic aldehydes was improved by incorporating E_s with a negative coefficient, indicating that as the site of glucuronidation becomes more hindered, the rate of glucuronidation increases, probably because the competing oxidative metabolism is more sterically demanding and becomes more strongly inhibited. This was not observed for secondary aliphatic alcohols, in which oxidation is already hindered.

Recent developments in *in vitro* assays for drug metabolism in drug discovery have tended toward the deployment of high throughput, standardized, assays within pharmaceutical and biotech sectors using isolated cells or subcellular fractions such as microsomes or S9 [29]. Routine use of these assays both provides the capability to generate large data sets of standardized data for model building and defines the framework within which predictions from models are judged. Chang *et al*. [66] generated models for intrinsic clearance (CL_{int}) in a rat hepatic microsomal incubation using an isocitrate dehydrogenase nicotinamide adenine dinucleotide phosphate (NADPH)-generating system. The model was developed using ~28,000 drug leads across a broad chemical space with a calculated $\log P$ normally distributed between −3.7 and 10, molecular weights normally distributed between 140 and 839, and intrinsic clearance strongly right-skewed between 1.57 and 989 μL/min/mg protein. The model was a two-way classifier to distinguish between those compounds with $CL_{int} \leq 87$ μL/min/mg and those with $CL_{int} > 87$ μL/min/mg.

The model was a random forest and had an overall predictive accuracy of 83% on an internal test set and 77% on an external test set that was in a distinct area of property space from the training set compounds. Descriptors important in the model were a double-bond E-state descriptor, $\log P$, and fractional hydrophobic and fractional hydrophilic van der Waal's surface areas (which are completely correlated). The result is a potentially useful model, but the study illustrates the challenge in modeling xenobiotic metabolism to obtain useful results, even when using very large training sets.

The difficulty in predicting CL_{int} is unsurprising. Not only are xenobiotic-metabolizing enzymes promiscuous with respect to substrate structure [68,69], but CL_{int} is itself a composite property. For an enzyme displaying Michaelis–Menten kinetics, it is defined as the ratio of the maximal rate, V_{max}, and the Michaelis constant, K_m:

$$CL_{int} = \frac{V_{max}}{K_m} \tag{8.1}$$

Many compounds can be metabolized by a number of P450 isoforms present in the microsomal fraction so that the total rate of metabolism, v, is the sum of metabolic rates of the individual isoforms. If each of n isoforms displays Michaelis–Menten

kinetics toward a compound,

$$v = \frac{V_{\max,1} \cdot S}{K_{\mathrm{m},1} + S} + \frac{V_{\max,2} \cdot S}{K_{\mathrm{m},2} + S} + \cdots + \frac{V_{\max,n} \cdot S}{K_{\mathrm{m},n} + S} = \sum_{i}^{n} \frac{V_{\max,i} \cdot S}{K_{\mathrm{m},i} + S}$$

where S is the drug concentration at the active site of the enzyme.

In this case, the total intrinsic clearance is given by

$$\mathrm{CL}_{\mathrm{int}} = \sum_{i}^{n} \frac{V_{\max,i}}{K_{\mathrm{m},i}}$$

assuming that $S < K_{\mathrm{m},i}$ for all isoforms. Thus, for a compound that is metabolized by n isoforms, $\mathrm{CL}_{\mathrm{int}}$ is a combination of $nV_{\max}$ and $n\mathrm{K}_{\mathrm{m},i}$. Even for a compound whose rate of metabolism is quantitatively dominated by a single isoform, $\mathrm{CL}_{\mathrm{int}}$ is a conflation of two composite parameters. Using the Briggs–Haldane definition of $V_{\max}$ and K_{m} for a single enzyme/isoform,

$$V_{\max} = E_0 \cdot k_2$$

and

$$K_{\mathrm{m}} = \frac{k_{-1} + k_2}{k_1}$$

where E_0 is the total concentration of the isoform, k_2 the rate constant for the irreversible catalytic step converting the enzyme–substrate complex to generate the product, k_1 the association rate constant of the substrate and enzyme, and k_{-1} the corresponding dissociation rate constant of the enzyme–substrate complex. Thus, the two composite parameters of the Michaelis–Menten equation correspond to four different fundamental properties. E_0 is, in principle, measurable and, for the specific case where the rate of catalytic conversion is much less than the rate of dissociation of the enzyme–substrate complex ($k_2 \ll k_{-1}$), then

$$K_{\mathrm{m}} \approx \frac{k_{-1}}{k_1}$$

which is the dissociation constant for the enzyme–substrate complex. Thus, in the simplest case, prediction of $\mathrm{CL}_{\mathrm{int}}$ requires the implicit prediction of an underlying association step, which could be expected to be dominated by bulk properties relating to hydrophobicity/hydrophilicity, polarity, and hydrogen bonding, and of a transformative step, which could perhaps be expected to depend more on electronic and steric properties. The significant descriptors identified by the Chang model reflect the importance of hydrophobicity/hydrophilicity, hydrogen bonding, and electronic but not steric factors.

In contrast to the complexity inherent in attempting to predict $\mathrm{CL}_{\mathrm{int}}$ in microsomes, predicting the properties of individual enzymes/isoforms offers a more tractable proposition. The difficulties should not be underestimated, because of the inherent promiscuity of drug-metabolizing enzymes, but the individual enzyme provides a far simpler system to understand than the complex microsomal fraction. In addition, the full

range of tools bringing in structural data about the isoforms concerned (Section 8.3) is available for studying these simpler systems, whereas the prediction of CL_{int} in multi-isoform systems, such as microsomes, is essentially limited to statistical methods, such as QSAR.

Ultimately, prediction of drug metabolism via the properties of individual isoforms should prove more valuable than prediction of, say, microsomal CL_{int}. This latter property can be measured in relatively high throughput screens, requiring a relatively small amount of compound [29]. Data on the effects of individual drug-metabolizing enzymes on a potential drug (such as the relative rates of metabolism, inhibitibility by other drugs, and nonlinearity of metabolism), while being increasingly required by regulatory agencies and/or to predict likely population pharmacokinetics of clinical candidates compounds, are more expensive, time consuming, and/or require more compound to obtain. This combination of potentially providing greater cost savings, at the same time as offering problems that are more soluble indicates that prediction of the properties of individual drug-metabolizing enzymes is the route that offers greatest potential in drug discovery. That said, the advantage of developing models for microsomal CL_{int} is the very availability of data for model generation. The work of Chang *et al*. implies that a large "global" model for CL_{int} using data generated across many areas of chemistry has potential benefit in screening new chemistry and could thus be used to virtually screen libraries or hits for lead identification.

8.4.2 Drug Interaction with Specific Enzyme Isoforms

Changing the focus of study from composite systems (such as isolated cells, microsomes, or S9 subcellular fractions) to simpler systems, such as individual enzyme isoforms, increases both the toolset available for model development and the number of properties that can potentially be predicted. The additional tools include (Section 8.3) docking and molecular dynamics methods (where the crystal structures are available), field alignment methods, such as CoMFA and COMSIA, and pharmacophore methods. The additional properties that can determined, and hence potentially predicted, with *in silico* models include various measures of the propensity of a particular enzyme isoform to metabolize a given compound or, alternatively, the propensity of a given compound to inhibit the metabolism of other compounds by a particular isoform. These are two different problems, though with common underlying bases that relate to the complementarity—in ways that need to be defined and determined—between a compound's structure and the structure of active/allosteric sites on the enzyme form in question. Each of these problems must be tackled during drug discovery, and we address each in turn in the following sections.

8.4.2.1 Isoform-Specific Metabolism. The identities of the enzyme isoforms responsible for metabolizing a given drug, and the distribution of metabolism between them, determine the pharmacokinetics of a drug in a heterogeneous population, the potential for nonlinear pharmacokinetics as a consequence of enzyme saturation, and the potential for enzyme inhibition and induction to affect the drug's pharmacokinetics, toxicity, and therapeutic efficacy. Experimental methods are available for identifying the enzymes, such as measuring metabolism in expressed enzyme systems or inhibiting metabolism in microsomes by isoform-specific inhibitors. Since metabolism plays

such an important role in determining the efficacy and toxicity of a compound, it is necessary to understand the metabolic fate of potential drug entities early on in the drug discovery process. A large number of compounds presented at this stage would render a full metabolic screen of them all unfeasible and therefore there is a clear need to develop reliable predictive models of metabolism instead.

First, we focus our attention to simply predicting whether a compound is a substrate for metabolism. This problem has been tackled using a wide variety of molecular descriptors and machine learning methods and we outline a few studies [70–73] in Table 8.1. The predictive accuracies for classifying whether a compound is a substrate of an enzyme isoform span the range 65–97%, but estimated accuracies depend on the chemical domain spanned by the training and validation sets, so it not a straightforward process to rank and compare models from different studies. However, given a compound is predicted to be a substrate, the conditional probability it is actually a

TABLE 8.1 Summary of Selected Studies that Predict Whether a Compound Is a Substrate/Nonsubstrate for Metabolism

References	Enzyme System	Method	Accuracy	Notes
Sorich *et al*. [70]	Human UGT	Support vector machine	76%	Averaged accuracies across all isoforms
		Partial least squares determinant analysis	66%	
		Bayesian regularized artificial neural network	71%	
Sorich *et al*. [71]	Human UGT	Partial least squares determinant analysis	69% and 71%	Averaged accuracies across all isoforms
		K-means clustering plus genetic algorithm	72%	
Balakin *et al*. [72]	CYP450	Kohonen self-organizing maps	CYP3A4 91% (high K_m predictions) 97% (low K_m predictions) All CYP isoforms 82% (high K_m predictions) 65% (low K_m predictions)	Accuracies calculated from compound training set
Yap and Chen [73]	CYP450	Support vector machine	CYP3A4 95% CYP2D6 95% CYP2CP 97%	

The predictive accuracy of each of the methods used is given.

substrate can be shown to be

$$P(\text{sub}|\text{pred}) = \frac{1}{1 + \frac{1-\text{Sp}}{\text{Se}} \times \frac{1-f}{f}}$$

The specificity (Sp) is the probability a nonsubstrate is correctly identified as being a nonsubstrate by the model. Likewise, the sensitivity (Se) is the probability a substrate is correctly identified to be a substrate by the model. Finally, f is the fraction of actual substrates in our total set of compounds. Taking the worst case scenario that the model accuracy is only 65%, and assuming this represents both the specificity and sensitivity,

$$P(\text{sub}|\text{pred}) = \frac{1}{1 + 0.538 \times \frac{1-f}{f}}$$

Therefore, if overall only 10% of the compounds are actually enzyme substrates ($f = 0.1$), then the model will predict a subset of these compounds to be substrates, 17% of which will actually be substrates. Thus, the model will have increased the proportion of substrates by almost a factor of 2, from 10% of the total set to 17% of the predicted subset. Using the model accuracy in this way can help judge whether applying a model for screening purposes would result in worthwhile savings in time and expense.

In silico docking is a computationally expensive method of predicting drug metabolism and therefore cannot be routinely used to screen large numbers of compounds. It is particularly valuable, however, for providing an atom-level understanding of the drug–enzyme binding complex and has been used to study drug binding to CYP2D6 [74]. This study revealed that metoclopramide can adopt two positions in the enzyme active site, but only one of these binding conformations had been observed experimentally. Analysis of the second binding conformation implied the production of a second metabolite that had not previously been identified, but which was later verified by experiment.

Molecular dynamics simulations, like *in silico* docking methods, are computationally demanding but can provide insight into the mode of binding for a small number of drugs. Seifert *et al*. [75] analyzed the dynamics of warfarin in the active site of CYP2C9. The crystal structure of this warfarin-bound complex cannot by itself predict the subsequent metabolites, as the warfarin sits too far away from the catalytic heme group. Molecular dynamics simulations, on the other hand, revealed two ways of binding to this catalytic center, resulting in the formation of the two experimentally observed metabolites. In a more recent study [76], the role of water within the active site of CYP2D6 was investigated using molecular dynamics simulations. Hydration sites were identified and these water molecules, when included in an *in silico* docking protocol, improved the reliability in predicting the sites of metabolism for methylenedioxy-*N*-methylamphetamine.

Both docking and molecular dynamics methods predict the site of metabolism on a compound by identifying the energetically favorable binding conformations in the active site of an enzyme. The compound atom that predominantly interacts with the catalytic group of the active site is identified as being the site of metabolism. In the absence of an enzyme structure, an alternative way of estimating possible sites of metabolism is to consider the electronic environment surrounding the hydrogen

atoms of the compound [77,78]. If it is assumed that removal of the hydrogen is a rate-limiting step, then hydrogen atoms requiring least energy to dissociate from the molecule are most likely to be at the sites of metabolism. Calculating dissociation energies requires expensive quantum mechanical calculations, but semiempirical methods have been developed, which reduce the computation time with minimal compromise on accuracy [79]. Although these methods are far too costly to be used in early stage drug discovery, it has been proposed that there is potential value in employing these methods to assess metabolic transformations of lead compounds in the preclinical phase [78].

8.4.2.2 Inhibition of Metabolism. While some drugs are metabolized by CYP450 or UGT enzymes, other drugs will inhibit this metabolism, potentially resulting in drug–drug interactions that may compromise the success of clinical trials. Prediction of enzyme inhibition in early drug discovery is becoming increasingly important and we summarize several studies of this kind in Table 8.2 [43,73,80–82].

The studies summarized in Table 8.2 show that compounds can be predicted to be strong or weak enzyme inhibitors with an accuracy ranging from 66% to 96%, with the support vector machine method developed by Yap and Chen [73] achieving accuracies at the higher end of this range.

Kriegl *et al*. [80] applied partial least squares methods to a combination of 2D, 3D, and quantum mechanical descriptors resulting in an overall accuracy of 66%. In addition, a model based purely on quantum descriptors (accuracy of 54%) did not outperform a model composed purely of 2D descriptors (accuracy = 58.3%). This is surprising as unlike 2D descriptors, quantum mechanical descriptors are derived from the 3D chemical structure, and a molecule's electronic structure offers great insight into its binding potential. However, on this evidence, it seems the considerable computational overhead required to calculate quantum mechanical descriptors does not translate to increased predictive ability.

Hudelson *et al*. [81] built four models to predict compounds that inhibit CYP2C9, the predictive ability ranging from 74% to 90%. To increase reliability, it was proposed that a prediction should only stand when all four models agreed whether a compound was an enzyme inhibitor. Surprisingly, there was agreement across all four models for only two thirds of the compounds tested, and so in this case, one-third of compounds could not be classified. This stringent rule was then relaxed so that a result would stand if at least three of the models agreed, increasing coverage to the majority of compounds and resulted in a predictive accuracy of 91%.

Vasanthanathan *et al*. [82] tested a variety of models on a large test set of 7000 compounds, resulting in predictive accuracies in the range 67–76%, and since the test set is so large, these values can be taken to be a reliable measure of the predictive accuracy of enzyme inhibition by machine learning methods. The nonlinear support vector machine and random forest methods gave slightly superior accuracies, 75% and 76%, respectively. At the other end of the scale, a decision tree based on Lipinski's rule-of-five descriptors managed to classify 67% of test compounds correctly. However, this decision tree is easy to interpret and provides structural insight into CYP1A2 inhibitors—hydrophobicity and hydrogen bond donors/acceptors being important features for classifying CYP1A2 inhibitors. This structural interpretation is not as transparent from the machine learning methods. On balance, one must decide

TABLE 8.2 Summary of Selected Studies that Predict Whether a Compound Is an Enzyme Inhibitor

References	Enzyme System	Method	Accuracy	Notes
Kriegl *et al*. [80]	CYP3A4	Partial Least Squares:	66%	—
Yap and Chen [73]	CYP450	Support vector machine	CYP3A4 92–96% CYP2D6 92–94% CYP2C9 95%	—
Marechal *et al*. [43]	CYP3A4	Docking	Strong ($k_i < 10\,\mu$M): 5/7 or 7/7 Weak ($k_i < 100\,\mu$M): 5/5	Depends on the crystal structure used
Hudelson *et al*. [81]	CYP2C9	Line walking recursive partitioning	88%	—
		Normal equation recursive partitioning	90%	
		Gravity	86%	
		Substructure discovery using examples	74%	
Vasanthanathan *et al*. [82]	CYP1A2	Support vector machine (linear)	73% (77%)	Accuracies calculated from compound training set in brackets
		Support vector machine (nonlinear)	75% (82%)	
		Random forest	76% (100%)	
		k-nearest neighbors	74% (83%)	
		Decision tree	71% (97%)	
		Lipinski-decision tree model	67%	

The predictive accuracy of each of the methods used is given.

whether the additional accuracy obtained from machine learning methods offsets the ease of use and straightforward interpretation of the Lipinski's rule of five.

Finally, as mentioned in Section 8.4.2.1, the computational expense of the docking method restricts its application to a small number of compounds. That said, Marechal *et al*. [43] were able to discriminate well between strong ($k_i < 10\,\mu M$) and weak ($k_i < 100\,\mu M$) inhibitors using this technique, but its main selling point is in the detailed structural information it reveals regarding the mode of binding in the active site of the enzyme.

8.4.3 Transporters

8.4.3.1 Identifying Substrates for Transporters. To fully understand drug metabolism and drug–drug interactions, one must consider active transport into and out of cells as it may be significant compared with passive diffusion for certain drugs, and there has been corresponding interest in developing in silico prediction models to predict transporter activity [83].

In Table 8.3, we summarize results from selected studies that have developed QSAR models to predict the affinities for a range of transport proteins: P-glycoprotein, organic cation transporters, and nucleotide transporters [84–91]. Osterberg and Norinder [84] conclude that variables associated with hydrogen bonding and high polarizability correlate with high P-glycoprotein ATPase activity. Other studies also highlight the importance of hydrophobic features [90–92], which may represent aliphatic or aromatic hydrophobes [92].

TABLE 8.3 Summary of Selected Studies that Have Developed *In Silico* Models to Predict the Affinity of Drugs to Transporter Proteins

References	Transporter Protein	Method	Predictive Accuracy	Correlation Coefficient Between Experimental and Observed Values
Osterberg and Norinder [84]	P-Glycoprotein	Partial least squares	—	0.718
Bednarczyk *et al*. [85]	Human organic cation transporter	Forward stepwise regression	—	Pharmacophore based model: 0.86 Molecular descriptor based model: 0.95
Chang *et al*. [86]	Human nucleotide transporters	CoMFA	—	
	hCNT1			0.98
	hCNT2			0.83
	hENT1			1.00
Gombar *et al*. [87]	P-Glycoprotein	Stepwise discriminant analysis	86.2%	—
Xue *et al*. [88]	P-Glycoprotein	k-nearest neighbors	70.8%	—
		Probabilistic neural network	74.4%	
		C4.5 decision Tree	71.5%	
		Support vector machine	79.4%	
Suhre *et al*. [89]	Organic ion transporters	Forward stepwise regression	—	2D-QSAR: 0.81 CoMFA: 0.97
Chang *et al*. [90]	P-Glycoprotein	Pharmacophore model	—	0.87
Zalloum and Taha [91]	P-Glycoprotein	Genetic partial least squares	—	Training set: 0.8 Validation set: 0.6

8.4.4 Induction of Metabolizing Enzymes or Transport Proteins

The PXR is a nuclear receptor that regulates the expression of metabolizing enzymes and P-glycoprotein efflux transporters. A drug activating PXR can therefore influence the metabolism and transport of itself, or a second drug dosed simultaneously, and may lead to drug–drug interactions. A pharmacophore for PXR ligands has been identified by Ekins and Erickson [93] and consists of hydrogen bond acceptor and hydrophobic features, similar to the pharmacophore for CYP3A4 substrates and inhibitors, but in different positions. The pharmacophore alone was able to differentiate very potent activators from weak activators. More involved methods are required to build more reliable predictive models for PXR activation, some of which are summarized in Table 8.4 [94–96].

A comparison of the predictive accuracies obtained by Ung *et al*. [94] and Khandelwal *et al*. [95] reveals that the support vector machine method is superior to other methods investigated. The predictive accuracies obtained by Ung *et al*. were roughly 10% greater than those achieved by Khandelwal *et al*., even when they both used the support vector machine technique. The most likely explanation for this discrepancy is perhaps in the calculation of the accuracy statistics, rather than the way in which the methods were applied in the two studies. Khandelwal *et al*. used a 145-compound test set to measure the predictive accuracy, whereas Ung *et al*. adopted a 10-fold cross-validation study in which the test sets each contained only 20 compounds. The larger test set of Khandelwal *et al*. could conceivably cover a greater area of chemical space

TABLE 8.4 Prediction of PXR Activators and Nonactivators

References	Transporter	Method	Accuracy	Notes
Ung *et al*. [94]	Human PXR	k-nearest neighbors	75.0%	Recursive frequency elimination (Guyon *et al*. 2002) used to select relevant molecular descriptor
		Probabilistic neural network	77.7%	
		Support vector machine	79.6%	
Khandelwal *et al*. [95]	Human PXR	Recursive partitioning	63.45% (87.50%)	Accuracies calculate from compound training set in brackets
		Random forest	65.52% (73.45%)	
		Support vector machine	66.90% (94.35%)	
Ekins *et al*. [96]	Human PXR	Docking	47.06–57.98%	The range of predictive accuracies obtained from docking into six different crystal structures

and would therefore be a greater test of the model's ability to predict human PXR activators.

In contrast to the machine learning methods, *in silico* docking as employed by Ekins *et al*. [96] resulted in a much worse predictive accuracy, barely greater than a 50–50 guess. However, the authors also investigated the performance of 3D-, 4D- and 5D-QSAR. Surprisingly, 3D-QSAR methods, such as CoMFA and CoMSIA, also performed poorly, a reason being that the PXR active site is flexible, allowing small molecules to bind in several orientations. In CoMFA and CoMSIA, the compounds must be aligned and so these two methods would not capture the multiple orientations, which contribute to the overall observed enzyme activation. 4D-QSAR, on the other hand, does consider multiple conformations, orientations, and stereoisomers of each compound and improved the predictions for the training set compounds, although the prediction of the test set compounds was still poor. 5D-QSAR is an extension of 4D-QSAR, also considering the induced fit of the compound in the active site of the protein, and was the only QSAR method capable of making test set predictions of steroidal PXR activation with a reasonable level of success.

8.4.5 Physiologically Based Models

The modeling approaches discussed hitherto only provide information about the existence or extent of the interaction between a compound and a protein. These data may be useful in early stage drug development, providing an indication of the likelihood of such an interaction, but do not provide any indication of the clinical significance of this interaction: crudely, if significant interactions occur only when the test compound concentration is greater than some lower bound, we also need to know the time for which the compound concentration is greater than this lower bound after the compound is dosed. To obtain information about the clinical significance of a compound–protein interaction, it is necessary to integrate interaction data with pharmacokinetic data. Indeed, interactions between a compound and metabolizing enzymes or transport proteins determine the compound's pharmacokinetics.

8.4.5.1 Single-Compound Pharmacokinetics. The use of physiologically based pharmacokinetic (PBPK) models to predict or model pharmacokinetics has been widely discussed [97–100] and the prediction or modeling of single-compound pharmacokinetics is a prerequisite of the prediction or modeling of interactions between compounds. A number of commercial products exist or it is possible to implement basic PBPK models in-house.

The gross structure of a whole-body PBPK model is shown in Fig. 8.1. Dosed compound flows with blood from the arteries through the organs into which it can distribute or from which it may be cleared from the body. After passing through the organs, compound arrives in the venous blood, from which it enters the pulmonary circulation and then flows back into the arteries. While this gross structure is generally agreed, the variations that are observed in the details of PBPK models mean that there is as yet no well-defined and agreed generic structure for a PBPK model that can be relied on in all situations. Different groups model different organs, and sometimes organs, particularly the liver, are supposed to be segmented [101]. Other models include the dynamics of plasma protein binding and transport effects in the

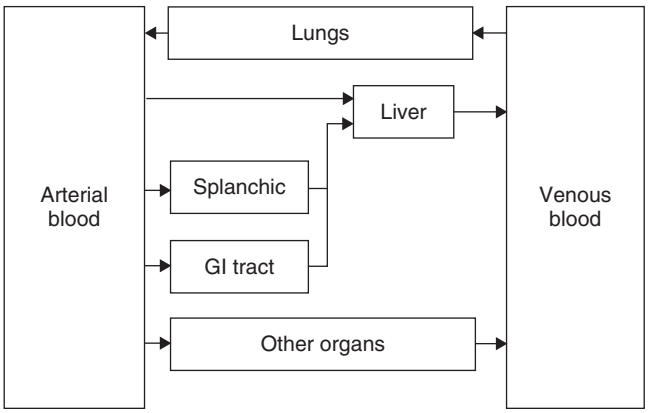

Figure 8.1 The structure of a generic PBPK model. Boxes ("compartments") represent blood pools (arterial and venous), organs, and tissues. Arrows represent blood flows between compartments. Different models differ primarily in the number of organs and tissues that are explicitly represented.

liver [102,103]. Sometimes organs are lumped into "central" and "peripheral" compartments [104,105]; this makes a model less physiological, but may make it easier to parameterize from *in vivo* data. Organs are commonly modeled as being well-stirred, with rapid equilibration of free compound between plasma and tissue; but taking typical values for compound permeability, it seems that this rapid equilibration may be an unrealistic assumption [105–107]. In any case, to model transporter effects, it is necessary to include permeability barriers in the organ model; organs are sometimes subdivided into blood, interstitial, and intracellular compartments [108]. The inclusion of transporter effects is a relatively recent development [109–115].

The modeling of absorption from and transport along the GI tract is an art in itself. Physiological models of the GI tract consist of a sequence of segments through which flow dissolved and undissolved compound (Fig. 8.2), with some modeling of dissolution, and with absorption of the dissolved compound [12,107,116,117]. Metabolism in the walls of the GI tract is also sometimes included [118,119].

The extent of distribution in PBPK models is predicted using a competition between partitioning into tissue lipids and binding to plasma proteins. A few variants on the idea are reported in the literature [120–122]; as yet there is no definitive method. The octanol–water partition coefficient is commonly used as a measure of lipophilicity; some groups use, instead, affinity to phospholipid membranes [123,124], and there does not seem to be agreement on the best measure of lipophilicity to use in general. It should be noted that these models rely on the partition coefficient directly rather than on its logarithm: an error of half a log unit in the logarithm translates to a threefold error in the partition coefficient and—for lipophilic compounds—a threefold error in the volume of distribution even before difficulties in the measurement of plasma protein binding are taken into account.

It is also necessary to parameterize the rate of metabolism in a PBPK model. If *in vivo* data are available, one can just optimize the model to fit the measured data, but if they are unavailable, one must predict the rate of metabolism using *in vitro* or *in silico* data. The prediction of metabolic clearance from *in vitro* data has recently

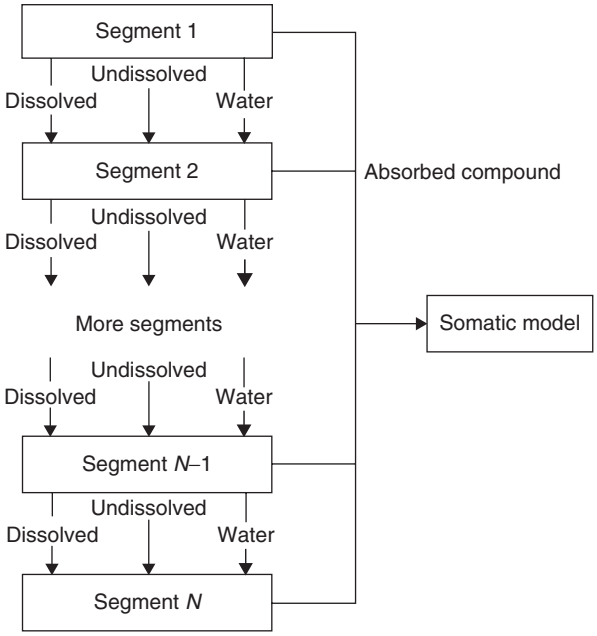

Figure 8.2 The generic structure of gastrointestinal tract absorption modeling in PBPK. The GI tract is treated as a series of segments (compartments) in sequence. Dissolved and undissolved compound plus water (more precisely the gastric or intestinal fluid) flows from a compartment to the next in the series. Dissolved compound is available for absorption across the GI tract epithelium into the body, which can be represented by a PBPK model (as in Fig. 8.1). Different models differ in the number of segments that are used to represent each part of the tract.

been reviewed [125]. Large uncertainties and variabilities in the *in vitro* data mean that, generally, predictions of *in vivo* clearance based on *in vitro* data alone can be statistical at best.

Some success is reported in using generic PBPK models for statistical predictions of pharmacokinetic parameters such as clearance [108,126] and, until the details of mechanistic extrapolation from *in vitro* to *in vivo* are resolved, this is perhaps the most that can be expected from predictive PBPK modeling in early drug discovery. Later in drug discovery and in drug development more effort can be expended on developing a customized and optimized model for a particular compound, when *in vivo* data are available for the calibration of such models. These optimized models can be used for explanatory purposes [127] and to design detailed *in vivo* trials that assess specific aspects of pharmacokinetics [128]. The detailed analysis of such models allows the determination of the controlling parameters and can give good mechanistic understanding of *in vivo* pharmacokinetics [127,129].

8.4.5.2 Interactions Between Compounds. The interaction most commonly modeled mechanistically is the inhibition of CYP-mediated metabolism. PBPK models are separately developed for each interacting compound using *in vitro* or *in vivo* data. These models are used to predict compound concentration at the critical sites for drug–drug interaction such as the liver or intestinal mucosa. *In vitro* interaction data

can then be inserted to link the two models. A relatively detailed model of metabolism is necessary: in general, compounds may be metabolized by multiple enzymes, so to model correctly the effect of the inhibition of a given enzyme, one must account for the relative contributions of all enzymes that significantly metabolize the compounds in question. For compounds that can be modeled with single-enzyme metabolism, this step is unnecessary, but this will not in general be the case. Interactions between more than two compounds can be handled in a similar way. Irreversible inhibition, in which a metabolizing enzyme is permanently inactivated, demands a different type of model to reversible inhibition. To predict irreversible inhibition, some idea is needed about protein turnover [130]; the actual protein concentration must be evolved as part of the simulation.

Successes for this approach have been reported in the literature (Table 8.5) [101,104,105,119,123,130–137]. Some use mechanistic PBPK modeling to predict single-compound pharmacokinetics; others use a more empirical approach in which physiological or semiphysiological models of single-compound pharmacokinetics are fit to *in vivo* data. The overwhelming majority study CYP3A, presumably because CYP3A-family enzymes are the most common drug metabolizers.

In principle, this general approach could also be used for interactions controlled by the inhibition of active transport; all that is needed is a mechanistic model of transporter activity *in vivo* [109–112,114] coupled to some *in vitro* transporter inhibition data.

TABLE 8.5 Physiological and Semiphysiological Metabolism Inhibition Models Found in the Literature

Model Type	Compounds	Inhibited Enzyme(s)	References
Whole body	Midazolam/itraconazole	CYP3A	Vossen *et al*. [123]
Whole body	Midazolam/clarithromycin	CYP3A	Quinney *et al*. [104]
Whole body	Midazolam/diltiazem	CYP3A	Zhang *et al*. [105]
Whole body	Midazolam (cimetidine/itraconazole/ erythromycin)	CYP3A	Yamano *et al*. [131]
Whole body	Various CYP3A4 substrates/various CYP3A4 inhibitors	CYP3A4	Fenneteau *et al*. [119]
Whole body	Midazolam/phase 1 compound	CYP3A4	Chenel *et al*. [132,133]
Whole body	Tesofensine/itraconazole	CYP3A4	Lehr *et al*. [134]
Whole body	CYP3A4 substrates/ketoconazole, verapamil	CYP3A4	Perdaems *et al*. [134]
Whole body	Diltiazem/triazolam/*N*-desmethyldiltiazem	CYP3A	Rowland Yeo *et al*. [130]
Whole body	Midazolam/ketoconazole	CYP3A	Chien *et al*. [136]
Whole body	Midazolam/macrolide antibiotics	CYP3A	Ito *et al*. [137]
Isolated perfused liver	*R*-bufuralol/ bunitrol/debrisoquine	CYP2D	Haddad *et al*. [101]

TABLE 8.6 Physiological and Semiphysiological Mechanistic Induction Models Found in the Literature

Model Type	Compound	Enzyme Induced	References
Whole body	2,3,7,8-Tetrachloro-dibenzo-*p*-dioxin	CYP1A1, CYP1B1	Kohn *et al*. [144]
Whole body	2,3,7,8-Tetrachloro-dibenzo-*p*-dioxin	CYP1A2	Emond *et al*. [140]
			Wang *et al*. [141]
			Santostefano *et al*. [142]
			Leung *et al*. [142]
Semiphysiological	Isoniazid	CYP2E1	Chien *et al*. [145]
Semiphysiological	Phenobarbital	Various	Magnusson *et al*. [146]
Semiphysiological	Artemisinin	Unspecified	Gordi *et al*. [147]
Semiphysiological	Cyclophosphamide	Unspecified	Hassan *et al*. [148]

But while these are all promising successes, it must of course be noted that, as discussed above, the quantitative prediction of even single-compound pharmacokinetics is challenging. It would be surprising if prediction of drug–drug interactions was simpler than the prediction of plasma clearance: a mechanistic model of metabolic drug–drug interaction requires, to some extent, an understanding of metabolic clearance. Given the challenges in the quantitative prediction of metabolic clearance from *in vitro* data [125], it is easy to see how difficult it is to accurately predict *in vivo* inhibition of metabolism.

Heuristic models of CYP induction relate the extent of *in vitro* induction to the rate of *in vivo* metabolism [138]. This essentially applies sensitivity analysis to the single-compound pharmacokinetics: develop a working model for the victim compound in base conditions and just rerun the model using various hypothetical rates of metabolism following induction.

Mechanistic models of enzyme induction require some concept of protein turnover and gene expression [139] and require the enzyme concentrations to be evolved as part of the simulation. There are two mechanisms for CYP induction that have been modeled. Both result from a change to the balance between protein synthesis and degradation: the more common mechanism is an increase in protein synthesis following nuclear receptor binding, the less common is a decrease in protein degradation due to stabilization caused by xenobiotic binding to the protein. Some of the mechanistic induction models found in the literature are given in Table 8.6 for induction by dioxin [140–144], isoniazid [145], phenobarbital [146], artemisinin [147], and cyclophosphamide [148]; these are all customized models for a particular interaction and are not *a priori* predictive.

8.5 INCORPORATING *IN SILICO* METHODS IN DRUG DISCOVERY

In previous sections, we have described a number of *in silico* methods and discussed some specific applications to predict drug metabolism and drug–drug interactions. These methods perform an important function by helping increase our understanding

of the properties of compounds that determine their fates and interactions *in vivo*. To have maximum impact, however, in a drug discovery program, the methods must support and inform the selection of (virtual) compounds to be synthesized or (real) compounds to be selected for screening assays. In this section, we consider ways whereby the methods can be incorporated into the drug discovery process, with a view to inform and help direct medicinal chemistry efforts toward designing compounds with required characteristics. In this respect, the methods so far described do not stand alone. The capability to beneficially employ them in drug discovery depends on

- the availability of necessary supporting software;
- appropriately skilled and experienced staff;
- effective communication between modelers, drug metabolism/pharmacokinetics (DMPK) scientists, and medicinal chemists.

This can be illustrated with a simple example in which systemic clearance is to be predicted in order, for instance, to estimate plasma half-life. In this case, the effect on predicted *in vivo* plasma clearance of uncertainty in a prediction of hepatic clearance diminishes in relative importance as the predicted hepatic clearance increases toward hepatic blood flow. This is intuitively apparent from the effect of blood flow on limiting clearance, but sensitivity and error analyses of PBPK models can quantify the effect, can process large numbers of compounds in a short time, and can do so for those less obvious relationships that can exist between model inputs and outputs. Use of sensitivity and error analysis can thus indicate the acceptable bounds in uncertainty and bias on any given prediction or set of predictions.

As a simple, illustrative, example, we can first consider the case of predicting lipophilicity, in the form of the octanol/water partition coefficient, P. The partition process can be viewed as the creation of cavities in the two solvents (octanol and water)—an endoergic process—and the insertion of a solute into the solvents, which gives rise to various solute/solvent interactions (hydrogen bonding, electrostatic, van der Waal's, etc.), which may be endo- or exoergic. The balance of these processes will determine the overall energy change, and hence equilibrium position, of the partitioning (Fig. 8.3). Many *in silico* methods exist for this property (reviewed in Refs 149 and 150), but the method of Abraham and coworkers [151] has made the above relationships explicit:

$$\log P_{\text{oct}} = a + b \cdot E - c \cdot Z + d \cdot A - e \cdot B + f \cdot V$$

where $\log P_{\text{oct}}$ is the logarithm (base 10) of P. E, S, A, B, and V are properties of the particular solute:

- E is the *e*xcess molar refraction.
- Z is the polari*z*ability/dipolarity.
- A is the hydrogen bond *a*cidity (donor capability).
- B is the hydrogen bond *b*asicity (acceptor capability).
- V is the McGowan's characteristic *v*olume.

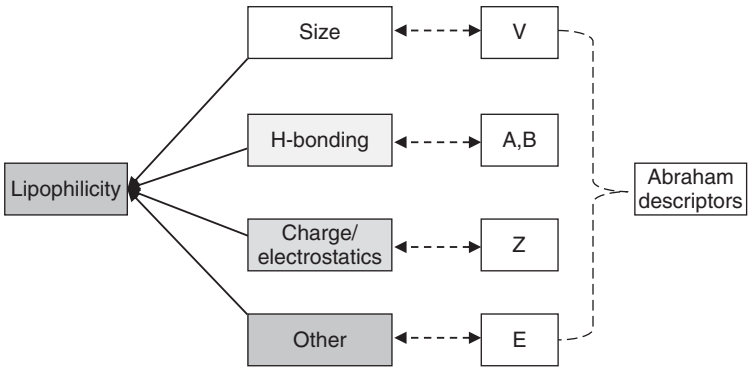

Figure 8.3 Relationships between Abraham descriptors, compound characteristics, and lipophilicity. V, A, B, Z, and E are the Abraham descriptors.

$a-f$ are constants with specific values for any partitioning system. For octanol/water, the partitioning is dominated by fV (i.e., molecular size, which favors octanol) and eB (hydrogen bond basicity, which favors water). E and V can be calculated for a compound directly from atom and bond contributions, whereas S, A, and B require manual summation from specified molecular fragments. While more accurate $\log P$ predictors currently exist [149,150], in terms of implementation in a drug discovery process, the Abraham method has the following points in its favor:

- The model involves a limited number of (six) unknown parameters, reducing the risk of being overtrained on reasonably sized training sets.
- The descriptors are readily interpretable in terms of fundamental properties of compounds.
- The small number, and the nature, of the descriptors facilitates their understanding by nonspecialists.
- The model is linear with respect to its descriptors, so that the way in which changes in any descriptor's value affects the predicted $\log P$ is clear.

The problems of interpreting and utilizing metabolism and drug interaction are harder than the analogous problem for $\log P$. Involving, as they do, interaction between a small molecule (or, in the case of drug interactions, two small molecules) and an enzyme or transporter protein active site, simple explanations are less frequent and correlations are weaker than for $\log P$. Additional properties, relating to molecular shape are more important and such properties are, of necessity, less amenable to simple interpretation. Lewis [152] has collated information from a number of sources to specify characteristics of individual human P450 isoforms, in a qualitative/semiquantitative manner in terms of planarity, acid/base characteristics, volume, and lipophilicity. Lewis [153] has also defined QSARs for substrates (and inhibitors) of the human P450 CYP2 family (2A6, 2B6, 2C8, 2C19, 2D6, and 2E1) in terms of a small number of readily definable descriptors, including lipophilicity, molecular weight, and hydrogen bonding capability. These QSARs have, however, been generated using small data sets so, while they can provide some useful insight into potentially important features that define

substrate specificity, the ability of the QSARs to extrapolate to other compounds is doubtful. Methods for predicting CYP450 substrate specificity have been reviewed by Ekins *et al*. [154] and de Groot [155].

Turning in more detail to the identification of CYP3A4 substrates, the key CYP450 isoform for drug metabolism in humans, a number of approaches, including pharmacophore [156], QSAR [72,73], and docking [157], have been attempted.

Balakin *et al*. [72] developed a model for the Michaelis constant, K_m, of CYP450 substrates. As we have discussed in Section 8.4.1, the Michaelis constant does not, by itself, predict the rate of metabolism of a compound. For this, the V_{max} is also required. The Michaelis constant does, however, provide information about binding to the active site, potential for nonlinearity in metabolism, and potential for drug–drug interactions mediated by inhibition of metabolism. For CYP3A4, Balakin *et al*. had experimental K_m values for 126 compounds, categorizing the values as low ($< 10\,\mu M$), medium ($10–100\,\mu M$), or high ($> 100\,\mu M$). For differentiating between low and high (ignoring medium) K_m values, six descriptors were found to be important. Listed in order of decreasing importance, these are as follows:

1. Zagreb (a measure of topological complexity).
2. Number of H-bond acceptors.
3. Number of rotatable bonds (a measure of flexibility and also correlated to size).
4. A measure of partial negative surface area (PNSA-1).
5. Number of H-bond donors.
6. $\log P_{oct}$.

For all except the two hydrogen bonding descriptors, the differences were statistically significant by a two-tailed t test (Table 8.7). It can thus be seen that the most important determinant between low and high K_m is a complex descriptor (Zagreb), whose interpretation is not obvious—and therefore whose utilization in guiding medicinal chemistry is not trivial. Consequently, to use the results of this model to guide chemistry, medicinal chemists would require appropriate computational tools and/or databases, either to screen libraries for compounds or to design compounds having appropriate properties [158–160]. Three of the properties (numbers of hydrogen bond acceptors and donors and number of rotatable bonds) are more amenable to quick assessment, as is $\log P_{oct}$ (on the assumption that $\log P_{oct}$ predictive software is almost ubiquitous and/or chemists would rapidly gain experience in estimating the lipophilicity

TABLE 8.7 Important Descriptors for Differentiating Between Strong and Weak Binders to Human CYP3A4

Property	Mean Value for Strong Binders[a]	Mean Value for Weak Binders[a]
Zagreb index	~ 210	~ 120
Number of rotatable bonds	9	4
PNSA-1	~ 260	~ 175
$\log P_{oct}$	~ 4.8	~ 2.6

[a] Strong and weak binders are distinguished by low ($<10\,\mu M$) and high ($> 100\,\mu M$) K_m, respectively.

within a particular series). Furthermore, by inspection of Table 8.7, it can be seen that the model predicts marked differences in $\log P_{\text{oct}}$ and the number of rotatable bonds for strong and weak binders, so there are some, at least, clear, readily identifiable differences between strong and weak binders.

Yap and Chen [73] adopted a different approach, identifying compounds as CYP3A4 substrates or nonsubstrates (368 and 334, respectively) on the basis of reports—or lack of such—in the literature of a compound being metabolized by a particular P450. The groups of descriptors that were selected as important during the modeling process were mostly composite, and therefore difficult to interpret, namely

- 3D-MoRSE—encode 3D features such as van der Waal's volume, electronegativities, and polarizabilities.
- Radial distribution function (RDF)—encode information about bond lengths, ring types, planar and nonplanar systems, and atom types.
- Randic molecular profiles—encode interactions between atoms in a molecule and information on molecular shape.

The simpler, noncomposite, descriptors ($\log P$, hydrogen bond acceptors and donors, etc.) provided less information about the substrate/nonsubstrate (and, analogously, the inhibitor/noninhibitor) classification of substrates with respect to CYP3A4. Thus, substrates had mean Moriguchi $\log P$ of 2.2 ± 1.9 compared with 1.6 ± 2.06 for nonsubstrates. This contrasts with the findings of Balakin [72], where low and high K_{m} compounds were differentiated by statistically significant mean $\log P$ of $\sim$4.8 and $\sim$2.6, respectively (Table 8.7). The difference arises both from the nature of the problem being addressed and the amount of information being sought. Identifying whether a compound is a substrate for metabolism of CYP3A4 is more difficult than identifying whether it binds tightly to the protein's active site, in which bulk-phase hydrophobic interactions can exert a significant effect. Yap and Chen were, furthermore, seeking a classification for all possible compounds into substrate/nonsubstrate, whereas Balakin *et al*. had restricted themselves to the widely separated classes of low and high K_{m} compounds and were not attempting to address the intermediate category of medium K_{m} (between 10 and 100 μM).

The study of Ung *et al*. on the activation of PXR also provides interesting insight into the interpretation of molecular descriptors and their potential for informing medicinal chemistry [94]. They point out the tremendous range of compounds that are known to be activators, and the range of structural features they exhibit, including aromatic and saturated multiring structures, halogen- and oxygen-rich compounds, and compounds with chain-like structures, such as discodermolide and β-carotene. This is consistent with PXR's flexible, predominantly hydrophobic, binding site that permits binding of a wide range of substrate sizes in multiple conformations [161]. Despite the flexibility of the site, which might lead to the expectation that complex descriptors would be preferred in the model, certain molecular descriptors were selected that are related to identifiable structural features in compounds. Table 8.8 lists some of these descriptors and their values for activators and nonactivators. These descriptors relate to size, polar surface area, flexibility, and the number/electronic and topological state of certain types of heteroatoms (nitrogen and halogens). The number of rotatable bonds and Kier molecular flexibility index indicate that, on average, activators are slightly smaller and less flexible than nonactivators. Polar surface area is slightly lower for activators

TABLE 8.8 Selected Descriptors for Differentiating Between Activators and Nonactivators of Human Pregnane X Receptor

Property	Activators[a]	Nonactivators[a]
Number of rotatable bonds	4.45	5.99
Kier molecular flexibility index	5.84	6.24
Polar surface area	61.9	69.5
Number of halogens	1.16	0.8
Number of chlorines	1.02	0.27
Number of nitrogen atoms	0.8	1.79
Atom-type E-state sum for chlorine	6.33	1.63
Atom-type E-state sum for >NH	0	0.45
Atom-type E-state sum for = N-	0.31	1.15
Atom-type E-state sum for >N-	0.15	0.94

[a]Figures are the mean values for compounds predicted by the model to be activators or nonactivators.

(although it might not be so when normalized for molecular size). Activators tend to have more halogen and less nitrogen atoms than nonactivators. However, for simple descriptors (numbers of rotatable bonds, halogens, chlorines, and nitrogens), the differences are small, particularly when taking into account that the values are limited to integers; these descriptors may not be particularly informative. Thus, the presence of four or five rotatable bonds is more likely to indicate an activator than nonactivator, while six is more likely to indicate an activator. One chlorine is more likely to indicate an activator than a nonactivator, but a significant proportion of nonactivators must have at least one chlorine. For halogens, in general, there is very little difference between the mean values for activators and nonactivators, both of which are close to 1. Two nitrogen atoms in a molecule are more likely to indicate a nonactivator, while one is more likely to indicate an activator.

While being less amenable to immediate interpretation, the E-state indices [162], which encode information about the electronic and topological environment of particular atoms, offer the potential for greater discrimination on a continuous scale. Thus, the E-state values for chlorine and for different bonding configurations of nitrogen (see the last five rows of Table 8.8) provide such information for discriminating between activators and nonactivators of human PXR, in which the nonintegral mean values and values between them can have a true meaning. Thus, it can be immediately inferred from the mean number of chlorine atoms in predicted nonactivators (0.27) that a significant proportion of them must have at least one chlorine atom. In contrast, the nonintegral values of the E-state sums of chlorines in predicted activators and nonactivators (6.33 and 1.63, respectively) provide information about the number of chlorines, and the modulating electronic field effects of the other atoms in the molecule, weighted by their topological distance (i.e., the number of bonds) from the chlorine(s). Similarly, the mean numbers of nitrogen atoms in predicted activators (0.8) and nonactivators (1.8) indicate some uncertainty around the status of compounds having one nitrogen atom. Again, the relative values of the E-state indices for the different configurations of nitrogen are well separated, helping distinguish between the two classes of compounds. The drawback in using such descriptors has already been addressed in reference to the model of Balakin *et al*. for CYP3A4 binding.

A further means of exploiting predictions of drug metabolism and drug interactions is to use them as inputs to appropriate PBPK models, linking the prediction of a drug's properties to its expected behavior *in vivo*. Sensitivity/error analysis of a PBPK model (i.e., the quantification of the dependence of its outputs on the value of any of its inputs) will provide information about how uncertainty and/or bias in the predictions of a drug's property will affect its expected behavior *in vivo*. This type of analysis makes use of the ease of rerunning simulation models with appropriate variation of the input and can be used in one of two related ways:

1. By scanning an input value through a (potentially hypothetical) range and recording the predictions of the PBPK model for each input value, the sensitivity of the predictions to the value of the input can be ascertained. This is known as *sensitivity analysis*.
2. By sampling the predicted distribution of the input (which can be generated from the error statistics of the model that generates the input), the corresponding distributions in the PBPK model's predictions can be determined. This is an aspect of *error analysis*.

Because PBPK models require quantitative inputs on the whole, the most effective use of this approach requires that the models for the input properties generate quantitative values or classifications with classes with well-defined upper and lower limits. As we have discussed, many models for predicting drug metabolism and drug–drug interactions are binary classification models (i.e., for substrate/non–substrate and inhibitor/noninhibitor), so there is still significant progress to be made in model development in this area to realize the prediction of *in vivo* behavior from *in silico* metabolism and drug interaction models.

We complete this section by emphasizing that DMPK represents just one part of a drug's property set and that the significance of its DMPK properties—whether determined by experiment or by *in silico* modeling—has to be viewed in the context of its other properties, particularly its on-target and off-target pharmacology. Of the properties we have discussed in this chapter, it is particularly important to view those that can cause drug–drug interaction (i.e., inhibition of metabolism and induction of metabolism/transport) in this context. The propensity of any potential new drug to actually cause such interactions *in vivo* is dependent on the concentrations it is required to achieve *in vivo*. These are, in turn, governed primarily by its activity against its target(s). Low activities against targets will require high concentrations *in vivo*, increasing the likelihood of interactions with other drugs. In a similar manner, the most effective use of PBPK models that can predict plasma concentrations is dependent on knowledge of the expected therapeutic window of the compound, which is in turn determined by its primary pharmacology and its dose-limiting toxicity. The most efficient screening strategies will be those that bring together useful data in pharmacology, toxicity, and DMPK, each having similar costs of information relevant to the particular stage of the discovery process. This could entail a combination of *in vitro* and *in silico* data, the balance between the two depending on the cost of information. High throughput *in vitro* screens for drug metabolism properties [66,163] are able to generate the large amounts of data required for producing reliable models and also reduce the cost of generating primary assay data, potentially reducing the need for predictive models. Set against this, the use of systems for automated QSAR model development [164] can

drastically reduce the time and cost of model development and their regeneration from evolving data sets. If target activity can be predicted *in silico*, so permitting hits to be identified virtually, then the ability to predict metabolic and *in vivo* drug–drug interaction properties can result in a significant recovered opportunity cost by permitting deferral of the synthesis of potentially unsuitable compounds [66].

8.6 CONCLUSIONS

In silico methods are making significant contributions to our growing knowledge of the factors that determine how drugs and other xenobiotics are processed *in vivo*. Numerous methods have been developed and employed over the past 30 years. In this chapter, we have endeavored to do justice to a number of these methods, to their contributions to the furthering of knowledge and their practical applications in the drug discovery process. We have described two major areas of application. Each area is extremely complex in its own right; this must be borne in mind when assessing current capabilities and future progress.

The first is concerned with the prediction of interactions between drugs and individual proteins (drug-metabolizing enzymes, transporters, and nuclear receptors), including the simultaneous interactions of multiple drugs with an enzyme or transporter that can give rise to mutual or unilateral inhibition of metabolism or transport. The complexity in this area arises from the well-documented promiscuity of the proteins involved. A given enzyme/transporter/receptor can metabolize/transport (or be inhibited by)/be activated by multiple xenobiotics. A given xenobiotic can be metabolized by/transported by (or inhibit)/activate multiple proteins. The challenge of *in silico* modeling is to identify those interactions that will/will not occur, and the strength of those interactions, as measured by rates of metabolism or transport, binding constants, inhibition constants, and so on. For many of the applications we have described, a large part of the challenge entails the collation of a reliable data set of a large enough size for model development. For any given problem, it is not guaranteed that a suitable data set can be generated, because of the huge number of combinations of xenobiotic/protein interactions that must be considered. For example, development of a QSAR model to discriminate between the metabolism of compounds by (say) four isoforms of CYP450 requires reliable data for each compound and for each isoform. It frequently arises that the absence of reported metabolism of a compound by a particular isoform is taken to indicate that the compound is not metabolized by that isoform but this is not, of course, the same thing. Without definitive data, such assumptions can compromise the modeling process from the start. Extend this to the myriad other interactions of interest—with UGT enzymes, ABC transporters, nuclear receptors, and so on, and the whole model development and application structure can be fatally undermined. Improvements in reliable, high throughput, low volume assay methodologies will significantly increase the availability of reliable data for model development in the future.

In assessing the progress and utility of *in silico* model development in this area, the differences between practice in the drug discovery industry and the majority of reports in the scientific literature must be borne in mind. Many the papers we have discussed in Section 8.4 have shown modest predictive ability for a range of properties related

to drug metabolism and drug–drug interactions. We should not be surprised with these results, given the data available, the relatively small data sets, and the diversity of the compounds that can be processed by, or bind to, the proteins in question. These types of models are generally termed *global* models, in that they are generated using data for compounds scattered across a wide range of chemistry—almost invariably using data on marketed drugs. This is not necessarily because of the desire to generate global models, but because of the limitations of data availability in the public domain. The practicalities of drug discovery are different. Discovery projects focus on a small number of specific areas of chemistry, so chemical diversity (and the problems introduced by this) is much reduced. The need for a model to extrapolate or interpolate into an area of chemical space that is poorly represented in the training set is consequently reduced. This matter has been discussed by Weaver and Gleeson [49] in relation to the development of QSPR models for predicting plasma protein binding and inhibition of CYP3A4-mediated metabolism. It seems reasonable to regard the prediction statistics of global models as a lower bound on what should be achievable with models on more focused areas of chemistry.

The second application area we have considered is that of PBPK modeling, by means of which data (whether from *in vitro* or *in silico* sources) on drug metabolism and drug–drug interactions can be integrated to predict *in vivo* outcomes. The key modeling challenges for the prediction of drug–drug interactions are (i) to predict the concentrations of drugs at the active sites of the relevant enzymes/transporters and (ii) to be able to appropriately scale, if necessary, kinetic parameters (e.g., V_{max}, K_m, and K_i) to the *in vivo* case. These two are dependent, as the ability to predict concentrations depends, in part, on the scaling of kinetic parameters of drug transport and metabolism. The complexity in this case arises from the multiple potential *in vivo* processes (absorption, distribution, metabolism and elimination, each potentially via multiple competing routes) to which a xenobiotic can be subject and which affect prediction of drug concentration at the active site(s). In most cases, in drug discovery and development, appropriate data required to model some—maybe most—of these processes will be lacking. Therefore, predictions must be interpreted in the light of what is, and what is not, known about the processes affecting the fate(s) of a particular compound *in vivo*. The evolution of predictive PBPK models is absolutely dependent on the development of appropriate assays for these processes and the knowledge to enable data to be successfully scaled to the *in vivo* case. This knowledge frequently lags significantly behind assay development and is frequently a significant impediment to new applications of PBPK modeling. Having said that, we have described a number of applications where the technique has already been used to provide valuable information and interpretation concerning *in vivo* drug metabolism and drug–drug interactions.

It can be seen that *in silico* techniques are able to provide valuable insight and predictivity regarding drug metabolism and interactions *in vitro* and *in vivo*. Significant work remains to be done on improving predictability for known routes of metabolism and transport, and the challenge is continually growing as further routes are discovered, and hence opportunities for drug–drug interactions to arise are identified. In fact, this very increase in the known complexities involved in drug metabolism and interactions requires that the contribution of *in silico* methods in this area increases to help keep drug discovery costs under control and to inform and guide the discovery selection process.

REFERENCES

1. Wen B, Ma L, Zhu M. Bioactivation of the tricyclic antidepressant amitriptyline and its metabolite nortriptyline to arene oxide intermediates in human liver microsomes and recombinant P450s. Chem Biol Interact 2008;173:59–67.
2. Dockens RC, Salazar DE, Fulmor IE, *et al*. Pharmacokinetics of a newly identified active metabolite of buspirone after administration of buspirone over its therapeutic dose range. J Clin Pharmacol 2006;46:1308–1312.
3. Liederer BM, Borchardt RT. Enzymes involved in the bioconversion of ester-based prodrugs. J Pharm Sci 2006;95:1177–1195.
4. Wiltshire H, Hirankarn S, Farrell C, *et al*. Pharmacokinetic profile of ganciclovir after its oral administration and from its prodrug, valganciclovir, in solid organ transplant recipients. Clin Pharmacokinet 2005;44:495–507.
5. Wrighton SA, Stevens JC. The human hepatic cytochromes P450 involved in drug metabolism. Crit Rev Toxicol 1992;22:1–21.
6. Gibson GG, Skett P. Introduction to drug metabolism. Cheltenham: Nelson Thornes; 2001.
7. Hamada H, Tsuruo T. Purification of the 170- to 180-kilodalton membrane glycoprotein associated with multidrug resistance. 170- to 180-kilodalton membrane glycoprotein is an ATPase. J Biol Chem 1988;263:1454–1458.
8. Makhey VD, Guo A, Norris DA, *et al*. Characterization of the regional intestinal kinetics of drug efflux in rat and human intestine and in Caco-2 cells. Pharm Res 1998;15: 1160–1167.
9. Cordon-Cardo C, O'Brien JP, Casals D, *et al*. Multidrug-resistance gene (P-glycoprotein) is expressed by endothelial cells at blood-brain barrier sites. Proc Natl Acad Sci U S A 1989;86:695–698.
10. Shitara Y, Horie T, Sugiyama Y. Transporters as a determinant of drug clearance and tissue distribution. Eur J Pharm Sci 2006;27:425–446.
11. Kalliokoski A, Niemi M. Impact of OATP transporters on pharmacokinetics. Br J Pharmacol 2009;158:693–705.
12. Bolger MB, Lukacova V, Woltosz WS. Simulations of the nonlinear dose dependence for substrates of influx and efflux transporters in the human intestine. AAPS J 2009;11:353–363.
13. Lee W, Glaeser H, Smith LH, *et al*. Polymorphisms in human organic anion-transporting polypeptide 1A2 (OATP1A2): implications for altered drug disposition and central nervous system drug entry. J Biol Chem 2005;280:9610–9617.
14. Kalliokoski A, Neuvonen M, Neuvonen PJ, *et al*. The effect of SLCO1B1 polymorphism on repaglinide pharmacokinetics persists over a wide dose range. Br J Clin Pharmacol 2008;66:818–825.
15. Xiang X, Jada SR, Li HH, *et al*. Pharmacogenetics of SLCO1B1 gene and the impact of *1b and *15 haplotypes on irinotecan disposition in Asian cancer patients. Pharmacogenet Genom 2006;16:683–691.
16. Thummel KE, Wilkinson GR. *In vitro* and *in vivo* drug interactions involving human CYP3A. Annu Rev Pharmacol Toxicol 1998;38:389–430.
17. Zhang L, Zhang YD, Zhao P, *et al*. Predicting drug-drug interactions: an FDA perspective. AAPS J 2009;11:300–306.
18. Marroum PJ, Uppoor RS, Parmelee T, *et al*. *In vivo* drug-drug interaction studies–a survey of all new molecular entities approved from 1987 to 1997. Clin Pharmacol Ther 2000;68:280–285.
19. Aronov AM. Common pharmacophores for uncharged human ether-a-go-go-related gene (hERG) blockers. J Med Chem 2006;49:6917–6921.
20. Waring MJ, Johnstone C. A quantitative assessment of hERG liability as a function of lipophilicity. Bioorg Med Chem Lett 2007;17:1759–1764.

21. Wandel C, Kim RB, Guengerich FP, *et al*. Mibefradil is a P-glycoprotein substrate and a potent inhibitor of both P-glycoprotein and CYP3A *in vitro*. Drug Metab Dispos 2000;28:895–898.
22. Honkakoski P, Negishi M. Regulation of cytochrome P450 (CYP) genes by nuclear receptors. Biochem J 2000;347:321–337.
23. Back DJ, Orme ML. Pharmacokinetic drug interactions with oral contraceptives. Clin Pharmacokinet 1990;18:472–484.
24. Backman JT, Olkkola KT, Neuvonen PJ. Rifampin drastically reduces plasma concentrations and effects of oral midazolam. Clin Pharmacol Ther 1996;59:7–13.
25. Kalliokoski A, Backman JT, Kurkinen KJ, *et al*. Effects of gemfibrozil and atorvastatin on the pharmacokinetics of repaglinide in relation to SLCO1B1 polymorphism. Clin Pharmacol Ther 2008;84:488–496.
26. Simonson SG, Raza A, Martin PD, *et al*. Rosuvastatin pharmacokinetics in heart transplant recipients administered an antirejection regimen including cyclosporine. Clin Pharmacol Ther 2004;76:167–177.
27. Lau YY, Okochi H, Huang Y, *et al*. Pharmacokinetics of atorvastatin and its hydroxy metabolites in rats and the effects of concomitant rifampicin single doses: relevance of first-pass effect from hepatic uptake transporters, and intestinal and hepatic metabolism. Drug Metab Dispos 2006;34:1175–1181.
28. Lau YY, Huang Y, Frassetto L, *et al*. effect of OATP1B transporter inhibition on the pharmacokinetics of atorvastatin in healthy volunteers. Clin Pharmacol Ther 2007;81:194–204.
29. Wan H, Holmen AG. High throughput screening of physicochemical properties and *in vitro* ADME profiling in drug discovery. Comb Chem High Throughput Screen 2009;12:315–329.
30. Ekins S, Mestres J, Testa B. In silico pharmacology for drug discovery: applications to targets and beyond. Br J Pharmacol 2007;152:21–37.
31. Ekins S, Mestres J, Testa B. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. Br J Pharmacol 2007;152:9–20.
32. Lagorce D, Reynes C, Camproux AC, *et al*. Silico ADME/Tox predictions. In: Tsaioun K, Kates SA, editors. ADMET for medicinal chemists: a practical guide. Hoboken (NJ): John Wiley and Sons, Inc.; 2011. pp. 29–124.
33. van De Waterbeemd H, Gifford E. ADMET in silico modelling: towards prediction paradise? Nat Rev Drug Discov 2003;2:192–204.
34. van De Waterbeemd H, Smith DA, Beaumont K, *et al*. Property-based design: optimization of drug absorption and pharmacokinetics. J Med Chem 2001;44:1313–1333.
35. Penzotti JE, Landrum GA, Putta S. Building predictive ADMET models for early decisions in drug discovery. Curr Opin Drug Discov Devel 2004;7:49–61.
36. Khakar PS. Two-dimensional (2D) in silico models for absorption, distribution, metabolism, excretion and toxicity (ADME/T) in drug discovery. Curr Top Med Chem 2010;10:116–126.
37. Matthews EJ, Contrera JF. In silico approaches to explore toxicity end points: issues and concerns for estimating human health effects. Expert Opin Drug Metab Toxicol 2007;3:125–134.
38. Gleeson MP. Plasma protein binding affinity and its relationship to molecular structure: an in-silico analysis. J Med Chem 2007;50:101–112.
39. Yamashita F, Hashida M. In silico approaches for predicting ADME properties of drugs. Drug Metab Pharmacokinet 2004;19:327–338.
40. Kortagere S, Krasowski MD, Ekins S. The importance of discerning shape in molecular pharmacology. Trends Pharmacol Sci 2009;30:138–147.
41. Feenstra KA, De Graaf C, Vermeulen NPE. Cytochrome P450 protein modelling and ligand docking. In: Rodrigues AD, editor. Drug-drug interactions. New York: Informa Healthcare; 2008. pp. 435–470.

42. Kemp CA, Marechal JD, Sutcliffe MJ. Progress in cytochrome P450 active site modeling. Arch Biochem Biophys 2005;433:361–368.
43. Marechal JD, Yu J, Brown S, *et al*. *In silico* and *in vitro* screening for inhibition of cytochrome P450 CYP3A4 by comedications commonly used by patients with cancer. Drug Metab Dispos 2006;34:534–538.
44. Hu Y, Unwalla R, Denny RA, *et al*. Development of QSAR models for microsomal stability: identification of good and bad structural features for rat, human and mouse microsomal stability. J Comput Aided Mol Des 2010;24:23–35.
45. Li H, Sun J, Fan X, *et al*. Considerations and recent advances in QSAR models for cytochrome P450-mediated drug metabolism prediction. J Comput Aided Mol Des 2008;22:843–855.
46. Crivori P, Poggesi I. Computational approaches for predicting CYP-related metabolism properties in the screening of new drugs. Eur J Med Chem 2006;41:795–808.
47. Ekins S, Kim RB, Leake BF, *et al*. Three-dimensional quantitative structure-activity relationships of inhibitors of P-glycoprotein. Mol Pharmacol 2002;61:964–973.
48. Todeschini R, Consonni V. Handbook of molecular descriptors. Weinheim, Germany: WILEY-VCH Verlag GmbH; 2000.
49. Weaver S, Gleeson MP. The importance of the domain of applicability in QSAR modeling. J Mol Graph Model 2008;26:1315–1326.
50. Hawkins DM, Basak SC, Shi X. QSAR with few compounds and many features. J Chem Inf Comput Sci 2001;41:663–670.
51. Guner OF. History and evolution of the pharmacophore concept in computer-aided drug design. Curr Top Med Chem 2002;2:1321–1332.
52. Bordas B, Komives T, Lopata A. Ligand-based computer-aided pesticide design. A review of applications of the CoMFA and CoMSIA methodologies. Pest Manage Sci 2003;59:393–400.
53. Murumkar PR, Zambre VP, Yadav MR. Development of predictive pharmacophore model for in silico screening, and 3D QSAR CoMFA and CoMSIA studies for lead optimization, for designing of potent tumor necrosis factor alpha converting enzyme inhibitors. J Comput Aided Mol Des 2010;24:143–156.
54. Verma J, Khedkar VM, Coutinho EC. 3D-QSAR in drug design–a review. Curr Top Med Chem 2010;10:95–115.
55. Sippl W. 3D-QSAR-applications, recent advances and limitations. In: Puzyn T, Leszczynski J, Cronin MT, editors. Recent advances in QSAR studies: methods and applications. Heidelberg: Springer Science + Business Media B.V; 2010. pp. 103–125.
56. van De Waterbeemd H, Carter RE, Grassy G, *et al*. Glossary of terms in computational drug design (IUPAC Recommendations 1997). Pure Appl Chem 1997;69:1137–1152.
57. Mason JS, Morize I, Menard PR, *et al*. New 4-point pharmacophore method for molecular similarity and diversity applications: overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. J Med Chem 1999;42:3251–3264.
58. Lukacova V, Balaz S. Multimode ligand binding in receptor site modeling: implementation in CoMFA. J Chem Inf Comput Sci 2003;43:2093–2105.
59. Klebe G. Comparative molecular similarity indices analysis: CoMSIA. In: Kubinyi H, Folkers G, Martin YC, editors. 3D QSAR in drug design. Netherlands: Springer; 2002.
60. Suh M, Park S, Lee H. Comparison of QSAR methods (CoMFA, CoMSIA, HQSAR) of anticancer 1-N-substituted imidazoquinoline-4,9-dione derivatives. Bull Korean Chem Soc 2002;23:417–422.
61. Kubinyi H. Comparative molecular field analysis (CoMFA). In: Schleyer PVR, Allinger NL, Clark T, *et al*., editors. Encyclopedia of Computational Chemistry. Chichester: John Wiley and Sons; 1998.

62. Schwede T, Kopp J, Guex N, *et al*. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res 2003;31:3381–3385.
63. Shan Y, Kim ET, Eastwood MP, *et al*. How does a drug molecule find its target binding site? J Am Chem Soc 2011;133:9181–9183.
64. Kitchen DB, Decornez H, Furr JR, *et al*. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 2004;3:935–949.
65. Sousa SF, Fernandes PA, Ramos MJ. Protein-ligand docking: current status and future challenges. Proteins 2006;65:15–26.
66. Chang C, Duignan DB, Johnson KD, *et al*. The development and validation of a computational model to predict rat liver microsomal clearance. J Pharm Sci 2009;98:2857–2867.
67. Hansch C, Lien EJ, Helmer F. Structure–activity correlations in the metabolism of drugs. Arch Biochem Biophys 1968;128:319–330.
68. Veith H, Southall N, Huang R, *et al*. Comprehensive characterization of cytochrome P450 isozyme selectivity across chemical libraries. Nat Biotechnol 2009;27:1050–1055.
69. Bertz RJ, Granneman GR. Use of *in vitro* and *in vivo* data to estimate the likelihood of metabolic pharmacokinetic interactions. Clin Pharmacokinet 1997;32:210–258.
70. Sorich MJ, Miners JO, McKinnon RA, *et al*. Comparison of linear and nonlinear classification algorithms for the prediction of drug and chemical metabolism by human UDP-glucuronosyltransferase isoforms. J Chem Inf Comput Sci 2003;43:2019–2024.
71. Sorich MJ, Miners JO, McKinnon RA, *et al*. Multiple pharmacophores for the investigation of human UDP-glucuronosyltransferase isoform substrate selectivity. Mol Pharmacol 2004;65:301–308.
72. Balakin KV, Ekins S, Bugrim A, *et al*. Kohonen maps for prediction of binding to human cytochrome P450 3A4. Drug Metab Dispos 2004;32:1183–1189.
73. Yap CW, Chen YZ. Prediction of cytochrome P450 3A4, 2D6, and 2C9 inhibitors and substrates by using support vector machines. J Chem Inf Model 2005;45:982–992.
74. Yu J, Paine MJ, Marechal JD, *et al*. *In silico* prediction of drug binding to CYP2D6: identification of a new metabolite of metoclopramide. Drug Metab Dispos 2006;34:1386–1392.
75. Seifert A, Tatzel S, Schmid RD, *et al*. Multiple molecular dynamics simulations of human P450 monooxygenase CYP2C9: the molecular basis of substrate binding and regioselectivity toward warfarin. Proteins 2006;64:147–155.
76. Santos R, Hritz J, Oostenbrink C. Role of water in molecular docking simulations of cytochrome P450 2D6. J Chem Inf Model 2010;50:146–154.
77. Singh SB, Shen LQ, Walker MJ, *et al*. A model for predicting likely sites of CYP3A4-mediated metabolism on drug-like molecules. J Med Chem 2003;46:1330–1336.
78. Zheng M, Luo X, Shen Q, *et al*. Site of metabolism prediction for six biotransformations mediated by cytochromes P450. Bioinformatics 2009;25:1251–1258.
79. McNamara JP, Hillier IH. Semi-empirical molecular orbital methods including dispersion corrections for the accurate prediction of the full range of intermolecular interactions in biomolecules. Phys Chem Chem Phys 2007;9:2362–2370.
80. Kriegl JM, Eriksson L, Arnhold T, *et al*. Multivariate modeling of cytochrome P450 3A4 inhibition. Eur J Pharm Sci 2005;24:451–463.
81. Hudelson MG, Ketkar NS, Holder LB, *et al*. High confidence predictions of drug-drug interactions: predicting affinities for cytochrome P450 2C9 with multiple computational methods. J Med Chem 2008;51:648–654.
82. Vasanthanathan P, Taboureau O, Oostenbrink C, *et al*. Classification of cytochrome P450 1A2 inhibitors and noninhibitors by machine learning techniques. Drug Metab Dispos 2009;37:658–664.
83. Chang C, Pang KS, Swaan PW, *et al*. Comparative pharmacophore modeling of organic anion transporting polypeptides: a meta-analysis of rat Oatp1a1 and human OATP1B1. J Pharmacol Exp Ther 2005;314:533–541.

84. Osterberg T, Norinder U. Theoretical calculation and prediction of P-glycoprotein-interacting drugs using MolSurf parametrization and PLS statistics. Eur J Pharm Sci 2000;10:295–303.
85. Bednarczyk D, Ekins S, Wikel JH, *et al*. Influence of molecular structure on substrate binding to the human organic cation transporter, hOCT1. Mol Pharmacol 2003;63:489–498.
86. Chang C, Swaan PW, Ngo LY, *et al*. Molecular requirements of the human nucleoside transporters hCNT1, hCNT2, and hENT1. Mol Pharmacol 2004;65:558–570.
87. Gombar VK, Polli JW, Humphreys JE, *et al*. Predicting P-glycoprotein substrates by a quantitative structure-activity relationship model. J Pharm Sci 2004;93:957–968.
88. Xue Y, Yap CW, Sun LZ, *et al*. Prediction of P-glycoprotein substrates by a support vector machine approach. J Chem Inf Comput Sci 2004;44:1497–1505.
89. Suhre WM, Ekins S, Chang C, *et al*. Molecular determinants of substrate/inhibitor binding to the human and rabbit renal organic cation transporters hOCT2 and rbOCT2. Mol Pharmacol 2005;67:1067–1077.
90. Chang C, Bahadduri PM, Polli JE, *et al*. Rapid identification of P-glycoprotein substrates and inhibitors. Drug Metab Dispos 2006;34:1976–1984.
91. Zalloum HM, Taha MO. Development of predictive *in silico* model for cyclosporine- and aureobasidin-based P-glycoprotein inhibitors employing receptor surface analysis. J Mol Graph Model 2008;27:439–451.
92. Ekins S, Kim RB, Leake BF, *et al*. Application of three-dimensional quantitative structure-activity relationships of P-glycoprotein inhibitors and substrates. Mol Pharmacol 2002;61:974–981.
93. Ekins S, Erickson JA. A pharmacophore for human pregnane X receptor ligands. Drug Metab Dispos 2002;30:96–99.
94. Ung CY, Li H, Yap CW, *et al*. *In silico* prediction of pregnane X receptor activators by machine learning approaches. Mol Pharmacol 2007;71:158–168.
95. Khandelwal A, Krasowski MD, Reschly EJ, *et al*. Machine learning methods and docking for predicting human pregnane X receptor activation. Chem Res Toxicol 2008;21:1457–1467.
96. Ekins S, Kortagere S, Iyer M, *et al*. Challenges predicting ligand-receptor interactions of promiscuous proteins: the nuclear receptor PXR. PLoS Comput Biol 2009;5:e1000594.
97. Nestorov I. Whole body pharmacokinetic models. Clin Pharmacokinet 2003;42:883–908.
98. Jones HM, Gardner IB, Watson KJ. Modelling and PBPK simulation in drug discovery. AAPS J 2009;11:155–166.
99. De Buck SS, Sinha VK, Fenu LA, *et al*. Prediction of human pharmacokinetics using physiologically based modelling: a retrospective analysis of 26 clinically tested drugs. Drug Metab Dispos 2007;35:1766–1780.
100. De Buck SS, Sinha VK, Fenu LA, *et al*. The prediction of drug metabolism, tissue distribution, and bioavailability of 50 structurally diverse compounds in rat using mechanism-based absorption, distribution, and metabolism prediction tools. Drug Metab Dispos 2007;35:649–659.
101. Haddad S, Poulin P, Funk C. Extrapolating *in vitro* metabolic interactions to isolated perfused liver: Predictions of metabolic interactions between R-bufuralol, bunitrolol, and debrisoquine. J Pharm Sci 2010;99:4406–4426.
102. Weisiger RA. Dissociation from albumin: a potentially rate-limiting step in the clearance of substances by the liver. Proc Natl Acad Sci U S A 1985;82:1563–1567.
103. Baker M, Parton T. Kinetic determinants of hepatic clearance: plasma protein binding and hepatic uptake. Xenobiotica 2007;37:1110–1134.
104. Quinney SK, Zhang X, Lucksiri A, *et al*. Physiologically based pharmacokinetic model of mechanism-based inhibition of CYP3A by clarithromycin. Drug Metab Dispos 2010;38:241–248.

105. Zhang X, Quinney SK, Gorski JC, *et al*. Semiphysiologically based pharmacokinetic models for the inhibition of midazolam clearance by diltiazem and its major metabolite. Drug Metab Dispos 2009;37:1587–1597.
106. Huang L, Berry L, Ganga S, *et al*. Relationship between passive permeability, efflux, and predictability of clearance from *in vitro* metabolic intrinsic clearance. Drug Metab Dispos 2010;38:223–231.
107. Metcalfe PD, Thomas S. Challenges in the prediction and modeling of oral absorption and bioavailability. Curr Opin Drug Discov Devel 2010;13:104–110.
108. Brightman FA, Leahy DE, Searle GE, *et al*. Application of a generic physiologically based pharmacokinetic model to the estimation of xenobiotic levels in rat plasma. Drug Metab Dispos 2006;34:84–93.
109. Poirier A, Lave T, Portmann R, *et al*. Design, data analysis, and simulation of *in vitro* drug transport kinetic experiments using a mechanistic *in vitro* model. Drug Metab Dispos 2008;36:2434–2444.
110. Fenneteau F, Turgeon J, Couture L, *et al*. Assessing drug distribution in tissues expressing P-glycoprotein through physiologically based pharmacokinetic modeling: model structure and parameters determination. Theor Biol Med Model 2009;6:2.
111. Poirier A, Funk C, Scherrmann JM, *et al*. Mechanistic modeling of hepatic transport from cells to whole body: application to napsagatran and fexofenadine. Mol Pharm 2009;6:1716–1733.
112. Poirier A, Cascais AC, Funk C, *et al*. Prediction of pharmacokinetic profile of valsartan in human based on *in vitro* uptake transport data. J Pharmacokinet Pharmacodyn 2009;36:585–611.
113. Poirier A, Cascais AC, Funk C, *et al*. Prediction of pharmacokinetic profile of valsartan in humans based on *in vitro* uptake-transport data. Chem Biodivers 2009;6: 1975–1987.
114. Fenneteau F, Li J, Nekka F. Assessing drug distribution in tissues expressing P-glycoprotein using physiologically based pharmacokinetic modeling: identification of important model parameters through global sensitivity analysis. J Pharmacokinet Pharmacodyn 2009;36:495–522.
115. Watanabe T, Kusuhara H, Sugiyama Y. Application of physiologically based pharmacokinetic modeling and clearance concept to drugs showing transporter-mediated distribution and clearance in humans. J Pharmacokinet Pharmacodyn 2010;37:575–590.
116. Willmann S, Schmitt W, Keldenich J, *et al*. A physiological model for the estimation of the fraction dose absorbed in humans. J Med Chem 2004;47:4022–4031.
117. Thomas S, Brightman F, Gill H, *et al*. Simulation modelling of human intestinal absorption using Caco-2 permeability and kinetic solubility data for early drug discovery. J Pharm Sci 2008;97:4557–4574.
118. Peters SA. Identification of intestinal loss of a drug through physiologically based pharmacokinetic simulation of plasma concentration-time profiles. Clin Pharmacokinet 2008;47:245–259.
119. Fenneteau F, Poulin P, Nekka F. Physiologically based predictions of the impact of inhibition of intestinal and hepatic metabolism on human pharmacokinetics of CYP3A substrates. J Pharm Sci 2010;99:486–514.
120. Yokogawa K, Nakashima E, Ishizaki J, *et al*. Relationships in the structure-tissue distribution of basic drugs in the rabbit. Pharm Res 1990;7:691–696.
121. Poulin P, Theil FP. 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 2000;89:16–35.
122. Rodgers T, Rowland M. Mechanistic approaches to volume of distribution predictions: understanding the processes. Pharm Res 2007;24:918–933.

123. Vossen M, Sevestre M, Niederalt C, *et al*. Dynamically simulating the interaction of midazolam and the CYP3A4 inhibitor itraconazole using individual coupled whole-body physiologically-based pharmacokinetic (WB-PBPK) models. Theor Biol Med Model 2007;4:13.
124. Small H, Gardner I, Jones HM, *et al*. Measurement of binding of basic drugs to acidic phospholipids using surface plasmon resonance and incorporation of the data into mechanistic tissue composition equations to predict steady-state volume of distribution. Drug Metab Dispos 2011;39:1789–1793.
125. Hallifax D, Houston JB. Methodological uncertainty in quantitative prediction of human hepatic clearance from *in vitro* experimental systems. Curr Drug Metab 2009;10: 307–321.
126. Brightman FA, Leahy DE, Searle GE, *et al*. Application of a generic physiologically based pharmacokinetic model to the estimation of xenobiotic levels in human plasma. Drug Metab Dispos 2006;34:94–101.
127. Peters SA. Evaluation of a generic physiologically based pharmacokinetic model for lineshape analysis. Clin Pharmacokinet 2008;47:261–275.
128. Kuentz M, Nick S, Parrott N, *et al*. A strategy for preclinical formulation development using GastroPlus as pharmacokinetic simulation tool and a statistical screening design applied to a dog study. Eur J Pharm Sci 2006;27:91–99.
129. Peters SA, Hultin L. Early identification of drug-induced impairment of gastric emptying through physiologically based pharmacokinetic (PBPK) simulation of plasma concentration-time profiles in rat. J Pharmacokinet Pharmacodyn 2008;35:1–30.
130. Rowland Yeo K, Jamei M, Yang J, *et al*. Physiologically based mechanistic modelling to predict complex drug-drug interactions involving simultaneous competitive and time-dependent enzyme inhibition by parent compound and its metabolite in both liver and gut - the effect of diltiazem on the time-course of exposure to triazolam. Eur J Pharm Sci 2010;39:298–309.
131. Yamano K, Yamamoto K, Katashima M, *et al*. Prediction of midazolam-CYP3A inhibitors interaction in the human liver from *in vivo*/*in vitro* absorption, distribution, and metabolism data. Drug Metab Dispos 2001;29:443–452.
132. Chenel M, Bouzom F, Cazade F, *et al*. Drug-drug interaction predictions with PBPK models and optimal multiresponse sampling time designs: application to midazolam and a phase I compound. Part 2: clinical trial results. J Pharmacokinet Pharmacodyn 2008;35:661–681.
133. Chenel M, Bouzom F, Aarons L, *et al*. Drug-drug interaction predictions with PBPK models and optimal multiresponse sampling time designs: application to midazolam and a phase I compound. Part 1: comparison of uniresponse and multiresponse designs using PopDes. J Pharmacokinet Pharmacodyn 2008;35:635–659.
134. Lehr T, Staab A, Trommeshauser D, *et al*. Semi-mechanistic population pharmacokinetic drug-drug interaction modelling of a long half-life substrate and itraconazole. Clin Pharmacokinet 2010;49:53–66.
135. Perdaems N, Blasco H, Vinson C, *et al*. Predictions of metabolic drug-drug interactions using physiologically based modelling: Two cytochrome P450 3A4 substrates coadministered with ketoconazole or verapamil. Clin Pharmacokinet 2010;49:239–258.
136. Chien JY, Lucksiri A, Ernest CS, *et al*. Stochastic prediction of CYP3A-mediated inhibition of midazolam clearance by ketoconazole. Drug Metab Dispos 2006;34:1208–1219.
137. Ito K, Ogihara K, Kanamitsu S, *et al*. Prediction of the *in vivo* interaction between midazolam and macrolides based on *in vitro* studies using human liver microsomes. Drug Metab Dispos 2003;31:945–954.
138. Grime K, Ferguson DD, Riley RJ. The use of HepaRG and human hepatocyte data in predicting CYP induction drug-drug interactions via static equation and dynamic mechanistic modelling approaches. Curr Drug Metab 2010;11:870–885.

139. Yang J, Liao M, Shou M, *et al*. Cytochrome P450 turnover: regulation of synthesis and degradation, methods for determining rates, and implications for the prediction of drug interactions. Curr Drug Metab 2008;9:384–394.
140. Emond C, Birnbaum LS, DeVito MJ. Use of a physiologically based pharmacokinetic model for rats to study the influence of body fat mass and induction of CYP1A2 on the pharmacokinetics of TCDD. Environ Health Perspect 2006;114:1394–1400.
141. Wang X, Santostefano MJ, Evans MV, *et al*. Determination of parameters responsible for pharmacokinetic behavior of TCDD in female Sprague-Dawley rats. Toxicol Appl Pharmacol 1997;147:151–168.
142. Santostefano MJ, Wang X, Richardson VM, *et al*. A pharmacodynamic analysis of TCDD-induced cytochrome P450 gene expression in multiple tissues: dose-and time-dependent effects. Toxicol Appl Pharmacol 1998;151:294–310.
143. Leung HW, Paustenbach DJ, Murray FJ, *et al*. A physiological pharmacokinetic description of the tissue distribution and enzyme-inducing properties of 2,3,7,8-tetrachlorodibenzo-p-dioxin in the rat. Toxicol Appl Pharmacol 1990;103:399–410.
144. Kohn MC, Walker NJ, Kim AH, *et al*. Physiological modeling of a proposed mechanism of enzyme induction by TCDD. Toxicology 2001;162:193–208.
145. Chien JY, Thummel KE, Slattery JT. Pharmacokinetic consequences of induction of CYP2E1 by ligand stabilization. Drug Metab Dispos 1997;25:1165–1175.
146. Magnusson MO, Karlsson MO, Sandstrom R. A mechanism-based integrated pharmacokinetic enzyme model describing the time course and magnitude of phenobarbital-mediated enzyme induction in the rat. Pharm Res 2006;23:521–532.
147. Gordi T, Xie R, Huong NV, *et al*. A semiphysiological pharmacokinetic model for artemisinin in healthy subjects incorporating autoinduction of metabolism and saturable first-pass hepatic extraction. Br J Clin Pharmacol 2005;59:189–198.
148. Hassan M, Svensson US, Ljungman P, *et al*. A mechanism-based pharmacokinetic-enzyme model for cyclophosphamide autoinduction in breast cancer patients. Br J Clin Pharmacol 1999;48:669–677.
149. Mannhold R, Ostermann C. Prediction of LogP with substructure-based methods. In: Mannhold R, editor. Molecular drug properties: measurement and prediction. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2008. pp. 357–379.
150. Tetko IV, Poda GI. Prediction of LogP with property-based methods. In: Mannhold R, editor. Molecular drug properties: measurement and prediction. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2008. pp. 381–406.
151. Abraham MH, Ibrahim A, Zissimos AM, *et al*. Application of hydrogen bonding calculations in property based drug design. Drug Discov Today 2002;7:1056–1063.
152. Lewis DF. On the recognition of mammalian microsomal cytochrome P450 substrates and their characteristics: towards the prediction of human P450 substrate specificity and metabolism. Biochem Pharmacol 2000;60:293–306.
153. Lewis DF. Quantitative structure-activity relationships (QSARs) for substrates of human cytochromes P450 CYP2 family enzymes. Toxicol *In vitro* 2004;18:89–97.
154. Ekins S, De Groot MJ, Jones JP. Pharmacophore and three-dimensional quantitative structure activity relationship methods for modeling cytochrome P450 active sites. Drug Metab Dispos 2001;29:936–944.
155. De Groot MJ. Designing better drugs: predicting cytochrome P450 metabolism. Drug Discov Today 2006;11:601–606.
156. Ekins S, Bravi G, Wikel JH, *et al*. Three-dimensional-quantitative structure activity relationship analysis of cytochrome P-450 3A4 substrates. J Pharmacol Exp Ther 1999;291:424–433.
157. Keseru GM. A virtual high throughput screen for high affinity cytochrome P450cam substrates: implications for in silico prediction of drug metabolism. J Comput Aided Mol Des 2001;15:649–657.

158. Agrafiotis DK, Myslik JC, Salemme FR. Advances in diversity profiling and combinatorial series design. Mol Divers 1998;4:1–22.
159. Bender A, Mussa HY, Glen RC, *et al*. Similarity searching of chemical databases using atom environment descriptors (MOLPRINT 2D): evaluation of performance. J Chem Inf Comput Sci 2004;44:1708–1718.
160. Balakin KV, Tkachenko SE, Lang SA, *et al*. Property-based design of GPCR-targeted library. J Chem Inf Comput Sci 2002;42:1332–1342.
161. Watkins RE, Wisely GB, Moore LB, *et al*. The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity. Science 2001;292:2329–2333.
162. Ivanciuc O. Electrotopological state indices. In: Mannhold R, editor. Molecular drug properties: measurement and prediction. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2008. pp. 85–109.
163. Di L, Kerns EH. Profiling drug-like properties in discovery research. Curr Opin Chem Biol 2003;7:402–408.
164. Cartmell J, Enoch S, Krstajic D, *et al*. Automated QSPR through competitive workflow. J Comput Aided Mol Des 2005;19:821–833.