DDT Vol. 9, No. 6 March 2004

Predicting undesirable drug interactions with promiscuous proteins *in silico*

Sean Ekins

Although computational tools have been used to predict toxic responses resulting from molecules binding either as substrates or inhibitors to proteins, there are complications to be resolved. Some proteins appear promiscuous in their ability to bind a diverse array of hydrophobic molecules. This promiscuity arises from the binding site simultaneously accommodating more than one molecule, multiple separate binding sites, protein flexibility, or a combination of all these properties. With the availability of more crystal structures for these non-target proteins, we should be able to predict binding *in silico* with a greater accuracy, thus avoiding or managing toxic side effects, therefore ultimately improving the success of drug discovery.

Sean Ekins
Concurrent Pharmaceuticals
502 West Office Center Drive
Fort Washington
PA 19034, USA
Present address: GeneGo
500 Renaissance Drive
Suite 106
St. Joseph
MI 49085, USA
e-mail: ekinssean@yahoo.com

▼ We live in an age where genomic technologies promise to revolutionize the way in which drugs are discovered and developed. Many treatments for chronic diseases will be needed, perhaps more so with an ageing demographic. The safe delivery of a drug to its target in people that are probably taking many medications concomitantly, thus increasing the potential for drug–drug interactions, is challenging. In reality, we are all daily consumers of a variety of medicinal compounds, such as dietary supplements and over the counter medications, for example, St. John's Wort [1]. Understanding the potential for endobiotic and xenobiotic interactions of pharmaceuticals with specific enzymes, transporters, receptors and ion channels is moving into a more predictive realm through the use of a combination of *in vitro* and *in silico* methods. This predictive approach is necessary to ensure that molecules in pharmaceutical development have the highest probability of reaching the clinic and the market. One of the key problems associated with drug discovery is to ensure that the drug molecules not only have adequate physical properties to sustain therapeutic levels to facilitate their oral dosing, but that they also (preferably) avoid interactions with non-target proteins and other probable concomitantly dosed drugs. So much emphasis has been placed on predicting the solubility, absorption and blood–brain barrier penetration of a drug that the roles of metabolism, possible drug–drug, drug–non-target interactions and other key toxicity endpoints have received far less attention from the broader computational chemistry community.

Computational algorithms can be exploited to make predictions of the interactions between vast numbers of virtual molecules and non-target proteins (with or without their crystal structures). Approaches using crystal structures have identified molecules with a binding affinity for proteins that on complex formation could result in toxic side effects [2,3]. However, the majority of computational studies have been completed in the absence of the crystal structure of the protein of interest. Common computational tools such as Quantitative Structure Activity Relationships (QSAR), and pharmacophore and homology modeling have all been used to generate predictions for molecules binding either as substrates or inhibitors to an array of proteins that might be important for human toxicity.

During early drug discovery, a computational approach that produces a fingerprint of the predicted interactions of a drug with different proteins, as well as other physical properties, can be used alongside efficacy data in a multi-criteria decision making process [4]. Some of the key non-target proteins for drug discovery include the human ether-a-go-go related gene

1359-6446/04/$ – see front matter PII: S1359-6446(03)03008-3

(hERG) potassium channel, the pregnane X-receptor (PXR), cytochrome P450s (CYPs), P-glycoprotein (P-gp) and Phase II metabolizing enzymes. All of these proteins appear promiscuous in their binding interactions with small hydrophobic ligands, which proves problematic when predicting the potential for undesirable drug–protein interactions. Although the hydrophobicity of a drug is a necessity for desirable potency towards therapeutic targets, it could also be partly responsible for the difficulty in developing drugs with adequate bioavailability. The importance of hydrophobic character for ligand–protein binding has been known for many years, yet the recognition that it is important for drug metabolism rests with the work of Hansch [5], among others. The inherent promiscuity of the diverse non-target proteins makes it difficult to predict ligand-binding conformation, let alone 'the relative effects on substituents on the metabolism of drugs', as Hansch desired [5]. However, there have been considerable modeling efforts for these promiscuous proteins, which are summarized in this review.

Human ether-a-go-go related gene

The tetrameric hERG protein is a potassium channel that represents a key part of the cardiac action potential and has received considerable attention in recent years. hERG contributes to phase three repolarization by opposing the depolarizing Ca^{2+} influx during the plateau phase [6]. Several drugs, including cisapride, terfenadine, astemizole, sertindole and grepafloxacin have been withdrawn from the market in recent years, partly as a result of cardiovascular toxicity associated with undesirable blockade of this channel. In a small percentage of individuals, subsequent prolongation of a part of the cardiac action potential termed the 'QT interval' results in potentially life-threatening ventricular arrhythmia [6]. Therefore, in drug discovery, it is important to understand the structural requirements of molecules that bind to this potassium channel. To date, the *in vitro* assessment of the drug mediated interaction with these channels uses cell systems that express the hERG channel and uses methods that include patch clamping, radioligand binding, fluorescent probes and rubidium flux studies. These methods produce data of varying quality and reliability that can be computationally modeled.

An initial homology model for hERG was generated from the template bacterial KcsA channel, which together with site directed mutagenesis work has provided an insight into the key amino acid residues involved in drug binding to the hERG channel [7]. Other computational approaches, for example, 3D-QSAR, which is a ligand-based approach, have been used to evaluate potential hERG inhibitors with some degree of predictive success [8–11]. A preliminary pharmacophore was derived from a literature set of 15 molecules that are known to inhibit hERG in patch clamping studies with HEK 293 cells and contained four hydrophobes and one positive ionizable feature [8]. In addition, this model was assessed using a test set of 22 molecules that included known inhibitors of hERG, such as antipsychotics and their metabolites. In an attempt to predict classes of hERG inhibitors [using data from Chinese hamster ovary (CHO) cells expressing the channel], a second computational analysis found a striking difference in calculated logP (ClogP, a measure of hydrophobicity). ClogP was higher in potent hERG inhibitors (Figure 1) compared with non-potent inhibitors [10]. A further pharmacophore using a comparative molecular field analysis (CoMFA) model for 31 inhibitors found that three aromatic (hydrophobic) functions and a central nitrogen were key for binding to hERG [11]. Recently, a comparative molecular similarity analysis (CoMSiA) using 28 molecules with patch clamping data was dominated by sertindole analogs tested in CHO cells expressing hERG [9]. This model also suggested that hydrophobic and positive ionizable features were important for binding. Furthermore, the model was placed in the context of the protein using a hERG homology model based on the MthK channel [9] to illustrate key interactions with Phe656 and Tyr652 residues [7], the same residues previously identified by Mitcheson *et al.* [7]. Attempts at building a QSAR model with a recursive partitioning algorithm and based on patch clamping data for hERG with 66 molecules have also been described, along with a diverse test set of 25 molecules [12]. Furthermore, an updated model built using 99 literature molecules [12] has been recently used to rank the 23 sertindole analogs generated by Pearlstein *et al.* (Spearman's rho = 0.74, $P < 0.0001$, $r^2 = 0.53$) [9]. This type of model is perhaps less visually interpretable than a pharmacophore because it is based on 2D descriptors, for example, atom path lengths, but it enables a high-throughput for scoring virtual libraries of molecules. More recently, another descriptor-based approach using a partial least squares algorithm and a 55 molecule training set was found to be predictive for a 13 compound test set ($r^2 = 0.81$) [13].

A further approach that determines the probability of interaction of a drug with hERG uses the inhibition of [^{3}H]-dofetilide binding, which is a higher-throughput method compared with the 'gold standard' patch clamping. Researchers at Merck (http://www.merck.com) have published papers containing binding data for the 5-HT_{2A} receptor antagonists [14,15], 3-aminopyrrolidinone farnesyltransferase inhibitors [16] and phospodiesterase-4 inhibitors, in which hERG binding was reduced to some extent by structural modification. A simple alignment of pharmacophores generated with the Merck data suggests common areas of

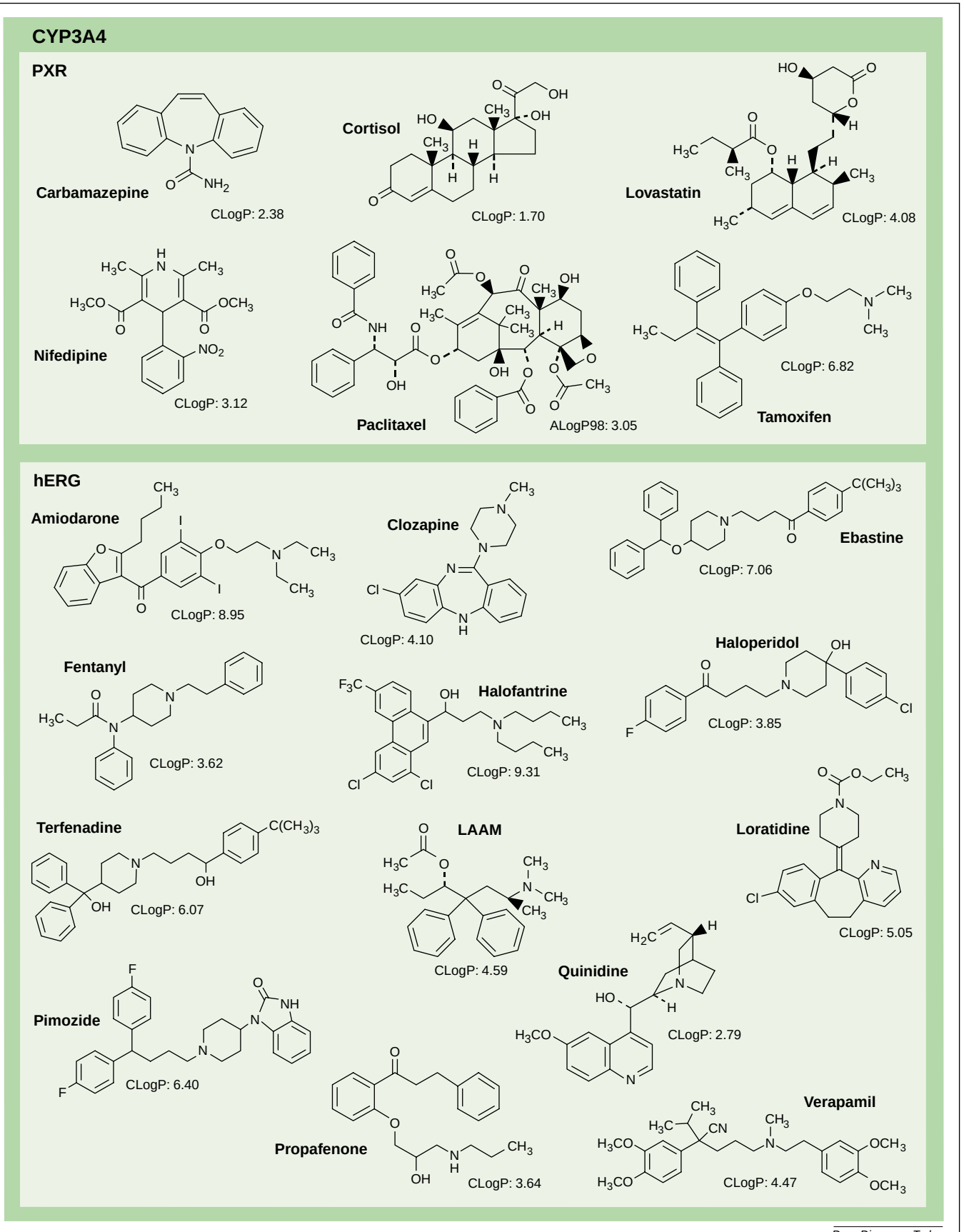
CYP3A4
PXR
Carbamazepine
CLogP: 2.38
Cortisol
CLogP: 1.70
Lovastatin
CLogP: 4.08
Nifedipine
CLogP: 3.12
Paclitaxel
ALogP98: 3.05
Tamoxifen
CLogP: 6.82
hERG
Amiodarone
CLogP: 8.95
Clozapine
CLogP: 4.10
Ebastine
CLogP: 7.06
Fentanyl
CLogP: 3.62
Halofantrine
CLogP: 9.31
Haloperidol
CLogP: 3.85
Terfenadine
CLogP: 6.07
LAAM
CLogP: 4.59
Loratidine
CLogP: 5.05
Pimozide
CLogP: 6.40
Quinidine
CLogP: 2.79
Propafenone
CLogP: 3.64
Verapamil
CLogP: 4.47
Drug Discovery Today

Figure 1. Summary of selected ligands with overlapping affinity for CYP3A4 and PXR or hERG. ClogP values were calculated in ChemDraw (CambridgeSoft; http://www.camsoft.com) for illustration of the hydrophobicity. The AlogP98 value for paclitaxel was calculated using Cerius² (Accelrys; http://www.accelrys.com). Abbreviations: AlogP98, atom-based logP; ClogP, calculated logP; CYP, cytochrome P450; hERG; human ether-a-go-go protein; PXR, pregnane X-receptor.

positive ionizable features and hydrophobicity (from aromatic rings) (Figure 2).

The inhibition of hERG is not limited to small therapeutic molecules. Large naturally occurring molecules, for example, the scorpion toxins BeKm-1 (4092 Da) [17] and CnErg1 (Ergotoxin, 4757 Da) act as potent and selective inhibitors of hERG (IC_{50} 3.3 nM and K_d 8.5 nM, respectively) [18]. The sea anemone toxin APETx1 (4552 Da) is also a specific inhibitor of hERG (IC_{50} 34 nM) [19]. These hERG binding values are in a similar range to some of the small molecule inhibitors reported [8–11]. The surface of the scorpion toxins has an area of positive charge surrounded by multiple hydrophobic areas [20], which is generally similar to the pharmacophores defined for the small molecule inhibitors described. It is suggested that these large molecules bind in the pore or turret region of the protein channel [21–23], rather than in the channel itself as is the case with small molecules (Figure 3). However, for binding, some of these protein ligands require electrostatic interaction, others rely on hydrophobic interaction, whereas BeKm-1 appears to bind to the channel during the closed state [24]. Because large and small hydrophobic compounds bind hERG at different regions of the channel, there are implications for the testing of protein therapeutics for potential interactions.

The similarity in potency between large and small molecules for binding to hERG also complicates the prediction of the binding site of a particular molecule. The expression of a key pharmacophore on the surface of the large ligand molecule appears to be important in determining whether it binds the channel in the open or closed state. Therefore, the hERG channel displays a remarkable ability to bind molecules from diverse structural classes and size, which suggests that computational models generated with homologous molecular series and diverse structures are necessary. The screening of large diverse virtual libraries of molecules could benefit from global hERG models that can classify the potency of compounds. It might be necessary to generate similar computational models for protein ligands. In the case of small protein ligands, the modeling of hERG-small protein interactions would be desirable. The computational modeling of this channel is in its infancy and predictive capability could be limited to classification, but we will see more widely applicable models in the future using advanced algorithms and descriptors.

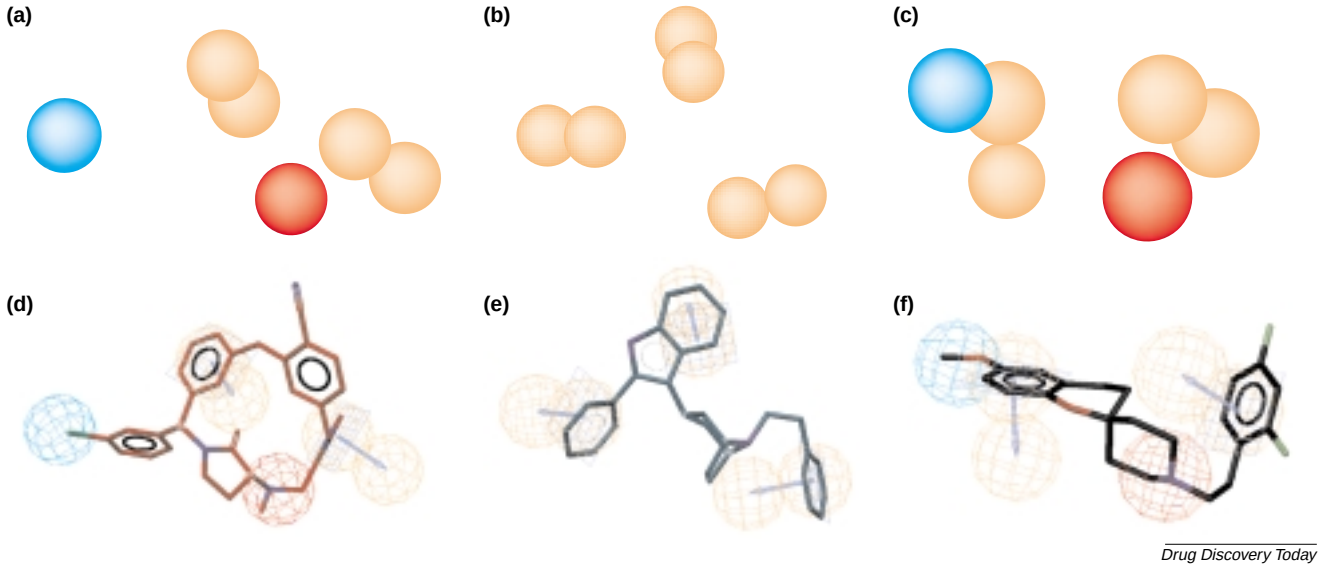

Figure 2. Three hERG pharmacophores generated with Catalyst® (version 4.7) based on *in vitro* hERG [³H]-dofetilide binding data published by Merck [cyan spheres = hydrophobe; orange = aromatic ring (hydrophobic); red = positive ionizable]: **(a, d)** hERG pharmacophore for farnesyltransferase inhibitors from Bell *et al*. [16]. Observed versus predicted r = 0.87 for the training set. Molecule shown had an 82 nM inflection point; **(b, e)** hERG pharmacophore for 5-HT_{2A} inhibitors from Rowley *et al.* [14]. Observed versus predicted r = 0.87 for the training set. Molecule shown had an IC_{50} of 19 nM; **(c, f)** hERG pharmacophore for 5-HT_{2A} inhibitors from Fletcher *et al.* [15]. Observed versus predicted r = 0.93. Molecule shown had an IC_{50} of 20 nM. Combination of all three hERG pharmacophores suggests a similar arrangement of common features. Abbreviation: hERG; human ether-a-go-go protein.

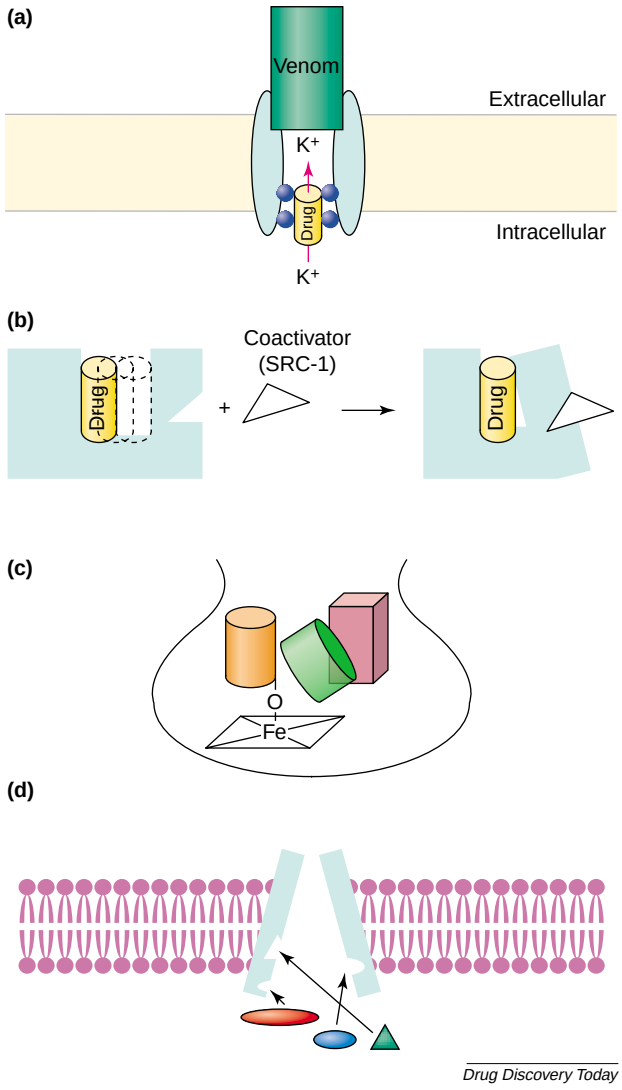

Figure 3. Diagrams to illustrate the potential multiple ligand–protein interactions for **(a)** hERG, **(b)** PXR, **(c)** CYP3A4 and **(d)** P-gp, based on data from a combination of different computational and empirical sources. hERG could be inhibited by venoms from the extracellular surface or small molecules that form hydrophobic interactions (blue circles) inside the channel. Ligands could bind in multiple positions within PXR (dotted lines), but when the coactivator SRC-1 binds PXR, the location for ligand binding is restricted. The CYP3A4 binding site appears large enough to simultaneously accommodate multiple ligands simultaneously. P-gp also appears to bind either large hydrophobic molecules or multiple molecules at different binding sites, as illustrated by the different colored shapes. Abbreviations: CYP; cytochrome P450; hERG; human ether-a-go-go protein; P-gp, P+glycoprotein; PXR, pregnane X-receptor.

Cytochrome P450

CYPs are hemoproteins that catalyze the hydroxylation, epoxidation and oxidation reactions that are found in every phylla and most mammalian tissues [25]. This large family of enzymes is important owing to its role in the endobiotic and xenobiotic metabolism of many commonly prescribed drugs across therapeutic areas. Drug-mediated interactions with these enzymes *in vivo* pose a potential clinically relevant event. The late discovery of a clinically significant drug–drug interaction can be financially costly. Therefore, it is important to screen for potential interactions [26] to enable rejection of suspect molecular series early or suggest the most appropriate *in vivo* studies later on. The CYPs could be involved in many drug–drug interactions because of their affinity for hydrophobic molecules of varying sizes. Varied computational approaches have been used to attempt to predict drug metabolism and interactions. Many of the key early studies in this area by Hansch [5], Lewis [27] and co-workers used QSAR to understand substrate–structure activity relationships. In more recent years, homology models of the CYPs based either on the crystal structures for bacterial CYPs or rabbit CYP2C5 enzymes have been widely described [28]. Early CoMFA models were used with hepatic microsomal data for CYP1A2 [29], CYP2A5 [30] and CYP2C9 [31] and used relevant conformers of inhibitors to identify key functional groups, the geometry of selected structural features and regions of electrostatic and steric interactions. More recent QSAR models have been built after analysis of literature data, including recombinant-derived kinetic values for the major CYPs [32,33]. Recent reviews have also compared the many pharmacophores for CYPs [32,33] and there have been several groups that combined pharmacophores with homology models based on either bacterial CYPs [31,34–36] or CYP2C5/3 [37–40] to provide further structural insight.

In terms of human drug metabolism, the CYP3A family of enzymes is the most important [41] because they metabolize many different classes of molecules of diverse sizes (Figure 1). Assessment of the inhibitory potency of CYP3A4 and CYP2D6 is often included as a first tier screen in early drug discovery. Computational models have been described that use the recursive partitioning algorithm and path length descriptors for this type of high-throughput CYP interaction data. The models consisted of over 1700 diverse molecules and were used to rank the inhibition for test sets of nearly 100 molecules [42]. In addition, computational modeling of CYP3A4 ligands has identified hydrogen bonding and hydrophobicity as important determinants for substrate binding [43,44]. Computational inhibitor models of CYP3A4 differ depending on the substrate probe used *in vitro*, although they do maintain the trend of highlighting multiple hydrophobic and hydrogen bonding functional groups [45–47]. The interaction between CYPs and their substrates *in vitro* is complex and Michaelis–Menten kinetics are the exception rather than the rule, as exemplified by CYP3A4 [48,49]. Furthermore, inhibition

of the CYP3A4 catalysis is substrate dependent [50–52], which adds to the overall complexity of interpreting *in vitro* screening data. Data from various studies are consistent with the active site of CYP3A4 simultaneously accommodating multiple substrate and inhibitor molecules [48,53,54] (Figure 3). This makes accurate computational and *in vitro* predictions difficult, and is far removed from the traditional lock and key hypothesis for ligand–protein interactions.

Electronic models for CYP-mediated metabolism have been produced [55,56] that have combined aliphatic and aromatic oxidation reactions to generate predictions for regioselectivities of substrates. More recently, this approach has been used to identify the major sites of human CYP3A4 metabolism [57]. In an attempt to incorporate all the hepatic enzymes involved in drug metabolism, metabolic stability data generated from human liver homogenates was used [58]. One study measured % turnover data at a single drug concentration to generate a *k*-nearest neighbors model that incorporated 500 molecules. This approach was able to differentiate between stable and unstable molecules to an accuracy between 77–84% [58]. Similarly, a recursive partitioning model for human liver microsomal metabolic stability [12] and pharmacophores based on microsomal and hepatocyte data [59] have been used to predict human clearance. In these cases, there was considerable complexity in modeling multiple enzyme systems using a single approach, therefore, methods that capture a more complete biological system and models that are based on kinetic half-life measurements rather than % turnover data are worth pursuing. One group used a large database that included most of the published CYP-mediated metabolic pathways, ligands and metabolites to produce a Kohonen self-organizing map neural network [60]. This approach was able to differentiate between the physicochemical properties of substrates and metabolites, and on average 60% of the CYP substrates and 65% of metabolites were classified correctly. This study demonstrated the difficulty of using a combination of data for different CYPs that each probably have unique optimal molecular descriptors for binding. Based on the published data, computational modeling of CYPs has a long way to go before being universally accepted. Studies to date imply that no single modeling approach or descriptor type will be optimal for determining the complexities of CYP-mediated metabolism. Deconvoluting metabolism into the separate enzymes responsible might be unavoidable, unless an alternative algorithm that can handle such a complex non-linear system is found. The recent crystallization of human CYP2C9 [61] by researchers at Astex (http://www.astex-technology.com/index.html) represents an opportunity to validate some of the available computational models that have been generated over the years. Additionally, this structure could aid in understanding more clearly how multiple molecules can interact in the active site. From this knowledge, we should be able to build improved computational models for the other major human CYP enzymes that still await crystallization.

P-glycoprotein

P-gp is a large membrane-bound protein that is expressed at the interface of many important organs with their environment, and acts as a barrier to limit the exposure of the organs to xenobiotics [62]. In addition, P-gp is present in the canalicular domain of hepatocytes, brush border of proximal tubule cells and capillary endothelial cells in the central nervous system (CNS) [62] where it might reduce oral drug absorption, enhance renal and biliary excretion or limit CNS entry of many substrate drugs. Computational models for P-gp have been reviewed previously [63,64], but in the past few years several new pharmacophore-based models have appeared for substrates [65] and inhibitors [66–69]. The recent ensemble pharmacophore model for P-gp substrates was based on 144 molecules, and correctly classified 80% of the training set, but only 63% of the test set [65]. The pharmacophores used contained hydrophobic and aromatic ring features and functional groups that could act as hydrogen-bond donors and/or acceptors. Research indicated that the hydrogen-bond acceptor was the most important feature of the pharmacophore. Other Catalyst™ (Accelrys; http://www.accelrys.com) pharmacophore-based approaches generated multiple models for inhibitors of transport, binding or accumulation in different cell systems. These models indicated large distances between multiple hydrophobic and hydrogen-bond acceptor features [68,69]. Further pharmacophores using genetic algorithm similarity program [GASP™ (Tripos; http://www.tripos.com)] -based alignments of vinblastine and rhodamine123 have been carried out to elucidate the binding interactions between verapamil and P-gp [66]. Again, multiple hydrophobic and hydrogen bonding interactions were described [66]. Other groups have superimposed a small number of P-gp ligands with SYBYL™ (Tripos) and MOLCAD™ (Tripos) to generate pharmacophores [67] that were validated by aligning two additional compounds. Presently, the computational models produced for P-gp have been valuable for identifying the molecular features required for binding, but quantitative predictions have involved relatively small numbers of molecules so it is difficult to assess their applicability. As larger datasets are generated with high-throughput screening (HTS) data for different substrate probes, such as those for CYP interactions, this

situation will change and more complex computational models will be required.

Pregnane X-receptor

PXR is a member of the nuclear receptor family of ligand-activated transcriptional factors that regulates induction of several drug-metabolizing enzymes (e.g. CYP3A, CYP2B6, CYP2C8/9) and transporters (P-gp, multi-drug resistance protein-1 and multi-drug resistance protein-2), which are responsible for eliminating conjugates of toxic molecules and biliary efflux of endogenous molecules, respectively [70]. Analogous to other members of the nuclear receptor family, PXR has DNA and ligand-binding domains that enable the binding of PXR as a heterodimer with the retinoid X receptor to response elements on target genes. There is considerable structural diversity in the molecules binding to human PXR *in vitro*, and some of these ligands are also substrates for CYP3A4 (Figure 3). The range of PXR ligands includes, some bile acids, statins, herbal components, a selection of HIV protease inhibitors, calcium channel modulators, numerous steroids, plasticizers and monomers, organochlorine pesticides, a peroxisome proliferator-activated receptor-γ antagonist, as well as other xenobiotics and endobiotics [70,71]. A more complete understanding of the clinical implications of the role of PXR could result in a more accurate prediction of the potential drug–drug interactions that result from the co-administration of molecules that are known to be metabolized or transported by any of the genes upregulated by this receptor.

Initially, X-ray crystallography of the ligand-binding domain of PXR [72] suggested that it was a large, flexible hydrophobic site that incorporated a few polar residues and three distinct binding sites for the co-crystallized ligand SR12813. A crystal structure for PXR with hyperforin bound indicated that the volume of the binding site can change and that the ligand can form other interactions with the protein [73]. Nine hydrophobic and three polar interactions with hyperforin were observed. However, hyperforin did not appear to bind in the multiple conformations associated with SR12813, which contacted fewer hydrophobic residues overall [73]. The most recent crystallography studies by Watkins *et al.* [70] have shown that the binding of the steroid receptor coactivator-1 to the outer surface of PXR limits the ligand SR12813 to a single conformation, which is probably necessary for transcriptional activation. This work is compelling because it shows that the ligand becomes 'trapped' in the binding site, and as a consequence eleven hydrophobic interactions are formed with the binding site. It remains to be ascertained whether or not all ligands, or even larger molecules, will behave similarly and become trapped in the same region of the ligand-binding domain. The investigation and understanding of the binding characteristics of molecules to a promiscuous protein such as PXR could require the determination of the binding location for many ligands across numerous structurally similar and diverse series of compounds. This would facilitate more accurate computational estimations of binding conformation and affinity by presumably highlighting subtle and drastic changes in ligand structure and binding interactions.

Once again, computational pharmacophore approaches were used in an early attempt to understand the key features of PXR ligands. A model based on a small set of reasonably diverse ligands defined key features responsible for interaction with the PXR binding site [74]. The pharmacophore implicates at least four hydrophobic features and one hydrogen bonding feature that could be avoided in future drug candidate molecules. Furthermore, the pharmacophore was validated by positioning it in the human PXR ligand-binding domain X-ray structure, and illustrated good predictions for 28 other PXR ligands in a test set [74]. This pharmacophore alone might have difficulty predicting all possible PXR ligands that can bind to such a promiscuous receptor. Assessment of considerably more molecules will enable the generation of predictive algorithms for classifying molecules as binders with a higher degree of accuracy. As a result of the complexity of PXR binding, it will be difficult to determine whether or not a small molecule will be a potent or less potent ligand *in silico* using only rigid docking in the available crystal structures. Docking systems that enable flexible docking of the ligand and protein have been used for virtual HTS for therapeutic targets [75] and could be applied to PXR, together with more dynamic pharmacophore models [76].

Models for Phase II enzymes

Phase II metabolism, commonly termed conjugation reactions, has been less widely pursued by research groups compared with research on the Phase I enzymes (e.g. CYPs) [77,78]. For example, the formation of sulfate conjugates of endogenous and exogenous molecules occurs via the cytosolic enzyme sulfotransferase superfamily, of which there could be ten or more members in humans. SULT1A3 is a sulfotransferase that is expressed in the intestine and is important for the sulfation of amine neurotransmitters, drugs and other xenobiotics [78]. A QSAR model generated for this enzyme identified that hydrogen bonding and hydrophobicity are important for substrate binding to the enzyme, which is consistent with information obtained from the crystal structure of the enzyme [79]. The recently described crystal structure for a second member of the sulfotransferases, SULT1A1, revealed that two molecules of

p-nitrophenol were bound in the substrate site [80]. This is analogous to the CYP2C9 structure that has a second hypothesized binding site [61] and might suggest that the simultaneous binding of multiple ligands in the enzyme is not limited to the CYPs or P-gp [61].

Glucuronidation of small lipophilic molecules is an important metabolic process for the clearance of drugs, endobiotics and xenobiotics in all mammalian species [77]. The uridine diphospho-glucuronosyltransferases (UGT) family can be separated into at least two distinct sub-families by sequence similarity. The UGT1 and UGT2 family of isoforms are known to be encoded by individual genes [81]. A recent study described the glucuronidation of simple 4-substituted phenols by the recombinant human UGT1A6 and UGT1A9 isoenzymes. Subsequently, this data was used to generate genetic function approximation computational models for these isoenzymes with a range of molecular surface and atomic descriptors [82]. Furthermore, glucuronidation of 19 indolocarbazoles in human liver microsomes indicated that molecules with a diameter greater than 14.5 Å measured perpendicular to the site of the phenolic O-position are not glucuronidated [83]. Further studies with the recombinant enzyme suggested that UGT1A9 is responsible for catalyzing this glucuronidation [83]. Some encouraging studies have generated 2D-QSAR and 3D-QSAR pharmacophore models for human UGT1A1 [84] and UGT1A4 substrates [85]. These models contained similar orientations of features, at least two hydrophobes and one glucuronidation feature [84]. Because metabolism and drug–drug interaction screening to date has focused on CYPs, the substrate and inhibitor selectivity for Phase II enzymes will need to be studied in considerably more detail to generate improved metabolic clearance predictions. At present, there are few computational models for Phase II enzymes, and most have been validated with small test sets. The generation of larger quantities of *in vitro* data and crystal structures for Phase II enzymes will aid further computational model building, which is ultimately desirable.

Conclusions

All of the proteins described in this review have the ability to bind to structurally diverse hydrophobic xenobiotics, which makes the development of potent ligands for disease targets more difficult. Protein promiscuity is facilitated by multiple binding sites and/or a flexible dynamic binding site, which in most cases seems essential for their biological function (Figure 3). In the case of PXR, the receptor increases the levels of transporters and metabolizing enzymes to remove potentially toxic molecules from the body. CYPs and Phase II enzymes need to be able to metabolize and rapidly clear the large number of diverse molecules we are exposed to in our diet and environment. Similarly, P-gp recognizes and exports large and small hydrophobic molecules. By contrast, the promiscuity of hERG is a double-edged sword because it can be blocked with diverse and relatively rigid hydrophobic drugs or toxins. This is advantageous for the development of new classes of antiarrhythmics, but represents a severe hindrance for developing non-cardiac drugs that need to avoid hERG.

Clearly, the ultimate avoidance of undesirable interaction requires validated computational models for proteins besides the therapeutic target, which can then be used to predict large virtual libraries before synthesis. To date, few computational approaches have been attempted and there is still a great deal of fundamental research to be carried out for each of the proteins described to improve our modeling predictions. Hence, we are just at the beginning of the model building process for these proteins and are still determining the trade-offs required for selecting those molecules predicted to bind to targets versus non-targets *in silico*. To empower earlier prospective screening *in silico*, the understanding of protein affinity or promiscuity towards xenobiotic ligands from protein sequence alone would be highly desirable. However, this is complicated by the flexibility of the proteins. Ultimately, we move closer to computers improving the efficiency of the pharmaceutical industry by avoiding molecules that form undesirable interactions with promiscuous proteins. However, this advance is not without an associated loss of interpretability of the models as a result of the requirement for increasing complexity in the algorithms used and considerable investment in more empirical data.

Acknowledgements

Maggie A. Z. Hupcey and Peter W. Swann are gratefully acknowledged for insightful discussions.

References

1 Zhou, S. *et al.* (2003) Interactions of herbs with cytochrome P450. *Drug Metab. Rev.* 35, 35–98
2 Rockey, W.M. and Elcock, A.H. (2002) Progress toward virtual screening for drug side effects. *Proteins* 48, 664–671
3 Chen, Y.Z. and Ung, C.Y. (2001) Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand-protein inverse docking approach. *J. Mol. Graph. Model.* 20, 199–218
4 Shimada, J. *et al.* (2002) Integrating computer-based *de novo* drug design and multidimensional filtering for desirable drugs. *Targets* 1, 196–205
5 Hansch, C. (1972) Quantitative relationships between lipophilic character and drug metabolism. *Drug Metab. Rev.* 1, 1–14
6 Crumb, W. and Cavero, I. (1999) QT interval prolongation by non-cardiovascular drugs: issues and solutions for novel drug development. *Pharm. Sci. Technol. Today* 2, 270–280
7 Mitcheson, J.S. *et al.* (2000) A structural basis for the drug-induced log QT syndrome. *Proc. Natl. Acad. Sci. U. S. A.* 97, 12329–12333
8 Ekins, S. *et al.* (2002) Three dimensional quantitative structure activity

relationship for the inhibition of the hERG (human ether-a-gogo related gene) potassium channel. *J. Pharmacol. Exp. Ther.* 301, 427–434
9 Pearlstein, R.A. *et al.* (2003) Characterization of HERG Potassium channel inhibition using CoMSiA 3D QSAR and homology modeling approaches. *Bioorg. Med. Chem. Lett.* 13, 1829–1835
10 Roche, O. *et al.* (2002) A virtual screening method for the prediction of the hERG potassium channel liability of compound libraries. *Chembiochem* 3, 455–459
11 Cavalli, A. *et al.* (2002) Toward a pharmacophore for drugs inducing the long QT syndrome: insights from a CoMFA study of HERG K(+) channel blockers. *J. Med. Chem.* 45, 3844–3853
12 Ekins, S. (2003) In silico approaches to predicting metabolism, toxicology and beyond. *Biochem. Soc. Trans.* 31, 611–614
13 Kesuru, G.M. (2003) Prediction of hERG potassium channel affinity by traditional and hologram QSAR methods. *Bioorg. Med. Chem. Lett.* 13, 2773–2775
14 Rowley, M. *et al.* (2001) 3-(4-Fluoropiperidin-3-yl)-2-phenylindoles as high affinity, selective, and orally bioavailable h5-HT2A receptor antagonists. *J. Med. Chem.* 44, 1603–1614
15 Fletcher, S.R. *et al.* (2002) 4-(phenylsulfonyl)piperidines: novel, selective, and bioavailable 5-HT2A receptor antagonists. *J. Med. Chem.* 45, 492–503
16 Bell, I.M. *et al.* (2002) 3-Aminopyrrolidinone farnesyltransferase inhibitors: design of macrocyclic compounds with improved pharmacokinetics and excellent cell potency. *J. Med. Chem.* 45, 2388–2409
17 Korolkova, Y.V. *et al.* (2001) An ERG channel inhibitor from the scorpion Buthus eupeus. *J. Biol. Chem.* 276, 9868–9876
18 Torres, A.M. *et al.* (2003) Solution structure of CnErg1 (Ergtoxin), a HERG specific scorpion toxin. *FEBS Lett.* 539, 138–142
19 Diochot, S. *et al.* (2003) APETx1, a new toxin from the sea anemone Anthopleura elegantissima, blocks voltage-gated human ether-a-go-go-related gene potassium channels. *Mol. Pharmacol.* 64, 59–69
20 Korolkova, Y.V. *et al.* (2002) New binding site on common molecular scaffold provides HERG channel specificity of scorpion toxin BeKm-1. *J. Biol. Chem.* 277, 43104–43109
21 Rodriguez de la Vega, R.C. *et al.* (2003) Novel interactions between K+ channels and scorpion toxins. *Trends Pharmacol. Sci.* 24, 222–227
22 Xu, C-Q. *et al.* (2003) Turret and pore block of the K+ channels: what is the difference? *Trends Pharmacol. Sci.* 24, 446–448
23 Possani, L.D. and Rodrigues de la Vega, R.C. (2003) Response to Xu *et al.*: hypothesis-driven science paves the way for new discoveries. *Trends Pharmacol. Sci.* 24, 448–449
24 Milnes, J.T. *et al.* (2003) Preferential closed channel blockade of HERG potassium currents by chemically synthesized BeKm-1 scorpion toxin. *FEBS Lett.* 547, 20–26
25 Nelson, D.R. *et al.* (1996) P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* 6, 1–42
26 Wrighton, S.A. and Silber, P.M. (1996) Screening studies for metabolism and toxicology. In *Scientific Strategies for Accelerated Drug Discovery*, pp. 4.1–4.18, Technomic Publishing AG
27 Lewis, D.F.V. (2000) Structural characteristics of human P450s involved in drug metabolism: QSARs and lipophilicity profiles. *Toxicology* 144, 197–203
28 Kirton, S.B. *et al.* (2002) Comparative modelling of the cytochromes P450. *Adv. Drug Deliv. Rev.* 54, 385–406
29 Fuhr, U. *et al.* (1993) Quinolone antibacterial agents: relationship between structure and *in vitro* inhibition of human cytochrome P450 isoform CYP1A2. *Mol. Pharmacol.* 43, 191–199
30 Poso, A. *et al.* (1995) Comparative molecular field analysis of compounds with CYP2A5 binding affinity. *Quant. Struct.-Act. Relatsh.* 14, 507–511
31 Jones, J.P. *et al.* (1996) Three-dimensional quantitative structure–activity relationship for inhibitors of cytochrome P4502C9. *Drug Metab. Dispos.* 24, 1–6
32 de Groot, M.J. and Ekins, S. (2002) Pharmacophore modeling of cytochromes P450. *Adv. Drug Deliv. Rev.* 54, 367–383
33 Ekins, S. *et al.* (2001) Pharmacophore and three dimensional quantitative structure–activity relationship methods for modeling cytochrome P450 active sites. *Drug Metab. Dispos.* 29, 936–944
34 de Groot, M.J. *et al.* (1999) Novel approach to predicting P450-mediated drug metabolism: Development of a combined protein and pharmacophore model for CYP2D6. *J. Med. Chem.* 42, 1515–1524
35 de Groot, M.J. *et al.* (1999) A novel approach to predicting P450 mediated drug metabolism. CYP2D6 catalyzed N-dealkylation reactions and qualitative metabolite predictions using a combined protein and pharmacophore model for CYP2D6. *J. Med. Chem.* 42, 4062–4070
36 Rao, S. *et al.* (2000) A refined 3-dimensional QSAR of P4502C9. *J. Med. Chem.* 43, 2789–2796
37 Snyder, R. *et al.* (2002) Three dimensional quantitative structure–activity relationship for CYP2D6 substrates. *Quant. Struct.-Act Relatsh.* 21, 357–368
38 Afzelius, L. *et al.* (2001) Competitive CYP2C9 inhibitors: enzyme inhibition studies, protein homology modeling, and three dimensional quantitative structure activity relationship analysis. *Mol. Pharmacol.* 59, 909–919
39 Wang, Q. and Halpert, J.R. (2002) Combined three-dimensional quantitative structure–activity relationship analysis of cytochrome P450 2B6 substrates and protein homology modeling. *Drug Metab. Dispos.* 30, 86–95
40 de Groot, M.J. *et al.* (2002) Development of a combined protein and pharmacophore model for cytochrome P450 2C9. *J. Med. Chem.* 45, 1983–1993
41 Wrighton, S.A. *et al.* (2000) The human CYP3A subfamily: practical considerations. *Drug Metab. Rev.* 32, 339–361
42 Ekins, S. *et al.* (2003) Generation and validation of rapid computational filters for CYP2D6 and CYP3A4. *Drug Metab. Dispos.* 31, 1077–1080
43 Ekins, S. *et al.* (1999) Three dimensional quantitative structure activity relationship (3D-QSAR) analysis of CYP3A4 substrates. *J. Pharmacol. Exp. Ther.* 291, 424–433
44 Lewis, D.F.V. (2000) On the recognition of mammalian microsomal cytochrome P450 substrates and their characteristics. *Biochem. Pharmacol.* 60, 293–306
45 Riley, R.J. *et al.* (2001) development of a generalized, quantitative physicochemical model of CYP3A4 inhibition for use in early drug discovery. *Pharm. Res.* 18, 652–655
46 Ekins, S. *et al.* (1999) Three and four dimensional-quantitative structure activity relationship analyses of CYP3A4 inhibitors. *J. Pharmacol. Exp. Ther.* 290, 429–438
47 Ekins, S. *et al.* (2003) *In vitro* and pharmacophore insights into CYP3A enzymes. *Trends Pharmacol. Sci.* 24, 161–166
48 Korzekwa, K.R. *et al.* (1998) Evaluation of atypical cytochrome P450 kinetics with two-substrate-models: evidence that multiple substrates can simultaneously bind to cytochrome P450 active sites. *Biochemistry* 37, 4137–4147
49 Ekins, S. *et al.* (1998) Autoactivation and activation of cytochrome P450s. *Int. J. Clin. Pharmacol. Ther.* 36, 642–651
50 Stresser, D.M. *et al.* (2000) Substrate-dependent modulation of CYP3A4 catalytic activity: analysis of 27 test compounds with four fluorimetric substrates. *Drug Metab. Dispos.* 28, 1440–1448
51 Wang, R.W. *et al.* (2000) Human cytochrome P-450 3A4: *in vitro* drug–drug interaction patterns are substrate-dependent. *Drug Metab. Dispos.* 28, 360–366
52 Kenworthy, K.E. *et al.* (1999) CYP3A4 drug interactions: correlation of ten *in vitro* probe substrates. *Br. J. Clin. Pharmacol.* 48, 716–727
53 Domanski, T.L. *et al.* (2001) Phenylalanine and tryptophan scanning mutagenesis of CYP3A4 substrate recognition site residues and effect on substrate oxidation and cooperativity. *Biochemistry* 40, 10150–10160
54 Hosea, N.A. *et al.* (2000) Elucidation of distinct ligand binding sites for the cytochrome P450 3A4. *Biochemistry* 39, 5929–5939
55 Jones, J.P. *et al.* (2002) Computational models for cytochrome P450: a

predictive electronic model for aromatic oxidation and hydrogen abstraction. *Drug Metab. Dispos.* 30, 7–12
56 Jones, J.P. and Korzekwa, K.R. (1996) *Predicting the Rates and Regioselectivity of Reactions Mediated by the P450 Superfamily*, Academic Press
57 Singh, S.B. *et al.* (2003) A model for likely sites of CYP3A4-mediated metabolism on drug-like molecules. *J. Med. Chem.* 46, 1330–1336
58 Shen, M. *et al.* (2003) Development and validation of k-nearest neighbour QSPR models of metabolic stability of drug candidates. *J. Med. Chem.* 46, 3013–3020
59 Ekins, S. and Obach, R.S. (2000) Three dimensional-quantitative structure–activity relationship computational approaches of prediction of human *in vitro* intrinsic clearance. *J. Pharmacol. Exp. Ther.* 295, 463–473
60 Korolev, D. *et al.* (2003) Modeling of human cytochrome P450-mediated drug metabolism using unsupervised machine learning approach. *J. Med. Chem.* 46, 3631–3643
61 Williams, P.A. *et al.* (2003) Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* 424, 464–468
62 Wandell, C. *et al.* (1999) P-glycoprotein and cytochrome P-450 3A inhibition: dissociation of inhibitory potencies. *Cancer Res.* 59, 3944–3948
63 Zhang, E.Y. *et al.* (2002) Modeling of active transport systems. *Adv. Drug Deliv. Rev.* 54, 329–354
64 Stouch, T.R. and Gudmundsson, O. (2002) Progress in understanding the structure–activity relationships of P-glycoprotein. *Adv. Drug Deliv. Rev.* 54, 315–328
65 Penzotti, J.E. *et al.* (2002) A computational ensemble pharmacophore model for identifying substrates of P-glycoprotein. *J. Med. Chem.* 45, 1737–1740
66 Pajeva, I.K. and Wiese, M. (2002) Pharmacophore model of drugs involved in P-glycoprotein multidrug resistance: explanation of structural variety (hypothesis). *J. Med. Chem.* 45, 5671–5686
67 Garrigues, A. *et al.* (2002) Characterization of two pharmacophores on the multidrug transporter P-glycoprotein. *Mol. Pharmacol.* 62, 1288–1298
68 Ekins, S. *et al.* (2002) Three dimensional quantitative structure–activity relationships of inhibitors of P-glycoprotein. *Mol. Pharmacol.* 61, 964–973
69 Ekins, S. *et al.* (2002) Application of three dimensional quantitative structure–activity relationships of P-glycoprotein inhibitors and substrates. *Mol. Pharmacol.* 61, 974–981
70 Watkins, R.E. *et al.* (2003) Coactivator binding promotes the specific interaction between ligand and the pregnane X receptor. *J. Mol. Biol.* 331, 815–828
71 Mankowski, D.C. and Ekins, S. Prediction of human drug metabolizing enzyme induction. *Curr. Drug Metab.* (in press)
72 Watkins, R.E. *et al.* (2001) The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity. *Science* 292, 2329–2333
73 Watkins, R.E. *et al.* (2003) 2.1A crystal structure of human PXR in complex with the St John's Wort compound hyperforin. *Biochemistry* 42, 1430–1438
74 Ekins, S. and Erickson, J.A. (2002) A pharmacophore for human pregnane-X-receptor ligands. *Drug Metab. Dispos.* 30, 96–99
75 Doman, T.N. *et al.* (2002) Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J. Med. Chem.* 45, 2213–2221
76 Carlson, H.A. *et al.* (2000) Developing a dynamic pharmacophore model for HIV-1 integrase. *J. Med. Chem.* 43, 2100–2114
77 Tukey, R.H. and Strassburg, C.P. (2000) Human UDP-glucuronosyltransferases: metabolism, expression, and disease. *Annu. Rev. Pharmacol. Toxicol.* 40, 581–616
78 Duffel, M.W. *et al.* (2001) Enzymatic aspects of the phenol (aryl) sulfotransferases. *Drug Metab. Rev.* 33, 369–395
79 Dajani, R. *et al.* (1999) X-ray crystal structure of human dopamine sulfotransferase, SULT1A3. *J. Biol. Chem.* 274, 37862–37868
80 Gamage, N.U. *et al.* (2003) Structure of a human carcinogen-converting enzyme, SULT1A1. *J. Biol. Chem.* 278, 7655–7662
81 Mackenzie, P.I. *et al.* (1997) The UDP glycosyltransferase gene superfamily: recommended nomenclature based on evolutionary divergence. *Pharmacogenetics* 7, 255–269
82 Ethell, B.T. *et al.* (2002) Quantitative structure activity relationships for the glucuronidation of simple phenols by expressed human UGT1A6 and UGT1A9. *Drug Metab. Dispos.* 30, 734–738
83 Takenaga, N. *et al.* (2002) Structure-activity relationship in O-glucuronidation of indocarbazole analogs. *Drug Metab. Dispos.* 30, 494–497
84 Smith, P.A. *et al.* (2003) Pharmacophore and quantitative structure-activity relationship modeling: complementary approaches for the rationalization and prediction of UDP-glucuronosyltransferase 1A4 substrate selectivity. *J. Med. Chem.* 46, 1617–1626
85 Sorich, M. *et al.* (2002) Pharmacophore and quantitative structure activity relationship modelling of UDP-glucuronosyltransferase 1A1 (UGT1A1) substrates. *Pharmacogenetics* 12, 635–645