Toward *in silico* structure-based ADMET prediction in drug discovery

Gautier Moroy[1], Virginie Y. Martiny[1], Philippe Vayer[2], Bruno O. Villoutreix[1] and Maria A. Miteva[1]

[1] Université Paris Diderot, Sorbonne Paris Cité, Molécules Thérapeutiques *In Silico*, Inserm UMR-S 973, 35 rue Helene Brion, 75013 Paris, France
[2] BioInformatic Modelling Department, Technologie Servier, 45007 Orléans Cedex 1, France

Quantitative structure–activity relationship (QSAR) methods and related approaches have been used to investigate the molecular features that influence the absorption, distribution, metabolism, excretion and toxicity (ADMET) of drugs. As the three-dimensional structures of several major ADMET proteins become available, structure-based (docking-scoring) computations can be carried out to complement or to go beyond QSAR studies. Applying docking-scoring methods to ADMET proteins is a challenging process because they usually have a large and flexible binding cavity; however, promising results relating to metabolizing enzymes have been reported. After reviewing current trends in the field we applied structure-based methods in the context of receptor flexibility in a case study involving the phase II metabolizing sulfotransferases. Overall, the explored concepts and results suggested that structure-based ADMET profiling will probably join the mainstream during the coming years.

Introduction

The success of a drug is determined not only by good efficacy but also by an acceptable ADMET profile. Although a large variety of medium- and high-throughput *in vitro* ADMET screens are available, being able to predict some of these properties *in silico* is valuable. Today, it is recognized that employing computational ADMET, in combination with *in vivo* and *in vitro* predictions as early as possible in the drug discovery process, helps to reduce the number of safety issues [1]. Moreover, there is a pressure to reduce the number of animal experiments (e.g. the REACH project). Traditionally, data modeling methods, such as expert systems and quantitative structure–activity (property) relationships (QSARs/QSPRs) [2,3], have been used to investigate ADMET properties. These methods use statistical and learning approaches, molecular descriptors and experimental data to model complex biological processes (e.g. oral bioavailability, intestinal absorption, permeability and mutagenicity [2,4]). The rules for drug-likeness or lead-likeness or metabolite-likeness [5,6] relying on simple physicochemical properties are also well-known and implemented in commercially and freely available packages [4,7,8]. However, limitations of all these approaches come from the fact that high quality experimental data are seldom available [9], and that the approaches tend to neglect direct structural information about the ADMET proteins. *In silico* approaches based on the 3D structures of these proteins could therefore be an attractive alternative or could complement ADMET data-modeling techniques [10].

The first attempt to predict ADMET taking into account the protein structures at the atomic level started about ten years ago with the early homology models of human cytochrome P450 (CYP) [11,12]. Several new studies have recently been reported that exploit the 3D structures of ADMET proteins, molecular docking and different strategies for taking into account protein flexibility during the process. They all highlight that these proteins are difficult to investigate – in part because of the presence of large and flexible ligand-binding cavities that can interact with diverse ligands. Most of these investigations focus on phase I metabolizing enzymes such as CYP (for recent key reviews, see Refs [10,13,14]). To date, predictions of interactions between drug candidates and phase II metabolizing enzymes based on 3D protein structures are still essentially missing.

Corresponding author:. Miteva, M.A. (maria.miteva@univ-paris-diderot.fr)

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2011.10.023

Here, we synthesize recently reported *in silico* studies aiming at predicting small molecules binding to ADMET-related proteins based on the knowledge of the 3D structures of these macromolecules with a special emphasis on metabolizing enzymes. Numerous advantages of direct 3D approaches compared with QSAR or expert systems have been mentioned such as limitations of the applicability domain owing to the small number of compounds used to develop the models [15]. Also, these models cannot usually explain the molecular mechanisms taking place because a full understanding requires analysis of the reactions at the atomic level in the context of a flexible ligand and protein 3D structure. Further, only a few reviews on ADMET predictions using the 3D structure of relevant proteins have been reported to date, although the field is gaining momentum, in particular for phase II metabolizing enzymes that have been overlooked despite their importance. Clearly, additional efforts are needed to advance the exploitation of the growing amount of 3D structural information; this will also require receptor flexibility to be taken into account more efficiently, and to tailor the scoring functions for these promiscuous ADMET-related proteins. Recent developments in *in silico* structure-based methods enable the tackling of some of these issues as discussed in this review. Along this line, we apply structure-based methods combined with receptor flexibility simulations [16–18] to study sulfotransferases (SULTs). Indeed, sulfonation can be crucial for ADMET processes because it can increase the elimination of drugs or, in some cases, induce toxic effects through the formation of highly reactive intermediates [19] or provoke genotoxicity [20].

Metabolizing enzymes

CYPs

CYP enzymes are the major, and most studied, phase I drug-metabolizing enzymes mediating oxidation of a variety of compounds involved in various physiological and pathophysiological processes, including detoxification of xenobiotic compounds and sometimes bioactivation of nontoxic to toxic reactive intermediates and procarcinogens (Table 1). CYPs are also implicated in drug–drug interactions (DDIs) mediated by drug inhibition or induction [21,22]. It has been estimated that ~75% of the marketed drugs are metabolized by CYPs, with five major CYP isoforms

TABLE 1

Drugs known to interact with ADMET proteins

ADMET proteins	Drugs
CYP	R–H + O_2 + NADPH + H^+ → R–OH + $NADP^+$ + H_2O
3A4	Antiallergic (terfenadine[a]), antibiotic (cyclosporin[b], troleandomycin[a], clarithromycin[a], erythromycin[a]), anti-VIH (indinavir[a], amprenavir[a], ritonavir[a], fluvoxamine[b]), antifungal (itraconazole[a], miconazole[b]), anticancer (tamoxifen[a]), CNS (desipramine[c])
2D6	Basic substrates (bufuralol[a], metoprolol[a]); atypical substrates: acidic (pactimibe[c]) or neutral (i.e. not containing a basic nitrogen) (spirosulfonamide, steroids) Antiarrhythmics (aprindine[a], flecainidine[b], propafenone[a]), beta-blockers (bufuralol[a], propranolol[c]), antiallergic (terfenadine[a], azelastine[c]), analgesics (quinine[b]), antidepressants (fluoxetine[a])
2C9	Compounds containing dipole or negative charge, lipophilic groups: piroxicam[a], ibuprofen[a], chlorpropamide[a], nateglinide[a], phenytoin[a], warfarin[a], diclofenac[a]
2B6	Anticancer (cyclophosphamide[a], ifostamide[c]), analgesic (propofol[a]), antiretroviral (efavirenz[a]), andidepressant (bupropion[a])
SULT	R–OH + PAPS → R–O–SO_3H + PAP SULT1A1: uncharged phenolic compounds, hydrophobic, small flat aromatic; SULT1A3: positively charged, simple aromatic phenol; SULT1B: thyroid hormones; SULT1C: arylhydoxylamine; SULT1E1: steroids, estrogen; SULT2A1: hydoxysteroids, DHEA; SULT2B: cholesterol Analgesic (diflunisal[b]), anti-inflammatory non-steroidal (ibuprofen[b]), antihypertensive (minoxidil[c]), antiulcerative (naringenin[a]), antineoplastic (tangeretin[b]), antibiotic (triclosan[a]), antifungal (salicylic acid[b]), antioxidant (naringin[b]), antidiabetic (troglitazone[c]), anticonvulsant (chrysin[c]), antiasthmatic (epicatechin[a])
UGT	R–OH + UDP-glucuronic acid → R–O–glucuronic acid + UDP Aromatic or aliphatic hydroxyls, carboxylic acid and amines: anticancer (epirubicin[c]), opioid-based pain medicines (codeine[a], cocaine[b], morphine[a]), anti-inflammatory non-steroidal (ibuprofen[a], flurbiprofen[a], indomethacin[a]), antihistaminic (cetirizine[b], astemizole[b], diphenhydramine[a], loratadine[b]), anxiolytic (diazepam[b], medazepam[b], flunitrazepam[b]), antibiotic (triclosan[a], sulfamethoxazole[b])
PXR	Large and hydrophobic: antibiotic (dicloxacillin, rifapentine, mevastatin), antidiabetic (mevastatin, troglitazone, pioglitazone), anti-HIV (ritonavir), anti-inflammatory non-steroidal (celecoxib, flurbiprofen, phenylbutazone), anti-inflammatory steroidal (betamethasone, pregnenolone, dexamethasone, hydrocortisone), anticoagulant (warfarin)
CAR	Large and hydrophobic: antifungal (clotrimazole, ketoconazole), antipsychotic (thioridazine), myorelaxant (temazepam), antineoplastic (PK11195), androstenol, coumestrol
HSA	Anti-inflammatory non-steroidal (ibuprofen, diflunisal, azapropazone), anticoagulant (warfarin), anxiolytic (diazepam), analgesic (propofol)
AGP	Apolar, basic or lipophilic compounds: alpha blocker (nicergoline, prazosin), analgesic (methadone, fentanil), antiarrhythmic (aprindine, disopyramide), antibiotic (erythromycin), anticoagulant (warfarin), beta-blocker
Herg	Astemizole, cisapride, haloperidol, sertindole, thioridazine, terfenadine
P-gp	Antidepressant (desvenlafaxine, nefazodone, fluoxetine, phenelzine), antiviral (atazanavir, ritonavir), antibiotic (retapamulin, romidepsin, erythromycin), antipsychotic (tetrabenazine, trifluopromazine, haloperidol), antidiabetic (dapagliflozin, sitagliptin, repaglinide), antineoplastic (vinblastine, tamoxifen)

Information sourced from Aureus Sciences database http://www.aureus-pharma.com/. The catalyzed reactions for CYP, SULT and UGT are schematically shown for the enzymes and ligand action is noted as follows: Substrate and Inhibitor[a]; Inhibitor[b]; Substrate[c].

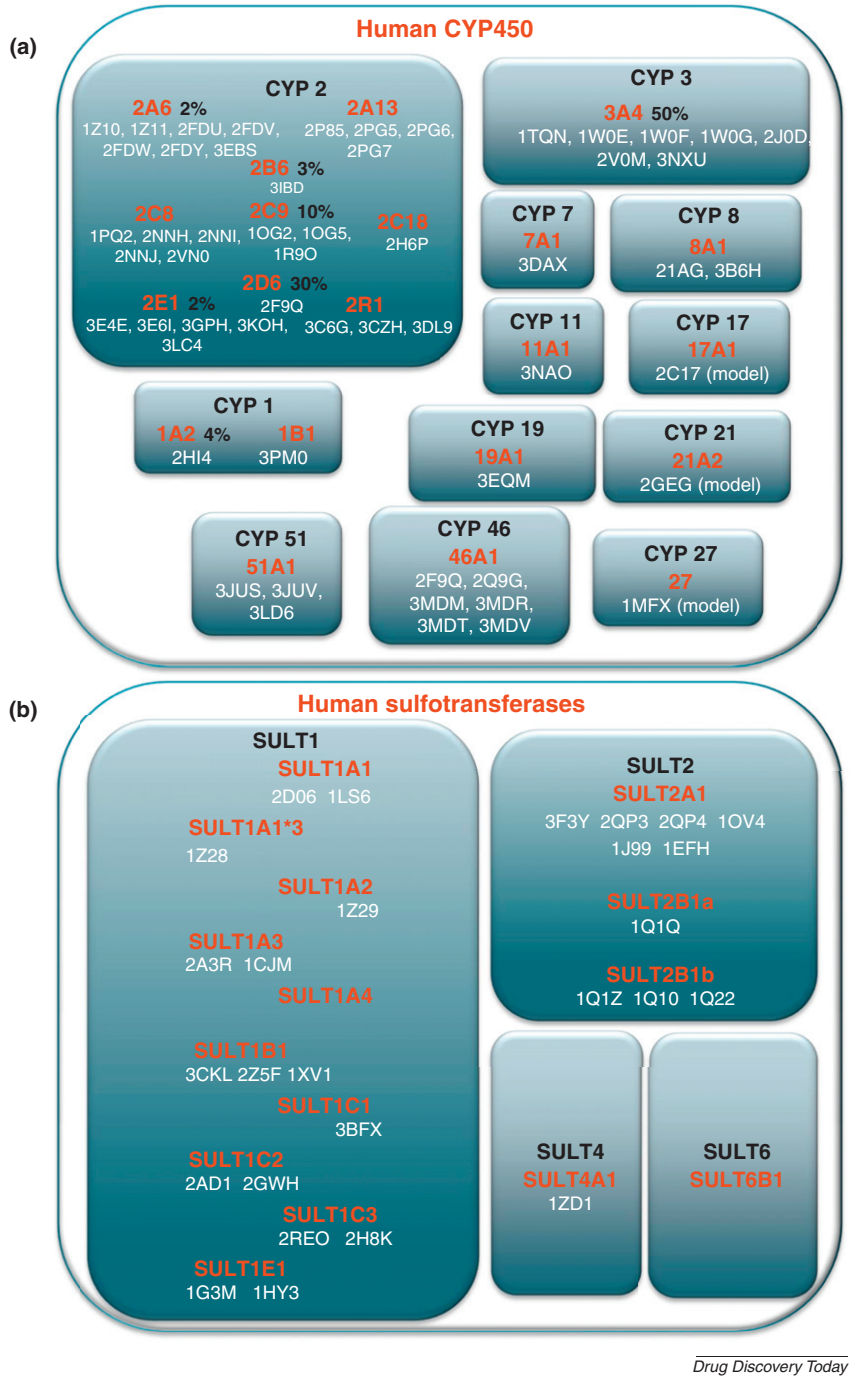


FIGURE 1

Available 3D structures from the PDB for **(a)** CYPs and **(b)** SULTs.

involved in 75–90% of CYP-related metabolism [23] (Fig. 1a). Interestingly, a variability of the drug metabolization rate is observed in several CYPs, among them CYP2D6 and CYP2C9, owing to their high polymorphism.

Today, numerous structures of human CYPs are available in the Protein Data Bank (PDB) (Fig. 1a), and they all share a similar fold (Fig. 2). The active site is generally large and flexible and sometimes more than one ligand can bind simultaneously. Yet, important differences are observed when conducting in-depth analyses of the active sites of the different CYP enzymes.

CYP3A4 metabolizes ~50% of all drugs [24] and displays a large and flexible active site. Many studies reported during the past decade outlined the importance of accounting for the flexibility of proteins involved in the ADMET processes [10,25–27]. Modeling the binding of small molecules to ADMET-related proteins without considering flexibility can lead to many artifacts, in particular those due to relatively large conformational changes potentially induced by different ligands, as seen for example from the X-ray structures of CYP3A4 bound to ritonavir or metyrapone. On the basis of the CYP3A4 experimental structure, several possible

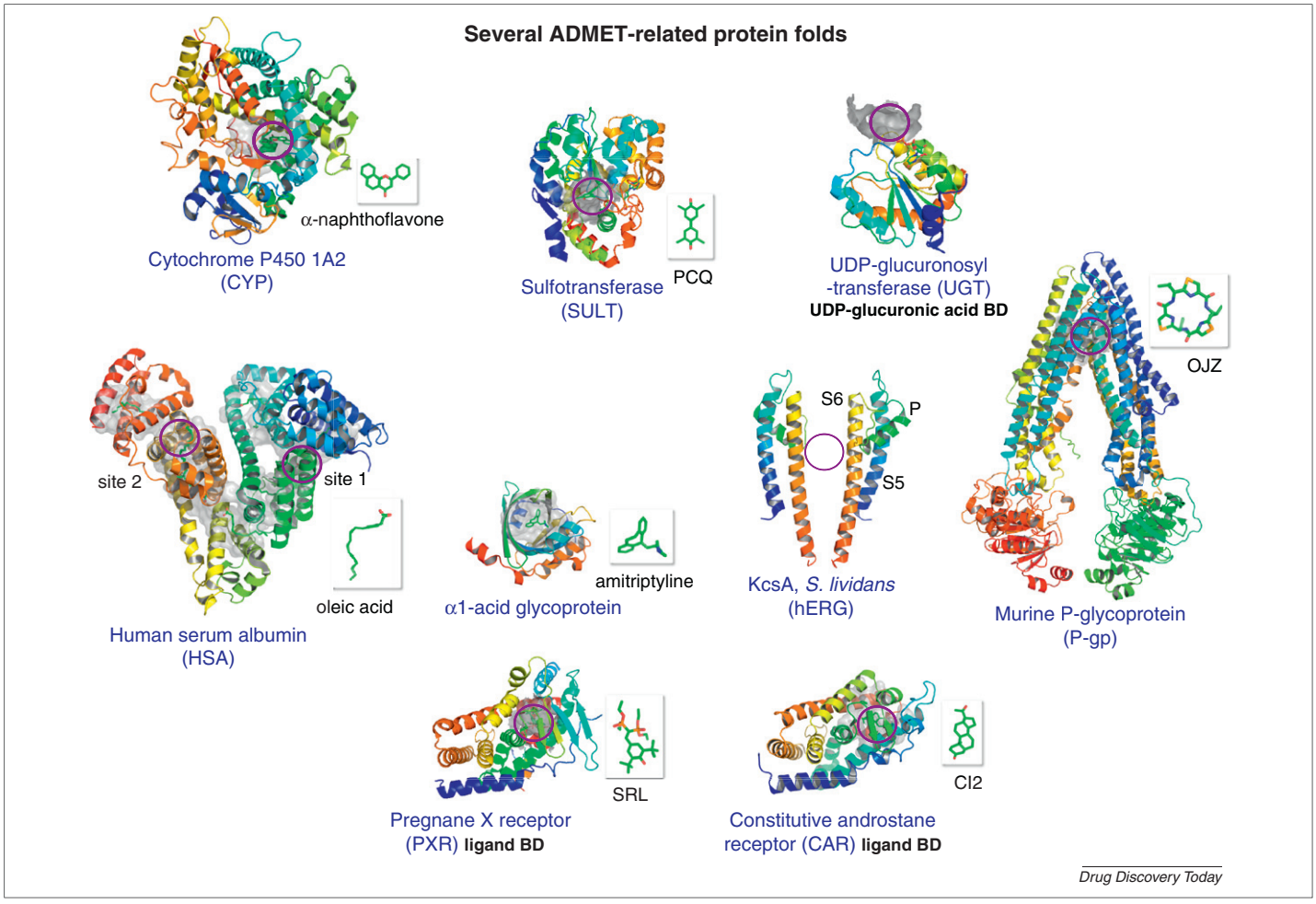


FIGURE 2

Several ADMET-related protein folds and bound ligands. The ligand-binding sites are highlighted as a grey surface and pink circle: human CYP (PDB ID: 2HI4); human SULT (PDB ID: 1G3M); UDP-glucuronic acid binding domain (BD) of human UGT (PDB ID: 2O6L); human HSA (PDB ID: 1GNI); human AGP (PDB ID: 3APV); hERG – a schematic representation of S5, S6 and P helices (KcsA; PDB ID: 1K4C); murine P-gp (PDB ID: 3G60); ligand BD of human PXR (PDB ID: 1NRL); ligand BD of human CAR (PDB ID: 1XV9).

binding modes were investigated by Vedani and Smiesko [10]. The authors combined flexible docking and multidimensional QSAR to evaluate the inhibitory potential of 48 compounds. This approach was validated on experimental holo structures and experimental metabolism data for CYP3A4. A promising strategy was recently designed to predict regioselectivity of some ligands of CYP3A4 through a combination of docking, molecular dynamics (MD) simulations and quantum-chemistry-based calculations of the activation energy [14].

CYP2D6 is the second-most studied drug-metabolizing enzyme. CYP2D6 shows the largest phenotypic variability among the CYPs, largely owing to genetic polymorphism. Although the crystal structure of CYP2D6 was released in 2006 [28], structure-based methods initially made use of homology models to investigate CYP2D6 interaction with its ligands. For instance, Kemp *et al.* applied homology modeling, docking with GOLD (see Supplementary material Table I) and scoring with ChemScore, and they successfully identified several compounds from the National Cancer Institute database as being CYP2D6 inhibitors [29]. Yet, no sufficient correlation between the ChemScore values and the experimental log IC_{50} has been obtained ($r^2 = 0.61$). Later, MD simulations and simulated annealing protocols were performed to generate 20 different conformations of CYP2D6 [30]. On the basis of the docking scores, the authors used a neural network model to identify different CYP2D6 conformations relevant for the binding affinity prediction. Another study demonstrated that the accuracy of the docking and virtual screening on a homology model of CYP2D6 can be improved by adding water molecules to the active site [31]. In this direction, MD simulations [32] suggested there were 12 hydration sites in the active site of CYP2D6 that could be exploited during docking and virtual screening experiments. In a recent study, the flexibility of the CYP2D6 active site was analyzed with the aim of carrying out virtual screening computations [33]. Sixty-five substrates were docked into 2500 structures extracted from MD simulations and a binary decision tree was used to find the three most essential structures enabling the accurate prediction of the metabolism site for most of the ligands. At the end, 80% of the sites of metabolism were correctly predicted by this approach. Recently, homology modeling and a docking study with Glide highlighted the importance of taking into account induced-fit adaptations upon ligand binding [34]. Indeed, the authors obtained an 85% success rate for identifying the site of metabolism when they docked CYP2D6 substrates into a homology model based on the holo CYP2C5 crystal structure,

whereas a lower success rate was obtained on the apo crystal structure of CYP2D6.

Several *in silico* studies attempted to predict drug interactions with CYP2C9 based on the three human crystal structures available at the PDB – one being ligand-free (apo) and the two others complexed with either warfarin or flurbiprofen. A recent investigation into the interaction mechanism between CYP2C9 and proton pump inhibitors (PPIs) [35] highlighted the importance of a hydrogen-bond network involving PPIs, water molecules and some binding site residues. The importance of including explicit water molecules in docking exercises is often discussed in the literature because they can mediate the substrate–enzyme reaction. In fact, the positions of the water molecules can be crucial, because it has been observed when predicting metabolic sites of CYP-mediated metabolic reactions [25]. To improve the prediction of ligand affinity toward CYP2C9, Stjernschantz and Oostenbrink developed a protocol combining docking, MD simulations and free energy calculations with the linear interaction energy (LIE) approach [36]. Rossato *et al*. combined MD simulations, docking experiments and a QSAR modeling scheme that included a term corresponding to the predicted binding energies of compounds against CYP2D6 and CYP2C9 [26].

Several *in silico* studies predicting ligand binding at the atomic level for other CYP isoforms have also been reported [37]. Taking into account that CYP metabolism involves binding and substrate chemical modification driven by atom reactivity toward the oxygen–heme complex, a combination of binding prediction based on the similarity between molecular interaction fields of the active site and substrates with substrate reactivity [38] is a valuable approach. This enabled MetaSite [38] to give high success rates in terms of the prediction of CYP-specific metabolites.

Together, these studies demonstrate several crucial issues that need to be solved to predict potent CYP binders accurately, such as the role of water molecules and how to incorporate protein flexibility more efficiently during the docking process.

UDP-glucuronosyltransferases

UDP-glucuronosyltransferases (UGTs) are phase II drug-metabolizing enzymes responsible for glucuronidation leading to covalent addition of the glucuronic moiety from UDP-glucuronic acid (UDPGA) to endogenous compounds and drugs. This is a major pathway for detoxification of numerous carcinogens such as polycyclic aromatic hydrocarbons (PAHs) and aryl- and hetero-cyclic amines [39]. UGT-catalyzed glucuronidation is thought to account for up to 35% of the phase II reactions. Three main isoforms: UGT2B7, UGT1A4 and UGT1A1, are responsible for drug modification of 35%, 20% and 15% of the drugs metabolized by UGTs, respectively [40]. Computational modeling of human xenobiotic glucuronidation has only started in the past decade using classication, 2D-(3D)-QSAR or regression methods [41,42].

The experimentally known crystal structure of human UGT (isoform 2B7) contains only the C-terminal UDPGA-binding domain [43] but the catalytic ligand-binding domain is not resolved yet (Fig. 2). Homology modeling of UGT2B7 based on the related plant flavonoid glucosyltransferases [44] suggests that the human UGTs share a common catalytic mechanism and this introduces the possibility of studying potential interactions with drug candidates at the atomic level [45].

Nuclear receptors

Nuclear receptors (NRs) are ligand-regulated transcription factors that control the expression of numerous genes and are generally composed of a DNA-binding domain and a ligand-binding domain. Triggering the upregulation of metabolizing-enzyme transcription, some NRs (i.e. pregnane X receptor, constitutive androstane receptor) can indirectly induce undesirable DDIs. Other NRs, such as androgen receptor, estrogen receptor, glucocorticoid receptor, thyroid receptor, bind endocrine disruptors which interfere with the function of the endocrine system and might affect reproductive, developmental, immunological and neurobiological functions. The ligand-binding and induced-fit effects have been modeled based on the 3D structures of several NRs [10,27,46].

Pregnane X receptor

Pregnane X receptor (PXR) is implicated in the regulation of the CYP3A, UGT and ABC transporter genes [47]. PXR is activated by a wide range of compounds, such as pregnane compounds, antifungals and glucocorticoids. Two structures of the human apo PXR and seven structures of the human ligand-bound PXRs are now known and have been found to be similar, but only the ligand-binding domain is solved.

Recently, docking of small organic molecules was used to investigate the structural basis for PXR xenobiotic recognition [48]. Five hotspots with a major contribution in the binding free energy between PXR and ligands have been identified at the surface of the ligand-binding domain. Ekins and co-workers combined docking, hybrid scoring strategies and 3D-QSAR modeling on several PXR structures to improve the prediction of PXR agonists among the ToxCast™ database and a steroid database [49]. In a recent study, Chinese herbal molecules were docked into the ligand-binding cavity of PXR. The results suggest that some molecules can interact with PXR and therefore act on CYP3A4 [50]. The large number of potential arrangements within the binding site seem to explain why PXR can accommodate a large variety of compounds, whereas structural analysis and molecular modeling suggest a unique signal transduction mechanism between the PXR homodimerization interface and its coactivator binding site [48].

Constitutive androstane receptor

Constitutive androstane receptor (CAR) reveals *in vitro* activity even in the nonliganded state. *In vivo* CAR is quiescent in the cytoplasm and, upon treatment with an inducer, it can translocate to the nucleus where it activates the transcription of genes including some CYP members, UGT, SULTs and ABC transporters, thus influencing drug response [51].

Two structures of the human CAR ligand-binding domain have been solved in complex with CITCO (see Supplementary material Figure I) or 5β-pregnanedione [52]. The general fold is similar to the PXR fold (Fig. 2). On the basis of homology models, complexed with ligands or in the uncomplexed state, the flexibility of the CAR binding site has been investigated by MD simulations combined with molecular docking [53]. Küblbeck and co-workers performed virtual screening of the LeadQuest® database on a human CAR structure [54]. Among the 66 compounds tested *in vitro*, 19 substituted sulfonamides and thiazolidin-4-one derivatives were identified as CAR agonists. Recently, the models of two human CAR

splicing variants were built using human CAR structures [55]. Alternative splicing of the CAR gene seems to increase the diversity of compounds that can activate CAR. Although the ligand-binding domains are similar overall, the presence of specific residues results in small changes in the binding site leading to different affinities for the CAR isoforms and possible different binding modes [55].

Estrogen receptors

Estrogen receptors (ERs) exert a set of effects on the differentiation, growth and maintenance of many tissues. Two distinct ERs, ERα and ERβ, compete to bind 17β-estradiol, which plays an important part in mediating sexual development. Vedani and co-workers built a model to predict molecules that can trigger toxic effects through their binding to ERs [56]. They docked 106 compounds into the ER binding pocket and showed that retaining up to four of the best binding modes is helpful for a successful creation of a 6D-QSAR model.

Androgen receptor

Androgen receptor (AR) is known to be involved in the growth of the prostate gland as well as the development of prostate cancer. Identification of potential drugs that can disrupt hormonal systems via binding into AR is important for their ADMET evaluation. To develop a system for the prediction of ligand binding to AR, 119 AR ligands were analyzed by flexible docking, MD, LIE and multi-dimensional QSAR methods [57]. Indeed, the built model could predict the binding energy of AR ligands correctly. Further, marketed oral drugs were docked into multiple structures of AR using the software ICM [58] and three antipsychotic drugs, acetophenazine, fluphenazine and periciazine, have been identified as being weak nonsteroidal AR antagonists, correlating well with endocrine side effects observed in individuals taking these medications.

Finally, a recent modeling study on 13 NRs [27], such as AR, ER, glucocorticoid receptor and mineralcorticoid receptor, has been reported. To consider the protein plasticity, multiple NR conformations collected from crystallographic structures were taken and 157 diverse NR ligands were docked into the multiple ensembles. Although the near-native ligand-binding geometry was reproduced in 89% of the cases of employed ensemble docking, 78% cases were correctly reproduced when a single receptor was used, underlining the gain in accuracy when taking into account the receptor flexibility.

Plasma-binding proteins

Several blood plasma proteins, such as human serum albumin (HSA), α1-acid glycoprotein (AGP), lipoproteins and α-, β-, γ-globulins, bind a wide variety of drugs and endogenous compounds. HSA and AGP are known to bind ~300 drugs. Because only the free drug concentration determines the pharmaceutical activity, possible interaction of compounds with plasma protein binding capability has to be considered during the drug discovery process.

Human serum albumin

HSA represents 60% of total human plasma protein, with a concentration ~0.6 mM [59]. It has a crucial role in the transport of relatively insoluble endogenous compounds such as fatty acids, hormones and vitamins. More than 70 experimental structures of HSA, free or bound with endogenous compounds or drugs, are available in the PDB. HSA is an α-helix single-chain protein formed by three homologous domains (I–III) arranged in a heart-shaped fashion [60] (Fig. 2). Seven fatty-acid-binding sites are distributed throughout the protein, whereas most drugs bind to one of the two primary binding sites located in domains IIA and IIIA [60], and allosteric mechanisms can also be involved [61]. Ligands that can bind to the highly flexible site I (the so-called warfarin site) are usually negatively charged large heterocyclic and/or dicarboxylic acid compounds (warfarin, azapropazone and dansylamide) [62]. The smaller site II (the diazepam site) often binds aromatic carboxylic acids (i.e. diazepam, ibuprofen and arylpropionic acids). Other molecules, endogenous (bilirubin, etc.) and some drugs (propofol, oxyphenbutazone, among others) bind within the five other sites identified in the three domains. Interestingly, several compounds can bind to at least two sites, such as azapropazone, indomethacin and fatty acids.

The knowledge of the 3D HSA structure was successfully used to optimize potential therapeutic molecules for preventing their binding to HSA, like methionine aminopeptidase-2 inhibitors [63] and Bcl-2 inhibitors [64]. Molecular docking studies have been used to examine the interaction of creatinine with HSA and attempt to shed new light onto the mechanism of uremic toxin disposition in renal disease state [65]. Other *in silico* approaches [i.e. molecular docking, MD simulations and molecular mechanics/generalized born surface area (MM/GBSA) analysis] have been employed to predict the binding modes of flavonic compounds to HSA [66]. In a recent study, the authors exploited the LIE method in addition to the aforementioned ones to evaluate the binding free energy of complexes formed by HSA and two perfluorooctanoic acid compounds [67]. Despite the numerous crystal structures of HSA it is still challenging to predict ligand binding because of the different binding sites present on the protein and the possible subtle or more-significant conformational changes [68]. An online service using support vector machines and automated docking with the Autodock Vina program has recently been reported and might help to predict ligand binding to HSA [69].

α-1-Acid glycoprotein

The biological function of the acute phase AGP is not completely understood yet [70]. AGP has a lipocalin fold and is able to bind and transport several hundreds of endogenous or exogenous molecules, predominantly apolar, basic or lipophilic compounds. Interestingly, in human AGP there is a mixture of two or three genetic variants. Although many drugs have similar binding constants, others (e.g. promethazine, warfarin and dipyridamole) demonstrate considerable differences between the AGP variants [71].

Whereas AGP has been extensively studied since the 1980s, a first structural model was only built in 2003 [72]. The authors used a combination of vibrational spectroscopies, homology modeling based on the crystal structure of the bilin-binding protein and molecular docking to propose a model of the AGP-progesterone interaction. Interestingly, although AGP is one of the most glycosylated proteins in the human body, the oligosaccharide moieties do not appear to play a part in ligand binding [62]. The first

high-resolution structure of the F1*S variant of AGP was solved in 2008 [73] with a (2R)-2,3-dihydroxypropyl acetate compound bound in the upper portion of the binding site. To gain insights into the AGP binding mode, the authors performed molecular docking with diazepam and progesterone. Recently, high-resolution structures of the A variant of AGP alone and bound with amitriptyline, chlorpromazine and disopyramide have been released [70]. The structural comparison of AGP has shown that the drug-binding pocket of the A variant is narrower than that of the F1*S variant. Hence, for some drugs the selectivity of AGP variants has been explained by differences in the nature and shape of the binding pockets.

The human ether-a-go-go-related gene potassium channel

QT prolongation is related to ventricular repolarization that is under the control of Na^+ and Ca^{2+} and K^+. The rapid and slow delayed rectifier currents, IKr and IKs, are produced by the hERG (human ether-a-go-go-related gene) and the KCNQ1 K^+ channels. The QT prolongation regulatory document [74] recommends to study hERG with *in vitro* binding, electrophysiology or *in vivo* studies in the guinea pig, recording ECG. Several drugs are known to prolong the QT/QTc interval without blocking the hERG channel, and vice versa. Thus, there are other unknown mechanisms or multiple ion (Na^+ and Ca^{2+} and K^+) channels involved in QT prolongation [75]. A more global assay using Rb flux, which reflects the resulting effect on both ion channels, has been proposed in QT studies [76].

Still, at present, hERG blockers are the most explored ones with *in silico* approaches. A wide variety of compounds prolong the QT interval through actions on the hERG channel. Among them, several drugs, such as astemizole, cisapride, grepafloxacin, haloperidol, pimozide, sertindole, thioridazine and terfenadine, have been restricted or withdrawn from the market [77].

hERG is a tetramer, with each subunit containing six transmembrane helices. The helices S1–S4 form the voltage-sensor domain, and S5–S6 form the pore domain. The drug-binding site is within the cavity formed by four S6 helices of the hERG tetramer (Fig. 2). To date, only an extracellular loop and the cytoplasmic N-terminal domain involved in the regulation of the channel opening are experimentally known. Different homology models of the pore domain have been proposed based on known structures of other eukaryotic voltage-sensitive channels or bacterial potassium channels [78–80].

Several homology models have been exploited to elucidate drug blocking of the hERG channel [81]. Based on docking computations, several studies suggested that Y652 and F656 located in S6 are implicated in the hERG binding mechanism forming π stacking interactions with the ligands [78,82]. The residues T623, S624 and V625 localized near the pore helix have also been shown to be important for drug interactions [61,62]. The residues T623 and S624 interact with the polar tails present in several hERG ligands. MD simulations were also carried out to optimize the best docking solutions and to evaluate relative binding affinities by the LIE approach [80]. Insights of the drug blocking of hERG were further investigated using docking on refined homology models [83]. A set of 20 hERG blockers has been docked and a good agreement with the experimental data was obtained. Additional accuracy was achieved by rotating the helix S6 in a different model, underlining that the model quality is crucial to obtain relevant binding mode predictions. Recently, the relative importance of key residues for ligand binding was quantified (F656 > Y652 > T623 > S649) and a model to predict drug blockage of hERG was derived by combining 3D-QSAR and docking computations [84].

In the future it would be of interest to set up *in silico* models for other ion channels or, when available, on Rb flux data. The combination of these *in silico* models should result in more-complete prediction of QT prolongation. In this direction, a systems chemical biology approach has been proposed for the prediction of new possible targets related to cardiac arrhythmias [85]. In addition, a recent *in silico* study going beyond the most explored hERG has been performed on two K^+ channels, hERG (IKr) and KCNQ1 (IKs), by combining docking simulations with 3D-QSAR modeling [86]. Introducing the obtained results in electrophysiological models suggested the importance of multiscale prediction systems for preliminary screening in lead discovery.

ABC transporters

P-glycoprotein

The ATP-binding cassette (ABC) transporters form a large superfamily of proteins that transport several molecules across the membrane bilayers. To date, 48 members of the ABC superfamily have been found in humans [87]. ABC transporters are within the main focus of research interest [88], because of their involvement in multidrug resistance (MDR) and reducing the exposure of drugs through the drug transport. P-glycoprotein (P-gp), also called multidrug resistance protein 1 (MDR1), is the first ABC transporter discovered and it is the most explored transport protein. It acts as an active transporter expelling molecules out of the cell in an energy-dependent manner. P-gp has broad substrate specificity and it can recognize diverse compounds: charged or neutral, linear or circular, aromatic or non-aromatic.

The eukaryotic ABC transporters are generally composed of two membrane domains and two nucleotide-binding domains (NBDs). Experimental data suggest that they undergo large conformational changes during the transport cycle. According to the most popular ATP switch model, proposed by Higgins and Linton [89], the transporter substrate binds to its high-affinity binding site on the membrane domains from the inner leaflet of the membrane. It seems that P-gp is able to bind simultaneously to more than one ligand in up to seven binding sites [90]. The human P-gp structure has not been solved yet. Thus, during the past decade, numerous homology models of different parts of human P-gp were built based on the structure of the bacterial MsbA lipid transporter [91] or based on the Sav1866 crystal structure, an ABC protein from *Staphylococcus aureus* [92]. Recently, docking on human P-gp models in several different states (i.e. nucleotide-bound, closed and open nucleotide-free models) was used to explore structure–function relationships of the putative ligand-binding sites [93]. Two potential pathways formed by a chain of interacting residues and involved in the propagation of a signal upon ATP binding throughout the membrane domains were proposed. Recently, the X-ray structure of murine P-gp, sharing 87% sequence identity to human P-gp, has been reported in bound and unbound states [94] (Fig. 2) and several new human P-gp models were developed. On the basis of such models, Pajeva and co-workers used docking

to propose a binding mode for quinazolinones and other ligands [95]. The results also confirmed the possibility of multiple binding sites for the binders. Recently, five propafenone derivatives were docked with GOLD (and rescored via the empirical scoring function XSCORE into P-gp homology models built from the murine structures [96]. The binding poses in closed and open structures seem to be in good agreement with the available experimental data.

The high conformational flexibility of the ABC transporters, the large binding cavities composed of multiple binding sites, the ability to accommodate more than one ligand simultaneously, the change in the level of affinity of the binding sites during the transport cycle – all these factors render the modeling task extremely difficult at present [87]. Furthermore, often DDIs related to ABC transporters are due to substrate overlap with drugs targetting other proteins [21]. Several studies focused on searches of selective inhibitors for P-gp and CYP3A4 [97,98], very important for circumventing MDR effects efficiently [97].

Case study: toward prediction of drug modifications by SULTs

SULTs are a supergene family of enzymes catalyzing the transfer of a sulfonate group from 3′-phosphoadenosine 5′-phosphosulfate (PAPS) to a hydroxyl or amino group of the substrate. Sulfonation is a phase II metabolic mechanism increasing the molecular solubility and decreasing the molecular bioactivity enabling a detoxification process by a rapid excretion. However, in some cases sulfonation can lead to highly reactive metabolites that are mutagenic and carcinogenic [19,20]. The known experimental 3D structures of SULTs (Fig. 1b) show multiple ligands in the binding site or alternative orientations of the ligands in some isoforms. Despite the large number of known experimental structures for SULTs, only a few 3D *in silico* studies have been reported, probably owing to the complexity of dealing with a flexible binding site. To explain the substrate specificities for different isoforms, several SULT families have been studied by hierarchical clustering of the binding site structures [99]. Two recent works examined ligand binding to SULTs with docking methods. Stjernschantz and coworkers [100] screened 34 potential endocrine-disrupting compounds on the murine and human SULT1E1 to find selective inhibitors of the human enzyme. The active compounds discovered were docked with GOLD, and subsequent MD simulations of the docked complexes were performed explaining in part the selectivity of some of the inhibitors.

To the best of our knowledge the only reported work aiming at distinguishing between binders and non-binders for SULTs at the atomic level is a virtual screening study performed with Glide on SULT1A3 and SULT1E1 [101]. The authors explored the substrate selectivity profiles for SULT1A3 and SULT1E1 and demonstrated that docking and/or virtual screening could distinguish preferential substrate classes for each SULT. However, all results were obtained on rigid X-ray protein structures and, as suggested by the authors, it would be important to take into account receptor flexibility in such 3D profiling more efficiently.

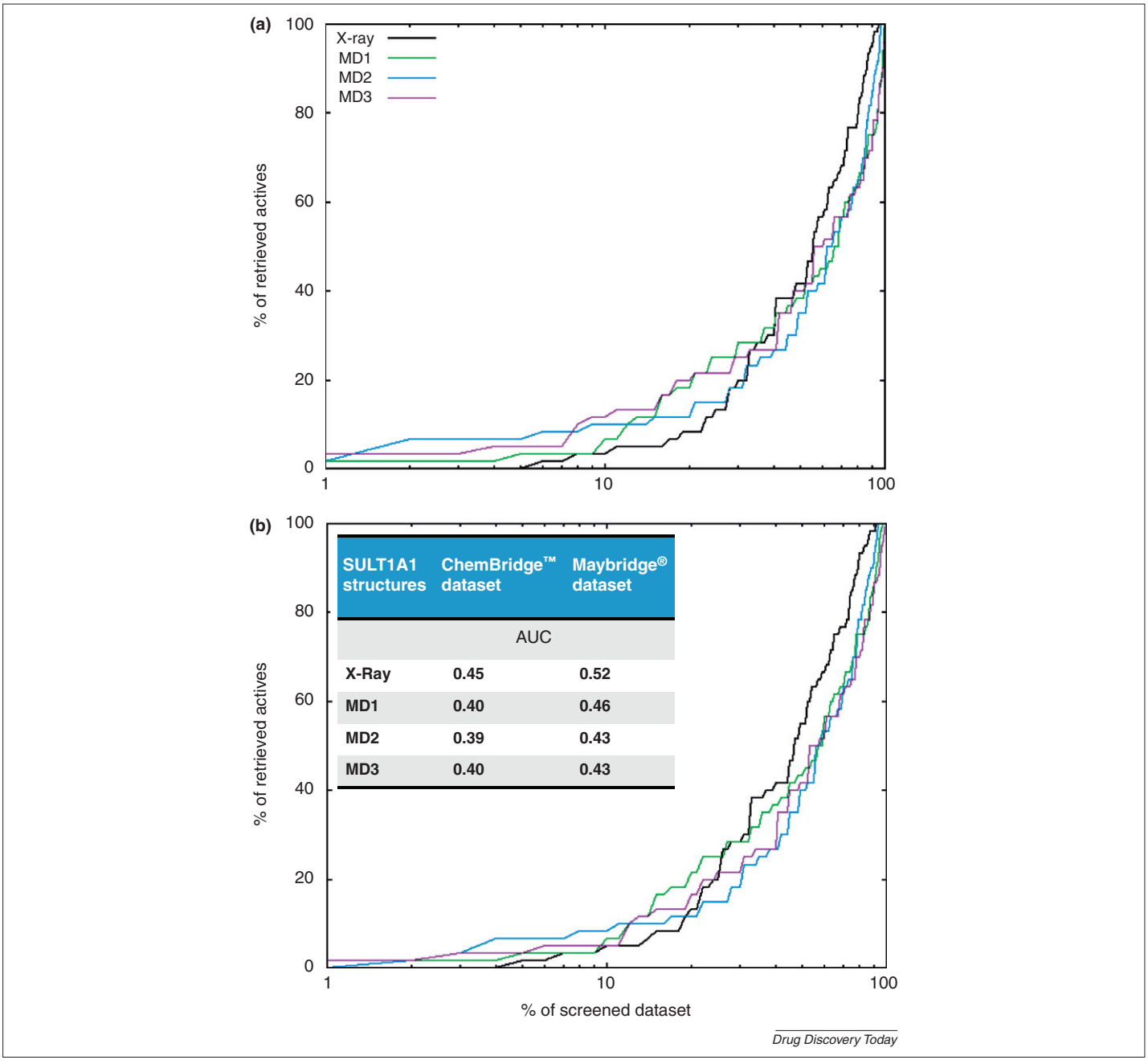
We explored the receptor flexibility of SULT1A1 to take it into consideration for prediction of ligand binding. We took the X-ray structure of the protein co-crystallized with the cofactor 3′-phosphoadenosine 5′-phosphate (PAP) and p-nitrophenol (pNP) (PDB ID: 1LS6). We ran three MD simulations of 2 ns using the CHARMM program [102] on SULT1A1 in the presence of PAP and in the absence of ligand. We extracted 4500 structures from the MD that were then clustered based on the active site with the hierarchical ascendant classification (HAC) approach implemented in the R software.

The median structures from the six obtained clusters were chosen to define a representative set of protein conformations. For the virtual screening experiments we collected 157 known substrates of SULT1A1 ([99]; databases: BRENDA, Aureus Sciences). We clustered the active molecules using the fingerprint FCFP_4 available in Pipeline Pilot v.7.5 (SciTegic, Inc/Accelrys). As decoys we took the diverse ChemBridge™ PremiumSet™ and Maybridge® HitFinder™ sets. All actives and decoys were filtered with a soft drug-like filter using the FAF-Drugs 2 server [8]. Finally, we performed virtual screening experiments with Vina 1.0 [103] on the representative MD structure set and on the X-ray structure of SULT1A1 and the 60 diverse actives were merged with the 49,496 putative decoys from the ChemBridge™ collection or with 13,088 molecules from the Maybridge® collection.

Figure 3 represents the enrichment graphs obtained for the X-ray and MD protein conformations. Although the AUC (area under curve) for the ROC (receiver operating characteristic) curves (not shown) are better for the virtual screening experiments performed on the X-ray SULT1A1 than on the selected MD structures, early enrichment is better on some of the MD structures (up to 30% for ChemBridge™, and up to 50% for Maybridge®). Obtaining earlier enrichments with some MD extracted structures suggests that it is important to take into account the flexibility of the binding site of SULTs. However, further improvements could be achieved for instance by employing induced-fit approximations, development of tuned scoring functions and/or the use of interaction fingerprints. For difficult proteins, combination of docking-scoring, QSAR and network pharmacology or related approaches [104] would seem valuable.

Future trends and conclusions

Current *in silico* ADMET predictions cannot fully replace well-established *in vitro* cell-based approaches or *in vivo* assays but they can provide significant insights. QSAR ADMET models are widely used but are limited within the training set chemical space. Regarding ADMET predictions based on the 3D structures of the relevant proteins, improvements are still required owing to ambiguities in experimental structures and in the biological data used for validation and inaccuracies in several force-field parameters and terms. Obviously, the known limitations of docking-scoring methods are also valid for ADMET proteins. For instance, the difficulty in taking the contribution of water molecules into account accurately [25], and known problems with docking-scoring algorithms, and more specifically with scoring functions. Although it is a common practice to select the docked poses and to rank compounds using simple scoring functions [29,31], insufficient correlation between the docked scores and experimental binding energies are generally observed [29], although some promising results have also been reported [105,106]. In fact, different protocols to improve scoring or to compute the free energy of binding have been investigated and compared to experimental binding data for structurally similar ligands (e.g. for



SULT1A1 structures	ChemBridge™ dataset	Maybridge® dataset
	AUC	
X-Ray	0.45	0.52
MD1	0.40	0.46
MD2	0.39	0.43
MD3	0.40	0.43

FIGURE 3

Enrichment graphs for retrieved actives with VINA docking-scoring performed on the X-ray (in black) and three selected MD structures (in green, blue, violet) for SULT1A1. 100% refers to all screened compounds including the selected 60 actives and the ChemBridge™ decoys **(a)** or the Maybridge® decoys **(b)**. AUC for both datasets are shown in (b).

CYP2C9) [36]. In addition, a recent study suggests that rigorous thermodynamic approaches can be useful to predict binding free energies of structurally diverse ligands for ER [107]. However, although some approaches are relatively efficient for predicting binding energy, they tend to be time consuming and are thus generally applied to a short list of compounds.

It is also worthwhile to note that in the case of enzymatic reactions most of the experimental data (e.g. K_m, K_i) include kinetic components, whereas only a few parameters, poorly documented in the literature (e.g. K_d, K_s), purely reflect ligand binding. Further, ADMET proteins seem to be even more challenging than many other targets because they are often promiscuous, with flexible and sometimes multiple binding sites. The presently implemented flexibility approaches (e.g. multiconformational pocketome [108]) tend to have difficulties in handling this family of proteins. In addition, for many experimental observations useful for clinical extrapolation, the involved proteins are unknown. For instance, intestinal absorption and brain permeability involve many different compound-specific mechanisms. For such *in silico* predictions global QSAR approaches can be more appropriate than protein-based methods.

Despite these limitations, mechanistic understanding of numerous processes through structure-based ADMET investigation should be within reach in the coming years and efforts are presently

undertaken in this direction by several groups as illustrated by the recent study by Obiol-Pardo *et al.* [86]. A combination of molecular docking-scoring and MD simulations can help to improve ADMET properties of drug candidates. We have shown that virtual screening and MD can be used to predict whether a compound is likely to bind to SULTs. Moreover, in some cases it is crucial to have access to protein atomic details for example AGP which exists as a mixture of two or three genetic variants in the plasma of most individuals. Indeed, molecules demonstrate considerable differences in binding the different AGP variants, and distinguishing between the AGP variants can be done properly only by taking into account the slight structural and physicochemical properties of the binding site. ADMET predictions based on combined *in silico* 3D docking and modeling with QSAR methods or pharmacophore screening can help improve the prediction success, as shown in the recently developed VirtualToxLab concept [10,26] or in the MetaSite matching fingerprints of the binding receptor structure and ligands using flexible molecular interaction fields [38].

It is worth mentioning that other *in silico* approaches relevant for ADMET prediction can be noted as 'emerging' because no extensive studies have been reported but they seem to be clearly promising for the future. Quantum mechanical and hybrid quantum mechanical/molecular mechanical (QM/MM) methods are powerful approaches for predicting the rates of reactions in drug metabolism. Early studies started on homology models of CYP [109] and lately the resolved 3D structures for several CYP isoforms enabled the inclusion of the protein environment more precisely. For instance, the mechanism of benzene hydroxylation was investigated in the realistic enzyme environment of the human CYP2C9 by using QM/MM calculations of the whole reaction profile [110]. Recently, MD and QM/MM approaches were used to study the factors influencing the reactivity of compound I for different CYP450 isoforms [111].

Other approaches that could be relevant for ADMET prediction involve proteochemometric modeling [112], in which the ligand and protein descriptors for a series of ligands and a series of proteins are considered. With this approach it might be possible to extrapolate the interactions of unknown compounds on known proteins and vice versa. For instance, proteochemometric modeling was applied on 14 CYP isoforms and 375 inhibitors [37] resulting in the prediction of new potential inhibitors of multiple CYP isoforms with good accuracy. It is possible that such investigations could be combined with docking-scoring protocols.

Definitively, advances in structural genomics [113] [i.e. recently the structure of the human mitochondrial ABC transporter ABCB10 has been resolved by the Structural Genomics Consortium (http://www.sgc.ox.ac.uk)] and major progression of the structure-based methods will open new avenues toward ADMET predictions at the atomic level.

Conflicts of interest

The authors have no conflicts of interest to declare.

Acknowledgements

Support from the French National Research Institute Inserm, the University Paris Diderot and Servier is greatly appreciated. V.Y.M. thanks the doctoral school 'MTCE' of the Universities Paris Descartes and Paris Diderot.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.drudis.2011.10.023.

References

1 Merlot, C. (2010) Computational toxicology – a tool for early safety evaluation. *Drug Discov. Today* 15, 16–22

2 Michielan, L. and Moro, S. (2010) Pharmaceutical perspectives of nonlinear QSAR strategies. *J. Chem. Inf. Model.* 50, 961–978

3 Gleeson, M.P. *et al.* (2011) *In silico* ADME models: a general assessment of their utility in drug discovery applications. *Curr. Top. Med. Chem.* 11, 358–381

4 van de Waterbeemd, H. and Gifford, E. (2003) ADMET *in silico* modelling: towards prediction paradise? *Nat. Rev. Drug Discov.* 2, 192–204

5 Dobson, P.D. *et al.* (2009) 'Metabolite-likeness' as a criterion in the design and selection of pharmaceutical drug libraries. *Drug Discov. Today* 14, 31–40

6 Lipinski, C.A. *et al.* (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 46, 3–26

7 Lagorce, D. *et al.* (2008) FAF-Drugs2: free ADME/tox filtering tool to assist drug discovery and chemical biology projects. *BMC Bioinformatics* 9, 396

8 Lagorce, D. *et al.* (2011) The FAF-Drugs2 server: a multistep engine to prepare electronic chemical compound collections. *Bioinformatics* 27, 2018–2020

9 Bhogal, N. *et al.* (2005) Toxicity testing: creating a revolution based on new technologies. *Trends Biotechnol.* 23, 299–307

10 Vedani, A. and Smiesko, M. (2009) *In silico* toxicology in drug discovery – concepts based on three-dimensional models. *Altern. Lab. Anim.* 37, 477–496

11 de Groot, M.J. *et al.* (1996) A three-dimensional protein model for human cytochrome P450 2D6 based on the crystal structures of P450 101, P450 102, and P450 108. *Chem. Res. Toxicol.* 9, 1079–1091

12 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

13 Stoll, F. *et al.* (2011) Utility of protein structures in overcoming ADMET-related issues of drug-like compounds. *Drug Discov. Today* 16, 530–538

14 Sun, H. and Scott, D.O. (2010) Structure-based drug metabolism predictions for drug design. *Chem. Biol. Drug Des.* 75, 3–17

15 Tetko, I.V. *et al.* (2006) Can we estimate the accuracy of ADME-Tox predictions? *Drug Discov. Today* 11, 700–707

16 Cavasotto, C.N. and Singh, N. (2008) Docking and high throughput docking: successes and the challenge of protein flexibility. *Curr. Comput. Aid. Drug Des.* 4, 221–234

17 Sperandio, O. *et al.* (2010) How to choose relevant multiple receptor conformations for virtual screening: a test case of Cdk2 and normal mode analysis. *Eur. Biophys. J.* 39, 1365–1372

18 Isvoran, A. *et al.* (2011) Exploring NMR ensembles of calcium binding proteins: perspectives to design inhibitors of protein–protein interactions. *BMC Struct. Biol.* 11, 24

19 Bojarova, P. and Williams, S.J. (2008) Sulfotransferases, sulfatases and formylglycine-generating enzymes: a sulfation fascination. *Curr. Opin Chem. Biol.* 12, 573–581

20 Shimada, T. (2006) Xenobiotic-metabolizing enzymes involved in activation and detoxification of carcinogenic polycyclic aromatic hydrocarbons. *Drug Metab. Pharmacokinet.* 21, 257–276

21 Bode, C. (2010) The nasty surprise of a complex drug–drug interaction. *Drug Discov. Today* 15, 391–395

22 U.S. Department of Health and Human Services, F.D.A., (1999) *Guidance for Industry, Drug Metabolism/Drug Interaction Studies in the Drug Development Process: Studies In Vitro*. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm072104.pdf

23 Guengerich, F.P. (2008) Cytochrome p450 and chemical toxicology. *Chem. Res. Toxicol.* 21, 70–83

24 Clark, S.E. and Jones, B.C. (2002) Human cytochromes P450 and their role in metabolism-based drug–drug interactions. In *Drug–Drug Interactions* (Rodrigues, A.D., ed.), pp. 55–88, Marcel Dekker

25 Zhou, D. *et al.* (2006) Comparison of methods for the prediction of the metabolic sites for CYP3A4-mediated metabolic reactions. *Drug Metab. Dispos.* 34, 976–983

26 Rossato, G. *et al.* (2010) Probing small-molecule binding to cytochrome P450 2D6 and 2C9: an in silico protocol for generating toxicity alerts. *ChemMedChem* 5, 2088–2101

27 Park, S.J. *et al.* (2010) Improved docking, screening and selectivity prediction for small molecule nuclear receptor modulators using conformational ensembles. *J. Comput. Aid. Mol. Des.* 24, 459–471

28 Rowland, P. *et al.* (2006) Crystal structure of human cytochrome P450 2D6. *J. Biol. Chem.* 281, 7614–7622

29 Kemp, C.A. *et al.* (2004) Validation of model of cytochrome P450 2D6: an *in silico* tool for predicting metabolism and inhibition. *J. Med. Chem.* 47, 5340–5346

30 Bazeley, P.S. *et al.* (2006) Synergistic use of compound properties and docking scores in neural network modeling of CYP2D6 binding: predicting affinity and conformational sampling. *J. Chem. Inf. Model.* 46, 2698–2708

31 de Graaf, C. *et al.* (2006) Catalytic site prediction and virtual screening of cytochrome P450 2D6 substrates by consideration of water and rescoring in automated docking. *J. Med. Chem.* 49, 2417–2430

32 Santos, R. *et al.* (2010) Role of water in molecular docking simulations of cytochrome P450 2D6. *J. Chem. Inf. Model.* 50, 146–154

33 Hritz, J. *et al.* (2008) Impact of plasticity and flexibility on docking results for cytochrome P450 2D6: a combined approach of molecular dynamics and ligand docking. *J. Med. Chem.* 51, 7469–7477

34 Unwalla, R.J. *et al.* (2010) Using a homology model of cytochrome P450 2D6 to predict substrate site of metabolism. *J. Comput. Aid. Mol. Des.* 24, 237–256

35 Shi, R. *et al.* (2011) Exploration of the binding of proton pump inhibitors to human P450 2C9 based on docking and molecular dynamics simulation. *J. Mol. Model.* 17, 1941–1951

36 Stjernschantz, E. and Oostenbrink, C. (2010) Improved ligand-protein binding affinity predictions using multiple binding modes. *Biophys. J.* 98, 2682–2691

37 Kontijevskis, A. *et al.* (2008) Generalized proteochemometric model of multiple cytochrome p450 enzymes and their inhibitors. *J. Chem. Inf. Model.* 48, 1840–1850

38 Cruciani, G. *et al.* (2005) MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist. *J. Med. Chem.* 48, 6970–6979

39 Daly, A.K. (2003) Pharmacogenetics of the major polymorphic metabolizing enzymes. *Fundam. Clin. Pharmacol.* 17, 27–41

40 Williams, J.A. *et al.* (2004) Drug-drug interactions for UDP-glucuronosyltransferase substrates: a pharmacokinetic explanation for typically observed low exposure (AUCi/AUC) ratios. *Drug Metab. Dispos.* 32, 1201–1208

41 Smith, P.A. *et al.* (2004) Towards integrated ADME prediction: past, present and future directions for modelling metabolism by UDP-glucuronosyltransferases. *J. Mol. Graph Model.* 22, 507–517

42 Cucurull-Sanchez, L. (2010) Successful identification of key chemical structure modifications that lead to improved ADME profiles. *J. Comput. Aid. Mol. Des.* 24, 449–458

43 Miley, M.J. *et al.* (2007) Crystal structure of the cofactor-binding domain of the human phase II drug-metabolism enzyme UDP-glucuronosyltransferase 2B7. *J. Mol. Biol.* 369, 498–511

44 Shao, H. *et al.* (2005) Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from Medicago truncatula. *Plant Cell* 17, 3141–3154

45 Wu, B. *et al.* (2011) Three-dimensional quantitative structure–activity relationship studies on UGT1A9-mediated 3-O-glucuronidation of natural flavonols using a pharmacophore-based comparative molecular field analysis model. *J. Pharmacol. Exp. Ther.* 336, 403–413

46 Krasowski, M.D. *et al.* (2011) Evolution of promiscuous nuclear hormone receptors: LXR, FXR, VDR, PXR, and CAR. *Mol. Cell. Endocrinol.* 334, 39–48

47 Savas, U. *et al.* (1999) Molecular mechanisms of cytochrome P-450 induction by xenobiotics: an expanded role for nuclear hormone receptors. *Mol. Pharmacol.* 56, 851–857

48 Ngan, C.H. *et al.* (2009) The structural basis of pregnane X receptor binding promiscuity. *Biochemistry* 48, 11572–11581

49 Ekins, S. *et al.* (2009) Challenges predicting ligand–receptor interactions of promiscuous proteins: the nuclear receptor PXR. *PLoS Comput. Biol.* 5, e1000594

50 Liu, Y.H. *et al.* (2011) Regulation of human pregnane X receptor and its target gene cytochrome P450 3A4 by Chinese herbal compounds and a molecular docking study. *Xenobiotica* 41, 259–280

51 Baes, M. *et al.* (1994) A new orphan member of the nuclear hormone receptor superfamily that interacts with a subset of retinoic acid response elements. *Mol. Cell. Biol.* 14, 1544–1552

52 Xu, R.X. *et al.* (2004) A structural basis for constitutive activity in the human CAR/RXRalpha heterodimer. *Mol. Cell* 16, 919–928

53 Windshugel, B. *et al.* (2005) Molecular dynamics simulations of the human CAR ligand-binding domain: deciphering the molecular basis for constitutive activity. *J. Mol. Model.* 11, 69–79

54 Küblbeck, J. *et al.* (2008) Discovery of substituted sulfonamides and thiazolidin-4-one derivatives as agonists of human constitutive androstane receptor. *Biochem. Pharmacol.* 76, 1288–1297

55 DeKeyser, J.G. *et al.* (2011) Selective phthalate activation of naturally occurring human constitutive androstane receptor splice variants and the pregnane x receptor. *Toxicol. Sci.* 120, 381–391

56 Vedani, A. *et al.* (2005) Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. *J. Med. Chem.* 48, 3700–3703

57 Lill, M.A. *et al.* (2005) Impact of induced fit on ligand binding to the androgen receptor: a multidimensional QSAR study to predict endocrine-disrupting effects of environmental chemicals. *J. Med. Chem.* 48, 5666–5674

58 Bisson, W.H. *et al.* (2007) Discovery of antiandrogen activity of nonsteroidal scaffolds of marketed drugs. *Proc. Natl. Acad. Sci. U. S. A.* 104, 11927–11932

59 Carter, D.C. and Ho, J.X. (1994) Structure of serum albumin. *Adv. Protein Chem.* 45, 153–203

60 He, X.M. and Carter, D.C. (1992) Atomic structure and chemistry of human serum albumin. *Nature* 358, 209–215

61 Ghuman, J. *et al.* (2005) Structural basis of the drug-binding specificity of human serum albumin. *J. Mol. Biol.* 353, 38–52

62 Otagiri, M. (2005) A molecular functional study on the interactions of drugs with plasma proteins. *Drug Metab. Pharmacokinet.* 20, 309–323

63 Sheppard, G.S. *et al.* (2006) Discovery and optimization of anthranilic acid sulfonamides as inhibitors of methionine aminopeptidase-2: a structural basis for the reduction of albumin binding. *J. Med. Chem.* 49, 3832–3849

64 Wendt, M.D. *et al.* (2006) Discovery and structure–activity relationship of antagonists of B-cell lymphoma 2 family proteins with chemopotentiation activity *in vitro* and *in vivo*. *J. Med. Chem.* 49, 1165–1181

65 Varshney, A. *et al.* (2011) Elimination of endogenous toxin, creatinine from blood plasma depends on albumin conformation: site specific uremic toxicity & impaired drug binding. *PLoS ONE* 6, e17230

66 Liu, H. *et al.* (2010) Binding modes of flavones to human serum albumin: insights from experimental and computational studies. *J. Phys. Chem. B* 114, 12938–12947

67 Salvalaglio, M. *et al.* (2010) Determination of energies and sites of binding of PFOA and PFOS to human serum albumin. *J. Phys. Chem. B* 114, 14860–14874

68 Deeb, O. *et al.* (2010) Exploration of human serum albumin binding sites by docking and molecular dynamics flexible ligand–protein interactions. *Biopolymers* 93, 161–170

69 Zsila, F. *et al.* (2011) Evaluation of drug–human serum albumin binding interactions with support vector machine aided online automated docking. *Bioinformatics* 27, 1806–1813

70 Nishi, K. *et al.* (2011) Structural insights into differences in drug-binding selectivity between two forms of human alpha1-acid glycoprotein genetic variants, the A and F1*S forms. *J. Biol. Chem.* 286, 14427–14434

71 Herve, F. *et al.* (1998) Ligand specificity of the genetic variants of human alpha1-acid glycoprotein: generation of a three-dimensional quantitative structure–activity relationship model for drug binding to the A variant. *Mol. Pharmacol.* 54, 129–138

72 Kopecky, V., Jr *et al.* (2003) Structure of human alpha1-acid glycoprotein and its high-affinity binding site. *Biochem. Biophys. Res. Commun.* 300, 41–46

73 Schonfeld, D.L. *et al.* (2008) The 1.8-Å crystal structure of alpha1-acid glycoprotein (Orosomucoid) solved by UV RIP reveals the broad drug-binding activity of this human plasma lipocalin. *J. Mol. Biol.* 384, 393–405

74 European Medecines Agency. ICH S7B – The nonclinical Evaluation of the Poetential for delayed Ventricular Repolarisation (QT interval prolongation) by human pharmaceutics. CPMP/ICH/423/02 November 2005, http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002841.pdf

75 Tabrizchi, R. (2010) Molecular mechanisms of adverse drug reactions in cardiac tissue. *Handb. Exp. Pharmacol.* 196, 77–109

76 Netzer, R. *et al.* (2003) Critical appraisal of available models for the detection of proarrhythmic side-effects of compounds. In *Atrial Fibrillation: New Therapeutic Concepts* (Papp, J.G. *et al.* eds), IOS Press pp. 103–114

77 Aronov, A.M. (2005) Predictive *in silico* modeling for hERG channel blockers. *Drug Discov. Today* 10, 149–155

78 Mitcheson, J.S. *et al.* (2000) A structural basis for drug-induced long QT syndrome. *Proc. Natl Acad. Sci. U. S. A.* 97, 12329–12333

79 Pearlstein, R. *et al.* (2003) Understanding the structure–activity relationship of the human ether-a-go-go-related gene cardiac K^+ channel. A model for bad behavior. *J. Med. Chem.* 46, 2017–2022

80 Osterberg, F. and Aqvist, J. (2005) Exploring blocker binding to a homology model of the open hERG K^+ channel using docking and molecular dynamics methods. *FEBS Lett.* 579, 2939–2944

81 Inanobe, A. *et al.* (2008) *In silico* prediction of the chemical block of human ether-a-go-go-related gene (hERG) K^+ current. *J. Physiol. Sci.* 58, 459–470

82 Fernandez, D. *et al.* (2004) Physicochemical features of the HERG channel drug binding site. *J. Biol. Chem.* 279, 10120–10127

83 Stansfeld, P.J. *et al.* (2007) Drug block of the hERG potassium channel: insight from modeling. *Proteins* 68, 568–580

84 Durdagi, S. *et al.* (2011) Combined receptor and ligand-based approach to the universal pharmacophore model development for studies of drug blockade to the hERG1 pore domain. *J. Chem. Inf. Model.* 51, 463–474

85 Taboureau, O. and Jørgensen, F.S. (2011) *In silico* predictions of hERG channel blockers in drug discovery: from ligand-based and target-based approaches to systems chemical biology. *Comb. Chem. High Throughput Screen.* 14, 375–387

86 Obiol-Pardo, C. *et al.* (2011) A multiscale simulation system for the prediction of drug-induced cardiotoxicity. *J. Chem. Inf. Model.* 51, 483–492

87 Pajeva, I. and Wiese, M. (2011) Application of in silico methods to study ABC transporters involved in multidrug resistance. In *In Silico Lead Discovery* (Miteva, M.A., ed.), pp. 144–162, Bentham Science Publishers

88 Williamson, G. *et al.* (2007) Interaction of positional isomers of quercetin glucuronides with the transporter ABCC2 (Cmoat, MRP2). *Drug Metab. Dispos.* 35, 1262–1268

89 Higgins, C.F. and Linton, K.J. (2004) The ATP switch model for ABC transporters. *Nat. Struct. Mol. Biol.* 11, 918–926

90 Safa, A.R. (2004) Identification and characterization of the binding sites of P-glycoprotein for multidrug resistance-related drugs and modulators. *Curr. Med. Chem. Anticancer Agents* 4, 1–17

91 Seigneuret, M. and Garnier-Suillerot, A. (2003) A structural model for the open conformation of the mdr1 P-glycoprotein based on the MsbA crystal structure. *J. Biol. Chem.* 278, 30115–30124

92 O'Mara, M.L. and Tieleman, D.P. (2007) P-glycoprotein models of the apo and ATP-bound states based on homology with Sav1866 and MalK. *FEBS Lett.* 581, 4217–4222

93 Becker, J.P. *et al.* (2009) Molecular models of human P-glycoprotein in two different catalytic states. *BMC Struct. Biol.* 9, 3

94 Aller, S.G. *et al.* (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science* 323, 1718–1722

95 Pajeva, I.K. *et al.* (2009) Comparison of the inward- and outward-open homology models and ligand binding of human P-glycoprotein. *FEBS J.* 276, 7016–7026

96 Klepsch, F. *et al.* (2011) Exhaustive sampling of docking poses reveals binding hypotheses for propafenone type inhibitors of P-glycoprotein. *PLoS Comput. Biol.* 7, e1002036

97 Dantzig, A.H. *et al.* (1999) Selectivity of the multidrug resistance modulator, LY335979, for P-glycoprotein and effect on cytochrome P-450 activities. *J. Pharmacol. Exp. Ther.* 290, 854–862

98 Achira, M. *et al.* (1999) Comparative studies to determine the selective inhibitors for P-glycoprotein and cytochrome P4503A4. *AAPS Pharm. Sci.* 1, E18

99 Allali-Hassani, A. *et al.* (2007) Structural and chemical profiling of the human cytosolic sulfotransferases. *PLoS Biol.* 5, e97

100 Stjernschantz, E. *et al.* (2010) Comparison of murine and human estrogen sulfotransferase inhibition *in vitro* and *in silico* – implications for differences in activity, subunit dimerization and substrate inhibition. *Mol. Cell. Endocrinol.* 317, 127–140

101 Campagna-Slater, V. and Schapira, M. (2009) Evaluation of virtual screening as a tool for chemical genetic applications. *J. Chem. Inf. Model.* 49, 2082–2091

102 Brooks, B.R. *et al.* (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4, 187–217

103 Trott, O. and Olson, A.J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461

104 Achenbach, J. *et al.* (2011) Computational tools for polypharmacology and repurposing. *Future Med. Chem.* 3, 961–968

105 Ito, Y. *et al.* (2008) Analysis of CYP2D6 substrate interactions by computational methods. *J. Mol. Graph Model.* 26, 947–956

106 Afzelius, L. *et al.* (2007) State-of-the-art tools for computational site of metabolism predictions: comparative analysis, mechanistical insights, and future applications. *Drug Metab. Rev.* 39, 61–86

107 Michel, J. and Essex, J.W. (2008) Hit identification and binding mode predictions by rigorous free energy simulations. *J. Med. Chem.* 51, 6654–6664

108 Abagyan, R. and Kufareva, I. (2009) The flexible pocketome engine for structural chemogenomics. In *Chemogenomics, Methods in Molecular Biology*, (Vol. 575) (Jacoby, E., ed.), pp. 249–279, Springer

109 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

110 Bathelt, C.M. *et al.* (2008) QM/MM modeling of benzene hydroxylation in human cytochrome P450 2C9. *J. Phys. Chem. A* 112, 13149–13156

111 Lonsdale, R. *et al.* (2011) Does compound I vary significantly between isoforms of cytochrome P450? *J. Am. Chem. Soc.* 133, 15464–15474

112 van Westen, G.J.P. *et al.* (2011) Proteochemometric modeling as a tool to design selective compounds and for extrapolating to novel targets. *Med. Chem. Commun.* 2, 16–30

113 Lee, W.H. *et al.* (2009) SGC – structural biology and human health: a new approach to publishing structural biology results. *PLoS ONE* 4, e7675