Wiley Interdisciplinary Reviews: Computational Molecular Science / Volume 7, Issue 6

Advanced Review | Full Access

Recent advances in the prediction of non-CYP450-mediated drug metabolism

Vaibhav A. Dixit, ... See all authors ›

First published: 02 June 2017
https://doi.org/10.1002/wcms.1323
Cited by: 1

Conflict of interest: The authors have no conflicts of interest to declare.

Abstract

Computational models of drug metabolism prediction have focused mainly on cytochrome P450 enzymes, because drug–drug interactions, reactive metabolite formation, hepatotoxicity, idiosyncratic adverse drug interactions, and/or loss of efficacy of many drugs were the results of interactions with CYP450s. Metabolic regioselectivity and isoform specificity prediction models for CYP450-catalyzed reactions have reached approximately 95% accuracy. Thus, a new drug candidate is less likely to show unexpected metabolic profile due to metabolism via CYP450 pathways. For such candidates, secondary metabolic Phase I and II enzymes are likely to play an expected (or unexpected) role in drug metabolism. The importance of flavin monooxygenases (FMOs), aldehyde and alcohol dehydrogenase, monoamine oxidase from the Phase I and UDP-glucuronosyltransferase (UGT), sulfotransferase, glutathione *S*-transferase, and

methyltransferase from Phase II has increased and United States Food and Drug Administration guidelines on NDA have specific recommendations for *in vitro* and *in vivo* testing against these enzymes. Thus, there is an urgent requirement of reliable predictive models for drug metabolism catalyzed by these enzymes. In this review, we have classified drug metabolism prediction models (site of metabolism, isoform specificity, and kinetic parameter) for these enzymes into Phase I and II. When such models are unavailable, we discuss the Quantitative Structure Activity Relationship (QSAR), pharmacophore, docking, dynamics, and reactivity studies performed for the prediction of substrates and inhibitors. Recently published models for FMO and UGT are discussed. The need for comprehensive, widely applicable, sequential primary and secondary metabolite prediction is highlighted. Potential difficulties and future prospectives in the development of such models are discussed. *WIREs Comput Mol Sci* 2017, 7:e1323. doi: 10.1002/wcms.1323

This article is categorized under:

- Structure and Mechanism > Reaction Mechanisms and Catalysis
- Computer and Information Science > Chemoinformatics
- Software > Molecular Modeling

INTRODUCTION

Cytochrome P450s (CYP450s) and non-CYP450 enzyme systems are involved in the homeostasis and xenobiotic (drug) metabolism. Ten CYP450 isoforms are known to be involved in drug metabolism, namely CYP1A1, CYP1A2, CYP2B6, CYP2E1, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP3A4, and CYP3A5.[1] These CYP450s and UDP-glucuronosyltransferases (UGTs) contribute to the Phase I and Phase II metabolism, respectively, catalyze majority of the metabolic reactions, and may lead to drug–drug interactions, reactive metabolite (RM) formation, hepatotoxicity, and associated drug withdrawals (Figure 1).[2] Thus scientists, involved in drug discovery efforts for the last two decades, have designed compounds to control the CYP450-mediated drug metabolism. This resulted in the development of small-molecule drugs, which either escape the CYP450 and downstream enzyme-catalyzed metabolic pathways or are metabolized by a selected set of CYP450s.[3] This has indirectly led to an increase in the demand for the assessment of non-CYP450-mediated Phase I and non-UGT-mediated Phase II drug metabolism.[1] Therefore, screening of metabolism by non-CYP450 Phase I metabolic enzymes, such as flavin monooxygenases (FMOs), monoamine oxidase (MAO), alcohol dehydrogenase (ADH), aldehyde dehydrogenase (ALDH), and hydrolyases, has become a routine drug discovery step.[4] While due to the formation of Phase I metabolites from these non-CYP450 metabolic systems and the design of novel

prodrugs, understanding the role of the non-UGT Phase II enzymes such as glutathione S-transferase (GST) and sulfotransferases (STs) in drug metabolism has also gained importance.[4]

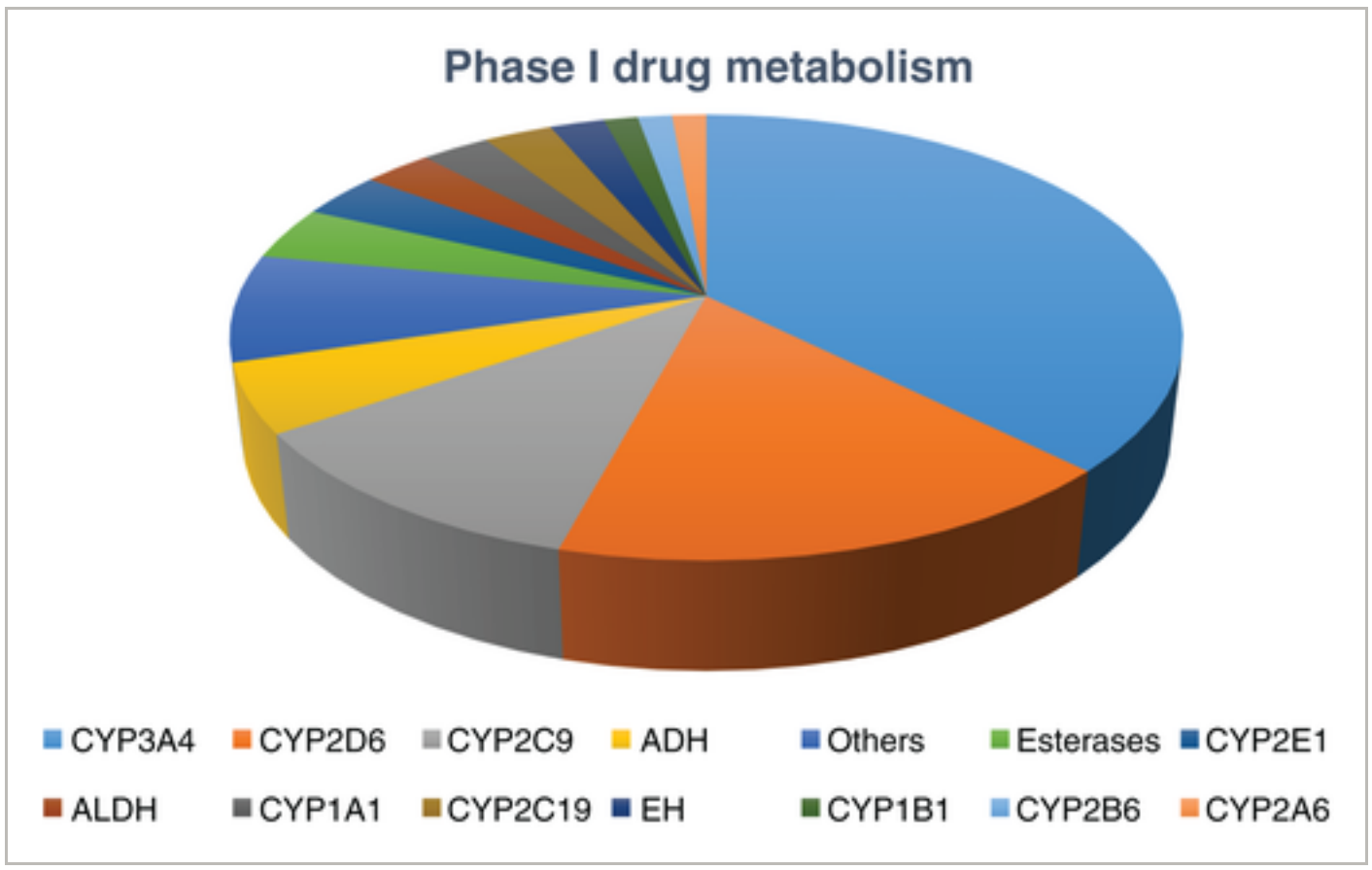

Figure 1

Open in figure viewer | PowerPoint

Relative contribution of different Phase I metabolic enzymes toward drug metabolism. Flavin monooxygenases, quinone oxidoreductases (NQ0), and dihydropyrimidine dehydrogenase (DPD) are included in others.

Caption

Cerny has recently reviewed the contribution of CYP450 and non-CYP450 enzymes in the metabolism of 125 drugs approved by the United States Food and Drug Administration (US-FDA) during 2006–2015.[3] Non-CYP450 enzymes were found to consistently metabolize approximately 30% of the drugs in the dataset during this period. The percentage of drugs metabolized by CYP450 showed a modest change (43–58%) during these two 5-year periods, i.e., 2006–2010 and 2011–2015. For non-CYP450-mediated metabolism, the percentages of drugs metabolized were similar (27 and 31%) for these 5-year periods. It is noteworthy that during this time, the percentage of drugs for which no major metabolite (>10% of the total dose) is found decreased from 28 to 11%. This may reflect the better understanding and control that medicinal chemists and Drug Metabolism and Pharmacokinetics (DMPK) scientist have on the metabolic fate of drug-like molecules. Figures 1 and 2, and Table 1

show the relative contributions of various Phase I and Phase II metabolic enzymes toward drug metabolism. Despite the fact that non-CYP450 enzyme metabolize over 30% of the approved drugs, the scientific community has paid considerably less attention to these classes of oxidative enzyme. A search in 'Google Scholar' for keywords for the various Phase I and Phase II enzymes and 'drug metabolism' shows that scientific interest in the CYP450 research steadily increased after the discovery of relationships between CYP450s and drug metabolism (Table 2), whereas, other Phase I metabolic enzymes attracted only moderate to low attention. A similar situation is apparent from the data for Phase II metabolic enzymes. More than 10,000 research articles, reviews, and patents exist which involve the concept of drug metabolism and study the Phase II metabolic enzyme UGT and methyltransferase (MT). But the other Phase II metabolic enzymes, in comparison, have not been studied to a similar extent.

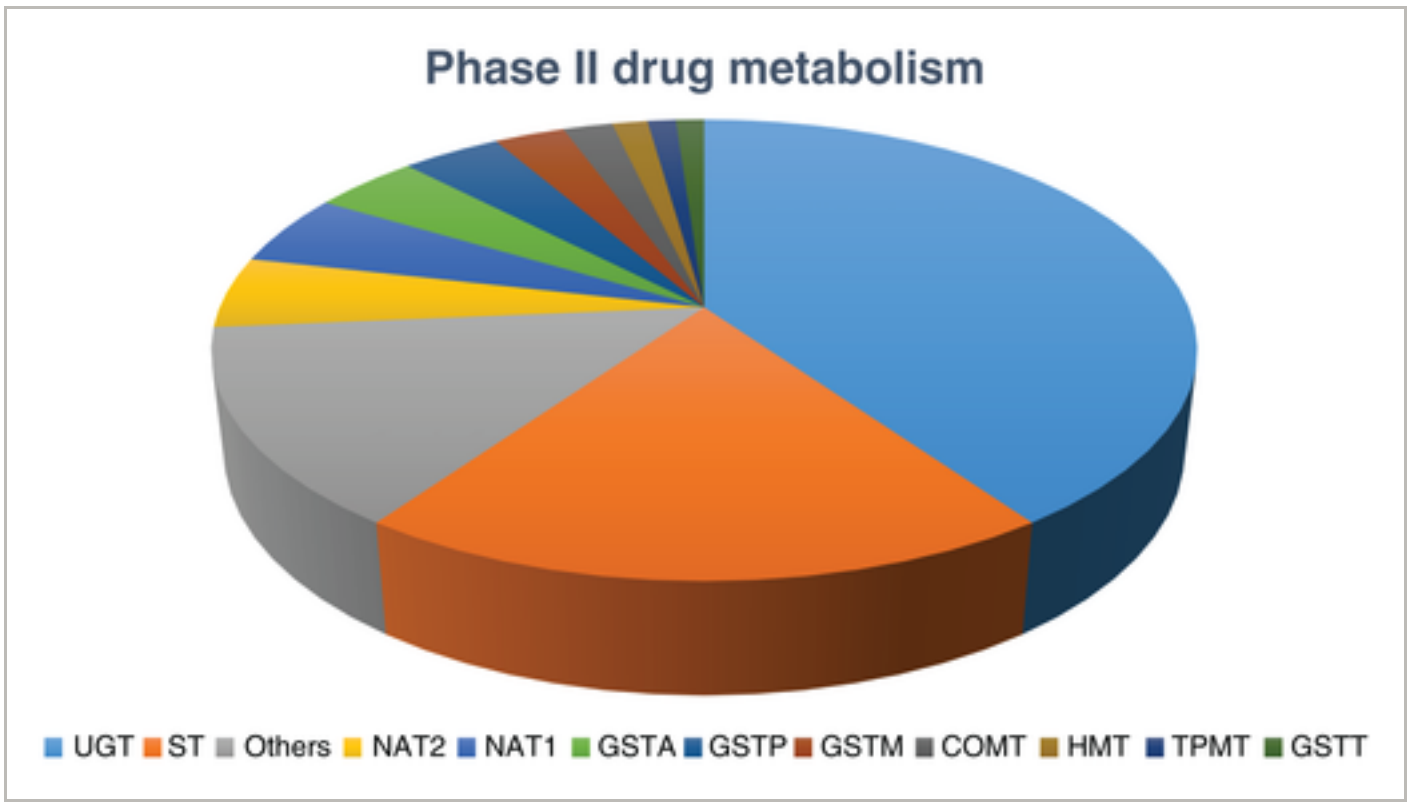

Figure 2

Open in figure viewer | PowerPoint

Relative contribution of different Phase II metabolic enzymes toward drug metabolism. UDP-glucuronosyltransferase (UGT), sulfotransferase (SULT), *N*-acetyltransferase-1 (NAT1), *N*-acetyltransferase -2 (NAT2), glutathione *S*-transferase (GST), GST-α (GSTA), GST-π (GSTP), GST-μ (GSTM), GST-θ (GSTT), catechol *O*-methyltransferase (COMT), histamine *N*-methyltransferase (HMT), thiopurine *S*-methyltransferase (TPMT).

Caption ⌄

Table 1. Percentage Contribution of Various Phase I and Phase II Metabolic Enzymes in Drug

Metabolism

Phase I	Percentage Contribution	Phase II	Percentage Contribution
CYP3A4	37.50	UGT	40.28
CYP2D6	16.67	SULT	19.44
CYP2C9	11.11	Others	13.89
ADH	5.00	NAT2	5.00
others	7.78	NAT1	5.00
Esterases	3.89	GSTA	4.17
CYP2E1	3.33	GSTP	3.89
ALDH	2.78	GSTM	2.78
CYP1A1	2.78	COMT	1.94
CYP2C19	2.78	HMT	1.39
EH	2.22	TPMT	1.11
CYP1B1	1.39	GSTT	1.11
CYP2B6	1.39		
CYP2A6	1.39		

See caption of Figures 1 and 2 for full form of common enzyme names.

Table 2. Number of Articles Published During Different 5-Year Periods that Contain the Concept of 'Drug Metabolism' and Various Important Phase I and Phase II Metabolic Enzymes

Year	ALDH	FMO	CYP450	AD	AO	EH	UGT	SULT	GST	MT
1941–1945	0	0	0	0	0	0	0	0	0	0

1946–1950	0	0	0	0	1	0	0	1	0	0
1951–1955	1	0	2	2	4	0	0	0	0	0
1956–1960	2	0	2	3	4	0	0	0	0	0
1961–1965	4	1	6	9	27	0	0	0	0	5
1966–1970	14	1	82	39	40	2	6	2	1	18
1971–1975	42	0	317	149	74	25	28	17	12	74
1976–1980	101	2	695	199	135	79	103	63	69	123
1981–1985	171	6	1040	301	138	510	153	172	179	178
1986–1990	238	44	1700	360	149	735	280	252	421	230
1991–1995	314	36	3960	370	138	811	440	412	489	392
1996–2000	372	103	7220	433	119	954	651	507	606	545
2001–2005	713	103	13,500	849	234	1290	1430	1050	730	1700
2006–2010	997	195	16,600	1080	248	1550	2720	1710	1060	2810
2011–2015	1500	234	17,300	1510	288	1540	3620	2020	1180	3770
2016–current	306	84	5240	286	46	273	657	460	193	857
Total	4775	809	67,664	5590	1645	7769	10,088	6666	4940	10,702
Percentage of total across all enzyme	3.96	0.67	56.08	4.63	1.36	6.44	8.36	5.53	4.09	8.87

The data were collected from Google scholar on November 16, 2016.
AD, alcohol dehydrogenase; ALDH, aldehyde dehydrogenase; AO, amine oxidase; CYP450, cytochrome P450; EH, epoxide hydrolyase; FMO, flavin monooxygenase; GST, glutathione *S*-transferase; MT, methyltransferase; SULT, sulfotransferase; UGT, UDP-glucuronosyl transferase.

There are multiple reasons for this bias toward CYP450 and UTG metabolic enzymes. Since the discovery of CYP450 monooxygenases, a great deal of research has focused on isolation, purification, characterization, and applications of these enzymes. Initially, microbial (especially bacterial) CYP450s were studied by various experimental methods, such as X-ray

crystallography, electron spin resonance spectroscopy, and computational quantum chemical (QC) methods. In the last two decades, many crystal structures have become available for human CYP450s, facilitating studies, such as homology modeling, structure-based QSAR, pharmacophore modeling, drug metabolism prediction, including site of metabolism (SOM), and isoform specificity prediction. As a result, an almost complete understanding of the catalytic cycle, identity of the rate-determining step, and factors affecting the rate of reaction, enzyme–substrate binding affinity, and so on, exists. The total number of articles published on CYP450 (67,664) reflects this fact. Similarly UGT, discovered a decade later, received significant interest (total articles = 10,088, ~8.36%), albeit less than CYP450s.

In silico models based on *in vivo* data can be developed, but considering the time and cost investment associated with the development of such models, applications to a set of chemically diverse drug-like molecules are very difficult. Thus, development of generalized models with greater accuracy (using *in vivo* data) is currently not feasible.5 The aim of the computational prediction of drug metabolism is not a definite prediction. Instead, *in silico* models are generally used to guide the selection of drug-like molecules with high probability of having desirable pharmacokinetic and metabolic outcomes. Once a candidate (or a small set of molecules) is chosen using *in silico* methods, more accurate and expensive *in vitro* drug metabolism predictions are taken up to estimate the human *in vivo* outcomes.

The computational prediction of drug metabolism generally falls into three main categories as follows: (1) prediction of pharmacokinetic parameters, e.g., hepatic clearance Cl_H and metabolic rate constants; (2) isoform specificity prediction, i.e., prediction of isoforms involved in the metabolism; and (3) prediction of potential soft-spots, i.e., SOM in the given xenobiotic/endobiotic. As mentioned above, xenobiotics and drug metabolism are mostly classified into Phases I and II. This classification is mainly based on the type of enzymatic reaction that the molecule undergoes, mainly oxidation in Phase I and conjugation reactions in Phase II. The use of term ‘SOM’ for all these reactions, although common in the literature, may lead to confusion especially when one is describing more than one step in drug metabolism. Thus, in this review, we have introduced a comprehensive classification of metabolic sites for both Phase I and Phase II based on the fact that most Phase I reactions involve either oxidation, reduction, or hydrolysis, whereas Phase II reactions involve mostly conjugation reactions. Thus, the SOMs can be further subclassified as shown in Figure 3. Once universally accepted, the use of this terminology will help in avoiding the cumbersome description of different SOMs and any confusion about the identity of the reaction. It will also help the scientists and programmers in quickly classifying the data into different categories for descriptor, statistical calculations, and predictions. At present, a handful of

SOM prediction methodologies and softwares focus only on a limited number of SOM types which mostly form from the CYP450 and/or UGT-catalyzed metabolic reactions.

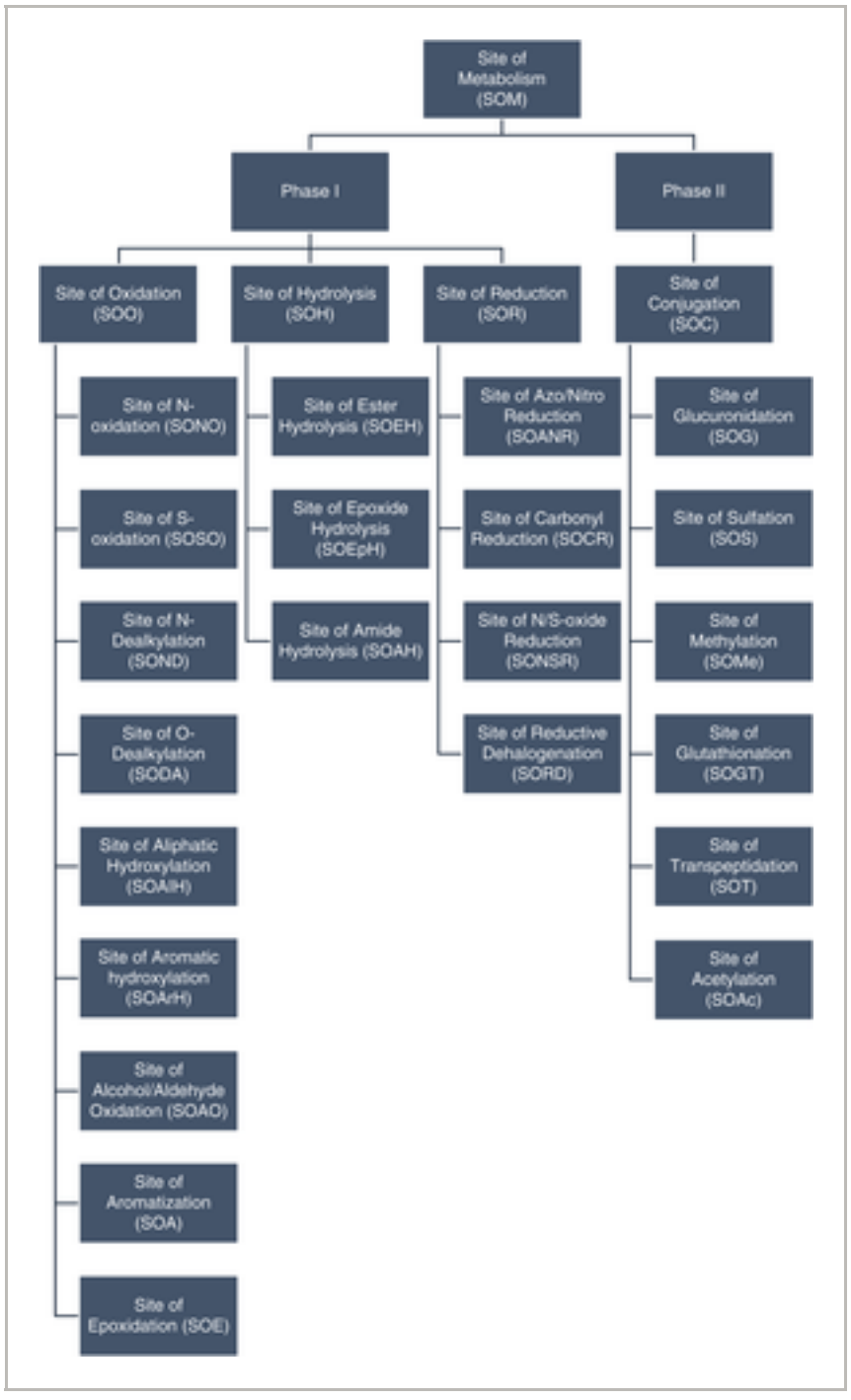

Figure 3

Open in figure viewer | PowerPoint

A comprehensive and chemically intuitive classification of site(s) of metabolism (SOMs) based on the phase and type of metabolic reactions.

Caption ⌄

Some of these methodologies/softwares (e.g., XenoSite, MetaSite, SMART-Cyp, RS-predictor, and ADME-predictor) can predict the site of N-oxidation (SONO), site of S-oxidation (SOSO), site of N-dealkylation (SOND), site of O-dealkylation (SODA), site of epoxidation (SOE),6 site of aliphatic hydroxylation (SOAlH), site of aromatic hydroxylation (SOArH), and site of glucuronidation (SOG) (e.g., XenoSite-UGT). To the best of our knowledge, methodologies for

the prediction of other SOM categories such as site of alcohol/aldehyde oxidation (SOAO), all the sites of hydrolysis (SOHs), site of reduction (SORs) from Phase I and site of sulfonation (SOS), site of glutathionation (SOGT), site of methylation (SOMe), site of acetylation (SOAc), site of transpeptidation (SOT) from Phase II are either rare or nonexistent. As discussed above, contribution to drug metabolism by non-CYP450 enzymatic systems, although less than CYP450s, is important to account for many of idiosyncratic drug–drug interactions, and a complete coverage of all the possible SOMs (shown in Figure 3) is thus necessary. This fact has partially been appreciated in the recent studies by Swamidass group, where they have developed models for prediction of potential of different double bonds to get epoxidized, i.e., prediction of SOE.6 But the prediction of other SOM types has remained largely ignored despite the developments in the predictions of SOMs related to CYP450, UGT, and recently FMO3-mediated metabolisms.

In addition, very few prediction methodologies have the feature to make sequential metabolic predictions to enumerate all the possible primary, secondary, and higher order metabolites, i.e., metabolites of metabolites till the metabolic potential of the molecule is exhausted. In an ideal situation, one would like to associate a probability of occurrence with all the enumerations resulting from such an iterative metabolite generation procedure to eliminate improbable and less probable metabolites. Systematic Generation of potential Metabolites (SyGMa)7 and Meteor Nexus8 are capable of doing this to some extent, but suffer from lower accuracy, do not cover all the possible SOM mentioned above and may have limited applicability outside the molecular space of the training dataset.

An integral part of a drug discovery programs is the optimization of properties of drug candidates within the design space.9 This optimization process is a compromise between the ideally expected profile and those which can be realistically achieved. Such an ideal and realistic profile varies widely as per the therapeutic areas of interest. It is thus, idealistic to design molecules with zero metabolic potential such that they enter the human body, show pharmacological effect, and get excreted unchanged. However, in a realistic drug discovery setting, metabolism also serves as an additional pathway to terminate the pharmacological action of the drug. Thus, the main aim is to predict known and control (by design) the metabolic profile of the most promising candidate drugs. Medicinal chemists can additionally ask questions like, metabolism by which Phase I or Phase II enzymes will give me a realistic and safe metabolic profile? What structural changes do I need to perform to increase recognition by these selected set of metabolic enzymes to achieve a controlled metabolism and excretion?

The field of prediction of CYP450-mediated drug metabolism has matured to a great extent in the last decade.10, 11 As a result, SOM prediction (predicting which atom/group of atoms

are involved in the metabolic reaction(s)), and isoform specificity prediction (prediction of CYP450 isoforms involved in the metabolism), for CYP450-mediated metabolic reactions is relatively well established. The state-of-the-art prediction tools (e.g., for SOM: MetaSite,[12, 13] XenoSite,[14] StarDrop (Optibrium Ltd., Cambridge, UK), SMARTCyp,[15] RS-predictor,[16] MetaPrint2D,[17] ADME Predictor,[18] JChem Metabolizer[19]; and for isoform specificity prediction: WhichCyp,[20] CypRules,[21] MetaSite,[12, 13] and VirtualToxlab[22]) can be employed for a good to excellent estimation of metabolic fates and enzymes responsible. This is only a representative list of useful tools (both free and commercial) available to the user. For an exhaustive list of softwares, research articles, and discussion on them, some of the recent reviews are highly recommended.[10, 11, 23]

This review focuses on the software tools and methodologies available for prediction of drug/xenobiotic metabolism catalyzed by non-CYP450, UGTs and non-UGTs. The content is divided into two major sections for Phase I and Phase II metabolic enzymes. Within each section, introduction to specific enzyme(s), substrates, their pharmacophoric features, and isoform specificities, (Q)SAR, SOM prediction tools (where available), and prediction of metabolic reaction kinetic parameters are discussed.

PREDICTION OF PHASE I (NON-CYP450) DRUG METABOLISM

FMO-Mediated Drug Metabolism

CYP450s-catalyzed reactions make important contributions to endobiotic, drug metabolism, and drug–drug interactions. Another major class of monooxygenases detected in the liver, lungs, kidney, small intestine, and brain are the NADPH-dependent FMOs.[24] FMOs are known to catalyze almost 6% of the Phase I metabolic reactions,[13] but the exact number of drugs metabolized by FMOs is not clear due to overlapping reaction types catalyzed by CYP450s.[25, 26] FMOs have five major isoforms, among which FMO3 has been reported to be the most abundant isoform in the liver. FMO isoforms 1, 2, and 4 are found in small amounts. Recent reports tend to agree with these findings, except for Cashman et al. who report FMO5 as the major isoform in the liver.[24, 25] In addition, the functional roles of these isoforms at various stage of human development have not been understood completely.[24] It is proposed that they play an important role in homeostasis by oxidation of important biogenic amines.[27]

The yeast (*Schizosaccharomyces pombe*) FMO structures reported by Alfieri et al. in 2008 shows that NADPH and substrate (methimazole) bind sequentially.[28] In the proposed mechanism, first step involves binding of the NADPH and dioxygen near the FAD site leading

to the formation of FADOOH peroxide active oxidizing species (Figure 4(a)). $NADP^+$ then is believed to dissociate leaving space for the substrate (methimazole, Figure 4(b)) binding. Substrate-bound structure suggests that this displacement of $NADP^+$ is the result of competition for $NADP^+$ binding site. This is followed by the oxidation of the nucleophilic centers (N, S, P, or Se) in the substrate by FADOOH. This oxidation step is a single-step, two-electron transfer reaction unlike the CYP450-mediated oxidations, which involve two single-electron transfer steps. Since representative crystal structures for none of the human FMOs are known till date, the actual catalytic mechanism is still open for further experimentation and debate. During more recent attempts to model FMO3-based drug metabolism by Cruciani et al. suggested that NADPH, dioxygen, and the substrate are added before the oxidation step and that release of $NADP^+$ and water are the rate-limiting step.13 Both these studies reported important role for the Asn residues in the stabilization of the active oxidizing species. In contrast to the CYP450s, this active oxidizing species in FMOs, flavin hydroperoxide, is a relatively stable intermediate which transfers an oxygen atom to the substrate's nucleophilic center (N, S, P, or Se). Molecular dynamics (MD) studies have shown that the shape and volume of active site (1000–1200 Å3) of FMO3 is relatively stable in comparison to some of the large CYP450s such as CYP3A4.13

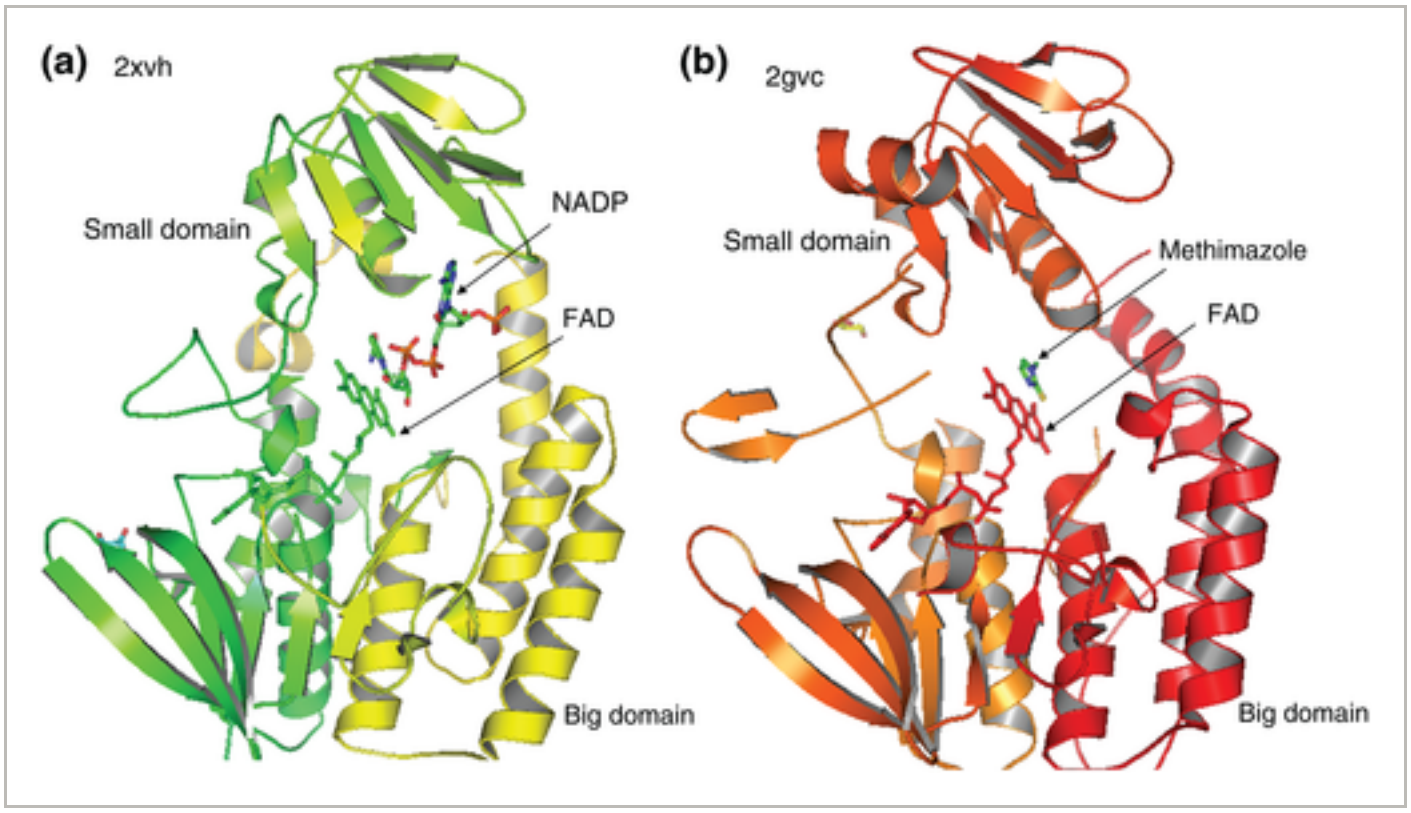

Figure 4

Open in figure viewer | PowerPoint

Crystal structures of yeast FMO. Panel (a) shows the FAD cofactor bound in the big domain while NADP is seen interacting with the small domain. Panel (b) shows the FAD cofactor in the big domain, while NADP has been displaced by the substrate (methimazole). Electron density analysis has shown the presence of dioxygen in both these crystal structures.

Caption ⌄

In contrast to the CYP450 family of drug-metabolizing enzymes, the FMOs have been overlooked until recently with respect to their importance in drug metabolism and drug–drug interactions. Many drug metabolism reactions (especially N and S oxidations) which were earlier ascribed to CYP450 have now been confirmed to be catalyzed mainly by the FMO especially the isoform FMO3.[13, 25] Moclobemide, arbidol, itopride, zimelidine, pargyline, ranitidine, benzydamine, olanzapine, and xanomeline are some of the well-known examples of drugs metabolized mainly by FMO3.[13] Study on FMO1 by Chen et al. had earlier shown that Bayer-Villinger-type sp^2 carbon oxidations can be catalyzed by this family of metabolic enzymes.[29] Recently, Fiorentini et al. have shown that FMO5 can catalyze a similar Bayer-Villinger oxidations of many small aliphatic and aromatic carbonyl compounds (aldehydes and ketones).[30] Since FMOs are rarely inhibited mostly due to narrow substrate specificities and rare reports of induction,[31] they are being considered as potential targets for selective drug-elimination pathways which can minimize drug–drug interactions and associated toxicities.[13, 27] With the increasing awareness about the contribution of FMOs toward drug metabolism, the FMOs are getting the attention of medicinal and

computational chemists, DMPK experts, and regulatory bodies.[4, 25] Thus, efforts are being directed toward the dissection of relative contributions of CYP450s, FMOs, and other non-CYP450s toward drug metabolism.

FMO3 versus CYP450s

A recent study by Wagmann et al. has used the concept of relative activity factor for identification of substrates for which FMO3 makes a major contribution in metabolism and clearance.[25] This study showed that for amitriptyline and clozapine, FMO3 plays only a minor role in hepatic N-oxidation, while for 3-bromomethcathinone (3-BMC), *N*,*N*-diallyltryptamine (DALT), DMT, methoxypiperamide (MeOP), and methamphetamine, N-oxidations were mainly catalyzed by FMO3 and CYP3A4 or 2D6. The case of 3-BMC and imipramine are peculiar as they showed major contribution to metabolism by FMO3-mediated N-oxidation. However, zopilcone N-oxide was formed exclusively by CYP450-mediated metabolism. Dextromethorphan, glaucine, and methadone were found not be substrates for FMO3, most likely due to steric hindrance. For amitriptyline, FMO3-mediated metabolism at high substrate concentration (high K_m and V_{max} values) might serve as an 'emergency metabolism route' for fast clearance of high drug concentrations.

Although, FMOs are known to perform hetero-atom oxidations giving N-oxides, sulfone, and sulfoxide metabolites, rare carbon oxidations have been observed. For example, Boersma et al. investigated the oxidation of 4-fluoro-*N*-methylaniline catalyzed by rat FMO.[32] They proposed the insertion of oxygen atom at the fourth carbon on the aniline ring with subsequent loss of HF and formation of quinone imine intermediate which was confirmed by trapping with *N*-acetylcysteine. FMOs have also been observed to catalyze Bayer-Villinger-type carbon oxidation reactions leading to formation of ester metabolites from drugs and drug candidates containing carbonyl groups. Oxazolidinone-based investigational antibacterial drug candidate MRX-I has been recently shown by Meng et al. to undergo FMO5-catalyzed oxidation. The 2,3-dihydropyridin-4-one ring is converted into the corresponding ester by FMO5, ester hydrolysis then gives the aldehyde which is then reduced by ADH to give an alcohol. The authors claimed that due to this dominant metabolic pathway, drug–drug interactions with CYP450 inhibitors are not expected. Similarly, FMO5 was observed by Lai et al. to catalyze sequential carbon oxidation of the 4-hydroxypiperidine ring in E7016 (an investigational anticancer agent) to the ketone and ester.[33] Thus, FMO3 and FMO5 have been well established to metabolize drug-like compounds and prediction of potential metabolism by these enzymes becomes important step in lead optimization and avoiding drug–drug interactions.

Computational Predictions of FMO/FMO3-Mediated Drug Metabolism

Over the years, FMOs, especially FMO3, have been recognized to be involved in drug metabolism. But, drug metabolism prediction tools in this area have lagged behind those for CYP450s and only a handful of studies directly probe the kinetic parameters, isoform specificity, or SOM for reactions catalyzed by FMOs. Following is the classification of studies based on the type of prediction they make for FMO-catalyzed reactions.

QSAR Studies for Prediction of FMO/FMO3 Substrates

Recently, Pirovano et al. have used QSAR studies for the prediction of K_m and V_{max} for the human, horse, rat, mouse, rabbit, and pig metabolic enzymes.[34] ALDH, ADH, FMO, and CYP450 enzyme kinetics experimental data for all available species and isoforms were collected, merged from the BRENDA enzyme database and other literature studies. Due to heterogeneity in the dataset, rate constants (K_{cat}) were converted into V_{max} using conversion factors such as weight of the protein when V_{max} were not directly reported in the literature. Genetic-algorithm-based methodology was used to extract six important molecular descriptors from a set of 2000 potential ones. For FMO model, these descriptors were (1) RHSA (sum of solvent-accessible surface area for atoms with partial charge less than 0.2) descriptors, (2) molecular bond E-state index, (3) dragon-6-derived atom-centered fragments, (4) dragon-6-derived hydrophilic factor, (5) number of heteroatoms, and (6) autocorrelation function based on Sanderson electronegativity. Out of these, the RHSA and E-state index (a functional group descriptor) were found to be the most important one. The FMO and CYP450 models were the least predictive (R^2_{adj}: 0.48, 0.52 for K_m values and 0.21 and 0.63 for V_{max} values, respectively), but the authors claimed that this is expected and the best that one can get from a model containing heterogeneous datasets.[35] The low explained variance in the FMO model was ascribed to the 'unusual' feature of the FMO catalytic cycle that substrate binding has no effect on the reaction velocity and the rate-determining step is the formation of the flavin hydroperoxy (FADOOH) species. This is indirectly in contrast with the substrate specificities discussed by Cruciani et al., who suggest that FMO3-catalyzed hetero-atom oxidation reactions do depend on the substrate–protein binding interactions.[13] An earlier study by Guoand Ziegler and coworkers have also shown that the V_{max} for microsomal FMO1-catalyzed oxidation of thiourea derivatives do depend on the substrate structure and molecular surface area.[36] These models with relatively smaller correlations are applicable for the prediction of kinetic parameters for novel class of compounds. This is due to the inclusion of data for a wide variety of compounds while developing the model.

Lin and Tsai have recently developed a linear QSAR model, by combing the feature selectable support vector machine (SVM) and genetic algorithm (GA), for the classification of human and pig FMOs (hFMO and pFMO, respectively).[37] In this method, the molecular

feature selection was incorporated as a part of model training process. SVM parameters and feature subset was simultaneously optimized by GA to improve the classification accuracy. A feature mask (selected feature) is inserted into the GA chromosome during the training phase, such that SVM parameter and features are coevolved. The feature mask set was then optimized for computing classification accuracy of both training and test sets. This optimized feature mask set was then used for the construction of a linear QSAR mode. K_m values were used to cluster the 71 and 151 substrates of hFMO and pFMO into three and four clusters, respectively. Calculated descriptors were reassigned binary values and employed to regroup the substrates. The authors found that 30 and 60 top-ranked features were sufficient to generate useful models for hFMO and pFMO. For both the enzymes, these features represented properties like electronegativity, polarizability, molecular volume, and shape. These molecular properties were calculated using E-DRGON web tool. Accumulated |w| values computed by the GFSVM methodology proved an effective feature selection scheme. Prediction of hFMO classification was found more difficult than the pFMO. This model was then applied on a set of 10 external compounds using descriptor generation using conformations derived from QM and MM methods. The predictions were better for QM (80%) than MM (70%)-based method of descriptor generation.

SOM Prediction and Substrate Specificity for FMO3-Mediated Drug Metabolism

Considering the small set of functional groups (amines and thiols) metabolized by FMOs (especially FMO3), one might think that the prediction of SOM should be straightforward. But, as shown by Cruciani et al.,[13] this is not the case, and prediction of SOM for FMO3-mediated reactions is complicated by factors such as the three-dimensional (3D) FMO3–substrate interactions, flexibility of the substrate (in the active site), accessibility, and reactivity of the potential site. When two substrates having very similar structures are allowed to react with FMO3, the most reactive (nucleophilic) site gets oxidized. When the structures are significantly different, then prediction of SOM is difficult as the relative importance of the deciding factors (reactivity and accessibility) may change from case to case. Even when structures of two or more compounds are very similar but have multiple oxidizable nucleophilic sites, the accurate prediction of SOM is difficult. For example, as shown in Figure 5 molecules **1** and **2** have very similar structures but are metabolized differently due to differences in the nucleophilicity of pyrrolidine nitrogen and thiophene sulfur atoms. Molecule **3** is metabolized by FMO3, while molecule **4** is not metabolized due to lower flexibility and hence accessibility of the amine group. There is no total charge preference observed for FMO3, i.e., substrates of either positive (e.g., xanomeline), negative (e.g., sulindac) charge, or even neutral species (e.g., albendazole) are known to be good substrates for this enzyme.

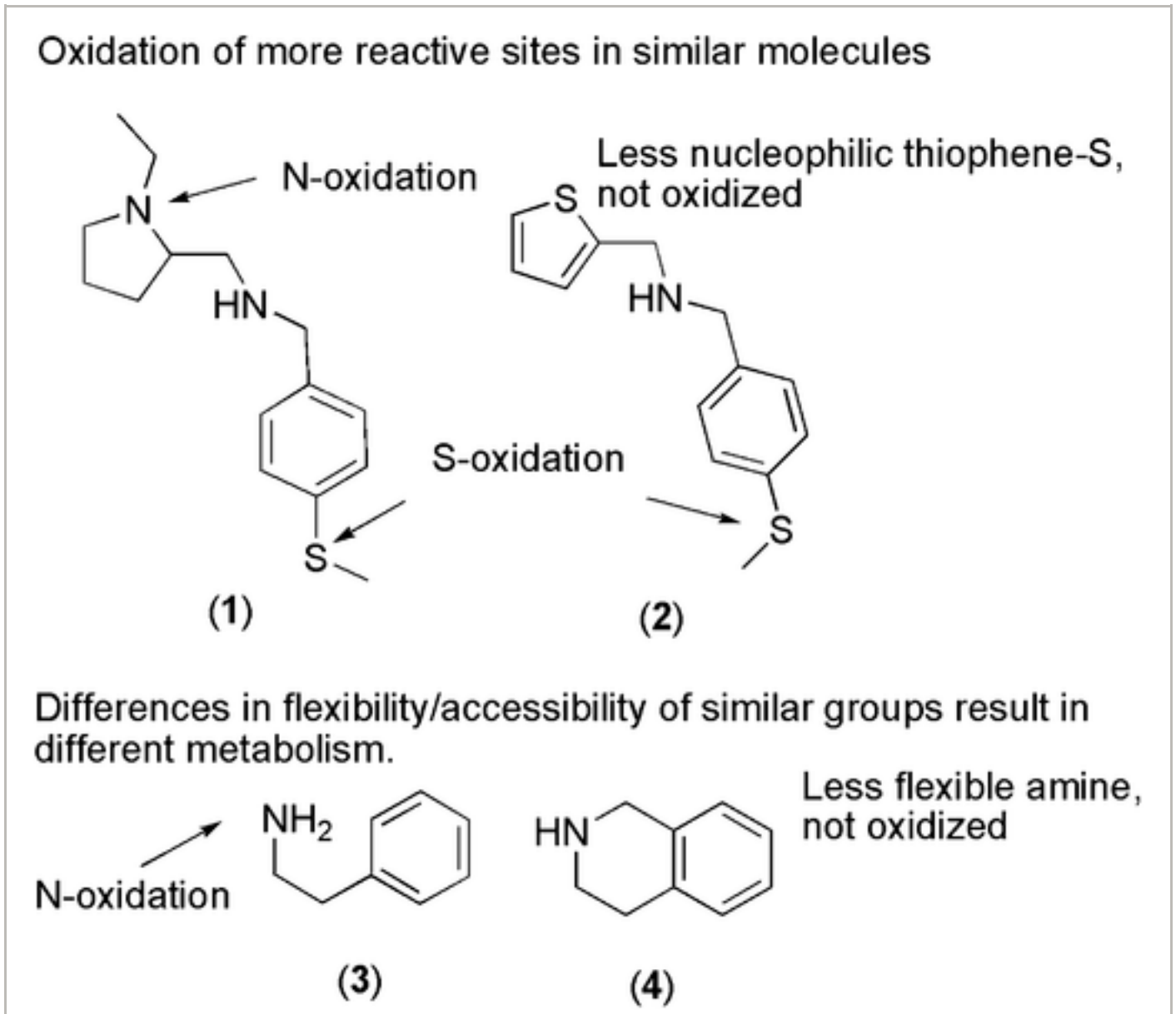

Figure 5

Open in figure viewer | PowerPoint

Accessibility and reactivity are important factors determining FMO3-mediated drug metabolism.

Caption

Cruciani et al. have modified the MetaSite12 program to include SOM and substrate specificity predictions for FMO3 using the FLAP algorithm.13, 38 In this method, both accessibility and reactivity (nucleophilicity) of potential sites are considered during prediction. For estimating accessibility, the molecular interaction fields (MIFs) were calculated dynamically for the protein and ligand using GRID-based force fields. The protein structure required for this was obtained by homology modeling based on bacterial enzyme crystal structures followed by MD simulations. Then FLAP algorithm is used to describe the protein and ligand structures using 4-point pharmacophoric models and molecular cavity shape. Ligand conformations in which MIFs are superimposable with those of the protein are considered energetically favorable and are recorded. For estimating the nucleophilicity, QC parameters (HOMO energy, polarizability, and atomic charge) were calculated. Partial

least squares (PLS) method was then applied for developing a chemometric model for the prediction of experimental nucleophilicity values. As in the older version of MetaSite, the reactivity and accessibility are combined in a probability function P_{SoM}, which estimates the likelihood of a site to be metabolized. In the MetaSite5 version, the reactivity and accessibility are recalculated using different weights such that for small- to medium-sized ligand's reactivity is given higher weightage, while for larger ligands accessibility contributes more to the final score for each site in the molecule. The method was able to predict 90% of known substrate and 85% of the nonsubstrates correctly from the known literature. For SOM prediction, the success rate was 92% as assessed by top-1 metric.

Recently, Fu and Lin have developed a ligand-based classification model for the prediction of SOM for FMO-catalyzed xenobiotic metabolism.39 They have utilized condensed atomic Fukui function and circular fingerprints (Molprint2D) for the classification of 85 molecules as substrates (73) and nonsubstrates (12) for N- and S-oxidative metabolism. Potential sites (N and S atoms, 228 in all) in the molecules were marked first, and then randomized into training and test sets. Circular fingerprints (37 attributes) were generated considering the atom types and types of their corresponding neighbors in layers. LIBSVM40 in the WEKA package was used for developing classification models, where parameter values for soft margin cost function (C) and Gaussian radial basis function (γ) were altered. The results showed that the atomic Fukui function correlated well with known SOMs than the atomic charge. But some of the nonsubstrates also had high Fukui function values, thus circular fingerprints were included in model building for more accurate predictions. Performance of the model was tested using standard Mathew's coefficient (MCC) (0.7), accuracy (0.86), sensitivity (0.85), and specificity (0.86) measures. Failure in correct prediction for some substrates phenothiazine was explained by the authors, as resulting from the FMO-specific interactions. The results from MetaPrint2D were compared for true negative rates, and the SVM-based model was found to be better (specificity = 0.91 vs 0.61).

Thus, interest in developing FMO-mediated metabolic predictions is slowly rising and there is enough scope in this area for better predictions. Current models make predictions in isolation, i.e., do not make predictions of relative contribution to metabolism by CYP450s and FMOs. For example, models for CYP2D6 (SMARTCyp) and FMO3 (MetaSite) can be used separately to estimate the substrate likelihood and SOM within the molecule of interest, but there is no direct way to rerank or compare the relative importance of these predictions.

Alcohol and Aldehyde Dehydrogenase

The ADH, ALDH, and amine oxidase (AO) family of enzymes metabolizes a small percentage of drugs and xenobiotic, because they recognize xenobiotics and some endogenous

compounds with very specific chemical features. For example, the AO enzymes recognize compounds with amine and aromatic groups separated by aliphatic spacer of different lengths,41 whereas ADH and ALDH specifically recognize compounds with alcoholic and aldehyde functional groups. Thus, the probability of a new xenobiotic, selected randomly, being a substrate for AO, AD, or ALDH is low. But a diverse set of compounds can act as inhibitors of these enzymes and thus lead to drug–drug, drug–food, or drug–xenobiotic interactions of clinical significance. Therefore, research in this area has so far focused on identification of potential inhibitors for either treatment of some diseases (AO: neuropsychiatric disorders,42 AD: alcohol poisoning,43 and ALDH: ischemia and cancer44) or avoidance of drug interactions. But whenever a xenobiotic meets the pharmacophoric requirements for these enzymes, it gets metabolized by these secondary Phase I enzymes. For example, almotriptan is metabolized by ALDH, and MAO-A by different oxidative pathways (oxidation and deamination).45 Very few computational studies and methods directly probe the substrate potential of any given compounds for these enzymes.46 Thus, predictive models for these enzyme systems are often limited to inhibitor prediction and there is a need for development of regioselective and isoform-specific substrate prediction models for these systems.

Alcohol Dehydrogenase

ADH is an NAD^+-dependent enzyme which catalyzes the oxidation of ethanol, retinol, and other biogenic alcohols into aldehydes or ketones. Six genes (*ADH1-6*) are known to express the corresponding isoforms which are classified into three main classes.47 Two of these classes are found in tissue-specific manner (especially liver and intestine), while the class III ADHs are found ubiquitously. The class I enzymes are known to form homo as well as heterodimers. There is only 50–60% sequence identify in the active site region despite higher overall sequence similarity in class I ADHs. Gibbons and Hurley have crystallized three isoforms of class I ADHs in complex with formamide derivatives.48 The αα-isoform has a slightly closed catalytic domain and in all the isoforms carbonyl oxygen of the ligand was in close contact with the Zn atom in the catalytic center and formed hydrogen bonds with either Ser48 or Thr48.

Generally, drug-like molecules having one or two primary (or secondary) hydroxyl groups are substrates for ADH. For example, Abacavir which has one hydroxyl group is metabolized by ADH (Figure 6). Walsh et al. studied the metabolism of Abacavir *in vitro* and identified the ADH αα and γ1γ1 isoforms responsible for the metabolism to acid metabolite and formation of protein adducts via possible migration of double bond to give Michael acceptor.49 The presence of double bond seems to be necessary for oxidation by ADH covalent adduct formation, as analogues with saturated rings (didanosine and emtricitabine) are not oxidized

by ADH and directly undergo glucuronidation by UGT. More recently, Gunness et al. have shown that nephrotoxicity of acyclovir is mediated by oxidation to the aldehyde which is subsequently oxidized by ALDH to 9-carboxymethoxymethylguanine (Figure 7).[50]

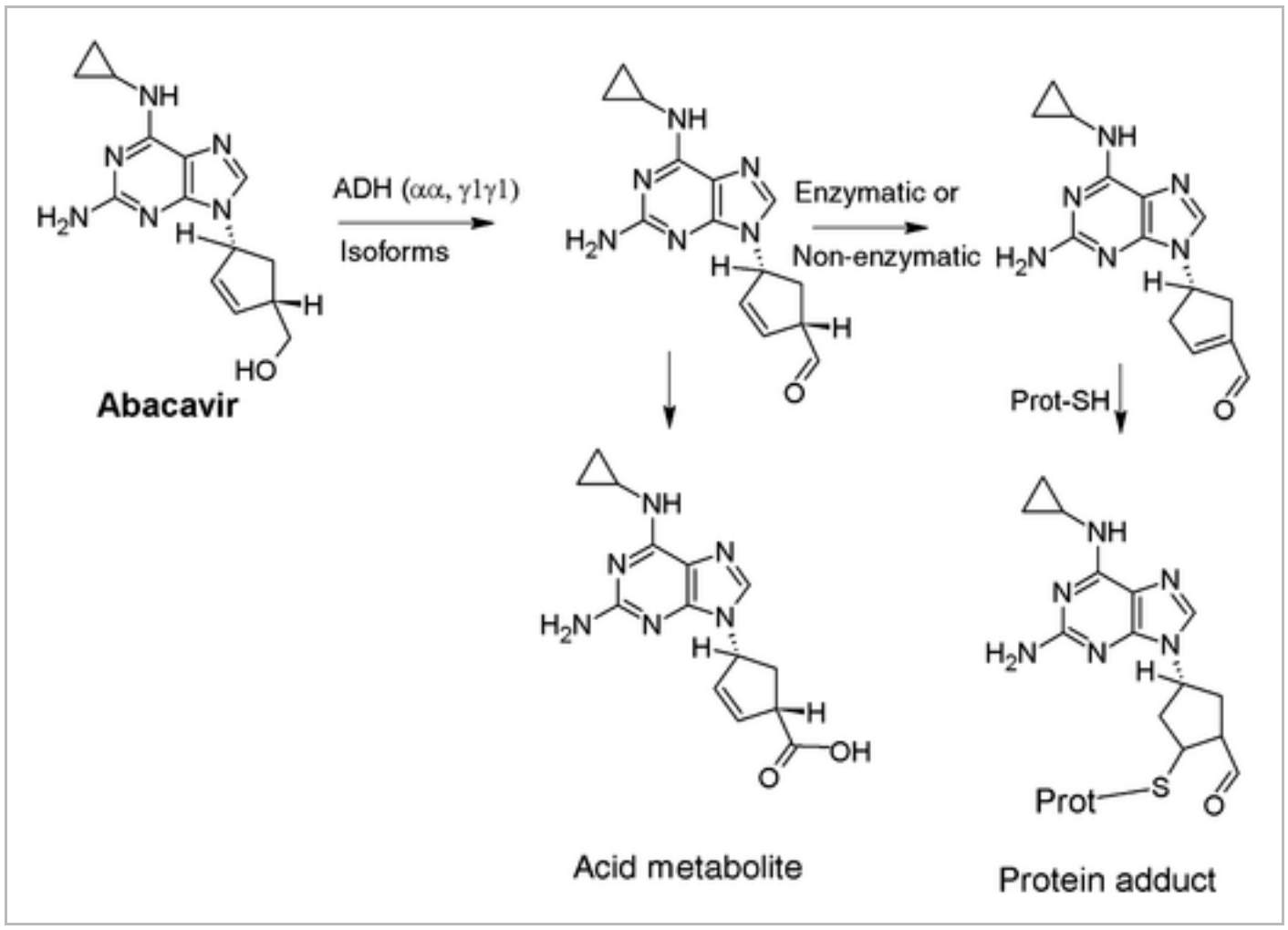

Figure 6

Open in figure viewer | PowerPoint

Abacavir metabolism by alcohol dehydrogenase (ADH) isoforms to reactive metabolites.

Caption ⌄

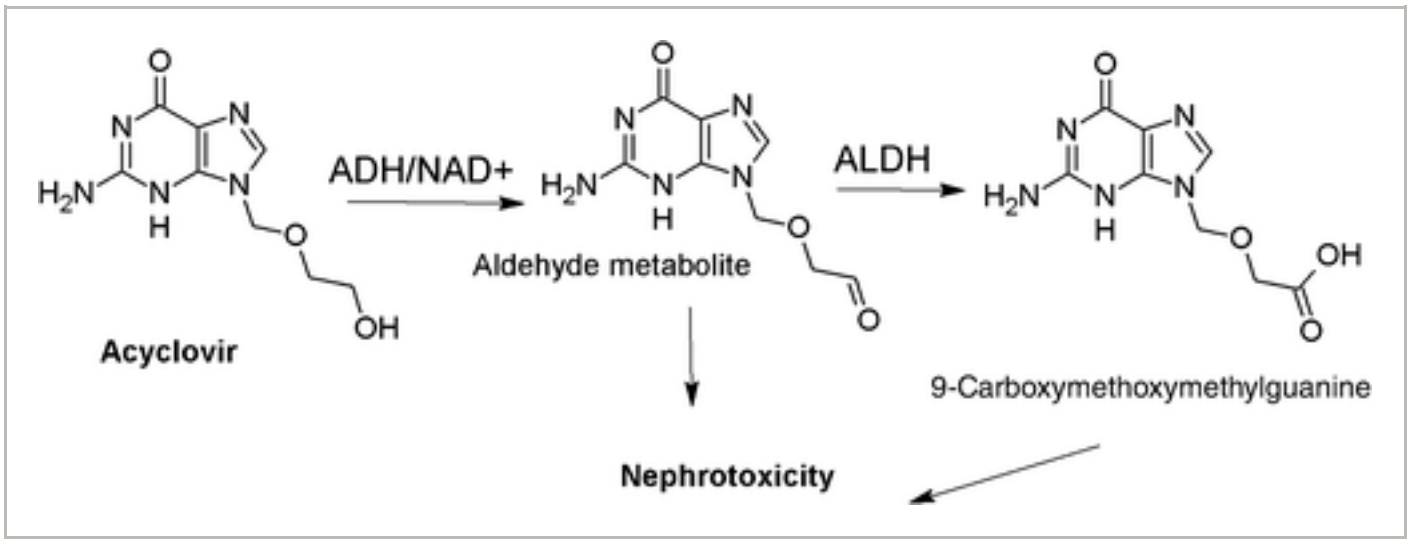

Figure 7

Open in figure viewer | PowerPoint

Possible mechanisms of nephrotoxicity of acyclovir.

Caption ⌄

Due to the narrow substrate binding site, straight-chain, aliphatic-alcohol-containing compounds are generally oxidized by ADH51 and desolvation appears to be a critical step before substrate binding. Nevertheless, identifying SOM is not as straightforward in these compounds as one would imagine. This is exemplified by another nucleoside analogue zidovudine, which contains a terminal hydroxyl group, but it is not a substrate for ADH and instead is glucuronidated by UGT. Another important question that needs to be answered is whether a given compound is a substrate or inhibitor for ADH or not? There is a descent amount of literature available on the prediction of inhibitor potential, but for substrate prediction very few studies are available in the public domain.

QSAR Studies for ADH Inhibitor Prediction

Hansch et al. developed QSAR models for the prediction of ADH inhibition constants.51 They found that octanol/water partition coefficient (log *P*) alone is good enough to make reasonably accurate predictions for the inhibition constants for a series of pyrazole, acetamide, and benzylamine derivatives. The hypothesis for desolvation was supported by the fact that the coefficients for log *P* in most of the QSAR equations were approximately 1. Cases where smaller or larger coefficients were found were explained by invoking factors like induced-fit conformational changes in the receptor, possibility of different binding modes for some compounds within the series, and unfavorable steric interactions with Phe93 and Ser48.

Molecular Docking Studies to Understand Receptor Interactions and Substrate Predictions

Earlier, Moreno et al. have performed docking studies on the class I and IV (gastric) human

ADHs.[52] Models for class IV isoform were developed from class I isoform crystal structures. Class IV isoform active site was found to have larger surface area than the class I isoform, partially explaining substrate-specificity differences. Based on the binding energy estimations and distances from the Zn atom in the docked poses, all-trans retinol was found to bind tightly with class IV isoform in comparison to class I enzymes. 11-*cis* Retinol was also predicted to have higher binding affinity for the class IV isoforms. Docking simulations with coenzyme showed difference in the binding mode between the two enzymes. In class I isoform, the docking reproduced the crystal structure pose, whereas for class IV isoform different conformations of the Arg47 resulted in different orientation of the adenine ring, explaining the decreased binding affinity and higher K_{cat} values for this isoform.

H2-receptor antagonists Cimetidine was observed by Stone et al. to inhibit alcohol oxidation by ADH isoforms.[53] Molecular docking studies suggested good fit for Cimetidine in the active site, but other analogues, such as Nizatidine, Ranitidine, and Famotidine, did not dock well due to steric clashes between dimethylaminomethyl and guanidine substitution on the five-membered rings. A conflicting report has been reported in a later study by Brown.[54] They have used ethanol concentrations relevant to real-life settings and found that cimetidine, ranitidine, and omeprazole have little impact on the ethanol peak concentrations, area under the time–blood ethanol curve and first-pass metabolism. More recently, Lai et al. have reinvestigated the topic using recombinant enzyme systems (eight ADH isoforms).[55] In agreement with both the references cited above, they found that Cimetidine inhibits ethanol metabolism at low ethanol (10 mM) concentrations, while at physiologically relevant concentrations the inhibition is minimal.

Mutational studies by Plapp and coworkers[56] and Adolph et al.[57] on horse liver ADH have earlier established the role of active site shape on the substrate specificity. They found that the enzyme works as a dimer of either the E or S monomers, giving three isoforms, EE, ES, and SS. The EE and ES isomers catalyze the oxidation of ethanol, while the SS isoform can oxidize ethanol and steroidal substrates also. This was also supported by the interconversion of E into S subunits by mutations of selected residues in the active site. Similar yeast studies have been reported, where mutation of active site restudies (T48S/W93F) has been shown to alter substrate specificities.[58]

Aldehyde Dehydrogenase

ALDH is another important enzyme involved in the $NAD(P)^+$-dependent oxidation of biogenic and xenobiotic aldehydes to carboxylic acids or CoA esters. It plays an important role in the cell proliferation and homeostasis by regulating the level of retinal, retinoic acid in the cytosol, and neurotransmitter GABA (γ-aminobutyric acid). Nineteen functional genes have

been identified so far in humans. Biogenic aldehydes generated from the lipid peroxidation (e.g., 4-hydroxyhexenal, malondialdehyde, and 4-hydroxynonenal) are substrates of ALDH.[59] Mutations in several ALDH isoforms have been associated with neurological disorders, mental retardation, embryonic developmental disorders, and might be responsible for the development of clef palate syndrome.[60-62] ALDH1 isoforms share approximately 70% sequence identity, while other members of the family have around 50% sequence similarity among them.

Since aldehydes are generally very reactive species due to the terminal position of electrophilic carbonyl carbon, very few drugs and drug candidates are designed with this functional group. Thus, ALDHs are mostly involved in the metabolism of primary metabolites resulting from the action of other Phase I metabolic enzymes. As a result, very little effort has so far been devoted to the prediction of drug metabolism catalyzed by ALDHs.

A relatively simple but important structure–activity relationship (SAR) study was performed by Haynes et al. to understand the toxicity of 4-hydroxy-2-nonenal (4HNE).[63] They studied the relative importance of each structural feature in 4HNE, namely the carbonyl, double bond, and hydroxyl group by testing the toxicity of substrates with and without these features. The order of toxicity in compounds with these individual chemical features was C1 aldehyde > double bond > hydroxyl group. This was consistent with growth inhibition and apoptosis assays in macrophage cell lines. Interestingly, cell overexpressing ALDH were resistant to the toxic effects of 4HNE suggesting a major role in detoxification pathways.

Cyclophosphamide, an anticancer agent, is converted into the active metabolite 4-hydroxycyclophosphamide. This metabolite tautomerizes to cytotoxic aldophosphamide which is detoxified by ALDH to carboxyphosphamide. Thus, inhibitors of ALDH have the potential to increase the efficacy of such cytotoxic compounds in the treatment of cancer. One of these purine derivatives has been used in other studies to investigate the physiological and pathological roles of ALDH1A1 especially in ovarian cancer.[64]

ALDH isoforms have also been shown to have $NAD(P)^+$-independent esterase activity and the same active site is known to catalyze the ester hydrolysis.[65] There has been a renewed interest in the inhibitors of ALDH isoforms, especially ALDH1A1, ALDH2, and ALDH3A1 for the discovery of novel compounds for diseases and conditions like alcoholism, cocaine addiction, and cancer.[44] Parajuli et al. have found, through high-throughput screening, that vinyl ketones generated in the first step of the enzymatic reaction can covalently add to the catalytic Cys residue of ALDH1A1, ALDH2, and ALDH3A1.[44] Later on in the same group, Kimble-Hill et al. have developed selective inhibitors for these enzymes by modulating the structure of indole-2,3-dione derivatives.[66] These compounds exploit the fact that sequence

identity between ALDH2 and ALDH3A1 is low (~30%) and they have differences in the active site residue composition. Although these isoforms share active Cys residues in the active site, only ALDH3A1 isoform formed covalent adduct with compound **IV** via the 3-keto group. These inhibitors were found to display noncompetitive kinetics for the coenzyme binding site, while for the substrate binding site the kinetics data supported competitive binding hypothesis. Substitution at C5 and N1 gave selective inhibitors as shown in Figure 8.

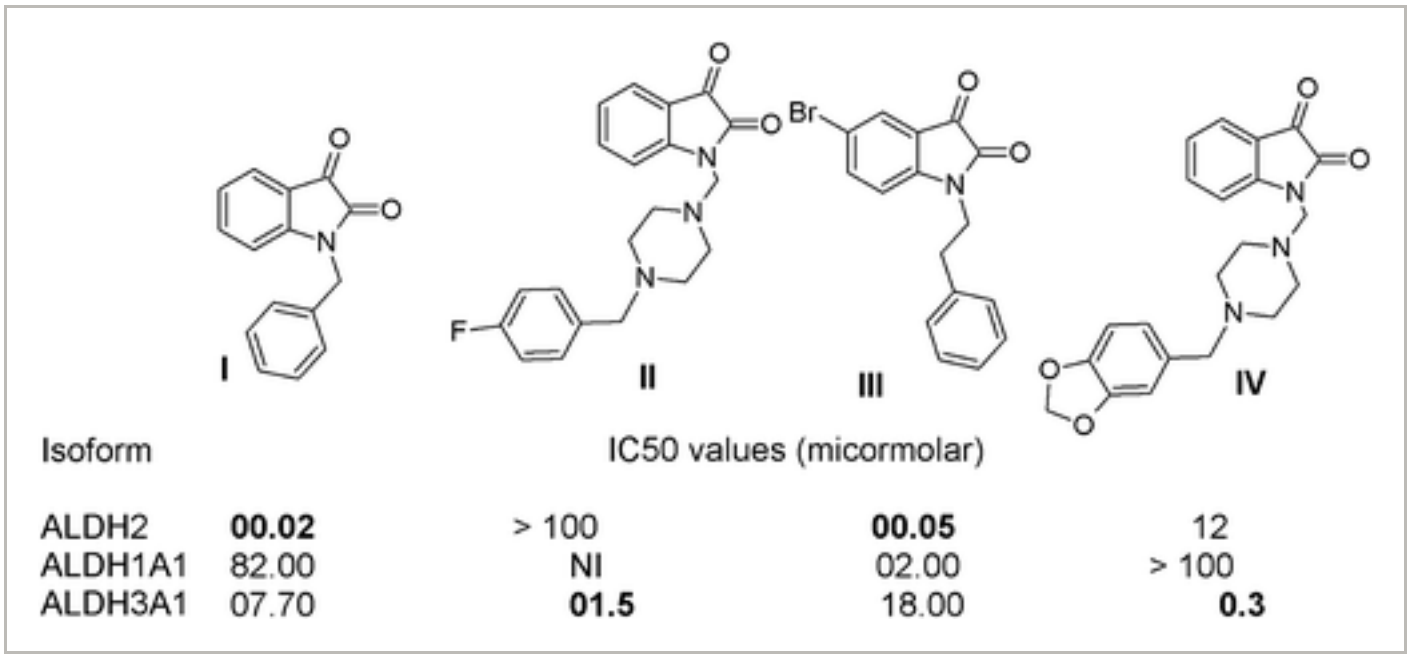

Figure 8

Open in figure viewer | PowerPoint

Indole-2,3-dione derivatives with selectivity toward the isoforms of aldehyde dehydrogenase (ALDH). The bold values indicate that the activity of the structures above are high for the corresponding isoforms.

Caption

Morgan and Hurley have developed two classes of novel inhibitors for ALDH1A1 using an esterase assay, as there is a descent overlap between the active site for ALDH and esterase activity.67 This assay gave the advantage of noninterference from the absorption characteristics and elimination of compounds that might compete with the cofactor for binding. SAR studies showed that purine nucleus alone is not sufficient to show inhibitory activity as theophylline and caffeine did not inhibit any of the isoforms of the enzyme. Crystal structure analysis gave vital clue about the difference in the binding pocket of ALDH1A1, ALDH2, and ALDH3A1. ALDH1A1 was found to have a wide opening leading to the active site, thus allowing the CM037 and CM026 to binding selectively to this isoform.

Mitochondrial ALDH2 is known to play an important role in the efficacy of nitroglycerin to act as an antianginal drug.68 It is also involved in the metabolism of acetaldehyde (derived from

ethanol oxidation) and other biogenic aldehydes such as 3,4-dihydroxyphenylacetaldehyde.[69] Thus, inhibition of ALDH2 in the brain can increase the risk of Parkinson's disease (PD).[70]

Prediction of metabolism of ADH and ALDH systems will play an important role in sequential metabolism prediction models, i.e., models which consider the probability of formation of secondary metabolites from the action of these enzymes on primary metabolic products formed by the action of CYP450, FMO, or MAO. The review of CYP450 predictive models[23] and the recent interest in the prediction of non-CYP450 Phase I and Phase II metabolic processes hint that development of such sequential models is within the grasp of the current generation of drug metabolism scientists.

Monoamine Oxidases (MAO-A and MAO-B)

MAOs are flavoenzymes which catalyze the oxidation of a wide variety of biogenic, xenobiotic amines, and are located on the outer membrane of mitochondria.[71] MAO genes are located on chromosome X and mutations are associated with a rare Brunner syndrome.[72] Two isoforms MAO-A and MAO-B are well known and in contrast to FMOs are distributed in a wide variety of organ and tissue systems except the blood platelets and spleen. Distribution of these isoforms within the brain is neurotransmitter dependent and MAO-A is found in dopamine-secreting neurons, while MAO-B is present predominantly in serotonergic neurons. Sequence analysis shows that they have around 70% sequence identity and have 526 and 520 amino acids, respectively. Secondary structure for human MAO-B (1GOS) shows 15 α-helices, 20 β-sheet, and 9 3/10 helices, whereas the secondary structure for MAO-A (4A7A) shows 17 α-helices, 21 β-sheet, and 8 3/10 helices which binds rosiglitazone by noncovalent interactions (PDB ID: 4A7A). MAO-A has a distinct loop region (residues 210–216) and Glu151 is replaced by Lys, thus partly explaining its substrate specificities.[73] The active site volume of MAO-B (~700 $Å^3$) is larger than MAO-A (~550 $Å^3$). Both isoforms have the typical fold and have highly conserved regions for (1) ADP, (2) substrate binding, (3) covalent attachment of FAD, and (4) a transmembrane helix at the C-terminus. Earlier, Cesura et al. identify His382, Cys397, and Thr158 by site-directed mutagenesis to be involved in the catalytic cycle of MAO-B.[74] Later, Geha et al. identified that a different set of amino acid residues in the active site play an important role in the catalytic cycle (MAO-A: Lys305, Trp397, Tyr407, and Tyr444 and MAO-B: Lys296, Trp388, Tyr398, and Tyr435).[75]

For catalytic activity, presence of two hydrogens on the α-carbon of amines is essential, thus both MAOs can catalyze the oxidation of primary, secondary, and tertiary amines. Over the last two decades, evidence for the polar mechanism has built up against the radical

pathways.71, 76 The radical pathway involves initial hydrogen abstraction to give carbon center radical followed by combination with active site radical and subsequent formation of iminium ion which is hydrolyzed nonenyzmatically to aldehyde and ammonium ion, whereas the polar pathway involves formation of the FAD adduct via nucleophilic attack of the amine nitrogen and subsequent proton abstraction by a basic residue in the active site (Figure 9). Recent computational studies also support the polar nucleophilic pathway.77, 78

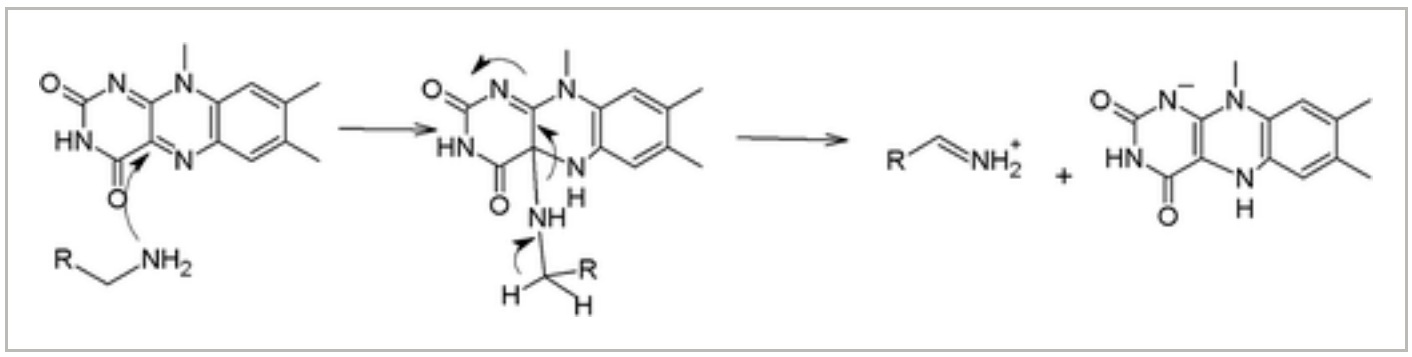

Figure 9

Open in figure viewer | PowerPoint

Proposed polar mechanism for the oxidation of amine with hydrogens on α-carbon atoms.

Caption ⌄

SAR studies have shown that MAO-A has specificity for secondary amines, while MAO-B has moderate specificity for the tertiary amines in the phenylethylene amine series. MAOs are also known to oxidize aliphatic amines containing 5–10 carbon atoms such as *N*-acetylspermidine, but it selectively drops sharply outside this chain length range.71 Substrate recognition also fails when branched chains are present as in enantiomers of amphetamines which act as inhibitors. The selectivity is thought to result from different recognition of substrates/inhibitor due to the presence of Ile335 in MAO-A and Tyr326 in MAO-B.79

As with FMOs, the MAOs initially were almost neglected in many drug metabolism studies and CYP450s were ascribed as major enzymes responsible for the metabolism of these xenobiotics.46 For example, oxidative deamination of citaprolam was attributed to CYP450-catalyzed metabolic reactions.80 But later studies by Kosel et al. have found that citaprolam is also metabolized in rat and human brain by MAO-A and MAO-B via deamination reactions.81

Because of intense interest in the MAO inhibitors for various pharmacological applications, e.g., treatment of Parkinson's, Alzheimer's, and other neurological disorders,82 majority of the literature focuses on prediction of inhibitory activity and associated parameters. These models might be of indirect use in the substrate prediction, since many of these inhibitors

are themselves substrates of good to moderate activity, e.g., tranylcypromine, pargyline, phenelzine, and others.[83] But the direct use of these models in metabolism prediction is difficult due to differences in chemical properties of substrates and inhibitors and narrow chemical space used to develop these models. Nevertheless, a brief summary of the currently available models for substrate/inhibitors is given below. The interested readers are directed toward two recent reviews on the topic.[42, 84]

Ligand-Based Models for MAO Substrate/Inhibitor Prediction

Vilar et al.[84] have reviewed the details of the ligand-based (Q)SAR models for the prediction of inhibitors of MAO isoforms with an aim to get insights into their SAR and mechanism. A short summary of these studies is given in Table 3. Pyridazines were found more selective for MAO-A and pyrimidines are more selective toward MAO-B isoforms. Potency of MAO-B inhibitors not only depends on V_w, lipophilicity, and hammet constant of the substituents attached to C-3 of the phenyl ring, but it also varies with the substituents attached to C-4 because of bulkiness and lipophilicity. Coumarins showed more affinity and selectivity for MAO-B isoform. Short compounds with structural similarity showed similar biological activity. All the ligand scaffolds mentioned in Table 3 are aromatic compounds (except thiosemicarbazide derivatives). They contain nitrogen either in their heterocyclic ring (pyrrole, pyrimidine, caffeine, or pirindole) or outside the ring (phenyl alkylamines or phenyl isopropylamines) except coumarin and xanthones which do not contain nitrogen. For indole and isatin analogues, MAO-A inhibitors were generally of larger molecular volume and molecular weight than the MAO-B selective inhibitors.

Table 3. Summary of the QSAR and Pharmacophore Studies Performed on the MAO Inhibitors

S. No	Modeling Method	MAO Isoform	Ligand Scaffold	Descriptors	Performance Metric	Reference
1	2D QSAR with MLR	A	Xanthone	E-state indices (S_i), molecular connectivity (χ) and shape (k) index	$r^2 = 0.847$, $\sigma = 8.06$, $n = 34$	85
2		A	Pyrrole	Topological: connectivity, eccentricity,	$r^2 = 0.904$, $n = 32$	86

				superpendentic, Wiener, and flexibility indices		
3		B	Pyrrole	Taft steric parameter (Es) and Swain–Lupton electronic constant (*F*) for phenyl ring substituents	r^2 = 0.970, *F* = 72.2	87
4		B	Pyridazine and pyrimidine	HPLC parameters	r^2 = 0.82, q^2 = 0.704, *n* = 14	88
5		B	8-Benzyloxycaffeine	Hansch lipophilicity (π), Hammet electronic (σ) constants and van der wals (V_w)	C-3 substitution: r^2 = 0.99, *F* = 109 C-4 substitution: r^2 = 0.99, *F* = 608	89
6	2D QSAR with PLS	A	Phenyl alkylamines	Hammet electronic (σ) constants, Swain–Lupton field (*F*), Hansch lipophilicity (π), molecular refractivity (MR)	*n* = 29, r^2 = 0.73	90, 91
7	3D QSAR	A	Xanthone	Energy of highest, second	*n* = 42,	92

				highest, lowest, second lowest and occupied π orbital for ring 1, ring 2 and ring 3	$r^2 = 0.825$, $F = 14.633$, $S = 0.392$, $q^2 = 0.6322$.	
8		A	Phenyl isopropylamines	Energy of highest occupied molecular orbital (E_{HOMO}), charge descriptors	$n = 18$, $r^2 = 0.96$, $q^2 = 0.89$, $S^2 = 0.05$, $F = 56.89$ $n = 15$, $r^2 = 0.77$, $q^2 = 0.63$, $S^2 = 0.32$, $F = 8.33$	93
9		A and B	Aryloxyacetohydrazides	Electrical topological state index (T_q)	$n = 24$, $r^2 = 0.791$, $S = 0.20$, $F = 76.4$	94
10	3D QSAR	A	Xanthone	Steric and electrostatic fields	$n = 34$, $r^2 = 0.87$, $q^2 = 0.74$	95
11 12		A	Indole and isatin analogues	Molecular volume and electron density	$r^2 = 0.743$ $r^2 = 0.94$, $q^2 = 0.66$	96, 97
		B	Indole and isatin analogues		$r^2 = 0.603$ $r^2 = 0.97$, $q^2 = 0.85$	96, 97
13		A	Pirlindole analogues	AM1 charges, hydrophobicity (log *P*), CoMFA model	$n = 5$, $r^2 = 0.444$, $F = 65$, $S = 0.19$	98

14	B	Pirlindole analogues		n = 4, r^2 = 0.525, F = 131, S = 0.15.	
15	A	Indolylmethylamine derivatives	Steric and electrostatic properties at the 5 position of the indole ring and solvation	n = 16, r^2 = 0.993, q^2 = 0.895, F = 220.86	99
16	B	Indolylmethylamine derivatives		n = 16, r^2 = 0.998, q^2 = 0.859, F = 674.083	
17	A	Phenethylamine derivatives	Steric and electrostatic fields	n = 38, r^2 = 0.92, q^2 = 0.72	100, 101
				n = 31, r^2 = 0.994, q^2 = 0.614	
18	A	Coumarin derivatives	CoMFA-GOLPE, steric, electrostatic and lipophilic fields. Docking.	n = 38, r^2 = 0.819, q^2 = 0.727	102
19	B	Coumarin derivatives		n = 38, r^2 = 0.821, q^2 = 0.692	102
				n = 58, r^2 = 0.856,	

					$q^2 = 0.605$	
20		A	Hydrazothiazoles	CoMFA	$n = 45$, $r^2 = 0.752$, $q^2 = 0.449$	103
21		B	Hydrazothiazoles		$n = 45$, $r^2 = 0.678$, $q^2 = 0.356$	103
22		A	Pyrrazole derivatives	Glide docking and CoMFA	$n = 10$, $r^2 = 0.978$, $q^2 = 0.671$	104
23		B	Pyrrazole derivatives		$n = 8$, $r^2 = 0.912$, $q^2 = 0.555$	104
24	3D pharmacophore	B	Thiazole and thiosemicarbazide derivatives	HYP1: HB donor, positive ionizable and aromatic ring or HYP2: two HB acceptors, one hydrophobic and one aromatic ring	$n = 100$	105
25		A	Coumarin analogues	Two HBA, three hydrophobic groups, and GOLD docking	$n = 64$, $r^2 = 0.95$	106
26	3D pharmacophore	A	Tricyclic derivatives	Three HBA and one aromatic	$r^2 = 0.959$ $q^2 = 0.623$	107

	with 3D QSAR			ring		
27		B	Indole and pyrrole derivatives	HBD, two hydrophobic groups, and an aromatic ring	NA	108

NA, not applicable.

Hypertensive side effects of MAO inhibitors have forced their use as second line of treatment for neurological disorders when other agents fail to give the desired pharmacological effect. As mentioned in *Alcohol and Aldehyde Dehydrogenase* section, very few studies focus on substrate, isoform specificity, or regioselectivity predictions for MAO enzymes. Nevertheless, the proportion of drugs that are metabolized by these enzymes is not negligible. Given the fact that modern drug design methods are designed to bypass CYP450-catalyzed metabolic pathways, drug metabolism is likely to depend more on these secondary Phase I enzymes. Thus, there is an urgent need for the development of reliable predictive models for MAO, ADH, and ALDH class of enzymes.

PREDICTION OF PHASE II (UGT AND NON-UGT) DRUG METABOLISM

Although the Phase I metabolic reactions result in polar metabolites, their water solubility may still be low enough to preclude excretion through urine. Therefore, many of these primary Phase I metabolites are further metabolized by Phase II metabolic enzymes. This second phase of drug metabolism generally consists of enzymatic reactions catalyzed by UGT, STs, GST, and MTs and *N*-acetyltransferases (NATs). Occasionally, other Phase II reaction like transpeptidation can also take place. These reactions generally further increase the polarity of the drugs or Phase I metabolites by incorporating groups, such as glucuronic acid, sulfate, and glutathione or cysteine at nucleophilic sites in the molecules. Such sites can also be called as 'site of conjugation' (SOC), in analogy to the SOM for Phase I enzymatic reaction.109 Based on the identity of the reaction and the groups added, these conjugations can be further classified (Figure 3) into (1) SOG; (2) SOS; (3) SOGT, (4) SOMe, and (5) SOT. The increase in polarity of Phase II metabolites increases their water solubility and excretion via urine. In addition, incorporation of these groups may increase efflux/transport (Phase III metabolism) by membrane transporters in the liver and kidney.110 Generally, conjugation

terminates the pharmacological action of the parent molecule. A well-known exception is the glucuronidation of morphine to morphine-6-*O*-glucuronide, which has a larger analgesic activity than the parent molecule.[111] Although the role of Phase II metabolites in toxicity has been recognized since late 1990s,[112] only a handful of models for prediction of kinetic parameters, substrate, isoform specificity, or SOC haven been developed till date, and these have limited applicability domains. Recently, there has been some interest in the development of predictive models, especially for UGT-mediated metabolic reactions, which are the most important family of Phase II metabolic enzymes. A few studies do focus on the substrate prediction for sulfotransferase (SULT) and GST families of enzymes. But other Phase II metabolic enzymes remain neglected in this regard. This fact is apparent from the number of research papers containing the concepts of these enzymes and drug metabolism (Table 2, exception is MT due to its role in epigenetic regulation of gene expression).

In the following sections, a brief introduction to different Phase II metabolic enzymes and discussion of currently available predictive models is given. Whenever feasible a classification of these studies based on the type of prediction (substrate, kinetic parameters, isoform specificity, or regioselectivity) is provided.

UDP-Glucuronosyltransferase (UGT)

UGTs (EC 2.4.1.17) are a family of glucuronosyltransferases, which catalyze the transfer of glucuronic acid group from high-energy UDP-glucuronic acid to nucleophilic sites in many xenobiotics, drugs, and a few important endobiotics (e.g., dopamine and 5-hydroxytryptamine). This metabolic reaction can significantly decrease the bioavailability of many drugs and natural product, cause activation of some drugs (e.g., morphine), and lead to drug–drug interactions.[113] This reaction is also known as glucuronidation and fall under the general class of Phase II conjugation reactions (Figure 3). The reaction takes place by a serine-hydrolase such as S_N2-type displacement of UDP from UDP-glucuronic acid by a nucleophilic center in the substrate.[114] Products of CYP450-catalyzed reactions, containing nucleophilic hydroxyl, amino, phenol, carboxylic acids, and thiols, are good substrates for UGTs. A few substrates with nucleophilic carbon atoms are also known to form C-glucuronides.[115, 116] The intracellular localization of CYP450 and UGTs on the microsomal members indicate a coupling of these enzyme systems for effective biotransformation of their substrates. Multiple isoform of UGTs (1A, 2A, 2B, 3A, and 8A) are known to be associated with inherited disorders of bilirubin metabolism and involved in xenobiotic/drug metabolism.[117] Although liver contains the largest amount of these enzymes, UGTs are also found in significant amounts in other organs, such as kidney, brain, lungs, intestinal tract, and so forth.[118]

In the only crystal structure of human UGT known so far, Miley et al. found that residues in the C-terminal UDP-glucuronic acid binding domain play an important role in the catalytic activity.[114] The N-terminal domain contains the substrate binding site, and a cryptic membrane binding region.[119] A homology model based on plant flavonoid glucosyltransferases suggested a reaction mechanism similar to serine hydrolases. N-terminal domain has the largest variation among the UGTs, explaining the substrate promiscuity and overlap. Androsterone, estriol, and retinoic acid are some of the endogenous substrates, whereas furosemide, levothyroxine, and paracetamol are some of the well-known drug substrates.[120] Drug–drug interactions involving UGT substrates are well known in the literature,[120] but seldom cause variations in drug concentrations greater than twofolds. Such drug pairs include valproic acid/lorazepam, valproic acid/zidovudine, and valproic acid/lamotrigine. An analysis of the *in vitro* K_m and K_i values for UGT substrates and inhibitors, respectively, reveals that the expected pharmacokinetics is similar to the CYP450s. However, *in vitro–in vivo* correlation is not always straightforward as structurally different substrates and different organs will have different scaling factors for the extrapolation of *in vitro* data to *in vivo* clearance.[121, 122] Additionally, the experimental conditions chosen during *in vitro* experiments (e.g., pH, buffer system, detergent, or any endogenous activators) are known to alter glucuronidation rates.[109] Both the incubation period and pH are known to affect the glucuronidation rate, isoform specificity, and result in regioselective glucuronidation depending on the identity of the flavonoid.[123]

SARs for glucuronidation of flavonoids were reviewed by Wong et al. in 2009.[124] Quercetin and luteolin were primarily metabolized by glucuronidation at O7 and O3 positions by the UGT1A family. Isoflavonoids such as daidzein, genistein, and prunetin were glucuronidated at the O7 position by the UGT1A isoforms. Other UGT isoforms did not make significant contribution to the metabolism of these molecules. UGT isoforms differ in their selectivity toward transferring the glucuronic acid moiety to hydroxyl or amino groups and Table 4 shows examples of typical UGT substrates and inhibitors. For example, in phenolic compounds, nucleophilic attack by oxygen is preferred by UGT1A1 and UGT1A9 isoforms, while for compounds containing a nucleophilic nitrogen atom, *N*-glucuronidation by the UGT1A4 isoform is the preferred pathway.[134] The position of the hydroxyl and substitution pattern was also found to affect the regioselectivity of this metabolic reaction. Planar flavonoids are known to be poor substrates of UGTs. In addition, flavonoid diglucuronides and flavonoid glucoside glucuronides have also been detected *in vivo*.[135] Thus, the *in vitro* screening for UGT activities targeting only monoglucuronidation might not give a complete picture of the complexity of this Phase II biotransformation process. Substitution of hydroxyl groups with methoxyl group decreases the metabolic activity of UGTs for flavonoids.

Table 4. Typical Substrates and Inhibitors of UGT Isoforms

UGT Isoform	Substrates	Inhibitors	Reference
UGT1A1	Ethinylestradiol, estradiol	Sorafenib	125
UGT1A3	Sulindac sulfone	Ritonavir, quinidine	126, 127
UGT1A4	Imipramine, trifluoperazine	Diclofenac, probenecid	128
UGT1A6	Serotonin, planar molecules such as simple phenols (e.g., naphthol and indoles)	Small molecular weight phenolic compounds and diclofenac	129
UGT1A9	Propofol, sorafenib	Axitinib, vandetanib	125, 130
UGT1A10	Flavonoids with C6 or C7 hydroxyl group on A ring, OTS167	Emodin	131, 132
UGT2B7	Zidovudine, epirubicin, morphine	Ketoconazole	133

Different populations (Chinese vs Japanese) have been found to metabolize quercetin and luteolin differently, suggesting potential for effect of genetic polymorphism on glucuronidation rates.123 But establishing a clear difference is not always straightforward as exemplified by the anticancer drug flavopiridol. Clinical studies for this molecule showed variations in metabolism in different populations, but the study of most common polymorphic forms of UGT1A9 did not correlate with the clinical data.136 Age has been found to be yet another factor affecting the glucuronidation capacity and hence the ability of different individuals to metabolize flavonoids and related compounds.137 Importance of efflux pumps such as P-glycoprotein in the transport and metabolism of flavonoids was shown by Lan et al.138 The development of robust SAR for UGT substrates is generally hampered by factors, such as (1) insufficient dataset for SAR or QSAR studies (not enough structural varieties); (2) lack of high-resolution 3D structure of UGTs, (3) difficulties in isolating functional UGTs from cells and thus estimation of relative levels via methods such as immunosorbant assays, and (4) inter-laboratory variations.

Past attempts to classify UGT substrates based on simple chemical characteristics such as functional groups have not been very successful. This is mostly due to the promiscuous nature of the UGTs such as the CYP450s. Thus, modern methods of *in silico* drug metabolism

prediction have been applied to UGTs. They can be broadly classified into (1) 2D, 3D QSAR, and pharmacophore studies for the substrate/inhibitor prediction, (2) regioselective prediction of SOG, and (3) hybrid models for substrate/kinetic parameter prediction.

2D, 3D QSAR, and Pharmacophore Studies for the Substrate Predictions

Dong and Wu have developed a QSAR for the prediction of UGT1A10 substrate selectivity using VolSurf descriptors and PLS analysis.[139] Thirty two compounds representing different chemical classes were considered for building models, while 11 representative compounds were kept in the test set. Descriptors were generated by considering different probe atoms (e.g., O and water) for modeling the interactions with the target. K_m values for these compounds were collected from different literature sources. The initially developed models were improved by performing a fractional factorial design (FFD) procedure to reduce the number of descriptor to 33. The authors found the models useful in categorizing substrate as either good or poor. The models could make predictions within one log unit of the experimental values, which can be considered as a good predictive power. Shape, size, and hydrophobicity descriptors were found crucial for the categorization of substrates and estimation of K_m values. The descriptors encoding the hydrophilicity of the molecules were also found important, but the magnitude of their contribution was not clear from the analysis since three showed positive and one showed a negative value.

A similar, pharmacophore-3D QSAR and VolSurf-3D QSAR approach was employed by Ako et al. for the prediction of K_m values for UGT2B7 substrates.[140] Modeling UGT2B7 substrate selectivity is inherently difficult, especially in the absence of crystallographic data, as it can accept both phenolic, alcoholic hydroxyl (e.g., estrogen and lipids),[141] carboxylic acid (e.g., ibuprofen),[142] and amino (e.g., carbamazepine) groups as SOG. For 53 substrates (36 in training and 17 in test set), experimental binding affinities were measured using recombinant enzyme systems. The best pharmacophoric model included a glucuronidation site (for compounds containing OH, COOH, primary, and secondary amines), three hydrophobic regions, and one hydrogen bond acceptor (HBA). This model gave an accuracy of 82% for the prediction of K_m values within one log unit. Importance of steric interactions with the receptor and the possibility of multiple binding modes are ignored in such pharmacophore models. Another model was built using VolSurf-3D QSAR methodology. In this model, descriptors derived from VolSurf are mapped on the 2D structure of the compounds and thus 3D alignment of the substrates is not required. Performance of this model increased after application of the FFD and PLS methods. Only two principal components were required to categorize substrates as good (K_m < 100 μM) or poor (K_m > 100 μM). Log *P*, polarizability, and molecular weight were found to correlate positively, while diffusivity and fractional hydrophilic molecular area were found to have negative

relationship with the activity.

The same group have recently developed very similar VolSurf-3D QSAR and Pharmacophore-3D QSAR models for estimating clearance of UGT1A3 substrates.[143] This study used a dataset of 34 compounds in the training set and 16 compounds in the test set for which clearance data were collected from the literature reports on recombinant enzyme. Using a very similar methodology, they found that the test set prediction accuracies for 81% of the UGT1A3 substrates prediction were within one log unit for a four-log unit range. The pharmacophoric features contained a glucuronidation site and two HBA sites at 6.5 Å distance. The model was found useful in distinguishing good (buprenorphine) from poor substrates (lamotrigine). As expected, molecular size and shape were found to be important parameters as per the pharmacophore model, while molecular diffusivity was found additionally crucial in the VolSurf-based model. Interestingly, hydrophobic regions were found to have negative correlations with the activity.

Earlier, Sorich et al. have developed a multiple pharmacophore-based model for the substrate prediction of 12 UGT isoforms.[144] Such models try to account for the fact that many UGTs may have multiple binding sites as evident from their biphasic reaction kinetics and formation of multiple metabolites from polyfunctional substrates, e.g., flavonoids and aminoalcohols or aminophenols. Pharmacophore fingerprinting and pattern recognition techniques were used to identify pharmacophoric features associated with substrates and nonsubstrates (55–216 for different isoforms). Three models were generated using a flowchart similar to that shown in Figure 10. The models were evaluated for prediction accuracy on a test set consisting of approximately 30% of the dataset. Substrate likelihood was predicted for each isoform using linear equations involving scaled contributions from each pharmacophoric feature. Filtering the initially generated pharmacophores based on the presence of nucleophilic feature improved the performance of the model and increased their interpretability. All the three models were found to make predictions better than by chance and the third model involving GA, PLSDA, and nucleophilic feature gave the best average predictions (Model 1: 69%, Model 2: 71%, and Model 3: 72% for the test set, Table 5). Models for UGT1A6 and 1A7 suggested that there is probably only one binding site in these isoforms. Comparison with the 2D-SVM and 2D-PLSDA-based models developed earlier[145] showed that pharmacophore-PLSDA model performs inferior and better, respectively. Lower performance for other isoforms was interpreted to be the result of other factors such as electrostatic interactions playing an important role in determining substrate specificities. As expected, the pharmacophoric features for UGT2B4 substrates (majority being hydroxysteroids) were distinct from those for the UGT1A family of isoforms.

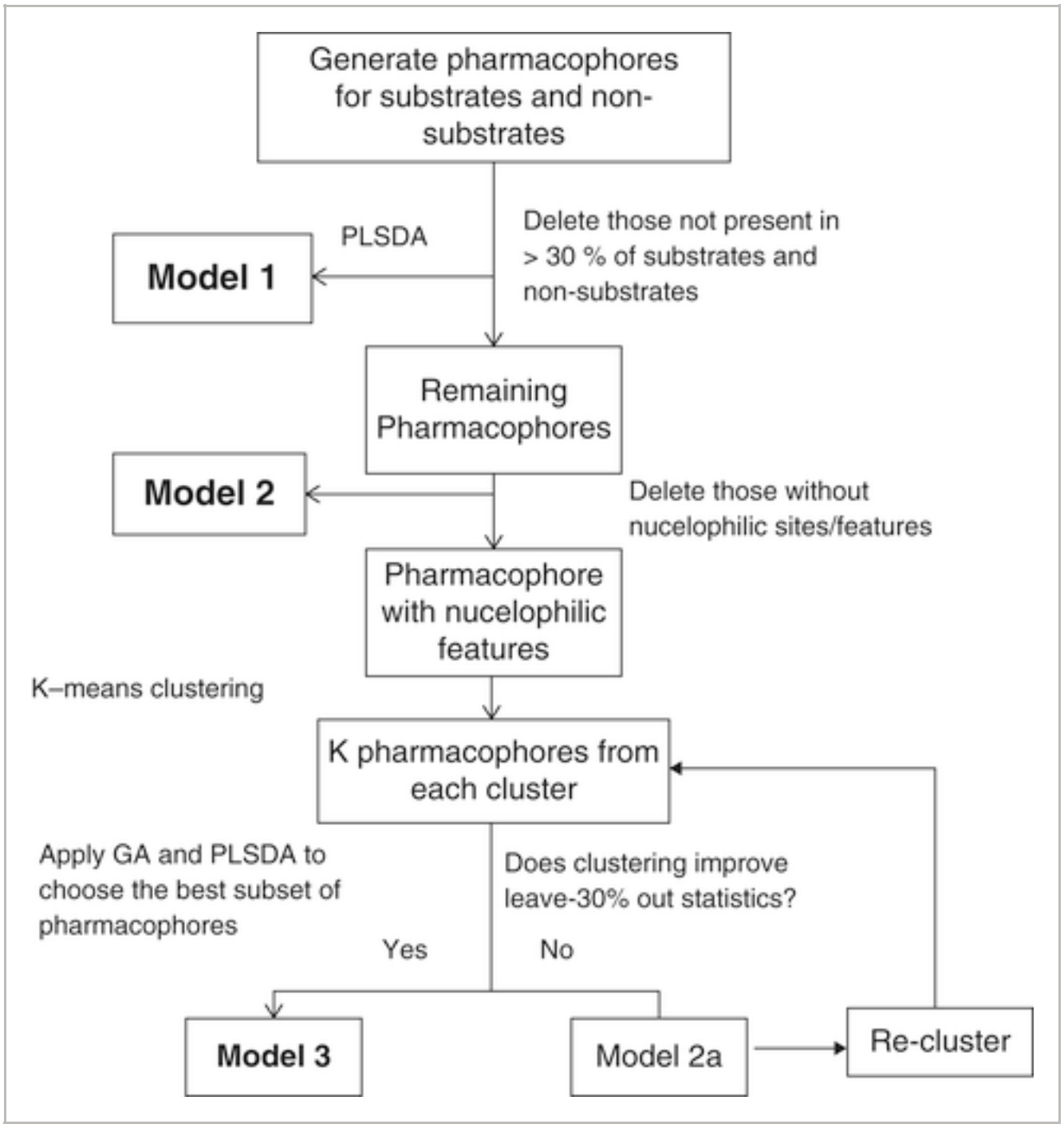

Figure 10

Open in figure viewer | PowerPoint

Flow-chart employed by Sorich et al. for the generation of multiple pharmacophores for the prediction of substrates and nonsubstrates of UGTs.

Caption

Table 5. Summary of the Best Results Obtained by Sorich et al. Using Multiple Pharmacophore Modeling for UGT Isoform Substrate Predictions

UGT Isoform	Pharmacophore	Description	Prediction Accuracy (%)		
			Substrate	Nonsubstrate	Test Set

1A6	P1 for substrate	A nucleophile approximately 2 Å from an aromatic ring	72	80	78
	P2 for substrate	Two hydrophobic regions (2 and 6 Å) away from nucleophile			
	P3 for nonsubstrate	Hydrophobic regions (6 and 8 Å) away from nucleophile)			
1A7	P1 for substrate	A nucelophile attached to aromatic ring	72	92	85
	P2 for substrate	HBA 2 Å away from nucleophile			
	P3 for nonsubstrate	HBD 8 Å away from nucleophile			
1A9	P1 substrates	Nucleophile attached to aromatic ring	83	74	89
	P2 for substrates	Aromatic ring 8 Å away from the nucleophile			
	P3 for nonsubstrates	Hydrophobic region and HBA 3 Å away from nucleophile			
2B4	P1 for substrates	A nucleophile within 2–4 Å of hydrophobic region	92	83	86
	P2	Nucleophile and HBD separated by 10 Å			

An earlier pharmacophore model developed by Michael et al. showed that UGT1A4 and UGT1A1 ligands have similar features.5 A typical ligand contains two hydrophobic features at a distance of 6–8 Å from each other and an HBA region (C4 carbonyl).

In the above-mentioned study pharmacophore models achieved reasonable prediction accuracy for only 4 of 12 UGT (1A6, 1A7, 1A9, and 2B4) isoforms. Thus, Sorich et al. developed a hybrid model consisting of QC descriptors (e.g., pK_a, molecular, and atomic nucleophilicity indices) derived from electronegativity equalization methods and SVM.146 Another set of models using the same dataset were developed by employing the 'Cluster Analysis–Genetic Algorithm–Partial Least Squares Discriminant Analysis' (Cluster-GA-PLSDA) since this method is known to generate linear models with small number of descriptors. The QC-SVM methodology gave average predictions of 78%, a marginal improvement compared

to the earlier study. The results further showed that in addition to the mechanistically anticipated nucleophilicity descriptors, those representing the molecular recognition are equally important (Table 6). A combination of 2D and QC descriptors with SVM modeling was attempted but was found not to result in a significant improvement in the predictive ability. Thus, a consensus approach was applied, which labeled the class as 'uncertain' if the 2D and QC predictions did not agree. This consensus model gave an average accuracy of 84% across. Lack of sufficiently large datasets (≥500) and noise were cited as the major reasons for misclassifications of the remaining compounds especially with UGT2B7.

Table 6. Summary of the Sorich et al.'s Results Obtained Using 2D and QC Descriptors Along with SVM and Cluster-GA-PLSDA Methodology

UGT Isoform	Test Set Prediction Accuracy			
	SVM-2D	SVM-QC	Cluster-GA-PLSDA	Consensus
1A1	64	85	68	88
1A3	84	89	71	90
1A4	88	83	85	90
1A6	78	67	71	83
1A7	74	79	74	81
1A8	68	77	61	77
1A9	82	80	64	87
1A10	80	80	70	86
2B4	88	83	83	88
2B7	71	64	71	73
2B15	64	67	72	70
2B17	73	80	80	90
Average	76	78	73	84

Regioselective Prediction of UGT-Mediated Drug Metabolism (SOG)

Regioselective prediction of SOG was made by Peng et al. recently.[147] They have built four classification models using 7 global and 25 physicochemical descriptors derived quantum chemically. Geometrical and lipophilicity descriptors were calculated separately using CODESSA and Open Babel,[148] respectively. Phase II glucuronidation reactions were retrieved from MDL metabolite database and SOG were identified by matching the SMARTS patterns. A subgraph isomorphism algorithm was applied to identify the SOG from the graph representations of substrate and their known metabolite structures. SOGs were further divided into glucuronidation sites for aliphatic and aromatic hydroxyl groups, carboxyl and nitrogen. A GA-based feature and SVM-classifier selection procedure involving 10-fold cross-validation was used. A preliminary analysis of the literature data showed that different nucleophilic sites showed different tendency to undergo glucuronidations. Balanced accuracy, MCC, and area under the ROC curve were calculated as measures of model's predictive abilities. The AUC for ROC curve were 0.89, 0.88, 0.90, and 0.96 for aliphatic, aromatic, carboxyl, and nitrogen glucuronidations in the training set. For the test set of 54 external compounds, the model showed an overall accuracy of 87%. In general, small lipophilic compounds were found to be more susceptible to glucuronidation by UGTs. Descriptors representing bond strength and local reactivity (Fukui function) were found important after GA-based feature selection. Predictions were incorrect mostly for molecules containing multiple potential SOGs and unaccounted steric factors were proposed to be the reasons for these inaccurate predictions. The results agreed with earlier reports by Sorch et al. of relative tendency of different functional groups to undergo glucuronidation. For example, nucleophilic groups attached to aromatic ring are glucuronidated with higher frequency than those attached to aliphatic groups. The authors found that developing models specific for each isoform is not practical as sufficient dataset is unavailable for the development of statistically valid models and such models would not be able to predict the overall metabolic profile with respect to UGT-mediated metabolism. Instead, models based on different properties of functional groups were constructed using QC descriptors.

Rudik et al. have developed a ligand-based ‘site of metabolism prediction’ (SOMP) tool for the prediction of SOG for five major CYP450s and UGT-catalyzed metabolic reactions.[149] For model development, experimental SOG data collected from Biovia (Accelrys) Metabolite database were used to label atoms as positives, while other atoms within the molecules were labeled as negatives. These were used to generate structures with one labeled atom (SoLA) and labeled multilevel neighborhoods of atom (LMNA) descriptors for training the model (592 molecules). SOMP requires an SDF, smiles, or MDL formatted input and gives a ranked probability of all the heavy atoms to be SOM. An approach similar to Rudik et al.[150] including Bayesian methodology and LMNA were used. Probabilities of SoLA to be the actual SOM (Pt) and false SOM (Pf) were calculated and the difference of these two (ΔP) was used

to calculate the average IAP (invariant accuracy of prediction) by leave-one-out procedure was used to estimate the accuracy. IAP is metric similar to ROC AUC, and has an advantage (over Top-N metric) of being independent of the size of the molecule. The model showed 95% accurate prediction in Top-3 and IAP metric for UGT-mediated metabolism. A major limitation of this methodology is that it cannot distinguish between enantiomers which could potentially have different metabolic profiles. SOMP has been implemented as a server-based online application and displays numbered molecules, along with a table of atom numbers ranked as per Δ*P* values for CYPs and UGT.

XenoSite UGT is a regioselective model for the prediction of UGT-mediated xenobiotic metabolism. In this study, Dang et al. have employed 2989 unique substrates involving 4557 SOGs derived from the AMD metabolic reaction database.151 XenoSite methodology is different from the one discussed above, as it makes predictions for all atoms within a given molecule and ranks these sites according to the potential for being SOG. SOGs were further divided into typical (similar to Peng et al.)147 and atypical sites/groups based on the frequency of occurrence in the database. This database shows that potential SOG are metabolized 80.0, 76.3, 49.6, 8.5, and 0.15% of the times for carboxylic acids, aromatic, and aliphatic hydroxyls, nitrogen and atypical (e.g., carbon). Even though the atypical sites were rarely observed, their number was sufficient for development of acceptable models. Three external sets (containing 141, 54, and 20 substrates) were used for the validation of the models. A binary heuristic model based on the statistics of the database was developed to serve as a baseline. In this model, all molecules with at least one SOG got a scaled score of 1, else they got a 0 score. In more complex models, numerical vectors derived from topological, molecular, and QC descriptors and additional descriptors representing the number of aliphatic, aromatic hydroxyls, carboxyl, and amino groups were used. Models were developed using neural network and logistic regressor with appropriate weighing factors (to account for rare elements) and cross-validation procedures. Top-2 and area under the ROC curve were chosen as performance metrics. The Top-2 metric is suitable for ranking sites within a molecule, while the ROC metric tests a method's ability to identify true positives and negatives. For the training set accuracies, the heuristic and neural-network-based models had comparable performance (91 and 97%, respectively), but the latter was required for better predictions of atypical SOG. All the UGT models performed about 10% better than the Xenosite P450 models, suggesting that modeling UGT metabolism is less complicated. Descriptor importance was assessed by randomly permuting/exchanging data columns and noting the drop in the performance of the models generated after the permutation. This procedure identified two-bond topological, number of hydrogens bonded to an atom, heavy atoms, size of the ring, and heuristic score as the most important descriptors. Interestingly, QC descriptors were not found important in the accurate

prediction. Accuracy trends on external datasets were similar to the training set (86% in Top-1 metric) and neural network model was better for atypical SOG. Comparisons with SOMP and XenoSite UGT model showed that the differences in performance are not statistically significant (Table 7). Peng et al.'s model is slower than SOMP and Xenosite UGT which required around 1 second per molecule.

Table 7. Comparison of XenoSite UGT, Peng Et al.'s Model and SOMP on a Set of 49 Compounds Not Correctly Predicted by the Heuristic Model

Metric/Regioselectivity Model	XenoSite UGT	Peng Et al.147	SOMP
Top-3	98	64	80
Average AUC[1]	79	51	75

1 Includes AUC values for aliphatic, aromatic, carboxylic, nitrogen, and atypical SOGs.

Hybrid Models for the Prediction of UGT Substrates/Inhibitors

A hybrid methodology including homology modeling, docking, SVM, and CoMSIA were recently developed by Ghemtio et al. for the modeling of UGT1A6–ligand interactions.129 In this study, the inhibitory potential of the test ligands against the glucuronidation of 1-naphthol by UGT1A6 was tested and modeled using glucuronidation rates. A dataset of 46 compounds (37 planar and 9 almost planar) was used to build and test the models. The training set (n = 31) and test set (n = 15) were generated by randomly distributing the compounds in these two sets. PaDEL descriptors (905) generated for these compounds were filtered by excluding descriptors with >80% zero values and having standard deviations of <3% giving a final set of 712 descriptors. A homology model for UGT1A6 was built using multiple sequence alignment, UDP-glucose coordinates, and Profiles-3D protocol.152 Molecular docking was performed by applying distance constrained between ligand phenolic-OH groups and HIS12 and UDP-glucose. Only for 39 compounds docked poses could be generated. Five poses per compounds were used for CoMSIA alignment and modeling MIFs. Known SOG for 10 compounds were used as validation criteria for nine different alignment methods chosen. SVM models were developed to classify compounds as actives (% probe glucuronidation > 50) and inactives (<50%) and classified 12 test compounds correctly. The results showed that binding and inhibition of probe (1-naphthol) glucuronidation is not a simple function of presence or absence of OH or COOH groups, but

it also depends on the details of the UGT–ligand interactions. Good correlations were found between Hammett's σ meta/para constants and inhibitory activity for a subset of 2-hydroxybenzoic acid derivatives. An increase in the size of substituent increased the inhibitory activity, but this might not reflect the substrate glucuronidation efficiency. An analysis of the descriptor importance showed that electronic, constitutional, and topological parameters are sufficient to explain the activity.

Sulfotransferases

SULTs are the second most important Phase II drug-metabolizing enzymes. They also play an important role in the regulation of activity of many hormones (thyroxine and dopamine) and phenolic compounds. SULTs transfer a sulfate groups to mostly compounds containing hydroxyl or amino groups. This transfer occurs via an S_N2-like mechanism and requires the presence of 3′-phosphoadenosine 5′-phosphosulfate (PAPS) as a cofactor (Figure 11).[153, 154] Recent studies have suggested that binding of PAPS in the active sites of dimers regulates the affinity and selectivity of SULT1 class of enzymes.[155] Forty six crystal structures for human SULTs are known till date. These belong to either of the three isoform types, namely SULT1 (most abundant in liver), SULT2, and SULT4. Study of these crystal structures has established that SULT family of enzymes has very flexible active site and can accommodate more than one ligand in different orientations.[156] This flexibility and plasticity is responsible for the promiscuous nature of SULTs.[154] The differences in the gene expression are usually noticed at the 5′ promoter and intron sequences and 1A1 and 1A2 show 93% sequence similarity and 1A3 comparatively shows 60% similarity. For majority of the drugs, e.g., paracetamol and dopamine, sulfonation leads to the formation of hydrophilic metabolites which are excreted safely via urine. But for some molecules (e.g., minoxidil),[153] efficacy is dependent on the sulfonation step and in other cases it can lead to RM formation and toxicity (e.g., troglitazone).[157-159] Since the recognition of the importance of non-CYP450 metabolic enzymes in drug metabolism and toxicity, the research interest in SULTs has been increasing.[1] As seen in Table 2, the number of papers containing the concept of ‘drug metabolism’ and ‘sulfotransferase’ has doubled at the turn of the 21st century.

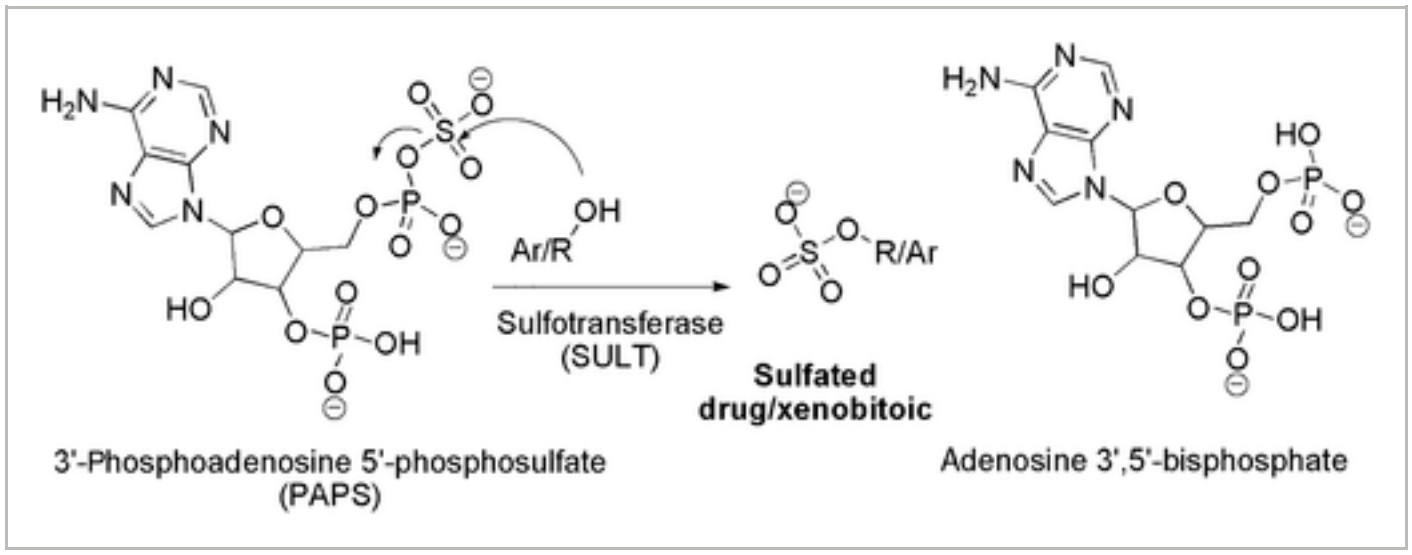

Figure 11

Open in figure viewer | PowerPoint

S_N2 type mechanism of sulfotransferase reaction with hydroxyl group containing drugs and xenobiotics.

Caption ⌄

Even then prediction (regioselectivity, isoform specificity, or kinetic parameters) for sulfonation reactions catalyzed by SULTs has received limited attention. In the following section, we discuss currently available structure-based models for the prediction of substrates, attempts toward understanding regioselectivity of SULT-catalyzed reactions and QSAR studies.

Structure-Based Substrate Prediction for SULTs

Martiny et al. have developed a protocol for the estimation of SULT–ligand interactions using a combination of MD, docking, and QSAR studies. MDs simulations on three isoforms, namely SULT1A1, SULT1A3, and SULT1E, with bound cofactor PAP and without ligands were performed to study the flexibility and the plasticity using potential energy and the solvent-accessible surface area values. Of the 4500 structure for each isoform, hierarchical ascendant classification was used to reduce the conformations to 11 for SULT1A1, 7 for SULT1A3, and 7 for SULT1E1. The best receptor conformations of the isoforms were identified by performing virtual screening (VS) using docking method. Twenty-four VS for SULT1A1, 16 VS each for SULT1A3, SULT1E1 conformations suggested that two MD-generated conformations of SULT1A1 performed better than its X-ray structure. The performance of conformations was low in case of SULT1A3, while the X-ray and conformations of SULT1E1 show similar results. This resulted in better classification of the substrates for SULT1A1 and SULT1E1 than for SULT1A3. Conformations which could discriminate binders from nonbinders also had high druggability score (estimated using

DogSiteScorer).

Furthermore, two QSAR models, using extended connectivity fingerprint descriptors dependent upon the topological information of compound structures and the second using the binding energies of MD conformations were developed. SVM, random forest, and naïve Bayes methodologies were used to train classifiers using the model information. The performance of SULT1E1 and SULT1A3 QSAR models was high compared to SULT1A1 which might be due to the high flexibility and promiscuity, giving high chemical diversity to this isoform. This approach might be useful in the prediction of ligand binding to the three isoforms and *in silico* toxicity predictions of the small molecules which interact with SULTs.

This method was earlier followed by Moroy et al. for determining the receptor flexibility of SULT1A1 using its rigid X-ray structure cocrystallized with cofactors, PAP and *p*-nitrophenol.156 They performed three MD simulations in CHARMM program and 4500 conformations extracted were clustered using HAC before performing VS using Vina 1.0. BRENDA and Aureus Sciences databases were referred to identify 157 substrates for SULT1A1. Maybridge, Chembridge, PremiumSet, and HitFinder datasets were used for the decoys, which along with the active sites were filtered using FAF-Drugs 2 server. X-ray structure and MD conformation both performed reasonably well at predicting substrates from Maybride, but screening of the complete dataset was required to retrieve maximum hits, i.e., enrichment was lower.

Cook et al. have performed a thorough study to explore substrate specificity for the SULT1A1 using MDs and proposed a molecular clamp mechanism which explains the increase in the enzyme activity toward positive synergy substrates.155 These studies helped in the classification of the substrates into two classes: one class for which catalytic efficiency (K_{cat}/K_m) is affected by the presence of PAPS and those unaffected by it. Reaction rates and K_m values were determined by fluorescence titration. GOLD was used for generating ligand poses which were used to perform MD using GROMACS with Amber force field. The benzyl groups of active site residues F81 and F84 sandwiched the substrate ring forming a molecular clamp. This rearrangement brings the nucleophilic hydroxyl at a position where both activation by general base H108 and the attack by sulfuryl group are possible. The MD results showed that for the substrates with negative synergy (estradiol), the nucleophilic site was close to the substrate for about 2% of the time during the 1 ns simulation, while the substrate with positive synergy (acetaminophen) was almost stationary near the nucleophilic site.

Rakers et al. developed a method by initially using MD simulations to find out enzyme flexibility and protein conformation.160 Molecular docking simulations were clubbed with

collection of active SULT1E1 ligands and clusters were generated which served as the basis for 3D pharmacophore development. Later, individual ligand–protein complex was then evaluated qualitatively for pharmacophore generation. Eight 3D pharmacophores were then prepared that represents different modes of binding between enzymes and substrates or inhibitors. These eight pharmacophores were then validated using set of active ligands and 2829 decoys and showed overall sensitivity of 62% and individual sensitivities 56% (inhibitors), 33% (substrates), and 80% (CLDs). For further improving the ligand identification, additionally machine learning models were used like SVM for substrate and inhibitor classification. SVM model for substrates correctly predicted 91% of test set molecules and MCC was found to be 0.38 along with sensitivity of 0.87 and specificity of 0.91, while for inhibitor model sensitivity was 1, specificity of 0.84, and prediction accuracy was 85% with MCC of 0.32.

Studies for Understanding the Regioselective Sulfonation by SULTs

Iyer et al. worked on the regioselective and stereospecific phenolic sulfonation of 4′-methoxyfenoterol (MF), a potential candidate for the treatment of congestive heart failure.161 Experimental results showed that sulfonation of MF is stereoselective. Sulfonation rate was fastest for the *S*,*R* isomer, followed by *S*,*S* and *R*,*S* (*R*,*R*) isomers. Sulfonation of *R*,*S* isomer was inhibited by the *S*,*R* isomer, thus having potential to increase its bioavailability. Molecular docking with Molegro virtual docker using crystal structures of SULT1A3 (1CJM, 2A3R); and SULT1A1 (1LS6), respectively, was performed to explain the observed stereoselectivity. The change in the chirality from *S* to *R* at the β-carbon atom, changes the hydroxyl group orientation which enables additional H-bond formation with the site Glu146. The results for SULT1A3 showed electrostatic interaction of the 3-OH group of MF with PAPS and Lys106 and His108. MolDockScore values of the four stereoisomers of MF suggested that the binding affinity was highest for *S*,*R* isomer than *R*,*S* isomer. Thus, it was concluded that *S*,*R* isomer might serve as an inhibitor of the Phase II metabolism of the more active *R*,*S* isomer and thus increases the bioavailability.

In 2016, Ning et al. determined the regioselective and stereoselective sulfation of six bufadienolide derivatives (bufalin, resibufogenin, cinobufagin, bufotalin, telocinobufagin, and deacetylcinobufagin) extracted from the skin secretions of the amphibians using experiments and molecular docking simulations.162 Initially, experiments with recombinant SULT isoforms were conducted which showed that SULT2A1 isoform is responsible for the metabolism to 3-*O*-sulfates. Molecular docking using Surflex-Dock on the crystal structure (1J99) revealed the possible interactions between the bufadienolides derivatives and SULT2A1 active site. His99 residue showed H-bonding interactions with the dehydroepiandrosterone (DHEA) and the bufadienolide derivatives. The affinities of these

molecules toward SULT2A1 were determined using chemscore and were correlated with K_m values. The study concluded that SULT2A1 enzyme is stereoselective toward substrates which have β-hydroxyl group at C3. Hydroxyl groups at other carbon atoms such as C11, C14, and C16 positions reduce the affinity, whereas in compounds with the C5 position hydroxyl group (due to Phase I reaction) the affinity is increased.

QSAR Studies for SULT Substrate Predictions

In 2002, Sharma et al. worked on rat aryl sulfotransferase (AST IV) which is homologous to human SULT1A1 and developed a comparative molecular field analysis (CoMFA) model which predicted the catalytic efficiency expressed as log (K_{cat}/K_m). A homologous model of AST IV was generated from mouse estrogen sulfotransferase (mEST, 1BO6) by using SYBL 6.5. Thirty five substrates were used for the development of CoMFA models with and without the active site information (docking, Figure 12). For the CoMFA model based on the active site information, the positions in the active site residues were fixed in addition to a constrained OH group in a transition-state-like structure. The ligand structures were minimized in the presence of the constrained protein and then extracted for alignment. The cross-validated q^2 value (0.007) for CoMFA model without active site information showed that interactions of the substrates within the active site are important for modeling activity. The q^2 value for the conformations based on active site model was 0.691. Predictions for 4-pentylbenzyl alcohol were in good agreement with the experiments, while benzyl alcohol showed the largest deviation. The analysis showed that the experimental or calculated values do not correlate with the type of the functional group (hydroxyl benzylic or aromatic or *N*-hydroxy arlyamino). The external validation of the generated CoMFA model on 4-(1-adamantyl)phenol, 4-biphenylmethanol, and 3-indolemethanol showed that the experimental kinetic constants were within 0.3–0.8 log units of the predicted values.

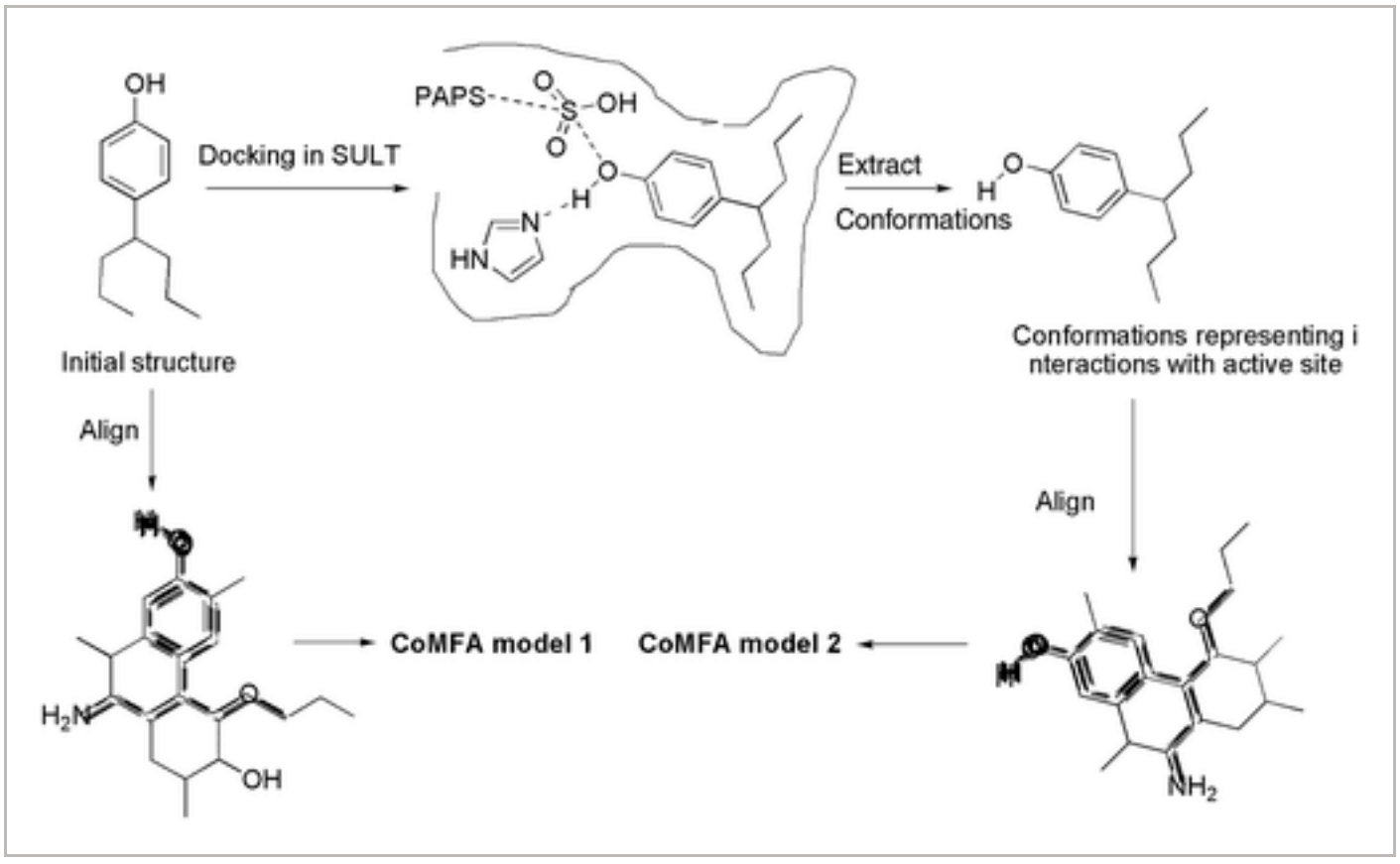

Figure 12

Open in figure viewer | PowerPoint

Schematic representation of the protocol employed by Sharma et al. for generating CoMFA models for prediction of SULT substrates.

Caption

These authors then applied a similar strategy for substrate prediction on SULT2A3.163 The homology model rSULT2A3 was prepared using coordinates of crystal structure of hSULT2A1 (human hydroxysteroid sulfotransferase). PAPS from PAP-vanate-estrogen sulfotransferase complex and DHEA were extracted from crystal structure (3DHE) and were attached to the active site of rSULT2A3. The resulting energy minimized rSULT2A3 model had high degree of structural similarity with hSULT2A1. CoMFA models were developed using the schematic procedure shown in Figure 12. Leave-one-out cross-validation procedure gave a model with $q^2 = 0.664$ and r^2 value was 0.948. This model was then tested on external set of molecules for which $\log(K_{cat}/K_m)$ values were measured experimentally to check the predictive ability. For 2-methylcyclohexanol, predicted value matched the experimental value, while for other molecules predictions were within 0.01–0.3 log units. The final CoMFA model for all the molecules was developed and showed r^2 value of 0.972 and $q^2 = 0.719$. Table 8 shows the theoretical and predicted value of five compounds.

Table 8. Experimentally Determined and Catalytic Efficiencies Predicted Using CoMFA

Compounds	Experimental	Predicted	q^2
SULT1A1			
4-Propylphenol	6.38	6.43	0.691
4-Nonylphenol	4.45	4.40	
2-Bromo-*N*-(4-hydroxyphenyl)acetamide	5.17	5.18	
4-Pentylbenzyl alcohol	5.27	5.26	
L-Tyrosine methyl ester	3.47	3.44	
rSULT2A3			
Hexan-1-ol	2.73	2.43	0.719
(4-Propylphenyl)methanol	3.79	3.78	
Cis-4-Methylcyclohexylmethanol	2.61	1.86	
1-Methylcyclohexanol	1.58	1.58	
(1*R*)-1-(1-Napthyl)ethanol	3.33	3.63	

Values of K_{cat}/k_m are expressed in min^{-1} M^{-1}.

Earlier, Dajani et al attempted to derive a relationship between the properties of various substrates analogues and kinetics (K_m) of the SULT1A3.164 Earlier literature studies suggested that SULT1A3 catalyzed the conjugation of phenols and catechols, thus 39 different substituted phenols were used in this study to predict experimental K_m values. The properties such as partition coefficient (*c*log *P*); calculated log *D*; calculated molar refraction (CMR); calculated molar volume (V_m); flexibility (ratio of total rotatable bonds and total nonrotatable bonds); aromatic rings (ArRngs); number of hydrogen bond donor groups (HBD); HBA; and ionization potential (IP) were determined. Two linear regression models were developed using these parameters and were in excellent agreement with the observed K_m data. Equation (1) depicted an inverse relationship between the K_m and hydrogen bonding groups plus hydrophobicity (determined by the three conserved Phe residues). It also indicated that the substrate specificity is favored toward narrow substituted groups in the para position of the catechols. Equation (2) predicted lower K_m values for molecules with lower V_m and recognized *c*log*D* as an important parameter in determining the affinity of the

drugs toward SULT1A3. Based on results, the authors concluded that the enzyme prefers catalyzing the catechols, especially catecholamines, more compared to phenols and these conclusions were supported by the QSAR equations. The SULT1A1 isoform showed lower K_m values for dopamine and tyramine, whereas SULT1A3 showed decreased K_m values for other four-substituted catechols. Molecular modeling predicted that this decrease is due to the hydrogen bonding between the amino acid residue Glu146 and the amine nitrogen present at the fourth position.

$$\begin{aligned}\mathrm{Log}(1/K_m) = {} & 0.82(\pm 0.32) + 0.27(\pm 0.07)\mathrm{HBA} \\ & + 1.95(\pm 0.16)\mathrm{HBD} + 0.58(\pm 0.06)c\mathrm{Log}D \\ & - 0.31(\pm 0.09)Y\end{aligned} \tag{1}$$

$$\begin{aligned}\mathrm{Log}(1/K_m) = {} & 3.68(\pm 0.28) + 1.87(\pm 0.18)\mathrm{Icat} \\ & + 3.89(\pm 0.67)\mathrm{Iami} + 0.61(\pm 0.12)c\mathrm{Log}D \\ & - 0.014(\pm 0.004)V_m\end{aligned} \tag{2}$$

As seen in Figure 2 and Table 1, UGTs and SULTs make a total of ~60% contribution to Phase II metabolism of marketed drugs. But currently available models cannot distinguish between the substrates for these two enzymes, i.e., identifying drugs that are glucuronidated and those preferentially sulfonated at the hydroxyl groups. Additionally for molecules with multiple functional groups, current methods cannot predict SOG and SOS.

Glutathione *S*-Transferase

GST constitutes a family of important Phase II metabolic enzymes, which catalyze the transfer of glutathione group to a variety of electrophilic endogenous and xenobiotic substrates. They also play a critical role in the control of the cellular redox processes.165 GSTs are known to metabolize approximately 12% of the marketed drugs (Figure 2 and Table 1). The human GST family of enzymes consists of eight major isoforms, four of these [GST-α (GSTA), GST-μ (GSTM), GST-π (GSTP), GST-θ (GSTT)] have been implicated in human drug metabolism.166 X-ray structure for most of the GSTs shows that they are homodimeric proteins, consisting of α/β domains (Figure 13).167 The α-domain contains a glutathione binding site and the interface between the two domains constitutes the substrate binding site. Residues lining this interface are mostly responsible for the substrate specificity. These enzymes, similar to other metabolic enzyme systems, have overlapping substrate specificities, e.g., 1-chloro-2,4-dinitrobenzene is a substrate for most of them and is thus

used in spectrophotometric assays. Additional details and comparison about different crystal structures reported up till recently can be found in the review by Wu et al.[168] A typical mechanism for aromatic substrates involves the activation of GSH to thiolate ion by hydrogen bonding with selected residues in the active site.[169] This is followed by the attack of the thiolate ion on the electrophilic substrate to form a Meisenheimer complex, which loses the leaving group to form the glutathionated product. As the electrophilicity of the substrate decreases, the rate-determining step shift from the thiolate generation to the formation of the Meisenheimer complex.

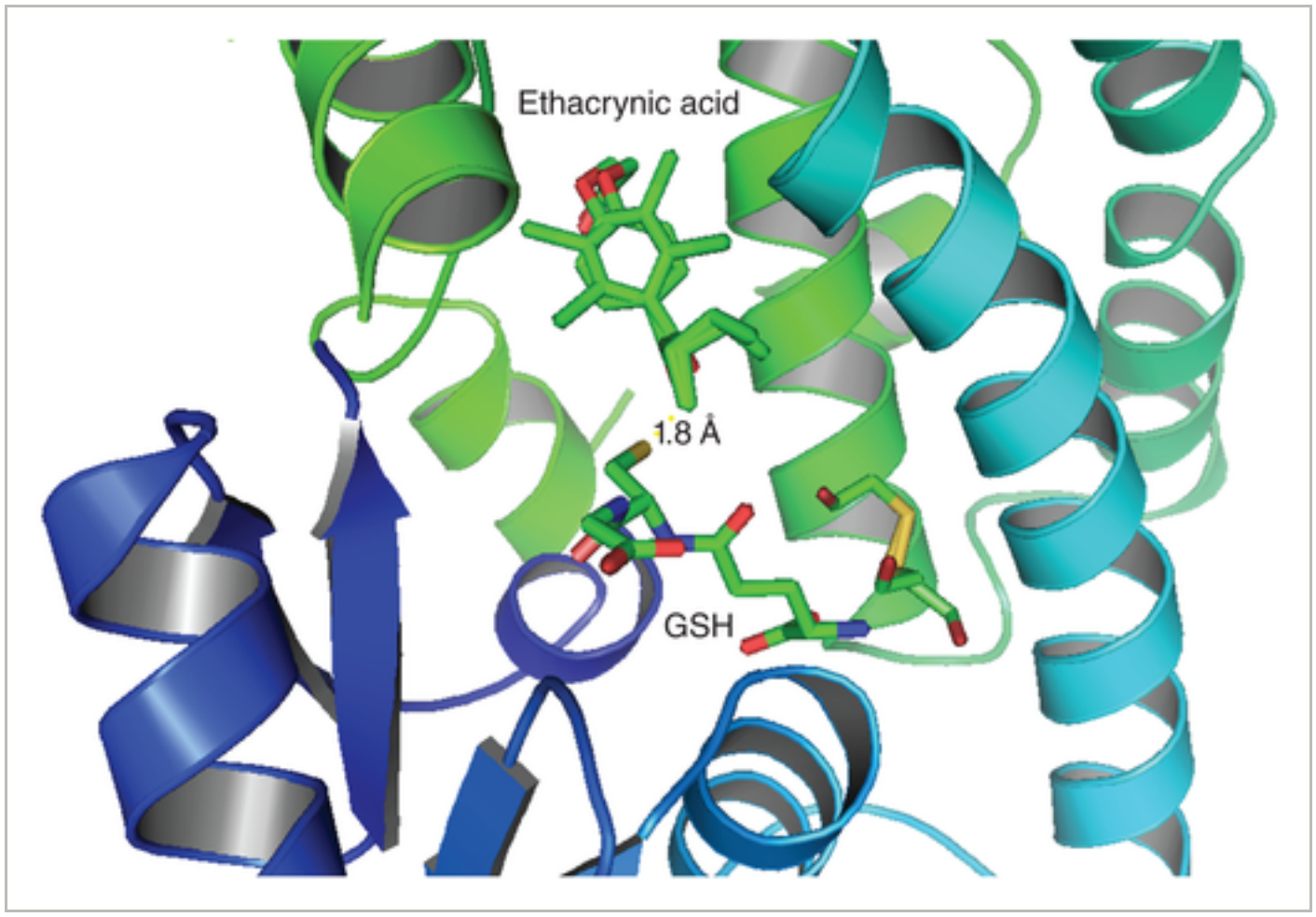

Figure 13

Open in figure viewer | PowerPoint

Crystal structure (1GSE) of ethacrynic acid conjugated to GSH bond in the GST binding site. Glutathione is covalently bonded to ethacrynic acid.

Caption ⌄

GSTs are mostly involved in the detoxification of xenobiotics and drugs, but cases of metabolic activation leading to RM formation have also been reported. Dihalo-substituted aliphatic and bis-electrophilic compounds are well-known examples of GST-mediated metabolic activation and interestingly haloforms are devoid of such effects.[170] Typical xenobiotic substrates are, e.g., chemotherapeutic agents (pro-drug of 6-mercaptopurine and azathioprine),[171] epoxides, α,β-unsaturated aldehydes, and polycyclic aromatic

hydrocarbons.172 Since these enzymes make a smaller contribution to drug metabolism than UGTs and SULTs, research interest (Table 2) and development of predictive models has lagged behind for GSTs. Nevertheless, a handful of QSAR and docking studies for identification of substrates or inhibitors have been reported in the literature and these are discussed in the following sections.

QSAR Studies on GST Substrates and Inhibitors

SAR between the energy of the lowest occupied molecular orbital (E_{LUMO}) and glutathinoation rates by microsomal GST (MGST) and cytosolic GST (CGST) was tested earlier by Jolivette and Anders for nephrotoxic haloalkenes.173 Use of purified fractions of MGST was found to be necessary to avoid confusion related to the contribution made by CGST enzymes as observed in some of the earlier studies with conflicting results. Metabolic reaction rates were generally smaller for the MGST than the CGST. E_{LUMO} calculated using semi-empirical QC methods correlated with reaction rates for the MGST enzyme but not for the CGST. Haloalkenes with E_{LUMO} lower than −0.73 eV were found to be good substrates for MGST, while those with E_{LUMO} higher than −0.44 eV were not metabolized by MGST. The authors highlighted that a smaller HOMO-LUMO gap gives an indication of the strength of the interaction between the donor (GSH) and acceptor (alkene). Interestingly, the rank order of the haloalkenes for both enzymes was different, thus suggesting that the same ligand-based parameter might not be suitable for prediction of activity at both enzymes and structural information about the enzyme can improve prediction.

Regioselectivity and QSAR were studied by Soffers et al. for the conjugation of fluoronitrobenzene derivatives by rat and human GST.174 The purpose of this study was to test, if QSAR's developed from chemical descriptors and log(reaction rates) for one species could be used to predict reaction rates for another one. E_{LUMO} and relative heats of formation (of Meisenheimer complex with MeS^-) for four compounds 2,4-difluoronitrobenzene, 2-fluoronitrobenzene, 2,4,6-trifluoronitrobenzene, and 2,3,4,6-tetrafluoronitrobenzene were calculated using AM1 semi-empirical QC method. Results showed that the SOGT (site of attack by GSH) was the ortho position of the fluoronitrobenzenes in most of the cytosolic GST catalyzed reactions. The V_{max} values increased, while the K_m values decreased as the number of fluorine atoms and log *P* values increased for a substrate. The V_{max} values had an excellent correlation with E_{LUMO} and $\Delta\Delta H$ values (r^2 = −0.97 to −0.99). The slopes of these regression lines were all almost parallel to each other, thus suggesting a common mechanism for these conjugation reaction. Thus, for this series of compounds the QSAR models can be used across the species for prediction of SOGT and kinetic parameters.

Earlier Van der Aar et al.[175] have performed a QSAR study based on rat liver GST catalyzed glutathionation of 1-chloro-4-nitrobenzene derivatives. For eleven 2-substituted 1-chloro-4-nitrobenzene (CNB) derivatives K_{cat}, K_m, and K_s (second-order rate constant) values were experimentally determined. These substrates are believed to undergo a second-order nucelophilic aromatic substitution (SNAr) reaction. Substituent properties (steric, electronic, lipophilic parameters, and Hammett constants), calculated molecular descriptors (charge, E_{LUMO}, and E_{HOMO}) of substrates and the Meisenheimer complex with methanethiolate anion were used to build models. The K_s values for base catalyzed reactions were found 100-fold lower than the GST catalyzed reactions. Liner regression analysis showed that K_m values did not correlated with the physicochemical parameters and calculated descriptors. Whereas, for the base catalyzed reactions, log K_s values correlated well with the Hammett constant, suggesting that electron withdrawing substituents increase the rate of the reaction. The slope of these plots suggested that the reaction mechanisms are similar for base catalyzed and GST 1–1 and 7–7 isoforms. Inhibition studies ruled out the possibility of the product release being a rate limiting factor. This, study lead to the finding that GST 3–3 and 4–4 isoform have a different mechanism (different rate determining step) for the formation of glutathione adducts. Intermediate involving radical–radical anion electron-transfer complexes were proposed. Predictions on the unsubstituted 1-chloro-4-nitrobenzene suggested it to be a substrate for only GST 3–3 isoform. These predictions were then confirmed by experiments.

The reaction of CNB with GST was utilized by Appiah-Opong et al. to develop a QSAR for curcumin analogues.[176] Thirty-four curcumin analogues were synthesized and their inhibitor potential was determined against purified GSTM, GSTA, GSTP, human, and rat liver cytosolic fractions. Compounds shown in Figure 14 had higher inhibitory activity than curcumin against the purified enzymes. Molecular descriptors generated by MOE were used in model generation and routine parameters ($r^2 = 0.78$ and $q^2 = 0.62$) were noted after removing an outlier from the dataset. Molecular surface area and contribution of Van der Waals surface area to log *P* were found as important descriptors for prediction of inhibitory potential.

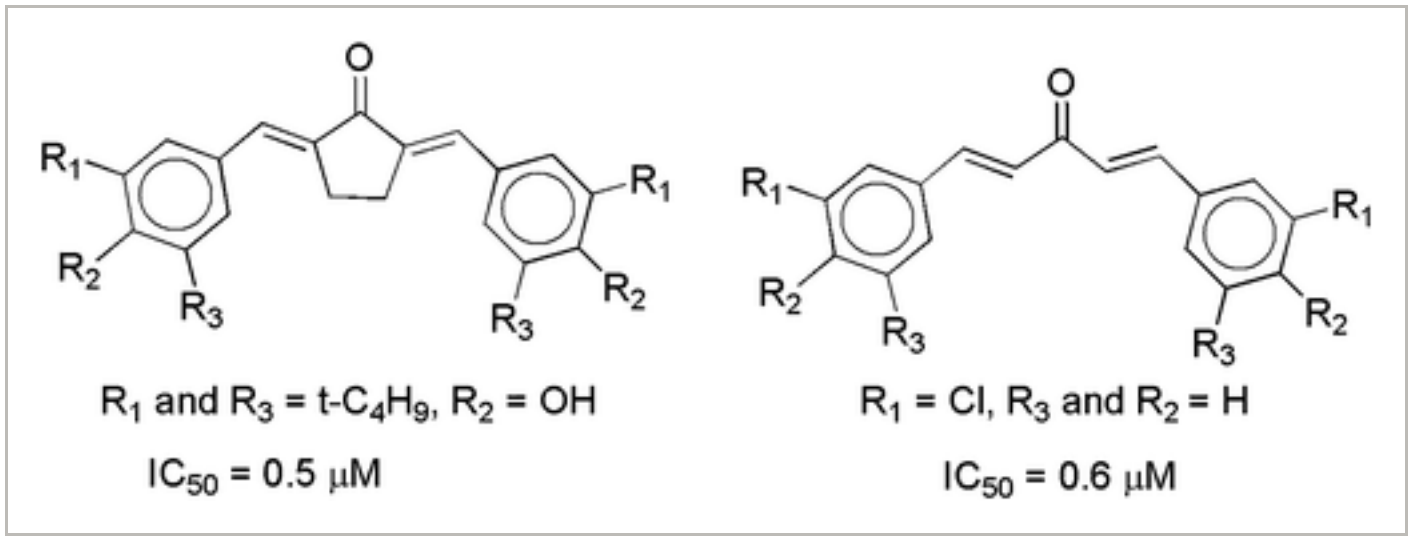

Figure 14

Open in figure viewer | PowerPoint

Curcumin analogues with inhibitory potential against purified GST enzymes.

Caption ⌄

Docking Studies for the Prediction of GST Substrates

Dong et al. have utilized a covalent docking approach for the prediction of GSH substrates.177 The advantage of covalent docking is that it decreases the conformational space to be sampled and thus leads to a better performance than the conventional docking approach. In this study, they have compared the accuracy of X-ray structures, homology and consensus models for 14 different GSH isoforms and 11 different metabolic reactions catalyzed by GSH. SMARTs pattern matching was used to find potential substrates (from KEGG, MetaCyc databases, and literature) from 6738 molecules. Product structures were generated using OECHEM library generation and Epik. Dimeric GST protein structures (extracted from PDB) with two GSH and substrate binding site were used for covalent docking. Homology models for each GST protein target were generated using MODELLER and the best model for each isoform was selected based on the lowest discrete optimized protein energy score. Each potential product was docked into the X-ray structure and homology models, keeping the original GSH coordinates and docking the remaining product atoms with OMEGA and POLP. All of the rotatable bonds corresponding to the substrate and CA-CB and CB-SG bonds in GSH were considered while generating conformations for the products. Docking in native structure and cross-docking experiments gave the native product conformation within 3 Å for majority of the GSTs (8 of 12 cases). Analysis of the docking results showed that correctly scoring ligands is more difficult than finding the native binding conformation. No correlations or weak correlations were found between the binding site RMSD error in the homology model and the RMSD error of the top dock pose, i.e., the best homology model was not always best at predicting native pose. The holo and apo

structure had little difference in performance. Residues close to GSH were postulated to play a major role in determining the docking accuracy.

Lannutti et al. have developed a cross-docking protocol for identification of GSTP substrates.178 Crystal structures of nine GST enzyme, their B factors, and active site atomic fluctuations were used to construct an ensemble of nine protein models representing flexibility. GSH conjugates were seen to stabilize the fluctuations (calculated using in house script) of the active site residues in comparison to the apo form. Glide based cross-docking with GSH conjugates was performed with these and X-ray crystal structures either with or without the crystallographic waters and standard precision (SP) and extra precision (XP) options were selected sequentially to estimate the binding scores. Nonhydrogen RMSD from the crystal structure coordinates of the ligands was used to evaluate the performance of the cross-docking experiment. Interaction frequencies were calculated by observing the interactions graphically using default settings in Maestro visualizer. Classical docking showed that for most of the ligands the native pose is correctly predicted (i.e., with RMSD < 2°, except for the GSH conjugate of 2,4-dinitrobenzene and ethacrynic acid). The performance of the cross-docking simulations was testing by using a docking accuracy function (DAF) defined as the sum of fraction of poses with RMSD <2 Å and the half of the fraction of poses with RMSD within 2–3 Å of the native pose (*Equation 3*). RMSD analysis showed that any of the GST crystal structure can produce a near native pose except for the GSH conjugates of benzyl bromide, ethacrynic acid and 2,4-dinitrobenzene. This was mostly ascribed to the lack of correct modeling of π-stacking interactions by the scoring function. Gln64, Leu52, Lys44, and Trp38 were mostly involved in hydrogen bonding interactions with GSH moiety of the conjugates, while hydrophobic interactions were spread over a large set of active site residues. Six crystallographic waters were selected in ligands binding by examining RMSD values for waters (oxygen atoms) in the aligned complex structures. Docking and cross-docking simulations led to marginal improvement in the performance assessed by DAF (Table 9). Two of the waters formed hydrogen bonding interactions with the negatively charged carboxylate of the GSH moiety in the conjugates and were considered important for prediction of poses. In comparison, GOLD docking experiments gave less accurate results and required additional hydrogen bonding constraints to be included for similar accuracy.

Table 9. DAF Values for Glide and GOLD-based Docking Performed by Lannutti Et al.

Protocol	Glide	GOLD
Without waters	0.88	0.96

With waters	0.86	0.64

Formula for docking accuracy function (DAF)

$$\mathrm{DAF} = f_{\mathrm{RMSD}<2\,\text{Å}} + \frac{(f_{\mathrm{RMSD}<3\,\text{Å}} - f_{\mathrm{RMSD}<2\,\text{Å}})}{2} \quad (3)$$

Methyltransferases

MTs are an important group of Phase II enzymes responsible for methylation of xenobiotics and drugs, i.e., incorporating methyl group using S-adenosylmethionine (SAM) as cofactor.179, 180 As shown in Table 2, a significant amount of research has been performed on MTs, mostly due to their role in the regulation of endogenous neurotransmitters amines and some drugs. But their overall contribution to drug metabolism is small when compared to other Phase II metabolic enzymes (Table 1 and Figure 2). MTs are classified based on the reactions (O-methylation, N-methylation, and S-methylation) catalyzed, i.e., catechol *O*-methyltransferase (COMT), histamine *N*-methyltransferase (HNMT), and thiopurine *S*-methyltransferase (TPMT), respectively. Crystal structures for human COMT (13), HNMT (2), and TPMT (2) are available in the protein databank. Most of the endogenous and some xenobiotic [3,4-dihydroxymandelic acid (DOMA), 3,4-dihydroxybenzoic acid (DOPAC), ascorbic acid, carbidopa, benserazide, fenoldopam, isoprenaline, and quercetin]181 substrates of MTs are polar and soluble in water. Methylation of these substrates would thus decrease the polarity (and thus water solubility) of the metabolites marginally (Figure 15). Thus, the main aim of methylation of these substrates seems to be termination of biological activity rather than increase in water solubility.

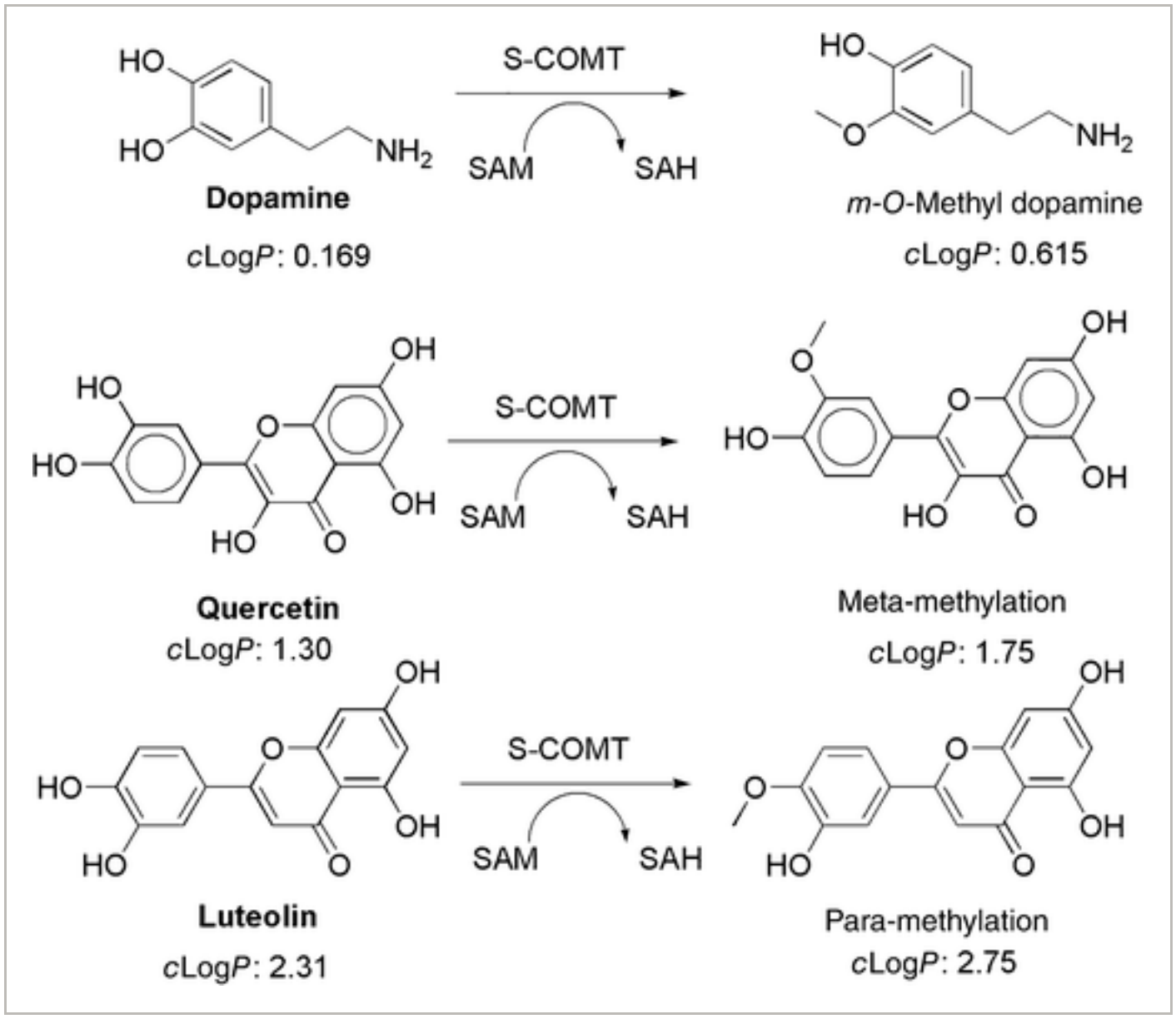

Figure 15

Open in figure viewer | PowerPoint

COMT catalyzed methylation of dopamine, quercetin, and luteolin. SAM, *S*-adenosyl-L-methionine; SAH, *S*-adenosylhomocysteine. *c*Log *P* values for the substrates and metabolites show that methylation decreases the polarity of substrates marginally.

Caption

COMT catalyzes the addition of methyl group from *S*-adenosylmethionine to catecholamines bound to Mg^{2+} in the active site. The reaction follows an S_N2-type mechanism where a Lys residue acts as a base.[182] The substrates of this enzyme include neurotransmitter such as norepinephrine and dopamine and also catechol containing compounds which are used in the treatment of diseases related to oxidative stress like PD, hypertension, and asthma.[183] The methylation reaction is known to be regioselective, with the meta position getting methylated preferentially over the para hydroxyl groups in many catecholamines.[184] It is an intracellular enzyme found in post-synaptic neuron and exists in two forms: soluble cytosolic COMT (S-COMT) and membrane-bound COMT (MB-COMT) present on cytoplasm side of rough endoplasmic reticulum (RER). These two forms are structurally different with respect

to the N-termini which in case of MB-COMT extended to 50 amino acids. The soluble form is comparatively expressed more. Flavanoids quercetins, entacapone, and tolcapone act as COMT inhibitors and are administered along with L-dopa. Genetic polymorphism has been discovered in the COMT gene at the 158 and 108 codon positions for MB-COMT and S-COMT, respectively. At these sites 'G' is replaced with 'A,' expressing valine instead of methionine, this substitution results in decrease in the activity of the enzyme by three to four times.[179]

TPMT is abundantly found in liver and kidney and metabolizes immunosuppressive and anticancer thiopurines, such as 6-mercaptopurine (6-MP), 6-thioguanine (6-TG), and azathioprine. Oselin and Anier found that certain nonsteroidal anti-inflammatory drugs such as naproxen, mefanamic acid, and tolfenamic acid show noncompetitive inhibition of TPMT.[185] Genetic polymorphism has been found in TPMT and might lead to inter-individual variations in thiopurine drug toxicity and therapeutic efficacy.[186] HNMT catalyzes the degradation of histamine and drug molecules structurally similar to histamine.[187] It is inhibited by competitive mechanism by amine-containing drugs such as H1 and H2 histamine receptor antagonist.

MTs make a small relative contribution (<5%, Figure 2 and Table 2) toward drug metabolism, but have received greater attention from the scientific community (Table 1). This is mostly due to their role in the degradation of endogenous amines and drugs used for the treatment of neurological disorders. Kiss and Soares-da-Silva have recently given a comprehensive medicinal chemistry account of COMT inhibitors.[181] They have classified COMT inhibitors into (1) competitive inhibitors (catechol and pyrogallol), (2) tight-binding inhibitors (3,5-dinitrobenzene-1,2-diol), (3) meta-nitrated catechols (vinylic derivatives: entacapone, acetylated derivatives: tolcapone), (4) heterocyclic nitrocatechol derivatives (opicapone), (5) diphenyl sulfone scaffolds, (6) ortho-nitrated catechols (3,4-dihydroxy-5-nitrophenyl phenyl ketone), (7) trisubstituted catechol derivatives [3-(3,4-dihydroxy-2,5-dinitrophenyl)propanoic acid], (8) fused nitrocatechol derivatives [5,6-dihydroxy-4-nitroisobenzofuran-1(3*H*)-one], and other miscellaneous categories. Most of the inhibitors are substrates with high affinity but low catalytic efficiency. Thus, majority of the literature till date has focused on the discovery of the inhibitors of MTs, especially for COMT.[188, 189] Nevertheless, computational studies have been performed for understanding the regioselectivity of methylations of COMT substrates. These are discussed briefly below.

MD Simulations to Understand Regioselectivity in COMT-mediated Methylations

Tsao et al. have studied the COMT-catalyzed methylation of levodopa and dopamine.[182] The authors suggested that the biological function of COMT is to degrade the

catecholamines and this could be achieved by methylating any of the hydroxyl group. Thus, the regioselectivity of the reaction helps the enzyme increase the reaction rate and achieve a faster degradation. The simple reaction of catecholamine and trimethylsulfonium was studied in the gas phase, water phase, and protein environment to understand if the meta-selectivity is the result of ligand properties or due to the enzyme. For modeling protein environment molecular docking, dynamics and QM/MM calculations were performed on the reactants, TS, and products. The calculations showed that in the gas and water phase there is a marginal difference in the activation barrier for meta versus para methylation, whereas, in the protein environment the barrier difference is increased to almost 6 kcal/mol. Thus, it is the protein environment that favors the meta-methylation of catecholamines catalyzed by COMT. Active site residue Trp143 was involved in π-stacking interactions with the ligand and Lys144 acted as a base during the reaction.

Cao et al. have recently studied the regioselective methylation of quercetin and luteolin.[184] Experiments with recombinant S-COMT showed that luteolin is para-*O*-methylated, whereas quercetin is methylated at the meta position (Figure 15). Two conformations for both compounds were considered for generating eight complexes with COMT for performing MD simulations using AMBER 11. Ligand binding free energy and per-residue decomposition calculations were performed. Atomic charges and population analysis was performed for each set of conformations of the ligands. Conformations with highest binding energies were selected for QM/MM calculations (at B3LYP/6-311 + G(d,p)//AMBER level) for estimating the reaction barriers. Distances between catechol oxygen atoms to be methylated and sulfur atom of *S*-adenosyl-L-methionine (2.95−3.10 Å) and the O🍎C🍎S angle (160–170°) were constrained for selection of starting geometries form the MD trajectories.

Overall, the calculations showed that quercetin has highest binding energy in the pose that favors methylation at the meta position, whereas, for luteolin, the binding mode for the para methylation reaction had higher binding energy than the meta position. Thus, in short the regioselectivity of COMT-catalyzed methylation of flavonoids is driven by the balance between the thermodynamics and kinetics of the reaction.

Molecular Docking Studies to Understand COMT Substrate/Inhibitor Interactions

Palma et al. studied the interactions of BIA 3-202 (Figure 16) with COMT using flexible docking simulations and interpreted the regioselectivity of metabolic O-methylation.[190] *In vivo* experiments had shown that only meta-O-methylation takes place, but *in vitro* data showed formation of both meta and para O-methylation products. Docking simulations using GOLD v1.1 were performed to get insight into this observation. Ligand structures were optimized at PM3 semi-empirical level using SpartanPro v1.03 program. Twenty docking

runs were done for each ligand (BIA 3-202, 3,5-dinitrocatechol, and L-DOPA) using S-COMT crystal structure (1VID) and 20 optimized configurations were obtained. The docking studies showed that in order to have an effective binding into the catalytic pocket of COMT the carbonyl group should adopt a planar conformation with respect to catechol aromatic ring. The docking simulation showed poses that for both meta and para hydroxyl groups the oxygen of catechol and carbon of SAM were close to bonding distances. This explained methylation at both positions seen *in vitro*. Additional experiments with microsomal enzymes showed preferred dimethylation of the para-O-methylated product. The results were successfully validated using L-DOPA and 3,5-DNC as ligands. These results although not conclusive hinted that the overall balance is toward the accumulation of the meta-O-methylated metabolites over the para regioisomers. The same group later, worked on the effect of nitro group position on the regioselectivity of COMT-catalyzed methylation.191 Experimental incubations showed that BIA 3-228 and BIA 8-176 (Figure 16) were methylated slowly with opposite regioselectivity and could serve as inhibitors of the COMT. A similar docking procedure and MOPAC calculations were used to explain the experimental findings. Interactions of the nitro group within the active site (Trp143 and Lys144) were found more important than interactions with the benzoyl group for determining regioselectivity. Docking in the ortho conformations were found preferable, but reactivity studies at MOPAC level suggested that for BIA 3-228 the reactivity for ortho hydroxyl group is lower, thus explaining the formation of meta-O-methylation product.

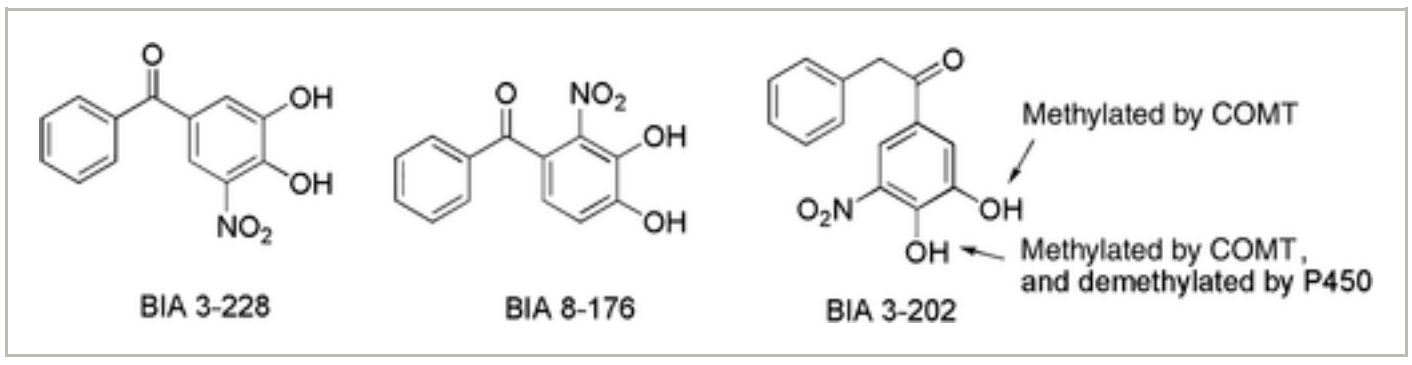

Figure 16

Open in figure viewer | PowerPoint

Regioselectivity of methylation and demethylation of nitrocatechol derivatives by COMT and P450.

Caption ⌄

Chen et al. worked on the inhibitory activity of tea catechins and their metabolites on COMT enzyme.192 Epigallocatechin gallate (EGCG) is methylated to 4',4"-di-*O*-methyl-EGCG by COMT which inhibits the degradation of L-DOPA. Experiments and molecular modeling were performed to understand the SAR of tea catechins. The homology model of human COMT

enzyme was generated using crystal structure of rat COMT (PDB ID =1VID). Mg^{2+} and the surrounding residue coordinates were deduced from the rat COMT. AM1 level electronic changes were determined using Spartan 02 program. Flexible docking was performed using FlexX software. The reaction involves removal of a proton from enzyme–catechol complex by Lys144, thus the distance between the Lys144 and hydroxyl groups was monitored. Furthermore, the binding energy of all the 10 ligands was calculated using molecular mechanics procedures. The results enabled the understanding of the inhibitory activity of the EGCG and (−)-epicatechin (EC). Docking predicted that the inhibition is more due to the presence of galloyl group, either of the two rings (D and B) can occupy the binding pocket of COMT and the distance calculations suggested that the affinity is more when D-ring is accommodated in the binding site. The model also depicted that the stabilization of the D-ring is more due to molecular π–π interactions with the side chain of hydrophobic residue. This stabilization effect is not seen in metabolite of EGCG and indicated that EGCG has better inhibitory activity.

Molecular docking and 3D-QSAR techniques were employed by Tervo et al. to understand the SAR of the COMT enzyme with 92 inhibitors.193 Flexible docking was carried out using the program FlexX with the X-ray crystal structure of COMT (1VID). The conformations from docking were used in the QSAR studies (using CoMFA and GRID/GOLPE) and the results were subjected to statistical analysis using PLS and cross-validation method. In the GOLPE-based QSAR, initially obtained variables were reduced using smart region definition and factorial selection procedure. The results obtained showed that there is a no correlation between the docking score and IC_{50} values, but the CoMFA and GOLPE gave satisfactory results (Table 10). The 3D QSAR models thus generated supported the previous findings that replacement of the nitro substituent with more electronegative group will have a positive effect on inhibitory activity of entacapone derivative. Interactions with polar (Lys144, Asn170, and Glu199) and hydrophobic (Trp38, Pro174, and Leu198) residues were found important in the docking studies. The QSAR studies also suggest that an increase in the steric volume at the side diethylamine chain would enhance the inhibition activity. The inhibitors of COMT developed from 1990 to 2010 can be divided into two generations.181, 194 First generation includes pyrogallol, tropolone, and gallic acid which have toxic side effect and less potency, while second generation included entacapone, tolcapone, nitecapone, and CGP.

Table 10. Summary of the QSAR Results Obtained by Tervo Et al.

S. No.	Method	r^2	q^2
1	FlexX docking score	0.299	–

2	CoMFA	0.903	0.483
3	GOLPE	0.830	0.591

In 2007, Ai et al. used a dataset of 36 COMT-inhibitors which included entacapone, tolcapone, and 34 other compounds considering 3-nitrocatechol as a pharmacophore. Twenty eight molecules were used as a training set and remaining 8 formed the test set. 3D QSAR studies were performed using CoMFA and CoMSIA and PLS analysis was performed. Table 11 shows the summary of the COMFA and CoMSIA results. The CoMSIA model derived from steric and electrophilic field is inferior among the three models. From the analysis of CoMFA, contour maps showed that inhibitory activity decreases if O atom of carbonyl is replaced by steric bulk. Naphthyl substitution tends to increased inhibition of COMT due to aryl conjugation. The CoMSIA contour map shows that phenyl group substitution at N position of piperazinyl enhances the bioactivity.

Table 11. Summary of QSAR Results Obtained by Ai Et al.

Parameters	CoMFA	CoMSIA	
	(Steric and Electrostatic)	Steric and Electrostatic	Hydrophobic
PLS components	6	4	6
r^2_{cv}	0.585	0.528	0.544
r^2_{ncv}	0.979	0.891	0.930
r^2_{pred}	0.569	0.199	0.779

COMBINED/HYBRID MODELS FOR DRUG METABOLISM PREDICTION

Experience in modeling metabolic outcome of CYP450-catalyzed reactions has led to the realization in the scientific community that use of combined or hybrid models is useful (sometimes necessary) for achieving high prediction accuracies.23 Such hybrid models can make use of two or more conventional modeling methodologies (e.g., docking, QSAR, and reactivity) along with machine learning methods such as SVM, random forest, *k*-nearest

neighbor, and others. Examples of applications of such methodologies to non-CYP450 Phase I and Phase II enzymes are rare (e.g., SOMP and XenoSite for UGT predictions). It is expected that in the near future such model will be developed by both academia and industry. Till date, very few models can make predictions for multiple Phase I and Phase II enzymes using such hybrid methodologies. One of the major hurdles in developing such models is the lack of sufficient datasets (collected under identical or similar experimental settings) for these enzymes. Many times, it is difficult to estimate the relative contributions from different enzymes for the formation of various metabolites.3 This can further complicate the data curation step for the development of reliable predictive models. We found only one study reported so far in which models for prediction of multiple SOM types, mostly from Phase I metabolism, were developed.

He et al. have recently developed hybrid models for the prediction of SOM for six metabolic Phase I reactions namely aldehyde oxidation to ketone and acid, double bond formation, oxidative deamination, N-dealkylation, aliphatic and aromatic hydroxylation each involving 483, 221, 317, 87, 104, 372, molecules, respectively.195 Data on metabolic reactions were extracted from BKM-react which consists of 28,042 unique biochemical reactions. Molecules involving only H, C, O, N, F, S, P, Cl, Br, and I were considered for model building. Positive SOMs (i.e., possible metabolic sites observed experimentally) and negative SOMs (i.e., possible metabolic sites not observed experimentally) were classified as labeled and unlabeled by identifying chemical bonds undergoing transformation to form products from the biotransformation database. Machine learning methods involving feature selection by four different techniques (CHI, IG, GR, Relief) and seven data classification procedures (KStar, BN, IBK, J48, RF, SVM, AdaBoostM1, Bagging) were used. Descriptor calculations were based on quantum chemically derived parameters for chemical bonds, molecular topology, and physicochemical properties. In the first step, feature selection was performed using scores obtained from feature selection methods mentioned above (Figure 17). Thus, features with nonzero scores representing features suitable for classification were kept. In step 2, feature reduction was performed by building models with classification procedures and then eliminating features with lowest scores. As step 3, models developed in step 2 were classified based on the features used and four series of models were developed for each reaction type. Most of the models performed well with accuracies of over 0.9, except the C H oxidative hydroxylation models for which the balanced accuracies were 0.892 and 0.894 for the training and test sets, respectively (Table 12).

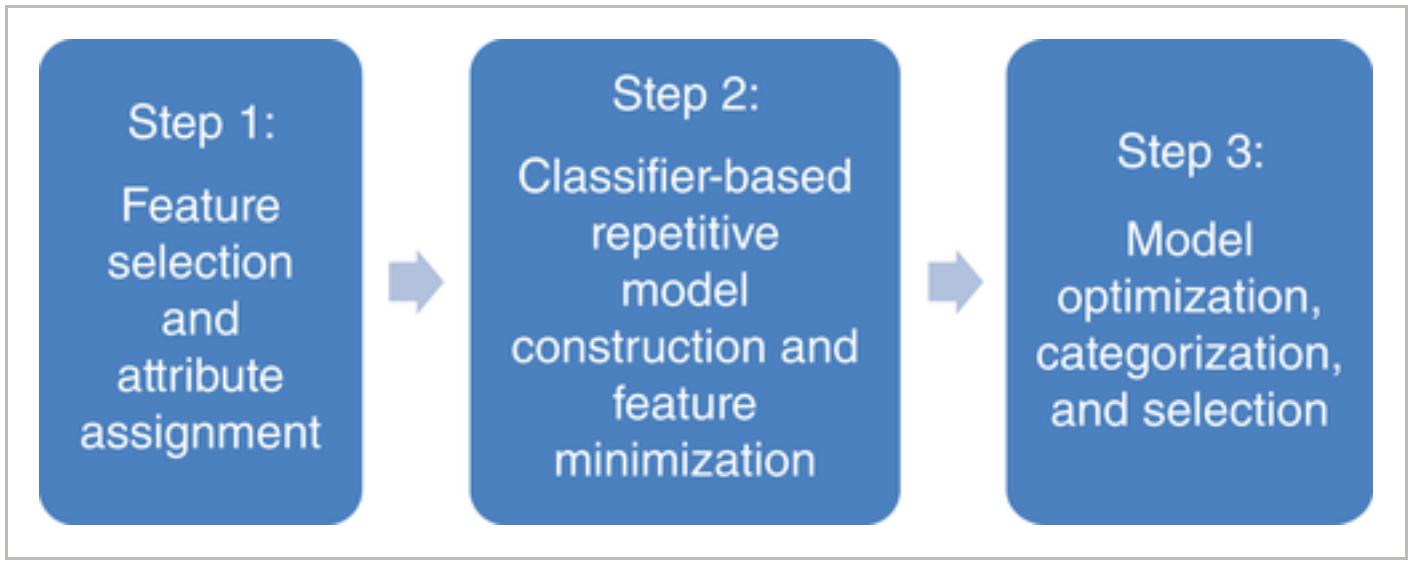

Figure 17

Open in figure viewer | PowerPoint

Summary of work flow used for SOM prediction model development for seven metabolic reaction types.

Caption ⌄

Table 12. Summary of SOM Data, Optimum Models and Results from He Et al.[195]

Reaction Type	SOM			Optimum Model			Balanced Accuracy	
	Positive	Negative	Number of Enzymes	Classifier	FS	Features	Training Set	Test Set
Alcohol to ketone	651	1678	269	KStar	CHI/IG	12	0.921	0.905
Alcohol to aldehyde	142	440	71	KStar	CHI/GR/IG	10	0.937	0.954
Double bond formation	708	1897	123	Random Forest	IG	45	0.968	0.975
Oxidative deamination	103	138	23	KStar	CHI/GR/IG	8	0.996	0.983
N-Dealkylation	86	250	38	KStar	Relief	26	0.940	0.955

C–H oxidation	861	2112	218	KStar	CHI	12	0.892	0.894

Recently, there have been attempts to simultaneously predict the biological activity and ADMET properties using a quantitative structure–biological effect relationship (QSBER) model.[196, 197] These models known as multitasking QSBER (mkt-QSBER) have been proposed to serve as complementary approaches to high-throughput screening technologies and reduce the chemical space that needs to be actually tested for finding the most promising clinical candidate. Speck-Planche and Cordeiro have used the CHEMBL database of 23,705 chemicals containing 37,834 biological endpoints collected under different conditions. The biological endpoints were classified as either positives or negatives based on a set of predetermined cutoff values. Descriptors were calculated using TOMOCOMD-CARDD framework and represented the polar surface area, hydrophobicity, atomic weights, and molecular refractivity. The dataset was randomly divided into training (75%) and test (25%) sets each comprising approximately 46% positive and 54% negative cases. The best model, as assessed by sensitivity, specificity, accuracy, and MCC was built using linear discriminant analysis and showed excellent predictive power. The model was tested on avarofloxacin (an investigational antimicrobial agent) showed that predictions for anti-*E. coli* activities and ADMET properties were in good agreement with the literature reports.

The same group later performed similar studies for the simultaneous estimation of anti-*Pseudomonas* [198], anti-*Staphylococcus* [199]activity, and ADMET properties. The best mkt-QSBER model used six molecular descriptors. Quality assessment parameters showed that the accuracy was >90% in both training and test sets. These models were used to understand the effect of different molecular features (ring size, hydrophobicity, and hydrophilicity) on the biological activity. These two studies respectively made predictions for delafloxacin and an investigational drug (JNJ-Q2) to demonstrate the practical applicability of the model and found good agreement with the experimental positive results.

CONCLUSION AND FUTURE DIRECTIONS

US-FDA guidelines for NDA include *in vitro* and *in vivo* drug metabolism studies for the Phase I and Phase II enzymes covered in this review. Such studies are resource intensive and require considerable amount of time. Recently, Bohnert et al. have given comprehensive schemes for the *in vivo* estimation of metabolic drug–drug interactions based on a set of *in vitro* sequential screening with different Phase I and Phase II recombinant enzymes and

hepatocyte fractions.200 Similar workflows or schemes utilizing computational models for sequential estimation of Phase I and Phase II-mediated drug metabolism predictions from molecular structure will be of great value for academic and industrial research and may help in increasing the productivity of drug discovery processes. Thus, computational models for the prediction of kinetic parameters, isoform specificity, and regioselectivity drug metabolism are in high demand, but are available only for CYP450, FMO, and to some extent UGT enzymes.

Recently, SOM and isoform specificity prediction models have been published for FMO and UGT class of Phase I and II enzymes, respectively. But considering the control that medicinal chemists and drug metabolism scientist have on CYP450-mediated metabolic profile of drug-like molecules, importance of better predictions for other Phase I and II enzymes has increased recently. For ADH, ALDH, MAO, SULT, GST, and MT enzyme systems, literature till date has focused on either prediction of inhibitors, substrates, understanding reaction mechanism, QSAR, kinetic parameters, or regioselectivity on a case-to-case basis. Thus, in contrast to the CYP450s, generalized predictive models for SOM, isoform specificity, and kinetic parameters for these enzymes are almost missing from the literature. Moreover, majority of the currently available models (including those for CYP450) make predictions for only drug candidates. But majority of the Phase II enzymes act on the primary metabolites resulting from Phase I metabolism. These primary metabolites, themselves might be subjected to another round of Phase I and II metabolism leading to toxicity (e.g., epoxides formed from the Phase I oxidation of polyaromatic hydrocarbons).

Therefore, there is an urgent need for the development of more comprehensive and sequential metabolic prediction models. Ideally one would like to not only make regioselective, isoform specific, and kinetic predictions for individual enzymes, but also to compare and rerank these predictions (based on probability) across different Phase I and Phase II enzymes. Additionally, estimation of the probability of formation of secondary metabolite from these enzymes will give an almost complete picture of these metabolic processes. At present, this task is complicated by facts such as lack of established *in vitro* metabolic systems suitable for drug metabolism studies (e.g., for ADH, ALDH, GSH, and MT), lack of sufficient data to generate statistically significant models, unclear contribution from other metabolic enzymes such as CYP450s. As observed by Swamidass et al. when sufficient data is available, generating such models for Phase II enzymes like UGTs can be simpler than for CYP450s.151 This is expected for enzymes which have specific functional group requirements for the metabolic reactions to occur (e.g., hydroxyl and amino for UGTs, amines for MAO, FMO, alcohol, and aldehyde for ADH and ALDH, respectively). This raises hopes for the development of useful and widely applicable models. But the development of

comprehensive metabolic prediction, as mentioned above, will require development of novel modeling (mathematical) techniques to tackle multidimensional, heterogeneous, incomplete nature of the data, noise, and associated variables. Currently, our group is working on developing regioselectivity models for SULT isoforms. Overall, we expect that in the near future scientific community will pay increasing attention to all these aspects toward the development of comprehensive and holistic models for drug metabolism prediction.

REFERENCES

Notes :

Conflict of interest: The authors have no conflicts of interest to declare.

See caption of Figures 1 and 2 for full form of common enzyme names.

The data were collected from Google scholar on November 16, 2016.

AD, alcohol dehydrogenase; ALDH, aldehyde dehydrogenase; AO, amine oxidase; CYP450, cytochrome P450; EH, epoxide hydrolyase; FMO, flavin monooxygenase; GST, glutathione *S*-transferase; MT, methyltransferase; SULT, sulfotransferase; UGT, UDP-glucuronosyl transferase.

NA, not applicable.

1 Includes AUC values for aliphatic, aromatic, carboxylic, nitrogen, and atypical SOGs.

Values of K_{cat}/k_{m} are expressed in min^{-1} M^{-1}.

Citing Literature

Connect with Wiley

WILEY