Regioselectivity Prediction of CYP1A2-Mediated Phase I Metabolism

Jihoon Jung,[†] Nam Doo Kim,[†,‡] Su Yeon Kim,[§] Inhee Choi,[||] Kwang-Hwi Cho,[⊥] Won Seok Oh,[†] Doo Nam Kim,[#] and Kyoung Tai No*[,†,#]

Department of Biotechnology, Yonsei University, 120-749, Seoul, Korea, Drug Design Team, Equispharm Inc., 864-1, Gyeonggi, Korea, BK21 Yonsei Biomolecule Research Initiative, Yonsei University, 120-749, Seoul, Korea, Institute of Life Science & Biotechnology, Yonsei University, 120-749, Seoul, Korea, Department of Bioinformatics and CAMDRC, Soongsil University, 156-743, Seoul, Korea, and Bioinformatics & Molecular Design Research Center (BMDRC), 120-749, Seoul, Korea

Received January 4, 2008

A kinetic, reactivity-binding model has been proposed to predict the regioselectivity of substrates meditated by the CYP1A2 enzyme, which is responsible for the metabolism of planar-conjugated compounds such as caffeine. This model consists of a docking simulation for binding energy and a semiempirical molecular orbital calculation for activation energy. Possible binding modes of CYP1A2 substrates were first examined using automated docking based on the crystal structure of CYP1A2, and binding energy was calculated. Then, activation energies for CYP1A2-mediated metabolism reactions were calculated using the semiempirical molecular orbital calculation, AM1. Finally, the metabolic probability obtained from two energy terms, binding and activation energies, was used for predicting the most probable metabolic site. This model predicted 8 out of 12 substrates accurately as the primary preferred site among all possible metabolic sites, and the other four substrates were predicted into the secondary preferred site. This method can be applied for qualitative prediction of drug metabolism mediated by CYP1A2 and other CYP450 family enzymes, helping to develop drugs efficiently.

INTRODUCTION

Since the majority of xenobiotics are metabolized by cytochrome P450 (CYP) enzymes to develop drugs efficiently, it is necessary to know the CYP450-mediated metabolic profiles of compounds during drug discovery and development. However, in early stages of drug discovery due to the large number of compounds to be studied, there is a limitation in measuring the CYP450-mediated metabolic profiles of these compounds. Therefore, to develop new and safer drugs faster, several in silico approaches have been introduced to predict metabolites of xenobiotics.[1]

The purpose of in vitro and in vivo xenobiotic metabolism studies is to obtain the activities and specificities of CYP450 enzymes to target compounds (substrates) and to identify their metabolites.

On the basis of the selectivity and activity of the enzyme toward a certain compound, the compound can be classified as a substrate, nonsubstrate, or inhibitor. The quantitative information on such selectivity and activity can be used for the prediction of (i) the pharmacokinetic (PK) profile of the chemical such as the concentration of the chemical in the blood as a function of time and (ii) drug−drug interactions as well as undesired side effects.

The regioselectivity of a compound by specific CYP450 isozyme-mediated metabolism contains information about the selectivity, activity, and steric information of the compound in the active site of the enzyme. It can be used to optimize lead compounds to reduce the rate of metabolism and avoid toxic metabolite formation.

To predict the phase I metabolites, several computational methods have been proposed. These methods can be classified into three categories (i) use the reactivity of the substrate, ligand-based, (ii) use the binding of the substrate to a CYP450 enzyme, docking-based, and (iii) use both reactivity and binding, reactivity-docking-based.

In ligand-based approaches, empirical models were developed that were able to predict activation energies involved in CYP-mediated metabolic reactions using semiempirical quantum mechanical (QM) calculations.[2–6] These models predicted the sites of biotransformation based on the estimated activation energy of each atom in a compound. A statistical model was also developed that was able to predict the most probable metabolic sites on the basis of a statistical analysis of various metabolic transformations.[7] The models to predict CYP450 isoform specificity for substrates were investigated. Since CYP450 substrates are likely to be metabolized by several CYP450 isoforms, it was the aim of these models to predict the isoform responsible for the primary route of metabolism.[8–10]

Docking-based approaches rely on the binding of compounds in the active site of the CYP enzyme. There have been a number of accounts in which the regioselectivity of known substrates were predicted based on the atomic site oriented toward the ferric atom of the heme of the corresponding CYP enzyme.[11–21]

* To whom correspondence should be addressed. Tel: +82-2-2123-5882. Fax: +82-2-362-7265. E-mail: ktno@yonsei.ac.kr.
[†] Department of Biotechnology, Yonsei University.
[‡] Equispharm Inc.
[§] BK21 Yonsei Biomolecule Research Initiative, Yonsei University.
[||] Institute of Life Science & Biotechnology, Yonsei University.
[⊥] Soongsil University.
[#] Bioinformatics & Molecular Design Research Center (BMDRC).

10.1021/ci800001m CCC: $40.75
Published on Web 04/16/2008

Reactivity-docking-based approaches take into account the reactivity of the compound as well as the shape of the active site. Therefore, these approaches provide better predictions of regioselectivity for CYP substrates. In these approaches, metabolic sites were presumed based on substrate reactivity used in combination with the accessibility of each atom toward the heme of the enzyme.[22–28]

CYP1A2 is an important enzyme of the CYP1 family expressed in human liver; it is responsible for the metabolism of aromatic and heterocyclic amines, as well as polycyclic aromatic hydrocarbons (PAHs).[29,30] A well-known feature of the binding of CYP1A2 substrates is that these substrates have a relatively flat conformation within the planar cavity, which is made by both the Gly316-Ala317 peptide bond and the side chain of Phe226.[31]

The aim of the present work is to predict the metabolic sites of CYP1A2 substrates, including drugs such as caffeine, tacrine, and theophylline. We suggest a prediction model that combines binding energy, activation energy, and spatial arrangement of the atom that participates in the metabolic reaction to determine the regioselectivity for these CYP1A2 substrates.

METHODS

Reactivity-Binding-Based Kinetic Model. We propose a model based on the reactivity-binding concept. In the model, both the chemical reactivity of the rate-determining step of the phase I metabolic reaction and the binding energy of the substrate to the enzyme and the local conformation at the reactive site are considered. Although the phase I metabolic cycle is complicated,[32] we represent it by two main steps as follows.

$$[S] + [E] \underset{k_r}{\overset{k_f}{\rightleftharpoons}} [SE] \xrightarrow[k_p]{} [M] \quad K_{eq} = \frac{k_f}{k_r} = \frac{[SE]}{[S][E]} \tag{1}$$

S, E, SE, and M represent the substrate, CYP450 enzyme, substrate−CYP450 complex, and metabolite, respectively. Since the activation energy of the rate-determining step is usually larger than 17 kcal·mol^{-1},[2] it is assumed that the substrate and CYP450 enzyme form a complex, SE, and at equilibrium, as in eq 1, the equilibrium constant is K_{eq}. The forward and backward reaction rate constants are k_f and k_r, respectively. From the Boltzmann distribution law, the concentration ratio, K_{eq}, is proportional to the free energy difference at equilibrium.

$$\Delta\Delta G_b = \Delta G(SE) - (\Delta G(S) + \Delta G(E)) \tag{2}$$

where ΔG (SE), ΔG(S), ΔG(E), and $\Delta\Delta G_b$ are the free energy of the complex, substrate, enzyme, and the binding free energy of the complex, respectively.

The equilibrium constant K_{eq} becomes

$$K_{eq} = e^{-\Delta\Delta G/RT} \tag{3}$$

The Arrhenius equation was used for the rate-determining step, and the rate constant is k_p.

$$k_p = Ae^{-E_a/RT} \tag{4}$$

Since the calculation of the free energies in eq 2 requires much effort even for one substrate−enzyme pair, in our calculation, the free energy change upon binding is approximated by AutoDock empirical free energy function,[33] ΔG_b.

$$K_{eq} \cong e^{-\Delta G_b/RT} \tag{5}$$

The rate of metabolite formation is shown as

$$\frac{d[M]}{dt} = k_p[SE] \tag{6}$$

which results in

$$\frac{d[M]}{dt} = k_p K_{eq}[S][E] = k_{app}[S][E] \tag{7}$$

where k_{app} is the apparent reaction rate,

$$k_{app} = A'e^{-\Delta G_b/RT}e^{-E_a/RT} = A'e^{-(\Delta G_b + E_a)/RT} \tag{8}$$

Thus, the relative probability of the formation of the ith metabolite (P_i) among all the possible metabolites of a substrate is

$$P_i = \frac{e^{-(\Delta G_b^i + E_a^i)/RT}}{\sum_{j=1}^{N} e^{-(\Delta G_b^j + E_a^j)/RT}} \tag{9}$$

where N is the number of all possible metabolites of the substrate.

Computation. For the calculation of G_b, the AutoDock program[33] was used, and E_a was calculated using an empirical activation energy calculation method developed by both Korzekwa and Jones et al.[2,4]

CYP1A2 Enzyme Model. The crystal structure of human CYP1A2 in complex with the inhibitor α-naphthoflavone (PDB ID: 2HI4) was determined at 1.95 Å resolution by Sansen et al.[31] The α-naphthoflavone was positioned within the active site above the distal surface of the heme prosthetic group. Since the binding pocket is relatively narrow and planar, CYP1A2 has the right size and shape specificity for a substrate.

Data Collection of CYP1A2-Mediated Metabolism. Twelve CYP1A2 substrates were identified from the Fujitsu ADME database.[34–36] Most well-known and representative test/marker CYP1A2 substrates were included in our data set, such as caffeine, tacrine, and theophylline. The major metabolic site of each compound is shown in Figure 1.

Binding Energy, ΔG_b, Calculation. AutoDock was used for the binding energy calculation. The coefficients of AutoDock free energy function were empirically determined using linear regression analysis from a set of protein–ligand complexes with known binding constants.[33] During the docking simulation, the enzyme structure was kept rigid but the substrate was left fully flexible. For the initial structure of the substrate in the docking simulation, the substrate structure was obtained using energy minimization by the conjugate gradient method implemented in Cerius2.[37] Since Gasteiger−Marsili net atomic charges were used for the calibration of the AutoDock empirical free energy function, the Gasteiger−Marsili net atomic charges were used for the energy calculation of the substrates.

For the preparation of the CYP1A2 enzyme, Amber charge parameters were used to assign partial charges with Insight II.[38] Amber charge parameters of nonstandard protein residues, such as the heme prosthetic group, were not determined. For heme partial charges of the CYP1A2

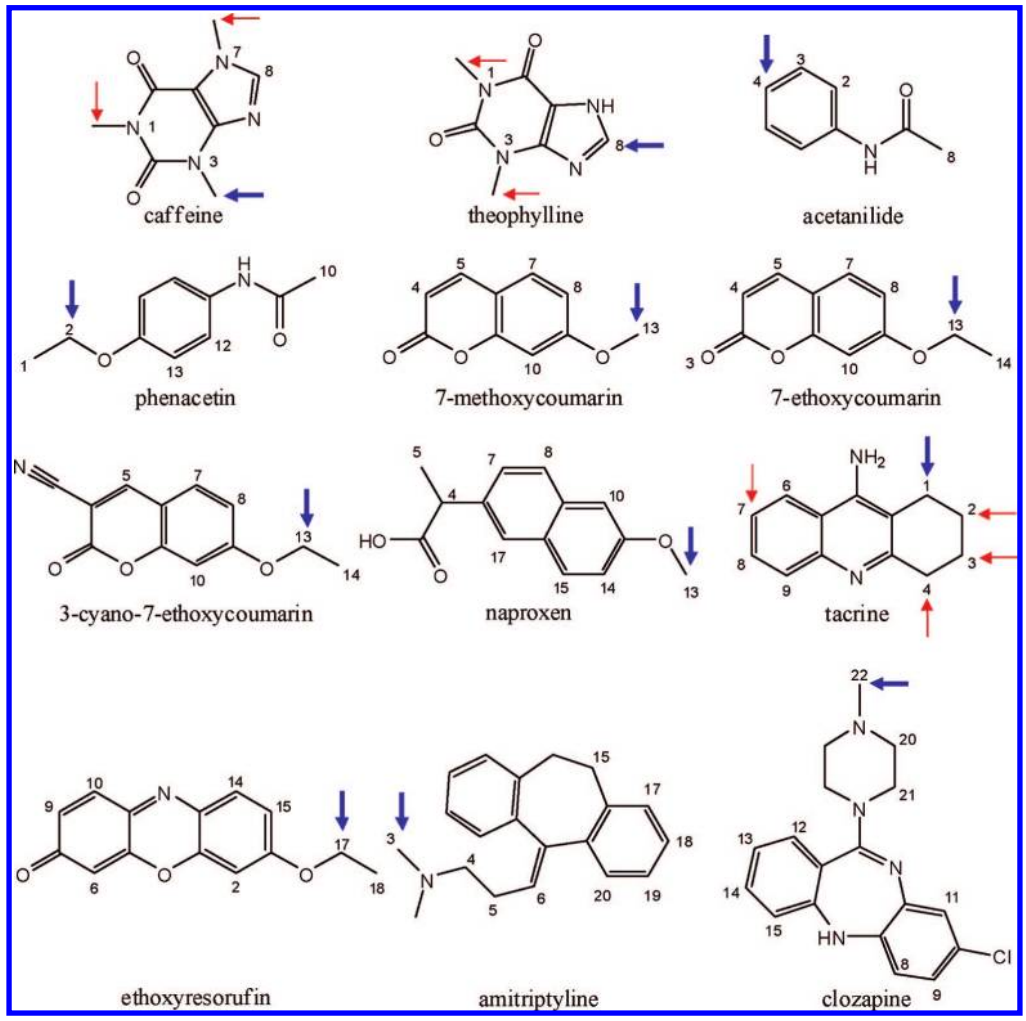

Figure 1. Twelve CYP1A2 substrates in the data set. The chemical structures of cytochrome P450 1A2 substrates are shown. The site of metabolism by CYP1A2 is indicated by an arrow. Major (blue) and minor (red) metabolic sites are indicated. The numbers on atoms are used to distinguish between atom sites.

enzyme, therefore, RESP charges determined by quantum chemical calculations were used,[39] and the r and epsilon values were changed for the M and X atoms of the iron and tetrapyrrole nitrogens in the protein GPF files, respectively.

To reproduce the binding energies in aqueous solution, atomic solvation parameters and fragmental volumes were assigned to the protein atoms using the AddSol application of the AutoDock program. During the docking simulation, all atoms of the CYP1A2 substrates were allowed to move. The AutoTors application of the AutoDock program was used to define the rotatable bonds in the CYP1A2 substrates. The interaction energies between enzyme and substrate were evaluated using atom affinity potentials that were precalculated on grid maps calculated using AutoGrid. The grid maps had dimensions of $10 \times 10 \times 10$ Å^3 with 0.125 Å spacing between the grid points.

Using a Lamarckian genetic algorithm (LGA), the minimum energy conformations of the substrate inside the binding pocket of CYP1A2 were generated. LGA parameters used for the docking were as follows: population size of 50; random starting position; maximal mutation of 2 Å in translation and 50 degrees in rotation; elitism of 1; mutation rate of 0.02; crossover rate of 0.8; and local search rate of 0.06. Simulations were performed with a maximum of 25,000 energy evaluations and a maximum of 27,000 generations. The pseudo-Solis and Wets local search method was included with the default parameters. For each substrate, 100 docking trials were performed, and final docked conformations of substrates were clustered using a tolerance of 1 Å rmsd. The docked conformation that was properly oriented toward the heme group was selected, and its binding free energy (G_b) was estimated.

To determine the possible metabolic sites of the substrates, the distances (r) between the heme iron of CYP1A2 and the atoms of the substrates were measured for all docking results. A catalytically reactive distance from the heme iron of CYP450 is generally known to be within 5 Å; however, the range of distances used in this model was rather large due to ligand-induced conformational changes of the active site.[32] Thus, the atom sites within a catalytically reactive distance from the heme iron (~5.5 Å) were selected as possible metabolic sites. The binding free energies of the possible metabolic sites were determined from those of top-ranked docking poses where those atoms were located within the cutoff distance. The distances from the heme iron to the aromatic carbon and hydrogen were also measured for aromatic hydroxylation and hydrogen atom abstraction, respectively.

Activation Energy, E_a, Calculation. For the estimation of the activation energies of CYP-mediated hydrogen

abstraction (aliphatic) and aromatic oxidation, both Korzekwa and Jones et al.[2,4] proposed empirical methods which can be used to compare the reactivity between the two. The scaled activation energies based on regioselectivity were calculated using eq 10.

$$\Delta\Delta G = 1.22(\Delta H_{act(Habs)}) - 1.1(\Delta H_{act(arom)}) - 17.5 \quad (10)$$

Thus, scaled activation energy in our model for hydrogen atom abstraction ($\Delta E_{a(Habs)}$) and aromatic hydroxylation ($\Delta E_{a(arom)}$) can be calculated using eqs 11a and 11b.

$$\Delta E_{a(Habs)} = 1.22\Delta H_{act(Habs)} \quad (11a)$$

$$\Delta E_{a(arom)} = 1.1\Delta H_{act(arom)} + 17.5 \quad (11b)$$

AM1 ground-state energies of parent compounds and their radicals were calculated for the activation energy ($\Delta H_{act(Habs)}$) of hydrogen abstraction. The enthalpy of the reaction was also calculated for the activation energy ($\Delta H_{act(arom)}$) of aromatic hydroxylation. AM1 calculations were performed using the Gaussian 03 program.[40]

RESULTS AND DISCUSSION

In Table 1, all the aliphatic and aromatic oxidation sites of each substrate are listed. The binding free energy of each complex, corresponding to the minimal conformational space of the complex, obtained with docking calculations is also shown in Table 1. The possible metabolic sites of a substrate are assigned to those atoms located within 5.5 Å from the heme Fe in CYP1A2 from all the binding modes obtained with docking calculations. In the case of aliphatic oxidation, the hydrogen atoms with sufficient solvent exposure (solvent accessible surface area ≥ 8 Å^2) can be abstracted by the activated oxygen radical of the enzyme.[5] Among the possible metabolic sites assigned by our model, all hydrogen atoms have sufficient solvent exposure. Otherwise, the possible metabolic site is not assigned and is considered not to be a candidate for the metabolic reaction. This is because the sites can be metabolized when they come close enough to the heme. Thus, binding energies of only the possible metabolic sites are shown in Table 1.

The binding energy differences between the substrates are not large enough to provide specificity for a certain reaction site in the substrate. The binding energy differences are smaller than 0.6 kcal·mol^{-1} except for amitriptyline which is 2.3 kcal·mol^{-1}.

The activation energies (E_a) of all sites including the possible metabolic sites of each substrate, calculated with the method proposed by both Korzekwa and Jones et al., are shown in Table 1. The differences in the E_a values of candidate reaction sites in a molecule are large enough to find major metabolites. With the sum, $\Delta G_b + E_a$, the relative contribution of each metabolic site was calculated with the Boltzmann distribution law and normalized (eq 9). The probabilities (P_i) of the metabolic sites are also shown in Table 1.

Table 1 shows the ranking order of the metabolic sites in a molecule calculated using the method described by Korzekwa and Jones et al. as well as our method. Major metabolites are indicated by "M" and minor metabolites are indicated by "m". The docking poses where minor metabolic sites of two substrates (caffeine and theophylline) were close to the heme had not been found. One of the reasons for this is that rigid docking was performed with the crystal structure which typically displays a compact conformation.

In Table 2, the predictabilities of regioselectivity obtained only with E_a (method of Korzekwa et al.) and with ΔG_b and E_a (our method) are summarized. Only with E_a, the major metabolite of 3 substrates (25%) out of 12 agreed with experimental results, whereas using ΔG_b and E_a, those of 8 substrates (67%) agreed with experimental results.

In some cases, there were some intrinsic origins of error in the in the binding energy calculation method. The energy function had an error of about 2.2 kcal·mol^{-1}.[33] However, this value is a standard error for binding free energies of many different protein–ligand complexes. Therefore, the error for those energies of various binding modes of the same protein–ligand complex may be smaller, and the error of the differences in these energies may be much smaller. The E_a calculation model also had an error of about 0.7 kcal·mol^{-1}; a 1 kcal·mol^{-1} difference in $\Delta G_b + E_a$ may not play an important role in ranking metabolite formation. The force field, flexibility of the enzyme, and the role of water also may be the origins of the error. Therefore, in estimations of predictability, it may be reasonable to count regioselectivity prediction as correct if the most probable or second most probable metabolites match the major experimental metabolites. As shown in Table 2, with the lenient criteria mentioned above, the predictability only with E_a and with ΔG_b and E_a are 67%, and 100%, respectively.

Figure 2 reveals that all substrates introduced in our study were positioned parallel within the narrow and planar cavity that is made by both the Gly316-Ala317 peptide bond and the side chain of Phe226 in the active site, as reported previously.[31]

CONCLUSIONS

Since CYP enzymes play key roles in drug metabolism, the accurate prediction of CYP-mediated metabolic profiles is necessary for efficient drug discovery. With the aid of accurate prediction, one can design compounds that have moderate metabolic rates that avoid fast phase I drug metabolism and compounds can be designed in which their metabolites have no or weak toxicity, or carcinogenicity.

Metabolism reactions depend on physicochemical and structural characteristics of chemical compounds as well as on binding affinity and conformation of the active site of the CYP-metabolizing enzymes. The use of the docking method has allowed the determination of both the binding of the CYP enzyme–substrate complex and the atom sites oriented toward the heme iron of the CYP enzyme. Using only the docking method, however, cannot always predict the sites of metabolism accurately because it solely considers the binding affinity and steric effects without considering the reactivity of the substrates. On the other hand, semiempirical molecular orbital calculations can provide the activation energy when metabolism reactions occur; thus, one can know the atom sites where reactions happen easily through this method. There is also a limitation to this approach. The predicted regioselectivity of large compounds is not as good as that of small compounds since the approach does not take into account the steric effects of the active site of the enzyme. This limitation can be overcome using the docking method which considers the structural features of the active site of the enzyme.

Table 1. Probabilities for All Possible Metabolic Sites within Each CYP1A2 Substrate

compound	atom site[a]	r^b	$\Delta G_b{}^c$	$E_a{}^d$	$P_i{}^e$	R1[f]	R2[g]	exp[h]
caffeine	N1-CH_3	3.74	−7.07	26.40	0.09	3	2	m
	N3-CH_3	3.26	−7.22	25.16	0.86	1	1	M
	N7-CH_3			25.95		2		m
	C8H	5.41	−7.19	26.85	0.05	4	3	
theophylline	N1-CH_3			26.46		3		m
	N3-CH_3	3.24	−6.88	25.39	0.77	1	1	m
	C8H	5.47	−6.76	26.01	0.23	2	2	M
acetanilide	C2H	4.46	−6.07	27.00	0.79	1	1	
	C3H	3.81	−6.07	29.39	0.02	3	3	
	C4H	4.76	−6.07	27.87	0.19	2	2	M
	$C8H_3$	4.05	−6.10	29.89	0.01	4	3	
phenacetin	$C1H_3$	3.58	−6.70	28.89	0.01	4	2	
	$C2H_2$	3.17	−6.70	26.03	0.96	1	1	M
	$C10H_3$	4.29	−6.81	28.98	0.01	5	2	
	C12H			27.18		2		
	C13H	4.84	−6.60	28.45	0.02	3	2	
7-methoxycoumarin	C4H			35.9		6		
	C5H			34.76		5		
	C7H			27.93		2		
	C8H			29.11		4		
	C10H			26.49		1		
	$C13H_3$	5.10	−7.06	28.58	1.00	3	1	M
7-ethoxycoumarin	C4H			35.37		7		
	C5H	3.83	−7.19	34.63	0.00	6	5	
	C7H	4.71	−7.19	27.96	0.13	3	2	
	C8H	5.16	−7.22	29.09	0.02	4	3	
	C10H			26.18		1		
	$C13H_2$	4.79	−7.22	26.82	0.84	2	1	M
	$C14H_3$	2.80	−7.22	29.66	0.01	5	4	
3-cyano-7-ethoxycoumarin	C5H	4.20	−7.83	35.69	0.00	6	4	
	C7H			27.96		3		
	C8H	5.27	−8.38	29.42	0.05	4	2	
	C10H			26.34		1		
	$C13H_2$	5.01	−8.38	27.57	0.95	2	1	M
	$C14H_3$	2.97	−8.38	30.52	0.01	5	3	
naproxen	C4H	4.03	−7.78	26.13	0.78	5	1	
	$C5H_3$	5.01	−8.07	28.30	0.04	9	3	
	C7H			27.17		8		
	C8H			24.29		3		
	C10H			22.65		1		
	$C13H_3$	4.44	−7.88	27.11	0.19	7	2	M
	C14H			26.16		6		
	C15H			24.22		2		
	C17H			24.67		4		
tacrine	$C1H_3$	4.83	−10.72	21.51	0.94	1	1	M
	$C2H_3$	3.66	−10.69	25.68	0.00	5	4	m
	$C3H_3$	4.03	−10.69	25.78	0.00	7	4	m
	$C4H_3$	4.43	−10.73	24.02	0.02	4	3	m
	C6H	4.34	−11.08	23.83	0.04	3	2	
	C7H	4.66	−11.08	25.68	0.00	5	4	m
	C8H			27.19		8		
	C9H			23.59		2		
ethoxyresorufin	C2H			23.89		2		
	C6H	5.22	−9.06	33.74	0.00	8	4	
	C9H	4.79	−8.60	32.82	0.00	6	4	
	C10H	4.89	−9.07	32.82	0.00	6	4	
	C14H			22.52		1		
	C15H	5.33	−8.78	26.41	0.35	4	2	
	$C17H_2$	4.69	−8.78	26.05	0.63	3	1	M
	$C18H_3$	4.82	−8.58	28.12	0.02	5	3	
amitriptyline	$C3H_3$	4.79	−11.50	22.04	1.00	2	1	M
	$C4H_2$			20.94		1		
	$C5H_2$			22.25		3		
	C6H			30.50		9		
	$C15H_2$			23.35		4		
	C17H			28.43		7		
	C18H			27.88		5		
	C19H	5.46	−9.19	28.27	0.00	6	2	
	C20H			29.23		8		
clozapine	C8H			25.71		5		
	C9H			26.77		7		
	C11H			25.63		4		

Table 1. Continued

compound	atom site[a]	r[b]	ΔG_b[c]	E_a[d]	P_i[e]	R1[f]	R2[g]	exp[h]
C12H			29.32		10			
C13H			26.86		8			
C14H			27.92		9			
C15H			26.28		6			
$C20H_2$	4.64	−10.84	21.20	0.89	1	1		
$C21H_2$			21.66		2			
$C22H_3$	5.13	−10.84	22.47	0.11	3	2	M	

[a] Atom sites are all candidates of the metabolic reaction. [b] r (Å) indicates the distance between the site of metabolism and the heme iron of CYP1A2. [c] ΔG_b (kcal·mol^{-1}) is the lowest value among binding energies in which the corresponding atom site is located within 5.5 Å from the heme iron. The values for atom sites that were not detected from docking results are shown as blank. [d] E_a (kcal·mol^{-1}) is the activation energy in which metabolism reactions occur on each atom site. [e] P_i (probability of each atom site in a substrate) indicates the potential that metabolism reaction can occur at the atom site calculated using eq 9. [f] Ranked by the activation energy. [g] Ranked by the P_i. [h] Major (M) and minor (m) metabolic sites are indicated.

Table 2. Predictability of Regioselectivities

	prediction model by E_a only			reactivity-binding-based kinetic model		
	first	second	others	first	second	others
predictability	3	5	4	8	4	0

In this study, we constructed the reactivity-binding-based kinetic model using both compound reactivity and binding energy. As a result, the predicted major metabolic sites of 8 out of 12 CYP1A2 substrates tested were in accordance with experimental results. For the other four substrates, this model predicted the metabolic site as the secondary preferred site. Therefore, the current model predicted metabolism well because it considered the reactivity of the compound as well as the binding affinity and steric effects of the enzyme.

This method can be easily expanded for qualitative predictions of drug metabolism meditated by not only CYP1A2 but also by other CYP450 family enzymes, helping to develop drugs more efficiently. Furthermore, this method can be a better model for regioselectivity prediction by developing more accurate methods of E_a and G_b calculation.

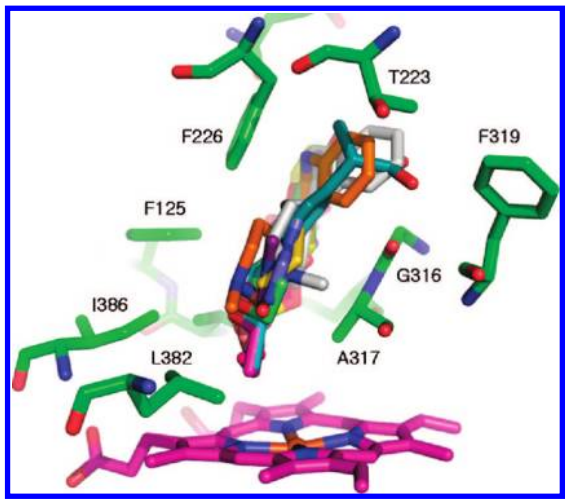

Figure 2. Superimposed CYP1A2 substrates within the active site of the enzyme. The 12 substrates docked within the active site are positioned parallel between the flat surface formed by Gly316 and Ala317 and the other side produced by Phe226. Amino acid residues in the active site are indicated by their number and one-letter code. These residues and substrates are shown as sticks and are colored according to atom type (oxygen, red; nitrogen, blue). Carbon atoms of the residues in the active site and heme group are colored green and purple, respectively. The iron atom in heme is colored orange.

ACKNOWLEDGMENT

This work was supported by the Hyperstructured Organic Materials Research Center (HOMRC) at Seoul National University, KOSEF, Korea, and by the Korea Research Foundation Grant funded by the Korean Government (MOE-HRD) (KRF-2006-005-J04501). This work was also supported in part by the brain Korea 21(BK21) Program (Su Yeon Kim is a fellowship awardee by the BK21 program).

REFERENCES AND NOTES

(1) Crivori, P.; Poggesi, I. Computational approaches for predicting CYP-related metabolism properties in the screening of new drugs. *Eur. J. Med. Chem.* **2006**, *41*, 795–808.
(2) Korzekwa, K. R.; Jones, J. P.; Gillette, J. R. Theoretical studies on cytochrome P-450 mediated hydroxylation: a predictive model for hydrogen atom abstractions. *J. Am. Chem. Soc.* **1990**, *112*, 7042–7046.
(3) Korzekwa, K. R.; Grogan, J.; DeVito, S.; Jones, J. P. Electronic models for cytochrome P450 oxidations. *Adv. Exp. Med. Biol.* **1996**, *387*, 361–369.
(4) Jones, J. P.; Mysinger, M.; Korzekwa, K. R. Computational models for cytochrome P450: a predictive electronic model for aromatic oxidation and hydrogen atom abstraction. *Drug Metab. Dispos.* **2002**, *30*, 7–12.
(5) Singh, S. B.; Shen, L. Q.; Walker, M. J.; Sheridan, R. P. A model for predicting likely sites of CYP3A4-mediated metabolism on drug-like molecules. *J. Med. Chem.* **2003**, *46*, 1330–1336.
(6) Sheridan, R. P.; Korzekwa, K. R.; Torres, R. A.; Walker, M. J. Empirical Regioselectivity Models for Human Cytochromes P450 3A4, 2D6, and 2C9. *J. Med. Chem.* **2007**, *50*, 3173–3184.
(7) Borodina, Y.; Rudik, A.; Filimonov, D.; Kharchevnikova, N.; Dmitriev, A.; Blinova, V.; Poroikov, V. A new statistical approach to predicting aromatic hydroxylation sites. Comparison with model-based approaches. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1998–2009.
(8) Manga, N.; Duffy, J. C.; Rowe, P. H.; Cronin, M. T. Structure-based methods for the prediction of the dominant P450 enzyme in human drug biotransformation: consideration of CYP3A4, CYP2C9, CYP2D6. *SAR QSAR Environ. Res.* **2005**, *16*, 43–61.
(9) Yap, C. W.; Chen, Y. Z. Prediction of cytochrome P450 3A4, 2D6, and 2C9 inhibitors and substrates by using support vector machines. *J. Chem. Inf. Model.* **2005**, *45*, 982–992.
(10) Terfloth, L.; Bienfait, B.; Gasteiger, J. Ligand-Based Models for the Isoform Specificity of Cytochrome P450 3A4, 2D6, and 2C9 Substrates. *J. Chem. Inf. Model.* **2007**, *47*, 1688–1701.
(11) Koymans, L.; Vermeulen, N. P.; van Acker, S. A.; te Koppele, J. M.; Heykants, J. J.; Lavrijsen, K.; Meuldermans, W.; Donne-Op, den; Kelder, G. M. A predictive model for substrates of cytochrome P450-debrisoquine (2D6). *Chem. Res. Toxicol.* **1992**, *5*, 211–219.
(12) Lewis, D. F.; Eddershaw, P. J.; Goldfarb, P. S.; Tarbit, M. H. Molecular modelling of CYP3A4 from an alignment with CYP102: identification of key interactions between putative active site residues and CYP3A-specific chemicals. *Xenobiotica* **1996**, *26*, 1067–1086.

(13) Lewis, D. F.; Lake, B. G. Molecular modelling of CYP1A subfamily members based on an alignment with CYP102: rationalization of CYP1A substrate specificity in terms of active site amino acid residues. *Xenobiotica* **1996**, *26*, 723–753.
(14) de Groot, M. J.; Vermeulen, N. P.; Kramer, J. D.; van Acker, F. A.; Donne-Op, den; Kelder, G. M. A three-dimensional protein model for human cytochrome P450 2D6 based on the crystal structures of P450 101, P450 102, and P450 108. *Chem. Res. Toxicol.* **1996**, *9*, 1079–1091.
(15) Modi, S.; Paine, M. J.; Sutcliffe, M. J.; Lian, L. Y.; Primrose, W. U.; Wolf, C. R.; Roberts, G. C. A model for human cytochrome P450 2D6 based on homology modeling and NMR studies of substrate binding. *Biochemistry* **1996**, *35*, 4540–4550.
(16) Lewis, D. F.; Eddershaw, P. J.; Goldfarb, P. S.; Tarbit, M. H. Molecular modelling of cytochrome P4502D6 (CYP2D6) based on an alignment with CYP102: structural studies on specific CYP2D6 substrate metabolism. *Xenobiotica* **1997**, *27*, 319–339.
(17) Lozano, J. J.; Lopez-de-Brinas, E.; Centeno, N. B.; Guigo, R.; Sanz, F. Three-dimensional modelling of human cytochrome P450 1A2 and its interaction with caffeine and MeIQ. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 395–408.
(18) Payne, V. A.; Chang, Y. T.; Loew, G. H. Homology modeling and substrate binding study of human CYP2C9 enzyme. *Proteins* **1999**, *37*, 176–190.
(19) Lewis, D. F.; Bird, M. G.; Dickins, M.; Lake, B. G.; Eddershaw, P. J.; Tarbit, M. H.; Goldfarb, P. S. Molecular modelling of human CYP2E1 by homology with the CYP102 haemoprotein domain: investigation of the interactions of substrates and inhibitors within the putative active site of the human CYP2E1 isoform. *Xenobiotica* **2000**, *30*, 1–25.
(20) Lewis, D. F. Molecular modeling of human cytochrome P450-substrate interactions. *Drug Metab. Rev.* **2002**, *34*, 55–67.
(21) Lewis, D. F.; Lake, B. G.; Dickins, M.; Goldfarb, P. S. Homology modelling of CYP3A4 from the CYP2C5 crystallographic template: analysis of typical CYP3A4 substrate interactions. *Xenobiotica* **2004**, *34*, 549–569.
(22) de Groot, M. J.; Ackland, M. J.; Horne, V. A.; Alex, A. A.; Jones, B. C. A novel approach to predicting P450 mediated drug metabolism. CYP2D6 catalyzed N-dealkylation reactions and qualitative metabolite predictions using a combined protein and pharmacophore model for CYP2D6. *J. Med. Chem.* **1999**, *42*, 4062–4070.
(23) de Groot, M. J.; Ackland, M. J.; Horne, V. A.; Alex, A. A.; Jones, B. C. Novel approach to predicting P450-mediated drug metabolism: development of a combined protein and pharmacophore model for CYP2D6. *J. Med. Chem.* **1999**, *42*, 1515–1524.
(24) de Groot, M. J.; Alex, A. A.; Jones, B. C. Development of a combined protein and pharmacophore model for cytochrome P450 2C9. *J. Med. Chem.* **2002**, *45*, 1983–1993.
(25) Park, J. Y.; Harris, D. Construction and assessment of models of CYP2E1: predictions of metabolism from docking, molecular dynamics, and density functional theoretical calculations. *J. Med. Chem.* **2003**, *46*, 1645–1660.
(26) Zamora, I.; Afzelius, L.; Cruciani, G. Predicting drug metabolism: a site of metabolism prediction tool applied to the cytochrome P450 2C9. *J. Med. Chem.* **2003**, *46*, 2313–2324.
(27) Cruciani, G.; Carosati, E.; De Boeck, B.; Ethirajulu, K.; Mackie, C.; Howe, T.; Vianello, R. MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist. *J. Med. Chem.* **2005**, *48*, 6970–6979.
(28) Zhou, D.; Afzelius, L.; Grimm, S. W.; Andersson, T. B.; Zauhar, R. J.; Zamora, I. Comparison of methods for the prediction of the metabolic sites for CYP3A4-mediated metabolic reactions. *Drug Metab. Dispos.* **2006**, *34*, 976–983.
(29) Anzenbacher, P.; Anzenbacherova, E. Cytochromes P450 and metabolism of xenobiotics. *Cell. Mol. Life Sci.* **2001**, *58*, 737–747.
(30) Danielson, P. B. The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans. *Curr. Drug Metab.* **2002**, *3*, 561–597.
(31) Sansen, S.; Yano, J. K.; Reynald, R. L.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. Adaptations for the Oxidation of Polycyclic Aromatic Hydrocarbons Exhibited by the Structure of Human P450 1A2. *J. Biol. Chem.* **2007**, *282*, 14348–14355.
(32) de Graaf, C.; Vermeulen, N. P.; Feenstra, K. A. Cytochrome p450 in silico: an integrative modeling approach. *J. Med. Chem.* **2005**, *48*, 2725–2755.
(33) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
(34) Rendic, S. Fujitsu Kyushu System Engineering Ltd. ADME Database. http://jp.fujitsu.com/group/fqs/services/lifescience/english/asp/admedb/index.html(accessed Feb 8, 2007).
(35) Rendic, S.; Di Carlo, F. J. Human cytochrome P450 enzymes: a status report summarizing their reactions, substrates, inducers, and inhibitors. *Drug Metab. Rev.* **1997**, *29*, 413–580.
(36) Rendic, S. Summary of information on human CYP enzymes: human P450 metabolism data. *Drug Metab. Rev.* **2002**, *34*, 83–448.
(37) *Cerius2*, version 4.10; Accelrys Software Inc.: San Diego, CA, 2005.
(38) *Insight II*, version 2000; Accelrys Software Inc.: San Diego, CA, 2000.
(39) Oda, A.; Yamaotsu, N.; Hirono, S. New AMBER force field parameters of heme iron for cytochrome P450s determined by quantum chemical calculations of simplified models. *J. Comput. Chem.* **2005**, *26*, 818–826.
(40) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

CI800001M