E_aMEAD: Activation Energy Prediction of Cytochrome P450 Mediated Metabolism with Effective Atomic Descriptors

Doo Nam Kim,[†,‡] Kwang-Hwi Cho,[§] Won Seok Oh,[†] Chang Joon Lee,[||] Sung Kwang Lee,[⊥] Jihoon Jung,[||] and Kyoung Tai No*[,†,||]

Bioinformatics and Molecular Design Research Center, Seoul 120-749, Korea, Graduate Program in Functional Genomics, Yonsei University, Seoul 120-749, Korea, Department of Bioinformatics and CAMD Research Center, Soongsil University, Seoul 156-743, Korea, Department of Biotechnology, Yonsei University, Seoul 120-749, Korea, and Department of Chemistry, Hannam University, 461-6 Jeonmin-Dong, Yuseong-Gu, Daejeon 305-811, Korea

Received January 12, 2009

In an effort to improve drug design and predictions for pharmacokinetics (PK), an empirical model was developed to predict the activation energies (E_a) of cytochrome P450 (CYP450) mediated metabolism. The model, E_aMEAD (Activation energy of Metabolism reactions with Effective Atomic Descriptors), predicts the E_a of four major metabolic reactions of the CYP450 enzyme: aliphatic hydroxylation, N-dealkylation, O-dealkylation, and aromatic hydroxylation. To build and validate the empirical model, the E_a values of the substrates with diverse chemical structures (394 metabolic sites for aliphatic hydroxylation, 27 metabolic sites for N-dealkylation, 9 metabolic sites for O-dealkylation, and 85 metabolic sites for aromatic hydroxylation) were calculated by AM1 molecular orbital (MO). Empirical equations, Quantitative Structure Activity Relationship (QSAR) models, were derived using effective atomic charge, effective atomic polarizability, and bond dipole moments of the substrates as descriptors. E_aMEAD is shown to accurately predict E_a with a correlation coefficient (R) of 0.94 and root-mean-square error (RMSE, unit is kcal/mol) of 0.70 for aliphatic hydroxylation, N-dealkylation, and O-dealkylation, and R of 0.83 and RMSE of 0.80 for aromatic hydroxylation, respectively. Physical origin and the role of the effective atomic descriptors of the models are presented in detail. With this model, the E_a of the metabolism can be rapidly predicted without any experimental parameters or time-consuming QM calculation. Regioselectivity prediction with our model is presented in the case of CYP3A4 metabolism. The reliability and ease of use of this model will greatly facilitate early stage PK predictions and rational drug design. Moreover, the model can be applied to develop the E_a prediction model of various types of chemical reactions.

INTRODUCTION

The prediction of the drug metabolism is crucial for the optimization of the metabolic properties of test compounds,[1] the elimination of compounds with inappropriate pharmacokinetic (PK) properties in early drug discovery stages,[2] the identification of precursor-product relationships,[3] and delineating the pharmacological effect.[4] Most of the drug metabolism occurs by cytochrome P450 isoenzymes.[5,6] Experimentally obtained regioselectivity of the CYP450 mediated metabolism contains information on the selectivity, reactivity, and steric hindrance of a compound bound to the enzymes. The regioselectivity provides valuable information on the preferentially oxidized site of a target molecule. Providing this data to medicinal chemists would vastly improve their ability to design prodrugs and more stable molecules *in vivo* by blocking the metabolism/oxidation of the target molecule. In this regard, the regioselectivity prediction is crucial in the rational design of good PK profile molecules.

To predict the regioselectivity of the CYP450 isozymes, both accessibility and reactivity are important, and several computational methods have been introduced.[6−24] Cruciani et al. presented a widely distributed package named Metasite, which used a combined method of accessibility and reactivity. They give the physical meaning in the prediction with structural information and radical formation energy.[12] Oh et al. proposed the combined model for the prediction of regioselectivity in cytochrome P450/3A4 mediated metabolism.[13] They used an ensemble catalyticphore-based docking method for the accessibility prediction and the QM E_a estimation of Korzekwa et al.[14] for the reactivity prediction. Jung et al. predicted the regioselectivity of CYP1A2-mediated phase I metabolism by a docking simulation for binding energy and a semiempirical molecular orbital (MO) calculation of Korzekwa et al.[14] for activation energy.[6]

To predict the reactivity for regioselectivity, there are three QM based empirical approaches. They are Jones and Korzekwa et al.,[7,8,14,21,25,26] Singh et al.,[17] and Olsen et al.[27] Jones and Korzekwa et al.[7,8,14,21,25,26] introduced p-nitrosophenoxy radical (PNR) and methoxy radical as surrogates

* Corresponding author phone: 82-2-393-9551; fax: 82-2-393-9554; e-mail: ktno@yonsei.ac.kr.
† Bioinformatics and Molecular Design Research Center.
‡ Graduate Program in Functional Genomics, Yonsei University.
§ Soongsil University.
|| Department of Biotechnology, Yonsei University.
⊥ Hannam University.

10.1021/ci900011g CCC: $40.75
Published on Web 06/22/2009

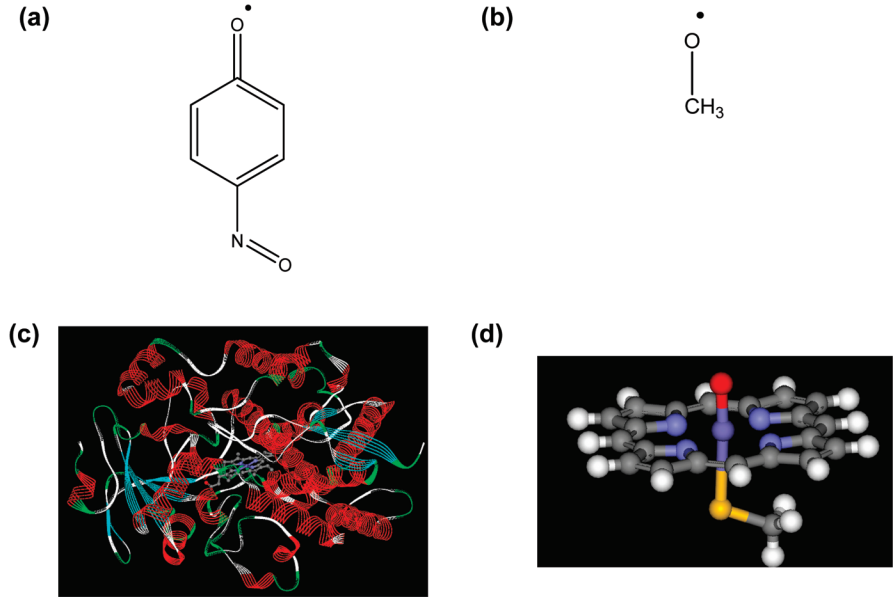


Figure 1. Oxygenating species for researching the metabolism by cytochrome P450: a) para-nitrosophenoxy radical, b) methoxy radical, c) human CYP 3A4,[28] d) compound I.

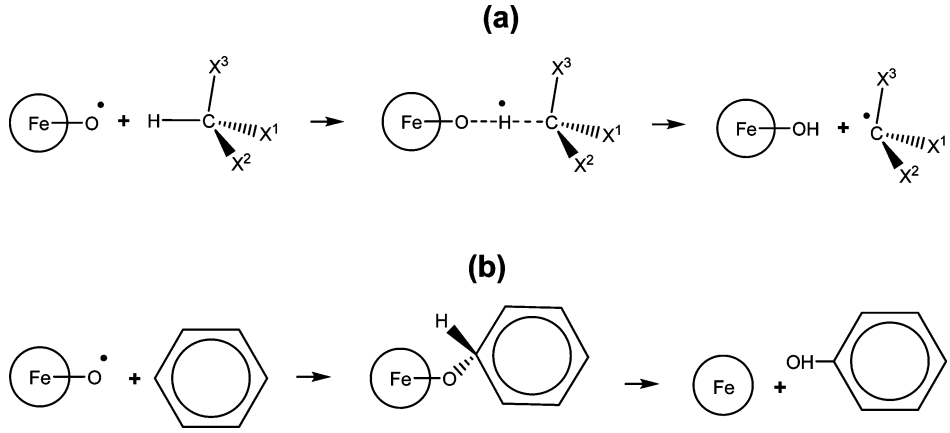


Figure 2. Schematic diagrams of metabolic reaction mediated by cytochrome P450: a) a hydrogen abstraction (one of Xs could be a heteroatom) and b) an aromatic hydroxylation. Each case contains reactant complex, transition state, and product complex sequentially. The circle with Fe−O represents the oxyferryl intermediate.

of the CYP450 for the hydrogen abstraction and aromatic hydroxylation, respectively (Figure 1). The enthalpy of reaction of hydrogen abstraction, $\Delta H_{rxn(Habs)}$, and aromatic hydroxylation, $\Delta H_{rxn(Habs)}$, were obtained using AM1 MO calculations. Empirical equations for the E_a estimation of the hydrogen abstraction, $\Delta H_{act(Habs)}$, and aromatic hydroxylation, $\Delta H_{act(Arom)}$, were given by

$$\Delta H_{act(Habs)} = 2.60 + 0.22 \quad \Delta H_{rxn(Habs)} + 2.38 \quad IP \tag{1}$$

$$\Delta H_{act(Arom)} = 21.91 + 0.61 \quad \Delta H_{rxn(Arom)} \tag{2}$$

where IP is the ionization potential of the product radical. By combining experimental data and estimated E_as for hydrogen abstraction and aromatic hydroxylation, the energy difference in E_as ($\Delta\Delta G_{measured}$) is given by

$$\Delta\Delta G_{measured} = 1.22\Delta H_{act(Habs)} - 1.1\Delta H_{act(Arom)} - 17.5 \tag{3}$$

Singh et al.[17] treated the reactivity of a hydrogen atom within a substrate as the reaction energy of hydrogen abstraction. The reaction energy of hydrogen abstraction on an isolated substrate is given by

$$\Delta H_{rxn} = \Delta H_2 - \Delta H_1 \tag{4}$$

The hydrogen abstraction reaction energy (ΔH_{rxn}) is the difference between the heat of formation of the substrate (ΔH_1) and that of its radical (ΔH_2).

Olsen et al.[27] developed an empirical E_a estimation model with QM descriptors, such as bond dissociation energies, for the hydrogen abstraction reaction of small organic molecules.

Li et al.[29] showed that compound I, a part of CYP450, represents the catalytic activity of CYP450. Hazan et al.[11] elucidated the factors that influence the regioselectivity of toluene hydroxylation, including hydrogen abstraction and aromatic hydroxylation, by modeling compound I (Figure 1d). Park et al.[30] predicted regioselectivity of CYP2E1 by calculating the E_a and showed that the TS geometry of the compound I-substrate complex is similar to that of the PNR-substrate complex.

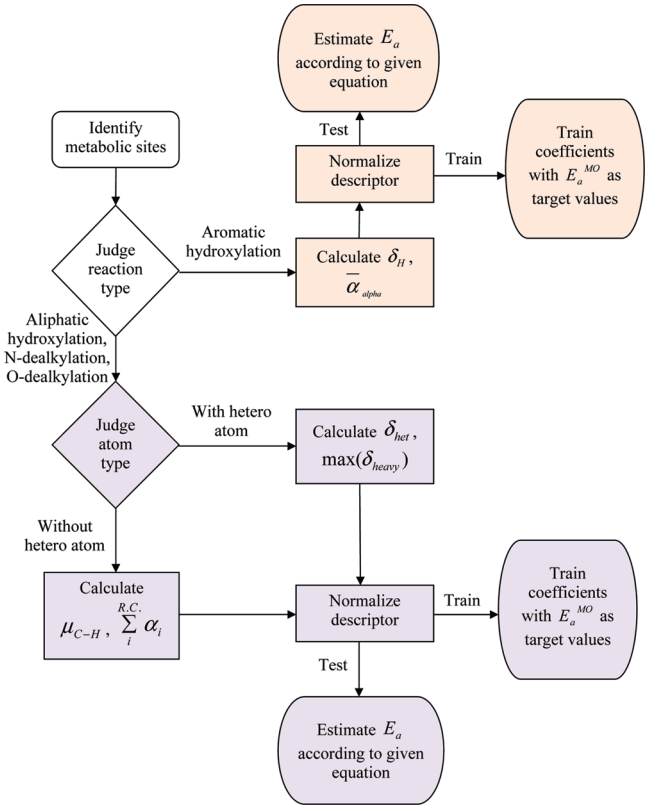


Figure 3. Flowchart of modeling for E_a.

Since the methods mentioned above included MO calculation in the E_a calculation in any form, it cannot be used for library design or virtual screening with millions of compounds.

The purpose of this work is to develop E_a estimation empirical models for major metabolic reactions. To predict AM1 QM values for E_a of aliphatic hydroxylation, N-dealkylation, O-dealkylation, and aromatic hydroxylation, ligand-based empirical models were derived from MPEOE (Modified Partial Equalization of Orbital Electronegativity)-derived effective atomic descriptors. Then another empirical equation that correctly scales between the E_a of aliphatic hydroxylation, N-dealkylation, and O-dealkylation and the E_a of aromatic hydroxylation was generated for regioselectivity prediction for molecules which has multiple metabolic sites of various reaction types. Finally this scaling equation was applied to regioselectivity prediction of 3A4 substrates and compared with experimental data.

Table 2. Correlations of Various Effective Atomic Descriptors with E_a^{MO} for Aliphatic Hydroxylation, N-Dealkylation, and O-Dealkylation

	training set		test set	
descriptors	R[a]	RMSE (kcal/mol)[b]	R[a]	RMSE (kcal/mol)[b]
δ_H[c]	0.75	0.88	0.69	1.00
δ_C[c]	0.52	1.15	0.47	1.22
$\delta_C - \delta_H$[c]	0.49	0.49	0.44	1.23
$\Sigma_i^{R.C.}\delta_i$[c]	0.37	1.25	0.39	1.27
μ_{C-H}[c]	0.88	0.63	0.85	0.72
$\Sigma_i^{R.C.}\alpha_i$[c]	0.67	1.00	0.60	1.10
δ_{het}[d]	0.82	1.51	0.96	0.77
$\max(\delta_{heavy})$[d]	0.57	2.16	0.55	2.43

[a] R is a correlation coefficient. [b] RMSE stands for root mean squared error. [c] Correlation of this descriptor with E_a^{MO} came from the data set without a heteroatom in the reaction center. [d] Correlation of this descriptor with E_a^{MO} came from the data set with a heteroatom in the reaction center.

METHOD

A. Modeling Procedures.

Overall Modeling Procedures. In our empirical model, E_aMEAD, the metabolic reaction was described with effective atomic descriptors of the atoms in the reaction center, but the influence of the steric hindrance, caused by the repulsion between the substrate and CYP450 protein, was not taken into account. The steric factor could be obtained by the docking simulation between the substrate and CYP450 protein with empirical potential energy functions.

Aliphatic hydroxylation, N-dealkylation, and O-dealkylation have a different TS from aromatic hydroxylation as shown in Figure 2. Therefore, two different E_aMEAD models, one for aliphatic hydroxylation, N-dealkylation, and O-dealkylation and the other for aromatic hydroxylation were developed. In the models, the E_a was described by multilinear equations with the effective descriptors derived from the atomic properties alone. These effective descriptors were constructed with in-house software. The coefficients of the equations were optimized to generate the least amount of error between the E_a determined from the models and that from the QM calculation of a given data set by a linear regression. The detailed procedures are presented in Figure 3.

Molecular Orbital (MO) Calculation. The TS can be determined by searching for first-order saddle points on the potential energy surface. The saddle point is a point of a function or surface which is a stationary point but not an

Table 1. Descriptors Used for Developing and Applying E_aMEAD

reaction type	class	descriptor	meaning
aliphatic hydroxylation, N-dealkylation, O-dealkylation	no heteroatom in reaction center	μ_{C-H}[a]	dipole moment of C−H bond
		$\Sigma_i^{R.C.}\alpha_i$	the sum of polarizabilities of the reaction center atoms
	with a heteroatom in reaction center	δ_{het}	a net atomic charge of heteroatom in the reaction center
		$\max(\delta_{heavy})$	a maximum charge among heavy atoms, which are alpha to the reaction carbon
aromatic hydroxylation	all classes[b]	δ_H	the net atomic charge of hydrogen
		$\bar{\alpha}_{alpha}$	mean polarizabilities of C_α

[a] $(\Sigma_{n=x,y,z}\{(\mu_C + \mu_H)_{in\ n\ axis}\}^2)^{1/2}$ with μ_i = (distance between center of mass of the bond and i) × (δ_i), where i could be carbon and hydrogen. [b] Aromatic carbons at the ortho and para positions of a functional group, aromatic carbons at the meta positions of a functional group, and aromatic carbons having zero, two, or three functional groups.

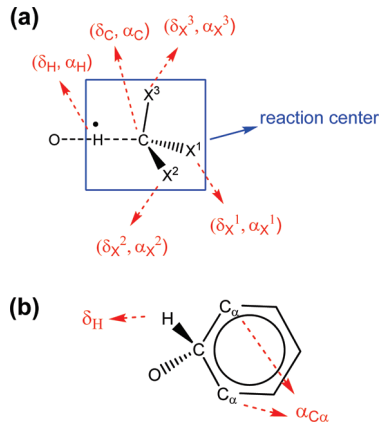


Figure 4. Descriptors for E_aMEAD: a) a hydrogen abstraction (one of Xs could be a heteroatom) and b) an aromatic hydroxylation. Each descriptor (in red) representing the character of each atom constitutes effective descriptor solely for in combination with other descriptors. For example, δ_H and δ_C constitute μ_{C-H}. α_H, α_C, α_{X^1}, α_{X^2}, and α_{X^3} constitute $\Sigma_i^{R.C.}\alpha_i$. δ_X that has a heteroatom of X, is δ_{het}. max(δ_{heavy}) is a maximum charge among δ_X that has a heavy atom of X. Mean value of two $\alpha_{C\alpha}$ is $\bar{\alpha}_{alpha}$.

Table 3. Correlations of Various Effective Atomic Descriptors with E_a^{MO} of Aromatic Hydroxylation

	training set		test set	
descriptors	R	RMSE (kcal/mol)	R	RMSE (kcal/mol)
$\delta_H{}^a$	0.69	0.95	0.39	1.20
$\delta_C{}^a$	0.53	1.11	0.20	1.09
$\delta_C - \delta_H{}^a$	0.49	1.14	0.17	1.28
$\bar{\alpha}_{alpha}{}^a$	0.05	1.31	0.14	1.29
$\Sigma_i^{substituents}\alpha_i{}^a$	0.58	1.06	0.02	1.30
$\Sigma_i^{substituents}\delta_i{}^a$	0.28	1.25	0.08	1.30

[a] Correlation of this descriptor with E_a^{MO} came from the data set of aromatic carbons with 0, 2, and 3 substituents.

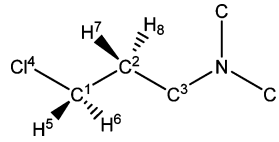


Figure 5. Structure for example of effective descriptors for E_aMEAD.

extremum. All TS structures we used were obtained by the quadratic synchronous transit (QST)[31] with AM1[32] MO calculation implemented in Gaussian03.[33] For QST calculation, the reactant complex and the product complex should be defined together with a presumed TS structure. All the TS structures we used were confirmed by frequency analysis. It is known that AM1 is less accurate than *ab initio* or DFT. However, with the following three reasons, adopting AM1 to calculate TS of druglike molecules was regarded as reasonable in a relatively small amount of time. First, Korzekwa et al. calculated E_a with AM1, and their results are correlated highly with experimental results.[14,25] Second, the usage of AM1 for TS calculation showed a high correlation (0.97 R, statistics with 23 atoms without 1 outlier) with that of DFT.[27] Third, it is assumed that the systematical errors hopefully have about the same magnitude for the calculation of the activation energy of different metabolic sites.

As the CYP450 enzyme was too large to be effectively treated with QM calculations, the para-nitrosophenoxy radical (PNR), Figure 1a, was used as the surrogate molecule of the oxygenating species, compound I, which was responsible for the catalytic property of CYP450. Using PNR and AM1, as a surrogate molecule and computational method, respectively, have been shown to be a proper approximation.[14,25,30] This usage is further corroborated by the fact that the E_a of AM1 calculations with PNR exhibited a high correlation (0.95 R, statistics with 24 atoms) to that of B3LYP calculations with compound I.[27]

A Model for Aliphatic Hydroxylation, N-Dealkylation, and O-Dealkylation. The E_a of the hydrogen abstraction is a good criterion for estimating the regioselectivity of aliphatic hydroxylation, N-dealkylation, and O-dealkylation of metabolism reactions (Figure 2a).[7,30] Because the hydrogen abstraction is the rate determining step in those reactions. There is a great influence by the heteroatom in the C−H bond breakage; therefore, hydrogen abstraction reaction type was divided into two classes dependent upon whether the bond had a heteroatom in the reaction center or not. Subsequently, linear equations with two normalized effective atomic descriptors for each class were derived to estimate the E_a in various chemical environments (averages and standard deviations to normalize each descriptor are presented in Table S1 in the Supporting Information).

The E_a of hydrogen abstraction without heteroatoms in the reaction center, E_a^{Habs}, was obtained as

$$E_a^{Habs} = 28.50 - 2.22\mu_{C-H} + 1.12\sum_i^{R.C.}\alpha_i \qquad (5)$$

where μ_{C-H} was the representative of the EI among δ_H, δ_C, $\delta_H - \delta_C$, $\Sigma_i^{R.C.}\alpha_i$, and μ_{C-H} (Table 1, Table 2). The prediction

Table 4. Correlation between E_a^{MO} and Estimated E_a of the Metabolism Reactions

		train				test			
class	standard beta	N	R	RMSE[a]	P value	N	R	RMSE[a]	P value
				H Abstraction					
without hetero (eq *5*)	−1.65, 0.84	224	0.95	0.43	<0.0001	116	0.92	0.55	<0.0001
with hetero (eq *6*)	0.72, 0.39	61	0.91	1.14	<0.0001	29	0.92	1.12	<0.0001
				Aromatic Hydroxylation					
ortho, para (eq *7*)	−0.59, −0.98	16	0.71	0.95	0.009	6	0.40	0.95	0.437
meta (eq *8*)	−1.32, −1.60	8	0.88	0.30	0.026	3	0.59	0.52	0.597
0, 2, 3 substituents (eq *9*)	−1.16, −0.72	33	0.87	0.65	<0.0001	19	0.80	0.78	<0.0001

[a] The unit of RMSE is kcal/mol.

model for the reaction center with high electronegative atoms with μ_{C-H} and $\Sigma_i^{R.C.}\alpha_i$ (as for the previous class) gave a low correlation, 0.67 R, resulting in the necessity to introduce the descriptor for the heteroatom. Therefore, the net atomic charge of the heteroatom (δ_{het}) and max (δ_{heavy}) were introduced in the model. δ_{het} is highly correlated, and the max (δ_{heavy}) is moderately correlated with E_a^{MO}, E_a which is quantum mechanically calculated (Table 2).

$$E_a^{Habs\_Het} = 25.94 + 1.88\delta_{het} + 1.03 \max(\delta_{heavy}) \quad (6)$$

As the coefficients of δ_{het} and max(δ_{heavy}) indicate, negative charge of heavy neighbor atoms lowers the E_a barrier.

A Model for Aromatic Hydroxylation. Studies[8,14] have determined that the tetrahedral (Td) intermediate formation is a good indicator for estimating the regioselectivity of aromatic hydroxylation of metabolism reactions (Figure 2b).

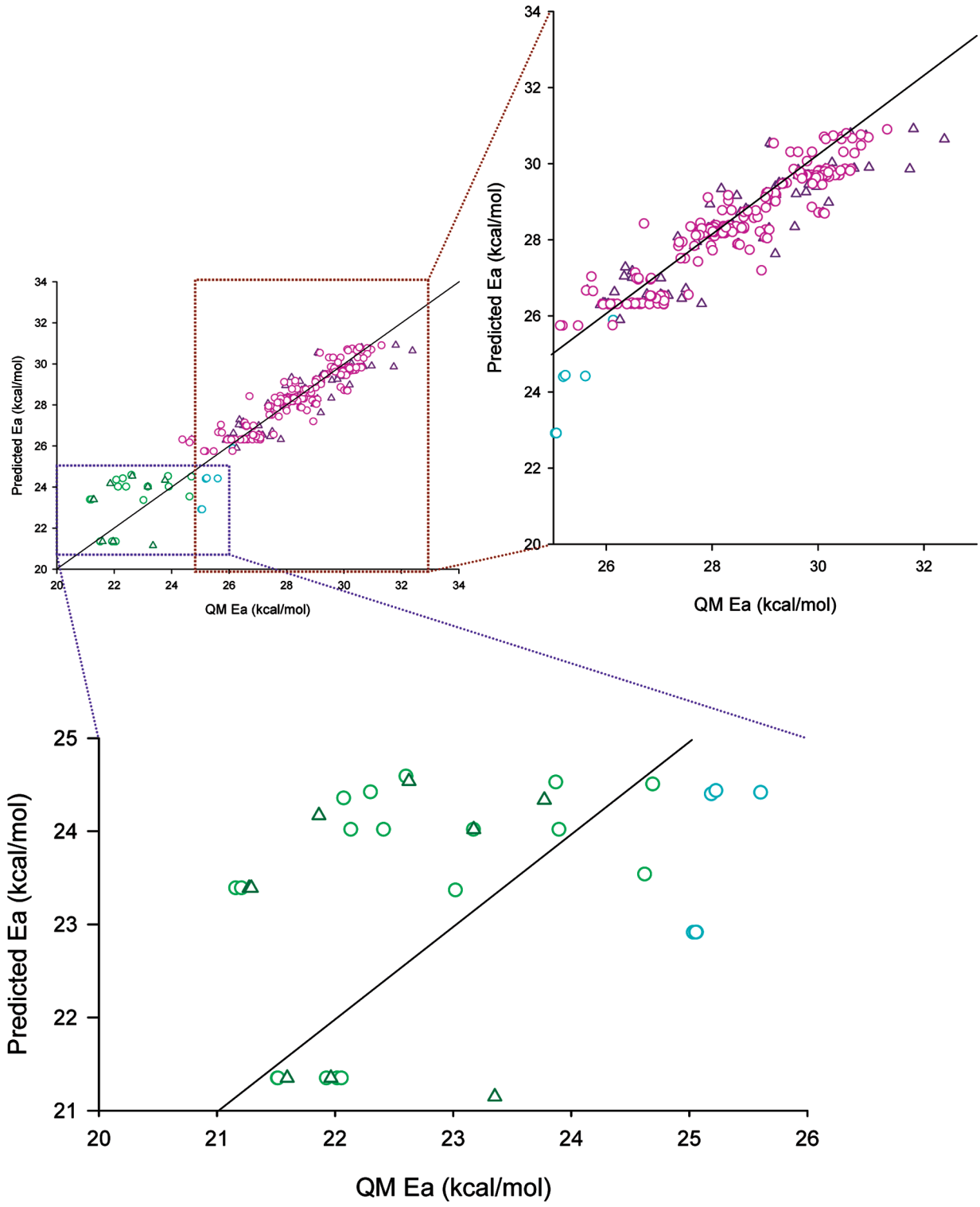


Figure 6. Correlation between E_a^{MO} and estimated E_a of the hydrogen abstraction. Circle and triangle mean training and test sets, respectively (aliphatic hydroxylation: pink and dark pink; N-dealkylation: green and dark green; O-dealkylation: cyan and dark cyan).

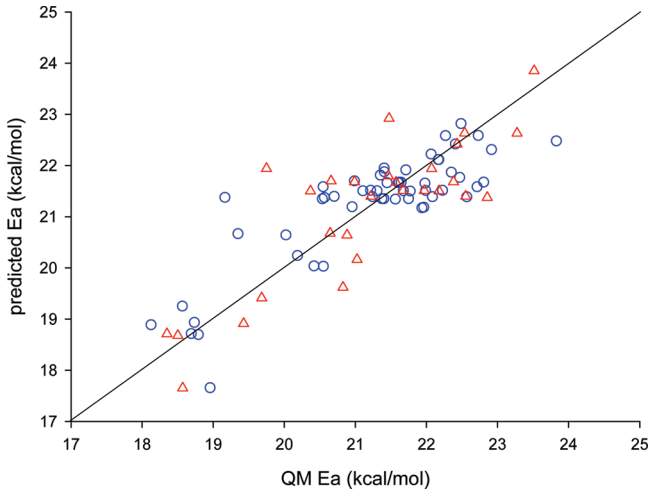


Figure 7. Correlation between E_a^{MO} and estimated E_a of aromatic hydroxylation. Blue circle and red triangle mean training and test sets, respectively.

Therefore, the methodology which we used to model aliphatic hydroxylation, N-dealkylation, and O-dealkylation was applied to model this Td formation. It is important to note that functional groups attached to the aromatic carbons influence the molecule in various modes based on their position. Aromatic carbons at the ortho and para positions of a functional group influence the reaction differently to the aromatic carbons at meta positions of a functional group. Therefore, the data sets were divided into three classes: aromatic carbons at the ortho and para positions of a functional group, aromatic carbons at the meta positions of a functional group, and aromatic carbons having zero, two, or three functional groups.

The E_a models for aromatic carbons having a substituent at the ortho or para positions ($E_a^{ar\_o,p}$), the meta positions ($E_a^{ar\_m}$) from the reaction center, and 0, 2, or 3 substituents ($E_a^{ar\_0,2,3}$) were obtained as follows

$$E_a^{ar\_o,p} = 21.34 - 0.75\delta_H - 1.24\bar{\alpha}_{alpha} \quad (7)$$

$$E_a^{ar\_m} = 22.14 - 0.68\delta_H - 0.83\bar{\alpha}_{alpha} \quad (8)$$

$$E_a^{ar\_0,2,3} = 21.02 - 1.49\delta_H - 0.92\bar{\alpha}_{alpha} \quad (9)$$

A Combined Model. To predict the regioselectivity of substrates that have metabolic sites of both hydrogen abstraction and Td formation, it is necessary to combine the models for each reaction. The reason for this is that the difference in mechanism of these two reactions gives different E_a magnitude. A combined model, as in Korzekwa et al.,[14] is suggested in the Results and Discussion section. The coefficients of the combined equation were obtained using the experimental regioselectivities of metabolism of ten organic molecules that have both reaction types. These molecules are relatively free from steric hindrance due to their small size, as in the research of Higgins et al.[21] The regioselectivities of these molecules reflect their individual reaction rates which were in inverse proportion to the E_a as the Arrhenius equation indicates.

Regioselectivity Prediction Based on E_a only. The applicability of the E_aMEAD method on the regioselectivity prediction of druglike molecules was demonstrated by comparing it with the experimental data.[17,27] Since the intrinsic reactivity of the various sites is our main interest, we focused on metabolites of CYP3A4. This enzyme has a very broad specificity (over 50% of the marketed drugs are metabolized by this enzyme) due to the fact that it does not have severe restrictions on the orientation of the ligand in the active site.[34,35] For example, the heme of CYP3A4 has greater accessibility to the active site than does that of CYP2C9.[35]

B. Further Explanation of the Modeling.

Physical Assumptions of Models. The E_a, a measure of reactivity of a certain type of reaction, can be described by the electron redistribution at the reaction center and the steric energy barrier between reactants. The reaction center consists of the center carbon and its neighboring atoms. Theoretically, the E_a of a reaction can be obtained from the difference in energy between the reactant complex and the TS. The energies of each state could be calculated quantum mechanically. However, the MO calculations are extremely time demanding, even with the use of a semiempirical method, particularly in regard to the TS calculation. Therefore, it is necessary to develop an empirical method for E_a prediction.

Figure 2a,b shows the schematic diagrams of the hydrogen abstraction and aromatic hydroxylation mediated by CYP450, respectively. It was assumed that the location of the TS of the hydrogen abstraction could be described with i) the Electrostatic Interaction (EI) between reactants and ii) the Electron ReDistribution (ERD) mainly among the atoms in the reaction center.[36] To describe the EI and ERD, the net atomic charges and polarizabilities of the reaction center atoms were derived by MPEOE.[37−39] As shown in Figure 4a, the reaction center atoms are the reactive hydrogen atom (H), reactive carbon atom bonded to the H (C), and the atoms bonded to the C (X^1, X^2, and X^3). It was assumed that the atoms that are not bonded to the reactive carbon influence the reaction center through the connecting bonds as the MPEOE charge calculation method follows. The net atomic charges and polarizabilities were used to obtain the effective descriptors to represent each metabolism chemical reaction (Table 1). Various physically realistic functional forms have been tested for selecting the descriptors, and the correlation coefficients were calculated to test the discriminating power of the descriptors (Table 2 and 3). For aliphatic hydroxylation, N-dealkylation, and O-dealkylation, μ_{C-H}, $\Sigma_i^{R.C.}\alpha_i$, δ_{het}, and max(δ_{heavy}) were used for representing the strength of the C−H bond, stabilization of the reaction center, and the effects heteroatom in the reaction center (latter two descriptors), respectively. The MO effect on aromatic hydroxylation was considered by introducing electrostatic force and stabilization of the TS. For aromatic hydroxylation, δ_H and $\bar{\alpha}_{alpha}$ were used each representing the measure of the electronegativity of the highest occupied molecular orbital (HOMO) of the aromatic systems and stabilization of the TS, respectively.

MPEOE (Modified Partial Equalization of Orbital Electronegativity). All effective descriptors for E_aMEAD were constructed with atomic properties derived from MPEOE (Modified Partial Equalization of Orbital Electronegativity).[37−39] MPEOE is a modified method for the determination of net atomic charges, based on an existing scheme utilizing the PEOE (partial equalization of orbital electronegativity method).[40] This method introduces (as

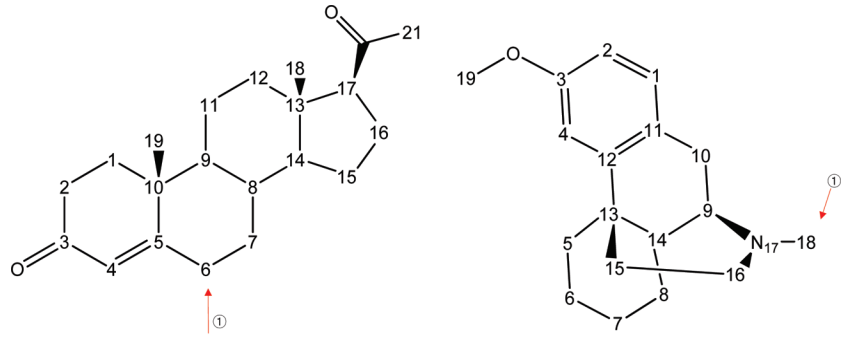


Figure 8. Chemical structures of two CYP3A4 substrates used for comparison of E_aMEAD with the QM descriptors based prediction. Arrows indicate the major metabolic sites.

Table 5. Application of E_aMEAD to Predict the Regioselectivity of Two CYP3A4 Substrates—A Comparison of E_aMEAD Based Prediction with QM Descriptors Based Prediction

		QM based prediction of Olsen et al.[a]		E_aMEAD based prediction	
no.[b]	metabolic site	E_a^{DFT} [c]	E_a^{AM1} [d]	$E_a^{E_aMEAD}$ [e]	$\Delta\Delta G_{pred.}^{rel.}$ [f]
			Progesterone		
1		13.4, 14.1	13.4, 13.7	29.2	5.3
2		10.7, 16.1	14.0, 15.8	26.7	2.8
6	major[42]	8.3, 15.1	10.4, 14.1	23.9	0
7		14.1, 14.8	13.8, 14.1	28.6	4.7
8		11.3	12.1	26.8	2.9
9		10.5	10.0	26.8	2.9
11		13.0, 13.2	12.8, 13.0	28.5	4.6
12		14.2, 15.1	13.6, 15.2	28.4	4.5
14		9.7	9.6	26.6	2.7
15		13.8, 14.7	14.4, 14.8	28.5	4.5
16		13.4, 14.4, 14.5, 15.4	14.1, 14.2, 14.2, 15.7	29.0	5.1
17		6.3, 11.1	11.3, 12.8	28.3	4.4
18		16.3, 16.8, 17.8	15.8, 15.8, 16.7	29.7	5.8
19		16.6, 16.9, 17.2	16.0, 16.3, 16.5	29.9	6.0
21		14.4, 16.0, 20.6	17.2, 18.0, 18.2	25.7	1.8
			Dextromethorphan		
1		N/A[g]	N/A[g]	21.8	7.2
2		N/A[g]	N/A[g]	20.0	5.4
4		N/A[g]	N/A[g]	19.8	5.3
5		13.6, 14.1	12.8, 13.7	28.9	7.3
6		14.3, 14.6	13.5, 13.7	28.5	6.9
7		14.3, 14.5	13.1, 13.8	28.4	6.9
8		13.9, 14.1	13.2, 13.8	28.7	7.1
9		10.6	8.8	24.9	3.6
10		6.9, 10.9	9.0, 11.2	23.5	2.2
14		12.2	11.8	25.7	4.3
15		14.3, 15.6	14.0, 14.7	29.4	7.8
16		6.3, 11.9	5.9, 10.1	24.3	3.0
18	major[43]	7.1, 14.1, 15.8	8.0, 12.3, 13.8	21.1	0
19		12.1, 12.2, 17.7	13.9, 13.9, 17.4	23.3	2.0

[a] The method of Olsen et al. gives more than one value of E_a^{DFT} or E_a^{AM1} when the reaction carbon has more than one hydrogen or the molecule is in various conformations. [b] no. indicates the atom numbers designated in Figure 8. [c] Predicted E_a (in kcal/mol) using QM descriptors derived from B3LYP calculation. [d] Predicted E_a (in kcal/mol) using QM descriptors derived from AM1 calculation. [e] The E_a, which was predicted by E_aMEAD. [f] Relative $\Delta\Delta G_{predicted}$. $\Delta\Delta G_{pred.}^{rel.}$ of a given metabolic site was the difference between the E_aMEAD predicted E_a and the lowest E_a among all metabolic sites in each molecule. The lowest relative energy (in the 6th column) of each molecule is 0 kcal/mol. [g] Not applicable since the approach of Olsen et al. could not cover the aromatic hydroxylation.

constraints) experimental dipole and, where possible, quadrupole moments of a set of molecules chosen to resemble the different atomic environments found in proteins. As a consequence, the atomic charges reproduce accurately the experimental values of molecular dipole and quadrupole moments. The electrostatic potential surface, calculated with the new net atomic charge set, compares well to a surface obtained from a moderate level *ab initio* calculation (6-31G**). However, it is observed that the new charges are of twice the magnitude of those obtained from investigations in which the charges are determined from a least-squares fit to the calculated *ab initio* electrostatic potential (potential-derived method). The methodology leads to a direct calculation of the permanent molecular charge distribution represented as a set of distributed monopoles, dependent only on the connectivity of the molecule. Hence, the method circumvents assumptions concerning the transferability of charges and can therefore approximately describe any

Table 6. Data Used to Derive the Combined Model Described in Eq 10

molecule	positions	E_a^{Habs} or $E_a^{Habs\_Heta}$	E_a^{arb}	$\Delta\Delta G_{measured}$ [c]	$\Delta\Delta G_{predicted}$ [d]	res.[e]
ethylbenzene	para/benzylic	25.5	21.7	−3.8[44]	−2.9	−0.9
o-xylene	para/benzylic	26.5	21.6	−2.8[45]	−1.9	−0.9
methylbenzene	meta/benzylic	26.4	22.2	−2.5[46]	−2.5	0.0
p-xylene	ortho/benzylic	26.1	21.7	−2.4[45]	−2.3	−0.1
m-xylene	aromatic/benzylic	26.2	21.2	−1.0[45]	−1.8	0.8
methylbenzene	ortho/benzylic	26.4	21.9	−0.8[46]	−2.2	1.4
methoxybenzene	meta/o-dealkyl	28.8	22.4	−0.6[46]	−0.4	−0.2
1-methoxy-2-methyl-benzene	para/o-dealkyl	28.4	21.0	0.3[21]	0.5	−0.2
methoxybenzene	ortho/o-dealkyl	28.8	21.5	0.3[46]	0.4	−0.1
methoxybenzene	para/o-dealkyl	28.8	21.2	1.0[46]	0.7	0.3

[a] E_a^{MO} of hydrogen abstraction. [b] E_a^{MO} of aromatic hydroxylation. [c] Measured difference in energies based on regioselectivity which is derived by the Arrhenius relationship. [d] Predicted difference in energies based on regioselectivity which is derived by eq 10. [e] The residual between $\Delta\Delta G_{measured}$ and $\Delta\Delta G_{predicted}$.

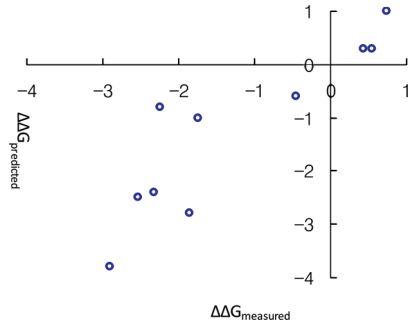


Figure 9. Correlation between the measured difference in energies based on regioselectivity and the predicted difference in energies, based on regioselectivity. The correlation coefficient is 0.89.

redistribution of charge density upon assembly of amino acid residues to form a polypeptide.

Examples of Effective Descriptors. To predict the E_a for aliphatic hydroxylation at C^2 in (3-chloropropyl)dimethylamine, μ_{C-H} and $\Sigma_i^{R.C.}\alpha_i$ are needed (Figure 5, Table 1). To derive μ_{C-H}, the charge of C^2, 0.0214, and the charge of H^7, 0.0117, are needed. These two values derive an initial μ_{C-H}. The initial μ_{C-H} is normalized by averages and standard deviations of training set (Table S1) and then became a final μ_{C-H}, −0.5524. To derive $\Sigma_i^{R.C.}\alpha_i$, polarizability values of C^1, C^2, C^3, H^7, and H^8 are summed into initial $\Sigma_i^{R.C.}\alpha_i$. The initial $\Sigma_i^{R.C.}\alpha_i$ is normalized like the above way and became a final $\Sigma_i^{R.C.}\alpha_i$, 0.0954. Therefore with the eq 5, the predicted E_a at C^2 is 29.83 kcal/mol.

To predict the E_a for aliphatic hydroxylation at C^1 in (3-chloropropyl)dimethylamine, δ_{het} and max(δ_{heavy}) are needed (Table 1). The charge of Cl^4, −0.1958, becomes an initial δ_{het} which is normalized like the above way to become a final δ_{het}, 0.8444. To derive max(δ_{heavy}), charges of C^2, 0.0550, and Cl^4, −0.1958, are compared. The higher value (0.0550) is normalized like the above way and became a final max(δ_{heavy}), 0.4984. Therefore with eq 6, the predicted E_a at C^1 is 28.04 kcal/mol.

Compound Sets for Training and Test. In order to model the aliphatic hydroxylation, O-dealkylation, and N-dealkylation, the E_a of various aliphatic carbons were considered for the MO calculation (Table 4). The molecules contained methyl, primary, secondary, and tertiary carbons with diverse functional groups such as adjacent sp^2 carbon, aromatic, thiol, amines, hydroxyl, and halogens. These compounds cover more diverse chemical environments than previous predictions.[7,14,27] For the modeling of the aromatic hydroxylation, the E_a of the variously substituted benzenes were considered for the MO calculation (Table 4). The substituted functional groups are methyl, ethyl, propyl, amine, hydroxyl, carbonyl, methoxyl, thiol, cyano, and halogens. The compounds used in this study are listed in part 2 of the Supporting Information.

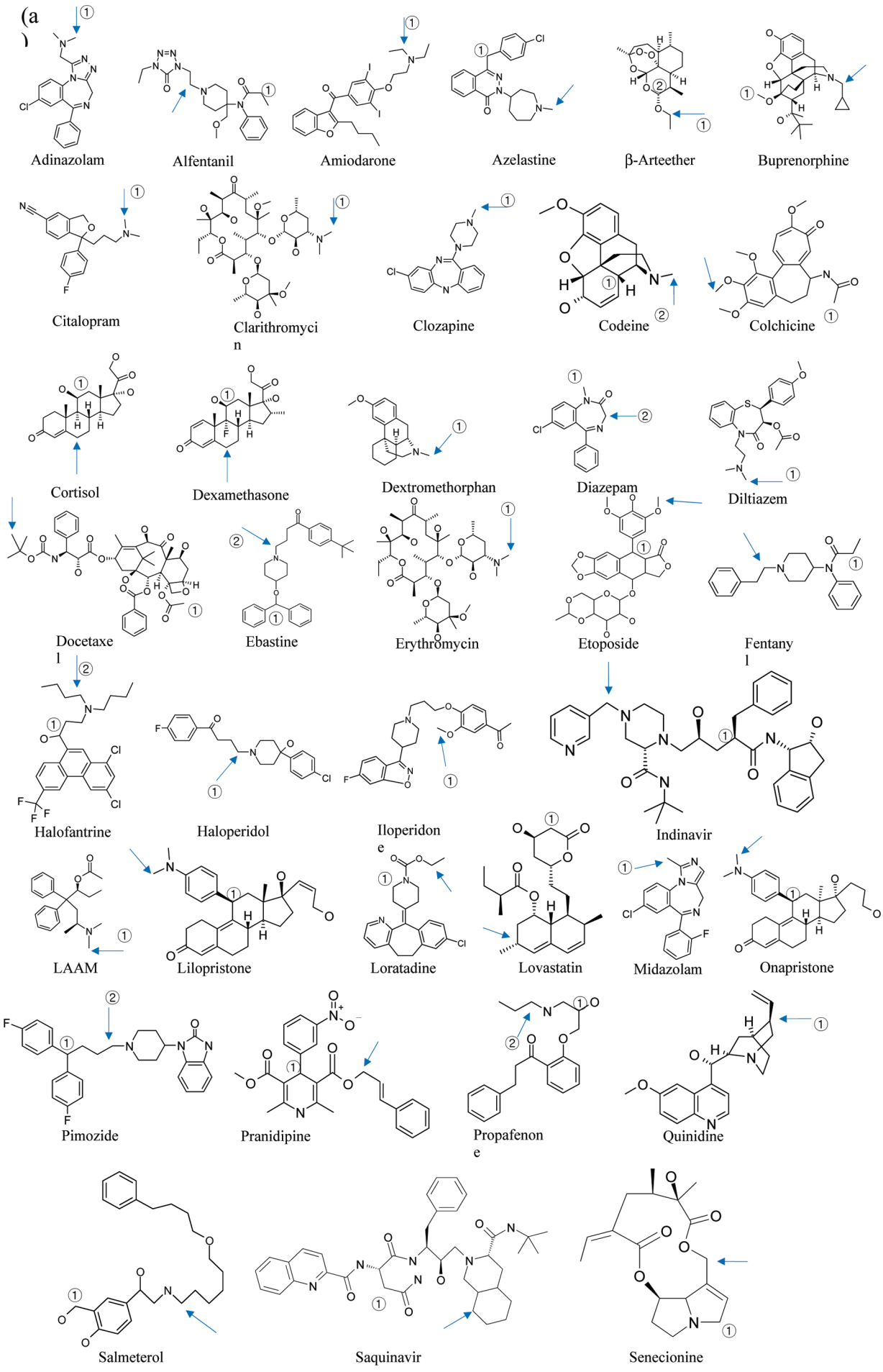
RESULTS AND DISCUSSION

***E_a* of Aliphatic Hydroxylation, N-Dealkylation, O-Dealkylation (without Heteroatom).** The predicted E_a shows a good correlation with the E_a obtained with the MO calculation, E_a^{MO}, as presented in Figure 6. The number of metabolic sites (N), the correlation coefficients of the training and test sets, the relative importance of each descriptor in the model (Std. beta), and the probability that a variable would assume a value greater than or equal to the observed value strictly by chance (P values) are summarized in Table 4. Of 340 metabolic sites, including training and test sets, 325 are within the chemical accuracy (i.e., 1 kcal/mol).[41]

During the hydrogen abstraction process, the C−H bond is elongated at the TS. As the C−H bond weakens, the TS barrier decreases, which consequently lessens the E_a. This strength of the C−H bond was represented by the dipole moment of the C−H bond, μ_{C-H}. The greater the polarity which means a bigger dipole, the easier the bond is broken. So as the bond becomes weaker, the reaction will be easier. This theory of vulnerability of the bond explains why μ_{C-H} has the negative coefficient sign in the eq 5. The sum of the effective atomic polarizabilities of reaction center atoms, $\Sigma_i^{R.C.}\alpha_i$, shows a high correlation with the E_a^{Habs}. As $\Sigma_i^{R.C.}\alpha_i$ increases, the reaction center is stabilized (the reactant is also stabilized) and at the same time the C−H bond energy increases. As a result, E_a increased as the $\Sigma_i^{R.C.}\alpha_i$ increases. The selected descriptors, μ_{C-H} and $\Sigma_i^{R.C.}\alpha_i$, are physically realistic in their description of the E_a of the aliphatic hydroxylation, N-dealkylation, and O-dealkylation of those aliphatic molecules without high electronegative heteroatoms in the reaction center.

***E_a* of Aliphatic Hydroxylation, N-Dealkylation, and O-Dealkylation (with Heteroatom).** The heteroatom in the reaction center is seen to weaken the C−H bonds. The predicted $E_a^{Habs\_Het}$ is highly correlated with the E_a^{MO} (Table 4 and Figure 6). Of 90 metabolic sites, including training and test sets, 61 are within the chemical accuracy.

***E_a* of Aromatic Hydroxylation.** For the aromatic hydroxylation, the oxygenation radical, Figure 1d, approaches the reaction center, forms a Td complex, and


(a)
Adinazolam
Alfentanil
Amiodarone
Azelastine
β-Arteether
Buprenorphine
Citalopram
Clarithromycin
Clozapine
Codeine
Colchicine
Cortisol
Dexamethasone
Dextromethorphan
Diazepam
Diltiazem
Docetaxel
Ebastine
Erythromycin
Etoposide
Fentanyl
Halofantrine
Haloperidol
Iloperidone
Indinavir
LAAM
Lilopristone
Loratadine
Lovastatin
Midazolam
Onapristone
Pimozide
Pranidipine
Propafenone
Quinidine
Salmeterol
Saquinavir
Senecionine

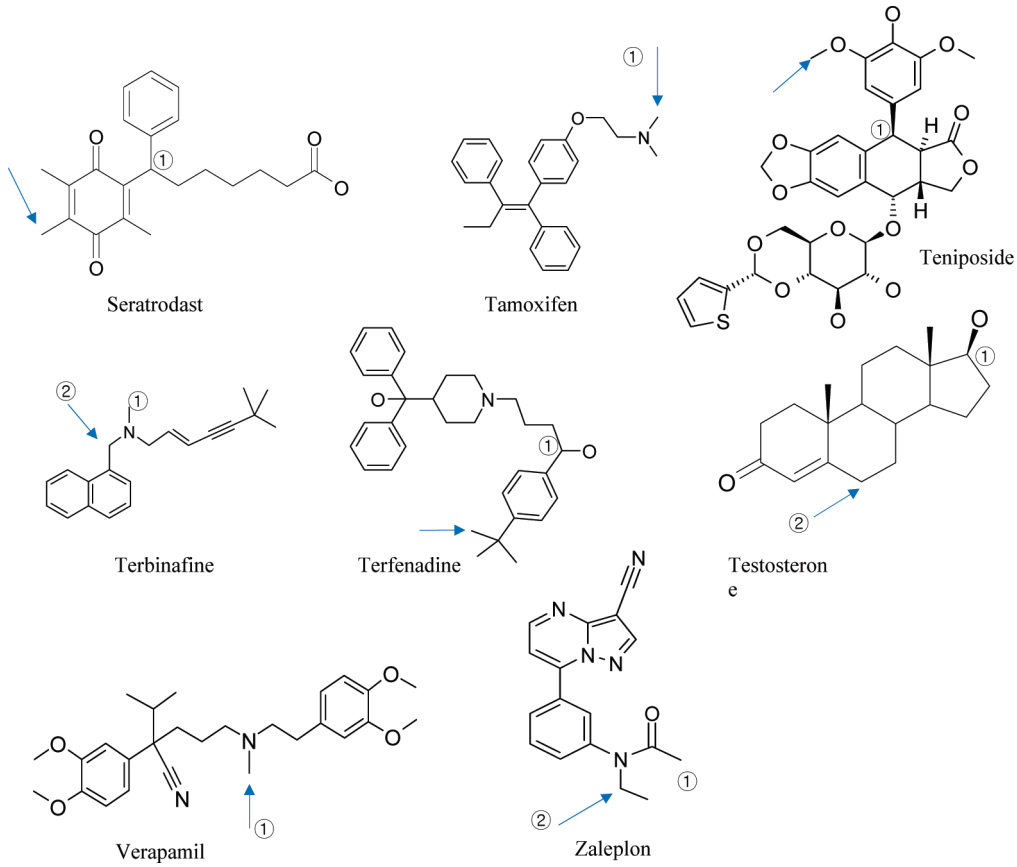


Figure 10. Chemical structures of 46 3A4 substrates used for comparison of E_aMEAD based prediction with reaction energy based prediction. Blue arrows indicate the major experimental metabolic sites.

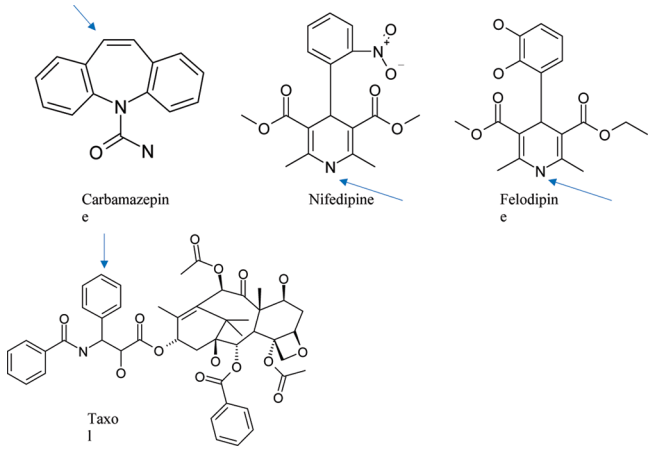


Figure 11. Chemical structures of four CYP3A4 substrates which were not used for comparison of E_aMEAD based prediction with reaction energy based prediction. These molecules were excluded because these are not metabolized by hydrogen abstraction from sp^3 carbon. Blue arrows indicate the major experimental metabolic sites.

then crosses the reaction energy barrier, where the minimum energy barrier is the E_a of the forward reaction. Since the reaction center is a part of the conjugate system, the electron population of hydrogen (δ_H) bonded to the center carbon of the Td shows the measure of the electronegativity of the highest occupied molecular orbital (HOMO) of the aromatic systems. δ_H shows the highest correlation with E_a by that descriptor alone (Table 3). As the δ_H increases, the E_a decreases. $\bar{\alpha}_{alpha}$ shows the stabilization of the TS. As the $\bar{\alpha}_{alpha}$ increases, the E_a decreases. An energy barrier as shown in eqs 7−9, the E_a

Table 7. Application of E_aMEAD to Predict the Regioselectivity of 46 CYP3A4 Substrates—A Comparison of QM Reaction Energy Based Prediction with E_aMEAD Based Prediction

	reaction energy[a]		E_aMEAD[b]	
order of lowest energy	Nc[c]	ratio (%)[d]	Nc[c]	ratio (%)[d]
with the 1st lowest energy	10	21.7	17	37.0
with the 1st and 2nd lowest energy[e]	20	43.5	28	60.9

[a] QM calculated reaction energy which was used for training the trend vector model by Singh et al. [b] E_aMEAD predicts E_a of all these atoms in all 46 molecules within 12 s.[47] [c] Nc = number of correctly predicted substrates. [d] Percentage of correct predictions. [e] 6 atoms out of 11 atoms which are predicted correctly with the 2nd lowest E_a by E_aMEAD have less than 1.5 kcal/mol energy difference between the 1st and the 2nd lowest E_a.

of the aromatic system decreases as the δ_H and $\bar{\alpha}_{alpha}$ increase. These results demonstrate a good agreement between the E_a^{ar} and the E_a^{MO} (Figure 7, Table 4). Seventy metabolic sites out of 85, including training and test sets, are within the chemical accuracy. Figure 7 depicts that only two metabolic sites of the 85 screened resulted in a residual larger than 1.5 kcal/mol. Both metabolic sites have an amino group at the para position. From this information, it was determined that the model does not properly predict the influence of atoms at the para position to δ_H, which might be explained by including the through space effect.

A Combined Model. As mentioned in the Method section, the combined equation for hydrogen abstraction and Td formation was made

$$\Delta\Delta G_{predicted} = -5.76 + 0.95(E_a^{Habs} \text{ or } E_a^{Habs\_Het}) - 0.99E_a^{ar} \quad (10)$$

This combined model was used to predict the regioselectivity of druglike molecules (Figure 8, Table 5). A plot of the predicted versus measured differences in energies based on regioselectivity for the oxidation listed in Table 6 is presented in Figure 9. A linear relationship exists between the predicted and measured energies.

Regioselectivity Prediction of Druglike Molecules. Olsen et al.[27] predicted regioselectivity for hydrogen abstraction reaction of progesterone and dextromethorphan with QM calculated bond dissociation energies for electronic reactivity and solvent-accessible surface area (SASA) for steric consideration. In contrast, E_aMEAD predicts not only the regioselectivity for hydrogen abstraction but also for the Td formation without any QM calculation and steric consideration. E_aMEAD predicts correctly the major experimentally metabolized atoms in faster (within 2 s)[47] and much simpler manner (Figure 8, Table 5). While E_aMEAD considered that the hydrogens bonded to the same carbon were equivalent, the Olsen et al. treated these hydrogens differently. Consequently, they gave more than one value for E_a^{DFT} or E_a^{AM1} when the reaction carbon had more than one hydrogen or the molecule was in various conformations.

For progesterone, seven suggested methods of Olsen et al. predicted qualitatively that atom number 2, 6, 17, and 21 are more reactive than other atoms. In constrast, E_aMEAD predicted that the carbon number 6 is the major metabolic site as it had the lowest $\Delta\Delta G_{pred.}^{rel.}$ (0 kcal/mol) among all metabolic sites. This is in good agreement with experimental result reported by Yamazaki et al.[42] They showed that carbon 6 is the most metabolized one (quantitatively, more than 4 times than the secondly metabolized carbon) with CYP3A4. E_aMEAD predicted also that the carbon number 21 was the second major metabolic site as it had the second lowest $\Delta\Delta G_{pred.}^{rel.}$ (1.8 kcal/mol). This is in good agreement with experimental result. Yamazaki et al.[42] also showed that carbon 21 is the most metabolized one (quantitatively, more than 4 times than the secondly metabolized carbon again) with CYP2C9 and CYP2C19.

For dextromethorphan, qualitative methods of Olsen et al. predicted that carbon 18 is the most reactive, followed by carbon 19. This is in good agreement with the experimental result of Yu et al.[43] which showed that carbon 18 is the only atom that is metabolized by CYP3A4, and carbon 19 is the most metabolized one by CYP2D6. E_aMEAD also predicted that the carbon number 18 is the major metabolic site as it has the lowest $\Delta\Delta G_{pred.}^{rel.}$ (0 kcal/mol) followed by the carbon number 19 which has the second lowest $\Delta\Delta G_{pred.}^{rel.}$ (2.0 kcal/mol).

The reaction energy of metabolism, which is the difference between reactant and product (i.e., the energy to totally remove a H radical) rather than the reactant and TS, could also be used to predict the regioselectivity of CYP3A4 substrates. In an effort to predict the regioselectivity, Singh et al.[17] used the reaction energy and exposed surface area of hydrogens. To produce this reaction energy rapidly, they built a trend vector model, which is the partial-least-squares (PLS) QSAR method with chemical descriptors. These descriptors capture the local topological environment of the hydrogen such as the element, the number of non-hydrogen neighborhoods, and atoms being in bridgeheads, etc. However, they listed only 40 components of the 3,748 components of the trend vector model in their paper, which was not enough to generate a reproducible result. For this reason, the QM calculated reaction energy was used to predict regioselectivity, rather than the reaction energy from the empirical model of Singh et al. The result using reaction energy was compared with that using E_aMEAD, as shown in Figures 10 and 11 and Table 7. The E_aMEAD based regioselectivity prediction method gave better results than that based on reaction energy and enabled much better physical interpretations. It seemed that the regioselectivity of a reaction could be better predicted with E_a rather than reaction energy, as the E_a can give information on kinetic property.

CONCLUSIONS

As CYP450 enzymes play key roles in drug metabolism, the accurate prediction of CYP450-mediated metabolic reactions is necessary for efficient drug discovery. With the accurate prediction of the regioselectivity of metabolic reactions, which can be predicted with E_a, one can develop rationally designed compounds with the desirable PK properties. The E_aMEAD method described here enables early prediction of these two critical properties in a practical and efficient manner. This method is particularly useful as validated herein to CYP3A4, which is responsible for more than half of the marketed drugs. It does not require any experimental parameters or QM calculations, as often seen with other models. Since our model is based on atomic property, it is expected that this kind of method can be expanded to develop the prediction model of various types of chemical reactions. In conclusion, this new method predicts the E_a of four major metabolic reactions (aliphatic hydroxylation, N-dealkylation, O-dealkylation, and aromatic hydroxylation) reasonably well in a fast and simple manner. This effective method can be applied to practical usage in rapid drug discovery, especially for the high throughput screening (HTS) of the many metabolic characteristics, which are directly or indirectly related toE_a.

ACKNOWLEDGMENT

This work was partly supported by the 21st Century Frontier R&D program (CBM31-B2000-01-00-00) of the Center for Biological Modulators and the Brain Korea 21 (BK21) program funded by the Ministry of Education, Science and Technology, and the IT R&D program of MKE/KEIT [2008-F-029-01, Development of e-Organ system based on Cyber Computing], Korea. We thank Dr. S. J. Lee of Korea Institute of Science and Technology Information for his support of parallel computers that we used for QM calculation.

Supporting Information Available: Statistical values to normalize descriptors, distribution of E_as for hydrogen abstraction, and molecules used in modeling. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Ahlstrom, M. M.; Ridderstrom, M.; Zamora, I.; Luthman, K. CYP2C9 structure-metabolism relationships: Optimizing the metabolic stability of cox-2 inhibitors. *J. Med. Chem.* **2007**, *50*, 4444–4452.
(2) 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.
(3) Klopman, G.; Dimayuga, M.; Talafous, J. Meta. 1. A program for the evaluation of metabolic transformation of chemicals. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1320–1325.

(4) Talafous, J.; Sayre, L. M.; Mieyal, J. J.; Klopman, G. Meta. 2. A dictionary model of mammalian xenobiotic metabolism. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1326–1333.
(5) 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.
(6) Jung, J.; Kim, N. D.; Kim, S. Y.; Choi, I.; Cho, K.; Oh, W. S.; Kim, D. N.; No, K. T. Regioselectivity prediction of CYP1A2-mediated phase I metabolism. *J. Chem. Inf. Model.* **2008**, *48*, 1074–1080.
(7) 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.
(8) Dowers, T. S.; Rock, D. A.; Rock, D. A.; Perkins, B. N. S.; Jones, J. P. An analysis of the regioselectivity of aromatic hydroxylation and N-oxygenation by cytochrome P450 enzymes. *Drug Metab. Dispos.* **2004**, *32*, 328–332.
(9) 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.
(10) Boyer, S.; Arnby, C. H.; Carlsson, L.; Smith, J.; Stein, V.; Glen, R. C. Reaction site mapping of xenobiotic biotransformations. *J. Chem. Inf. Model.* **2007**, *47*, 583–590.
(11) Hazan, C.; Kumar, D.; de Visser, S. P.; Shaik, S. A density functional study of the factors that influence the regioselectivity of toluene hydroxylation by cytochrome P450 enzymes. *Eur. J. Inorg. Chem.* **2007**, *2007*, 2966–2974.
(12) Cruciani, G.; Carosati, E.; Boeck, B. D.; 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.
(13) Oh, W. S.; Kim, D. N.; Jung, J.; Cho, K.; No, K. T. New combined model for the prediction of regioselectivity in cytochrome P450/3A4 mediated metabolism. *J. Chem. Inf. Model.* **2008**, *48*, 591–601.
(14) 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.
(15) de Graaf, C.; Vermeulen, N. P. E.; Feenstra, K. A. Cytochrome P450 *in silico*: An integrative modeling approach. *J. Med. Chem.* **2005**, *48*, 2725–2755.
(16) 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.
(17) 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.
(18) de Graaf, C.; Oostenbrink, C.; Keizers, P. H. J.; van der Wijst, T.; Jongejan, A.; Vermeulen, N. P. E. Catalytic site prediction and virtual screening of cytochrome P450 2D6 substrates by consideration of water and rescoring in automated docking. *J. Med. Chem.* **2006**, *49*, 2417–2430.
(19) Keizers, P. H. J.; de Graaf, C.; de Kanter, F. J. J.; Oostenbrink, C.; Feenstra, K. A.; Commandeur, J. N. M.; Vermeulen, N. P. E. Metabolic regio- and stereoselectivity of cytochrome P450 2D6 towards 3,4-methylenedioxy-n-alkylamphetamines: in silico predictions and experimental validation. *J. Med. Chem.* **2005**, *48*, 6117–6127.
(20) Kemp, C. A.; Flanagan, J. U.; van Eldik, A. J.; Marechal, J.; Wolf, C. R.; Roberts, G. C. K.; Paine, M. J. I.; Sutcliffe, M. J. Validation of model of cytochrome P450 2D6: An in silico tool for predicting metabolism and inhibition. *J. Med. Chem.* **2004**, *47*, 5340–5346.
(21) Higgins, L.; Korzekwa, K. R.; Rao, S.; Shou, M.; Jones, J. P. An assessment of the reaction energetics for cytochrome P450-mediated reactions. *Arch. Biochem. Biophys.* **2001**, *385*, 220–230.
(22) Sykes, M. J.; McKinnon, R. A.; Miners, J. O. Prediction of metabolism by cytochrome P450 2C9: Alignment and docking studies of a validated database of substrates. *J. Med. Chem.* **2008**, *51*, 780–791.
(23) Ahlstrom, M. M.; Ridderstrom, M.; Zamora, I. CYP2C9 structure-metabolism relationships: Substrates, inhibitors, and metabolites. *J. Med. Chem.* **2007**, *50*, 5382–5391.
(24) Vasanthanathan, P.; Hritz, J.; Taboureau, O.; Olsen, L.; Jorgensen, F. S.; Vermeulen, N. P. E.; Oostenbrink, C. Virtual Screening and Prediction of Site of Metabolism for Cytochrome P450 1A2 ligands. *J. Chem. Inf. Model.* **2009**, *49*, 43–52.
(25) Yin, H.; Anders, M. W.; Korzekwa, K. R.; Higgins, L.; Thummel, K. E.; Kharasch, E. D.; Jones, J. P. Designing safer chemicals: Predicting the rates of metabolism of halogenated alkanes. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 11076–11080.
(26) Harris, J. W.; Jones, J. P.; Martin, J. L.; LaRosa, A. C.; Olson, M. J.; Pohl, L. R.; Anders, M. W. Pentahaloethane-based chlorofluorocarbon substitutes and halothane: Correlation of in vivo hepatic protein trifluoroacetylation and urinary trifluoroacetic acid excretion with calculated enthalpies of activation. *Chem. Res. Toxicol.* **1992**, *5*, 720–725.
(27) Olsen, L.; Rydberg, P.; Rod, T. H.; Ryde, U. Prediction of activation energies for hydrogen abstraction by cytochrome P450. *J. Med. Chem.* **2006**, *49*, 6489–6499.
(28) Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. The structure of human microsomal cytochrome P450 3A4 determined by x-ray crystallography to 2.05-Å resolution. *J. Biol. Chem.* **2004**, *279*, 38091–38094.
(29) Li, C.; Zhang, L.; Zhang, C.; Hirao, H.; Wu, W.; Shaik, S. Which oxidant is really responsible for sulfur oxidation by cytochrome P450. *Angew. Chem.* **2007**, *119*, 8316–8318.
(30) Park, J.; 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.
(31) Young, D. C. *Computational Chemistry*; John Wiley & Sons, Inc.: 2001; p 153.
(32) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
(33) 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.
(34) Rendic, S. Summary of information on human CYP enzymes: Human P450 metabolism data. *Drug Metab. Rev.* **2002**, *34*, 83–448.
(35) Williams, P. A.; Cosme, J.; Vinkovic, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H. Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science* **2004**, *305*, 683–686.
(36) Denisov, E. T. Free radical addition: Factors determining the activation energy. *Russ. Chem. Rev.* **2000**, *69*, 153–164.
(37) No, K. T.; Grant, J. A.; Scheraga, H. A. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 1. Application to neutral molecules as models for polypeptides. *J. Phys. Chem.* **1990**, *94*, 4732–4739.
(38) No, K. T.; Grant, J. A.; Jhon, M. S.; Scheraga, H. A. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 2. Application to ionic and aromatic molecules as models for polypeptides. *J. Phys. Chem.* **1990**, *94*, 4740–4746.
(39) No, K. T.; Cho, K. H.; Jhon, M. S.; Scheraga, H. A. An empirical method to calculate average molecular polarizabilities from the dependence of effective atomic polarizabilities on net atomic charge. *J. Am. Chem. Soc.* **1993**, *115*, 2005–2014.
(40) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity-a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219–3288.
(41) Claeyssens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schutz, M.; Thiel, S.; Thiel, W.; Werner, H. High-accuracy computation of reaction barriers in enzymes. *Angew. Chem., Int. Ed.* **2006**, *45*, 6856–6859.
(42) Yamazaki, H.; Shimada, T. Progesterone and testosterone hydroxylation by Cytochromes P450 2C19, 2C9, and 3A4 in human liver microsomes. *Arch. Biochem. Biophys.* **1997**, *346*, 161–169.
(43) Yu, A.; Haining, R. L. Comparative contribution to dextromethorphan metabolism by Cytochrome P450 isoforms in vitro: Can dextromethorphan be used as a dual probe for both CYP2D6 and CYP3A activities. *Drug. Metab. Dispos.* **2001**, *29*, 1514–20.
(44) White, R. E.; Miller, J. P.; Favreau, L. V.; Bhattacharyya, A. Stereochemical dynamics of aliphatic hydroxylation by cytochrome P-450. *J. Am. Chem. Soc.* **1986**, *108*, 6024–6031.
(45) Tassaneeyakul, W.; Birkett, D. J.; Edwards, J. W.; Veronese, M. E.; Tassaneeyakul, W.; Tukey, R. H.; Miners, J. O. Human cytochrome P450 isoform specificity in the regioselective metabolism of toluene and o-, m- and p-xylene. *J. Pharmacol. Exp. Ther.* **1996**, *276*, 101–108.
(46) Hanzlik, R. P.; Hogberg, K.; Judson, C. M. Microsomal hydroxylation of specifically deuterated monosubstituted benzenes. Evidence for direct aromatic hydroxylation. *Biochemistry* **1984**, *23*, 3048–3055.
(47) A PC, Intel Pentium D, CPU 2.8 GHz, 0.99GB RAM, Microsoft Windows XP.

CI900011G