Prediction of Activation Energies for Hydrogen Abstraction by Cytochrome P450

Lars Olsen,*,† Patrik Rydberg,‡ Thomas H. Rod,§ and Ulf Ryde‡

Department of Medicinal Chemistry, The Danish University of Pharmaceutical Sciences, 2 Universitetsparken, DK-2100 Copenhagen, Denmark, Department of Theoretical Chemistry, University of Lund, Sweden, and Nano-Science Center, Atomistix A/S, Niels Bohr Institute, Juliane Maries Vej 30, DK-2100 Copenhagen, Denmark

Received May 11, 2006

We have estimated the activation energy for hydrogen abstraction by compound I in cytochrome P450 for a diverse set of 24 small organic substrates using state-of-the-art density functional theory (B3LYP). We then show that these results can be reproduced by computationally less demanding methods, for example, by using small organic mimics of compound I with both B3LYP and the semiempirical AM1 method (mean absolute error of 3−4 kJ/mol) or by calculating the bond dissociation energy, without relaxation of the radical (B3LYP) or estimated from three-point fit to a Morse potential (AM1; errors of 4 and 5 kJ/mol, respectively). We can assign activation energies of 74, 61, 53, 47, and 30 kJ/mol to primary carbons, secondary/tertiary carbons, carbons with adjacent sp^2 or aromatic groups, ethers/thioethers, and amines, respectively, which gives a very simple and predictive model. Finally, some of the less demanding methods are applied to study the CYP3A4 metabolism of progesterone and dextromethorphan.

Introduction

The cytochromes P450 form a ubiquitous protein family with functions including synthesis and degradation of many physiologically important compounds, as well as degradation of xenobiotic compounds, for example, drugs.[1] Much effort has been put into the study of these enzymes, because they influence the transformation of prodrugs into their active form and the bioavailability and degradation of many drugs.

The cytochrome P450 enzymes catalyze several different types of reactions, of which the hydroxylation of CH groups is the most studied.[2,3] Recent investigations suggest that the hydroxylation follows a two-step mechanism, in which the first step involves a hydrogen abstraction from the substrate by an Fe^{V}═O species (formally), denoted compound I, and in the second step, the radical rebounds to the Fe^{IV}−OH intermediate. This gives an Fe^{III}−alcohol complex from which the hydroxylated product may dissociate. Density functional theory (DFT) calculations indicate that the highest energy barrier along this pathway is observed for the hydrogen-abstraction step,[4−7] and this is also supported by experimental results.[8]

Considering the importance of the cytochromes P450 in the metabolism of drugs, it would be highly desirable to have a method that could predict if and in what way a drug candidate will be metabolized by these enzymes. Most previous studies have been focused on how a compound is metabolized, based on quantum chemical studies on isolated substrates, pharmacophore models, docking, molecular dynamics simulations, chemical rules, or quantitative structure−activity relationships (QSAR) from physicochemical, topological, or 3D structures.[9−13] The consensus is that not one single computational approach can reliably predict metabolism by the P450s, but for most enzymes, a combination of both the intrinsic reactivity of various parts of the substrate (electronic factors) and the accessibility of the groups to the reactive Fe^{V}═O group in the enzyme (steric effects) need to be taken into account.

The intrinsic reactivity of the various groups have normally been estimated by quantum mechanical (QM) methods at the Hartree−Fock, semiempirical, or DFT levels. It has typically been estimated from the stability of the radicals formed after hydrogen abstraction,[14] from the ionization potential of these radicals (estimated from the energy of the highest occupied molecular orbital, HOMO),[11,15,16] or from the density of the HOMO on the various atoms.[17] More sophisticated methods involve the direct estimation of the activation energy for the hydrogen abstraction by use of simplified models of the Fe^{V}═O state of cytochrome P450, for example, an isolated O atom,[18,19] a methoxy radical,[16,20,21] or a *p*-nitrosophenoxy radical.[15,22] Recently, it has even become possible to calculate activation energies with DFT and full models of the active porphyrin species, giving near-quantitative results.[3,23,24] However, such calculations are still are quite time-consuming, especially for molecules of the size of a typical drug (weeks of CPU time). On the other hand, they can be used to develop and calibrate more approximate methods. A first such attempt has recently been published,[6] but it used only 11 substrates, without any external validation, and tested relatively few methods. It should also be mentioned that Park and Harris[22] have compared relative reactivity of two sites in methoxyflurane predicted from a full DFT optimization of the transition state and from simpler models, showing that both the radical stability and the *p*-nitrosophenoxy radical model gave excellent results at the DFT level but not at the semiempirical AM1 level.[22]

In this paper, we extend this work. We have used DFT calculations on a full model of compound I and determined the activation energy for the hydrogen abstraction from 24 substrates (cf. Figure 1), involving primary, secondary, and tertiary aliphatic carbon atoms and nitrogen, oxygen, and sulfur, as well as sp^2 or aromatic carbon atoms next to the reactive atom. This covers the most common types of atoms in druglike molecules. The 24 substrates were divided into training (14 compounds) and test sets (10 compounds), which were used to validate the ability of different computationally less demanding methods to predict the activation energies in a systematic manner. These models include DFT calculations with different basis sets and calculations with small radical models of compound I at both

* To whom correspondence should be addressed: phone (+45) 35 30 63 05; fax (+45) 35 30 60 40; e-mail lo@dfuni.dk.
† The Danish University of Pharmaceutical Sciences.
‡ University of Lund.
§ Atomistix A/S.

10.1021/jm060551l CCC: $33.50
Published on Web 10/04/2006

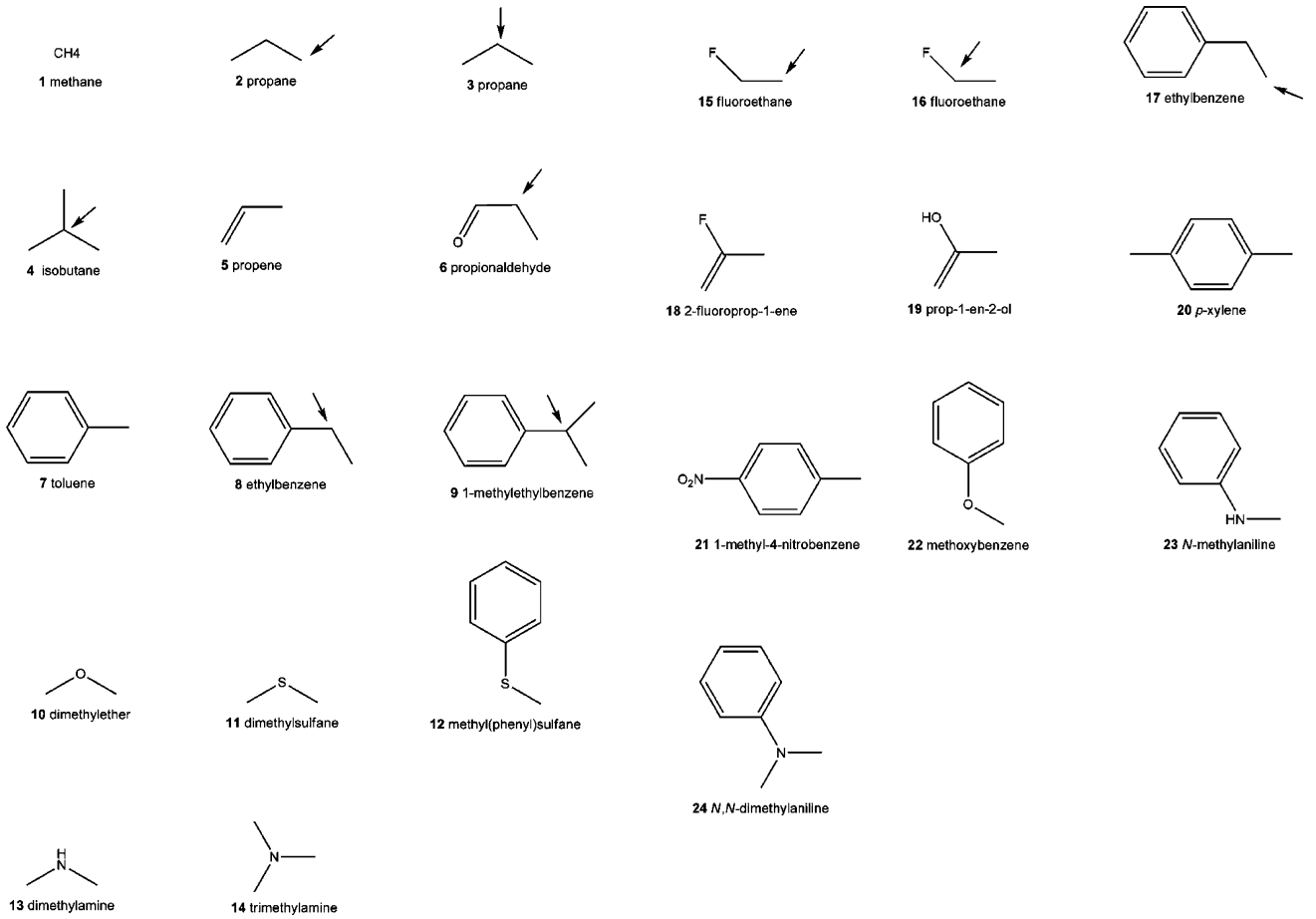

Figure 1. Compounds used to study the activation energies for the aliphatic hydrogen abstraction reaction. Arrows indicate the carbon atom from which the hydrogen was abstracted for compounds with more than one possible aliphatic site of abstraction. Compounds **1**−**14** were used as a training set and compounds **15**−**24** as a test set for models of the activation energies.

the DFT and semiempirical levels, as well as a set of 20 different descriptors of the substrate or the radical (e.g., bond dissociation and orbital energies). The results show that it is possible to predict the DFT activation energies with an accuracy of 2−5 kJ/mol at a modest computational cost.

Methods

Model Complexes. We have modeled the compound I species of cytochromes P450 as iron (formally Fe^{V}) porphine (i.e., a porphyrin without side chains) with CH_3S^- and O^{2-} (formally) as axial ligands (cf. Figure 2). Shaik and co-workers[23,25] have argued that SH^- is a better model of Cys than CH_3S^-. However, a recent QM/MM study,[26] where a larger reference model for Cys was used [$HCONHCH(CH_2S)CONH_2$], showed that SCH_3^- gives better geometries than does SH^-. SH^- gives better spin densities in a vacuum, but the opposite is true in the protein. Therefore, we have used SCH_3^- as a model of the Cys ligand. We also tested modeling of compound I by a methoxy, phenoxy, or *p*-nitrosophenoxy radical (Figure 2), as has been suggested before.[15,16,20−22,27]

We have studied hydrogen abstraction from 24 different substrates (Figure 1). The studied substrates involve primary, secondary, and tertiary aliphatic carbon atoms, which have nitrogen, oxygen, and sulfur, as well as sp^3, sp^2, or aromatic carbon atoms next to them. For the test of various simplified methods to predict the activation energy, the substrates were divided into a training set **(1**−**14)** and test set **(15**−**24),** which both still contain all the various functional environments.

Three different states along the hydrogen-abstraction reaction were studied (Figure 2), viz., the isolated compound I and substrate, the transition state (TS) for the hydrogen abstraction, and the intermediate after the hydrogen abstraction, where the substrate radical weakly interacts with the Fe^{IV}−OH complex (protonated compound II), that is, the intermediate before the radical-rebound step. All energies are given relative to the sum of the energies of the isolated compound I model and the isolated substrate.

All three states were studied in the quartet (intermediate-spin) state. Previous studies have shown that the low- and intermediate-spin states are very close in energy for both compound I and the hydrogen-abstraction TS.[6,28] Therefore, it is enough to study one of the spin states, and we have selected the quartet for computational reasons. The next step in the reaction, the radical-rebound step, is barrierless in the doublet state but it has a small barrier for the quartet state, 4−12 kJ/mol.[4,5,7,23,29] This means that the TS for hydrogen abstraction normally has the highest energy on the potential energy surface.[4,5,7] Therefore, it is enough to study only the TS of the hydrogen-abstraction step.

The quantum chemical calculations were performed with the density functional method B3LYP[30−32] (unrestricted formalism for open-shell systems) or with the semiempirical AM1 method.[33] In the B3LYP calculations, we have used for iron the double-ζ basis set of Schäfer et al.,[34] enhanced with a p function with the exponent 0.134915. For the other atoms, the 6-31G(d) basis set[35] was used. The final energies were determined at the B3LYP/6-311++G(2d,2p) level. These energies also include the zero-point vibrational energy, calculated at the B3LYP/6-31G(d) level. The frequency calculations also verified that the structures represented true minima or transition states.

Molecular Descriptors. We have extracted 20 descriptors for the substrates and radicals from the DFT and AM1 calculations. These are the Mulliken charges on the carbon (q_C) and hydrogen (q_H) atoms involved in the reaction, the spin on the carbon atom in

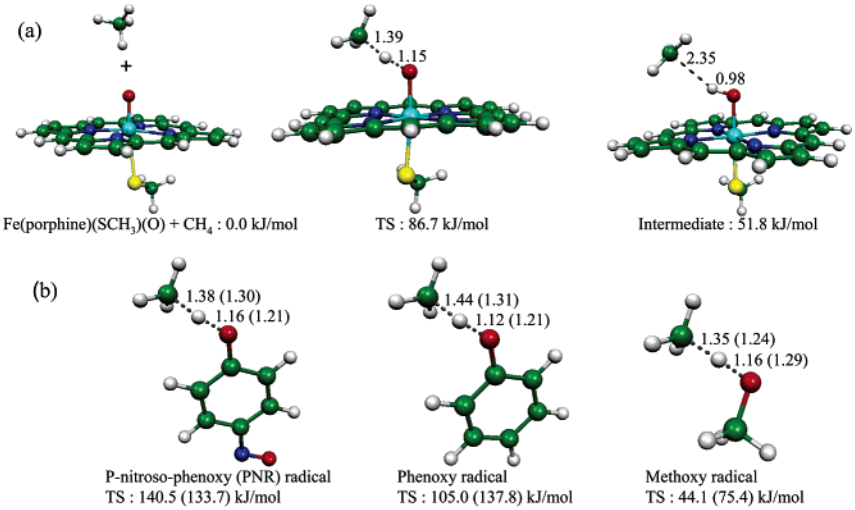

Figure 2. Energies and O−H and C−H distances (angstroms) for the hydrogen atom abstraction of methane by (a) a Fe(porphine)(SCH_3)O model and (b) smaller organic radicals at the B3LYP/6-311++G(2d,2p)//B3LYP/6-31G(d) level with the zero-point vibrational energy at the B3LYP/6-31G(d) level included. Numbers in parentheses are the corresponding AM1 energies.

the radical (S_C), the energies of the HOMO and LUMO (lowest unoccupied molecular orbital), and the energy difference between these two orbitals, as well as the coefficients for the hydrogen 1s and carbon 2p atomic orbitals in the HOMO and LUMO.

Moreover, we studied the bond dissociation energies (BDEs), calculated in three different ways: First, we have used the standard definition of the BDE, that is, the difference in energy between the substrate and the sum of the radical and hydrogen atom, using optimized geometries for all three species:

$$\mathrm{BDE} = E(\mathrm{substrate}) - E(\mathrm{radical}) - E(\mathrm{H}) \qquad (1)$$

Of course, this is also a direct measure of the stability of the radical (the energy of a hydrogen atom is a constant). We can save some computer time by omitting the optimization of the radical [i.e., by calculating the *E*(radical) for the geometry of the substrate but with the hydrogen atom removed].[6] This will be denoted $\mathrm{BDE_{fr}}$ (frozen radical) in the following. The structures were optimized by use of B3LYP/6-31G(d), and the final energies were determined with B3LYP/6-311++G(2d,2p) with the zero-point vibrational energy included (or with AM1). Finally, we also tested to obtain the BDE by a fit to a Morse potential, as has been proposed by Lewin and Cramer:[36] The energy of the substrate was calculated in three points (the equilibrium C−H distance and ±0.3 Å from the equilibrium C−H distance), with the rest of the geometry fully optimized, and they were then used to fit the dissociation energy by a Morse potential:

$$E(r) = \mathrm{BDE_{Morse}}[1 - \exp(\alpha_{\mathrm{Morse}}(r - r_{\mathrm{eq}}))]^2 \qquad (2)$$

The calculations were performed with the B3LYP/MIDI![37] and AM1 methods.

Programs. The nonlinear fits to the Morse potential were done with the GraFit program.[38] Partial least squares of latent variables (PLS) were generated in the Simca program,[39] and all B3LYP and AM1 calculations were performed with the Gaussian 03 software.[40] The solvent-accessible surface area was determined with the SAVOL program.[41]

Results and Discussion

Calculated Activation Energies. We have studied 24 different hydrogen-abstraction reactions (Figure 1) using the Fe(porphine)(SCH_3)O model of compound I (Figure 2a). In Table 1, the B3LYP/6-311++G(2d,2p)//B3LYP/6-31G(d) energies, with the zero-point vibrational energy included (relative to the isolated substrate and compound I), and O−H and C−H bond lengths in the TS and the intermediate are shown. The amines

Table 1. Relative Energies and C−H and O−H Bond Lengths in the Transition State for Hydrogen Abstraction and Following Intermediate[a]

		transition state			intermediate		
	substrate[b]	energy (kJ/mol)	C−H (Å)	O−H (Å)	energy (kJ/mol)	C−H (Å)	O−H (Å)
1	methane	86.7	1.39	1.15	51.8	2.35	0.98
2	propane (1)	73.9	1.35	1.18	34.2	2.25	0.98
3	propane (2)	62.0	1.32	1.21	15.0	2.23	0.98
4	isobutane	59.7	1.31	1.24	4.5	2.33	0.98
5	propene	53.9	1.31	1.25	−27.2	2.44	0.98
6	propionaldehyde	47.9	1.35	1.21	−22.4	2.59	0.97
7	toluene	54.6	1.31	1.24	−15.0	2.43	0.98
8	ethylbenzene (2)	50.6	1.29	1.27	−29.5[c]	2.63	0.97
9	1-methylethylbenzene	55.8	1.30	1.26	−34.0[c]	2.70	0.97
10	dimethylether	50.9	1.31	1.24	3.8	2.24	0.98
11	dimethylsulfane	45.9	1.33	1.22	−4.0	2.27	0.98
12	methyl(phenyl)sulfane	45.4	1.34	1.21	−2.6	2.32	0.98
13	dimethylamine	31.9	1.27	1.30	−14.7	2.11	0.99
14	trimethylamine	27.9	1.27	1.31	−18.4	2.12	0.99
15	fluoroethane (2)	77.2	1.38	1.15	35.2	2.70	0.97
16	fluoroethane (1)	61.6	1.34	1.19	20.3	2.24	0.98
17	ethylbenzene(1)	72.2	1.35	1.18	34.8	2.29	0.98
18	2-fluoroprop-1-ene	55.2	1.33	1.23	−19.4	2.48	0.97
19	prop-1-en-2-ol	49.1	1.32	1.25	−52.5[d]	3.43	0.97
20	*p*-xylene	53.0	1.31	1.25	−17.8	2.40	0.97
21	1-methyl-4-nitrosobenzene	49.5	1.33	1.21	−20.8[c]	2.52	0.97
22	methoxybenzene	54.5	1.34	1.21	9.6	2.34	0.98
23	*N*-methylaniline[d]	31.9	1.30	1.26	−12.4[e]		
24	*N*,*N*-dimethylaniline	28.9	1.29	1.28	−21.2	2.20	0.98

[a] The energies were determined at B3LYP/6-311++G(2d,2p)//B3LYP/6-31G(d) level with the zero-point vibrational energy at the B3LYP/6-31G(d) level included. All energies are relative to the sum of the energies of the isolated substrate and the compound I model. [b] Numbers in parentheses for the substrates indicate the position from which the hydrogen is abstracted. [c] RMS forces were converged to at least 0.000030 au (the maximum and RMS distances were not converged). [d] In the optimized structure, a hydrogen bond between the OH groups of the substrate and compound II was formed. [e] The intermediate was not stable, and the result of the optimization is the corresponding imine. The energy is calculated as the sum of the isolated radical and protonated compound II.

(**13**, **14**, **23**, and **24**) have the lowest activation energies for the hydrogen abstraction, 28−32 kJ/mol. The ethers and thioethers (**10**−**12** and **22**) have activation energies of 45−55 kJ/mol, and compounds with sp^2-hybridized (**5**, **6**, **18**, and **19**) or aromatic (**7**−**9**, **20**, and **21**) groups next to the reactive carbon have activation energies in the 48−56 kJ/mol range. Substrates with only sp^3-hybridized atoms have the highest barriers. Abstractions from secondary (**3**) and tertiary (**4**) carbon atoms or from carbons with F atoms bound (**16**) have activation energies of 60−62

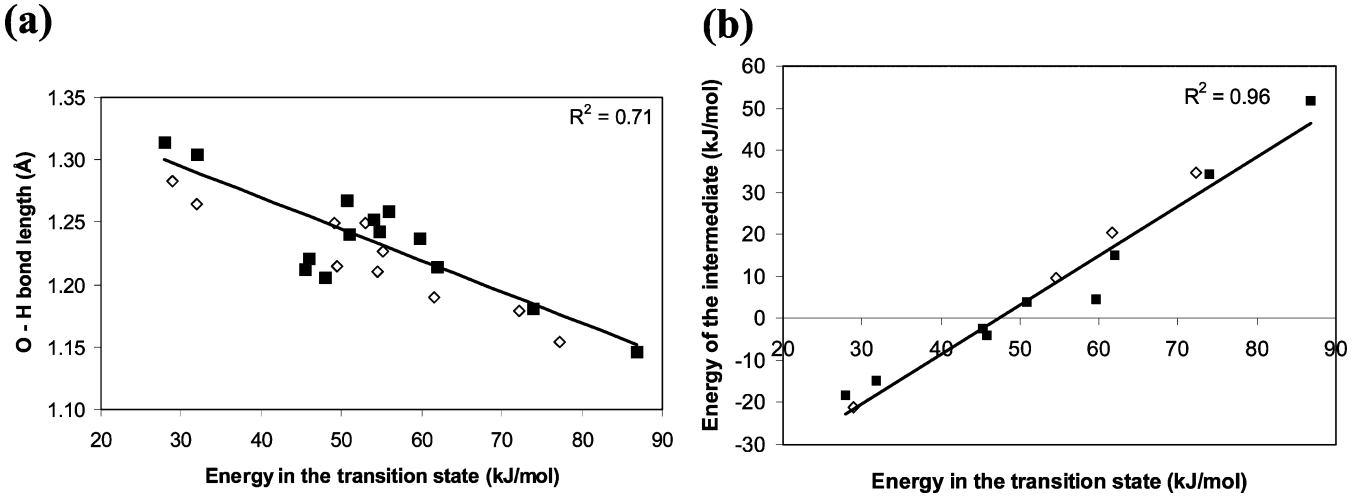

Figure 3. Activation energies of the Fe(porphine)(SCH_3)O model related to (a) O−H bond length in the TS and (b) energy of the intermediate (compounds with sp^2-hybridized atoms, aromatic atoms, or a F atom next to the reactive carbon atom omitted, that is, with nine and four compounds in the training and test sets, respectively). Solid and open symbols represent training and test sets, respectively. R^2 is given for the training set.

kJ/mol, whereas abstractions from primary carbon atoms (**1**, **2**, **15**, and **17**) have even larger activation energies (>72 kJ/mol). This follows the qualitative stability of the radicals, as has been noted before.[15,16]

The C−H and O−H bond lengths in the TS are clearly correlated with the activation energy: the longer the C−H and the shorter the O−H (cf. Figure 3a) bond lengths, the larger the activation energy. One extreme is methane, which has the highest activation energy and a very late TS with an O−H bond length of 1.15 Å, indicating that the H atom is almost completely transferred to the O atom already in the TS. The other extremes are the amines (**13**, **14**, **23**, and **24**), which have O−H bond lengths of ∼1.3 Å and the lowest activation energies. A similar relationship between the bond lengths and activation energies for hydrogen-abstraction reactions has been previously noted.[6,15]

Seven of the substrates (**1**−**3**, **5**, **7**, **17**, and **24**) have previously been studied by Shaik and co-workers[6,28] using similar methods (structures at the B3LYP/LACVP level and energies at the B3LYP/LACVP3+* level, with a SH^- model of the Cys ligand). The two sets of activation energies differ by less than 10 kJ/mol, which shows that the model of the Cys ligand and the other details of the calculations are not crucial in terms of energies.

Simplified DFT Calculations. The energies in Table 1 constitute state-of-the-art DFT estimates of the activation energy for hydrogen abstraction from various substrates by any cytochrome P450. They should provide a consistent set of best computational estimates of the intrinsic reactivity for various substrates with compound I. Similar methods have been used to study many different enzymes, with absolute errors in the activation energy of less than ∼20 kJ/mol and even smaller relative errors.[42,43] In particular, such methods have been successfully applied also to cytochromes P450,[3] giving kinetic isotope effects[28] and regioselectivity[22] in good accordance with experimental results. The aim of this paper is to see if we can predict these energies with computationally less demanding methods (the DFT calculations took several CPU-weeks for each substrate; cf. Table 2.)

A first step in this direction is to see if the zero-point energies are really needed, because the frequency calculations take almost as much time as the geometry optimizations. The results in Table 2 show that this is not really the case: The activation energies with zero-point corrections correlate almost perfectly with those without the corrections [$R^2 = 1.00$; mean absolute error (MAE) between energies predicted from the correlation line and the activation energies with zero-point corrections, 0.7 kJ/mol]. Likewise, the time-consuming calculation with the large 6-311++G(2d,2p) basis set can also be avoided (it also takes about the same time as the geometry optimization), but the effect is appreciably larger: $R^2 = 0.78$ and MAE = 6 kJ/mol for the test set if the B3LYP/6-31G(d) energies are used without both the zero-point energies and the calculation with the big basis set. De Visser et al.[6] reached a similar conclusion.

Table 2. Correlation of Activation Energy Calculated by the Full Method with Calculations with Smaller Basis Sets or without Zero-Point Energy Corrections[a]

	Fe(porphine) (CH_3S)O	methoxy	phenoxy	BDE_{fr}
	B3LYP/6-31G(d)[b]			
R^2 (training set)	0.90	0.91	0.91	0.84
MAE (training set) (kJ/mol)	3.81	2.84	3.59	4.73
R^2 (test set)	0.78	0.89	0.96	0.90
MAE (test set) (kJ/mol)	6.31	4.09	2.88	3.60
CPU time	6 days 10 h	1 h 30 min	5 h 36 min	20 min
	B3LYP/6-311++G(2d,2p)[c]			
R^2 (training set)	1.00	0.89	0.89	0.86
MAE (training set) (kJ/mol)	0.72	3.50	3.98	4.32
R^2 (test set)	1.00	0.94	0.96	0.94
MAE (test set) (kJ/mol)	0.74	3.70	3.36	3.21
CPU time	12 days 15 h	2 h 12 min	7 h 19 min	48 min
	B3LYP/6-311++G(2d,2p) and ZPE corrections[d]			
CPU time	17 days 17 h	3 h 40 min	13 h 5 min	1 h 50 min

[a] CPU estimates for 1-methyl-4-nitrosobenzene (compound **21**) are also included for each method. [b] Geometry optimization at B3LYP/6-31G(d) level. An initial frequency calculation at HF/3-21G(d) was carried out before the TS optimizations. [c] TS optimization as described above and single-point energy calculation at B3LYP/6-311++G(2d,2p) level. [d] TS optimization as described above, followed by a frequency calculation with B3LYP/6-31G(d) level and single-point energy calculation at B3LYP/6-311++G(2d,2p) level.

Another possibility is to study the intermediate after the hydrogen abstraction (but before the radical rebound; it is easier to optimize than the TS). The results in Figure 3b show that for most of the substrates, the energies of the intermediates correlate well with that of the TS ($R^2 = 0.96$). However, for substrates with aromatic or sp^2-hybridized groups next to the reactive carbon (**5**−**9** and **18**−**21**), this correlation disappears TS ($R^2 = 0.50$). For these substrates, the interaction between the radical and compound II is weak (C−H distance larger than 2.40 Å), and the radical is essentially dissociated from the

Table 3. Correlation between Activation Energies and O−H and C−H Bond Lengths Obtained with the Fe(porphine)(SCH_3)(O) Model and with Small Organic Radical Models of Compound I[a]

		training set		test set	
		R^2	MAE (kJ/mol)	R^2	MAE (kJ/mol)
		B3LYP			
p-nitrosophenoxy	energy	0.89	3.50	0.94	3.68
	O−H	0.90	0.01	0.81	0.01
	C−H	0.82	0.01	0.63	0.01
phenoxy	energy	0.89	3.98	0.96	3.36
	O−H	0.92	0.01	0.97	0.01
	C−H	0.85	0.01	0.95	0.01
methoxy	energy	0.89	3.50	0.94	3.70
	O−H	0.93	0.01	0.98	0.01
	C−H	0.77	0.01	0.92	0.01
		AM1			
p-nitrosophenoxy[b]	energy	0.86	4.24	0.31	9.08
	O−H	0.79	0.02	0.77	0.02
	C−H	0.39	0.02	0.48	0.01
phenoxy	energy	0.88	3.96	0.89	3.95
	O−H	0.80	0.02	0.84	0.01
	C−H	0.37	0.02	0.74	0.01
methoxy	energy	0.87	4.55	0.84	4.43
	O−H	0.86	0.01	0.79	0.01
	C−H	0.07	0.02	0.38	0.02

[a] Regression lines were obtained by fitting the activation energies by using smaller radical mimics at B3LYP and AM1 level with those obtained for the Fe(porphine)(CH_3S)O model at B3LYP level. [b] If compound **21** is omitted from the test set, $R^2 = 0.91$ and MAE = 3.78.

compound II model (whereas it forms a stable interaction with the Fe^{IV}−OH group, with C−H distances of 2.11−2.35 Å for the other substrates). This is probably because the radical is delocalized, which stabilizes it. Thus, there is little gain to study the intermediates.

Smaller Models of Compound I. In previous work, simplified organic radicals have been used to model compound I, for example, methoxy[16,20,21] or *p*-nitrosophenoxy radicals.[15,22] Therefore, we also tested how well activation energies calculated with *p*-nitrosophenoxy, phenoxy, and methoxy radicals (Figure 2b) correlate with the activation energies obtained with the Fe(porphine)(SCH_3)O model. The activation energies and the O−H and C−H bond lengths in the TS obtained with these models at both the B3LYP and AM1 levels are collected in Tables S1 and S2 in the Supporting Information. The correlations with the full DFT activation energies are shown in Table 3 for both the training and test sets. It can be seen that these models work well at both B3LYP and AM1 levels: The MAE of the predicted activation energies is only 3−4 kJ/mol for B3LYP and ∼4 kJ/mol for AM1 (with one outlier excluded for the *p*-nitrosophenoxy radical). However, most of the models give a large systematic error in the absolute activation energies (see Figure 4 and Tables S1 and S2). The phenoxy radical gives both the smallest MAEs and the best absolute energies.

As for the full porphyrin model, significant computer time can be saved by omitting the frequency and big basis set calculations (Table 2), and even though the phenoxy radical gives slightly better results, much computer time can be saved by using the methoxy radical. Even more computer time can be saved by using AM1 for these calculations (Table 4).

Molecular Descriptors. All calculations up to now have involved the optimization of TS geometries. This is quite tedious and not guaranteed to always converge. It would be preferable if a correlation could be found between the activation energies and the properties of the isolated substrates or radicals alone (geometry optimizations of ground states are much easier and more stable). To this aim, we have studied the correlation between the activation energies and various molecular descriptors of the substrates and their corresponding radicals. The descriptors include the Mulliken charges on the carbon (q_C) and hydrogen (q_H) atoms involved in the reaction, the spin on the carbon atom in the radical (S_C), the energies of the HOMO (which is an approximation of the ionization potential, according to Koopmans' theorem) and LUMO and their energy difference, and the coefficients for the hydrogen 1s and carbon 2p atomic orbitals in the HOMO and LUMO, as well as the BDE of the reactive C−H bond, calculated by three different methods. The results (R^2 and MAE for the training and test sets) are collected in Tables 5 (B3LYP) and 6 (AM1).

Most of the descriptors correlate quite well with the activation energies for the compounds that do not have a sp^2-hybridized or aromatic atom next to the reactive carbon (substrates **5**−**9** and **18**−**21**), as is illustrated in Figure 5a. However, including the latter substrates strongly weakens the correlations. The best correlations for all compounds are obtained for the descriptors related to the BDE. Interestingly, with B3LYP the best results are obtained for BDE_{fr} (MAE = 4 kJ/mol; Table 5 and Figure 5b), which is the cheapest BDE descriptor, whereas the other two gave twice as large errors. De Visser et al.[6] have obtained similar results for a smaller set of substrates (e.g., not involving N, O, and S), and the models were not externally validated. As before, the B3LYP results are similar if the frequency and big basis set calculations are omitted (cf. Table 2). With AM1, all three BDE descriptors gave similar MAEs for the test set (6−8 kJ/mol), but α_{Morse} gave slightly better results (MAE = 5 kJ/mol). α_{Morse} is associated with the width of the Morse potential, that is related to the force constant and BDE_{Morse}, which may explain why there is a correlation with the height of the energy barrier.

All the other descriptors gave much worse correlations and predictions (MAE > 8 kJ/mol). Descriptors based on the radicals gave slightly better results than those based on the substrates. In general, the HOMO energy of the radical (Figure 5c), the HOMO−LUMO energy gap, the carbon 2p coefficient in HOMO, and the spin density on the radical carbon atom (AM1) gave the best results. The ionization potential of the radical has previously been related to calculated activation energies (together with the radical stability, i.e., the BDE).[15,16,20,22,27] Our results in Table 4 and Figure 5c confirm that this is a reasonable approximation.

Finally, we applied the partial least squares of latent variables (PLS) method to make models of the activation energies using all the molecular descriptors at either B3LYP or AM1 level. The score plot in Figure 6a shows that the first component separates the compounds according to the activation energy, for example, compounds **1** and **13**/**14** are found in different parts of the plot. The second component separates the compounds that contain sp^2-hybridized or aromatic atoms (compounds **5**−**9**) next to the reactive carbon. The loadings plot in Figure 6b shows that the BDE descriptors, the 2p coefficient in HOMO, the HOMO−LUMO energy gap, and the spin on the carbon of the radical [2p (HOMO), E(HOMO) − E(LUMO), and S_C, respectively] are found relatively close to the activation energy, and are therefore positively correlated with it. This means that the larger the bond dissociation energy, the larger the activation energy. On the other hand, the HOMO and LUMO energies and α_{Morse} are negatively correlated with activation energy, because these descriptors are found in the opposite part of the plot. The fact that the LUMO energies are located in the lower region of the loadings plot is also one of the reasons why the second component separates the compounds with and without

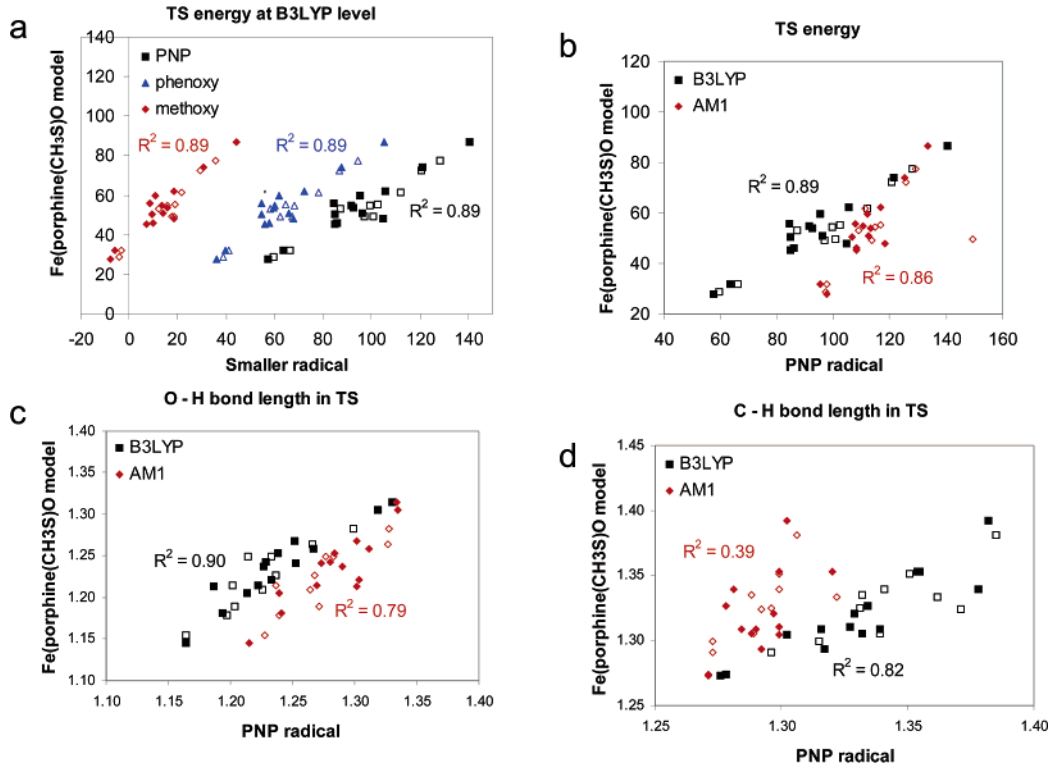

Figure 4. (a) Comparison of activation energies (kilojoules/mole) for *p*-nitrosophenoxy (PNP), phenoxy, or methoxy radicals to abstract the H atom with Fe(porphine)(SCH_3)O as model for compound I. Also shown are comparisons of (b) energies and (c) O−H and (d) C−H bond lengths (angstroms) in the TS for a *p*-nitrosophenoxy radical instead of the Fe(porphine)(SCH_3)O model at B3LYP or AM1 level. Solid and open symbols represent the training and test sets, respectively. R^2 is given for the training set.

Table 4. CPU Time Estimates at AM1 Level[a]

	CPU time (min)
methoxy	8
phenoxy	7
α_{Morse}	5

[a] The CPU estimates are with 1-methyl-4-nitrosobenzene (compound **21**) as substrate. The CPU estimates for the TS calculations include an initial frequency calculation.

aromatic rings in the score plot (LUMO is more stable for the aromatic compounds).

At the B3LYP level, a PLS model that contains all descriptors with VIP > 0.8 (variable influence on projection) is of the same quality as BDE_{fr} alone (the MAEs of the test set are both ∼4 kJ/mol); see Table 7. Likewise, at the AM1 level, the PLS model is not significantly better than α_{Morse} alone (MAE of the test set are 4 and 5 kJ/mol, respectively). Therefore, we see no reason to employ the more complicated and computationally demanding PLS models.

A Simple Qualitative Model. As mentioned above, we observe a clear grouping of the activation energies according to the chemical function and environment of the reactive carbon atom. On the basis of this observation and the average energies in the training set, we can set up the following very simple predictive model for the activation energies (energies in parentheses are the averages obtained for both the training and test sets together): (1) Primary carbon atoms in a sp^3 environment; no F atom bound to reactive C (**2**; **15**, **17**): 73.8 (74.4) kJ/mol. (2) Secondary or tertiary carbon in a sp^3 environment or F directly bound to reactive C (**3**, **4**; **16**): 60.9 (61.1) kJ/mol. (3) Aromatic or sp^2-hybridized atoms next to the reactive C (**5**−**9**; **18**−**21**): 52.6 (52.2) kJ/mol. (4) O or S atoms next to the reactive C (**10**−**12**; **22**): 47.3 (49.1) kJ/mol. (5) N atoms next to the reactive C (**13**, **14**; **23**, **24**): 29.8 (30.1) kJ/mol

Table 5. Correlations of Various Descriptors at the B3LYP Level with Activation Energies Obtained with the Full Porphine Model at the B3LYP Level[a]

	training set		test set	
	R^2	MAE (kJ/mol)	R^2	MAE (kJ/mol)
Descriptors Related to BDE				
BDE_{Morse}	0.65	6.53	0.76	7.30
α_{Morse}	0.73	6.21	0.57	8.60
BDE	0.37	9.85	0.58	7.69
BDE_{fr}	0.86	4.32	0.94	3.21
Substrate Descriptors				
q_C	0.08	11.06	0.02	12.20
q_H	0.01	10.87	0.05	10.77
E (HOMO)	0.80	5.69	0.55	8.97
E (LUMO)	0.10	9.95	0.17	11.13
2p (HOMO)	0.50	8.88	0.33	11.83
1s (HOMO)	0.04	10.09	0.01	10.89
2p (LUMO)	0.12	10.93	0.14	12.53
1s (LUMO)	0.08	9.80	0.16	11.66
E(HOMO) − *E*(LUMO)	0.49	8.32	0.44	10.24
Radical Descriptors				
q_C	0.00	10.46	0.01	10.86
S_C	0.43	9.10	0.42	9.70
E (HOMO)[b]	0.41	8.65	0.64	8.25
2p (HOMO)	0.40	8.81	0.53	9.83
E (LUMO)	0.04	10.67	0.02	11.84
2p (LUMO)	0.25	8.69	0.16	12.55
E(HOMO) − *E*(LUMO)	0.28	9.80	0.55	8.40

[a] The following descriptors were used: BDE_{Morse} and α_{Morse}, bond dissociation energy and α parameter from eq 2; BDE, defined in eq 1, where the geometry of the radical is relaxed; BDE_{fr}, same as BDE except that the geometry of the radical is not relaxed; q_C and q_H, Mulliken charges on the reactive C and H atoms; 2p and 1s, orbital coefficients for either the HOMO or LUMO on the reactive C or H atom; *E*, energies of the HOMO or LUMO orbitals; S_C, spin on the reactive C atom. [b] If propionaldehyde (**6**) is omitted from the training set, $R^2 = 0.82$ and MAE = 5.02 kJ/mol.

Interestingly, this very simple model gave the best predictions obtained in this paper: the MAE for the test set was 2.5 kJ/

Table 6. Correlations of Various Descriptors at AM1 Level with Activation Energies Obtained Using the Full Porphine Model at the B3LYP Level[a]

	training set		test set	
AM1	R^2	MAE (kJ/mol)	R^2	MAE (kJ/mol)
Descriptors Related to BDE				
BDE_{Morse}	0.72	6.30	0.76	6.16
α_{Morse}	0.73	5.17	0.82	5.19
BDE	0.43	9.16	0.71	7.56
BDE_{fr}	0.68	7.34	0.68	6.25
Substrate Descriptors				
q_C	0.02	10.82	0.14	10.96
q_H	0.05	10.24	0.24	10.80
E (HOMO)	0.63	7.68	0.52	8.73
E (LUMO)	0.14	9.38	0.23	11.08
2p (HOMO)	0.45	8.48	0.26	11.28
1s (HOMO)	0.09	9.52	0.03	10.69
2p (LUMO)	0.08	10.40	0.26	10.29
1s (LUMO)	0.37	8.25	0.26	9.76
E(HOMO) − *E*(LUMO)	0.37	8.81	0.42	10.01
Radical Descriptors				
q_C	0.05	9.36	0.44	10.18
S_C	0.54	7.75	0.60	7.69
E (HOMO)[b]	0.53	7.61	0.60	8.30
2p (HOMO)	0.32	9.42	0.45	9.55
E (LUMO)	0.01	10.35	0.04	11.28
2p (LUMO)	0.13	9.46	0.11	10.97
E(HOMO) − *E*(LUMO)	0.25	9.73	0.67	8.46

[a] The descriptors used are defined in Table 5, footnote *a*. [b] If propionaldehyde (**6**) is omitted from the training set, $R^2 = 0.86$ and MAE = 5.09 kJ/mol.

mol, with the largest error arising from methoxybenzene (**22**; 7.2 kJ/mol; MAE = 2.1 kJ/mol without this compound). Thus, the data in this paper allow for a prediction of the activation energy by visual inspection. Of course, such predictions can be easily automatized. However, for most typical drug candidates with many nearby heteroatoms and complicated ring systems, this qualitative model may be hard to apply. If so, B3LYP calculations of BDE_{fr} or AM1 calculations of the α_{Morse} descriptor may be needed. Even more accurate results are obtained with explicit TS calculations for a methoxy or phenoxy radical model at AM1 level.

Applications on Druglike Molecules. Our previous results have established correlations between various theoretical descriptors and activation energies for hydrogen abstraction calculated by state-of-the-art DFT methods. In this section, we will illustrate the applicability of these methods on druglike molecules and compare with experimental data. Since we study only the intrinsic reactivity of the various sites, we will concentrate on metabolites of cytochrome P450 CYP3A4, because this enzyme has a very broad specificity (over 50% of the marketed drugs are metabolized by this enzyme) and therefore does not put severe restrictions on the orientation of the ligand in the binding pocket.[44,45] In addition, we combine the calculated activation energies with estimates of the solvent-accessible surface area (SASA)[41] of the various hydrogen atoms, which has been used as a measure of the intrinsic accessibility of the various sites.[46] We have studied two CYP3A4 substrates, progesterone[47] and dextromethorphan[48] (cf. Figure 7).

There are 30 hydrogen atoms in progesterone, all of which may react by H-atom abstraction except H4, which is expected to react via epoxide formation instead. We have studied all these 29 atoms with seven of our suggested methods (it is too time-demanding to do a full DFT optimization of all the TS energies). Our simple qualitative model directly suggests that the four C atoms next to sp^2-hybridized atoms, positions 2, 6, 17, and 21, should be more reactive than the tertiary (in positions 8, 9, and

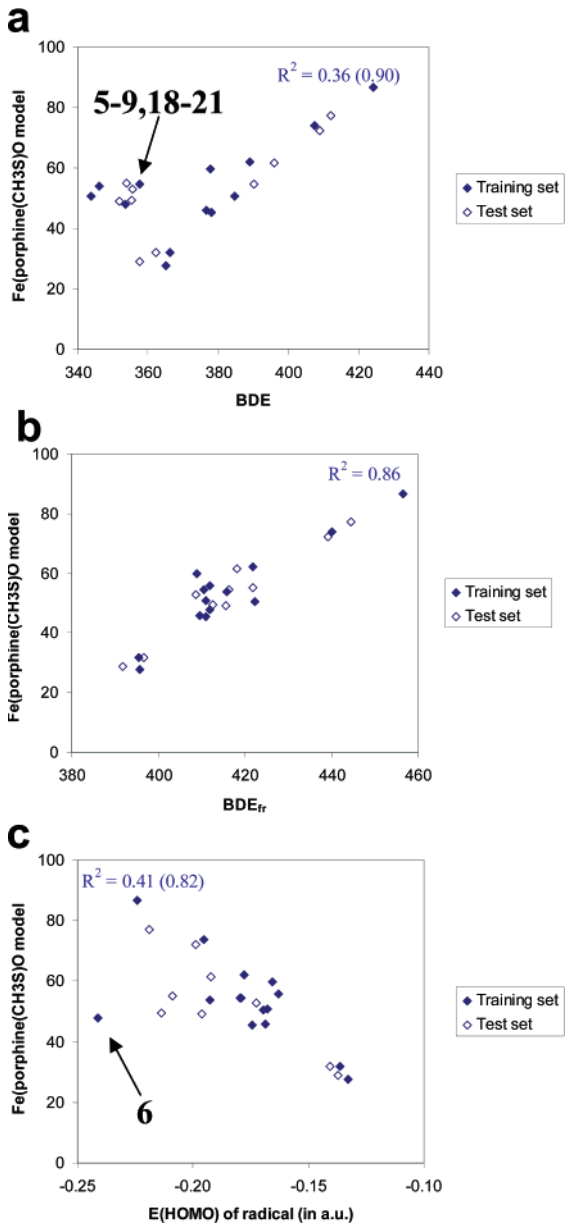

Figure 5. Activation energy for the Fe(porphine)(CH_3S)O model versus (a) BDE, (b) BDE_{fr}, and (c) *E*(HOMO) of the radical determined at B3LYP level. R^2 for the training set is also indicated and without the compounds that are highlighted in the plot in brackets.

14), secondary (in positions 1, 7, 11, 12, 15, and 16), and primary (in positions 18 and 19) carbon atoms in a sp^3 environment. This is in good agreement with experimental results, which show that CYP3A4 performs 2β, 6β, and 16α hydroxylations, whereas other CYP2C9 and CYP2C19 are responsible for hydroxylation of the H atoms in the 17 and 21 positions.[42]

However, the qualitative model cannot discriminate between the α and β positions. The intrinsic accessibility may be the explanation why H16α is more reactive than H16β (the SASA is ~55% larger for this atom, cf. Table 8), but for H2 and H6, the difference is smaller and goes in the wrong direction.

Apparently, the qualitative model is not accurate enough to correctly predict such details in the metabolism. However, all the other descriptors (the activation energy predicted with the methoxy radical and BDE_{fr} calculated at the B3LYP level or AM1 levels, as well as BDE_{Morse} or α_{Morse} calculated at the AM1 level; see Table 8), predict that the H2β and H6β atoms are

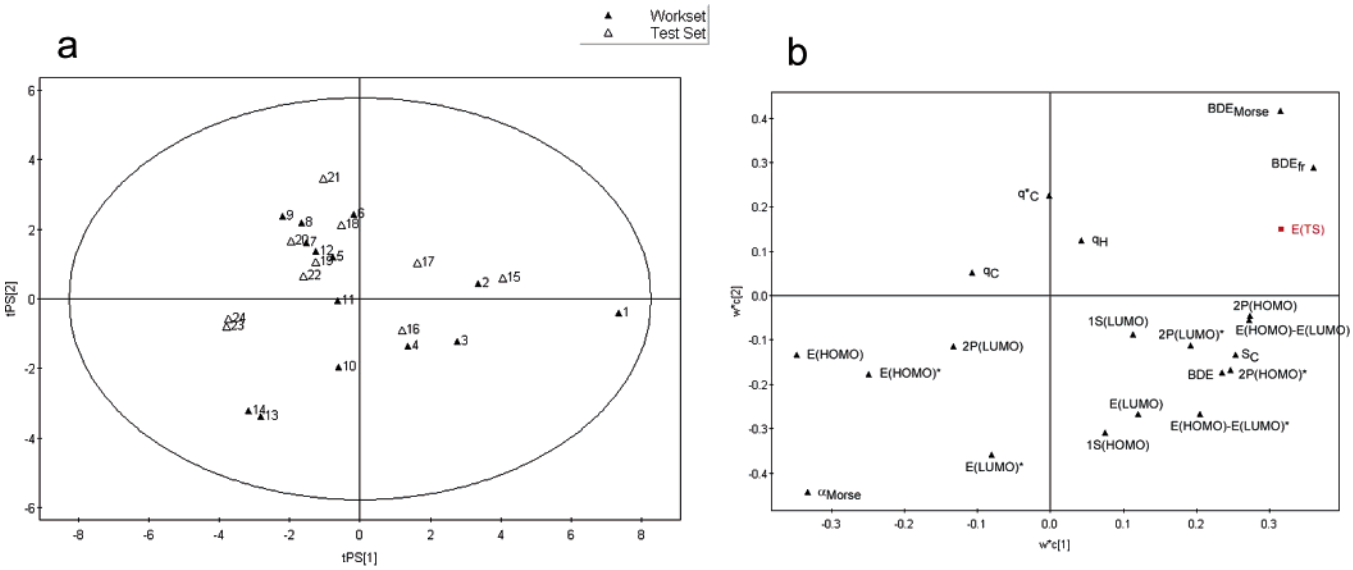

Figure 6. PLS analysis of B3LYP descriptors: (a) score plot and (b) loadings plot.

Table 7. Results of the PLS Model, with Either B3LYP or AM1 Descriptors[a]

		training set			test set
	no. of PC[b]	R^2	Q^2	MAE	MAE
		B3LYP			
all	2	0.90	0.79	3.66	4.39
all (VIP > 0.8)[c]	2	0.92	0.87	3.28	3.95
BDE_{fr}		0.86		4.32	3.21
		AM1			
all	2	0.88	0.76	3.40	4.38
all (VIP > 0.8)[c]	2	0.91	0.85	3.36	3.98
α_{Morse}		0.73		5.17	5.19

[a] The descriptors are those listed in Table 5. [b] Number of principal components in the model. [c] VIP, variable influence on projection.

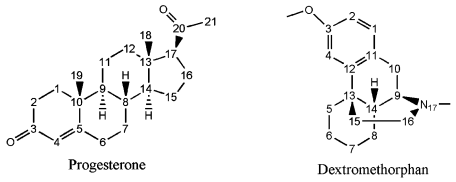

Figure 7. Atom numbers in progesterone and dextromethorphan [3-methoxy-17-methyl-(9α,13 α,14 α)-morphinan].

appreciably more reactive than the corresponding α atoms. These H atoms are almost perpendicular to the π electron system, which probably stabilize the TS and intermediate, since the radical would be in conjugation with the π system.

The secondary and tertiary C atoms are other possible sites of hydroxylation. The tertiary C atoms (positions 8, 9, and 14) have smaller predicted TS barriers than the secondary C atoms, but they are not very accessible (the SASA is less than 8 Å^2), which may explain why they are not hydroxylated.[43] The secondary sp^3-hybridized C atoms have larger predicted TS energies, but they are significantly more accessible (>10 Å^2). The most accessible secondary sp^3-hybridized H atom is that in the 16α position, which may explain why it is metabolized by CYP3A4.

Dextromethorphan contains 25 hydrogen atoms, three of which are in the benzene ring and are not expected to react by H abstraction. Our qualitative model predicts that the carbon atoms next to the amine (N17) are most reactive (C9, C16, and the free CH_3 group), followed by the C atom of the methoxy group, the C atom next to the aromatic ring (C10), and the tertiary C atom (C14). Experimentally, only the H atoms of the free methyl group that is next to the amine (N17) are metabolized by CYP3A4. This can be explained by the higher accessibility of these atoms (Table 9). One hydrogen atom in the methyl group bound to N17 has a significantly smaller activation barrier than the other two hydrogen atoms. This is the hydrogen atom that is pointing in the opposite direction of the lone pair of N17. Interestingly, CYP2D6 catalyzes O-dealkylation, which is also predicted by the qualitative model, and the methoxy hydrogen atoms have rather low predicted activation energies and a high solvent accessiblity.

Conclusions

With DFT calculations, we have determined the TS for the reaction of a realistic model of compound I in cytochromes P450 with 24 different substrates. We have then tested in a systematic way if these activation energies can be predicted by computationally less demanding methods. To this aim, the 24 substrates were divided into a training and a test set, which both contain primary, secondary, and tertiary aliphatic carbon atoms and nitrogen, oxygen, and sulfur, as well as sp^2 or aromatic carbon atoms next to the reactive atom.

First, we showed that the zero-point energy corrections and the big basis set [B3LYP/6-311++G(2d,2p)] energies normally employed in such calculations change the activation energies by only 1 and 6 kJ/mol on the average, but they increase the computational effort by a factor of 2−3. On the other hand, it is not possible to study the intermediate after the hydrogen abstraction rather than the TS for the hydrogen abstraction.

Second, we have evaluated the use of small radicals instead of a full iron−porphine model for compound I. We have tested three such models: *p*-nitrosophenoxy, phenoxy, and methoxy radicals. All give excellent correlations with the activation energies obtained with the full compound I model, with mean absolute errors of 3−4 kJ/mol for B3LYP and 4 kJ/mol with AM1. The phenoxy radical gave the best correlation and the smallest absolute errors. Furthermore, the O−H and C−H bond lengths of the atoms directly involved in the reaction correlate well with those obtained by using the full porphine model, which means that optimizations that contain compound **I** can be started from reasonable structures that will converge faster.

Third, we have tried to find a correlation between the activation energies and molecular descriptors of the substrate

Table 8. Activation Energies for the Abstraction of Various H Atoms of Progesterone, Estimated by Six Different Methods, and Solvent-Accessible Surface Areas[a]

		B3LYP		AM1				
atom no.		methoxy[b]	BDE_{fr}[c]	BDE_{Morse}[d]	α_{Morse}[e]	methoxy[f]	BDE_{fr}[g]	SASA[h]
1	α	61.3	58.8	57.9	63.9	71.1	55.9	13.9
1	*β*	57.4	55.9	59.2	65.2	65.7	57.3	11.8
2	α	61.3	67.4	64.1	70.2	68.9	66.2	21.8
2	*β*	54.8	44.7	51.1	56.1	64.6	58.6	16.4
6	α	63.1	63.0	60.5	66.8	66.5	59.0	17.9
6	*β*	41.4	34.6	46.1	52.4	53.0	43.7	15.7
7	α	60.6	61.9	57.5	61.3	63.9	57.6	15.0
7	*β*	62.3	59.1	62.4	67.7	66.7	59.2	13.4
8	*β*	65.9	47.2	53.2	60.3	76.6	50.8	3.2
9	α	55.0	43.8	44.7	59.7	57.5	41.7	6.3
11	α	59.3	54.4	55.4	58.8	62.6	53.7	12.3
11	*β*	60.8	55.3	56.8	62.4	61.0	54.5	6.6
12	α	54.0	59.6	57.1	60.6	57.9	57.1	12.5
12	*β*	64.7	63.0	63.9	70.3	70.1	63.6	9.5
14	α	51.9	40.7	39.5	53.1	49.8	40.2	7.3
15	α	57.4	57.9	63.2	63.2	62.9	60.2	16.2
15	*β*	60.8	61.4	68.1	67.7	66.2	61.8	12.9
18		63.2	68.4	65.2	61.5	67.1	66.2	9.5
18		74.0	74.4	65.7	66.2	74.7	70.0	11.6
18		63.2	70.1	65.4	63.6	67.1	66.9	13.2
19		70.8	72.0	66.1	65.1	74.6	67.1	10.6
19		68.6	69.5	65.0	62.7	73.6	68.0	14.3
19		66.3	70.8	67.6	64.8	71.7	68.9	15.4
21		65.5	66.8	67.8	65.6	80.1	75.0	18.5
21		61.2	60.4	61.3	60.9	74.1	71.9	25.3
21		63.4	86.0	69.9	70.2	77.3	76.3	14.5
				Conformation 1[i]				
16	α	56.0	60.5	66.2	65.9	64.2	59.6	21.8
16	*β*	69.6	64.4	61.0	62.9	70.8	65.5	13.7
17	α	44.4	46.3	55.7	64.8	59.6	53.4	9.4
				Conformation 2[i]				
16	α	56.7	60.3	66.2	66.3	61.4	59.5	17.3
16	*β*	53.1	56.2	62.7	63.2	60.2	59.1	11.2
17	α	45.8	26.2	44.5	56.8	58.7	47.3	10.6

[a] Estimates were based on the correlation lines obtained for the training set (compounds **1**−**14**). All activation energies are given in kilojoules per mole. [b] Correlation line: $E(TS) = 0.9081x + 31.98$, where x is the activation energy obtained with the methoxy radical for the reaction at B3LYP/6-31G(d) level. [c] Correlation line: $E(TS) = 0.8259x - 32.30$, where x is the BDE_{fr} descriptor at B3LYP/6-31G(d) level. [d] Correlation line: $E(TS) = 0.4475x - 252.92$, where x is the BDE_{Morse} descriptor determined at AM1 level. [e] Correlation line: $E(TS) = -761.83x + 1185.79$, where x is the α_{Morse} descriptor determined at AM1 level. [f] Correlation line: $E(TS) = 2.044x - 72.77$, where x is the activation energy obtained with the methoxy radical for the reaction at AM1 level. [g] Correlation line: $E(TS) = 0.7856x - 256.95$, where x is the BDE_{fr} descriptor at AM1 level. [h] Solvent-accessible surface area (square angstroms) was determined with the SAVOL program.[41] [i] The energy of conformation 1 is 4.1 kJ/mol lower in energy at B3LYP/6-31G(d) level than conformation 2. The torsional angle τ (defined by H17−C17−C20−C23) is 26.5° for conformation 1and −72.3° for conformation 2−.

Table 9. Activation Energies for the Abstraction of Various H Atoms of Dextromethorphan, Estimated by Six Different Methods, and Solvent-Accessible Surface Areas[a]

	B3LYP		AM1				
atom no.	methoxy[b]	BDE_{fr}[c]	BDE_{Morse}[d]	α_{Morse}[e]	methoxy[f]	BDE_{fr}[g]	SASA[h]
3 (OCH_3)	49.4	74.0	70.1	70.6	63.9	72.7	28.6
3 (OCH_3)	49.3	50.9	58.0	51.7	52.1	58.1	21.6
3 (OCH_3)	49.4	50.8	58.0	51.5	52.1	58.2	21.6
5	58.5	58.9	54.1	60.5	59.5	53.4	13.4
5	59.3	56.7	66.0	68.7	66.2	57.4	12.3
6	60.5	60.0	62.2	66.4	62.5	57.3	23.7
6	68.8	61.0	55.4	59.6	72.7	56.6	12.5
7	61.7	60.7	63.2	67.4	61.8	57.9	23.7
7	58.6	60.0	56.1	59.6	57.0	54.9	18.4
8	65.3	58.1	55.6	59.4	70.2	55.1	8.1
8	57.6	59.1	61.6	65.5	62.0	57.6	18.6
9	39.6	44.4	46.6	62.2	44.3	36.7	11.8
10	49.6	45.5	53.0	58.5	50.0	47.0	9.0
10	37.5	28.7	49.2	56.0	40.7	37.7	14.3
14	51.2	51.2	55.9	70.3	57.6	49.3	10.4
15	57.2	60.0	59.7	63.7	59.5	58.7	15.2
15	65.4	65.4	64.6	68.4	71.0	61.5	15.6
16	37.6	49.6	42.1	47.8	42.5	42.4	19.1
16	29.0	26.4	23.6	28.7	32.1	24.8	10.8
17 (NCH_3)	36.6	59.0	49.1	49.3	47.2	51.3	23.1
17 (NCH_3)	36.6	65.9	51.5	50.6	47.2	57.7	21.2
17 (NCH_3)	20.7	29.6	28.3	26.0	26.8	33.5	16.3

[a] Estimates were based on the correlation lines obtained for the training set (compounds **1**−**14**). All activation energies are given in kilojoules per mole. [b−g] Correlation lines are as described in the corresponding footnotes for Table 8. [h] Solvent-accessible surface area (square angstroms) was determined with the SAVOL program.[41]

or the corresponding radicals. The best results are obtained with the BDE obtained without relaxing the geometry of the radical (BDE_{fr}) with the B3LYP method, without zero-point and big basis set corrections (MAE = 4 kJ/mol). At the AM1 level, only slightly worse results (MAE = 5 kJ/mol) can be obtained from a three-point fit of the BDE, by use of a Morse potential (α_{Morse}). Interestingly, the results are not significantly improved by a PLS model involving all the 20 studied descriptors.

However, the fastest and most accurate results can actually be obtained directly from the DFT activation energies and a classification of the chemical function and environment of the reactive carbon atom: substrates with nitrogen atoms next to the reactive carbon have activation energies of 30 kJ/mol. Those with oxygen and sulfur atoms next to the reactive carbon have activation energies of 47 kJ/mol. The activation energy of substrates with aromatic or sp^2-hybridized atoms next to the reactive C is 53 kJ/mol. Finally, primary carbon atoms have an activation energy of 74 kJ/mol, whereas secondary and tertiary carbons, as well as carbons with bound F atoms, have activation energies around 61 kJ/mol. By use of these five groups, the activation energies in the test set can be predicted with a MAE of only 2.5 kJ/mol.

Thus, our investigation allows us to define a hierarchy of methods (in terms of accuracy and computational load) to estimate the intrinsic reactivity of each aliphatic carbon atom in a general drug candidate: For simple groups, an accurate estimate of the activation energy can be directly obtained by visual inspection of the chemical environment. This should be enough for most screenings of drugs. Only if the drug contains atoms with complicated or conflicting environments or functional groups not covered in this investigation does it seem to be justified to use more demanding methods to estimate the activation energy. If so, reasonably accurate results can be obtained from α_{Morse}, calculated at the AM1 level. These calculations can easily be automatized and takes a few minutes for a typical drug candidate. If a higher accuracy is needed or if the AM1 method is not accurate enough for the substrate of interest, the BDE_{fr} can be calculated at the B3LYP level in less than an hour for a typical drug. Even more accurate results are obtained with explicit TS calculations using a methoxy or phenoxy radical model, calculations that take ∼10 min at the AM1 level and a few hours at the B3LYP level. Finally, the most accurate results are obtained by a DFT optimization of the TS with a full porphyrin model. Such calculations take several weeks, however.

We have illustrated the applicability of our results on two druglike molecules, progesterone and dextromethorphan. We show that we can predict most experimental sites of metabolism with the simple qualitative model, combined with an estimate of the intrinsic accessibility of each site from the SASA.

However, for a more detailed view of the reactivities (e.g., the preference between α and β atoms), DFT or AM1 calculations are needed.

Another interesting result in this investigation is that all activation energies are quite low, <80 kJ/mol (excluding methane). This means that essentially all CH groups in a drug can be metabolized by cytochrome P450 within a reasonable time (an activation energy of 80 kJ/mol corresponds to a rate constant of $\sim 0.1\ s^{-1}$). This illustrates the extreme reactivity of compound I in cytochromes P450. This reduces the fundamental question whether and how a drug with be metabolized by these enzymes to two simpler ones: What groups are sterically accessible to the oxygen atom of compound I and which of these groups have the highest intrinsic reactivity. The present investigation solves the second issue and gives a few predictive rules that are directly applicable to nearly any drug candidate. The first issue is what remains to be solved. It involves steric factors of the enzyme, which of course differ between different types of cytochromes P450 and may be approached by docking and MD simulations.[49−52]

Acknowledgment. This work was supported by a grant from the Carlsberg Foundation and by funding from the research school in pharmaceutical science at Lund University (FLÄK) and the Swedish Research Council. It has also been supported by computer resources from Lunarc at Lund University.

Supporting Information Available: B3LYP and AM1 energies and C−H and O−H bond lengths in the TS with small organic radical models of compound **I**. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Guengerich, F. P. Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem. Res. Toxicol.* **2001**, *14*, 611−650.
(2) Meunier, B.; de Visser, S. P.; Shaik, S. Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. *Chem. Rev.* **2004**, *104*, 3947−3980.
(3) Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. *Chem. Rev.* **2005**, *105*, 2279−2328.
(4) Choe, Y. K.; Nagase, S. Effect of the axial cysteine ligand on the electronic structure and reactivity of high-valent iron(IV) oxoporphyrins (Compound I): A theoretical study. *J. Comput. Chem.* **2005**, *26*, 1600−1611.
(5) de Visser, S. P.; Ogliaro, F.; Sharma, P. K.; Shaik, S. What factors affect the regioselectivity of oxidation by cytochrome P450? A DFT study of allylic hydroxylation and double bond epoxidation in a model reaction. *J. Am. Chem. Soc.* **2002**, *124*, 11809−11826.
(6) de Visser, S. P.; Kumar, D.; Cohen, S.; Shacham, R.; Shaik, S. A predictive pattern of computed barriers for C−H hydroxylation by compound I of cytochrome P450. *J. Am. Chem. Soc.* **2004**, *126*, 8362−8363.
(7) Kumar, D.; de Visser, S. P.; Sharma, P. K.; Cohen, S.; Shaik, S. Radical clock substrates, their C−H hydroxylation mechanism by cytochrome P450, and other reactivity patterns: What does theory reveal about the clocks' behavior? *J. Am. Chem. Soc.* **2004**, *126*, 1907−1920.
(8) Guengerich, F. P.; Krauser, J. A.; Johnson, W. W. Rate-limiting steps in oxidations catalyzed by rabbit cytochrome P450 1A2. *Biochemistry* **2004**, *43*, 10775−10788.
(9) 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.
(10) 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.
(11) Jones, J. P.; Korzekwa, K. R. Predicting the rates and regioselectivity of reactions mediated by the P450 superfamily. *Methods Enzymol.* **1996**, 326−335.
(12) Vermeulen, N. P. E. Prediction of drug metabolism: The case of cytochrome P450 2D6. *Curr. Top. Med. Chem.* **2003**, *3*, 1227−1239.
(13) Zamora, I.; Afzelius, L.; Cruciani, G. Predicting drug metabolism: A site of metabolism prediction tool applied to the cytochrome P4502C9. *J. Med. Chem.* **2003**, *46*, 2313−2324.
(14) Collins, J. R.; Loew, G. H. Theoretical-Study of the Product Specificity in the Hydroxylation of Camphor, Norcamphor, 5,5-Difluorocamphor, and Pericyclocamphanone by Cytochrome-P-450Cam. *J. Biol. Chem.* **1988**, *263*, 3164−3170.
(15) 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.
(16) 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.
(17) Rietjens, I. M. C. M.; Soffers, A. E. M. F.; Veeger, C.; Vervoort, J. Regioselectivity of Cytochrome-P-450 Catalyzed Hydroxylation of Fluorobenzenes Predicted by Calculated Frontier Orbital Substrate Characteristics. *Biochemistry* **1993**, *32*, 4801−4812.
(18) De Groot, M. J.; Den Kelder, G. M. D. O.; Commandeur, J. N. M.; Van Lenthe, J. H.; Vermeulen, N. P. E. Metabolite Predictions for Parasubstituted Anisoles Based on Ab Initio Complete Active Space Self-Consistent-Field Calculations. *Chem. Res. Toxicol.* **1995**, *8*, 437−443.
(19) Pudzianowski, A. T.; Loew, G. H. Quantum-Chemical Studies of Model Cytochrome P450 Hydrocarbon Oxidation Mechanisms.1. A Mindo-3 Study of Hydroxylation and Epoxidation Pathways for Methane and Ethylene. *J. Am. Chem. Soc.* **1980**, *102*, 5443−5449.
(20) Higgins, L.; Korzekwa, K. R.; Rao, S.; Shou, M. G.; Jones, J. P. An assessment of the reaction energetics for cytochrome P450-mediated reactions. *Arch. Biochem. Biophys.* **2001**, *385*, 220−230.
(21) 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.
(22) 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.
(23) Ogliaro, F.; Harris, N.; Cohen, S.; Filatov, M.; de Visser, S. P.; Shaik, S. A model "rebound" mechanism of hydroxylation by cytochrome P450: Stepwise and effectively concerted pathways, and their reactivity patterns. *J. Am. Chem. Soc.* **2000**, *122*, 8977−8989.
(24) Yoshizawa, K.; Shiota, Y.; Kagawa, Y. Energetics for the oxygen rebound mechanism of alkane hydroxylation by the iron-ore species of cytochrome p450. *Bull. Chem. Soc. Jpn.* **2000**, *73*, 2669−2673.
(25) Ogliaro, F.; Cohen, S.; Filatov, M.; Harris, N.; Shaik, S. The high-valent compound of cytochrome P450: The nature of the Fe−S bond and the role of the thiolate ligand as an internal electron donor. *Angew. Chem., Int. Ed.* **2000**, *39*, 3851−3855.
(26) Schoneboom, J. C.; Lin, H.; Reuter, N.; Thiel, W.; Cohen, S.; Ogliaro, F.; Shaik, S. The elusive oxidant species of cytochrome P450 enzymes: Characterization by combined quantum mechanical/molecular mechanical (QM/MM) calculations. *J. Am. Chem. Soc.* **2002**, *124*, 8142−8151.
(27) Yin, H. Q.; 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.
(28) Li, C. S.; Wu, W.; Kumar, D.; Shaik, S. Kinetic isotope effect is a sensitive probe of spin state reactivity in C−H hydroxylation of *N*,*N*-dimethylaniline by cytochrome P450. *J. Am. Chem. Soc.* **2006**, *128*, 394−395.
(29) Yoshizawa, K.; Kamachi, T.; Shiota, Y. A theoretical study of the dynamic behavior of alkane hydroxylation by a compound I model of cytochrome P450. *J. Am. Chem. Soc.* **2001**, *123*, 9806−9816.
(30) Becke, A. D. Density-Functional Thermochemistry. 3. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(31) Becke, A. D. A New Mixing of Hartree−Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372−1377.
(32) Lee, C. T.; Yang, W. T.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula Into A Functional of the Electron-Density. *Phys. Rev. B* **1988**, *37*, 785−789.
(33) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. The Development and Use of Quantum-Mechanical Molecular-Models. 76. AM1−A New General-Purpose Quantum-Mechanical Molecular-Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(34) Schafer, A.; Horn, H.; Ahlrichs, R. Fully Optimized Contracted Gaussian-Basis Sets for Atoms Li to Kr. *J. Chem. Phys.* **1992**, *97*, 2571−2577.
(35) Hehre, W.; Radom, L.; Schleyer, P.; Pople, J. *Ab initio molecular orbital theory*; Wiley-Interscience: New York: 1986.

(36) Lewin, J. L.; Cramer, C. J. Rapid Quantum Mechanical Models for the Computational Estimation of C−H Bond Dissociation Energies as a Measure of Metabolic Stability. *Mol. Pharmacol.* **2005**, *1*, 128−135.
(37) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. The MIDI! basis set for quantum mechanical calculations of molecular geometries and partial charges. *Theor. Chim. Acta* **1996**, *93*, 281−301.
(38) Leatherbarrow, R. J. *GraFit, Version 5*; Erithacus Software Ltd., Horley, U.K.
(39) Simca-P, Version 10.0, Umetrics AB, Umeå, Sweden.
(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, 2003.
(41) Pearlman, R. S.; Skell, J. M.; Deanda, F. *SAVOL3: A Program for the Atomic Partitioning of the Surface Area and Volume of Molecular Structures*; The University of Texas: Austin, TX, 1999.
(42) Himo, F.; Siegbahn, P. E. M. Quantum chemical studies of radical-containing enzymes. *Chem. Rev.* **2003**, *103*, 2421−2456.
(43) Siegbahn, P. E. M.; Blomberg, M. R. A. Transition-metal systems in biochemistry studied by high-accuracy quantum chemical methods. *Chem. Rev.* **2000**, *100*, 421−437.
(44) Rendic, S. Summary of information on human CYP enzymes: Human P450 metabolism data. *Drug Metab. Rev.* **2002**, *34*, 83−448.
(45) 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.
(46) 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.
(47) 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.
(48) Yu, A. M.; 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*, 1514−1520.
(49) de Graaf, C.; Oostenbrink, C.; Keizers, P. H. J.; van der Wijst, T.; Jongejan, A.; Vemleulen, 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.
(50) 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.
(51) Kemp, C. A.; Flanagan, J. U.; van Eldik, A. J.; Marechal, J. D.; 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.
(52) Zhou, D. S.; 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.

JM060551L