A Model for Predicting Likely Sites of CYP3A4-mediated Metabolism on Drug-like Molecules

Suresh B. Singh,* Lucy Q. Shen,# Matthew J. Walker, and Robert P. Sheridan

Department of Molecular Systems Merck Research Laboratories, 126 E. Lincoln Avenue, RY50SW-100, Rahway, New Jersey 07065-0900

Received September 12, 2002

We have developed a rapid semiquantitative model for evaluating the relative susceptibilities of different sites on drug molecules to metabolism by cytochrome P450 3A4. The model is based on the energy necessary to remove a hydrogen radical from each site, plus the surface area exposure of the hydrogen atom. The energy of hydrogen radical abstraction is conventionally measured by AM1 semiempirical molecular orbital calculations. AM1 calculations show the following order of radical stabilities for the hydrogen atom abstractions: sp2 centers > heteroatom sp3 centers > carbon sp3 centers. Since AM1 calculations are too time intensive for routine work, we developed a statistical trend vector model, which is used to estimate the AM1 abstraction energy of a hydrogen atom from its local atomic environment. We carried out AM1 and trend vector calculations on 50 CYP3A4 substrates whose major sites of metabolism are known in the literature. A plot of the lowest hydrogen radical formation energy versus its sterically accessible surface area exposure for these 50 substrates shows that only those hydrogen atoms with solvent accessible surface area exposure ≥ 8.0 Å^2 are susceptible to CYP3A4-mediated metabolism. This approach forms the basis for our general model, which predicts sites on drugs that are susceptible to cytochrome P450 3A4-mediated hydrogen radical abstraction followed by a hydroxylation reaction. This model, in conjunction with specific enzyme site binding requirements, can aid in identifying possible sites of metabolism catalyzed by other cytochrome P450 enzymes.

Introduction

One of the major requirements for a successful drug is its ability to be administered orally as a therapeutic intervention. Oral bioavailability of a drug largely depends on its ability to withstand degradation by intestinal and hepatic metabolizing enzymes.[1]

Mammalian liver expresses a wide variety of metabolizing enzymes including cytochrome P450 3A. Mammalian small intestines also express CYP3A, which metabolizes xenobiotic substances that attempt to cross the intestinal lining into the bloodstream.[1] Drugs that enter the bloodstream from the digestive tract get transported to the liver via the portal vein and are subjected to "first-pass" metabolism.[1] Cytochrome P450 3A4 (CYP3A4) has a broad range of substrate specificity and metabolizes nearly 50% of the marketed drugs.[2] The set of 50 CYP3A4 substrates used in our study is shown in Figure 1. CYP3A4 catalyzes the following reactions: alkyl carbon hydroxylation, O- and N-dealkylations, epoxidation, and less frequently, aromatic ring hydroxylation.[3] Due to CYP3A4's broad substrate specificity, it has not been possible to predict a priori which site on a drug is susceptible to metabolism by that enzyme.

A semiempirical quantum mechanical model was first applied to small organic compounds and halogenated alkanes to predict rates and sites of metabolism.[4,5] More recently, a model for aromatic oxidation has been proposed.[6]

Cytochrome P450 enzymes metabolize substrates by various mechanisms, including hydrogen atom abstraction by an activated oxygen intermediate in the enzyme active site. The support for the presence of the incipient heme-iron-oxo radical species comes from the crystallographic studies of the redox cycle of P450cam.[7]

The approach presented here relies on the assumption that CYP3A4 substrates, in the absence of specific binding forces that orient them, tumble in the enzyme active site until the hydrogen atom with lowest activation barrier and sufficient solvent exposure gets abstracted by the activated oxygen radical, thereby creating a radical on the substrate. The radical so-created would rapidly recombine with hydroxy radical to yield the final hydroxylated major metabolite. The rate-determining step for the hydroxylation reaction is hydrogen atom abstraction,[8] and this is the basis for our approach.

Here we show that a model that relies solely on the electronics and intramolecular sterics of a set of drug-like molecules can predict the sites of CYP3A4 metabolism to a useful degree of accuracy.

Theory

Our model is based on the assumption that the CYP3A4 susceptibility is largely dependent on the electronic environment surrounding the hydrogens in any drug molecule and that the extraction of a hydrogen atom is the rate determining step.

* Corresponding author: Suresh B. Singh, RY50SW 100, P. O. Box 2000. Rahway, NJ 07065. Email: suresh_singh@merck.com. Tel: 732-594-4954. Fax: 732-594-4224.

Current address: Harvard Medical School, HST Class of 2004, Boston, MA 02115.

10.1021/jm020400s CCC: $25.00
Published on Web 03/15/2003

An example of hydrogen atom abstraction mediated metabolism is the N-demethylation reaction catalyzed by CYP3A4 (Figure 2). The C−H bond of the methyl substituent on the tertiary amine undergoes a homolytic cleavage with the assistance of the heme-iron-oxo radical species in the enzyme active site to yield two highly reactive species: the methylamine and the hydrogen radicals. The methylamine radical undergoes an addition of an HO· to form an unstable hydroxylated intermediate followed by the elimination of formaldehyde to form a desmethyl secondary amine. Note that our approach does not take into account that electrons can be abstracted from the nitrogen lone-pair leading to the loss of proton from the attached methyl group.

The energy to extract the hydrogen atom on the methyl group is given by

$$\Delta H_{act} = \Delta H_2 - \Delta H_1$$

Thus, the energy for a given hydrogen atom is given by the difference between the heat of formation of the native substrate (ΔH_1) and that of its radical (ΔH_2). Henceforth, we will call the ΔH_{act} "hydrogen abstraction energy."

Methodology

AM1 Calculations. For any given molecule, we calculated the hydrogen abstraction energy for each hydrogen using the AM1 Hamiltonian with AMPAC V. 6.0 software.[9] In all cases, the UHF Hamiltonian was used with a SCF convergence parameter SCFCRT = 0.001. The geometry was optimized in XYZ mode with the convergence parameter GNORM = 0.1. To facilitate convergence and to decrease the time necessary for the AM1 calculations, the molecule was trimmed such that nonhydrogen atoms further than five bonds from the hydrogen to be abstracted were removed and hydrogens added back to make up the valence. However, all atoms in conjugated or aromatic ring systems were retained if any atom in the ring was within five bonds. We saw that the removal of extra atoms, for the 50 molecules examined, did not appreciably change the hydrogen abstraction energy (<0.5 kcal/mol). Topologically equivalent hydrogens have nearly identical heats of formation, so it was not necessary to calculate every hydrogen in a molecule.

These AM1 calculations were done for two sets of compounds in their neutral state. The first set consisted of 50 drug molecules for which the major site of metabolism is known (Figure 1).[2] The second set, used for the calibration of the trend vector model, consisted of about 400 randomly selected drug-like compounds from the MDDR (www.mdli.com) database, CYP3A4 substrates from Ekins et al.,[10] and small organic compounds.[4] In the second set, there were about 3984 topologically distinct hydrogens. The energies ranged from 10 to 95 kcal/mol.

Sterically Accessible Surface Area. The solvent accessible surface, a measure of steric accessibility, was calculated using the Shrake and Rupley algorithm[11] using a probe size of 1.4 Å and the following set of atomic radii: H 1.0, C 1.6, N 1.4, O 1.35, F 1.45, P 1.8, S 2.0, Cl 1.75, Br 1.85, I 1.98. The surface was taken as the average over up to 25 low energy conformations generated by the flexibase procedure.[12] Topologically equivalent hydrogens (e.g., the three hydrogens on the same methyl group) may not have identical surface areas. The surface area of a hydrogen was taken as the maximum over all topologically equivalent hydrogens, the assumption being that CYP3A4 would attack the most exposed hydrogen.

Calibration of the Trend Vector Model. Even with the approach of trimming molecules, AM1 calculations take too long to be a practical method for many applications. Therefore, it was desirable to find a faster way of predicting hydrogen abstraction energy from the local atomic environment.

Trend vector analysis[13,14] is an implementation of the partial-least squares (PLS) QSAR method wherein an "activity" can be modeled on the presence or absence of chemical descriptors. In this case, the "activity" is the AM1 hydrogen abstraction energy for a given hydrogen. We need to invent descriptors that capture the local topological environment of the hydrogen. The descriptors are of the form H-AT1, H-AT1-AT2, H-AT1-AT2-AT3, H-AT1-AT2-AT3-AT4, where AT1 is the type of atom to which the hydrogen is directly bonded, AT2, the atom two bonds away, etc. Thus, any given hydrogen would have a large number of descriptors associated with it. The atom type includes the element, the number of nonhydrogen neighbors, and the hybridization. An atom type might also include special labels such as ar5 (aromatic five-membered ring), ar6 (aromatic six-membered ring), br5 (bridgehead atom in a five-membered ring), and br6 (bridgehead atom in a six-membered ring). Thus, the descriptor H-CX3sp3br6 would indicate that the atom to which the hydrogen is attached is an sp3 carbon with three nonhydrogen neighbors and is a bridgehead carbon in a six-membered ring.

The special labels (ar5, br5, etc.) were found to be important for getting a good fit to the data. They make physical sense as well. For example, carbon radicals formed when the hydrogen radical is abstracted prefer to be planar, and anything that prevents planarity, such as the carbon being in a bridgehead, will tend to raise the hydrogen abstraction energy. We used the PATTY methodology[15] to assign descriptors for all the hydrogens in our study.

As with any PLS method, the trend vector method works as follows: (i) fit a first component (a linear combination of the original descriptors) to the data, (ii) fit a second component to the residuals (differences between the predicted and observed activities) from the first fit, and (iii) fit a third component to the residuals from the first and second fit, etc. To avoid overfitting, a randomization method is used to ensure that the correlation between descriptors and residuals is statistically significant for each new component. If the statistical significance for a component is < 3, it is not added.[14]

Since the final trend vector model **T** has the form of a vector, the predicted abstraction energy for a hydrogen is the projection of its descriptors onto **T**:

$$\text{hydrogen abstraction energy} = \sum_{d} T(d)f(d) + C$$

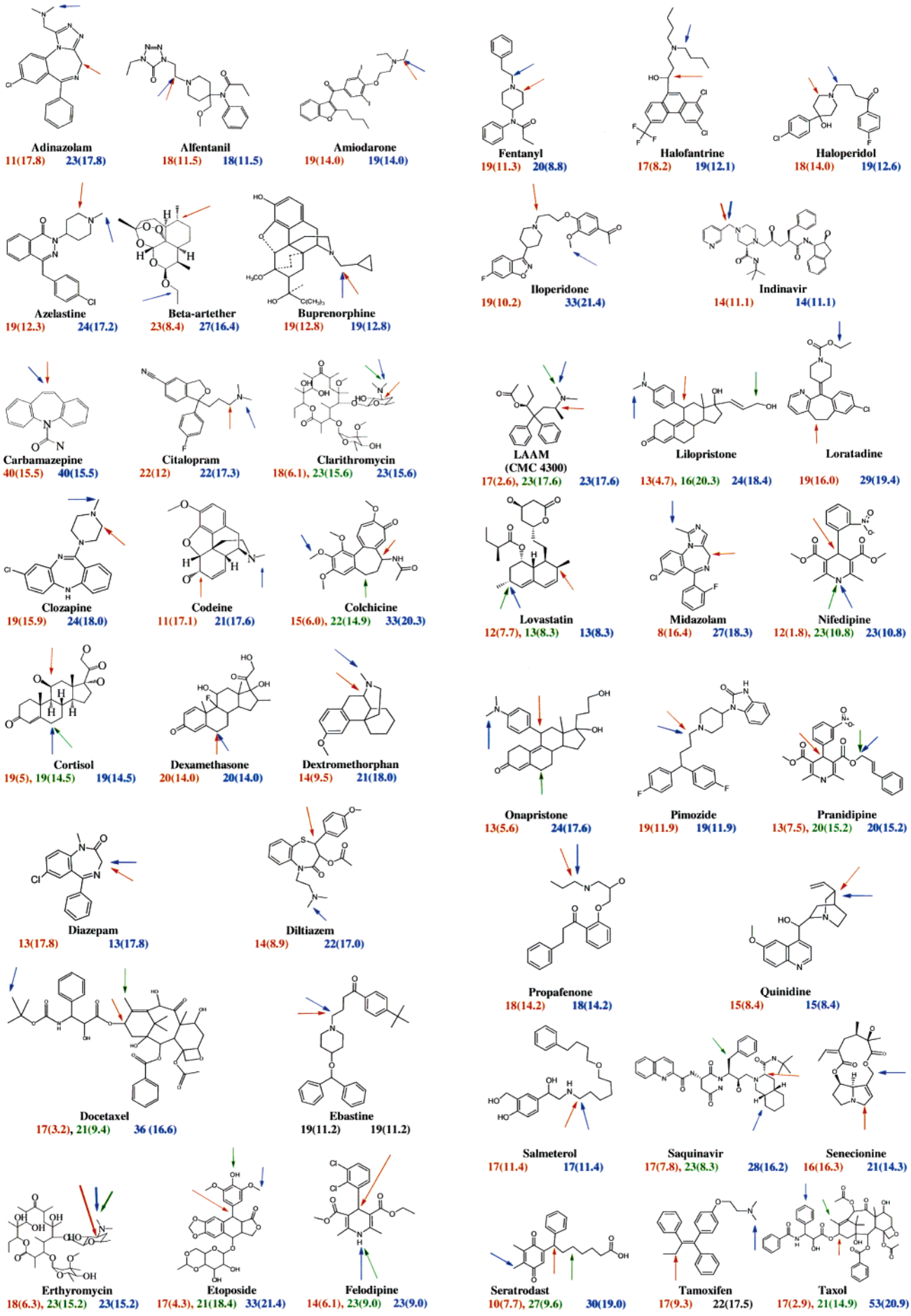
where d goes over the unique descriptors in the training set and f(d) = 1 if the descriptor is present for the hydrogen and 0 if it is absent. C is a constant.

Our main goal in developing this methodology was to facilitate rapid visualization of sites in a molecule that could be susceptible to CYP3A4 metabolism. Therefore, we chose to implement relatively less accurate trend vector methodology for rapid calculation of hydrogen abstraction energy.

Results

Fit of Trend Vector to AM1 Energies. A total of 14 significant PLS components were allowed, indicating that an excellent fit to the AMPAC/AM1 energies could be obtained with our descriptors. The R^2 for the fit of 3984 hydrogens was 0.98, and the standard error of the prediction was 2.06 kcal/mol

The highest and lowest 20 components of trend vector (**T**) are shown in Table 1. The intercept C = 40.6 kcal/mol. Figure 3 shows the correlation plot of trend vector predicted hydrogen abstraction energy vs AMPAC/AM1 computation for the hydrogens at major metabolic sites in the 50 substrates. This plot shows that the trend vector is able to predict AM1 hydrogen abstraction energies quite reliably. This constitutes a real prediction for the trend vector model because the 50 substrates


Adinazolam
11(17.8) 23(17.8)
Alfentanil
18(11.5) 18(11.5)
Amiodarone
19(14.0) 19(14.0)
Azelastine
19(12.3) 24(17.2)
Beta-artether
23(8.4) 27(16.4)
Buprenorphine
19(12.8) 19(12.8)
Carbamazepine
40(15.5) 40(15.5)
Citalopram
22(12) 22(17.3)
Clarithromycin
18(6.1), 23(15.6) 23(15.6)
Clozapine
19(15.9) 24(18.0)
Codeine
11(17.1) 21(17.6)
Colchicine
15(6.0), 22(14.9) 33(20.3)
Cortisol
19(5), 19(14.5) 19(14.5)
Dexamethasone
20(14.0) 20(14.0)
Dextromethorphan
14(9.5) 21(18.0)
Diazepam
13(17.8) 13(17.8)
Diltiazem
14(8.9) 22(17.0)
Docetaxel
17(3.2), 21(9.4) 36 (16.6)
Ebastine
19(11.2) 19(11.2)
Erthyromycin
18(6.3), 23(15.2) 23(15.2)
Etoposide
17(4.3), 21(18.4) 33(21.4)
Felodipine
14(6.1), 23(9.0) 23(9.0)
Fentanyl
19(11.3) 20(8.8)
Halofantrine
17(8.2) 19(12.1)
Haloperidol
18(14.0) 19(12.6)
Iloperidone
19(10.2) 33(21.4)
Indinavir
14(11.1) 14(11.1)
LAAM
(CMC 4300)
17(2.6), 23(17.6) 23(17.6)
Lilopristone
13(4.7), 16(20.3) 24(18.4)
Loratadine
19(16.0) 29(19.4)
Lovastatin
12(7.7), 13(8.3) 13(8.3)
Midazolam
8(16.4) 27(18.3)
Nifedipine
12(1.8), 23(10.8) 23(10.8)
Onapristone
13(5.6) 24(17.6)
Pimozide
19(11.9) 19(11.9)
Pranidipine
13(7.5), 20(15.2) 20(15.2)
Propafenone
18(14.2) 18(14.2)
Quinidine
15(8.4) 15(8.4)
Salmeterol
17(11.4) 17(11.4)
Saquinavir
17(7.8), 23(8.3) 28(16.2)
Senecionine
16(16.3) 21(14.3)
Seratrodast
10(7.7), 27(9.6) 30(19.0)
Tamoxifen
17(9.3) 22(17.5)
Taxol
17(2.9), 21(14.9) 53(20.9)

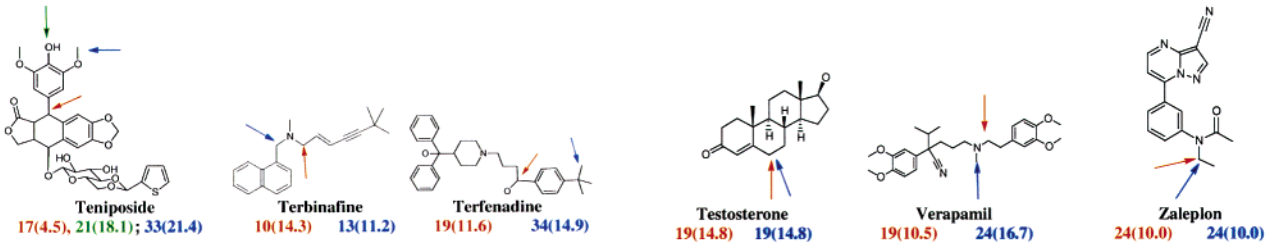


Figure 1. Comparison of the trend vector predicted hydrogen abstraction energy (red) versus experimentally (blue) known major metabolic site of metabolism of CYP3A4 substrates. The number in the parenthesis is the largest solvent accessible surface area exposure for that site. If the surface area exposure of the site marked in red is < 8 Å cutoff (see Figure 5), then the hydrogen atom next lowest in hydrogen abstraction energy, with exposure ≥ 8 Å, is marked with a green arrow and its energy and surface exposure shown in green next to the lowest energy number.

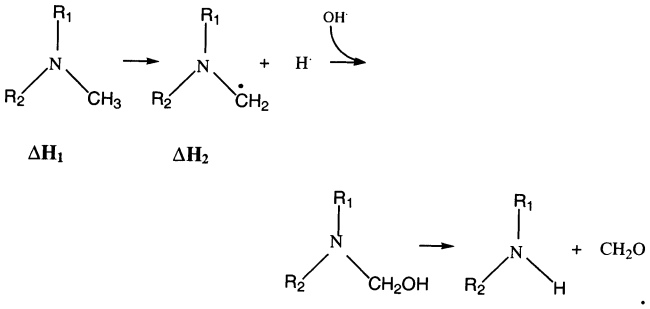


Figure 2. An example of hydrogen atom abstraction mediated metabolism by CYP3A4.

were not included in the 400 compounds used to calibrate the trend vector model.

Correlation of Metabolic Site with Hydrogen Abstraction Energy. The histograms in Figure 4 show the following: the AM1 and trend vector predicted hydrogen abstraction energies of (i) all hydrogens, (ii) the hydrogen with the lowest energy in the molecule, and (iii) the hydrogen corresponding to the single major metabolic site in a molecule.

The plot for all hydrogens shows two large peaks centered around 25 and 50 kcal/mol, respectively. These correspond to the hydrogens attached to the sp3 centers and the sp2 centers, respectively. The lowest energy histogram peaks around 15 kcal/mol with the majority of values between 10 and 20 kcal/mol. The metabolic site energies have significant overlap with the lowest energy peak but are shifted by about 6 kcal/mol. The second peak appears to be shifted by 10 kcal/mol for methyl and methylene moieties that are attached to oxygen and *tert*-butyl methyl group. This shift is evident from the energies of the metabolic sites given in Figure 1. Thus, in summary, the plot of the hydrogen abstraction energies for these substrates shows that hydrogens with abstraction energies lower than 27 kcal/mol are most likely to be susceptible for CYP3A4 mediated metabolism. Since the trend vector predicted hydrogen abstraction energies are quite similar to those computed by AM1 (Figure 4), our analyses from here on will only deal with trend vector predicted energies.

Surface Area Consideration for Internal Sterics. We see from Figure 4 that using the lowest energy criterion alone will not predict the major site of metabolism. A hydrogen atom can be abstracted only if it has sufficient solvent exposure. A plot of the trend vector predicted lowest energy sites (circles) and metabolic site energies (asterisks) vs surface area is shown in Figure 5. The trend vector predicted lowest energy sites agree

Table 1. 20 Highest and Lowest Energy Trend Vector Components

No.	Trend vector component	Energy (kcal/mol)
1	H−CX1sp	18.711490
2	H−CX1sp-CX2sp	18.711490
3	H−CX2sp2a5	16.834734
4	H−CX2sp2a6	12.943820
5	H−CX2sp2a5-CX2sp2a5-OX2sp3a5	11.404792
6	H−OX1sp3-CX3sp2-OX1sp2	9.795876
7	H−NX2sp3-CX3sp2-OX1sp2	9.235794
8	H−CX2sp2a5-CX2sp2a5-CX3sp2a5-OX2sp3a5	8.105881
9	H−CX2sp2a5-NX2sp2a5	7.497335
10	H−CX1sp-CX2sp-CX1sp3	6.944467
11	H−CX2sp2a5-CX2sp2a5-OX2sp3a5-CX2sp2a5	6.624751
12	H−CX1sp-CX2sp-CX2sp3	6.488254
13	H−CX1sp-CX2sp-CX2sp3-CX3sp3	6.488254
14	H−NX1sp3-CX3sp2-OX1sp2	6.407460
15	H−NX2sp3-CX3sp2-NX3sp3	6.172007
16	H−OX1sp3-CX3sp2	6.036877
17	H−CX2sp2-CX3sp2-OX1sp2	5.899733
18	H−CX2sp2a5-CX2sp2a5-NX3sp3a5	5.593336
19	H−CX2sp2a5-CX2sp2a5-NX3sp3a5-CX2sp2a5	5.358986
20	H−OX1sp3-CX3sp3	5.341077
	●	
	●	
	●	
3729	H−OX1sp3-CX3sp2-NX3sp3-OX2sp3	−5.549718
3730	H−OX1sp3-CX3sp2-NX3sp3	−5.549718
3731	H−CX2sp3-NX3sp3	−5.906686
3732	H−CX2sp3-SX2sp3	−5.912649
3733	H−CX2sp2-CX3sp2-CX1sp3	−6.074765
3734	H−CX3sp3-CX2sp2	−6.132203
3735	H−OX1sp3-NX3sp3-CX3sp2-OX1sp2	−6.297911
3736	H−OX1sp3-NX3sp3	−6.297911
3737	H−OX1sp3-NX3sp3-CX3sp2	−6.297911
3738	H−OX1sp3-NX3sp3-CX2sp3	−6.297911
3739	H−CX2sp3-CX2sp2	−6.385551
3740	H−CX2sp3-NX2sp3	−6.431753
3741	H−CX2sp2-CX2sp2-CX2sp2	−6.836477
3742	H−NX2sp3	−7.589312
3743	H−CX3sp3-NX1sp3	−8.821259
3744	H−OX1sp3-CX3sp2a6	−9.552199
3745	H−CX1sp3	−9.752637
3746	H−CX2sp3-NX1sp3	−10.301351
3747	H−CX2sp3	−12.677226
3748	H−CX3sp3	−15.388554

with those calculated with AMPAC in 62% (31/50) of the cases. The asterisks are shifted to the right of the circles, consistent with Figure 4. The points that have overlapping circles and asterisks in Figure 5 correspond to the major metabolic sites with the lowest hydrogen abstraction energies. Thus, about 28% of the major metabolic sites are predicted by the lowest hydrogen abstraction energy. It is evident from Figure 5 that none of the hydrogens with surface area exposure < 8 Å^2 are susceptible to metabolism (no overlap of circles and

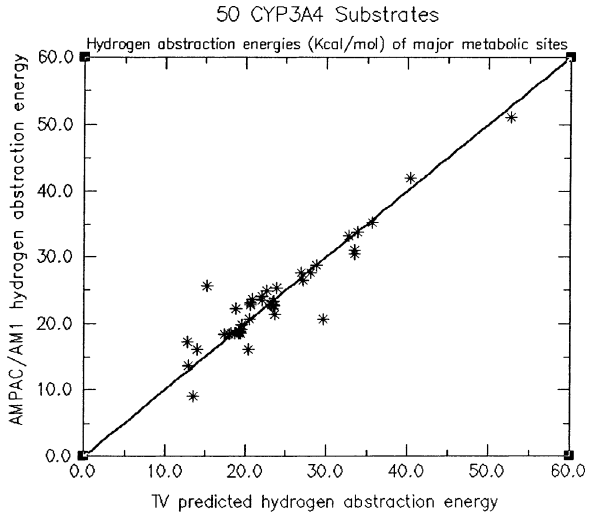


Figure 3. Correlation plot of the trend vector predicted hydrogen abstraction energy versus the AM1 calculated energy.

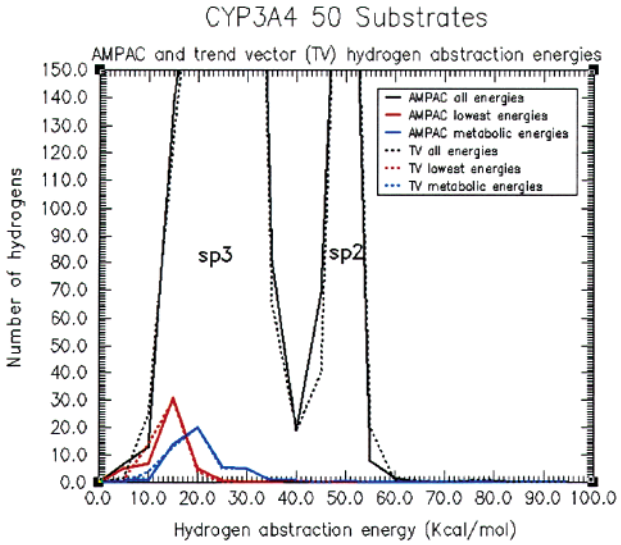


Figure 4. Histogram plot of hydrogen abstraction energies. Two peaks corresponding to all the hydrogens have peak heights 372 and 317, respectively.

asterisks). About 32% of the hydrogens with lowest energies lack sufficient exposure to be abstracted by the enzyme (see Figures 1 and 4). If we consider the lowest energy site with surface exposure ≥ 8 Å^2, the percentage of correct prediction of metabolic sites increases to 42%. Given that there is uncertainty in the prediction of hydrogen abstraction energy, and there may be other relevant factors not taken into account, it probably is reasonable to consider low energy hydrogens other than the very lowest. It is interesting to note that in all the nine cases where *N*-demethylated product is the major metabolite, the adjacent sites at the secondary and tertiary carbons have lower hydrogen abstraction energies. An examination of these near misses could allow us to formulate an empirical rule that CYP3A4 would metabolize the *N*-methyl group more easily than the more favorable adjoining sites.

If we consider all hydrogens with hydrogen abstraction energies ≤ 27 kcal/mol and surface area exposure ≥ 8 Å^2 (362 out of 1523 hydrogens for 50 substrates) these hydrogens include a major metabolic site in the

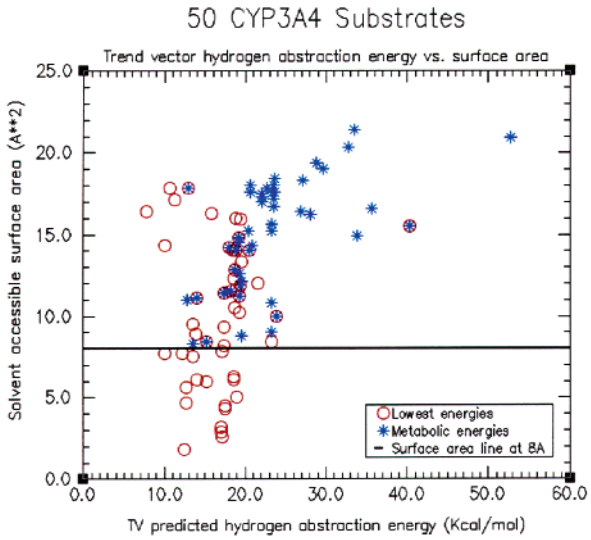


Figure 5. Plot of trend vector predicted hydrogen abstraction energy versus the surface area.

CYP3A4 substrates 78% of the time. So, in the absence of the experimental data, use of this model would require us to examine all hydrogens with abstraction energies ≤ 27 kcal/mol and surface area exposure ≥ 8 Å^2. However, this criterion still will not allow us to predict the major site of metabolism for the following eight cases: docetaxel, etoposide, loratadine, saquinavir, seratrodast, taxol, teniposide, and terfenadine. It is possible more detailed examination of exceptions such these would allow us to identify features in the molecule that prevent the lowest energy site from being a major site of metabolism for CYP3A4. However, the number of such cases is too small at present to make any firm rules.

Visualization of Results. The primary reason the trend vector model was developed was to provide chemists with an easily accessible, intuitive, and interpretable visual representation of CYP3A4 metabolic susceptibility that can be rapidly generated. Such a graphic representation is shown in Figure 6. We found that interpretability was greatly enhanced by having the information about hydrogen abstraction energy and surface area of a hydrogen transferred to the nonhydrogen atom to which it is attached. Hydrogen abstraction energies (now on the nonhydrogen atom) are presented as a color spectrum scaled relative to the range of hydrogen abstraction energies within a given molecule. This is implemented using the embedded ChemDrawPro plug-in with the Internet Explorer browser. The lowest energy sites with hydrogen surface area exposure ≥ 8 Å^2 are shown in bright red with increasing shade of blue for increasing energies. The highest energy site is shown in bright blue. Thus the energy spectrum spans from bright red to purple to bright blue. The atoms that do not have any hydrogen atoms or those atoms whose hydrogens have surface area exposure < 8 Å^2 are shown in black. For topologically equivalent hydrogens, the energy and the surface area of the hydrogen with the highest surface area is given.

In the first three cases presented in Figure 6 the site of metabolism corresponds to the lowest energy site, rendered in brightest red. In the next three cases the

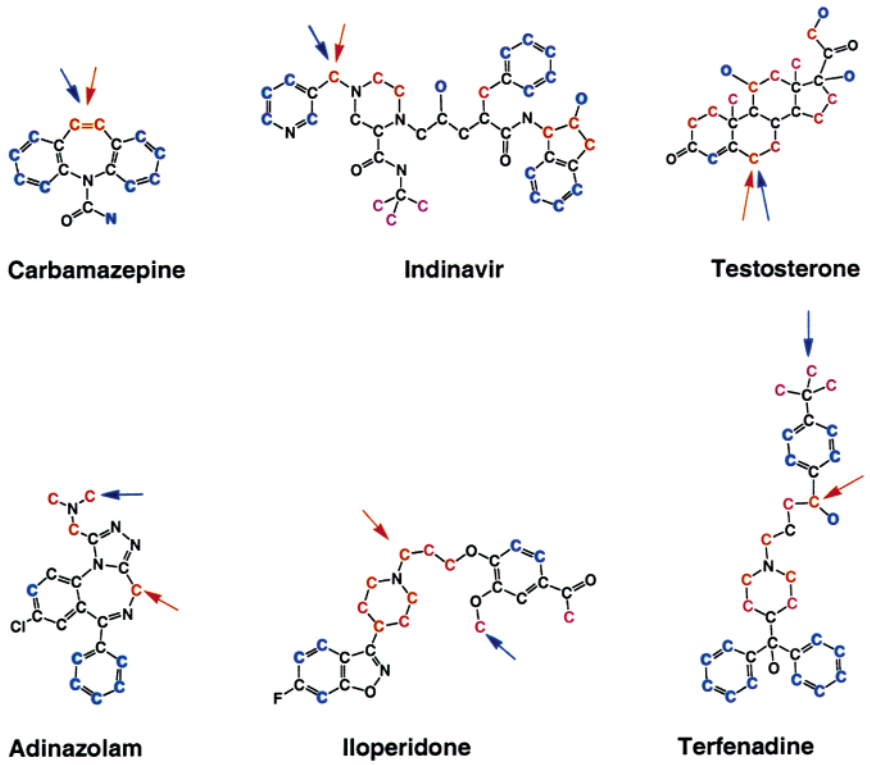


Figure 6. Graphic visualization of trend vector predicted hydrogen abstraction energies. The color spectrum for each molecule is relative to its range of hydrogen abstraction energies. The lowest energy site is colored red and increasing shades of blue is used to show sites with increasing energy. The highest energy site is bright blue. The red arrow corresponds to the lowest energy site and the blue arrow corresponds to the major metabolic site.

lowest energy site does not coincide with the major metabolic site. Without knowledge of specific binding requirements for CYP3A4, we are unable to provide an explanation for why the sites with the lowest energies and sufficient solvent exposure are not major sites of metabolism.

Discussion

We presented here a method that allows a rapid and relatively accurate measurement of hydrogen atom abstraction energy, which is used in assessing the CYP3A4 mediated metabolic susceptibility of any given site in a drug molecule. This model identifies a set of labile hydrogens that can be extracted without regard to orientation of the molecule in the enzyme active site. To our knowledge, this is the first general model that suggests likely sites of metabolism for compounds that could potentially be CYP3A4 substrates.

The model for hydrogen abstraction energy is based on semiempirical molecular orbital calculations. It is feasible to do such calculations at a much higher level than AM1. We chose to stay with this type of calculation for the following reasons: (i) earlier work showed that the hydrogen abstraction energy was predictive of the site of metabolism,[4] (ii) our work seems to confirm the adequacy of these calculations, and (iii) potentially more accurate, but more compute-intensive, calculations were not practical for the number of molecules considered in our investigation.

Note that our model applies only to the problem of deciding where in a molecule CYP3A4 is likely to act. It does not help in deciding whether the molecule will be a substrate; we have found that the distribution of hydrogen abstraction energies for substrates and non-substrates are very similar. Obviously, enzyme binding affinity information will be needed in deciding whether a given molecule is a substrate. Since hydrogens attached to sp2 centers have the highest hydrogen abstraction energies (Figure 4) our model cannot predict aromatic ring hydroxylation by CYP3A4, which is quite rare.

If we consider only those hydrogens with hydrogen abstraction energies $\leq$ 27 kcal/mol and surface area exposure $\geq$ 8 Å^2 we are still are unable to predict the major site of metabolism in an appreciable number of cases. This may be due to a number of reasons: (i) We may not be able to predict hydrogen abstraction energy to a high enough accuracy, (ii) It is possible that the lowest hydrogen abstraction energy may correspond to a secondary site of metabolism rather than the major site, (iii) Our assumption that the iron-oxo species in CYP3A4 can access all nooks and crannies of a molecule just as easily as a solvent molecule is an oversimplification, (iv) The rate-limiting step for metabolism in some cases may not be the hydrogen abstraction as we have assumed here. Clearly, our goal is not to definitively predict the major site of metabolism but to highlight "hot-spots" in a molecule. A chemist may wish to modify these sites synthetically to avoid metabolism.

It is known in the literature that the other isoforms, such as CYP2C9 and CYP2D6, interact with specific functional groups in the substrate (e.g., an acid for CYP2C9 and a basic nitrogen for CYP2D6) and help orient them.[16,17] These binding requirements impose an additional geometric constraint on which hydrogen will be abstracted. Thus, our method could be used to predict a superset of metabolically susceptible sites (excluding

those for aromatic or heteroatom oxidations) that are further constrained by the orientation.

Acknowledgment. The authors would like to thank the valuable discussions with the members of PK/ADME modeling team, the Molecular Systems department, and the Drug Metabolism groups in West Point and Rahway. We also would like to thank the members of the following departments at worldwide Merck Research Laboratories: Molecular Modeling, Basic Chemistry, and Drug Metabolism.

References

(1) Shen, D. D.; Kunze, K. L.; Thummel, K. E. Enzyme-catalyzed processes of first-pass hepatic and intestinal drug extraction. *Adv. Drug Delivery Rev*. **1997**, *27*, 99−127.

(2) Rendic S. Summary of information on human CYP enzymes: human P450 metabolism data. *Drug Metab. Rev*. **2002**, *34*, 83−448.

(3) Rendic S.; Di Carlo FJ. Human cytochrome P450 enzymes: a status report summarizing their reactions, substrates, inducers, and inhibitors. *Drug Metab. Rev*. **1997**, *29*(1−2): 413−580.

(4) 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.

(5) 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.

(6) 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.

(7) Schlichting I.; Berendzen J.; Chu K.; Stock A. M.; Maves S. A.; Benson D. E.; Sweet R. M.; Ringe D.; Petsko G. A.; Sligar S. G. The catalytic pathway of cytochrome p450cam at atomic resolution. *Science* **2000**, *287*, 1615−1622.

(8) (a) Shaik, S.; de Visser, S. P.; Ogliaro, F.; Schwarz, H.; Schröder, D. Two-state reactivity mechanisms of hydroxylation and epoxidation by cytochrome P-450 revealed by theory. *Curr. Opin. Chem. Biol*. **2002**, *6*, 556−567. (b) Lightfoot, T.; Ellis, S. W.; Mahling, J.; Ackland, M. J.; Blaney, F. E.; Bijloo, G. J.; De Groot, M. J.; Vermeulen, N. P.; Blackburn, G. M.; Lennard, M. S.; Tucker, G. T. Regioselective hydroxylation of debrisoquine by cytochrome P4502D6: implications for active site modelling. *Xenobiotica* **2000**, *30*, 219−233.

(9) AMPAC Version 6.0. Semichem 1997−2001.

(10) Ekins, S.; Bravi, G.; Wikel, J. H.; Wrighton, S. A. Three-dimensional-quantitative structure activity relationship analysis of cytochrome P-450 3A4 substrates. *J. Pharm. Exp. Ther*. **1999**, *291*, 424−433.

(11) Shrake, A.; Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol*. **1973**, *79*, 351−371.

(12) Kearsley S. K.; Underwood D. J.; Sheridan, R. P.; Miller, M. D. Flexibases: a way to enhance the use of molecular docking methods. *J. Comput. Aided Mol. Des*. **1994**, *8*, 565−582.

(13) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom pairs as molecular features in structure−activity studies: definition and application. *J. Chem. Inf. Comput. Sci*. **1985**, *25*, 64−73.

(14) Sheridan, R. P.; Nachbar, R. B.; Bush, B. L. Extending the trend vector: the trend matrix and sample-based partial least squares. *J. Comput.-Aided Mol. Des*. **1994**, *8*, 323−340.

(15) Bush, B. L.; Sheridan, R. P. PATTY: A programmable atom typer and language for automatic classification of atoms in molecular databases. *J. Chem. Inf. Comput. Sci*. **1993**, *33*, 756−762.

(16) de Groot, M. J.; Alex, A. A.; Jones, B. C. Development of a combined protein and pharmacophore model for cytochrome P450 2C9. *J. Med. Chem*. **2002**, *45*, 1983−1993.

(17) de Groot, M. J.; Ackland, M. J.; Horne, V. A.; Alex, A. A., Jones, B. C. Novel approach to predicting P450-mediated drug metabolism: development of a combined protein and pharmacophore model for CYP2D6. *J. Med. Chem*. **1999**, *42*, 1515−1524.

JM020400S