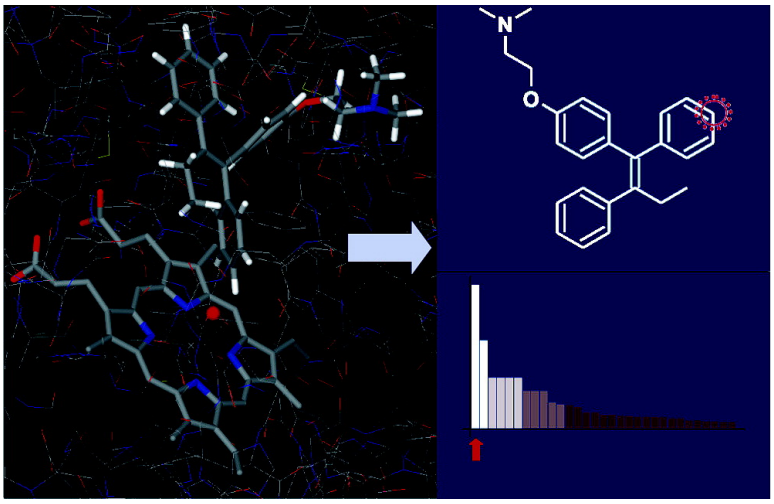
Journal of Medicinal Chemistry

Article

MetaSite: Understanding Metabolism in Human Cytochromes from the Perspective of the Chemist

Gabriele Cruciani, Emanuele Carosati, Benoit De Boeck, Kantharaj Ethirajulu, Claire Mackie, Trevor Howe, and Riccardo Vianello

J. Med. Chem., **2005**, 48 (22), 6970-6979• DOI: 10.1021/jm050529c • Publication Date (Web): 30 September 2005



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 24 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

Journal of Medicinal Chemistry is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

MetaSite: Understanding Metabolism in Human Cytochromes from the Perspective of the Chemist

Gabriele Cruciani,*,† Emanuele Carosati,† Benoit De Boeck,‡ Kantharaj Ethirajulu,§ Claire Mackie,§ Trevor Howe,# and Riccardo Vianello||

Laboratory for Chemometrics and Cheminformatics, Department of Chemistry, University of Perugia, Via Elce di Sotto, 10, I-06123, Perugia, Italy, Departments of Medicinal Chemistry, ADME, and Molecular Informatics, Johnson and Johnson Pharmaceutical Research and Development, A Division of Janssen Pharmaceutica, Turnhoutseweg 30, 2340 Beerse, Belgium, and Molecular Discovery Ltd., 215 Marsh Road, Pinner, Middlesex, HA5 5NE London, U.K.

Received June 6, 2005

Identification of metabolic biotransformations can significantly affect the drug discovery process. Since bioavailability, activity, toxicity, distribution, and final elimination all depend on metabolic biotransformations, it would be extremely advantageous if this information could be produced early in the discovery phase. Once obtained, this information can help chemists to judge whether a potential candidate should be eliminated from the pipeline or modified to improve chemical stability or safety of new compounds. The use of in silico methods to predict the site of metabolism in phase I cytochrome-mediated reactions is a starting point in any metabolic pathway prediction. This paper presents a new method, specifically designed for chemists, that provides the cytochrome involved and the site of metabolism for any human cytochrome P450 (CYP) mediated reaction acting on new substrates. The methodology can be applied automatically to all the cytochromes for which 3D structure is known and can be used by chemists to detect positions that should be protected in order to avoid metabolic degradation or to check the suitability of a new scaffold or prodrug. The fully automated procedure is also a valuable new tool in early ADME-Tox assays (absorption, distribution, metabolism, and excretion toxicity assays), where drug safety and metabolic profile patterns must be evaluated as soon, and as early, as possible.

Introduction

The experimental elucidation of the site of metabolism (i.e., the place in a molecule where the metabolic reaction occurs) is usually a high-resource-demanding task that requires an identifiable isotope in the drug and several experimental techniques and consumes a considerable amount of compound. Nevertheless, the recognition of the site of metabolism could be a significant advance in designing new compounds with a better pharmacokinetic profile.

Labile compounds can be stabilized when the site of metabolism is known by adding stable groups at a metabolically susceptible position. It is sometimes possible to remove, replace, or protect metabolically susceptible groups. Steric shields can be added to prevent a certain cytochrome P450 (CYP) enzyme causing metabolism in a susceptible group. Toxic metabolites in drug candidates can be avoided by chemically protecting the labile moieties. In scaffold hopping and scaffold optimization, the recognition of the site of metabolism is crucial in avoiding chemical modification inducing substrate selectivity toward some human cytochromes. Altering the configuration of an asymmetric center can sometimes make it impossible for a specific human CYP to recognize certain drugs. Finally, knowing the site of metabolism is essential in prodrug design, where the compound needs to be metabolized in order to become active, and also in drugs showing excessively long half-lives.

Knowledge of where functional groups are metabolized can help in designing more stable drugs. Figure 1a reports a very potent and selective h5-HT_{2A} receptor antagonist developed by Rowley et al.[1] The bioavailability of fluoropiperidinphenylindole was 18%, and the terminal half-life was 1.4 h. The poor pharmacokinetic behavior was examined carefully, and a major metabolite, the 6-hydroxyindole, was isolated. By blockage of the major site of metabolism for the compound in Figure 1a, using the 6-fluoro derivative reported in Figure 1b, the pharmacokinetics were dramatically improved when bioavailability was increased to 80% and half-life was increased to 12.0 h. To this end the work of chemists in research laboratories and drug industries may be largely facilitated by computational predictive methods able to identify the potential site of metabolism of given drug candidates as early in the drug discovery process as possible.

Although several papers report the development of computational models to predict the position of metabolism,[2−11] the methods are not designed to fulfill the needs of chemists and are far from being used in a "wet chemistry" laboratory. Moreover, the computational predictions are very poor in most cases, and so far, those

* To whom correspondence should be addressed. Phone: +39-075-5855550. Fax: +39-075-45646. E-mail: gabri@chemiome.chm.unipg.it.
† University of Perugia.
‡ Department of Medicinal Chemistry, Johnson and Johnson Pharmaceutical Research and Development.
§ Department of ADME, Johnson and Johnson Pharmaceutical Research and Development.
# Department of Molecular Informatics, Johnson and Johnson Pharmaceutical Research and Development.
|| Molecular Discovery Ltd.

10.1021/jm050529c CCC: $30.25
Published on Web 09/30/2005

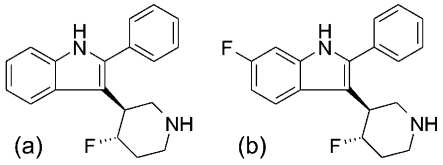


Figure 1. 3-(4-F-Piperidin-3-yl)-2-phenyl-1*H*-indole (a) and the 6-fluoro derivative (b) used by Rowley et al.[1]

methods have not improved the confidence of chemists in their ability to predict the site of metabolism.

The aim of the present paper is to describe a method, evolved from a recently published one,[12,13] that is fast, easy to use, computationally inexpensive, and able to predict human CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4 regioselective metabolism using only the 3D structure of the potential substrates. The method has been specifically designed so that chemists can simply provide the 2D structure of the potential substrate. Then its 3D structure is automatically generated in the CYP cavity, and the method automatically determines the interaction of potential substrates with the CYP enzymes. The site of metabolism probability is then provided in numerical or graphical output formats.

The "State of the Art"

Nowadays different computational approaches are used to predict the position of metabolism,[2−11] which can be grouped into quantitative structure−activity relationship (QSAR) based, pharmacophore-based, structure-based (docking), reactivity-based, and rule-based methods. Since the superfamily of cytochrome P450 enzymes catalyzes a wide variety of oxidative and reductive reactions with an enormous variety of compounds, classical QSAR methods cannot be applied correctly because they require the same mechanism of action and similar molecules (or homologues).

Pharmacophore-based substrate methods[4,5] are training-set-dependent and give a static picture of metabolic recognition and reaction where neither reactivity nor cytochrome active site shape plays any role in the overall process. For example, they state that CYP2D6 binds compounds with a basic nitrogen and/or positive charge and oxidizes atoms at a distance of 5−7 Å from the nitrogen. However, several substrates do exist that have a larger distance between the site of oxidation and the basic nitrogen, e.g., tamoxifen (>10 Å).

The knowledge of 3D structural information concerning important human CYPs (such as 2C9 and 3A4)[14,15] has revitalized attempts to use docking methods to predict the position of metabolism for drug candidates. However, these methods are still affected by imprecise scoring functions and by the great flexibility of cytochrome structures, so they have not yet improved the ability to predict the site of metabolism for xenobiotics.

Reactivity-based ab initio calculations on substrate molecules[8−10] are generally very slow and do not take substrate−enzyme recognition and orientation into account. Figure 2 shows the three most reactive positions in carteolol and ondansetron (CYP2D6 and CYP3A4 substrates, respectively), computed using the ab initio method.[16] Unfortunately, for carteolol the only site of metabolism is at position 8, and the 8-hydroxycarteolol is the only metabolite formed,[17] while for ondansetron

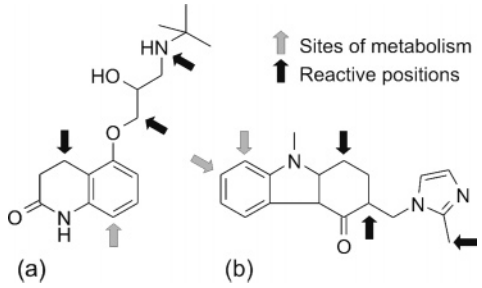


Figure 2. Black arrows indicate the three most reactive positions toward radical abstraction of carteolol (a CYP2D6 substrate on the left) and ondansetron (a CYP3A4 substrate on the right), computed with the ab initio method. Gray arrows indicate the experimental sites of metabolism.

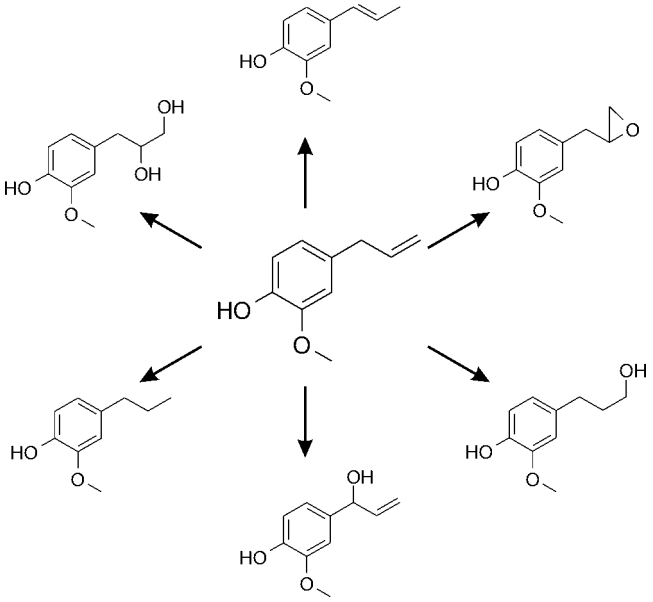


Figure 3. Phase I transformations of eugenol, predicted using a rule-based method. However, experimental findings[20] show that human CYP2D6 mainly catalyzes O-demethylation to produce hydroxychavicol, and this reaction is not included in the transformations above.

the only sites of metabolism are at positions 7 and 8.[18] Chemical reactivity alone, without taking into consideration the orientation of the compound in the reactive active site, is far from able to predict the correct site of metabolism.[19]

Rule-based methods[11] are based on metabolic transformation rules extracted from the literature and stored in a suitable database, assuming metabolic regularities. These methods ignore enzymes and 3D structure of compounds. Rules are assembled with appropriate logic to work in template molecules. As expected, much depends on the type and number of rules, on the training set, on the quality of data from the literature, and on molecular recognition. The methods generally overpredict the metabolic transformations, giving back hundreds of possible metabolites, and often fail to predict some significant pathways. Moreover, they sometimes fail to identify important minor metabolites.

Figure 3 shows the possible phase I transformations of eugenol, predicted using a rule-based method. Each metabolite generated through phase I reactions may undergo subsequent biotransformations and then phase II reactions, thus producing hundreds of possible final metabolites. However, experimental findings often show

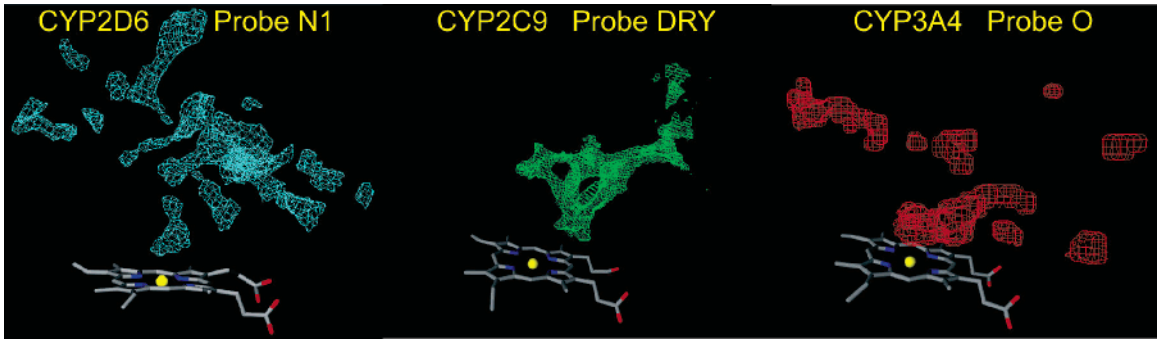


Figure 4. Some of the molecular interaction fields obtained from the various cytochromes studied are compared. On the left, CYP2D6 shows the highest hydrogen-bond acceptor region volume (about 50% of the cavity volume), while CYP2C9 is the cytochrome showing the largest hydrophobic regions (about 20% of its cavity is hydrophobic). On the right, CYP3A4 shows the highest hydrogen-bond donor region volume (about 25% of its cavity volume).

that only one path is populated or just a few of them. So the question is how the right ones are found.

Results

The Chemist Needs a New Approach. If a drug is to be orally active, it should be both chemically and metabolically stable. Metabolism normally only takes place at a specific position of a molecular skeleton, and unfortunately, metabolic regularities are exceptions. Experienced chemists also find it very difficult to predict where metabolism occurs in a molecule.

Researchers have recently focused on developing faster robotic systems and more sensitive analytical metabolite identification tools.[21−25] However, such techniques are usually particularly resource-demanding tasks, consuming a considerable amount of compound, and cannot be used before synthesis. Moreover, owing to the increasing abundance of potential candidates, experimental metabolite identification remains a huge challenge.

Chemists need a simple method able to answer three main questions: Which CYP isoform(s) is (are) involved in metabolic degradation(s)? Where is (are) the most likely position(s) of metabolism in a molecular skeleton? And finally, what can be done to prevent metabolism? From that, chemists possess the skills to synthesize and make a variety of modifications for the purpose of preventing metabolism. Mooers' law states that an information system will not tend to be used whenever it is more painful and troublesome for a customer to have information than for him not to have it.[26] Therefore, we developed an easy to use and fast procedure. This procedure requires a few seconds per molecule, is automatically performed when a molecule or a set of molecules are provided by the user in SMILES notation, 2D SDF format, or 3D coordinates, and is specifically designed to answer the questions of chemists stated above.

Although very simple to use, the method consists of complex algorithms, which are only briefly reported here but explained in depth in the Experimental Section. The proposed methodology, called MetaSite (site of metabolism prediction), involves the calculation of two sets of descriptors, one for the CYP enzyme and one for the potential substrate, respectively representing the chemical fingerprint of the enzyme and the substrate. The set of descriptors used to characterize the CYP enzyme is based on GRID *flexible* molecular interaction fields (GRID-MIFs).[27−29] Flexible molecular interaction fields, reported in Figure 4, are independent of the initial side chain position of the cytochrome 3D structure and better suited to simulate the adaptation of the enzyme to the substrate structure.

The descriptors developed to characterize the substrate chemotypes are obtained from GRID probe pharmacophore recognition. All the substrate atoms are classified into GRID probe categories depending on their hydrophobic, hydrogen-bond donor, hydrogen-bond acceptor, or charge capabilities. Their distances in space are then binned and transformed into clustered distances. One set of descriptors is computed for each atom type category (hydrophobic, hydrogen-bond acceptor, hydrogen-bond donor, and charged), which yields a fingerprint for each atom category in the molecule.

The two sets of descriptors are then used to compare the fingerprint of the cytochrome with the fingerprint of the substrate (see Figure 5). The comparison yields the accessibility of each molecular group toward the reactive heme. The accessibility, called E_i, represents the recognition between the specific CYP protein and the ligand when the ligand is positioned in the CYP protein and exposes the atom i toward the heme.

The substrate reactivity is computed from a mixture of molecular orbital calculations and fragment recognition (see Experimental Section for details). When R_i is the reactivity of atom i in the appropriate reaction mechanism, it represents the activation energy required to produce the reactive intermediate. The reactivity of atom i depends on the ligand 3D structure and on the mechanism of reaction. Therefore, R_i represents a score proportional to the reactivity of the ligand atom i in a specific reaction mechanism.

Once the accessibility and reactivity components are calculated, the site of metabolism can be described by a probability function P_{SM} (probability for the site of metabolism),

$$P_{SM},(i) = E_i R_i \qquad (1)$$

which is correlated to, and can be considered to be, an approximation of the free energy of the overall process.[30]

For the same ligand and the same cytochrome, the P_{SM} function assumes different values for different ligand atoms according to the E_i and R_i components. When a ligand atom i is well exposed to the reaction center of the heme (E_i has a high score) but its reactivity

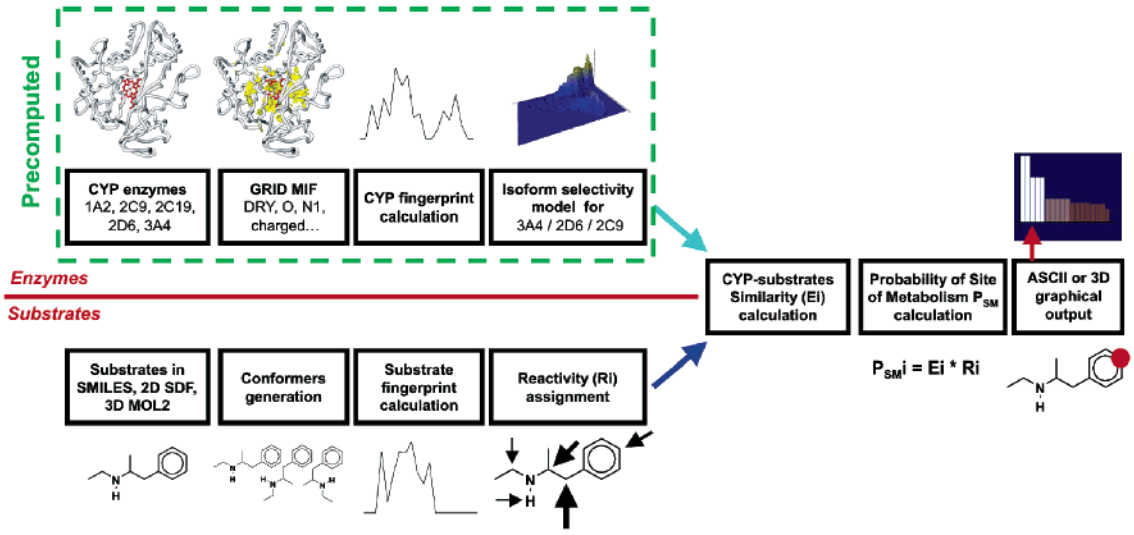


Figure 5. Flowchart of MetaSite procedure. The isoform selectivity models and the GRID-based representations for the main human cytochrome enzymes are precomputed and stored. However, any cytochrome structure can be imported, with MIF computed and then stored. The ligand pharmacophoric recognition, descriptor handling, and similarity computations are performed automatically once the structure(s) of the compound(s) has been provided. The calculation for radical abstraction energy is only performed when molecular fragments are not present in the MetaSite database. The user only needs to introduce the structure of the ligand in SMILES, 2D SDF, or 3D mol2 file. All remaining processes are automated.

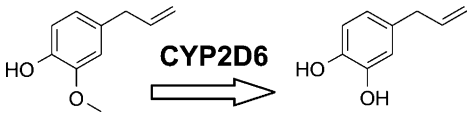


Figure 6. CYP2D6 catalyzed O-demethylation of eugenol to hydroxychavicol.

is very low (R_i reports a very low score), the probability of metabolism in atom i will be very low or zero. Similarly, when a ligand atom i is very reactive in the considered mechanism (R_i reports high score) but atom i is not exposed to the reaction center of the heme (E_i has a very low score), the probability of metabolism in atom i will be close to zero. Therefore, to be the site of metabolism, an atom i should possess significant accessibility and reactivity components related to the heme.

Applications. Figure 3 reports the possible phase I transformations of eugenol predicted by a rule-based method. The molecular structure is relatively simple, but the number of possible metabolites generated through phase I reactions is relatively high. However, experimental findings[20] show that human CYP2D6 catalyzes mainly O-demethylation to produce hydroxychavicol (see Figure 6), and this reaction is not included in any of the transformations in Figure 3. Conversely, MetaSite, in agreement with experimental findings, predicts the path leading to hydroxychavicol as being the most probable in human 2D6 cytochrome.

Rule-based methods are not human-specific, so they produce a relatively large number of possible phase I reactions. On the other hand, being based on human cytochrome structures, MetaSite dramatically reduces the number of reactions that may undergo subsequent phase II biotransformations and, hence, the number of possible final metabolites.

In the discovery setting, metabolite identification of a compound is usually performed using LC−MS−MS.[22,31] However, this is often limited to giving only a crude indication of which part of a molecule is subject to oxidation, reduction, or hydrolysis by the cytochrome

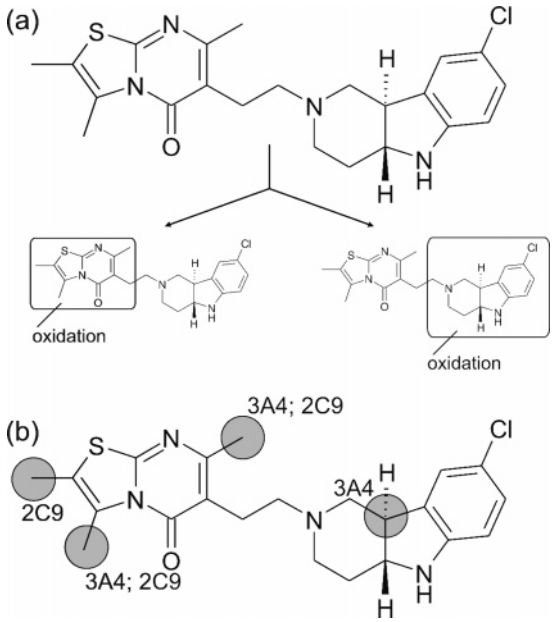


Figure 7. (a) LC−MS−MS gives only a hint as to which positions are metabolically oxidized. (b) MetaSite predictions for the CYP3A4 and 2C9 (the most probable sites are indicated for each isozyme involved). The combination of LC−MS−MS and MetaSite predictions enables chemists to limit the number of possible sites that must be modified in order to achieve metabolic stability for this compound.

P450 isozymes. In this context, MetaSite can provide a refinement of the results obtained by the LC−MS−MS experiments. As shown in Figure 7a, the metabolic pathways identified by LC−MS−MS for a thiazolopyrimidin-5-one derivative are only very crude, with the metabolism-prone sites located on either one or the other half of the molecule without any clues of the specific positions that are affected by oxidation.[32] The MetaSite isoform selectivity model (see below) indicates mainly CYP3A4 and with minor extent CYP2C9 as the major isoforms involved in the oxidation, and Figure 7b shows the most probable metabolic positions suggested by

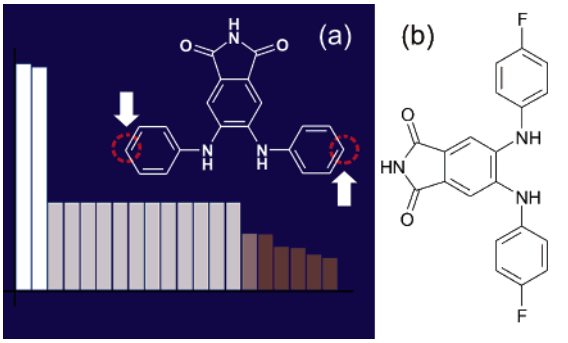


Figure 8. (a) The predicted sites of oxidation are ranked according to probability values and reported in the histogram. White bars highlight the higher probability values that correspond to the para position of the molecule (indicated by white arrows in the 2D structure). (b) Compound CGP53353, a metabolically stable EGF-receptor kinase inhibitor.

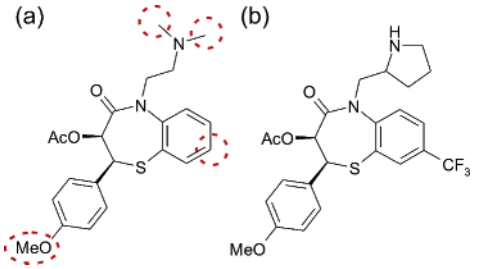


Figure 9. (a) MetaSite prediction of labile sites for (+)-*cis*-diltiazem, compared with a derivative (b) designed to increase its metabolic stability.

MetaSite for each of the cytochromes. As shown in Figure 7b, combining the information provided by LC−MS−MS and MetaSite contributes to identification of the sites where molecular metabolism could occur. This can then guide the chemists in choosing which derivatives to synthesize in order to improve the metabolic stability of the compound.

Figure 8a shows a dianilinophthalimide compound, a potent and selective inhibitor of epidermal growth factor (EGF) receptor kinase.[33] It is well absorbed orally, but it is also rapidly metabolized in man. Drug metabolism studies were carried out to discover the site of metabolism in man. Para-hydroxylation on phenylamino moieties was then followed by glucuronylation and excretion. The MetaSite procedure indicates CYP3A4 as the major isoform involved in the oxidation, and the predicted sites of oxidation are reported in the histogram in Figure 8a. The higher white bars correspond to the para position in the molecule (as highlighted by white arrows in the 2D structure). To prevent drug metabolism, fluorine substituents were placed at the para positions to act as metabolic blockers. The resulting fluoro derivative, also referred to as CGP53353 (Figure 8b), had similar potency but was metabolically stable, in agreement with MetaSite findings.

Diltiazem (see Figure 9a) is extensively metabolized along distinct pathways, and so its duration of action is relatively short (the plasma elimination half-life of diltiazem is approximately 3.0−4.0 h). CYP3A4 is responsible for the major route represented by N-demethylation. In the search for more stable compounds, the benzothiazepinone ring was modified with the addition of the trifluoromethyl group.[34] However, the two classes of compounds showed similar metabolic

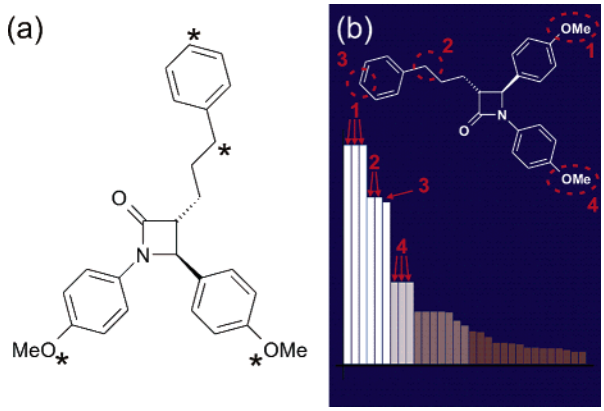


Figure 10. Prediction of the site of metabolism for a cholesterol absorption inhibitor. (a) The asterisks report experimental positions of metabolism. (b) MetaSite ranking of the probability of metabolism for all of the different molecular positions; the first four predicted positions (ranked by probability value) are highlighted.

routes. The structure was then further modified by replacing the dimethylethylamine group; the resulting new derivative was much more stable than (+)-*cis*-diltiazem (Figure 9b). The improved stability can be explained by the presence of the N-1 pyrrolidinyl derivative, designed to decrease radical stability and to increase steric hindrance at the amino position. To our knowledge, no direct effect on metabolic pathways of the trifluoromethyl group was reported.[35]

Interestingly, the MetaSite prediction shows that the trifluoromethyl is blocking aromatic oxidation and thus the conversions to glucuronide and sulfate conjugates that follow.

An inhibitor of intestinal cholesterol absorption (3*R*)-(3-phenylpropyl)-1,(4*S*)-bis(4-methoxyphenyl)-2-azetidinone (Figure 10a) has been demonstrated to lower total plasma cholesterol in man. The potential sites of metabolism in this compound were studied, resulting in a complex metabolite mixture.[36] Further studies confirmed that the mixture was composed of at least four different metabolites, obtained from two different demethylation reactions, plus one benzylic oxidation and one phenyl oxidation (see asterisks in Figure 10a); all these are due to CYP3A4. Figure 10b reports the probability values for the site of metabolism for all the atoms of the molecule obtained by using MetaSite and the 3A4 CYP isoform. The first four ranked positions in Figure 10b correspond to the circled positions and exactly match the experimental sites of metabolism reported by asterisks in Figure 10a. The complex metabolic profile was predicted well, thus showing the great potential impact of this procedure.

Finally, MetaSite can also provide the molecular contributions to the exposure of the reactive atom toward the heme. By alteration of these molecular moieties, the metabolic pattern can be modified. For example, bunitrolol is hydroxylated by cytochrome 2D6 at the aromatic ring,[37] as reported by the gray circle in Figure 11a. It is logical to assume that bunitrolol exposes the aromatic carbon atom in position 4 to the CYP2D6 heme group. Starting from this assumption, MetaSite orients the bunitrolol in the CYP cavity and computes the bunitrolol−CYP hydrophobic complementarity and the complementarity of charges and H-bonds

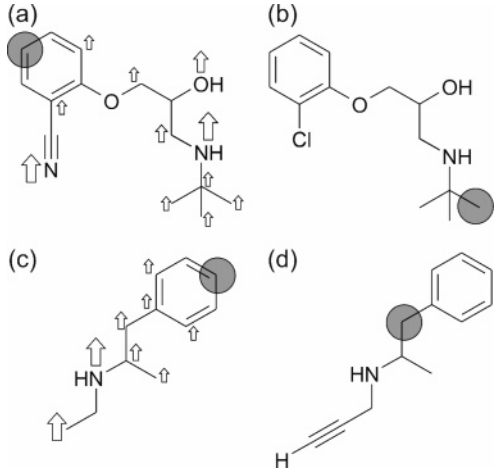


Figure 11. Experimental site of metabolism for two similar compounds (bunitrolol and desmethylbupranolol, (a) and (b), respectively), and the score contribution (white arrows in (a)) due to the substrate−CYP cavity complementarity as calculated by MetaSite. NEA site of metabolism and score contribution (c), compared to NPyA derivative (d), which shows a new experimental site of metabolism due to the modification of the starting NEA substrate molecule.

between bunitrolol and CYP2D6. Such complementarities are then used to assign a contribution score to the different atoms in the substrate. Owing to the computation mechanism, the score is proportional to the contribution made by the molecular moieties to the exposure of the experimental reactive atom toward the heme. Thus, the contribution scores reported in Figure 11a highlight the molecular group that influences the 4-hydroxylation reaction most. Chemical modifications of such molecular groups may induce a different site of metabolism. For the case under consideration, desmethylbupranolol (see Figure 11b) can be considered to be a derivative of bunitrolol. As predicted by the computation, by replacement of the cyano group in bunitrolol with a chlorine atom in desmethylbupranolol, the metabolic reaction changes from 4-hydroxylation to *N*-methyl oxidation.

N-Ethylamphetamine (NEA) is experimentally[38] metabolized by cytochrome 2D6 at the position reported in Figure 11c. Figure 11c also reports the molecular moieties that, by attractive fit with the CYP enzyme cavity, contribute most to the exposure of the aromatic carbon toward hydroxylation. Figure 11d reports an NEA derivative, the *N*-3-propynylamphetamine (NPyA). According to the MetaSite recognition hypothesis, changing some of the atoms that largely contribute to the atom exposure results in a modification of the site of metabolism. Experimentally, NEA is hydroxylated in the aromatic ring while NPyA is only subjected to a dealkylation reaction (as reported by the gray circles).

Discussion

Figures 6−11 illustrate the two driving forces operating on test ligand molecules, substrates of human CYP enzymes. Calculations show that for all the atoms of the test molecules the probability of being the site of metabolism depends on the components of accessibility and reactivity. The accessibility component, called E_i, represents the recognition between the specific CYP protein and the ligand when the ligand is positioned in the CYP protein and exposes the atom i toward the heme. It depends on the ligand 3D structure, conformation, and chirality and on the 3D structure and side chain flexibility of the CYP enzyme. Thus, the E_i score is proportional to the exposure of the ligand atom i toward the heme group of a specific CYP enzyme. In contrast, R_i is the reactivity of atom i in the appropriate reaction mechanism and represents the activation energy required to produce the reactive intermediate. It depends on the ligand 3D structure and on the mechanism of reaction. Therefore, the score R_i is proportional to the reactivity of the ligand atom i in a specific reaction mechanism.

Table 1. Results Obtained with MetaSite on Carefully Selected Data from the Literature, on Compounds Showing Different Metabolic Pathways and Considerable Structural Diversity, Flexibility, and Lipophilicity

human CYP	no. of substrates	E_i/R_i [a]	static MIFs,[b] %	flexible MIF,[b] %
2C9	152	80/20	68	86
2C19	125	80/20		81
2D6	200	80/20	62	86
3A4	340	65/35	64	78
1A2	135	70/30		75

[a] Contribution of recognition (E_i) and reactivity (R_i) is reported. [b] Predictions obtained within the first two selected atoms are compared for static and flexible GRID MIFs.

The methodology was validated on carefully selected data from the literature: 150 metabolic reactions catalyzed by CYP1A2, 160 by CYP2C9, 140 by CYP2C19, 200 metabolic reactions catalyzed by CYP2D6, and 350 metabolic reactions catalyzed by CYP3A4. These data were used together with information concerning their sites of metabolism. Table 1 summarizes the results obtained after the tests. It is important to note that the compounds used showed various metabolic pathways. Some are metabolized at one single site only; others present two sites of metabolism, with only a few presenting three. Moreover, substrates show a large structural diversity, including both rigid compounds (e.g., steroids) and very flexible ones with more than 10 rotatable bonds, not to mention a wide range of molecular weight and lipophilicity.

In more than 70% of CYP2C9 reactions, the first atom selected by the methodology matches the experimental reactive one. Moreover, in more than 16% of cases, the second atom fits the experimental one. Therefore, Table 1 shows that in considering the overall ranking list for the single and multiple sites of metabolism, the methodology predicts the site of metabolism for CYP2C9 within the first two atoms selected in approximately 86% of the reactions, independent of the conformer used (see Table 1, last column). Similar results were obtained with the other human cytochromes.

Comparison with experimental data demonstrated that the recognition component (or exposition E_i) is by far the most important component in eq 1. E_i alone accounts for more than 80% of the experimental result, leaving the reactivity component with only limited impact. However, the situation is a little different for CYP3A4. Its broad substrate specificity, probably due to its large cavity, suggests that its substrates may adopt more than one orientation in the active site. Consequently, it has been reported (but not proved) that CYP3A4 attack on ligand positions is mainly deter-

Table 2. Percentage of Correct Predictions Obtained from an Investigation Carried Out by Various Pharmaceutical Companies Using MetaSite and Internal Proprietary Compounds for Which Information Concerning Their Sites of Metabolism in Human CYPs Is Known

pharmaceutical company	CYP	N [a]	CP,[b] %
Sanofi-Aventis, Bridgewater, NJ[c]	2C9	90	84
Pfizer, Sandwich, U.K.[d]	2D6	14	85
Pfizer, Sandwich, U.K.[d]	3A4	55	86
Johnson & Johnson PRD, Beerse, Belgium[e]	2C9, 2D6, 3A4	50	85
AstraZeneca, Molndal, Sweden[f]	3A4	13	85

[a] N = number of substrates. [b] CP = percent correct predictions. [c] Drug Design Group.[40] [d] PDM Group.[41] [e] ADME and Med Chem Group.[42]. [f] DMPK & Biochem Group.[43]

Table 3. Percentage of Correct Predictions Made by MetaSite Using CYP2C9, CYP2D6, and CYP3A4 Isoform Selectivity To Recognize Each Substrate[a]

substrate selectivity	N [b]	CP,[c] %
substrates for 2C9	48	83
substrates for 2D6	30	62
substrates for 3A4	140	90
substrates for 2C9 and 2D6	3	67
substrates for 2C9 and 3A4	30	86
substrates for 2D6 and 3A4	100	83

[a] Experimental data from Johnson & Johnson PRD.[42] [b] N = number of substrates. [c] CP = percent correct predictions.

mined by the chemical reactivity of these ligand positions. However, the results presented here suggest that although the reactivity component relative to CYP3A4 is the largest when compared with the other cythocromes, the recognition component is still the site-determining step of the reaction.[39]

Table 1 also shows that when static molecular interaction fields are used, the predictive ability of the method decreases by 15−20%. The explanation put forward is that the cytochromes may use side chain flexibility to allocate more space when necessary or to orient the substrate closer to the reactive heme. It seems that the prediction of side chain movement due to side chain flexibility or substrate binding is essential in order to compile the CYP−substrate interactions correctly.

The methodology was also tested by various pharmaceutical companies on proprietary molecules, data for which information concerning their sites of metabolism in human CYPs is known.[40−43]

The quantitative results reported in Table 2 show trends similar to those reported in Table 1 for the compounds selected from the literature. This demonstrates that the method is generally applicable because it is not biased by training or by local models. It thus demonstrates general predictivity, a fundamental requirement any method must possess in order to be applied to the metabolism arena.

Table 3 shows the results of MetaSite calculations to differentiate between human P450 substrates and CYP2C9, CYP2D6, and CYP3A4 isoform selectivity. More than 300 substrates, about which experimental data from a common source are known, were computationally analyzed. On average, more than 80% of substrates were well predicted by the procedure.

Although the current software was prepared to work with the five most important human cytochromes, this procedure was designed to work with any cytochrome structure, and so it can be applied to humans, fish, plant, and bacteria cytochromes. There are more than 120 P450 families and more than 1000 P450 enzymes. In theory all these structures can be imported, processed, and used for the prediction of the site of metabolism. The procedure is totally automatic, does not require any user assistance, and only requires the 3D structure of the enzyme to be available. Once the 3D structure of the compound has been provided, the semiempirical calculations of charges and radical abstraction energy assignment, pharmacophoric recognition, descriptor handling, and similarity computation are all carried out automatically.

Conclusion

We have developed a methodology to predict the site of metabolism, the isoform selectivity, and the ligand−cytochrome complementarity for substrates of the most important human cytochromes. On average, in about 85% of the cases the method predicted the correct site of metabolism within the first two choices in the ranking list, and in 80% of the cases the method predicted the correct isoform(s) involved in the metabolic path. The method also seems to be able to predict the regions in a molecule that contribute most to the exposition of a certain chemical group toward the reactive heme.

The methodology works for the most important human cytochromes but can be automatically applied to all the cytochromes for which 3D structures are known. It can be used by chemists to suggest positions that should be protected in order to avoid metabolic degradation or to check the suitability of substituent insertion to modify the site of metabolism or to transform a molecule into a prodrug.

It is important to stress that the method highlighted here does not require training or docking procedures and associated scoring functions and does not require 2D or 3D QSAR models. The methodology does not use any training set or supervised or unsupervised technique. In contrast, the method relies on flexible molecular interaction fields generated by the GRID force field on the CYP homology modeling structures that were treated and filtered in order to extract the most relevant information.

The methodology is easy to use and fast. The method only requires a few seconds per molecule to predict a site of metabolism for druglike substrates. It is noteworthy that the method has proven to be predictive for the very different validation sets examined by various pharmaceutical companies.

The 3D structure of the substrate to be analyzed (the starting conformation) has an impact on the outcome of the method. Satisfactory results were obtained using the in-house conformer generation, included in MetaSite, which depends on the MIFs and the flexible shape of the active site of the enzymes. The latter procedure is automatically performed when a molecule or a set of molecules are provided by the user in SMILES, 2D SDF, or 3D coordinates.

Owing to the increasing abundance of potential candidates at the time of writing, experimental metabolite identification remains a huge challenge. We have clearly demonstrated that the use of this new in silico approach to predict a hypothetical metabolite structure

can speed the process of metabolite identification by focusing experimental work on specific target structures, thus improving the method of metabolite structure confirmation and elucidation.

The fully automated computational procedure is also a valuable new tool in early ADME-Tox assays (absorption, distribution, metabolism, and excretion toxicity assays), where drug safety and metabolic profile patterns must be evaluated in order to enhance and streamline the process of developing new drug candidates.

However, we clarify once more that the methodology was not intended for and cannot be used to predict the relative rate of metabolism of a compound across the CYP isoforms.

Experimental Section

Molecular Interaction Field Generation and Transformation in CYP Fingerprint. GRID force field[27−29] was used to produce molecular interaction fields inside the cytochrome active site. The program GRID is calibrated in a water environment to obtain chemically specific information about a (macro)molecule (in this case the human cytochrome). An electrostatic potential does not normally allow favorable binding sites to be differentiated for a primary, secondary, or tertiary amine cation, for pyridinium, or for a sodium cation, and the GRID method is an attempt to compute analogous potentials that do have some chemical specificity. The object used to measure the potential at each point is given the generic name "Probe". Many different Probes can be used on the same macromolecule one after another, and each represents a specific chemical group. A great deal of chemically specific information can therefore be accumulated concerning the way in which the macromolecule might interact favorably with other ligand molecules.

The molecular interaction fields (MIF) in the binding sites of the cytochromes were obtained using the flexible mode in GRID.[29] With the flexible option, some of the amino acid side chains can automatically move in response to attractive or repulsive interactions with the chemical probe. The side chain flexibility in GRID can mimic the amino acid movements that occur in the CYP active site to accommodate different substrates according to their sizes, shapes, and interaction patterns.

As shown by Zamora et al., the MIFs obtained from cytochrome enzymes are subsequently transformed and simplified.[12] In a cytochrome, where a catalytic reaction must take place, all of the 3D map information can be compressed and referenced to the catalytic center of the enzyme, that is, the oxene atom of the protoporphyrin group.

The selected 3D interaction points are used to calculate enzyme fingerprints using the GRIND technology.[44] For each CYP−Probe interaction map, this approach transforms the interaction energies at a certain spatial position (the MIF descriptors) into a number of histograms that capture the 3D pharmacophoric interactions of the flexible protein. Such histograms are called correlograms. The correlograms represent the distance between the reactive center of the cytochrome (the oxene in the heme moiety) and the different chemical regions inside the enzyme active site.

3D Structure of Substrates and Fingerprint Generation. The majority of CYP substrates contain flexible moieties. Since conformation is relevant to recognition of the substrate and binding to the CYP and has a noticeable impact on the outcome of the method, each substrate is modeled by using a population of diverse low-energy minimum conformations, obtained by means of in-house software integrated in the computational procedure. The 3D structures of obtained conformers were induced by the interaction fields and shape of the CYP active site.

The descriptors developed to characterize the substrate chemotypes are obtained from a mixture of molecular orbital calculations and GRID probe−pharmacophore recognition. Molecular orbital calculations to compute the substrate's electron density distribution are the first to be performed. All atom charges are determined using the AM1 Hamiltonian. Then the computed charges are used to derive a 3D pharmacophore based on the molecular electrostatic potential (MEP) around the substrate molecules.

Moreover, all the substrate atoms are classified into GRID probe categories depending on their hydrophobic, hydrogen-bond donor, or hydrogen-bond acceptor capabilities. Their distances in space are then binned and transformed into clustered distances. One set of descriptors is computed for each atom type category (hydrophobic, hydrogen-bond donor, hydrogen-bond acceptor, negatively and positively charged), thus yielding a fingerprint for each atom category in the molecule. The distances between the different atomic positions classified using the previous criteria are then transformed into binned distances. In this case, the distances between the different atoms are calculated and a value of 1 or 0 is assigned to each bin distance, respectively indicating the presence or absence of such a distance in the substrate.

Substrate-CYP Enzyme Comparison: Recognition Component. Once the protein interaction pattern is translated from Cartesian coordinates into distances from the reactive center of the enzyme, and the structure of the ligand has been described with similar fingerprints, both sets of descriptors can be compared.[12,13] The hydrophobic complementarity, the complementarity of charges and hydrogen bonds for the protein and the substrates are all computed using Carbó similarity indices.[45] The prediction of the site of metabolism is based on the hypothesis that the distance between the reactive center on the protein (oxene atom in protoporphyrin group) and the interaction points in the protein cavity (GRID-MIF) should correlate with the distance between the reactive center of the molecule (i.e., positions of hydrogen atoms and heteroatoms) and the position of the different atom types in the molecule.[46−48]

Finally, each atom in each substrate is assigned a similarity score. Owing to the mechanism of computation, the score is proportional to the exposure of such substrate atoms toward the reactive heme and represents the accessibility component.

The accessibility component, called E_i, represents the recognition between the specific CYP protein and the ligand when the ligand is positioned in the CYP protein and exposes the atom i to the heme. It depends on the ligand 3D structure, conformation, and chirality and on the 3D structure and side chain flexibility of the CYP enzyme. Thus, the E_i score is proportional to the exposure of the ligand atom i to the heme group of a specific CYP enzyme.

Reactivity Component. Cytochromes P450 catalyze oxidative and reductive reactions. Oxidative biotransformations are more frequent and include aromatic and side chain hydroxylation, N-, O-, S-dealkylation, N-oxidation, sulfoxidation, N-hydroxylation, deamination, dehalogenation, and desulfuration. The majority of these reactions require the formation of radical species, which is usually the rate-determining step for the reactivity process.[49]

When R_i is the reactivity of atom i in the appropriate reaction mechanism, it represents the activation energy required to produce the reactive intermediate. It depends on the ligand 3D structure and on the mechanism of reaction. Therefore, R_i is a score proportional to the reactivity of the ligand atom i in a specific reaction mechanism.

Furthermore, in this reaction mechanism, R_i does not depend on the P450 enzyme but is only related to the molecular topology and 3D structure. There is only limited experimental data available that report the R_i component for druglike compounds. However, the quantification of the R_i component can be approximated using ab initio methods. The problem is that on-line calculations using ab initio methods take too long to be of any practical use. Therefore, we have developed a faster procedure that has three steps. The first step involved collecting the large majority of druglike substrates for human cytochromes and dissecting them in non-

redundant chemical fragments, which involved the selection of hundreds of fragments. In the second step ab initio calculations[16] simulating hydrogen abstraction processes were carried out in all the fragments. Although this process is long and time-consuming, once completed, it does not need to be repeated. Fragment atomic positions were classified according to their liability as stable, nonreactive, medium, moderate, or very reactive and were then ranked in a quantitative reactivity scale ranging from 0.0 (stable) to 1.5 (strongly reactive). In the third step, a software routine was produced that recognizes the constitutive fragments of the fragmentized substrate when a potential cytochrome substrate is given. After recognition, the reactivity component R_i can be assigned to the atomic positions. In the case of recognition failure due to a lack of fragment, an AM1 calculation simulating hydrogen abstraction is carried out. However, to save computational time, radical formation energy is computed only at atomic positions with the relevant E_i component.

Substrate Selectivity Prediction. Figure 4 reports the molecular interaction fields obtained from the various cytochromes studied. CYP2D6 shows the highest H-bond acceptor region volume (about 50% of its cavity volume). Although the movements of the flexible side chains are considered, graphical inspection shows that the H-bond acceptor cyan regions are clustered in particular 3D regions of the enzyme. From the spatial locations obtained, a pharmacophore pattern for the CYP2D6 enzyme can be created. The procedure is then repeated using the maps derived from the hydrophobic probe, H-bond donor, and charged probes. The global pharmacophore pattern produced is then compared with the pharmacophore of a substrate, left conformationally free to fit the enzyme pharmacophore positions. A global similarity between the two matching pharmacophores is then computed. If the procedure is repeated for all CYP enzymes, each CYP similarity score may then be used to rank the substrate selectivity for human P450 enzymes (see Supporting Information for more information).

Human Cytochrome Structures. The major xenobiotic-metabolizing cytochromes P450 in humans belong to families 1, 2, and 3 and include CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4. The crystal structure of human 2C9 and 3A4 cytochromes were recently resolved and deposited in the Brookhaven Protein Data Bank.[14,15] These structures were also submitted to GRID computation in order to produce molecular interaction fields. The human structure for P450 2C9 was used as a template in homology modeling of the CYP2C19 enzyme. In fact, this enzyme has a high degree of similarity and identity with CYP2C9.

The initial 3D structures of the CYP2D6 and CYP1A2 enzymes were kindly provided by DeRienzo et al.,[50] with 3D models being built using restraint-based comparative modeling of the X-ray crystallographic structures of bacterial cytochromes P450 BM3, CAM, TERP, and ERYF, all used as templates (PDB entries 2bmh, 3cpp, 1cpt, and 1oxa). After this, secondary structure predictions were obtained using the method of Rost and Sander.[51]

The heme molecule, with the iron in its ferric oxidation state, was extracted from the structure of CYPBM3 and fitted into the active site of each of the two cytochromes. Last, dynamic runs were carried out on the starting structures, without any ligands, to select an average bioconformation for all the isoenzymes.

All P450 cytochromes contain a protoporphyrin group with a central iron atom that is normally hexacoordinated in ferric form. The substrates bind reversibly to the enzyme, and the complex undergoes reduction to the ferrous state. This allows molecular oxygen to bind as a third partner. Molecular oxygen is transformed into oxene, an electrophilic and reactive species, which normally pulls a hydrogen radical away from the substrate and transfers a formal hydroxyl group back.[52] After release of the product, the regenerated cytochrome P450 is ready for a new cycle.

Software Package. The procedure is called MetaSite (site of metabolism prediction).[30] The MetaSite procedure is fully automated and does not require any user assistance. All the work can be handled and submitted in batch queue. The molecular interaction field for CYPs obtained from the GRID package are precomputed and stored inside the software. Once the structures of the compounds are provided, the semiempirical calculations, pharmacophoric recognition, descriptor handling, similarity computation, and reactivity computation are all carried out automatically. The complete calculation is performed in a few seconds on IRIX SGI machines and is even faster in the Linux or Windows environment. For example, processing a database of 100 compounds, starting from 3D molecular structures, takes about 3 min at full resolution using an R14000 Silicon Graphics 500 MHz CPU, less than a minute on a Windows Pentium machine, and about 30 s using a Linux Pentium machine. Starting from SMILES notation, processing a database of 100 compounds (each in 20 conformations) takes about 6 min at full resolution with a R14000 Silicon Graphics 500 MHz CPU, 3 min using a Windows Pentium IV machine, and about 1 min on a Linux Pentium machine.

The MetaSite software is available to nonprofit organizations free of charge and can be downloaded from www.moldiscovery.com.

Acknowledgment. The authors thank I. Zamora for the inspiration, M. Engels (J&J), L. Afzelius (AZ), G. Cianchetta, R. Vaz (Sanofi-Aventis), and M. Howard (Pfizer) for the tests made with the software, Y. Aristei and P. Benedetti for the compilation of the figures and ab initio calculations, and S. Clementi for the critical revision of the manuscript.

Supporting Information Available: Structures of compounds presented in all the examples in the SMILES format and additional details on the calculations of E_i and R_i components and substrate selectivity model. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Rowley, M.; Hallett, D. J.; Goodacre, S.; Moyes, C.; Crawforth J.; Sparey, T. J.; Patel, S.; Marwood, R.; Patel, S.; Thomas, S.; Hitzel, L.; O'Connor, D.; Szeto, N.; Castro, J. L.; Hutson, P. H.; MacLeod, A. M. 3-(4-Fluoropiperidin-3-yl)-2-phenylindoles as high affinity, selective, and orally bioavailable h5-HT_{2A} receptor antagonists. *J. Med. Chem.* **2001**, *44*, 1603−1614.

(2) Korzekwa, K. R.; Grogan, J.; DeVito, S.; Jones, J. P. Electronic models for cytochrome P450 oxidations. *Adv. Exp. Med. Biol.* **1996**, *38*, 361−369.

(3) Lewis, D. F.; Dickins, M.; Eddershaw, P. J.; Tarbit, M. H.; Goldfarb, P. S. Cytochrome P450 substrate specificities, substrate structural templates and enzyme active site geometries. *Drug Metab. Drug Interact.* **1999**, *15*, 1−49.

(4) De Groot, M. J.; Ackland, M.; Horne, V.; Alexander, A.; Barry, J. A novel approach to predicting P450 mediated drug metabolism. CYP2D6 catalyzed N-dealkylation reactions and qualitative metabolite predictions using a combined protein and pharmacophore model for CYP2D6. *J. Med. Chem.* **1999**, *42*, 4062−4070.

(5) Jones, B. C.; Hawksworth, G.; Horne, V. A.; Newlands, A.; Morsman, J.; Tute, M. S.; Smith, D. A. Putative active site template model for cytochrome P4502C9 (tolbutamide hydroxylase). *Drug Metab. Dispos.* **1996**, *24*, 260−266.

(6) Mancy, A.; Broto, P.; Dijols, S.; Dansette, P. M.; Mansuy, D. The Substrate binding site of human liver cytochrome P450 2C9: An approach using designed tienilic acid derivatives and molecular modeling. *Biochemistry* **1995**, *34*, 10365−10375.

(7) Riddestrom, M.; Zamora, I.; Fjäström, O.; Andersson, T. B. Analysis of selective regions in the active sites of human cytochromes P450, 2C8, 2C9, 2C18, and 2C19 homology models using GRID/CPCA. *J. Med. Chem.* **2001**, *44*, 4072−4081.

(8) Singh, S. B.; Shen, L. Q.; Walker, M. J.; Sheridan, R. A model for predicting likely sites of CYP3A4-mediated metabolism on drug-like molecules. *J. Med. Chem.* **2003**, *46*, 1330−1336.

(9) Chen, H.; de Groot, M.; Vermulen, N.; Hanzlik, R. P**.** Oxidative N-dealkylation of *p*-cyclopropyl-*N,N*-dimethylaniline. A substituent effect on a radical-clock reaction rationalized by ab initio calculations on radical cation intermediates. *J. Org. Chem.* **1997**, *62*, 8227−8230.

(10) 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.

(11) Darvas, F.; Marokhazi, S.; Kormos, P.; Kulkarmi, G.; Kalasz, H.; Papp, A. MetabolExpert: Its Use in Metabolism Research and in Combinatorial Chemistry. In *Drug Metabolism*; Erhardt, P. W., Ed.; Blackwell Science: Oxford, 1999; pp 237−270.
(12) Zamora, I.; Afzelius, L.; Cruciani, G. Predicting drug metabolism: A site of metabolism prediction tool applied to the cytochrome P450 2C9. *J. Med. Chem.* **2003**, *46*, 2313−2324.
(13) Berellini, G.; Cruciani, G.; Mannhold, R. Pharmacophore, Drug metabolism, and pharmacokinetics models on non-peptide AT_1, AT_2, and AT_1/AT_2 angiotensin II receptor antagonists. *J. Med. Chem.* **2005**, *48*, 4389−4399.
(14) Wester, M. R.; Yano, J. K.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. http://www.pdb.org, 2004, 1R9O entry.
(15) Yano, J. K.; Wester, M. R.; Schoch, G. A., Griffin, K. J.; Stout, C. D.; Johnson, E. F. http://www.pdb.org, 2004, 1TQN entry.
(16) Open-shell radicals were optimized at the AM1 semiempirical level. Single-point energy evaluations were performed by DFT at the B3LYP/6-311G** level of theory because correlation between experimental and calculated radical stabilities resulted in reasonable agreement for this level of theory.
(17) Kudo, S.; Uchida, M.; Odomi, M. Metabolism of carteolol by cDNA-expressed human cytochrome P450. *Eur. J. Clin. Pharmacol.* **1997**, *52*, 479−485.
(18) Pritchard, J. F. Ondansetron metabolism and pharmacokinetics. *Semin. Oncol.* **1992**, *19*, 9−15.
(19) 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.
(20) Katsuhisa, S.; Yuji, I.; Shinji, O.; Yusuke, H.; Shosuke, K. Copper-mediated oxidative DNA damage induced by eugenol: possible involvement of O-demethylation. *Mutat. Res.* **2004**, *565*, 35−44.
(21) Thomas, S. R.; Gerhard, U. Open-access high-resolution mass spectrometry in early drug discovery. *J. Mass Spectrom.* **2004**, *39*, 942−948.
(22) Kantharaj, E.; Tuytelaars, A.; Proost, P.; Ongel, Z.; Assouw, H. P.; Gilissen, R. A. Simultaneous measurement of drug metabolic stability and identification of metabolities using ion-trap mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2661−2668.
(23) Kostiainen, R.; Kotiano, T.; Kuuranne, T.; Auriola, S. Liquid chromatography−atmospheric pressure ionization/mass spectrometry (LC−API/MS) in Drug Metabolism Studies. *J. Mass Spectrom.* **2003**, *38*, 357−372.
(24) Corcoran, O.; Spraul, M. LC−NMR−MS in drug discovery. *Drug Discovery Today* **2003**, *8*, 624−631.
(25) Nassar, A. E. F.; Talaat, R. E. Strategies for dealing with metabolite elucidation in drug discovery and development. *Drug Discovery Today* **2004**, *9*, 317−327.
(26) Mooers, C. N. Why some retrieval systems are used and others are not. *Bull. Am. Soc. Inf. Sci.* **1996**, *23*, 22−23.
(27) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.
(28) Carosati, E.; Sciabola, S.; Cruciani, G. Hydrogen bonding interactions of covalently bonded fluorine atoms: From crystallographic data to a new angular function in the GRID force field. *J. Med. Chem.* **2004**, *47*, 5114−5125.
(29) Goodford, P. J. Atom Movements during Drug−Receptor Interactions. In *Rational Molecular Design in Drug* Research; Liljefors, T., Jorgensen, F. S., Krogsgaard-Larsen, P., Eds.; Alfred Benzon Symposium 42; Munkgaard: Copenhagen, 1998; pp 215−230.
(30) *MetaSite*, version 2.5; Molecular Discovery Ltd., 2004 (http://www.moldiscovery.com).
(31) Cai, Z.; Sinhababu, A. K.; Harrelson, S. Simultaneous quantitative cassette analysis of drugs and detection of their metabolites by high performance liquid chromatography/ion trap mass spectrometry. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1637−1643.
(32) Kantharaj, E.; Ehmer, P. B.; Tuytelaars, A.; Van Vlaslaer, A.; Mackie, C.; Gilissen, R. A. H. J. Simultaneous measurements of metabolic stability and metabolite identification of 7-methoxymethylthiazolo[3,2-*a*]pyrimidin-5-one derivatives in human liver microsomes using liquid chromatography/ion-trap mass spectrometry. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1069−1074.
(33) Trinks, U.; Buchdunger, E.; Furet, P.; Kump, W.; Mett, H.; Meyer, T.; Muller, M.; Regenass, U.; Rihs, G.; Lydon, N. Dianilinophthalimides: Potent and selective, ATP-competitive inhibitors of the EGF-receptor protein tyrosine kinase. *J. Med. Chem.* **1994**, *37*, 1015−1027.
(34) Floyd, D. M.; Kimball, S. D.; Krapcho, J.; Das, J.; Turk, C. F.; Moquin, R. V.; Lago, M. W.; Duff, K. J.; Lee, V. G.; White, R. E.; Ridgewell, R. E.; Moreland, S.; Brittain, R. J.; Normandin, D. E.; Hedberg, S. A.; Cucinotta, G. G. Benzazepinone calcium channel blockers. 2. Structure activity and drug metabolism studies leading to potent antihypertensive agents. Comparison with benzothiazepinones. *J. Med. Chem.* **1992**, *35*, 756−772.
(35) Smith, D. A.; Waterbeemd, H.; Walker, D. K. In *Pharmacokinetic and Metabolism in Drug Design*; Mannhold, R., Kubinyi, H., Timmerman, H., Eds.; Wiley-VCH: Weinheim, Germany, 2001; Vol. 13, p 84.
(36) Rosemblum, S. B.; Huynh, T.; Afonso, A.; Davis, H. R.; Yumibe, N.; Clader, J. W.; Burnett, D. Discovery of 1-(4-fluorophenyl)-(3*R*)-[3-(4-fluorophenyl)-(3*S*)-hydroxypropyl]-(4*S*)-(4-hydroxyphenyl)-2-azetidinone (SCH 58235): A designed, potent, orally active inhibitor of cholesterol absorption. *J. Med. Chem.* **1998**, *41*, 973−980.
(37) Narimatsu, S.; Masubuchi, Y.; Hosokawa, S.; Ohmori, S.; Kitada, M.; Suzuki, T. Involvement of a cytochrome P4502D subfamily in human liver microsomal bunitrolol 4-hydroxylation. *Biol. Pharm. Bull.* **1994**, *17*, 803−807.
(38) Bach, M. V.; Coutts, R. T.; Baker, G. B. Involvement of CYP2D6 in the in vitro metabolism of amphetamine, two *N*-alkylamphetamines and their 4-methoxylated derivatives. *Xenobiotica* **1999**, *29*, 719−732.
(39) For 600 substrates of 2C9, 2C19, and 2D6 the average relative importance of the recognition component and reactivity components was 80% and 20%, respectively. For 350 substrates of 3A4 the average relative importance of the recognition component and reactivity components was 65% and 35%, while for 150 substrates of 1A2, it was 70% and 30%, respectively.
(40) Cianchetta, G.; Vaz, R. Drug Design I Group, Sanofi-Aventis, Bridgewater, NJ. Personal communication, 2004.
(41) Howard, M. PDM Group, Pfizer, Sandwich, U.K. Personal communication, 2005. Cinato, F.; Cocchiara, G. Nerviano Medical Science, Milano, Italy. Personal communication, 2005.
(42) Ethirajulu, K.; De Boeck, B. Departments of ADME and Medicinal Chemistry, Johnson and Johnson PRD, Beerse, Belgium. Personal communication, 2004.
(43) Afzelius, L. DMPK & Biochem Group, AstraZeneca, Molndal, Sweden. Personal communication, 2005.
(44) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. GRid-INdependent descriptors (GRIND): A novel class of alignment-independent three-dimensional molecular descriptors. *J. Med. Chem.* **2000**, *43*, 3233−3243.
(45) Amat, L.; Carbó-Dorca, R. Fitted electronic density functions from H to Rn for use in quantum similarity measures: *cis*-Diamminedichloroplatinum(II) complex as an application example. *J. Comput. Chem.* **1999**, *20*, 911−920.
(46) Afzelius, L.; Zamora, I.; Ridderström, M.; Kalén, A.; Andersson, T. B.; Masimirembwa, C. Competitive CYP2C9 inhibitors: Enzyme Inhibition studies, protein homology modeling, and three-dimensional quantitative structure−activity relationship analysis. *Mol. Pharmacol.* **2001**, *59*, 909−919.
(47) Jones, J. P.; He, M.; Trager, W. F.; Rettie, A. E. Three-dimensional quantitative structure−activity relationship for inhibitors of cytochrome P4502C9. *Drug Metab. Dispos.* **1996**, *24*, 1−6.
(48) Ekins, S.; Bravi, G.; Binkley, S.; Gillespie, J. S.; Ring, B.; Wikel, J.; Wrighton, S. Three- and four-dimensional-quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2C9 inhibitors. *Drug Metab. Dispos.* **2000**, *28*, 994−1002.
(49) Guengerich, F. P. Cytochrome P450 oxidations in the generation of reactive electrophiles: epoxidation and related reactions. *Arch. Biochem. Biophys.* **2003**, *409*, 59−71.
(50) De Rienzo, F.; Fanelli, F.; Menziani, M. C.; De Benedetti, P. G. Theoretical investigation of substrate specificity for cytochromes P450 IA2, P450 IID6 and P450 IIIA4. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 93−116.
(51) Rost, B.; Sander, C. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* **1994**, *19*, 55−72.
(52) Magdalou, J.; Fournel-Gigleux, S.; Testa, B.; Ouzzine, M. Biotransformation Reactions. In *The Practice of Medicinal Chemistry*, 2nd ed.; Wermuth, C. G., Ed.; Academic Press: Amsterdam, 2003; pp 517−544.

JM050529C