DOI: 10.1002/cmdc.200800384

CypScore: Quantitative Prediction of Reactivity toward Cytochromes P450 Based on Semiempirical Molecular Orbital Theory

Matthias Hennemann,[a] Arno Friedl,[b] Mario Lobell,[c] Jörg Keldenich,[d] Alexander Hillisch,[c] Timothy Clark,*[a] and Andreas H. Göller*[c]

CypScore is an in silico approach for predicting the likely sites of cytochrome P450-mediated metabolism of druglike organic molecules. It consists of multiple models for the most important P450 oxidation reactions such as aliphatic hydroxylation, N-dealkylation, O-dealkylation, aromatic hydroxylation, double-bond oxidation, N-oxidation, and S-oxidation. Each of these models is based on atomic reactivity descriptors derived from surface-based properties calculated with ParaSurf™ and based on AM1 semiempirical molecular orbital theory. The models were trained with data derived from Bayer Schering Pharma's in-house MajorMetabolite Database with more than 2300 transformations and more than 800 molecules collected from the primary literature. The models have been balanced to allow the treatment of relative intramolecular, intra-chemotype, and inter-chemotype reactivities of the labile sites toward oxidation. The models were evaluated with promising hit rates on three public datasets of varying quality in the annotation of the experimental positions. For 39 well-characterized compounds from 14 in-house lead optimization programs, we could detect at least one major metabolite for the three highest-ranked positions in 87% of the compounds and overall more than 62% of all major metabolites, with promising true- to false-positive ratios of 0.9.

Introduction

Recent decades have seen a shift in the drug-discovery process from a more or less one-dimensional sequential optimization strategy with the main focus on affinity to a parallel multi-dimensional optimization strategy that also takes ADMET (absorption, distribution, metabolism, excretion, toxicity) properties into account. The advent of powerful in vitro tests has facilitated this approach, and so has the introduction and iterative optimization of in silico prediction tools based on data generated with these assays. The combination of in silico and in vitro results will clearly have synergistic benefits relative to any method alone. Optimally, only experimental data should be considered for final decision making. On the other hand, these data are often not available because of limitations in resources. As an alternative, in silico tools allow the assessment of large numbers of molecules, for example, profiling large hit clusters, selecting subsets for in vitro and in vivo testing, gaining maximum SAR/SPR (structure–activity relationship/structure–property relationship) information, and profiling virtual compounds.

The "M" in ADMET stands for metabolism, a difficult aspect to predict because it involves properties of reactivity rather than equilibrium. Phase I metabolism makes molecules more polar, and thus soluble, mainly by oxidation reactions catalyzed by cytochrome P450 enzymes.[1] These oxidations are often followed by phase II conjugation reactions. These metabolic processes are usually mediated by the liver and decrease drug concentrations in the body, but can also form reactive or toxic metabolites or P450 inhibitors and inducers. The products of these reactions must therefore be known precisely before a compound can enter development.[2] The situation is complicated by factors such as species, age, race, and genetic background, which can influence rates, regioselectivity, inhibition, and induction. Thus, metabolic results are often not transferable from in vitro to in vivo experiments.

Experimental techniques for predicting metabolic clearance range from relatively fast and inexpensive assays to extremely complex and expensive procedures. The in vitro assessment of metabolic stability by incubation of liver cells or microsomes is a broadly applied technique. Intrinsic bioavailability is assessed from the amount of compound remaining after a certain

[a] Dr. M. Hennemann, Prof. Dr. T. Clark
Computer-Chemie-Centrum and
Interdisciplinary Center for Molecular Materials
Friedrich-Alexander-Universität Erlangen-Nürnberg
Nägelsbachstraße 25, 91052 Erlangen (Germany)
Fax: (+49) 9131-8526565
E-mail: tim.clark@chemie.uni-erlangen.de

[b] Dr. A. Friedl
Pharma BSP GD-GRA-INT International Coord.&Supp.
Bayer Healthcare AG, Aprather Weg 18a, 42096 Wuppertal (Germany)

[c] Dr. M. Lobell, Dr. A. Hillisch, Dr. A. H. Göller
Pharma Chemical Research BSP GDD-LGO-MCW-MC VI
Bayer Healthcare AG, Aprather Weg 18a, 42096 Wuppertal (Germany)
Fax: (+49) 202-365461
E-mail: andreas.goeller@bayerhealthcare.de

[d] Dr. J. Keldenich
Pharma BSP GDD-GED-DMPK-RP-RPW, Research Pharmacokinetics
Bayer Healthcare AG, Aprather Weg 18a, 42096 Wuppertal (Germany)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cmdc.200800384.

period of time. The outcome is a classification into high, medium, or low clearance classes. Metabolites can be identified from the so-called Metabol-ID assay,[3] in which the reaction solution containing the metabolites after in vitro incubation of liver microsomes or hepatocytes is usually analyzed by LC–MS–MS. After metabolite fragmentation in the first mass spectrometer, the second determines the fragment masses, which are then used to propose likely metabolites. Radioligand experiments are analogous to Metabol-ID, but more complex and allow the precise identification of metabolized positions. Finally, the same procedure can be applied in vivo, with analysis of blood, urine, or bile.

In lead generation and optimization, only the first two assays are usually applicable, with Metabol-ID being a scarce resource. These assays are often unable to provide medicinal chemists with the precise and rapid answers they need for early-phase optimization cycles. Therefore, a fast and reliable in silico method is needed that 1) identifies the major metabolites in project compounds, 2) allows synthesis proposals and virtual compounds to be ranked, 3) allows parallel in silico affinity and metabolic stability optimization, and 4) ranks the lability of alternative metabolism sites, rather than just identifying the first weak position.

The potential importance of computational methods for the prediction of cytochrome P450-related pharmacological effects was recognized early, so that the earliest techniques were developed almost as soon as the roles of the cytochromes P450 were understood.[4] A summary of QSAR-, pharmacophore-, and knowledge-based approaches can be found in a recent review by Crivori and Poggesi.[5]

Some approaches for the prediction of regioselectivity (i.e., the substrate positions attacked by metabolizing enzymes) use quantum-mechanical calculations. Korzekwa and co-workers[6] first described AM1 semiempirical molecular orbital (MO) calculations of hydrogen abstraction energies (HAE). This technique was later used in a modified form as a trend-vector model (a QSPR for AM1-derived HAEs) by Singh et al.[7] METASITE takes a mixed approach of reactivity prediction through quantum-mechanical HAEs and cavity fitting via GRID representations.[8] Approaches were recently introduced that encode reaction patterns and the corresponding probabilities of metabolic reactions as found in databases into rule-based models.[9] Similarly, Sheridan et al. described atom-neighborhood encoding fingerprints for an atom-centered reactivity QSAR model.[10] All these methods provide a complete description of the encoded (i.e., observed) metabolic events, but they can only predict weakly reactive sites on the basis of inferences drawn indirectly from the training data. None of these methods are likely to be able to make qualitative predictions of rank orders of reactivity within one molecule or between different compounds.

Early on, various ab initio techniques[11] were used in conjunction with experimental work to study the reaction mechanisms involved in substrate oxidation by cytochromes P450 in detail. These calculations, although too slow and cost-intensive to be used routinely, give very important hints for mechanistic understanding and act as benchmarks for fast models. In this case, they allow us to rationalize the meanings of the descriptors used in our models.

The CypScore approach described below applies to a hypothetical "P450 super-enzyme": a nonspecific oxidation enzyme, which might be compared with the outcome of a microsomal Metabol-ID assay. This is further justified in that the main enzyme responsible for metabolic clearance is often the quite unspecific 3A4 isoform, followed by 2C9 and 2D6, all three of which often act together. Furthermore, once a model for a "super-enzyme" has been constructed, other QSAR- or pharmacophore-based methods can be used to predict isoform specificity.[5]

The model was designed with the goal of providing medicinal chemists with a tool to identify the exact atomic positions of cytochrome P450 oxidative attack on small drug molecules. The chemists would then modify the compounds to decrease oxidative reactivity. It uses as much experimental information on microsomal, hepatocyte, and hepatic clearance as possible. The model described herein is entirely ligand based, and does not use any cavity information.

A major feature of the CypScore system is that competing metabolic reactions are united in one reactivity scale, so that various metabolic sites and various molecules can be compared directly. We show that although CypScore ignores the specific contribution of a given protein cavity on the orientation of the substrate, the model is highly predictive. It is now a routine tool in drug discovery at Bayer Schering Pharma.

Results and Discussion

Datasets

An in-house database (MajorMetaboliteDB, MMDB) was created that contains the training data for the CypScore models. The database currently consists of 844 parent compounds with 2336 metabolic transformations, which were carefully selected from the primary literature[12] and the Compendium of Biotransformations.[13] Metabolic transformations were only included if their major metabolites were precisely defined.[14] The database contains the ligands labeled by reactive position, metabolic reaction, data source, isoform, species, percentage metabolite (if applicable), and the type of experiment (in vivo, in vitro, ex vivo). The database covers all important reaction types and isoforms in a spectrum similar to that observed for metabolites in in-house projects. The main reaction types are listed in Table 1. These reactions are consistent with our findings in in-house projects.

The distribution of metabolites from various isoforms is 59.5% from 3A4/3A5, 18.5% from 2D6, 17.2% from 2C9, 16.9% from 1A2, 14.3% from 2C19, 7.8% from 2E1, and 28.3% from other isoforms, and reflects the relative importance of the various isoforms. The species and experimental distributions reflect our interests for pharmacologically relevant species, with 44% human data (504 in vivo, 518 in vitro), 38.9% rat, and 7.3% mouse.

We extracted subsets of the MMDB for each reaction class for which a model was to be created. The metabolic outcomes of these subsets are considered to be independent of species

Table 1. Dataset compositions for the MMDB, the public test and validation sets, and the in-house validation set. The numbers of compounds, numbers of all and reactive heavy atoms, and the numbers of the different metabolic reactions covered are given (percent values in parentheses).

	MMDB	Set 0 Public test set	Set 1 Public validation set	Set 2 Public validation set	Set 3 In-house validation set
Number of compounds	844	55	226	25	39
Number of heavy atoms	56 195	1401	5525	527	1325
Number of metabolic sites	2336	108	378	41	85
Percentage metabolic sites	4.2%	7.7%	6.8%	7.8%	6.4%
Metabolic sites/compound	2.76[a]	1.96	1.67[b]	1.64	2.3
Aliphatic hydroxylation	396[a]	25 (23)	102 (27)	7 (28)	26 (31)
N-dealkylation	300[a]	34 (31)	105 (28)	8 (32)	15 (18)
O-dealkylation	122[a]	13 (12)	55 (15)	6 (24)	19 (22)
Aromatic hydroxylation	286[a]	23 (21)	80 (21)	12 (48)	20 (24)
Double-bond oxidation	43[a]	2 (2)	5 (1)	1 (4)	2 (2)
N-oxidation	64[a]	10 (9)	6 (2)	2 (8)	0 (0)
S-oxidation	57[a]	1 (1)	18 (5)	0 (0)	1 (1)
Others	1068	0	8(2)	2 (8)	2 (2)

[a] Numbers are for the 1268 cytochrome P450-mediated reactions. [b] Ratio not corrected for symmetry-equivalent aliphatic and equivalent aromatic hydroxylation positions.

and isoform. This is not strictly true, but represents a reasonable a priori assumption, because we need datasets of sufficient size for meaningful models. Each non-hydrogen atom in these molecules is labeled as reactive (=1) or unreactive (=0). A binary descriptor was chosen, as the extent of reactivity cannot be captured from published data.

Set 0: public test set

The public test set (Set 0) consists of 55 molecules selected from the publications of Singh et al.[7] and Lewis,[15] shown in Supporting Information scheme S1 and table S1. The dataset is also provided as SDF format structures with annotated reactive positions. It contains 108 reactive atoms within a total of 1401 heavy atoms (reactivity labels are always assigned to the appropriate heavy atom), which results in a mean of 1.96 and a range from 1 to 5 reactive positions per molecule (see Table 1). The reactive positions in each molecule were corrected for equivalence by symmetry. For instance, two equivalent methyl groups in an N-dealkylation reaction were counted as one reactive position and labeled accordingly in Supporting Information scheme S1. Similarly, aromatic hydroxylation reactions that can occur at two equivalent aromatic carbon atoms are defined as one reactive position (for details, see the description of the model for aromatic hydroxylation).

Despite the information provided by the original publications of Singh et al.[7] and Lewis,[15] most of the compounds are not exclusive substrates of specific isoforms. The dramatic breakdown in predictivity of the Lewis model if all non-CYP 3A4 reactions are taken into account has been described by Sheridan et al.[10]

The reactions observed include 25 aliphatic hydroxylations, 34 N-dealkylations, 13 O-dealkylations, 23 aromatic hydroxylations, two double-bond oxidations, ten N-oxidations, and one S-oxidation, which is a fair reflection of the relative importance of these reactions in our in-house projects. Most of the compounds in this set occurred in the training set for at least one reactivity model. We use them as a semidependent test set for two reasons: First, because each "raw model" was created for exactly one specific reaction and centered and scaled afterward, a good prediction for any test-set molecule will only be achievable by using the entire palette of CypScore models. Second, our knowledge of the metabolic fate of these compounds is much more complete than for the two published validation sets described below.

Set 1: public validation set 1

The training-set compounds of Sheridan et al.[10] with 521 isoform-specific observations (305 compounds metabolized by CYP 3A4, 92 by CYP 2C9, and 124 by CYP 2D6) were transformed into 404 unique compounds for prediction with the isoform-unspecific CypScore program. We removed 168 compounds that are present in our MMDB and were therefore part of model training sets. In this step, we became aware of two Sheridan-set compounds with the wrong chemical structures and one with the wrong stereochemistry, which suggests that there may be similar problems for other compounds that we did not detect. We next removed all molecules outside an "extended druglike" molecular weight range of 200–600 Da. The remaining 226 compounds constitute the public validation set 1 (Set 1) with experimental information and structures as provided by the original authors. We expect this dataset (derived from a commercial database) to be incomplete with respect to metabolic positions, as the ratio between reactive positions and molecules is quite low, at 1.67, relative to the significantly higher ratios for Sets 0 and 3. Moreover, our MMDB often reports additional metabolites for the 168 compounds that were removed. The dataset compounds are provided as structures in SDF format with annotated reactive positions and citations from the original publications in the Supporting Information. The dataset is described in Table 1.

Set 2: public validation set 2

Public validation set 2 (Set 2) consists of the 25 validation set compounds from the work of Sheridan et al.,[10] condensed from the 40 compounds from the three isoform-specific datasets provided. None of these compounds are contained in our MMDB.

Set 2 consists of 41 symmetry-corrected reactive heavy atoms out of a total of 527. The mean value of 1.64 reactive positions per molecule is again lower than for Sets 0 and 3. The problem with both external validation sets is the quality and completeness of the experimental data provided. As discussed below, a brief comparison of the primary publications with the datasets reveals that they are incomplete with respect to the isoforms involved and the reactive positions, and some reactive positions are ambiguous. We tried to correct these ambiguities as much as possible for Set 2 using the original publications. The dataset compounds are provided as structures in SDF format with annotated reactive positions and the original literature citations in the Supporting Information.

Set 3: Bayer Schering Pharma validation set

Finally, we carefully selected a further validation set (Set 3) from a larger set of in-house Bayer Schering Pharma compounds. The selection was based on the availability of the experimental Metabol-ID and clearance experiments. We consider this dataset to be the only real proof of the CypScore concept because it is the one with the best-defined reactivity data. However, we can only provide statistics and general performance data but no chemical structures because all compounds stem from current projects.

The Bayer Schering Pharma validation set consists of 39 compounds from 14 in-house projects for various medicinal indications and with non-overlapping structural classes. The numbers of major metabolites in these compounds range from one to five, with a mean value of 2.26. The profile of P450-mediated phase I transformations is similar to that of the public validation set and the MMDB set, as shown in Table 1, again with aliphatic hydroxylations including N- and O-dealkylations as the most important, followed by aromatic oxidations. In one project, S-dealkylation occurs as one major pathway; we do not have an S-dealkylation model because we lack training data.

Model application

Once the individual CypScore models were available, a complete CypScore application was built to enable automatic use and validation. This application is a fully automated workflow that starts from 2D SD files,[16] which are transformed into 3D coordinates by Corina.[17] The resulting coordinates are used for subsequent AM1[18] optimization, which leads to a ParaSurf™[19] calculation. The descriptors calculated by VAMP and ParaSurf™ are then used in the appropriate CypScore model for each atom. Automatic atom-type recognition decides which model is appropriate. An example CypScore output for diazepam is shown in Figure 1.

In the following sections, we report the results for the various validation sets described above. We discuss the applicability of the model and the specific features of the data for this application that warrant special consideration.

Model validation and tests

A complex "atom event model" like CypScore must be validated stringently simply because of the number of individual regression models which exacerbates the danger of over-fitting. We therefore created the validation sets described above to provide as complete a test of the method as possible. To assess how well the model performs for compounds that are metabolized by several competing oxidation reactions, the dataset must contain a representative mixture of compounds that are metabolized at multiple sites via multiple reactions

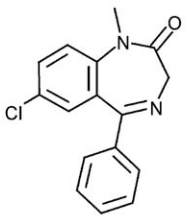
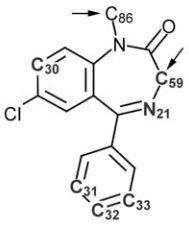
Compound Name	Structure	CypScore Structure model2 singleconf	METABOLIC SCORE	METABOLIC BOND SCORE
Diazepam			1 0.000 2 0.000 3 0.000 4 20.756 5 0.000 6 0.000 7 0.000 8 0.000 9 59.310 10 0.000 11 9.568 12 3.107 13 30.614 14 33.059 15 0.000 16 85.633 17 29.660 19 31.742	2 5 0.000 2 6 0.000 3 7 0.000 3 8 0.000 5 11 4.784 6 12 1.554 7 13 15.307 8 14 16.530 11 17 19.614 12 17 16.383 13 19 31.178 14 19 32.401

Figure 1. CypScore output visualization in the "pharmacophore informatics application". The CypScore structure column shows the molecule with the scores labeled for the atoms indicated in bold (arrows indicate the two highly reactive atoms in this case; the other numbered atoms here are moderately reactive). The metabolic bond score lists the atom numbers that define a given bond and the mean score for that bond. The metabolic score lists the score for each heavy atom.

and compounds with just one metabolite. The performance criteria are then:

1. the percentage of compounds with at least one major metabolite[14] identified,
2. the overall fraction of metabolites identified,
3. the ratio of true and false positives, and
4. the mean rank of major metabolites.[14]

The number of false positives in a molecule is not as well defined as the false negatives. Although MMDB contains as much information as can be gleaned from the literature, this information is necessarily incomplete because, for instance, not all metabolites are reported for some compounds. Additionally, the published data do not usually provide the rank order between different metabolic reactions within a molecule, and the major metabolite may not necessarily have been identified because of species and assay differences. However, false positives do not necessarily reflect an error in the model because of the structure of the data, as is illustrated in Figure 2.

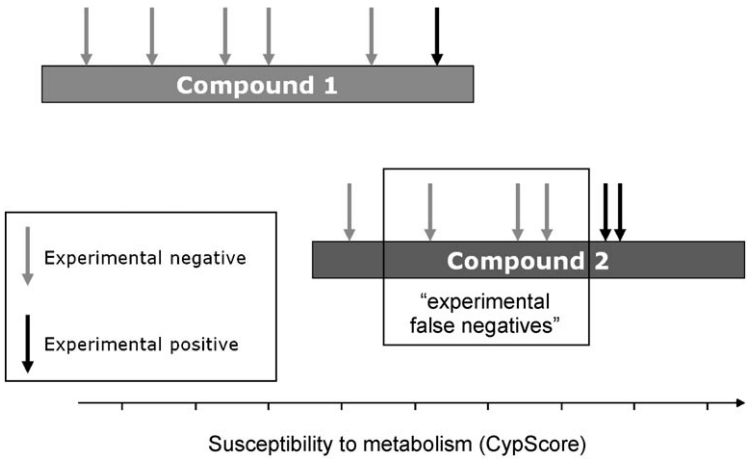


Figure 2. Schematic diagram of the occurrence of "experimental false negatives". The black arrows indicate metabolic products that are observed experimentally, and the gray arrows those that are not. The horizontal axis indicates the lability of the individual reaction sites. See the text for a detailed discussion.

Compound 1 is considerably more stable (less labile) than compound 2 toward metabolism by cytochrome P450. Therefore, its most reactive position is actually less labile than four potential reaction sites in compound 2 and is comparably reactive to one more. However, it is likely that only products from, say, the two most reactive sites will be found for compound 2 in the experimental assay (black arrows). This means that three potential reaction sites in compound 2 will be identified as "experimental negatives" although they are of higher or similar reactivity to the one site identified for compound 1; i.e., the rank order predicted is still the same as experimentally observed for compound 2. Because CypScore models use multiple linear regression (MLR), they can partially compensate for this effect. The situation illustrated in Figure 2 introduces a high level of noise into the training data. MLR can, however, ideally ignore (i.e., be unable to fit) the noise and thus be able to reproduce some of the hidden continuous reactivity scale inherent in the real nature of the data. In this case, however, it is of paramount importance to avoid overtraining (partially fitting the noise).

The false negatives, on the other hand, are instructive, because the experimental techniques used do not allow for "experimental false positives" as they do for "experimental false negatives". False negatives therefore indicate real weaknesses in the models. They are also useful in identifying ranking errors between different possible metabolic reactions.

Our ultimate aim is to construct a composite model that is able to rank the metabolites from the molecular structure in order of importance, taking the species and type of assay into account and considering the selectivity of individual isoforms. Our initial model is only designed to rank order potential reaction sites in and between molecules, and within and across chemical classes. The current version of CypScore has achieved this goal, as we show in the following. The ability to compare the reactivities of various positions and molecules in a purely ligand-based approach ignores any effect of ligand binding to cytochrome P450. In practice, CypScore has proven its ability to predict the relative stabilities of different molecules, probably because it has been used exclusively for druglike moieties. CypScore does not yet attempt to treat cytochrome P450 isoform selectivity because of the nature of the data provided by the main assay, which uses microsome- or hepatocyte- mediated phase I metabolism and therefore does not distinguish between isoforms.

Set 0: public test set

Set 0 was selected to contain druglike compounds of general interest along with some challenging small compounds with a mean value of 1.96 reactive positions per molecule. The compounds are metabolized by a variety of CYP isoforms and are mostly isoform-unspecific substrates. The structures of Set 0 together with a detailed listing of their CypScore results are given in Supporting Information scheme S1 and table S1.

Set 0 is a valid test set, because to the best of our knowledge it provides a complete picture of the most prominent cytochrome P450-mediated oxidation reactions, whereas the other public validation sets appear to be relatively incomplete. This is manifested by the significantly higher ratio of reactive positions per molecule for this dataset and Set 3 compared with Sets 1 and 2. Almost every compound was included in the training set for at least one model. However, because the individual models are weighted and centered to a common scale, this test set provides some validation of this weighting and scaling process.

The following results are based on the reactivity cutoff of 38 (percent scale) and 55 metabolized compounds. The Sheridan-adapted[10] plot in Figure 3 a shows the distribution of atom scores per molecule for all atoms for which models are avail-

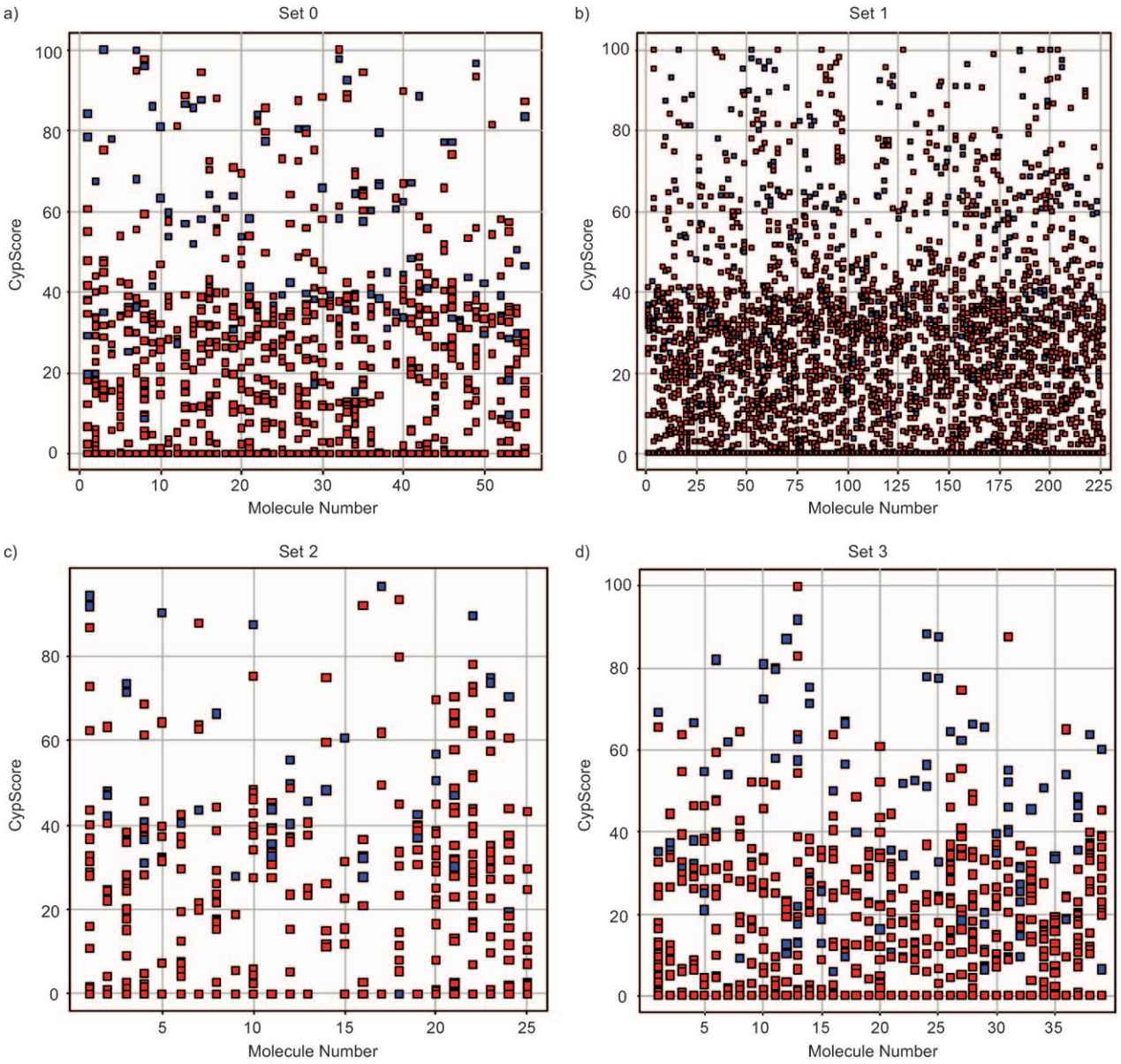


Figure 3. Prediction plots for the four datasets. The *y* axes give the scores ranging in percent per atom. The *x* axes give the molecules in the dataset order. Atoms in one molecule fall in a single column. Experimental oxidation sites are shown as blue squares and nonreactive sites as red squares.

able. Models are available for all observed products except for a single S-demethylation.

CypScore is able to rank at least one major metabolite in 62% (34) of the compounds highest. If we extend the prediction to the two highest-ranked positions, the success rate rises to 85% (47), and to 91% (50) for the three highest-ranked positions. Overall, with a reactivity cutoff of 38 we find 74% of all known reactive positions, with the ratio between true and false positives being 0.6. Note, however, that the effect shown in Figure 2 is particularly important for this dataset.

No reactive position was found for four compounds. One of them, loratadine **S24**, is actually stable toward P450 oxidation reactions, but is only cleaved by carbamate hydrolysis, contrary to reported results.[7] CypScore therefore correctly predicts no P450-reactive positions for this molecule.

The predictivity for most of the compounds is better than expected. For ten compounds, **S14**, **S15**, **S22**, **S24**, **S37**, **S38**, **S39**, **S47**, **S51**, and **S55**, we find exactly the correct metabolites and no further positives, even though visual inspection often suggests other potential sites of metabolism. For another 24 compounds, CypScore finds all experimental positions plus additional reactive atoms (i.e., false positives or "experimental false negatives"), some of which are likely minor experimental metabolites. For some compounds, such as verapamil **S35** or etoposide **S17**, many additional sites are predicted.

The current models give poor results for two difficult chemotypes: dihydropyridines (DHP) and steroids. CypScore consistently fails to predict the N-oxidation in DHP central rings as is the case, for example, with felodipine **S18** (and other examples not part of this dataset). We are aware that the current

N-oxidation model is the weakest because of the low number of examples in the original training sets. In the case of the steroids, which are less flexible than most of our training compounds, the orientation of the substrate in the P450 active site, rather than inherent reactivity, may determine the active positions. In **S23**, **S27**, and **S55**, one major metabolite is found, but not in **S12**.

The model also performs less well for crowded aliphatic functionalities such as *tert*-butyl and isopropyl groups or aliphatic 5-membered (heterocyclic) rings, which are predicted to be too stable relative to the rest of the molecule because of the low polarity of the C–H bonds and the small accessible surface areas of the hydrogen atoms. We were surprised by the generally high performance for small relatively non-drug-like molecules such as **S37**, **S38**, **S40**, **S41**, and **S51**, for which we find all or at least one major metabolite.

Set 1: public validation set

Set 1 consists of 226 compounds extracted from the training set of Sheridan. The dataset is large enough to be statistically significant and has a similar distribution of metabolic reactions. On the other hand, it also has some shortcomings. As outlined above, we often find more reactive positions in our database than in the Sheridan files for the 168 molecules that overlap with our MMDB. Therefore, we expect a significant number of "false positive" predictions by CypScore for important metabolites not annotated in the Set 1 original sources. This is also exemplified by the low mean value of reactive positions per molecule of only 1.67, compared with 2.76 for the MMDB and 2.3 for our in-house Set 3. We therefore consider this dataset to be a benchmark for the minimum capacity of our approach to identify major metabolites.

As we expected from these considerations of the quality of the dataset, the highest-ranked CypScore only accounts for one major metabolite in 39% of the compounds. For the two highest-ranked positions, we find 59%, and 66% (50) for the three highest-ranked positions. Overall, with a reactivity cutoff of 38 we find 56% of all known reactive positions. If we relax our threshold we will be able to find more and more major metabolites, but with a significant increase in the number of false positives.

The results for the separate isoform-specific datasets were similar, with the 2C9 subset performing best, at 45, 65, and 75% for the highest, two highest, and three highest-ranked positions, respectively. The "false positives" rate rises significantly for all the subsets as expected if the alternative reactive positions from the other isoforms are ignored, as CypScore is, by design, isoform unspecific.

Apart from missing reactive sites in the dataset, there are some more reasons for the lower performance. Set 1 contains 18 steroids and five DHPs, cores known to be problematic. Additionally, there are three compounds, parathion, phorate, and malathion, that are claimed to be oxidized at the phosphorus atom of a phosphate group. There are other quite unusual reactions such as the 3A4-mediated oxidations of carbonyls or thiocarbonyls, the N-oxidation of a primary amine or of a urea to an N-nitrosourea. At least one compound, phenformin, is a strong base. The models work on neutral structures because the data basis for the creation of charge-state-specific models is too small. Generally, we expect the model performance to be lower in these cases.

Set 2: public validation set

The Sheridan "calibration set" was condensed to the 25 isoform-unspecific compounds shown in Figure 4 with a molecular weight range of 108–494 Da. It has 41 reactive positions after combination of symmetry-equivalent atoms and aromatic edges, that is, again a low mean value of only 1.64 reactive positions per molecule. The distribution of metabolic reactions differs significantly from that of the other datasets (see Table 1) with a dominance of aromatic over aliphatic hydroxylation reactions.

Overall, we are able to identify 68% of all experimental metabolites; 15 compounds (60%) have one major metabolite ranked highest, 18 compounds (72%) have positions that are ranked 1 or 2, and for 84% of the compounds we find major metabolites ranked 1–3.

We calculated the rank statistics based on the annotated positions from the original source. We followed up on the primary citations for some unusual annotations, and these are discussed below.

Two compounds, methyleugenol **7** and *trans*-phenylpropene **19**, are oxidized via an intermediate allylic system. In one case, the label is on the methylene unit, and in the other, on all atoms of the propenyl, yielding CypScore prediction ranks of 3 and 1 based on the annotations, but prediction ranks of 1 based on chemical intuition, as one can expect both compounds to be metabolized via the same mechanism and at the same primary position. The product structures may differ nevertheless. This is a general observation for all double-bond oxidations, whether in alkenes or aromatic ring systems.

Compound Flu-1 **9** is one major metabolite of flutamide. The annotated reactions are an aromatic hydroxylation and the oxidation of a primary amine to a hydroxylamine. CypScore gives a score of 28 (moderately reactive) to the first and a score of 0 to the second reaction. The primary literature[20] indicates that the aromatic hydroxylation is indeed more important and that both reactions occur only at high enzyme concentrations (i.e., are at most moderately likely).

The published structure[21] for compound **25** of the source dataset, fluticasone, is in reality fluticonase propionate, i.e., it bears a propionate group instead of the hydroxy group. We did not exclude this compound from the dataset, although it is probably better to ignore it.

Compound **18**, R-95913, one major metabolite of prasugrel, is oxidized by CYP 3A to a carbonic acid via electrophilic-like attack of the Michael system. This new type of reaction, which we also find in Set 1 for carbonyl and thiocarbonyl groups, can be considered a candidate for a future reactivity model. Currently, it is predicted by the double-bond oxidation model.

There are some compounds for which CypScore is rather unspecific, especially **10**, **21**, and **22**, partially due to model defi-

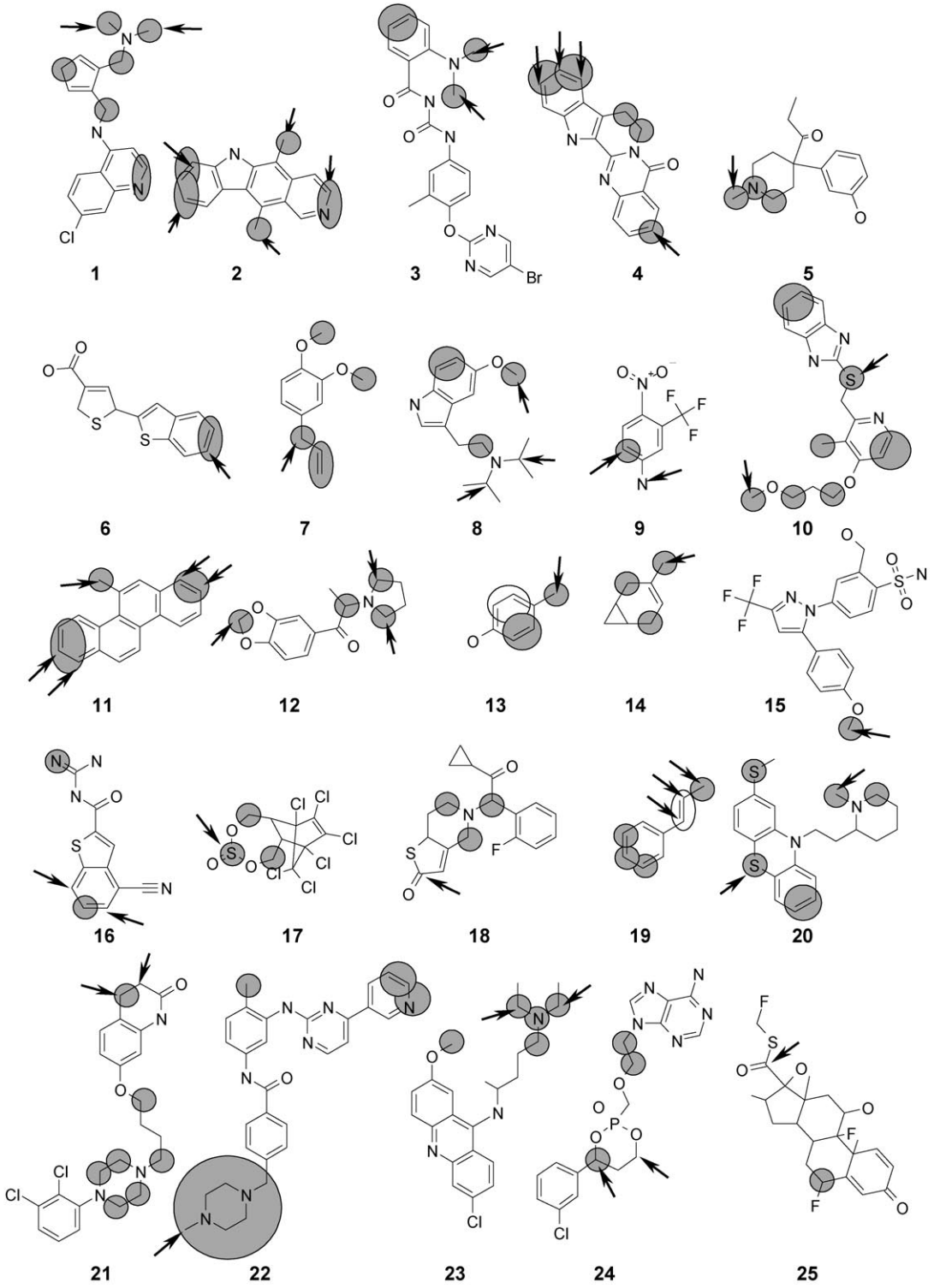


Figure 4. Structures of Set 2. CypScore predictions are shaded, and experimental positions are indicated with arrows.

ciencies and partly due to missing annotations or undetected reactive positives. Overall, CypScore is powerful for predicting labile positions and can guide the chemist for the next steps, and there is a clear path forward for future improvement. One anticipated improvement is to create multi-conformation models. The necessity becomes clear for a compound as simple as **13**, in which the equivalent aromatic edges have slightly different scores due to the relative orientations of the two substituents, which leads to an asymmetric electron distribution in the molecule.

Set 3: Bayer Schering Pharma validation set

The final proof of principle is the application to our daily lead optimization problems in pharmaceutical research. CypScore was used in 14 in-house projects during the first year after introduction. All 14 projects required support to help solve metabolic clearance problems. Microsomal stability and, to a very limited extent, Metabol-ID data were available in all projects. The problems to be addressed were:

1. to explore the exact weak positions in the compounds,
2. to provide chemical modifications that, if possible, do not interfere with the overall SAR, and
3. to select the most promising compounds from virtual libraries for the next synthesis steps.

Whenever possible, the CypScore predictions were compared with the available Metabol-ID data, and subsequently a few to several hundred compounds were designed for these projects based on combined in silico SAR and CypScore proposals. The iterative lead optimization process then used CypScore predictions together with the data from the microsomal stability assay and less often from Metabol-ID. In two of the projects, metabolic stability and affinity were optimized in parallel over several iterations, leading to the final drug candidates. In eleven other projects, the exact positions for metabolic clearance were identified and more stable compounds were designed. CypScore failed to give clear and useful predictions in just one out of fourteen projects (too many higher-ranked false positives were found). Our working hypothesis for this particular project is that the compounds are exclusively CYP 2C9 substrates (they fit published pharmacophore models for the 2C9 isoform selectivity almost perfectly).[22] If we then rule out all positions that do not fit into the pharmacophore models, the remaining predicted metabolites are significant.

We cannot show the chemical structures for these compounds from current projects. Instead, we provide performance statistics and graphs (Figure 3 d) analogous to those given for the public sets.

Between one and eight (mean 3.14) compounds per project were selected for which Metabol-ID data were available and unambiguously interpretable, including two examples from the one project for which CypScore was not predictive. The numbers of major metabolites in these compounds range from 1 to 5, with a mean value of 2.26. The profile of P450-mediated phase I transformations is different from those of the public and MMDB sets, as shown in Table 1. In one project, S-dealkylation occurs as one major pathway. Because we lack training data, we do not have an S-dealkylation model, and so we have no results for this reaction.

CypScore finds 1.41 (rank 1–3) of 2.26 reactive positions (mean values) in the 39 compounds. The mean number of false positives is 0.95. This true- to false-positives ratio of 0.9 is far better than for the public test set (0.6).

The accumulated CypScore performance is 59% of the major metabolites predicted for the most reactive ranked position, 80% for the two most reactive, and 87% for the three most reactive, a little lower than for the public validation set. On the other hand, both the metabolic stability and the Metabol-ID assay show very low clearances for three compounds, in accordance with the mean reactivity values given by CypScore. If these positions of moderate reactivity are also counted, the rank order performance is even better than for the public validation set. This is probably because the public set covers a broader spectrum of structural classes (including some small compounds) and metabolic reactions than we expect for our compounds. Thus, the models have proven their effectiveness, especially in combination with the metabolic stability assay.

Performance over all datasets

Table 2 shows a summary of the overall performance of CypScore for the validation datasets, and Figure 5 shows plots of the percentages of reactive and stable positions with the corresponding predictions as reactive (Score $\geq$ 38), moderately reactive (38 $>$ Score $\geq$ 22), and stable (Score $<$ 22).

Except for Set 1, we are always able to identify more than 60% of all known metabolites with our current empirical threshold of 38. Using the same threshold, the percentages of false positives (i.e., stable positions predicted to be reactive) are impressively low. This is especially true for our in-house compounds, for which the rate is less than 6%. The false positives rate is also quite low considering that only about 6–8% of all heavy atom moieties are reactive. Our current prediction rates imply that the ratios between true and false positives are 0.60, 0.37, 0.37, and 0.90 for Sets 0, 1, 2, and 3, respectively. One of the future goals is to enhance these rates further if appropriate experimental data become available. The models are, however, predictive enough that the "outliers" are worth examining.

One of the reasons why the in-house dataset performs so well might be the higher ratio of reactive positions per molecule, due to the nature of the compounds from lead optimization projects relative to marketed drugs and late-stage compounds in the other datasets. The lead-optimization project compounds have many reactive positions, because they were submitted to the assay and became part of the dataset for exactly this reason. On the other hand, the superior performance

Table 2. Summary of the CypScore performance for the four different validation datasets.

	Set 0	Set 1	Set 2	Set 3
Compounds with a metabolic position ranked highest [%]:	62	39	60	59
Compounds with a metabolic position ranked 1 or 2 [%]:	85	59	72	80
Compounds with a metabolic position ranked 1, 2 or 3 [%]:	91	66	84	87
Reactive positions with a score $>$ 38 [%]:	74	56	68	62
Ratio between true and false positives:	0.60	0.37	0.37	0.90

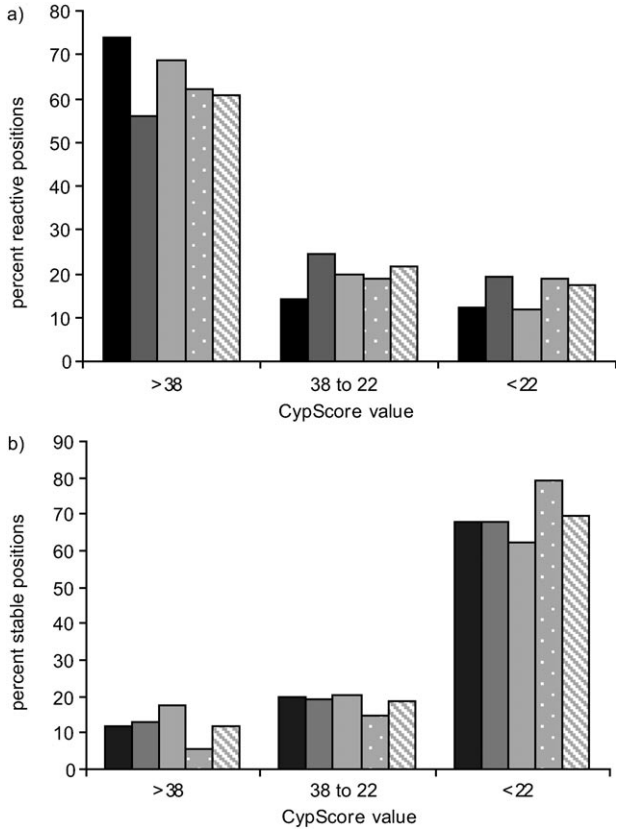


Figure 5. Percentages of a) reactive and b) stable positions in the datasets predicted to be reactive (Score $\geq$ 38), medium reactive (38 > Score $\geq$ 22), and stable (Score < 22). Set 0: ■, Set 1: ■, Set 2: ■, Set 3: dotted gray bars, combined performance: diagonally striped bars.

of CypScore for these compounds relative to the others manifests the method's suitability to the task for which it was built.

Conclusions

We present herein a novel and effective in silico tool for predicting the exact metabolic positions in drug molecules. CypScore is based on the concept of having separate models for the specific chemical oxidation reactions that occur in cytochromes P450. The models are designed to mimic as closely as possible the enzymatic reactivity for such diverse reactions as aliphatic, aromatic, double bond, N- and O-oxidations. The models use atomic reactivity descriptors derived from semiempirical MO calculations and have been trained on the basis of subsets of a large and well-reviewed metabolite database established in house with about 2400 transformations on more than 800 molecules from the literature, covering a large variety of P450 isoforms and species. The models are carefully balanced against each other to allow for intra-molecule, intra-chemotype, and cross-chemotype prediction and ranking.

CypScore is a purely reaction-based system that ignores the influence of the protein surroundings of the active site. This design was chosen based on the strategy of first establishing a high-quality representation for inherent reactivity, and in a second step ruling out inaccessible positions.

We have provided evidence of the validity of the approach based on four differently designed datasets. We are able to identify more than 60% of all known phase I major metabolites (except for data set 1, which is probably incomplete regarding experimental data) with good accuracy for the true- to false-positives ratio, and in more than 85% of the compounds we identify at least one major metabolite among the three highest-ranked positions. CypScore is now a robust tool, and there is a clear path forward for stepwise improvements beyond its current basis.

CypScore predictions are strictly based on atomic descriptors from quantum-mechanics wave functions. The wave-function-derived descriptors reflect the influence of the electron distribution over a whole molecule on the reactivity of a specific atomic position, in contrast to strictly rule-, substructure-, or molecular-descriptor-based QSAR approaches. CypScore applies distinct models for each metabolic reaction, and can therefore be modularly extended to cover new reaction types for phase I, phase II, or even plasma-mediated reactions. The models will probably further improve, if quantitative metabolite data would be used for training instead of binary information from the literature. Complete experimental data will also further improve the true- to false-positives ratio. If larger datasets with known ligand protonation states are available, the models will further improve as we can deduce from tests with substrate protonation.

The CypScore ansatz described herein uses one starting conformation derived from Corina as input to the semiempirical calculation which might not be the active conformation metabolized by a P450 enzyme. Therefore, the extension to a multi-conformer approach is a clear step forward.

Another influence on the reactivity currently ignored by CypScore but to be explored is the protein cavity, with its influence on ligand orientation and accessibility constraints for the metabolic sites. Nevertheless, in its current state, CypScore has already proven useful in parallel SAR and metabolic optimization in 13 out of 14 lead optimization projects to which it was applied. It is now used routinely by Bayer Schering Pharma computational chemists in parallel with other ADME and SAR optimization techniques.

Experimental Section

Model generation

The 2D structures in the MMDB were extracted as SD files and translated to 3D structures with the addition of hydrogen atoms and neutralization by Corina.[17] Molecular geometries were then optimized[23] in the package VAMP 10.0[24] with the semiempirical AM1 Hamiltonian.[18] Atomic and molecular descriptors derived directly from the AM1 wave function were bond orders, Coulson charges,[25] π- and lone-pair charges as the sum of the two natural atomic orbital point charges (NAO-PCs)[26] that correspond to the occupation of the appropriate π or sp^x orbitals, atomic solvent-accessible surfaces (*SASA*)[24] obtained by partitioning the molecular accessible surfaces (van der Waals + 1.4 Å solvent radius), and the eigenvalues and the percent p character[27] of the localized lone-pair orbital.

Atomic surface properties were calculated with ParaSurf[19] on the atomic isodensity surface at 3×10^{-4} e$^-$·Å^{-3} using multipole electrostatics.[28] The most important atomic surface properties from ParaSurf™ are the local ionization energy IE_L,[29] the local electron affinity EA_L,[30] the molecular electrostatic potential V,[31] and the local polarizability α_L,[30] among others listed in Table 3. For these properties, ParaSurf™ calculates mean, minimum, and maximum values for each atom. Values for atoms with zero accessible surface area are set to zero.

The various reaction models were created by MLR using TSAR 3.3.[32] Although it might appear unusual, or perhaps even incorrect, to train simple MLR models on binary data, this approach is theoretically justified if the underlying variable is continuous and not binary,[33] which is the case for products of metabolism, and is an important feature of our models. This is because the experimental data are not consistent from molecule to molecule, but rather create the impression of binary (yes/no) data. This point is illustrated in Figure 2 and in the discussion above, but the result is that the experimental results only indicate the most reactive position(s) in the molecule being tested. The absolute reactivity of the sites found in various molecules may differ widely.

The final step is to establish a common scale of reactivity for the various reaction classes. This step is partially empirical, may be repeated iteratively based on new knowledge, and can be considered successful if the global model obtained is able to rank the various potential metabolic reactions in an independent test set. First, we must define a common reference point for all the models. For each model, the crossing points of the enrichment curves for false positives and false negatives for the various reactions are all set to zero by adjusting the constants in the MLR equations. The zero value is then the reference parameter that describes the same level of reactivity for all the reactions. The equations are then scaled to a common value range from -10 (extremely labile) to $+10$ (perfectly stable). This is achieved by determining a scaling factor so that the largest absolute CypScore is 10. If extreme outliers were present, these were ignored. The scaled curves of false positives and negatives are shown in Supporting Information figure S1. Note that this scaling is based on the performance of the models in predicting correctly and not on any absolute reactivities, which may not cover the same range for each reaction.

Finally, because we found from validation set compounds that the significant value range observed in almost all compounds is between -7 and $+3$, we decided to generate a more easily understood score that we call percent lability. This score is obtained by linearly transforming the 10 reactivity units value scale into a percent range from 100% (-7) to 0% ($+3$). Any value outside this range is transformed into the closest border value. Within this scale, an empirical cutoff of 38 has been found appropriate as the boundary between reactive and stable sites. A second cutoff of 22 is used to indicate sites of medium reactivity.

Models

CypScore currently provides six different models for seven generic P450-mediated oxidation reactions. These are aliphatic hydroxylation, N-dealkylation, O-dealkylation, double-bond oxidation, aromatic hydroxylation, N-oxidation, and S-oxidation. Details of the proposed reaction mechanisms have been published,[4,34] but this knowledge is not used to derive the models. Nevertheless, we find that the automatically selected descriptors reflect the important parameters for these reaction mechanisms. In the following, we describe each model and its descriptors. The descriptors used in the various models are given in Table 3.

Aliphatic hydroxylation, N-dealkylation, O-dealkylation

These three reactions share the same primary step, the insertion of the active oxygen from the so-called compound I into a C–H bond. Two reaction mechanisms have been discussed: the concerted and the radical abstraction mechanisms.[11] There is some evidence that both or even a mixture may apply, depending on the compound to be oxidized. Training was based on 180 compounds and 286 hydroxylation reactions.

The descriptors for the aliphatic hydroxylation model (with the atoms to which the descriptors apply given in brackets) are the atomic solvent-accessible surface area *SASA*, the Coulson charge *q*, the maximum local ionization energy IE_L^{max}, and the bond order *BO* of the C–H bond, which give regression Equation (1):

Table 3. Descriptors used in the various CypScore models.

Symbol	Description	Calculated from:	Calculated for:
BO	Bond order	VAMP	C–H
SASA	Atomic solvent-accessible surface area (solvent radius 1.4 Å)	VAMP	all atoms
Valence	Atomic valence	VAMP	all atoms
q	Coulson charge	VAMP	all atoms
q_π	Sum of the two NAO-PCs belonging to the p orbital of an atom in a π system	VAMP	C sp^2, NR_2
q_{lp}	Sum of the two NAO-PCs belonging to the lone-pair orbital of nitrogen atoms	VAMP	N
E_{LMO}	Eigenvalue of the localized lone-pair orbital	VAMP	N
%p	Percent p character in the localized lone-pair orbital	VAMP	N
Area	Atomic surface area	ParaSurf™	all atoms
V_{max}	Maximum molecular electrostatic potential on atomic surface	ParaSurf™	all atoms[a]
V_{min}	Minimum molecular electrostatic potential on atomic surface	ParaSurf™	all atoms[a]
IE_L^{max}	Maximum local ionization energy on the atomic surface	ParaSurf™	all atoms[a]
IE_L^{min}	Minimum local ionization energy on the atomic surface	ParaSurf™	all atoms[a]
EA_L^{max}	Maximum local electron affinity on the atomic surface	ParaSurf™	all atoms[a]
EA_L^{min}	Minimum local electron affinity on the atomic surface	ParaSurf™	all atoms[a]
$\bar{\alpha}_L$	Mean local polarizability on the atomic surface	ParaSurf™	all atoms[a]

[a] The atomic surface properties of atoms with a surface area of zero are undefined and thus also set to zero. These atoms are considered as metabolically stable and are ignored during model generation.

$$Y = 0.00906 \cdot SASA(\mathrm{H}) + 0.00213 \cdot IE_{\mathrm{L}}{}^{\max}(\mathrm{H}) - 6.00 \cdot q(\mathrm{H}) - 6.67 \cdot BO(\mathrm{C-H}) + 5.792 \quad (1)$$

Aromatic hydroxylation

The proposed first reaction step for the aromatic hydroxylation is the electrophilic addition of an activated iron-oxo species to the aromatic ring to give a tetrahedral intermediate, which, after [1,2]-hydride shift, leads to hexadienone, which then tautomerizes to the phenol.[11] For this mechanism to occur, two neighboring aromatic atoms, most likely both sp^2 carbon atoms, are required. The mechanism results in an ambiguity with respect to the annotated metabolites in the experimental literature. The hydroxylated position in the metabolite will often not be the most reactive position in an aromatic ring, but instead one of the neighbors because of the hydride shift.

The reactivity can be described by four atomic descriptors, the accessibility of the carbon atom (*SASA*), maximum local ionization energy $IE_{\mathrm{L}}{}^{\max}$, the atomic polarizability α, and the π charge q_π. The training dataset contains 7377 aromatic C sp^2 atoms that undergo 286 aromatic hydroxylations:

$$Y = 0.0108 \cdot SASA(\mathrm{C}) - 0.000154 \cdot IE_{\mathrm{L}}{}^{\max}(\mathrm{C}) + 0.397 \cdot \overline{\alpha_{\mathrm{L}}}(\mathrm{C}) - 0.140 \cdot q_\pi(\mathrm{C}) - 0.213 \quad (2)$$

Double-bond epoxidation/oxidation

These reactions are relatively uncommon, but are nevertheless important. The dataset is small and unbalanced and consists of 344 C=C double bonds, only 16 of which are metabolized.

We found that regression equations that use only the descriptors of the two carbon atoms in the double bond give poor reactivity predictions for terminal or exocyclic double bonds. Our final model uses the mean descriptor values for the two carbon atoms, that is, the mean values of the solvent-accessible surface area *SASA*, the maximum localized electron affinity $EA_{\mathrm{L}}{}^{\max}$, and the minimum localized electron affinity $EA_{\mathrm{L}}{}^{\min}$ in Equation (3). We find that this model also has some predictive capacity for the oxidation of carbonyl groups and the lability of amides toward amide hydrolases.

$$Y = 0.0254 \cdot \overline{SASA}(\mathrm{C}) - 0.00270 \cdot \overline{EA_{\mathrm{L}}}{}^{\max}(\mathrm{C}) + 0.00231 \cdot \overline{EA_{\mathrm{L}}}{}^{\min}(\mathrm{C}) + 0.026 \quad (3)$$

N-oxidation

N-oxidation of imino and amino groups cannot be handled with a single regression equation, because of the different hybridizations and accessible surface areas. We therefore developed two individual models for R^1–N(–R^2)–R^3 and R^1=N–R^2 groups, based on 1017 and 442 observations, with 42 and 15 metabolic positions, respectively.

Amines:

$$Y = 0.00323 \cdot SASA(\mathrm{N}) - 0.000797 \cdot IE_{\mathrm{L}}{}^{\max}(\mathrm{N}) + 0.515 \quad (4a)$$

Imines:

$$Y = 0.00462 \cdot SASA(\mathrm{N}) + 0.00867 \cdot \%\,\mathrm{p}(\mathrm{N}) - 0.267 \quad (4b)$$

S-oxidation

The dataset contained 120 sulfur atoms (5 S=R, 94 R^1–S–R^2, and 21 R^1–S(=O)–R^2) with 47 metabolic positions. MLR gave three sulfur-atom descriptors: the minimum electrostatic potential $V_{\min}$, the minimum localized ionization potential $IE_{\mathrm{L}}{}^{\min}$, and the maximum localized electron affinity $EA_{\mathrm{L}}{}^{\max}$.

$$Y = -0.00606 \cdot V_{\min}(\mathrm{S}) + 0.00340 \cdot IE_{\mathrm{L}}{}^{\min}(\mathrm{S}) - 0.0171 \cdot EA_{\mathrm{L}}{}^{\max}(\mathrm{S}) - 0.903 \quad (5)$$

Acknowledgements

We are grateful to Thomas Krämer for his support, which allowed us to start the project, and Michael Boberg, Dieter Lang, Martin Radtke, and Klemens Lustig from the pharmacokinetics department for fruitful discussions and scientific advice in the interpretation and understanding of the experimental setup and data. Work at the University of Erlangen was funded by Bayer HealthCare.

Keywords: AM1 · cytochromes · metabolism · semiempirical calculations · surface properties

[1] D. A. Smith, H. van de Waterbeemd, D. K. Walker, *Pharmacokinetics and Metabolism in Drug Design* (Eds.: R. Mannhold, H. Kubinyi, H. Timmerman), Wiley-VCH, Berlin, **2001**, pp. 75–97.
[2] M. Bertrand, P. Jackson, P. Walther, *Eur. J. Pharm. Sci.* **2000**, *11 (2 Suppl.)*, 61–72.
[3] R. B. van Breemen, D. Nikolic, J. L. Bolton, *Drug Metab. Dispos.* **1998**, *26*, 85–90.
[4] a) M. Sono, M. Roach, E. Coulter, J. Dawson, *Chem. Rev.* **1996**, *73*, 2841–2887; b) B. Testa in *Biochemistry of Redox Reactions (Metabolism of Drugs and Other Xenobiotics)* (Eds.: B. Testa, J. Caldwell), Academic Press, New York, **1994**, pp. 1–40.
[5] P. Crivori, I. Poggesi, *Eur. J. Med. Chem.* **2006**, *41*, 795–808.
[6] J. P. Jones, M. Mysinger, K. R. Korzekwa, *Drug Inf. News Drug Met. Disp.* **2002**, *30*, 7–12.
[7] S. B. Singh, L. Q. Shen, M. J. Walker, R. P. Sheridan, *J. Med. Chem.* **2003**, *46*, 1330–1336.
[8] G. Cruciani, E. Carosati, B. D. Boeck, K. Ethirajulu, C. Mackie, T. Howe, R. Vianello, *J. Med. Chem.* **2005**, *48*, 6970–6979.
[9] a) S. Boyer, I. Zamora, *J. Comput. Aided Mol. Des.* **2002**, *16*, 403–413; b) S. Boyer, C. H. Arnby, L. Carlsson, J. Smith, V. Stein, R. C. Glen, *J. Med. Chem.* **2007**, *50*, 583–590; c) L. Ridder, M. Wagener, *ChemMedChem* **2008**, *3*, 821–832.
[10] R. P. Sheridan, K. R. Korzekwa, R. A. Torres, M. J. Walker, *J. Med. Chem.* **2007**, *50*, 3173–3184.
[11] a) G. H. Loew, D. L. Harris, *Chem. Rev.* **2000**, *77*, 407–419; b) S. Shaik, D. Kumar, S. E. de Visser, A. Altun, W. Thiel, *Chem. Rev.* **2005**, *82*, 2279–2328; c) B. Meunier, S. E. de Visser, S. Shaik, *Chem. Rev.* **2004**, *81*, 3947–3980.
[12] a) *Drug Met. Disp.* **1987** to **03/2005**; b) *Xenobiotica* **1997** to **11/2004**; c) *J. Clin. Pharm.* **1999** to **11/2004**; d) *Biopharm. Drug. Disp.* **1998** to **09/2004**; e) *Clin. Pharm. Therapeutics* **1999** to **02/2005**; f) *Biochem. Pharm.* **2003** to **12/2004**.
[13] *Biotransformations* **1987** to **1995**.
[14] We considered major metabolites as such if they accounted for at least 10% in the final product spectrum collected experimentally; that is, we used only published sources that quantified the metabolites or defined their metabolic classification analogously.
[15] D. F. V. Lewis, *Curr. Med. Chem.* **2003**, *10*, 1955–1972.
[16] MDL file formats: http://www.mdl.com/solutions/white_papers/ctfile_formats.jsp (accessed February 3, 2009).

[17] J. Sadowski, C. Schwab, J. Gasteiger, Corina 3.4: Molecular Networks GmbH, **2007**.
[18] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
[19] T. Clark, J. Lin, A. H. C. Horn, ParaSurf 07™: CEPOS InSilico Ltd., The Old Vicarage 132 Bedford Road, Kempston, Bedford, MK42 8BQ (UK) **2007**, http://www.ceposinsilico.com (accessed February 3, 2009).
[20] R. Goda, D. Nagai, Y. Akiyama, K. Nishikawa, I. Ikemoto, Y. Aizawa, K. Nagata, Y. Yamazoe, *Drug Metab. Dispos.* **2006**, *34*, 828–835.
[21] R. E. Pearce, J. S. Leeder, G. L. Kearns, *Drug Metab. Dispos.* **2006**, *34*, 1035–1040.
[22] S. Ekins, M. J. de Groot, J. P. Jones, *Drug Metab. Dispos.* **2001**, *29*, 936–944.
[23] No symmetry constraints, gradient norm = 0.4 mol^{-1} $Å^{-1}$, keywords: AM1 IIS SDINP XYZ NOZ ALLVECTORS LOCALIZE AREAS RSOLV = 1.4 *T* = 240H.
[24] T. Clark, A. Alex, B. Beck, F. Burckhardt, J. Chandrasekhar, P. Gedeck, A. Horn, M. Hutter, B. Martin, G. Rauhut, W. Sauer, T. Schindler, T. Steinke, VAMP 10.0: Erlangen (Germany), **2007**, available from Accelrys Inc., San Diego, CA (USA).
[25] C. A. Coulson, H. C. Longuet-Higgins, *Proc. R. Soc. (London)* **1947**, *A191*, 39–60.
[26] a) G. Rauhut, T. Clark, *J. Comput. Chem.* **1993**, *14*, 503–509; b) B. Beck, G. Rauhut, T. Clark, *J. Comput. Chem.* **1994**, *15*, 1064–1073.
[27] M. Hennemann, T. Clark, *J. Mol. Model.* **2002**, *8*, 95–101.
[28] A. H. C. Horn, J.-H. Lin, T. Clark, *Theor. Chem. Acc.* **2005**, *113*, 159–168; Erratum: *Theor. Chem. Acc.* **2007**, *117*, 461–465.
[29] P. Sjoberg, J. S. Murray, T. Brinck, P. A. Politzer, *Can. J. Chem.* **1990**, *68*, 1440–1443.
[30] B. Ehresmann, B. Martin, A. H. C. Horn, T. Clark, *J. Mol. Model.* **2003**, *9*, 342–347.
[31] *Chemical Applications of Atomic and Molecular Electrostatic Potentials: Reactivity, Structure, Scattering, and Energetics of Organic, Inorganic, and Biological Systems* (Eds.: P. Politzer, D. G. Truhlar), Plenum Press, New York, **1981**.
[32] TSAR 3.3: Oxford Molecular Ltd., Oxford (UK), **2000**, available from Accelrys Inc., San Diego, CA (USA).
[33] G. S. Madalla, *Limited-dependent and Qualitative Variables in Econometrics*, Econometric Society Monographs, Cambridge University Press, Cambridge, **1983**.
[34] R. B. Silverman, *The Organic Chemistry of Drug Design and Drug Action, 2nd ed.*, Elsevier Academic Press, Amsterdam, **2004**, pp. 405–456.

Received: November 13, 2008
Revised: January 24, 2009
Published online on February 26, 2009

