

Model based on GRID-derived descriptors for estimating CYP3A4 enzyme stability of potential drug candidates

Patrizia Crivori^{a,*}, Ismael Zamora^c, Bill Speed^a, Christian Orrenius^b & Italo Poggesi^a

^aPharmacokinetics, Dynamics and Metabolism and ^bChemistry Department, Pharmacia, Gruppo Pfizer Inc., Viale Pasteur 10, I-20014 Nerviano (Mi), Italy; ^cGrupo de Recerca Informatica Biomedica IMIM, Pompeu Fabra University, Dr. Aiguader 80, E-08003 Barcelona, Spain

Received 22 January 2004; accepted in revised form 31 March 2004

Key words: CYP3A4 *in silico* screening, CYP3A4 stability, GRIND descriptors, Partial Least Squares Discriminant PLSD, Quantitative Structure Property Relationships (QSPR), VolSurf descriptors

Summary

A number of computational approaches are being proposed for an early optimization of ADME (absorption, distribution, metabolism and excretion) properties to increase the success rate in drug discovery. The present study describes the development of an *in silico* model able to estimate, from the three-dimensional structure of a molecule, the stability of a compound with respect to the human cytochrome P450 (CYP) 3A4 enzyme activity. Stability data were obtained by measuring the amount of unchanged compound remaining after a standardized incubation with human cDNA-expressed CYP3A4. The computational method transforms the three-dimensional molecular interaction fields (MIFs) generated from the molecular structure into descriptors (VolSurf and Almond procedures). The descriptors were correlated to the experimental metabolic stability classes by a partial least squares discriminant procedure. The model was trained using a set of 1800 compounds from the Pharmacia collection and was validated using two test sets: the first one including 825 compounds from the Pharmacia collection and the second one consisting of 20 known drugs. This model correctly predicted 75% of the first and 85% of the second test set and showed a precision above 86% to correctly select metabolically stable compounds. The model appears a valuable tool in the design of virtual libraries to bias the selection toward more stable compounds.

Abbreviations: ADME – absorption, distribution, metabolism and excretion; CYP – cytochrome P450; MIFs – molecular interaction fields; HTS – high throughput screening; DDI – drug-drug interactions; 3D – three-dimensional; PCA – principal components analysis; CPCA – consensus principal components analysis; PLS – partial least squares; PLSD – partial least squares discriminant; GRIND – grid independent descriptors; GRID – software originally created and developed by Professor Peter Goodford.

Introduction

In recent years, combinatorial chemistry and high throughput screening (HTS) approaches have been actively pursued in pharmaceutical companies to increase the probability to gain quick access to novel drugs. Since the number of viable leads obtained using these approaches is still low, significant effort

is being spent to generate more specialized chemical libraries enriched with compounds having ‘drug-like’ properties [1]. In this context, virtual screening and filtering methods focused on specific pharmacological targets and relevant pharmacokinetic properties are increasingly used in the drug design process. Such methods have the advantage to be useful in the selection and design of potential candidates and to reduce the cost and resources dedicated to compounds that are unlikely to reach the market [2–4].

*To whom correspondence should be addressed. Fax: +39-02-48383965; e-mail: Patrizia.Crivori@Pharmacia.com

Metabolic stability is important for ensuring both low clearance and acceptable oral bioavailability through minimization of the first pass metabolism. Thus, drug stability towards the cytochrome P450 (CYP) family of enzymes has been recognized as one of the most important properties to be measured and estimated early in drug research [5]. The CYP enzymes include a large superfamily of heme-coupled monooxygenases that catalyze a wide variety of oxidative reactions [6, 7]. A high metabolic clearance may prevent the achievement of therapeutically relevant plasma levels. Also, since many drug-metabolizing enzymes are genetically polymorphic, this may cause high variability in the exposure and, in turn, difficulties in the clinical use of a drug. Finally, molecules extensively metabolized by a single CYP have the potential to undergo drug–drug interactions (DDI) [8], which may require extensive clinical monitoring.

Metabolic stability screens are usually based on liver preparations (hepatocytes, microsomes) that express the whole range of CYPs. However, since relatively few isoforms (2A1, 2D6, 2C9, 3A4, 2E1, 2A6) are involved in the metabolism of most drugs [9], other screens rely on the evaluation of stability toward an individual CYP. In particular CYP3A4, which is the major CYP enzyme expressed in human liver and small intestine, is responsible for the oxidative metabolism of a large number of drugs (approximately one-third of all the marketed drugs) [10, 6]. Since the fraction of the oral dose that reaches the systemic circulation is influenced by absorption and both hepatic and intestinal first-pass metabolism, the effect of an extensive CYP3A4-mediated oxidation has a critical impact on the oral bioavailability. Also, due to its importance, metabolism mediated solely by CYP3A4 may imply a high risk of a DDI. Undoubtedly, screening based on liver preparations is likely to be more predictive of the *in vivo* outcome. A small number of false negatives, due to metabolism of compounds by other cytochromes or other elimination mechanisms, would be expected in a CYP3A4-based screening. For example, propranolol is classified as stable in the CYP3A4 screening assay because it is not extensively metabolized by this isoenzyme. Despite this, it has to be considered unstable *in vivo*, as its high clearance (1120 ml min^{-1}) is due to CYP2D6 and CYP1A2 [11]. However, compared with screens in which the whole range of oxidative enzymes are used, the approach of focusing on an individual CYP has the advantage of facilitating the possibility to move from an *in vitro* to an *in silico* screen. In this case, the structural fea-

tures responsible for the instability will be more easily identifiable and the interpretation of a computational model will be simpler.

In the present study we describe a computational model, extensively validated using two external sets and comparing several criteria of robustness, for predicting CYP3A4 metabolic stability of potential drug candidates. The approach was based on principal component (PCA) and partial least squares discriminant (PLSD) analysis of the descriptors derived from three-dimensional molecular interaction fields (MIFs) [12, 13], calculated from the compounds under study. Such a model could be applied as a filter in virtual screening to bias libraries toward more stable compounds, allowing a fast and cost-effective elimination of molecules that unlikely will be pursued in later stages of drug development. Applied in combination with computational methods able to identify the most likely site(s) of biotransformation in a molecule, it would also be a valuable tool in the lead optimization phase [14, 15].

Methods

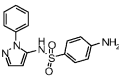
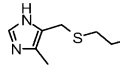

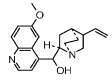
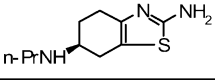
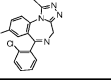
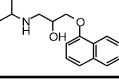
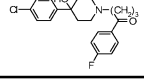

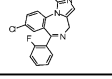
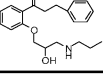
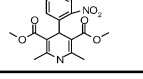
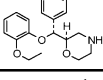
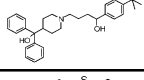

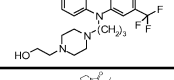
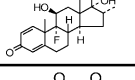
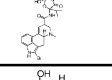
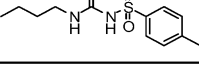
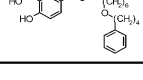
Datasets

The starting set used for model development consisted of 1800 compounds taken from the Pharmacia collection. Of the two test sets used for model validation, the former consisted of 825 Pharmacia compounds (first test set) and the latter (second test set) of 20 known drugs. The structures of the second test set are reported in Table 1. The entire dataset came from different therapeutic areas and included several structurally diverse chemical classes. The two external sets were selected in a way that the descriptor space explored by them was well covered by the training molecules used to develop the model. This avoided making unreliable predictions for compounds that differed substantially from the training set.

Experimental metabolic stability

In vitro CYP3A4 metabolic stability of the studied compounds was assessed using the following protocol. In the medium-throughput screening assay, each compound was incubated at a concentration of $2 \mu\text{M}$ for 60 min with human CYP3A4 cDNA expressed microsomal preparations (Gentest; 6 pmol) at 37°C . The assay was performed in a 96-well plate format. The incubation was stopped by the addition of $100 \mu\text{L}$ of acetonitrile. After centrifugation the supernatant

Table 1. 2D structures of the 20 compounds included in the second test set.

Name	Structure	Name	Structure
Sulfaphenazole		Cimetidine	
Imipramine		Quinidine	
Pramipexole		Triazolam	
Propranolol		Haloperidol	
Desipramine		Midazolam	
Propafenone		Nifedipine	
Reboxetine		Terfenadine	
Dextromethorphan		Fluphenazine	
Dexamethasone		Bromocriptine	
Tolbutamide		Salmeterol	

was analyzed using automated high throughput generic LC/MS/MS methodology, with a cycle time of <2 min per sample. The percent of metabolic stability was then calculated as the ratio between the corresponding compound concentrations in the supernatant and in the control sample (i.e. incubated with insect control, Gentest) by the following equation:

$$\% \text{ Metabolic Stability} =$$

$$100 \frac{[\text{compound A in supernatant}]}{[\text{compound A in control sample}]}$$

Therefore, each compound was experimentally assigned a value of metabolic stability ranging from 0% to 100%. Considering the relevance of the metabolic stability data in our screening funnel, we divided the compounds in two classes: unstable (U) class (% metabolic stability <40) and stable (S) class (% metabolic stability \geq 40).

Computational procedure

The molecules, modeled in their neutral form, were converted into three-dimensional (3D) structures using the program Corina [16]. From the 3D structures, molecular descriptors were calculated using both the Volsurf [17] and Almond [18] programs. A range of statistical approaches, i.e. principal component analysis (PCA) [19], consensus principal component (CPCA) [20] and partial least squares discriminant (PLSD) analysis [21], were then applied to identify correlations between the calculated descriptors and the two predefined classes of stable and unstable compounds using the GOLPE software package [22]. Below the details of these computational procedures are reported.

Table 2. VolSurf descriptors.

Descriptors Obtained from the Hydrophilic (H₂O) Interaction Fields	
V	Volume of the water molecule interaction field at 0.2 kcal/mol energy level
S	Surface of the water interaction field at 0.2 kcal/mol energy level
R	Rugosity: ratio between the volume (V) and the surface (S)
G	Globularity: ratio between the surface (S) and the surface of a sphere with the same volume (V)
W1OH2-W8OH2	Volumes of the water molecule interaction fields at eight different energy levels: -0.2, -0.5, -1.0, -2.0, -3.0, -4.0, -5.0 and -6.0 kcal/mol
Iw1-Iw8	Integy moments: distances between the center of mass of the molecule and the center of the hydrophilic regions calculated at the same eight energy levels as W1OH2-W8OH2
Cw1-Cw8	Capacity factors: ratio between the hydrophilic regions (W1OH2-W8OH2) and the molecular surface (S)
Emin1, Emin2, Emin3	Energy values for the three lowest energy minima
d12, d13 and d23	Distances between the three lowest energy minima
Descriptors Obtained from the Hydrophobic (DRY) Interaction Fields	
D1-D8	Volumes of the hydrophobic interaction fields at eight energy levels: -0.2, -0.4, -0.6, -0.8, -1.0, -1.2, -1.4 and -1.6 kcal/mol
ID1-ID8	Integy moments: distances between the center of mass of the molecule and the center of the hydrophobic interaction regions calculated at the same 8 energy levels as D1-D8
HI-HL2	Hydrophilic-Lipophilic balances: ratio between the hydrophilic regions measured at -3.0 and -4.0 kcal/mol and the hydrophobic regions measured at -0.6 and -8.0 kcal/mol.
Descriptors Obtained from the Water and DRY Interaction Fields	
A	Amphiphilic moment: distance between the center of the hydrophobic domain and the center of the hydrophilic domain.
CP	Critical packing
Descriptors Obtained from the Carbonyl and Water Interaction Fields	
W10-W80	Volumes of the carbonyl molecule interaction fields at eight different energy levels: -0.2, -0.5, -1.0, -2.0, -3.0, -4.0, -5.0 and -6.0 kcal/mol
HB1-HB8	Hydrogen bond donor capability: differences between the volumes of the water interaction fields and the carbonyl interaction fields at 8 different energy levels: -0.2, -0.5, -1.0, -2.0, -3.0, -4.0, -5.0 and -6.0 kcal/mol
Descriptor Obtained Directly from the Molecular Structure	
POL	Polarizability

VolSurf descriptors calculation

The GRID [23, 24] 3D molecular interaction fields (MIFs) were calculated between the molecules in the dataset and three different probes (water, hydrophobic and carbonyl). MIFs were automatically converted into molecular descriptors using the standard procedure called VolSurf, which has been presented in detail elsewhere [12]. A grid box with a 0.5 Å spacing that extended 5 Å beyond the analyzed molecules was used.

The VolSurf method is simple to apply and is specifically designed to produce descriptors relevant to pharmacokinetic properties. These descriptors, which

refer to molecular size and shape, hydrophilic and hydrophobic regions, hydrogen bond donor/acceptor capacity, etc., are summarized in Table 2.

GRIND Independent Descriptors (GRIND) calculation

GRIND are alignment independent descriptors which represent favorable energy regions where groups of a potential protein would interact favorably with the ligand. GRIND generation involves three steps, all performed automatically by the program Almond: calculation of a set of molecular interaction fields, MIF filtering and transformation of the extracted MIF regions into the GRIND variables. This procedure is

Table 3. Physicochemical interpretation of the six correlograms in Almond analysis.

Correlogram no.	Probe 1	Probe 2	Interactions between nodes of type:
1	DRY	DRY	Hydrophobic
2	O	O	Hydrogen bond donor
3	N1	N1	Hydrogen bond acceptor
4	DRY	O	Hydrophobic and hydrogen bond donor
5	DRY	N1	Hydrophobic and hydrogen bond acceptor
6	O	N1	Hydrogen bond donor and hydrogen bond acceptor

reported in detail by Pastor et al. [13]. The MIFs were generated with three probes: the DRY probe to represent hydrophobic interactions, amide (N1) and carbonyl (O) probes representing respectively hydrogen bond donor and acceptor groups. All MIFs were computed using a grid spacing of 0.5 Å (grid unit) with the grid extending 5 Å beyond the molecule. The Almond procedure extracts from each MIF a fixed number of intense favourable (negative) energies of interaction (nodes). For this calculation, the maximum number of extracted nodes was set to 150 and the importance of the MIFs decreased to 35%. Six correlograms, containing 78 variables each, were calculated analyzing node–node interactions belonging to the same MIF (auto-correlograms) or to different pairs of MIFs (cross-correlograms). The explanation of the six correlograms obtained in this analysis is summarized in Table 3.

Chemometric analysis

Principal Component Analysis (PCA), Partial Least Squares (PLS) and Partial Least Squares Discriminant (PLSD) are chemometric tools for extracting and rationalizing the information from any multivariate model. The difference between PLS and PLSD analysis is that in the former analysis the dependent variable is a continuous variable, while in the latter case the dependent variable is transformed into a number of specific classes (generally two; e.g. stable/unstable or active/inactive). In our PLSD analysis, for the dependent variable two categorical scores were used, i.e. −1 and 1, indicating respectively the unstable and stable classes as defined in the Experimental section. PCA and PLSD have the advantage to reduce the

multivariate space into few principal components or latent variables (PCs or LVs) describing most of the variance in the data. These tools condense the overall information into two smaller matrices, namely the score plot (which shows the pattern of compounds) and the loading plot (which shows the pattern of descriptors). Because the chemical interpretation of score and loading plots is simple and straightforward, PCA and PLSD are usually preferred to other linear multivariate methods, especially when the noise is relatively high and the descriptors are not fully uncorrelated. Consensus PCA belongs to the hierarchical PCA and is useful when the descriptors are clustered in blocks. In this case block scores and block loadings are calculated, providing unique information about the importance of a particular block for principal component. Cross-validation techniques [25] have been applied to identify the number of significant components or latent variables and to validate internally the statistical models generated. These methods use part of the dataset for training and leave a number of samples out to be used during prediction. In this procedure, non-overlapping sets of samples are set aside and are predicted in turn, such that each sample is predicted once. PCA, CPCA and PLSD are least square methods and for this reason their results depend on data scaling. The initial variance of a column variable partly determines its importance in the model. In order to avoid this problem before analysis, VolSurf variables were centered and scaled to unit variance [26] and then the two blocks of descriptors, GRIND and Volsurf, submitted to blocked unscaled weights [20] (BUW). All applied chemometric tools are implemented in the GOLPE software.

Results

CPCA and PCA analysis of the training set

In a first analysis, CPCA was performed on VolSurf and GRIND descriptors computed for the training set of 1800 Pharmacia compounds. This set was originally selected to have a good exploration of the chemical space of our interest and an equal proportion of CYP3A4 stable and unstable compounds (900 stable and 900 unstable). The CPCA was done to evaluate the relevance and the relative contribution of each block of descriptors in discriminating molecules with *in vitro* metabolic stability or instability to the CYP3A4 enzyme. Two principal components explained respectively 47% and 41% of the variance of the VolSurf and

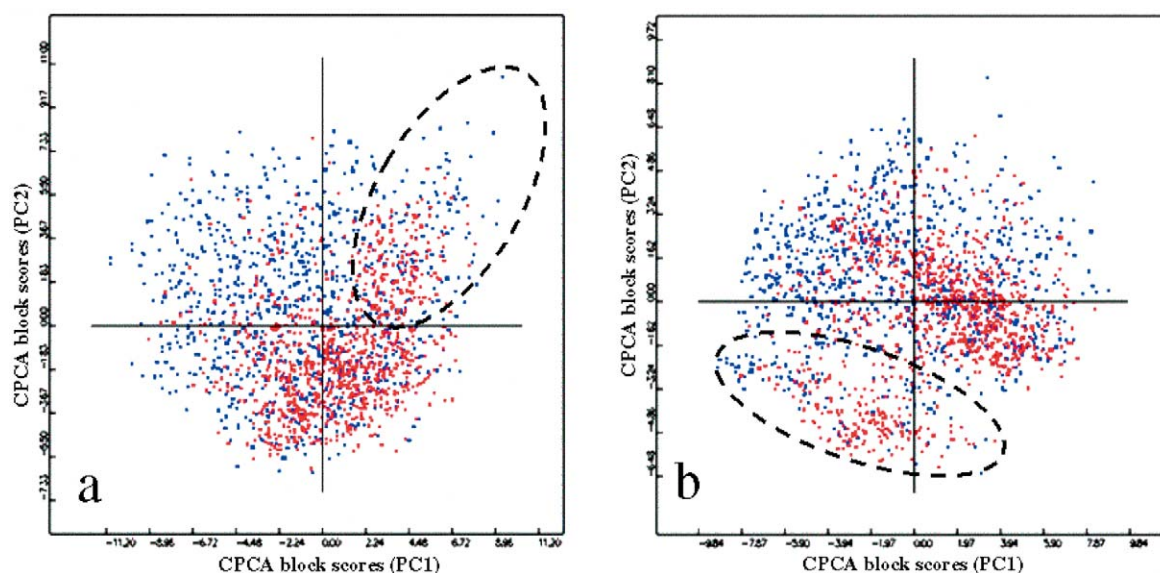


Figure 1. CPCA block scores plot of the training set of 1800 compounds for the calculated (a) VolSurf descriptors and (b) GRIND descriptors. Blue circles represent the CYP3A4 stable compounds and red circles the unstable compounds as defined in the Methods section. In each block score, the first two PCs cluster the compound according to their CYP3A4 metabolic stability class memberships. Compounds included in the ellipsoid are better clustered by GRIND (b) compared to VolSurf descriptors (a).

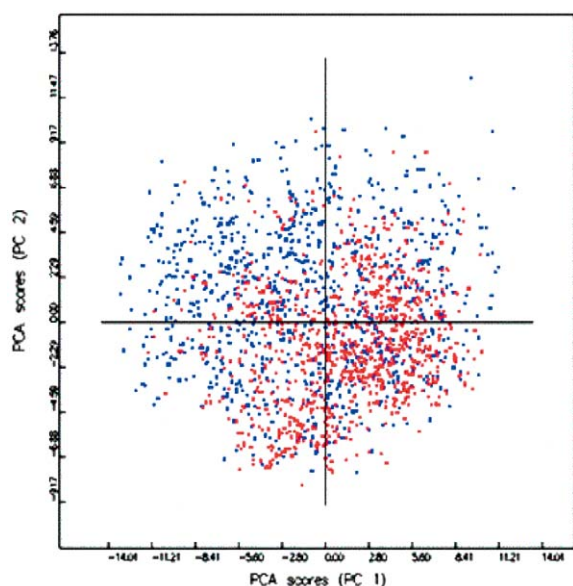


Figure 2. PCA score plot derived from the analysis of the VolSurf and GRIND descriptors calculated for the entire training set of 1800 compounds.

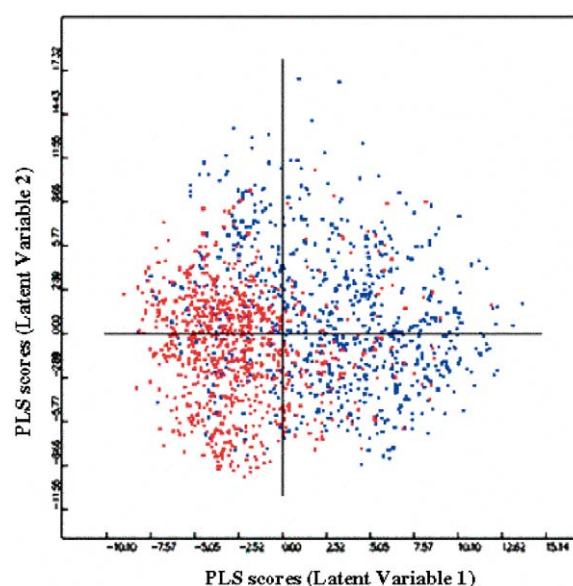


Figure 3. PLS discriminant t1-t2 score plot for the final model. Blue circles represent the CYP3A4 stable compounds and red circles the unstable compounds.

GRIND descriptor matrices. The block score plots for the first two PCs are reported in Figure 1a (referring to VolSurf block) and 1b (referring to GRIND block). Each symbol in the plots corresponds to a molecule in the training set and is color-coded by its associated

CYP3A4-stability class membership (blue stable, red unstable) as defined in the Methods section. It can be observed that the VolSurf descriptor block is able to discriminate between clusters of stable and unstable molecules (Figure 1a); however, a region in which

the two classes overlap is also present. In the GRIND block score plot a separation between the two groups of compounds is still notable (Figure 1b). Moreover, the latter descriptors seem to be more efficient to model a subset of CYP3A4 unstable compounds not so well described by the VolSurf block (see subset highlighted in Figure 1a and 1b). Indeed, the super weight for the first two PCs, which express the relative importance of each block in the overall model, shows that each set of descriptors are important for both principal components and, in turn, in discriminating stable and unstable molecules. VolSurf parameters seem to capture well the relevant molecular physical-chemical properties to reach and interact with the cytochrome active site. These descriptors have already been used extensively to model different ADME properties, like passive permeability through the gastrointestinal tract or blood-brain barrier and membrane partitioning of oligopeptides, etc. [12, 27–29]. Moreover, the orientation of a compound in the cytochrome active site and the subsequent reaction with the heme also implies the presence of chemical functions in specific positions and orientations. For describing the latter properties GRIND seems to be more suited, as this approach was designed for modeling ligand-receptor binding, as reported by Pastor et al. [13].

Once the importance of both sets of descriptors was confirmed, principal components analysis was performed on the entire matrix of variables calculated for the training set. Three significant principal components were found by a cross-validation technique. These components explained 51% of the total variance of the matrix. As already seen in CPCA, the first two principal components of the PCA model (Figure 2) distinguished relatively well the two classes of compounds. However, based on the assumption that compounds close in the descriptor space may also have similar properties, it was found that some compounds were misclassified. The result was still of high interest since neither classification of compounds nor any training information was given to the PCA model. A more detailed inspection of the misclassified compounds revealed that about 64% of the false negative and 47% of the false positive were characterized by an aqueous solubility at pH 7 of less than 5 μM . In this respect, there is a risk of experimental error if a compound is poorly soluble. For instance, a compound not in solution would not have the opportunity to interact with the CYP, so that it will likely result in a false negative. This situation is further complicated by the fact that acetonitrile is used to terminate the incubation

and may result in dissolution of previously insoluble compounds. Hence, the insoluble compounds (defined as compounds with an aqueous solubility at pH 7 less than 5 μM) were removed from the initial training set and the reduced set of 1556 molecules used to develop the final model.

PLS discriminant model generation and external validations

PLSD analysis was performed on the reduced set to identify a relationship between the calculated GRID-derived descriptors (72 VolSurf descriptors including polarizability and molecular weight and 424 GRIND) and the two CYP3A4 stability classes. This procedure is equivalent to those used by other researchers [30, 31] when training neural nets. However, in contrast to neural network procedure, we checked, as discussed previously, the existence of relevant information in the calculated descriptor matrix before building and training the model. The final PLS discriminant model, after three random-groups cross-validation, had two significant LVs and the PLS t1–t2 score plot is shown in Figure 3. Several parameters were calculated to evaluate the statistical significance of the model developed. The model showed a total accuracy of 81% (calculated as the total number of compounds correctly assigned to each class divided by the total number of compounds), interclass accuracy of 74%, for the stable class, (calculated as the ratio between true negatives and true negatives plus false positives) and 86% for the unstable class (calculated as the ratio between true positives and true positives plus false negatives) and a precision of 83% for the stable (calculated as the ratio between true negatives and true negatives plus false negatives) and 79% for the unstable classes (calculated as the ratio between true positives and true positives plus false positives). These parameters are reported in Table 4a.

To test the predictive value of the PLSD model, two external data sets not included in the original training set were used. Metabolic stability membership class evaluation of external sets is quite straightforward with the PLS discriminant; compounds predicted to have a score greater than 0 are classified as stable, whilst those with score less than 0 are classified as unstable. Predicted classes are then compared with the actual ones. The first test set, including 825 proprietary structures, was selected to cover the chemical space of our interest. In addition, only molecules with experimental aqueous solubility greater than 5 μM

Table 4. Accuracy analysis of the PLSD model based on GRID-derived descriptors.

(a) Cross-validated (three random groups) training set, 1556 compounds					
Actual class	Predicted class		% Interclass accuracy	% Precision	% Total accuracy
	Stable	Unstable			
Stable (48%)	554	188	74	83	80
Unstable (52%)	113	701	86	79	
(b) First test set, 825 compounds					
Actual class	Predicted class		% Interclass accuracy	% Precision	% Total accuracy
	Stable	Unstable			
Stable (62%)	344	168	67	91	75
Unstable (38%)	35	278	89	62	
(c) Second test set, 20 compounds					
Actual class	Predicted class		% Interclass accuracy	% Precision	% Total accuracy
	Stable	Unstable			
Stable (65%)	12	1	92	86	85
Unstable (38%)	2	5	71	83	

at pH 7 were included in the set. The PLSD model correctly classified 75% of the entire first set with an interclass accuracy of 67% for the stable and 89% for the unstable class and a precision of 91% for the stable and 62% for the unstable compounds. All results are summarized in Table 4b. The second test set consisted of 20 known drugs of different therapeutic classes; e.g. antidepressants, antipsychotics, hormones, anti-hypertensives, antiarrhythmics, anti-histamines, etc. The total accuracy in predicting the second test set was 85%, with 12 out of 13 experimentally stable compounds correctly classified (interclass accuracy of 92%). Of the 7 compounds unstable *in vitro*, the PLSD model correctly classified 5 of them (interclass accuracy of 71%). The precision in selecting acceptable (stable) compounds and eliminating unstable molecules was of 86% and 83% respectively (all results summarized in Table 4c). Predicted versus actual metabolic stability data and classes for the second test set are also reported in Table 5.

Discussion

The developed PLSD model has a very good predictive power when applied to external datasets. The precision of the model when used to select or bias libraries for acceptable (i.e. metabolic stable) compounds ranges from 86 to 91% and the accuracy for eliminating ‘unstable’ compounds ranged from 62 to 83%. The model is thus a valuable tool in library design. In general, we found that the total accuracy in predicting other independent datasets (results not shown) ranged within the limits of 75 and 85% derived from the two test sets, depending on the chemical classes analyzed. This is typical of the accuracy observed for statistical models developed on medium-throughput screening data. In the lead optimization phase, in which the research is mainly focused on specific chemical templates, more accurate local models may be preferable and can be developed if well distributed experimental metabolic stability data are generated in sufficient amount. This PLSD model has the advantage to be broadly applicable and valid for predicting compounds covering a relatively wide molecular space.

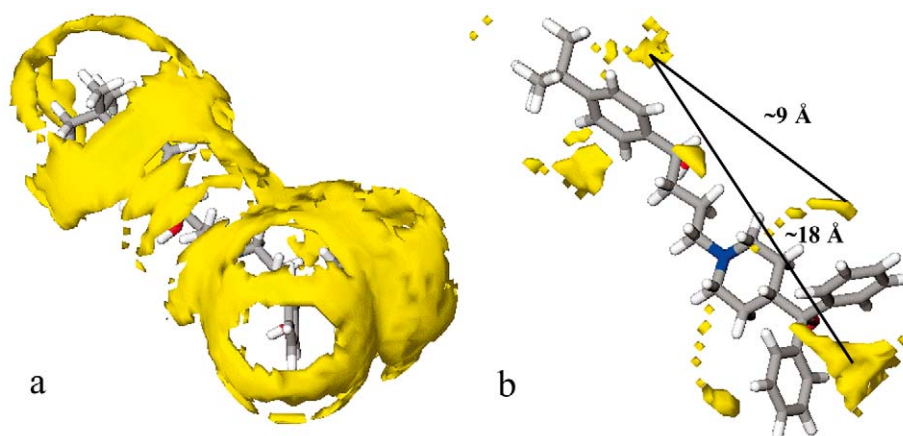


Figure 4. Graphical representations of some VolSurf and GRIND descriptors characterizing the CYP3A4-metabolic unstable terfenadine. (a) GRID 3D molecular interaction field calculated with a DRY probe. The zones are the hydrophobic regions contoured at -0.4 kcal/mol; (b) interaction between hydrophobic-hydrophobic regions at distances of ~ 9 Å and 18 Å.

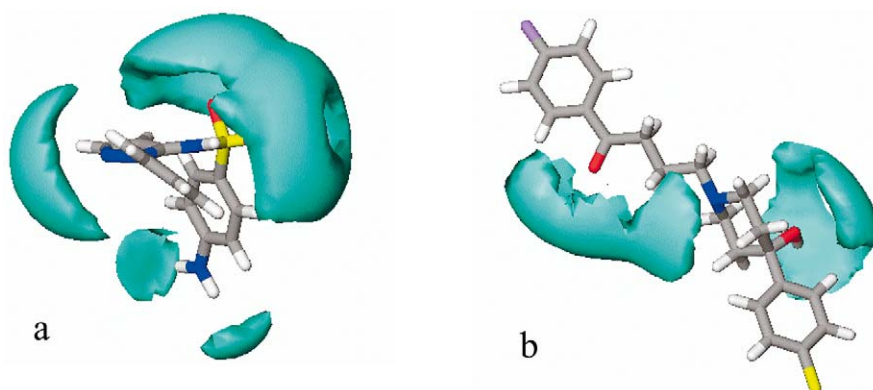


Figure 5. GRID 3D molecular interaction fields of sulfaphenazole (a) and haloperidol (b) calculated with a water probe at -3 kcal/mol.

After testing the predictive ability of the model it was also important to identify the most relevant molecular features that could modulate the presence of CYP3A4-mediated instability. The PLS coefficients that mostly contribute to explain the PLSD model are reported in Table 6. Based on the coefficients of the PLSD, unstable compounds are characterized by intense and wide hydrophobic interactions (D1-6 descriptors). The descriptors of the DRY autocorrelogram (DRYDRY from 23 to 44) represent the optimal distances (from ~ 9 Å to 18 Å) that should separate the aromatic interacting regions (or hydrophobic parts) of an unstable compound. For instance, terfenadine (unstable compound included in the second test set) is characterized by a wide hydrophobic region (Figure 4a) and aromatic interaction patches placed at the optimal distances as shown in Figure 4b. Amphiphilic moment (representing the distribution of

the polar and hydrophobic interaction regions), polarizability (A and POL), size and shape descriptors (VOL, S, MW, R, and G) are also descriptors positively correlated with unstable molecules. The presence of favourable interacting regions around an aromatic ring and a H-bond acceptor group separated by a distance of ~ 13 – 15 Å, also characterizes unstable compounds (indicated by the PLS coefficients of the DRYN1 descriptors in Table 6). Stable compounds are instead characterized by dense and localized polar regions (Iw1-8), larger polar surface per surface unit (Cw1-7) and high hydrophilic-lipophilic balances (HL1-2) in which polar regions dominate the hydrophobic part. As an example of the interpretative value of this model, Figures 5a and 5b illustrate a visual comparison of water molecular interaction fields calculated at -3 kcal/mol that is larger for the stable sulfaphenazole compound (5a), compared to the un-

Table 5. Predicted versus experimental classes and data for the second test set.

Name	Experimental % metabolic stability	Experimental class	Predicted class
Sulfaphenazole	96	S	S
Imipramine	90	S	U
Pramipexole	89	S	S
Propranolol	98	S	S
Desipramine	95	S	S
Propafenone	96	S	S
Reboxetine	78	S	S
Dextromethorphan	88	S	S
Dexamethasone	81	S	S
Tolbutamide	68	S	S
Cimetidine	81	S	S
Quinidine	61	S	S
Triazolam	56	S	S
Haloperidol	39	U	U
Midazolam	10	U	U
Nifedipine	13	U	S
Terfenadine	27	U	U
Fluphenazine	18	U	U
Bromocriptine	27	U	U
Salmeterol	4	U	S

stable haloperidol (5b). Positive and high values for the descriptors related to H-bond donor regions (WO3-7) are also indicative of stable compounds. In general, stable compounds seem also to be characterized by intense O-O (i.e. the regions around H-bond donor groups, see PLS coefficients of OO from 10 to 17 and OO from 30 to 34 in Table 6) and O-N1 (i.e. the regions around the H-bond donor and acceptor groups; see PLS coefficients of ON1 34-40 in Table 6) interaction energies at small (from ~ 4 Å to 7 Å) and medium (from ~ 12 Å to 16 Å) distances. These descriptors probably only reflect the general trend of stable compounds being more hydrophilic compared to the unstable ones, as these particular H-bond interaction regions would be unfavorable to interact with the CYP3A4 binding site. This interpretation, even if simplistic, captures most of the relevant information present in the model. Part of these findings is consistent with what is already known in the literature on CYP3A4 active site and on its substrates [32–36]. The CYP3A4 binding site is large and is constituted predominantly of hydrophobic amino acids. Some neutral amino acids are also present with a small number

Table 6. The PLS coefficients mostly contributing to explain the final PLD model developed on 1556 compounds.

PLS coeff.	Variable name	PLS coeff.	Variable name
−0.02736	D1	0.02376	W4O
−0.02678	POL	0.02314	W5O
−0.02625	VOL	0.02225	Cw2
−0.02516	MW	0.02171	Cw1
−0.02514	S	0.02110	Cw4
−0.02502	R	0.02070	Cw3
−0.02362	D2	0.02032	Cw5
−0.02166	G	0.02029	W6O
−0.01858	D3	0.01947	W3O
−0.01834	DRYDRY-35	0.01888	Iw8
−0.01800	DRYDRY-34	0.01783	Iw1
−0.01796	DRYDRY-33	0.01774	HL1
−0.01719	DRYDRY-36	0.01751	Iw7
−0.01704	DRYDRY-32	0.01748	Cw6
−0.01695	W1OH2	0.01730	OO-13
−0.01665	DRYDRY-31	0.01724	OO-14
−0.01653	W1O	0.01641	OO-15
−0.01608	DRYDRY-37	0.01593	OO-12
−0.01592	DRYDRY-30	0.01580	W7O
−0.01552	DRYDRY-29	0.01506	Iw2
−0.01531	A	0.01438	Iw6
−0.01504	DRYDRY-38	0.01438	Iw4
−0.01487	DRYDRY-28	0.01379	Iw3
−0.01434	DRYDRY-27	0.01378	Iw5
−0.01425	DRYDRY-39	0.01360	OO-16
−0.01337	D4	0.01331	HL2
−0.01305	DRYDRY-26	0.01292	OO-11
−0.01281	DRYDRY-40	0.01149	ON1-36
−0.01169	DRYDRY-25	0.01085	ON1-37
−0.01071	DRYDRY-41	0.01070	Cw7OH2
−0.01024	DRYDRY-24	0.01026	OO-17
−0.00936	DRYDRY-42	0.01020	ON1-35
−0.00880	D5	0.01007	ON1-38
−0.00879	DRYN1-33	0.00985	OO-10
−0.00876	DRYN1-36	0.00952	ID1
−0.00869	DRYN1-34	0.00848	OO-33
−0.00856	D6	0.00832	OO-32
−0.00846	DRYN1-37	0.00816	ON1-34
−0.00837	DRYDRY-23	0.00815	OO-34
−0.00829	DRYDRY-43	0.00810	OO-31
−0.00825	DRYN1-35	0.00804	ON1-39
−0.00818	DRYN1-32	0.00782	OO-30
−0.00765	DRYDRY-44	0.00780	ON1-40

of polar side chains. Thus, the presence of aromatic side chains should allow π - π interactions with aromatic substrates. Conversely, highly polar compounds should hardly reach the enzyme reaction site. These characteristics make the CYP3A4 binding pocket able to accommodate and interact preferably with neutral or basic, relatively large, hydrophobic molecules.

In conclusion, based on the results obtained, VolSurf and GRIND descriptors are shown to be relevant in discriminating molecules with different CYP3A4 metabolic stability properties. The combination of these two sets of descriptors, which accounts not only for the physicochemical (i.e. lipophilicity, hydrophilicity etc.) but also for the pharmacophoric characteristics of a compound, resulted in an accurate and interpretable model.

Quantitative structure-property relationships for predicting metabolic stability were recently reported in the literature [37]. These models, which were developed based on the metabolic turnover rates obtained in human liver S9 homogenates, i.e. preparations that express the whole range of CYPs, are characterized by a similar predictive power to ours if a comparative classification scheme is applied. Whilst their approach has the likely advantage of being more predictive of the *in vivo* situation, the use of metabolic stability data based on a single CYP allowed our model to have improved interpretability. Thus, the structural features responsible for the metabolic liability of a molecule can be identified, providing indications on how a candidate within a series can be modified to improve its stability.

Since some drugs classified as CYP3A4 unstable in this study are actually on the market, it could be argued that CYP3A4 stability should not be considered a critical property to be optimized. However, CYP3A4 drug instability has critical impact on the *in vivo* outcome that could result in undesirable high clearance, poor and/or dose-dependent bioavailability and possible drug-drug interactions, all causing problems in the design and optimization of dosages regimens. Therefore, the application of this computational model for producing libraries biased towards more stable compounds could efficiently speed up the identification of an enhanced drug.

Acknowledgements

We are grateful to Prof. Gabriele Cruciani (Laboratory for Chemometrics, University of Perugia, Italy)

for valuable discussions and advice and to Anna Moscone, Maddalena Gallina, Silvia Portolan, Flavio Cinato and Daniele Pezzetta for the experimental data.

References

1. Clark, D.E. and Pickett, S.D., *Drug Discov. Today*, 5 (2000) 49.
2. van de Waterbeemd, H., Smith, D.A., Beaumont, K. and Walker, D.K., *J. Med. Chem.*, 44 (2001) 1313.
3. van de Waterbeemd, H., *Curr. Opin. Drug Discov. Dev.*, 5 (2002) 33.
4. van de Waterbeemd, H. and Gifford, E., *Nat. Rev. Drug Discov.*, 2 (2003) 192.
5. Boobis, A., Gundert-Remy, U., Kremers, P., Macheras, P. and Pelkonen, O., *Eur. J. Pharm. Sci.*, 17 (2002) 183.
6. Guengerich, F.P., *Chem. Res. Toxicol.*, 14 (2001) 611.
7. Testa, B., In Wolff, M. (Ed) *Drug Metabolism*, in Burger's Medicinal Chemistry and Drug Discovery. Principles and Practice, Vol. 1. John Wiley & Son, New York, 1994, pp. 129–180.
8. Bertrand, M., Jackson, P. and Walther, P., *Eur. J. Pharm. Sci.* 11 (2 Suppl) (2000) S61.
9. Lewis, D.F.V., *Xenobiotica*, 28 (1998) 617.
10. Wrighton, S.A., Schuetz, E.G., Thummel, K.E., Shen, D.D., Korzekwa, K.R. and Watkins, P.B., *Drug Metab. Rev.*, 32 (2000) 339.
11. Thummel, K.E. and Shen, D.D., In Hardman, J.G., Limbird, L.E., Molinoff, P.B. and Gilman, A.G. (Eds), Appendix II. Design and Optimization of Dosage Regimens: Pharmacokinetic Data. In Goodman and Gilman's. The Pharmacological Basis of Therapeutics, 10th Edition, McGraw-Hill, New York, 2001, pp. 1917–2023.
12. Cruciani, G., Crivori, P., Carrupt, P.-A. and Testa, B., *Theochem*, 503 (2000) 17.
13. Pastor, M., Cruciani, G., McLay, I., Pickett, S. and Clementi, S., *J. Med. Chem.*, 43 (2000) 3233.
14. Singh, S.B., Shen, L.Q., Walker, M.J. and Sheridan, R.P., *J. Med. Chem.*, 46 (2003) 1330.
15. Zamora, I., Afzelius, L. and Cruciani, G., *J. Med. Chem.*, 46 (2003) 2313.
16. CORINA. Molecular Networks, GmbH, Computerchemie, Erlangen, Germany, 1998.
17. VolSurf v.3.0, Molecular Discovery Ltd., Oxford, UK, 2002.
18. Almond v.3.0, Multivariate Infometric Analysis S.r.l., Perugia, Italy, 2002.
19. Wold, S., Esbensen, K. and Geladi, P., *Chemom. Intell. Lab. Syst.*, 2 (1987) 37.
20. Kastenholtz, M.A., Pastor, M., Cruciani, G., Haaksma, E.E.J. and Fox, T., *J. Med. Chem.*, 43 (2000) 3033.
21. Dunn, W.J. and Wold, S., Pattern Recognition Techniques in drug design. In Hansch, C., Sammes, P.G. and Taylor, J.B. (Eds), *Comprehensive Medicinal Chemistry*, Vol. 4, Pergamon Press, Oxford, UK, 1990, pp. 691–714.
22. GOLPE 4.5, Multivariate Infometric Analysis S.r.l., Perugia, Italy, 2002.
23. GRID v. 20, Molecular Discovery Ltd., Oxford, UK, 2002.
24. Goodford, P.J., *J. Med. Chem.*, 28 (1985) 849.
25. Wold, S., *Technometrics*, 20 (1979) 379.
26. Wold, S., Albano, C., Dunn, W. J. III, Edlund, U., Esbensen, K., Geladi, P., Helberg, S., Johansson, E., Lindberg, W. and

- Sjostrom, M., Multivariate data analysis in chemistry. In Kowalsky, B.R. (Ed), *Chemometrics Mathematics and Statistics in Chemistry*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1983, pp. 17–96.
27. Cruciani, G., Pastor, M. and Guba, W., *Eur. J. Pharm. Sci.*, 11(Suppl. 2) (2000) S29.
 28. Crivori, P., Cruciani, G., Carrupt, P.-A. and Testa, B., *J. Med. Chem.*, 43 (2000) 2204.
 29. Alifrangis, L.H., Christensen, I.T., Berglund, A., Sandberg, M., Hovgaard, L. and Frokjaer, S., *J. Med. Chem.*, 43 (2000) 103.
 30. Ajay, W., Walters, P. and Murck, M.A., *J. Med. Chem.*, 41 (1998) 3314.
 31. Sadowski, J. and Kubinyi, H., *J. Med. Chem.*, 41 (1998) 3325.
 32. Guengerich, F.P., *Annu. Rev. Pharmacol. Toxicol.*, 39 (1999) 1.
 33. Lewis, D.F.V., Eddershaw, P.J., Goldfarb, P.S. and Tarbit, M.H., *Xenobiotica*, 26 (1996) 1067.
 34. Smith, D.A., Jones, B.C. and Walker, D., *Med. Res. Rev.*, 16(3) (1996) 243.
 35. Smith, D.A., Ackland, M.J. and Jones, B.C., *Drug Discov. Today*, 2(10) (1997) 406.
 36. Smith, D.A., Ackland, M.J. and Jones, B.C., *Drug Discov. Today*, 2(11) (1997) 479.
 37. Shen, M., Xiao, Y., Golbraikh, A., Gombar, V.K. and Tropsha, A., *J. Med. Chem.*, 46 (2003) 3013.