

Conformer- and Alignment-Independent Model for Predicting Structurally Diverse Competitive CYP2C9 Inhibitors

Lovisa Afzelius,^{*,†,‡,§} Ismael Zamora,^{||} Collen M. Masimirembwa,[†] Anders Karlén,[‡] Tommy B. Andersson,[†] Silvio Mecucci,[§] Massimo Baroni,[§] and Gabriele Cruciani[§]

DMPK and Bioanalytical Chemistry, AstraZeneca R&D, S-431 83 Mölndal, Sweden,

Department of Organic Pharmaceutical Chemistry, Uppsala University, Post Office Box 574, S-751 23 Uppsala, Sweden,

Lead Molecular Design, Valles 96-102, Loc 27, 08190 Sant Cugat de Valles, Barcelona, Spain, and

Laboratory for Chemometrics, Università degli Studi di Perugia, Via Elce di Sotto, 12 I-06100 Perugia, Italy

Received July 24, 2003

A conformer- and alignment-independent three-dimensional structure–activity relationship (3D-QSAR) model has been derived that is based on flexible molecular interaction fields calculated in GRID and the subsequent description of these fields by use of alignment-independent descriptors derived in ALMOND. The training set consisted of 22 diverse and flexible competitive inhibitors of the drug-metabolizing enzyme CYP2C9 and generated a model with r^2 of 0.81 and q^2 of 0.62. The predictive capacity of the model was externally evaluated with a test set of 12 competitive inhibitors and 11 out of 12 were predicted within 0.5 log unit. The most relevant points of interaction in the model correlated well to the amino acids involved in CYP2C9–substrate/inhibitor binding in the active site of a CYP2C9 homology model, further validating the mechanistic sense of our model. This approach offers the possibility to derive predictive 3D-QSAR models without the need for an alignment rule for chemically diverse ligands and in the absence of target protein crystal structure information.

Introduction

Computational methods are increasingly being used in all phases of drug discovery and development to understand ligand–receptor interactions. Their initial use was in understanding and improving pharmacological potency and selectivity. Lately, however, these methods have been extended to investigations of how new chemical entities interact with proteins important for ADME-Tox (Absorption, Distribution, Metabolism, Excretion–Toxicology) properties.¹ Many 3-D structure–activity relationship (3D-SAR) models have been proposed for the interaction of drugs with the major drug-metabolizing enzyme system, cytochrome P450 (CYP).^{2–16}

The success or failure of a 3D-SAR depends on a number of factors such as the quality of the target protein crystal structure or homology model, the accuracy with which one can describe the compound's bioactive conformer(s), and the ability to find superimposition rules of the bioactive conformer(s) consistent with ligand–protein interactions. Further development of the 3D-SAR to a quantitative model, 3D-QSAR, depends on additional factors such as quality of biological measures of ligand–receptor interaction (K_s , K_i , or K_m) and on finding chemical descriptors that are related to these constants. All these factors have proved challenging in understanding 3D-SAR and 3D-QSAR in predicting compounds that would be substrates or inhibitors of CYP450s.⁷ In our previous work, we have addressed some of these challenges by using a homo-

logy model of one of the human CYPs, CYP2C9, for conformer selection³ and by using an alignment-independent approach to derive a 3D-QSAR for CYP2C9 inhibitors.² In those studies, we also put emphasis on high-quality data by uniform derivation of K_i s and determination of the mechanism of inhibition.

The work on CYP2C9 benefited from the first reported mammalian CYP450 crystal structure, the CYP2C5 from the rabbit, with which it shows a high degree of similarity.¹⁷ Recently CYP2C9 has been crystallized (<http://www.astex-technology.co.uk.servlet/astex>), and a higher resolution of CYP2C5 cocrystallized with benzenesulfonamide (DMZ) has been published.¹⁸ Coordinates for these crystals or co-complexes with substrates and inhibitors will be invaluable in deriving good 3D-SAR models. The challenge of conformer selection still remains for CYPs for which there are no crystal structures or co-complexes, hence the need to develop methodologies that can simultaneously handle the lack of good protein structural information and the ligand diversity characteristic of the CYPs. In this study, we explored such an approach using competitive inhibitors of CYP2C9. We have used a data set of which we have extensive knowledge, so as to simultaneously validate our proposed approach.

Goodford¹⁹ recently published work on a GRID algorithm for flexible molecules to handle flexible protein side chains anchored to a protein backbone called the CORE. We adapted this method for small molecules and assigned the most rigid part as the CORE. The method does not select a bioactive conformer. Instead the conformational space for a ligand is explained as a probability-of-interaction map. The resulting GRID interaction fields are then analyzed by use of alignment-independent descriptors in ALMOND. An approach based on flexible ligand movements mimics the changes

* To whom correspondence should be addressed: e-mail Lovisa. Afzelius@astrazeneca.com; tel +46 31 776 21 06; fax +46 31 776 37 86.

[†] AstraZeneca R&D.

[‡] Uppsala University.

^{||} Lead Molecular Design.

[§] Università degli Studi di Perugia.

Table 1. Training Set^a

nicardipine (1) $K_i = 0.28$ 	sulphaphenazole (2) $K_i = 0.5$ 	dicoumarol (3) $K_i = 1.9$ 	<i>R,S</i> -fluvastatin (4) $K_i = 2.2$
<i>S,R</i> -fluvastatin (5) $K_i = 3.3$ 	progesterone (6) $K_i = 5.5$ 	phenytoin (7) $K_i = 6$ 	<i>S</i> -miconazole (8) $K_i = 6$
<i>R</i> -warfarin (9) $K_i = 13.6$ 	<i>R,S</i> -pyranocoumarin (10) $K_i = 16$ 	phenylbutazone (11) $K_i = 19$ 	<i>S</i> -warfarin (12) $K_i = 20$
quercetin (13) $K_i = 27$ 	quinine (14) $K_i = 32$ 	omeprazole sulfone (15) $K_i = 35$ 	<i>R</i> -rabeprazole (16) $K_i = 36$
<i>S</i> -rabeprazole (17) $K_i = 37$ 	pyrimethanine (18) $K_i = 51.5$ 	<i>S</i> -pantoprazole (19) $K_i = 64$ 	SFID (20) $K_i = 125$
<i>R</i> -pantoprazole (21) $K_i = 145$ 	thiabendazole (22) $K_i = 245$ 		

^a Experimental K_i values are given as micromolar.

that occur when a ligand binds to a protein, where energetically favorable effects such as dispersion and induction interactions and hydrogen bonding tend to compensate for any loss in energy due to the decrease in freedom. Application of this method to CYP2C9 inhibitors resulted in a predictive 3D-QSAR model. The model was further validated by comparison to the homology model of CYP2C9. The inhibitor-enzyme interactions were found to be consistent with the enzyme active-site geometry and chemistry.

Materials and Methods

Equipment and Software. Enzyme kinetic analysis was done with GraFit 4.0.12 (Erithacus Software Limited, Middlesex, U.K.) and SIMFIT 5.3.²³ Molecular interaction fields were calculated in an Irix environment on a Silicon Graphics O₂ workstation (Silicon Graphics Inc., Mountain View, CA) and in a Linux environment on a 32 Mb PC. The software utilized in the computational analysis was GRID v20 and ALMOND3.2.0 (Molecular Discovery Ltd., <http://www.moldiscovery.com>) and SYBYL 6.5.3 and CONCORD (Tripos Associates Inc., St Louis, MO). Chemical structures were imported from the ISIS-BASE database or drawn in ISIS-Draw (MDL Information Systems Inc., San Leandro, CA).

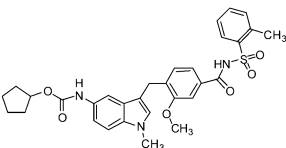
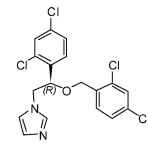
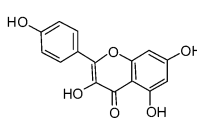
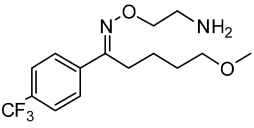
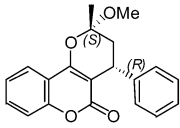
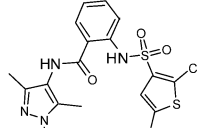
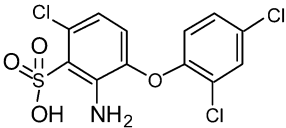
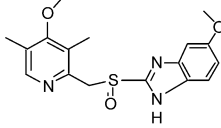
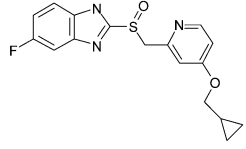
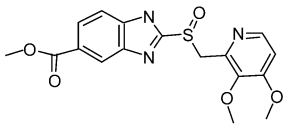
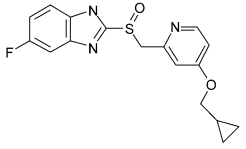
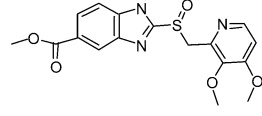
CYP2C9 Biological Data. Biological data were determined for a set of 34 inhibitors with K_i values ranging from 0.28 to 245 μ M as described by ref 3. By use of the CYP2C9-catalyzed diclofenac 4-hydroxylation, the mechanism of inhibition of the compounds was found to be competitive. The K_i data was randomly divided into a training set of 22 compounds (Table 1) and a test set of 12 different compounds (Table 2). The model

was validated by comparison with a previously reported homology model of CYP2C9.³

Conformer Independence by Calculating Flexible GRID Interaction Fields. GRID is a well-known methodology for calculating molecular interaction fields.²⁰ For a selected conformer, favorable sites of interaction are explored with different probes that are moved over a grid lattice that surrounds the molecule. The program takes into account the movement of tautomeric hydrogens, such as those in the imidazole ring of histidine, and the torsional rotation of aliphatic or phenolic hydroxyl or amino hydrogens but keeps the heavy atoms of the target rigid throughout the calculations. In a more recent version, a new option called flexible-GRID was added.¹⁹ Flexible-GRID takes into account the flexibility of some substituents and the associated movements of rigid groups at the end of those chains (e.g., the terminal group of an arginine side chain). The flexible algorithm was originally designed to handle amino acid side chains anchored to a protein backbone. In this study, we will explore its possible application to small molecules.

This flexible option allows both the probe and the target to respond to the environmental changes that occur upon moving the probe between the different grid points. For each grid point, the energetically optimal interaction of the flexible side chain is calculated on the basis of interactions with the probe and the entropic contribution due to the geometrical rearrangement of the side chain. The extent of the interactions with the probe is evaluated by electrostatic, van der Waals, and hydrogen-bond interactions after entropic effects are taken into account. The resulting map for each probe describes the most energetically favorable possibilities that a ligand has when it is allowed to adjust to the surroundings. In this way,

Table 2. Test Set^a

zafirlukast (22) $K_i = 2.5$ 	0 <i>R</i> -miconazole (23) $K_i = 6$. 	kaempferol (24) $K_i = 6.0$ 
fluvoxamine (25) $K_i = 8.5$ 	<i>S,R</i> -pyranocoumarin (26) $K_i = 16$ 	D-62126 (27) $K_i = 17.0$ 
A-4438 (28) $K_i = 20$ 	<i>R</i> -omeprazole (29) $K_i = 45.0$ 	<i>S</i> -H259/31 (30) $K_i = 60$ 
<i>S</i> -H287/23 (31) $K_i = 64$ 	<i>R</i> -H259/31 (32) $K_i = 93$ 	<i>R</i> -H287/23 (33) $K_i = 140$ 

^a Experimental K_i values are given as micromolar.

the method mimics the adjustments that occur when a ligand binds to a receptor.

GRID Setup. The molecules are assigned flexibility via the GRIN directive MOVE = 1. This is an automatic procedure that is applied to the .mol2 files as they are converted into .kout files, the GRID input format. Three different possibilities are considered for each atom: (1) assignment to the rigid CORE of the ligand, (2) assignment to a "BEAD" (which is also rigid but smaller in size than the CORE), or (3) assignment to flexibility (Figure 1, far left image of each panel). The CORE does not change its position in space during the calculations and each target can only have one CORE. A BEAD is allowed to move to an extent determined by the length and flexibility of the chains linking it to the CORE. Such movements can result in the BEAD changing its position and orientation with respect to the CORE in order to obtain an energetically favorable state. There can be many BEADs in each target molecule. The flexible atoms are allowed to move during the calculations and will be positioned in the most energetically favorable positions.

The GRID calculations were performed in a grid cage in increments of 0.5 Å. The number of planes per angstrom for the grid box (NPLA) was set to 2. The GRID directive set at MOVE = 3 was used to expand the grid cage for each individual compound to cover the increased chemical space into which flexible chains could extend. The DRY probe was used to describe hydrophobic interactions, the carbonyl (O) probe was used to evaluate the hydrogen-bond donor capacity of the target, and the amide (N1) probe was used to describe the hydrogen-bond acceptor properties. The ALMD directive, which memorizes the specific atoms influencing the main interactions, was set to 1.

Validating the Flexible Fields. For each compound, a random conformational search was performed in SYBYL with defined chirality. The Tripos force field was used for at least

1000 iterations to obtain a maximum of 100 conformers/compound. The conformers of each molecule were aligned in SYBYL by the CORE atoms that were defined in the flexible GRID calculations. This gave each conformer the same orientation with regard to the CORE so that each individual grid point could be compared between conformers. Thereafter, GRID fields for all conformers were calculated with exactly the same settings as for the flexible GRID calculations (DRY, N1, and O probes; 0.5 Å grid box size) except for the MOVE directive that was set to 0 (rigid mode). Finally, for each grid point, the most favorable energy obtained from any of the 100 conformers was put into a single "maximum interaction" file referred to as the merged file.

The shape of the GRID fields based on the merged rigid fields could then be compared to the field generated from the flexible search. A preliminary comparison was made on the basis of visual inspection of the GRID interaction fields to see if any major features were treated differently in the different cases (Figure 1). Alignment-independent descriptors were then computed in ALMOND²¹ and compared for the three different fields (DRY, O, and N1) by use of different similarity indices,²⁴ the Carbó index (cosine similarity coefficient), the Hodgkin index (Dice similarity coefficient), and the Tanimoto similarity coefficient (Jaccard coefficient). Euclidean distances were also calculated to evaluate the distribution in a PCA score plot based on the ALMOND descriptors.

Alignment Independence by Use of GRIND Descriptors in ALMOND. To obtain alignment independence, the information from the flexible GRID calculations was compressed into grid-independent descriptors by use of ALMOND.²¹ The degrees of freedom in the system are decreased by using descriptors derived from auto- and cross-correlograms. The interaction fields (Figure 2a) are described as the distances between the most favorable GRID interaction points, which are selected according to their energy of interaction and

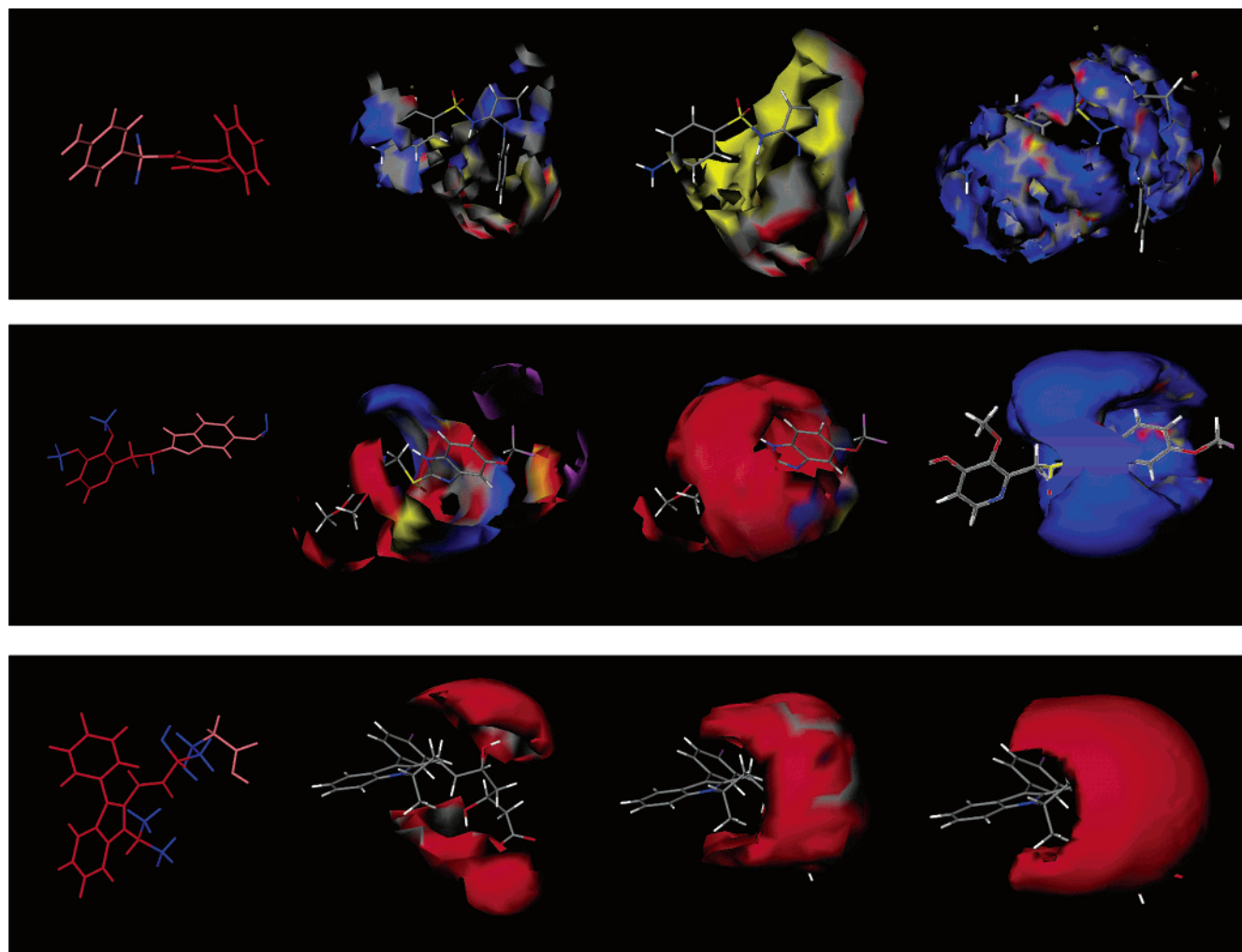


Figure 1. Molecular interaction fields for (top panel) sulfaphenazole, calculated by use of the DRY probe; (middle panel) (*R*)-pantoprazole, calculated by use of the O probe; and (bottom panel) (*3R,5S*)-fluvastatin, calculated by use of the N1 probe. For each compound, the image on the far left visualizes the flexibility assigned (CORE = red, BEAD = pink, and flexible = blue) for each compound. The following fields show (from left to right) the grid molecular interaction fields for a rigid compound with MOVE = 0, a flexible compound with MOVE = 3, and the merged fields based on 100 conformers, where each conformer has been described with MOVE = 0.

separation distance (Figure 2c). Only negative values are considered (Figure 2b), which results in some information being lost during the compression. A smoothing window was set as the default value of 0.8 grid unit. Since the gridbox is 0.5 Å, this generated bins of 0.4 Å each (0.8 × 0.5 Å). Distances between all grid points were calculated, and each distance was positioned within the correct bin. The most favorable energy product within each bin was selected and plotted against the distance, yielding the correlograms. Energies of interactions were derived from three different probes (DRY, O, and N1), generating the autocorrelograms explaining the DRY–DRY, O–O, and N1–N1 interactions and the cross-correlograms explaining the DRY–O, DRY–N1, and O–N1 interactions (Figure 2e). The selection of points was made on the basis of two criteria: the highest energy level and the greatest diversity according to distance in space with each parameter given equal importance (50%/50%). The number of points selected in the filtering process can be optimized for the individual data set.^{2,21}

When the flexible GRID interaction fields were calculated, certain defined side chains or single atoms were allowed to move, which generated more extended fields of interaction compared to the typical (rigid) GRID interaction fields. To better describe the interactions available for a compound, a directive called ALMD has been introduced in GRID. ALMD registers the atom from which the largest contribution to the energy originates for each selected point (Figure 2b). The

identity of that atom is stated in the output file and can be visualized in the GRID package visualization interface GVIEW. When the ALMOND program derives the grid-independent descriptors, distances between points originating from the same atom according to the ALMD directive are discarded. Distances within the same field will thus not be reselected. For example, the hydrogen in a hydroxyl group can only be donated once. In other instances, it could be argued that certain atoms can give rise to dual interactions such that the two unshared electron pairs of a hydroxyl oxygen could accept one hydrogen bond each. The benefit of discarding this possible interaction is, however, considered greater than the loss when disregarding cases such as that for the unshared electron pair.

ALMOND Model Derivation. The flexible fields were imported into ALMOND. The default values were used except that point selection was biased to select points based on higher energy rather than the diversity of the points in space (75%/25%). The number of points selected was set to 50 since this generated the most predictive model. The number of planes per angstrom for the grid box (NPLA) was set to 2. NOR scaling was applied to normalize the values within each block between 0.5 and 2 by use of the maximum value of product of interaction found within the block (Figure 2f). The model was built from the 22 compounds in the training set and their distribution in space, and the variance was examined in a PCA score plot together with the test set (Figure 3). Thereafter the *y*-variable was introduced and a PLS analysis was performed.

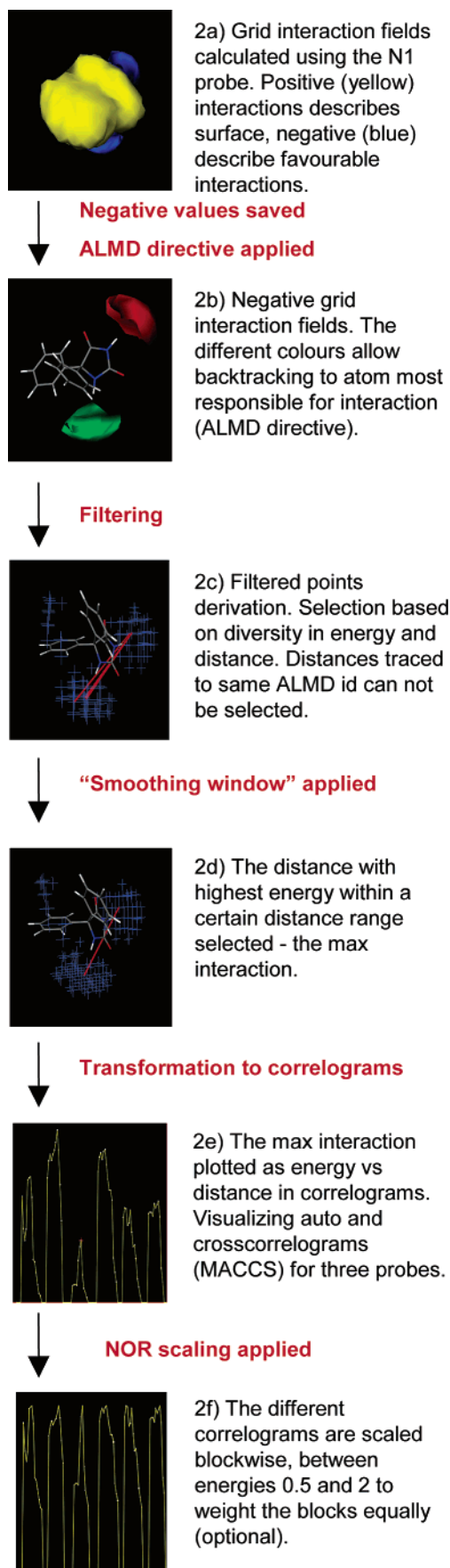


Figure 2. Flowchart describing the derivation of alignment-independent descriptors in ALMOND.

A fractional factorial design was applied, where all uncertain

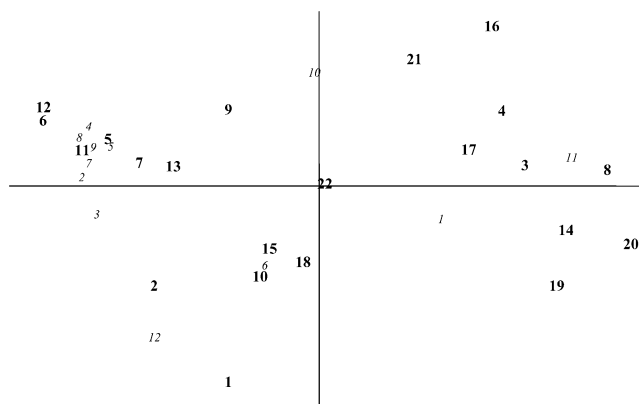


Figure 3. x -Variance of the ALMOND descriptors, visualized in a PCA score plot for the first and second component. The training set is printed in boldface type and the test set is shown in italic type, with the numbering from Tables 1 and 2. The plot show that the test set is spread over the range of the training set.

variables were removed. The model was then externally validated with a test set of 12 molecules.

Model Interpretation. Further validation of the model was made by comparison to a homology model of CYP2C9.³ A CYP2C9 substrate, phenytoin, was docked into the active site by use of GOLD. Selection of docked solutions was based on knowledge of phenytoin's orientation in the active site and its limited conformational freedom. A solution with a favorable distance and orientation between the site of oxidation and the iron in the heme group ($\sim 4 \text{ \AA}$) was chosen. The Flexible-Grid/ALMOND model was recomputed with the conformer from the active-site docking procedure. This generated an identical model with one compound, phenytoin, aligned to the protein, which enabled a direct comparison of the pharmacophoric distances to the active site.

The correlogram energies for phenytoin were multiplied by the PLS coefficients of the model to identify distances that were the main contributors to the activity of this compound. The GRID interaction fields were then calculated for the protein with the exact same settings as for the pharmacophore (DRY, N1, and O probes, NPLA = 2, MOVE = 1). The most favorable distances were then visualized in the protein together with the GRID interaction fields calculated for the active site.

Results and Discussion

CYP2C9 Biological Data. The measured inhibition constants for the present data set are given in Tables 1 and 2. They describe a large chemical space³ and include rigid and flexible compounds, some with more than 20 rotatable bonds. It is thus a challenging data set to model, but characteristic of CYP450 substrates and inhibitors.

Conformer Independence by Use of Flexible Fields. Since the flexible GRID algorithm was originally designed to handle amino acid side chains anchored to a protein backbone, an important part of our present study was to evaluate its use for small molecules.

The automatically assigned flexibility was checked for all molecules to ensure that the algorithm treated similar cases in a comparable way. The flexibility assigned to 21 CONCORD-generated structures showed compound class consistency. Compounds belonging to the same compound class, e.g., omeprazole derivatives, were all assigned the same degree of flexibility.

The flexible fields cannot be interpreted in the same way as rigid GRID interaction fields. The flexible fields

Table 3.

		MW	rot. bonds	Tanimoto ^a	Hodgkin ^b	Carbó ^c	Euclidean
1	(3 <i>R</i> ,5 <i>S</i>)-fluvastatin sodium	411	8	0.76	0.86	0.88	2.57
2	(3 <i>S</i> ,5 <i>S</i>)-fluvastatin sodium	411	8	0.63	0.77	0.84	4.05
3	(<i>R</i>)-rabeprazole	359	8	0.78	0.87	0.87	1.82
4	(<i>S</i>)-rabeprazole	359	8	0.73	0.84	0.85	2.41
5	(<i>S</i>)-pantoprazole	383	7	0.54	0.70	0.79	5.09
6	(<i>R</i>)-pantoprazole	383	7	0.55	0.71	0.79	4.88
7	(<i>S</i>)-miconazole	416	6	0.44	0.61	0.70	3.71
8	phenylbutazone	308	5	0.86	0.92	0.95	0.71
9	omeprazolesulfone	361	5	0.68	0.81	0.83	2.08
10	(<i>R</i>)-warfarin	308	4	0.81	0.90	0.90	1.79
11	(<i>S</i>)-warfarin	308	4	0.84	0.91	0.92	1.91
12	quinine	324	4	0.61	0.76	0.85	3.78
13	sulfaphenazole	314	3	0.56	0.72	0.76	5.40
14	(3 <i>R</i> ,13 <i>S</i>)-pyranocoumarin	322	3	0.81	0.90	0.90	1.03
15	dicoumarol	336	2	0.84	0.91	0.92	1.59
16	phenytoin	252	2	0.77	0.87	0.90	2.03
17	pyrimethamine	249	2	0.87	0.93	0.93	1.65
18	quercetin	302	1	0.88	0.94	0.95	1.85
19	progesterone	314	1	0.69	0.82	0.84	1.45
20	SFID	250	0	0.73	0.84	0.86	1.96
21	thiabendazole	203	0	0.74	0.85	0.86	1.29
	average	328	4	0.72	0.83	0.86	2.53

^a Range of the similarity coefficient was from -0.33 to 1. ^b Range of the similarity coefficient was from -1 to 1. ^c Range of the similarity coefficient was from 0 to 1.

are, in general, smoother and more extended since flexible side chains may be able to stretch out in several directions from the CORE of the molecule. When all the lowest detectable interactions are visualized the fields are almost spherical. Both rigid and flexible contours were displayed at positive and negative energies, and they were studied by use of the program GVIEW. The rigid positive contours showed the arbitrary shape of each molecule in the particular rigid conformation chosen for display on GVIEW. However, the flexible positive contours outlined the shape of the rigid core, and this was, in itself, informative since the rigid CORE is one part of a molecule whose shape cannot adapt to protein active-site interaction constraints.

To evaluate how well the flexible algorithm could generate an overview of the interactions available to a molecule over its conformational space, the flexible field maps were compared to the merged field of 100 conformers of the same compound. These fields should not be identical since they do not represent exactly the same information. For any particular rigid conformer, the interactions at each grid point are not optimized as they are in the flexible algorithm. Instead the most favorable interaction for each grid point is taken from one or another of the generated conformers. This particular geometry does not have to be presented to the same grid point in the flexible fields since another, possibly more favorable, conformation might be available. Nevertheless, comparing the overall shape of these fields should indicate whether the new flexible method can describe the conformational space comparable to 100 individual runs.

The validation of the flexible fields generated for small molecules was accomplished in the following steps: First, the overall properties of the fields were inspected visually (Figure 1). The overall pattern was in good agreement for most compounds. In some cases (Figure 1, middle panel), different atoms seemed to be responsible for the most important energies of interaction.

Second, alignment-independent descriptors calculated for the flexible fields and the merged fields were compared via different measures of similarity. The

similarities between descriptors generated from both fields, in this case, alignment-independent ALMOND descriptors, should correlate well since they extract the most valuable information from the fields. The results from the analysis are presented in Table 3. The similarity was high for all analyses with an average of 0.86 ± 0.06 (Carbó), 0.83 ± 0.09 (Hodgkin), and 0.72 ± 0.12 (Tanimoto). None of the similarity indices reflected changes in flexibility based on the number of rotational bonds. This clearly shows that the differences do not originate from changes in flexibility but reflect the differences in calculated energy levels in the two methods. The compounds were also evaluated in a PCA score plot and the Euclidean distances between them were determined in a four-dimensional space (Table 3). The calculations based on merged versus flexible fields do not group together in the PCA plot. Instead, for the compounds that are explained by the first two components the results for the flexible fields are often close to their merged fields in space.

Quantitative Model. The flexible fields were calculated in GRID for the entire data set of 34 compounds. The flexibility assigned to the molecules was carefully checked and shown to be consistent throughout the data set. These fields were then imported into ALMOND where the ALMD directive was turned on.

The training set consisted of 22 compounds (Table 1). The competitive inhibition constant, K_i , was introduced as the y -variable. In the PLS analysis that followed, the model was refined by a single factorial design variable selection that removed uncertain variables and generated a predictive model explained by two components ($r^2 = 0.81$ and $q^2 = 0.62_{\text{LOO}}$, $q^2 = 0.59_{\text{randomgroups}}$) (Figure 4). The model was externally evaluated with a data set of 12 compounds (Table 2) with K_i values ranging from 2.5 to 140 μM . The test set was representative of the PCA space described by the model according to the PCA plot (Figure 3). The predictions yielded an SDEP value of 0.42. In the validation set, nine out of 12 compounds were predicted within 0.3 log unit and an additional two

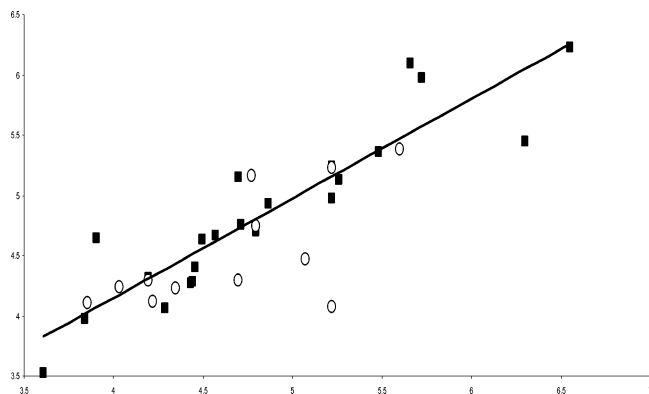


Figure 4. Correlation between model and predicted biological data in logarithmic units, with $r^2 = 0.8$ and $q^2 = 0.6$. (■) Training set; (○) test set.

within 0.5 log unit. One compound, kaempferol, had an error of 0.97 log unit.

Model Interpretation. The most important distances (according to the model) for phenytoin were visualized inside the protein together with the GRID interaction fields calculated for the protein. Since the pharmacophoric points are derived from the most favorable positions for the interacting GRID probe, these points should be positioned at the same coordinates as interacting amino acids. Exploring the dry interaction fields together with points derived from auto- and cross-correlograms revealed three distinct patches, one showing a direct overlap to Ala297 and the second and third within 1 Å distance from Phe114 and Leu366, respectively (Figure 5). According to experimental data, Phe114 is a major site of interaction in this isoform²² and the other amino acids are likely to have an importance for phenytoin affinity to CYP2C9 and could be suggested as targets for site-directed mutagenesis studies. Another hydrophobic patch was found pointing toward Phe476, which has also been shown to be important for binding.²⁰ This amino acid is, however, too distant (5 Å) in its current conformation for significant interactions with phenytoin, but the flexible interaction fields show that the side chain could move and facilitate binding. Based on the model, there are also two hydrogen-bond acceptor areas for interaction that

are important for the activity of the compounds. One of the areas corresponds best to Asp293 (3.77 Å) and a backbone interaction with Val292 (2.80 Å), while the other one is in proximity to Glu300 (3.62 Å). Neither the flexible side chain of Glu300 nor that of Asp293 points in a favorable orientation toward the pharmacophore. The flexible fields calculated for the active site, however, clearly show the possibility of an interaction, demonstrating that there might be room for the side chains to adapt different conformations.

Conclusion

By using this novel approach, neither a bioactive conformer nor an alignment rule needs to be defined. This way, initial and possibly erroneous assumptions in modeling can be avoided. This method is of particular value in cases where the lack of biological data prevents one from making any valid assumptions about possible bioactive conformations and where structural diversity hampers alignment assignment. The validation shows that these new flexible molecular interaction fields can explain the conformational space described by 100 conformers.

An internally predictive model has been derived that is externally able to predict other flexible compounds based only on the alignment-independent descriptors derived from the conformer-independent flexible interaction fields generated by GRID. A particularly important advantage is the way in which one can backtrack all the interactions to the original filtered GRID fields.

Several compounds predicted by the model are clinically important inhibitors of CYP2C9, such as miconazole, warfain, and zafirlukast, indicating a potential for a wider usage.

Furthermore, the impressive correlation between the pharmacophoric points and amino acids in the active site, some of which are known to be involved in the interactions, show that the model corresponds very well with experimental evidence.

Acknowledgment. We thank Dr. Peter Goodford and Dr. Richard Thompson for valuable comments and critical reading of the manuscript.

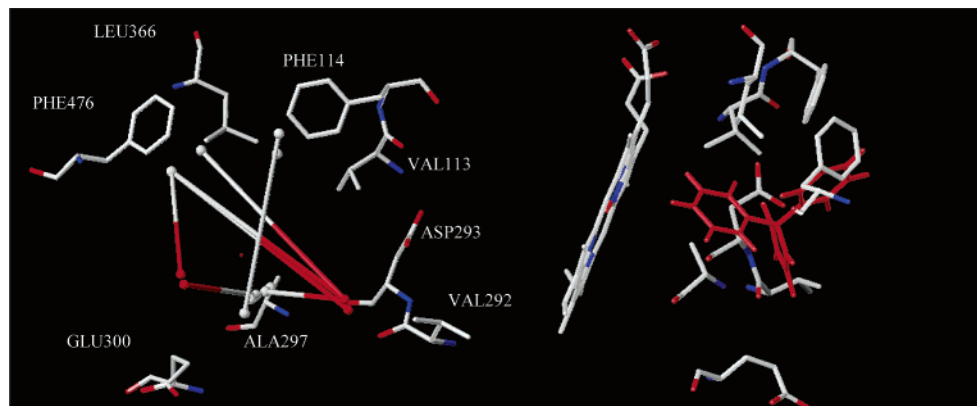


Figure 5. Docking of phenytoin into the homology model of CYP2C9. (Left) In gray are the most favorable hydrophobic interaction points, and in red are the most important polar interaction points. They are derived on the basis of the most important interactions for phenytoin according to the model. The heme group (not visualized) provides the rear wall in the projection. (Right) Docked solution of phenytoin used for the projection of pharmacophore distances into the model. The illustration is rotated 90° compared to the left picture, and here the heme group make up the left wall of the active site.

Note Added after ASAP Posting. This manuscript was released ASAP on 1/13/2004 with errors in the designation of Glu300 in the Model Interpretation paragraph before the Conclusion. The correct version was posted on 1/22/2004.

References

- (1) van de Waterbeemd, H. High-throughput and in silico techniques in drug metabolism and pharmacokinetics. *Curr. Opin. Drug Discovery Dev.* **2002**, *5* (1), 33–43.
- (2) Afzelius, L., et al. Discriminant and quantitative PLS analysis of competitive CYP2C9 inhibitors versus noninhibitors using alignment independent GRIND descriptors. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 443–458.
- (3) Afzelius, L., et al. Competitive CYP2C9 inhibitors: enzyme inhibition studies, protein homology modeling, and three-dimensional quantitative structure–activity relationship analysis. *Mol. Pharmacol.* **2001**, *59* (4), 909–19.
- (4) De Groot, M. J., et al. A Novel Approach to Predicting P450 Mediated Drug Metabolism. CYP2D6 catalyzed N-dealkylation Reactions and Qualitative Metabolite Predictions Using a Combined Protein and Pharmacophore Model for CYP2D6. *J. Med. Chem.* **1999**, *42*, 4062–4070.
- (5) de Groot, M. J., et al. Novel approach to predicting P450-mediated drug metabolism: development of a combined protein and pharmacophore model for CYP2D6. *J. Med. Chem.* **1999**, *42* (9), 1515–24.
- (6) de Groot, M. J.; Alex, A. A.; Jones, B. C. Development of a combined protein and pharmacophore model for cytochrome P450 2C9. *J. Med. Chem.* **2002**, *45* (10), 1983–93.
- (7) de Groot, M. J.; Ekins, S. *Pharmacophore modeling of cytochromes P450. Advanced Drug Delivery Reviews.* **2002**, *54* (3), 367–83.
- (8) De Groot, M. J.; Vermeulen, N. P. Modeling the active sites of cytochrome P450s and glutathione S-transferases, two of the most important biotransformation enzymes. *Drug Metab. Rev.* **1997**, *29* (3), 747–99.
- (9) Ekins, B. G.; W. S. Three and four dimensional-quantitative structure activity relationship analyses of CYP2D6 inhibitors. *Pharmacogenetics* **1999**, *9*, 477–489.
- (10) Ekins, S., et al. Three and four dimensional-quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2D6 inhibitors. *Pharmacogenetics* **1999**, *9* (4), 477–89.
- (11) Ekins, S., et al. Three- and four-dimensional-quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2C9 inhibitors. *Drug Metab. Dispos.* **2000**, *28* (8), 994–1002.
- (12) Ekins, S., et al. Three-dimensional quantitative structure activity relationship analyses of substrates for CYP2B6. *J. Pharmacol. Exp. Ther.* **1999**, *288* (1), 21–9.
- (13) Ekins, S.; de Groot, M. J.; Jones, J. P. Pharmacophore and three-dimensional quantitative structure activity relationship methods for modeling cytochrome p450 active sites. *Drug Metab. Dispos.* **2001**, *29* (7), 936–44.
- (14) Ekins, S.; Obach, R. S. Three-dimensional quantitative structure activity relationship computational approaches for prediction of human in vitro intrinsic clearance. *J. Pharmacol. Exp. Ther.* **2000**, *295* (2), 463–73.
- (15) Jones, J. P., et al. Three-dimensional quantitative structure–activity relationship for inhibitors of cytochrome P4502C9. *Drug Metab. Dispos.* **1996**, *24* (1), 1–6.
- (16) Rao, S., et al. A refined 3-dimensional QSAR of cytochrome P450 2C9: computational predictions of drug interactions. *J. Med. Chem.* **2000**, *43* (15), 2789–96.
- (17) Williams, P. A., et al. Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol. Cell* **2000**, *5* (1), 121–31.
- (18) Wester, M. R., et al. Structure of a Substrate Complex of Mammalian Cytochrome P450 2C5 at 2.3 Å Resolution: Evidence for Multiple Substrate Binding Modes. *Biochemistry* **2003**, *42* (21), 6370–6379.
- (19) Goodford, P. J. Atom movement during drug–receptor interactions. *Rational Mol. Des. Drug Res.* **1998**, *42*, 215–230.
- (20) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
- (21) Pastor, M., et al. GRIND-INdependent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. *J. Med. Chem.* **2000**, *43* (17), 3233–43.
- (22) Melet, A., et al. Substrate selectivity of human cytochrome P450 2C9: importance of residues 476, 365, and 114 in recognition of diclofenac and sulfaphenazole and in mechanism-based inactivation by tienilic acid. *Arch. Biochem. Biophys.* **2003**, *409* (1), 80–91.
- (23) Bardsley, W. G., Ed; *SIMFIT. A computer package for simulation, curve fitting and statistical analysis using life science models*, 5.3 ed.; Modern Trends in Biothermokinetics; Schuster, S., et al., Eds.; Plenum Publishing Corp.: New York, 1993; pp 455–458.
- (24) Leach, A. R. *Molecular Modelling—Principal and Applications*, 2nd ed.; Pearson Education Limited: Essex, U.K., 2001.

JM030972S