Analysis of Selective Regions in the Active Sites of Human Cytochromes P450, 2C8, 2C9, 2C18, and 2C19 Homology Models Using GRID/CPCA

Marianne Ridderström,*,† Ismael Zamora,† Ola Fjellström,‡ and Tommy B. Andersson†

Departments of DMPK & Bioanalytical Chemistry and Medicinal Chemistry, AstraZeneca R&D Mölndal, Mölndal, Sweden

Received April 27, 2001

This study demonstrates a selectivity analysis using the GRID/CPCA strategy on four human cytochrome P450 2C homology models (CYP2C8, 2C9, 2C18, and 2C19). Although the four enzymes share more than 80% amino acid sequence identity, the substrate specificity differs. To investigate the selectivity of the enzymes and the amino acids that determine the specificity of each CYP2C enzyme, a selectivity analysis was made using GRID/CPCA. In the GRID calculations 10 probes were used covering hydrophobic, steric, and hydrogen bond acceptor and donor interactions. The selectivity analysis showed that the most important determinants of selectivity among the CYP2C models are the geometrical features of the active sites and the hydrophobic interactions. The selectivity analysis singled out CYP2C8 as the most different of the four CYP2C enzymes with amino acids with distinct properties in positions 114, 205, and 476 (Ser, Phe, and Ile, respectively) compared to the other enzymes. An inverse pharmacophore model for CYP2C9 was constructed from the selective regions, and the model agreed with the docking of diclofenac where the properties of the ligand overlapped with the pharmacophoric points in the model.

Introduction

The cytochrome P450 enzymes (CYPs) play a central role in the metabolism of a wide variety of xenobiotics including clinically important drugs. The drug metabolizing CYPs are mainly located in the liver, and the CYP2C subfamily accounts for about 18% of the total CYP content.[1] The different CYP enzymes have a broad specificity for different chemical structures, which could be both substrates and inhibitors.[1−3] Therefore, a drug can inhibit the metabolism of another concomitantly administrated drug, which may lead to increased plasma levels and serious side effects. The CYP2Cs substrate warfarin, an anticoagulantium, is sensitive to drug interactions, which may generate serious bleeding complications as side effects. Hence, the possibility of predicting inhibition and metabolism by human CYPs is important in the discovery and development of new drugs.

Although CYP2C8, 2C9, 2C18, and 2C19 share high sequence identity (more than 80%), they have rather distinct substrate specificity.[3,4] Typical substrates for CYP2C9 are diclofenac (Figure 1a),[5] tolbutamide, and *S*-warfarin (Figure 1b).[6] Specific substrates for CYP2C19 involves *R*-omeprazole and *S*-mephenytoin.[3] Overlap of substrate specificity can be seen among the members of the CYP2C subfamily such as the metabolism of diclofenac and warfarin, although the warfarin stereoisomers are metabolized differently by different CYP2C enzymes (Figure 1b). Recently, a few drugs that are substrates for CYP2C8 were published, i.e., paclitaxel, troglizone, and zopiclone.[7−9] So far, no drugs that are substrates for CYP2C18 have been reported.[10]

Different approaches such as 3D-QSAR and pharmacophore modeling have been used to predict inhibition and metabolism of compounds. Protein modeling of human CYPs has so far been difficult because the templates available are of prokaryotic origin and have low amino acid sequence identity to human CYPs. The most commonly used template, the CYP102 crystal structure from *Bacillus megaterium*,[11−12] shares about 20% amino acid sequence identity to the human CYP2C subfamily.

The CYP2C5 crystal structure from rabbit is the first mammalian CYP structure solved.[13] This enzyme was modified via deletion of residues 3−21 in the N-terminal membrane binding domain and via site-directed mutagenesis of five amino acids (D2A, Q22K, N23T, G25S, and R26K) in the N-terminus as well as site-directed mutagenesis of five amino acids (N202H, R206E, I207L, S209G and S210T) to allow expression, purification, and crystallization. This engineered enzyme has one-third of the 21-hydroxylating activity of progesterone compared to the CYP2C5 wild-type. The amino acid sequence identity between CYP2C5 and the human CYP2C8, 2C9, 2C18, and 2C19 varies between 77% and 83% (Table 1). The high degree of sequence similarity makes the CYP2C5 crystal structure a more suitable template than the previously used CYP102 enzyme from *B. megaterium*. Recently, a 3D-QSAR analysis based on a homology model of CYP2C9 was used for compound docking to guide the conformer selection and alignment of inhibitors. The obtained model was able to predict an external data set within 0.5 log units of the experimental K_i values.[14]

Presented here are the first protein homology models of CYP2C8, 2C18, and 2C19, in addition to the previously published CYP2C9 model, based on the mammalian CYP2C5 crystal structure. The CYP2C models

* To whom correspondence should be sent. E-mail: Marianne.Ridderstrom@astrazeneca.com. Phone: +46 31-776 27 23. Fax: +46 31-776 37 86.
† Department of DMPK & Bioanalytical Chemistry.
‡ Department of Medicinal Chemistry.

10.1021/jm0109107 CCC: $20.00
Published on Web 10/20/2001

Table 1. Amino Acid Sequence Identities and Similarities between CYP2C5 and the Modeled Human CYPs[a]

	CYP2C5	CYP2C8	CYP2C9	CYP2C18	CYP2C19	
CYP2C5	*1.00*	0.79	0.83	0.77	0.83	relative identity (above the
CYP2C8	0.79	*1.00*	0.82	0.77	0.83	diagonal of italic numbers)
CYP2C9	0.83	0.82	*1.00*	0.82	0.93	
CYP2C18	0.83	0.82	0.87	*1.00*	0.81	
CYP2C19	0.83	0.83	0.93	0.87	*1.00*	
	relative similarity (below the diagonal of italic nubmers)					

[a] BestFit in the GCG program package, with default parameters, was used for the alignments.

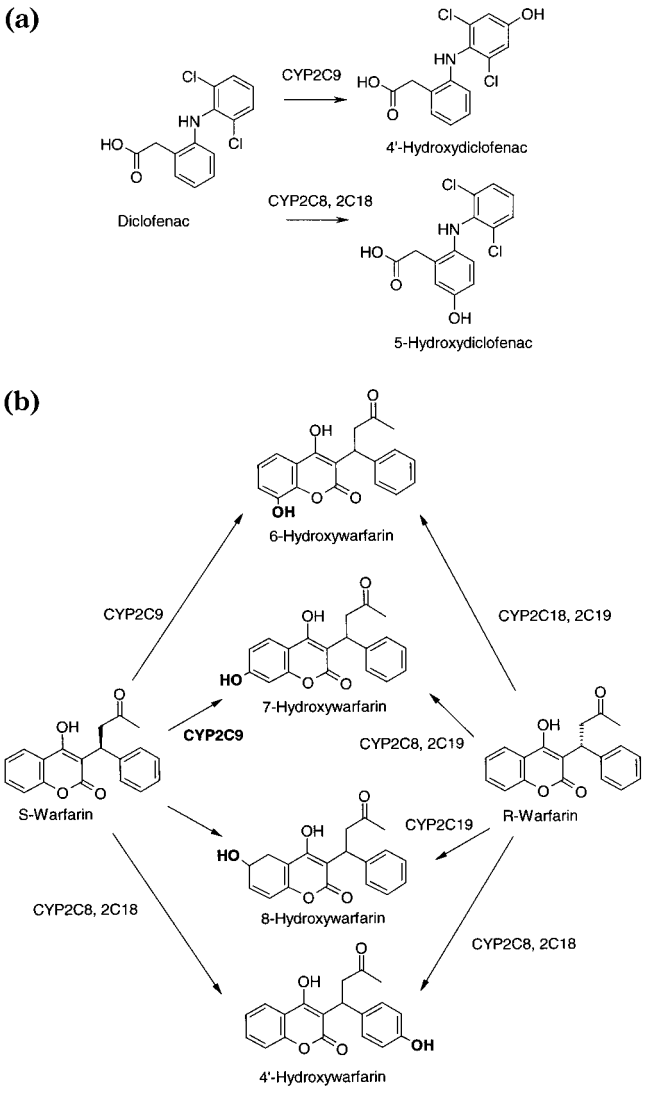

Figure 1. Metabolism of diclofenac and warfarin by CYP2C enzymes: (a) main route of diclofenac metabolism by CYP2C9 via 4′-hydroxylation and by CYP2C8 and 2C18 by 5-hydroxylation; (b) main routes of *S*- and *R*-warfarin metabolism. The major metabolite from *S*-warfarin is hydroxylation in position 7 mainly carried out by CYP2C9, marked in bold.

have been used in a selectivity analysis using the GRID/CPCA strategy[15] to elucidate the selective interactions within the CYP2C subfamily. GRID/CPCA is a method that allows the study of a family of proteins and has been applied to serine proteases previously.[15] This is the first time it is applied to the cytochrome P450 family. The selective regions from the analysis, transformed into pharmacophoric maps, could guide the synthesis of drugs with better pharmacokinetic properties, avoiding CYP interactions.

Table 2. Probes Used in the GRID Calculations and Their Chemical Properties

		no. of H bonds	
probe	chemical group	accepted	donated
DRY	hydrophobic	0	0
C3	methyl group	0	0
N1	neutral flat NH (e.g., amide)	0	1
NH=	sp^2 NH with lone pair	1	1
N:	sp^3 with lone pair	1	0
$N1^+$	sp^3 amine NH cation (+1)	0	3
NM3	trimethylammonium cation (+1)	0	0
O	sp^2 carbonyl oxygen	2	0
OH	phenol or carboxy OH	1	1
COO^-	carboxylic acid anion (−1)	2	0

Experimental Section

Methods. 1. Homology Modeling. The rabbit CYP2C5 crystal structure was used as a template in the modeling.[13] This crystal structure lacks the first 29 N-terminal amino acids and has 5 mutations compared to the wild-type sequence. The corresponding N-terminal amino acids in the human CYP sequences were therefore also removed prior to amino acid sequence alignment. ClustalW was used for sequence alignment.[16] The amino acid sequences of the alignment were imported into the Insight II 98.0 Homology module (Molecular Simulations, Inc., San Diego, CA) or Look, version 3 (Molecular Application Group, Palo Alto, CA) and manually adjusted. The default values in the programs (Insight II and Look) were used for modeling. Several homology models were generated, in some cases using loop optimization. On the basis of analyses using the SYBYL 6.5.3 ProTable module (Tripos, Inc., St. Louis, MO) and Prostat in the Insight II Homology module, an optimal homology model was selected for the selectivity analysis.

2. GRID Calculations. The C_α atoms of the protein structures were aligned using homology alignment in SYBYL 6.5.3 (Tripos, Inc., St. Louis, MO). The PDB files were adjusted to be accepted by GRIN (version 17).[17] The proteins were neutralized by adding counterions in the positions suggested by the program MINIM and FILMAP and always outside the grid box considered in the selectivity analysis. GRIN produced the output files used in the GRID calculation.[17] The setup for the GRID calculation defined a box of 18.5 Å × 20 Å × 15 Å including the active site for each of the CYP models. Two GRID calculations were made, one where the amino acids in the active sites were considered rigid and another where they were considered flexible. In the flexible calculation the MOVE directive was set to 1 (MOVE = 0 in the rigid calculation) in the GRID command file to allow for flexible side chains to respond to the interactions with the probe. The GRID step was set to 1 plane/Å. Default values were used for the other parameters. The following single-atom probes were used in the calculations: DRY, C3, N1, N:, NH=, $N1^+$, NM3, O, OH, and the multiprobe COO^- (Table 2). Hydrophobic interactions are calculated with the DRY probe. The C3 and NM3 probes both describe steric interactions. The $N1^+$ and the COO^- probes are charged. The polar probes consist of the N1, N:, NH=, O, and OH. The ability of hydrogen bonding for each probe is described in Table 2.

3. Consensus Principal Component Analysis (CPCA). The molecular interaction fields (MIFs) from the GRID calculations were imported into GOLPE, version 4.5.11.[18] The

files from the calculation with the single and multiprobe files were merged. The following pretreatment was done: (a) The maximum cutoff was set to zero to consider only the positive interactions (negative energy values) and to enable direct interpretation of the results. (b) Block unscaled weights (BUW) were used to normalize the interaction energies between the different probes. (c) Variables with values smaller than 0.01 kcal/mol and those with a standard deviation below 0.02 kcal/mol were removed in order to eliminate the noisy variables. (d) Analyses were made both on the originally defined box volume (18.5 Å × 20 Å × 15 Å) and on a limited volume defined by the docked diclofenac molecule including a radius of 4 Å using the cutout tool in the pretreatment. The pretreated data were used in CPCA modeling.

Kastenholz et al.[15] developed the algorithm for the consensus principal component analysis (CPCA). This is a modified principal component analysis (PCA) made over a set of data that is intrinsically organized in blocks, like the one obtained from GRID using different probe types. The aim of this algorithm is to capture both the structure of the information in each block and the global structure for all the **X** matrix like in a PCA. This analysis can be seen as a PCA at two different levels. On one level there are the different blocks of structure data, and on the other level there is the combination of these blocks to yield an analysis for the overall data (results similar to the result from usual PCA).

Each of these levels has loading and score vectors that summarize the information like in a usual PCA, and they will be called CPCA loading and score plots, respectively. At the combination level, it is possible to analyze the entire **X** matrix, obtaining a result similar to that of the usual PCA; here there are also loading and score vectors. The connection between the two levels can be done using the CPCA superweight plot that highlights the influence of the different blocks into the principal components.

4. Docking. GOLD, version 1.1,[19] was used in the docking of CYP enzyme specific substrates. The compounds were built and energy-minimized in vacuo using the Tripos force field and Gasteiger−Marsili atom charges. The active-site cavity was defined with a radius of 10 Å from the oxygen atom in the water coordinated to the heme. The genetic algorithm implemented in GOLD was used to optimize the orientation of the ligand in the active site. During this optimization process the ligand was considered flexible (movement of the flexible bonds) while the active site of the enzyme was considered rigid (no movement of the amino acid side chains was allowed). For each ligand 10 dockings were allowed with an early termination if the root-mean-square distances (rmsd) were within 1.5 Å for the top three solutions.

5. Inverse Pharmacophore. A set of selective points (high loading values in the CPCA) for each interaction type are found from the selectivity analysis. The coordinates of these points were collected and merged. Clusters of those in the 3D space were found, and a pharmacophoric point was assigned to each cluster by considering the characteristics (hydrophobic, hydrogen bond acceptor or donor) of the molecular interaction field.

Results and Discussion

CYP2C Homology Models. The homology models were based on the structure of the crystallized CYP2C5 where 29 N-terminal amino acids are absent.[13] The final alignment after ClustalW and manual adjustments is shown in Figure 2. There are three additional amino acids in the human CYP2C amino acid sequences present in the loop region between helices H and I. Insertions were adjusted to minimize the disturbances in the two helices. The C_α atoms of the homology models had an rmsd between 0.18 and 0.23 relative to the CYP2C5 crystal structure (Table 3). The $\Phi-\psi$ angles in the core region of the models that were present in the allowed region were higher in the models (79−83%) compared to that of the template (67%) because of the optimization of the angles by MODELER. The MatchMaker average energy score was reasonable for the models and varied between $-0.13kT$ and $-0.10kT$ (CYP2C5, $-0.13kT$).

Table 3. Analysis of the CYP2C5 Template and Human CYP Homology Models

CYP structure	rmsd[a] (Å)	$\Phi-\psi$ angles$_{core}$ (%)	MMAES[b] (kT)
2C5	0	67	−0.13
2C8	0.19	80	−0.13
2C9	0.18	83	−0.11
2C18	0.23	83	−0.11
2C19	0.18	79	−0.10

[a] Root-mean-square distance of the C_α atoms of the model in relation to CYP2C5. [b] MatchMaker avarage energy score.

Chemometric Analysis and Probe Selection. Four different types of analyses were performed on the basis of the rigid or flexible GRID calculations, with and without using the cutout tool in the pretreatment that defines the 4 Å radius around the docked diclofenac molecule (see Experimental Section). Table 4 shows the number of variables and the percentage of variance explained by each of these analyses. The relationships between the different block variables for each analysis type (chemical probes) are shown in Figure 3. The probes are grouped into four or five clusters, depending on the type of analysis, including probes DRY, the steric C3 and NM3, acceptors N:, O, and COO^-, donor/acceptor NH═ and OH, and donors N1 and $N1^+$ for the rigid map. The flexible map looked similar to the map of the rigid analysis with cutout pretreatment. The DRY probe was clearly distinct from the other nine probes in the rigid analysis with the cutout pretreatment, while it was close to the other probes when the cutout was not applied. This effect is due to the fact that the cutout was made close to the heme as defined by the docked diclofenac molecule and that most of the amino acids characteristic of this region are hydrophobic. However, when the analysis is extended to other parts of the binding site, i.e., the full box volume, as in the case without the cutout tool in the pretreatment, other types of amino acids appear to be relevant. Nevertheless, the clustering of the probes is not affected by the type of pretreatment, observing the same clusters for the rigid and flexible maps with and without cutout. Therefore, the selected probes for further analysis were DRY, C3, one donor (N1), and one acceptor (O) because they are representatives of each cluster in the CPCA superweight plot for the rigid analysis with the cutout pretreatment. The DRY probe describes the hydrophobic interactions. Steric interactions are described by the interactions of the C3 probe with the protein active site. The charged and polar probes map the hydrogen bonding properties and the electrostatic interactions as described in Table 2.

Geometry of the CYP2C Active Sites. The shape of the active sites can be described from the steric interactions of the C3 probe. The first component in all four analyses discriminates between CYP2C8 and the other three enzymes as seen in the C3 score plot in Figure 4a, meaning that CYP2C8 is the most different among the CYP2C enzymes. The second component in all analyses discriminates between CYP2C18 and the rest of the CYP2C enzymes with a different variance

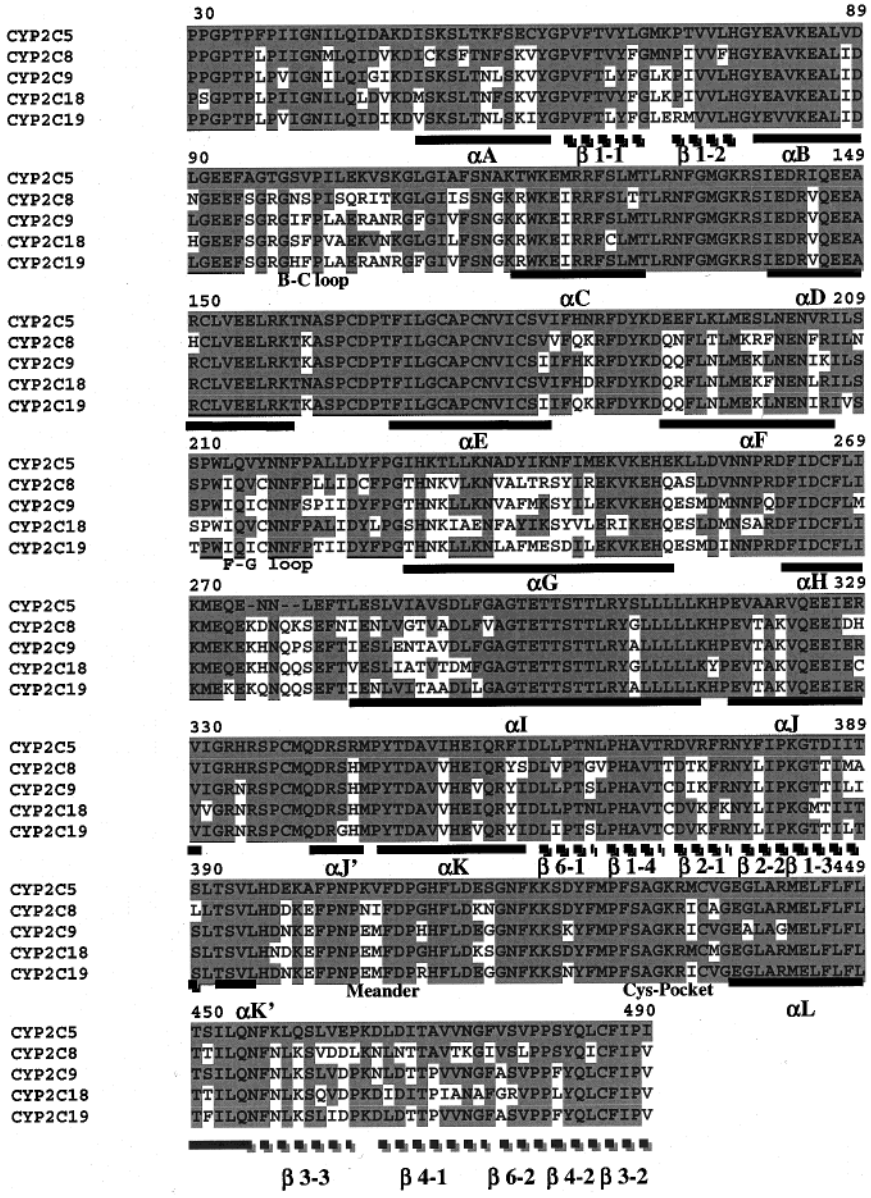

Figure 2. Amino acid sequence alignment of CYP2C5 with human CYP2C enzymes. The amino acid sequence starts with residue number 30. Helices, β strands, the meander region, the Cys pocket, B−C loop, and F−G loop are marked. Identical amino acids are marked in gray.

Table 4. *X* Variables Explaining Each Component in the Rigid CPCA Model with Cutout Pretreatment ($x = 5471$)[a]

component	XVEtot (%)	XVE (%) DRY	XVE (%) C3	XVE (%) N1	XVE (%) N:	XVE (%) NH=	XVE (%) $N1^+$	XVE (%) NM3	XVE (%) O	XVE (%) OH	XVE (%) COO^-
1	40	18	40	49	52	48	45	45	55	52	54
2	13	33	10	9	9	8	5	8	8	8	8
3	46	49	50	43	39	44	51	47	37	40	39

[a] XVE corresponds to *X* variables explained and is listed for the total *x* variables and for each probe.

explained in each case. Finally, the third component discriminates between CYP2C9 and the rest of the proteins (not shown). Consequently, the first component will highlight selectivity toward CYP2C8, the second component toward CYP2C18, and the third against CYP2C9. Identification of selective regions for the C3 probe interactions with the CYP2C8 active site is exemplified in Figure 4. The variables conferring selectivity to CYP2C8 are selected in the loading plot (Figure 4b) and visualized in the GRID plot (Figure 4c). The molecular interaction fields (MIFs) for the C3 probe interactions with CYP2C8 can be added to the GRID plot (Figure 4d), and if the variables are selective, there is an overlap between the variables and the MIFs. Then, if the CYP2C8 structure is added to the GRID plot, selective regions and amino acids can be identified in the active site. Another option is to use the active plot option, such as the CPCA differential plot, in GOLPE where it is possible to look at the differences between two proteins.

The analysis of the rigid and flexible molecular interaction fields reveals two binding pockets for the different enzymes (Figure 5). The pockets differ in size and amino acid composition. The A pocket is formed by

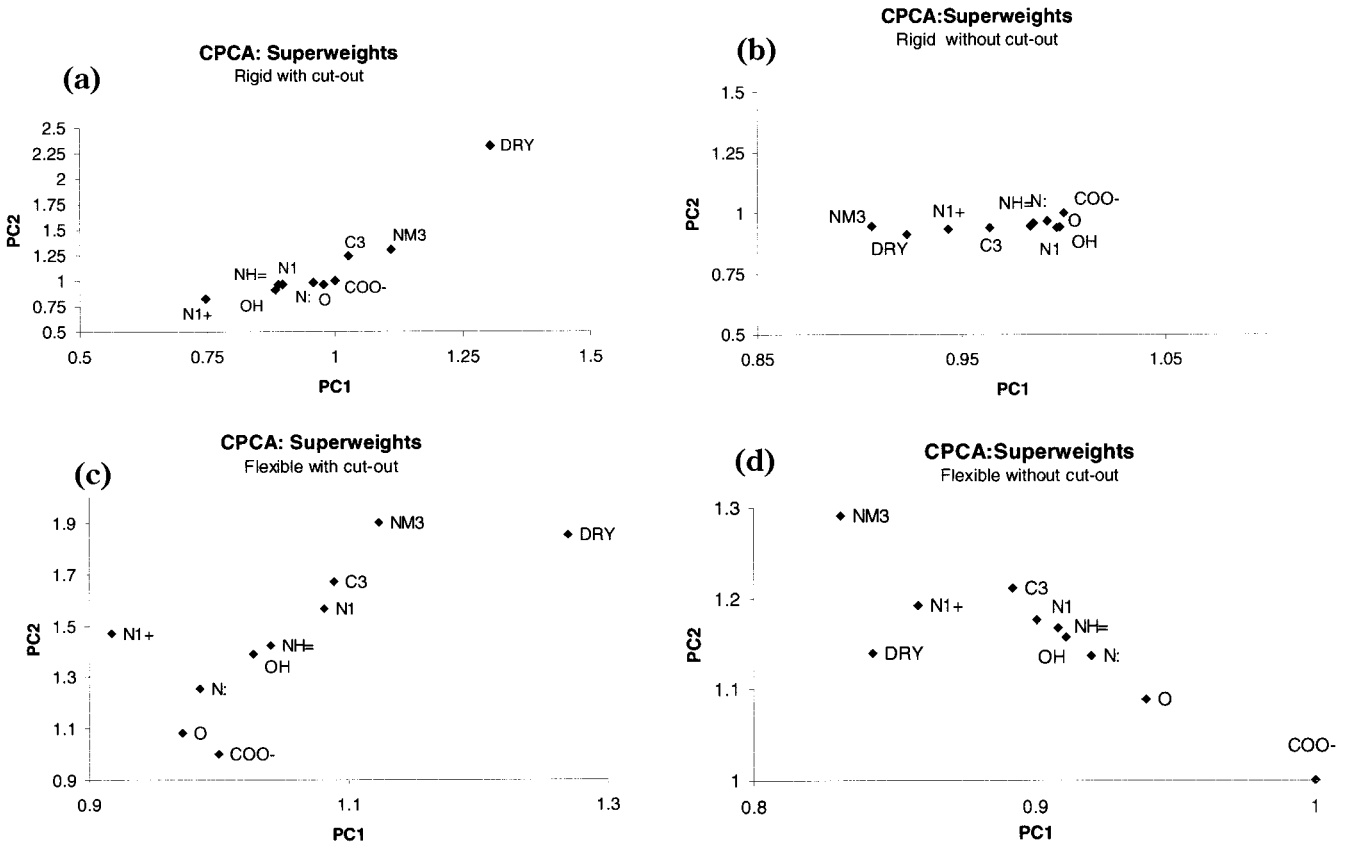

Figure 3. CPCA superweights for the four analyses, with rigid GRID calculation with (a) and without cutout (b) and with flexible GRID calculation with (c) and without cutout (d).

amino acids (Table 5) from helices F, I, and L, β-strand 2, the B−C loop, and the F−G loop. The B pocket, built up by amino acids from helices F, G, I, and the B−C loop, is more elongated, and the access to it seems to be determined mainly by the side chain in positions 114, 205, and 296. The size of the B pocket varies substantially if the analysis is based on the limited binding site where the cutout pretreatment of the variables is used or if the analysis is extended without cutout pretreatment. Because diclofenac is docked close to the heme and the cutout is 4 Å around diclofenac, the B pocket is not covered by this pretreatment. CYP2C18 has bulkier residues in positions 237 and 240 that make the B pocket more shallow compared to the models of the other three enzymes.

We will start by describing the selective regions in the A pocket for the C3 probe (steric) interactions. The amino acids that confer selectivity in each CYP2C enzyme are listed in Table 5. The first component that discriminates CYP2C8 from the other CYP2C forms (Figure 4) has a selective region with amino acid 476 in the rigid analysis, with the addition of amino acids in positions 475 and 478 in the flexible analysis. The second selective region is close to position 362 in the rigid analysis with the addition of residue 361 in the flexible analysis. The second component that discriminates CYP2C18 from the other three CYP forms is also selective in the region around positions 475 and 476 where an Ala and Phe is present. In the flexible analysis, Ala475 and Gly477 are also selective for CYP2C18. CYP2C9 selectivity, described by the third component, has Phe476 and Ala477 that are selective in the first region and Leu 362 in a second region. In the flexible analysis Lys206 and Ala477 are selective for CYP2C9.

Also, pocket B has variations of amino acids in different positions important for hydrophobicity, volume, and polarity. The regions relevant for CYP2C8 selectivity are Ile102, Phe201, Val233, and Val237 where the Ile102 and Phe201 are also found to be selective for CYP2C8 in the flexible analysis. CYP2C18 has a selective region with Val102, Phe201, Ile233, Phe237, Ile240 in this B pocket region where CYP2C18 is very divergent compared to the other CYP enzymes. All of these amino acids except Ile240 are seen in the flexible analysis. The third component, selective against CYP2C9, has a selective region with Leu102, Leu201, Val237, Met240, and Val292.

The middle region in the first component that is selective for CYP2C8 has a selective region including Ile113, Phe205, Val296, and Val366 (pocket A). The flexible analysis shows that Asn99, Ile113 to Phe114 in the B−C loop, Phe205, Leu208, and Val296 are selective for CYP2C8. From the CPCA differential plot in the second component (CYP2C18 discrimination) two selective regions are highlighted: (a) two regions close to the Leu205 that is Ile in CYP2C9 and 2C19; (b) another region close to Ser99 that is a His in CYP2C19, an Asn in CYP2C8, and an Ile in CYP2C9. The third component is selective in the region consisting of Ile99, Phe114, Ile205, Leu208, and Glu300.

Selective Regions Identified with the DRY Probe. The hydrophobic regions are the most important in the selectivity of the CYP2C subfamily of enzymes based on the interactions with the DRY probe in the GRID/CPCA analysis. The block score plot for the DRY probe

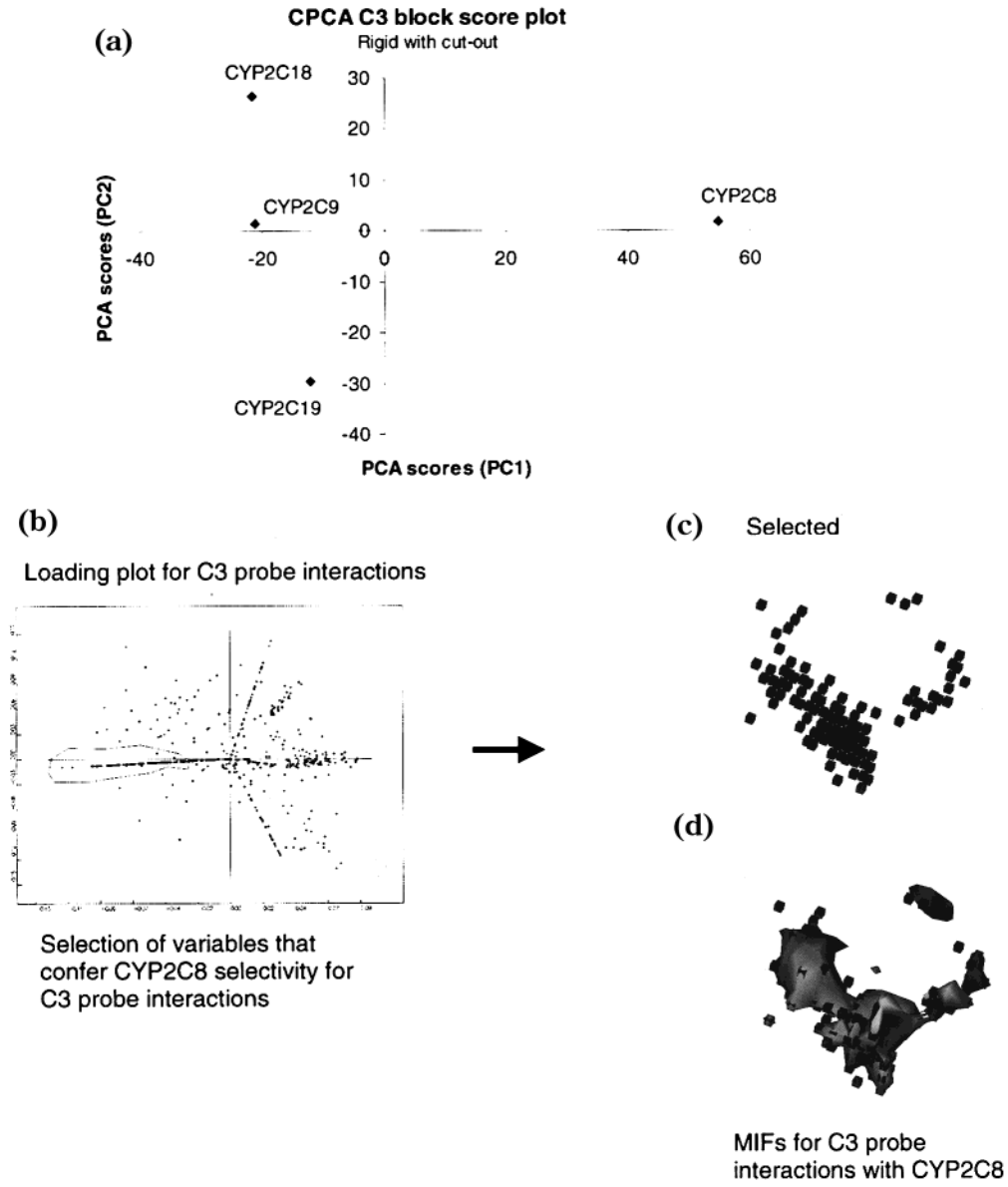

Figure 4. Identification of selective regions for the C3 probe interactions with the CYP2C8 active site. (a) CPCA C3 block score plot is shown for the rigid GRID calculation with cutout in the pretreatment. (b) Variables that confer selectivity for C3 probe interactions with CYP2C8 are selected (circled) in the corresponding loading plot. (c) The selected variables are visualized in the GRID plot (cubes), and when the molecular interaction fields for the C3 probe interactions with CYP2C8 are added (d), they cover the selected variables because they are selective.

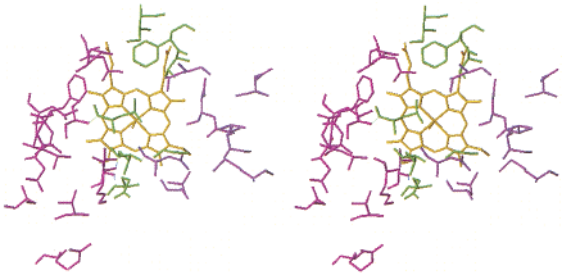

Figure 5. Schematic representation of the active sites of CYP2C enzymes with pockets A and B and the middle region. Shown are the heme (in yellow) and the amino acids located in pockets A (magenta) and B (purple) and the middle region (green).

is very similar to the one obtained with the C3 probe (not shown). The amino acids that are important for the selectivity from the GRID calculations with the DRY probe are listed in Table 5 and marked in bold if the interaction with the DRY probe is energetically favorable. The first principal component discriminates between CYP2C8 and the rest of the enzymes, and the main selective regions are (a) a region in the A pocket including amino acids Ile113 (middle region), Val366, Val362, Ile476, Leu361, Val477, and Thr301 in CYP2C8 that are less hydrophobic than the other proteins; (b) a region in the A pocket close to Ser210 in CYP2C8 that is more hydrophilic than the other CYPs; (c) in the middle a region close to the amino acid in position 205. The amino acids that are selective for CYP2C8 compared to CYP2C9 from the CPCA differential plot are shown in Figure 6. The lowest energies for the molecular interaction fields in the GRID plot identify eight amino acids that are important for the selectivity with the DRY probe. Some of the amino acids in these positions, such as 114, 205, and 206, are also contributing to the shape

Table 5. Amino Acids Conferring Selectivity in the Different CYP2C Enzymes Identified from the GRID/CPCA with the C3, DRY, N1, and O Probes[a]

amino acid no.	CYP2C8	CYP2C9	CYP2C18	CYP2C19
	Pocket A, DRY			
361	***Leu***	Leu	**Leu**	**Leu**
362	***Val***	**Leu**	**Leu**	**Ile**
366	***Val***	Leu	**Leu**	**Leu**
475	*Gly*	Gly	Ala	Gly
476	**Ile**	**Phe**	**Phe**	**Phe**
477	**Val**	**Ala**	*Glt*	**Ala**
	Pocket A, Acceptor or Donor			
51	Cys	Ser	Ser	**Ser**
206	Arg	*Lys*	Arg	Arg
209	**Asn**	**Ser**	Ser	**Ser**
210	**Ser**	**Ser**	Ser	Thr
214	**Gin**	**Gin**	**Gin**	**Gln**
301	**Thr**	Thr	Thr	Thr
304	Thr	Thr	**Thr**	Thr
305	**Thr**	Thr	Thr	Thr
307	**Arg**	**Arg**	Arg	**Arg**
365	Gly	**Ser**	Asn	Ser
474	**Lys**	**Asn**	Asn	**Asn**
478	*Ser*	**Ser**	Arg	Ser
	Pocket B, DRY			
102	**Ile**	**Leu**	**Val**	**Leu**
201	**Phe**	Leu	Phe	Leu
233	*Val*	Leu	**Ile**	**Leu**
237	**Val**	**Val**	Phe	**Leu**
240	Thr	**Met**	Ile	**Met**
289	Gly	Asn	Ala	**Ile**
292	**Ala**	**Val**	**Thr**	Ala
	Pocket B, Acceptor or Donor			
103	**Ser**	Ala	Ala	Ala
104	Asn	**Glu**	Glu	**Glu**
107	Thr	**Asn**	Asn	**Asn**
200	Arg	**Lys**	**Lys**	Lys
204	Asn	**Asn**	**Asn**	**Asn**
240	Thr	**Met**	Ile	**Met**
241	Arg	Lys	Lys	**Glu**
288	Val	**Glu**	Ile	Val
289	Gly	**Asn**	Ala	Ile
292	Ala	**Val**	**Thr**	Ala
293	**Asp**	**Asp**	Asp	**Asp**
	Middle, DRY			
110	Leu	**Phe**	Leu	Phe
113	**Ile**	**Val**	**Leu**	Val
114	*Ser*	**Phe**	**Phe**	**Phe**
205	**Phe**	*Ile*	*Leu*	**Ile**
208	*Leu*	**Leu**	**Leu**	**Val**
296	**Val**	Gly	Gly	Gly
	Middle, Acceptor or Donor			
99	**Asn**	Ile	**Ser**	**His**
114	**Ser**	Phe	Phe	Phe
300	**Glu**	**Glu**	**Glu**	**Glu**

[a] Listed are the amino acids divided into pockets A and B and the middle region for the four CYP enzymes in positions where they are found to be important for the selectivity in one of the CYP2C enzymes. The C3 probe is a steric probe and includes amino acids, which can be identified with the DRY, donor, or acceptor probes. The amino acids conferring selectivity with the C3 probe are underlined, and the DRY and donor (N1) or acceptor (O) probes are marked in bold. In italics are amino acids that were selective in the analysis from the flexible calculation in addition to the residues from the rigid GRID calculation.

of the binding pocket because they are located between the two pockets.

The second component discriminates between CYP2C18 and the remaining enzymes. There is mainly a selective region close to Leu113 and Phe114 in CYP2C18 that has a bigger hydrophobic cavity than in the other enzymes; also, CYP2C9 has a Phe in position 110 that is in the middle of the region. The third component, discriminating CYP2C9 from the other enzymes, is selective in the region close to Leu102 and Phe114.

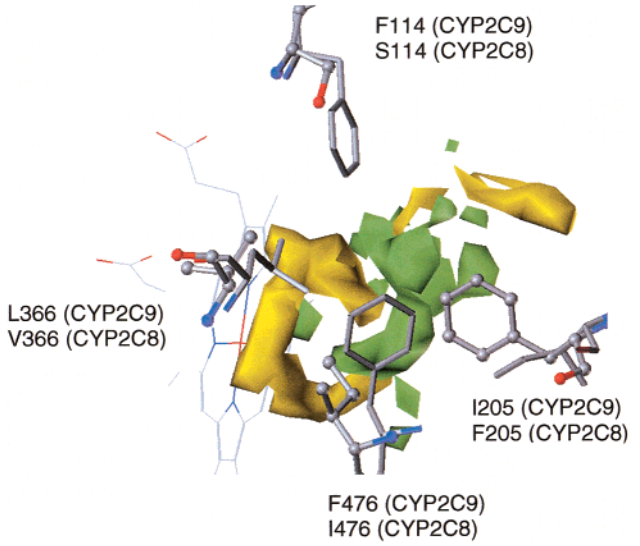

Figure 6. CPCA pseudofield plot of the field differences between CYP2C8 (in ball-and-stick) and CYP2C9 (in stick) for the GRID DRY probe. Shown are selective amino acids and molecular interaction fields that are favorable (in yellow) and less favorable (in green) for the interaction with the active site of the CYP2C9 model compared to that of the CYP2C8 model.

The serine in CYP2C8, instead of the Phe in position 114 present in the other three CYP2Cs, is a drastic difference that would have impact on the binding. Replacement of Phe in position 114 against Leu in CYP2C9 by site-directed mutagenesis results in a large increase in the K_m value for *S*-warfarin-7-hydroxylation as well as for 4′-hydroxylation of diclofenac (Figure 1b).[20] However, lauric acid 11-hydroxylation was not affected to the same extent by this mutation. Hence, it is believed that Phe114 is important for the interaction with the aromatic parts of *S*-warfarin and diclofenac. However, this is probably only one possible interaction with the amino acids in the active site because the Phe in position 205 in CYP2C8 is also likely to interact with the substrate, and in CYP2C9 this amino acid is replaced by a Ile that might affect the orientation of diclofenac such as in the docking in Figure 7. The 4′-hydroxydiclofenac reaction is rather specific for CYP2C9 compared to the metabolism of diclofenac by the other three CYP2C subfamily enzymes where the activity of 4′-hydroxylation of diclofenac is lower and the 5 position of diclofenac is favored by CYP2C8 and 2C18.

Polymorphism in position 359 (Ile or Leu) in CYP2C9 is suggested to be important for the regioselective hydroxylation of warfarin by CYP2C9.[21] However, the residue in position 359 is not identified as important for the selectivity in our study and it is located further away from the active site.

Leu362 is located close to the heme and identified in the selectivity analysis. This amino acid was mutated to Ala in CYP2C9, yielding a 10 times higher K_m value with diclofenac, while the L362I mutant was largely unaffected by the replacement (unpublished results).

Hydrogen Bond Acceptor and Donor Amino Acids in the CYP2C Subfamily. The contributions

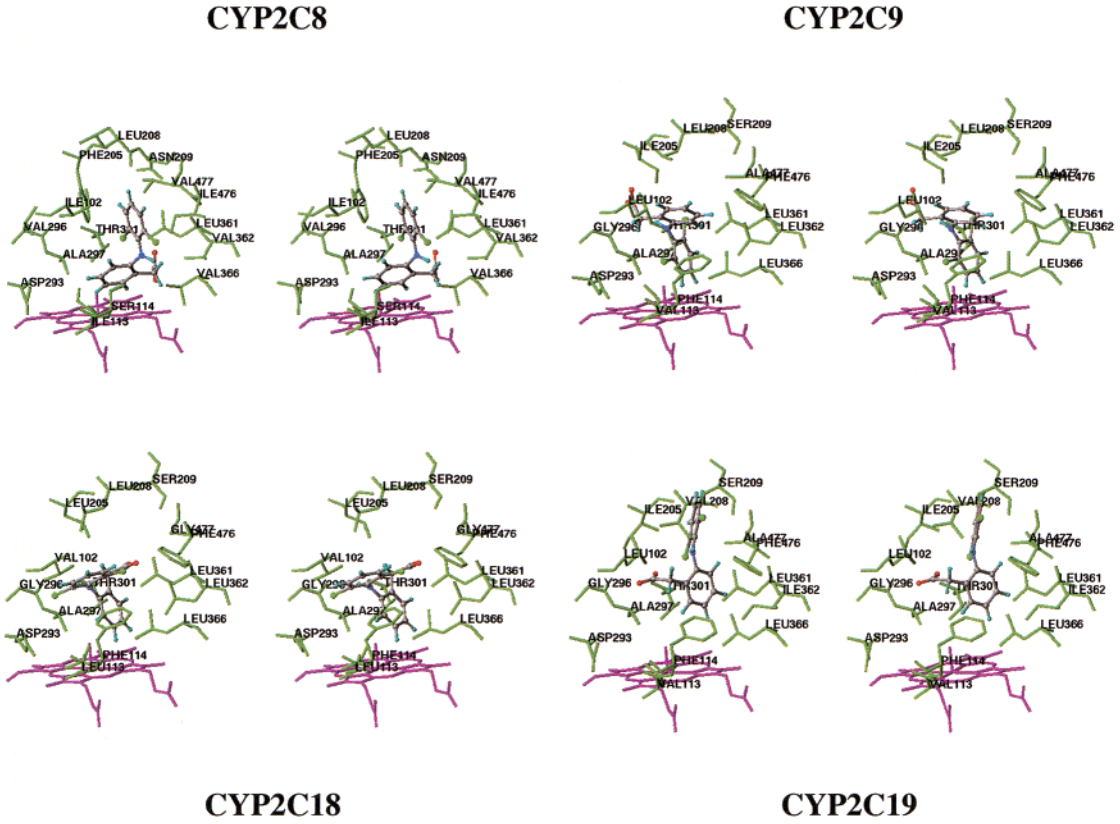

Figure 7. Docking of diclofenac to CYP2C8, 2C9, 2C18, and 2C19. Stereoview of the active-site residues in green, heme in magenta, and diclofenac. Amino acids close to the docked ligand are shown and labeled (positions 102, 113, 114, 205, 208, 209, 293, 296, 301, 362, 366, 476, and 477).

from charged or polar amino acids in the binding and hence the selectivity of the CYP2C enzymes are less pronounced in the plot of the *x* variables profile (not shown) where the DRY probe contributes with the most negative values. The identified amino acids in the active sites of the models with the acceptor and donor probes are listed in Table 5. The first component for the interaction with the N1 probe identifies two selective regions in the A pocket of CYP2C8: one in the back of the A pocket close to residue Gln214 and a second region close to Lys474 and Asn209. The first component of the calculation with the O probe is also selective in the two regions mentioned for CYP2C8. The second component (CYP2C18) of the calculation with the O probe has two selective regions in the A pocket close to Glu300 (middle) and Thr304 (A pocket), a second region close to Gln214 (A pocket) and Ser99 (middle) and a selective region in the B pocket close to Asn204, Thr292, and Lys200. The third component from the calculations with the N1 probe is selective in two regions in CYP2C9; one in the B pocket close to Asp293 and another in the middle region close to Glu300. The O probe identifies two regions in CYP2C9 as selective in the B pocket: one close to Asn204 and a second region close to Glu300 (middle region), Ser303, and Arg307.

Mutation of I99H in CYP2C9 has previously been shown to increase omeprazole 5-hydroxylation to 51% of the CYP2C19 wild-type activity.[22] The 5-hydroxylation of omeprazole was further increased by the mutations S200P and P221T in CYP2C9. The interpretation by the authors is that either an unidentified basic residue is involved directly in hydrogen bonding of the substrate or they change the configuration of the active site. In a recent study, the three mutations (I99H, S200P, and P221T) with the addition of S286N, V292A, and F295L were shown to confer *S*-mephenytoin 4′-hydroxylase activity in CY2C9 (6% of the CYP2C19 wild-type activity).[23] The His99 is identified as conferring selectivity with the O probe. Position 292 is identified as selective with the DRY probe, but positions 200 and 221 are located far away from the active site and not likely to have direct interactions with the substrate. The rest of the selective charged or polar amino acids are located rather far away from the heme in the active site and are not likely to directly contribute to binding. Other kinds of interactions such as water bridges are possible, but they are not considered in this analysis.

Docking of progesterone to the CYP2C5 crystal structure has identified several of the amino acids listed in Table 5 as important for interaction with the ligand.[13] The amino acids correspond to positions 102, 113, 114, 205, 208, 293, 296, 301, 361, 362, 366, 476, and 477 in Table 5. Among the listed residues are amino acids 201, 206, 209, and 210, which are found to be selective in CYP2C8 and CYP2C9. These four amino acids have been mutated in CYP2C5 and found to have no effect on the 21-hydroxylation of progesterone.[24] The reasons for the additional amino acids in Table 5 identified from the selectivity analysis, compared to those previously identified, could be that (1) the binding pocket without the cutout pretreatment is larger compared to binding pocket with the cutout pretreatment around the docked diclofenac molecule and covers regions far from the heme and (2) additional amino acids are identified by flexible analysis.

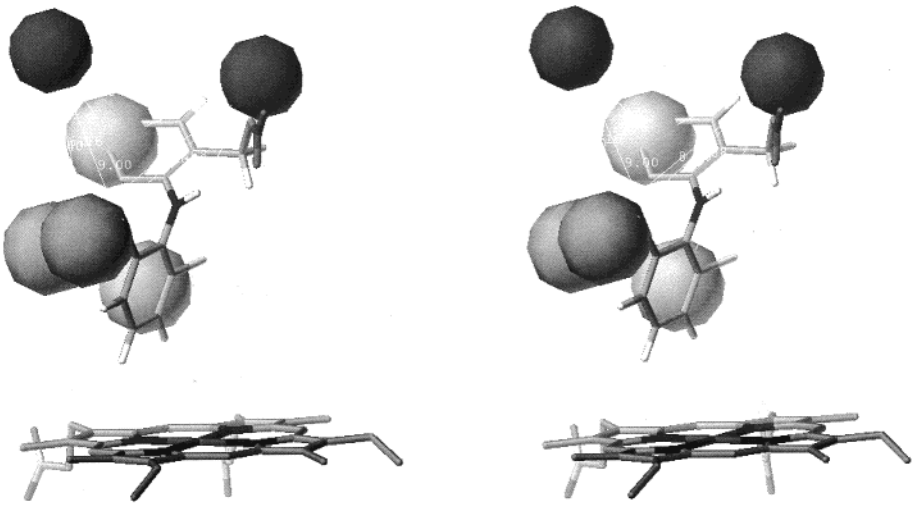

Figure 8. Inverse pharmacophore model of CYP2C9 with diclofenac docked into the active site. The pharmacophore points resemble the variables from the GRID calculations with the DRY probe (gray) and COO^- probe (dark gray).

Docking of Ligands to the CYP2C Models. Diclofenac is metabolized by the four CYP2C enzymes but in two different positions and to different extents (Figure 1a).[5] CYP2C9 is the major enzyme catalyzing 4′-hydroxylation of diclofenac. CYP2C18, CYP2C19, and CYP2C8 also generate the 4′-hydroxy metabolite, in decreasing order of activity, with k_{cat}/K_m values that are a hundred times lower than that of CYP2C9. The major route for diclofenac metabolism by CYP2C8 and CYP2C18 are oxidation in the 5 position. CYP2C19 has a 10-fold lower k_{cat}/K_m value for the 5-hydroxylation of diclofenac compared to CYP2C8 and 2C18. Results of docking of diclofenac into the active sites of the CYP2C models are shown in Figure 7. The 4′-position of the substrate was in a more energetically favorable position in CYP2C9 compared to the other CYP2C enzyme models, which is in agreement with the product from the reaction catalyzed by CYP2C9. CYP2C8, 2C18, and 2C19 had the 5 position closest to the heme, as suggested by GOLD. This coincides with the major products from the reactions catalyzed by the three enzymes. In the active sites of the models, diclofenac seems to be positioned differently and the interaction with the amino acids in positions 114, 205, and 476, identified as selective in the GRID/CPCA analysis, seems to be important for the orientation of the molecule in the active site. Because the GOLD program does not take flexibility of the active site amino acids into account, other docking methods are currently explored to investigate the effect of flexibility in the protein.

Inverse Pharmacophore Model of CYP2C9. The regions conferring selectivity in CYP2C9 from the analysis were used as pharmacophoric points in the model. These regions consisted of variables from the analysis of the GRID calculation with the DRY probe and COO^- probe. The model (Figure 8) is in agreement with the properties of the docked diclofenac molecule.

Conclusions

The selectivity analysis of the four CYP2C homology models shows that the most important determinants of selectivity for the enzymes are hydrophobicity and the geometry of the active sites. CYP2C8 is the most distinct among the four proteins because of big differences in the size and properties of the amino acids in the active site of the model. On the basis of the regions conferring CYP2C9 selectivity, an inverse pharmacophore model could be constructed that resembles the interactions of the docked diclofenac molecule.

References

(1) Rendic, S.; di Carlo, F. J. Human Cytochrome P450 Enzymes: A Status Report Summarizing their Reactions, Substrates, Inducers, and Inhibitors. *Drug Metab. Rev.* **1997**, *29*, 413−580.
(2) Smith, D. A.; Ackland, M. J.; Jones, B. C. Properties of Cytochrome P450 Isoenzymes and Their Substrates Part 1: Active Site Characteristics. *Drug Discovery Today* **1997**, *2*, 406−414.
(3) Smith, D. A.; Ackland, M. J.; Jones, B. C. Properties of Cytochrome P450 Isoenzymes and Their Substrates Part 2: Properties of Cytochrome P450 Substrates. *Drug Discovery Today* **1997**, *2*, 479−486.
(4) Miners, J. O.; Birkett, D. J. Cytochrome P4502C9: An Enzyme of Major Importance in Human Drug Metabolism. *Br. J. Clin. Pharmacol.* **1998**, *45*, 525−538.
(5) Mancy, A.; Antignac, M.; Minoletti, C.; Dijols, S.; Mouries, V.; Duong, N.-T. H.; Battioni, P.; Dansette, P. M.; Mansuy, D. Diclofenac and Its Derivatives as Tools for Studying Human Cytochromes P450 Active Sites: Particular Efficiency and Regioselectivity of P450 2Cs. *Biochemistry* **1999**, *38*, 14264−14270.
(6) Kaminsky, L. S.; de Morais, S. M. F.; Faletto, M. B.; Dunbar, D. A.; Goldstein, J. A. Correlation of Human Cytochrome P4502C Substrate Specificities with Primary Structure: Warfarin as a Probe. *Mol. Pharmacol.* **1993**, *43*, 234−239.
(7) Sonnichsen, D. S.; Liu, Q.; Schuetz, E. G.; Schuetz, J. D.; Pappo, A.; Relling, M. V. Variability in Human Cytochrome P450 Paclitaxel Metabolism. *J. Pharmacol. Exp. Ther.* **1995**, *275*, 566−75.
(8) Yamazaki, H.; Shibata, A.; Suzuki, M.; Nakajima, M.; Shimada, N.; Guengerich, F. P.; Yokoi, T. Oxidation of Troglitazone to a Quinone-Type Metabolite Catalyzed by Cytochrome P-450 2C8 and P-450 3A4 in Human Liver Microsomes. *Drug Metab. Dispos.* **1999**, *27*, 1260−1266.
(9) Becquemont, L.; Mouajjah, S.; Escaffre, O.; Beaune, P.; Funck-Brentano, C.; Jaillon, P. Cytochrome P-450 3A4 and 2C8 Are Involved in Zopiclone Metabolism. *Drug Metab. Dispos.* **1999**, *27*, 1068−1073.
(10) Minoletti, C.; Dijols, S.; Dansette, P. M.; Mansuy, D. Comparison of the Substrate Specificities of Human Liver Cytochrome P450s 2C9 and 2C18: Application to the Design of a Specific Substrate of CYP 2C18. *Biochemistry* **1999**, *38*, 7828−7836.
(11) Ravichandran, K. G.; Boddupalli, S. S.; Hasemann, C. A.; Peterson, J. A.; Deisenhofer, J. Crystal Structure of Hemoprotein Domain of P450BM-3, a Prototype for Microsomal P450's. *Science* **1993**, *261*, 731−736.

(12) Li, H.; Poulos, T. L. The Structure of the Cytochrome P450BM-3 Haem Domain Complexed with the Fatty Acid Substrate, Palmitoleic Acid. *Nat. Struct. Biol.* **1997**, *4*, 140−146.
(13) Williams, P. A.; Cosme, J.; Sridhar, V.; Johnson, E. F.; McRee, D. E. Mammalian Microsomal Cytochrome P450 Monooxygenase: Structural Adaptations for Membrane Binding and Functional Diversity. *Mol. Cell* **2000**, *5*, 121−131.
(14) Afzelius, L.; Zamora, I.; Ridderström, M.; Andersson, T. B.; Karlén, A.; Masimirembwa, C. M. Competitive CYP2C9 Inhibitors: Enzyme Inhibition Studies, Protein Homology Modeling, and Three-Dimensional Quantitative Structure−Activity Relationship Analysis. *Mol. Pharmacol.* **2001**, *59*, 909−919.
(15) Kastenholz, M. A.; Pastor, M.; Cruciani, G.; Haaksma, E. E. J.; Fox, T. GRID/CPCA: A New Computational Tool To Design Selective Ligands. *J. Med. Chem.* **2000**, *43*, 3033−3044.
(16) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Res.* **1994**, *22*, 4673−4680.
(17) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.
(18) Massimo, B.; Cruciani, G.; Costantino, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating Optimal Linear PLS Estimations (GOLPE): An Advanced Chemometric Tool for Handling 3D-QSAR Problems. *Quant. Struct.−Activity Relat.* **1993**, *12*, 9−20.
(19) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(20) Haining, R. L.; Jones, J. P.; Henne, K. R.; Fisher, M. B.; Koop, D. R.; Trager, W. F.; Rettie, A. E. Enzymatic Determinants of the Substrate Specificity of CYP2C9: Role of B′−C Loop Residues in Providing the π-Stacking Anchor Site for Warfarin Binding. *Biochemistry* **1999**, *38*, 3285−3292.
(21) Yamazaki, H.; Inoue, K.; Shimada, T. Roles of Two Allelic Variants (Argl44Cys and Ile359Leu) of Cytochrome P4502C9 in the Oxidation of Tolbutamide and Warfarin by Human Liver Microsomes. *Xenobiotica* **1998**, *28*, 103−115.
(22) Ibeanu, G. C.; Ghanayem, B. I.; Linko, P.; Li, L.; Pedersen, L. G.; Goldstein, J. A. Identification of Residues 99, 220, and 221 of Human Cytochrome P450 2C19 as Key Determinants of Omeprazole Hydroxylase Activity. *J. Biol. Chem.* **1996**, *271*, 12496−12501.
(23) Tsao, C.-C.; Wester, M. R.; Ghanayem, B.; Coulter, S. J.; Chanas, B.; Johnson, E. F.; Goldstein, J. A. Identification of Human CYP2C19 Residues that Confer *S*-Mephenytoin 4′-Hydroxylation Activity to CYP2C9. *Biochemistry* **2001**, *40*, 1937−1944.
(24) Cosme, J.; Johnson, E. F. Engineering Microsomal Cytochrome P450 2C5 To Be a Soluble, Monomeric Enzyme. *J. Biol. Chem.* **2000**, *275*, 2545−2553.

JM0109107