# Regioselectivity of CYP2B6: Homology modelling, molecular dynamics simulation, docking

Christine Bathelt, Rolf D. Schmid, Jürgen Pleiss[1]

Institute of Technical Biochemistry, University of Stuttgart, Allmandring 31, D-70569 Stuttgart, Germany

[1]Correspondence to: Dr. Jürgen Pleiss
Institute of Technical Biochemistry
University of Stuttgart
Allmandring 31
D-70569 Stuttgart, Germany
Phone: (++49)-711-685-3191
Fax: (++49)-711-685-3196
Email: Juergen.Pleiss@po.uni-stuttgart.de

**Abstract**

Human cytochrome P450 (CYP) 2B6 activates the anticancer prodrug cyclophosphamide (CPA) by 4-hydroxylation. In contrast, the same enzyme catalyzes N-deethylation of a structural isomer, the prodrug ifosfamide (IFA) thus causing severe adverse drug effects. To model the molecular interactions leading to a switch in regioselectivity, the structure of CYP2B6 was modelled based on the structure of rabbit CYP2C5. We modelled the lacking 22 residue loop in CYP2C5 between helix F and G (F-G loop) which is not resolved in the X-ray structure by molecular dynamics (MD) simulations using a simulated annealing protocol. The modelled conformation of the loop was validated by unconstrained MD simulations of the complete enzymes (CYP2C5 and CYP2B6) in water for 70 and 120 ps, respectively. The simulations were stable and led to a backbone r.m.s. deviation of 1.7 Å between the two CYPs.

The shape of the substrate binding site of CYP2B6 was further analyzed. It consists of three well-defined hydrophobic binding pockets adjacent to the catalytic heme. Size, shape and hydrophobicity of these pockets was compared to the shapes of the two structurally isomeric substrates. In their preferred orientation in the binding site both substrates fill all three binding pockets without repulsive interactions. The distance to the heme iron is short enough for 4-hydroxylation and N-deethylation to occur for CPA and IFA, respectively. However, if the substrates are docked in the non-preferred orientation (such that 4-hydroxylation and N-deethylation would occur for IFA and CPA, respectively), one pocket is left empty, and clashes were observed between the substrates and the enzyme.

**Keywords**

loop
MD simulation
P450 monooxygenase

**Introduction**

The cytochrome P450 enzymes (CYP) are a superfamily of hemoproteins involved in the metabolic biotransformation of a variety of endogenous compounds, such as steroids and prostaglandins, and exogenous compounds, such as drugs and carcinogens [1]. More than 20 distinct forms of mammalian liver P450s [2] have already been characterized. Individual CYPs exhibit unique substrate specifity, regio-and stereoselectivity, and most of them play a role in the oxidative metabolism of drugs.

Human CYP2B6 metabolizes about 3% of drugs in clinical use [3]. It catalyzes the metabolism of two widely used anticancer prodrugs, cyclophosphamide (CPA) and its structural isomer ifosfamide (IFA) [4]. These compounds can be metabolized via two alternative pathways. 4-Hydroxylation corresponds to a drug activation pathway yielding an active alkylating species, while N-dechloroethylation generates inactive, neurotoxic metabolites. The regioselectivity of CYP2B6 has been studied experimentally [4]. With CPA as a substrate CYP2B6 catalyzes primarily 4-hydroxylation whereas with IFA N-dechloroethylation is preferred. For CPA, IFA, and other drugs metabolized by CYP2B6 the knowledge of the regioselectivity can be valuable to avoid metabolic pathways which lead to severe adverse drug effects when administered to patients.

To understand the experimental findings and predict substrate specifity of CYP2B6 it is necessary to know the three-dimensional structure of this enzyme. Several structures of bacterial CYPs are available, however only one crystal structure of a mammalian CYP, CYP2C5 from rabbit [5], has been solved up to date. Due to the high sequence similarity between CYP2C5 and CYP2B6 a homology model based on the crystal structure of CYP2C5 can be helpful in the absence of an experimental structure of CYP2B6. However the crystal structure of CYP2C5 lacks the residues 212 to 222 of the F-G loop. This loop is similar to the corresponding loop in CYP2B6 . Therefore, prior to homology modelling of CYP2B6, the structure of this lacking loop has to be modelled. Peterson et al [6] report that the F-G loop plays a role for membrane-association of the enzyme by forming part of a hydrophobic surface. It is also known from the crystal structures of the bacterial CYPBM-3 that the F-G loop region is involved in the opening motion of a substrate access channel leading from the surface of the protein to the active site [7].

The difficulty in modelling of the F-G loop arises because the corresponding loops in all known structures of the bacterial enzymes, which are not membrane-associated, are shorter and therefore do not provide good templates. In this study, we applied MD simulations with simulated annealing to obtain a structure for the missing F-G loop in CYP2C5. Unconstrained MD simulations in water served as a criterion to evaluate the stability of the modelled structure. This is a very sensitive test for the quality of a structure as a protein structure denatures in totally unconstrained MD simulation if it is not folded into its native state [8]. The completed CYP2C5 structure then served as a template for generating a model of CYP2B6, which was used to investigate the regioselectivity of CYP2B6.

## Methods

### Hard-and software

Molecular dynamics (MD) simulations and energy minimizations were carried out on a PC (AMD Athlon, 1.33 GHz) with OS linux Mandrake 8.0 using the Amber 6.0 [9] suite of programs. Substrate building and geometry optimization were performed with InsightII (Accelrys Inc., San Diego, CA) implemented on Silicon Graphics workstations (Octane2, Irix 6.5).
For visualization of structures and trajectories the SwissPdb Viewer 3.7 (b2) [10] and VMD1.6.1[11] respectively were used.

### Structures and sequences

The X-ray structures of CYP2C5 and CYP BM-3 in the substrate-bound and unbound state were taken from the Brookhaven Protein Data Bank (PDB) [12] with the PDB codes 1DT6[5] , 1FAG[13] and 1BU7[14], respectively. The sequence of recombinant CYP2C5, which was engineered for solubility by removing the N-terminal membrane anchor, was taken from the PDB entry 1DT6, the sequence of CYP2B6 was obtained from Swiss-Prot [15] with the accession number P20813.

### MD simulations

All systems were minimized prior to simulation. Hereby the steepest descent method followed by the conjugate gradient method was used. MD simulations were carried out with a time step of 1 fs up to 300 K and 0.5 fs for temperatures over 300 K. SHAKE [16] was applied to all bonds involving H-atoms. For simulations in water a truncated octahedral box of TIP3P water molecules was built around the structure, and simulations were run with the particle mesh Ewald method [17]. Solvated systems were heated with constant volume, and coupled to a temperature bath with a coupling constant of 0.05 ps. In the subsequent equilibration phase the conditions were changed to constant pressure and the coupling constants were increased gradually. The production phase was then run with temperature and pressure coupling constants of 0.8 ps and 1.0 ps, respectively.

### F-G loop modelling

1) Starting structure for F-G loop
A starting structure was built for residues 208 to 226 of CYP2C5. This segment contains residues 212 to 222 of the F-G loop, which have not been resolved experimentally, and the residues with B-factors between 86 and 100 $Å^2$ adjacent to the missing structure part.
The F-G loop of CYP2C5 has an insertion of 4 residues compared to CYP BM-3 as shown in the structural alignment (Fig.1). To accomodate the additional 4 residues structural information from two different conformations of CYP BM-3 was combined. In the closed conformation the F-helix and the F-G loop are more extended, while the open conformation has a more compact F-G loop region. 4 residues of the contracted F-helix of CYP BM-3 in the N-terminal region were used twice and combined with a fragment of the F-G loop in the extended

CYP BM-3 structure (Fig.1). In the C-terminal region 2 residues of CYP BM-3 in the open conformation were used as the F-G loop reaches out further to the side in this conformation, which gives the constructed loop room to expand. The amino acid sequence of this patchwork structure was mutated to the corresponding sequence of the F-G loop in CYP2C5. Two turns of the adjacent helices F and G from the PDB structure of CYP2C5 (residues 201 to 206 and 228 to 233) were added as anchor regions on either end of the constructed loop. The initial structure was energy minimized. During the minimization and all simulations of the structure the anchor regions were constrained in position with a harmonic force constant of 150 kcal/mol $Å^2$.

2) Simulated annealing of F-G loop fragment

Simulated annealing was performed in vacuum. The structure fragment was heated stepwise to 500, 600 and 700 K allowing 100 ps for each 100 K step. From each temperature step the structure was cooled down slowly allowing 20 ps for every 10 K step and applying a loose temperature coupling with a coupling constant of 0.5 ps. To compare the potential energies the 3 resulting structures were cooled down to 10 K, equilibrated at 10 K for 10 ps, and the average energies were calculated.

For the structure, which was cooled down from 700K, an average structure was calculated after heating it to 300 K, equilibrating it at this temperature for 20 ps and averaging the structures of the trajectory over the last 10 ps.

3) Simulated annealing of F-G loop fragment inserted in CYP2C5

Simulated annealing with the modelled loop inserted in CYP2C5 was performed with constraints applied to the backbone of residues 1 to 206 and 228 to 489. Heating and cooling were carried out as for the structure fragment.

4) Simulations of initial model

The complete structure model of CYP2C5 was heated to 300 K in water over 5 ps with constraints applied to all backbone atoms followed by 5 ps of simulation where only the residues 207 to 227 of the F-G loop were free to move. Then all constraints were gradually removed over a period of 10 ps decreasing the force constant from 150 to 0 kcal /mol $Å^2$. Subsequently the system was equilibrated with free backbone for 20 ps, and a production phase of 30 ps followed.

Homology modelling and MD simulation of CYP2B6

The sequences of CYP2C5 and CYP2B6 were aligned with ClustalX [18]. Using this alignment a model of CYP2B6 was generated by Swiss-Model [10,19] based on the completed CYP2C5 structure. The heme group was copied from the PDB structure of CYP2C5 into the CYP2B6 model.

Parameters for all atoms were taken from the Amber force field [20], for the heme group as published previously [21].

The generated CYP2B6 structure model was heated and equilibrated as described for CYP2C5. The production period was 100 ps.

Docking

The substrates were geometry optimized with the semi-empirical program MOPAC(MNDO) as implemented in InsightII and docked manually into the active site of CYP2B6. The S-enantiomer of IFA (Fig.2) was used as suggested by the experimental data[22]. For CPA no experimental data on the stereochemistry are available. Therefore the R-enantiomer (Fig.2) was used which results from S-IFA by changing the position of one chloroethyl side chain. Each substrate was docked in two orientations:

a) CPA was placed for hydroxylation at the C4 ring carbon. The ring was placed in pocket A (Ala298, Thr302, Leu363, Phe206) with a C4-Fe distance of 4.8 Å. Then the molecule was rotated such that one chloroethyl side chain was located in pocket B (Phe297, Phe115, Ser294, Ile114). Finally the position of the second chloroethyl side chain was adjusted to reach into pocket C (Ile209, Val477, Leu216, Val367).

b) For N-deethylation of CPA the N2 chloroethyl side chain, which is in a position equivalent to the N2 chloroethyl side chain in S-IFA, was placed in pocket A with a distance of 5.9 Å between the carbon atom neighboring to N2 and the heme iron atom. A closer distance would have caused clashes of the other chloroethyl side chain with the heme group. Then the ring was placed in pocket B. With these two pockets filled the possible orientations of the second chloroethyl side chain are limited. It can not be placed in pocket C and points towards the side chain of Val367.

c) For N-deethylation of IFA the N2 chloroethyl side chain was chosen in accordance to the experimental data [22]. This side chain was placed in pocket A with a distance of 5.0 Å between the carbon atom neighboring to N2 and the heme iron atom. Then the molecule was rotated such that the ring was located in pocket B. Finally the N3 chloroethyl side chain was adjusted to reach into pocket C.

d) IFA was placed for hydroxylation at the C4 ring carbon. The ring was placed in pocket A with a C4-Fe distance of 4.9 Å Then the molecule was rotated such that the N3 chloroethyl side chain was located in pocket B. With these two moieties bound to pockets A and B the N2 chloroethyl side chain points toward Val367 and can not be adjusted to occupy pocket C.

**Results**

F-G loop modelling

A starting structure was built for residues 208 to 226 of CYP2C5, which is based on the structures of CYP BM-3 in the open and closed conformation and includes parts of the F-and G-helices of CYP2C5 as anchor regions. During minimization of this initial structure the helical structure element from CYP BM-3 at the N-terminal end of the F-G loop unfolded (Fig.3). The structure was relaxed by a simulation in water at 300 K over 1.5 ns. Immediately after heating to 300 K a right-handed one-turn helix was formed comprising the residues Trp212, Leu213, Gln214 and Val215. This one-turn helix stayed stable. Starting from this conformation, *loop1* (Fig.3), a larger conformational space was explored by simulated annealing in vacuum to find other stable structures. *Loop1* was heated to 500, 600 and 700 K where the helix turn was completely unfolded, then each high temperature structure was slowly cooled down to 300 K, and the energies of the resulting 3 structures were calculated. The structure which was cooled down from 700 K, *loop2* (Fig.3), is about 70 kcal/mol lower in energy than *loop1* and is the most favorable conformation. While *loop1* and *loop2* both contain a right-handed one-turn helix its position and orientation differ. To account for the influence of the whole protein, especially the N-terminal β-sheet region opposite to the F-G loop, on the conformation of the F-G loop the simulated annealing in vacuum was repeated with *loop2* inserted into the PDB structure of CYP2C5. Again only the F-G loop was allowed to move. During heating a considerable movement of the loop toward the N-terminal β-sheets was observed. The isolated loop conformation *loop3* (Fig.3) is higher in energy than *loop2.* However in the protein it remained stable in subsequent unconstrained MD simulations in water, whereas *loop2* did not give a stable CYP2C5 model.

Modelling of CYP2B6

The model of CYP2B6 was generated based on the sequence alignment with CYP2C5 shown in Figure 4. The minimized CYP2B6 model has identical backbone coordinates with the experimental structure of CYP2C5 except for the modelled F-G loop and the H-I loop, which differs in length by 3 amino acids (Fig.4). The CYP2B6 model was heated to 300K in water with constrained backbone. After removing the constraints, 100 ps of unconstrained MD simulation followed. The root mean square deviation (RMSD) values of the backbone atom positions between trajectory structures and the minimized CYP2B6 structure was calculated. They rapidly increased to 1.1 Å within the first 10 ps, then they slowly increased to 1.5- 1.7 Å over a period of 100 ps (Fig.5).
The F-G loop of CYP2B6 contains mainly hydrophobic residues with bulky side chains such as phenylalanine and leucine. In this model (Fig.6) the N-terminal residues of the F-G loop are buried in the protein blocking the putative substrate access channel with their side chains. The hydrophobic residues of the C-terminal half of the F-G loop form a hydrophobic patch on the protein surface together with the N-terminal β-strand region. The active site above the heme group is located in the core of the protein and is not in contact with the F-G loop.

Geometry of the binding site

The residues of the binding site form three main pockets (Fig.7). Pocket A, the “heme pocket”, is formed by the side chains of Ala298, Thr302, Leu363 and Phe206. This pocket encloses the heme iron from all sides except one. The substrate moiety, where hydroxylation occurs, can only reach the heme iron if it enters this pocket. In addition two more pockets further away from the heme iron can be identified in the binding cavity. A narrow

pocket B, the “side pocket”, is formed by Ser294 and Ile114 and the two opposing side chains of Phe297 and Phe115, and a pocket C, the “roof pocket”, is located between Ile209, Val477, Leu216 and Val367 and builds the ceiling of the binding cavity at a level further above the heme plane.

Docking

Cyclophosphamide (CPA) and ifosfamide (IFA) (Fig.2) were placed into the active site of CYP2B6.
Both substrates consist of an identical ring moiety and two chloroethyl side chains, but the position of one chloroethyl side chain differs and results in a different geometry of the substrates. In CPA the two chloroethyl side chains are bound to nitrogen N2, which is located axial to the ring plane, and build a fork (Fig.7a and 7b). In IFA the second chloroethyl side chain is bound to the ring nitrogen N3, and is located equatorial to the ring plane (Fig.7c and 7d). Both substrates were placed in two orientations, for C4 ring hydroxylation and N-deethylation. For each substrate we searched for the orientation where all three pockets in the binding cavity were filled by the substrate moieties (Tab. 1).
An appropriate orientation for C4 hydroxylation of CPA is shown in Figure 7a. The ring moiety binds to pocket A with the C4 atom above the heme iron. This places one chloroethyl side chain in pocket B, and the second chloroethyl side chain points upwards into pocket C such that the substrate optimally fills the available space in the binding cavity preventing unfavorable contacts. In contrast, if one of the chloroethyl side chains of CPA is placed in pocket A for N-deethylation to occur, the ring moiety is locked in pocket B, and the second chloroethyl side chain is forced to adopt a position close to the heme plane, where space is limited by the side chain of Val367 (Fig.7b). In this orientation the distance between the chlorine atom of CPA and the methyl group of Val367 is only 1.7 Å (Fig.7b).
If IFA is docked in an analogous way for N-deethylation, the second chloroethyl side chain adopts an ideal position to reach into pocket C as it is situated on the ring nitrogen atom pointing upwards (Fig.7c). However with the ring moiety of IFA placed in pocket A for C4 hydroxylation and one chloroethyl side chain in pocket B, the second chloroethyl side chain can not reach up into pocket C but instead clashes with Val367 in the lower level of the binding cavity (Fig.7d). In this position the distance between the chlorine atom of IFA and the methyl group of Val367 is only 2 Å. (Fig.7d).

**Discussion**

F-G loop modelling

In the structure of CYP2C5 a 19 residue segment of the FG-loop was modelled. Generally the limit for reliable loop structure prediction by ab initio and database search techniques is 5 residues [23]. To overcome this problem we used an approach that relies on molecular dynamics (MD) simulations with simulated annealing. This method was described previously by Fiser et al [23] as part of an automated modelling technique, however only applied to loops of up to 14 residues. During the MD simulations the position of the whole loop underwent considerable changes, however the one-turn helix comprising Trp, Leu, Gln and Val always formed. The pattern of a one-turn helix in the N-terminal region of the F-G loop also appears in the corresponding loop of the bacterial CYP51 (PDB code 1E9X [24]), which has a length of 14 residues. Here, the one-turn helix comprises the 4 residues Leu, Ala, Tyr and Val. In contrast, the helical structure taken from the CYP BM-3 structure, which was used for the starting structure, unfolded at the N-terminal end of the F-G loop after minimization and did not reappear later. This is in accordance with the low resolution X-ray structure of CYP2C5 [5] in this region, which also does not show a helix.

The conformation of a loop is also influenced by the rest of the protein as a segment can adopt different conformations in different proteins [23]. Therefore simulated annealing of the loop was performed in the protein environment. Indeed the conformation changed and the loop moved towards the N-terminal β-sheet domain. As suggested by the low resolution of the X-ray structure [5] in this region the loop is flexible and possibly takes on different conformations, especially in solution where the location of the loop is not limited by spatial constraints in the crystal lattice. Therefore the modelled loop conformation represents one of different alternatives.

CYP2B6 model

The high sequence identity of 48% with CYP2B6 makes CYP2C5 an appropriate template for modelling CYP2B6. The stability of the constructed CYP2B6 model was assessed by 100 ps of unconstrained MD simulation in water. Unconstrained MD simulations in water are a sensitive tool to measure stability: When an automatically modelled loop was used, unconstrained MD simulations in water caused an unfolding of the whole protein after only 10 ps (data not shown). In contrast, the model of CYP2B6 described here changed by 1.1 Å within this critical time. The low RMSD of less than 1.7 Å of the equilibrated model indicates that the CYP2B6 structure only shows small differences to the structure of CYP2C5. This is in accordance with the expected value of 1.4 Å given by the correlation between sequence identity and RMSD [25].

The structure of CYP2B6 shows the expected features of mammalian CYPs [5]. The F-G loop and the N-terminal domain contain predominantly hydrophobic amino acids suggesting insertion of this region in the membrane. Furthermore, it is remarkable that the F-G loop of CYP2B6 contains a large number of bulky hydrophobic residues. This is similar in some other mammalian CYP enzymes like CYP2C5 and CYP2E1 but there are also examples like CYP2A6 and CYP2D6 which show a smaller number of bulky hydrophobic residues in the F-G loop. As postulated for the mammalian CYP7A1 [26] these residues might play a structural role for the protein by preventing it from getting inserted too deeply in the membrane. It is also conceivable that these residues contribute to substrate recognition [26] as hydrophobic residues, especially phenylalanine and leucine residues, dominate in the F-G loop as well as in the binding site.

The F-G loop is located at the surface of the protein and its conformation does not influence the geometry of the

active site, which is in the interior of the protein. One can assume that the active site of CYP2B6 can be reliably modelled as the structure of CYP2C5 is very well resolved in the region around the heme [27], thus providing a good template.

Docking

The shape of the binding site explains the regioselectivity of CYP2B6 towards the substrates CPA and IFA. It has been experimentally observed that the difference in the position of one chloroethyl side chain leads to inverse regioselectivity[4] in the active site of CYP2B6. By manual shape fitting it could be shown why this difference between the two structural isomers has an effect on the orientations of the substrates in the binding cavity. Fitting the shapes of substrates and binding site assumes that steric repulsions for substrate moieties are minimal if they bind to well-defined binding pockets. Three pockets were analyzed, a “heme pocket” A, which the substrate moiety to be hydroxylzed has to enter, a “side pocket” B and a “roof pocket” C. For both substrates all three pockets could be filled without any clashes for the orientation leading to the experimentally observed product while for the non-preferred orientation pocket C could not be filled and instead unfavorable contacts with Val363 occurred. This study represents the first docking results for CPA and IFA using a CYP2B6 model which is based on a mammalian CYP. In an earlier study [28] a structural model of CYP2B6 constructed from the coordinates of CYP BM-3 was used. CYP BM-3 is more distantly related to CYP2B6, and no explanation for the experimentally observed regioselectivity was given. Recently a model of CYP2B6 based on the crystal structure of CYP2C5 was reported [29]. In this study QSAR analysis was combined with homology modeling to predict the $K_m$ values of a number of substrates but did not include the two drugs CPA and IFA.
Our complete model of CYP2B6, which is stable in unconstrained MD simulations in water, can be used for further docking studies and for generating homology models of other CYP family 2 enzymes.

**Conclusion**

The lacking F-G loop in the X-ray structure of cytochrome P450 CYP2C5 was modelled by a simulated annealing approach. The completed 2C5 structure was then used as a template to model the structure of the homologous human CYP2B6. Both structures were stable during molecular dynamics simulations in water. As the sequences of CYP2C5 and CYP2B6 are similar to other mammalian CYPs of family 2, they can be further used as templates for homology modelling.

The model was further validated by studying the interaction of two substrates, CPA and IFA, with the binding site of CYP2B6. Experimentally observed regioselectivity could be explained by size, hydrophobicity, and spatial arrangement of three distinct binding pockets. Thus, a limited number of residues was identified as major structural determinants of regioselectivity.

**Acknowledgement**

This work was supported by the German Federal Ministry of Education and Research (project PTJ 31/0312702).

**References**


1. Dai R, Pincus MR, Friedman FK (2000) Cell Mol Life Sci 57:487-499
2. Yu LJ, Drewes P, Gustafsson K, Brain EGC, Hecht JED, Waxman DJ (1999) J Pharmacol Exp Ther 288**:**928-937
3. Rendic S, Di Carlo FJ (1997) Drug Metab Rev 29:413-580
4. Huang Z, Roy P, Waxman DJ (2000) Biochem Pharmacol 59:961-972
5. Williams PA, Cosme J, Sridhar V, Johnson EF, McRee DE (2000) Mol Cell 5:121-131
6. Graham SE, Peterson JA (1995) In: Physiological functions of cytochrome P450 in relation to structure and regulation, Jai Press Inc., Greenwich, CT, pp 57-79
7. Li H, Poulos TL (1997) Nat Struct Biol 4:140
8. Chang Y, Loew GH (1997) Protein Engng 9:755-766
9. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, DeBolt SE, Ferguson DM, Seibel GL, Kollman PA (1995) Comp Phys Lett, in press
10. Guex N, Peitsch MC (1997) Electrophoresis 18:2714-2723
11. Humphrey W, Dalke A, Schulten K (1996) J Mol Graph 14:33-38
12. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) J Mol Biol 112:535-542
13. Li H, Poulos TL (1997) Nat Struct Biol 4: 140
14. Sevrioukova IF, Li H, Zhang H, Peterson JA, Poulos T (1999) Proc Natl Acad Sci USA 96:1863
15. Bairoch A, Apweiler R (1997) J Mol Med 75:312-316
16. Ryckaert JP, Ciccotti G, Berendson JC (1977) J Comput Phys 23:327
17. Darden T, York D, Pederson L (1993) J Chem Phys 98:10089
18. Thompson JD, Higgins DG, Gibson TJ (1994) Nucleic Acids Res 22:4673-80
19. Peitsch MC (1996) Biochem Soc Trans 24:274-279;
    Peitsch MC (1995) Bio/Technology 13:658-660
20. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) J Amer Chem Soc 117:5179-5197
21. Harris D, Loew GJ (1995) Am Chem Soc 117:2738-2746
22. Granvil CP, Madan A, Sharkawi M, Parkinson A, Wainer IW (1999) Drug Metab Dispos 27:533-541
23. Fiser A, Gian Do RK, Sali A (2000) Protein Sci 9:1753-1773
24. Podust LM, Poulos TL, Waterman MR (2001) Proc Nat Acad Sci USA 98:3068
25. Wood TC, Pearson WR (1999) J Mol Biol 291:977-995
26. Nakayama K, Puchkaev A, Pikuleva IA (2001) J Biol Chem 276:31459-31465
27. Williams PA, Cosme J, Sridhar V, Johnson EF, McRee DE (2000) J Inorg Biochem 81:183-190
28. Lewis DF, Lake BG, Dickins M, Eddershaw PJ, Tarbit MH, Goldfarb PS (1999) Xenobiotica 29:361-393
29. Wang Q, Halpert JR (2002) Drug Metab Dispos 30:86-95

**Tables and figure captions**

**Table 1.** Bindig pockets A, B, and C occupied by the substrate moieties of CPA and IFA in both orientations: all pockets are occupied in the preferred orientation of the substrates (C4-hydroxylation for CPA, N-demethylation for IFA), one pocket remains empty in the non-preferred orientation.

| | Pocket A: A-la298,Thr302, Leu363, Phe206 | Pocket B: Ser294, Ile114, Phe297, Phe115 | Pocket C: Ile209, Val477, Leu216, Val367 |
|---|---|---|---|
| CPA C4-hydroxylation | ring moiety | chloroethyl side chain | second chloroethyl side chain |
| CPA N-demethylation | chloroethyl side chain | ring moiety | empty |
| IFA N-demethylation | chloroethyl side chain | ring moiety | second chloroethyl side chain |
| CPA C4-hydroxylation | ring moiety | chloroethyl side chain | empty |

**Figure captions**

**Figure 1**. Structural alignment between CYP2C5 and CYP BM-3 in the F-G loop region. The structure segment of CYP2C5, which is not resolved, is shown in red. The residues underlined in red form a one-turn helix in the MD simulations. Green residues in the sequence of CYP BM-3 belong to the closed conformation, blue and violet residues belong to the open conformation of CYP BM-3, blue residues were used twice.

**Figure 2**. Position of preferred hydroxylation/N-deethylation by CYP2B6 [4] is pointed out by arrows. Binding pockets A, B and C are depicted schematically.

(a) cyclophosphamide (CPA), R-enantiomer

(b) ifosfamide (IFA), S-enantiomer

**Figure 3**. Overview of the MD simulations performed for modelling the F-G loop in CYP2C5 and energies of the displayed loop structures in vacuum.

Starting structure: Helices from crystal structure of CYP2C5 are shown in yellow, structure segments from CYP BM-3 in the open conformation in red, from CYP BM-3 in the closed conformation in green.

**Figure 4**. Sequence alignment between CYP2B6 and the soluble, recombinant form of CYP2C5. The insertion in CYP2B6 is coloured in red, the segments of CYP2C5 without experimental structure are shown in blue.

**Figure 5**. Root mean square deviation (RMSD) between the backbone atoms of CYP2B6 model and the minimized CYP2B6 model after heating and releasing the backbone

**Figure 6**. Homology model of CYP2B6. Hydrophobic side chains of F-G loop are shown in green, the heme group is coloured in blue, the insertion in the H-I loop is marked in orange.

**Figure 7.** Active site of CYP2B6 with docked substrates. Amino acid side chains are coloured (pocket A: orange; pocket B: blue; pocket C: violet). Position in the substrates, where reaction can occur, is pointed out by the arrows.
(a) CPA in a position for C4 hydroxylation
(b) CPA in a position for N-deethylation
(c) IFA in a position for N-deethylation
(d) IFA in a position for C4 hydroxylation

**Supplementary material**

The coordinates of the model of CYP 2B6 (average structure of MD trajectory, energy minimized) is available as supplementary material.

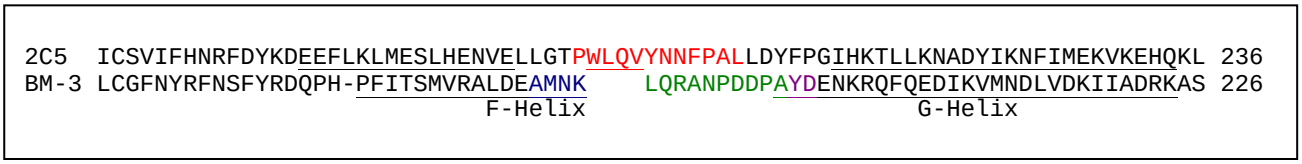
2C5 ICSVIFHNRFDYKDEEFLKLMESLHENVELLGTPWLQVYNNFPALLDYFPGIHKTLLKNADYIKNFIMEKVKEHQKL 236
BM-3 LCGFNYRFNSFYRDQPH-PFITSMVRALDEAMNK LQRANPDDPAYDENKRQFQEDIKVMNDLVDKIIADRKAS 226
F-Helix
G-Helix

Figure 1

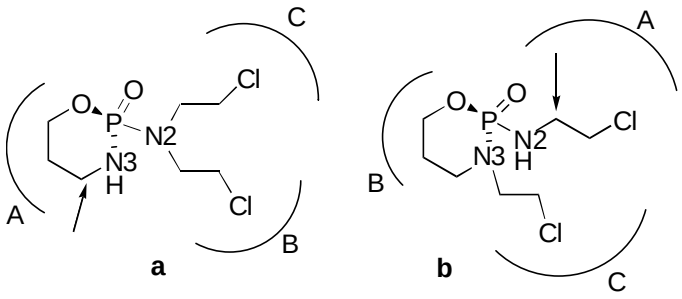


Figure 2

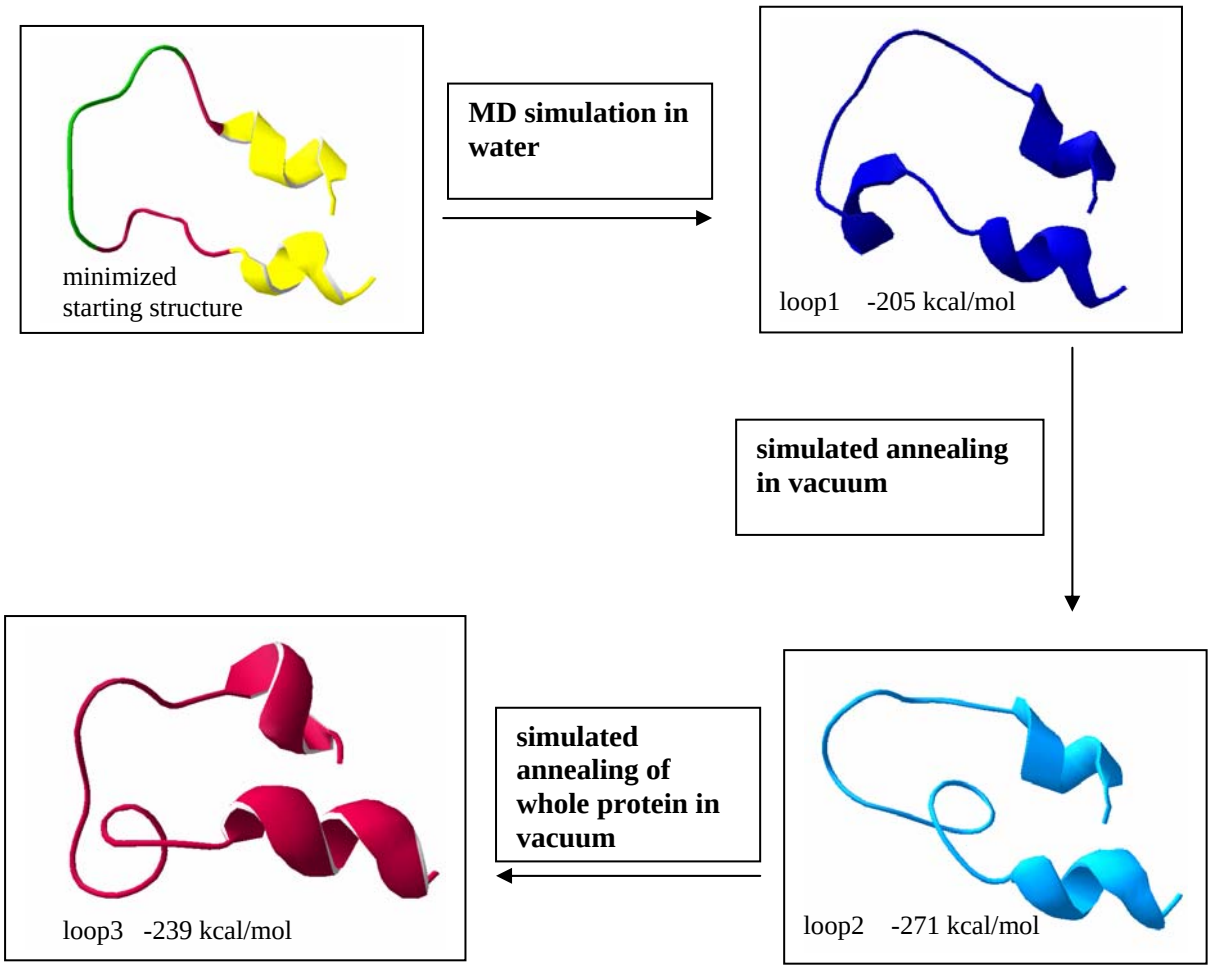
minimized
starting structure
MD simulation in water
loop1 -205 kcal/mol
simulated annealing in vacuum
loop2 -271 kcal/mol
simulated annealing of whole protein in vacuum
loop3 -239 kcal/mol

Figure 3

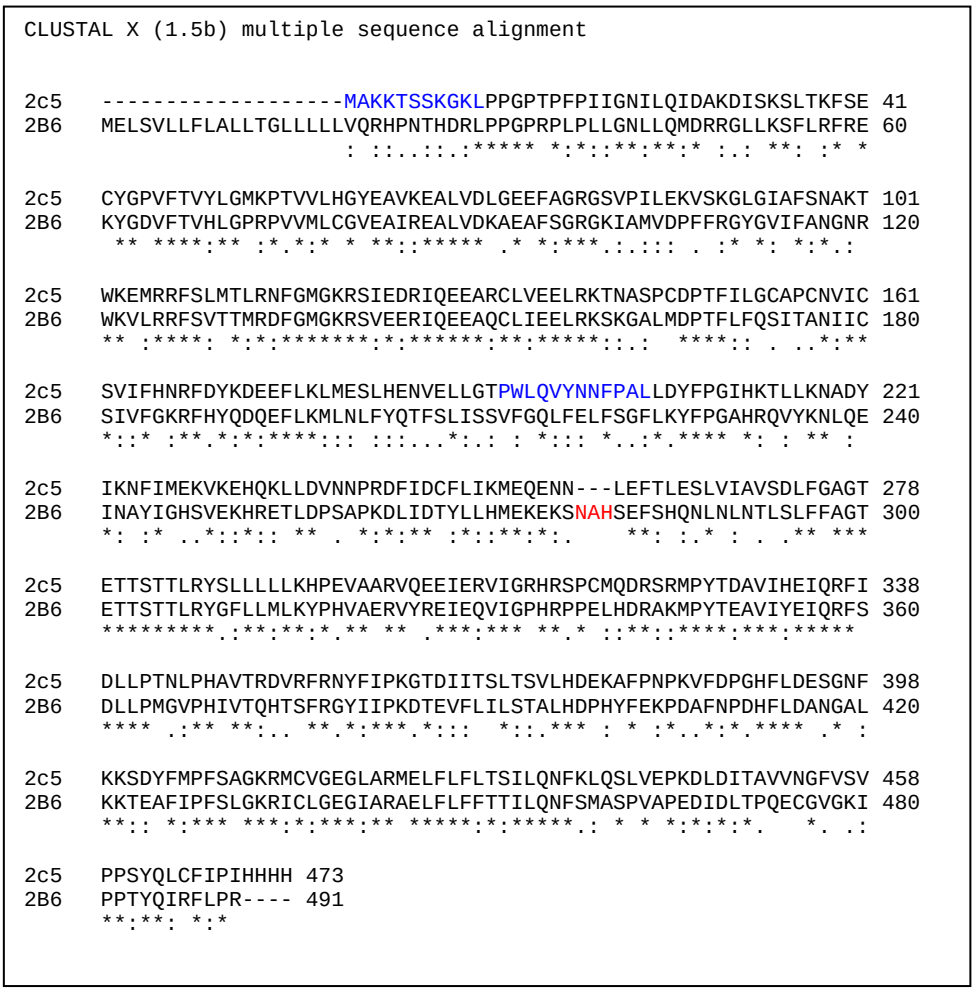
CLUSTAL X (1.5b) multiple sequence alignment
2c5 ------------------MAKKTSSKGKLPPGPTPFPIIGNILQIDAKDISKSLTKFSE 41
2B6 MELSVLLFLALLTGLLLLLVQRHPNTHDRLPPGPRPLPLLGNLLQMDRRGLLKSFLRFRE 60
2c5 CYGPVFTVYLGMKPTVVLHGYEAVKEALVDLGEEFAGRGSVPILEKVSKGLGIAFSNAKT 101
2B6 KYGDVFTVHLGPRPVVMLCGVEAIREALVDKAEAFSGRGKIAMVDPFFRGYGVIFANGNR 120
2c5 WKEMRRFSLMTLRNFGMGKRSIEDRIQEEARCLVEELRKTNASPCDPTFILGCAPCNVIC 161
2B6 WKVLRRFSVTTMRDFGMGKRSVEERIQEEAQCLIEELRKSKGALMDPTFLFQSITANIIC 180
2c5 SVIFHNRFDYKDEEFLKLMESLHENVELLGTPWLQVYNNFPALLDYFPGIHKTLLKNADY 221
2B6 SIVFGKRFHYQDQEFLKMLNLFYQTFSLISSVFGQLFELFSGFLKYFPGAHRQVYKNLQE 240
2c5 IKNFIMEKVKEHQKLLDVNNPRDFIDCFLIKMEQENN---LEFTLESLVIAVSDLFGAGT 278
2B6 INAYIGHSVEKHRETLDPSAPKDLIDTYLLHMEKEKSNAHSEFSHQNLNLNTLSLFFAGT 300
2c5 ETTSTTLRYSLLLLLKHPEVAARVQEEIERVIGRHRSPCMQDRSRMPYTDAVIHEIQRFI 338
2B6 ETTSTTLRYGFLLMLKYPHVAERVYREIEQVIGPHRPPELHDRAKMPYTEAVIYEIQRFS 360
2c5 DLLPTNLPHAVTRDVRFRNYFIPKGTDIITSLTSVLHDEKAFPNPKVFDPGHFLDESGNF 398
2B6 DLLPMGVPHIVTQHTSFRGYIIPKDTEVFLILSTALHDPHYFEKPDAFNPDHFLDANGAL 420
2c5 KKSDYFMPFSAGKRMCVGEGLARMELFLFLTSILQNFKLQSLVEPKDLDITAVVNGFVSV 458
2B6 KKTEAFIPFSLGKRICLGEGIARAELFLFFTTILQNFSMASPVAPEDIDLTPQECGVGKI 480
2c5 PPSYQLCFIPIHHHH 473
2B6 PPTYQIRFLPR---- 491

Figure 4

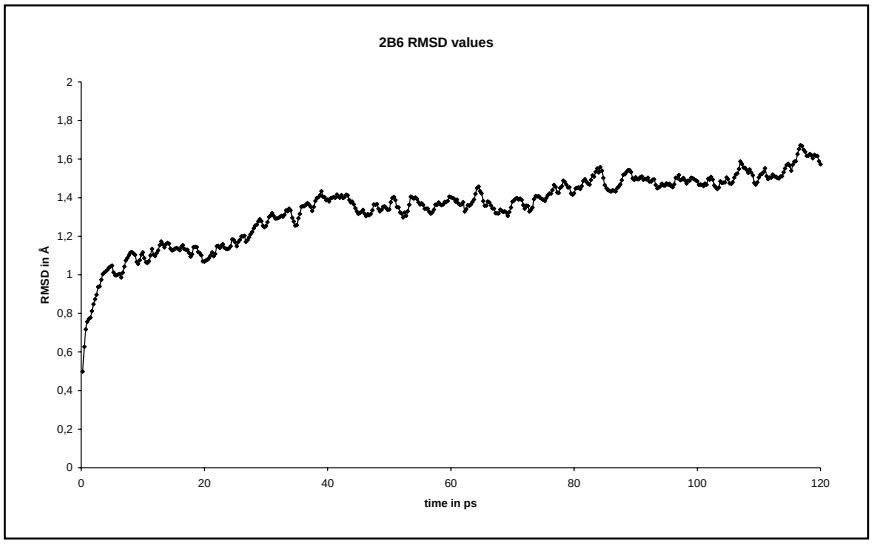
2B6 RMSD values
RMSD in Å
2
1,8
1,6
1,4
1,2
1
0,8
0,6
0,4
0,2
0
0
20
40
60
80
100
120
time in ps

Figure 5

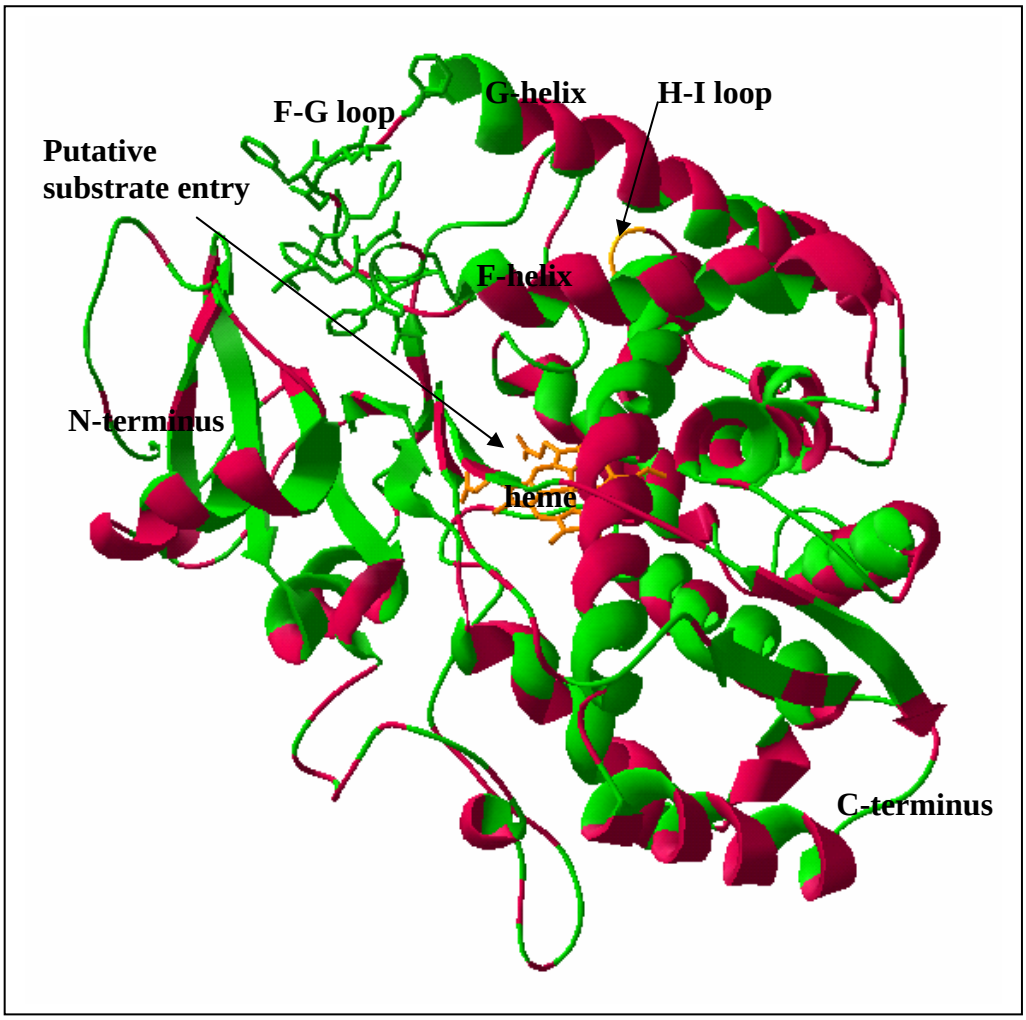
G-helix
H-I loop
F-G loop
Putative
substrate entry
F-helix
N-terminus
heme
C-terminus

Figure 6

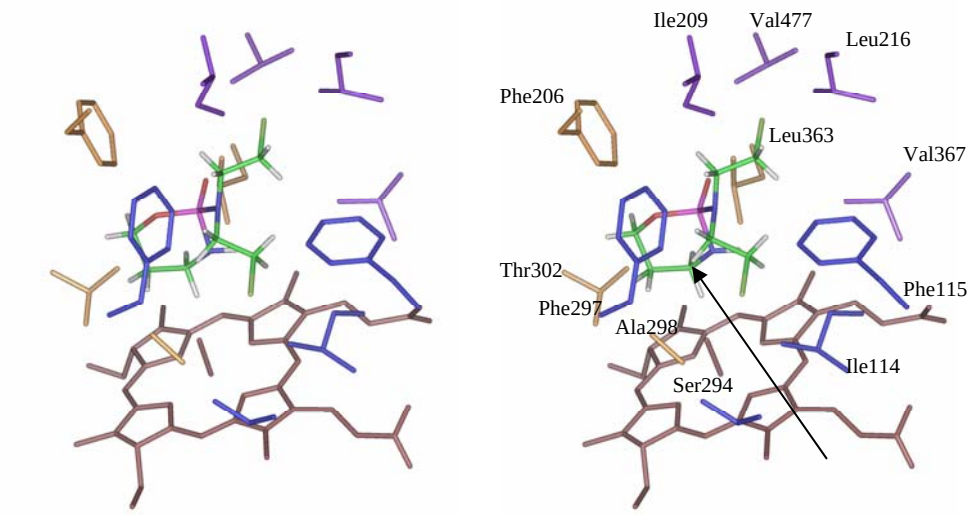
Ile209
Val477
Leu216
Phe206
Leu363
Val367
Thr302
Phe115
Phe297
Ala298
Ile114
Ser294

Figure 7a

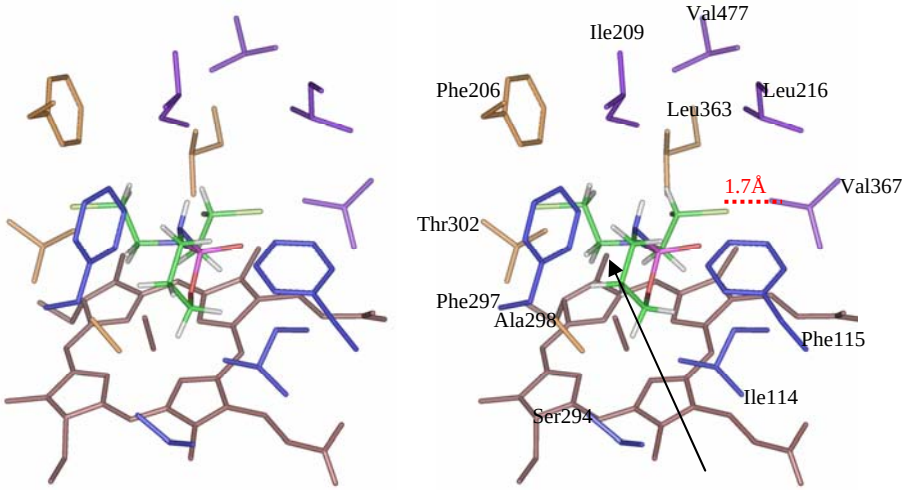
Val477
Ile209
Phe206
Leu216
Leu363
1.7Å
Val367
Thr302
Phe297
Ala298
Phe115
Ile114
Ser294

Figure 7b

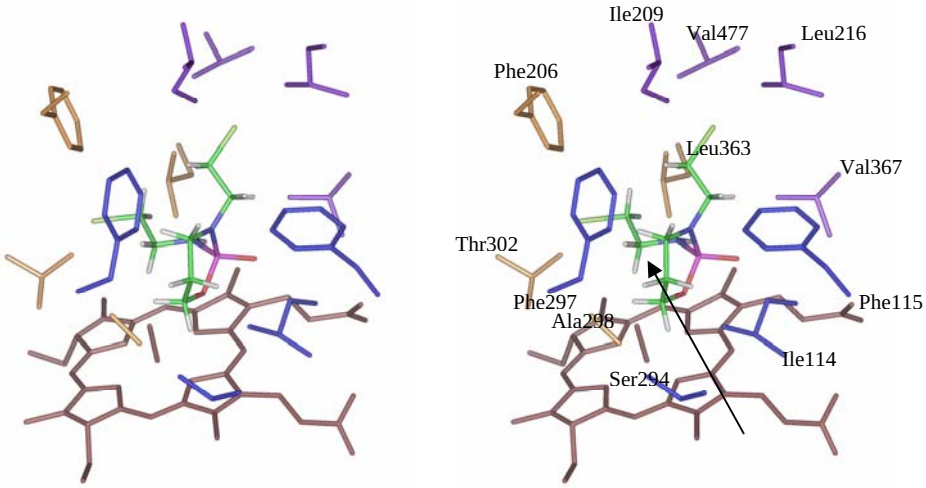
Ile209
Val477
Leu216
Phe206
Leu363
Val367
Thr302
Phe297
Ala298
Phe115
Ile114
Ser294

Figure 7c

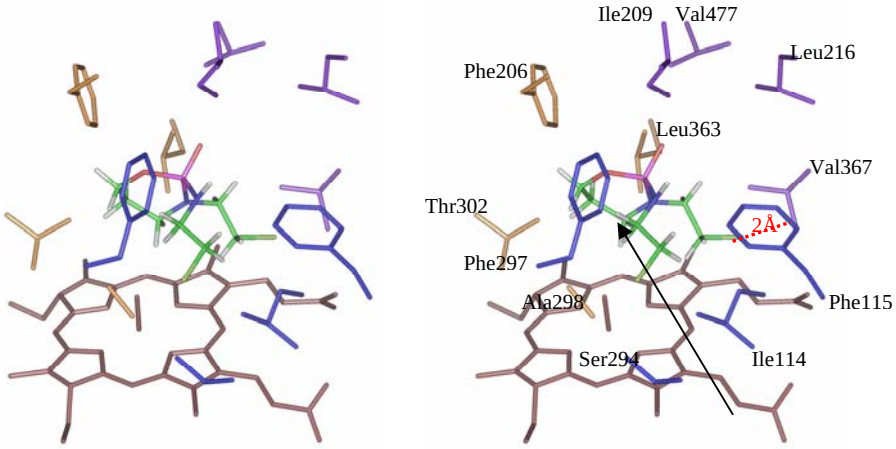
Ile209
Val477
Leu216
Phe206
Leu363
Val367
Thr302
2Å
Phe297
Ala298
Phe115
Ser294
Ile114

Figure 7d