A Model of Specificity and Selectivity of Mammalian Cytochrome P450 Monooxygenases

Alexander Seifert, Michael Krahn, Stephan Tatzel, Rolf D. Schmid, Jürgen Pleiss

Institute of Technical Biochemistry
University of Stuttgart
Allmandring 31
70569 Stuttgart
Alex_Seifert@gmx.de
Michael.Krahn@itb.uni-stuttgart.de
Stephan.Tatzel@itb.uni-stuttgart.de
Rolf.D.Schmid@itb.uni-stuttgart.de
Juergen.Pleiss@itb.uni-stuttgart.de

Abstract: Multiple molecular dynamics simulations and a systematic analysis of sequence and structure of mammalian cytochrome P450 monooxygenases were performed to investigate the structural basis of their specificity and selectivity. While the substrate binding cavity is mobile, the protein core and the access funnel to the heme are rigid. High mobility of the substrate binding pocket is consistent with the broad substrate profile observed for these enzymes, while the rigid core mediates regioselectivity by controlling substrate access to the heme. For enzymes with narrow heme access funnels, only highly accessible positions in substrates are accepted, while for enzymes with a more exposed heme regioselectivity is driven by chemical reactivity of the substrate.

1 Introduction

Cytochrome P450 monooxygenases (E.C.1.14.14.1, E.C.1.14.13, and others) are involved in the metabolism of physiologically important compounds in microorganisms,

plants, and animals. In mammals these enzymes participate in the detoxification of a broad range of xenobiotics such as environmental toxins and drugs. Understanding the factors involved in P450 substrate selectivity is of considerable interest, particularly to the pharmaceutical industry [Sm98], [SAJ97a], [SAJ97b] where an early prediction of likely drug metabolism pathways in *Homo sapiens* can aid in the design and development of drugs. It has been found that some compounds are substrates for more than one P450 isoform, while regioselectivity might differ: dextromethorphan, an antitussive, is metabolized by both CYP2B6 and CYP2D6. CYP2B6 metabolizes dextromethorphan to methoxymorphinan by catalyzing the *N*-demethylation, whereas CYP2D6 catalyzes the *O*-demethylation to dextrorphan [YH01]. In 2003 the crystal structure of the human microsomal cytochrome P450 CYP2C9 was published [Wi03]. This enzyme metabolizes 16% of all therapeutics in current clinical use. Human CYP2C9 has been crystallized in the absence and presence of its substrate (*S*)-warfarin. While the comparison of the substrate free structure and the complex with (*S*)-warfarin did not show major conformational changes, for other P450s the comparison of free and substrate-complexed crystal structures revealed that helices F and G, the F-G and the B-C loop undergo large, adaptive changes in the presence of substrates and inhibitors.

To understand the molecular basis of substrate specificity, stereo- and regioselectivity, the enzyme-substrate complex has been studied by X-ray analysis, docking, and molecular dynamics simulations [WSS04], [Ya04]. In all P450s, the cofactor heme is deeply buried inside the protein at the bottom of a large, internal binding cavity. It has been suggested that substrate access into the binding cavity and the orientation of the bound substrate relative to the heme determine specificity and selectivity. The fact that human P450s metabolize substrate molecules of various size and shape implies a large, highly flexible substrate binding cavity. However, in most cases the catalyzed reaction is highly regioselective which would require a more rigid enzyme. To solve this contradiction and to identify the molecular basis of substrate specificity and regioselectivity, we combined extensive molecular dynamics simulations and a systematic comparison of sequence and structure of mammalian P450s.

2 Molecular dynamics simulations of human CYP2C9

Molecular dynamics simulations (MD) were used to locate the rigid and the mobile parts of the enzyme. The rigid regions of the protein are assumed to have a decisive effect on the regioselectivity of the enzyme. To this aim 6 MD simulations (3 ns each) of CYP2C9 in water as well as 16 simulations (5 ns) of CYP2C9 in complex with its substrate *(S)*-warfarin in water were carried out with different distributions of the initial velocities. As initial structure for the simulations of the free enzyme and the enzyme-substrate complex, the energy minimized crystal structures of free enzyme and the substrate complex were used [Wi03]. For the setup the systems and the calculation of the trajectories the AMBER 7.0 [Ca02] program package was used. The protonation states at pH 7.0 of all histidines, glutamic and aspartic acids were calculated by the program MCCE [AG97], of the other titratable amino acids by the program xleap of the AMBER 7.0 program [Ca02]. The partial charges of the substrate and the heme group in the oxyferryl state (Fe^{3+}=O) were derived by fitting partial charges using the RESP program of AMBER 7.0 to the electrostatic potential derived by ab initio geometry optimization on an HF/6-31G* level by using the GAUSSIAN 98 program (Florian Barth, personal communication). The oxyferryl state (Fe^{3+}=O) is considered to be the main reactive species of the P450 catalytic cycle. The PMEMD program of AMBER 7.0 was used for minimization and molecular dynamics simulations. The initial structures were first energy minimized for 2000 steps (1000 steepest descent and 1000 conjugate gradient) and then heated to a temperature of 300 K. The SHAKE algorithm [RCB77] was applied to all bonds containing hydrogen atoms, and a time step of 1 fs was used. The Berendsen method was used to couple the system to constant temperature and pressure [Be84]. For each system, multiple MD simulations (6 and 16 simulations for the free and complexed enzyme, respectively) were performed with different initial random velocity distributions. For the free enzyme, the C_α atoms were restrained for 6 ps using a harmonic potential with a gradually decreasing force constant from 10 to 0.1 kcal/mol followed by a unrestrained simulation of 3 ns. For the complex, the C_α atoms were restrained for 6 ps, the substrate atoms for 600 ps with a gradually decreasing force constant from 10 to 0.1 kcal/mol followed by a unrestrained simulation of 4.5 ns.

During each simulation snapshots of the system were taken every 500 fs and stored to a trajectory file. To analyse the trajectories, the calculation of the root mean square deviation (RMSD) of the backbone atoms between each conformer and the X-ray structure and between all conformers (2D-RMSD), measurements of atom-atom distances, and the generation of average structures were done with the ptraj program of AMBER 7.0 by fitting each conformer to the initial structure.

To evaluate the deviation of the trajectory from the initial structure, the RMSD was monitored along the trajectory. For all simulations, the protein core (residues 48-57, 64-68, 72-92, 117-133, 141-155, 162-166, 285-328, 335-339, 347-371, 385-401, 406-412, 421-457, 469-473, 477-490) was stable during the simulations (Fig. 1).

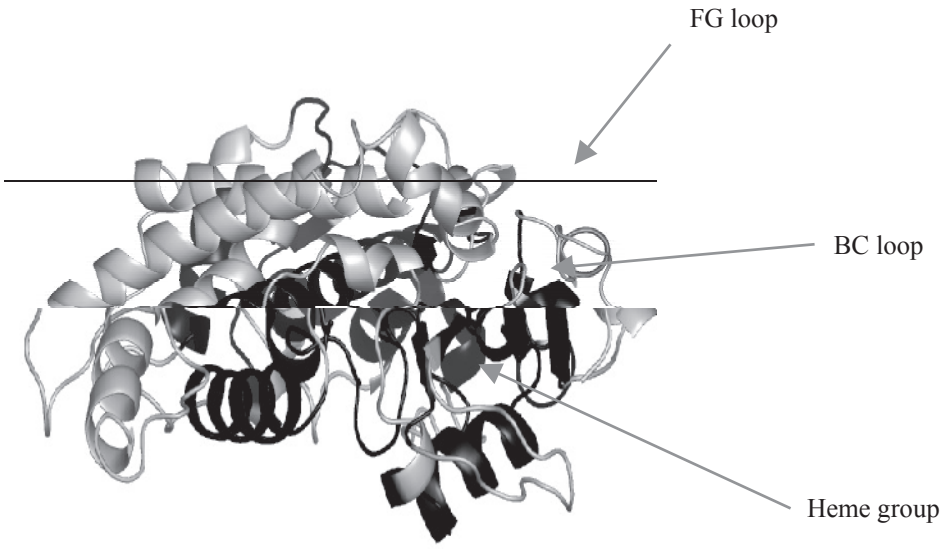

Figure 1: Location of the mobile regions (grey) of CYP2C9 identified during simulations with and without substrate. The stable protein core is shown in black. The active site (heme group) is embedded in the rigid protein core.

After fitting the core region the backbone RMSD for all simulations was between 1.5 and 2.0 Å. 2D-RMSD plots of each simulation revealed that this deviation occurred during the first ns of the simulations; after 1.5 ns the core became equilibrated as the deviations between the conformers were less than 1.5 Å for each simulation. The average structure of all simulations of the free enzyme only deviates by 1.6 Å and for the complex 1.2 Å from the X-ray structure, much less than the individual simulations. In contrast, the deviation of all backbone atoms during the simulation was considerably larger: 2.0 - 2.5 Å and 2.0 - 3.0 Å for the free enzyme and the substrate complex, respectively, which indicates that there are highly mobile regions outside the core.

The program VOIDOO [KJ94] was used to calculate the solvent accessible surface of the substrate binding cavity during snapshots of the different simulations. The comparison of the shape of the cavity reveals that the region embedding the heme and a funnel leading from the binding cavity to the heme remained unchanged during the different simulations. This funnel is formed mainly by amino acids found to belong to the stable protein core (Ala297, the highly conserved Thr301, Leu362, Leu366, and Cys435). However, the conformation of the rest of the substrate binding cavity is massively changing between the different trajectories. This region is formed mainly by helices F and G and the loop between them (FG loop) as well as the loop between the B and C helices (BC loop) which contains the B´ helix. These structural elements were found to belong to the mobile regions of the protein. After equilibration, the RMSD values for the BC loop were 1.2- 3.5 Å and 1.8- 2.9 Å for the free enzyme and substrate complex, respectively, upon fitting the core region. For the FG loop values between 2.0 Å and 4.5 Å (enzyme substrate complex) and 2.0-4.0 Å (free enzyme) were measured. Thus, these loops are highly mobile. Additional mobile elements involved in the formation of substrate binding cavity are the N-terminal residues, the turn in the C-terminal antiparallel β-sheet, and the turn in β_1 sheet.

Consistent with the high flexibility of the substrate binding cavity, the human P450s metabolize substrate molecules of various size and shape. On the other hand, they are highly regioselective which in turn implies a more rigid enzyme. CYP2C9 as a member of the human drug metabolising enzyme family possesses both properties: the enzyme has mobile elements at the protein surface which are involved in the formation of the substrate binding cavity. These mobile elements might be able to adapt for substrates of different size and shape. On the other hand the enzyme possesses a stable and more rigid protein core which contains the active site. Core amino acids surrounding the active site are forming a narrow funnel, which allows the substrate to reach the active heme oxygen in selected orientations only, resulting in a defined regioselectivity.

3 Structure comparison: the P450 Engineering Database

The P450 Engineering Database (PED) integrates information on sequences, structure, and function of P450 monooxygenases. The design of the relational data model is based on the Lipase Engineering Database [FP03]. Sequence data on 2413 protein entries are assigned to 934 homologous families which are grouped into 462 superfamilies, with 128 experimental crystal structures of 21 proteins from 18 different homologous families. The internal protein sequence classification is based on sequence similarity. The PED has been applied to systematic analysis of sequence-structure-function relationships of the vast and diverse P450 enzyme class to identify functional relevant residues and regions contributing to substrate specificity and regioselectivity.

P450 crystal structures show that the heme is embedded between two structural elements, the conserved I helix and the variable substrate recognition site 5 (SRS5) [Go92]. A systematic comparison of P450 binding cavity architecture using VOIDOO [KJ94] in combination with our database approach allowed to assign each mammalian P450 to either of two classes with a short or a long β1-4 strand [Me05]. In enzymes with a short β1-4 strand (3 residues), it is preceded by a loop region (Fig. 2) which narrows the funnel and thus reduces heme accessibility. In enzymes with a long β1-4 strand (4-6 residues), the heme is fully exposed to the substrate binding cavity without steric hindrance.

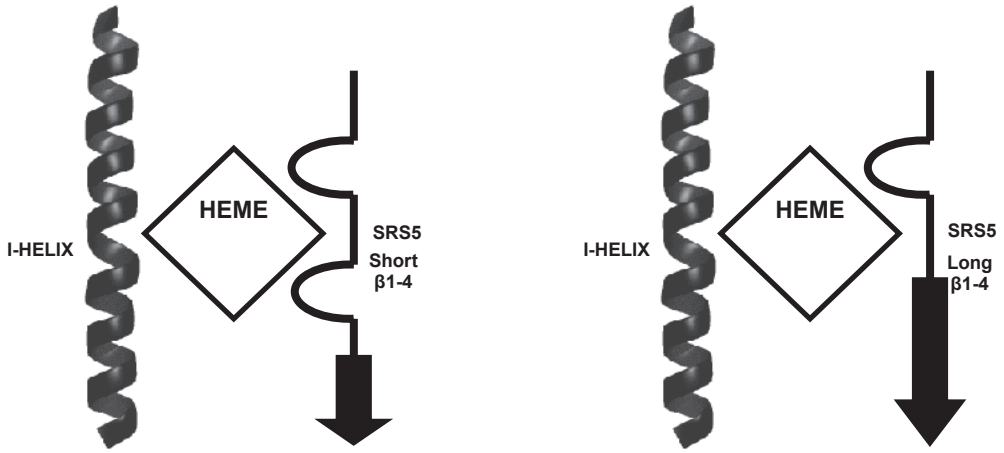

Figure 2: Effect of the β1-4 strand on heme accessibility. The β1-4 is part of the SRS5 and can extend from 3 to 6 amino acids.

While residues within the strand orient their side chains away from the funnel, residues in the preceding loop region point towards the catalytic centre and consequently determine the diameter of the funnel. Therefore, in enzymes with a short β1-4 strand the accessibility to the reactive heme group is determined by the size of four hydrophobic residues within the active site cavity. Molecular dynamics simulations showed they are part of the rigid protein core. In fact, two of these four residues, an alanine and a catalytically active threonine, are part of the I-Helix and are highly conserved in most mammalian and bacterial P450 monooxygenases [Mo95], while the two other positions in SRS5 are variable [PG01].

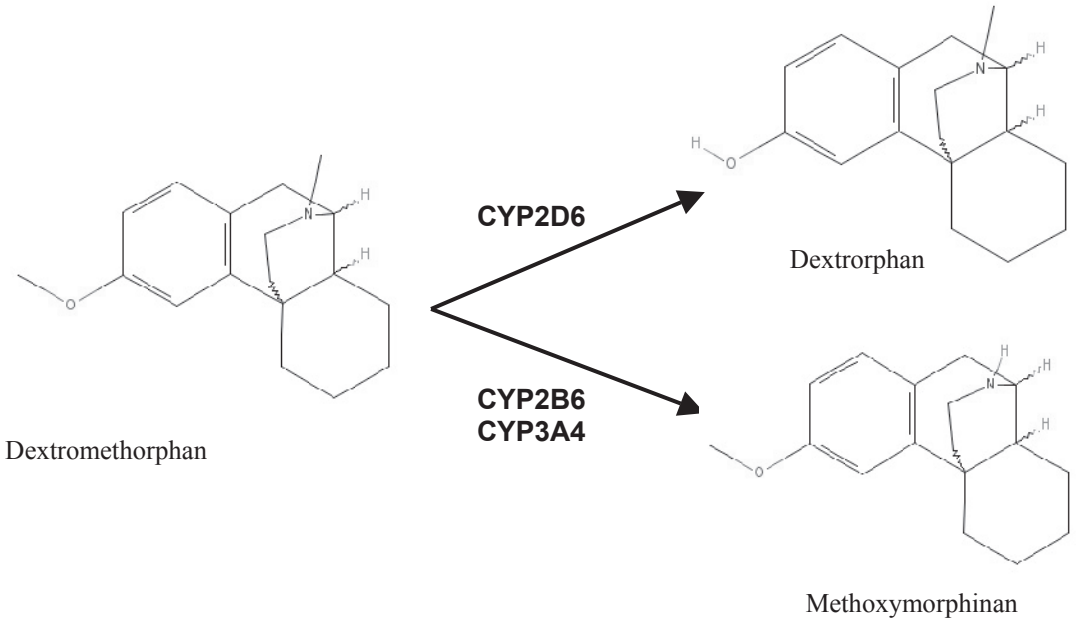

Figure 3: Reactions of dextromethorphan catalyzed by human P450s.

According to their heme accessibility, there are two types of enzymes: P450s with a narrow (<5 Å) heme access channel (e.g. CYP2D6), and P450s with a wide (>6 Å) heme access channel (e.g. CYP2B6) or with a fully exposed heme (e.g. CYP3A4). To explain the experimental observation that CYP2D6 catalyses *O*-demethylation of dextrometorphan, while CYP2B6 and CYP3A4 catalyse *N*-demethylation (Fig. 3), the analysis of binding site geometry was combined with an estimation of chemical reactivity of dextrometorphan by calculating the bond dissociation energy (BDE) using the PETRA program [Ga88]. According to the BDE calculation, the highest reactivity is predicted at positions 9 and 10 (342 and 353 kJ/mol, respectively) (Fig. 4). However, none of the three P450s metabolise the molecule in these positions, supposedly because of steric hindrance. Positions 18 and 17, the sites of *O*- and *N*-demethylation, respectively, are less sterically hindered. They differ in their reactivity and in their accessibility: While position 18 is less reactive than position 17 (418 and 364 kJ/mol, respectively), its methyl group is more accessible. As a consequence, the two enzymes with a wide funnel (CYP2B6) or a fully exposed heme (CYP3A4) prefer demethylation at position 17, while the enzyme with the narrow funnel (CYP2D6) demethylates the substrate at position 18. Thus, a combination of accessibility of the heme and the chemical reactivity of the substrate are sufficient to explain regioselectivity. This classification scheme seems to be generally applicable and has been applied to explain regioselectivety of verampamil and propafenon metabolism by other human P450s.

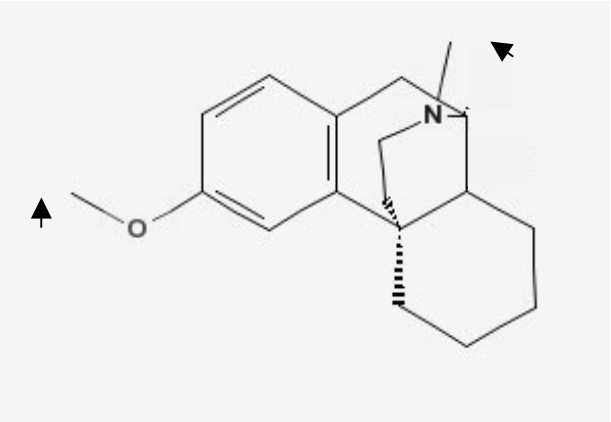

Figure 4: Reactivity of dextromethorphan.

References

[AG97] Alexov, E. G.; Gunner, M. R.: Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. Biophys J, 1997; 72: 2075-93

[Be84] Berendsen, H. J. C., et al.: Molecular dynamics with coupling to an external bath. J Chem Phys, 1984; 81: 3684-3690

[Ca02] Case, D. A., et al.: AMBER 7. University of California, San Francisco. 2002.

[FP03] Fischer, M.; Pleiss J.: The Lipase Engineering Database: a navigation and analysis tool for protein families. Nucleic Acids Res, 2003; 31: 319-321

[Ga88] Gasteiger, J.: Empirical Methods for the Calculation of Physicochemical Data of Organic Compounds. Physical Property Prediction in Organic Chemistry, 1988; 119-138

[Go92] Gotoh, O.: Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. J Biol Chem, 1992; 267: 83-90

[KJ94] Kleywegt, G. J.; Jones, T. A.: Detection, delineation, measurement and display of cavities in macromolecular structures. Acta Crystallogr D Biol Crystallogr, 1994; 50: 178-185

[Me05] Mestres, J.: Structure conservation in cytochromes P450. Proteins, 2005; 58: 596-609

[Mo95] Montellano, O.: Cytochromes P450: Structure, Mechanism and Biochemistry. 1995.

[PG01] Pei, J.; Grishin, N. V.: AL2CO: Calculation of positional conservation in a protein sequence alignment. Bioinformatics, 2001; 17: 700-712

[RCB77] Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C.: Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. Computat Phys, 1977; 23: 327-341

[Sm98] Smith, D. A., et al.: Human cytochrome P450s: selectivity and measurement in vivo. Xenobiotica, 1998; 28: 1095-1128

[SAJ97a] 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

[SAJ97b] 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

[WSS04] Wade, C. W. W., P.J. Schlichtling, I. Sudarko,: A survey of active site access channels in cytochromes P450. Journal of Inorganic Biochemistry, 2004; 98: 1175-1182

[Wi03] Williams, P. A., et al.: Crystal structure of human cytochrome P450 2C9 with bound warfarin. Nature, 2003; 424: 464-468

[Ya04] Yano, J. K., et al.: The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-A resolution. J Biol Chem, 2004; 279: 38091-38094

[YH01] Yu, A.; Haining, R. L.: Comparative contribution to dextromethorphan metabolism by cytochrome P450 isoforms in vitro: can dextromethorphan be used as a dual probe for both CTP2D6 and CYP3A activities? Drug Metab Dispos, 2001; 29: 1514-1520