Identification of Selectivity Determinants in CYP Monooxygenases by Modelling and Systematic Analysis of Sequence and Structure

Alexander Seifert and Jürgen Pleiss*

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

Abstract: Cytochrome P450 monooxygenases (CYPs) form a large, ubiquitous enzyme family and are of great interest in red and white biotechnology. To investigate the effect of protein structure on selectivity, the binding of substrate molecules near to the active site was modelled by molecular dynamics simulations. From a comprehensive and systematic comparison of more than 6300 CYP sequences and 31 structures using the Cytochrome P450 Engineering Database (CYPED), residues were identified which are predicted to point close to the heme centre and thus restrict accessibility for substrates. As a result, sequence-structure-function relationships are described that can be used to predict selectivity-determining positions from CYP sequences and structures. Based on this analysis, a minimal library consisting of bacterial CYP102A1 ($P450_{BM3}$) and 24 variants was constructed. All variants were functionally expressed in *E. coli*, and the library was screened with four terpene substrates. Only 3 variants showed no activity towards all 4 terpenes, while 11 variants demonstrated either a strong shift or improved regio- or stereoselectivity during oxidation of at least one substrate as compared to CYP102A1 wild type. The minimal library also contains variants that show interesting side products which are not generated by the wild type enzyme. By two additional rounds of molecular modelling, diversification, and screening, the selectivity of one of these variants for a new product was optimised with a minimal screening effort. We propose this as a generic approach for other CYP substrates.

Keywords: CYPED, MD-simulations, protein engineering, minimal mutant library.

INTRODUCTION

Cytochrome P450 monooxygenases (CYPs) catalyse the oxidation of inert hydrocarbons under environmentally friendly conditions utilising molecular oxygen [1, 2]. Members of this vast and diverse class of enzymes participate in the detoxification of a broad range of xenobiotics, including drugs. The participation of human CYP isoforms in drug metabolism makes them prime targets for pharmacokinetic studies for the pharmaceutical industry. In drug development, it is desired to predict possible oxidation products of a drug candidate early in the process. In human liver, three aspects affect the CYP mediated metabolism of a drug: CYP activity, amount of enzyme, and interaction with the electron delivering reductase. Only a hand full of different CYP isoforms metabolises the majority of drugs [3]. These isoforms differ in their range of accepted substrates and in their products (regio-, chemo-, and stereoselectivity). Understanding the molecular basis of selectivity and specificity control in CYPs is a prerequisite for the prediction of likely drug metabolism pathways in *Homo sapiens*. In personalised medicine, the impact of naturally occurring single nucleotide polymorphisms (SNPs) on drug metabolism is taken into account. Methods have been described that predict whether a nonsynonymous SNP (nsSNP) in a CYP coding region is deleterious or neutral to protein function [4, 5]. However, to predict from CYP sequences whether nsSNPs have an impact on substrate specificity and selectivity is still difficult.

The fact that CYPs catalyse a broad variety of reactions [6-8] implies a great potential for applications in white biotechnology, e.g. for the production of fine chemicals, drugs and drug metabolites, flavour and aromatic compounds, and for bioremediation [9, 10]. However, the application of CYPs in preparative synthesis is still limited due to their limited stability and activity [11]. Most of the CYPs depend on a reductase for oxygen activation [12] and, for the majority of them, the corresponding reductase is not known. As an alternative, self-sufficient CYP102A1 ($P450_{BM3}$) is a fusion protein composed of a monooxygenase and a reductase domain [13]. It is relatively stable under process conditions [14] and several mutations have been described that widen the substrate scope [15, 16]. However, the wild type and many of the variants often convert substrates with low selectivity or the product of interest is not obtained. Here, the challenge is to engineer the enzyme's regio-, chemo-, and stereoselectivity to generate variants that are able to produce the desired oxidation products selectively. CYPs possess a large, variable and flexible substrate binding cavity [17-19]. This fact is also reflected by the existence of 6 substrate recognition sites (SRS) with lengths of up to 10 amino acids [20]. The challenge is to identify among many potential substrate interacting residues those positions that are in contact with every substrate (independently of its shape and size) during the attack of the activated oxygen, since residues in these position control substrate orientation during catalysis and, therefore, selectivity and specificity. Here we show that studying enzyme-substrate interactions by molecular dynamic simulations, followed by the systematic analysis of CYP sequence and structure, can unveil sequence-structure-function relationships which can be used to predict selectivity-determining positions from sequence and structure. Focussing mutagenesis on these hotspots is a very effective route to selective biooxidation catalysts.

THE CYTOCHROME P450 ENGINEERING DATABASE (CYPED)

A still increasing number of CYP sequences and structures have been accumulated in public databases such as NCBI (http://www.ncbi.nlm.nih.gov) and the PDB (http://www.pdb.org/pdb/) [21]. The CYPED (http://www.cyped.uni-stuttgart.de) was designed as a navigation tool for a comprehensive and systematic analysis of CYP sequences and structures [22]. An extended version additionally contains biochemical information on substrates and inhibitors, information on the effect of single nucleotide polymorphisms, as well as the annotation of structurally and functionally relevant residues [23]. The latest version of the database contains information on 8613 sequences and 47 structures of CYPs, which are grouped according to Nelson's classification [24]. The database was used as a source for CYP sequence and structure information in the following systematic analysis.

*Address correspondence to this author at the Institute of Technical Biochemistry, University of Stuttgart, Allmandring 31, 70569 Stuttgart, Germany; Tel: (+49) 711-68563191; Fax (+49) 711-68563196;
E-mail: Juergen.Pleiss@itb.uni-stuttgart.de

1875-5453/12 $58.00+.00

THE STRUCTURAL BASIS OF CYP SELECTIVITY CONTROL

Molecular dynamics (MD) simulation is a powerful tool to study enzyme-substrate-interactions in atomic detail. To identify those regions of the enzyme that are involved in selectivity control, we carried out multiple MD simulations of human CYP2C9 in complex with its substrate warfarin[25]. In the crystal structure of a CYP2C9-warfarin-complex, the substrate was bound in a large substrate binding cavity 10 Å away from the heme centre (active site), a distance unlikely to permit oxidation[18]. In subsequent MD simulations, starting from this complex the substrate approached the active site heme in an orientation that is in agreement with the experimentally observed hydroxylation at carbons C6 and C7 (Fig. **1**). Two different regions of the substrate binding cavity can be distinguished. One region is composed by mobile and flexible loops that shield the substrate binding cavity from the bulk solvent enabling the entrance of differently shaped and sized substrates and the release of products. Such conformational plasticity of structural elements involved in the formation of the substrate entrance region has been described for many different CYP enzymes.[26] The other part is the funnel-like hydrophobic heme access region which is embedded in the rigid protein core and connects the pocket containing the heme group with the rest of the substrate binding cavity (Fig. **1**). The shape and rigidity of this region restricts the accessibility of the heme and forces the Y-shaped substrate warfarin during the approach to the heme into a defined orientation, exposing selectively the most accessible atoms of the molecule (C6 and C7) to the active site.

The same approach was applied to study enzyme-substrate interactions of mutants of bacterial CYP102A1 in complex with its hydrophobic substrate α-pinene[27]. In contrast to human CYP2C9, the CYP102A1 variants A74G/F87G/L188Q and A74G/F87V/L188Q possess a rather wide and unrestricted heme access region (Fig. **2**). However, amino acid variation in position 87, which is part of the heme access region, strongly affects regio- and chemoselectivity. While variant A74G/F87G/L188Q preferentially converted α-pinene to pinene oxide, variant A74G/F87V/L188Q preferentially catalysed hydroxylation at allylic position C4. Multiple MD simulation showed that the access of α-pinene to the heme centre is not restricted by the heme access region, but the substrate is stabilised in defined orientations by hydrophobic interactions with the side chain of the amino acid in position 87. These defined orientations expose only parts of the substrate molecule to the active site, which leads to the observed selectivity. In summary, the modelling results indicate that the broad substrate specificity observed for human CYPs is conferred by mobile structural elements that shield the substrate binding cavity from the bulk solvent. However, a major determinant for selectivity is the heme access region because of its strong impact on the orientation of the substrate in the immediate vicinity of the active site heme. Owing to the close distance of this region to the heme, this dominant role is expected to apply for all potential substrates, independent of their size and shape.

PREDICTING SELECTIVITY-DETERMINING RESIDUES FROM CYP STRUCTURES

The systematic analysis of 31 CYP crystal structures (Table **1**) indicated that in all CYPs the heme access region is formed by 3 structural elements: the I-helix, the loop between the B-helix and the C-helix (BC-loop) and substrate recognition site 5 (SRS5) which extends from the conserved ExxR motif to the strand β1-4 (Fig. **2**). Amino acids positioned in this region, with their side chains pointing towards the heme centre, are expected to be in contact with the substrate molecule during the attack of the activated oxygen due to their spatial closeness to the active site heme. These protruding residues control the orientation of all potential substrates and are therefore predicted to be hotspots for determining regio-, chemo-, and stereoselectivity. As a consequence, substitutions in these positions are expected to change, decrease, or increase selectivity. The I-helix contains two such selectivity-determining residues (alanine and threonine), which are part of the conserved sequence motif AGxxT[28]. The threonine is known to be involved in proton transfer during oxygen activation[29], therefore, upon substitution of this residue a decrease in turnover rate or coupling efficiency is generally observed[30]. The alanine is not known to be involved in oxygen activation, however the size of this residue is limited due to its immediate adjacency to the heme centre and its ability to shield the heme centre from dioxygen and/or substrate. SRS5 possesses 1 - 2 selectivity-determining residues (Fig. **2**). While the SRS5 is variable in sequence and structure, it is adjacent to the highly conserved ExxR motif (Fig. **2**) and thus it can be easily located in a CYP sequence. With one exception (CYP158A2), the first selectivity-determining residue was observed in position 5 after the highly conserved ExxR motif in all CYP crystal structures[31]. A second selectivity-determining residue was observed at position 7, 8, or 9.

In CYP crystal structures, the BC-loop hosts one additional residue involved in the formation of the heme access region (Fig. **2**)[31]. The comparison of 31 crystal structures revealed that this residue of the structurally highly diverse BC-loop is conserved in its position relative to the heme centre[32]. As an example, F87 in CYP102A1 corresponds to V113 in CYP2C9 and F120 in CYP2D6.

In total, up to 5 selectivity-determining amino acid positions involved in the formation of the heme access region can be predicted from CYP crystal structures.

PREDICTING SELECTIVITY-DETERMINING RESIDUES FROM CYP SEQUENCES

For the great majority of CYP enzymes only sequence information is available. Hence, it is of broad interest to identify selectivity-determining residues from sequence alone. Owing to the high degree of sequence conservation, selectivity-determining residues positioned in the I-helix can be assigned by multiple sequence

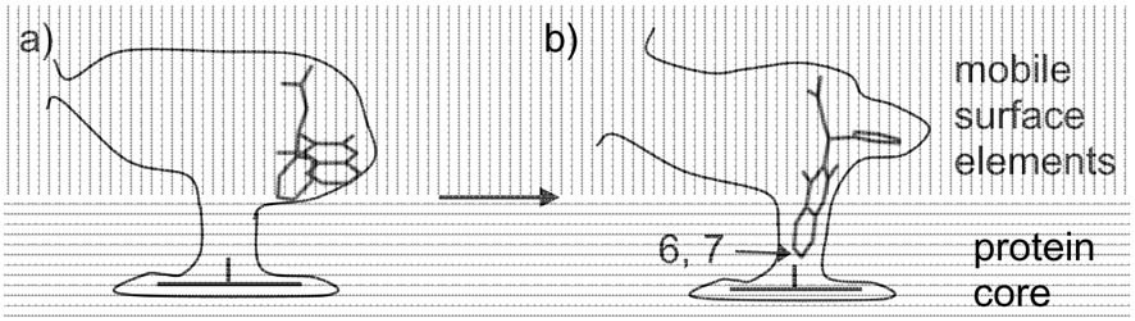

Fig. (1). Shape of the substrate binding cavity of the CYP2C9-warfarin complex in a) the crystal structure and b) after approach of the substrate to the active site heme.

Table 1. CYP Crystal Structures Analysed in this Study

Name	Organism	PDB Entry
CYP1A2	*Homo sapiens*	2HI4
CYP2A6	*Homo sapiens*	1Z10
CYP2B4	*Oryctolagus cuniculus*	1SUO
CYP2C5	*Oryctolagus cuniculus*	1NR6
CYP2C8	*Homo sapiens*	1PQ2
CYP2C9	*Homo sapiens*	1OG2
CYP2D6	*Homo sapiens*	2F9Q
CYP2R1	*Homo sapiens*	2OJD
CYP3A4	*Homo sapiens*	1TQN
CYP8A1	*Homo sapiens*	2IAG
CYP51	*Mycobacterium tuberculosis*	1E9X
CYP55A	*Fusarium oxysporum (NO2 Reductase)*	1CL6
CYP101D	*Pseudomonas putida (P450cam)*	1CP4
CYP102A1	*Bacillus megaterium*	1BU7
CYP105	*Nonomuraea recticatena*	2Z36
CYP107A	*Saccharopolyspora erythraea*	1EGY
CYP107L1	*Streptomyces venezuelae(PikC)*	2BVJ
CYP108A	*Pseudomonas sp.*	1CPT
CYP119A	*Sulfolobus solfactaricus*	1F4T
CYP121	*Mycobacterium tuberculosis*	1N40
CYP152A	*Bacillus subtilis*	1IZO
CYP154A	*Streptomyces coelicolor*	1ODO
CYP154C	*Streptomyces coelicolor*	1GWI
CYP158A2	*Streptomyces coelicolor a3(2)*	2D09
CYP165B	*Amycolatopsis orientalis*	1LFK
CYP165C	*Amycolatopsis orientalis*	1UED
CYP167A	*Polyangium cellulosum*	1Q5D
CYP175A	*Thermus thermophilus*	1N97
CYP176A	*Citrobacter braakii*	1T2B
CYP199A2	*Rhodopseudomonas palustris*	2FR7
CYP245A1	*Streptomyces sp. TP-A0274*	2Z3U

alignments including sequences with known tertiary structure. In contrast to the I-helix, the SRS5 region is highly variable in sequence, hence an unequivocal assignment of selectivity-determining residues based on sequence alignments is not feasible. However, these residues can by identified based on their proximity to the highly conserved ExxR-motif. Based on the analysis of 31 CYP crystal structure and over 6300 CYP sequences, it was predicted that in 98.4% of all CYPs, the residue at position 5 after the highly conserved ExxR motif is preferentially involved in substrate binding during oxidation and, therefore, crucial for selectivity control. The analysis of CYP crystal structures and sequences showed that a second preferentially substrate interacting residue in the SRS5 region can be found at position 9 after ExxR in 27% of all CYPs. This subgroup can be identified by the PROSITE pattern E-X-X-R-X(7)-{P}-x-P-[HKR].

A further selectivity-determining residue involved in heme access region formation originates from the BC-loop. Although easily identifiable from crystal structures, the residue is located in a highly diverse part of the protein sequence and due to a lack of an adjacent conserved sequence motif is hardly identifiable from sequence alone. In general, a way to enhance the accuracy of (multiple) alignments of sequences with low similarity is to include structure information in the alignment process. Recently, the comparison of crystal structures of 31 different CYPs revealed that this class of enzymes share a highly similar structural core[32]. Owing to the structural similarities, the core of these CYP structures could be properly aligned leading to an accurate alignment of the underlying protein sequences, which shared an average sequence identity of only 25%. From this structure-guided multiple sequence alignment, a sequence profile was derived to assign structurally conserved elements in CYP sequences. With the help of this sequence profile, the selectivity-determining residue in the BC-loop was identified in CYPs without structure information.

The ability to identify selectivity-determining positions from CYP sequence further allowed us to analyse the distribution of amino acids in these positions throughout the CYP family. The analysis of the two hotspots in the SRS5 region as well as the hotspot located in the BC-loop in several thousand CYP sequences clearly reveals a strong preference for hydrophobic residues (alanine, valine, leucine, isoleucine, and phenylalanine), while charged residues are rare exceptions[31, 32].

THE IMPACT OF SNPS ON CYP SELECTIVITY AND SPECIFICITY

Currently, it is impossible to predict the impact that small nsSNPs in CYP-coding-regions have on enzyme properties like selectivity and specificity. However, the ability to identify positions involved in the control of selectivity and specificity from protein sequences can aid to estimate the impact of nsSNPs in human CYPs on drug metabolism. A survey of experimental literature on mutational studies on a variety of different bacterial and mammalian CYPs confirmed that amino acid substitutions in the 3 positions introduced in this work: one in the BC-loop (corresponding to F87 in CYP102A1) [27, 33-37] and the two hotspots positioned in the SRS5-region (position 5 and 9 after the ExxR motif) [38-42] have a strong impact on selectivity, specificity, and activity. Although the enzyme variants described in these reports are not naturally observed, this clearly indicates that if a nsSNP is observed in a CYP-coding-region that relates to one of the positions identified as hotspots, this SNP is likely linked to a distinct phenotype (e.g. shift of substrate or metabolite spectrum), while SNPs in other regions of the gene often influence activity.

FOCUSSING MUTAGENESIS ON HOTSPOTS FOR SELECTIVITY AND SPECIFICITY: THE CONCEPT OF MINIMAL CYP MUTANT LIBRARIES IN CATALYST DEVELOPMENT

Besides estimating the impact of CYP mutations observed in nature on enzyme properties like selectivity and specificity, the ability to predict selectivity-determining residues from protein structures and sequences can aid in the process of developing selective CYP-catalysts for biooxidation. The major disadvantage of approaches that are based on random mutagenesis, such as directed evolution[43], is the huge amount of variants that have to be gener-

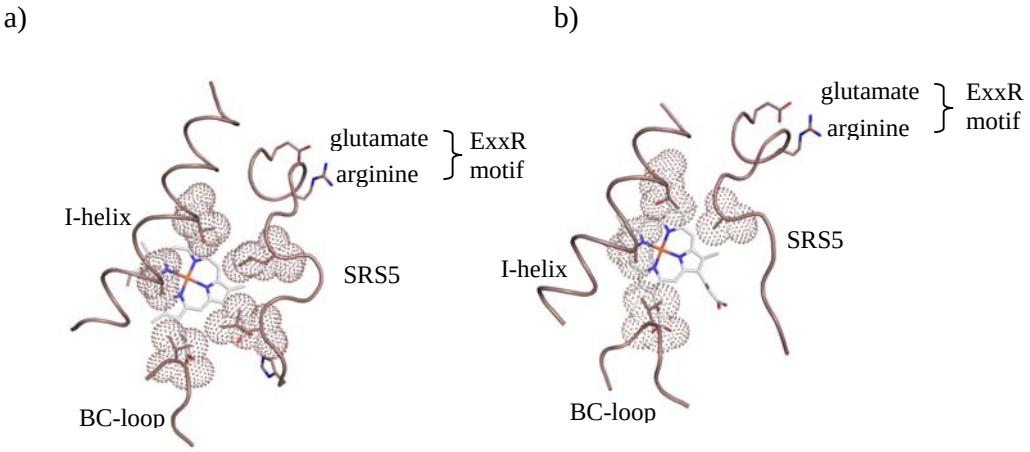

Fig. (2). Heme access region in **a**) human CYP2C9 and **b**) bacterial CYP102A1-F87V. Active site heme centre is shown in orange. Van der Waals radii of heme pointing amino acids are depicted as dot-clouds.

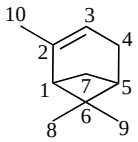

Fig. (3). (-)-α-pinene.

ated and screened. To apply such methods in the quest of changing or improving CYP regio-, chemo-, and stereoselectivity is time and resources intensive, since high-throughput screening assays for selectivity would typically require the extensive quantification of oxidation products by the use of chromatographic methods. Therefore, it is promising to reduce the screening effort by using small, highly enriched mutant libraries by focussing mutagenesis on selectivity-determining positions. Reducing drastically the number of mutated positions allows to exhaustively explore combinations of different amino acids in these positions. Apart from focussing on a few positions, library size can be reduced by limiting the amino acids that are used. Instead of taking into account all proteinogenic amino acids as substituents in hotspot positions, one can omit residues that are not observed in naturally occurring CYP sequences (thus avoiding potentially inactivating substitution) and focus on the most abundant residues. Based on the results of the analysis of thousands of CYP sequences we restricted substitutions to leucine, isoleucine, valine, alanine, and phenylalanine[31, 32]. These hydrophobic residues are different in size and shape. Thus, substitutions to these residues in hotspot positions maintain the hydrophobic character of the heme access region, while its size and shape is strongly varied. By applying this strategy, we constructed a minimal mutant library of the well known CYP102A1 from *Bacillus megaterium*, a fusion protein of a monooxygenase and a CPR-type reductase[44]. The variants plus wild type embody 25 combinations of 5 hydrophobic residues in two hotspots (F87 in the BC-loop and A328 in SRS5) (Fig. **4**). Judging from the presence of the Soret-band at 450 nm upon measuring of CO-difference spectra, all 24 variants and the wild type incorporated the heme group during expression. Thus all 24 variants are expressed as functional enzymes. Initially, this minimal library was screened with differently shaped and sized terpenes (geranylacetone, nerylacetone, (4*R*)-limonene, and (+)-valencene). The CYP102A1 wild type enzyme converted all 4 terpenes with poor regioselectivity. While only 3 out of 24 variants showed no activity towards any of the tested terpenes, 11 variants demonstrated either a strong shift or improved regio- or stereoselectivity during oxidation of at least one substrate as compared to CYP102A1 wild type.

In a subsequent study, the minimal library was screened with inert cyclic and acyclic alkanes and identified variants with high selectivity for 2-hydroxylation of octane[45].

A more detailed look at the results of the (4*R*)-limonene screening shows two variants that highly selectively form (4*R*)-limonene-8,9-epoxide, while the CYP102A1 wild type enzyme is highly unselective (4 products resulting from oxidation at activated positions). However, there are two additional activated positions in the

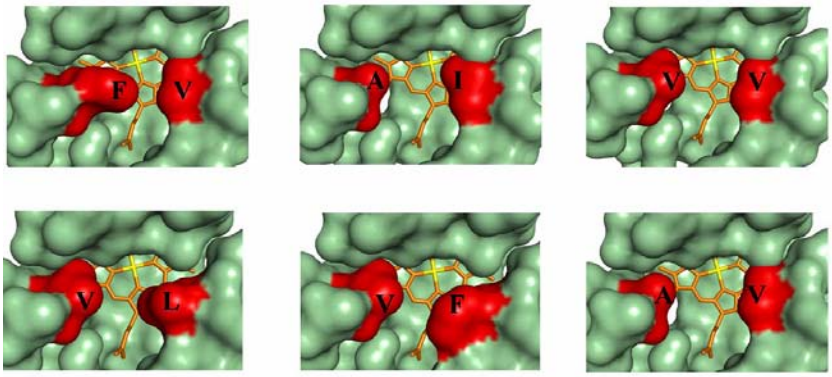

Fig. (4). Shape of the substrate binding site near the heme (yellow) of selected CYP102A1 minimal library variants. The mutated hotspot positions (87 (left) and 328 (right)) are shown in red. Variants A328V (FV), F87A/A328I (AI), F87V/A328V (VV), F87A/A328V (AV), F87V/A328F (VF), and F87V/A328L (VL) possess highly diverse substrate binding site shapes. Models are based on the CYP102A1 crystal structure (PDB entry 1bu7 chain A) and were generated with the Pymol v0.99 program [47].

substrate (4*R*)-limonene that are not addressed by the wild type enzyme. The minimal library also contains variants that form these new oxidation products, however with low selectivity. The selectivity of these variants can be improved by additional rounds of molecular modelling (identification of additional selectivity determining positions), diversification (allowing again only substitution to hydrophobic residues) and screening. By applying this approach we were able to generate a triple mutant of CYP102A1 which highly selectively converts (4*R*)-limonene to the potential anti-cancer drug perillyl alcohol, a product not formed by the wild type enzyme, by screening in total only 29 variants[46].

CONCLUSION

The structural basis of substrate specificity and regioselectivity was investigated by combining two approaches: a systematic comparison of sequences and structures to derive simple rules for predicting the regioselectivity of CYPs, and molecular modelling of substrate binding to investigate the effect of shape and flexibility. As a result, a small number of selectivity-determining positions were predicted based on the sequence of CYPs. The prediction was applied to interpret the effect of mutations in human CYPs on regioselectivity and to design a minimal mutant library of 24 CYP102A1 variants. All mutations resulted in stable and catalytically active enzymes. For all substrates tested yet, the library contained variants with either a strong shift or an improved regio- or stereoselectivity as compared to CYP102A1 wild type enzyme. Since selectivity-determining residues can be identified from protein sequence, the concept of constructing a minimal mutant library is generic and constitutes an efficient route to improved, highly selective biooxidation catalysts.

ACKNOWLEDGEMENT

Financial support by the Deutsche Forschungsgemeinschaft ("Sonderforschungsbereich 706") and by the Federal Ministry of Education and Research ("Systems biology in Pseudomonas for industrial biocatalysis") is gratefully acknowledged.

REFERENCES

[1] Panke, S.; Witholt, B.; Schmid, A.; Wubbolts, M. G. Towards a biocatalyst for (S)-styrene oxide production: Characterization of the styrene degradation pathway of Pseudomonas sp. strain VLB120. *Appl. Environ. Microbiol.*, 1998, *64*(6), 2032-2043.

[2] van Hellemond, E. W.; Janssen, D. B.; Fraaije, M. W. Discovery of a novel styrene monooxygenase originating from the metagenome. *Appl. Environ. Microbiol.*, 2007, *73*(18), 5832-5839.

[3] Ingelman-Sundberg, M. Human drug metabolising cytochrome P450 enzymes: properties and polymorphisms. *Naunyn. Schmiedebergs Arch. Pharmacol.*, 2004, *369*(1), 89-104.

[4] Ramensky, V.; Bork, P.; Sunyaev, S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.*, 2002, *30*(17), 3894-900.

[5] Ng, P. C.; Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.*, 2003, *31*(13), 3812-4.

[6] Guengerich, F. P. Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem. Res. Toxicol.*, 2001, *14*(6), 611-650.

[7] Isin, E. M.; Guengerich, F. P. Complex reactions catalyzed by cytochrome P450 enzymes. *Biochim. Biophys. Acta*, 2007, *1770*(3), 314-29.

[8] Meunier, B.; de Visser, S. P.; Shaik, S. Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. *Chem. Rev.*, 2004, *104*(9), 3947-3980.

[9] Urlacher, V. B.; Eiben, S. Cytochrome P450 monooxygenases: perspectives for synthetic application. *Trends Biotechnol.*, 2006, *24*(7), 324-30.

[10] Bernhardt, R. Cytochromes P450 as versatile biocatalysts. *J. Biotechnol.*, 2006, *124*(1), 128-45.

[11] O'Reilly, E.; Kohler, V.; Flitsch, S. L.; Turner, N. J. Cytochromes P450 as useful biocatalysts: addressing the limitations. *Chem. Commun.*, 2011, *47*(9), 2490-501.

[12] McLean, K. J.; Sabri, M.; Marshall, K. R.; Lawson, R. J.; Lewis, D. G.; Clift, D.; Balding, P. R.; Dunford, A. J.; Warman, A. J.; McVey, J. P.; Quinn, A. M.; Sutcliffe, M. J.; Scrutton, N. S.; Munro, A. W. Biodiversity of cytochrome P450 redox systems. *Biochem. Soc. Trans.*, 2005, *33*(Pt 4), 796-801.

[13] Narhi, L. O.; Fulco, A. J. Identification and characterization of two functional domains in cytochrome P-450BM-3, a catalytically self-sufficient monooxygenase induced by barbiturates in Bacillus megaterium. *J. Biol. Chem.*, 1987, *262*(14), 6683-90.

[14] Kuehnel, K.; Maurer, S. C.; Galeyeva, Y.; Frey, W.; Laschat, S.; Urlacher, V. B. Hydroxylation of dodecanoic acid and (2R,4R,6R,8R)-tetramethyldecanol on a preparative scale using an NADH-dependent CYP102A1 mutant. *Adv. Synth. Catal.*, 2007, *349*(8-9), 1451-1461.

[15] Kang, J. Y.; Kim, K. H.; Kim, D. H.; Jung, H. C.; Pan, J. G.; Ahn, T.; Yun, C. H. Generation of the human metabolites of simvastatin and lovastain by bacterial cytochrome P450 BM3 enzymes. *Drug Metab. Rev.*, 2009, *41*, 47-47.

[16] van Vugt-Lussenburg, B. M. A.; Stjernschantz, E.; Lastdrager, J.; Oostenbrink, C.; Vermeulen, N. P. E.; Commandeur, J. N. M. Identification of critical residues in novel drug metabolizing mutants of cytochrome P450BM3 using random mutagenesis. *J. Med. Chem.*, 2007, *50*(3), 455-461.

[17] Li, H. Y.; 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*(2), 140-146.

[18] Williams, P. A.; Cosme, J.; Ward, A.; Angove, H. C.; Matak Vinkovic, D.; Jhoti, H. Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature.*, 2003, *424*(6947), 464-8.

[19] Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.0 angstrom resolution. *Drug Metab. Rev.*, 2004, *36*, 345-345.

[20] 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*(1), 83-90.

[21] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.*, 2000, *28*(1), 235-42.

[22] Fischer, M.; Knoll, M.; Sirim, D.; Wagner, F.; Funke, S.; Pleiss, J. The Cytochrome P450 Engineering Database: a navigation and prediction tool for the cytochrome P450 protein family. *Bioinformatics*, 2007, *23*(15), 2015-7.

[23] Sirim, D.; Wagner, F.; Lisitsa, A.; Pleiss, J. The cytochrome P450 engineering database: Integration of biochemical properties. *BMC Biochem.*, 2009, *10*, 27.

[24] Nelson, D. R. Cytochrome P450 nomenclature, 2004. *Methods Mol. Biol.*, 2006, *320*, 1-10.

[25] Seifert, A.; Tatzel, S.; Schmid, R. D.; Pleiss, J. Multiple molecular dynamics simulations of human p450 monooxygenase CYP2C9: the molecular basis of substrate binding and regioselectivity toward warfarin. *Proteins Struct. Funct. Bioinf.*, 2006, *64*(1), 147-55.

[26] Pochapsky, T. C.; Kazanis, S.; Dang, M. Conformational plasticity and structure/function relationships in cytochromes P450. *Antioxid. Redox Signal.*, *13*(8), 1273-96.

[27] Branco, R. J.; Seifert, A.; Budde, M.; Urlacher, V. B.; Ramos, M. J.; Pleiss, J. Anchoring effects in a wide binding pocket: the molecular basis of regioselectivity in engineered cytochrome P450 monooxygenase from B. megaterium. *Proteins Struct. Funct. Bioinf.*, 2008, *73*(3), 597-607.

[28] Mestres, J. Structure conservation in cytochromes P450. *Proteins Struct. Funct. Bioinf.*, 2005, *58*(3), 596-609.

[29] Vidakovic, M.; Sligar, S. G.; Li, H.; Poulos, T. L. Understanding the role of the essential Asp251 in cytochrome p450cam using site-directed mutagenesis, crystallography, and kinetic solvent isotope effect. *Biochemistry*, 1998, *37*(26), 9211-9.

[30] Clark, J. P.; Miles, C. S.; Mowat, C. G.; Walkinshaw, M. D.; Reid, G. A.; Daff, S. N.; Chapman, S. K. The role of Thr268 and Phe393 in cytochrome P450 BM3. *J. Inorg. Biochem.*, 2006, *100*(5-6), 1075-90.

[31] Seifert, A.; Pleiss, J. Identification of selectivity-determining residues in cytochrome P450 monooxygenases: a systematic analysis of the substrate recognition site 5. *Proteins Struct. Funct. Bioinf.*, 2009, *74*(4), 1028-35.

[32] Sirim, D.; Widmann, M.; Wagner, F.; Pleiss, J. Prediction and analysis of the modular structure of cytochrome P450 monooxygenases. *BMC Struct. Biol.*, 2010, *10*.

[33] Kronbach, T.; Kemper, B.; Johnson, E. F. A hypervariable region of P450iic5 confers progesterone 21-hydroxylase activity to P450iic1. *Biochemistry.*, 1991, *30*(25), 6097-6102.

[34] Urlacher, V. B.; Makhsumkhanov, A.; Schmid, R. D. Biotransformation of beta-ionone by engineered cytochrome P450BM-3. *Appl. Microbiol. Biot.*, 2006, *70*(1), 53-59.

[35] Xu, F.; Bell, S. G.; Rao, Z.; Wong, L. L. Structure-activity correlations in pentachlorobenzene oxidation by engineered cytochrome P450(cam). *Protein Eng. Des. Sel.*, 2007, *20*(10), 473-480.

[36] Hasler, J. A.; Harlow, G. R.; Szklarz, G. D.; John, G. H.; Kedzie, K. M.; Burnett, V. L.; He, Y. A.; Kaminsky, L. S.; Halpert, J. R. Site-Directed Mutagenesis of Putative Substrate Recognition Sites in Cytochrome-P450-2b11 - Importance of amino-acid residue-114, residue-290, and residue-363 for substrate-specificity. *Mol. Pharmacol.*, 1994, *46*(2), 338-345.

[37] Szklarz, G. D.; He, Y. Q.; Kedzie, K. M.; Halpert, J. R.; Burnett, V. L. Elucidation of amino acid residues critical for unique activities of rabbit cytochrome P450 2B5 using hybrid enzymes and reciprocal site-directed mutagenesis with rabbit cytochrome P450 2B4. *Arch. Biochem. Biophys.*, 1996, *327*(2), 308-318.

[38] Born, S. L.; John, G. H.; Harlow, G. R.; Halpert, J. R. Characterization of the progesterone 21-hydroxylase activity of canine cytochrome P450 PBD-2/P450 2B11 through reconstitution, heterologous expression, and site-directed mutagenesis. *Drug Metab. Dispos.*, 1995, *23*(7), 702-7.

[39] Kerdpin, O.; Elliot, D. J.; Boye, S. L.; Birkett, D. J.; Yoovathaworn, K.; Miners, J. O. Differential contribution of active site residues in substrate recognition sites 1 and 5 to cytochrome P450 2C8 substrate selectivity and regioselectivity. *Biochemistry.*, 2004, *43*(24), 7834-42.

[40] Schalk, M.; Croteau, R. A single amino acid substitution (F363I) converts the regiochemistry of the spearmint (-)-limonene hydroxylase from a C6- to a C3-hydroxylase. *PNAS.*, 2000, *97*(22), 11948-53.

[41] Lentz, O.; Feenstra, A.; Habicher, T.; Hauer, B.; Schmid, R. D.; Urlacher, V. B. Altering the regioselectivity of cytochrome P450 CYP102A3 of Bacillus subtilis by using a new versatile assay system. *ChemBioChem.*, 2006, *7*(2), 345-50.

[42] Chen, C. K. J.; Berry, R. E.; Shokhireva, T. K.; Murataliev, M. B.; Zhang, H. J.; Walker, F. A. Scanning chimeragenesis: the approach used to change the substrate selectivity of fatty acid monooxygenase CYP102A1 to that of terpene omega-hydroxylase CYP4C7. *J. Biol. Inorg. Chem.*, 2010, *15*(2), 159-174.

[43] Kuchner, O.; Arnold, F. H. Directed evolution of enzyme catalysts. *Trends Biotechnol.*, 1997, *15*(12), 523-530.

[44] Seifert, A.; Vomund, S.; Grohmann, K.; Kriening, S.; Urlacher, V. B.; Laschat, S.; Pleiss, J. Rational design of a minimal and highly enriched CYP102A1 mutant library with improved regio-, stereo- and chemoselectivity. *ChemBioChem.*, 2009, *10*(5), 853-61.

[45] Weber, E.; Seifert, A.; Antonovici, M.; Geinitz, C.; Pleiss, J.; Urlacher, V. B. Screening of a minimal enriched P450 BM3 mutant library for hydroxylation of cyclic and acyclic alkanes. *Chem. Commun.*, 2011, *47*(3), 944-946.

[46] Seifert, A.; Antonovici, M.; Hauer, B.; Pleiss, J. An efficient route to selective bio-oxidation catalysts: an iterative approach comprising modeling, diversification, and screening, based on CYP102A1. *ChemBioChem.*, 2011, *12*(9), 1346-51.

[47] DeLano, W. L., The PyMOL Molecular Graphics System. 2002.

Received: February 28, 2011 **Revised: June 28, 2011** **Accepted: July 6, 2011**