Cytochrom P450-Monooxygenasen: Modellierung, Datenbankanalyse und experimentelle Charakterisierung neuer Enzymvarianten

Von der Fakultät Energie-, Verfahrens- und Biotechnik
der Universität Stuttgart zur Erlangung der Würde eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigte Abhandlung

Vorgelegt von

Alexander Seifert

geboren in Greiz

Hauptberichter: Prof. Dr. Jürgen Pleiss
Mitberichter: Prof. Dr. Rolf D. Schmid
Tag der mündlichen Prüfung: 04.03.2009

Institut für Technische Biochemie der Universität Stuttgart

2008

Teile der vorliegenden Arbeit wurden bereits veröffentlicht:

Seifert A., Vomund S., Grohmann K., Kriening S., Urlacher V. B., Laschat S. and Pleiss J. (2009). "Rational Design of a Minimal and Highly Enriched CYP102A1 Mutant Library with Improved Regio-, Stereo- and Chemoselectivity." *Chembiochem* 10(5): 853-61.

Seifert A. and Pleiss J. (2009). "Identification of selectivity-determining residues in cytochrome P450 monooxygenases: a systematic analysis of the substrate recognition site 5." *Proteins* 74(4): 1028-35.

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

Zusätzliche Publikationen des Autors, die nicht in dieser Arbeit enthalten sind:

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

Burns B. P., Seifert A., Goh F., Pomati F., Jungblut A. D., Serhat A. and Neilan B. A. (2005). "Genetic potential for secondary metabolite production in stromatolite communities." *FEMS Microbiol Lett* 243(1): 293-301.

Inhaltsverzeichnis

Zusammenfassung **4**

Abstract **7**

1 Einleitung **9**

1.1 Cytochrom P450-Monooxygenasen 9
1.2 Nomenklatur 10
1.3 Katalysemechanismus 11
1.4 Katalysierte Reaktionen 13
1.5 Redoxpartner von P450-Monooxygenasen 14
1.6 Struktur von P450-Monooxygenasen 15

2 Ergebnisse und Diskussion **19**

2.1 Untersuchung der molekularen Grundlagen für Substratbindung und Regioselektivität der humanen Cytochrom P450-Monooxygenase 2C9 durch multiple molekulardynamische Simulationen 19
2.2 Die systematische Analyse der Substraterkennungsstelle 5 zur Identifizierung selektivitätsbestimmender Aminosäuren in Cytochrom P450-Monooxygenasen . . . 23
2.3 Rationales Design einer minimalen und hoch angereicherten Mutantenbibliothek zur Steigerung der Regio-, Stereo- und Chemoselektivität von CYP102A1 27

3 Publikationen **34**

3.1 Multiple Molecular Dynamics Simulations of Human P450 Monooxygenase CYP2C9: The Molecular Basis of Substrate Binding and Regioselectivity toward Warfarin 34
3.1.1 Abstract 34
3.1.2 Introduction 35
3.1.3 Materials and Methods 37
3.1.4 Results 39

3.1.5 Discussion . . . 42
3.1.6 Conclusion . . . 46
3.1.7 References . . . 46
3.1.8 Supplementary material . . . 56

3.2 Identification of selectivity-determining residues in cytochrome P450 monooxygenases: a systematic analysis of the substrate recognition site 5 . . . 63
3.2.1 Abstract . . . 63
3.2.2 Introduction . . . 64
3.2.3 Materials and Methods . . . 66
3.2.4 Results . . . 66
3.2.5 Discussion . . . 70
3.2.6 Conclusion . . . 73
3.2.7 References . . . 74
3.2.8 Supplementary material . . . 81

3.3 Rational design of a minimal and highly enriched CYP102A1 mutant library with improved regio-, stereo-, and chemoselectivity . . . 83
3.3.1 Abstract . . . 84
3.3.2 Introduction . . . 84
3.3.3 Materials and Methods . . . 88
3.3.4 Results . . . 94
3.3.5 Discussion . . . 99
3.3.6 References . . . 102

4 Gesamtliteraturverzeichnis 110

5 Danksagungen 118

6 Erklärung 119

Zusammenfassung

Die vorliegende Arbeit befasst sich mit Cytochrom P450-Monooxygenasen. Vertreter dieser Enzymsuperfamilie sind aufgrund ihrer Beteiligung am Medikamentenstoffwechsel des Menschen, sowie ihrer Fähigkeit eine Vielzahl chemischer Verbindungen stereo- und regioselektiv zu oxidieren Gegenstand intensiver Forschung. Die genaue Vorhersage des von P450-Monooxygenasen verursachten Medikamentenstoffwechsels ist von großer Bedeutung für die Entwicklung neuer Wirksubstanzen. Weiterhin sind einige Vertreter dieser Enzymsuperfamilie für biotechnologische Anwendungen interessant. Aufgrund ihrer hohen Aktivität gegenüber verschiedenen Substraten, einem breiten Substratspektrum und relativer Prozessstabilität erweisen sich dabei Varianten der bakteriellen P450-Monooxygenase CYP102A1 (P450 BM-3) als besonders vielversprechend. Für die effiziente Anwendung von P450-Monooxygenasen ist jedoch häufig die Verbesserung bestimmter biochemischer Eigenschaften, wie der Regio-, Stereo- und Chemoselektivität unerlässlich. Eine wichtige Voraussetzung für die gezielte Verbesserung dieser Eigenschaften sowie für genaue Vorhersagen des von P450-Monooxygenasen verursachten Medikamentenstoffwechsels ist das Verständnis der molekularen Grundlagen von Aktivität, Spezifität und Selektivität.

Der erste Teil der vorliegenden Arbeit widmet sich daher der Untersuchung von Substrat-Enzym-Wechselwirkungen von Cytochrom P450-Monooxygenasen zur Identifizierung von selektivitätsbestimmenden Regionen. Hierzu wurde beispielhaft die Dynamik des CYP2C9-Warfarin-Komplexes mittels multipler molekulardynamischer Simulationen untersucht. Die Simulationsexperimente zeigten stark bewegliche Strukturelemente, welche die Bildung von Kanälen von der Proteinoberfläche ins Innere bewirken. Diese Kanäle ermöglichen den Austausch von Substraten und Produkten zwischen der Proteinumgebung und dem aktiven Zentrum. Die Beweglichkeit dieser Strukturelemente erlaubt darüber hinaus die Adaption des Enzyms an Substrate verschiedener Größe und Form. Die Regioselektivität wiederum wird durch einen engen trichterförmigen Hämzugangskanal kontrolliert, welcher vom starren Proteinkern gebildet wird. Dieser gewährt nur den Teilen des Substratmoleküls, die durch den Trichter passen, einen Zugang zum aktivierten Häm-Sauerstoff. Aminosäuren, die den Hämzugangskanal bilden, sind folglich von großer Bedeutung für die Orientierung des Substrats in der Nähe des Häms. Diese Erkenntnis führte zur Vorhersage von Aminosäuren mit starkem Einfluss auf die Regioselektivität in CYP2C9.

Daraufhin wurden verschiedene P450-Monooxygenasen in dem für die Kontrolle der Regioselektivität wichtigen Bereich verglichen. Wegen ihrer direkten Nachbarschaft zu dem in der Sequenz und in der Struktur hoch konservierten ExxR-Motiv wurde die Region um die Substraterkennungsstelle 5 (SRS-5) für eine systematische Analyse von 31 Kristallstrukturen und über 6300 Sequenzen ausgewählt. Die Ergebnisse dieser Analyse zeigten, dass P450-Monooxygenasen in dieser Region in Sequenz und Struktur variabel sind. Es wurde eine positiv geladene mit der Hämgruppe interagierende Aminosäure am C-terminalen Ende der SRS-5 Region in 29 Strukturen sowie in 97,7% aller Sequenzen identifiziert. Die Konformation der Region ist abhängig von der Position dieser Aminosäure nach dem konservierten ExxR-Motiv. Dabei zeigte sich, dass P450-Monooxygenasen hinsichtlich der Position dieser positiv geladenen Aminosäuren klassifiziert werden können. Die beobachteten Konformationen unterscheiden sich auch hinsichtlich der Position von zum Hämzentrum ausgerichteten Aminosäuren. Diese Aminosäuren beeinflussen die Orientierung von Substraten in der Nähe der Hämgruppe und haben damit einen starken Einfluss auf die Spezifität und Selektivität. Mit Hilfe der hier aufgeklärten Sequenz-Struktur-Funktionsbeziehungen können solche Aminosäuren ausschließlich anhand der Proteinsequenz ohne Kenntnis der Proteinstruktur identifiziert werden. Dies erlaubte das Aufstellen von Regeln zur einfachen Identifizierung zweier Positionen mit großem Einfluss auf Spezifität und Selektivität nur anhand von Sequenzinformationen. Die Prädiktivität dieser Regeln wurde mit Hilfe experimenteller Literaturdaten bestätigt.

Im letzten Teil der Arbeit wurden die bis dahin gewonnenen Erkenntnisse genutzt, um neue Varianten der für biotechnologische Prozesse vielversprechenden bakteriellen P450-Monooxygenase CYP102A1 zu generieren. Ziel war es hierbei die Regio, Stereo- und Chemoselektivität des Enzyms gegenüber verschiedenen Substraten zu erhöhen, sowie das Produktspektrum zu verändern. Zu diesem Zweck wurde eine Bibliothek von CYP102A1-Varianten erstellt, welche Mutationen in nur zwei Positionen aufweisen. Eine dieser Positionen wurde im Rahmen dieser Arbeit als *hotspot* für Selektivität und Spezifität identifiziert und kann in der Sequenz fast aller P450-Monooxygenasen lokalisiert werden. Die zweite Position wurde in vorangegangen Arbeiten als *hotspot* für Selektivität und Spezifität in CYP102A1 identifiziert. Durch Kombination von 5 hydrophoben Aminosäuren in beiden *hotspot* Positionen wurde eine Mutantenbibliothek minimaler Größe bestehend aus 24 Einfach- und Doppelmutanten generiert. Ziel war es dabei, kooperative Effekte von Mutationen in beiden *hotspot* Positionen zu induzieren, welche aufgrund der räumlichen Nähe beider Positionen erwartet wurden. Diese Mutantenbibliothek wurde mit vier Terpenen ((4*R*)-

Limonen, (+)-Valencen, Nerylaceton und Geranylaceton) unterschiedlicher Größe und Form durchmustert. Nach der Entwicklung einer umfassenden Analytik zur Identifizierung der gebildeten Oxidationsprodukte konnte gezeigt werden, dass 11 Enzymvarianten ein verändertes Produktspektrum und/oder verbesserte Regio- und Stereoselektivität aufweisen. Zu den gebildeten Oxidationsprodukten gehört der wertvolle Aromastoff (+)-Nootkaton. Weiterhin werden hier erstmalig CYP102A1-Varianten vorgestellt, welche bevorzugt die 11- und 12-Hydroxyprodukte von Geranyl- und Nerylaceton produzieren. Dabei handelt es sich um wertvolle Ausgangsstoffe für die Synthese von Naturstoffen. Es zeigte sich, dass kooperative Effekte von Mutationen in beiden *hotspots* entscheidend für die Erhöhung der Selektivität von CYP102A1 gegenüber den sperrigeren Substraten (+)-Valencen und (4*R*)-Limonen sind.

Summary

The present work deals with cytochrome P450 monooxygenases. Members of this enzyme superfamily are subject of intense research due to their participation in the human drug metabolism, as well as their ability to oxidise a wide variety of chemical compounds regio- and enantioselectively. An accurate prediction of the CYP related drug metabolism in *Homo sapiens* can aid in the design and development of new drugs. Furthermore, some members of this enzyme superfamily are promising candidates for applications in biotechnology. Variants of CYP102A1 from *Bacillus megaterium* (P450 BM-3) are particularly promising due to a high activity towards various substrates, a broad substrate acceptance and relative stability under process conditions. However, for the application of CYP enzymes in biotechnological processes regio- and stereoselectivity often has to be improved. A prerequisite for rational protein engineering, as well as accurate predictions of the CYP related human drug metabolism is the understanding of the molecular basis of activity, specificity and selectivity.

In the first part of this work enzyme-substrate interactions were investigated to identify selectivity determining regions in cytochrome P450 monooxygenases. Therefore, the protein dynamics of the CYP2C9-warfarin complex were examined by molecular dynamics simulations. The results show that mobile structural elements cause the formation of channels between the substrate binding cavity and the protein surface, which facilitate the transfer of substrates and products between the active site and the bulk solvent, as well as the adaptation of the enzyme to substrate molecules with different size and shape. However, regioselectivity is determined by a narrow funnel like region, which is formed by the rigid protein core. This funnel greatly reduces the number of possible substrate orientations above the activated heme oxygen. As a consequence, amino acids which are involved in the formation of this funnel like region were predicted to have a pronounced effect on regioselectivity in CYP2C9.

In the second part of this work, the regions which correspond to the selectivity determining region in CYP2C9 were identified in different cytochrome P450 monooxygenases and compared. Due to its direct proximity to the structurally and sequentially conserved ExxR motif it was possible to systematically analyse the substrate recognition site 5 (SRS-5) in 31 crystal structures and more than 6300 protein sequences. The results of this analysis showed that the SRS-5 region is variable in sequence and structure. A positively charged heme-interacting residue was identified at the C-terminal end of the SRS-5 region in 29 structures and in 97.7% of all sequences. The conformation of the SRS-5 region depends on the position

of this amino acid after the conserved ExxR motif. The results presented here show that P450 monooxygenases can be classified according to the position of this positively charged heme-interacting residue. The observed conformations also differ in terms of the position of amino acids, whose side chains point towards the heme centre. These amino acids influence the orientation of substrates close to the activated heme oxygen and thus have a strong influence on specificity and selectivity. With the help of the sequence-structure-function relationships unveiled in this work, these amino acids can be identified from sequence even in the absence of protein structure information. These findings allowed us to derive rules on how to readily identify two positions with strong influence on selectivity and specificity from protein sequence alone. The predictivity of these rules was validated with experimental data from the literature.

The knowledge derived in this work was used to create new variants of CYP102A1, since this enzyme is one of the most promising monooxygenases for applications in biotechnology. The aim was to improve regio-, stereo-, and chemoselectivity of this enzyme towards different substrates. Therefore, a small library of CYP102A1 variants with mutations in only two positions was created. One of these positions was identified as a hotspot for selectivity and specificity in this work and can be localised in the sequence of almost all P450 monooxygenases. The second position was previously identified as a hotspot for selectivity and specificity in CYP102A1. The combination of 5 hydrophobic amino acids in both hotspot positions resulted in a mutant library of minimal size, which consists of 24 single and double mutants. The aim was to induce cooperative effects of mutations in both hotspot positions, which were expected due to the spatial proximity of both positions. This mutant library was screened with four differently sized and shaped terpenes ((4*R*)-limonene, (+)-valencene, nerylacetone, and geranylacetone). After developing a comprehensive analysis to identify the formed oxidation products we could show that 11 enzyme variants revealed a changed product profile and strongly improved regio- and stereoselectivity. The valuable flavour compound (+)-nootkatone is among the formed oxidation products. Furthermore, for the first time it was shown that CYP102A1 variants preferentially oxidise geranyl- and nerylacetone at allylic positions to form 11- and 12-hydroxyproducts. These compounds are valuable starting materials for the total synthesis of natural products. The results presented here indicate that cooperative effects induced by combined mutations in both hotspot positions are required to improve selectivity towards the more bulky substrates (4*R*)-limonene and (+)-valencene.

1 Einleitung

Im Folgenden werden die Grundlagen der in dieser Arbeit abgehandelten Themen beschrieben. Dazu gehört eine Einführung zu Funktion, Vorkommen und biochemischen Eigenschaften der Enzymklasse der Cytochrom P450-Monooxygenasen (P450, CYP). Weiterhin wird auf spezielle Aspekte der Proteinstruktur und Sequenz eingegangen. Ein besonderes Augenmerk der hier vorgestellten Ergebnisse gilt dem im Medikamenten-stoffwechsel des Menschen involvierten CYP2C9 sowie dem für biotechnologische Prozesse vielversprechenden bakteriellen CYP102A1.

1.1 Cytochrom P450-Monooxygenasen

Die zu den Oxidoreduktasen gehörenden cytochrom P450-Monooxygenasen (E.C. 1.14. x.y) umfassen eine ubiquitär in allen Formen des Lebens vorkommende Klasse von Enzymen. Zum Zeitpunkt der Entstehung dieser Arbeit sind mehr als 8000 CYP codierende Gene bekannt. Diese verteilen sich auf 2872 benannte Tier-CYP, 2867 Pflanzen-CYP, 912 bakterielle CYP, 239 Protisten-CYP, sowie 1238 Pilz-CYP (http://drnelson.utmem.edu/p450stats.Feb2008.htm). CYP sind Hämproteine, die Eisenporphyrin als prosthetische Gruppe tragen. Die Bindung des Eisenporphyrins an das Thiolat eines Cysteins des sogenannten Apoproteins erfolgt über das Eisenatom. Namensgebend für P450-Systeme ist das charakteristische Absorptionsspektrum des reduzierten, kohlenstoffmonoxidgebundenen Enzyms mit einem Maximum bei 450 nm (Omura und Sato 1964). Das P in P450 steht für Pigment. CYP haben eine wichtige Rolle im Metabolismus praktisch aller Organismen. Allgemein katalysieren sie die Oxidation nicht aktivierter aliphatischer und aromatischer Verbindungen unter Verwendung von Luftsauerstoff, wobei als Reduktionsmittel meist das Coenzym NADH oder NADPH verwendet wird. Folgende Redoxgleichung beschreibt die zugrundeliegende Gesamtreaktion (1):

$$RH + NADPH + H^+ + O_2 \rightarrow ROH + H_2O + NADP^+ \qquad (1)$$

Dabei können CYP katalysierte Reaktionen sowohl hoch regio- und stereoselektiv sein, wie bei der Biosynthese von Hormonen, Signalüberträgern sowie vieler Sekudärmetabolite in Pflanzen, als auch unselektiv ablaufen, wie zum Beispiel bei der Detoxifikation von Fremdstoffen (Xenobiotika), zu denen auch pharmazeutische Wirkstoffe gehören. Hier dient

das eher unselektive Einfügen von polaren Gruppen in die meist hydrophoben Substanzen der Erhöhung der Löslichkeit und damit der rascheren Ausscheidung. Aber auch die Bioaktivierung vieler Xenobiotika durch P450-Monooxygenasen wird beobachtet. Dadurch werden einige Medikamente wie z.B. Cyclophosphamid und Ifosfamid in ihre therapeutisch aktive Form überführt. Die Bioaktivierung kann weiterhin zur Bildung von Epoxid- oder Aldehydverbindungen führen, die aufgrund ihrer Wechselwirkung mit Proteinen oder der DNA toxisch und mutagen wirken. Für die Medikamentenentwicklung ist es daher von großem Interesse den oxidativen Metabolismus des Menschen vorhersagen zu können, um eventuelle Gesundheitsrisiken früh im Entwicklungsprozess zu erkennen. Eine Voraussetzung für solche Vorhersagen ist das Wissen um die molekularen Faktoren, welche Selektivität und Spezifität beeinflussen.

1.2 Nomenklatur

In den späten 80iger Jahren wurde eine systematische Nomenklatur und Einteilung der CYP vorgeschlagen, die bis zum heutigen Tag weithin verwendet wird (Nebert et al. 1989; Nebert und Nelson 1991; Nelson et al. 1993; Nelson et al. 1996). Dabei wird als Einteilungskriterium die Aminosäuresequenzidentität herangezogen. CYP mit einer Aminosäuresequenzidentität >40% werden in Familien zusammengefasst und CYP mit einer Sequenzidentität von >55% werden der gleichen Unterfamilie zugeordnet. Darauf aufbauend setzt sich ein CYP-Name aus dem Kürzel CYP gefolgt von einer Nummer für die Proteinfamilie, einem Buchstabe für die Unterfamilie und abschließend einer Zahl welche die jeweilige Isoform bezeichnet, zusammen. Weiterhin zeigt eine kursive Schreibweise an, dass der jeweilige Name für ein Gen steht, während eine nicht-kursive Darstellung ein Genprodukt (mRNA, cDNA, Protein) bezeichnet. Zur Benennung der CYP Familien waren anfangs die Nummern 1 bis 100 für Eukaryoten vorgesehen, während bakteriellen CYP Nummern >100 zugewiesen wurden. Im Detail wurden die ersten 100 Familien unter Tieren (CYP1-50), niederen Eukaryoten (CYP51-70) sowie Pflanzen (CYP71-100) aufgeteilt (Nelson et al. 1993). Die schnell ansteigende Anzahl an P450 Sequenzen machte jedoch eine Erweiterung des Nomenklatursystems notwendig (Bakterien CYP101-299, CYP1001-2999; Tiere CYP301-499, 3001-4999; niedere Eukaryoten CYP501-699, CYP5001-6999, und Pflanzen CYP701-999, CYP7001-9999) (Nelson 2006).

1.3 Katalysemechanismus

Die katalytische Wirkung von Cytochrom P450-Monooxygenasen besteht in der Aktivierung von Luftsauerstoff und der darauffolgenden Oxidation eines Substrats. Das zweite Sauerstoffatom wird zu Wasser reduziert. Der zugrundeliegende Katalysezyklus wurde bereits im Jahre 1968 vorgeschlagen (Katagiri et al. 1968). Dieser konnte in den darauffolgenden Jahren durch eine Vielzahl von Untersuchungen weitestgehend bestätigt und verfeinert werden und ist heute allgemein anerkannt (Abbildung 1) (Makris 2005).

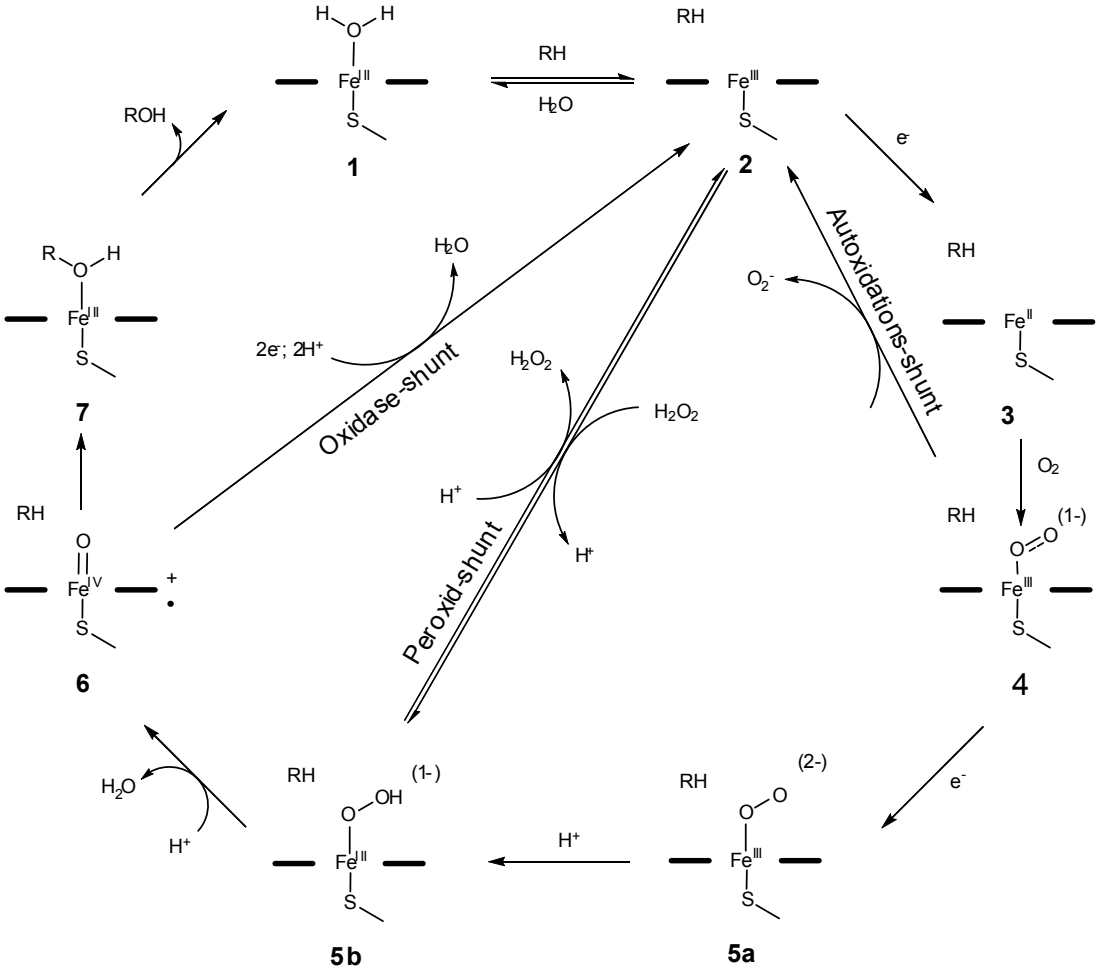


Abbildung 1: Katalysezyklus der P450-Monooxygenasen nebst möglichen Entkopplungsreaktionen („*shunt pathways*“). Das Porphyrinringsystem der Hämgruppe wird durch die beiden das Eisenatom flankierenden Balken dargestellt.

Im ersten Schritt verdrängt das eintretende Substrat das als sechster Ligand am Häm gebundene Wassermolekül. Dies überführt das Hämeisen (Fe^{3+}) vom *low-spin* in den *high-spin* Zustand (2). Durch diesen Schritt wird das Redoxpotential des Eisens positiver, was die Reduktion von Fe^{3+} zu Fe^{2+} durch die zugehörige CYP Reduktase ermöglicht (3). Im nächsten

Schritt bindet ein Sauerstoffmolekül koordinativ an das Hämeisen und bildet einen Oxy-Eisen Komplex (4). Durch die Übertragung eines weiteren Elektrons kommt es zur Bildung einer Peroxoeisenspezies (5a), die durch Protonierung in einen Hydroperoxoeisenkomplex überführt wird (5b). Die weitere Protonierung und subsequente Abspaltung von Wasser führt zur Bildung des Oxo-Eisenkomplexes (6). Letzterer wird oft als „*Compound I*" bezeichnet und stellt die katalytisch aktive Spezies dar. Der genaue Mechanismus der Substratoxidation durch den „*Compound I*" wird noch immer kontrovers diskutiert. Am wahrscheinlichsten gilt jedoch für die CYP katalysierte Hydroxylierung von C-H Gruppen der radikalische „*rebound*"-Mechanismus (Abbildung 2). Demzufolge wird in einem ersten Schritt ein Wasserstoffatom durch den „*Compound I*" (6) abstrahiert. Die homolytische Spaltung der C-H Bindung führt zur Bildung eines Radikals im Substrat. Im zweiten Schritt bindet das Fe^{4+}-OH-Intermediat über den Sauerstoff an das Substrat, woraufhin das Hydroxyprodukt abdissoziieren kann (Meunier et al. 2004; Shaik et al. 2005).

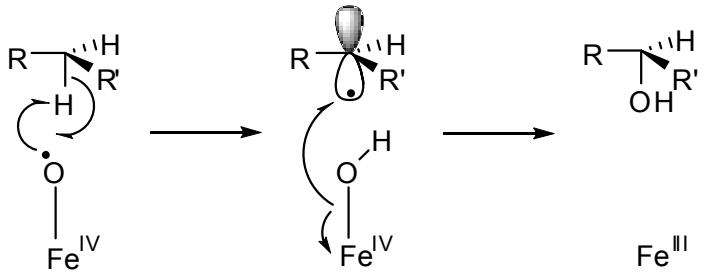


Abbildung 2: „*Rebound*"-Mechanismus

Wie quantenmechanische Berechnungen (de Visser et al. 2002; de Visser et al. 2004) sowie experimentelle Befunde (Guengerich et al. 2004) zeigten, ist die größte Energiebarriere bei diesem Mechanismus mit der Wasserstoffabstraktion assoziiert. Daher wurde in dieser Arbeit für den Vergleich der chemischen Reaktivität verschiedener Positionen in einem Substratmolekül die Bindungsdissoziationsenergie der zugehörigen C-H Bindungen herangezogen.

Neben der gezeigten Übertragung des aktivierten Sauerstoffs auf das Substrat gibt es drei weitere mögliche Reaktionsverläufe (Abbildung 1). So beschreibt die Abspaltung von Wasserstoffperoxid von Intermediat (5) den „Peroxid-*shunt*". Die Dissoziation von Intermediat (6) zu Intermediat (2) und Wasser nach Aufnahme zweier zusätzlicher Elektronen wird als „Oxidase-*shunt*" bezeichnet. Der „Autoxidations-*shunt*" beschreibt den direkten Übergang von Intermediat (4) zu Intermediat (2) durch die Abspaltung eines

Superoxidanions. Eine Gemeinsamkeit dieser drei alternativen Reaktionsverläufe besteht in der unerwünschten Entkopplung des Cofaktorverbrauchs von der Substratoxidation. Eine Konsequenz dieser Entkopplung ist die Bildung von reaktiven Nebenprodukten wie Wasserstoffperoxid, dessen Akkumulation zur Inaktivierung des Enzyms führt (Karuzina und Archakov 1994; Karuzina und Archakov 1994).

1.4 Katalysierte Reaktionen

Cytochrom P450-Monooxygenasen katalysieren ein großes Spektrum verschiedener Reaktionen (Juchau 1990), wie z. B. Hydroxylierung von aliphatischen und aromatischen Kohlenwasserstoffen, oxidative N- und O-Dealkylierung, oxidative Desaminierung und Desulfurylierung sowie Epoxidierungen von -C=C-Doppelbindungen (Abbildung 3). Darüber hinaus existieren eine Menge weiterer, komplexer P450-Reaktionen (Guengerich 2001; Isin und Guengerich 2007).

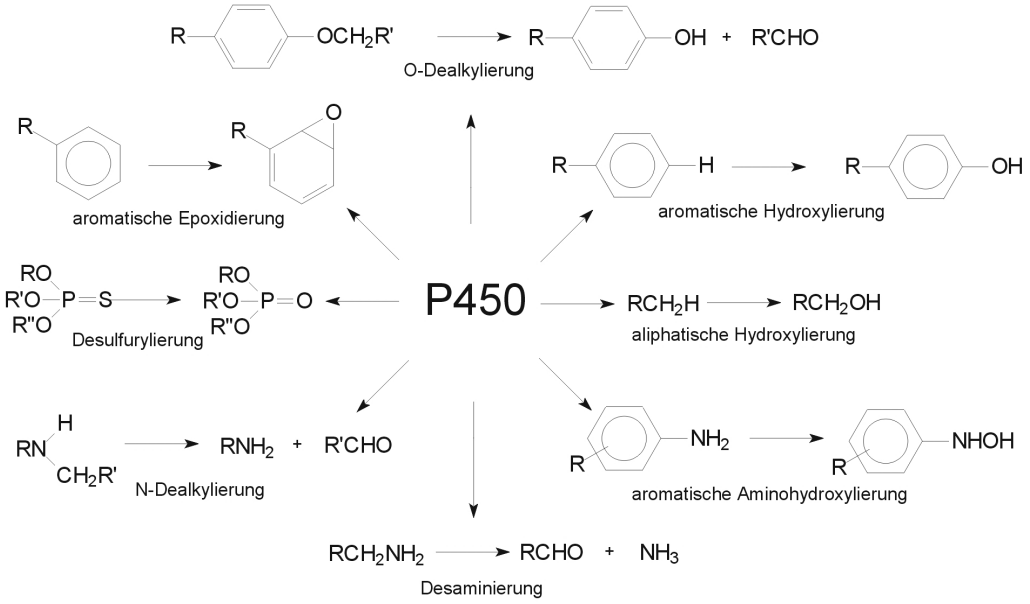


Abbildung 3: Wichtige von P450-Monooxygenasen katalysierte Reaktionen. R, R' und R'' können aromatische oder aliphatische Kohlenwasserstoffe sowie Heteroaromaten oder Halogene sein.

Die am häufigsten beobachtete von P450-Monooxygenasen katalysierte Reaktion ist die Hydroxylierung nicht-aktivierter C-H Bindungen aliphatischer Verbindungen. So hydroxyliert zum Beispiel CYP71D18 aus *Mentha spicata* (-)-(4*S*)-Limonen hoch regioselektiv zu (-)-trans-Carveol (Lupien et al. 1999). Die Hydroxylierung von Aromaten durch P450-

Monooxygenasen wird häufig im humanen Medikamentenstoffwechsel beobachtet, wie z.B. beim Abbau des Antikoagulans *(S)*-Warfarin (Rettie et al. 1992). Aber auch biosynthetische Stoffwechselwege beinhalten P450 katalysierte aromatische Hydroxylierungen, so z.B. die Biosynthese des Chitin Synthase-Inhibitors Nikkomycin in *Streptomyces tendae* (Bruntner et al. 1999). Die so gebildeten Alkohole können von einigen P450-Monooxygenasen in einem zweiten Schritt zu Ketonen (bei sekundären Alkoholen), wie im Falle der Oxidation von (+)-Valencen zu (+)-Nootkaton durch $P450_{cam}$-Mutanten (Sowden et al. 2005), oder zu Aldehyden und sogar Carbonsäuren (bei primären Alkoholen) oxidiert werden. Neben der Detoxifizierung von Xenobiotika (Fremdstoffe) sind P450-Monooxygenasen aber auch verantwortlich für die Bioaktivierung vieler Chemikalien. So können die aus Fremdstoffen gebildeten Metabolite karzinogene Wirkung haben, wie z.B. das durch Epoxidierung entstehende DNA-alkylierende Styroloxid (Ioannides und Lewis 2004).
Die oxidative Dealkylierung ist ein weiterer häufig von P450-Monooxygenasen katalysierter Reaktionstyp, im speziellen die O-, N- und S-Dealkylierung (Karki und Dinnocenzo 1995). Dabei werden Kohlenstoffatome in Nachbarschaft zu Sauerstoff oder Stickstoffatomen unter Bildung instabiler Halbacetale (bzw. den entsprechenden stickstoff- oder schwefelhaltigen Verbindungen) hydoxyliert, welche dann in ein Aldehyd und ein Alkohol (bzw. Amin oder Thiol) zerfallen.
Selbst oxidative C-C Kopplung gehört zu den von P450-Monooxygenasen katalysierten Reaktionen. So katalysiert CYP245A1 (StaP) bei der Biosynthese des Antitumorwirkstoffes Staurosporin durch Wasserstoffabstraktion von zwei Kohlenstoffatomen und subsequenter Radikalkopplung die Bildung einer Aryl-Aryl Bindung (Howard-Jones und Walsh 2007).

1.5 Redoxpartner von P450-Monooxygenasen

Wie der CYP-Katalysezyklus zeigt (Abbildung 1), beinhaltet die Aktivierung des Sauerstoffs die schrittweise Übertragung zweier Elektronen. Als Elektronenquelle dienen dabei die Cofaktoren NADH oder NADPH. Für die zeitlich präzise Übertragung der Elektronen von der Elektronenquelle auf das Hämeisen ist ein System von Oxidoreduktasen verantwortlich. Es wurden eine Reihe verschiedener Elektronentransfersysteme mit unterschiedlichen Redoxpartnern beschrieben (McLean et al. 2005). Der erste Schritt, nämlich die Übertragung eines Elektronenpaares (Hydridions) von NAD(P)H auf eine NAD(P)H-Oxidoreduktase ist dabei allen diesen Reduktasesystemen gemeinsam. Von der NAD(P)H-Oxidoreduktase

wandern die Elektronen zu einem Elektronentransferprotein, das seinerseits die Reduktion des Hämeisens durch Einelektronenübertragungen katalysiert.

Die häufigsten Klassen von Elektronentransportsystemen sind in Abbildung 4 gezeigt. Die Redoxpartner der Klasse I bestehen aus einer FAD-enthaltende NAD(P)H-abhängigen Reduktase und einem Eisen-Schwefel-Ferredoxin (v. a. in Pflanzen und Bakterien), diejenigen der Klasse II besitzen eine membrangebundene Cytochrom P450-Reduktase, die aus einer FAD und FMN-Domäne zusammengesetzt ist (z.B. humane P450 Systeme).

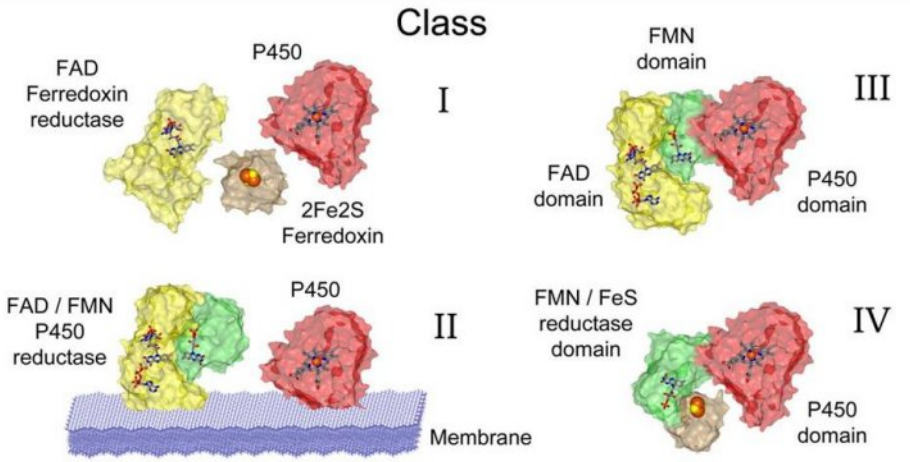


Abbildung 4: Schematische Darstellung der vier häufigsten Klassen von Elektronentransportsystemen nach S. K. Chapman (http://www.chem.ed.ac.uk/chapman/p450.html).

Die Klasse III beinhaltet Fusionsproteine aus einer FAD/FMN Reduktase und einer P450-Monooxygenase (z.B. CYP102A1). Fusionsproteine, bestehend aus einer FMN-Domäne, einer Eisen-Schwefel-NAD(P)H-Oxidase sowie einer P450-Domäne, sind in Klasse IV zusammengefasst. Darüber hinaus existieren noch weitere Wege des Elektronentransfers auf P450-Monooxygenasen (McLean et al. 1998; Seth-Smith et al. 2002; Munro et al. 2007).

1.6 Struktur von P450-Monooxygenasen

Seit der Veröffentlichung der ersten Kristallstruktur einer P450-Monooxygenase im Jahre 1987 wurden bis zum Zeitpunkt der Erstellung dieser Arbeit weitere 164 Strukturen von 31 verschiedenen P450 Enzymen durch Röntgenstrukturanalyse bestimmt und publiziert. Im

Unterschied zu den löslichen bakteriellen P450-Monooxygenasen liegen Säugetier-P450 und damit auch die des Menschen membranassoziiert vor und blieben daher lange Zeit der Methode der Röntgenstrukturanalyse unzugänglich. Eine Erhöhung der Löslichkeit durch Entfernung des Membranankers und gezielte Aminosäureaustausche ermöglichten erstmalig die Bestimmung der Kristallstruktur eines Säugetier-P450-Enzyms (Cosme und Johnson 2000). Durch eine ähnliche Vorgehensweise gelang es im Jahr 2003 die Kristallstruktur einer P450-Monooxygenase des Menschen (CYP2C9) zu bestimmen (Williams et al. 2003). Aktuell sind elf Säugetier-CYP bekannt, wovon neun humanen Ursprungs sind.

Wie der Vergleich vorhandener Strukturen zeigt, sind trotz geringer Aminosäure-sequenzidentität von teilweise unter 20% allgemeine Aspekte der dreidimensionalen Struktur konserviert (Denisov et al. 2005). Dies gilt besonders für den die prosthetische Hämgruppe beherbergenden Proteinkern, welcher durch die Helices D, E, I und L (*four-helix bundle*), die Helices J und K, zwei β-Faltblattstrukturen und den sogenannten *meander-loop* gebildet wird (Abbildung 5).

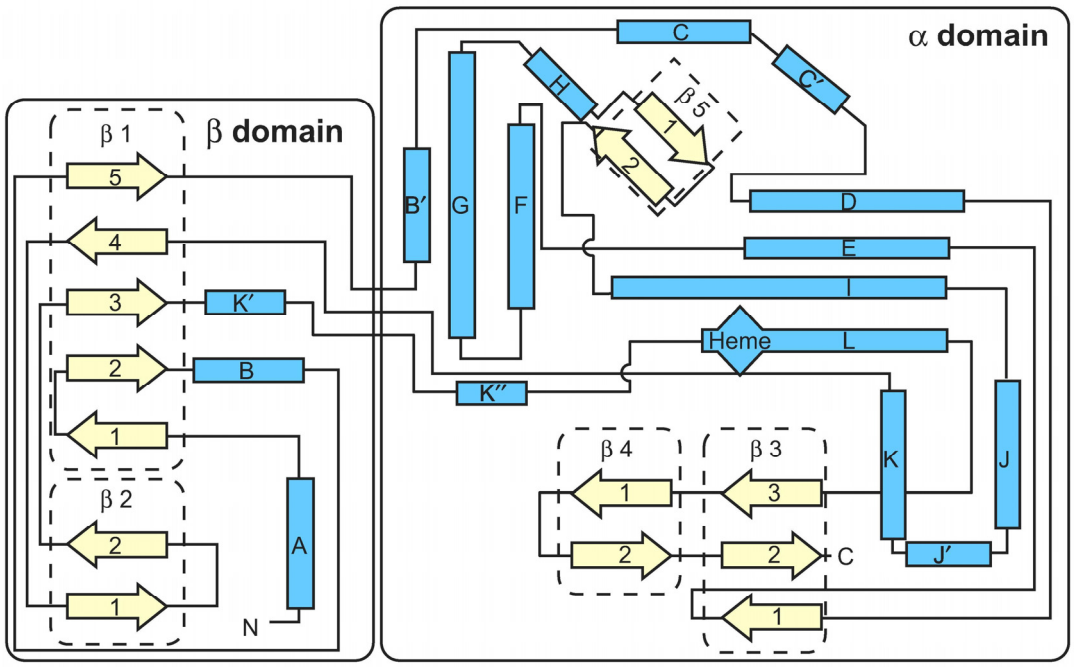


Abbildung 5: Schematische Darstellung konservierter Sekundärstrukturelemente von P450-Monooxygenasen aus Werck-Reichhart (Werck-Reichhart und Feyereisen 2000). Blaue Balken symbolisieren α-Helices, weiße Pfeile β-Faltblattstrukturen.

Grund dafür ist der dieser Enzymklasse gemeinsame Mechanismus der Elektronen-übertragung und Sauerstoffaktivierung.

Die dreidimensionale Anordnung der Sekundärstrukturelemente wird in Abbildung 6 für das humane CYP2C9 gezeigt. Die Hämgruppe ist über das Hämeisen an den strukturell

konservierten und in der Sequenz absolut konservierten Cysteinrest des Apoproteins gebunden. Die beiden negativ geladenen Propionatgruppen des Häms werden über Wasserstoffbrücken zu basischen Aminosäuren des Apoproteins stabilisiert. Bereiche der Struktur, in denen sich P450-Monooxygenasen unterscheiden, sind z. B. der N-Terminus, welcher bei den membranassoziierten Enzymen einen Membrananker aufweist. Die starke Variabilität der Substratspezifität unter P450-Monooxygenasen spiegelt sich auch in Unterschieden in den Bereichen der Tertiärstruktur wider, die für Substratbindung verantwortlichen sind. So kann die Form und Größe der die Hämgruppe umgebenden Substratbindekavität (Abbildung 7) zwischen verschieden Vertretern dieser Enzymklasse beträchtlich variieren.

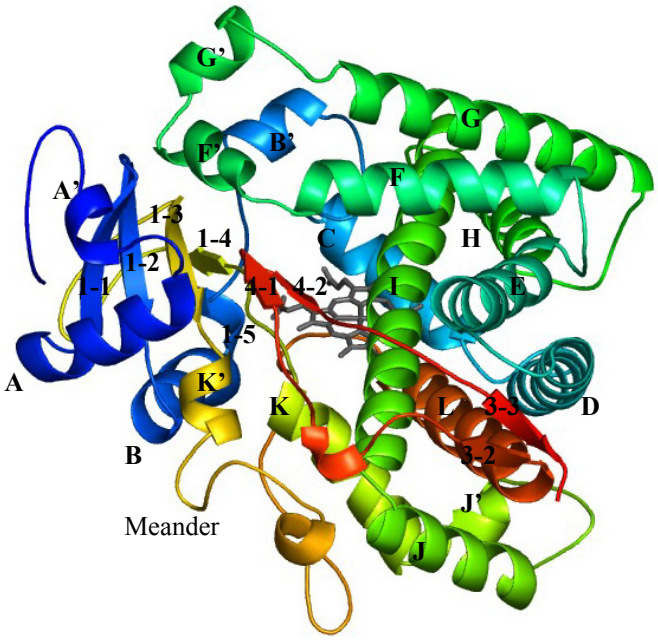


Abbildung 6: Kristallstruktur von CYP2C9 (PDB Eintrag 1og2) (Williams et al. 2003). Die prosthetische Hämgruppe ist grau dargestellt.

Zu den an der Formung dieser Kavität beteiligten Strukturelementen gehören die Helices I, F und G, der F-G *loop*, der B-C *loop*, sowie das C-terminale antiparallele β-Faltblatt. Durch Sequenzvergleiche auf Nukleotidbasis konnte gezeigt werden, das auf diese Strukturelemente sechs Sequenzbereiche (*substrate recognition site* (SRS) 1-6) entfallen, die Aminosäuren enthalten, welche für Substratbindung verantwortlich sind (Gotoh 1992). Es konnte für viele verschiedene P450-Monooxygenasen gezeigt werden, dass Aminosäureaustausche in diesen Bereichen zu Änderungen der Substratspezifität und Regioselektivität führen (Schalk und

Croteau 2000; Melet et al. 2003; Keizers et al. 2004; Sherman et al. 2006). Obwohl das Wissen um die Substraterkennungsstellen die Anzahl an potentiell Spezifität und Selektivität beeinflussenden Positionen reduziert, ist für die Suche nach Enzymvarianten mit verbesserten Eigenschaften aufgrund der Ausdehnung dieser 6 Bereiche auf Längen von jeweils bis zu 26 Aminosäuren dennoch eine immense Anzahl an kombinatorischen Möglichkeitenn zu durchmustern.

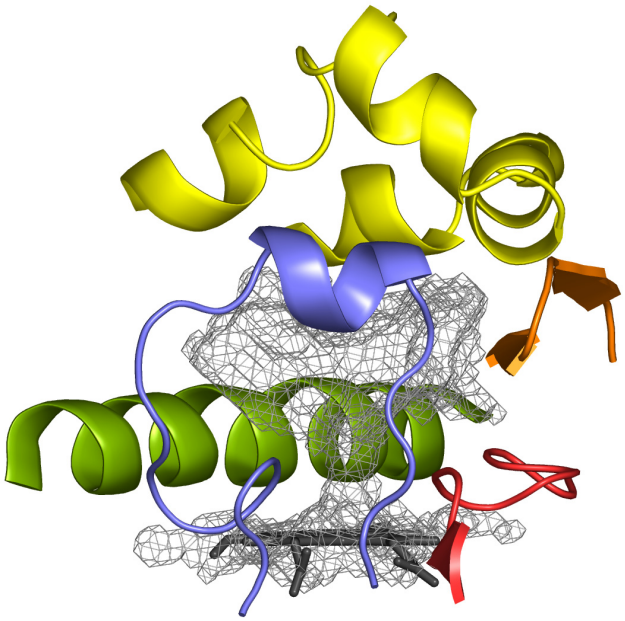


Abbildung 7: Substratbindkavität (hellgrau) (Krahn 2004) im Inneren der Kristallstruktur von CYP2C9 (PDB-Eintrag 1og2). Die Kavität wird von der Helix I (grün), den Helices F, G einschließlich F-G *loop* (gelb), dem B-C *loop* (blau), der SRS-5 Region (rot) sowie dem C-terminalen antiparallelen β-Faltblatt (orange) begrenzt. Die prosthetische Hämgruppe ist in grau wiedergegeben.

Ein weiterer für das Verständnis der Funktionsweise von P450-Monooxygenasen wichtiger Punkt ist die Identifikation von Kanälen, welche die Kavität im Inneren des Enzyms mit dessen Außenwelt verbinden. Entlang solcher Kanäle wandern sowohl Substrate vom Lösungsmittel zum aktiven Zentrum als auch die entstehenden Produkte aus der Kavität zurück ins Lösungsmittel. Die an der Formung dieser Kanäle beteiligten Aminosäuren haben folglich einen direkten Einfluss auf Enzymeigenschaften wie z.B. die Substratspezifität.

2 Ergebnisse und Diskussion

2.1 Untersuchung der molekularen Grundlagen für Substratbindung und Regioselektivität der humanen Cytochrom P450-Monooxygenase 2C9 durch multiple molekulardynamische Simulationen

(siehe Publikation: "Multiple Molecular Dynamics Simulations of Human P450 Monooxygenase CYP2C9: The Molecular Basis of Substrate Binding and Regioselectivity Toward Warfarin" Abschnitt 3.1)

Die genaue Vorhersage des von P450-Monooxygenasen verursachten Medikamentenstoffwechsels ist von großer Bedeutung für die Entwicklung neuer Wirksubstanzen (Smith et al. 1997; Smith et al. 1998). Die Kenntnis der molekularen Grundlagen biochemischer Eigenschaften wie Aktivität, Substratspezifität und Regioselektivität ist eine wichtige Voraussetzung für solche Vorhersagen. Mit der Veröffentlichung der ersten Kristallstruktur einer P450-Monooxygenase des Menschen (CYP2C9) (Williams et al. 2003) konnten erstmals Untersuchungen zu Struktur-Funktionsbeziehungen dieses für den Medikamentenstoffwechsel so wichtigen Enzymes (Rendic und Di Carlo 1997) durchgeführt werden. Es zeigte sich, dass dieser humane Vertreter über eine vergleichsweise große Substratbindekavität verfügt, die relativ große, aber auch mehrere Substratmoleküle gleichzeitig aufnehmen kann. Weiterhin wurde die Kristallstruktur des Enzyms zusätzlich in Anwesenheit des Substrats (*S*)-Warfarin aufgelöst. Im Enzym-Substrat-Komplex wurde das Substratmolekül allerdings in einem Abstand von 10 Å von der Hämgruppe gefunden, was einem nicht produktiven Komplex entspricht. Dieser Umstand machte es unmöglich jene Aminosäuren zu identifizieren, die das Substrat im reaktiven Zustand binden. Darüber hinaus zeigte die Kristallstruktur keine Kanäle, die Aufschluss darüber hätten geben können, wie potentielle Substrate ins Innere des Enzyms gelangen, oder die in der Folge gebildeten Produkte das Enzym verlassen. Es wurde deutlich, dass die Dynamik des CYP2C9-(*S*)-Warfarin-Komplexes untersucht werden musste, um Aufschluss über den produktiven Komplex zu gewinnen. Zu diesem Zweck wurde die Methode der molekulardynamischen Simulation (MD-Simulation) gewählt. Als Vorarbeit hierzu wurde im Rahmen meiner Diplomarbeit (Seifert 2005) das Substrat (*S*)-Warfarin parametrisiert und erste stabile Simulationen im Nanosekundenmaßstab durchgeführt. Es zeigte sich, dass aufgrund der hohen Flexibilität des Enzyms multiple MD-Simulationen

erforderlich waren, um den großen Konformationsraum dieses Enzym-Substrat-Komplexes auszuleuchten.
Dieses Projekt wurde im Rahmen meiner Doktorarbeit fortgesetzt, um letztendlich eine Gesamtzahl von 6 MD-Simulationen des freien Enzyms (je 3 ns) und 16 MD-Simulationen des Enzym-Substrat-Komplexes (je 5 ns) einer eingehenden Analyse unterziehen zu können. Die Analyse der erzeugten Trajektorien zeigte einen stabilen und starren Proteinkern. Im Gegensatz dazu wurden stark bewegliche Bereiche außerhalb des Proteinkerns beobachtet. Ingesamt wurden vier Zustände des Enzym-Substrat-Komplexes beobachtet, die sich hinsichtlich der Form der Substartbindekavität und der Position des Substratmoleküls unterscheiden.

Kanäle in CYP2C9:
Besonderen Einfluss auf die Form der Substratbindekavität hatten die Strukturelemente B-C *loop* und F-G *loop*. Die starke Beweglichkeit beider *loops* führte besonders in den MD-Simulationen des Enzym-Substrat-Komplexes zur Bildung zweier Kanäle von der Proteinoberfläche zur Substratbindekavität. Ein Kanal entsteht im Bereich zwischen B-C *loop*, F-G *loop* und dem *turn* im β_1-Faltblatt. Dieser Kanal wird durch das anwesende Substrat stabilisiert. Durch Hydrophobizitätsanalyse und Antikörpermarkierung konnte gezeigt werden, dass der B-C *loop* und der F-G *loop* von P450-Monooxygenasen der Familie 2 mit der Membran des endoplasmatischen Retikulum interagiert (Williams et al. 2000). Es wird angenommen, dass durch diesen Kontakt der direkte Übergang der hydrophoben Substrate von der Membran in das Enzym gewährleistet wird. In einer vorangegangenen MD Simulationsstudie des substratfreien CYP2C9 Enzyms wurden ebenfalls Fluktuationen im genannten Bereich beobachtet, jedoch konnte die Existenz eines Kanals mit geeigneter Größe für den Transfer von Substraten nicht gezeigt werden (Afzelius et al. 2004). Unsere MD Simulationen zeigen jedoch, dass in Anwesenheit eines Substrates solche Kanäle stabilisiert und geweitet werden können. Die Stabilisierung einer geöffneten Konformation durch ein zweites Molekül konnte ebenfalls für Säugetier-CYP2B4 beobachtet werden. In Abwesenheit eines Inhibitors zeigt die Kristallstruktur einen weit geöffneten Spalt, welcher in erster Linie durch die Helices B´ bis C and F bis G gebildet wird (Scott et al. 2003). In diesem Fall wird die weitgeöffnete Struktur durch Dimerisierung im Kristall generiert, wobei sich 2 Proteinmoleküle durch hydrophobe Wechselwirkungen gegenseitig stabilisieren. Dies zeigt, dass der Kanal und Teile der Substratbindekavität sowohl in Kristallstrukturen als auch in unseren MD Simulationen eine hohe Plastizität aufweisen und durch molekulare Interaktionen

stabilisiert werden können. Die Plastizität erlaubt die Adaption des Enzyms an Substrate verschiedener Größe und Form, was im Einklang mit dem breiten Substratspektrum von CYP2C9 und anderen P450-Monooxygenasen steht (Rendic und Di Carlo 1997). Auch in CYP101A1, CYP102A1, CYP107A1 und Säugetier-CYP2C5 konnte ein Kanal im Bereich zwischen B-C *loop*, F-G *loop* und dem *turn* im β_1-Faltblatt durch *random expulsion* MD Simulationen identifiziert werden (Winn et al. 2002; Wade 2004). Ein Lösungsmittelkanal im Bereich zwischen Helix F, Helix I und dem C-terminalen β-Faltblatt ist in der Kristallstruktur von CYP2C9 sowie in einer Anzahl weiterer P450-Monooxygenasen in substratgebundener als auch substratfreier Form zu erkennen (Wade 2004). Da dieser Kanal direkt vom aktiven Zentrum (Hämgruppe) in das das Enzym umgebende wässrige Lösungsmittel führt, wäre dies ein geeigneter Weg für das hydrophile Oxidationsprodukt die Substratbindekavität zu verlassen. In der Kristallstruktur von CYP2C9 ist dieser Kanal jedoch zu eng, um dem Oxidationsprodukt den Durchtritt zu gewähren. Das Binden des Substrates in der Nähe der Hämgruppe erweitert den Durchmesser dieses Kanals beträchtlich, sodass das sperrige Produkt den Kanal passieren kann. Beide hier beobachteten Kanäle wurden auch in Säugetier-CYP2C5 identifiziert (Schleinkofer et al. 2005). Unsere Ergebnisse implizieren, dass Substratzugang und Produktfreisetzung über zwei verschiedene Kanäle erfolgen. Dieses Szenario wurde in anderen Arbeiten als Grund für die extrem hohe katalytische Aktivität von Acetylcholinesterasen vorgeschlagen (Gilson et al. 1994; Bartolucci et al. 1999).

Strukturelle Grundlage der Regioselektivität in CYP2C9:

Die Berechnung der Bindungsdissoziationsenergie von C-H Gruppen (J. Gasteiger 1988) in (*S*)-Warfarin deutet auf eine ausreichend hohe intrinsische Reaktivität für eine Hydroxylierung in mehr als 10 Positionen hin. Die CYP-katalysierte Hydroxylierung von (*S*)-Warfarin ist jedoch hoch regioselektiv und wird von der Form der Substratbindekavität beeinflusst. Während CYP2C9 die Hydroxylierung an Position 7, 6 und 4 mit einem Produktverhältnis von 71%: 22%: 7% katalysiert (Rettie et al. 1992), hydroxyliert CYP3A4 in Position 4 und 10 (Ngui et al. 2001). Die experimentell beobachtete Regioselektivität von CYP2C9 kann jedoch nicht anhand der Kristallstruktur des Enzym-(*S*)-Warfarin-Komplexes vorhergesagt werden, da sich darin das Substrat in einem Abstand von 10 Å vom Hämeisen befindet (Williams et al. 2003). Die Berechnung der Dynamik des Enzym-Substrat-Komplexes zeigte das Wandern des Substrates in der Substratbindekavität. In einer der 16 durchgeführten Simulationen wurde das Wandern des Substrates zur Hämgruppe beobachtet. Dadurch kommen selektiv die an die Kohlenstoffe C6 und C7 gebundenen Wasserstoffatome

des (*S*)-Warfarin Moleküls in einen Abstand <3 Å zum aktivierten Häm-Sauerstoff, was den ersten Schritt der Hydroxylierungsreaktion - die Wasserstoffabstraktion - ermöglicht (Shaik et al. 2005). Im Folgenden wird dieser Zustand des Enzym-Substrat-Komplexes als 6-7-Hydroxylierungszustand bezeichnet. In diesem Zustand ist C7 von (*S*)-Warfarin dreimal häufiger in einem Abstand <3 Å als C6. Diese klare geometrische Präferenz ist das Ergebnis einer definierten Orientierung und Positionierung des (*S*)-Warfarin-Moleküls in dem engen trichterförmigen Hämzugangskanal, der Teil des starren Proteinkerns ist. Zwei MD-Simulationen zeigten das Substrat (*S*)-Warfarin in einer Orientierung, in der sich dessen Phenylgruppe bis auf 7 Å an den aktivierten Häm-Sauerstoff annähert. In der Mehrzahl der Trajektorien verblieb das Substrat allerdings in der hydrophoben Bindetasche, in der es in der Kristallstruktur zu finden war. Das hier vorgestellte Modell spiegelt die in Experimenten beobachtete Regioselektivität von CYP2C9 gegenüber (*S*)-Warfarin wider. Weiterhin lässt die Identifizierung des 6-7-Hydroxylierungszustands Vorhersagen über die Regioselektivität beeinflussende Aminosäuren zu. Wie in einer vorangegangenen Studie gezeigt werden konnte, ändert die Substitution von Phe476 die Regioselektivität von CYP2C9 (Melet et al. 2003). Diese Aminosäure ist Teil der Substratbindekavität und beeinflusst trotz ihrer relativ großen Entfernung von der Hämgruppe (>14 Å) die Orientierung des Substrates im 6-7-Hydroxylierungszustand. Die 4 Aminosäuren, die den engen Zugangskanal formen, haben jedoch einen viel stärkeren Einfluss auf die Orientierung des Substrates in der Nähe des aktiven Zentrums. Daraus wird gefolgert, dass die Mutagenese von Ala297, Thr301, Leu362 und Leu366 zu einer drastischen Änderung der Regioselektivität führen sollte. Die selektivitätsbestimmende Rolle des engen Zugangskanals, welcher die Zugänglichkeit des Substrates zur Hämgruppe beschränkt, wird auch dadurch bestätigt, dass Hydroxylierung in Position 4 von (*S*)-Warfarin kaum vorkommt. Position 6 und 7 befinden sich an der Basis des Y-förmigen Substratmoleküls, während Position 4 am sperrigen Ende liegt. Der limitierte Hämzugang erschwert demnach die Annäherung von Position 4 an den aktivierten Häm-Sauerstoff, während die Positionen 6 und 7 den engen Zugangskanal passieren können. Eine produktive Substratkonformation in einem Abstand <3Å von der Hämgruppe wurde zuvor in der Kristallstruktur des CYP2C9-Flurbiprofen-Komplexes beobachtet (Wester et al. 2004). Das stabförmige Fluorbiprofen-Molekül passt darin gut in den engen Trichter und kontaktiert die Hämgruppe auf ähnliche Weise wie (*S*)-Warfarin im 6-7-Hydroxylierungszustand. Die Überlagerung der Durchschnittsstruktur des 6-7-Hydroxylierungszustandes und des CYP2C9-Flurbiprofen-Komplexes (PDB-Eintrag 1R9O) zeigt, dass die bevorzugten Hydroxylierungs-stellen beider Substrate um nur 1,4 Å voneinander abweichen. Die den Trichter formenden

Aminosäuren beider Komplexe weichen nur 1,1 Å von einander ab. Während die Struktur beider Komplexe in dem trichterförmigen Bereich in der Nähe der Hämgruppe sehr ähnlich ist, unterscheiden sie sich in anderen Bereichen der Substratbindekavität beträchtlich. Diese Abweichungen können für Unterschiede in der katalytischen Aktivität verantwortlich sein. Die Regioselektivität von CYP2C9 kann dagegen mit der Existenz des starren, engen trichterförmigen Bereichs erklärt werden.

Die Rolle der Bindestelle für (*S*)-Warfarin in der Kristallstruktur in Enzymkinetiken wie Aktivierung und Autoaktivierung:
Die große Substratbindekavität, welche mehrere Substratmoleküle gleichzeitig beherbergen kann (Hummel et al. 2004), sowie die Existenz einer affinen Bindetasche für (*S*)-Warfarin weit entfernt vom Hämzentrum, wird in der Literatur als ein Grund für die Aktivierung der Umsetzung anderer Substrate in Gegenwart von (*S*)-Warfarin diskutiert (Williams et al. 2003). Vergleicht man die Position des Substrats in der Kristallstruktur mit der nahe am Häm-Sauerstoff zeigt sich, dass beide überlappen, d.h. beide Positionen können wahrscheinlich nicht von zwei (*S*)-Warfarin Molekülen gleichzeitig eingenommen werden. Die Beobachtung, dass (*S*)-Warfarin selbst keiner Autoaktivierung unterliegt (Stresser et al.), ist demnach wahrscheinlich darauf zurückzuführen, das die aktivierende Bindestelle und die produktive Bindestelle überlappen.

2.2 Die systematische Analyse der Substraterkennungsstelle 5 zur Identifizierung selektivitätsbestimmender Aminosäuren in Cytochrom P450-Monooxygenasen

(siehe Publikation: „Identification of selectivity-determining residues in cytochrome P450 monooxygenases: a systematic analysis of the substrate recognition site 5" Abschnitt 3.2)

Die gezielte Verbesserung von P450-Monooxygenasen für den Einsatz in der Biokatalyse setzt das Verständnis der molekularen Grundlagen biochemischer Eigenschaften wie Substratspezifität und Regioselektivität voraus. Von besonderem Interesse ist dabei die Identifizierung von Aminosäurepositionen, nach Möglichkeit auf Sequenzebene, die Einfluss auf diese Eigenschaften haben. Mit dem Wissen um solche *hotspots* ist es möglich, mit relativ geringem Aufwand gezielt neue Enzymvarianten mit verbesserten Eigenschaften herzustellen. In vorangegangenen Arbeiten wurden mehr als 6300 Proteinsequenzen sowie Kristall-

strukturen von 31 verschiedenen P450-Monooxygenasen zu einer Datenbank vereint und so einer systematischen Analyse zugänglich gemacht (Fischer et al. 2007). Auf Sequenzebene ist in P450-Monooxygenasen nur eine geringe Anzahl an konservierten Aminosäuren bekannt. Dazu gehören: das die Hämgruppe bindende Cystein, das essenziell für die Funktion von P450-Monooxygenasen ist, ein Phenylalanin, das 7 Aminosäuren N-terminal vom konservierten Cystein zu finden ist, das Glutaminsäure/Arginin Paar des ExxR-Motivs (Ravichandran et al. 1993) sowie Alanin, Glycin und Threonin des AGxxT-Motivs der Helix I (Mestres 2005), wobei das Threonin am Transfer von Protonen zum Hämzentrum beteiligt ist (Vidakovic et al. 1998). Darüber hinaus ließ die Analyse von 4 Kristallstrukturen und 200 Proteinsequenzen die Existenz eines funktionell konservierten, mit der Hämgruppe integrierenden Arginins in P450-Monooxygenasen vermuten (Oprea et al. 1997). Ein vorangegangener Strukturvergleich von P450-Monooxygenasen aus 9 verschiedenen Familien offenbarte mehrere konservierte Elemente: die Helix E, die C-terminale Hälfte der Helix I, Helices J und K (Helix K enthält das konservierte ExxR-Motiv), den β1-3 Strang, Helices K´ und K´´, die Cystein-Tasche, Helix L, sowie den β3-2 Strang (Mestres 2005).

Das aktive Zentrum von P450-Monooxygenasen, die Hämgruppe, befindet sich tief im Inneren des Enzyms, am Grund der Substratbindekavität. Wie die Simulationen der Dynamik des CYP2C9-Warfarin-Komplexes (Abschnitt 2.1) gezeigt haben, sind die Aminosäuren des Proteins, die den Zugang zur Hämgruppe beeinflussen, von großer Bedeutung für die Orientierung des Substrates während der Katalyse und haben damit großen Einfluss auf Substratspezifität und Regioselektivität. Für die gezielte Verbesserung der Substratspezifität und Regioselektivität ist es daher entscheidend Aminosäuren zu identifizieren, die in der Nähe der Hämgruppe liegen und deren Seitenketten zum Hämzentrum ausgerichtet sind. Drei Strukturelemente bilden die Substratbindekavität in direkter Umgebung der Hämgruppe. Dies sind Helix I, der B-C *loop* und die Substraterkennungsstelle 5 (SRS-5), welche sich vom strukturell hoch konservierten ExxR-Motiv bis in den β1-4 Strang erstreckt (Gotoh 1992). Während die Helix I konserviert ist, sind der B-C *loop* und die SRS-5 Region in Sequenz und Struktur variabel. Durch die direkte Nachbarschaft der SRS-5 Region zum hoch konservierten ExxR-Motiv war es möglich, diese Region in über 6300 verschiedenen CYP Sequenzen zu identifizieren und zu analysieren. Neben der Sequenzanalyse wurden auch strukturelle Unterschiede und Gemeinsamkeiten der SRS-5 Region untersucht. Dazu wurde dieser Bereich in Kristallstrukturen von 31 verschiedenen P450-Monooxygenasen verglichen.

Die Ergebnisse der systematischen Analyse von 31 Kristallstrukturen und 6300 Sequenzen zeigten, dass zusätzlich zu den bekannten hoch konservierten Aminosäuren 97,7 % aller P450-Monooxygenasen eine positiv geladene Aminosäure am C-terminalen Ende der SRS-5 Region besitzen. Diese Aminosäure ist weder Teil eines Sequenzmotivs, noch kann sie durch Sequenzalignment aller P450-Monooxygenasen identifiziert werden. Sie ist strukturell konserviert und formt eine Salzbrücke zum 7' Propionatrest der Hämgruppe in Kristallstrukturen. Der Vergleich von 4 CYP-Kristallstrukturen und 200 Sequenzen deutete bereits auf die Existenz einer funktionell konservierten Aminosäure hin, welche mit der Hämgruppe interagiert und an der Eliminierung von Wasser aus dem aktiven Zentrum beteiligt ist (Oprea et al. 1997). Mutagenesedaten zeigen darüber hinaus einen starken Einfluss auf die Hämbindung und die Stabilität der Tertiärstruktur (He et al. 1997). Im Rahmen dieser Arbeit konnte gezeigt werden, das dieser mit der „Hämgruppe interagierende Rest" (HIR) in Position 9, 10, 11 oder 12 nach dem auf Strukturebene konservierten ExxR-Motiv vorkommen kann. Da das ExxR-Motiv auch auf Sequenzebene konserviert ist wurde es möglich, den HIR in 97,7% aller hier untersuchten Sequenzen (6379) zu identifizieren. So konnten wir unter Verwendung einer wesentlich größeren Datenbasis zeigen, dass der HIR ein gemeinsames Merkmal fast aller P450-Monooxygenasen ist. Zusätzlich dazu analysierten wir die 2,3% der P450-Monooxygenasen, die keinen HIR besitzen. Die meisten dieser Enzyme sind Fettsäurehydroxylasen und Fusionsproteine von Monooxygenase und Reduktase. Interessanterweise haben 98,3% aller „Nichtfusionsproteine" einen HIR, was die Relevanz des HIR für „Nichtfusionsproteine" bestätigt. Für Fusionsproteine wiederum scheint ein HIR weniger wichtig zu sein, da nur 34% der Fusionsproteine einen HIR aufweisen.

P450-Monooxygenasen verfügen im Inneren über eine große Substratbindekavität. Da die Größe und Form der Substrate vieler P450-Monooxygenasen variiert, können Aminosäuren aus verschieden Bereichen der Substratbindekavität die Substratorientierung in der Nähe der Hämgruppe beeinflussen und damit an der Kontrolle der Regioselektivität beteiligt sein (Melet et al. 2003; Keizers et al. 2004; Sherman et al. 2006). Aminosäuren in der direkten Umgebung des aktivierten Sauerstoffs der Hämgruppe sind jedoch wahrscheinlich mit jedem Substrat während der Oxidationsreaktion in Kontakt, unabhängig von dessen Größe und Form. In direkter Umgebung der Hämgruppe bilden 3 Strukturelemente die Substratbindekavität. Dies sind Helix I, der B-C *loop* und die sich vom strukturell hoch konservierten ExxR-Motiv bis einschließlich β1-4 Strang erstreckende SRS-5 Region (Gotoh 1992). Jedes dieser Strukturelemente beherbergt ein bis zwei Aminosäuren, deren

Seitenketten zum Hämzentrum hin ausgerichtet sind und damit den Zugang zum Häm direkt beeinflussen. Die beiden von der Helix I beherbergten Aminosäuren Alanin und Threonin sind Teil des konservierten AGxxT-Motivs. Es wurde gezeigt, dass Aminosäuresubstitutionen in diesem Bereich zum Verlust der Enzymaktivität führen (Clark et al. 2006). Der B-C *loop* enthält die SRS-1 Region und ist hoch variabel in Struktur und Sequenz. Daher ist es äußerst schwierig ohne Strukturinformationen Aminosäuren aus diesem Bereich zu identifizieren, die während der Katalyse mit den Substraten interagieren. Das dritte Element, die SRS-5 Region, enthält bis zu 11 Aminosäuren, die an der Substratbindung beteiligt sein können (Gotoh 1992). Aus den Ergebnissen der systematischen Analyse der SRS-5 Region in 31 Strukturen und über 6300 Sequenzen wurden Regeln zur einfachen Identifizierung von ein oder zwei Positionen in der hoch variablen SRS-5 Region abgeleitet, welche aufgrund ihrer Exponiertheit und großen Nähe zum Hämzentrum bevorzugt an der Substratbindung während der Oxidationsreaktion beteiligt sind. Daraus folgern wir, dass diese Positionen einen großen Einfluss auf Substratspezifität und Regioselektivität gegenüber allen Substraten haben. Beide Positionen können aufgrund ihrer Nähe zum hoch konservierten ExxR-Motiv auch ohne Proteinstrukturinformationen nur anhand der Sequenz identifiziert werden. Eine solche Aminosäure wird in 98.4% aller P450-Monooxygenasen für Position 5 nach dem ExxR-Motiv vorhergesagt. Für die verbleibenden 1,6% kann keine Vorhersage getätigt werden. Diese P450-Monooxygenasen tragen den HIR in Position 12 (jedoch nicht in Position 9, 10, 11) und können damit eindeutig identifiziert werden. Eine zweite exponierte und damit präferentiell mit Substraten interagierende Aminosäure wird nur für einen Teil der vorhandenen P450 Strukturen beobachtet. Aufgrund der in dieser Arbeit aufgeklärten Sequenz-Struktur-Beziehungen war es möglich jene P450-Monooxygenasen vorherzusagen, welche diese zweite präferentiell mit Substraten interagierende Aminosäure aufweisen. Es handelt sich dabei um P450-Monooxygenasen mit dem Sequenzmuster EXXR-X(7)-{P}-**x**-P-[HKR]. Das Sequenzmuster wurde in 27% aller P450-Monooxygenasen identifiziert. Diese Ergebnisse deuten darauf hin, dass die Orientierung der Aminosäure **x** durch ihre Nachbaraminosäure Prolin in Position 10 bestimmt wird. Ein solcher Einfluss von Prolin auf Nachbaraminosäuren wurde bereits für andere X-Pro Peptide beobachtet (Macarthur und Thornton 1991). Die Salzbrücke zwischen dem HIR in Position 11 und dem 7'-Propionatrest der Hämgruppe bewirkt eine weitere Restriktion der Konformation dieser Region. Interessanterweise scheint Prolin in Position 8 die Entfernung der Aminosäure in Position 9 vom Hämzentrum zu bewirken und wurde deshalb von dem Sequenzmuster ausgeschlossen.

Die Analyse der über 6300 CYP Sequenzen zeigte weiterhin, dass vornehmlich hydrophobe Aminosäuren unterschiedlicher Größe (im speziellen Leucin, Isoleucin, Valin und Alanin) in Position 5 und 9 zu finden sind. Da die Seitenketten der Aminosäuren in diesen Positionen zum Hämzentrum hin ausgerichtet sind, kann ihr Austausch durch sperrigere Aminosäuren die Anzahl möglicher Substratorientierungen beschränken. Dies führt zur Verringerung der Anzahl möglicher Oxidationsprodukte und damit zur Erhöhung der Regioselektivität (Fruetel et al. 1994; Bell et al. 2003). Darüber hinaus wird erwartet, dass ein limitierter Zugang zur Hämgruppe die Umsetzung sperriger Moleküle verhindert, was wiederum entscheidend für die Substratspezifität von P450-Monooxygenasen sein kann. Die Richtigkeit der Vorhersagen zum Einfluss der Position 5 nach dem ExxR-Motiv sowie der Position 9 in P450-Monooxygenasen mit dem Sequenzmuster EXXR-X(7)-{P}-**x**-P-[HKR] auf Substratspezifität und Regioselektivität konnte mit Hilfe experimenteller Literaturdaten validiert werden (Born et al. 1995; Schalk und Croteau 2000; Kerdpin et al. 2004; Liu et al. 2004; Lentz et al. 2006). Dabei zeigte sich, das Aminosäuresubstitutionen in diesen Positionen bei P450-Monooxygenasen verschiedener Familien (Sequenzidentität <55%) zu starken Änderungen der Regioselektivität und Substratspezifität führte.

2.3 Rationales Design einer minimalen und hoch angereicherten Mutantenbibliothek zur Steigerung der Regio-, Stereo- und Chemoselektivität von CYP102A1

(siehe Publikation: „Rational design of a minimal and highly enriched CYP102A1 mutant library with improved regio-, stereo-, and chemoselectivity" Abschnitt 3.3)

CYP102A1 ist eine der am besten untersuchten Cytochrom P450-Monooxygenasen. Das Enzym ist auch unter dem Namen P450 BM-3 bekannt, was es der Tatsache verdankt, die dritte Monooxygenase zu sein, die aus *Bacillus megaterium* isoliert wurde. Es handelt sich hierbei um ein lösliches, d. h. nicht membrangebundenes Fusionsprotein einer P450-Monooxygenase und einer FAD/FMN enthaltenden Reduktase (Narhi und Fulco 1987). CYP102A1 ist relativ stabil unter Prozessbedingungen (Kuehnel et al. 2007). Diese Eigenschaften machen es zu einer der vielversprechendsten P450-Monooxygenasen für den Einsatz in biotechnologischen Prozessen. Das Wildtypenzym ist eine hoch aktive Fettsäure-Hydroxylase (Munro et al. 2002), jedoch kann das Substratspektrum durch Mutationen

erweitert werden (Urlacher et al. 2004). Das Wildtypenzym weist eine geringe Selektivität auf. In einer Reihe von Arbeiten konnte gezeigt werden, dass die Regio- und Stereoselektivität gegenüber verschiedenen Substraten durch Mutationen erhöht werden kann. Im Rahmen dieser Arbeit wurden 4 Terpene ((4*R*)-Limonen, (+)-Valencen, Nerylaceton und Geranylaceton) untersucht, deren Oxidation zu interessanten und wertvollen Produkten führen kann. (4*R*)-Limonen und (+)-Valencen bilden die Hauptbestandteile ätherischer Öle aus den Schalen der Zitrusfrüchte. Sie sind billig und in großen Mengen verfügbar. Es wurde bereits gezeigt, dass die P450-Monooxygenase CYP102A7 (4*R*)-Limonen umsetzt. Die Oxidation geschieht jedoch unselektiv und führt zur Bildung von (4*R*)-Limonen-1,2-epoxid, (4*R*)-Limonen-8,9-epoxid und Carveol (Dietrich et al. 2008). Ein weiteres interessantes Terpen ist (+)-Valencen, welches durch selektive Oxidation am C2 Kohlenstoff zu (+)-Nootkaton umgewandelt werden kann. Bei letzterem handelt es sich um einen wertvollen Aromastoff der Grapefruit. Durch chemische Oxidation sowie durch Biotransformation mittels *Chlorella* und *Mucor* Spezies kann (+)-Nootkaton aus (+)-Valencen gebildet werden (Furusawa et al. 2005). Darüber hinaus wurde gezeigt, dass Mutanten von CYP101A1 (P450cam) und CYP102A1 die Oxidation von (+)-Valencen katalysieren (Sowden et al. 2005). P450cam Mutanten zeigten dabei eine vergleichsweise hohe Regioselektivität für die Oxidation am C2 Atom (85%), die gemessene Aktivität war jedoch gering. Im Gegensatz dazu zeigten CYP102A1 Mutanten eine höhere Aktivität, waren jedoch unselektiv. Weiterhin wurde gezeigt, dass CYP102A1 Geranyl- und Nerylaceton unselektiv oxidiert (Watanabe et al. 2007). In der gleichen Arbeit wurde eine Dreifachmutante dieses Enzyms vorgestellt (R47L/Y51F/F87V), die Geranylaceton hochselektiv zu 9,10-Epoxygeranylaceton umsetzt. Eine Verbesserung der Regioselektivität von CYP102A1 gegenüber Nerylaceton wurde dagegen nicht erreicht. Hydroxylierung in Position C11 und C12 beider Substrate führt zu wertvollen Ausgangsstoffen für die Synthese von Naturstoffen wie Indol-Alkaloiden (Clark et al. 2005), Furanocembran (Marshall und Dubay 1994), aus Braunalgen gewonnenen C18 Terpenoiden (Li et al. 1994), den makrozyklischen Terpenoiden Humulen, Flexibilen und Helminthogermacren (Mcmurry und Kocovsky 1984; Mcmurry und Kocovsky 1985; Mcmurry et al. 1987), sowie von Cyclopropan abgeleitete Substanzen (Charette et al. 1996; Charette et al. 1998). Hydroxyprodukte in Position C11 und C12 wurden jedoch bisher bei der Umsetzung von Geranylaceton nicht gebildet. Das Anliegen dieser Arbeit bestand nun darin, einen Biokatalysator auf der Basis von CYP102A1 zu entwickeln, der verschiedene Substrate mit hoher Selektivität oxidiert.

Enzymeigenschaften wie Aktivität, Selektivität und Spezifität können durch eine Vielzahl verschiedener Methoden verbessert werden. Die Methode der gerichteten Evolution, d. h. mehrere Runden zufälliger Mutationen des gesamten Gens mit anschließender Isolierung der besten Mutante, die wiederum als Ausgangspunkt für die Erzeugung einer neuen Generation von Mutanten dient, führte in vielen Fällen zur Identifizierung von Enzymvarianten mit verbesserten Eigenschaften (Kuchner und Arnold 1997). Informationen über die Struktur und mechanistische Details eines Enzyms können dazu verwendet werden, die Mutagenese auf bestimmte Sequenzbereiche zu beschränken, wodurch eine deutliche Verkleinerung der zu durchsuchenden Mutantenbibliothek erreicht wird. Diese Strategie wurde bereits erfolgreich angewendet, um die Substratspezifität von CYP102A1 zu erweitern (Li et al. 2001). Auch die Durchführung von Sättigungsmutagenesen an bestimmten aus der Struktur identifizierten Positionen wurde erfolgreich eingesetzt, um CYP102A1 für die enantioselektive Epoxidierung von Alkenen zu optimieren (Kubo et al. 2006). Allerdings kann selbst unter Verwendung von Strukturinformation die Anzahl an potentiell Selektivität beeinflussenden Positionen relativ hoch sein. Eine umfassende Analyse möglicher kooperativer Effekte zwischen den verschiedenen Positionen ist dann aufgrund der großen Menge an kombinatorischen Möglichkeiten nicht durchführbar. Eine schrittweise Verbesserung biochemischer Eigenschaften eines Enzyms kann durch iterative Zyklen kombinatorischer Sättigungsmutagenesen an aus der Proteinstruktur gewählten Positionen erreicht werden (Reetz et al. 2006). Dadurch wird die Wahrscheinlichkeit erhöht, kooperative Effekte zwischen verschiedenen Positionen zu identifizieren.

Die Limitationen der genannten Methoden bestehen jedoch darin, dass (1) kooperative Effekte zwischen Mutationen in verschiedenen Positionen nur dann identifiziert werden, wenn mindestens eine der zugrundeliegenden Einzelmutanten bereits verbesserte Eigenschaften aufweist (2) für das effiziente Durchsuchen von großen Mutantenbibliotheken für jedes Substrat ein geeignetes Durchmusterungsverfahren zu entwickeln ist (3) auch bei der Suche mit verschiedenen Substraten nach verbesserten Varianten des gleichen Enzyms Mutantenbibliotheken zumindest teilweise neu erstellt werden müssen. Daher war es Ziel dieser Arbeit, eine einzige Mutantenbibliothek minimaler Größe zu konstruieren, welche hoch angereichert ist an Enzymvarianten mit erhöhter oder veränderter Regio-, Stereo- und Chemoselektivität gegenüber verschiedenen Substraten, anstatt für jedes neue Substrat eine neue Mutantenbibliothek zu generieren. Dazu wurden die in den vorliegenden Modellierungsarbeiten gesammelten Erkenntnisse zur molekularen Grundlage von Selektivität in P450-Monooxygenasen genutzt und sich auf den die Hämgruppe direkt

umgebenden Bereich konzentriert. Wie im Rahmen dieser Arbeit gezeigt werden konnte, hat dieser Bereich in der humanen P450-Monooxygenase CYP2C9 einen entscheidenden Einfluss auf die Orientierung des Substrates (*S*)-Warfarin in der Nähe des Häms und kontrolliert dadurch die Regioselektivität der Oxidationsreaktion (Abschnitt 2.1). Aus dem entsprechenden Bereich in CYP102A1 wurden 2 Positionen gewählt (Position 87 und 328), die aufgrund ihrer Nähe zum Hämzentrum mit jedem potentiellen Substrat während der Oxidationsreaktion interagieren. Position 328 entspricht der Position 5 nach dem hoch konservierten ExxR-Motiv und wurde basierend auf der systematischen Analyse von 31 Kristallstrukturen, über 6300 Sequenzen sowie experimentellen Mutagenesedaten als *hotspot* für Substratspezifität und Regioselektivität in fast allen P450-Monooxygenasen vorhergesagt (Abschnitt 2.2). Ein starker Einfluss von Position 328 in CYP102A1 auf die Regio- und Stereoselektivität der Oxidation von Alkanen und Alkenen wurde darüber hinaus bereits experimentell nachgewiesen (Kubo et al. 2006; Meinhold et al. 2006). Auch der Einfluss von Position 87 in CYP102A1 auf die Regio- und Stereoselektivität der Oxidation verschiedener Substrate wurde bereits gezeigt (Graham-Lorence et al. 1997; Carmichael und Wong 2001; Urlacher et al. 2006; Li et al. 2008). Um mögliche kooperative Effekte beider Positionen zu untersuchen, wurden alle möglichen Kombinationen von 5 hydrophoben Aminosäuren (Alanin, Valin, Leucin, Isoleucin und Phenylalanin) in diesen *hotspot* Positionen generiert. Die Beschränkung auf 5 hydrophobe Aminosäuren basiert auf der Erkenntnis, dass in P450-Monooxygenasen Aminosäuren in unmittelbarer Umgebung der Hämgruppe, deren Seitenketten zum Hämzentrum ausgerichtet sind, vorwiegend diese hydrophoben Aminosäuren zu finden sind (Abschnitt 2.2). Die aus dieser Vorgehensweise resultierende Mutantenbibliothek hat eine minimale Größe. Die Substratbindetaschen der darin enthaltenen 25 Enzymvarianten zeigen in dem für Selektivität und Spezifität wichtigen Bereich in der Nähe des Häms eine große Vielfalt an Formen.

Diese Mutantenbibliothek wurde mit den vier genannten Terpenen durchmustert. Durch die Entwicklung einer geeigneten Analytik konnten die gebildeten Oxidationsprodukte identifiziert und quantifiziert werden. Einige der gebildeten Oxidationsprodukte sind kommerziell nicht als Standards erhältlich. Deshalb mussten die Epoxy- und Hydroxyverbindungen des Geranyl- und Nerylacetons sowie das (4*R*)-Limonen-8,9-Epoxid synthetisiert werden. Diese Arbeiten wurden in der Arbeitgruppe von Prof. Dr. Laschat am Institut für organische Chemie der Universität Stuttgart durchgeführt. Die Durchmusterung der minimalen Mutantenbibliothek ergab 12 Enzymvarianten mit erweitertem Produktspektrum und erhöhter Regio- und Stereoselektivität gegenüber mindestens einem der

4 Terpene. Nur 3 Enzymvarianten waren inaktiv gegenüber allen 4 Terpenen. Zu den gebildeten Oxidationsprodukten gehört der wertvolle Aromastoff (+)-Nootkaton. Weiterhin werden hier erstmalig CYP102A1-Varianten vorgestellt, die im Gegensatz zum Wildtypenzym 11- und 12-Hydroxygeranylaceton produzieren. In einer vorangegangen Arbeit wurden das CYP102A1 Wildtypenzym sowie 25 Varianten mit Mutationen in 5 Positionen (inklusive Position 87) mit Geranyl- und Nerylaceton durchmustert (Watanabe et al. 2007). Dabei wurde eine Dreifachmutante identifiziert, welche Geranylaceton mit hoher Regio- und Stereoselektivität zu 9,10-Epoxygeranylaceton umsetzt. In dieser Dreifachmutante war Phe87 durch Valin ersetzt. Im Einklang damit zeigen unsere Ergebnisse, dass der Austausch vom Phe87 durch Valin, aber auch durch Leucin und Isoleucin, eine starke Erhöhung der Selektivität für die Bildung von 9,10-Epoxygeranylaceton bewirkt. Darüber hinaus zeigen wir, dass durch Aminosäureaustausch in Position 328 das Substratspektrum zugunsten von Hydroxyprodukten erweitert werden kann. Kombinierte Mutationen in Position 87 und 328 führten zu einer CYP102A1-Variante, die fast ausschließlich Hydroxyprodukte bildet und eine ausgeprägte Regioselektivität für das C12 Atom aufweist. Für die Umsetzung von Nerylaceton hatte Position 328 den stärksten Einfluss auf die Regio- und Chemoselektivität. Obwohl Geranyl- und Nerylaceton Strukturisomere sind, zeigte im Gegensatz zu Geranylaceton, interessanterweise nur eine einzige Mutante eine starke Erhöhung der Regioselektivität gegenüber Nerylaceton. Es ist allgemein anerkannt, dass eine erhöhte Regio- und Stereoselektivität das Ergebnis einer reduzierten Anzahl von Substratorientierungen in der Nähe des aktiven Zentrums ist (Raag und Poulos 1991; Bell et al. 2003; Branco et al. 2008). Unsere Ergebnisse zeigen, dass die systematische Kombination hydrophober Aminosäuren unterschiedlicher Größe und Form in der direkten Umgebung der Hämgruppe ein präferenziell epoxidierendes Enzym in ein hydroxylierendes Enzym verwandeln kann. Folglich kann neben Regio- und Stereoselektivität auch die Chemoselektivität von CYP102A1 durch Mutationen verändert werden, die die Orientierung des Substrats in der Nähe des Hämzentrums beeinflussen.

Die selektive Oxidation von (+)-Valencen am C2-Kohlenstoff führt zur Bildung von (+)-Nootkatol. Durch einen weiteren Oxidationsschritt am selben Kohlenstoffatom entsteht der wertvolle Aromastoff (+)-Nootkaton. In einer vorangegangenen Arbeit wurde eine CYP102A1-Dreifachmutante (R47L/Y51F/F87A) vorgestellt, die im Gegensatz zur Doppelmutante R47L/Y51F und zum Wildtypenzym (+)-Nootkaton in geringen Mengen produziert. Diese Dreifachmutante produziert jedoch eine Reihe zusätzlicher Oxidationsprodukte (Sowden et al. 2005). Im Rahmen der vorliegenden Arbeit konnte

bestätigt werden, dass Position 87 einen Einfluss auf die Orientierung des Substrates in der Nähe der Hämgruppe hat. Der Austausch von F87 durch Alanin erhöhte die C2-Regioselekivität von 22% im Wildtypenzym auf 55% in der F87A Mutante. Darüber hinaus hatte dieser Aminosäureaustausch die Bildung geringer Mengen an (+)-Nootkaton zur Folge. Die Doppelmutante F87A/A328I zeigte jedoch eine deutlich stärkere Erhöhung der C2-Regioselektivität (95%). Die Kombination der Mutation in Position 87 mit einer Mutation in Position 328 bewirkt demzufolge eine wesentlich stärkere Beschränkung möglicher Substratorientierungen in der Nähe der Hämgruppe. Auch die CYP102A1-Varianten F87V/A328I, F87A/A328V und F87V/A328V zeigten eine stark erhöhte C2-Regioselektivität, demnach kann die Kombination verschiedener Aminosäuren in beiden Positionen zur Erhöhung der C2-Regioselektivität führen. Interessanterweise setzten die Einzelmutanten A328V und A328I (+)-Valencen nicht um, was bedeutet, dass die Auswirkungen der Mutationen in Position 87 und 328 nicht additiv sind, sondern einen kooperativen Charakter haben. Das macht es unmöglich die hohe Selektivität der Doppelmutanten aufgrund der Eigenschaften der zugrundeliegenden Einzelmutanten zu finden. Ein iteratives Verfahren würde die Varianten F87A/A328V, F87A/A328I, F87V/A328I und F87V/A328V finden, da die Einzelmutanten F87A und F87V im Vergleich zum Wildtypenzym eine leicht erhöhte C2-Regioselektivität aufweisen. Die Präferenz der Doppelmutante F87V/A328L für C12-Hydroxylierung von Geranylaceton könnte jedoch nicht mittels eines iterativen Verfahrens gefunden werden, da die Variante F87V bevorzugt epoxidiert und die Variante A328L Geranylaceton nicht umsetzt.

Im Allgemeinen zeigten mehr Einzelmutanten als Doppelmutanten eine verbesserte Regio- und Chemoselektivität bei der Oxidation der kleineren azyklischen Terpene Neryl- und Geranylaceton, während alle CYP102A1-Varianten mit erhöhter Selektivität gegenüber (4*R*)-Limonen und (+)-Valencen Doppelmutanten waren. Das zeigt, dass kombinierte Mutationen in beiden Positionen erforderlich sind, um die Selektivität gegenüber den sperrigeren Substraten (4*R*)-Limonen und (+)-Valencen zu verbessern.

Die Ergebnisse dieser Arbeit zeigen, dass die systematische Kombination von nur 5 hydrophoben Aminosäuren in Position 87 und 328 ein effektiver Ansatz zur Verbesserung der Regio-, Stereo- und Chemoselektivität ist. Die zum Durchmustern unserer minimalen Mutantenbibliothek verwendeten Verbindungen haben unterschiedliche Größe, Form und Polarität. Der Einfluss beider *hotspot* Positionen auf die Selektivität von CYP102A1 ist darüber hinaus nicht auf Terpene beschränkt. Daher erwarten wir, dass unsere Mutantenbibliothek auch eine reichhaltige Quelle für Biokatalysatoren mit verbesserter

Selektivität gegenüber anderen Substraten ist. Ein idealer CYP-Katalysator müsste sowohl ein breites Substratspektrum als auch eine hohe Selektivität aufweisen. Die Ergebnisse dieser Arbeit zeigen, dass diese Ansprüche nicht von einer einzelnen Enzymvariante erfüllt werden kann, sondern eher von einer möglichst kleinen Sammlung von Varianten. Die minimale Mutantenbibliothek ist ein erster Schritt in Richtung eines solchen idealen CYP Katalysators, da sie eine kleine Sammlung von aktiven CYP102A1-Varianten mit verbesserter Selektivität gegenüber verschiedenen Substraten zur Verfügung stellt. Es wird erwartet, dass der hier vorgestellte Ansatz auch für andere P450-Monooxygenasen angewendet werden kann, da in allen CYP mit bekannter Struktur und dazu homologen CYP die der Position 87 entsprechende Position identifiziert werden kann. Die der Position 328 entsprechende Position kann darüber hinaus wegen ihres definierten Abstands zum hoch konservierten ExxR-Motiv in fast allen CYP allein anhand der Proteinsequenz identifiziert werden (Abschnitt 2.2).

3 Publikationen

3.1

Publikation erschienen in *Proteins* 64(1): 147-155

Multiple Molecular Dynamics Simulations of Human P450 Monooxygenase CYP2C9: The Molecular Basis of Substrate Binding and Regioselectivity toward Warfarin.

Alexander Seifert, Stephan Tatzel, Rolf D. Schmid and Jürgen Pleiss[1]

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

[1]Corresponding author:
E-mail: Juergen_Pleiss@itb.uni-stuttgart.de
Fax (+49) 711-685-3196
Telephone (+49) 711-685-3191

Short title: Modelling human CYP2C9

Key words: enzyme dynamics, loop movements, channels

3.1.1 Abstract

To examine the molecular basis of activity and regioselectivity of the clinically important human microsomal cytochrome P450 (CYP) monooxygenase 2C9 towards its substrate warfarin, 22 molecular dynamics (MD) simulations (3-5 nanoseconds each) were carried out in the presence and absence of warfarin. The resulting trajectories revealed a stable protein

core and mobile surface elements. This mobility leads to the formation of two surface channels in the region between F-G loop, B´ helix/B-B´ loop, β_1 sheet, and between helices F and I and the turn in the C-terminal antiparallel β-sheet in the presence of warfarin.
Besides the non-productive state of the CYP2C9 warfarin complex captured in the crystal structure, three additional states were observed. These states differ in the shape of the substrate binding cavity and the position of the warfarin molecule relative to heme. In one of these states the 7- and 6- positions of warfarin contact the heme with a marked geometrical preference for position 7 over position 6. This modelling result is consistent with experimentally determined regioselectivity (71 % and 22 % hydroxylation in positions 7 and 6, respectively). Access to the heme group is limited by the core amino acids Ala297, Leu362, Leu366, and Thr301, which therefore are expected to have a major impact on regioselectivity. In addition, modelling predicts that auto-activation of warfarin is sterically hindered. Our study demonstrates how the combination of mobile surface and rigid core leads to interesting properties: a broad substrate profile and simultaneously a high regioselectivity.

Abbreviations: cytochrome P450 monooxygenase, CYP; molecular dynamics, MD; random expulsion molecular dynamics, REMD; root mean square deviation, RMSD

3.1.2 Introduction

Cytochrome P450 monooxygenases (CYPs) are involved in the metabolism of physiologically important compounds in many species of 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 CYP substrate selectivity is of considerable interest, particularly to the pharmaceutical industry (Smith et al. 1997; Smith et al. 1998) where an early prediction of likely drug metabolism pathways in *Homo sapiens* can aid in the design and development of drugs. CYP2C9 is one of the major drug metabolising isoforms. It contributes to the oxidative metabolism of 16% of all therapeutics in current clinical use (Rendic and Di Carlo 1997) and is involved in adverse drug effects. Human CYPs metabolize substrate molecules of different size and shape which implies a large, highly flexible substrate binding cavity. The common anti-coagulant drug warfarin is metabolised by CYP2C9 to form 7-hydroxywarfarin (71%), 6-hydroxywarfarin

(22%), and 4-hydroxywarfarin (7%) (Rettie et al. 1992). Like many other CYPs (Hlavica and Lewis 2001) CYP2C9 has been shown to have a large substrate binding site which can simultaneously bind several substrate molecules (Williams et al. 2003; Wester et al. 2004). There is experimental evidence for a "two-site binding model" for CYP2C9 mediated metabolism (Hutzler et al. 2001), however, there is no auto-activation by warfarin (Hemeryck et al. 1999). On the other hand, most of the catalysed reactions are highly regioselective which would require a more rigid enzyme.

In effort to understand the molecular basis of substrate specificity and regioselectivity the free enzyme and the enzyme-substrate complexes of several mammalian CYPs have been studied by X-ray analysis, docking, and molecular dynamics simulations (Wade 2004) (Wester et al. 2004). The comparison of free and substrate-complexed crystal structures of mammalian CYPs revealed that helices F and G and the loop between helices B and C undergo adaptive changes in the presence of substrates and inhibitors (Scott et al. 2003). Dramatic conformational changes were observed when the crystal structure of mammalian CYP2B4 was solved (Scott et al. 2004). The structure of the inhibitor and substrate-free enzyme exhibits a large open cleft which is formed mainly by helices F, F´, G´ and G on one side and helices B´ and C on the opposite side without loss of the general architecture of the CYP fold. The cleft extends from the protein surface directly to the heme. However, the crystal structure of CYP2B4 in presence of a specific inhibitor revealed a closed conformation of the protein (Scott et al. 2004). In contrast, the crystal structure of the human microsomal CYP2C9 (Williams et al. 2003) which has been crystallized in the absence and presence of its substrate warfarin did not show major conformational changes between the substrate-free structure and the enzyme-substrate complex. Furthermore, the analysis of channels in P450 crystal structures as well as random expulsion molecular dynamics simulations of bacterial CYP101, CYP102A1, CYP107A1, and mammalian CYP2C5 predicted for soluble CYPs one predominant ligand pathway situated between F-G loop, B´ helix/B-B´ loop, and β_1 sheet, while for mammalian membrane-bound CYP2C5 one additional route for substrates or products was suggested in the region between B´ helix/B-C loop, G and I helices (Wade 2004) (Schleinkofer et al. 2005). However, the only channel in the crystal structure of CYP2C9 is a solvent channel between helices F and I and the turn in the C-terminal antiparallel β-sheet, which is too small for substrate molecules to pass through (Wade 2004). How potential substrates access and products leave the substrate binding cavity remains elusive.

In all CYPs, the cofactor heme is deeply buried inside the protein at the bottom of a large, internal binding cavity. It has been suggested that the orientation of the bound substrate in the substrate binding cavity relative to the heme determine regioselectivity (Sason Shaik 2005). However, the CYP2C9 crystal structure in complex with warfarin does not provide information about the molecular basis of regioselectivity, because the substrate lies in a predominantly hydrophobic pocket at a distance of 10 Å from the heme (Williams et al. 2003). In contrast, in a 2.0 Å resolution crystal structure of the CYP2C9 in complex with flurbiprofen, the substrate is close to the heme in an orientation consistent with the experimentally determined regioselectivity of the enzyme (Wester et al. 2004) However, in this structure residues 214-220 between helices F and G and residues 38-42 are lacking.
We carried out extensive molecular dynamics simulations to locate the regions of human CYP2C9, where substrate access and product exit might occur, and to identify the structural basis of regioselectivity of this enzyme.

3.1.3 Materials and Methods

Initial structures:
Multiple MD simulations of CYP2C9 in water and complexed with the substrate (*S*)-warfarin in water were carried out. The initial structures were created from X-ray structures of free CYP2C9 (Protein Data Bank (Berman et al. 2000) entry 1OG2) and of a complex with (*S*)-warfarin (PDB entry 1OG5) at a resolution of 2.6 Å. Both crystal structures contain two protein molecules in the asymmetric unit cell with an all atom deviation of less than 0.65 Å. For the simulations chain A was used as initial structure. It should be noted that the sequence of 1OG2 and 1OG5 deviate from human CYP2C9. To improve crystallisation the protein was truncated by removing amino acids 1-29 and 7 amino acids were exchanged (Lys206Glu, Ile215Val, Cys216Tyr, Ser220Pro, Pro221Ala, Ile222Leu, and Ile223Leu). It has been shown that none of the amino acid substitutions leads to changes in activity, specificity, and the ability to metabolise 4-hydroxylation of diclofenac and 6- and 7-hydroxylation of (*S*)-warfarin (Williams et al. 2003).
The protonation states at pH 7.0 of all histidines, glutamic and aspartic acids were calculated by the program *MCCE* (Alexov and Gunner 1997), of the other titratable amino acids by the program xleap of the *AMBER 7.0* program (Case et al. 2002). The partial charges of (*S*)-warfarin 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* (Frisch et al. 1998) program (Florian Barth, personal communications; heme and warfarin parameters are given in the supplementary material). The oxyferryl state (Fe^{3+}=O) is considered to be the main reactive species of the CYP catalytic cycle. The all-atom *AMBER* force field "ff99" was used to represent the protein system. Bonded and non-bonded parameters for (*S*)-warfarin and the heme group were derived from the *AMBER* libraries. The *xleap* program was used to solvate the free enzyme and the CYP2C9-(*S*)-warfarin complex in a truncated octahedron box of TIP3P water with a minimal distance of 12.0 Å between the box boundary and the protein, and a closeness of 0.42 Å. One sodium ion was added to neutralize the system.

MD simulation:

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 simulated at a temperature of 300 K. The Shake algorithm (Ryckaert et al. 1977) 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 (Berendsen et al. 1984). 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. Each distribution was generated by a random number generator. A specific IG value was used as initial seed for the generator. 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.

Analysis of 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 the backbone atoms of each conformer to the initial structure. The RMSD between the substrate conformers generated by the 16

simulations of the enzyme-substrate complex were calculated using the *PROFIT* program (McLachlan 1982). Warfarin is bound to the B-C loop via hydrogen bonds in the crystal structure. Since this loop is one of the mobile regions of the protein the average structures were fitted in this region (AS 91-117) before RMSD calculation to exclude changes in the position of warfarin relatively to its position in the crystal structure because of movements of the B-C loop. The RMSD values were used to construct a distance-based tree with the program *NEIGHBOR* of the *PHYLIP* program package (Felsenstein 2004) applying the unweighted pair group method with arithmetic mean (UPGMA). Trees were visualized with *TreeView 1.6.1* (Page 1996). The protein structure was visualized using the *PyMOL 0.93* program (DeLano 2002). The *VOIDOO* (Kleywegt and Jones 1994) program was used to calculate the solvent accessible surface of the substrate binding cavity. In order to monitor the distance of potential sites of hydroxylation of the substrate to the active site, the distances between the oxygen bound to the heme iron and hydrogens bound to potential sites of hydroxylation of the substrate were measured.

Estimation of chemical reactivity of warfarin:
An estimation of chemical reactivity of warfarin was done by calculating the bond dissociation energy of all C-H bonds using the PETRA program (J. Gasteiger 1988).

3.1.4 Results

Stability of the simulations:
6 MD simulations (3 ns each) of CYP2C9 in water and 16 simulations (5 ns each) of CYP2C9 in complex with its substrate 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 warfarin complex, respectively, were used. To evaluate the deviation of the trajectory from the initial structure, the root mean square deviation (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, and 477-490 according to CYP2C9 wild-type sequence numbering) was stable during the simulations. After fitting the core region, the backbone RMSD from the crystal structure 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 (data not shown); after 1.5 ns the core region was equilibrated as the deviations between the conformers of one simulation were less than 1.5 Å, and the deviation between any two simulations was less than 2.4 Å (calculated from average structures of the last 100 ps of each simulation). In contrast, the deviation of all backbone atoms during the simulation was considerably larger: 2.0 and 2.5 Å and 2.0 - 3.0 Å deviation from the crystal structure for the free enzyme and the substrate complex, respectively. The deviation between any two simulations was up to 3.8 Å which indicates that there are highly mobile regions outside the core. Furthermore, the 2D-RMS plots demonstrated that these mobile regions of the protein are not yet in equilibrium. The backbone RMSD of the average structure of all simulations to the X-ray structure is 1.6 Å and 1.3 Å for the complex and the free enzyme, respectively, which is much less than the deviations between any two simulations or the deviation between each simulation and the crystal structure. Therefore the different simulations did not converge but scatter around the crystal structure.

We focused our analysis to those structure elements which are involved in the formation of the substrate binding cavity, since conformational changes in this region of the protein have been identified to change the shape of the active site containing cavity and lead to the formation of channels suitable for substrate access or product exit. From the 16 simulations of the enzyme-substrate complex 4 distinct states were identified, which differ in the shape of the substrate binding cavity and the position of the substrate: the *crystal structure state*, the *transfer state*, the *6-7-hydroxylation state,* and the *nearby 4-hydroxylation state* (Fig. 1). We observed transition from *crystal structure state* to the other three states and from the *nearby 4-hydroxylation state* back to the *crystal structure state*. The substrate binding cavity is formed mainly by helices F and G and the loop between them (F-G loop) as well as the loop between the B and C helices (B-C loop), which contains the B´ helix. These structural elements belong to the mobile regions of the protein, as the RMSD values for the B-C loop deviated at the end of the simulations of the free enzyme and substrate complex 1.8- 2.9 Å and 1.2- 3.5 Å from the crystal structure after fitting in the core region, respectively. For the F-G loop values between 2.0 Å and 4.5 Å for the enzyme-substrate complex and 2.0-4.0 Å for the free enzyme were measured. 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. In contrast, the heme pocket and a narrow funnel (Ala297, Leu362, Leu366, Thr301) leading from the substrate binding cavity to the heme belong to the core region and were structurally similar in all states.

Four states:

The four states were identified by analysing the relative positions of substrate in the binding cavity. RMSD values of warfarin between any two simulations (averaged over 100 ps) and to the crystal structure were calculated and used to construct a distance tree (Fig. 2). The c*rystal structure state* was observed during 12 out of 16 simulations. The *transfer state* was observed in one simulation where the shape of the substrate binding site cavity is massively changing: a channel from the protein surface to the binding cavity (Fig. 3a) is generated by a shift of the B-C loop towards the I-helix and a slight turn of the β_1 sheet towards the N-terminus (Fig. 4a). Furthermore, the side chains of Phe100 and Lys72 moved out of the region between B´B loop and β_1 sheet and thereby enlarge the channel. The substrate warfarin binds to this channel (Fig. 4b). To examine if the presence of the substrate in this region facilitates the formation of the channel, the distance between B-B´ loop (C_α of Ile99) and the adjacent β_1 sheet (C_α of Pro73) was measured during all simulations. In the *transfer state* this distance increased to up to 19 Å as compared to 12.5 Å and 12.3 Å in the two crystal structures. In all other simulations, the distance varied between 11 Å and 15 Å. Thus, the substrate seems to stabilize the channel.

The *6-7-hydroxylation state* was observed in one simulation. In this state the substrate points toward heme with its 6- and 7-position near to the active heme oxygen (Fig. 1d). The distance between the hydrogen bound to carbon 7 of warfarin (Fig. 5) and the active heme oxygen decreased to 2-3 Å, while in the other simulations this distance was between 8 and 14 Å (Fig. 6). In this state warfarin is stabilized by van der Waals contacts to Phe114, Ala297, Thr301, Leu362, Leu366, and Phe476. Remarkably the 7-position is 3 times more frequently at a distance of less than 3 Å from the active oxygen than the 6-position (Fig. 7a). For a distance less than 2.5 Å this ratio even increases to 1:8. Analysing the time course of these distances indicates that the substrate performs a diffusional motion in a single energy minimum rather than jumping between two distinct, equilibrated substates (Fig.7b). Four different regions can be assigned: both distances less than 3 Å (area 1), one of the distances less than 3 Å (area 2 and 3), or both distances larger than 3 Å (area 4). A comparison of area 2 and area 3 reveals that the 7-hydroxylation site is considerably more frequent in a distance at which hydroxylation can occur than the 6-hydroxylation site. This clear geometrical preference is caused by the well-defined position and orientation of the substrate in the narrow funnel and the binding cavity. Therefore the regioselectivity is a consequence of the overall shape of the protein rather than caused by local interaction near the heme oxygen.

In this state an additional channel from the bulk solvent to the binding cavity is observed (Fig. 3b) which is different from the channel observed in the *transfer state*. It is located between helices F and I and the turn in the C-terminal antiparallel β-sheet and could already be seen as a narrow solvent channel in the crystal structure, just wide enough for water molecules to enter. The increase of the channel is mainly caused by backbone and side chain movements of Ile205 and Ser209 (mobile F helix) as well as side chain movements of Glu300 (stable I helix). The distance between C_α atoms of Ile205 (mobile F helix) and Glu300 increased to 14 Å as compared to 7.6 Å in the crystal structure. In all other simulations this channel slightly opened with a distance between 9 and 11 Å.

The *nearby 4-hydroxylation state* of the substrate enzyme complex was observed during 2 simulations. In one of these simulations the system even moved back to the *crystal structure state*. In *nearby 4-hydroxylation state* the warfarin orients its phenyl group which contains the 4-position toward the active heme oxygen. The measured minimum distance of 7 Å between the 4-position and the active heme oxygen is considerably less than in the other states (11-17 Å), but not close enough to enable hydroxylation (Fig.8).

The *crystal structure state* is as a stable conformation which itself is not active but has been suggested to play an important role activation of other substrates by warfarin. However the positions of warfarin in the *6-7-hydroxylation state* or the *nearby 4-hydroxylation state* are overlapping to its position in the *crystal structure state* (Fig.8). Thus two of these states can not simultaneously be occupied by warfarin.

3.1.5 Discussion

CYP2C9 channels:

How do hydrophobic substrates enter the binding cavity, and how do more hydrophilic products leave it? In the X-ray structures of free and complexed human CYP2C9, the binding cavity was permanently connected to the bulk solvent by a narrow solvent channel between helices F and I and the turn in the C-terminal antiparallel β-sheet (Williams et al. 2003), which however is too narrow to allow a substrate to pass through. Therefore, we analysed the effect of substrate binding to the occurrence of channels between the binding pocket of CYP2C9 and the bulk solvent by multiple molecular dynamics simulations. In the free enzyme narrow channels between cavity and bulk solvent were formed temporarily mainly by fluctuations of the FG and BC loops. In the presence of substrate the shape of the binding

cavity changed. We observed four distinct positions of the substrate and thus four distinct shapes of the binding cavity. In two of them, the *transfer state* and the *6-7-hydroxylation state*, stable channels are formed wide enough for substrate or product.

For several family 2 CYPs, FG and BC loops have been identified by antibody binding data and hydrophobicity analysis to interact with the membrane of the endoplasmatic reticulum (ER) (Williams et al. 2000), from where the hydrophobic substrates are assumed to access the enzyme preventing contact to the bulk solvent. Fluctuations of this putative substrate access channel were observed for the free enzyme and the enzyme-substrate complex during our simulations. A recent MD simulation study of the free CYP2C9 enzyme also revealed fluctuations in this region of the protein, however, the existence of a channel with suitable size for substrate access was not detected (Afzelius et al. 2004). However our MD simulations indicate that these narrow channels are stabilized and widened in the presence of substrate. The stabilisation of an open conformation via a second molecule was also observed for mammalian CYP2B4. In absence of an inhibitor the crystal structure reveals a large open cleft, which is primarily formed by helices B´ to C and F to G (Scott et al. 2003). In this case, the wide open structure of the substrate-free CYP2B4 is stabilized in the crystal by dimerization, where two protein molecules stabilize each other through hydrophobic interactions. Thus, in X-ray structures as well as in our MD simulations the channel and the binding cavity have a high plasticity and are stabilized by molecular interactions. This plasticity is consistent with the wide range of substrates which are accepted by CYP2C9 and other CYPs (Rendic and Di Carlo 1997). The induction of a channel between F-G loop, B´ helix/B-B´ loop and β_1 sheet has also be found previously by random expulsion molecular dynamics simulations of bacterial CYP101, CYP102A1, CYP107A1, and mammalian CYP2C5, where a bulky substrate was pushed out of the binding cavity by an external random force (Winn et al. 2002; Wade 2004).

A solvent channel in the region between helices F and I and the turn in the C-terminal antiparallel β-sheet is visible in the free and complexed crystal structure of CYP2C9 as well as in several bacterial and mammalian CYPs (Wade 2004). Because this channel directly leads from the heme group into the bulk solvent it would be favourable for the hydrophilic product to leave the binding cavity via this pathway. However, in the X-ray structure this channel is too narrow for products to pass through. Binding of the substrate increased its size considerably: in the *6-7-hydroxylation state* the channel became wide enough to allow bulky product to pass through. Three different channels have been described for mammalian CYP2C5 (Schleinkofer et al. 2005). In addition to the two channels which we have observed

in our study, standard MD and REMD simulations revealed an additional channel between B´ helix/B-C loop, G and I helices. Also for a completely different enzyme class, the acetylcholinesterases (Gilson et al. 1994; Bartolucci et al. 1999), two different pathways have been suggested to be the structural basis of their extremely high catalytic activity.

Structural basis of the regioselectivity of CYP2C9:

A calculation of bond dissociation energy (J. Gasteiger 1988) of C-H groups in warfarin predicted that the reactivity of more than ten positions would be sufficiently high for hydroxylation to occur. However, CYP-catalyzed hydroxylation of warfarin is highly regioselective and depends on the shape of the substrate binding cavity. While CYP2C9 catalyses hydroxylation at the 7, 6, and 4 position with a product ratio of 71 : 22 : 7, respectively (Rettie et al. 1992), CYP3A4 hydroxylates warfarin at position 4 and 10 (Ngui et al. 2001). However, the experimentally observed regioselectivity of CYP2C9 cannot be predicted from the X-ray structure of its complex with warfarin, because warfarin binds at a hydrogen - oxygen distance of 10 Å from the heme iron (Williams et al. 2003). Since a contact between the substrate hydrogen and the oxygen of the oxyferryl heme (compound I) is prerequisite to hydroxylation (Sason Shaik 2005), we considered only conformations as reactive if their H···O distance was less than 3 Å, the distance of a C-H···O hydrogen bond (Steiner and Desiraju 1998). While the conformation of the crystal structure was therefore considered to be non-productive, in one simulation a transition to the reactive *6-7-hydroxylation state* was observed with close contacts between the oxyferryl and the hydrogens at positions 6 and 7. Since the 7-hydroxylation site is much more frequent within a distance of 3 Å as compared to the 6-hydroxylation site we would predict a marked preference of the 7- hydroxylation over the 6- hydroxylation, which is consistent with the experimentally measured ratio of 75 : 25 for 7- to 6-hydroxywarfarin. Another consequence of the prediction of a *6-7-hydroxylation state* is a prediction of residues which mediate the regioselectivity. Phe476 has been shown to change the regioselectivity of diclofenac hydroxylation (Melet et al. 2003). It is located in the binding cavity and mediates the orientation of warfarin in the *6-7-hydroxylation state* despite its distance of more than 14 Å from the heme. Most relevant for regioselectivity, however, are the four residues which form a narrow, rigid funnel leading from the binding cavity to the heme. Thus we would predict that mutation of Ala297, Thr301, Leu362, and Leu366 should have a drastic effect on regioselectivity. This selectivity-determining role of a narrow funnel which limits access of substrate to the heme is consistent with the observation that 4-hydroxylation of warfarin is rare. The 6 and 7 sites are positioned

at the far end of the base of the Y-shaped warfarin molecule, while the 4 site is at its more bulky end. Due to the limited access to the heme, the 4 position is hindered, while the 6 and 7 positions reach down the narrow funnel (Fig. 1d). Therefore, replacing Leu362 or Leu366 by less bulky residues is expected to facilitate formation of 4-hydroxywarfarin. The existence of a productive substrate conformation at a distance of less than 3 Å from the heme has recently be confirmed in a crystal structure of flurbiprofen in complex with CYP2C9 (Wester et al. 2004). In contrast to the more bulky warfarin, this rod-shaped molecule fits well into the narrow funnel and contacts the heme similar to the *6-7-hydroxylation state*. The superimposition of an average structure of the *6-7-hydroxylation state* and 1R9O reveals that both substrates bind to the rigid, regioselectivity-determining funnel in a rather similar orientation with their respective hydroxylation sites deviating by only 1.4 Å. The funnel forming amino acids only deviate about 1.1 Å (all atom) from the 1R9O structure. While the two complexes are highly similar in the funnel region, they deviate considerably elsewhere. While these conformational differences may account for differences in catalytic activity, the regioselectivity can be easily explained by the existence of a rigid, narrow funnel.

Role of the binding site for warfarin in the crystal structure in enzyme kinetics such as activation and auto activation:

The large substrate binding cavity of CYP2C9 (470 $Å^3$) can accommodate more than one substrate molecule simultaneously. NMR titration and T_1 relaxation time measurements indicated that dapsone and flurbiprofen are simultaneously occupying the substrate binding cavity of CYP2C9 (Hummel et al. 2004). It has been suggested that in addition to the warfarin molecule in the crystal structure, a second warfarin molecule could be accommodated in the binding cavity of CYP2C9 (Williams et al. 2003). This is, however, in contrast to the experimental observation that auto activation is not observed, while binding of warfarin activates hydroxylation of 7-methoxy-4-trifluoromethylcoumarin (Stresser et al.). According to our simulation results simultaneous binding of warfarin to the *crystal structure state* and to the active *6-7-hydroxylation state* is sterically hindered, while the warfarin complex in the *crystal structure state* would still leave sufficient space for other substrate molecules like flurbiprofen to bind close to the active heme oxygen.

3.1.6 Conclusion

How can human CYPs manage to have simultaneously a broad substrate profile and a high regioselectivity? The substrate binding sites of CYPs consist of two elements: a large, mobile binding cavity connected to the protein surface via substrate-induced channels, and a narrow, rigid, and hydrophobic funnel connecting the binding cavity and the active-site heme. Substrate specificity and regioselectivity is a result of a delicate balance between different states of the substrate-enzyme complex. The dynamics of the substrate binding site is crucial for understanding the biochemical properties of the enzyme and for predicting the effect of mutations. Substrate recognition in the binding cavity is by no means static, but the result of adaptive motion of the enzyme. In contrast, regioselectivity is determined by the shape of the substrate and its highly restricted orientation by a narrow, rigid funnel between binding cavity and the active-site heme group.

Acknowledgements

The authors would like to thank Michael Krahn for the calculation and visualisation of the substrate binding site cavity with the program VOIDOO and Florian Barth for the parameterisation of the oxyferryl heme group. This work has been supported by the Bundesministerium für Bildung und Forschung (BMBF).

3.1.7 References

Afzelius L., Raubacher F., Karlen A., Jorgensen F. S., Andersson T. B., Masimirembwa C. M. and Zamora I. (2004). "Structural analysis of CYP2C9 and CYP2C5 and an evaluation of commonly used molecular modeling techniques." *Drug Metab Dispos* 32(11): 1218-29.

Alexov E. G. and Gunner M. R. (1997). "Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties." *Biophys J* 72(5): 2075-93.

Bartolucci C., Perola E., Cellai L., Brufani M. and Lamba D. (1999). ""Back door" opening implied by the crystal structure of a carbamoylated acetylcholinesterase." *Biochemistry* 38(18): 5714-9.

Berendsen H. J. C., Postma J. P. M., Gunsteren W. F. v., DiNola A. and Haak J. R. (1984). "Molecular dynamics with coupling to an external bath." *J Chem Phys* 81: 3684-3690.

Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I. N. and Bourne P. E. (2000). "The Protein Data Bank." *Nucleic Acids Res* 28(1): 235-242.

Case D. A., Pearlman D. A., Caldwell J. W., III T. E. C., Wang J., Ross W. S., Simmerling C. L., Darden T. A., Merz K. M., Stanton R. V., Cheng A. L., Vincent J. J., Crowley M., Tsui V., Gohlke H., Radmer R. J., Duan Y., Pitera J., Massova I., Seibel G. L., Singh U. C., Weiner P. K. and Kollman P. A. (2002). AMBER 7, University of California, San Francisco.

DeLano W. L. (2002). The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA, USA.

Felsenstein J. (2004). "PHYLIP (Phylogeny Inference Package) version 3.6." *Distributed by the author Department of Genome Sciences, University of Washington, Seattle*.

Frisch M. J., Trucks G. W., Schlegel H. B., Scuseria G. E., Robb M. A., Cheeseman J. R., Jr J. A. M., Vreven T., Kudin K. N., Burant J. C., Millam J. M., Iyengar S. S., Tomasi J., Barone V., Mennucci B., Cossi M., Scalmani G., Rega N., Petersson G. A., Nakatsuji H., Hada M., Ehara M., Toyota K., Fukuda R., Hasegawa J., Ishida M., Nakajima T., Honda Y., Kitao O., Nakai H., Klene M., Li X., Knox J. E., Hratchian H. P., Cross J. B., Adamo C., Jaramillo J., Gomperts R., Stratmann R. E., Yazyev O., Austin A. J., Cammi R., Pomelli C., Ochterski J. W., Ayala P. Y., Morokuma K., Voth G. A., Salvador P., Dannenberg J. J., Zakrzewski V. G., Dapprich S., Daniels A. D., Strain M. C., Farkas O., Malick D. K., Rabuck A. D., Raghavachari K., Foresman J. B., Ortiz J. V., Cui Q., Baboul A. G., Clifford S., Cioslowski J., Stefanov B. B., Liu G., Liashenko A., Piskorz P., Komaromi I., Martin R. L., Fox D. J., Keith T., Al-Laham M. A., Peng C. Y., Nanayakkara A., Challacombe M., Gill P. M. W., Johnson B., Chen W., Wong M. W., Gonzalez C. and Pople J. A. (1998). GAUSSIAN 98 (Revision A.7), Gaussian, Inc., Pittsburgh PA.

Gilson M. K., Straatsma T. P., McCammon J. A., Ripoll D. R., Faerman C. H., Axelsen P. H., Silman I. and Sussman J. L. (1994). "Open "back door" in a molecular dynamics simulation of acetylcholinesterase." *Science* 263(5151): 1276-8.

Hemeryck A., De Vriendt C. and Belpaire F. M. (1999). "Inhibition of CYP2C9 by selective serotonin reuptake inhibitors: in vitro studies with tolbutamide and (S)-warfarin using human liver microsomes." *Eur J Clin Pharmacol* 54(12): 947-51.

Hlavica P. and Lewis D. F. (2001). "Allosteric phenomena in cytochrome P450-catalyzed monooxygenations." *Eur J Biochem* 268(18): 4817-32.

Hummel M. A., Gannett P. M., Aguilar J. S. and Tracy T. S. (2004). "Effector-mediated alteration of substrate orientation in cytochrome P450 2C9." *Biochemistry* 43(22): 7207-14.

Hutzler J. M., Hauer M. J. and Tracy T. S. (2001). "Dapsone activation of CYP2C9-mediated metabolism: evidence for activation of multiple substrates and a two-site model." *Drug Metab Dispos* 29(7): 1029-34.

J. Gasteiger E. G. J., M. G. Hicks, J. Sunkel (1988). "Empirical Methods for the Calculation of Physicochemical Data of Organic Compounds." *Physical Property Prediction in Organic Chemistry Springer Verlag, Heidelberg*: 119-138.

Kleywegt G. J. and Jones T. A. (1994). "Detection, delineation, measurement and display of cavities in macromolecular structures." *Acta Crystallogr D Biol Crystallogr* 50(Pt 2): 178-85.

McLachlan A. D. (1982). "Rapid Comparison of Protein Structures." *Acta Cryst A* 38: 871-873.

Melet A., Assrir N., Jean P., Pilar Lopez-Garcia M., Marques-Soares C., Jaouen M., Dansette P. M., Sari M. A. and Mansuy D. (2003). "Substrate selectivity of human cytochrome P450 2C9: importance of residues 476, 365, and 114 in recognition of diclofenac and sulfaphenazole and in mechanism-based inactivation by tienilic acid." *Arch Biochem Biophys* 409(1): 80-91.

Ngui J. S., Chen Q., Shou M., Wang R. W., Stearns R. A., Baillie T. A. and Tang W. (2001). "In vitro stimulation of warfarin metabolism by quinidine: increases in the formation of 4'- and 10-hydroxywarfarin." *Drug Metab Dispos* 29(6): 877-86.

Page R. D. (1996). "TreeView: an application to display phylogenetic trees on personal computers." *Comput Appl Biosci* 12(4): 357-8.

Rendic S. and Di Carlo F. J. (1997). "Human cytochrome P450 enzymes: a status report summarizing their reactions, substrates, inducers, and inhibitors." *Drug Metab Rev* 29(1-2): 413-580.

Rettie A. E., Korzekwa K. R., Kunze K. L., Lawrence R. F., Eddy A. C., Aoyama T., Gelboin H. V., Gonzalez F. J. and Trager W. F. (1992). "Hydroxylation of warfarin by human cDNA-expressed cytochrome P-450: a role for P-4502C9 in the etiology of (S)-warfarin-drug interactions." *Chem Res Toxicol* 5(1): 54-9.

Ryckaert J. P., Ciccotti G. and Berendsen H. J. C. (1977). "Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes." *Computat Phys* 23: 327-341.

Sason Shaik D. K., Samue¨l P. de Visser, Ahmet Altun,| and Walter Thiel (2005). "Theoretical Perspective on the Structure and Mechanism of Cytochrome P450 Enzymes." *American Chemical Society*.

Schleinkofer K., Sudarko, Winn P. J., Ludemann S. K. and Wade R. C. (2005). "Do mammalian cytochrome P450s show multiple ligand access pathways and ligand channelling?" *EMBO Rep* 6(6): 584-9.

Scott E. E., He Y. A., Wester M. R., White M. A., Chin C. C., Halpert J. R., Johnson E. F. and Stout C. D. (2003). "An open conformation of mammalian cytochrome P450 2B4 at 1.6-A resolution." *Proc Natl Acad Sci U S A* 100(23): 13196-201.

Scott E. E., He Y. A., White M. A., Halpert J. R., Johnson E. F. and Stout C. D. (2004). "Structure of Mammalian Cytochrome P450 2B4 Complexed with 4-(4-chlorophenyl)imidazole at 1.9 Å Resolution: Insight into the Range of P450 Conformations and Coordination of Redox Partner Binding." *J Biol Chem* 279(26): 27294-27301.

Smith D. A., Abel S. M., Hyland R. and Jones B. C. (1998). "Human cytochrome P450s: selectivity and measurement in vivo." *Xenobiotica* 28(12): 1095-128.

Smith D. A., Ackland M. J. and Jones B. C. (1997). "Properties of cytochrome P450 isoenzymes and their substrates Part 1: active site characteristics." *Drug Discov Today* 2(10): 406-414.

Steiner T. and Desiraju G. R. (1998). "Distinction between the weak hydrogen bond and the van der Waals interaction." *Chem Commun*(8): 891-892.

Stresser D. M., Turner S. D. Ackermann J. M., Miller V. P. and Crespi C. L. "Flourometric cytochromes P450 2C8, 2C9 and 2C19 inhibition assays" (http://www.bdbiosciences.com/discovery_labware/gentest/products/pdf/post_015.pdf)

Wade C. W. W., P.J. Schlichtling, I. Sudarko, (2004). "A survey of active site access channels in cytochromes P450." *J Inorg Biochem* 98: 1175-1182.

Wester M. R., Yano J. K., Schoch G. A., Yang C., Griffin K. J., Stout C. D. and Johnson E. F. (2004). "The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-A resolution." *J Biol Chem* 279(34): 35630-7.

Williams P. A., Cosme J., Sridhar V., Johnson E. F. and McRee D. E. (2000). "Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity." *Mol Cell* 5(1): 121-31.

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

Winn P. J., Ludemann S. K., Gauges R., Lounnas V. and Wade R. C. (2002). "Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine." *Proc Natl Acad Sci U S A* 99(8): 5361-6.

Figures

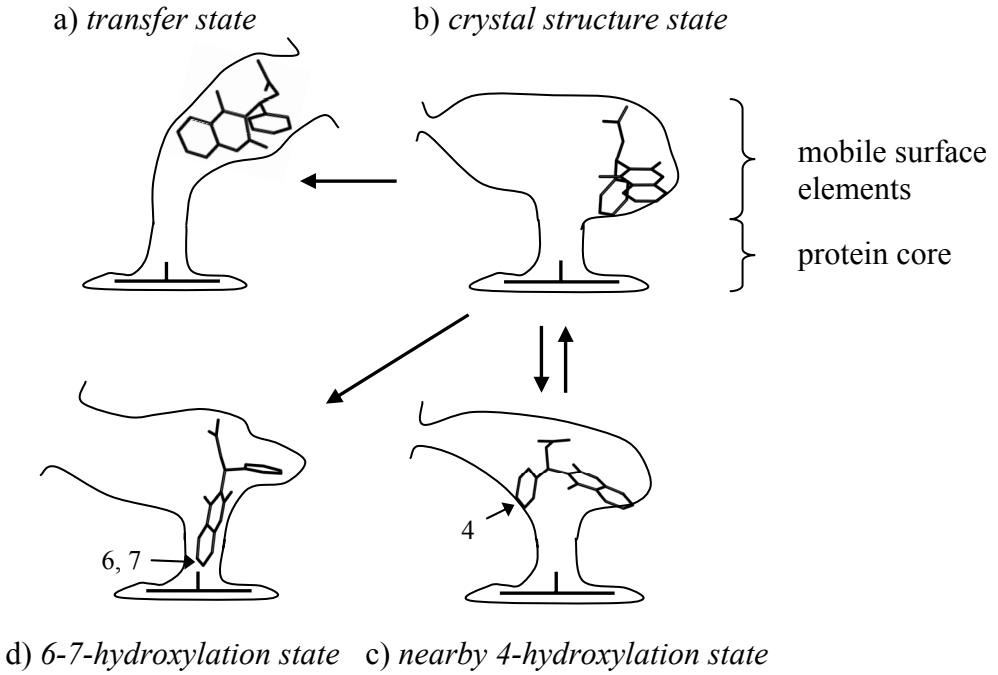


Fig. 1 Shape of the substrate binding cavity of the CYP2C9 warfarin complex in the a) *transfer state,* b) *crystal structure state,* c) *nearby 4-hydroxylation state,* and d) *6-7-hydroxylation state*. Arrows indicate the transitions observed by MD simulations.

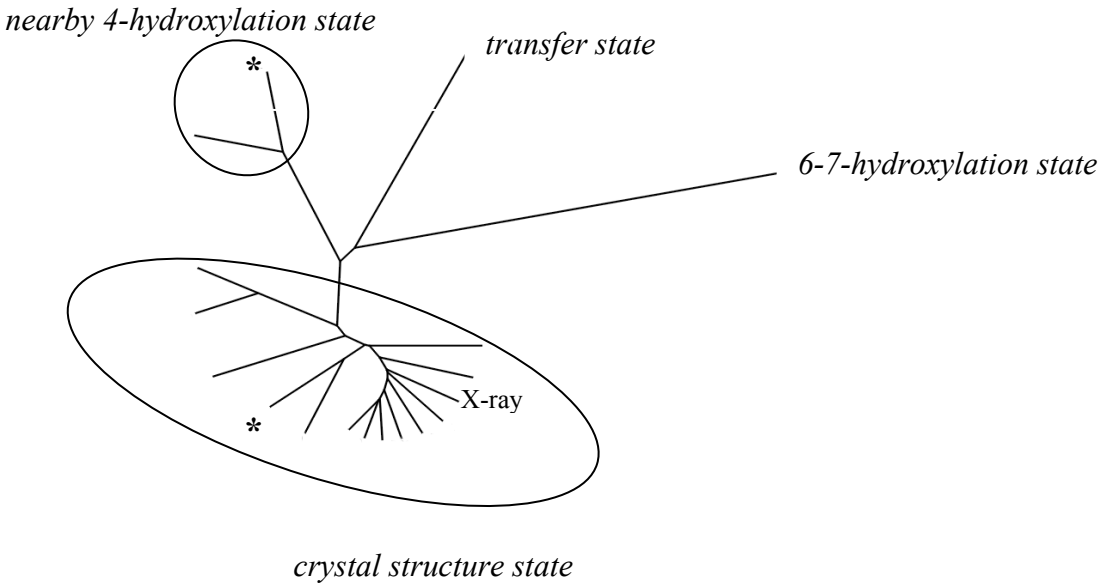


Fig. 2 Tree of warfarin conformers based on RMSD between all atoms except hydrogen observed during multiple MD simulations. "X-ray" indicates the crystal structure, which was the starting point of all simulations with substrate. The conformers marked with an asterisk were observed during one simulation, where the substrate first moved from its position in the *crystal structure state* to the *nearby 4-hydroxylation state* and then back to *crystal structure state*.

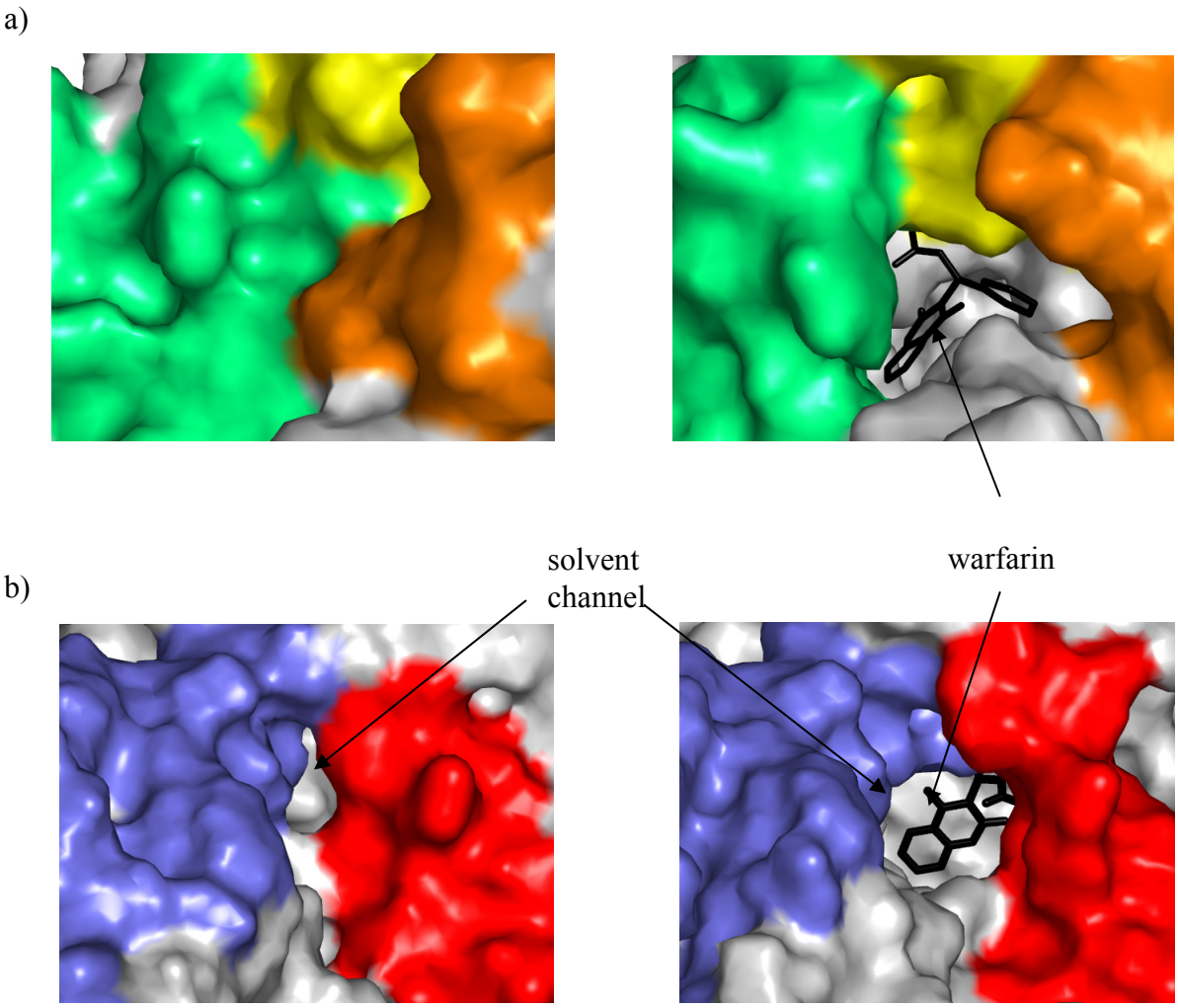


Fig. 3 Surface of CYP2C9 a) in the region between F-G loop (yellow), B´ helix/B-B´ loop (green), and β_1 sheet (orange): crystal structure (left), *transfer state* (right), and b) in the region of the solvent channel between the F helix (red) and the turn in the C-terminal antiparallel β-sheet (blue) : crystal structure (left), *6-7-hydroxylation state* (right).

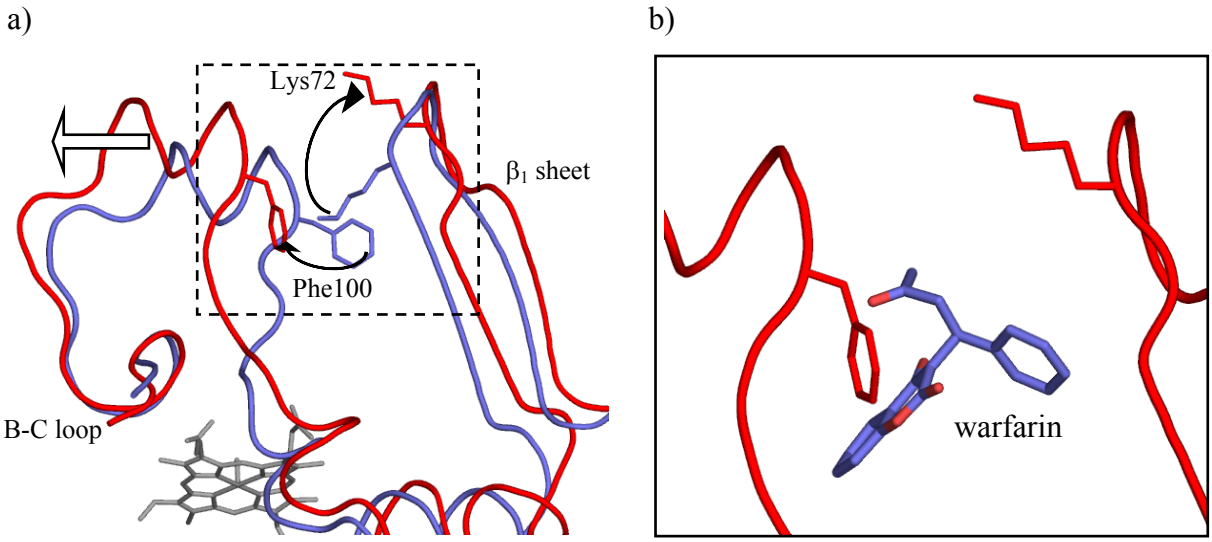


Fig. 4 a) A more open conformation was observed in the *transfer state* (red) compared to the *crystal structure* (blue). The open conformation was achieved due to movements of the B-C loop, β_1 sheet and the side chains of Phe100 and Lys72. Arrows indicate the direction of movements of the different structure elements. b) The opening was stabilized by the substrate warfarin.

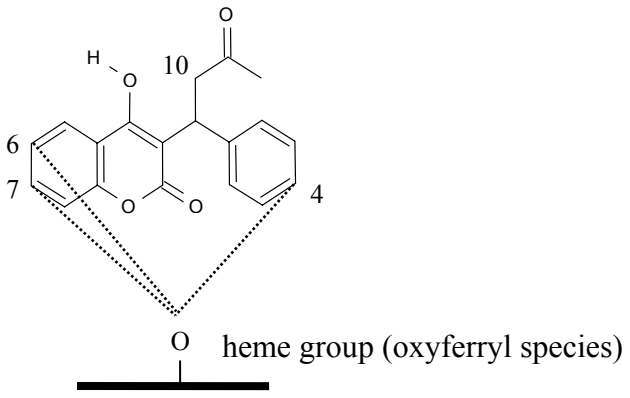


Fig. 5 Measured distances between warfarin and the active heme oxygen of CYP2C9; the 4-, 6-, 7-, and 10-hydroxylation sites of warfarin are labelled.

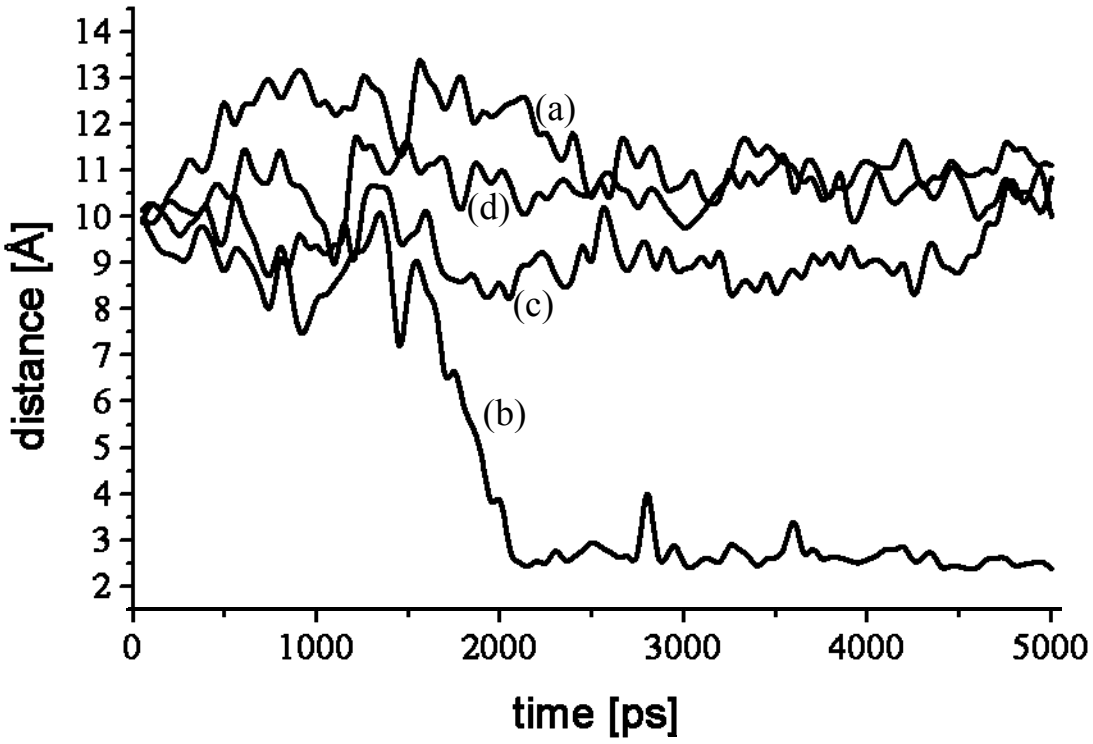
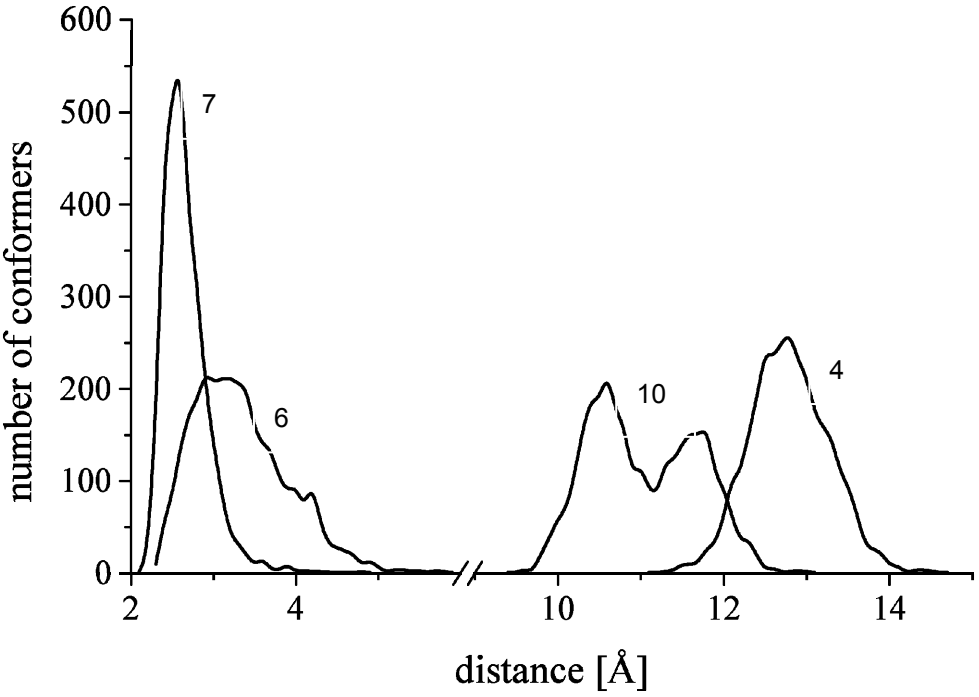


Fig. 6 Distance between the hydrogen atom bound to C7 of warfarin and the active oxygen of the heme group during the simulations of CYP2C9 with warfarin in water for representative trajectories in the *crystal structure state* (a), *6-7-hydroxylation state* (b), *transfer state* (c), and *nearby 4-hydroxylation state* (d).

a)



b)

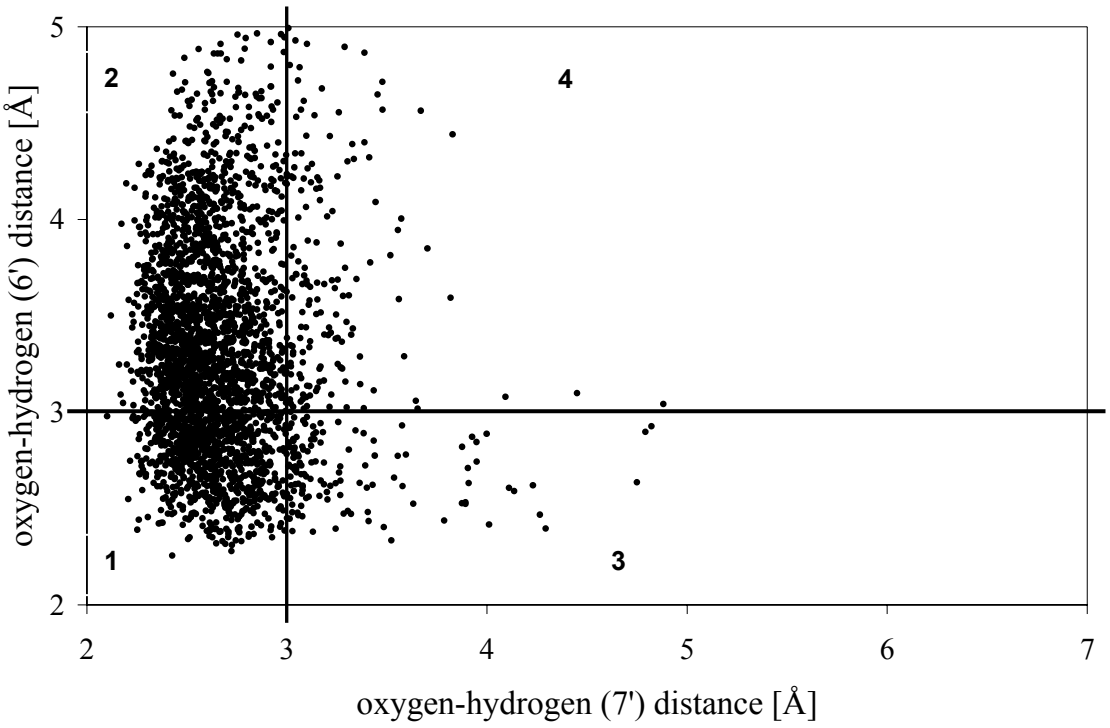


Fig. 7 a) Distribution of measured distances between 4-, 6-, 7-, and 10-hydroxylation site of warfarin, respectively, and the active oxygen in the *6-7-hydroxylation state* of the protein substrate complex. In contrast to position 6 and 7, the 4- and 10-hydroxylation site of warfarin do not contact the active heme oxygen. b) Two dimensional plot of distances between 6- and 7-hydroxylation site and the active heme oxygen. 1) both distances <3 Å; 2) distance FeO--HC7< 3 Å; FeO--HC6>3 Å; 3) distance FeO--HC7> 3 Å; FeO--HC6<3 Å; 4) both distances >3 Å.

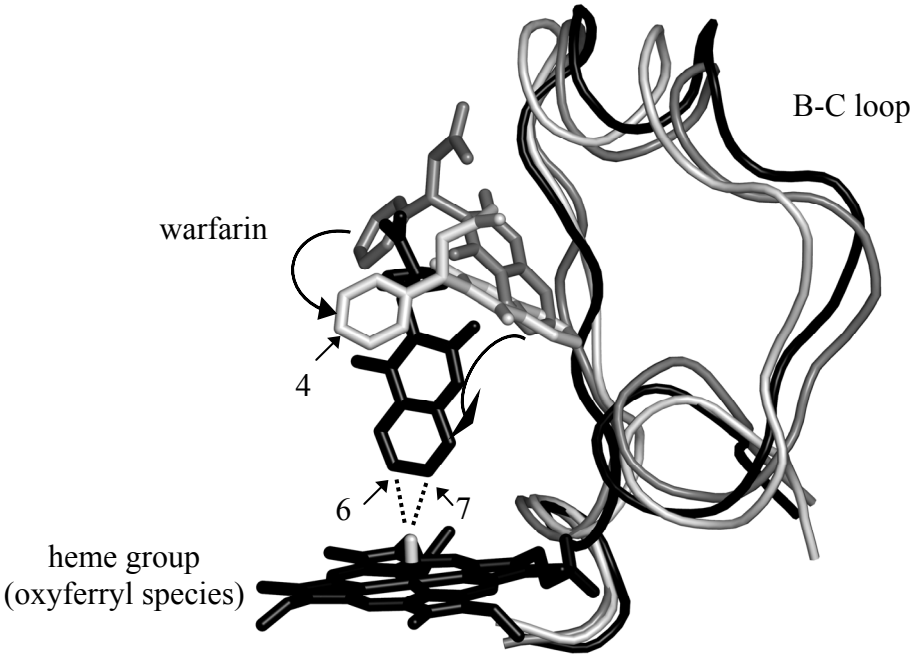
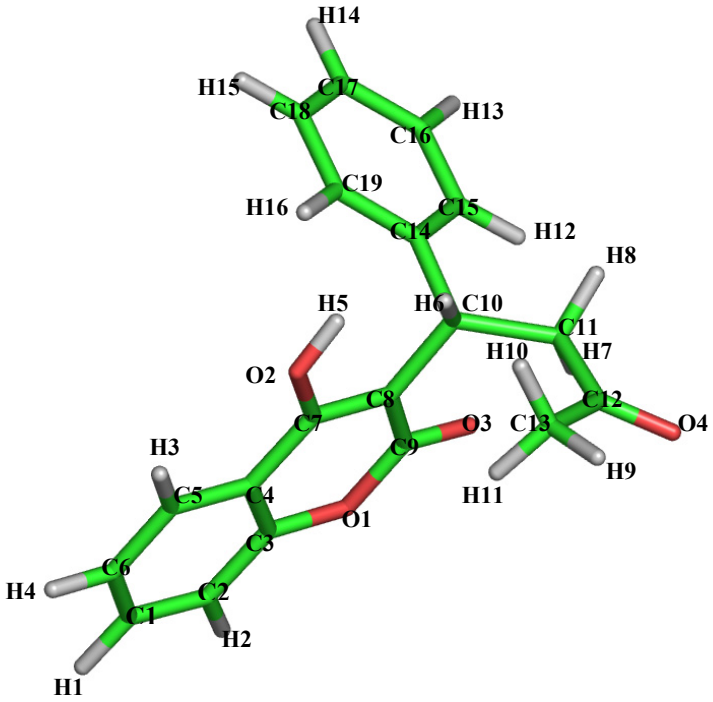


Fig. 8 Comparison of the warfarin conformers observed in *crystal structure state* (grey), *6-7-hydroxylation state* (black), and *nearby 4-hydroxylation state* (white). All conformers are partially overlapping. Arrows mark the 4-, 6-, and 7-hydroxylation site of warfarin.

3.1.8 Supplementary material

(*S*)-warfarin force field parameters:



(*S*)-warfarin atom charges:

atom name	atom type	charge
O3	(O):	-0.585
C9	(C):	0.765
O1	(OS):	-0.423
C3	(CA):	0.422
C2	(CA):	-0.319
H2	(HA):	0.189
C1	(CA):	-0.047
H1	(HA):	0.147
C6	(CA):	-0.271
H4	(HA):	0.166

atom name	atom type	charge
C5	(CA):	-0.026
H3	(HA):	0.144
C4	(CA):	-0.181
C7	(C):	0.361
O2	(OH):	-0.516
H5	(HO):	0.365
C8	(C):	-0.317
C10	(CT):	0.042
C11	(CT):	0.009
C12	(C):	0.675

atom name	atom type	charge
C13	(CT):	-0.600
H9	(HC):	0.164
H10	(HC):	0.164
H11	(HC):	0.164
O4	(O):	-0.536
H7	(HC):	-0.002
H8	(HC):	-0.002
H6	(HC):	0.014
C14	(CA):	0.091
C15	(CA):	-0.120

atom name	atom type	charge
H12	(HA):	0.118
C16	(CA):	-0.193
H13	(HA):	0.152
C17	(CA):	-0.105
H14	(HA):	0.135
C18	(CA):	-0.193
H15	(HA):	0.152
C19	(CA):	-0.120
H16	(HA):	0.118

Force field modifications for (*S*)-warfarin:

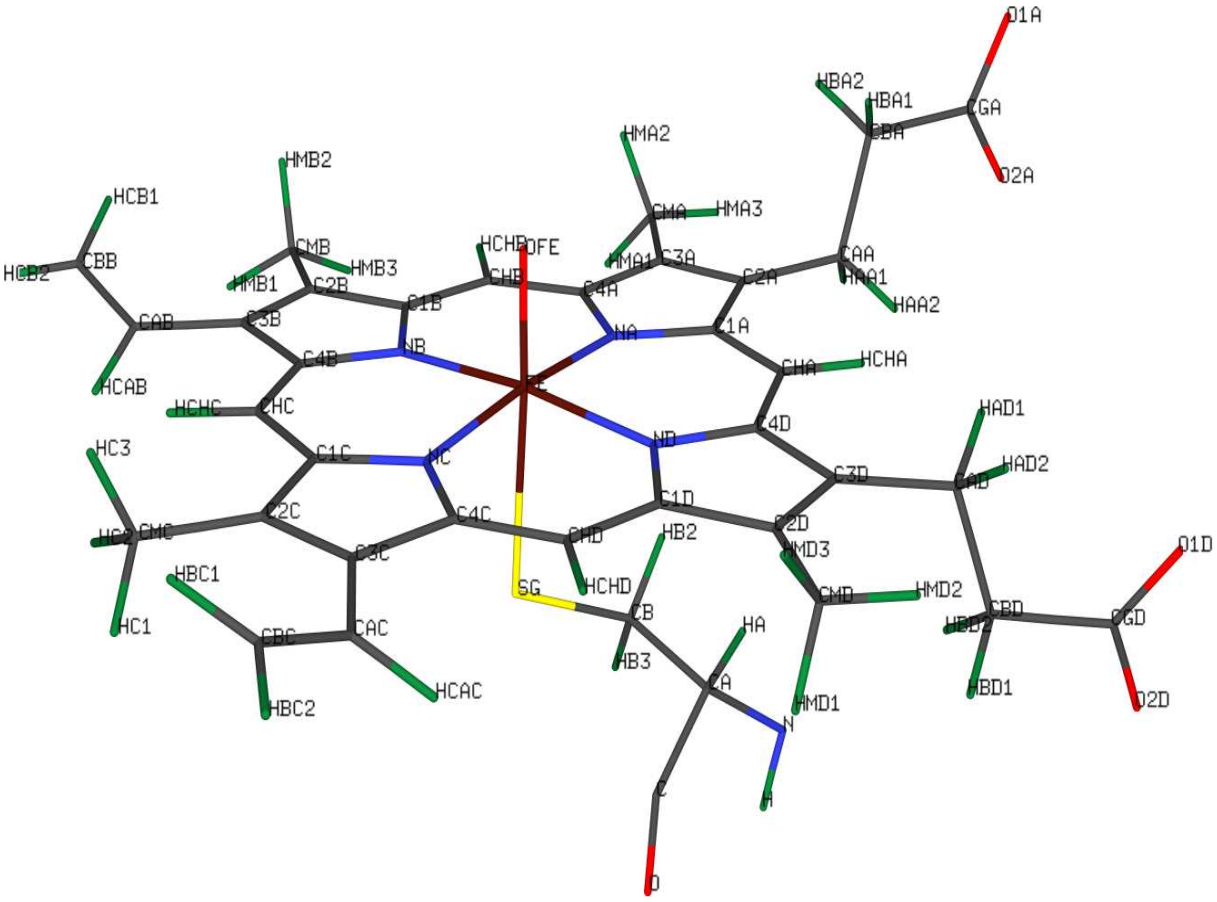
Bond	k_l [kcal/(mol · Å^2)]	l_0 [Å]
OS-CA	450.000	1.345

Angle	kθ [kcal/(mol · rad²)]	θ_0[rad]
C-CT-CA	63.000	114.250
C-C-CT	70.000	121.900
CA-C-C	63.000	121.950
OS-C-C	80.000	118.240
OS-CA-CA	80.000	120.560

C-OS-CA	60.000	124.000
C-C-C	63.000	117.420
C-C-CT	70.000	121.840

DIHEDRAL	IDIVF	PK [kcal/mol]	PHASE	PN
X -OS-CA-X	2	5.40	180.0000	2.0000

Heme force field parameters:



Heme atom charges:

atom name	atom type	charge
N	N	-0.416
H	H	0.272
CA	CT	0.043
HA	H1	0.077
CB	CT	0.113
HB2	H1	0.041
HB3	H1	0.041
SG	SH	-0.218
FE	FE	0.749
OFE	OF	-0.615
NA	N6	-0.079

C1A	C8	0.284
C2A	C1	-0.397
CAA	CT	0.320
HAA1	HC	-0.029
HAA2	HC	-0.029
CBA	CT	-0.230
HBA1	HC	0.037
HBA2	HC	0.037
CGA	C	0.851
O1A	O2	-0.830
O2A	O2	-0.830
C3A	C2	0.285
CMA	CT	-0.542
HMA1	HC	0.148
HMA2	HC	0.148
HMA3	HC	0.148
C4A	C9	-0.079
CHB	CA	-0.170
HCHB	HA	0.151
C1B	C9	-0.069
NB	N9	0.031
C2B	C3	0.141
CMB	CT	-0.312
HMB1	HC	0.087
HMB2	HC	0.087
HMB3	HC	0.087
C3B	C4	-0.044
CAB	CD	-0.058
HCAB	H4	0.118
CBB	CM	-0.491
HCB1	HA	0.186
HCB2	HA	0.186
C4B	C8	-0.114
CHC	CA	-0.089
HCHC	HA	0.156
C1C	C8	-0.184
NC	N8	-0.056
C2C	C3	0.202
CMC	CT	-0.294
HC1	HC	0.081
HC2	HC	0.081
HC3	HC	0.081
C3C	C4	-0.107
CAC	CD	-0.048
HCAC	H4	0.121
CBC	CM	-0.510
HBC1	HA	0.191
HBC2	HA	0.191
C4C	C9	0.091
CHD	CA	-0.266
HCHD	HA	0.166
C1D	C9	0.033
ND	N7	-0.160

C2D	C2	0.198
CMD	CT	-0.440
HMD1	HC	0.124
HMD2	HC	0.124
HMD3	HC	0.124
C3D	C1	-0.295
C4D	C8	0.290
CHA	CA	-0.528
HCHA	HA	0.322
CAD	CT	0.214
HAD1	HC	0.003
HAD2	HC	0.003
CBD	CT	-0.191
HBD1	HC	0.018
HBD2	HC	0.018
CGD	C	0.849
O1D	O2	-0.828
O2D	O2	-0.828
C	C	0.597
O	O	-0.568

Force field modifcations for activated ferryl-oxygen:

atom type	[g/mol]
N6	14.01
N7	14.01
N8	14.01
N9	14.01
C1	12.01
C2	12.01
C3	12.01
C4	12.01
C8	12.01
C9	12.01
FE	55.85
OF	16.00

Bond	k_l [kcal/(mol · Å^2)]	l_0 [Å]
SH-FE	150	2.540
FE-OF	220	1.640
FE-N6	150	2.000
FE-N7	150	2.000
FE-N8	150	2.020
FE-N9	150	2.020
N6-C8	450	1.370
N6-C9	450	1.370
N7-C8	450	1.370
N7-C9	450	1.370
N8-C8	450	1.372
N8-C9	450	1.362
N9-C8	450	1.365
N9-C9	450	1.368

C8-C1	400	1.462
C8-C3	400	1.445
C8-C4	400	1.458
C8-CA	400	1.389
C9-C2	400	1.451
C9-C3	400	1.445
C9-C4	400	1.458
C9-CA	400	1.388
C1-C2	500	1.368
C3-C4	450	1.380
C1-CT	317	1.504
C2-CT	317	1.500
C3-CT	317	1.498
C4-CD	400	1.455
CD-H4	340	1.090

Angle	kθ [kcal/(mol · rad²)]	θ_0[rad]
CT-SH-FE	50	109
SH-FE-OF	50	180
SH-FE-N6	50	90
SH-FE-N7	50	90
SH-FE-N8	50	90
SH-FE-N9	50	90
OF-FE-N6	40	90
OF-FE-N7	40	90
OF-FE-N8	40	90
OF-FE-N9	40	90
FE-N6-C8	40	127.8
FE-N6-C9	40	126.2
FE-N7-C8	40	127.8
FE-N7-C9	40	126.2
FE-N8-C8	40	127.3
FE-N8-C9	40	126.4
FE-N9-C8	40	127.3
FE-N9-C9	40	126.4
N6-FE-N7	40	89.4
N6-FE-N8	40	180
N6-FE-N9	40	90.6
N7-FE-N8	40	90.6
N7-FE-N9	40	180
N8-FE-N9	40	89.4
N6-C8-C1	60	110.8
N6-C8-CA	60	124.7
N6-C9-C2	60	110.6
N6-C9-CA	60	125.8
N7-C8-C1	60	110.8
N7-C8-CA	60	124.7
N7-C9-C2	60	110.6
N7-C9-CA	60	125.8
N8-C8-C3	60	110.7
N8-C8-CA	60	125.2
N8-C9-C4	60	110.6

N8-C9-CA	60	125
N9-C8-C4	60	110.5
N9-C8-CA	60	125.2
N9-C9-C3	60	110.9
N9-C9-CA	60	125
C8-N6-C9	60	106
C8-N7-C9	60	106
C8-N8-C9	60	106.3
C8-N9-C9	60	106.3
C8-C1-C2	60	105.7
C8-C1-CT	60	127.6
C8-C3-C4	60	106.4
C8-C3-CT	60	125.5
C8-C4-C3	60	106
C8-C4-CD	60	124.3
C8-CA-C8	60	125.6
C8-CA-HA	50	117.2
C9-C2-C1	60	106.9
C9-C2-CT	60	125.1
C9-C3-C4	60	106.3
C9-C3-CT	60	125.6
C9-C4-C3	60	106
C9-C4-CD	60	124.3
C9-CA-C9	60	126
C9-CA-HA	50	117
C1-C2-CT	70	128
C1-C8-CA	60	124.5
C1-CT-CT	60	113
C1-CT-HC	50	110.7
C2-C1-CT	70	126.7
C2-C9-CA	60	123.6
C2-CT-HC	50	111.5
C3-C4-CD	70	129.7
C3-C8-CA	60	124.1
C3-C9-CA	60	124.1
C3-CT-HC	50	111.5
C4-C3-CT	70	128.1
C4-C8-CA	60	124.3
C4-C9-CA	60	124.4
C4-CD-CM	60	128.1
C4-CD-H4	50	115.1
H4-CD-CM	50	116.8
CD-CM-HA	50	121.7

DIHEDRAL	IDIVF	PK [kcal/mol]	PHASE	PN
X -SH-FE-X	1	0	0	2
X -FE-N6-X	1	0	180	2
X -FE-N7-X	1	0	180	2
X -FE-N8-X	1	0	180	2
X -FE-N9-X	1	0	180	2
X -N6-C8-X	4	6	180	2
X -N6-C9-X	4	6	180	2

X -N7-C8-X	4	6	180	2
X -N7-C9-X	4	6	180	2
X -N8-C8-X	4	6	180	2
X -N8-C9-X	4	6	180	2
X -N9-C8-X	4	6	180	2
X -N9-C9-X	4	6	180	2
X -C8-C1-X	4	14.5	180	2
X -C8-C3-X	4	14.5	180	2
X -C8-C4-X	4	14.5	180	2
X -C8-CA-X	4	14.5	180	2
X -C9-C2-X	4	14.5	180	2
X -C9-C3-X	4	14.5	180	2
X -C9-C4-X	4	14.5	180	2
X -C9-CA-X	4	14.5	180	2
X -C1-C2-X	4	14.5	180	2
X -C3-C4-X	4	14.5	180	2
X -C1-CT-X	6	0	0	2
X -C2-CT-X	6	0	0	2
X -C3-CT-X	6	0	0	2
X -C4-CD-X	4	0	180	2

3.2

Publikation erschienenen in *Proteins* 74(4): 1028-1035

Identification of selectivity-determining residues in cytochrome P450 monooxygenases: a systematic analysis of the substrate recognition site 5

Alexander Seifert and Jürgen Pleiss[1]

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

[1]Corresponding author:
E-mail: Juergen.Pleiss@itb.uni-stuttgart.de
Fax (+49) 711-685-3196
Telephone (+49) 711-685-3191

Short title: Systematic analysis of SRS-5 in P450 monooxygenases

Key words: regioselectivity; substrate specificity; SRS-5; structure and sequence analysis

3.2.1 Abstract

The large and diverse family of cytochrome P450 monooxygenases was systematically analysed to identify selectivity- and specificity-determining residues in the substrate recognition site 5, which is located in close vicinity to the heme centre. A positively charged heme-interacting residue was identified in the structures of 29 monooxygenases and in 97.7 % of the 6379 CYP sequences investigated here. This heme-interacting residue restricts the

conformation of the substrate recognition site 5 and is preferentially located at position 10 or 11 after the conserved ExxR motif (in 94.4% of the sequences), in 3.3% of the sequences at position 9 or 12. As a result, a classification by the position of the heme-interacting residue allows to predict residues which are closest to the heme centre and restrict its accessibility. In 98.4% of all CYP sequences a preferentially hydrophobic residue is located at position 5 after the ExxR motif which is predicted to point close to the heme centre. Replacing this residue by hydrophobic residues of different size has been shown to change substrate specificity and regioselectivity for CYPs of different superfamilies. 27% of all CYPs are predicted to contain a second selectivity-determining residue at position 9 after the ExxR motif which can be identified by the pattern EXXR-X(7)-{P}-**x**-P-[HKR].

Abbreviations

Cytochrome P450 monooxygenase, CYP; substrate recognition site 5, SRS-5; heme interacting residue, HIR; cytochrome P450 engineering database, CYPED;

3.2.2 Introduction

Cytochrome P450 monooxygenases (CYPs) form a large, ubiquitous enzyme family with a high sequence diversity, and are involved in the metabolism of physiologically important compounds in microorganisms, plants, and animals. Due to their ability to regio- and enatioselectively oxidize a wide variety of compounds, CYPs are attractive biocatalysts for the production of fine chemicals, antibiotics and anti cancer drugs (Urlacher and Eiben 2006). In mammals, these enzymes contribute to the detoxification of a broad range of xenobiotics (Anzenbacher and Anzenbacherova 2001). Therefore, understanding the factors involved in CYP substrate specificity and selectivity is of considerable interest for drug development (Smith et al. 1997).

In the past, extensive sequence and structure analysis has been conducted for this class of enzymes. More than 6300 protein sequences and X-ray structures of 31 different CYPs are known (Fischer et al. 2007). All CYPs belong to the same fold family and have the same catalytic mechanism with a cysteine-ligated heme as active site. On sequence level, there are only few conserved residues: the heme-coordinating cysteine which is essential for CYP

oxygenase function, a phenylalanine located seven amino acids N-terminal to the conserved cysteine, a threonine in the I-helix, which is involved in proton transfer to the heme centre (Vidakovic et al. 1998), and a glutamic acid / arginine pair of the ExxR motif located in the K-helix (Ravichandran et al. 1993). The ExxR motif seems to be important for the stabilisation of the meander loop which in turn is assumed to be pivotal for the maintenance of the CYP tertiary structure and heme binding (Hasemann et al. 1995). However, it has been reported that few CYPs (e.g. CYP157C1 and CYP156A1) lack the conserved arginine, but still form a functional CYP ferrous-CO complex (Rupasinghe et al. 2006). In addition, based on the analysis of 4 CYP crystal structures and 200 sequences a heme interacting arginine was suggested to be functionally conserved in CYPs (Oprea et al. 1997). While the sequences of CYPs are highly diverse, their global structures are similar. Previous structure comparisons of CYPs from 9 different superfamilies revealed several structurally conserved elements: the helix E, the C-terminal half of helix I, helices J and K (containing the highly conserved ExxR motif), the strand β1-3, helices K´ and K´´, the Cys-pocket, helix L, and the strand β3-2 (Mestres 2005). The cofactor heme is deeply buried inside all CYP proteins at the bottom of a large, internal binding cavity. The structural elements near to the heme centre restrict the orientation of the substrate during catalysis and thus mediate substrate specificity and regionselectivity (Fruetel et al. 1994; Bell et al. 2003; Seifert et al. 2006). For protein engineering, it is pivotal to identify residues that are located close to the heme and point with their side chain toward the bound substrate, because these sites are hotspots for designing mutants with improved selectivity or specificity. In the close vicinity of the heme, the substrate binding cavity is confined by the I-helix, the BC loop, and the substrate recognition site 5 (SRS-5) which extents from the conserved ExxR motif to the strand β1-4 (Gotoh 1992). While the I-helix is conserved, the BC loop and the SRS-5 are diverse in sequence and structure. Substitutions in the SRS-5 have recurrently been shown to mediate selectivity, specificity, and activity of CYP enzymes (Born et al. 1995; Liu et al. 2004; Lentz et al. 2006; Meinhold et al. 2006; Urlacher et al. 2006). Hence, in the present work a comprehensive comparison of the SRS-5 region was performed which includes sequence conservation, backbone conformation, and orientation of side chains. Thus, substrate-interacting residues in the SRS-5 were identified and their role in regioselectivity and specificity is discussed.

3.2.3 Materials and Methods

Protein sequence entries were extracted from the Cytochrome P450 Engineering Database (CYPED) (Fischer et al. 2007). To annotate the highly conserved ExxR motif that precedes the substrate recognition site 5 (SRS-5), a multiple sequence alignment of all CYP sequences was performed. All sequences were clustered according to their length, resulting in 6 clusters containing sequences with <551, 551-650, 651-750, 751-850, 851-1000, 1001-1600 residues. Multiple sequence alignments of the 6 clusters were created by using the program CLUSTAL W 1.83 (Thompson et al. 1994). The GONNET series matrices were chosen for the construction of the multiple sequence alignment, a gap opening penalty of 20 and a gap extension penalty of 0.2 were applied. 6379 sequences showed properly aligned ExxR and heme binding motifs. Of each sequence, a short fragment starting with the ExxR motif and including 14 amino acids C-terminal to this motif was extracted, and the amino acid composition of all fragments was analysed.
The protein structures were visualized and analysed with the *PyMOL 0.99* program (DeLano 2002).

3.2.4 Results

Positively charged heme interacting residue:
In all CYP structures, the cofactor heme is embedded between the I-helix, the substrate recognition site 5 (SRS-5), and the BC loop. The SRS-5 extents from one end of the K-helix which contains the highly conserved ExxR motif up to the strand β1-4. In 29 out of 31 CYP crystal structures a heme interacting residue (HIR) (histidine or arginine) was identified in the SRS-5 region. The nearest distance between the negatively charged 7'-propionate of the heme and the side chain of the HIR (Figure 1) varies between 2.5 and 3.6 Å in 27 structures. Basic arginines (pK_a=12.5) are positively charged at pH 7. For histidines in close proximity to carboxylate groups, an increase of the pK_a has previously been observed experimentally (Inagaki et al. 1981; Anderson et al. 1990), hence we assume these histidines are positively charged at pH 7, too. The close distance and the positive charge of the HIR indicates the formation of a charge-assisted hydrogen bond (salt bridge) to the negatively charged 7'-propionate (Table 1). In the crystal structures of two CYPs, no salt bridge is formed between the 7'-propionate and a positively charged amino acid at a position similar to the HIRs in the

27 structures (1CL6, distance of 4.9 Å; 2IAG, distances of 7 Å and 5.6 Å for chain A and B, respectively). In addition, the side chain orientation of this positively charged amino acid is similar to the HIRs in the 27 structures. While its position and side chain orientation is conserved, the increase of the distance to the heme 7´-propionate is caused by a rotation of the propionate moiety. This indicates that for both proteins, a salt bridge might also be observed upon rotation of the propionate. The HIRs were preferentially found at position 10 or 11 after the conserved ExxR motif (in 8 and 17 structures, respectively), while HIRs at position 9 and 12 were found only in 3 and 1 structures, respectively. Although the position of the HIR varies between position 9 and 12 after the ExxR motif, the C_α–C_α distances between the HIR and the arginine of the ExxR motif (Figure 1) are nearly identical (21.4 ± 0.5 Å) (Table 1) which indicates a conserved distance between these two reference points. Interestingly, there are only two structures (CYP102A1 and CYP152A) where the heme does not interact with a positively charged amino acid originating from the SRS-5 region. Although in both crystal structures a lysine was found at position 13 after the ExxR motif, it is not interacting with the heme 7´-propionate (distances of 14.8 and 14.2 Å). Hence, the conserved distance between the ExxR motif and the heme interacting residue has two consequences: the position 12 after ExxR is the upper limit for a HIR in this region, and the position 9 seems to be the lower limit for such an interaction to occur. A positively charged residue outside this sequence window cannot reach the heme 7´ -propionate.

The SRS-5 regions of CYPs which contain the ExxR motif (6379 CYP sequences) were analysed to investigate whether the existence of a positively charged HIR is a general feature of CYPs. 97.7% of these CYPs possess a putative HIR (arginine, histidine, or lysine) in either position 9, 10, 11, or 12 after the ExxR motif. Although lysine was not observed in the crystal structures in these positions, the sequence data indicates the existence of CYPs where lysine acts as a HIR. Consistent with the observation that in the crystal structures arginine or histidine were preferentially found at position 10 and 11, in 94,4% of all sequences a positively charged residue was found at position 10 or 11(Figure 3). In only 142 CYP sequences (2.3%), this positively charged residue is lacking. Only a small fraction of these 142 CYPs have been biochemically characterised: predominantly fatty acid hydroxylases, but also a benzoate 4-monooxygenase, a 9- and 13-hydroperoxy fatty acid metabolizing CYP, a desoxycortisol hydroxylase, a sterol C22 desaturase, and enzymes of polyketide/polypeptide synthetase gene clusters from various producers of secondary metabolites (*Streptomyces*, cyanobacteria). 37 of the 142 CYPs are fusion proteins which corresponds to 66% of all CYPs annotated as fusion protein in the CYPED.

Architecture of the SRS-5 region:

Due to their close proximity to the heme centre, residues originating from the SRS-5 region have a high potential to be involved in substrate binding and hence in control of regioselectivity. In order to compare the potential contribution of each residue in the SRS-5 region to substrate binding, the distance of the closest atom of each amino acid after the ExxR motif to the heme iron was measured in the crystal structures (supplementary material Figure S1a). In 30 out of 31 CYPs, a protruding residue which points toward the heme was found at position 5. In 26 CYPs the side chain at position 5 is closest to the heme, in 4 CYPs (CYP2C8, CYP3A4, CYP152A, CYP167A) the residues at position 4 and 5 are equally close. In only one structure (CYP158A2), the residue at position 4 is considerably closer to the heme centre than the residue at position 5 (6.3 Å and 9.6 Å, respectively). For many CYPs, a second protruding residue pointing towards the heme was observed at position 7, 8, or 9 after the ExxR motif (supplementary material Figure S1a). To characterize the protein backbone conformation in respect to the heme, the distances between the C_α atoms of each amino acid in the SRS-5 region and the heme iron were compared (supplementary material Figure S1b). With only one exception (CYP158A2), in all CYP structures the protein backbone at positions 5 and 8 or 9 are closest to the heme centre which is caused by the formation of bends in the plane of the heme (Figure 2). These bends are a consequence of the conserved spatial distance between the ExxR motif and the HIR. Depending on the number of amino acids between these two conserved reference points, different courses of the protein backbone are observed. For the three structures with HIR at position 9 after the ExxR motif, a rather straight course of the backbone without bends is observed, hence a minimum chain length of 9 amino acids is required to bridge the two reference points (Figure 2a). In the structures with HIR at position 10 or 11, the additional residues cause the formation of a bend at position 5 and, in some CYPs, an additional bend at position 8 or 9 after the ExxR motif. In all structures with HIR at position 9, 10, and 11, and in structures without any HIR the residue at position 5 had the closest distance to the heme centre. However, for the only structure with the HIR at position 12 (CYP158A2), the residue at position 4 is closest. In this CYP, the backbone forms a long loop perpendicular to the heme plane instead of bends in the heme plane for CYPs with HIR at position 10 or 11 (Figure 2b), and thus the closest residue cannot be reliably predicted for CYPs with HIR at position 12. However, these CYPs are very rare (1.6%). Additionally, there are sequences with positively charged residues at both positions 11 and 12 (6.3%). In these CYPs, the HIR is at position 11, and position 5 is closest to the heme as indicated by the crystal structure of CYP2A6 (PDB entry 1Z10).

In addition to backbone conformation, also the amino acid side chain size and orientation adds considerably to the ability of a residue to reach close to the heme centre. The difference between the distances of the C_α atom and the closest atom to the heme iron of each amino acid in the SRS-5 reveals that the side chains at position 4 or 5 and either 7, 8, or 9 point towards the heme centre (supplementary material Figure S1c). In most structures, protruding residues closest to the heme centre are the result of a combination of bends in the heme plane and heme-pointing side chains located in the tip of the bend. However, there are also structures where side chains point to the heme centre in the absence of a bend (position 5 and 7 in CYP101D and CYP167A; position 8 in CYP245A1 and CYP199A2). In general, the formation of bends together with side chains oriented towards the heme centre lead to one or two exposed residues which are able to interact with the substrate near to the active site, and therefore are expected to play a dominant role in regioselectivity and activity.

Analysis of position 5 and 9 after the ExxR motif:

The amino acid composition of position 5 after the highly conserved ExxR motif in 6379 CYP sequences reveals predominantly hydrophobic residues which are predicted to be in contact with the hydrophobic substrates (Figure 4a). Valine (38%) and alanine (18%) are most abundant, followed by leucine (11%) and isoleucine (11%), while glycine (7%), threonine (5%), and serine (4%) are more rarely found. In contrast to position 5, the amino acid compositions of the two neighbouring positions 4 and 6 shows a different distribution (Figure 4c). Here the predominant amino acid is proline with 42% and 52%, respectively.

This data indicates that in all CYPs without a putative HIR (2.3 % of all CYPs) or with the putative HIR at position 9, 10, or 11 (96.1 % of all CYPs), position 5 is occupied by a residue which is suitable for substrate binding close to the heme centre.

The second protruding residue that points towards the heme centre can be found at position 7, 8, or 9. In a subgroup of 11 CYP crystal structures (CYP1A2, CYP2A6, CYP2B4, CYP2C5, CYP2C8, CYP2C9, CYP2D6, CYP2R1, CYP107L1, CYP121, and CYP167A), this residue is at position 9 (supplementary material Figure S1a). In these structures, the HIR is found at position 11. Additionally, in 6 out these 11 structures a proline at position 10 is observed (Table 1). Therefore, proline at position 10 followed by a positively charge residue at position 11 seems to direct the residue at position 9 towards the heme. Two additional CYPs have proline in 10 and the HIR in 11, however, here position 8 is equally close (CYP165B) or closer (CYP165C) than position 9 to the heme centre. In these CYPs a second proline was found at position 8 which causes a change in backbone conformation. The search for

sequences with proline at position 10, but not in 8, followed by a positively charged residue at position 11 (pattern EXXR-X(7)-{P}- **x**-P-[HKR]) resulted in 1740 hits (27% of all CYPs). In this subgroup, the amino acid composition of position 9 is very similar to that of position 5 (90% leucine, isoleucine, valine, or alanine), while for the remaining 73% of sequences a clear preference for hydrophobic residues in this position was not observed (Figure 4b). This indicates that for this subgroup of 27% of all CYPs a second substrate-interacting position besides position 5 can be identified which potentially mediates regioselectivity.

3.2.5 Discussion

Conserved positively charged amino acids in the SRS-5 region:

In addition to the small number of structurally and functionally relevant, highly conserved residues (heme-coordinating cysteine, a phenylalanine, a threonine in the I-helix, and the glutamic acid / arginine pair of the ExxR motif), a positively charged amino acid was identified which is conserved in 97.7% of all CYP sequences. This residue is located at the C-terminal end of the SRS-5 region, which is rather variable in sequence. It is not part of a sequence motif, nor can it be identified by sequence alignment of all CYPs. As indicated by crystal structures, it interacts with the 7´-propionate of the heme. Previously, a comparison of 4 CYP crystal structures and 200 sequences suggested a functionally conserved heme interacting arginine, being also involved in the elimination of water from the active site (Oprea et al. 1997). Mutagenesis data indicated strong effects on heme binding and tertiary structure stability of CYPs (He et al. 1997). By taking into account our finding that this HIR can occur in a defined sequence window of position 9-12 after the ExxR motif we were able to detect the potential HIR in 97.7% of a total of 6379 sequences, and confirmed based on a much broader data base that a heme interacting residue is a common feature among almost all CYPs. Additionally, we analysed the 2.3% of all CYPs that lack this residue. Most of them are fatty acid hydroxylases and fusion proteins of a monooxygenase and a reductase. Interestingly, a HIR is found in 98.3% of all non-fusion CYPs confirming its relevance for non-fusion CYPs, while it is only found in 34% of the fusion proteins and thus seems not to be a requirement for fusion proteins.

Identification of regioselectivity determining positions within SRS-5:

In all CYPs, the cofactor heme is deeply buried inside the protein at the bottom of a large, internal binding cavity. Residues involved in the formation of the substrate binding cavity are all potentially involved in substrate binding and hence in regionselectivity control. The substrates may vary in shape and size. Therefore, residues positioned throughout the entire substrate binding cavity may mediate the orientation of a substrate and thus are involved in regioselectivity control (Melet et al. 2003; Keizers et al. 2004; Sherman et al. 2006). However, the residues in the immediate vicinity of the activated oxygen are most likely in contact with every substrate in its transition state, independent of its size and shape. Three structural elements in close vicinity of the heme contain such residues. The first structural element, the I-helix, is structurally conserved and contains the conserved AGxxT motif (Mestres 2005) . The two highly conserved residues alanine and threonine point toward the heme centre and mediate heme accessibility. The threonine is assumed to be involved in proton transfer (Vidakovic et al. 1998), therefore its substitution generally results in a decrease of the turnover rate and the coupling efficiency (Clark et al. 2006). The second structural element, the BC loop, contains the SRS-1 region and is structurally and sequentially highly diverse, hence it is difficult to identify substrate-interacting residues in the absence of structure information. The third element, the SRS-5, extends from the highly conserved ExxR motif to the strand β1-4. This region contains up to 11 residues which are potentially involved in substrate binding (Gotoh 1992).

The systematic analysis of 31 CYP crystal structures and more than 6300 CYP sequences allowed us to derive rules on how to identify in this variable SRS-5 region only two positions which are due to their close proximity to the heme centre preferentially involved in substrate binding and thus in regioselectivity control, independently of substrate size and shape. They can be easily identified on sequence level due to their distance to the highly conserved ExxR motif, even in the absence of structure information. In 98.4% of all CYPs the residue at position 5 after the ExxR is closest to the heme centre and therefore putatively substrate interacting. The remaining 1.6 % CYPs comprise sequences with a positively charged residue at position 12, but not in 9, 10, or 11. A second putative substrate-interacting residue close to the heme centre resides at position 9 after ExxR and can be identified for 27% of all CYPs by the pattern EXXR-X(7)-{P}-**x**-P-[HKR]. These results indicate that the orientation of residue **x** at position 9 is determined by its neighbouring residue proline at position 10, as it has been observed for other X-Pro peptides (Macarthur and Thornton 1991). The salt bridge between HIR at position 11 and the heme 7´-propionate causes a further restriction of the conformation

in this region. Interestingly, proline at position 8 seems to move the side chain of residue at position 9 away from the heme centre and is therefore excluded from a general pattern to predict the second substrate-interacting residue in SRS-5.

The results of the sequence analysis indicate that hydrophobic residues of different size (leucine, isoleucine, valine, and alanine) are preferentially found at these two positions. Since they point toward the heme centre, they constrain the substrate orientations close to the activated oxygen and limit the accessibility to the active site. As a consequence, increasing the size of their side chains is expected to decrease the number of reaction products and thus to increase regioselectivity (Fruetel et al. 1994; Bell et al. 2003; Seifert et al. 2006). In addition, a limited accessibility of the activated heme oxygen is expected to prevent bulky molecules from being converted which leads to a shift in substrate specificity. Here we show that based on our results one or two positions within the SRS5 region can be readily identified from sequence alone that are involved in substrate binding close to the heme centre and hence mediate specificity and regioselectivity. Focusing to a small number of hotspots reduces greatly complexity and library size, when it comes to CYP mutant design.

Experimental validation:

Mutagenesis data on CYPs of different superfamilies confirm the importance of the residue at position 5 after the ExxR motif (in almost all CYPs) and of the residue at position 9 (in sequences containing the pattern EXXR-X(7)-{P}-**x**-P-[HKR]) for specificity and selectivity. For human CYP1A2 with HIR at position 11, site directed mutagenesis of position 5 resulted in a shifted substrate specificity (Liu et al. 2004). Depending on the size of the side chain at position 5, the enzyme preferentially catalysed O-dealkylation of either 7-methoxyresorufin or 7-ethoxyresorufin, where a correlation between the size of the side chain (valine or leucine) and the preferred substrate size (ethoxy or methoxy group, respectively) was observed. For human CYP2C8 which possesses proline and histidine at position 10 and 11, respectively, an exchange of small side chains at position 5 or 9 by more bulky side chains (V362L and V366L) led to a change of kinetic parameters depending on the size of the substrate (Kerdpin et al. 2004). The small substrate torsemide is converted by both mutants with an unchanged catalytic efficiency as compared to the wild type, while the bulky substrate paclitaxel is barely or not converted by both mutants. Hence, reducing the accessibility to the heme centre by increasing the size of amino acid side chains at position 5 or 9 shifts the substrate specificity of this enzyme towards smaller substrates. For canine CYP2B11 with a HIR at position 11, the substitution of leucine at position 5 after the ExxR motif by valine (L363V) drastically

shifts regioselectivity and converts the progesterone 21-hydroxylase into a highly active and specific progesterone 16-hydroxylase (Born et al. 1995). This can be explained by the increased accessibility of the active site in the mutant. While the wild type oxidises progesterone at its easily accessible C21 position, the mutant is able to oxidise the substrates at the less accessible C17 position. For bacterial CYP102A3 which has no HIR, a mutant identified by directed evolution showed a shift and increase in regioselectivity (Lentz et al. 2006). Upon substitution of alanine at position 5 by the more bulky valine (A330V), the selectivity of n-octane hydroxylation was shifted towards the ω–position. For plant CYP71D18 with a HIR at position 10, the amino acid at position 5 had a pivotal impact on regioselectivity. A single amino acid substitution (F363I) converted the spearmint (-)-limonene hydroxylase from a C6- to a C3-hydroxylase (Schalk and Croteau 2000). This example showed that subtle changes of the residue at position 5 can shift regioselectivity even between two equally accessible positions in a substrate. Thus, even if in some cases the effect of an amino acid substitution cannot be predicted by their bulkiness alone, our analysis allows to identify one or two positions out of several hundred which are hotspots of selectivity and specificity and which can be exchanged without substantial loss of activity.

3.2.6 Conclusion

Cytochrome P450 monooxygenases are highly diverse in sequence and differ in their shape of the binding site. By a systematic comparison of sequence and structure, one or two residues (for 98.4% or 27% of all CYPs, respectively) were identified which point toward the heme centre and therefore are expected to interact with all substrates. While the bulkiness of their side chain differs, their position in the substrate recognition site 5 is highly conserved. These two sites contribute to the orientation of the substrate near the heme centre and therefore are promising targets for protein engineering of enzymes with improved specificity and selectivity. The high conservation of position and orientation of these two residues in the otherwise highly variable SRS-5 region is caused by a structurally conserved heme-interacting residue at position 9, 10, or 11 after the highly conserved ExxR motif which restricts the backbone conformation of the SRS-5 region. Thus, for almost all CYPs one or two hotspots of specificity and selectivity can be identified on sequence level by simply counting amino acid positions after the ExxR motif.

Acknowledgements

This work was performed within the Collaborative Research Centre SFB 706 (Selective Catalytic Oxidations Using Molecular Oxygen; Stuttgart) and funded by the German Research Foundation.

3.2.7 References

Anderson D. E., Becktel W. J. and Dahlquist F. W. (1990). "Ph-Induced Denaturation of Proteins - a Single Salt Bridge Contributes 3-5 Kcal Mol to the Free-Energy of Folding of T4-Lysozyme." *Biochemistry* 29(9): 2403-2408.

Anzenbacher P. and Anzenbacherova E. (2001). "Cytochromes P450 and metabolism of xenobiotics." *Cell Mol Life Sci* 58(5-6): 737-47.

Bell S. G., Chen X., Sowden R. J., Xu F., Williams J. N., Wong L. L. and Rao Z. (2003). "Molecular recognition in (+)-alpha-pinene oxidation by cytochrome P450cam." *J Am Chem Soc* 125(3): 705-14.

Born S. L., John G. H., Harlow G. R. and Halpert J. R. (1995). "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* 23(7): 702-7.

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

DeLano W. L. (2002). The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA, USA.

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

Fruetel J. A., Mackman R. L., Peterson J. A. and Ortiz de Montellano P. R. (1994). "Relationship of active site topology to substrate specificity for cytochrome P450terp (CYP108)." *J Biol Chem* 269(46): 28815-21.

Gotoh O. (1992). "Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences." *J Biol Chem* 267(1): 83-90.

Hasemann C. A., Kurumbail R. G., Boddupalli S. S., Peterson J. A. and Deisenhofer J. (1995). "Structure and Function of Cytochromes-P450 - a Comparative-Analysis of 3 Crystal-Structures." *Structure* 3(1): 41-62.

He Y. A., He Y. Q., Szklarz G. D. and Halpert J. R. (1997). "Identification of three key residues in substrate recognition site 5 of human cytochrome P450 3A4 by cassette and site-directed mutagenesis." *Biochemistry* 36(29): 8831-9.

Inagaki F., Kawano Y., Shimada I., Takahashi K. and Miyazawa T. (1981). "Nuclear Magnetic-Resonance Study on the Micro-Environments of Histidine-Residues of Ribonuclease-T1 and Carboxymethylated Ribonuclease-T1." *J Biochem* 89(4): 1185-1195.

Keizers P. H., Lussenburg B. M., de Graaf C., Mentink L. M., Vermeulen N. P. and Commandeur J. N. (2004). "Influence of phenylalanine 120 on cytochrome P450 2D6 catalytic selectivity and regiospecificity: crucial role in 7-methoxy-4-(aminomethyl)-coumarin metabolism." *Biochem Pharmacol* 68(11): 2263-71.

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

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

Liu J., Ericksen S. S., Sivaneri M., Besspiata D., Fisher C. W. and Szklarz G. D. (2004). "The effect of reciprocal active site mutations in human cytochromes P450 1A1 and 1A2 on alkoxyresorufin metabolism." *Arch Biochem Biophys* 424(1): 33-43.

Macarthur M. W. and Thornton J. M. (1991). "Influence of Proline Residues on Protein Conformation." *J Mol Biol* 218(2): 397-412.

Meinhold P., Peters M. W., Hartwick A., Hernandez A. R. and Arnold F. H. (2006). "Engineering cytochrome P450BM3 for terminal alkane hydroxylation." *Adv Synth Catal* 348(6): 763-772.

Melet A., Assrir N., Jean P., Pilar Lopez-Garcia M., Marques-Soares C., Jaouen M., Dansette P. M., Sari M. A. and Mansuy D. (2003). "Substrate selectivity of human cytochrome P450 2C9: importance of residues 476, 365, and 114 in recognition of diclofenac and sulfaphenazole and in mechanism-based inactivation by tienilic acid." *Arch Biochem Biophys* 409(1): 80-91.

Mestres J. (2005). "Structure conservation in cytochromes P450." *Proteins* 58(3): 596-609.

Oprea T. I., Hummer G. and Garcia A. E. (1997). "Identification of a functional water channel in cytochrome P450 enzymes." *Proc Natl Acad Sci U S A* 94(6): 2133-8.

Ravichandran K. G., Boddupalli S. S., Hasermann C. A., Peterson J. A. and Deisenhofer J. (1993). "Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's." *Science* 261(5122): 731-6.

Rupasinghe S., Schuler M. A., Kagawa N., Yuan H., Lei L., Zhao B., Kelly S. L., Waterman M. R. and Lamb D. C. (2006). "The cytochrome P450 gene family CYP157 does not contain EXXR in the K-helix reducing the absolute conserved P450 residues to a single cysteine." *FEBS Lett* 580(27): 6338-42.

Schalk M. and Croteau R. (2000). "A single amino acid substitution (F363I) converts the regiochemistry of the spearmint (-)-limonene hydroxylase from a C6- to a C3-hydroxylase." *Proc Natl Acad Sci U S A* 97(22): 11948-53.

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

Sherman D. H., Li S., Yermalitskaya L. V., Kim Y., Smith J. A., Waterman M. R. and Podust L. M. (2006). "The structural basis for substrate anchoring, active site selectivity, and product formation by P450 PikC from Streptomyces venezuelae." *J Biol Chem* 281(36): 26289-97.

Smith D. A., Ackland M. J. and Jones B. C. (1997). "Properties of cytochrome P450 isoenzymes and their substrates Part 1: active site characteristics." *Drug Discov Today* 2(10): 406-414.

Thompson J. D., Higgins D. G. and Gibson T. J. (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." *Nucleic Acids Res* 22(22): 4673-80.

Urlacher V. B. and Eiben S. (2006). "Cytochrome P450 monooxygenases: perspectives for synthetic application." *Trends Biotechnol* 24(7): 324-30.

Urlacher V. B., Makhsumkhanov A. and Schmid R. D. (2006). "Biotransformation of beta-ionone by engineered cytochrome P450 BM-3." *Appl Microbiol Biotechnol* 70(1): 53-9.

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

Table 1. Conformation of SRS-5 in respect to the heme centre as determined from 31 crystal structures of CYPs.

Name	Organism	PDB entry	Position 5 after the ExxR motif	Position of HIR after the ExxR motif	Distance d_1 in [Å] **	Distance d_2 in [Å] **	Prolines after the ExxR motif at position
CYP1A2	*Homo sapiens*	2HI4	L382	11 (H388)	2.5	21.5	6, 10
CYP2A6	*Homo sapiens*	1Z10	I366	11 (R372)	2.8	21	6
CYP2B4	*Oryctolagus cuniculus*	1SUO	I363	11 (H369)	2.7	21.1	6, 10
CYP2C5	*Oryctolagus cuniculus*	1NR6	L359	11 (H365)	2.5	20.9	6, 10
CYP2C8	*Homo sapiens*	1PQ2	V362	11 (H268)	2.6	21	6, 10
CYP2C9	*Homo sapiens*	1OG2	L362	11 (H368)	2.7	20.7	6, 10
CYP2D6	*Homo sapiens*	2F9Q	V370	11 (H376)	2.9	21.5	6
CYP2R1	*Homo sapiens*	2OJD	V375	11 (H381)	3.2	21	6
CYP3A4	*Homo sapiens*	1TQN	A370	10 (R375)	2.8	21.7	3
CYP8A1	*Homo sapiens*	2IAG	P355	9 (R359)	7 or 5.6	22.3	5
CYP51	*Mycobacterium tuberculosis*	1E9X	L321	10 (R326)	2.8	21.5	3, 4
CYP55A	*Fusarium oxysporum (NO2 Reductase)*	1CL6	S286	11 (R292)	4.9	22	
CYP101D	*Pseudomonas putida (P450cam)*	1CP4	V295	9 (R299)	2.8	21.9	
CYP102A1	*Bacillus megaterium*	1BU7	A328				3, 6
CYP105	*Nonomuraea recticatena*	2Z36	A292	11 (R298)	2.9	21.2	
CYP107A	*Saccharopolyspora erythraea*	1EGY	P288	10 (R293)	2.8	21	4, 5
CYP107L1	*Streptomyces venezuelae(PikC)*	2BVJ	V290	11 (R296)	2.6	21.6	4
CYP108A	*Pseudomonas sp.*	1CPT	V314	10 (R319)	3.3	21.7	4
CYP119A	*Sulfolobus solfactaricus*	1F4T	V254	10 (R259)	2.9	20.9	3, 4
CYP121	*Mycobacterium tuberculosis*	1N40	F280	11 (R286)	2.7	21.6	10
CYP152A	*Bacillus subtilis*	1IZO	G290				4, 7
CYP154A	*Streptomyces coelicolor*	1ODO	V289	11 (R295)	2.8	21.8	3, 9
CYP154C	*Streptomyces coelicolor*	1GWI	T289	11 (R295)	2.8	22	4
CYP158A2	*Streptomyces coelicolor a3(2)*	2D09	R288	12 (R295)	3.6	22.7	3
CYP165B	*Amycolatopsis orientalis*	1LFK	P283	11 (R289)	2.8	21.5	5, 8, 10
CYP165C	*Amycolatopsis orientalis*	1UED	V292	11 (R298)	2.9	21.1	4, 8, 10
CYP167A	*Polyangium cellulosum*	1Q5D	L301	11 (R307)	2.8	20.9	
CYP175A	*Thermus thermophilus*	1N97	A268	10 (R274)	2.7	21.2	3, 4
CYP176A	*Citrobacter braakii*	1T2B	A285	9 (R289)	3.3	21	4
CYP199A2	*Rhodopseudomonas palustris*	2FR7	V298	10 (R303)	2.7	20.9	4
CYP245A1	*Streptomyces sp. TP-A0274*	2Z3U	V301	10 (R306)	2.8	21.3	3, 4

** see Figure 1

Figures

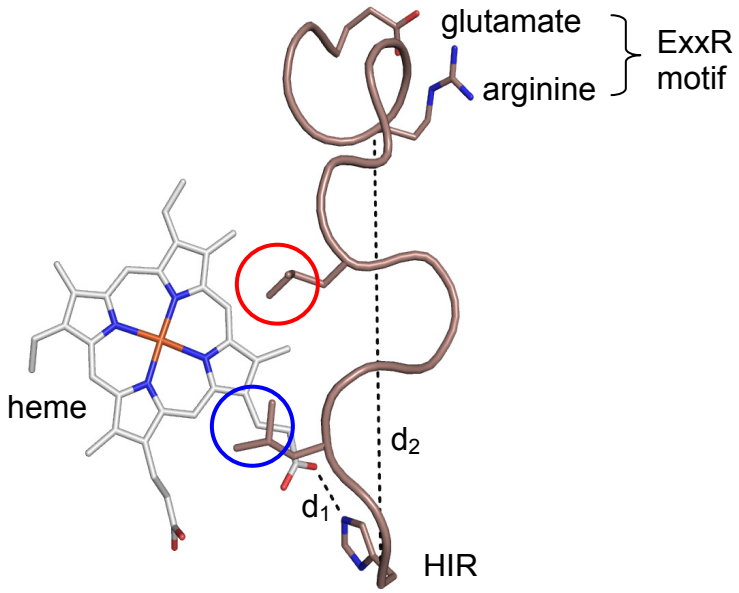


Figure 1. Heme and SRS5 region in CYP2C9: The circles indicate protruding heme pointing amino acids at position 5 (red) and 9 (blue) after the ExxR motif. Distances measured in this study: d_1 between 7´-propionate and HIR; d_2 Cα– Cα distance between the arginine of the structurally highly conserved ExxR motif and the HIR.

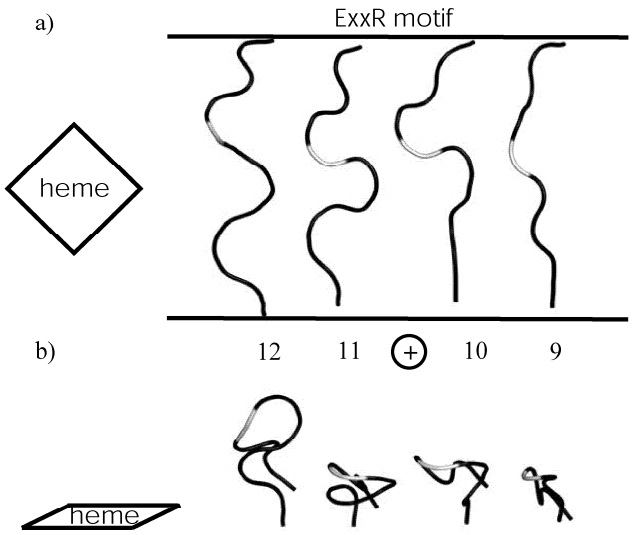


Figure 2. SRS-5 backbone architectures of CYPs. Depending on the position of the HIR after the ExxR motif (9, 10, 11, and 12 from the right to the left): no bend (CYP101D), one bend in the heme plane (CYP51), two bends in the heme plane (CYP2C9), and a loop perpendicular to the heme plane (CYP158A2) were observed, respectively. a) distal view to the SRS-5 region and the heme; b) side view to the SRS-5 region and the heme with the HIR in front and the arginine of the ExxR motif in the back. Position 5 after the ExxR motif is marked white.

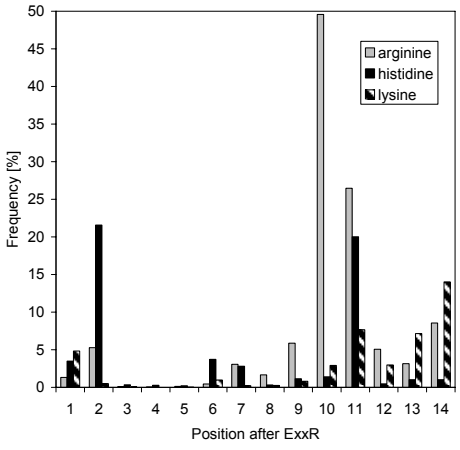


Figure 3. Occurrence of positively charged amino acids (histidine, arginine, lysine) at position 1 to 14 after the ExxR motif.

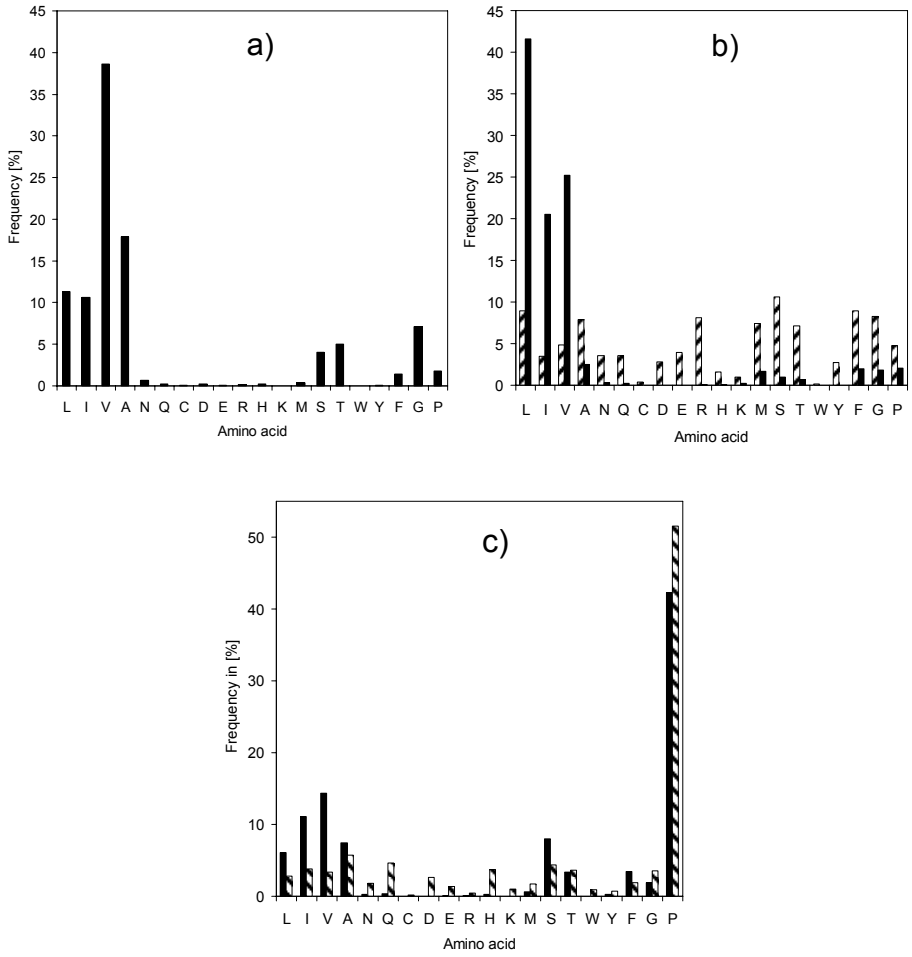
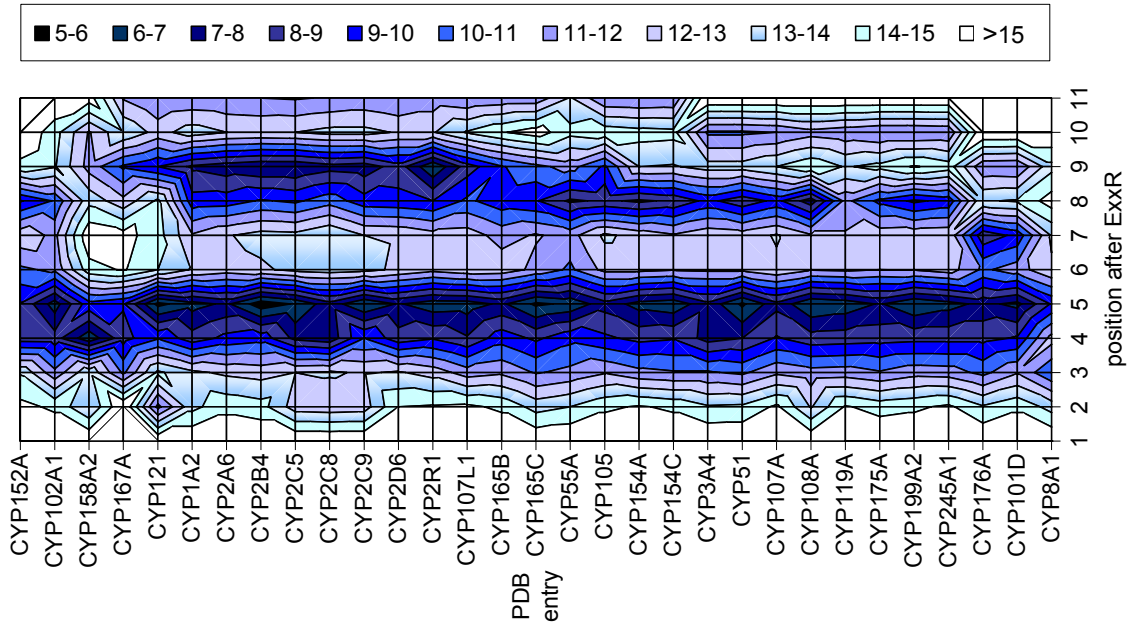


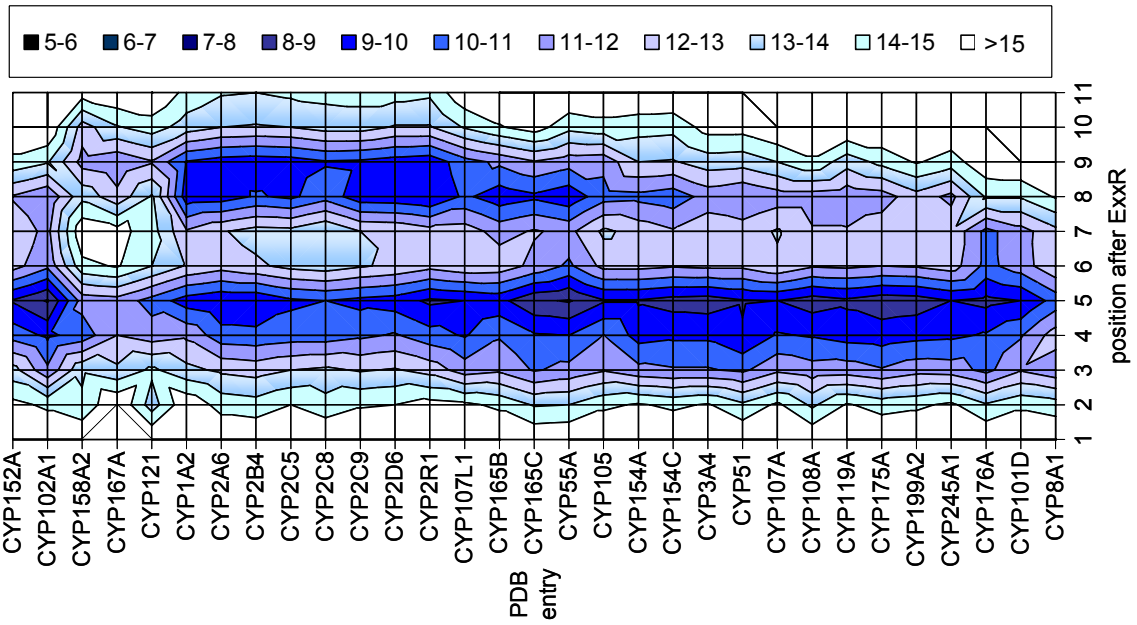
Figure 4. Amino acid composition of a) position 5 after the ExxR motif in all 6379 CYP sequences; b) position 9 in sequences containing the pattern EXXR-X(7)-{P}-x-P-[HKR] (black) and in CYPs without the pattern (dashed); c) position 4 (black) and 6 (dashed) after ExxR in all 6379 CYP sequences.

3.2.8 Supplementary material

a)



b)



c)

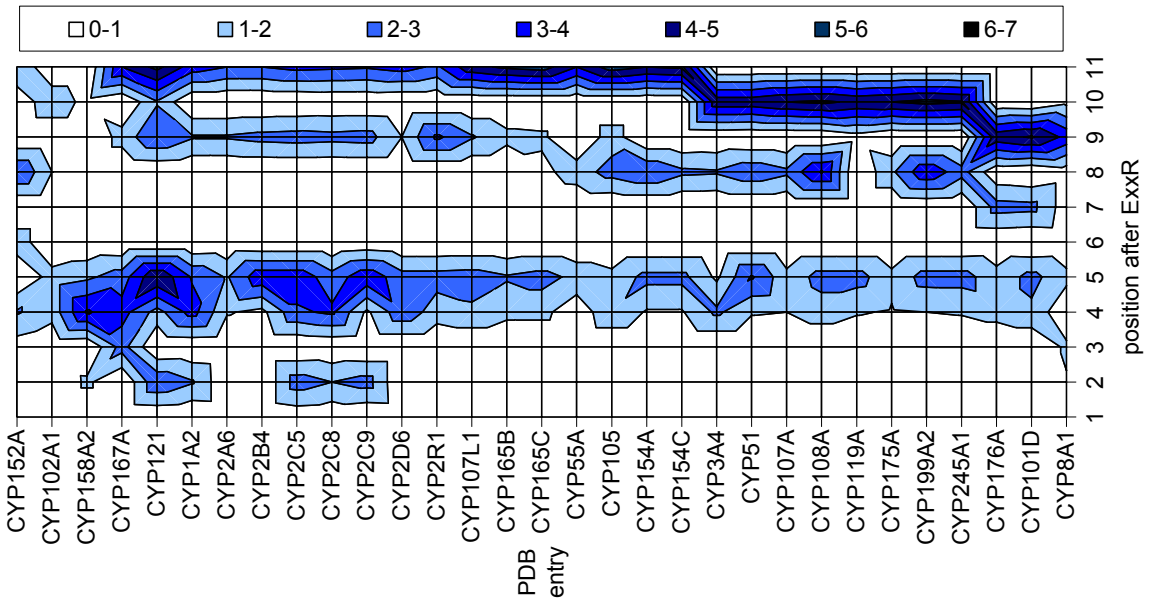


Figure 1S. Distances of SRS-5 residues to the heme centre in Å; a) distance of the closest atom of each residue at position 1 to 11 after the ExxR motif to the heme iron; b) distance of the C_α atoms to the heme iron c) difference of distances a) and b).

3.3

Publikation und Corrigendum erschienenen in *ChemBioChem* 10(5): 853-61.

Rational design of a minimal and highly enriched CYP102A1 mutant library with improved regio-, stereo-, and chemo-selectivity

Alexander Seifert[a], Sandra Vomund[a], Katrin Grohmann[b], Sebastian Kriening[b], Vlada B. Urlacher[a], Sabine Laschat[b] and Jürgen Pleiss[a]*

[a]Institute of Technical Biochemistry, University of Stuttgart, Allmandring 31, 70569 Stuttgart, Germany

[b]Institute of Organic Chemistry, University of Stuttgart, Pfaffenwaldring 55, 70569 Stuttgart, Germany

*Corresponding author:

E-mail: Juergen.Pleiss@itb.uni-stuttgart.de

Fax (+49) 711-685-3196

Telephone (+49) 711-685-3191

Short title: CYP102A1 minimal library

Key words: Cytochrome P450 monooxygenase, terpenes

3.3.1 Abstract

A minimal CYP102A1 mutant library of only 24 variants plus wild type was constructed by combining five hydrophobic amino acids (alanine, valine, phenylalanine, leucine and isoleucine) in two positions. Both positions are located close to the centre of the heme group. The first position 87 has been shown to mediate substrate specificity and regioselectivity in CYP102A1. The second hotspot, position 328 was predicted to interact with all substrates during oxidation and has previously been identified by systematic analysis of 31 crystal structures and 6300 sequences of cytochrome P450 monooxygenases. By systematically altering the size of the side chains, a broad range of binding site shapes was generated. All variants were functionally expressed in *E. coli*. The library was screened with four terpene substrates geranylacetone, nerylacetone, (4*R*)-limonene, and (+)-valencene. 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.

Abbreviations: CYP, cytochrome P450 monooxygenase

3.3.2 Introduction

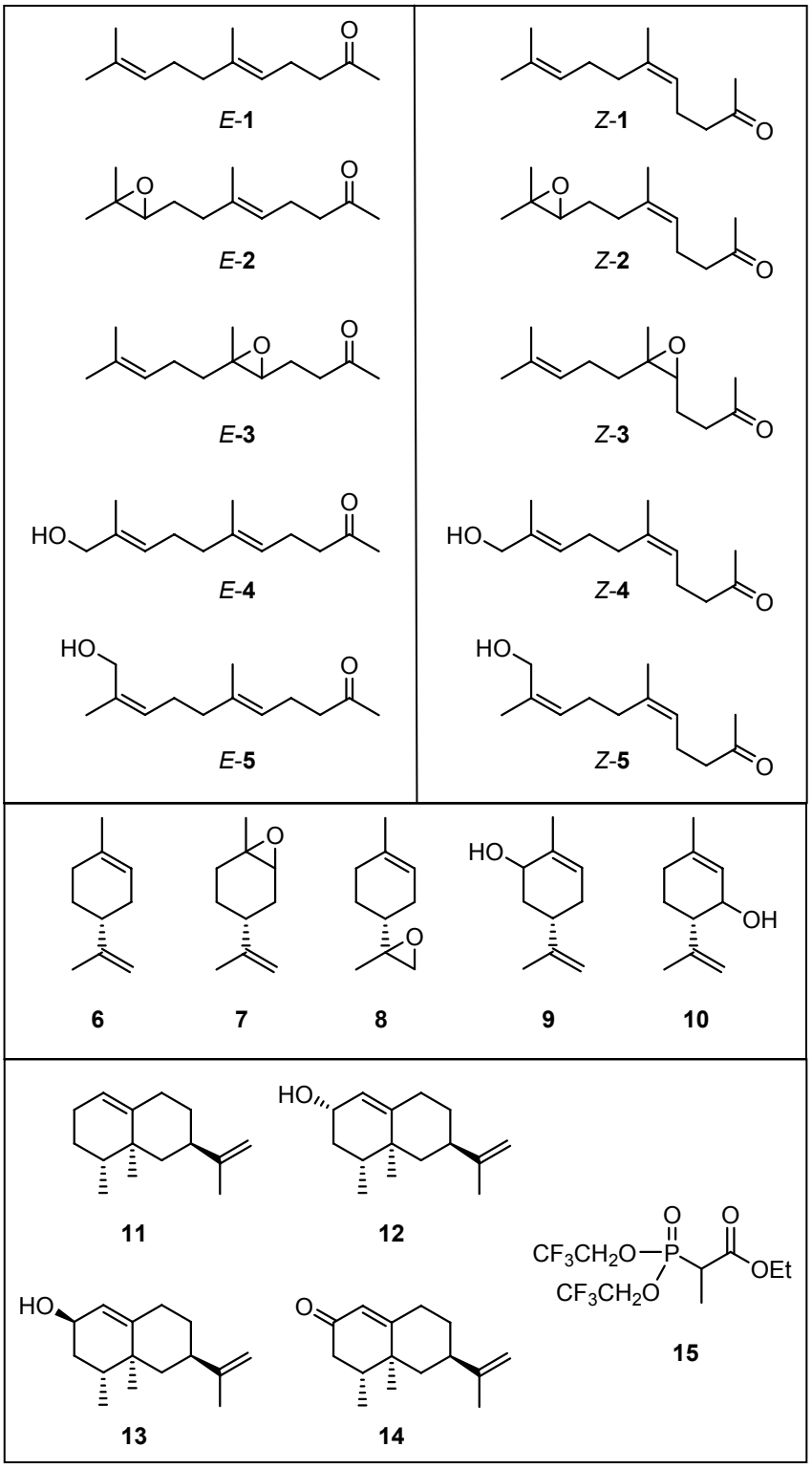
Cytochrome P450 monooxygenase CYP102A1 from *Bacillus megaterium* also known as P450 BM-3 is a widely used and well investigated monooxygenase. It is one of the most promising cytochrome P450 monooxygenases (CYPs) for the application in preparative synthesis, since it is a soluble fusion protein of the monooxygenase and a diflavin reductase (Narhi and Fulco 1987) and is relatively stable under process conditions (Kuehnel et al. 2007). The wild type enzyme is a highly active fatty acid hydroxylase (Munro et al. 2002) and its substrate profile has been widened by mutagenesis (Urlacher et al. 2004). While the wild type enzyme is unselective, mutants have been described with increased regio- and stereoselectivity for different substrates. Among others, position 87 which resides in the substrate recognition site 1 (Gotoh 1992) has a strong impact on substrate specificity and regioselectivity for different substrates (Graham-Lorence et al. 1997; Carmichael and Wong 2001; Urlacher et al. 2006; Li et al. 2008). The CYP102A1 crystal structure reveals that the

residue in this position is involved in the formation of the substrate binding cavity close to the heme centre. Molecular dynamics simulations of the complex of CYP102A1 with α-pinene revealed that different residues in this position cause changes in regioselectivity by stabilising the unpolar substrate in different binding conformations (Branco et al. 2008). A second hotspot for regioselectivity and substrate specificity in CYP102A1 (position 328) was identified by systematic analysis of 31 crystal structures and 6300 sequences (Seifert and Pleiss 2008). This study predicted that in 98 % of all CYP sequences the residue localised in position 5 after the highly conserved ExxR motif is located in the immediate vicinity of the heme group. Its side chain points towards the heme centre and therefore is expected to interact with all substrates during oxidation. The study further revealed that residues in this position are predominantly hydrophobic. Previously, position 328 in CYP102A1 has been shown to play an important role in enantio- and regioselectivity towards alkanes and alkenes (Kubo et al. 2006; Meinhold et al. 2006). A funnel like heme access region in human CYP2C9 has been previously shown by molecular dynamics simulations to restrict possible orientations of the substrate (*S*)-warfarin close to the heme centre, and hence to mediate regioselectivity (Seifert et al. 2006). Hence, the hotspot positions 87 and 328 in CYP102A1 are expected to influence the accessibility of the activated heme oxygen from opposite sides of the heme access channel (Figure 1). Due to their spatial closeness, we expect a cooperative effect of both positions on substrate access to the activated oxygen and therefore on activity, specificity, and selectivity.

Four differently sized and shaped terpenes ((4*R*)-limonene 6, nerylacetone *(Z)*-1, geranylacetone *(E)*-1, and (+)-valencene 11) were chosen to screen the library (Scheme 1). Oxidation of each compound can lead to interesting and valuable products. (4*R*)-limonene 6 the major constituent of citrus peel essential oils has been shown to be oxidised by CYP102A7 to racemic mixtures of (4*R*)-limonene-1,2-epoxide 7, (4*R*)-limonene-8,9-epoxide 8 and carveol 9 (Dietrich et al. 2008). It is further known to be regio- and stereoselectively converted by *Xanthobacter* sp. C20 to form (4*R*,8*R*)-limonene-8,9-epoxide 8 (van der Werf et al. 2001). It has been suggested that a CYP catalyses this oxidation reaction, however, the respective enzyme has not yet been isolated.

(+)-Nootkatone 13 is the most expensive aromatic of grapefruit which can be obtained by the selective oxidation of (+)-valencene 11 (a cheap constituent of orange oil). Besides chemical oxidation, the biotransformation of (+)-valencene 11 by *Chlorella* and *Mucor* species has been reported (Furusawa et al. 2005). Furthermore, cytochrome P450 monooxygenases P450cam, CYP102A1 and mutants thereof have been shown to oxidise (+)-valencene 11

(Sowden et al. 2005). P450cam mutants showed relatively high regioselectivity for C2 oxidation (85%), however, activity was rather low. On the other hand, CYP102A1 variants showed higher activity towards (+)-valencene 11, but were unselective. Hence, CYP102A1 variants with higher selectivity are needed. Furthermore, CYP102A1 has also been shown to unselectively oxidise geranylacetone to 3 products (Watanabe et al. 2007). In the same study a triple mutant (R47L/Y51F/F87V) was introduced that catalyse the epoxidation of geranylacetone *(E)*-1 to 9,10-epoxygeranylacetone *(E)*-2 with high regio- and stereoselectivity. The same mutant also oxidises nerylacetone *(Z)*-1, however, to 5 different products. However, allylic alcohols in position C11 and C12 of the two substrates have not (or only in minor amounts) been detected among the products of the corresponding terpenes. These hydroxy products are valuable starting materials for the total syntheses of natural products such as smenochromene D (Bruder and Moody 2008), indole alkaloids (Clark et al. 2005), pseudopteranes, furanocembranes (Marshall and Dubay 1994), brown algae-derived linear C18 terpenoids (Li et al. 1994), antitumor cembrane lactones crassin and isolobophytolide (Mcmurry and Dushin 1989; Mcmurry and Dushin 1990) and macrocyclic terpenoids humulene, flexibilene, helminthogermacrene and β-elemene (Mcmurry and Kocovsky 1984; Mcmurry and Kocovsky 1985; Mcmurry et al. 1987) as well as cyclopropane-derived materials (Charette et al. 1996; Charette et al. 1998). Thus, there is a need for CYP102A1 variants with changed chemoselectivity. By systematically combining 5 hydrophobic residues in both hotspot positions a minimal mutant library was constructed which can due to its small size be rapidly screened for variants with changed chemoselectivity and/or improved regio- and stereoselectivity.


E-1
Z-1
E-2
Z-2
E-3
Z-3
E-4
Z-4
E-5
Z-5
6
7
8
9
10
11
12
13
14
15
CF3CH2O
CF3CH2O
OEt

Scheme 1

3.3.3 Materials and Methods

Materials

Tryptone/peptone from caseine and yeast extract were purchased from Roth (Karlsruhe, Germany). NADPH tetrasodium salt was from Codexis (Jülich, Germany). 5,6-Epoxy- and 9,10-epoxygeranylacetone *(E)*-3, *(E)*-2, 5,6-epoxy- and 9,10-epoxynerylacetone *(Z)*-3, *(Z)*-2 were prepared as described previously (Watanabe et al. 2007), (4*R*)-limonene-8,9-epoxide 8 was prepared as described elsewhere (Almeida and Jones 2005). All other chemicals (except 11-hydroxy- and 12-hydroxy-geranylacetone *(E)*-4, *(E)*-5, 11-hydroxy- and 12-hydroxynerylacetone *(Z)*-4, *(Z)*-5) used in this work were purchased from Fluka (Buchs, Switzerland) or Sigma (Deisenhofen, Germany) and were of analytical grade or higher.

Synthesis of authentic samples

(*E,E*)-6,10-dimethyl-11-hydroxy-5,9-undecadien-2-one (11-hydroxygeranylacetone) (*E*-4): Geranylacetone *E*-1 (200 mg, 1.03 mmol) was added to a solution of selenium dioxide (46 mg, 0.41 mmol) and *tert*-butyl hydroperoxide (0.64 mL, 5 M solution in nonane, 3.2 mmol) in CH_2Cl_2 at 0 °C. After stirring under nitrogen at 0 °C for 5 h, the mixture was diluted with ethyl acetate (15 mL), and washed successively with water (2 x 10 mL), saturated $NaHCO_3$ (10 mL), water (10 mL) and brine (10 mL). The solution was then dried and evaporated under reduced pressure. Chromatographic purification on silica gel (hexanes/ethyl acetate 7 : 3) afforded alcohol *E*-4 (147 mg, 68%). ^{1}H NMR (300 MHz, $CDCl_3$), δ = 1.60 (s, 3 H), 1.62 (s, 3 H), 2.01 - 2.05 (m, 2 H), 2.12 (s, 3 H), 2.23 - 2.35 (m, 2 H), 2.44-2.50 (m, 1 H), 3.99 (br, 2 H), 5.06 (qt, 1 H, J = 1.3 Hz, J = 7.2 Hz), 5.34 (qt, 1 H, J = 1.4 Hz, J = 6.9 Hz); ^{13}C NMR (125 MHz, $CDCl_3$), δ = 13.7 (CH_3), 15.9 (CH_3), 22.5 (CH_2), 25.9 (CH_2), 29.8 (1-C), 39.2 (CH_2), 43.7 (3-C), 68.8 (4-C), 122.9 (CH), 125.5 (CH), 134.9 (C_q), 135.9 (C_q), 209.3 (2-C) ppm.

(*E,Z*)-6,10-Dimethyl-11-hydroxy-5,9-undecadien-2-one(11-hydroxynerylacetone)(*Z*-4): The alcohol *Z*-4 was prepared as described for *E*-4 from *Z*-1 (200 mg, 1.03 mmol). Chromatographic purification on silica gel (hexanes/ethyl acetate 7 : 3) gave alcohol *E*-4 (123 mg, 57%). ^{1}H NMR (300MHz, $CDCl_3$), δ = 1.58 (s, 3 H), 1.64 (s, 3 H), 2.10 – 2.12 (m, 2 H), 2.14 (s, 3 H), 2.21 – 2.33 (m, 2 H), 2.42 – 2.53 (m, 1 H), 4.00 (br, 2 H), 5.04 - 5.11 (m, 1 H) 5.39 – 5.44 (m, 1 H); ^{13}C NMR (125 MHz, $CDCl_3$), δ = 13.6 (CH_3), 22.4 (CH_2), 23.2 (CH_3),

25.7 (CH_2), 30.0 (1-C), 31.5 (CH_2), 43.9 (3-C), 68.6 (4-C), 123.7 (CH), 125.2 (CH), 135.1 (C_q), 136.1 (C_q), 209.2 (2-C) ppm.

(*E*)-4-Methyl-8-oxonon-4-enal (*E*-16): A solution of epoxide *E*-2 (750 mg, 3.57 mmol) in dry Et_2O (5 mL) was added to a solution of periodic acid dihydrate (976 mg, 4.28 mmol) in dry THF (20 mL) at 0 °C under N_2. After stirring for 0.5 h, saturated aqueous $NaHCO_3$ was added at 0 °C and the reaction mixture was stirred for another 15 min. The white suspension was filtered through a plug of silica and rinsed with Et_2O. The aqueous material was extracted with Et_2O. The combined organic layers were washed with H_2O (50 mL), saturated aqueous $NaHCO_3$ (50 mL) and brine (50 mL), dried and evaporated under reduced pressure. The crude product was purified by chromatography on silica gel (hexanes/ethyl acetate 5 : 1) to give the aldehyde *E*-16 as a colorless oil (345 mg, 61%). ^{1}H NMR (300 MHz, $CDCl_3$), δ = 1.63 (s, 3 H), 2.13 (s, 3 H), 2.22 – 2.35 (m, 4 H), 2.42 – 2.55 (m, 4 H), 5.11 (qt, 1 H, J = 1.3 Hz, J = 7.1 Hz), 9.74 (t, 1 H, J = 1.9) ppm.

(*Z*)-4-Methyl-8-oxonon-4-enal (*Z*-16): The aldehyde *Z*-2 was prepared as described for *E*-16 from *Z*-2 (1.70 g, 8.10 mmol). Chromatographic purification on silica gel (hexanes/ethyl acetate 6 : 1) gave the aldehyde *Z*-16 (756 mg, 56 %). ^{1}H NMR (300 MHz, $CDCl_3$), δ = 1.63 (s, 3 H), 2.13 (s, 3 H), 2.21 – 2.35 (m, 4 H), 2.42 – 2.55 (m, 4 H), 5.11 (qt, 1 H, J = 1.3 Hz, J = 7.0 Hz), 9.75 (t, 1 H, J = 1.9) ppm.

(*2Z,6E*)-Ethyl-2,6-dimethyl-10-oxoundeca-2,6-dienoate (*E*-17): To a solution of 18-crown-6 (1.6 g, 6.06 mmol) in dry THF (15 mL) was added a solution of 15 (671 mg, 2.02 mmol) in dry THF (3 mL) under N_2 atmosphere. The solution was cooled to -78 °C and KHMDS (4.5 mL, 0.5 M in toluene, 2.22 mmol) was added. After stirring for 0.5 h, a solution of the aldehyde *E*-16 (340 mg, 2.02 mmol) in dry THF (7 mL) was added drop wise. The reaction mixture was maintained at -78 °C for 1 h, then quenched with aqueous NH_4Cl (60 mL). The mixture was extracted with Et_2O (8 x 30 mL). The combined organic phases were washed with brine (20 mL), dried over $MgSO_4$ and concentrated in *vacuo* to give a pale yellow oil. Chromatographic purification on silica gel (hexanes/ethyl acetate 9 : 1) gave the unsaturated ethyl ester *E*-17 as a colourless oil (338 mg, 66%). ^{1}H NMR (300 MHz, $CDCl_3$), δ = 1.30 (t, 3 H, J = 7.1 Hz), 1.61 (s, 3 H), 1.88 (s, 3 H), 2.03- 2.10 (m, 2 H), 2.13 (s, 3 H), 2.22 - 2.31 (m, 2 H), 2.42 - 2.49 (m, 2 H), 2.50 - 2.60 (m, 2 H), 4.20 (q, 2 H, J = 7.1 Hz), 5.06 (qt, 1 H, J = 1.3 Hz, J = 7.1 Hz), 5.88 (qt, 1 H, J = 1.5 Hz, J = 7.3 Hz) ppm.

(*2Z,6Z*)-Ethyl-2,6-dimethyl-10-oxoundeca-2,6-dienoate (*Z*-17): The ethyl ester *Z*-17 was prepared as described for *E*-17 from *Z*-16 (1.01 g, 3.04 mmol). Chromatographic purification on silica gel (hexanes/ethyl acetate 5 : 1) gave the unsaturated ethyl ester *Z*-17 (484 mg, 63%). ^{1}H NMR (300 MHz, $CDCl_3$), δ = 1.30 (t, 3 H, J = 7.1 Hz), 1.68 (s, 3 H), 1.89 (s, 3 H), 2.10 - 2.17 (m, 2 H), 2.14 (s, 3 H), 2.22 - 2.31 (m, 2 H), 2.42 - 2.48 (m, 2 H), 2.49 - 2.59 (m, 2 H), 4.20 (q, 2 H, J = 7.2 Hz), 5.08 - 5.13 (m, 1 H), 5.90 (qt, 1 H, J = 1.5 Hz, J = 7.5 Hz) ppm.

(*2Z,6E*)-Ethyl-2,6-dimethyl-9-(2-methyl-1,3-dioxolan-2-yl)nona-2,6-dienoate (*E*-18): To a solution of *E*-17 (445 mg, 1.8 mmol) in dry CH_2Cl_2 (10 mL) was added a solution of bis-1,2-(trimethylsilyloxy)ethane (473 mg, 2.3 mmol) in dry CH_2Cl_2 (5 mL) under N_2 atmosphere. The solution was cooled to -78 °C and TMSOTf (6.4 µL, 35 µmol) was added dropwise. After 1 h, TLC control showed incomplete conversion of *E*-17. An additional portion of bis-1,2-(trimethylsilyloxy)ethane (40 mg, 0.19 mmol) was added to the reaction mixture. After stirring for 10 min, dry pyridine (4.5 mL) was added and stirred for another 3 min. The solvents were evaporated in *vacuo* to give a yellow oil. Chromatographic purification on silica gel (hexanes/ethyl acetate 15 : 1) yielded the protected unsaturated ethyl ester *E*-18 as a colourless oil (427 mg, 82%). ^{1}H NMR (300 MHz, $CDCl_3$), δ = 1.30 (t, 3 H, J = 7.2 Hz), 1.33 (s, 3 H), 1.61 (s, 3 H), 1.63 – 1.70 (m, 2 H), 1.87 - 1.90 (m, 3 H), 2.03 - 2.14 (m, 4 H), 2.51 - 2.60 (m, 2 H), 3.91 – 3.97 (m, 4 H), 4.20 (q, 2 H, J = 7.1 Hz), 5.15 (qt, 1 H, J = 1.3 Hz, J = 7.1 Hz), 5.90 (qt, 1 H, J = 1.5 Hz, J = 7.3 Hz) ppm.

(*2Z,6Z*)-Ethyl-2,6-dimethyl-9-(2-methyl-1,3-dioxolan-2-yl)nona-2,6-dienoate (*Z*-18): *Z*-18 was prepared as described for *E*-18 from *Z*-17 (438 mg, 1.90 mmol). Chromatographic purification on silica gel (hexanes/ethyl acetate 5 : 1) gave the unsaturated ethyl ester *Z*-18 (520 mg, 92%). ^{1}H NMR (300 MHz, $CDCl_3$), δ = 1.30 (t, 3 H, J = 7.2 Hz), 1.32 (s, 3 H), 1.54 (s, 3 H), 1.67 – 1.70 (m, 2 H), 1.87 - 1.90 (m, 3 H), 2.03 - 2.17 (m, 4 H), 2.50 - 2.59 (m, 2 H), 3.91 – 3.97 (m, 4 H), 4.22 (q, 2 H, J = 7.1 Hz), 5.12 – 5.19 (m, 1 H), 5.87 – 5.94 (m, 1 H) ppm.

(*2Z,6E*)-ethyl-2,6-dimethyl-9-(2-methyl-1,3-dioxolan-2-yl)nona-2,6-dien-1-ol (*E*-19): A solution of *E*-18 (338 mg, 1.14 mmol) in dry CH_2Cl_2 (10 mL) was cooled to -78 °C in a N_2 atmosphere. To this solution was added DIBAL (2.5 mL, 1 M in hexane, 2.5 mmol). After stirring for 1 h at -78 °C, a saturated aqueous potassium sodium tartrate solution (25 mL) was added. The reaction mixture was allowed to warm to room temperature and was stirred

subsequently over night. The aqueous phase was extracted with Et_2O (5 x 20 mL). The combined organic layers were dried and evaporated under reduced pressure. The crude product was purified by chromatography on silica gel (hexanes/ethyl acetate 3 : 1) to give the alcohol *E*-19 as a colorless oil (284 mg, 98%). ^{1}H NMR (300MHz, $CDCl_3$), δ = 1.33 (s, 3 H), 1.59 - 1.63 (m, 3 H), 1.63 – 1.70 (m, 2 H), 1.77 - 1.82 (m, 3 H), 1.96 - 2.06 (m, 2 H), 2.06 - 2.20 (m, 4 H), 3.90 – 3.97 (m, 4 H), 4.11 (dd, 2 H, J = 0.7 Hz, J = 5.8 Hz), 5.13 (qt, 1 H, J = 1.3 Hz, J = 7.0 Hz), 5.24 – 5.32 (m, 1 H) ppm.

(*2Z,6Z*)-Ethyl-2,6-dimethyl-9-(2-methyl-1,3-dioxolan-2-yl)nona-2,6-dien-1-ol (*Z*-19): *Z*-19 was prepared as described for *E*-19 from *Z*-18 (338 mg, 1.14 mmol). Chromatographic purification on silica gel (hexanes/ethyl acetate 5 : 1) gave the unsaturated ethyl ester *Z*-19 (278 mg, 96%). ^{1}H NMR (300 MHz, $CDCl_3$), δ = 1.32 (s, 3 H), 1.60 - 1.67 (m, 2 H), 1.67 – 1.70 (m, 3 H), 1.78 - 1.82 (m, 3 H), 1.99 - 2.19 (m, 6 H), 3.90 – 3.98 (m, 4 H), 4.10 (dd, 2 H, J = 0.7 Hz, J = 5.9 Hz), 5.13 – 5.21 (m, 1 H), 5.27 – 5.35 (m, 1 H) ppm.

(*E,E*)-6,10-Dimethyl-12-hydroxy-5,9-undecadien-2-one (12-hydroxygeranylacetone) (*E*-5): Amberlyst 15 (200 mg) was added to a solution of *E*-19 (280 mg, 1.1 mmol) in acetone/4 % H_2O (25 mL) and the mixture was stirred for 20 h at room temperature. The ion exchange resin was filtered off, the filtrate was washed with acetone and successively concentrated in *vacuo*. The residue was dried azeotropically with toluene. The crude product was purified by chromatography on silica gel (hexanes/ethyl acetate 3 : 1) to give the unprotected alcohol *E*-5, a light yellow oil (184 mg, 80%). ^{1}H NMR (300MHz, $CDCl_3$), δ = 1.42 (s, 1 H), 1.60 - 1.63 (m, 3 H), 1.78 – 1.82 (m, 3 H), 2.07 – 2.16 (m, 7 H), 2.22 - 2.29 (m, 2 H), 2.42 – 2.48 (m, 2 H), 4.10 (d, 2 H, J = 0.7 Hz), 5.07 – 5.14 (m, 1 H), 5.26 – 5.33 (m, 1 H); ^{13}C NMR (125 MHz, $CDCl_3$), δ = 16.0 (CH_3), 21.3 (CH_3), 22.4 (CH_2), 26.0 (CH_2), 30.0 (1-C), 39.7 (CH_2), 43.7 (3-C), 61.5 (4-C), 123.1 (CH), 127.9 (CH), 134.6 (C_q), 135.9 (C_q), 209.4 (2-C) ppm; HRMS (ESI): *m/z* calcd for $C_{13}H_{22}O_2Na$: 233.1512; found: 233.1510 $[M + Na]^+$.

(*E,Z*)-6,10-Dimethyl-12-hydroxy-5,9-undecadien-2-one (12-hydroxynerylacetone) (*Z*-5): *Z*-5 was prepared as described for *E*-5 from *Z*-19 (274 mg, 1.08 mmol). Chromatographic purification on silica gel (hexanes/ethyl acetate 3 : 1) gave the unprotected alcohol *Z*-5 (195 mg, 86%). ^{1}H NMR (300 MHz, $CDCl_3$), δ = 1.43 (s, 1 H), 1.67 - 1.69 (m, 3 H), 1.78 – 1.82 (m, 3 H), 2.07 – 2.16 (m, 7 H), 2.22 - 2.29 (m, 2 H), 2.42 – 2.48 (m, 2 H), 4.10 (d, 2 H, J = 0.7 Hz), 5.07 – 5.14 (m, 1 H), 5.26 – 5.33 (m, 1 H); ^{13}C NMR (125 MHz, $CDCl_3$), δ = 21.4 (CH_3), 22.2 (CH_3), 23.4 (CH_2), 30.1 (CH_2), 32.0 (1-C), 39.7 (CH_2), 43.8 (3-C), 61.6

(4-C), 123.9 (CH), 127.9 (CH), 134.9 (C_q), 136.1(C_q), 208.3 (2-C) ppm, HRMS (ESI): *m/z* calcd for $C_{13}H_{22}O_2Na$: 233.1512; found: 233.1501 [M + Na]$^+$.

Ethyl-2-(bis(2,2,2-trifluoroethoxy)phosphoryl)acetate (21) (Still and Gennari 1983; Malerich and Trauner 2003) was prepared analogously to Still and Gennari from Ethyl-2-(diethoxyphosphoryl)acetate (13.23 g, 64.5 mmol, 1 eq.) in toluene (80 mL), CF_3CH_2OH (13.55 g, 135.5 mmol, 2.1 eq.) and ^{i}Pr$_2$NEt in toluene (20 mL). ^{1}H NMR (300 MHz, $CDCl_3$), δ = 1.30 (t, 3 H, J = 7.2 Hz), 3.16 (d, 2 H, J = 21.1 Hz), 4.23 (q, 2 H, J = 7.2 Hz), 4.52 – 4.40 (m, 4 H) ppm.

Ethyl-2-(bis(2,2,2-trifluoroethoxy)phosphoryl)propanoate (15): A suspension of NaH (840 mg, 60%, 21 mmol) in dry THF (15 mL) was cooled to 0 °C under N_2. To this suspension was added a solution of 22 (6.8 g, 20.5 mmol) in dry THF (10 mL). The reaction mixture was allowed to warm to room temperature and was stirred for 3 h. The orange solution was cooled to -10 °C and a solution of MeI (1.4 mL, 21.7 mmol) was added. The reaction mixture was allowed to warm to 0 °C temperature, H_2O (30 mL) was added and the solution was stirred for 1 h. The reaction mixture was diluted with CH_2Cl_2 (100 mL) and brine (100 mL) was added. The aqueous phase was extracted with CH_2Cl_2 (2 x 50 mL). The combined organic layers were dried and evaporated under reduced pressure. The crude product was purified by chromatography on silica gel (hexanes/ethyl acetate 6 : 1) to give compound 15 as a pale orange oil (2.21 g, 31%). ^{1}H NMR (300 MHz, $CDCl_3$), δ = 1.30 (t, 3 H, J = 7.1 Hz), 1.52 (qd, 3 H, J = 7.1 Hz, J = 19.3 Hz), 3.19 (qd, 1 H, J = 7.4 Hz, J = 22.7 Hz), 4.24 (q, 2 H, J = 7.2 Hz), 4.51 – 4.36 (m, 4 H) ppm.

Mutant Expression

The 24 CYP102A1 single and double mutants plus wild type were heterologously expressed in *E. coli* as reported previously (Maurer et al. 2003). The pET22b and pET28+ derivative were used. Following the manufacturer's protocol, the QuikChange™ site-directed mutagenesis kit from Stratagene (La Jolla, California, USA) was used to introduce site directed mutations.

Enzyme Activity Measurements

CO-difference spectra measurements were used to determine CYP concentrations as described elsewhere (Omura and Sato 1964). Extinction coefficient of 91 $mM^{-1}cm^{-1}$ was used for calculations. A NADPH oxidation assay was applied to determine the initial activity of the mutants towards nerylacetone *(Z)*-1, geranylacetone *(E)*-1, (4*R*)-limonene 6 and (+)-valencene 11. One ml final reaction mixture contained 50 mM potassium phosphate buffer pH 7.5, 2% (v/v) DMSO, 0.2 mM substrate as well as 0.5 µM CYP enzyme. Reaction was started by adding 100 µl of 2 mM NADPH solution. The initial activity was determined from the absorption decrease at 340 nm, whereas the slope of the first 30 seconds was measured and an extinction coefficient of 6.22 $mM^{-1}cm^{-1}$ was applied.

GC-MS Analyses

The aqueous reaction mixture was extracted twice with diethyl ether (or ethyl acetate in case of (+)-valencene) and the organic phase was dried over magnesium sulfate. Analysis of reaction products and conversion were performed on a Shimadzu QP2010 GC/MS with EI-ionisation, the GC was equipped with a FS-supreme-5 capillary column (length: 30 m, internal diameter: 0.25 mm, film thickness: 0.25 µm). *Analysis of (4R)-limonene oxidation products.* The GC was programmed as follows: 40 °C, 1 min. iso; 1 °C/min. to 70 °C, 20 min. iso; 1 °C/min. to 100 °C; 10 °C/min. to 110 °C; 30 °C/min. to 250 °C; injector temperature 250 °C. The oxidation products carveol 9 and (4*R*)-limonene-8,9-epoxide 8 were identified by authentic samples. (4*R*)-limonene-1,2-epoxide 7 was identified by comparison of its characteristic mass fragmentation pattern with the NIST mass spectrometry data base (NIST05). Isopiperitenol was identified by comparison of the mass spectra to literature data (Lucker et al. 2004). *Analysis of neryl- and geranylacetone oxidation products.* The GC was programmed as follows: 120 °C, 30 sec. iso; 5 °C/min. to 165 °C, 2 °C/min. to 185 °C, 30 °C/min. to 280 °C, 1 min. iso; injector temperature 250 °C. The oxidation products 11-hydroxy- and 12-hydroxygeranylacetone *(E)*-4, *(E)*-5, 5,6-epoxy- and 9,10-epoxy-geranylacetone *(E)*-3, *(E)*-2, 11-hydroxy- and 12-hydroxynerylacetone *(Z)*-4, *(Z)*-5, 5,6-epoxy- and 9,10-epoxy-nerylacetone *(Z)*-3, *(Z)*-2 were identified using authentic samples. *Analysis of (+)-valencene oxidation products.* The GC was programmed as follows: 150 °C, 4 min. iso; 10 °C/min. to 250 °C, 5 min iso; 50 °C/min. to 300 °C, 1 min. iso; injector temperature 250 °C. (+)-nootkatone 14 was identified using an authentic sample and *cis*- and *trans*-nootkatol 12, 13 were identified by comparison of the mass spectra to literature data (Takahashi et al. 2007). To calculate the amount of converted substrate GC peak area ratios

between substrate and an internal standard (geraniol for geranyl- and nerylacetone conversions; citronellal for (4*R*)-limonene conversions; (-)-carvone for (+)-valencene conversions) were measured, before and after reactions. With an enzyme concentration of c µM, *a* µM of converted substrate and a time of *t* minutes to consume the NADPH, the substrate oxidation rate is defined as *a/c*t* µmol (µmol P450)$^{-}$1 min^{-}1. Regioselectivity was determined from the gas chromatograms by integrating the product peaks.

3.3.4 Results

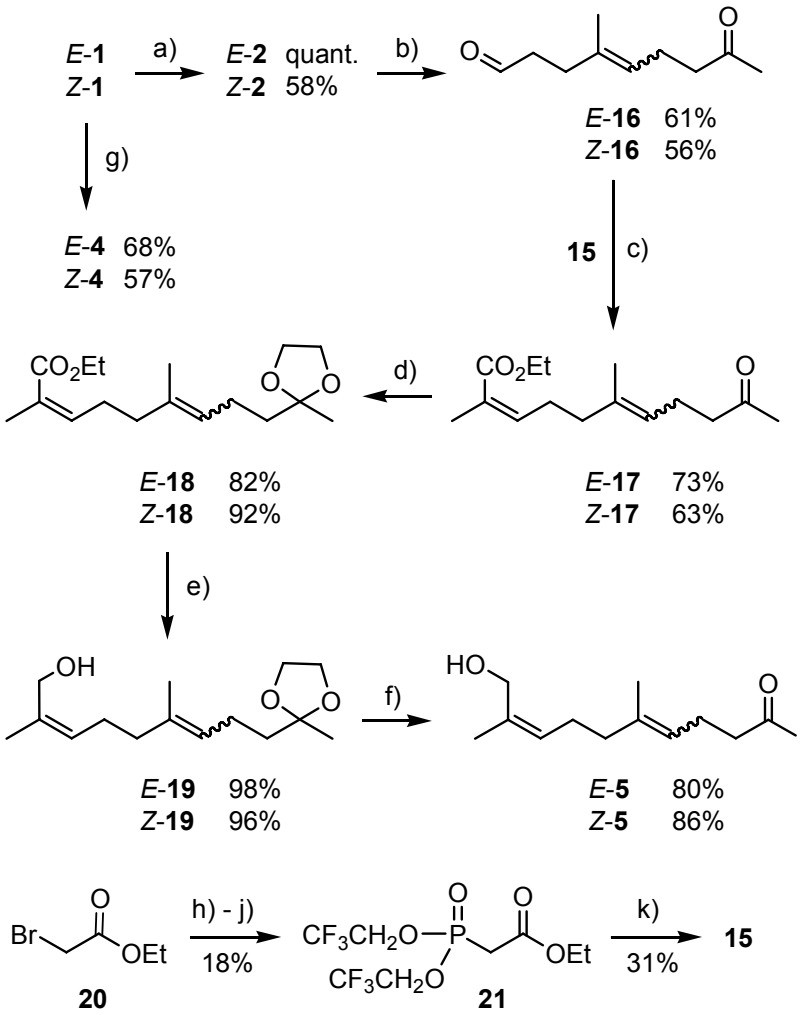
Terpene oxidation

A minimal library of 24 variants plus wild type was constructed by combination of 5 hydrophobic amino acids (alanine, valine, phenylalanine, leucine and isoleucine) in two positions (87 and 328), which have been previously identified as hotspots for selectivity. Altering the side chain size of these two amino acids drastically changes the shape of the substrate binding cavity in close vicinity of the heme group (Figure 1). The presence of the Soret-band at 450 nm upon measuring of CO-difference spectra indicated that all 24 variants and the wild type functionally incorporated the heme group during expression. The CYP102A1 wild type enzyme converted all terpenes (geranylacetone *(E)*-1, nerylacetone *(Z)*-1, (4*R*)-limonene 6 and (+)-valencene 11), however with poor regioselectivity and activity. The screening of the mutant library revealed 11 variants that convert at least one terpene to a new valuable product and possess strongly increased selectivity as compared to the wild type enzyme (Table 1). While 4 single mutants in positions 87 or 328 and one double mutant showed improved or strongly shifted regio- or chemoselectivity towards the small acyclic terpenes geranyl- or nerylacetone *(E)*-1, *(Z)*-1 , amino acid substitution in both positions were necessary to improve selectivity towards the more bulky substrates (4*R*)-limonene 6 and (+)-valencene 11. Only 3 variants (A328F, F87I/A328F and F87A/A338V) did not convert any of the four terpenes. In the following we focus only on the variants with the highest regio- and stereoselectivity, strongly shifted product profile, and variants which produce new and interesting products.

Synthesis of the oxidation products

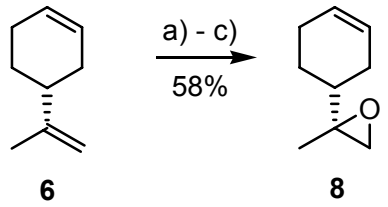
11-Hydroxygeranylacetone (*E*)-4 and 11-hydroxynerylacetone (*Z*)-4 were prepared from geranylacetone (*E*)-1 and nerylacetone (*Z*)-1 respectively via selenium dioxide oxidation

according to the method by McMurry (Mcmurry et al. 1987; Mcmurry and Dushin 1990), which gave the desired products (*E*)-4, (*Z*)-4 in 68 % and 57 % yield respectively (Scheme 2). The synthesis of 12-hydroxygeranylacetone (*E*)-5 commenced with treatment of geranylacetone (*E*)-1 with MCPBA in $CHCl_3$ at 0°C to give the corresponding 9,10-epoxygeranylacetone (*E*)-2 quantitatively [15]. Epoxide (*E*)-2 was submitted to periodate cleavage under acidic conditions following the method by Mori and Kogen (Muto et al. 1999; Tago et al. 2000) and the aldehyde (*E*)-16 was isolated in 61 % yield. Horner-Emmons olefination of compound (*E*)-16 using the Still-Gennari modification (Still and Gennari 1983; Malerich and Trauner 2003) with the phosphonate 15 in the presence of KHMDS and 18-crown-6 in THF at -78°C gave the α,β-unsaturated ester (*E*)-17 in 73 % yield. The Still-Gennari-phosphonate 15 was synthesized in 4 steps in an overall yield of 6 % (Still and Gennari 1983; Malerich and Trauner 2003). Sequential reaction of ethyl bromoacetate 20 with triethylphosphite at 155°C for 5 h followed by treatment of the phosphonate with PCl_3 and subsequent addition of 2,2,2-trifluoroethanol yielded the unbranched Still-Gennari-phosphonate 22 in 33 %. Methylation of 22 with methyl iodide and sodium hydide in THF gave compound 15 in 31 % yield. Protection of the carbonyl group in (*E*)-17 was achieved by transacetalisation employing the method by Noyori and Hashimoto (Tsunoda et al. 1980) with bis-1,2-(trimethylsilyloxy)ethane in the presence of TMSOTf in CH_2Cl_2 to yield the acetal (*E*)-18 in 82 %. After reduction of (*E*)-18 with DIBAL in CH_2Cl_2 at -78°C the allylic alcohol (*E*)-19 was obtained in 98 % yield. Subsequent deprotection with Amberlyst 15 ion exchange resin in water, acetone at room temperature gave the desired 12-hydroxygeranylacetone (*E*)-5 in 80 % yield. The synthesis of 12-hydroxynerylacetone (*Z*)-5 proceeded in a similar fashion, giving the desired product (*Z*)-5 in 6 steps and 16 % overall yield from nerylacetone (*Z*)-1.



Scheme 2 a) MCPBA, $CHCl_3$, 0°C; b) HIO_4, Et_2O, THF, 0°C; c) 15, KHMDS, 18-crown-6, THF, -78°C1.5 h; d) $(TMSOCH_2)_2$, TMSOTf, CH_2Cl_2, -78°C, 1 h; e) DIBAL, CH_2Cl_2, -78°C, 1 h; f) acetone, H_2O, Amberlyst 15, rt, 20 h; g) SeO_2, tBuOOH, CH_2Cl_2, 0°C, 5 h; h) $P(OEt)_3$, 150°C, 5 h; i) PCl_5, 80°C, 16 h; j) CF_3CH_2OH, toluene, iPr_2NEt, 0°C, 1 h; k) NaH, THF, 0°C, then MeI, -10°C, 1 h.

Following a method by Jones jr. (Almeida and Jones 2005) (4*R*)-limonene 6 was sequentially treated with NBS in MeOH at room temperature, followed by oxidation with MCPBA in CH_2Cl_2 and final treatment with zinc and NH_4Cl in ethanol to give (4*R*)-limonene-8,9-epoxide 8 in 58 % overall yield (Scheme 3).



Scheme 3 a) NBS, MeOH, rt, 1 d; b) MCPBA, CH_2Cl_2, 0°C to rt, 3 d; c) Zn, NH_4Cl, EtOH, rt, 1 d

Best geranylacetone-converting variants

Geranylacetone *(E)*-1 is converted by the CYP102A1 wild type enzyme to 2 products (Table 3). The main product is 9,10-epoxygeranylacetone *(E)*-2 (54%). The second product at a retention time of 12.3 min could not be identified. 11-Hydroxy- and 12-hydroxygeranylacetone *(E)*-4, *(E)*-5 are not among the products. The screening of our minimal library showed three single mutants (F87V, F87I, and F87L) with a strong preference for the formation of 9,10-epoxygeranylacetone *(E)*-2. Variant F87L produced 9,10-epoxygeranylacetone *(E)*-2 almost exclusively (97%). Furthermore, 9 variants were, in contrast to the CYP102A1 wild type, able to hydroxylate geranylacetone at allylic positions C11 and C12. Two of these 9 variants showed a strong shift in chemoselectivity. Variant A328V produced the highest amount of 11-hydroxygeranylacetone *(E)*-4 (37%) (Table 3). However, this variant was rather unselective, since it converted this acyclic terpene *(E)*-1 to a total of 3 products including 12-hydroxygeranylacetone *(E)*-5 (31%) and 9,10-epoxygeranylacetone *(E)*-2 (30%). Variant F87V/A328L in turn almost exclusively converted geranylacetone to 2 allylic alcohols, with a strong preference for position C12 (80%) over C11 (15%). To our knowledge, this is the first time a CYP is presented that oxidised geranylacetone *(E)*-1 to allylic alcohols in position 11 and 12. Except variant A328V, which showed 1.7 fold increased oxidation rate as compared to wild type, the remaining variants with improved selectivity were less active than the wild type (Table 2).

Best nerylacetone-converting variants

CYP102A1 wild type oxidised nerylacetone *(Z)*-1 unselectively to more than 4 different products (Table 3). Two products were identified as epoxides 9,10-epoxynerylacetone *(Z)*-2 (47%) and 5,6-epoxynerylacetone *(Z)*-3 (11%). Two minor products were allylic alcohols 11-hydroxy- and 12-hydroxynerylacetone *(Z)*-4, *(Z)*-5 (11% and 1%, respectively). Variant A328V showed a strong shift in chemoselectivity and a considerable improvement of regioselectivity. In contrast to the wild type enzyme this variant produces almost exclusively 2 allylic alcohols, with a marked preference for 11-hydroxynerylacetone *(Z)*-4 (64%) as compared to 12-hydroxynerylacetone *(Z)*-5 (34%). To our knowledge, for the first time a P450 catalyst is presented that oxidised nerylacetone *(Z)*-1 preferentially at allylic position 11 to form 11-hydroxynerylacetone *(Z)*-4. Variant A328V revealed 2.2 fold increased nerylacetone oxidation rate as compared to wild type (Table 2).

Best (4*R*)-limonene-converting variants

The CYP102A1 wild type converts (4*R*)-limonene 6 to four different products (racemic mixtures of (4*R*)-limonene-1,2-epoxide 7 (30%), (4*R*)-limonene-8,9-epoxide 8 (7%), isopiperitenol 10 (54%), and carveol 9 (9%) (Table 3). The screening of our minimal library revealed two variants (F87A/A328F and F87V/A328F) with a strong increase in regioselectivity. In contrast to the wild type enzyme both variants almost exclusively epoxidise at the C8-C9 double bond resulting in 94% and 97% (4*R*)-limonene-8,9-epoxide 8, respectively (Table 3). While F87A/A328F showed a slightly decreased NADPH turnover rate as compared to wild type, NADPH turnover in F87V/A328F was increased (Table 2). In total, considerably more (4*R*)-limonene 6 was converted to products by each of the two variant than by the wild type enzyme (Table 2).

Best (+)-valencene-converting variants

(+)-Valencene 11 is unselectively oxidised by CYP102A1 wild type enzyme to more than 4 products (Table 3). Only 29% of the products result from an oxidation at the C2 atom to (+)-nootkatol 13. (+)-Nootkatone 14, a possible product of further oxidation at C2 was not detected. The four CYP102A1 variants F87V/A328I, F87A/A328I, F87A/A328V, and F87V/A328V of our minimal library revealed 89%, 94%, 95%, and 86% preference for oxidation at the C2 atom, respectively. Thereof, variants F87A/A328I, F87V/A328I, and F87A/A328V showed the highest ratio of (+)-nootkatone 14 with 26%; 14%, and 11%, respectively, while F87V/A328V produced only 4% of (+)-nootkatone 14. These product distributions were measured in standard reactions containing the same amount of substrate and NADPH. An increase of the NADPH concentration led to increasing amounts of (+)-nootkatone 14 and decreasing amounts of (+)-nootkatol 13 for the four variants (up to 64 % for variant F87A/A328I). This indicates the need for two oxidation steps to arrive at (+)-nootkatone starting from (+)-valencene 11. The first step constitutes the oxidation of (+)-valencene 11 to (+)-nootkatol 13, which is then further oxidised by the CYP to (+)-nootkatone 14. This effect was smallest for variant F87V/A328V. This variant combines high regioselectivity and high stereoselectivity. F87V/A328V produces 82% (+)-nootkatol 13 with its *trans*-isomer 13 at an enantiomeric excess of 93%, while wild type produces only 29% (+)-nootkatol 13 with its *trans*-isomer at an enantiomeric excess of 79%. Variants F87V/A328V and F87V/A328I showed higher NADPH turnover rates than the wild type enzyme, while variants F87A/A328I and F87A/A328V showed decreased NADPH turnover rates (Table 2). For all four variants (+)-valencene 11 conversion was equal or higher than for the wild type.

Thereof, variants F87A/A328I and F87V/A328V showed the highest (+)-valencene conversion (Table 3). The fact that this effect is dependent on the CYP variant indicates that the second oxidation step is catalyzed by the CYP rather than other *E.coli* enzymes such as alcohol dehydrogenases.

3.3.5 Discussion

CYP102A1 from *Bacillus megaterium* is one of the most promising monooxygenases for the application in preparative synthesis. Hence, we chose this enzyme as a platform for the generation of a versatile toolbox of oxidation catalysts. Directed evolution by sequential rounds of random gene mutagenesis has been successfully applied to improve enzyme properties (Kuchner and Arnold 1997). All directed evolution approaches rely on an effective screening of huge combinatorial libraries. However, only a very small part of the theoretical library size can effectively be screened. Information about protein structure allows one to focus on a reduced number of positions leading to decreased library size. An approach that includes saturation mutagenesis at selected positions was previously applied to engineer CYP102A1 for enantioselective alkene epoxidation (Kubo et al. 2006). However, even in the presence of crystal structure information the number of potentially substrate interacting residues is quite high. Therefore, an exhaustive analysis of possible cooperative effects between the different positions is rarely feasible due to the large number of combinatorial possibilities. A stepwise improvement of properties can be achieved by iterative cycles of combinatorial active-site saturation mutagenesis, starting from different positions (Reetz et al. 2006). This strategy leads to a higher probability to identify cooperative effects between different positions.

The major limitations of all the mentioned strategies are: (1) only those combinations of mutations are found of which at least one of the underlying single mutants leads to improved properties; (2) for each substrate a suitable assay has to be developed to screen large mutant libraries efficiently; (3) the mutant libraries have to be at least partially reconstructed for each screening process. Hence, instead of generating new libraries for every new substrate we decided to generate a single minimal, highly enriched CYP102A1 mutant library which can be rapidly screened with different substrates. We focussed on two hotspot positions (87 and 328) which are located in the immediate vicinity of the activated oxygen and therefore are most likely in contact with every substrate in its reactive orientation, independent of its size

and shape. The generation of all possible combinations of 5 hydrophobic amino acids in both positions led to a great variety of substrate binding cavity shapes in the immediate vicinity of the heme (Figure 1). The screening of this minimal library with four differently sized and shaped terpenes identified variants with shifted or increased regio-, stereo- and chemoselectivity. It is widely accepted that increased regio- and stereoselectivity is the result of a reduced number of substrate orientations close to the heme (Raag and Poulos 1991; Bell et al. 2003; Branco et al. 2008). Here we show that the systematic variation of the size and shape of hydrophobic residues in the immediate vicinity of the heme can turn a preferentially epoxidating into an hydroxylating enzyme and vice versa, which provides evidence that besides regio- and stereoselectivity also chemoselectivity of CYP102A1 can be altered by mutations that induce changes in substrate orientations close to the heme centre. Some of the oxidation products described here are valuable and interesting compounds (e.g. (+)-nootkatone 14) and/or have been shown for the first time to be produced by CYPs (e.g. the hydroxy products 4, 5 of geranyl- and nerylacetone).

Previously, CYP102A1 wild type and mutants in 5 positions (including position 87) were screened with geranyl- and nerylacetone *(E)*-1, *(Z)*-1 (Watanabe et al. 2007). Geranylacetone *(E)*-1 was converted by a triple mutant to 9,10-epoxygeranylacetone *(E)*-2 with high regio- and stereoselectivity. The variant contained a substitution of phenyalanine in position 87 to valine. Consistently, our results show that substitutions of F87 by valine (but also by leucine and isoleucine) strongly increased the selectivity for the formation 9,10-epoxygeranylacetone *(E)*-2. However, the formation of hydroxy products in position C11 and C12 of the substrate was not catalysed by those mutants. Here we show that amino acid substitutions in position 328 extended the product spectra towards hydroxy products, while combined mutations in 87 and 328 led to a variant that almost exclusively produces hydroxy products with a pronounced regioselectivity for the C12 atom. The screening of our minimal library with nerylacetone *(Z)*-1 showed that position 328 had the strongest impact on regio- and chemoselectivity upon oxidation of this substrate. Only one mutant revealed a considerable increase in selectivity towards its *Z*-isomer nerylacetone *(Z)*-1, in contrast to geranylacetone *(E)*-1, although both substrates have a similar chemical reactivity.

(+)-Nootkatone 14 is a valuable oxidation product of (+)-valencene 11. Previously, a CYP102A1 mutant (R47L/Y51F/F87A) has been introduced which, in contrast to mutant R47L/Y51F and wild type, produces (+)-nootkatol 13 and (+)-nootkatone 14 in small amounts, however, the mutant also produced epoxides and other products (Sowden et al. 2005). This is in agreement with our results, showing mutant F87A to produce (+)-nootkatol

13 and (+)-nootkatone 14 among three additional oxidation products (C2-selectivity of 55%). This result confirms the important role of position 87 for substrate orientation close to the heme centre and hence for regioselectivity. However, double mutant F87A/A328I shows highly increased C2-regioselectivity (95%), which indicates a much stronger restriction of possible (+)-valencene 11 binding orientations by a combination of this mutation with a mutation in position 328. Also variants F87V/A328I, F87A/A328V and F87V/A328V showed strongly increased regioselectivity, hence different combinations of amino acids in these two positions can have similar effects. Interestingly, the single mutants A328V, and A328I showed no conversion of (+)-valencene 11, indicating that the effects of mutations in position 87 and 328 are non-additive but cooperative. This fact makes it impossible to predict the high selectivity of the double mutants based on the properties of the single mutants. An iterative approach would have found variants F87A/A328I, F87V/A328I, and F87V/A328V, because the single mutants F87A and F87V have slightly increased C2 regioselectivity as compared to wild type (data not shown). However, the preference for C12 hydroxylation of variant F87V/A328L could not be found by an iterative approach, since the respective single mutant F87V preferentially formed epoxides, and single mutant A328L was inactive towards geranylacetone.

In general, more single mutants than double mutants showed improved regio- and chemoselectivity towards the small acyclic terpenes neryl- and geranylacetone *(Z)*-1, *(E)*-1, while all variants with increased selectivity towards (4*R*)-limonene 6 and (+)-valencene 11 are double mutants. This demonstrates that combined mutations in both positions are required to improve selectivity towards the more bulky substrates (4*R*)-limonene 6 and (+)-valencene 11. It is instructive to analyse the three variants which were inactive towards all four terpene substrates. In variants F87/A328F, F87L/A328I and F87I/A328F, the amino acids side chains in the two positions are probably too bulky to permit heme access for the substrates used in this study.

Although it showed to be very effective to systematically combine only 5 hydrophobic amino acids in position 87 and 328, it might be promising to extent randomisation with serine, threonine, and glycine since these residues are also observed in CYP sequences at a position corresponding to position 328 in CYP102A1 (Seifert and Pleiss 2008). The compounds used to screen our minimal library have different shape, size, and polarity. Both hotspot positions also showed to influence selectivity of non-terpene substrates, hence we anticipate that this library is also a rich source for biocatalysts with improved selectivity towards non-terpene substrates. An ideal CYP catalyst would combine both a broad substrate specificity and high

selectivity. Our results indicate that these requirements can not be fulfilled by one variant only, but rather by a collection of variants. Our minimal library is a first step towards such an ideal CYP catalyst, since it provides a small collection of active variants with improved selectivity for different substrates. We expect that our approach is generic and can be applied to other CYPs. Position 87 can be identified in all CYPs with crystal structure information as well as in homologous CYPs. Position 328 can be identified from sequence alone in almost all CYPs, owing to its conserved distance from the highly conserved ExxR motif (Seifert and Pleiss 2008).

Acknowledgement

Financial support by the Deutsche Forschungsgemeinschaft (SFB 706), the Fonds der Chemischen Industrie, and the Ministerium für Wissenschaft, Forschung und Kunst des Landes Baden-Württemberg (Landesgraduierten fellowship for K. G.) are gratefully acknowledged. We would like to thank Dr. Alexandrine Busch for initial syntheses of reference compounds.

3.3.6 References

Almeida Q. A. R. and Jones J. (2005). "Chemoselective formation of 8,9-epoxy-limonene." *Synthetic Commun* 35(10): 1285-1290.

Bell S. G., Chen X. H., Sowden R. J., Xu F., Williams J. N., Wong L. L. and Rao Z. H. (2003). "Molecular recognition in (+)-alpha-pinene oxidation by cytochrome P450(cam)." *J Am Chem Soc* 125(3): 705-714.

Branco R. J., Seifert A., Budde M., Urlacher V. B., Ramos M. J. and Pleiss J. (2008). "Anchoring effects in a wide binding pocket: The molecular basis of regioselectivity in engineered cytochrome P450 monooxygenase from B. megaterium." *Proteins* 73(3): 597-607.

Bruder M. and Moody C. J. (2008). "Synthesis of (+/-)-smenochromene D (likonide B) using a regioselective Claisen rearrangement." *Synlett*(4): 575-577.

Carmichael A. B. and Wong L. L. (2001). "Protein engineering of Bacillus megaterium CYP102.The oxidation of polycyclic aromatic hydrocarbons." *Eur J Biochem* 268(10): 3117-25.

Charette A. B., Juteau H., Lebel H. and Deschenes D. (1996). "The chemo- and enantioselective cyclopropanation of polyenes: Chiral auxiliary vs chiral reagent-based approach." *Tetrahedron Lett* 37(44): 7925-7928.

Charette A. B., Juteau H., Lebel H. and Molinaro C. (1998). "Enantioselective cyclopropanation of allylic alcohols with dioxaborolane ligands: Scope and synthetic applications." *J Am Chem Soc* 120(46): 11943-11952.

Clark J. S., Myatt J., Roberts L. and Walshe N. (2005). "Investigation of the biomimetic synthesis of emindole SB using a fluorinated polyene cyclisation precursor." *Synlett*(4): 697-699.

DeLano W. L. (2002). The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA, USA.

Dietrich M., Eiben S., Asta C., Do T. A., Pleiss J. and Urlacher V. B. (2008). "Cloning, expression and characterisation of CYP102A7, a self-sufficient P450 monooxygenase from Bacillus licheniformis." *Appl Microbiol Biotechnol* 79(6): 931-940.

Furusawa M., Hashimoto T., Noma Y. and Asakawa Y. (2005). "Highly efficient production of nootkatone, the grapefruit aroma from valencene, by biotransformation." *Chem Pharm Bull* 53(11): 1513-1514.

Gotoh O. (1992). "Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences." *J Biol Chem* 267(1): 83-90.

Graham-Lorence S., Truan G., Peterson J. A., Falck J. R., Wei S., Helvig C. and Capdevila J. H. (1997). "An active site substitution, F87V, converts cytochrome P450 BM-3 into a regio- and stereoselective (14S,15R)-arachidonic acid epoxygenase." *J Biol Chem* 272(2): 1127-35.

Kubo T., Peters M. W., Meinhold P. and Arnold F. H. (2006). "Enantioselective epoxidation of terminal alkenes to (R)- and (S)-epoxides by engineered cytochromes P450 BM-3." *Chemistry* 12(4): 1216-20.

Kuchner O. and Arnold F. H. (1997). "Directed evolution of enzyme catalysts." *Trends Biotechnol* 15(12): 523-530.

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

Li H. M., Mei L. H., Urlacher V. B. and Schmid R. D. (2008). "Cytochrome P450BM-3 evolved by random and saturation mutagenesis as an effective indole-hydroxylating catalyst." *Appl Biochem Biotech* 144(1): 27-36.

Li Y., Li W. D. and Li Y. L. (1994). "An Alternative Short Synthesis of 6,10,14-Trimethyl-5e,9e-Pentadecadiene-2,13-Dione." *Synthetic Commun* 24(1): 117-121.

Lucker J., Schwab W., Franssen M. C., Van Der Plas L. H., Bouwmeester H. J. and Verhoeven H. A. (2004). "Metabolic engineering of monoterpene biosynthesis: two-step production of (+)-trans-isopiperitenol by tobacco." *Plant J* 39(1): 135-45.

Malerich J. P. and Trauner D. (2003). "Biomimetic synthesis of (+/-)-pinnatal and (+/-)-sterekunthal A." *J Am Chem Soc* 125(32): 9554-9555.

Marshall J. A. and Dubay W. J. (1994). "Synthesis of the Pseudopterane and Furanocembrane Ring-Systems by Intraannular Cyclization of Beta-Alkynyl and Gamma-Alkynyl Allylic Alcohols." *J Org Chem* 59(7): 1703-1708.

Maurer S. C., Schulze H., Schmid R. D. and Urlacher V. (2003). "Immobilisation of P450BM-3 and an NADP(+) cofactor recycling system: Towards a technical application of heme-containing monooxygenases in fine chemical synthesis." *Adv Synth Catal* 345(6-7): 802-810.

Mcmurry J. E. and Dushin R. G. (1989). "Total Synthesis of (+/-)-Crassin by Titanium-Induced Pinacol Coupling." *J Am Chem Soc* 111(24): 8928-8929.

Mcmurry J. E. and Dushin R. G. (1990). "Total Synthesis of (+/-)-Isolobophytolide and (+/-)-Crassin by Titanium-Induced Carbonyl Coupling." *J Am Chem Soc* 112(19): 6942-6949.

Mcmurry J. E. and Kocovsky P. (1984). "A Method for the Palladium-Catalyzed Allylic Oxidation of Olefins." *Tetrahedron Lett* 25(38): 4187-4190.

Mcmurry J. E. and Kocovsky P. (1985). "Synthesis of Helminthogermacrene and Beta-Elemene." *Tetrahedron Lett* 26(18): 2171-2172.

Mcmurry J. E., Matz J. R. and Kees K. L. (1987). "Synthesis of Macrocyclic Terpenoids by Intramolecular Carbonyl Coupling - Flexibilene and Humulene." *Tetrahedron* 43(23): 5489-5498.

Meinhold P., Peters M. W., Hartwick A., Hernandez A. R. and Arnold F. H. (2006). "Engineering cytochrome P450BM3 for terminal alkane hydroxylation." *Adv Synth Catal* 348(6): 763-772.

Munro A. W., Leys D. G., McLean K. J., Marshall K. R., Ost T. W. B., Daff S., Miles C. S., Chapman S. K., Lysek D. A., Moser C. C., Page C. C. and Dutton P. L. (2002). "P450BM3: the very model of a modern flavocytochrome." *Trends Biochem Sci* 27(5): 250-257.

Muto S., Nishimura Y. and Mori K. (1999). "Pheromone synthesis CXCVI - Synthesis of germacrene-B and its extension to the synthesis of (+/-)-9-methylgermacrene-B, the racemate of the male-produced sex pheromone of the sandfly Lutzomyia longipalpis from Lapinha, Brazil." *Eur J Org Chem*(9): 2159-2165.

Narhi L. O. and Fulco A. J. (1987). "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* 262(14): 6683-90.

NIST 05 Mass Spectral Library, National Institute of Standards and Technology, Gaithersburg, Maryland, USA.

Omura T. and Sato R. J. (1964). "The Carbon Monoxide-binding Pigment of Liver Microsomes. I Evidence for its Hemoprotein Nature." *J Biol Chem* 239: 2370-2378.

Raag R. and Poulos T. L. (1991). "Crystal-Structures of Cytochrome-P-450cam Complexed with Camphane, Thiocamphor, and Adamantane - Factors Controlling P-450 Substrate Hydroxylation." *Biochemistry* 30(10): 2674-2684.

Reetz M. T., Wang L. W. and Bocola M. (2006). "Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space." *Angew Chem Int Edit* 45(8): 1236-41.

Seifert A. and Pleiss J. (2008). "Identification of selectivity-determining residues in cytochrome P450 monooxygenases: A systematic analysis of the substrate recognition site 5." *Proteins* (DOI: 10.1002/prot.22242)

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

Sowden R. J., Yasmin S., Rees N. H., Bell S. G. and Wong L. L. (2005). "Biotransformation of the sesquiterpene (+)-valencene by cytochrome P450cam and P450BM-3." *Org Biomol Chem* 3(1): 57-64.

Still W. C. and Gennari C. (1983). "Direct Synthesis of Z-Unsaturated Esters - a Useful Modification of the Horner-Emmons Olefination." *Tetrahedron Lett* 24(41): 4405-4408.

Tago K., Arai M. and Kogen H. (2000). "A practical total synthesis of plaunotol via highly Z-selective Wittig olefination of alpha-acetal ketones." *J Chem Soc Perk T 1*(13): 2073-2078.

Takahashi S., Yeo Y. S., Zhao Y. X., O'Maille P. E., Greenhagen B. T., Noel J. P., Coates R. M. and Chappell J. (2007). "Functional characterization of premnaspirodiene oxygenase, a cytochrome P450 catalyzing regio- and stereo-specific hydroxylations of diverse sesquiterpene substrates." *J Biol Chem* 282(43): 31744-31754.

Tsunoda T., Suzuki M. and Noyori R. (1980). "Trialkylsilyl Triflates .6. Facile Procedure for Acetalization under Aprotic Conditions." *Tetrahedron Lett* 21(14): 1357-1358.

Urlacher V. B., Lutz-Wahl S. and Schmid R. D. (2004). "Microbial P450 enzymes in biotechnology." *Appl Microbiol Biotechnol* 64(3): 317-325.

Urlacher V. B., Makhsumkhanov A. and Schmid R. D. (2006). "Biotransformation of beta-ionone by engineered cytochrome P450 BM-3." *Appl Microbiol Biotechnol* 70(1): 53-9.

van der Werf M. J., Keijzer P. M. and van der Schaft P. H. (2001). "Xanthobacter sp. C20 contains a novel bioconversion pathway for limonene." *J Biotechnol* 84(2): 133-43.

Watanabe Y., Laschat S., Budde M., Affolter O., Shimada Y. and Urlacher V. B. (2007). "Oxidation of acyclic monoterpenes by P450 BM-3 monooxygenase: influence of the substrate E/Z-isomerism on enzyme chemo- and regioselectivity." *Tetrahedron* 63(38): 9413-9422.

Tables

Table 1: Comparison of terpene oxidation catalysed by CYP102A1 variants designed in this study. Black cells indicate variants that show the strongest shifts in regio- or chemoselectivity, or have the highest regio- or stereoselectivity as compared to wild type. Enzyme-substrate combinations without or only a minor improvement as compared to wild type are depicted in white. Grey cells indicate very little or no conversion of the respective substrates.

CYP102A1 variants	(4*R*)-(+)-limonene	Geranylacetone	Nerylacetone	(+)-Valencene
F87A / A328				
F87A / A328I				
F87A / A328L				
F87A / A328V				
F87A / A328F				
F87I / A328				
F87I / A328I				
F87I / A328L				
F87I / A328V				
F87I / A328F				
F87L / A328				
F87L / A328I				
F87L / A328L				
F87L / A328V				
F87L / A328F				
F87V / A328				
F87V / A328I				
F87V / A328L				
F87V / A328V				
F87V / A328F				
F87 / A328 *				
F87 / A328I				
F87 / A328L				
F87 / A328V				
F87 / A328F				

* CYP102A1 wild type

Table 2: Comparison of activity of CYP102A1 wild type and the most selective variants (Table 3) towards (4*R*)-(+)-limonene **6**, geranylacetone *(E)*-**1**, nerylacetone *(Z)*-**1**, and (+)-valencene **11**. The substrate oxidation rate is defined as $a/(c \cdot t)$ [µmol · (µmol CYP)$^{-1}$ · min^{-1}], with converted substrate *a* [µM], enzyme concentration c [µM], and time t to consume the NADPH [min].

CYP102A1 variants	Substrate	NADPH turnover (1/min)	Oxidation rate (1/min)	Conversion after 1h [%]
Wild type	**6**	112±7.3	54.7±4.6	48.7±1
F87A / A328F	**6**	88±12	60±7.4	68±1.3
F87V / A328F	**6**	277±8.3	181±4.2	65.3±1
Wild type	*(E)*-**1**	478±38	324±17	67.9±2
F87V	*(E)*-**1**	133±6	83±3	62.6±2.3
F87I	*(E)*-**1**	64±9	33±4	51.1±2.6
F87L	*(E)*-**1**	17±4	6±1	38.2±1.3
A328V	*(E)*-**1**	1145±105	550±72	48.1±4.4
F87V / A328L	*(E)*-**1**	300±23	152±11	50±3.8
Wild type	*(Z)*-**1**	301±27.4	155±6.9	51.6±2.6
A328V	*(Z)*-**1**	594±24.7	338±9.8	56.9±1
Wild type	**11**	43±3.3	10±2	22.9±3.1
F87A / A328I	**11**	25.3±5.6	12.8±3.7	50.3±3.9
F87V/ A328I	**11**	54.4±10.3	15.7±3.3	28.7±1.7
F87A / A328V	**11**	17.4±3.6	4.2±0.7	24.6±5.3
F87V / A328V	**11**	63.5±4.2	24.2±2.5	38.3±5.0

Table 3: Observed products of a) (4*R*)-(+)-limonene 6, b) geranylacetone *(E)*-1, c) nerylacetone *(Z)*-1, and d) (+)-valencene 11, converted by CYP102A1 wild type and the most selective variants. R_t indicates the GC retention time (in min) of unknown products.

CYP102A1 variants	Products in [%]											
a)	**7**	**8**	**10**	**9**	**R_t 73.5**	**R_t 74.8**						
Wild type	30.1±0.3	6.7±1.2	53.7±1	9.4±0.1								
F87A / A328F		94±0.2				3.5±0.6						
F87V / A328F	1.2±.3	97±0.2			1.9±0.1							
b)	***(E)*-3**	***(E)*-2**	**R_t 12.3**	**R_t 12.8**	**R_t 13.9**	***(E)*-5**	***(E)*-4**					
Wild type		54±0.2	46±0.2									
F87V	0.9±0.1	87.4±0.4	6.9±0.3	4.1±0.1			0.8±0.1					
F87I		84.9±0.6	7.4±0.1	1.0±0.1	2.1±0.2	3.0±0.2	1.6±0.2					
F87L		97.2±0.2	2.8±0.2									
A328V		30.1±0.4	2.0±0.2			31±0.8	36.9±0.2					
F87V / A328L		5.3±0.2				80±0.3	14.7±0.4					
c)	***(Z)*-3**	**R_t 8.7**	**R_t 9.2**	**R_t 9.7**	***(Z)*-2**	**R_t 10.5**	**R_t 11.8**	***(Z)*-5**	**R_t 13.9**	***(Z)*-4**		
Wild type	11.4±0.4	0.9±0.1	1.1±0.1	1.7±0.1	46.9±0.6	3.2±0.6	22.4±0.1	1.1±0.1		11.3±0.6		
A328V	0.7±0.1							33.5±0.2	2.1±0.3	63.7±0.3		
d)	**R_t 9.1**	**R_t 9.9**	**12**	**13**	**R_t 10.4**	**R_t 10.6**	**14**	**R_t 12.2**	**R_t 12.4**	**R_t 12.6**	**R_t 13.4**	**R_t 13.6**
Wild type	3.8±0.1	18.5±0.9	3.0±0.2	26.3±1				0.9±0.5	10.5±0.6	37.1±1.8		
F87A / A328I		0.9±0.4	54.8±1.7	12.6±0.3		0.7±0.2	26.1±0.9	0.5±0.2	1.9±0.1	1.0±0.2	0.7±0.1	0.9±0.3
F87V / A328I		6.2±0.2	13.7±1.5	61.4±1.2			13.6±1.1	0.8±0.6	1.2±0.8	2.3±0.1	0.2±0.2	0.6±0.6
F87A / A328V		2.0±0.3	62.4±0.2	22.2±0.3	2.8±0.8		10.7±0.5					
F87V/ A328V		8.9±0.6	2.9±0.4	79.5±1			3.9±0.3		1.3±0.0	3.5±0.5		

Figures

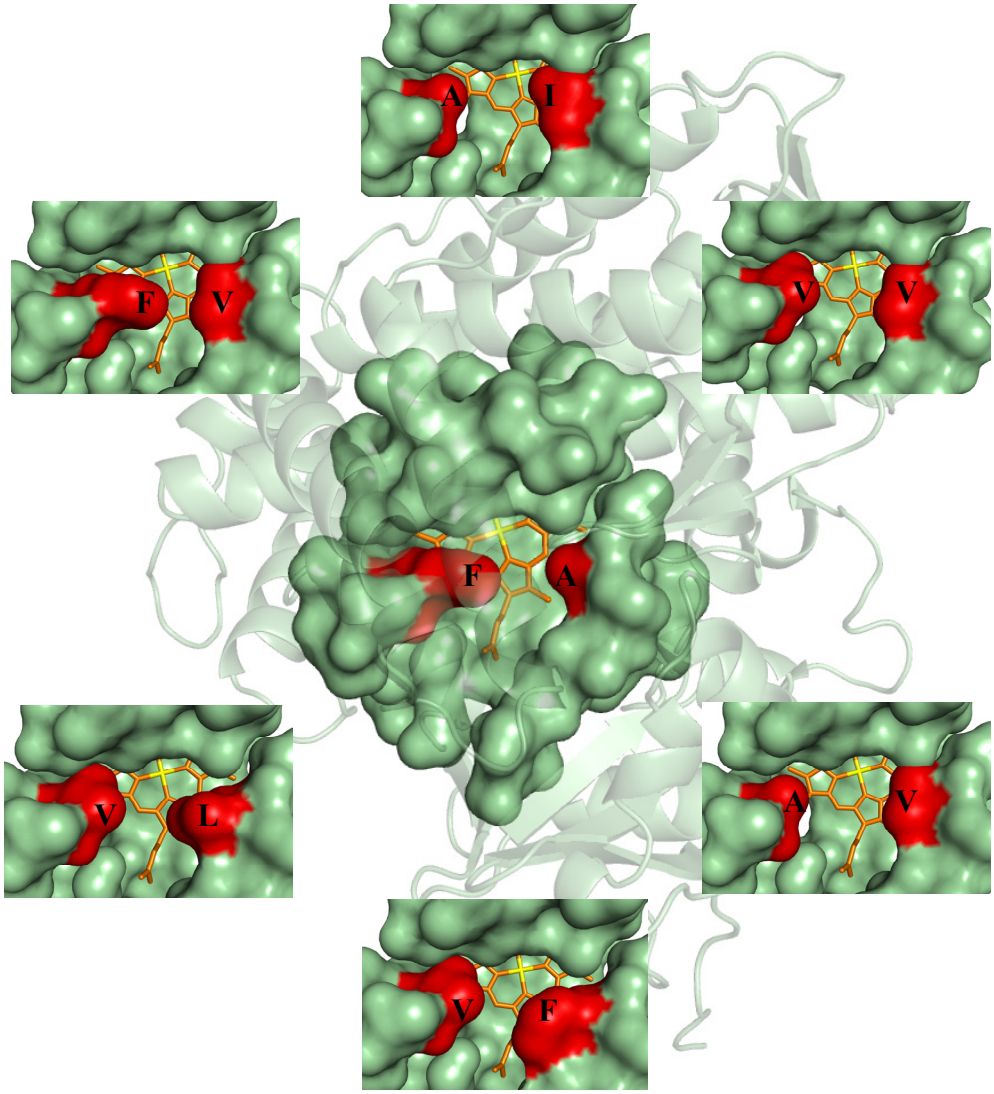


Figure 1. Shape of the substrate binding site in CYP102A1 wild type (centre) near to the active heme (yellow). The mutated positions 87 (left) and 328 (right) are shown in red. The high diversity of substrate binding site shapes of 6 highly selective variants A328V (FV), F87A/A328I (AI), F87V/A328V (VV), F87A/A328V (AV), F87V/A328F (VF), and F87V/A328L (VL) are visualised in the insets. Models are based on CYP102A1 crystal structure (PDB entry 1bu7 chain A) and were generated with the Pymol 0.99 program (DeLano 2002).

4 Gesamtliteraturverzeichnis

Afzelius L., Raubacher F., Karlen A., Jorgensen F. S., Andersson T. B., Masimirembwa C. M. and Zamora I. (2004). "Structural analysis of CYP2C9 and CYP2C5 and an evaluation of commonly used molecular modeling techniques." *Drug Metab Dispos* 32(11): 1218-29.

Bartolucci C., Perola E., Cellai L., Brufani M. and Lamba D. (1999). ""Back door" opening implied by the crystal structure of a carbamoylated acetylcholinesterase." *Biochemistry* 38(18): 5714-9.

Bell S. G., Chen X., Sowden R. J., Xu F., Williams J. N., Wong L. L. and Rao Z. (2003). "Molecular recognition in (+)-alpha-pinene oxidation by cytochrome P450cam." *J Am Chem Soc* 125(3): 705-14.

Born S. L., John G. H., Harlow G. R. and Halpert J. R. (1995). "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* 23(7): 702-7.

Branco R. J., Seifert A., Budde M., Urlacher V. B., Ramos M. J. and Pleiss J. (2008). "Anchoring effects in a wide binding pocket: The molecular basis of regioselectivity in engineered cytochrome P450 monooxygenase from B. megaterium." *Proteins* 73(3): 597-607.

Bruntner C., Lauer B., Schwarz W., Mohrle V. and Bormann C. (1999). "Molecular characterization of co-transcribed genes from Streptomyces tendae Tu901 involved in the biosynthesis of the peptidyl moiety of the peptidyl nucleoside antibiotic nikkomycin." *Mol Gen Genet* 262(1): 102-14.

Carmichael A. B. and Wong L. L. (2001). "Protein engineering of Bacillus megaterium CYP102.The oxidation of polycyclic aromatic hydrocarbons." *Eur J Biochem* 268(10): 3117-25.

Charette A. B., Juteau H., Lebel H. and Deschenes D. (1996). "The chemo- and enantioselective cyclopropanation of polyenes: Chiral auxiliary vs chiral reagent-based approach." *Tetrahedron Lett* 37(44): 7925-7928.

Charette A. B., Juteau H., Lebel H. and Molinaro C. (1998). "Enantioselective cyclopropanation of allylic alcohols with dioxaborolane ligands: Scope and synthetic applications." *J Am Chem Soc* 120(46): 11943-11952.

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

Clark J. S., Myatt J., Roberts L. and Walshe N. (2005). "Investigation of the biomimetic synthesis of emindole SB using a fluorinated polyene cyclisation precursor." *Synlett*(4): 697-699.

Cosme J. and Johnson E. F. (2000). "Engineering microsomal cytochrome P450 2C5 to be a soluble, monomeric enzyme. Mutations that alter aggregation, phospholipid dependence of catalysis, and membrane binding." *J Biol Chem* 275(4): 2545-53.

de Visser S. P., Kumar D., Cohen S., Shacham R. and Shaik S. (2004). "A predictive pattern of computed barriers for C-h hydroxylation by compound I of cytochrome p450." *J Am Chem Soc* 126(27): 8362-3.

de Visser S. P., Ogliaro F., Sharma P. K. and Shaik S. (2002). "What factors affect the regioselectivity of oxidation by cytochrome p450? A DFT study of allylic hydroxylation and double bond epoxidation in a model reaction." *J Am Chem Soc* 124(39): 11809-26.

Denisov I. G., Makris T. M., Sligar S. G. and Schlichting I. (2005). "Structure and chemistry of cytochrome P450." *Chem Rev* 105(6): 2253-77.

Dietrich M., Eiben S., Asta C., Do T. A., Pleiss J. and Urlacher V. B. (2008). "Cloning, expression and characterisation of CYP102A7, a self-sufficient P450 monooxygenase from Bacillus licheniformis." *Appl Microbiol Biotechnol* 79(6): 931-940.

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

Fruetel J. A., Mackman R. L., Peterson J. A. and Ortiz de Montellano P. R. (1994). "Relationship of active site topology to substrate specificity for cytochrome P450terp (CYP108)." *J Biol Chem* 269(46): 28815-21.

Furusawa M., Hashimoto T., Noma Y. and Asakawa Y. (2005). "Highly efficient production of nootkatone, the grapefruit aroma from valencene, by biotransformation." *Chem Pharm Bull* 53(11): 1513-1514.

Gilson M. K., Straatsma T. P., McCammon J. A., Ripoll D. R., Faerman C. H., Axelsen P. H., Silman I. and Sussman J. L. (1994). "Open "back door" in a molecular dynamics simulation of acetylcholinesterase." *Science* 263(5151): 1276-8.

Gotoh O. (1992). "Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences." *J Biol Chem* 267(1): 83-90.

Graham-Lorence S., Truan G., Peterson J. A., Falck J. R., Wei S., Helvig C. and Capdevila J. H. (1997). "An active site substitution, F87V, converts cytochrome P450 BM-3 into a regio- and stereoselective (14S,15R)-arachidonic acid epoxygenase." *J Biol Chem* 272(2): 1127-35.

Guengerich F. P. (2001). "Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity." *Chem Res Toxicol* 14(6): 611-50.

Guengerich F. P., Krauser J. A. and Johnson W. W. (2004). "Rate-limiting steps in oxidations catalyzed by rabbit cytochrome P450 1A2." *Biochemistry* 43(33): 10775-88.

He Y. A., He Y. Q., Szklarz G. D. and Halpert J. R. (1997). "Identification of three key residues in substrate recognition site 5 of human cytochrome P450 3A4 by cassette and site-directed mutagenesis." *Biochemistry* 36(29): 8831-9.

Howard-Jones A. R. and Walsh C. T. (2007). "Nonenzymatic oxidative steps accompanying action of the cytochrome p450 enzymes StaP and RebP in the biosynthesis of staurosporine and rebeccamycin." *J Am Chem Soc* 129(36): 11016-7.

Hummel M. A., Gannett P. M., Aguilar J. S. and Tracy T. S. (2004). "Effector-mediated alteration of substrate orientation in cytochrome P450 2C9." *Biochemistry* 43(22): 7207-14.

Ioannides C. and Lewis D. F. (2004). "Cytochromes P450 in the bioactivation of chemicals." *Curr Top Med Chem* 4(16): 1767-88.

Isin E. M. and Guengerich F. P. (2007). "Complex reactions catalyzed by cytochrome P450 enzymes." *Biochim Biophys Acta* 1770(3): 314-29.

J. Gasteiger E. G. J., M. G. Hicks, J. Sunkel (1988). "Empirical Methods for the Calculation of Physicochemical Data of Organic Compounds." *Physical Property Prediction in Organic Chemistry Springer Verlag, Heidelberg*: 119-138.

Juchau M. R. (1990). "Substrate specificities and functions of the P450 cytochromes." *Life Sci* 47(26): 2385-94.

Karki S. B. and Dinnocenzo J. P. (1995). "On the mechanism of amine oxidations by P450." *Xenobiotica* 25(7): 711-24.

Karuzina, II and Archakov A. I. (1994). "Hydrogen peroxide-mediated inactivation of microsomal cytochrome P450 during monooxygenase reactions." *Free Radic Biol Med* 17(6): 557-67.

Karuzina, II and Archakov A. I. (1994). "The oxidative inactivation of cytochrome P450 in monooxygenase reactions." *Free Radic Biol Med* 16(1): 73-97.

Katagiri M., Ganguli B. N. and Gunsalus I. C. (1968). "A soluble cytochrome P-450 functional in methylene hydroxylation." *J Biol Chem* 243(12): 3543-6.

Keizers P. H., Lussenburg B. M., de Graaf C., Mentink L. M., Vermeulen N. P. and Commandeur J. N. (2004). "Influence of phenylalanine 120 on cytochrome P450 2D6 catalytic selectivity and regiospecificity: crucial role in 7-methoxy-4-(aminomethyl)-coumarin metabolism." *Biochem Pharmacol* 68(11): 2263-71.

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

Krahn M. "Die P450 Engineering Database - systematische Analyse familienspezifischer Eigenschaften und der Sequenz-Struktur-Funktionsbeziehungen von cytochrom P450 Monooxygenasen." Diplomarbeit, Universität Stuttgart, Stuttgart, 2005.

Kubo T., Peters M. W., Meinhold P. and Arnold F. H. (2006). "Enantioselective epoxidation of terminal alkenes to (R)- and (S)-epoxides by engineered cytochromes P450 BM-3." *Chemistry* 12(4): 1216-20.

Kuchner O. and Arnold F. H. (1997). "Directed evolution of enzyme catalysts." *Trends Biotechnol* 15(12): 523-530.

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

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

Li H. M., Mei L. H., Urlacher V. B. and Schmid R. D. (2008). "Cytochrome P450BM-3 evolved by random and saturation mutagenesis as an effective indole-hydroxylating catalyst." *Appl Biochem Biotech* 144(1): 27-36.

Li Q. S., Schwaneberg U., Fischer M., Schmitt J., Pleiss J., Lutz-Wahl S. and Schmid R. D. (2001). "Rational evolution of a medium chain-specific cytochrome P-450 BM-3 variant." *Bba-Protein Struct M* 1545(1-2): 114-121.

Li Y., Li W. D. and Li Y. L. (1994). "An Alternative Short Synthesis of 6,10,14-Trimethyl-5e,9e-Pentadecadiene-2,13-Dione." *Synthetic Commun* 24(1): 117-121.

Liu J., Ericksen S. S., Sivaneri M., Besspiata D., Fisher C. W. and Szklarz G. D. (2004). "The effect of reciprocal active site mutations in human cytochromes P450 1A1 and 1A2 on alkoxyresorufin metabolism." *Arch Biochem Biophys* 424(1): 33-43.

Lupien S., Karp F., Wildung M. and Croteau R. (1999). "Regiospecific cytochrome P450 limonene hydroxylases from mint (Mentha) species: cDNA isolation, characterization, and functional expression of (-)-4S-limonene-3-hydroxylase and (-)-4S-limonene-6-hydroxylase." *Arch Biochem Biophys* 368(1): 181-92.

Macarthur M. W. and Thornton J. M. (1991). "Influence of Proline Residues on Protein Conformation." *J Mol Biol* 218(2): 397-412.

Makris T. M. D., I. G.; Schlichting, I.; Sligar, S. G. (2005). "Activation of Molecular Oxygen by Cytochrome P450." *In Cytochrome P450 : Structure, Mechanism, and Biochemistry* Ortiz de Montellano, P. R., Ed. Kluwer Academic/Plenum Publishers: New York: pp 149-182.

Marshall J. A. and Dubay W. J. (1994). "Synthesis of the Pseudopterane and Furanocembrane Ring-Systems by Intraannular Cyclization of Beta-Alkynyl and Gamma-Alkynyl Allylic Alcohols." *J Org Chem* 59(7): 1703-1708.

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. and Munro A. W. (2005). "Biodiversity of cytochrome P450 redox systems." *Biochem Soc Trans* 33(Pt 4): 796-801.

McLean M. A., Maves S. A., Weiss K. E., Krepich S. and Sligar S. G. (1998). "Characterization of a cytochrome P450 from the acidothermophilic archaea Sulfolobus solfataricus." *Biochem Biophys Res Commun* 252(1): 166-72.

Mcmurry J. E. and Kocovsky P. (1984). "A Method for the Palladium-Catalyzed Allylic Oxidation of Olefins." *Tetrahedron Lett* 25(38): 4187-4190.

Mcmurry J. E. and Kocovsky P. (1985). "Synthesis of Helminthogermacrene and Beta-Elemene." *Tetrahedron Lett* 26(18): 2171-2172.

Mcmurry J. E., Matz J. R. and Kees K. L. (1987). "Synthesis of Macrocyclic Terpenoids by Intramolecular Carbonyl Coupling - Flexibilene and Humulene." *Tetrahedron* 43(23): 5489-5498.

Meinhold P., Peters M. W., Hartwick A., Hernandez A. R. and Arnold F. H. (2006). "Engineering cytochrome P450BM3 for terminal alkane hydroxylation." *Adv Synth Catal* 348(6): 763-772.

Melet A., Assrir N., Jean P., Pilar Lopez-Garcia M., Marques-Soares C., Jaouen M., Dansette P. M., Sari M. A. and Mansuy D. (2003). "Substrate selectivity of human cytochrome P450 2C9: importance of residues 476, 365, and 114 in recognition of diclofenac and sulfaphenazole and in mechanism-based inactivation by tienilic acid." *Arch Biochem Biophys* 409(1): 80-91.

Mestres J. (2005). "Structure conservation in cytochromes P450." *Proteins* 58(3): 596-609.

Meunier B., de Visser S. P. and Shaik S. (2004). "Mechanism of oxidation reactions catalyzed by cytochrome p450 enzymes." *Chem Rev* 104(9): 3947-80.

Munro A. W., Girvan H. M. and McLean K. J. (2007). "Cytochrome P450--redox partner fusion enzymes." *Biochim Biophys Acta* 1770(3): 345-59.

Munro A. W., Leys D. G., McLean K. J., Marshall K. R., Ost T. W. B., Daff S., Miles C. S., Chapman S. K., Lysek D. A., Moser C. C., Page C. C. and Dutton P. L. (2002). "P450BM3: the very model of a modern flavocytochrome." *Trends Biochem Sci* 27(5): 250-257.

Narhi L. O. and Fulco A. J. (1987). "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* 262(14): 6683-90.

Nebert D. W. and Nelson D. R. (1991). "P450 gene nomenclature based on evolution." *Methods Enzymol* 206: 3-11.

Nebert D. W., Nelson D. R., Adesnik M., Coon M. J., Estabrook R. W., Gonzalez F. J., Guengerich F. P., Gunsalus I. C., Johnson E. F., Kemper B. and et al. (1989). "The P450 superfamily: updated listing of all genes and recommended nomenclature for the chromosomal loci." *DNA* 8(1): 1-13.

Nelson D. R. (2006). "Cytochrome P450 nomenclature, 2004." *Methods Mol Biol* 320: 1-10.

Nelson D. R., Kamataki T., Waxman D. J., Guengerich F. P., Estabrook R. W., Feyereisen R., Gonzalez F. J., Coon M. J., Gunsalus I. C., Gotoh O. and et al. (1993). "The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature." *DNA Cell Biol* 12(1): 1-51.

Nelson D. R., Koymans L., Kamataki T., Stegeman J. J., Feyereisen R., Waxman D. J., Waterman M. R., Gotoh O., Coon M. J., Estabrook R. W., Gunsalus I. C. and Nebert D. W. (1996). "P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature." *Pharmacogenetics* 6(1): 1-42.

Ngui J. S., Chen Q., Shou M., Wang R. W., Stearns R. A., Baillie T. A. and Tang W. (2001). "In vitro stimulation of warfarin metabolism by quinidine: increases in the formation of 4'- and 10-hydroxywarfarin." *Drug Metab Dispos* 29(6): 877-86.

Omura T. and Sato R. J. (1964). "The Carbon Monoxide-binding Pigment of Liver Microsomes. I Evidence for its Hemoprotein Nature." *J Biol Chem* 239: 2370-2378.

Oprea T. I., Hummer G. and Garcia A. E. (1997). "Identification of a functional water channel in cytochrome P450 enzymes." *Proc Natl Acad Sci U S A* 94(6): 2133-8.

Raag R. and Poulos T. L. (1991). "Crystal-Structures of Cytochrome-P-450cam Complexed with Camphane, Thiocamphor, and Adamantane - Factors Controlling P-450 Substrate Hydroxylation." *Biochemistry* 30(10): 2674-2684.

Ravichandran K. G., Boddupalli S. S., Hasermann C. A., Peterson J. A. and Deisenhofer J. (1993). "Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's." *Science* 261(5122): 731-6.

Reetz M. T., Wang L. W. and Bocola M. (2006). "Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space." *Angew Chem Int Edit* 45(8): 1236-41.

Rendic S. and Di Carlo F. J. (1997). "Human cytochrome P450 enzymes: a status report summarizing their reactions, substrates, inducers, and inhibitors." *Drug Metab Rev* 29(1-2): 413-580.

Rettie A. E., Korzekwa K. R., Kunze K. L., Lawrence R. F., Eddy A. C., Aoyama T., Gelboin H. V., Gonzalez F. J. and Trager W. F. (1992). "Hydroxylation of warfarin by human cDNA-expressed cytochrome P-450: a role for P-4502C9 in the etiology of (S)-warfarin-drug interactions." *Chem Res Toxicol* 5(1): 54-9.

Schalk M. and Croteau R. (2000). "A single amino acid substitution (F363I) converts the regiochemistry of the spearmint (-)-limonene hydroxylase from a C6- to a C3-hydroxylase." *Proc Natl Acad Sci U S A* 97(22): 11948-53.

Schleinkofer K., Sudarko, Winn P. J., Ludemann S. K. and Wade R. C. (2005). "Do mammalian cytochrome P450s show multiple ligand access pathways and ligand channelling?" *EMBO Rep* 6(6): 584-9.

Scott E. E., He Y. A., Wester M. R., White M. A., Chin C. C., Halpert J. R., Johnson E. F. and Stout C. D. (2003). "An open conformation of mammalian cytochrome P450 2B4 at 1.6-A resolution." *Proc Natl Acad Sci U S A* 100(23): 13196-201.

Seifert A. "Molekulardynamische Untersuchungen zu Substratspezifität der humanen cytochrom P450 Monooxygenase 2C9." Diplomarbeit, Universität Stuttgart, Stuttgart, 2004

Seth-Smith H. M., Rosser S. J., Basran A., Travis E. R., Dabbs E. R., Nicklin S. and Bruce N. C. (2002). "Cloning, sequencing, and characterization of the hexahydro-1,3,5-Trinitro-1,3,5-triazine degradation gene cluster from Rhodococcus rhodochrous." *Appl Environ Microbiol* 68(10): 4764-71.

Shaik S., Kumar D., de Visser S. P., Altun A. and Thiel W. (2005). "Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes." *Chem Rev* 105(6): 2279-328.

Sherman D. H., Li S., Yermalitskaya L. V., Kim Y., Smith J. A., Waterman M. R. and Podust L. M. (2006). "The structural basis for substrate anchoring, active site selectivity, and product formation by P450 PikC from Streptomyces venezuelae." *J Biol Chem* 281(36): 26289-97.

Smith D. A., Abel S. M., Hyland R. and Jones B. C. (1998). "Human cytochrome P450s: selectivity and measurement in vivo." *Xenobiotica* 28(12): 1095-128.

Smith D. A., Ackland M. J. and Jones B. C. (1997). "Properties of cytochrome P450 isoenzymes and their substrates Part 1: active site characteristics." *Drug Discovery Today* 2(10): 406-414.

Sowden R. J., Yasmin S., Rees N. H., Bell S. G. and Wong L. L. (2005). "Biotransformation of the sesquiterpene (+)-valencene by cytochrome P450cam and P450BM-3." *Org Biomol Chem* 3(1): 57-64.

Stresser D. M., Turner S. D. Ackermann J. M., Miller V. P. and Crespi C. L. "Flourometric cytochromes P450 2C8, 2C9 and 2C19 inhibition assays" (http://www.bdbiosciences.com/discovery_labware/gentest/products/pdf/post_015.pdf)

Urlacher V. B., Lutz-Wahl S. and Schmid R. D. (2004). "Microbial P450 enzymes in biotechnology." *Appl Microbiol Biotechnol* 64(3): 317-325.

Urlacher V. B., Makhsumkhanov A. and Schmid R. D. (2006). "Biotransformation of beta-ionone by engineered cytochrome P450 BM-3." *Appl Microbiol Biotechnol* 70(1): 53-9.

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

Wade C. W. W., P.J. Schlichtling, I. Sudarko, (2004). "A survey of active site access channels in cytochromes P450." *J Inorg Biochem* 98: 1175-1182.

Watanabe Y., Laschat S., Budde M., Affolter O., Shimada Y. and Urlacher V. B. (2007). "Oxidation of acyclic monoterpenes by P450 BM-3 monooxygenase: influence of the substrate E/Z-isomerism on enzyme chemo- and regioselectivity." *Tetrahedron* 63(38): 9413-9422.

Werck-Reichhart D. and Feyereisen R. (2000). "Cytochromes P450: a success story." *Genome Biol* 1(6): REVIEWS3003.

Wester M. R., Yano J. K., Schoch G. A., Yang C., Griffin K. J., Stout C. D. and Johnson E. F. (2004). "The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-A resolution." *J Biol Chem* 279(34): 35630-7.

Williams P. A., Cosme J., Sridhar V., Johnson E. F. and McRee D. E. (2000). "Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity." *Mol Cell* 5(1): 121-31.

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

Winn P. J., Ludemann S. K., Gauges R., Lounnas V. and Wade R. C. (2002). "Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine." *Proc Natl Acad Sci U S A* 99(8): 5361-6.

Danksagungen

Ich danke Herrn Prof. Dr. Rolf D. Schmid, MBA für die Möglichkeit, diese Arbeit unter exzellenten Arbeitsbedingungen an seinem Institut durchführen zu können.

Herrn Prof. Dr. Jürgen Pleiss danke ich für die Überlassung des Themas und die ständige Unterstützung und Motivation.

Meinen zeitweiligen wissenschaftlichen Hilfskräften Bettina Überle, Sandra Vomund und Mihaela Antonovici danke ich für die Assistenz bei der Generierung der Enzymvarianten sowie deren biochemischen Charakterisierung.

Weiterhin möchte ich Katrin Grohmann und Sebastian Kriening aus der Arbeitsgruppe um Frau Prof. Dr. Sabine Laschat am Institut für organische Chemie der Universität Stuttgart für die Synthese von Standards sowie für die Unterstützung bei der Entwicklung der Analytik danken.

Allen Mitgliedern der Arbeitsgruppe „Bioinformatik“ und „Biokatalyse“ des Instituts für Technische Biochemie möchte ich für die ständige Hilfsbereitschaft danken. Besonderer Dank gilt Frau Dr. Vlada B. Urlacher.

Der Deutschen Forschungsgemeinschaft danke ich für die Finanzierung meiner Arbeit im Rahmen des Sonderforschungsbereichs 706.

Ein sehr herzlicher Dank geht an meine Familie, die mich stets in allen Belangen unterstützte und mir zur Seite stand.

Erklärung

Hiermit versichere ich, die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Hilfsmittel und Literatur angefertigt zu haben.

Stuttgart, den 15. Dezember 2008

Alexander Seifert