STRUCTURAL AND FUNCTIONAL EXPLORATION OF AN ARTIFICIAL FAMILY OF CYTOCHROMES P450

Thesis by

Christopher R. Otey

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2007

(Defended October 30th, 2007)

ACKNOWLEDGEMENTS

The best feature of CalTech is the diversity and intelligence of the people one has an opportunity to work and interact with. The broad perspectives and vast knowledge base allows one to learn and grow in different ways each and every day. I have developed a lot of great friendships that I hope will last long into the future and I want to thank everyone I have come in contact with for how they have helped me grow as a person.

In the laboratory, everyone's assistance has been invaluable but I especially want to thank a few individuals. I want to start with Joff Silberg, who provided a great deal of mentorship during the early stages of my research and was invaluable in providing guidance and inspiration. I am thankful I could always go to him with questions, to get ideas and to get back on track. Kaori Hiraga was a joy to work with and developed SISDC and was extremely helpful in constructing the chimeric P450 library, both times we made it! I must thank the 'theoretical' folks in the lab; Jeff Endelman, Allan Drummond, Jesse Bloom, Chris Voigt and Chris Snow. Their assistance has proven to me once and for all that they do exist. I must especially thank Jeff Endelman and Chris Voigt for their direct contributions in the form of logistic regression tools and the SCHEMA algorithm. Michelle Meyer and I shared a lot of ideas and frustrations having very similar projects which was always a comfort. Matt Peters did a wonderful job of bringing a lot of us together in collaboration and provided invaluable chemistry knowledge and great support. I would like to thank all of the other P450 team members, Edgardo Farinas, Uli Schwaneberg, Pat Cirino, Jorge Rodriguez, Peter Meinhold, Matt Peters, and Mike Chen. The collaborative

nature of this team throughout the years is something we should all be proud of. Everyone here provided help in the form of knowledge, ideas, materials and friendship. Without their assistance I would have been unable to accomplish and learn all that I have with cytochromes P450. I also had the pleasure to work with some undergraduate and rotation students. James Ross for his help during a summer in which he assisted in the development of many HPLC methodologies and reminded me just how smart a young college student can be. I want to also thankNatalie Kruk, Ward Walkup and Kiowa Bower for assistance with high-throughput characterization of the chimeric P450 library.

A huge thanks goes to Geethani Bandara who fulfills so many roles in the lab. Not only does she fight the endless battle against lab entropy but also offers her support, friendship and an occasional helping hand in the lab. I was on the receiving end of each and every one of these and am very grateful for it. I also want to thank Marco Landwehr for his help at all levels. He is a great friend and colleague and was with me in the trenches on many of the experiments and analyses described here. He and his wife, Dani Dieterich, provided a ton of support outside the lab as well for which I will always be grateful.

I want to thank Frances Arnold, for allowing me the opportunity work in her lab. Frances supplied much of the inspiration for this project and through this process I developed a great respect for her ability to drive a project forward. Observing her management style and focused analytical ability has allowed me to take much more away from my graduate school experience than I think one gets in the average lab.

I would also like to thank the Croal family. Graciela and Tim have made me feel completely welcome in their family and have provided constant support as if I were a son. Megan and Jimmy have been another set of siblings for me and always chimed in to make me laugh or remind me that I am very important. All of their love and support will never be forgotten.

I would like to thank my brother, Clayton, for his inspiration and reminding me how great you can be at something when you truly love it. My Mother and Father have always provided constant love and support. They given me a solid foundation and provided a much-appreciated safety net which has allowed me to continue my education for so very long. They have served as role models with their own many accomplishments and are always there when I need them.

Finally I want to thank my lovely girlfriend, Laura, who has offered constant support, patience and love. She continues to be my muse.

ABSTRACT

Protein families are comprised of numerous sequences that adopt a similar three-dimensional structure and functional properties. The superfamily of cytochromes P450 are an excellent example of a common structural scaffold being utilized for a variety of biological functions. This functional diversity is achieved in Nature through millions of years of evolution to create new and diverse sequences. We have used site-directed recombination guided by the computation algorithm SCHEMA to create an artificial family of cytochromes P450 in the laboratory. Members of this family possess unique properties such as altered activity profiles, increased thermostability and the ability to accept new substrates.

We developed screening tools for the rapid analysis of hundreds of individual P450s. These high-throughput assays include the 4-aminoantipyrine (4-AAP) assay which is capable of detecting the hydroxylation of an aromatic ring. High-throughput carbon monoxide binding facilitates the rapid detection of P450s that correctly incorporate a heme cofactor and are thus properly folded and potentially functional. Finally, a substrate binding assay which measures a spectral shift that occurs when a substrate binds in a P450 active site is described.

Fourteen double-crossover chimeras created from the bacterial P450s CYP102A1 and CYP102A2 were constructed to calibrate the P450 scaffold for SCHEMA, a computational algorithm used to minimize structural disruption in chimeric proteins. We found that only chimeras with high levels of structural disruption as measured by SCHEMA were

unfolded. Among the fourteen chimeras we also observed three different activity profiles based on peroxygenase kinetic assays with the substrates p-nitrophenoxydodecanoic acid (12-pNCA), 2-phenoxyethanol and allyloxybenzene.

We applied this calibration to create an artificial family comprising ~3,000 chimeric heme P450 proteins that correctly fold and incorporate a heme cofactor by recombining three cytochromes P450 at seven crossover locations chosen to minimize structural disruption. Members of this protein family differ from any known sequence at an average of 72 and by as many as 109 amino acids. Most (>73%) of the properly folded chimeric P450 heme proteins are catalytically active peroxygenases; some are more thermostable than the parent proteins. A multiple sequence alignment of 955 chimeras, including both folded and not, was analyzed using logistic regression analysis (LRA) to identify key structural contributions to cytochrome P450 heme incorporation and peroxygenase activity and suggests possible structural differences between parents CYP102A1 and CYP102A2.

Thirty-four members of this artificial family were assayed for functional diversity on a set of eight substrates. P450 chimeras were able to exceed the parents in total activity on all eight substrates and were grouped into five different groups based on activity profiles using K-means clustering. Activity profiles on eight substrates were then performed in high throughput to produce a data set of 330 chimeras. The mean percent standard deviation of the activity assays showed the reproducibility of these high-throughput data and further analysis may reveal information about sequence-structure-function relationships.

The products of the catalytic reactions of four chimeric P450s with substrates of human P450s, some of which are drug compounds, were analyzed by HPLC in order to determine their identity. Chimeras were able to produce authentic human metabolites of chlorzoxazone, zoxazolamine and propranolol, showed peroxidase acitivty on 4-aminobiphenyl and produced an unknown product with tolbutamide. Finally, the peroxygenase activity of a mutant P450 heme domain is able to be further altered and enhanced using directed evolution. After two rounds of directed evolution and screening with the 4-AAP assay, we found mutants with altered substrate specificities and an overall enhancement of activity.

The design and high-throughput methodologies described here can be used to create artificial protein families and to discover new and useful protein sequences. Like natural protein families, artificial protein families can be used to identify regions of protein sequence and structure that are important for folding and function. This is especially useful for analyzing protein families with few members or for validating tools for structure prediction and for protein sequence-structure-function analysis. Artificial protein families are also rich in sequence diversity and can provide sources of novel protein function. Using the high-throughput methodologies described here, chimeric P450s with enhanced activity, altered activity profiles, and the ability to hydroxylate drug-like compounds to produce authentic human metabolites were discovered in our artificial family of P450s. These methodologies will hopefully be extended to the study of other protein families and to the creation and discovery of increasingly valuable protein catalysts.

Table of Contents

LIST OF TABLES

Table *Page*

LIST OF FIGURES

Number *Page*

ABBREVIATIONS

12-pNCA	p-nitrophenoxydodecanoic acid
4'OHP	4'-hydroxypropranolol
4-AAP	4-aminoantipyrine
5'OHP	5'-hydroxypropranolol
A1	CYP102A1 heme domain with F87A mutation
A2	CYP102A2 heme domain with F88A mutation
A3	CYP102A3 heme domain with F88A mutation
ABTS	2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonate)
CRFP	chimera reductase fusion protein
CE	combinatorial extension
DIP	desisopropylpropranolol
EPPS	N-[2-hydroxyethyl]piperazine-N'-[3-propanesulfonic acid]
FAD	flavin adenine dinucleotide
FMN	flavin mononucleotide
HPLC	high performance liquid chromatography
HRP	horse radish peroxidase
IPTG	isopropyl –β-D-thiogalactopyranoside
LRA	logistic regression analysis

MSA	multiple sequence alignment
PCR	polymerase chain reaction
RASPP	recombination as a shortest-path problem
RMSD	root-mean-square deviation
SISDC	sequence-independent site-directed chimeragenesis

Chapter 1

Introduction and thesis summary

Proteins are linear chains of amino acid that fold into specific three-dimensional structures (protein folds) that enable them to perform various biological functions critical for life. They have evolved over the past few billion years to comprise an enormous diversity of sequences, estimated at 5×10^{10} unique protein sequences [1]. However, there are postulated to only be on the order of one to ten thousand folds that are adopted by these many sequences [1-4]. These protein folds are therefore encoded by a multitude of amino acid sequences and each sequence is able to utilize that same fold for different biological functions. Sequences with similar folds are often grouped into protein families which are used to derive additional information about these protein folds and sequences. It is the functional diversity found within a single protein fold and the knowledge that can be extracted from protein families that has motivated the work described here.

An excellent example of functional diversity and sequence differentiation is found in the cytochromes P450 superfamily of enzymes. In Nature, cytochromes P450s are capable of a variety of biological functions such as the breakdown of xenobiotics, primary catabolism, and steroid and other secondary metabolite biosynthesis [5,6]. The P450s that perform these myriad functions evolved by natural selection from a common ancestor which appeared hundreds of millions of years ago [7]. There now exist an enormous number of unique P450 sequences with over 5,000 known members of this family [8]. These P450s are highly variable at the amino acid sequence level (often less than 15% amino acid identity) yet their three-dimensional structures are well conserved [9,10]. Figure 1.1 exemplifies the functional and sequence diversity found in cytochromes P450

and illustrates how Nature has utilized the P450 fold for a variety of biological functions by diversifying its amino acid sequence.

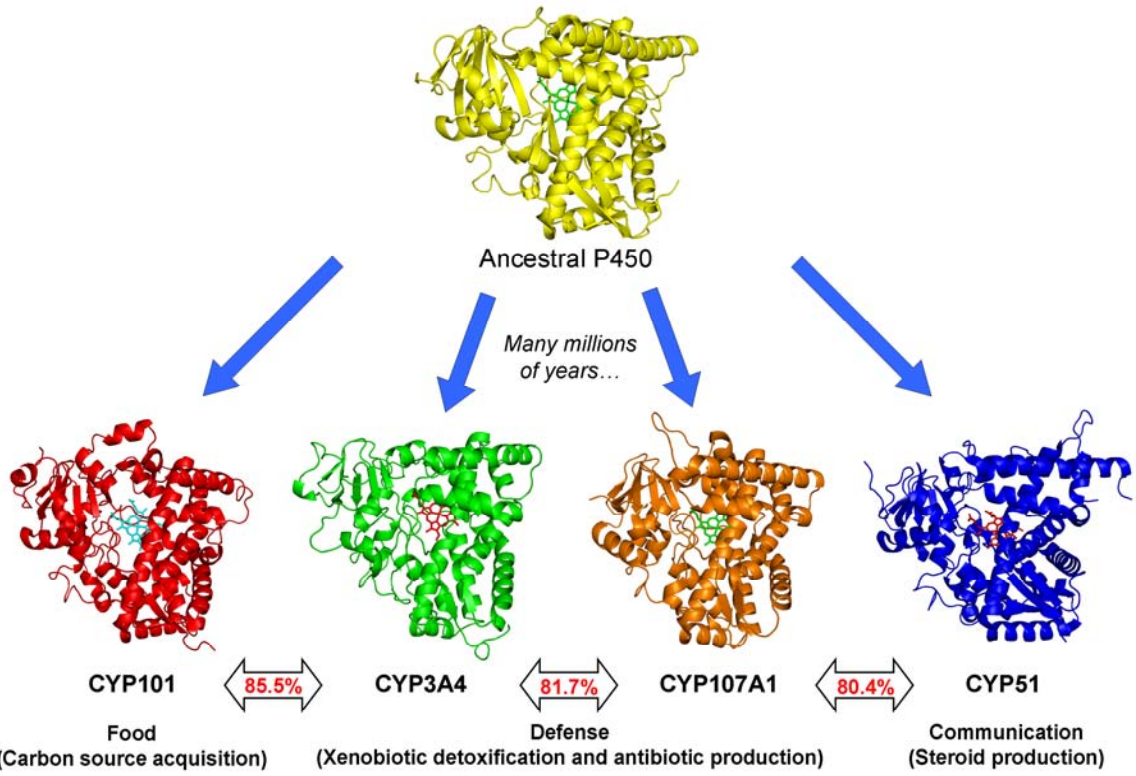

Figure 1.1. The diversification of an ancestral P450 sequence has led to new, divergent sequences with various biological functions but a well-conserved three-dimensional structure. CYP101 (red) is involved in the utilization of camphor as a carbon source in bacteria [11]. CYP3A4 (green) and CYP107A1 (orange) are involved in the defense of their host organisms by breaking down xenobiotics such as drugs or toxins and synthesizing antibiotics, respectively [12,13]. CYP51 (blue) is involved in cellular communication by synthesizing steroids [14]. The number in red shows the variability of the protein sequences as percent of amino acids which are different in a pairwise sequence alignment using ClustalW [15].

We sought to mimic the diverse functions found in a single protein fold by creating a large number of new P450 sequences in the laboratory using protein recombination

guided by the computational algorithm SCHEMA [16]. The result was the creation of an artificial family of cytochromes P450 which contains many new and diverse sequences that possess a variety of useful and interesting properties.

Chapter 2 provides background information on protein diversity, current methods for creating protein diversity in the laboratory, and fundamentals of cytochromes P450. Chapter 3 describes high-throughput assays we developed in order to rapidly characterize the P450s created by recombination. Methods for detecting aromatic ring hydroxylation, a substrate binding event and heme incorporation are described. In Chapter 4, SCHEMA structural disruption is calibrated against the P450 scaffold by creating seventeen individual double-crossover chimeras between the heme domains of CYP102A1 and CYP102A2 (A1 and A2). SCHEMA is a computational algorithm used to further improve the likelihood of creating folded proteins by minimizing structural disruption upon recombination. The double-crossover chimeras were assayed for their ability to correctly fold and incorporate a heme cofactor and only chimeras with a higher level of structural disruption were not folded. The chimeras also showed a wide range of thermostabilities and possessed novel activities and substrate specificities relative to the parental proteins.

Guided by the calibration in Chapter 4, we created an artificial family of cytochromes P450 using SCHEMA-guided, site-directed recombination of the three bacterial P450 heme domains, A1, A2 and CYP102A3 (A3) in Chapter 5. This family contains thousands of new and diverse chimeric P450 sequences which differ from any known

P450 by 72 amino acid substitutions on average. Using logistic regression analysis (LRA) of 955 of these sequences we determined which sequence elements contribute to a chimeric P450s ability to fold and correctly incorporate its heme cofactor. We found that some of the new P450s are also more thermostable than any of the starting parents. Activity measurements on 2-phenoxyethanol and p-nitrophenoxydodecanoic acid (12-pNCA) showed at least 73% of the chimeric P450s are catalytically active. LRA analysis of chimeras for which the activity on 12-pNCA was determined identified portions of the sequence contributing to this activity. Further analysis led to the construction of site-directed mutants that identified residues previously unknown to be important for catalytic activity.

Chapter 6 describes functional characterization of members of the artificial library on eight substrates and shows we can produce a large data set for further analysis. K-means clustering of 34 chimeras showed they separated into five groups based on activity profiles. This demonstrates that chimeric P450s can be functionally different from each other and from any of the parental proteins. The characterization of select chimeras towards substrates of human P450s, some of which are drug compounds, is described in Chapter 7. Chimeric P450s show activity on chlorzoxazone, propranolol, zoxazolamine, tolbutamide and 4-aminobiphenyl. The authentic human drug metabolites of chlorzoxazone, propranolol and zoxazolamine were shown to be shown to be produced by some of the chimeras using HPLC analysis. Chapter 8 describes the directed evolution of a mutant P450 peroxygenase for improved production of drug metabolites of propranolol. This chapter demonstrates that once activity on these important drug-like

molecules is found it can be further modified and improved using directed evolution. We were able to shift activity to the more active 5-hydroxypropranolol product as well as increase total production. We also demonstrate the potential utility of a bioconversion done with a P450 heme domain driven by hydrogen peroxide and its ability to produce comparable quantities of authentic drug metabolites to other methods.

By creating and analyzing an artificial family of cytochrome P450s, we have made advances in the field of protein engineering and further demonstrated the utility of the cytochrome P450 scaffold for biocatalysis. The use of site-directed recombination guided by SCHEMA demonstrates that thousands of new proteins can be created in the laboratory. These new sequences can be analyzed as a traditional protein family would to derive information about the new and starting proteins. Proteins created in this manner can also vary functionally and perform useful chemical reactions such as the hydroxylation of drug compounds into their authentic human metabolites. These properties can then be further improved from the newly discovered sequence using directed evolution. This general protein design, creation and discovery should be applicable in the creation of other new enzymes with desirable properties.

References

1. Koonin EV, Wolf YI, Karev GP (2002) The structure of the protein universe and genome evolution. Nature 420: 218-223.
2. Chothia C (1992) Proteins. One thousand families for the molecular biologist. Nature 357: 543-544.
3. Coulson AF, Moult J (2002) A unifold, mesofold, and superfold model of protein fold use. Proteins 46: 61-71.
4. Leonov H, Mitchell JS, Arkin IT (2003) Monte Carlo estimation of the number of possible protein folds: effects of sampling bias and folds distributions. Proteins 51: 352-359.
5. Lewis DFV (2001) Guide to cytochromes P450: structure and function. London; New York: Taylor & Francis.
6. Ortiz de Montellano PR (1995) Cytochrome P450: Structure, Mechanism, and Biochemistry. New York: Plenum Press.
7. Gotoh O, Fujii-Kuriyama Y (1989) Evolution, structure and gene regulation of cytochrome P450. In: Ruckpaul K, Rein H, editors. Frontiers in Biotransformation. Berlin: Akademie Verlag. pp. 195-243.
8. Nelson D (2007) Homepage <http://drnelson.utmem.edu/CytochromeP450.html>.
9. Mestres J (2005) Structure conservation in cytochromes P450. Proteins 58: 596-609.
10. Werck-Reichhart D, Feyereisen R (2000) Cytochromes P450: a success story. Genome Biol 1: REVIEWS3003.
11. Schlichting I, Jung C, Schulze H (1997) Crystal structure of cytochrome P-450cam complexed with the (1S)-camphor enantiomer. FEBS Lett. 415: 253-257.
12. Yano JK, Wester MR, Schoch GA, Griffin KJ, Stout CD, et al. (2004) The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-A resolution. J. Biol. Chem. 279: 38091-38094.
13. Nagano S, Cupp-Vickery JR, Poulos TL (2005) Crystal structures of the ferrous dioxygen complex of wild-type cytochrome P450eryF and its mutants, A245S and A245T: investigation of the proton transfer system in P450eryF. J. Biol. Chem. 280: 22102-22107.
14. Podust LM, Poulos TL, Waterman MR (2001) Crystal structure of cytochrome P450 14alpha -sterol demethylase (CYP51) from Mycobacterium tuberculosis in complex with azole inhibitors. Proc. Natl. Acad. Sci. U. S. A. 98: 3068-3073.
15. Thompson JD, Higgins DG, Gibson TJ (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: 4673-4680.
16. Voigt CA, Martinez C, Wang ZG, Mayo SL, Arnold FH (2002) Protein building blocks preserved by recombination. Nat. Struct. Biol. 9: 553-558.

Chapter 2

Background: Protein functional diversity and cytochromes P450

A. Natural protein diversity

Proteins are key building blocks of life and involved in virtually all aspects of cellular structure and function. They serve a variety of biological roles including the binding of nearly any conceivable molecule, catalyzing myriad chemical reactions, providing structure to living cells, allowing motion, and acting as sensors and switches. Their diversity stems from the nearly infinite number of ways these linear polymers can be assembled from the twenty amino acids. For instance, an average protein with length of 200 amino acids has 20^{200} different possible sequences. The space of all possible sequences, termed protein sequence space, is vast and for all intents and purposes infinitely large [1-3]. The frequency of sequences which form useful proteins in this space, however, is extremely low [4-6]. These useful sequences generally define a protein fold, a well-defined three-dimensional structure that gives a protein its properties.

Roughly three to four billion years ago [7], an early form of life happened upon a useful protein sequence which likely accelerated its rate of reproduction, possibly by enhancing the replication of a prebiotic nucleic acid [8,9]. Since then, this ancestral protein sequence has diversified and evolved through the process of natural selection to serve a variety of functions [10]. Protein sequences found in Nature are not randomly distributed throughout sequence space and this is used to group them into functionally or structurally related groups [11]. In the 1960s, the concept of protein families and superfamilies was introduced in order to better understand protein evolution and function [12]. A family of proteins is a group of closely related sequences and a superfamily is two or more related families. This has been expanded to more hierarchical systems which consider a proteins

three-dimenstional structure first, then sequence and functional similarity [13,14]. By grouping proteins into families, various bioinformatic tools can be applied to learn about the proteins in general as well as the individual members. Generally a set of sequences are aligned using multiple sequence alignment tools such as ClustalW [15,16]. Topics such as residues contributing to enzymatic activity [17,18], ligand binding [19], protein structure [20-22], and protein evolution [23,24] can be explored.

B. Artificial protein diversity

Artificial protein diversity consists of the new protein sequences created by humans. The number of sequences created this way pales in comparison to that found in Nature, but there has been much progress with the creation of ever more complicated proteins as well as those with folds not yet found in Nature [25]. To create new protein sequences, researchers must either modify existing sequences, design specific sequences from scratch, or search among random or semi-random sequences [26,27]. The former is generally the most successful and has found utility in the modification of enzymes for industrial purposes [28,29]. And indeed, biocatalysts are used in large-scale industrial applications including the production of acrylamide and nicotinamide with nitrile hydratases [30], pulp kraft-bleaching and recycling of paper [31], antibiotic synthesis [32], and the synthesis of biopolymers [33].

B.1 Modification of existing sequences

New protein sequences can be created from existing sequences by selectively mutating an amino acid to another specific amino acid or by randomly mutating portions of the entire sequence and screening or selecting for sequences with the desired properties. Generally, the goal of such experiments is to improve specific protein properties such as stability or activity. When targeting specific residues information from a variety of sources is employed such as structural information [34], comparison with other protein sequences [35] or from experimentally derived information (e.g. importance in substrate specificity) [36,37]. Successful inversion of enantioselectivity [38,39] and activity enhancements [34,40,41] are a few examples of successes achieved with targeted mutagenesis.

Computational protein design is another form of targeted mutagenesis that employs different algorithms to choose the specific amino acid residues to be altered. The mutations selected in this manner generally have a higher success rate as they are pre-screened by the computer. Early computational design efforts aimed to repack the hydrophobic cores of proteins to increase stability or to optimize other structural elements [42], eventually leading to fully automated computational sequence selection tools [43-45]. Since these early efforts, numerous other properties have been designed including the stabilization of normally unstable portions of proteins [46], optimization of interfaces [47,48], and the binding of metals [49] or other cofactors [50-53].

Unlike targeted mutagenesis, directed evolution does not require large amounts of information about structure or function to be successful. Instead an algorithm similar to

that used by Nature, natural selection, is employed [54]. Directed evolution involves the creation of a large number of variations of a gene of interest (a.k.a parental gene). These groups of genes with new sequences are termed ‘libraries’ and are screened or selected for an individual sequence with a desired property [55]. These improved variants are then used as parents for the next round of evolution and iterations of this cycle are repeated until the desired result is achieved. Numerous methods to create variant genes for directed evolution have been developed such as error-prone PCR, where a gene is randomly mutated, or DNA shuffling, where multiple genes are broken and reassembled into new sequences [27,56]. Directed evolution can be used to evolve a protein in unnatural conditions and for features which Nature may not evolve for but may be necessary for industrial or medical purposes [57,58]. The new proteins discovered in this fashion often contain very unpredictable sequence changes, for example mutations on the proteins surface can affect substrate specificity. Directed evolution has been used to alter and improve many properties of enzymes including thermostability [59-61], organic solvent resistance [62,63], enantioselectivity [64-66], increased activity [67,68], functional expression [69], and altered substrate specificity [70-72].

More recently, a synergy between targeted design and directed evolution has been demonstrated. By combining design and directed evolution, Park and co-workers were able to introduce β-lactamase activity into the αβ/βα metallohydrolase scaffold of glyoxalase II [73]. The design was used to choose mutations which gave the protein a very low level of the new activity, followed by rounds of directed evolution to find mutations to rapidly improve this activity. The creation of a biologically functioning

enzyme from a protein without any known enzymatic activity serves as another example of the potential [74-76]. In this experiment, Dwyer et al. began with ribose-binding protein and converted it to function as a triose phosphate isomerase using computational design. The designed molecule had a rate enhancement of 10^5 over the background reaction and after a few rounds of directed evolution by random mutagenesis was able to support *E. coli* growth under gluconeogenic conditions on glycerol [75].

B.2 Searching among random sequences and de novo design

The creation of libraries of random or semi-random protein sequences followed by screening or selection for folded or functional proteins has been successful in discovering interesting sequences with basic functions. For instance, isolation of proteins with simple functions such as the binding of adenosine triphosphate have been isolated from a pool of random sequences [77]. Methods which attempt to increase the frequency of useful sequences in these large libraries such as a 'binary patterning' have accelerated the discovery of *de novo* proteins. Binary patterning takes advantage of the structural periodicity of polar and nonpolar amino acids in protein secondary structures. The first success was in creating a four-helix bundle using amphipathic α-helices which have a repeating periodicity of 3.6 residues per turn [78]. This methodology was extended to beta-sheet proteins were this same strategy yielded proteins which self-assembled into large fibrils similar to the amyloid fibrils found in neurodegenerative diseases [79,80]. Another method to increase the frequency of promising sequences is to use a reduced set of amino acids that improves the frequency of secondary structural elements. For

example, isolation from random sequences consisting mainly of random combinations of glutamine, leucine, and arginine yielded folded structures with native-like folds [81].

Computational design can also be used to design sequences *de novo*. Instead of searching among a pool of sequences, a computational algorithm is used to design a single sequence that is predicted to fold or function. As this is a very computationally demanding problem, the complexity of sequences that can be designed is limited by accessible computational power. As such, some of the first computationally designed proteins were very simple proteins such as four-helix bundles that displayed circular dichroism spectra characteristic of α-helices but were not uniquely folded structures [82,83]. More recently, computational protein design has been used to design a protein fold not present in the Topology of Protein Structure server [25]

B.3 Protein recombination

Protein recombination is a powerful tool in the creation of new protein sequences. Since it will serve as a primary means here, the methods and background will be discussed in more detail. Recombination has been shown to be an effective search strategy in other fields such as animal breeding, computer programming and economics [84-86]. Many of the ideas behind the use of recombination to alter protein properties in the laboratory stem from the observation of alternative splicing and the presence of introns and exons in Nature [87-89]. Alternative splicing takes place when exons present in an organism's genome are differently assembled in order to modify a proteins sequence and potentially function. The exon theory of proteins takes this a step further and postulates that during

evolution, proteins acquired functional diversity by combining the exons encoded by early proteins [88,89]. A search for these protein building blocks or Nature's protein lego set still continues [90].

Protein recombination refers to the division and reassembly of the nucleotide sequence encoding two or more parent proteins and reassembling them into new sequences (Figure 2.1). The new proteins created in this manner are referred to as chimeras. The advantage of recombination over other methods such as random or targeted mutagenesis is that one is able to create a much larger number of sequence changes while still retaining a high frequency of folded and potentially interesting proteins. Using random mutations, one can make approximately five amino acid substitutions before the fraction of functional proteins in a library drops below 10% [91,92]. In contrast, recombination is much more conservative and allows at least an order of magnitude more substitutions while retaining the same frequency of functional proteins [93]. The main reason for this is the new proteins contain sequence changes which have effectively been prescreened by Nature to be compatible with a given structure [94]. The larger number of sequence changes allows one to explore more distant regions of protein sequence space and potentially functional space [94].

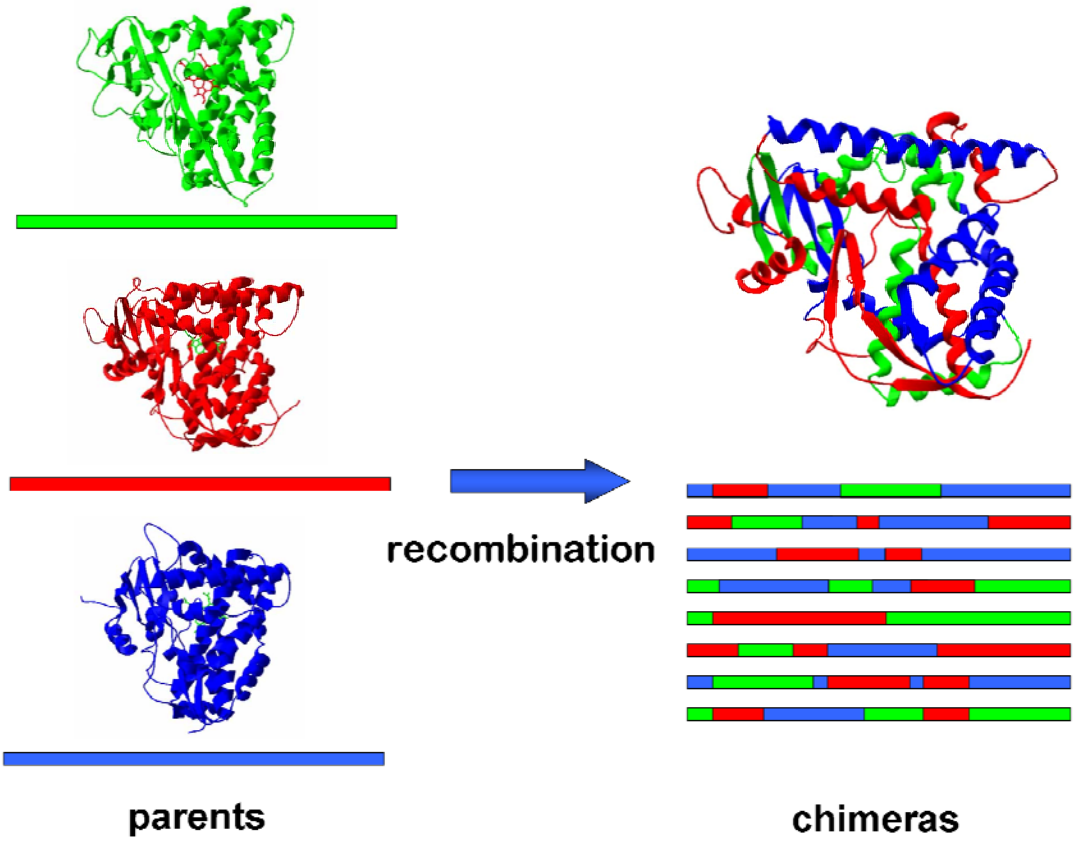

Figure 2.1. *In vitro* recombination showing three parental proteins and genes being recombined to form numerous chimeric offspring.

Numerous methodologies exist for *in vitro* protein recombination. One of the most well known, DNA shuffling, relies on stretches of DNA sequence identity to randomly reassemble gene fragments of parental genes into chimeras [68,95,96]. It works by breaking the DNA of different parent genes into pieces and reassembling them in a PCR reaction. Variant proteins with a few mutations each can be recombined in this manner in order to find the most favorable combinations of the mutations or to eliminate deleterious ones [97]. Alternatively, a method referred to as family shuffling uses naturally occurring protein sequences as the parent enzymes [94,98]. Family shuffling takes advantage of the

diversity found in homologous genes from nature and has assisted in the creation of enzymes with increased activity [94,99], higher stabilities[100], and altered substrate activities [101-103]. More distant regions of sequence space, that is regions that are not connected by single amino acid changes, are sampled by this approach. It is thought that this may allow for the discovery of other functional regions of sequence space. However, DNA shuffling and other annealing based methods suffer from only allowing the use of parental proteins with high levels of sequence identity (>70%) and crossover locations biased towards regions of highest sequence identity. Other annealing based methods include RACHITT [104], StEP [105], heteroduplex formation [106], synthetic shuffling [107], ADO [108] and DOGS [109].

To overcome the issues of crossover bias and high parental similarity requirements, sequence-independent recombination methods have been developed. Methods such as SHIPREC [110], ITCHY [111] and SCRATCHY [112] rely on direct DNA ligation rather than annealing, allowing for the creation of chimeras from very divergent parents but resulting in a higher frequency of deletions and frameshifts. Of these three methods, only SCRATCHY is able to create more than a single crossover in each chimera. Another method, SISDC, uses the insertion of tag sequences to specify crossover locations and is less susceptible to the problems resulting from direct ligation. It also allows the crossover locations to be specified and is able to incorporate multiple crossovers [113]. Finally, full gene synthesis can be used to assemble chimeric sequences [114].

To complement these construction methods are a number of design strategies by which libraries of chimeras can be further optimized. The methods FAMCLASH and OPTCOMB seek to minimize the number of unfavorable clashes in the resulting chimeras by examining evolutionary sequence information [115] and OPTCOMB [116]. IPRO uses energy-based scoring functions to improve overall library quality [117]. SCHEMA, which is used in the work described here, uses structural information from the parental proteins in order to design optimal libraries. Using the 3D structure of one of the parent proteins, SCHEMA identifies pairs of amino acids that are interacting (e.g., residues within a cutoff distance of 4.5 Å) and determines the net number of interactions broken when a chimeric protein inherits portions of its sequence from different parents [118]. SCHEMA has been shown to increase the frequency of folded proteins by choosing crossover points which minimize structural disruption [119,120].

C. Cytochromes P450

Cytochromes P450 are a fascinating family of enzymes due to their importance in fields such as the creation and administration of medicines, the evolution of every family of life and their future promise for biotechnology and ‘greener’ chemistry. P450s are ubiquitous in Nature and constitute one of the most diverse enzyme superfamilies. They are most well known for their role in oxidative chemistry whereby an oxygen atom is inserted into a C-H, C-C, C-N or N-H bond leading to epoxidations, S-oxidations, N-hydroxylations and dealkylations. Other reactions which are catalyzed include but are not limited to reductions, desaturations, ester cleavages, ring expansions, ring formations, aldehyde scission, dehydration, coupling reactions, *ipso* attacks, one-electron oxidations, *cis-trans*

bond rearrangements, oxidative cleavage and rearrangements [121-123]. The diverse functions of these enzymes stems from their heme cofactor (protoporphyrin IX) seen in Figure 2.2A. The heme is covalently bound to an axial cysteine ligand and is housed in a generally hydrophobic active site which allows for the binding of a diverse array of substrates. This heme cofactor is also what provides P450s with their name since P450 is derived from the unusual Soret band with a maximum absorption near 450 nm that is present upon CO difference spectroscopy (Figure 2.2B).

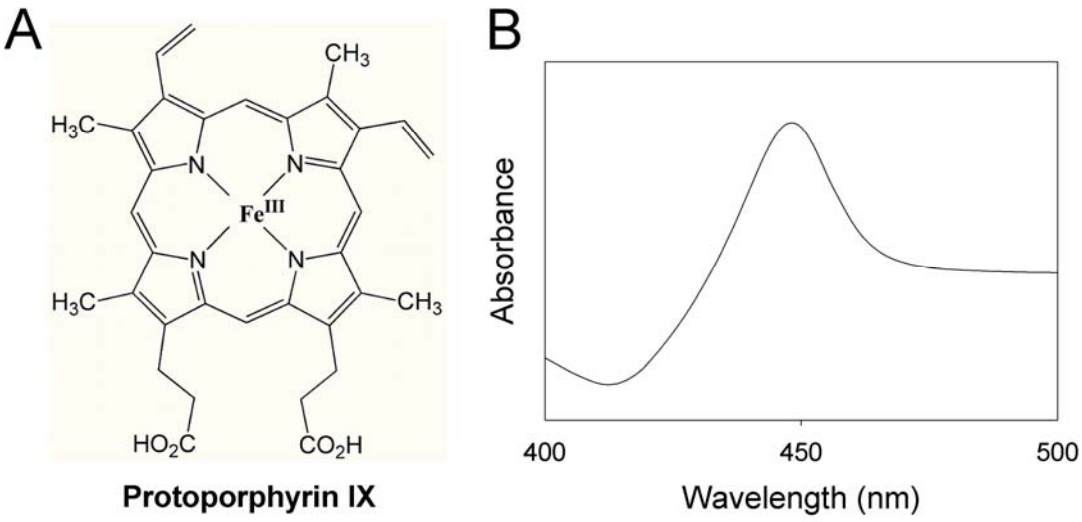

Figure 2.2. Basic features of a P450. A) Chemical structure of protoporphyrin IX, the heme cofactor found in cytochromes P450. B) Soret band at 450 nm observed using CO difference spectroscopy on the ferrous–CO complex.

C.1 P450 Nomenclature

A nomenclature system was designed for the large number of P450 sequences based on evolutionary divergence were P450 genes are grouped into families and subfamilies [124-128]. A P450 protein from one family generally shares ≤ 40% amino acid identity with a P450 protein from another family although there are exceptions [129]. Members of different families are though to have diverged from one another roughly 600 – 900 million years ago. P450s with 55% or greater amino acid sequence identity are then grouped into subfamilies and are approximated to have diverged from one another with the past 150 million years. Cytochromes P450 are abbreviated CYP which is then followed by their family and subfamily designations. For example, CYP3A4 is the fourth member of the A subfamily in family 3. Ranges of family numbers are assigned to different species with CYP1-49 belonging to animals, CYP51-70 belonging to fungi, CYP71-99 representing plant P450s and CYP101 and higher are for bacterial P450s. However, due to the rapid increase in known P450s this numbering system has now been extended to the hundreds and even thousands [128].

C.2 Diversity in the cytochromes P450 superfamily

Stemming from their ability to perform numerous chemical reactions and bind a variety of molecules, cytochromes P450 perform an enormous number of biological functions. In microbes, P450s function in secondary metabolism, signaling and catabolism. The most well known role of P450s in bacteria is in the catalysis of the first metabolic step in the metabolism of some carbon sources. Generally, a P450 hydroxylates an inactivated carbon compound which serves to solubilize the molecule as well as to initiate its

degradation. They facilitate a large number of molecules as carbon sources such as hydrocarbons [130], alkanes [131-135], and polycyclic hydrocarbons such as benzo[*a*]pyrene [136] and camphor [137,138]. Specific P450s that allow bacteria to utilize chemicals in the environment as energy sources include CYP101 [139,140], CYP108 [141], CYP111 [142,143] which are from different *Pseudomonas* sp. and CYP176A [144] from *Citrobacter braakii*.

The majority of P450 research is directed towards the 57 found in humans due to their critical roles in the synthesis of steroids and other signaling molecules as well as the breakdown of xenobiotics such as drugs and other environmental chemicals [145,146]. In fact, P450s are required to synthesize all steroids found in humans and assist in the metabolism of over 90% of currently prescribed drugs making them extremely interesting to the pharmaceutical industry [147,148]. Of the P450s important in drug metabolism, CYP3A4 does the majority of the work and is responsible for 30% of phase I metabolism [149]. Understanding P450s role in drug metabolism is critical to producing safe and effective drugs. Their variability in activity and expression from individual to individual can result in drastic differences in dosage requirements. For example, the expression of CYP2D6, which is important in drug metabolism, can vary over four orders of magnitude [150]. Inhibition of a P450 by a drug can lead to exaggerated pharmacological effects of another drug whose elimination or activation depends on the inhibited enzyme. This may cause build up of the drug in the patient or prevent the drug from being converted to its active form, both leading to therapeutic failure [151]. Understanding this metabolism can also lead to new and improved drugs since numerous drugs form active metabolites with

superior properties [152]. A well known example of this is with the anti-histamine drug terfenadine (Seldane) [153]. Approved by the FDA it went on the market in the early 1990's, however, it was soon discovered that it caused heart problems. It was then found that the active form of terfenadine was not the chemical that was ingested but the metabolite created by CYP3A4. This molecule turned out to also be less toxic and was approved by the FDA and introduced onto the market as fexofenadine, better know as Allegra™ [154]. Similar examples include the hypertension drug mibefradil and the asthma drug furafylline [155,156].

The genetic polymorphisms found in human P450s are one of the major contributors to individuals drastically different responses to drugs and other chemicals [157]. For instance, variable activities of CYP1A2 in humans is partially responsible for our different responses to caffeine since the rate of this enzyme can vary [158]. These variability's also make P450s of great interest to the medical community since numerous diseases are associated with defects in P450s. For example, a missing or damaged P450 involved in the processing of steroids or vitamins can have devastating health effects [159]. P450s are associated with diseases such as hypercholesterolaemia [160], coronary artery disease [161] and hypertension [162]. They are also key enzymes in cancer treatment and cancer formation since they can activate procarcinogens or inactivate carcinogens [163] as well as activate or inactivate cancer drugs [164].

In plants, P450s are involved in the biosynthesis of lignins, terpenes, alkaloids and a variety of other secondary compounds that act as plant defense agents [165-168]. Since

plants are sessile and cannot move away from danger, they require a large battery of P450s for defense. Whereas humans have 57 functional P450 genes, *Arabidopsis thalianis* has 246 (roughly 1% of genes) [169,170] and rice has a whopping 356 functional P450s [171]. An interesting example of insecticides and bactericides formed by P450s is the molecules DIBOA and DIMBOA, found in rye and maize, respectively [172-174]. Their biosynthesis involves four distinct P450s, each involved in a different synthetic. Additionally, plants use P450s in a variety of biosynthetic pathways of both central and secondary metabolites

Another form of diversity in P450s is their redox systems. There are numerous types of electron transfer components utilized by P450s as well [175,176]. Most bacterial and mitochondrial membrane P450s are termed class I P450s. Class I P450s have a reductase system with two components, a flavin adenine dinucleotide (FAD)-containing reductase and an iron-sulfur protein [177]. The class II P450s are typically membrane-bound eukaryotic microsomal P450s which interact with a single reductase containing both flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN) cofactors. There are some soluble class II P450s, as is the case with CYP102A1. Self-sufficient P450s, that is those that require no additional proteins since they utilize oxygen-containing substrates and do not require exogenous electrons, make up class III P450s. Class II P450s with all of the protein components fused into a single polypeptide chain are also sometimes grouped into this class [178,179]. A more recently discovered class IV consists of the catalytic heme domain fused to an FMN- and 2Fe-2S cluster-containing reductase similar to that found in dioxygenases [178,180,181]. In these systems, the flow

of electrons is from NADH to FMN to 2Fe2S cluster to heme domain. Artificial systems for supplying electrons have also been developed including electrochemical methods [182] and artificial cofactors such as cobaltocene [183] and cobalt (III) sepulchrate [184].

Due to the great interest in P450s, numerous x-ray crystal structures have been solved. Bacterial P450s were the first structures to be determined since they are highly soluble and easier to crystallize. To date the structures for CYP101A1 ($P450_{cam}$) [185-187], CYP102A1 ($P450_{BM3}$) [188-190], CYP108A1 ($P450_{Terp}$) [191,192], CYP107A1 ($P450_{eryF}$) [193-195], CYP119A1 [196], CYP154C1 [197], CYP152A1 ($P450_{BS\beta}$) [198], CYP175A1 [199], CYP121A1 [200], CYP107L1 ($P450_{PikC}$) [201], CYP154A1 [202], CYP167A1 ($P450_{epoK}$) [203], P450 OxyC [204], P450 OxyB [205], CYP199A2 [206] and CYP2R1 [207] have been solved. CYP55A1 ($P450_{Nor}$) [208] and CYP51 [209] from fungus have been determined. The structures of mammalian P450s include CYP2C5 [210-212], CYP2C9 [213,214], CYP3A4 [215,216], CYP2B4 [217,218], CYP2C8 [219], CYP2A6 [220], CYP2D6 [221].

C.3 P450 evolution

It is thought that the first ancestral P450 genes may have arisen in a prokaryote about 3.5 billion years ago [126]. There is strong evidence that all P450s are derived from this common ancestor due to their sequence similarity [222,223] and conservation of predicted secondary structure and hydrophobicity profiles [224]. The early function of P450s is still open to debate and has been postulated to be involved in the detoxification of oxygen as its levels began to increase to what we see today [225], a role in steroid

metabolism, assimilation of chemicals [129] or as peroxidases [224]. Regardless of their original biological function, P450s have rapidly diverged through gene duplications followed by divergence and ultimately fixation along with lateral gene transfer [129,226,227].

During more recent evolution over the past 800 million years, plant-animal and plant-insect warfare has aided the diversification of P450s [165]. As animals and insects began to ingest plants, the plants responded by synthesizing new metabolites that are toxic to their consumers. In response, animals and insects developed new enzymes capable of breaking down these toxins [129,228]. P450s were critical in this escalation. An example of how P450s affect plant-animal interactions is illustrated by CYP6B1 and CYP6B8 from the butterfly species *Papilio polyxenes* and *Helicoverpa zea*, respectively. The generalist, *H. zea,* is capable of consuming hundreds of types of plants compared to the specialist *P. polyxenes* which only feeds on furanocoumarin-containing plants. This is directly facilitated by their P450s with CYP6B8 showing a much broader substrate specificity allowing it to breakdown the insecticides present in a larger number of plants and therefore allow its host to feed on more plants. In contrast, CYP6B1 could only breakdown a small number of insecticides and therefore its host could only consume a small number of plants [229].

The rapid diversification of P450s has led to an extremely high variance in their primary sequences. The two most well conserved sequences among all P450s are within the K-helix and at the heme-binding site. The K-helix has an EXXR charged pair were X can be

any amino acid. The heme-binding site contains the 'signature sequence' of all P450s, FXXGXXXCXG which contains the cysteine that ligates the heme cofactor. With thousands of different P450s there are variations at each position except the critical cysteine. Other conserved features of different sub groups of P450s can be found but are less prominent [128,230]. However, despite having very low sequence identities ranging from 10 – 30 % between families [125], P450s possess the same tertiary structure and a well-conserved heme-binding core [231-233]. For example, a set of nine P450s which share between 10 and 27% amino acid sequence identity has over 28% of their C-alpha carbons withn a root-mean-square deviation (RMSD) of 2 Å demonstrating their high level of structural conservation [233].

C.4 P450 mechanism and catalytic cycle

The overall reaction catalyzed by cytochromes P450 can be seen in equation (1):

$$RH + O_2 + H^+ + NAD(P)H \rightarrow ROH + NAD(P)^+ + H_2O \quad \textbf{(1)}$$

The subsrate, RH, can be a tremendous number of molecules with a vast chemical diversity. The archetypal P450 reaction involves insertion of an oxygen atom into an unactivated C-H bond although many heteroatoms are possible targets [123]. P450s are capable of many other different types of reactions. The overall mechanism involves the combination of oxygen and substrate to produce a water molecule and a mono-oxygenated product facilitated by a protoporphyrin IX cofactor (heme). The central concepts of the catalytic cycle have been appreciated since early 1970 although specific

reaction intermediates are still up for debate. The structures of some of these reaction intermediates have been elucidated using x-ray crystallography in the CYP101 system [186,234-236] as well as with other spectroscopic methods [237-241].

The mechanism of cytochrome P450 catalysis has been the subject of numerous reviews [123,242-246] and is summarized here in Figure 2.3. The catalytic cycle begins with the enzyme in its resting state **[I]** with iron in the ferric state and the iron spin state typically low-spin (although some substrate-free P450s are high-spin [247,248]). The distal sixth ligand of the heme in this resting state is a water molecule [249]. When substrate binds, the water molecule is displaced, resulting in the iron spin-state equilibrium to become predominantly the high-spin ferric state **[II]** [250]. This results in the redox potential of the heme becoming less negative. In CYP101 and CYP102A1, for example, binding of substrate shifts the redox potential from -303 mV to -173 mV [250] and from -370 mV to > -270 mV [251], respectively. The redox potential shift allows the appropriate reductant, putidaredoxin (-240 mV) for CYP101 and the FMN- and FAD-reductase (-216 mV) for CYP102A1 for example, to provide reducing equivalents to the heme. The ferric iron moves to the ferrous state [**III**] while remaining high-spin. This is followed by preferential oxygen binding to the high-spin ferrous state with a rate constant of 1.7×10^6 M^{-1} s^{-1} producing a high-affinity complex (K_D = 0.6 μM in CYP101) [248,252]. This complex, however, is unstable and in equilibrium with a ferric-superoxide species, $Fe^{3+}O_2^{-}$ (**IV**) [253-256]. This ferric-superoxide species is favored by protonation of the superoxy anion and is facilitated in P450s by the presence of exchangeable protons in the heme environment. The second reduction yields a negatively charged ferric-peroxo and is

generally thought to be rate limiting in most P450s [**V**]. Protonation of this ferric-peroxo species is rapid, forming $Fe^{3+}OOH$ [**VI**]. A second protonation is quickly followed by heterolysis of the O-O bond forming reactive oxygen species equivalent to Compound I [**VII**] [257-259] and release of water. A thiolate-ligated nonheme oxoiron(IV) complex $[Fe^{+4}(O)(TMCS)]^{+}$ [260] as well as electronic paramagnetic resonance studies [261] supports the structure shown in [**VII**]. Interestingly it was found that in the absence of the thiolate ligand, the $[Fe^{+4}(O)(TMCS)]^{+}$ complex was an oxo-atom transfer agent (two-electron oxidant). The axial thiolate altered the reactivity into a hydrogen-atom abstraction agent (one-electron oxidant), consistent with P450 chemistry. This is an electron-deficient complex which is thought to abstract a hydrogen atom or an electron from the substrate resulting in collapse of this intermediate and generation of the product, ROH [**VIII**]. This product is then released, returning the enzyme to its resting state [**I**].

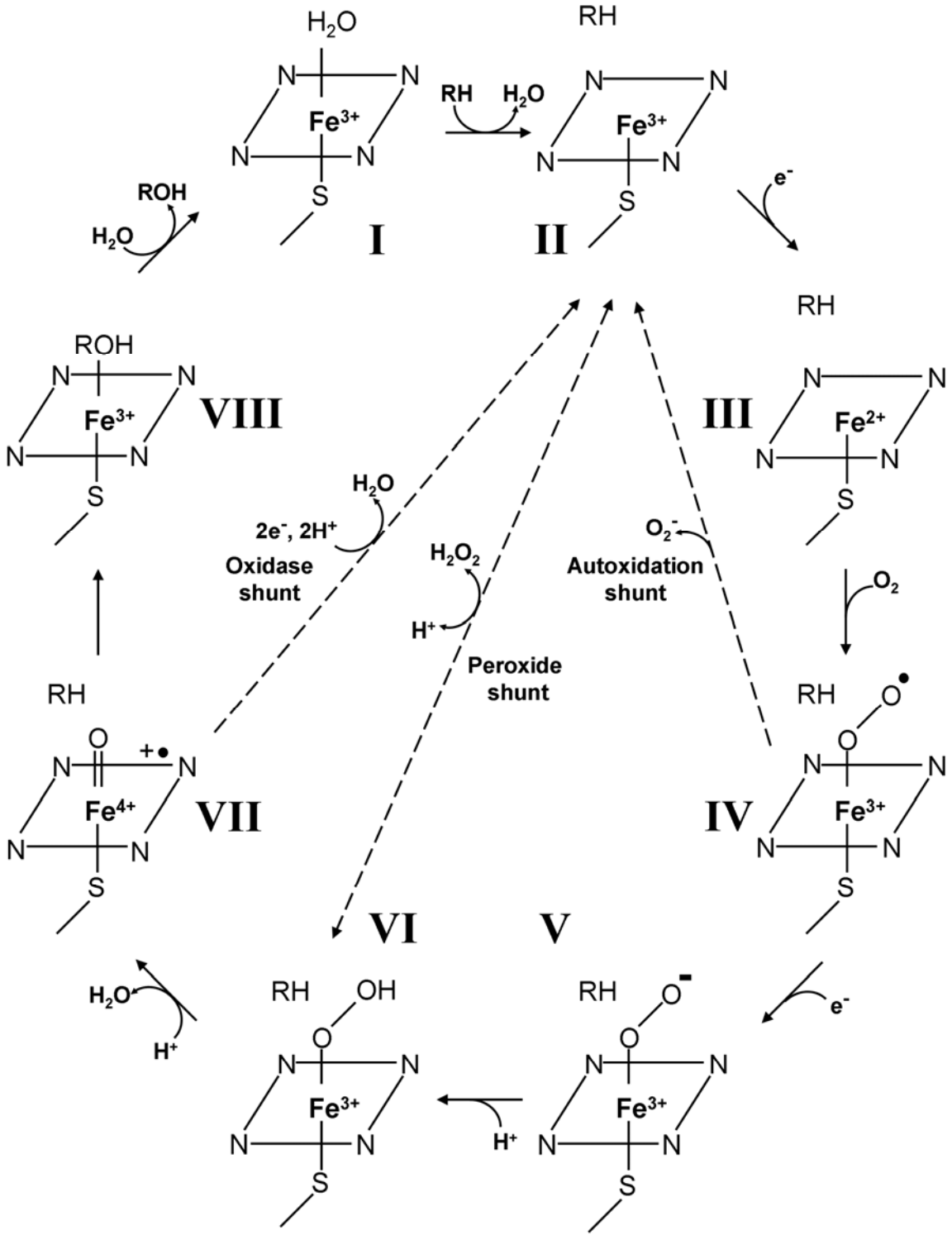

Figure 2.3. P450 catalytic cycle, see text for details.

The P450 catalytic cycle is not perfect, however, and there at least three branch points were side reactions may occur (Figure 2.3). These branch points are often referred to as uncoupling steps. The first uncoupling reaction that can occur during the cycle is the 'autoxidation shunt' were the oxy-ferrous species [**IV**] is autoxidized producing a superoxide anion and returning the enzyme to its resting state [**I**] [262,263]. In CYP101, the oxy-P450 complex autoxidizes with a rate constant of 0.01 s^{-1} at 20 °C [252]. The presence of substrate increases the half-life ten to twentyfold [252,253]. The second side reaction is the 'peroxide shunt' were the coordinate peroxide [**V**] or hydroperoxide anion [**VI**] dissociates from the iron forming hydrogen peroxide [264]. This completes an unproductive cycle in which the two-electron reduction of oxygen occurs without product formation. The peroxide shunt, however, may be exploited in the reverse direction eliminating the need for O_2 and NAD(P)H. This is discussed in more detail in the next section. A third side reaction the 'oxidase shunt' were the ferryl-oxo intermediate [**VII**] wherein the ferryl-oxo intermediate is oxidized to water instead of oxygenation of substrate [265]. This results in the four-electron reduction of dioxygen with the net formation of two water molecules.

C.5 The 'peroxide shunt' pathway

The two electrons used during the catalysis are generally supplied by one molecule of NAD(P)H, although the second electron transfer can be facilitated by cytochromes b_5 [266,267]. Alternatively, P450s are capable of utilizing oxygen and two electrons from peroxide to perform catalysis via the 'peroxide shunt' pathway (see equation 2).

$$RH + R'OOH \rightarrow ROH + R'OH \qquad (2)$$

Figure 2.3 shows that the peroxide shunt pathway bypasses a large part of the catalytic cycle. Numerous sources of peroxide such as organoperoxides or hydrogen peroxide may support the reaction, however this depends on the individual P450s. This pathway is thought to utilize a similar reactive oxygen species [242,268] produced by varying routes depending on the peroxide species, substrate and individual P450. Homolytic cleavage of peroxide has been shown to occur with peroxyphenylacetic acid [269], cumene hydroperoxide [270-272] while heterolytic cleavage has been shown with butylated hydroxytoluene [273]. The specific P450 and substrate also has an effect on how peroxide is utilized [272,274]. Additionally, a number of oxidants other than hydroperoxides such as peroxy acids, iodobenzene and sodium chlorite can be utilized [275-277].

NAD(P)H utilizing P450s generally have high K_m values for peroxides, with ranges from 15 to 250 mM have been reported [67,277-279]. Some P450s such as CYP152A1 and CYP152B1 ($P450_{BS\beta}$ and $P450_{SP\alpha}$, respectively) appear to naturally function as peroxygenases [198,280-284]. They hydroxylate near the carboxyl group of a fatty acid substrate which is thought to be stabilized by a nearby arginine residue similar to peroxidases. This presumably assists in peroxide activation by increasing the polarity of the active site. As such, CYP152A1 and CYP152B1 have low K_m's for H_2O_2 at 21 μM [285] and 72 μM [198], respectively.

The utilization of peroxide can cause different products to be made compared to the NAD(P)H driven reactions. For example, CYP1A2 catalyzes the activation of the heterocyclic aromatic amine 2-amino-3-methylimidazo[4,5-f]quinoline to genotoxic metabolites in the presence of peroxides [286]. P450s found in guinea pig liver were able to more efficiently catalyze hydroxylation at thermodynamically less reactive positions of arachidonic acid than the NADPH-dependent reaction [287]. P450s in rat liver microsomes showed different substrate specificities when using NADPH/O_2 versus H_2O_2 [288]. Reaction rates of CYP3A4 and CYP2D6 were able to be increased when using peroxide donors in place of the natural electron [289].

C.6 The CYP102 family

Members from the CYP102 family are employed in the studies here. Currently, there are thirteen known members of the CYP102 family with six members of the CYP102A subfamily (CYP102A1 – CYP102A6). CYP102's are class II P450s with a FAD- and FMN-reductase fused to the heme-binding domain [290]. The most well understood of the family and favored model system of our laboratory, is CYP102A1. The common name of CYP102A1 is P450 BM3, as it is the third P450 isolated from *Bacillus megaterium* [291]. Its sequence was first reported in 1989 causing much excitement in the field due to its potential use as a more accessible model system for understanding aspects of the membrane-associated human P450s [292]. CYP102A1 hydroxylates or epoxidates numerous substrates, but shows the highest activity (~17,000 min-1 [293]) on long-chain fatty acids at the omega-1-3 positions [293-295]. In Nature, its thought to play a role in the detoxification of polyunsaturated fatty acids such as linoleic acid [296,297].

Of the thousands of P450s, CYP102A1 is one of the best studied and frequently engineered. Several features make CYP102A1 such a popular P450 system [298]. The first is that it has the fastest known catalytic rate for P450 enzymes. This is thought to be a result of the fusion of the reductase domain and catalytic domain within the same polypeptide chain [299,300]. Although it can function when the heme and reductase domains are separated and mixed together, but at much lower efficiency [301]. Additionally, its large hydrophobic active site enables it to accept a variety of other, non-natural substrates [302]. Additionally, the numerous X-ray crystal structures that are available assist in understanding [188-190]. It is also easy to genetically manipulate and expresses well at levels of up to 1,000 mg/ml [298].

CYP102A1 has been engineered to accept and be more active on a variety of substrates. Early efforts aimed at improving activity on compounds similar to the natural fatty acid substrates. For instance, reversing the charge at the mouth of the active site allowed for increased activity on alkyltrimethyl ammonium compounds [303] and the activity towards shorter fatty acids was enhanced using targeted mutations [304]. Directed evolution of CYP102A1 led to a triple mutant (F87V/L188Q/A74G) able to oxidize a variety of substrates such as indole, alkanes, cycloalkanes, arenes, fluorene, acenaphthene and polycyclic, aromatic hydrocarbons [305-307]. Other triple mutants could hydroxylate polycyclic aromatic hydrocarbons such as fluoranthrene [308] and alkoxyresorufins [309]. Starting with wild-type CYP102A1, activity on octane was improved [310], followed by activity on gaseous alkanes such as propane and ethane [311,312]. Even

increases in the regioselectivity for the terminal position of octane [313] and epoxides [314,315]. Six amino acid mutations led to a 300-fold increase in activity on the carotenoid intermediate, β-ionone [316]. Engineering efforts have more recently targeted more commercial-like processes such as activity in organic solvents [62] and activity on valencene for fragrance production of (+)-nootkatone [317]

CYP102A2 and CYP102A3 are the next best characterized members of the CYP102 family [318,319]. Discovered in *Bacillus subtilus,* they share roughly 65% amino acid identity to CYP102A1 and are also fatty acid hydroxylases but prefer substituted fatty acids and have altered regioselectivity relative to CYP102A1 [319,320]. CYP102A2 has been evolved using error-prone PCR to exhibit modified substrate specificity with increased activity. The mutant discovered was proline 15 to serine, a residue likely not in direct contact with the substrate but in shaping the active site [321]. CYP102A3 was engineered using mutations that had been previously been made in CYP102A1 [307] to broaden its substrate specificity, especially towards aromatic compounds. Interestingly, one of these sequence changes was already present in CYP102A3. The mutations in CYP102A3 had similar effects, conferring activity on octane and naphthalene [322]. CYP102A3 has also been engineered for increased terminal hydroxylation of octane [307].

C.7 P450s and recombination

Breaking and reassembling P450s is not a new idea and is actually inspired by Nature were it is used to alter the function of or target a P450 to different cellular locations [323-

329]. For example, CYP3A7 is the most abundant P450 in human liver during fetal development and plays an important role in the metabolism of endogenous hormones, drugs and other potentially toxic chemicals [330,331]. Interestingly an alternately spliced variant comprised of exons from CYP3A7 and the pseudogene CYP3AP1 displays altered enzymatic activity and tissue distribution [329]. Another example of altered function is CYP4F3 which be modified by incorporating alternate exons [328]. Chimeric P450s can also be detrimental; unequal crossing over between CYP11B1 and CYP11B2 causes type I familial hyperaldosteronism [332].

Numerous studies have created chimeras of P450s for the exploration of novel functions as well as the study of the particular enzymes involved. For example, replacement of small stretches CYP102A1 with segments of roughly ten amino acids from the insect CYP4C7 produced enzymes capable of forming a new product from a previously accepted substrate [333]. Chimeric enzymes constructed by swapping domains between P450s and the heme enzyme nitric-oxide synthase proved to be able to either perform hydroxylation or produce nitric oxide [334,335]. Other examples of altered regioselectivities include chimeras between CYP2C2 and CYP2C1 which acquire 21-progesterone hydroxylase activity while retaining their activity on lauric acid [336]. Interestingly, replacing only the 28 carboxy-terminal residues of CYP2C2 with those from CYP2C14 created a chimeric P450 capable of novel 16β-hydroxylation of testosterone [337,338]. A study in which CYP1A1 and CYP1A2 were randomly recombined using DNA shuffling yielding a library in which 11.8% of the chimeras

analyzed were functional on naphthalene [96]. The DNA-shuffled chimeric P450s exhibited altered substrate specificities [339].

Numerous studies have used chimeragenesis to study P450s and identify functionally important regions of their sequence [340-348]. In fact, replacement of ten residues of CYP2C5 with those from CYP2C3 increased the solubility of CYP2C5 [349], leading to the determination of its crystal structure, the first ever of a microsomal P450 [210]. Similarly, the solubility of CYP1A2 was increased using recombination with CYP102A1, who share only 16% amino acid identity while retaining catalytic activity [110]. Chimeras between CYP2B5 and CYP2B4 aided in determining amino acid residues responsible for these enzymes regioselectivity on androstenedione [348]. Another example is were a segment of CYP2C5 important for substrate binding was identified by examining chimeras made using CYP2C5 and CYP2C4 [347]. Other chimeric P450s reported in the literature include those between CYP2E1 and CYP101 [350], CYP11A1 and CYP27 [351], and CYP2C9 and CYP101 [352], between CYPC17 from humans and rats [353].

Finally, numerous fusions between P450 heme domains and non natural reductases have been constructed [354-358]. Of notable interest is the fusion of 2C11 from rat with the reductase from CYP102A1 [355]. 2C11 was not able to hydroxylate arachidonic acid when mixed separately with the CYP102A1 reductase, however, upon fusion was able to hydroxylate this substrate. A fusion of human CYP2E1 with the CYP102A1 reductase showed levels of activity equivalent the wild-type enzyme without the addition of lipids

or detergents [354]. The utility of this enzyme was shown in its production of the human metabolite of chlorzoxazone.

C.8. Applications of cytochromes P450

Cytochromes P450 have numerous potential applications [302,359,360]. One of the most famous and well publicized examples is the incorporation of P450s from petunias into transgenic 'blue' carnations [361] and slightly less 'blue' roses [362]. More practically, the ability of P450s to perform difficult C-H bond hydroxylation gives them many uses in stereo- and regio-selective hydroxylation for the production of bulk and fine chemicals. Use of P450s in the synthesis of leukotoxin B [363] and construction of an engineered yeast for the biosynthesis of hydrocortisone from a simple carbon source [361] demonstrate specific examples. Cytochrome P450 proteins also show promise in bioremediation [364]. Since one natural function is to facilitate the breakdown of chemicals, new P450s could be discovered or engineered for specific applications. For instance, an engineered CYP101 variant was able to oxidize the pollutant polychlorinated benzene [365].

As mentioned, P450s have great importance in the synthesis and discovery of drugs. For example, cytochrome P450 Taxadiene 5α-hydroxylase was cloned and functionally expressed in yeast to perform the first oxygenation step in the biosynthesis of Taxol [366-368]. Similarly, the conversion of compactin to pravastatin, a hypocholesterolaemic compound, is catalyzed by a P450 from *Mucor hiemalis* [369]. In drug development, P450s may be used for combinatorial biocatalysis whereby they are used to produce

libraries of new molecules that can be screened for useful compounds, e.g. introduction of P450 variants into antibiotic producing strains and screening them for their ability to kill other microorganisms [302] or the generation of new protein kinase inhibitors based on the indigoid scaffold [370]. Novel metabolites of warfarin and an angiotensin II antagonist not typically produced by humans were produced using fungal P450s [371,372]. Such novel metabolites may possess enhanced drug properties. Finally P450s can be used to activate pro-drugs [373] at a specific target, such as a tumor cell. These serve as a few of the many examples of applications for natural and engineered cytochromes P450.

D. References

1. Rost B (2002) Did evolution leap to create the protein universe? Curr. Opin. Struct. Biol. 12: 409-416.
2. Holm L, Sander C (1996) Mapping the protein universe. Science 273: 595-603.
3. Smith JM (1970) Natural selection and the concept of a protein space. Nature 225: 563-564.
4. Koonin EV, Wolf YI, Karev GP (2002) The structure of the protein universe and genome evolution. Nature 420: 218-223.
5. Axe DD (2004) Estimating the prevalence of protein sequences adopting functional enzyme folds. J. Mol. Biol. 341: 1295-1315.
6. Xia Y, Levitt M (2004) Funnel-like organization in sequence space determines the distributions of protein stability and folding rate preferred by evolution. Proteins 55: 107-114.
7. Holland HD (1997) Evidence for life on Earth more than 3850 million years ago. Science 275: 38-39.
8. Eigen M, Schuster P (1982) Stages of emerging life--five principles of early organization. J. Mol. Evol. 19: 47-61.
9. Abkevich VI, Gutin AM, Shakhnovich EI (1996) How the first biopolymers could have evolved. Proc. Natl. Acad. Sci. U. S. A. 93: 839-844.
10. Wright BE (2000) A biochemical mechanism for nonrandom mutations and evolution. J. Bacteriol. 182: 2993-3001.
11. Murzin AG (1996) Structural classification of proteins: new superfamilies. Curr. Opin. Struct. Biol. 6: 386-394.
12. Dayhoff MO (1976) The origin and evolution of protein superfamilies. Fed. Proc. 35: 2132-2138.
13. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, et al. (1997) CATH--a hierarchic classification of protein domain structures. Structure 5: 1093-1108.
14. Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247: 536-540.
15. Thompson JD, Higgins DG, Gibson TJ (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: 4673-4680.
16. Wallace IM, Blackshields G, Higgins DG (2005) Multiple sequence alignments. Curr. Opin. Struct. Biol. 15: 261-266.
17. Saraf MC, Moore GL, Maranas CD (2003) Using multiple sequence correlation analysis to characterize functionally important protein regions. Protein Eng. 16: 397-406.
18. 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: 83-90.

19. Probst WC, Snyder LA, Schuster DI, Brosius J, Sealfon SC (1992) Sequence alignment of the G-protein coupled receptor superfamily. DNA Cell Biol. 11: 1-20.
20. Lockless SW, Ranganathan R (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. Science 286: 295-299.
21. Larson SM, Di Nardo AA, Davidson AR (2000) Analysis of covariation in an SH3 domain sequence alignment: applications in tertiary contact prediction and the design of compensating hydrophobic core substitutions. J. Mol. Biol. 303: 433-446.
22. Res I, Mihalek I, Lichtarge O (2005) An evolution based classifier for prediction of protein interfaces without using protein structures. Bioinformatics 21: 2496-2501.
23. Mihalek I, Res I, Lichtarge O (2006) Evolutionary and structural feedback on selection of sequences for comparative analysis of proteins. Proteins 63: 87-99.
24. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH (2005) Why highly expressed proteins evolve slowly. Proc. Natl. Acad. Sci. U. S. A. 102: 14338-14343.
25. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, et al. (2003) Design of a novel globular protein fold with atomic-level accuracy. Science 302: 1364-1368.
26. Butterfoss GL, Kuhlman B (2006) Computer-based design of novel protein structures. Annu. Rev. Biophys. Biomol. Struct. 35: 49-65.
27. Kaur J, Sharma R (2006) Directed evolution: an approach to engineer enzymes. Crit. Rev. Biotechnol. 26: 165-199.
28. May O, Nguyen PT, Arnold FH (2000) Inverting enantioselectivity by directed evolution of hydantoinase for improved production of L-methionine. Nat. Biotechnol. 18: 317-320.
29. Alcalde M, Ferrer M, Plou FJ, Ballesteros A (2006) Environmental biocatalysis: from remediation with enzymes to novel green processes. Trends Biotechnol. 24: 281-287.
30. Schmid A, Dordick JS, Hauer B, Kiener A, Wubbolts M, et al. (2001) Industrial biocatalysis today and tomorrow. Nature 409: 258-268.
31. Bajpai P (2004) Biological bleaching of chemical pulps. Crit. Rev. Biotechnol. 24: 1-58.
32. Sheldon RA, van Rantwijk F (2004) Biocatalysis for sustainable organic synthesis. Aust. J. Chem. 57: 281-289.
33. Steinbuchel A (2005) Non-biodegradable biopolymers from renewable resources: perspectives and impacts. Curr. Opin. Biotechnol. 16: 607-613.
34. Hill CM, Li WS, Thoden JB, Holden HM, Raushel FM (2003) Enhanced degradation of chemical warfare agents through molecular engineering of the phosphotriesterase active site. J. Am. Chem. Soc. 125: 8990-8991.
35. Socolich M, Lockless SW, Russ WP, Lee H, Gardner KH, et al. (2005) Evolutionary information for specifying a protein fold. Nature 437: 512-518.
36. Bloom JD, Meyer MM, Meinhold P, Otey CR, MacMillan D, et al. (2005) Evolving strategies for enzyme engineering. Curr. Opin. Struct. Biol. 15: 447-452.
37. Russ WP, Ranganathan R (2002) Knowledge-based potential functions in protein design. Curr. Opin. Struct. Biol. 12: 447-452.

38. Koga Y, Kato K, Nakano H, Yamane T (2003) Inverting enantioselectivity of Burkholderia cepacia KWI-56 lipase by combinatorial mutation and high-throughput screening using single-molecule PCR and in vitro expression. J. Mol. Biol. 331: 585-592.
39. Zha DX, Wilensek S, Hermes M, Jaeger KE, Reetz MT (2001) Complete reversal of enantioselectivity of an enzyme-catalyzed reaction by directed evolution. Chem. Commun.: 2664-2665.
40. Gabor EM, Janssen DB (2004) Increasing the synthetic performance of penicillin acylase PAS2 by structure-inspired semi-random mutagenesis. Protein Eng. Des. Sel. 17: 571-579.
41. Tang L, Torres Pazmino DE, Fraaije MW, de Jong RM, Dijkstra BW, et al. (2005) Improved catalytic properties of halohydrin dehalogenase by modification of the halide-binding site. Biochemistry 44: 6609-6618.
42. Desjarlais JR, Handel TM (1995) De novo design of the hydrophobic cores of proteins. Protein Sci. 4: 2006-2018.
43. Dahiyat BI, Mayo SL (1997) De novo protein design: fully automated sequence selection. Science 278: 82-87.
44. Jiang X, Bishop EJ, Farid RS (1997) A de novo designed protein with properties that characterize natural hyperthermophilic proteins. J. Am. Chem. Soc. 119: 838-839.
45. Jiang X, Farid H, Pistor E, Farid RS (2000) A new approach to the design of uniquely folded thermally stable proteins. Protein Sci. 9: 403-416.
46. Gibney BR, Rabanal F, Skalicky JJ, Wand AJ, Dutton PL (1999) Iterative protein redesign. J. Am. Chem. Soc. 121: 4952-4960.
47. Braisted AC, Wells JA (1996) Minimizing a binding domain from protein A. Proc. Natl. Acad. Sci. U. S. A. 93: 5688-5692.
48. Arndt KM, Pelletier JN, Muller KM, Alber T, Michnick SW, et al. (2000) A heterodimeric coiled-coil peptide pair selected in vivo from a designed library-versus-library ensemble. J. Mol. Biol. 295: 627-639.
49. Schnepf R, Hoerth P, Bill E, Wieghardt K, Hildebrandt P, et al. (2001) De novo design and characterization of copper centers in synthetic four-helix-bundle proteins. J. Am. Chem. Soc. 123: 2186-2195.
50. Robertson DE, Farid RS, Moser CC, Urbauer JL, Mulholland SE, et al. (1994) Design and synthesis of multi-haem proteins. Nature 368: 425-432.
51. Gibney BR, Dutton PL (1999) Histidine placement in de novo-designed heme proteins. Protein Sci. 8: 1888-1898.
52. Arnold PA, Shelton WR, Benson DR (1997) Peptide helix induction in a self-assembling hemoprotein model. J. Am. Chem. Soc. 119: 3181-3182.
53. Rojas NR, Kamtekar S, Simons CT, McLean JE, Vogel KM, et al. (1997) De novo heme proteins from designed combinatorial libraries. Protein Sci. 6: 2512-2524.
54. Arnold FH (2001) Combinatorial and computational challenges for biocatalyst design. Nature 409: 253-257.
55. Hibbert EG, Dalby PA (2005) Directed evolution strategies for improved enzymatic performance. Microb. Cell Fact. 4: 29.
56. Lutz S, Patrick WM (2004) Novel methods for directed evolution of enzymes: quality, not quantity. Curr. Opin. Biotechnol. 15: 291-297.

57. Johannes TW, Zhao H (2006) Directed evolution of enzymes and biosynthetic pathways. Curr. Opin. Microbiol. 9: 261-267.
58. Castle LA, Siehl DL, Gorton R, Patten PA, Chen YH, et al. (2004) Discovery and directed evolution of a glyphosate tolerance gene. Science 304: 1151-1154.
59. Zhao H, Arnold FH (1999) Directed evolution converts subtilisin E into a functional equivalent of thermitase. Protein Eng. 12: 47-53.
60. Giver L, Gershenson A, Freskgard PO, Arnold FH (1998) Directed evolution of a thermostable esterase. Proc. Natl. Acad. Sci. U. S. A. 95: 12809-12813.
61. Miyazaki K, Takenouchi M, Kondo H, Noro N, Suzuki M, et al. (2006) Thermal stabilization of Bacillus subtilis family-11 xylanase by directed evolution. J. Biol. Chem. 281: 10236-10242.
62. Seng Wong T, Arnold FH, Schwaneberg U (2004) Laboratory evolution of cytochrome p450 BM-3 monooxygenase for organic cosolvents. Biotechnol. Bioeng. 85: 351-358.
63. You L, Arnold FH (1996) Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide. Protein Eng. 9: 77-83.
64. Jaeger KE, Eggert T (2004) Enantioselective biocatalysis optimized by directed evolution. Curr. Opin. Biotechnol. 15: 305-313.
65. Park S, Morley KL, Horsman GP, Holmquist M, Hult K, et al. (2005) Focusing mutations into the P. fluorescens esterase binding site increases enantioselectivity more effectively than distant mutations. Chem. Biol. 12: 45-54.
66. Reetz MT (2004) Controlling the enantioselectivity of enzymes by directed evolution: practical and theoretical ramifications. Proc. Natl. Acad. Sci. U. S. A. 101: 5716-5722.
67. Cirino PC, Arnold FH (2003) A self-sufficient peroxide-driven hydroxylation biocatalyst. Angew. Chem. Int. Ed. Engl. 42: 3299-3301.
68. Stemmer WP (1994) Rapid evolution of a protein in vitro by DNA shuffling. Nature 370: 389-391.
69. Aharoni A, Gaidukov L, Yagur S, Toker L, Silman I, et al. (2004) Directed evolution of mammalian paraoxonases PON1 and PON3 for bacterial expression and catalytic specialization. Proc. Natl. Acad. Sci. U. S. A. 101: 482-487.
70. Matsumura I, Ellington AD (2001) In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. J. Mol. Biol. 305: 331-339.
71. Doyon JB, Pattanayak V, Meyer CB, Liu DR (2006) Directed evolution and substrate specificity profile of homing endonuclease I-SceI. J. Am. Chem. Soc. 128: 2477-2484.
72. Hsu CC, Hong Z, Wada M, Franke D, Wong CH (2005) Directed evolution of D-sialic acid aldolase to L-3-deoxy-manno-2-octulosonic acid (L-KDO) aldolase. Proc. Natl. Acad. Sci. U. S. A. 102: 9122-9126.
73. Park HS, Nam SH, Lee JK, Yoon CN, Mannervik B, et al. (2006) Design and evolution of new catalytic activity with an existing protein scaffold. Science 311: 535-538.
74. Dwyer MA, Looger LL, Hellinga HW (2003) Computational design of a Zn2+ receptor that controls bacterial gene expression. Proc. Natl. Acad. Sci. U. S. A. 100: 11255-11260.

75. Dwyer MA, Looger LL, Hellinga HW (2004) Computational design of a biologically active enzyme. Science 304: 1967-1971.
76. Looger LL, Dwyer MA, Smith JJ, Hellinga HW (2003) Computational design of receptor and sensor proteins with novel functions. Nature 423: 185-190.
77. Keefe AD, Szostak JW (2001) Functional proteins from a random-sequence library. Nature 410: 715-718.
78. Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH (1993) Protein design by binary patterning of polar and nonpolar amino acids. Science 262: 1680-1685.
79. Wei Y, Liu T, Sazinsky SL, Moffet DA, Pelczer I, et al. (2003) Stably folded de novo proteins from a designed combinatorial library. Protein Sci. 12: 92-102.
80. West MW, Wang W, Patterson J, Mancias JD, Beasley JR, et al. (1999) De novo amyloid proteins from designed combinatorial libraries. Proc. Natl. Acad. Sci. U. S. A. 96: 11211-11216.
81. Davidson AR, Sauer RT (1994) Folded proteins occur frequently in libraries of random amino acid sequences. Proc. Natl. Acad. Sci. U. S. A. 91: 2146-2150.
82. Hecht MH, Richardson JS, Richardson DC, Ogden RC (1990) De novo design, expression, and characterization of Felix: a four-helix bundle protein of native-like sequence. Science 249: 884-891.
83. Regan L, DeGrado WF (1988) Characterization of a helical protein designed from first principles. Science 241: 976-978.
84. Darwin C (1859) On the origin of species by means of natural selection, or the preservation of vavoured races in the struggle for life. London: John Murray.
85. Nelson RS, Winter S (1982) An Evolutionary theory of Economic Change. Cambridge, MA: Belknap Press.
86. Mitchell M (1998) An Introduction to Genetic Algorithms. Cambridge, MA: MIT Press.
87. Graveley BR (2001) Alternative splicing: increasing diversity in the proteomic world. Trends Genet. 17: 100-107.
88. Gilbert W, de Souza SJ, Long M (1997) Origin of genes. Proc. Natl. Acad. Sci. U. S. A. 94: 7698-7703.
89. de Souza SJ, Long M, Gilbert W (1996) Introns and gene evolution. Genes Cells 1: 493-505.
90. Das S, Smith TF (2000) Identifying nature's protein Lego set. Adv. Protein Chem. 54: 159-183.
91. Guo HH, Choe J, Loeb LA (2004) Protein tolerance to random amino acid change. Proc. Natl. Acad. Sci. U. S. A. 101: 9205-9210.
92. Bloom JD, Silberg JJ, Wilke CO, Drummond DA, Adami C, et al. (2005) Thermodynamic prediction of protein neutrality. Proc. Natl. Acad. Sci. U. S. A. 102: 606-611.
93. Drummond DA, Silberg JJ, Meyer MM, Wilke CO, Arnold FH (2005) On the conservative nature of intragenic recombination. Proc. Natl. Acad. Sci. U. S. A. 102: 5380-5385.
94. Crameri A, Raillard SA, Bermudez E, Stemmer WP (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. Nature 391: 288-291.

95. Stemmer WP (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc. Natl. Acad. Sci. U. S. A. 91: 10747-10751.
96. Abecassis V, Pompon D, Truan G (2000) High efficiency family shuffling based on multi-step PCR and in vivo DNA recombination in yeast: statistical and functional analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2. Nucleic Acids Res. 28: E88.
97. Kauffman S (1993) The origins of order. New York: Oxford University Press.
98. Ness JE, Del Cardayre SB, Minshull J, Stemmer WP (2000) Molecular breeding: the natural approach to protein design. Adv. Protein Chem. 55: 261-292.
99. Chang CC, Chen TT, Cox BW, Dawes GN, Stemmer WP, et al. (1999) Evolution of a cytokine using DNA family shuffling. Nat. Biotechnol. 17: 793-797.
100. Ness JE, Welch M, Giver L, Bueno M, Cherry JR, et al. (1999) DNA shuffling of subgenomic sequences of subtilisin. Nat. Biotechnol. 17: 893-896.
101. Raillard S, Krebber A, Chen Y, Ness JE, Bermudez E, et al. (2001) Novel enzyme activities and functional plasticity revealed by recombining highly homologous enzymes. Chem. Biol. 8: 891-898.
102. Zhang JH, Dawes G, Stemmer WP (1997) Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. Proc. Natl. Acad. Sci. U. S. A. 94: 4504-4509.
103. Hansson LO, Bolton-Grob R, Massoud T, Mannervik B (1999) Evolution of differential substrate specificities in Mu class glutathione transferases probed by DNA shuffling. J. Mol. Biol. 287: 265-276.
104. Coco WM, Levinson WE, Crist MJ, Hektor HJ, Darzins A, et al. (2001) DNA shuffling method for generating highly recombined genes and evolved enzymes. Nat. Biotechnol. 19: 354-359.
105. Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. Nat. Biotechnol. 16: 258-261.
106. Volkov AA, Shao Z, Arnold FH (1999) Recombination and chimeragenesis by in vitro heteroduplex formation and in vivo repair. Nucleic Acids Res. 27: e18.
107. Ness JE, Kim S, Gottman A, Pak R, Krebber A, et al. (2002) Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. Nat. Biotechnol. 20: 1251-1255.
108. Zha D, Eipper A, Reetz MT (2003) Assembly of designed oligonucleotides as an efficient method for gene recombination: a new tool in directed evolution. Chembiochem 4: 34-39.
109. Gibbs MD, Nevalainen KM, Bergquist PL (2001) Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling. Gene 271: 13-20.
110. Sieber V, Martinez CA, Arnold FH (2001) Libraries of hybrid proteins from distantly related sequences. Nat. Biotechnol. 19: 456-460.
111. Ostermeier M, Shim JH, Benkovic SJ (1999) A combinatorial approach to hybrid enzymes independent of DNA homology. Nat. Biotechnol. 17: 1205-1209.

112. Lutz S, Ostermeier M, Moore GL, Maranas CD, Benkovic SJ (2001) Creating multiple-crossover DNA libraries independent of sequence identity. Proc. Natl. Acad. Sci. U. S. A. 98: 11248-11253.
113. Hiraga K, Arnold FH (2003) General method for sequence-independent site-directed chimeragenesis. J. Mol. Biol. 330: 287-296.
114. Neylon C (2004) Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. Nucleic Acids Res. 32: 1448-1459.
115. Saraf MC, Horswill AR, Benkovic SJ, Maranas CD (2004) FamClash: a method for ranking the activity of engineered enzymes. Proc Natl Acad Sci U S A 101: 4142-4147.
116. Pantazes RJ, Saraf MC, Maranas CD (2007) Optimal protein library design using recombination or point mutations based on sequence-based scoring functions. Protein Eng. Des. Sel. 20: 361-373.
117. Saraf MC, Moore GL, Goodey NM, Cao VY, Benkovic SJ, et al. (2006) IPRO: an iterative computational protein library redesign and optimization procedure. Biophys. J. 90: 4167-4180.
118. Voigt CA, Martinez C, Wang ZG, Mayo SL, Arnold FH (2002) Protein building blocks preserved by recombination. Nat. Struct. Biol. 9: 553-558.
119. Silberg JJ, Endelman JB, Arnold FH (2004) SCHEMA-guided protein recombination. Methods Enzymol. 388: 35-42.
120. Meyer MM, Silberg JJ, Voigt CA, Endelman JB, Mayo SL, et al. (2003) Library analysis of SCHEMA-guided protein recombination. Protein Sci. 12: 1686-1693.
121. Ortiz de Montellano PR (1995) Cytochrome P450: Structure, Mechanism, and Biochemistry. New York: Plenum Press.
122. Guengerich FP (2001) Uncommon P450-catalyzed reactions. Curr. Drug Metab. 2: 93-115.
123. Guengerich FP (2001) Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. Chem. Res. Toxicol. 14: 611-650.
124. Nebert DW, Adesnik M, Coon MJ, Estabrook RW, Gonzalez FJ, et al. (1987) The P450 gene superfamily: recommended nomenclature. DNA 6: 1-11.
125. Nelson DR, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, et al. (1996) P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. Pharmacogenetics 6: 1-42.
126. Nelson DR, Kamataki T, Waxman DJ, Guengerich FP, Estabrook RW, 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: 1-51.
127. Nebert DW, Nelson DR (1991) P450 gene nomenclature based on evolution. Methods Enzymol. 206: 3-11.
128. Nelson DR (2006) Cytochrome P450 nomenclature. Methods Mol. Biol. 320: 1-10.
129. Nebert DW, Nelson DR, Feyereisen R (1989) Evolution of the cytochrome P450 genes. Xenobiotica 19: 1149-1160.
130. Kanaly RA, Hur HG (2005) Growth of Phanerochaete chrysosporium on diesel fuel hydrocarbons at neutral pH. Chemosphere.

131. van Beilen JB, Funhoff EG, van Loon A, Just A, Kaysser L, et al. (2006) Cytochrome P450 alkane hydroxylases of the CYP153 family are common in alkane-degrading eubacteria lacking integral membrane alkane hydroxylases. Appl. Environ. Microbiol. 72: 59-65.
132. Mauersberger S, Schunck WH, Muller HG (1984) The induction of cytochrome P-450 in the alkane-utilizing yeast Lodderomyces elongisporus: Alterations in the microsomal membrane fraction. Appl. Microbiol. Biotechnol. 19: 29-35.
133. Wiedmann B, Wiedmann M, Kargel E, Schunck WH, Muller HG (1986) N-alkanes induce the synthesis of cytochrome P-450 mRNA in Candida maltosa. Biochem. Biophys. Res. Commun. 136: 1148-1154.
134. Sanglard D, Chen C, Loper JC (1987) Isolation of the alkane inducible cytochrome P450 (P450alk) gene from the yeast Candida tropicalis. Biochem. Biophys. Res. Commun. 144: 251-257.
135. Lebeault JM, Lode ET, Coon MJ (1971) Fatty acid and hydrocarbon hydroxylation in yeast: role of cytochrome P-450 in Candida tropicalis. Biochem. Biophys. Res. Commun. 42: 413-419.
136. Woods LF, Wiseman A (1980) Benzo(a)pyrene hydroxylase from Saccharomyces cerevisiae. Substrate binding, spectral and kinetic data. Biochim. Biophys. Acta. 613: 52-61.
137. Grogan G, Roberts GA, Parsons S, Turner NJ, Flitsch SL (2002) P450(camr), a cytochrome P450 catalysing the stereospecific 6- endo-hydroxylation of (1 R)-(+)-camphor. Appl. Microbiol. Biotechnol. 59: 449-454.
138. Funk C, Koepp AE, Croteau R (1992) Catabolism of camphor in tissue cultures and leaf disks of common sage (Salvia officinalis). Arch. Biochem. Biophys. 294: 306-313.
139. Katagiri M, Ganguli BN, Gunsalus IC (1968) A soluble cytochrome P-450 functional in methylene hydroxylation. J. Biol. Chem. 243: 3543-3546.
140. Hedegaard J, Gunsalus IC (1965) Mixed function oxidation. IV. An induced methylene hydroxylase in camphor oxidation. J. Biol. Chem. 240: 4038-4043.
141. Peterson JA, Lu JY, Geisselsoder J, Graham-Lorence S, Carmona C, et al. (1992) Cytochrome P-450terp. Isolation and purification of the protein and cloning and sequencing of its operon. J. Biol. Chem. 267: 14193-14203.
142. Renganathan V, Madyastha KM (1984) Linalyl acetate is metabolized by *Pseudomonas incognita* with the acetoxy group intact. Appl. Environ. Microbiol. 45: 6-15.
143. Madyastha K, Bhattacharyya PK, Vaidyanathan CS (1977) Metabolism of monoterpene alcohol, linalool, by a soil pseudomonad. Can. J. Microbiol. 23: 230-239.
144. Hawkes DB, Adams GW, Burlingame AL, Ortiz de Montellano PR, De Voss JJ (2002) Cytochrome P450(cin) (CYP176A), isolation, expression, and characterization. J. Biol. Chem. 277: 27725-27732.
145. Guengerich FP, Wu ZL, Bartleson CJ (2005) Function of human cytochrome P450s: characterization of the orphans. Biochem. Biophys. Res. Commun. 338: 465-469.
146. Lewis DF (2004) 57 varieties: the human cytochromes P450. Pharmacogenomics 5: 305-318.

147. Rendic S (2002) Summary of information on human CYP enzymes: human P450 metabolism data. Drug Metab. Rev. 34: 83-448.
148. Rendic S, Di Carlo FJ (1997) Human cytochrome P450 enzymes: a status report summarizing their reactions, substrates, inducers, and inhibitors. Drug Metab. Rev. 29: 413-580.
149. Guengerich FP (2003) Cytochromes P450, drugs, and diseases. Mol. Interv. 3: 194-204.
150. Daly AK, Brockmoller J, Broly F, Eichelbaum M, Evans WE, et al. (1996) Nomenclature for human CYP2D6 alleles. Pharmacogenetics 6: 193-201.
151. Davit B, Reynolds K, Yuan R, Ajayi F, Conner D, et al. (1999) FDA evaluations using in vitro metabolism to predict and interpret in vivo metabolic drug-drug interactions: impact on labeling. J. Clin. Pharmacol. 39: 899-910.
152. Parkinson A (2003) Biotransformation of xenobiotics. In: Klaassen CD, Watkins JB, editors. Essentials of Toxicology. New York: McGraw-Hill Company, Inc.
153. von Moltke LL, Greenblatt DJ, Duan SX, Harmatz JS, Shader RI (1994) In vitro prediction of the terfenadine-ketoconazole pharmacokinetic interaction. J. Clin. Pharmacol. 34: 1222-1227.
154. Markham A, Wagstaff AJ (1998) Fexofenadine. Drugs 55: 269-276.
155. SoRelle R (1998) Withdrawal of posicor from market. Circulation 98.
156. Benedetti MS, Bani M (1998) Design of in vitro studies to predct in vivo inhibitory drug-drug interactions. Pharmacological Res. 38: 81-88.
157. Nelson DR (1999) Cytochrome P450 and the individuality of species. Arch. Biochem. Biophys. 369: 1-10.
158. Butler MA, Iwasaki M, Guengerich FP, Kadlubar FF (1989) Human cytochrome P-450PA (P-450IA2), the phenacetin O-deethylase, is primarily responsible for the hepatic 3-demethylation of caffeine and N-oxidation of carcinogenic arylamines. Proc. Natl. Acad. Sci. U. S. A. 86: 7696-7700.
159. Nebert DW, Russell DW (2002) Clinical importance of the cytochromes P450. Lancet 360: 1155-1162.
160. Pullinger CR, Eng C, Salen G, Shefer S, Batta AK, et al. (2002) Human cholesterol 7alpha-hydroxylase (CYP7A1) deficiency has a hypercholesterolemic phenotype. J. Clin. Invest. 110: 109-117.
161. Hennan JK, Huang J, Barrett TD, Driscoll EM, Willens DE, et al. (2001) Effects of selective cyclooxygenase-2 inhibition on vascular responses and thrombosis in canine coronary arteries. Circulation 104: 820-825.
162. Tuder RM, Cool CD, Yeager M, Taraseviciene-Stewart L, Bull TM, et al. (2001) The pathobiology of pulmonary hypertension. Endothelium. Clin. Ches.t Med. 22: 405-418.
163. Mueller GC, Miller JA (1953) The metabolism of methylated aminoazo dyes. II. Oxidative demethylation by rat liver homogenates. J. Biol. Chem. 202: 579-587.
164. Rodriguez-Antona C, Ingelman-Sundberg M (2006) Cytochrome P450 pharmacogenetics and cancer. Oncogene 25: 1679-1691.
165. Schuler MA (1996) The role of cytochrome P450 monooxygenases in plant-insect interactions. Plant Physiol. 112: 1411-1419.
166. Durst F, O'Keefe DP (1995) Plant cytochromes P450: an overview. Drug Metabol. Drug Interact. 12: 171-187.

167. Chou WM, Kutchan TM (1998) Enzymatic oxidations in the biosynthesis of complex alkaloids. Plant J. 15: 289-300.
168. Tudzynski B, Hedden P, Carrera E, Gaskin P (2001) The P450-4 gene of Gibberella fujikuroi encodes ent-kaurene oxidase in the gibberellin biosynthesis pathway. Appl. Environ. Microbiol. 67: 3514-3522.
169. Schuler MA, Werck-Reichhart D (2003) Functional genomics of P450s. Annu. Rev. Plant Biol. 54: 629-667.
170. Paquette SM, Bak S, Feyereisen R (2000) Intron-exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of Arabidopsis thaliana. DNA Cell. Biol. 19: 307-317.
171. Nelson DR, Schuler MA, Paquette SM, Werck-Reichhart D, Bak S (2004) Comparative genomics of rice and Arabidopsis. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. Plant Physiol. 135: 756-772.
172. Frey M, Chomet P, Glawischnig E, Stettner C, Grun S, et al. (1997) Analysis of a chemical plant defense mechanism in grasses. Science 277: 696-699.
173. Glawischnig E, Grun S, Frey M, Gierl A (1999) Cytochrome P450 monooxygenases of DIBOA biosynthesis: specificity and conservation among grasses. Phytochemistry 50: 925-930.
174. Persans MW, Wang J, Schuler MA (2001) Characterization of maize cytochrome P450 monooxygenases induced in response to safeners and bacterial pathogens. Plant Physiol. 125: 1126-1138.
175. McLean KJ, Sabri M, Marshall KR, Lawson RJ, Lewis DG, et al. (2005) Biodiversity of cytochrome P450 redox systems. Biochem. Soc. Trans. 33: 796-801.
176. De Mot R, Parret AH (2002) A novel class of self-sufficient cytochrome P450 monooxygenases in prokaryotes. Trends Microbiol. 10: 502-508.
177. Gunsalus IC, Pederson TC, Sligar SG (1975) Oxygenase-catalyzed biological hydroxylations. Annu. Rev. Biochem. 44: 377-407.
178. Roberts GA, Grogan G, Greter A, Flitsch SL, Turner NJ (2002) Identification of a new class of cytochrome P450 from a Rhodococcus sp. J. Bacteriol. 184: 3898-3908.
179. Miles CS, Ost TW, Noble MA, Munro AW, Chapman SK (2000) Protein engineering of cytochromes P-450. Biochim. Biophys. Acta 1543: 383-407.
180. Roberts GA, Celik A, Hunter DJ, Ost TW, White JH, et al. (2003) A self-sufficient cytochrome p450 with a primary structural organization that includes a flavin domain and a [2Fe-2S] redox center. J. Biol. Chem. 278: 48914-48920.
181. Liu L, Schmid RD, Urlacher VB (2006) Cloning, expression, and characterization of a self-sufficient cytochrome P450 monooxygenase from Rhodococcus ruber DSM 44319. Appl. Microbiol. Biotechnol. 72: 876-882.
182. Udit AK, Hill MG, Bittner VG, Arnold FH, Gray HB (2004) Reduction of dioxygen catalyzed by pyrene-wired heme domain cytochrome P450 BM3 electrodes. J. Am. Chem. Soc. 126: 10218-10219.
183. Udit AK, Arnold FH, Gray HB (2004) Cobaltocene-mediated catalytic monooxygenation using holo and heme domain cytochrome P450 BM3. J. Inorg. Biochem. 98: 1547-1550.

184. Schwaneberg U, Appel D, Schmitt J, Schmid RD (2000) P450 in biotechnology: zinc driven omega-hydroxylation of p-nitrophenoxydodecanoic acid using P450 BM-3 F87A as a catalyst. J. Biotechnol. 84: 249-257.
185. Schlichting I, Jung C, Schulze H (1997) Crystal structure of cytochrome P-450cam complexed with the (1S)-camphor enantiomer. FEBS Lett. 415: 253-257.
186. Schlichting I, Berendzen J, Chu K, Stock AM, Maves SA, et al. (2000) The catalytic pathway of cytochrome p450cam at atomic resolution. Science 287: 1615-1622.
187. Nagano S, Tosha T, Ishimori K, Morishima I, Poulos TL (2004) Crystal structure of the cytochrome p450cam mutant that exhibits the same spectral perturbations induced by putidaredoxin binding. J. Biol. Chem. 279: 42844-42849.
188. Ravichandran KG, Boddupalli SS, Hasermann CA, Peterson JA, Deisenhofer J (1993) Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. Science 261: 731-736.
189. Li H, Poulos TL (1997) The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. Nat. Struct. Biol. 4: 140-146.
190. Haines DC, Tomchick DR, Machius M, Peterson JA (2001) Pivotal role of water in the mechanism of P450BM-3. Biochemistry 40: 13456-13465.
191. Boddupalli SS, Hasemann CA, Ravichandran KG, Lu JY, Goldsmith EJ, et al. (1992) Crystallization and preliminary x-ray diffraction analysis of P450terp and the hemoprotein domain of P450BM-3, enzymes belonging to two distinct classes of the cytochrome P450 superfamily. Proc. Natl. Acad. Sci. U. S. A. 89: 5567-5571.
192. Hasemann CA, Ravichandran KG, Peterson JA, Deisenhofer J (1994) Crystal structure and refinement of cytochrome P450terp at 2.3 A resolution. J. Mol. Biol. 236: 1169-11685.
193. Cupp-Vickery JR, Garcia C, Hofacre A, McGee-Estrada K (2001) Ketoconazole-induced conformational changes in the active site of cytochrome P450eryF. J. Mol. Biol. 311: 101-110.
194. Cupp-Vickery JR, Poulos TL (1995) Structure of cytochrome P450eryF involved in erythromycin biosynthesis. Nat. Struct. Biol. 2: 144-153.
195. Cupp-Vickery JR, Poulos TL (1997) Structure of cytochrome P450eryF: substrate, inhibitors, and model compounds bound in the active site. Steroids 62: 112-116.
196. Yano JK, Koo LS, Schuller DJ, Li H, Ortiz de Montellano PR, et al. (2000) Crystal structure of a thermophilic cytochrome P450 from the archaeon Sulfolobus solfataricus. J. Biol. Chem. 275: 31086-31092.
197. Podust LM, Kim Y, Arase M, Neely BA, Beck BJ, et al. (2003) The 1.92-A structure of Streptomyces coelicolor A3(2) CYP154C1. A new monooxygenase that functionalizes macrolide ring systems. J. Biol. Chem. 278: 12214-12221.
198. Lee DS, Yamada A, Sugimoto H, Matsunaga I, Ogura H, et al. (2003) Substrate recognition and molecular mechanism of fatty acid hydroxylation by cytochrome P450 from Bacillus subtilis. Crystallographic, spectroscopic, and mutational studies. J. Biol. Chem. 278: 9761-9767.
199. Yano JK, Blasco F, Li H, Schmid RD, Henne A, et al. (2003) Preliminary characterization and crystal structure of a thermostable cytochrome P450 from Thermus thermophilus. J. Biol. Chem. 278: 608-616.

200. Leys D, Mowat CG, McLean KJ, Richmond A, Chapman SK, et al. (2003) Atomic structure of Mycobacterium tuberculosis CYP121 to 1.06 A reveals novel features of cytochrome P450. J. Biol. Chem. 278: 5141-5147.
201. Sherman DH, Li S, Yermalitskaya LV, Kim Y, Smith JA, et al. (2006) The structural basis for substrate anchoring, active site selectivity, and product formation by P450 PikC from Streptomyces venezuelae. J. Biol. Chem. 281: 26289-26297.
202. Podust LM, Bach H, Kim Y, Lamb DC, Arase M, et al. (2004) Comparison of the 1.85 A structure of CYP154A1 from Streptomyces coelicolor A3(2) with the closely related CYP154C1 and CYPs from antibiotic biosynthetic pathways. Protein Sci. 13: 255-268.
203. Nagano S, Li H, Shimizu H, Nishida C, Ogura H, et al. (2003) Crystal structures of epothilone D-bound, epothilone B-bound, and substrate-free forms of cytochrome P450epoK. J. Biol. Chem. 278: 44886-44893.
204. Pylypenko O, Vitali F, Zerbe K, Robinson JA, Schlichting I (2003) Crystal structure of OxyC, a cytochrome P450 implicated in an oxidative C-C coupling reaction during vancomycin biosynthesis. J. Biol. Chem. 278: 46727-46733.
205. Zerbe K, Pylypenko O, Vitali F, Zhang W, Rouset S, et al. (2002) Crystal structure of OxyB, a cytochrome P450 implicated in an oxidative phenol coupling reaction during vancomycin biosynthesis. J. Biol. Chem. 277: 47476-47485.
206. Xu F, Bell SG, Rao Z, Wong LL (2007) Crystal structure of CYP199A2, a para-substituted benzoic acid oxidizing cytochrome P450 enzyme from Rhodopseudomonas palustris. To be published.
207. Strushkevich NV, Min J, Loppnau P, Tempel W, Arrowsmith CH, et al. (2007) Structural analysis of CYP2R1 in complex with vitamin D3. To be published.
208. Shimizu H, Obayashi E, Gomi Y, Arakawa H, Park SY, et al. (2000) Proton delivery in NO reduction by fungal nitric-oxide reductase. Cryogenic crystallography, spectroscopy, and kinetics of ferric-NO complexes of wild-type and mutant enzymes. J. Biol. Chem. 275: 4816-4826.
209. Podust LM, Poulos TL, Waterman MR (2001) Crystal structure of cytochrome P450 14alpha -sterol demethylase (CYP51) from Mycobacterium tuberculosis in complex with azole inhibitors. Proc. Natl. Acad. Sci. U. S. A. 98: 3068-3073.
210. Williams PA, Cosme J, Sridhar V, Johnson EF, McRee DE (2000) Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. Mol. Cell. 5: 121-131.
211. Wester MR, Johnson EF, Marques-Soares C, Dansette PM, Mansuy D, et al. (2003) Structure of a substrate complex of mammalian cytochrome P450 2C5 at 2.3 A resolution: evidence for multiple substrate binding modes. Biochemistry 42: 6370-6379.
212. Wester MR, Johnson EF, Marques-Soares C, Dijols S, Dansette PM, et al. (2003) Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 A resolution: evidence for an induced fit model of substrate binding. Biochemistry 42: 9335-9345.
213. Williams PA, Cosme J, Ward A, Angove HC, Matak Vinkovic D, et al. (2003) Crystal structure of human cytochrome P450 2C9 with bound warfarin. Nature 424: 464-468.

214. Wester MR, Yano JK, Schoch GA, Yang C, Griffin KJ, et al. (2004) The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-A resolution. J. Biol. Chem. 279: 35630-35637.
215. Williams PA, Cosme J, Vinkovic DM, Ward A, Angove HC, et al. (2004) Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. Science 305: 683-686.
216. Yano JK, Wester MR, Schoch GA, Griffin KJ, Stout CD, et al. (2004) The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-A resolution. J. Biol. Chem. 279: 38091-38094.
217. Scott EE, White MA, He YA, Johnson EF, Stout CD, et al. (2004) Structure of mammalian cytochrome P450 2B4 complexed with 4-(4-chlorophenyl)imidazole at 1.9-A resolution: insight into the range of P450 conformations and the coordination of redox partner binding. J. Biol. Chem. 279: 27294-27301.
218. Scott EE, He YA, Wester MR, White MA, Chin CC, et al. (2003) An open conformation of mammalian cytochrome P450 2B4 at 1.6-A resolution. Proc. Natl. Acad. Sci. U. S. A. 100: 13196-13201.
219. Schoch GA, Yano JK, Wester MR, Griffin KJ, Stout CD, et al. (2004) Structure of human microsomal cytochrome P450 2C8. Evidence for a peripheral fatty acid binding site. J. Biol. Chem. 279: 9497-9503.
220. Yano JK, Hsu MH, Griffin KJ, Stout CD, Johnson EF (2005) Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen. Nat. Struct. Mol. Biol. 12: 822-823.
221. Rowland P, Blaney FE, Smyth MG, Jones JJ, Leydon VR, et al. (2006) Crystal structure of human cytochrome P450 2D6. J. Biol. Chem. 281: 7614-7622.
222. Gotoh O, Tagashira Y, Iizuka T, Fujii-Kuriyama Y (1983) Structural characteristics of cytochrome P-450. Possible location of the heme-binding cysteine in determined amino-acid sequences. J Biochem (Tokyo) 93: 807-817.
223. Gotoh O, Fujii-Kuriyama Y (1989) Evolution, structure and gene regulation of cytochrome P450. In: Ruckpaul K, Rein H, editors. Frontiers in Biotransformation. Berlin: Akademie Verlag. pp. 195-243.
224. Gotoh O (1993) Evolution and differentiation of P-450 genes. Cytochromes P-450. 2nd ed. Weinheim: VCH. pp. 255-272.
225. Wickramasinghe RH, Villee CA (1975) Early role during chemical evolution for cytochrome P450 in oxygen detoxification. Nature 256: 509-511.
226. Degtyarenko KN, Archakov AI (1993) Molecular evolution of P450 superfamily and P450-containing monooxygenase systems. FEBS Lett. 332: 1-8.
227. Nelson DR, Strobel HW (1987) Evolution of cytochrome P-450 proteins. Mol. Biol. Evol. 4: 572-593.
228. Gonzalez FJ, Nebert DW (1990) Evolution of the P450 gene superfamily: animal-plant 'warfare', molecular drive and human genetic differences in drug oxidation. Trends Genet. 6: 182-186.
229. Li X, Baudry J, Berenbaum MR, Schuler MA (2004) Structural and functional divergence of insect CYP6B proteins: From specialist to generalist cytochrome P450. Proc. Natl. Acad. Sci. U. S. A. 101: 2939-2944.
230. Kelly SL, Lamb DC, Jackson CJ, Warrilow AG, Kelly DE (2003) The biodiversity of microbial cytochromes P450. Adv. Microb. Physiol. 47: 131-186.

231. Hasemann CA, Kurumbail RG, Boddupalli SS, Peterson JA, Deisenhofer J (1995) Structure and function of cytochromes P450: a comparative analysis of three crystal structures. Structure 3: 41-62.
232. Werck-Reichhart D, Feyereisen R (2000) Cytochromes P450: a success story. Genome Biol 1: REVIEWS3003.
233. Mestres J (2005) Structure conservation in cytochromes P450. Proteins 58: 596-609.
234. Nagano S, Poulos TL (2005) Crystallographic study on the dioxygen complex of wild-type and mutant cytochrome P450cam. Implications for the dioxygen activation mechanism. J. Biol. Chem. 280: 31659-31663.
235. Nagano S, Cupp-Vickery JR, Poulos TL (2005) Crystal structures of the ferrous dioxygen complex of wild-type cytochrome P450eryF and its mutants, A245S and A245T: investigation of the proton transfer system in P450eryF. J. Biol. Chem. 280: 22102-22107.
236. Meilleur F, Dauvergne MT, Schlichting I, Myles DA (2005) Production and X-ray crystallographic analysis of fully deuterated cytochrome P450cam. Acta Crystallogr. D Biol. Crystallogr. 61: 539-544.
237. Davydov R, Makris TM, Kofman V, Werst DE, Sligar SG, et al. (2001) Hydroxylation of camphor by reduced oxy-cytochrome P450cam: mechanistic implications of EPR and ENDOR studies of catalytic intermediates in native and mutant enzymes. J. Am. Chem. Soc. 123: 1403-1415.
238. Davydov R, Macdonald IDG, Makris TM, Sligar SG, Hoffman BM (1999) EPR and ENDOR of catalytic intermediates in cryoreduced native and mutant oxy-cytochromes P450cam: mutation-induced changes in the proton delivery system. J. Am. Chem. Soc. 121: 10654-10655.
239. Denisov IG, Makris TM, Sligar SG (2001) Cryotrapped reaction intermediates of cytochrome p450 studied by radiolytic reduction with phosphorus-32. J. Biol. Chem. 276: 11648-11652.
240. Egawa T, Shimada H, Ishimura Y (1994) Evidence for compound I formation in the reaction of cytochrome P450cam with m-chloroperbenzoic acid. Biochem. Biophys. Res. Commun. 201: 1464-1469.
241. Kellner DG, Hung SC, Weiss KE, Sligar SG (2002) Kinetic characterization of compound I formation in the thermostable cytochrome P450 CYP119. J. Biol. Chem. 277: 9641-9644.
242. Sono M, Roach MP, Coulter ED, Dawson JH (1996) Heme-containing oxygenases. Chem. Rev. 96: 2841-2888.
243. Meunier B, de Visser SP, Shaik S (2004) Mechanism of oxidation reactions catalyzed by cytochrome p450 enzymes. Chem. Rev. 104: 3947-3980.
244. Denisov IG, Makris TM, Sligar SG, Schlichting I (2005) Structure and chemistry of cytochrome P450. Chem. Rev. 105: 2253-2277.
245. Makris TM, von Koenig K, Schlichting I, Sligar SG (2006) The status of high-valent metal oxo complexes in the P450 cytochromes. J. Inorg. Biochem. 100: 507-518.
246. Poulos TL (2005) Intermediates in P450 catalysis. Philos. Transact. A Math Phys. Eng. Sci. 363: 793-806; discussion 1035-1040.
247. Dong MS, Yamazaki H, Guo Z, Guengerich FP (1996) Recombinant human cytochrome P450 1A2 and an N-terminal-truncated form: construction,

purification, aggregation properties, and interactions with flavodoxin, ferredoxin, and NADPH-cytochrome P450 reductase. Arch. Biochem. Biophys. 327: 11-19.
248. Lewis DF, Pratt JM (1998) The P450 catalytic cycle and oxygenation mechanism. Drug Metab. Rev. 30: 739-786.
249. Raag R, Poulos TL (1992) X-ray crystallographic structural studies of cytochrome P450cam: factors controlling substrate metabolism. In: Ruckpaul K, Rein H, editors. Frontiers in Biotransformation. London: Taylor and Francis. pp. 1-43.
250. Sliger SG (1976) Coupling of spin, substrate and redox equilibria in cytochrome P450. Biochemistry 15: 5399-5406.
251. Daff SN, Chapman SK, Turner KL, Holt RA, Govindaraj S, et al. (1997) Redox control of the catalytic cycle of flavocytochrome P-450 BM3. Biochemistry 36: 13816-13823.
252. Archakov AI, Bachmanova GI (1990) Cytochrome P-450 and active oxygen. London: Taylor and Francis.
253. Guengerich FP, Ballou DP, Coon MJ (1976) Spectral intermediates in the reaction of oxygen with purified liver microsomal cytochrome P-450. Biochem. Biophys. Res. Commun. 70: 951-956.
254. Oprian DD, Gorsky LD, Coon MJ (1983) Properties of the oxygenated form of liver microsomal cytochrome P-450. J. Biol. Chem. 258: 8684-8691.
255. Estabrook RW, Hildebrandt AG, Baron J, Netter KJ, Leibman K (1971) A new spectral intermediate associated with cytochrome P-450 function in liver microsomes. Biochem. Biophys. Res. Commun. 42: 132-139.
256. Bonfils C, Debey P, Maurel P (1979) Highly purified microsomal P-450: the oxyferro intermediate stabilized at low temperature. Biochem. Biophys. Res. Commun. 88: 1301-1307.
257. Dunford HB, Stillman JS (1976) On the function and mechanism of action of peroxidases. Coord. Chem. Rev. 19: 187-251.
258. Kyte J (1995) Mechanism in Protein Chemistry. New York and London: Garland Publishing, Inc.
259. Kumar D, Hirao H, de Visser SP, Zheng J, Wang D, et al. (2005) New features in the catalytic cycle of cytochrome P450 during the formation of compound I from compound 0. J. Phys. Chem. 109: 19946-19951.
260. Bukowski MR, Koehntop KD, Stubna A, Bominaar EL, Halfen JA, et al. (2005) A thiolate-ligated nonheme oxoiron(IV) complex relevant to cytochrome P450. Science 310: 1000-1012.
261. Jung C, Schunemann V, Lendzian F, Trautwein AX, Contzen J, et al. (2005) Spectroscopic characterization of the iron-oxo intermediate in cytochrome P450. Biol. Chem. 386: 1043-1053.
262. Zhukov AA, Archakov AI (1982) Complete stoichiometry of free NADPH oxidation in liver microsomes. Biochem. Biophys. Res. Commun. 109: 813-818.
263. Sligar SG, Lipscomb JD, Debrunner PG, Gunsalus IC (1974) Superoxide anion production by the autoxidation of cytochrome P450cam. Biochem. Biophys. Res. Commun. 61: 290-296.
264. Kuthan H, Ullrich V (1982) Oxidase and oxygenase function of the microsomal cytochrome P450 monooxygenase system. Eur. J. Biochem. 126: 583-588.

265. Gorsky LD, Koop DR, Coon MJ (1984) On the stoichiometry of the oxidase and monooxygenase reactions catalyzed by liver microsomal cytochrome P-450. Products of oxygen reduction. J. Biol. Chem. 259: 6812-6817.
266. Noshiro M, Ullrich V, Omura T (1981) Cytochrome b5 as electron donor for oxy-cytochrome P-450. Eur. J. Biochem. 116: 521-526.
267. Hildebrandt A, Estabrook RW (1971) Evidence for the participation of cytochrome b 5 in hepatic microsomal mixed-function oxidation reactions. Arch. Biochem. Biophys. 143: 66-79.
268. Matsunaga I, Yamada A, Lee DS, Obayashi E, Fujiwara N, et al. (2002) Enzymatic reaction of hydrogen peroxide-dependent peroxygenase cytochrome P450s: kinetic deuterium isotope effects and analyses by resonance Raman spectroscopy. Biochemistry 41: 1886-1892.
269. White RE, Sligar SG, Coon MJ (1980) Evidence for a homolytic mechanism of peroxide oxygen--oxygen bond cleavage during substrate hydroxylation by cytochrome P-450. J. Biol. Chem. 255: 11108-11111.
270. Vaz AD, Coon MJ (1987) Hydrocarbon formation in the reductive cleavage of hydroperoxides by cytochrome P-450. Proc. Natl. Acad. Sci. U. S. A. 84: 1172-1176.
271. Ashley PL, Griffin BW (1981) Involvement of radical species in the oxidation of aminopyrine and 4-aminoantipyrine by cumene hydroperoxide in rat liver microsomes. Mol. Pharmacol. 19: 146-152.
272. Griffin BW, Marth C, Yasukochi Y, Masters BS (1980) Radical mechanism of aminopyrine oxidation by cumene hydroperoxide catalyzed by purified liver microsomal cytochrome P-450. Arch. Biochem. Biophys. 205: 543-553.
273. Thompson JA, Wand MD (1985) Interaction of cytochrome P-450 with a hydroperoxide derived from butylated hydroxytoluene. Mechanism of isomerization. J. Biol. Chem. 260: 10637-10644.
274. Lee WA, Bruice TC (1985) Homolytic and heterolytic oxygen-oxygen bond scissions accompanying oxygen transfer to Iron(III) porphyrins by percarboxylic acids and hydroperoxides. A mechanistic criterion for peroxidase and cytochrome-P450. J. Am. Chem. Soc. 107: 513-514.
275. Blake RC, 2nd, Coon MJ (1980) On the mechanism of action of cytochrome P-450. Spectral intermediates in the reaction of P-450LM2 with peroxy compounds. J. Biol. Chem. 255: 4100-4111.
276. Gustafsson JA, Rondahl L, Bergman J (1979) Iodosylbenzene derivatives as oxygen donors in cytochrome P-450 catalyzed steroid hydroxylations. Biochemistry 18: 865-870.
277. Nordblom GD, White RE, Coon MJ (1976) Studies on hydroperoxide-dependent substrate hydroxylation by purified liver microsomal cytochrome P-450. Arch. Biochem. Biophys. 175: 524-533.
278. Li QS, Ogawa J, Shimizu S (2001) Critical role of the residue size at position 87 in H2O2- dependent substrate hydroxylation activity and H2O2 inactivation of cytochrome P450BM-3. Biochem. Biophys. Res. Commun. 280: 1258-1261.
279. Li QS, Ogawa J, Schmid RD, Shimizu S (2005) Indole hydroxylation by bacterial cytochrome P450 BM-3 and modulation of activity by cumene hydroperoxide. Biosci. Biotechnol. Biochem. 69: 293-300.

280. Matsunaga I, Sumimoto T, Ayata M, Ogura H (2002) Functional modulation of a peroxygenase cytochrome P450: novel insight into the mechanisms of peroxygenase and peroxidase enzymes. FEBS Lett. 528: 90-94.
281. Imai Y, Matsunaga I, Kusunose E, Ichihara K (2000) Unique heme environment at the putative distal region of hydrogen peroxide-dependent fatty acid alpha-hydroxylase from Sphingomonas paucimobilis (peroxygenase P450(SPalpha). J. Biochem. (Tokyo) 128: 189-194.
282. Matsunaga I, Yokotani N, Gotoh O, Kusunose E, Yamada M, et al. (1997) Molecular cloning and expression of fatty acid alpha-hydroxylase from Sphingomonas paucimobilis. J. Biol. Chem. 272: 23592-23596.
283. Matsunaga I, Ueda A, Fujiwara N, Sumimoto T, Ichihara K (1999) Characterization of the ybdT gene product of Bacillus subtilis: novel fatty acid beta-hydroxylating cytochrome P450. Lipids 34: 841-846.
284. Matsunaga I, Sumimoto T, Ueda A, Kusunose E, Ichihara K (2000) Fatty acid-specific, regiospecific, and stereospecific hydroxylation by cytochrome P450 (CYP152B1) from Sphingomonas paucimobilis: substrate structure required for alpha-hydroxylation. Lipids 35: 365-371.
285. Matsunaga I, Ueda A, Sumimoto T, Ichihara K, Ayata M, et al. (2001) Site-directed mutagenesis of the putative distal helix of peroxygenase cytochrome P450. Arch. Biochem. Biophys. 394: 45-53.
286. Anari MR, Josephy PD, Henry T, O'Brien PJ (1997) Hydrogen peroxide supports human and rat cytochrome P450 1A2-catalyzed 2-amino-3-methylimidazo[4,5-f]quinoline bioactivation to mutagenic metabolites: significance of cytochrome P450 peroxygenase. Chem. Res. Toxicol. 10: 582-588.
287. McCallum GP, Weedon AC, Krug P, Bend JR (1996) Microsomal cytochrome P450 peroxygenase metabolism of arachidonic acid in guinea pig liver. J. Pharmacol. Exp. Ther. 278: 1188-1194.
288. Jonsson KH, Lindeke B (1990) Cytochrome P-455 nm complex formation in the metabolism of phenylalkylamines. XI. Peroxygenase versus monooxygenase function of cytochrome P-450 in rat liver microsomes. Chem. Biol. Interact. 75: 267-279.
289. Chefson A, Zhao J, Auclair K (2006) Replacement of natural cofactors by selected hydrogen peroxide donors or organic peroxides results in improved activity for CYP3A4 and CYP2D6. Chembiochem 7: 916-919.
290. Govindaraj S, Poulos TL (1997) The domain architecture of cytochrome P450BM-3. J. Biol. Chem. 272: 7915-7921.
291. Fulco AJ (1991) P450BM-3 and other inducible bacterial P450 cytochromes: biochemistry and regulation. Annu. Rev. Pharmacol. Toxicol. 31: 177-203.
292. Ruettinger RT, Wen LP, Fulco AJ (1989) Coding nucleotide, 5' regulatory, and deduced amino acid sequences of P-450BM-3, a single peptide cytochrome P-450:NADPH-P-450 reductase from Bacillus megaterium. J. Biol. Chem. 264: 10987-10995.
293. Noble MA, Miles CS, Chapman SK, Lysek DA, MacKay AC, et al. (1999) Roles of key active-site residues in flavocytochrome P450 BM3. Biochem. J. 339 (Pt 2): 371-379.

294. Truan G, Komandla MR, Falck JR, Peterson JA (1999) P450BM-3: absolute configuration of the primary metabolites of palmitic acid. Arch. Biochem. Biophys. 366: 192-198.
295. Narhi LO, Fulco AJ (1986) Characterization of a catalytically self-sufficient 119,000-dalton cytochrome P-450 monooxygenase induced by barbiturates in Bacillus megaterium. J. Biol. Chem. 261: 7160-7169.
296. Palmer CN, Axen E, Hughes V, Wolf CR (1998) The repressor protein, Bm3R1, mediates an adaptive response to toxic fatty acids in Bacillus megaterium. J. Biol. Chem. 273: 18109-18116.
297. Palmer CN, Causevic M, Wolf CR (1997) Modulation of fatty acid signalling by cytochrome P-450-mediated hydroxylation. Biochem. Soc. Trans. 25: 1160-1165.
298. Eiben S, Kaysser L, Maurer S, Kuhnel K, Urlacher VB, et al. (2006) Preparative use of isolated CYP102 monooxygenases -- a critical appraisal. J. Biotechnol. 124: 662-669.
299. Munro AW, Girvan HM, McLean KJ (2007) Cytochrome P450--redox partner fusion enzymes. Biochim. Biophys. Acta. 1770: 345-359.
300. Narhi LO, Fulco AJ (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: 6683-6690.
301. Boddupalli SS, Oster T, Estabrook RW, Peterson JA (1992) Reconstitution of the fatty acid hydroxylation function of cytochrome P-450BM-3 utilizing its individual recombinant hemo- and flavoprotein domains. J. Biol. Chem. 267: 10375-10380.
302. Guengerich FP (2002) Cytochrome P450 enzymes in the generation of commercial products. Nat. Rev. Drug Discov. 1: 359-366.
303. Oliver CF, Modi S, Primrose WU, Lian LY, Roberts GC (1997) Engineering the substrate specificity of Bacillus megaterium cytochrome P-450 BM3: hydroxylation of alkyl trimethylammonium compounds. Biochem. J. 327 (Pt 2): 537-544.
304. Ost TW, Miles CS, Murdoch J, Cheung Y, Reid GA, et al. (2000) Rational re-design of the substrate binding site of flavocytochrome P450 BM3. FEBS Lett. 486: 173-177.
305. Li QS, Ogawa J, Schmid RD, Shimizu S (2001) Engineering cytochrome P450 BM-3 for oxidation of polycyclic aromatic hydrocarbons. Appl. Environ. Microbiol. 67: 5735-5739.
306. Li QS, Schwaneberg U, Fischer P, Schmid RD (2000) Directed evolution of the fatty-acid hydroxylase P450 BM-3 into an indole-hydroxylating catalyst. Chemistry 6: 1531-1536.
307. Appel D, Lutz-Wahl S, Fischer P, Schwaneberg U, Schmid RD (2001) A P450 BM-3 mutant hydroxylates alkanes, cycloalkanes, arenes and heteroarenes. J. Biotechnol. 88: 167-171.
308. Carmichael AB, Wong LL (2001) Protein engineering of Bacillus megaterium CYP102.The oxidation of polycyclic aromatic hydrocarbons. Eur. J. Biochem. 268: 3117-3125.

309. Lussenburg BM, Babel LC, Vermeulen NP, Commandeur JN (2005) Evaluation of alkoxyresorufins as fluorescent substrates for cytochrome P450 BM3 and site-directed mutants. Anal. Biochem. 341: 148-155.
310. Glieder A, Farinas ET, Arnold FH (2002) Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. Nat. Biotechnol. 20: 1135-1139.
311. Peters MW, Meinhold P, Glieder A, Arnold FH (2003) Regio- and enantioselective alkane hydroxylation with engineered cytochromes P450 BM-3. J. Am. Chem. Soc. 125: 13442-13450.
312. Meinhold P, Peters MW, Chen MM, Takahashi K, Arnold FH (2005) Direct conversion of ethane to ethanol by engineered cytochrome P450 BM3. Chembiochem 6: 1765-1768.
313. Meinhold P, Peters MW, Hartwick A, Hernandez AR, Arnold FH (2006) Engineering cytochrome P450 BM3 for terminal alkane hydroxylation. Adv. Synth. Catal. 348: 763-772.
314. Kubo T, Peters MW, Meinhold P, Arnold FH (2006) Enantioselective epoxidation of terminal alkenes to (R)- and (S)-epoxides by engineered cytochromes P450 BM-3. Chemistry 12: 1216-1220.
315. Graham-Lorence S, Truan G, Peterson JA, Falck JR, Wei S, et al. (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: 1127-1135.
316. Urlacher VB, Makhsumkhanov A, Schmid RD (2006) Biotransformation of beta-ionone by engineered cytochrome P450 BM-3. Appl. Microbiol. Biotechnol. 70: 53-59.
317. Sowden RJ, Yasmin S, Rees NH, Bell SG, Wong LL (2005) Biotransformation of the sesquiterpene (+)-valencene by cytochrome P450cam and P450BM-3. Org. Biomol. Chem. 3: 57-64.
318. Budde M, Maurer SC, Schmid RD, Urlacher VB (2005) Cloning, expression and characterisation of CYP102A2, a self-sufficient P450 monooxygenase from Bacillus subtilis. Appl. Microbiol. Biotechnol. 66: 180-186.
319. Gustafsson MC, Roitel O, Marshall KR, Noble MA, Chapman SK, et al. (2004) Expression, purification, and characterization of bacillus subtilis cytochromes P450 CYP102A2 and CYP102A3: flavocytochrome homologues of P450 BM3 from Bacillus megaterium. Biochemistry 43: 5474-5487.
320. Budde M, Morr M, Schmid RD, Urlacher VB (2006) Selective hydroxylation of highly branched fatty acids and their derivatives by CYP102A1 from Bacillus megaterium. Chembiochem 7: 789-794.
321. Axarli I, Prigipaki A, Labrou NE (2005) Engineering the substrate specificity of cytochrome P450 CYP102A2 by directed evolution: production of an efficient enzyme for bioconversion of fine chemicals. Biomol. Eng. 22: 81-88.
322. Lentz O, Urlacher V, Schmid RD (2004) Substrate specificity of native and mutated cytochrome P450 (CYP102A3) from Bacillus subtilis. J. Biotechnol. 108: 41-49.
323. Lephart ED, Peterson KG, Noble JF, George FW, McPhaul MJ (1990) The structure of cDNA clones encoding the aromatase P-450 isolated from a rat Leydig cell tumor line demonstrates differential processing of aromatase mRNA in rat ovary and a neoplastic cell line. Mol. Cell. Endocrinol. 70: 31-40.

324. Lacroix D, Desrochers M, Lambert M, Anderson A (1990) Alternative splicing of mRNA encoding rat liver cytochrome P450e (P450IIB2). Gene 86: 201-207.
325. Miles JS, McLaren AW, Gonzalez FJ, Wolf CR (1990) Alternative splicing in the human cytochrome P450IIB6 gene: use of a cryptic exon within intron 3 and splice acceptor site within exon 4. Nucleic Acids Res. 18: 189.
326. Means GD, Kilgore MW, Mahendroo MS, Mendelson CR, Simpson ER (1991) Tissue-specific promoters regulate aromatase cytochrome P450 gene expression in human ovary and fetal tissues. Mol. Endocrinol. 5: 2005-2013.
327. Finta C, Zaphiropoulos PG (2000) The human cytochrome P450 3A locus. Gene evolution by capture of downstream exons. Gene 260: 13-23.
328. Christmas P, Jones JP, Patten CJ, Rock DA, Zheng Y, et al. (2001) Alternative splicing determines the function of CYP4F3 by switching substrate specificity. J. Biol. Chem. 276: 38166-38172.
329. Rodriguez-Antona C, Axelson M, Otter C, Rane A, Ingelman-Sundberg M (2005) A novel polymorphic cytochrome P450 formed by splicing of CYP3A7 and the pseudogene CYP3AP1. J. Biol. Chem. 280: 28324-28331.
330. Shimada T, Yamazaki H, Mimura M, Wakamiya N, Ueng YF, et al. (1996) Characterization of microsomal cytochrome P450 enzymes involved in the oxidation of xenobiotic chemicals in human fetal liver and adult lungs. Drug Metab. Dispos. 24: 515-522.
331. Ladona MG, Gonzalez ML, Rane A, Peter RM, de la Torre R (2000) Cocaine metabolism in human fetal and adult liver microsomes is related to cytochrome P450 3A expression. Life Sci. 68: 431-443.
332. Lifton RP, Dluhy RG, Powers M, Rich GM, Cook S, et al. (1992) A chimaeric 11 beta-hydroxylase/aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension. Nature 355: 262-265.
333. Murataliev MB, Trinh LN, Moser LV, Bates RB, Feyereisen R, et al. (2004) Chimeragenesis of the fatty acid binding site of cytochrome P450BM3. Replacement of residues 73-84 with the homologous residues from the insect cytochrome P450 CYP4C7. Biochemistry 43: 1771-1780.
334. Fuziwara S, Sagami I, Rozhkova E, Craig D, Noble MA, et al. (2002) Catalytically functional flavocytochrome chimeras of P450 BM3 and nitric oxide synthase. J. Inorg. Biochem. 91: 515-526.
335. Roman LJ, McLain J, Masters BS (2003) Chimeric enzymes of cytochrome P450 oxidoreductase and neuronal nitric-oxide synthase reductase domain reveal structural and functional differences. J. Biol. Chem. 278: 25700-25707.
336. Ramarao MK, Straub P, Kemper B (1995) Identification by in vitro mutagenesis of the interaction of two segments of C2MstC1, a chimera of cytochromes P450 2C2 and P450 2C1. J. Biol. Chem. 270: 1873-1880.
337. Uno T, Yokota H, Imai Y (1990) Replacing the carboxy-terminal 28 residues of rabbit liver P-450 (laurate (omega-1)-hydroxylase) with those of P-450 (testosterone 16 alpha-hydroxylase) produces a new stereospecific hydroxylase activity. Biochem. Biophys. Res. Commun. 167: 498-503.
338. Uno T, Imai Y (1992) Further studies on chimeric P450 2C2/2C14 having testosterone 16 beta-hydroxylase activity which is absent in the parental P450s. J. Biochem. (Tokyo) 112: 155-162.

339. Abecassis V, Urban P, Aggerbeck L, Truan G, Pompon D (2003) Exploration of natural and artificial sequence spaces: Towards a functional remodeling of membrane-bound cytochrome P450. Biocatal. Biotransform. 21: 55-66.
340. Bellamine A, Gautier JC, Urban P, Pompon D (1994) Chimeras of the human cytochrome P450 1A family produced in yeast. Accumulation in microsomal membranes, enzyme kinetics and stability. Eur. J. Biochem. 225: 1005-1013.
341. Jung F, Griffin KJ, Song W, Richardson TH, Yang M, et al. (1998) Identification of amino acid substitutions that confer a high affinity for sulfaphenazole binding and a high catalytic efficiency for warfarin metabolism to P450 2C19. Biochemistry 37: 16270-16279.
342. Johnson EF, Kronbach T, Hsu MH (1992) Analysis of the catalytic specificity of cytochrome P450 enzymes through site-directed mutagenesis. Faseb J. 6: 700-705.
343. Hanioka N, Gonzalez FJ, Lindberg NA, Liu G, Korzekwa KR (1992) Chimeric cDNA expression and site directed mutagenesis studies of cytochrome p450s CYP2A1 and CYP2A2. J. Steroid Biochem. Molec. Biol. 43: 1037-1043.
344. Zhou MY, Gomez-Sanchez CE, Xue D, Foecking MF (1994) The hybrid rat cytochrome P450 containing the first 5 exons of the CYP11B1 and last 4 exons from the CYP11B2 enzyme retains 11 beta-hydroxylase activity, but the alternative hybrid is inactive. Biochem. Biophys. Res. Commun. 199: 130-135.
345. Sakaki T, Shibata M, Yabusaki Y, Ohkawa H (1987) Expression in Saccharomyces cerevisiae of chimeric cytochrome P450 cDNAs constructed from cDNAs for rat cytochrome P450c and P450d. DNA 6: 31-39.
346. Pompon D, Nicolas A (1989) Protein engineering by cDNA recombination in yeasts: shuffling of mammalian cytochrome P-450 functions. Gene 83: 15-24.
347. Kronbach T, Larabee TM, Johnson EF (1989) Hybrid cytochromes P-450 identify a substrate binding domain in P-450IIC5 and P-450IIC4. Proc. Natl. Acad. Sci. U. S. A. 86: 8262-8265.
348. Szklarz GD, He YQ, Kedzie KM, Halpert JR, Burnett VL (1996) Elucidation of amino acid residues critical for unique activities of rabbit cytochrome P450 2B5 using hybrid enzymes and reciprocal site-directed mutagenesis with rabbit cytochrome P450 2B4. Arch. Biochem. Biophys. 327: 308-318.
349. Cosme J, Johnson EF (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: 2545-2553.
350. Sukumaran S, Atkins WM, Shanker R (2002) Engineering cytochrome P-450s: chimeric enzymes. Appl. Biochem. Biotechnol. 102-103: 291-302.
351. Pikuleva IA, Bjorkhem I, Waterman MR (1996) Studies of distant members of the P450 superfamily (P450scc and P450c27) by random chimeragenesis. Arch. Biochem. Biophys. 334: 183-192.
352. Shimoji M, Yin H, Higgins L, Jones JP (1998) Design of a novel P450: a functional bacterial-human cytochrome P450 chimera. Biochemistry 37: 8848-8852.
353. Brock BJ, Waterman MR (2000) The use of random chimeragenesis to study structure/function properties of rat and human P450c17. Arch. Biochem. Biophys. 373: 401-408.

354. Fairhead M, Giannini S, Gillam EM, Gilardi G (2005) Functional characterisation of an engineered multidomain human P450 2E1 by molecular Lego. J. Biol. Inorg. Chem. 10: 842-853.
355. Helvig C, Capdevila JH (2000) Biochemical characterization of rat P450 2C11 fused to rat or bacterial NADPH-P450 reductase domains. Biochemistry 39: 5196-5205.
356. Fisher CW, Shet MS, Estabrook RW (1996) Construction of plasmids and expression in Escherichia coli of enzymatically active fusion proteins containing the heme-domain of a P450 linked to NADPH-P450 reductase. Methods Enzymol. 272: 15-25.
357. Sakaki T, Shibata M, Yabusaki Y, Murakami H, Ohkawa H (1990) Expression of bovine cytochrome P450c21 and its fused enzymes with yeast NADPH-cytochrome P450 reductase in Saccharomyces cerevisiae. DNA Cell Biol. 9: 603-614.
358. Gilardi G, Meharenna YT, Tsotsou GE, Sadeghi SJ, Fairhead M, et al. (2002) Molecular Lego: design of molecular assemblies of P450 enzymes for nanobiotechnology. Biosens. Bioelectron. 17: 133-145.
359. Bernhardt R (2006) Cytochromes P450 as versatile biocatalysts. J. Biotechnol. 124: 128-145.
360. Gillam EM (2005) Exploring the potential of xenobiotic-metabolising enzymes as biocatalysts: evolving designer catalysts from polyfunctional cytochrome P450 enzymes. Clin. Exp. Pharmacol. Physiol. 32: 147-152.
361. Fukui Y, Tanaka Y, Kusumi T, Iwashita T, Nomoto K (2003) A rationale for the shift in colour towards blue in transgenic carnation flowers expressing the flavonoid 3',5'-hydroxylase gene. Phytochemistry 63: 15-23.
362. Holton TA, Brugliera F, Lester DR, Tanaka Y, Hyland CD, et al. (1993) Cloning and expression of cytochrome P450 genes controlling flower colour. Nature 366: 276-279.
363. Falck JR, Reddy YK, Haines DC, Reddy KM, Krishna UM, et al. (2001) Practical, enantiospecific syntheses of 14,15-EET and leukotoixn B (vernolic acid). Tetrahedron Letters 42: 4131-4133.
364. Guengerich FP (1995) Cytochrome P450 proteins and potential utilization in biodegradation. Environ. Health Perspect. 103 Suppl 5: 25-28.
365. Jones JP, O'Hare EJ, Wong LL (2001) Oxidation of polychlorinated benzenes by genetically engineered CYP101 (cytochrome P450(cam)). Eur. J. Biochem. 268: 1460-1467.
366. Dejong JM, Liu Y, Bollon AP, Long RM, Jennewein S, et al. (2006) Genetic engineering of taxol biosynthetic genes in Saccharomyces cerevisiae. Biotechnol. Bioeng. 93: 212-224.
367. Jennewein S, Long RM, Williams RM, Croteau R (2004) Cytochrome p450 taxadiene 5alpha-hydroxylase, a mechanistically unusual monooxygenase catalyzing the first oxygenation step of taxol biosynthesis. Chem. Biol. 11: 379-387.
368. Jennewein S, Park H, DeJong JM, Long RM, Bollon AP, et al. (2005) Coexpression in yeast of Taxus cytochrome P450 reductase with cytochrome P450 oxygenases involved in Taxol biosynthesis. Biotechnol. Bioeng. 89: 588-598.

369. Serizawa N, Nakagawa K, Hamano K, Tsujita Y, Terahara A, et al. (1983) Microbial hydroxylation of ML-236B (compactin) and monacolin K (MB-530B). J. Antibiot. (Tokyo) 36: 604-607.
370. Guengerich FP, Sorrells JL, Schmitt S, Krauser JA, Aryal P, et al. (2004) Generation of new protein kinase inhibitors utilizing cytochrome p450 mutant enzymes for indigoid synthesis. J. Med. Chem. 47: 3236-3241.
371. Cannell RJ, Knaggs AR, Dawson MJ, Manchee GR, Eddershaw PJ, et al. (1995) Microbial biotransformation of the angiotensin II antagonist GR117289 by Streptomyces rimosus to identify a mammalian metabolite. Drug Metab. Dispos. 23: 724-729.
372. Cannell RJ, Rashid T, Ismail IM, Sidebottom PJ, Knaggs AR, et al. (1997) Novel metabolites of warfarin produced by Beauveria bassiana and Streptomyces rimosus: a novel application of hplc-nmr. Xenobiotica 27: 147-157.
373. Chen G, Waxman DJ (1995) Complete reversal by thaliblastine of 490-fold adriamycin resistance in multidrug-resistant (MDR) human breast cancer cells. Evidence that multiple biochemical changes in MDR cells need not correspond to multiple functional determinants for drug resistance. J. Pharmacol. Exp. Ther. 274: 1271-1277.

Chapter 3

High-throughput assays for analysis of cytochromes P450

The majority of this chapter is adapted from two book chapters: Otey CR (2003) **High-throughput carbon monoxide binding assay for cytochromes P450**. *Methods Mol. Biol.* 230: 137-139, and Otey CR, Joern JM (2003) **High-throughput screen for aromatic hydroxylation**. *Methods Mol. Biol.* 230: 141-148. They are reprinted with permission from Humana press. The goal of these chapters is to provide a high level of detail so the methods can be easily reproduced. This detail is facilitated by frequent annotation of the methods in the form of 'notes' which are found at the end of this chapter. Additionally, I would like to thank John M. Joern for his work on the Gibbs' assay and assistance in adapting it to P450s and to Philip Romero for work on the substrate-binding assays.

A. Introduction

Central to any directed evolution experiment is an effective high-throughput screening or selection method. These methods allow an experimenter to test anywhere from one hundred to millions of clones per day for specific properties. Due to the great interest in the research and engineering of cytochromes P450 described in Chapter 2, a variety of methods amenable to high-throughput have been developed. Examples of these colorimetric, fluorescent and spectroscopic methods are listed in Table 3.1. The majority of these assays were developed to more efficiently test the inhibitory properties of potential drugs on human P450s, while others have been used in the study or alteration and improvement of a variety of P450 properties. However, most of these assays focus on a single substrate and do not offer information about P450s other than their enzymatic properties. In this chapter, new assays appropriate for the measurement of the ability of a P450 to fold, bind a substrate and hydroxylate a variety of substrates are discussed. These methods are applied herein and have subsequently been utilized in other studies [1-3].

Table 3.1. Cytochrome P450 assays amenable to high-throughput screening.

Substrate or reagent	Applicable P450s	Assay description[a]	References
NAD(P)H	Any NAD(P)H utilizing P450, applied to CYP102A1, CYP102A2	Monitor decrease in absorbance at 340 nm from conversion of NAD(P)H to NAD(P)+.	[4-7]
Naphthalene	CYP101	Phenolic products of naphthalene coupled by HRP to form fluorescent products, ex 350 em 450 .	[8,9]
4-Nitrophenol substituted substrates	CYP102A1	Substrates such as 12-*p*NCA; monitor increase of absorbance due to 4-nitrophenol formation at 398 or 410 nm.	[5,10-12]
7-Ethoxycoumarin	CYP1A1, CYP1A2	Measure fluorescence: ex 360, em 450.	[13,14]
3-Cyano-7-methoxycoumarin	CYP1A1, CYP1A2, CYP2C6, CYP2C11, CYP2E1, CYP2B6	Measure fluorescence: ex 410, em 460.	[15-17]
3-Cyano-7-ethoxycoumarin	CYP1A2, CYP1A1, CYP2A6	Measure fluorescence: ex 410, em 460.	[15,18-20]
7-Methoxy-4-trifluoromethylcoumarin	CYP2C9, CYP2E1, CYP2C6, CYP2C11	Measure fluorescence: ex 410, em 538.	[15,19,20]
7-Ethoxy-4-trifluouromethylcoumarin	CYP1A2, CYP2E1, CYP2B6	Measure fluorescence: ex 400, em 500.	[17,19,21]
7-Benzyloxy-4-trifluormethylcoumarin	CYP3A4, CYP2B1, CYP2B6, CYP3A1, CYP3A2, CYP3A4	Measure fluorescence: ex 410, em 538	[15,19,20]
7-Methoxy-4-(aminomethyl)-coumarin	CYP2D6	Measure fluorescence: ex 390, em 460.	[20]
3-[2-(*N*,*N*-diethyl-*N*-methylamino)ethyl]-7-methoxy-4-methylcoumarin	CYP2D6, CYP2D2	Measure fluorescence: ex 390 em 460.	[15,19]
7-Methoxyresorufin	CYP102A1	Measure fluorescence: ex 430, em 580.	[22,23]

7-Ethoxyresorufin	CYP102A1, CYP1A2	Measure fluorescence: ex 530, em 580.	[23-27]
7-Pentoxyresorufin	CYP102A1	Measure fluorescence: ex 530, em 580.	[23]
7-Benzyloxyresorufin	CYP102A1	Measure fluorescence: ex 530, em 580.	[23,26]
7-Benzylresorufin	CYP102A1, CYP1A1, CYP1A2, and CYP2B1	Measure fluorescence: ex 530, em 580.	[15,23,28]
3-Cyano-7-ethoxyresorufin	CYP1A2, CYP2C9, CYP2C19, and CYP2D6	Measure fluorescence: ex 420, em 485.	[28]
Dibenzylfluorescein	CYP2C8	Measure fluorescence: ex 485, em 538.	[15,19,23,26]
7-Benzyloxyquinoline	CYP1A1, CYP2D1, CYP3A1, and CYP3A2	Measure fluorescence: ex 410, em 538.	[15]
P450-Glo luminescent substrates	CYP1A1, CYP1A2, C1B1, CYP2C8, CYP2C9, CYP3A4, and CYP3A7	Substrates are conjugated to luciferin which is a substrate for luciferase that emmits a luminescent signal.	[29]
Purpald	CYP102A1	Detects formaldehyde which can be formed by hydroxylation of surrogate ether substruates, e.g. hexyl methyl ether and dimethyl ether. Measure absorbance at 550 nm.	[30]
4-(p-nitrobenzyl)pyridine (NBP)	CYP102A1	Detection of terminal epoxides. Measure absorbance at 580 nm.	[31,32]
Indole	2E6, 2C19, 2A6, CYP102A1 and CYP2W1	Indigo is converted to indole, indiribun, indigoids, etc. which are colored compounds.	[33-40]
2-Amino-3,5-dimethylimidazo[4,5-*f*]quinoline (MeIQ)	CYP1A2	Genotoxicity assay in which N-hydroxyation of MeIQ creates a mutagen which causes reversion of a LacZ- straint to LacZ+ allowing cell growth on minimal lactose media.	[22,25]

Diazo Blue B	CYP1A1 and CYP1A2	Detection of hydroxlated naphthalene or other phenolic prodcts. Measure absorbance at 490 nm.	[41,42]
11-Deoxycortisol	CYP11B1	Conversion of 11-deoxycortisol to cortsiol followed by addition of sulfuric and acetic acids which creates a yellow; fluorescent dye, ex 475 em 525.	[43]
LC/MS/MS	CYP1A2, CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1, and CYP3A4	Two substrate cocktails which are incubated with a mix of P450s and then pooled and analyzed using LC/MS/MS.	[44]
Carbon chains, e.g. fatty acids and alkanes	CYP102A3	Measure terminal hydroxylation of alkanes by monitoring the depletion of NADH at 340 nm caused by alcohol dehydrogenase activity, which is specific to primary alcohols.	[45]

[a] Abbreviations: horse radish peroxidase, HPR; p-nitrophenoxydodecanoic acid, 12-pNCA; excitation wavelength (nm), ex; emission wavelength (nm), em.

B. High-throughput carbon monoxide difference spectroscopy

The high-throughput carbon monoxide binding assay monitors the folding properties of a P450 by employing the spectroscopic properties of the noncovalently bound protoporphyrin IX in the P450 active site. This heme has an available sixth coordination site that is able to tightly bind carbon monoxide when the heme iron is in the ferrous state. Difference spectroscopy using carbon monoxide yields a spectral peak at approximately 450 nm characteristic of a correctly-folded cytochrome P450 when comparing bound and unbound forms [46,47]. This peak is referred to as a Soret band since it is in the blue region of the absorption spectrum of a heme protein. Binding of carbon monoxide in the presence of a catalytically inactive form of a cytochrome P450 yields a spectral peak at 420 nm [48-50]. Thus CO binding effectively assays for the presence of a correctly-folded cytochrome P450, an incorrectly-folded P420, or a lack of either. Carbon monoxide binding assays are typically performed in single cuvette format

[51] but are easily modified for high-throughput in microtiter plates (Figure 3.1A). This method is useful in assaying a library of cytochrome P450s for folded and possibly functional proteins while eliminating misfolded or low-expressing variants. It can also be used to rapidly determine the P450 concentration of multiple samples or the relative expression levels of individual clones in a library of P450 variants.

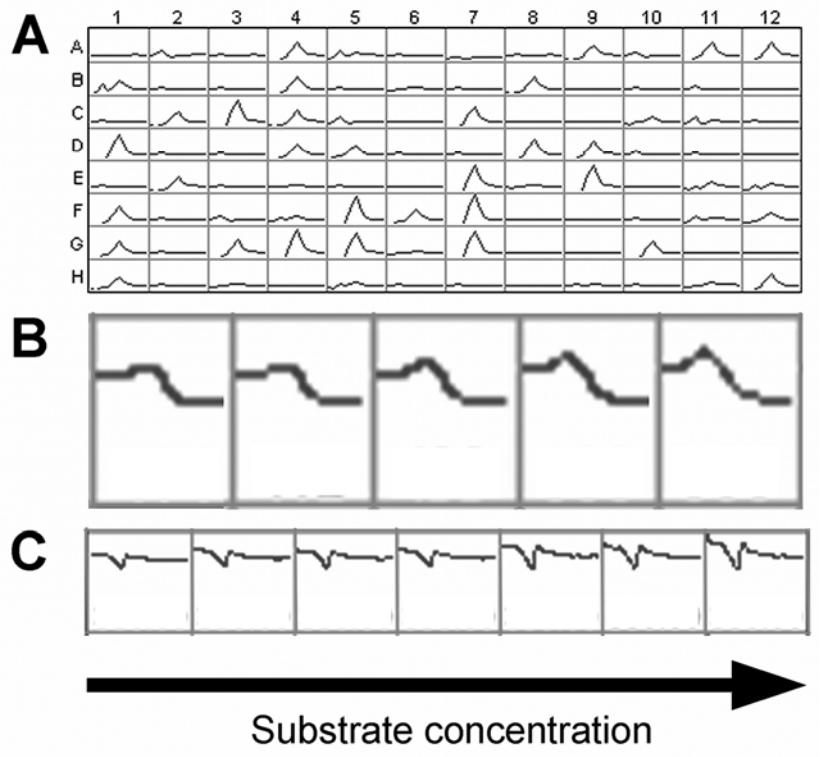

Figure 3.1. Spectroscopic assays developed for cytochromes P450. (A) A 96-well plate showing the CO difference spectra from a library of cytochromes P450. The characteristic Soret peak in the folded chimeras is present near 450 nm. The range 400 to 500 nm is shown with an absorbance range of -0.1 to 0.1. (B) Type I binding spectra resulting from the heme domain of wild-type CYP102A1 with 0, 94, 177, 333, and 632 μM lauric acid. The resulting spin-shift results in a decrease in absorbance near 415 nm with an increase near 395 nm. Range 350 to 500 nm with an absorbance range of 0.2 to 0.6 is shown. (C) Increasing difference spectra from Type II substrate binding with 3 μM of P450 mutant 2C11 (see chapter 6) titrated with 100, 200, 300, 333, 666, 1000 and 2000 μM serotonin. The Type II spectra results in an increase in absorbance near 430 nm with a concomitant decrease near 415 nm. Range 330 to 600 nm is shown with an absorbance range of -0.1 to 0.1.

B.1 Materials

1. 100 mM sodium hydrosulfite in 1.3 M phosphate buffer, pH 8.0 prepared fresh for each experiment.
2. 50 mM phosphate buffer, pH 7.3.
3. 96-well microtiter plates: R-96-OAPF-ICO (Rainin, Emeryville, CA).
4. Carbon monoxide tank with an appropriate regulator that allows for slow and controlled bubbling into liquids.
5. Carbon monoxide chamber: a container in which 96-well microtiter plates can be placed and where a vacuum can be created followed by a slight pressurization. It should be an easily sealed container with at least one hose connection or stopcock valve. A vacuum oven will work (Fisher Scientific) as well as a desiccator made of polycarbonate. A 3-way stopcock, C-clamps and vacuum grease may also be necessary (*see* **Note 1.1**).
7. Spectrophotometer/plate reader (Model Spectra max Plus 384, Molecular Devices, Synnyvale, CA). Software Softmax Pro 3.1.1.
8. Benchtop centrifuge that can accommodate 96-well microtiter plates: Allegra 25R Centrifuge (Beckman Coulter, Fullerton, CA).

B.2 Methods

1. Add 40 μL of the 100 mM sodium hydrosulfite solution into wells of a 96-well microtiter plate (*see* **Note 1.2**).
2. Add 160 μL of enzyme solution per well and mix (*see* **Note 1.3**). Enzyme solution can be either purified enzyme or extract from a cell lysis reaction after centrifugation

and removal of cell debris (*see* **Note 1.4**).

3. Blank plate reader, taking both baseline spectra and particular wavelengths (*see* **Note 1.5**). Put plate into the carbon monoxide chamber and pull a vacuum with a vacuum pump to roughly 350 mbar. Fill the container with carbon monoxide until a positive pressure is obtained and seal the container for at least eight minutes (*see* **Note 1.6**).
4. Remove the plate and record spectra from 400-500 nm and at specific wavelengths (*see* **Note 1.5**).
5. Using A_{450} and A_{490}, P450 concentrations can be determined using Beer's Law: $A = \varepsilon cl$ and the extinction coefficient, $\varepsilon_{450\text{-}490} = 91\ mM^{-1}\ cm^{-1}$. (*see* **Note 1.7**).

C. High-throughput substrate binding (spin-shift) assay

The second assay also employs the heme cofactor and monitors the spectral shifts that occur upon substrate binding by cytochromes P450. As discussed in chapter 1, when a substrate enters the P450 active site it displaces a water molecule as the sixth ligand of the heme cofactor causing a shift from low to high spin in the heme iron. This can result in three specific types of binding spectra; Type I, reverse Type I and Type II [52,53]. Type I binding spectra show a decrease in the absorption peak of the low-spin state at 414-420 nm and increase in the high-spin state at 385-395 nm (Figure 3.1B). Reverse Type I shows the opposite, with an increase around 420 nm and decrease around 390 nm. The cause is postulated to be displacement of the distal ligand and substrate binding to a hydrophobic region of the active site [54,55]. Type II binding spectra are thought to occur when a substrate ligates to the heme iron, a common mechanism of P450 inhibitors [56]. Type II binding causes a decrease in absorbance around 390 to 405 nm and an increase at

about 425-435 nm. These changes are caused by a simultaneous shift to longer wavelengths for each spin state and a change in the spin-state equilibrium from low to high (Figure 3.1C).

The goal of this assay is to obtain a binding curve and approximate a binding constant for multiple clones using either clarified cell lysate or purified protein and the substrate of your choice. To maximize detection one should use high concentrations of P450 (*see* **Note 2.1**) and large assay volumes (*see* **Note 2.2**). Initially one should determine the complete spectra resulting from substrate binding and determine specific wavelengths to monitor to reduce assay times. Within the typical bounds of concentrations obtained in deep-well plates, the P450 should be limiting and therefore one does not need to worry about relative enzyme concentrations if a binding curve is constructed.

C.1 Materials

1. Spectrophotometer/plate reader (Model Spectra max Plus 384, Molecular Devices, Sunnyvale, CA). Software Softmax Pro 3.1.1.
2. 96-well microtiter plates: R-96-OAPF-ICO (Rainin, Emeryville, CA) or 384-well microtiter plates (Nalge Nunc International, Rochester, NY).
3. Concentrated substrate solution.

C.2 Methods

1. Distribute cell lysate into a microtiter plate and read the baseline spectra or baseline individual wavelengths (*see* **Note 2.3**).

2. Add substrate solution and measure spectra or individual wavelengths (*see* **Note 2.2** and **Note 2.3**).
3. Continue step 2 until enough data is accumulated to produce a binding curve.
4. Construct binding curve for comparison of clones.

D. High-throughput screen for aromatic hydroxylation

The third method describes three different assay systems that detect phenolic compounds by producing a colored product that is easily detected in high throughput. These assays employ the Gibbs' reagent, 4-AAP or Fast Violet B (FVB) (Figure 3.2) and are useful for detecting the oxidation of aromatic compounds by P450s. The oxidation of aromatic compounds is generally important in producing chemical intermediates for the chemical and pharmaceutical industries [57,58]. Conventional aromatic oxidation reactions are prone to byproduct formation and often require heavy metal catalysts, extremes of temperature and pressure, and explosive reagents [59]. In contrast, biocatalysts such as P450s perform the same chemistry in water at ambient conditions, usually with higher regioselectivity than the analogous chemical process [60-62] and the methods discussed here may allow for the optimization of oxygenases to industrially relevant substrates and realistic process conditions. The use of these assays with P450s is described here but they may be applied to dioxygenases [63,64] and many other enzymes (e.g. oxidative dealkylases or dehalogenases).

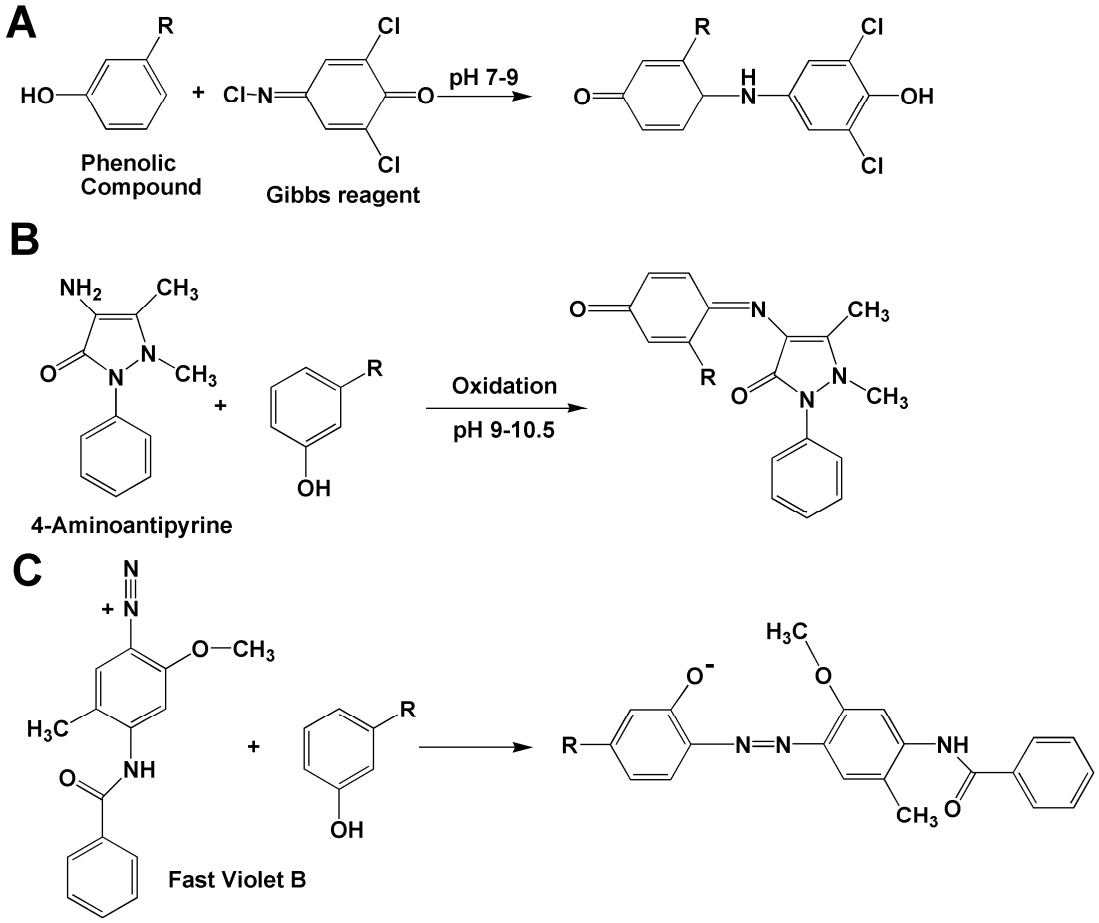

Figure 3.2. Chemistry of assay methods. (A) Coupling of Gibbs' reagent to a phenolic compound [65]. (B) Coupling of 4-aminoantipyrine to a phenolic compound [66]. (C) Coupling of Fast Violet B to a phenolic compound [67].

The 4-AAP and Gibbs' assays are very similar in the products they are able to detect (Table 3.2), as well as their sensitivity limits and coefficients of variation (*see* **Note 3.1**). These assays are linearly dependent on concentration for all of the phenols we have examined [64] (*see* **Note 3.2**) and give clear spectra with distinct λ_{max} in the presence of phenolic compounds, especially when compared to the FVB assay (Figure 3.3). In fact, FVB is typically less useful in general due to its lower sensitivity and reactivity with cells and various media, and the 4-AAP and Gibbs' assays are preferred.

Table 3.2 - The spectroscopic signals resulting from coupling of various phenols to Gibbs' reagent, Fast Violet B and 4-aminoantipyrine.

	Gibbs' assay		Fast Violet B assay		4-Aminoantipyrine assay	
Compound	λ_{max}	Max abs.	λ_{max}	Max abs.	λ_{max}	Max abs.
3-hydroxybenzaldehyde	670	0.06	n/a	< 0.05	580	0.18
2-hydroxybenzaldehyde	660	0.09	380	0.09	512	0.13
2-hydroxybenzamide	660	2.98	n/a	< 0.05	512	0.08
2,3-dihydroxybenzaldehyde	570	0.63	n/a	< 0.05	480	0.65
catechol	460	0.44	n/a	< 0.05	500	0.98
3-methylcatechol	460	0.49	n/a	< 0.05	540	0.25
3-fluorocatechol	450	0.38	n/a	< 0.05	535	0.49
phenol	630	0.10	n/a	< 0.05	500	2.57
o-cresol	610	0.31	n/a	< 0.05	505	1.85
m-cresol	620	0.17	n/a	< 0.05	500	1.76
2-aminophenol	600	0.76	440	0.08	520	1.07
3-aminophenol	570	1.61	480	0.82	470	2.61
2-chlorophenol	660	2.77	n/a	< 0.05	515	2.77
3-chlorophenol	670	1.25	n/a	< 0.05	545	0.78
1-naphthol	580	0.31 [a]	n.d.	n.d [b]	505	1.61
2-naphthol	n.d.	n.d.[b]	520	0.47	n.d.	n.d [c]
2,3-dihydroxynaphthalene	510	0.79	n.d.	n.d [b]	490	0.18
4-nitrophenol	n/a	< 0.05	n/a	< 0.05	n/a	< 0.05
2-hydroxypyridine	n/a	< 0.05	n/a	< 0.05	n/a	< 0.05
3-hydroxypyridine	600	0.39	n/a	< 0.05	n/a	< 0.05
o-coumaric acid	650	0.42	n/a	< 0.05	520	2.39
m-coumaric acid	670	0.24	n/a	< 0.05	540	0.87
p-coumaric acid	560	0.38	n/a	< 0.05	470	0.29
3-hydroxybenzoic acid	640	0.25	n/a	< 0.05	505	0.24
3,4-dihydroxybenzoic acid	460	0.29	n/a	< 0.05	495	0.75
3-hydroxy-4-methylbenzoic acid	610	1.21	n/a	< 0.05	465	0.90
2,3-dihydroxybenzoic acid	440	0.27	n/a	< 0.05	500	0.26

[a] Product slightly insoluble. [b].Product insoluble. [c].Color change to green that qickly fades to yellow with no λ_{max}.
n/a - not applicable, no λ_{max} n.d. - not determined due to insolubility

Compounds were diluted in M9 minimal media to a concentration of 0.25 mM and assayed as described. For the Gibbs' reagent, Fast Violet B and 4-aminoantipyrine assays, 0.1ml of phenol solution was assayed in a 96-well microtiter plate. Thirty minutes, 10 minutes and 10 minutes, respectively, were allowed for the reaction to occur before recording the visible spectra using a 96-well spectrophotometer.

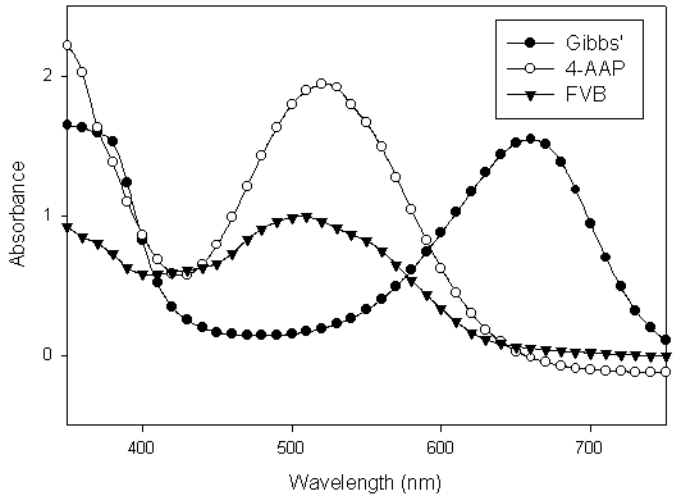

Figure 3.3. Spectra of potential assay products. 0.25 mM solutions of o-chlorophenol, o-coumaric acid and 2-naphthol in M9 media were assayed as described using Gibbs' reagent, 4-AAP and FVB, respectively.

The 4-AAP and Gibbs' assays are reported to react well with ortho- and meta-substituted phenolic compounds and with para-substituted compounds where the substituent is a halide or alkoxy group [68]. For example, if the substituent para to the site of hydroxylation is a carbon bond then the product will be less reactive with the 4-AAP and Gibbs' reagents. However, results to the contrary are shown with different regioisomers of hydroxynaphthalene in Figure 3.4. Substrates such as 2-naphthol are detectable with the 4-AAP assay even though the substituent para to the hydroxyl group is a tertiary carbon. These results also indicate that the 4-AAP assay is able to distinguish between regioisomers of hydroxynaphthalane by producing different spectra for each product. Another example of the value of the 4-AAP assay is its ability to detect 4-hydroxyatorvastatin, the main human metabolite of the drug atorvastatin (Lipitor). Finally, a critical difference between the assays, especially for the work here, is the 4-

AAP assay does not suffer from background signal due to H_2O_2 whereas the Gibbs' and FVB reagents do (Figure 3.5). This allows for the use of the 'peroxide-shunt' pathway to drive P450 catalysis in place of NADPH.

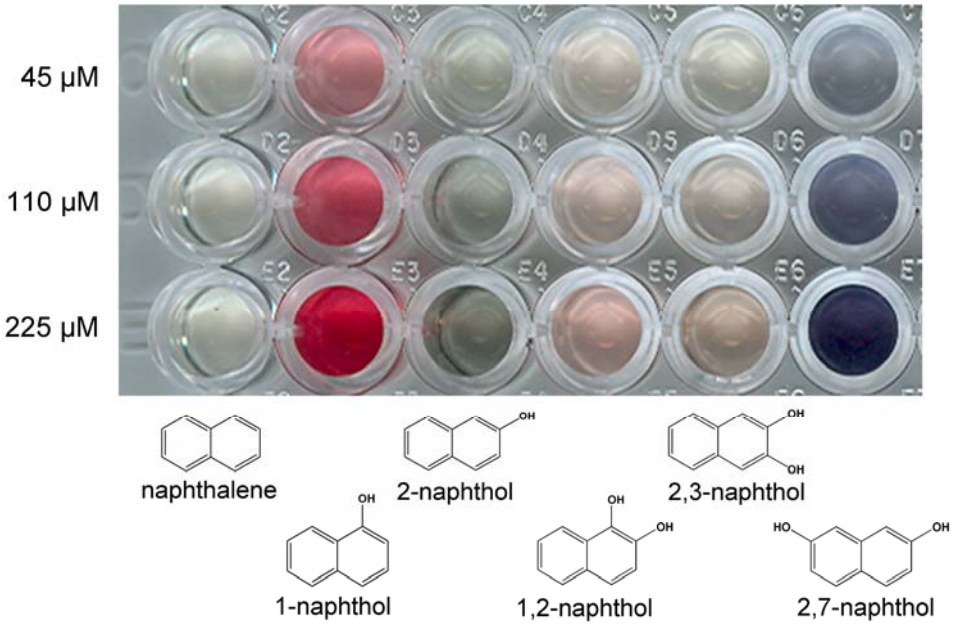

Figure 3.4. 4-aminoantipyrine assay is able to distinguish between regioisomers of hydroxynaphthalene.

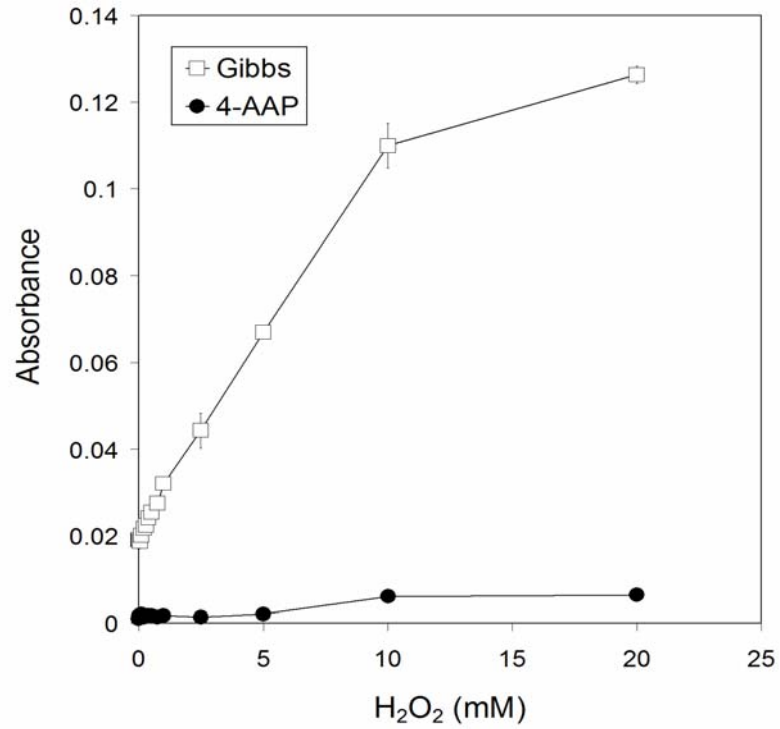

Figure 3.5. Assay background caused by H_2O_2 in the 4-AAP and Gibbs' assays. Absorbance readings for 4-AAP (500 nm) and Gibbs' (650 nm) in 100 mM Epps pH 8.2 with increasing amount of H_2O_2. The Gibbs' assay shows high level of background while 4-AAP shows almost none.

Some issues need to be considered when applying these assays to biotransformations using whole cells or cell extracts. When a whole cell system is used, careful consideration should be given to the method of supplying substrate to the enzyme. To access the enzyme, the substrate must be soluble and must readily permeate the cell membrane. Solubility can be increased in most cases by adding a nontoxic organic solvent [69]. Polymyxin B increases the permeability of many aromatic and non-aromatic substrates, including long chain fatty acids [70,71]. Though TB or LB-media are commonly used for whole cell growth, these rich broths contribute a significant amount of background to the assays discussed here (especially the Fast Violet B assay). This is easily remedied by

using a synthetic medium such as M9 minimal medium [72]. Supplying the substrate in the vapor phase is sometimes successful when the substrate is volatile and is particularly convenient when screening colonies using a solid-phase format [64].

In the assay descriptions below, "sample" refers to the solution containing the phenolic product to be determined and may be a cell extract or supernatant depending on which type of bioconversion is chosen. The absorbance value yields the total activity of the bioconversion. Times for color development are suggested below but this is another factor that varies from substrate to substrate and should be determined on an individual basis. Because every screening situation is unique, the following protocols should be adapted to each system.

D.1. Materials

1. 0.6% (w/v) 4-aminoantipyrine in H_2O. Store at 4°C and prepare fresh daily.
2. 0.6% (w/v) potassium persulfate in H_2O. Store at 4°C and prepare fresh daily.
3. 0.4% (w/v) 2,6-dichloroquinone-4-chloroimide (Gibbs' reagent) in ethanol. Store at 4°C and prepare fresh every 4 months.
4. 0.25% (w/v) Fast Violet B in H_2O. Prepare fresh every 2-3 days.
5. Quench: 100 mM NaOH and 4 M urea (See **Note 3.3**).
6. 96-well microtiter plates: R-96-OAPF-ICO (Rainin, Emeryville, CA).
7. Spectrophotometer/plate reader (Model Spectra max Plus 384, Molecular Devices, Sunnyvale, CA). Software Softmax Pro 3.1.1.
8. Benchtop centrifuge that can accommodate 96-well microtiter plates: Allegra 25R

Centrifuge (Beckman Coulter, Fullerton, CA).

9. Pipette robot: Multimek 96 Automated 96-Channel Pipetter (Beckman Instruments, Palo Alto, CA).

D.2. Phenol quantitation with Gibbs' reagent

1. To 100 μL of sample add 20 μL 0.4% (w/v) of Gibbs' reagent.
2. Mix and allow 3-30 minutes for color development (*see* **Note 3.4**).
3. Record spectrum or wavelength.

D.3. Phenol quantitation with 4-aminoantipyrine

1. To 100 μL of sample add 100 μL of quench (*see* **Note 3.5**).
2. Add 30 μL of 0.6% (w/v) 4-AAP. Mix thoroughly and incubate for 2 minutes.
3. Add 30 μL of 0.6% (w/v) potassium persulfate (*see* **Note 3.6**).
4. Mix and allow 10 minutes for color development (*see* **Note 3.7**).
5. Record spectrum or wavelength.

D.4. Phenol quantitation with Fast Violet B

1. To 100 μL of sample add 10 μL 0.25% of (w/v) Fast Violet B.
2. Mix and allow 10 minutes for color development (*see* **Note 3.8** and **Note 3.4**).
3. Record spectrum or wavelength.

D.5. Applying phenol detection with 4-AAP to cytochromes P450

1. In a 96-well microtiter plate, following whole cell or cell extract reaction, add an an

equivalent volume of quench (*see* **Notes 3.3, Note 3.5** and **Note 3.9**). If using a whole cell assay centrifuge at ~3500 *g* for 10 minutes and transfer the supernatant to a new microtiter plate (*see* **Note 3.10**).

2. Add 15 μL of 0.6% (w/v) 4-AAP per 100 μL. Mix and incubate for 2 minutes.
3. Add 15 μL of 0.6% (w/v) potassium persulfate per original 100 μL (*see* **Note 3.6**).
4. Record spectrum or wavelength after 10 minutes (*see* **Note 3.4**).

E. Notes

1.1. The positive pressures applied are not great and this piece of equipment can be improvised. To use a desiccator, vacuum grease is applied liberally around the edge where contact is made with the lid. When a plate is placed inside, the desiccator it is clamped shut with 4 c-clamps evenly distributed around the edges. This easily resists the positive pressure necessary for the assay. The desiccator is attached to a three-way stopcock with tubing. The three-way stopcock is connected to a vacuum pump on one end and the carbon monoxide tank on the other. This allows for switching between the vacuum pump and the carbon monoxide tank.

1.2. The high buffer concentration is to buffer against pH changes caused by sodium hydrosulfite, which lowers the pH of the solution. After adding the sodium hydrosulfite, time becomes more critical.

1.3. Volumes can be scaled up accordingly but should not exceed 300 μL since the maximum well volume of a 96-well microtiter plate is 320 μL. The assay can also be applied in 384-well plate format, which have a maximal working volume of approximately 110 μL. Enzyme can be diluted if necessary. Using the dilutions

described here, a well containing a total volume of 100 µl with 100 nM P450 has an A_{450} -A_{490} of approximately 0.003. Multiple plates can be assayed, but four is the suggested maximum at one time.

1.4. Bubbles due to pipetting/mixing can be removed by centrifugation at 3000 g or addition of a small amount of ethanol, using a pipette tip or spray bottle.

1.5. Spectra can be recorded every 10 nm from 400 to 500 nm or at smaller intervals. Specific points should be taken at the λ_{max} of your particular cytochrome P450, typically around 450 nm, and 490 nm [51]. Ultimately, the absorption change at 450 nm relative to the absorbance change at 490 nm is used to quanitate the amount of enzyme.

1.6. Alternatively, CO can be supplied at a slow rate equivalent to the bubbling used in cuvette format. This is approximately one to three bubbles per second from a Pasteur pipette attached to the carbon monoxide tank with tubing. This flow rate will quickly fill a sealed container and provide enough pressure to supply a sufficient amount of carbon monoxide.

1.7. The path length will vary based on volume and should be determined using standard cuvette assay. Path lengths have been determined to be approximately 0.37 and 0.71 cm for 100 and 200 µL, respectively. The path length can vary depending on the solution used due to the meniscus that forms. It is suggested that a control of known enzyme concentration be used in determining exact P450 concentrations.

2.1. Concentrations greater than 1 µM should be used. A typical concentration obtained from a 1 ml culture produced in 2 ml deep-well plates is 400 µl of a 1 to 10 µM solution.

2.2. Large total volumes will serve to maximize the path length in the microtiter plates. The maximum volumes allowed in 96-well and 384-well microtiter plates are 320 µl and 120 µl, respectively. Substrate solution should be added in a small but accurate volume for multichannel pipettes (e.g. 4 µl) although larger volumes may be used without concern of dilution since concentration is inversely proportional to path length. Volumes of 290 µl for 96-well plates and 90 µl for 384-well plates allow for addition of substrate seven to eight times.

2.3. The individual wavelengths that should be measured needs to be determined for each case and depends on the substrate being used, the individual P450 and whether there is type I or type II binding spectra. The wavelengths which give the maximum difference as substrate is added should be used.

3.1. Coefficients of variation tend to be approximately 8-14% for the Gibbs' and 4-AAP assays. The sensitivity limit for the Gibbs' reagent and 4-AAP is approximately 10 µM. Tris has been found to give a higher background level than other buffer systems. This should

be considered when the amount of product formed is near the detection limit.

3.2. Phenol, o-coumaric acid and m-coumaric acid have been checked for a linear dependence of concentration on absorbance with the 4-AAP assay. 3-methylcatechol and 3-fluorocatechol have been tested with the Gibbs' assay.

3.3. Since cytochrome P450s typically perform a single hydroxylation, no other preparative steps are necessary as they are with dioxygenases [64]. Reactions should be quenched in a suitable manner which terminates the enzymatic reaction allowing for reproducible time points. The NaOH/urea (quench) solution will provide this in

some systems and is effective with the P450s discussed here.

3.4. Optimal development time depends on the phenol assayed and, in some cases, accumulation of background absorbance over time. When assaying for improved enzyme function, only the wavelength of the product is taken and not the entire spectrum.

3.5. A pH between 9 and 10.5 is desired since 4-AAP will react with peroxides at a lower pH and produce a red color [66,73]. This also aids in solubilizing any precipitate from cell extract and/or substrate added. 4-AAP works in up to 75% DMSO in water.

3.6. Potassium persulfate serves to oxidize the NH_2 group of 4-aminoantipyrine, making it available for electrophilic attack by a phenol. Other oxidizing agents can be used such as ammonium persulfate and potassium hexacyanoferrate [66]. Potassium persulfate gave the least amount of background in the P450 system.

3.7. The reproducibility of the 4-AAP assay is not very sensitive to incubation times beyond 10 minutes. It is a good idea, however, to keep the incubation time constant from screen to screen in order to minimize variance. Centrifuging between addition of 4-AAP and potassium persulfate can be useful in case there is any residual liquid on the sides of the plate-wells. Ethanol can be used to remove any bubbles formed while pipetting, either from a spray bottle or on the tip of a pipette.

3.8. Increasing the pH to basic levels before addition of Fast Violet B can be useful in increasing the maximum absorbance value. It is not necessary, however.

3.9. A pipetting robot can be useful when doing multiple 96-well microtiter plate assays, but is not necessary.

3.10. Removal of cell debris is not necessary in the Gibbs' and 4-AAP assays, however it increases the reproducibility of the screens. It is necessary for FVB.

F. References

1. Wong TS, Wu N, Roccatano D, Zacharias M, Schwaneberg U (2005) Sensitive assay for laboratory evolution of hydroxylases toward aromatic and heterocyclic compounds. J. Biomol. Screen. 10: 246-252.
2. Meinhold P (2005) Engineering cytochrome P450 BM-3 for selective hydroxylation of alkanes. Pasadena: California Institute of Technology. 243 p.
3. Bloom JD, Labthavikul ST, Otey CR, Arnold FH (2006) Protein stability promotes evolvability. Proc. Natl. Acad. Sci. U.S.A. 103: 5869-5874.
4. Urlacher VB, Makhsumkhanov A, Schmid RD (2006) Biotransformation of beta-ionone by engineered cytochrome P450 BM-3. Appl. Microbiol. Biotechnol. 70: 53-59.
5. Glieder A, Farinas ET, Arnold FH (2002) Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. Nat. Biotechnol. 20: 1135-1139.
6. Glieder A, Meinhold P (2003) High-throughput screens based on NAD(P)H depletion. Methods Mol. Biol. 230: 157-170.
7. Axarli I, Prigipaki A, Labrou NE (2005) Engineering the substrate specificity of cytochrome P450 CYP102A2 by directed evolution: production of an efficient enzyme for bioconversion of fine chemicals. Biomol. Eng. 22: 81-88.
8. Joo H, Arisawa A, Lin Z, Arnold FH (1999) A high-throughput digital imaging screen for the discovery and directed evolution of oxygenases. Chem. Biol. 6: 699-706.
9. Joo H, Lin Z, Arnold FH (1999) Laboratory evolution of peroxide-mediated cytochrome P450 hydroxylation. Nature 399: 670-673.
10. Cirino PC, Arnold FH (2003) A self-sufficient peroxide-driven hydroxylation biocatalyst. Angew Chem. Int. Ed. Engl. 42: 3299-3301.
11. Farinas E (2003) Colorimetric screen for aliphatic hydroxylation by cytochrome P450 using p-nitrophenyl-substituted alkanes. Methods Mol. Biol. 230: 149-156.
12. Schwaneberg U, Schmidt-Dannert C, Schmitt J, Schmid RD (1999) A continuous spectrophotometric assay for P450 BM-3, a fatty acid hydroxylating enzyme, and its mutant F87A. Anal. Biochem. 269: 359-366.
13. Zitting A (1981) A sensitive liquid chromatographic assay of ethoxycoumarin deethylase with fluorescence detection. Anal. Biochem. 115: 177-180.
14. Waxman DJ, Chang TK (2006) Use of 7-ethoxycoumarin to monitor multiple enzymes in the human CYP1, CYP2, and CYP3 families. Methods Mol. Biol. 320: 153-156.
15. Stresser DM, Turner SD, Blanchard AP, Miller VP, Crespi CL (2002) Cytochrome P450 fluorometric substrates: identification of isoform-selective probes for rat CYP2D2 and human CYP3A4. Drug Metab. Dispos. 30: 845-852.
16. Marks BD, Smith RW, Braun HA, Goossens TA, Christenson M, et al. (2002) A high throughput screening assay to screen for CYP2E1 metabolism and inhibition using a fluorogenic vivid p450 substrate. Assay Drug Dev. Technol. 1: 73-81.
17. Marks BD, Goossens TA, Braun HA, Ozers MS, Smith RW, et al. (2003) High-throughput screening assays for CYP2B6 metabolism and inhibition using fluorogenic vivid substrates. AAPS PharmSci. 5: E18.

18. White IN (1988) A continuous fluorometric assay for cytochrome P-450-dependent mixed function oxidases using 3-cyano-7-ethoxycoumarin. Anal. Biochem. 172: 304-310.
19. Donato MT, Jimenez N, Castell JV, Gomez-Lechon MJ (2004) Fluorescence-based assays for screening nine cytochrome P450 (P450) activities in intact cells expressing individual human P450 enzymes. Drug Metab. Dispos. 32: 699-706.
20. Bapiro TE, Egnell AC, Hasler JA, Masimirembwa CM (2001) Application of higher throughput screening (HTS) inhibition assays to evaluate the interaction of antiparasitic drugs with cytochrome P450s. Drug Metab. Dispos. 29: 30-35.
21. Buters JT, Schiller CD, Chou RC (1993) A highly sensitive tool for the assay of cytochrome P450 enzyme activity in rat, dog and man. Direct fluorescence monitoring of the deethylation of 7-ethoxy-4-trifluoromethylcoumarin. Biochem. Pharmacol. 46: 1577-1584.
22. Kim D, Guengerich FP (2004) Enhancement of 7-methoxyresorufin O-demethylation activity of human cytochrome P450 1A2 by molecular breeding. Arch. Biochem. Biophys. 432: 102-108.
23. Lussenburg BM, Babel LC, Vermeulen NP, Commandeur JN (2005) Evaluation of alkoxyresorufins as fluorescent substrates for cytochrome P450 BM3 and site-directed mutants. Anal. Biochem. 341: 148-155.
24. Yun CH, Miller GP, Guengerich FP (2000) Rate-determining steps in phenacetin oxidations by human cytochrome P450 1A2 and selected mutants. Biochemistry 39: 11319-11329.
25. Parikh A, Josephy PD, Guengerich FP (1999) Selection and characterization of human cytochrome P450 1A2 mutants with altered catalytic properties. Biochemistry 38: 5283-5289.
26. Sakai-Kato K, Kato M, Homma H, Toyo'oka T, Utsunomiya-Tate N (2005) Creation of a P450 array toward high-throughput analysis. Anal. Chem. 77: 7080-7083.
27. Sieber V, Martinez CA, Arnold FH (2001) Libraries of hybrid proteins from distantly related sequences. Nat. Biotechnol. 19: 456-460.
28. Crespi CL, Miller VP, Penman BW (1997) Microtiter plate assays for inhibition of human, drug-metabolizing cytochromes P450. Anal. Biochem. 248: 188-190.
29. Cali J (2003) Screen for CYP450 inhibitors using P450-Glo luminescent cytochrome P450 assays. Cell Notes: 2-4.
30. Peters MW, Meinhold P, Glieder A, Arnold FH (2003) Regio- and enantioselective alkane hydroxylation with engineered cytochromes P450 BM-3. J. Am. Chem. Soc. 125: 13442-13450.
31. Kubo T, Peters MW, Meinhold P, Arnold FH (2006) Enantioselective epoxidation of terminal alkenes to (R)- and (S)-epoxides by engineered cytochromes P450 BM-3. Chemistry 12: 1216-1220.
32. Alcalde M, Farinas ET, Arnold FH (2004) Colorimetric high-throughput assay for alkene epoxidation catalyzed by cytochrome P450 BM-3 variant 139-3. J. Biomol. Screen 9: 141-146.
33. Gillam EM, Notley LM, Cai H, De Voss JJ, Guengerich FP (2000) Oxidation of indole by cytochrome P450 enzymes. Biochemistry 39: 13817-13824.

34. Li QS, Schwaneberg U, Fischer P, Schmid RD (2000) Directed evolution of the fatty-acid hydroxylase P450 BM-3 into an indole-hydroxylating catalyst. Chemistry 6: 1531-1536.
35. Wu ZL, Podust LM, Guengerich FP (2005) Expansion of substrate specificity of cytochrome P450 2A6 by random and site-directed mutagenesis. J. Biol. Chem. 280: 41090-41100.
36. Yoshioka H, Kasai N, Ikushiro S, Shinkyo R, Kamakura M, et al. (2006) Enzymatic properties of human CYP2W1 expressed in Escherichia coli. Biochem. Biophys. Res. Commun. 345: 169-174.
37. Guengerich FP, Sorrells JL, Schmitt S, Krauser JA, Aryal P, et al. (2004) Generation of new protein kinase inhibitors utilizing cytochrome p450 mutant enzymes for indigoid synthesis. J. Med. Chem. 47: 3236-3241.
38. Gillam EM, Aguinaldo AM, Notley LM, Kim D, Mundkowski RG, et al. (1999) Formation of indigo by recombinant mammalian cytochrome P450. Biochem. Biophys. Res. Commun. 265: 469-472.
39. Nakamura K, Martin MV, Guengerich FP (2001) Random mutagenesis of human cytochrome p450 2A6 and screening with indole oxidation products. Arch. Biochem. Biophys. 395: 25-31.
40. Celik A, Speight RE, Turner NJ (2005) Identification of broad specificity P450CAM variants by primary screening against indole as substrate. Chem. Commun. (Camb): 3652-3654.
41. Abecassis V, Pompon D, Truan G (2000) High efficiency family shuffling based on multi-step PCR and in vivo DNA recombination in yeast: statistical and functional analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2. Nucleic Acids Res. 28: E88.
42. Abecassis V, Urban P, Aggerbeck L, Truan G, Pompon D (2003) Exploration of natural and artificial sequence spaces: Towards a functional remodeling of membrane-bound cytochrome P450. Biocatal. Biotransform. 21: 55-66.
43. Appel D, Schmid RD, Dragan CA, Bureik M, Urlacher VB (2005) A fluorimetric assay for cortisol. Anal. Bioanal. Chem. 383: 182-186.
44. Kim MJ, Kim H, Cha IJ, Park JS, Shon JH, et al. (2005) High-throughput screening of inhibitory potential of nine cytochrome P450 enzymes in vitro using liquid chromatography/tandem mass spectrometry. Rapid Commun. Mass Spectrom. 19: 2651-2658.
45. Lentz O, Feenstra A, Habicher T, Hauer B, Schmid RD, et al. (2006) Altering the regioselectivity of cytochrome P450 CYP102A3 of Bacillus subtilis by using a new versatile assay system. Chembiochem 7: 345-350.
46. Omura T, Sato R (1964) The carbon monoxide-binding pigment of liver microsomes. II. Solubilization, purification, and properties. J. Biol. Chem. 239: 2379-2385.
47. Omura T, Sato R (1964) The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. J. Biol. Chem. 239: 2370-2378.
48. Martinis SA, Blanke SR, Hager LP, Sligar SG, Hoa GH, et al. (1996) Probing the heme iron coordination structure of pressure-induced cytochrome P420cam. Biochemistry 35: 14530-14536.

49. Wells AV, Li P, Champion PM, Martinis SA, Sligar SG (1992) Resonance Raman investigations of Escherichia coli-expressed Pseudomonas putida cytochrome P450 and P420. Biochemistry 31: 4384-4393.
50. Yu C, Gunsalus IC (1974) Cytochrome P-450cam. II. Interconversion with P-420. J. Biol. Chem. 249: 102-106.
51. Schenkman JB, Jansson I (1998) Spectral analyses of cytochromes P450. In: Phillips IR, Shephard EA, editors. Cytochrome P450 Protocols. Totowa, New Jersey: Human Press, Inc. pp. 25-33.
52. Schenkman JB (1970) Studies on the nature of the type I and type II spectral changes in liver microsomes. Biochemistry 9: 2081-2091.
53. Schenkman JB, Sligar SG, Cinti DL (1981) Substrate interaction with cytochrome P-450. Pharmacol. Ther. 12: 43-71.
54. Kumaki K, Sato M, Kon H, Nebert DW (1978) Correlation of type I, type II, and reverse type I difference spectra with absolute changes in spin state of hepatic microsomal cytochrome P-450 iron from five mammalian species. J. Biol. Chem. 253: 1048-1058.
55. Kumaki K, Nebert DW (1978) Spectral evidence for weak ligand in sixth position of hepatic microsomal cytochrome P-450 low spin ferric iron in vivo. Pharmacology 17: 262-279.
56. Lewis DFV (2001) Guide to cytochromes P450: structure and function. London; New York: Taylor & Francis.
57. Holland HL (1992) Organic synthesis with oxidative enzymes. New York.
58. Sheldrake GN (1992) Biologically derived arene cis-dihydrodiols as synthetic building blocks. New York: John Wiley & Sons.
59. Faber K (2000) Biotransformations in organic chemistry. Berlin: Springer-Verlag.
60. Wackett LP, Hershberger CD (2001) Biocatalysis and biodegradation: Microbial transformation of organic compounds. Washington, D. C.: ASM Press.
61. Jones JP, O'Hare EJ, Wong LL (2001) Oxidation of polychlorinated benzenes by genetically engineered CYP101 (cytochrome P450(cam)). Eur. J. Biochem. 268: 1460-1467.
62. Gibson DT, Parales RE (2000) Aromatic hydrocarbon dioxygenases in environmental biotechnology. Curr. Opin. Biotechnol. 11: 236-243.
63. Sakamoto T, Joern JM, Arisawa A, Arnold FH (2001) Laboratory evolution of toluene dioxygenase to accept 4-picoline as a substrate. Appl. Environ. Microbiol. 67: 3882-3887.
64. Joern JM, Sakamoto T, Arisawa A, Arnold FH (2001) A versatile high-throughput screen for dioxygenase activity using solid-phase digital imaging. J. Biomol. Screen. 6: 219-223.
65. Quintana MG, Didion C, Dalton H (1997) Colorimetric method for a rapid detection of oxygenated aromatic biotransformation products. Biotechnology Techniques 11: 585-587.
66. Fiamegos YC, Stalkikas CD, Pilidis GA, Karayannis MI (2000) Synthesis and analytical applications of 4-aminopurazolone derivatives as chromogenic agents for the spectrophotometric determination of phenols. Anal. Chim. Acta 403: 315-323.

67. Zollinger H (1991) Color chemistry: Syntheses, properties and applications of organic dyes and pigments. New York, NY: VCH Publishers, Inc.
68. Josephy PD, Van Damme A (1984) Reaction of 4-substituted phenols with benzidine in a peroxidase system. Biochem. Pharmacol. 33: 1155-1156.
69. Harrop AJ, Woodley JM, Lilly MD (1992) Production of naphthalene-cis-glycol by Pseudomonas putida in the presence of organic solvents. Enzyme Microb. Technol. 14: 725-730.
70. Vaara M (1992) Agents that increase the permeability of the outer membrane. Microbiol. Rev. 56: 395-411.
71. Schwaneberg U, Otey C, Cirino PC, Farinas E, Arnold FH (2001) Cost-effective whole-cell assay for laboratory evolution of hydroxylases in Escherichia coli. J. Biomol. Screen. 6: 111-117.
72. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: A laboratory manual. New York: Cold Springs Harbor Laboratory Press.
73. Faust SD, Mikulewicz EW (1967) Factors influencing the condensation of 4-aminoantipyrine with derivatives of hydroxybenzene-II. Influence of hydronium ion concentration on absorptivity. Water Research. Great Britain: Pergamon Press. pp. 509-522.

Chapter 4

Functional evolution and structural conservation in chimeric cytochromes P450: calibrating a structure-guided approach

Material from this chapter appears in: Otey, C.R., Silberg, J.J., Voigt, C.A., Endelman, J.B., Bandara, G., and Arnold, F.H. (2004) **Functional evolution and structural conservation in chimeric cytochromes P450: calibrating a structure-guided approach**. *Chem. Biol*. 11: 309-318, and is reprinted with permission from Elsevier publishing. Additional thanks to Claes von Wachenfeldt for kindly providing the CYP102A2 gene. This work was supported by the Army Research Office, the W. M. Keck Foundation, Maxygen Corporation, National Institutes of Health Grant R01 GM068664-01 and Fellowship F32 GM64949-01 (to J.J.S.), National Science Foundation (to C.A.V.), Burroughs-Welcome Fund (to C.A.V.) and National Defense Science and Engineering Fellowship (to J.B.E.).

A. Abstract

Recombination generates chimeric proteins whose ability to fold and function depends on minimizing structural perturbations that result when portions of the sequence are inherited from different parents. If the folded structure is retained, the new chimeric sequences can display functional properties characteristic of the parents. They can also acquire entirely new functions. Seventeen double-crossover chimeras were generated from two members of the CYP102 subfamily of the functionally diverse cytochrome P450 heme enzymes. These hybrid proteins cover a wide range of structural disruption, as defined by the number of residue-residue contacts broken upon recombination and computed using the SCHEMA algorithm, based on the protein's three-dimensional structure. Chimeras predicted to have limited structural disruption displayed CO-binding spectra characteristic of folded P450s, and many were catalytically active. Even this small population of chimeras exhibited significant functional diversity: chimeras displayed altered substrate specificity profiles, a wide range in thermostabilities, up to a 40-fold increase in peroxidase activity, and ability to hydroxylate a substrate towards which neither parent heme domain shows detectable activity. These results suggest that SCHEMA-guided recombination can be used to generate diverse P450s for exploring function evolution within the P450 structural framework.

B. **Introduction**

The cytochrome P450 superfamily of enzymes exhibits an impressive range of chemical activities and biological roles. Nature has utilized these diverse enzymes for everything from steroid biosynthesis to interspecies chemical warfare, drug detoxification and utilization of new food sources [1-5]. Individual members of the superfamily, however, show a much narrower range of catalytic activities (usually catalyzing oxygen insertion into C-H bonds) and substrate specificities. The heme prosthetic group recruited by cytochrome P450 to effect monooxygenation is also used by these and other proteins for oxygen transport, electron transfer, reduction, dealkylation and dehalogenation [6,7]. The highly versatile cytochrome P450 family offers unique opportunities to investigate the evolution of function within a single structural framework [8].

In this study we begin to explore the generation of P450s in the laboratory by recombination of homologous sequences. P450s typically exhibit low sequence identity, and annealing-based DNA-shuffling techniques [9-13] are not useful for creating highly diverse libraries of P450 chimeras. While several methods for making such shuffled gene libraries independent of sequence homology have been described [14-19], these approaches generate few crossovers and large numbers of inactive sequences, due to insertions, deletions, frameshifts, as well as disruptive crossover events. Functional characterization of such libraries is difficult without a selection to remove unfolded or nonfunctional sequences.

Recently, we reported a computational algorithm, SCHEMA, which can estimate the disruption caused by swapping different fragments among structurally-similar proteins and identify optimal crossover locations for making libraries by recombination [20]. Using the 3D structure of one of the parent proteins, the algorithm identifies pairs of amino acids that are interacting (e.g., residues within a cutoff distance of 4.5 Å) and determines the net number of interactions broken when a chimeric protein inherits portions of its sequence from different parents (*E*). By comparing SCHEMA disruption predictions to functional β-lactamases selected from a large library of chimeric sequences, we demonstrated that sequences retaining the parental protein fold and function tend to have low *E* values [21]. This criterion can be used to select crossover positions for individual chimeras or combinatorial libraries prepared by swapping elements from related parent sequences.

Here we explore the effects of recombination in a larger and more complex enzyme, the cytochrome P450. Seventeen double-crossover chimeras were made by swapping fragments between the heme domains of the soluble, bacterial enzyme A1 and A2, which are approximately 460 amino acids in length and share 63% amino acid identity. We have determined which sequences encode properly-folded heme domains and related those to the disruption calculated by SCHEMA. To probe the functional diversity of this small population, we have measured their stabilities and activities in different P450-catalyzed reactions. A subset of the heme domain chimeras has been reconstituted into holoenzymes by fusion with the CYP102A1 reductase domain and characterized. These

data are used in Chapter 4 to guide much larger efforts to explore the functional variation that is possible within the P450 scaffold.

C. Results

C.1 Chimera design

For this work we constructed chimeras of CYP102A1 [22] and CYP102A2 [23], homologs from *Bacillus megaterium* and *Bacillus subtilis*, respectively. These soluble fusion proteins, consisting of a catalytic heme domain and an FAD- and FMN-containing NADPH reductase [24], require dioxygen and a cofactor (NADPH) for monooxygenase activity. However, the P450 heme domain can also utilize hydrogen peroxide via the 'peroxide shunt' pathway to catalyze hydroxylation reactions. While this peroxygenase activity is low in CYP102A1, it is enhanced by the amino acid substitution F87A [25,26]; the equivalent F88A mutation in CYP102A2 has a similar effect. The P450 chimeras were constructed from the genes for the heme domains of CYP102A1 with the F87A mutation and CYP102A2 with the F88A mutation (abbreviated as A1 and A2). With these mutations, we can use the peroxygenase activity of the heme domain to explore substrate specificities in the chimeras, without having to supply cofactor or a reductase (which may or may not interact properly with a chimeric heme domain). The chimeric heme domain can also be fused to one of the parental reductase domains to regenerate a chimeric holoenzyme (see below).

Initial attempts to randomly recombine the A1 and A2 sequences were unsuccessful due to their 64% nucleotide identity, which places them below the limits for effective

recombination using annealing-based methods [9]. Both DNA shuffling with a fixed annealing temperature, as originally described [9], and with gradually decreasing annealing temperatures [27,28] yielded mostly reconstructed parental sequences rather than shuffled genes. In addition, random chimeragenesis on transient templates (RACHITT) [10,29], a method meant to increase the frequency of crossovers and decrease the sequence identity for a crossover to occur, was used. However, this method was also unsuccessful due to the inability to form single-stranded templates of the P450 genes containing uracil instead of thymine.

The effective levels of mutation (amino acid Hamming distance from the closest parent) and SCHEMA disruption (*E*) were calculated for all possible double-crossover chimeras of A1 and A2 with a minimum fragment size of ten amino acids. The distribution in the levels of disruption and effective mutation (*m*) for this population of chimeras can be seen in Figure 4.1. Fourteen chimeras were individually designed and constructed to encompass a broad range of *E* (2 to 42) and *m* (11 to 70) (Table 4.1). Crossovers were placed in regions of low and high sequence identity; in the 10 bp surrounding the crossovers, identity ranged from 20-90%. For three of the chimeras we generated both 'mirror' sequences, i.e. chimeras that derive sequences from opposite parents at every position. The other eleven sequences consisted of A1 with an internal fragment derived from A2 (Figure 4.2).

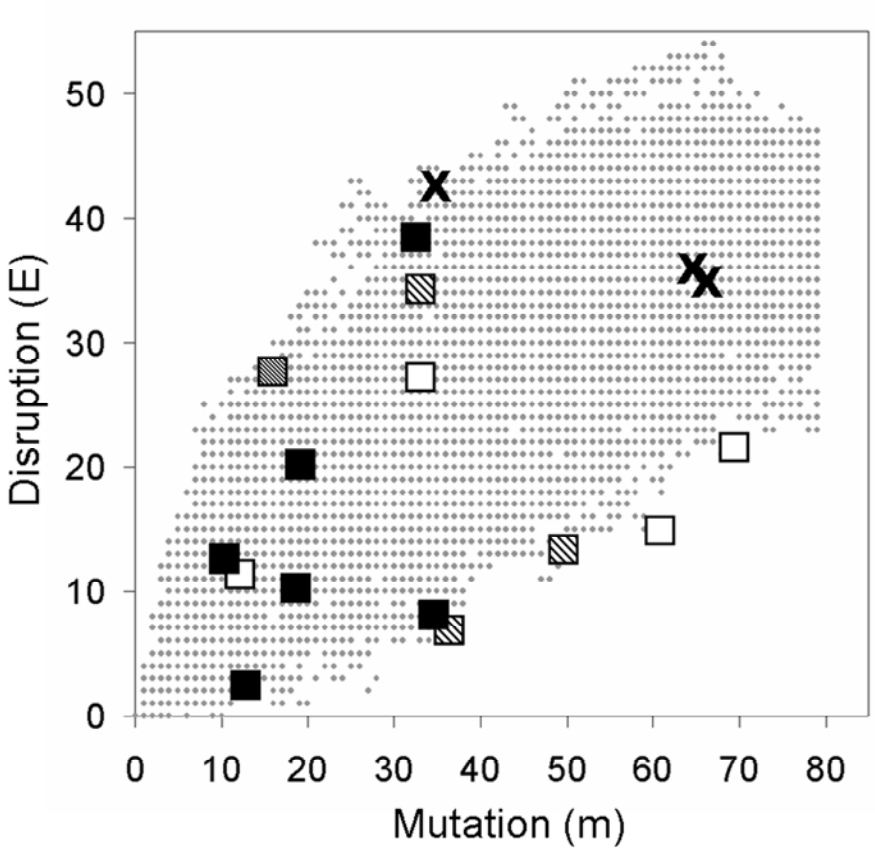

Figure 4.1. Effective mutation (*m*) and disruption (*E*) for cytochrome P450 heme-domain chimeras. Disruption values for all double-crossover chimeras with a minimum insert size of 10 amino acids between A1 and A2 were determined using the structure for the CYP102A1 heme-domain. Disruption values from the substrate-bound structure were used. For all double-crossover chimeras, the average *E* is 23.5±12.8, and average mutation *m* is 40.6±22.6. The 17 constructed chimeras were assayed for the ability to fold and hydroxylate 12-pNCA, 2-phenoxyethanol, and allyloxybenzene: Squares represent chimeras that retain the ability to fold, and X's indicate those that did not. Chimeras that fold but have little or no detectable peroxygenase activity (□), chimeras with parent-like substrate specificity profiles (■), chimeras with altered profiles (▨ , ▩), chimeras with altered profiles and (new) activity on allyloxybenzene (▨).

Table 4.1. Properties of designed A1-A2 chimeric P450s

Protein[a]	E (substrate-bound)[b]	E (substrate-free)[b]	m[c]	Folded (λ max)[d]	T_m (ºC)	Peroxidase activity[e]
A1	--	--	--	yes (448)	55	2.6±0.1
A2	--	--	--	yes (449)	44	0.4±0.1
364-403	2	2	13	yes (449)	51	1±0.2
165-256	7	7	36	yes (449)	48	16.1±1.4
165-256M	[7]	[7]	36	yes (448)	50	3.5±0.4
285-341	10	9	19	yes (448)	53	2.1±0.1
191-335	12	12	50	yes (449)	40	N.D.
169-197	12	10	11	yes (449)	52	100.3±3.1
169-197M	[12]	[10]	11	yes (449)	43	0.3±0.1
65-256	15	13	61	yes (447)	36	N.D.
118-194	20	19	20	yes (449)	47	34.3±1.2
70-299	21	19	70	yes (448)	42	0.8±0.3
46-73	27	30	16	yes (448)	55	6.8±0.5
277-365	27	28	33	yes (421)	39	N.D.
43-135	34	38	33	yes (448)	53	6.8±0.4
186-365	34	35	65	no	--	N.D.
186-365M	[34]	[35]	65	no	--	N.D.
50-140	38	39	32	yes (448)	52	10.4±0.6
345-448	42	43	34	no	--	N.D.

[a] A1 and A2 refer to the isolated heme domains of CYP102A1 (with the F87A substitution) and CYP102A2 (with the F88A substitution). Chimera names correspond to the first and last residue of A2 inserted into A1 according to the numbering of CYP102A1. 'M' indicates mirror chimeras where A1 is inserted into A2.

[b] SCHEMA-calculated disruption (see methods) based on substrate-bound (1JPZ) [30] and substrate-free structures (2HPD) [31]. Brackets [ ] indicate assumed disruption for mirror chimeras (due to lack of crystal structure of CYP102A2).

[c] Effective level of mutation (= amino acid Hamming distance to closest parent).

[d] Folding as assayed by reduced CO difference spectroscopy. λmax for Soret band is reported.

[e] Values reported in nmol product/nmol P450/min. Activities < 0.2 were not detectable (N.D.).

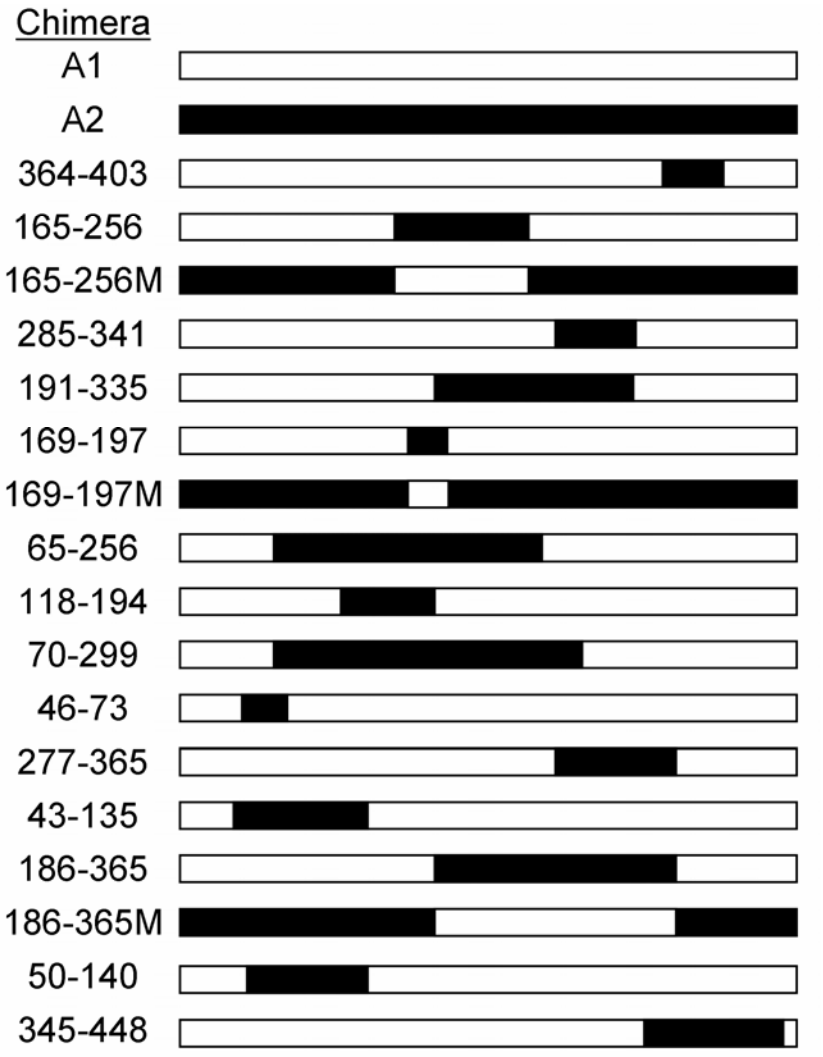

Figure 4.2. Location of crossovers in primary sequence of parent P450s, A1 and A2, and seventeen chimeras.

E values for the different chimeras were computed using the high-resolution structure for CYP102A1 with palmitoglycine bound in the active site [30]. Because previous studies have shown that substrate binding causes a large conformational change in CYP102A1 [32-34], we also calculated *E* using the substrate-free CYP102A1 structure [31]. As shown in Table 4.1, similar *E* values were obtained for the two calculations. Because

both parents contain the same heme cofactor, contacts between the heme and the protein cannot be broken upon recombination, at least in this simple model. It is assumed that chimeras retain parental heme contacts, and heme is not included in the calculation of *E*.

C.2 Folding of chimeric heme domains

The chimeras were constructed using SOEing [35], cloned into the IPTG-inducible pCWori vector [36], and sequenced to confirm the absence of point mutations. All proteins were overexpressed in a catalase-free strain of *E. coli* [37], which allows the peroxygenase activity of the heme domains to be monitored directly in cell extracts [38]. We used carbon monoxide difference spectroscopy to assess the level of structural disruption in the chimeras: a reduced CO difference spectrum producing a Soret band near 450 nm is indicative of heme incorporation and thus a correctly-folded P450 heme domain [39]. A Soret band near 420 nm is indicative of a folded protein that binds heme, but is catalytically inactive due to a disrupted heme environment [40,41]. Fourteen chimeras displayed detectable Soret bands: thirteen appeared at 450 nm and one at 420 nm (Figure 4.3). Chimeras with low calculated disruption (*E*) were most likely to retain folded structures: all with $E < 30$ were folded, but less than half with $E > 30$ yielded detectable Soret bands (Table 4.1).

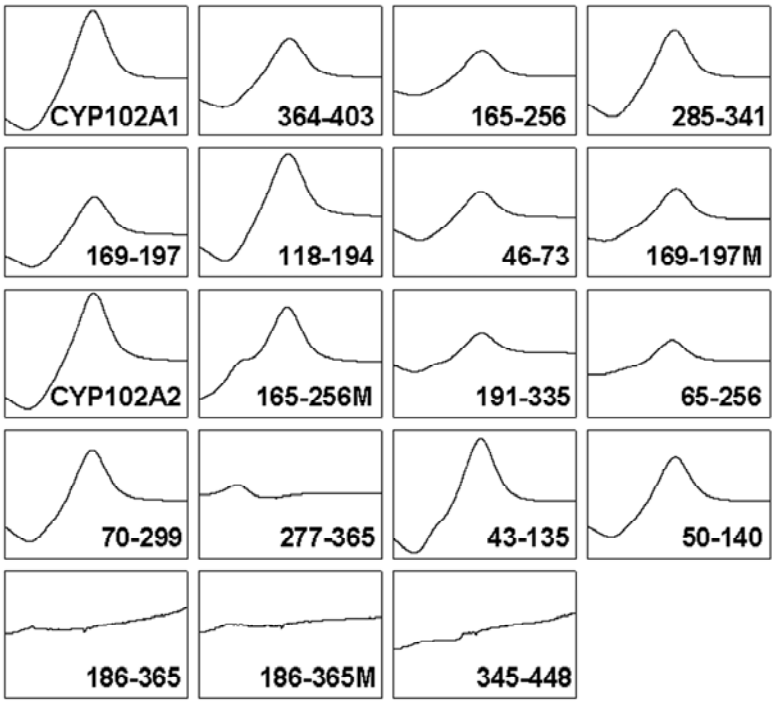

Figure 4.3. Reduced CO-difference spectra for A1, A2, and their (heme-domain) chimeras. Spectra were taken from 400 to 500 nm. The absorbance range for the first two rows is -0.5 to 0.6, the third and fourth rows are magnified 3X with a range of -0.17 to 0.2, and the last row is magnified 40X over the first row. Most chimeras exhibit a Soret band at 450±3 nm, characteristic of a folded P450 with correctly incorporated heme cofactor. Chimera 277-365 shows a Soret band at 420 nm, and no Soret band could be detected for chimeras 186-365, 186-365M and 345-448.

C.3 Peroxygenase activities of chimeric heme domains

We assayed the chimeric P450 heme domains for hydroxylation of *p*-nitrophenoxydodecanoic acid (12-pNCA), a fatty acid analog that is hydroxylated by A1 and A2 to yield 4-nitrophenol [42]. Initial rates were measured using a concentration of 12-pNCA (250 μM) significantly higher than the K_M of CYP102A1 for this substrate (K_M = 8.1 μM [43]). Activities on 2-phenoxyethanol and allyloxybenzene were also

determined, using the 4-AAP assay, which is sensitive to phenols and catechols (see Chapter 3). This assay yields a detectable product if hydroxylation occurs at the ortho or meta positions of the aromatic ring or when hydroxylation yields the hemiacetal, which decomposes to form phenol. At the maximum soluble concentrations of substrate, the parent CYP102 heme domains were active only on 2-phenoxyethanol; neither showed measurable peroxygenase activity towards allyloxybenzene.

Figure 4.4 compares the activities of the folded chimeric heme domains to those of the parent enzymes. Unfolded chimeras showed no activity towards any substrate, while all but two of the folded ones retained peroxygenase activity on at least one substrate. The P420 chimera (277-365) was inactive towards all substrates tested. Several chimeric heme domains were more active than the best parent, A1, on one or more substrate. Chimera 169-197 was the most active towards 12-pNCA, 46-73 had the highest activity on 2-phenoxyethanol, and 43-135 and 165-256 were the most active on allyloxybenzene. The chimeras also showed different specificities, falling roughly into three groups: 1) chimeras with little or no detectable activity towards any substrate; 2) parent-like chimeras, active on 12-pNCA and 2-phenoxyethanol; and 3) chimeras with altered substrate specificities relative to the parents, due to loss of activity towards 12-pNCA and/or acquisition of activity on allyloxybenzene. For the three chimeras with this ‘novel’ activity, one (165-256) had broadened specificity and was active on all three substrates. The remaining two (191-335 and 43-135) showed detectable activity on 2-phenoxyethanol and allyloxybenzene, but not 12-pNCA. The members of each pair of mirror chimeras had equivalent folding properties, but were not functionally equivalent.

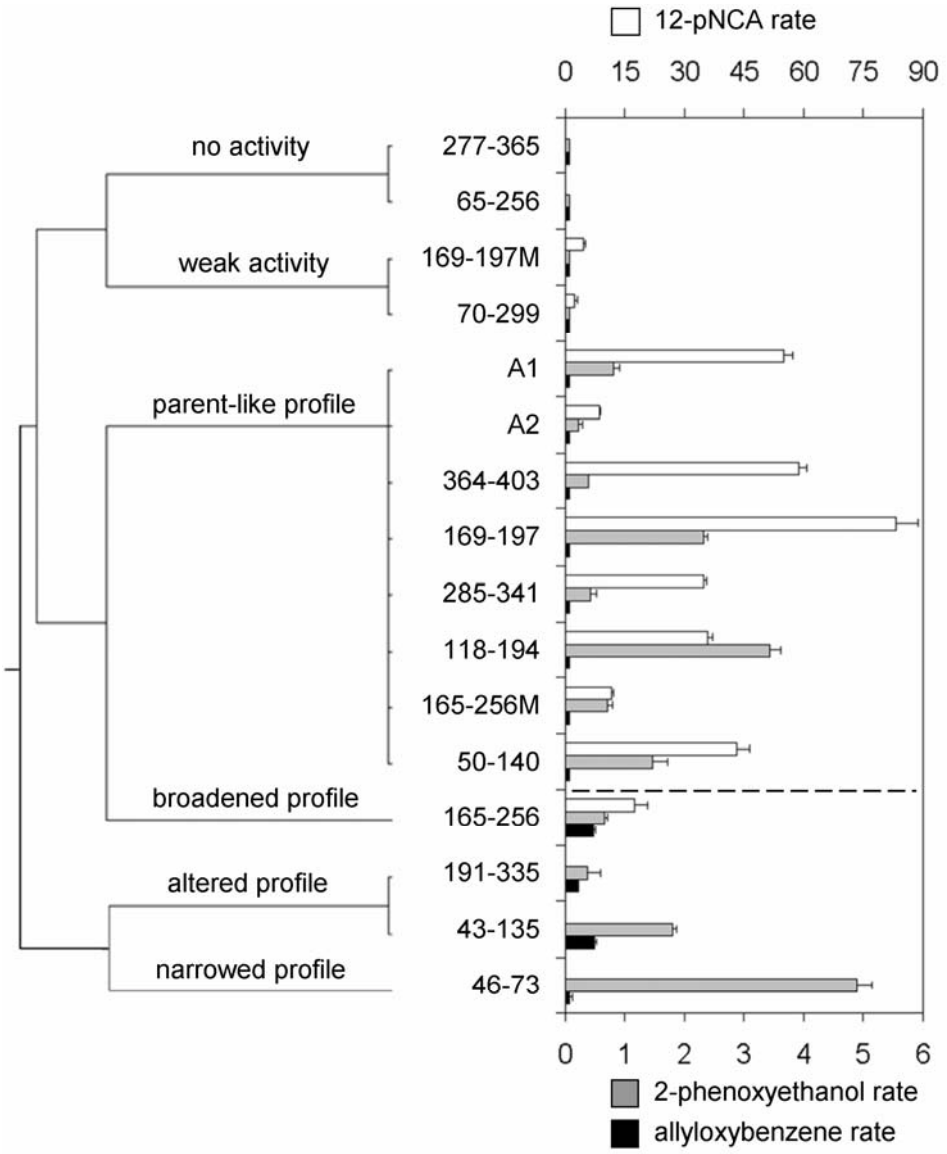

Figure 4.4. Substrate-activity profiles of A1, A2, and the folded chimeric heme domains. Chimeras were assayed for peroxygenase activity on 12-pNCA, 2-phenoxyethanol and allyloxybenzene, and hierarchical clustering analysis was used to group chimeras based on their functional properties. Three major categories are apparent: those with little or no detectable peroxygenase activity, those with parent-like profiles (activity on 12-pNCA and 2-phenoxyethanol), and those with altered profiles (below dashed line) relative to the parents (resulting from loss of activity on 12-pNCA and/or gain of activity on allyloxybenzene). The average amino acid Hamming distance (<*m*>) for the chimeras with parent-like profiles is 22, whereas the <*m*> of chimeras with altered profiles is 34. Initial rates are reported in nmol product/nmol P450/minute. Chimeras with no detectable activity are shown with values corresponding to the detection limits, which were 0.1, 0.06 and 0.08 for 12-pNCA, 2-phenoxyethanol and allyloxybenzene, respectively. Chimeras lacking detectable peaks in the CO difference spectra showed no activity on the substrates assayed.

C.4 Peroxidase activities of chimeric heme domains

P450s can reduce peroxide to water (and a proton) using a mechanism similar to that of peroxidases [44], although the intrinsic rate for P450s is orders of magnitude slower. It has been proposed that the earliest P450 function may have been as a peroxidase [2]. To investigate how recombination affects P450 peroxidase activity, we used the colorimetric substrate 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonate) (ABTS) to monitor this reaction [45]. A1 and A2 both show low, but detectable peroxidase activity (Table 4.1). Chimeras 50-140, 118-194, 165-256 and 169-197 have significantly higher peroxidase activities; 169-197 is approximately 40-fold more active than the most active parent. Mirror chimeras 169-197M and 165-256M do not show similarly enhanced levels of peroxidase activity. Three folded chimeras, 191-335, 65-256, and 277-335, showed no detectable peroxidase activity.

C.5 Thermostabilities of chimeric heme domains

Thermostability was assayed by monitoring the loss of the Soret band at increasing temperatures. Chimera melting temperatures ranged from 36°C to 55°C (Table 4.1), with none more stable than A1 (T_m = 55°C). More than half of the folded chimeric heme domains were more thermostable than A2 (T_m = 44°C); the rest were less stable. We found that chimeras less thermostable than the parents exhibited a wide range of *E* values, 12 to 27, and that stability does not correlate with calculated disruption, at least in this small population. However, thermostability may be important for retention of

catalytic activity: the two chimeras that lacked peroxygenase activity were also the least thermostable.

C.6 Folding and monooxygenase activities of chimera-reductase fusion proteins

We fused five of the functional chimeras (43-135, 46-73, 118-184, 165-256, and 169-197) to the N-terminus of the CYP102A1 reductase domain in order to investigate how the chimeric heme domains behave in the context of a P450 holoenzyme. CO difference spectra of the chimera-reductase fusion proteins (CRFPs) were used to monitor folding, and their activities on 12-pNCA, 2-phenoxyethanol and allyloxybenzene were measured in the presence of dioxygen and NADPH. All five CRFPs displayed a Soret band characteristic of a folded heme domain, and four of the five exhibited detectable activity on one or more substrates (Table 4.2). Monooxygenase activities of the fusion proteins were comparable to the peroxygenase activities of the respective heme domains for 12-pNCA and 2-phenoxyethanol.The fusion protein monooxygenases were roughly an order of magnitude more active towards allyloxybenzene than were the heme-domain peroxygenases.

Table 4.2. Monooxygenase activities of holoenzymes CYP102A1 (with F87A) and CYP102A2 (F88A) and chimera-reductase fusion proteins (CRFPs) on three substrates.

Protein[a]	12-pNCA[b]	2-phenoxyethanol[c]	allyloxybenzene[c]
CYP102A1	90.9±10.2	1.8±.3	12.0±0.8
CYP102A2	4.7±0.7	N.D.	N.D.
43-135-CRFP	N.D.	0.6±0.1	5.8±0.5
46-73-CRFP	N.D.	3.9±0.1	54.2±3.7
118-194-CRFP	87.6±4.1	0.8±0.1	6.6±0.9
165-256-CRFP	N.D.	N.D.	N.D.
169-197-CRFP	67.9±6.1	1.2±0.1	10.3±1.4

[a] CYP102A1 has the F87A substitution, CYP102A2 has F88A. Chimeric heme domains were fused to the N-terminus of the CYP102A1 reductase domain.

[b] Reported in nmol product/nmol P450/minute. Activity less than 0.1 was not detectable (N.D.).

[c] Reported in nmol product/nmol P450/minute. Activity less than 0.2 was not detectable (N.D.).

Overall, recombination appears to affect the function of the heme domains and the reconstituted holoenzymes in similar ways. For example, the specificities of heme-domain chimeras 118-194 and 169-197 are similar to A1 and A2 (Figure 4.4); their holoenzyme counterparts (118-194-CRFP and 169-197-CRFP) are also similar to the full-length parent with which they retain the most sequence similarity, CYP102A1 (Table 4.2). Furthermore, the chimeric heme domains 46-73 and 43-135, which showed altered peroxygenase activity profiles relative to the parents, also exhibited different oxygenase specificity (no activity toward 12-pNCA) when assembled as the holoenzyme. 165-256-CRFP, on the other hand, exhibited no detectable oxygenase activity on any of the test substrates, unlike 165-256, which as a peroxygenase hydroxylated all three.

The lower activity of CYP102A2 (having the F88A substitution) relative to CYP102A1 (with F87A) is more apparent as a monooxygenase than as a heme-domain peroxygenase. CYP102A1 was active on all three substrates tested, whereas the activity of CYP102A2 was measurable only on 12-pNCA, where it was 50 times less active than CYP102A1. If CYP102A2 has a substrate specificity similar to its A1 homolog, then the activities on allyloxybenzene and 2-phenoxyethanol would be below the detection limit of the assay. Although general features were similar, the CRFPs differed from their respective heme-domain chimeras in the details of the activities and specificities.

D. Discussion

D.1 Activities and specificities of recombined P450s

While a chimeric protein often equals the sum of its parts [46], it is also possible for a chimera to exceed its parents and find amino acid combinations that allow new properties to emerge [47-53]. Creating these beneficial amino acid combinations from different parental sequences that are "pre-screened" by nature is one goal of protein engineering by recombination [48]. We find that recombination is an effective way to alter the function of bacterial cytochrome P450s since more than half the folded P450 chimeric heme domains surpassed the parents in peroxidase or peroxygenase activity. In addition, nearly half had altered substrate specificities relative to the parents (Figure 4.4). Recombination yielded enzymes with detectable activity on only one or two of the substrates analyzed as well as a broadly-specific enzyme that hydroxylates all three. One specific heme-domain chimera (46-73) displayed 6-fold higher peroxygenase activity with 2-phenoxyethanol

than the parent most active on that substrate. Three chimeric heme domains hydroxylated allyloxybenzene, an activity not detectable in either of the parent heme domains.

Figure 4.5 shows how the functional properties of chimeras correlate well with their similarity to a parent. A1 is the most thermostable protein in this set of P450s at 55 ºC while A2 is significantly lower at 44 ºC. As the sequence from A2 is introduced into A1, the thermostability gradually decreases and eventually falls below that of A2 (Figure 4.5A). This is consistent with the additive nature of mutations with respect to protein stability shown in previous studies [54,55]. It is unclear, however, whether those chimeras with lower stability stem from this additivity or are due to more complex interactions. For example, these less stable sequences could result from a stabilizing residue from one parent being changed to a destabilizing residue from another parent. This is in contrast to a stabilizing residue being replaced by a less stabilizing one such that the thermostability would never fall below that of the least thermostable parent. On the other hand, the replacement of an amino acid involved in an important interaction, for example, the disruption of a hydrogen bond or hydrophobic interaction, could be more destabilizing than the individual contributions combined.

Similar to thermostability, the initial rates measured on 12-pNCA also correlate well with similarity to parents. As the sequence changes from the most active to least active parent, A1 to A2, the initial rate decreases linearly (Figure 4.5B). However, unlike thermostability, the initial rate on 12-pNCA was improved to a level greater than either parent, as in the case of chimera 169-197.

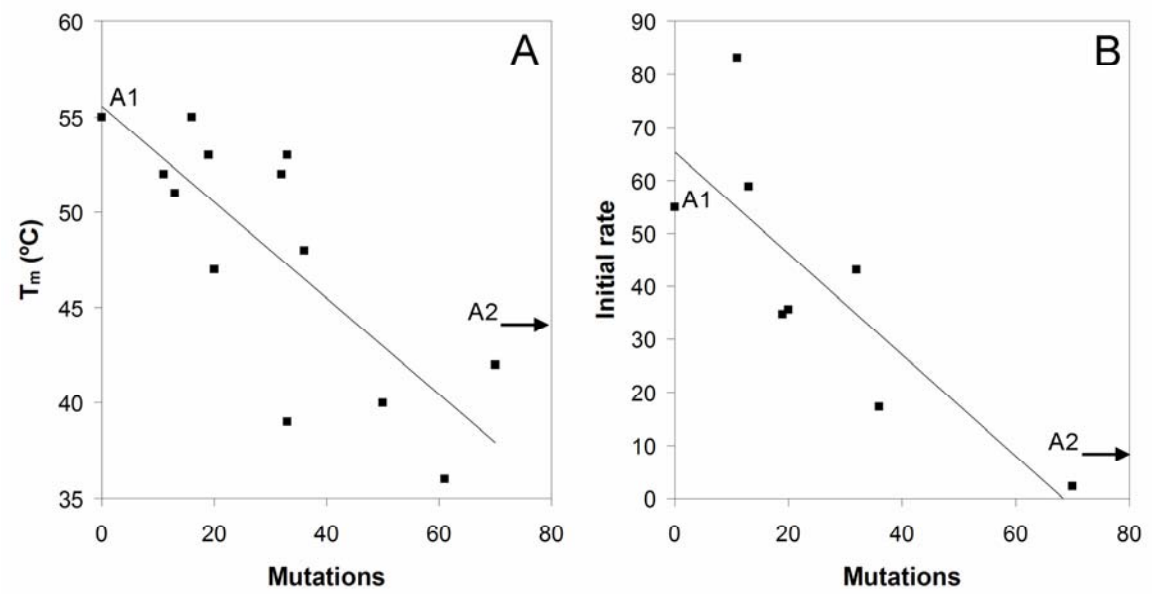

Figure 4.5. Activity and thermostability correlate linearly with sequence. A) As the sequence changes from the most to least thermostable parent (A1 to A2), the T_m linearly decreases with an R = 0.784. Unfolded and mirror chimeras not shown. B) As the sequence changes from the most to least active parent (A1 to A2) the initial rate on 12-pNCA decreases linearly with an R = 0.815. Inactive and mirror chimeras not shown. In both plots A1 is shown as zero mutations and the Tm or initial rate is show for A2 even though it does not lie on the graph with respect to mutation (A2 has 185 sequence changes relative to A1).

Using the heme domain's peroxygenase activity to monitor changes in substrate specificity allows us to explore the evolution of functional properties in this versatile enzyme upon recombination. Because the three-dimensional structure of the holoenzyme is not available, the SCHEMA algorithm can only be applied to the heme domain. The peroxygenase activity therefore provides a convenient way to screen chimeric enzyme libraries; it is also interesting in its own right for potential applications of this enzyme in chemical synthesis [38]. A chimeric heme domain can also be reconstituted into a holoenzyme by addition of the CYP102 reductase domain. Four of the five such CRFPs that were constructed in fact functioned as monooxygenases (Table 4.2). Furthermore, the

two active CRFPs whose heme domains showed altered substrate specificity relative to the parent heme domains were also different from the parent holoenzymes. The activity and specificity of a chimeric heme domain can be expected to change, however, when it is used in a CRFP as a monooxygenase, just as the parent enzymes differ in their peroxygenase and monooxygenase activities. Such differences were also reported in our previous study of A1 peroxygenase regioselectivity [26].

One of the functional heme-domain chimeras (165-256) generated an inactive CRFP. Upon recombination, the region of sequence that is derived from parent CYP102A2 introduces a glutamic acid residue in place of a lysine at position 241, located at the interface between the heme and reductase domains. We believe this impairs electron transfer by disrupting an electrostatic interaction between the reductase and heme domains of CYP102A1 [56]. None of the other CRFPs had this mutation. Because a chimeric heme domain may not in fact be compatible with a specific parental reductase, it is preferable to assay for the presence of function directly in the heme domain chimeras in order to assess the effects of recombination.

Our finding that recombination is effective in creating P450 chimeras with altered substrate specificities and novel activities is consistent with those reported for recombination of homologs in other enzyme families [48-52] and with mammalian P450s [53,57]. In most of these studies, closely-related proteins exhibiting distinct substrate specificities or activities were recombined. For example, Raillard and coworkers shuffled two triazine hydrolases, AtzA and TriA, which catalyze dechlorination and deamination

reactions, respectively, to obtain chimeras with enhanced activities and novel substrate specificities [50]. Our results demonstrate that recombination of functionally-similar enzymes can also yield functionally-diverse chimeras. In a previous study in which functionally-similar cephalosporinases were shuffled [48], the high levels of point mutation made it impossible to deconvolute the effects of recombination and point mutation. Here we show that residues that appear to be functionally neutral in the parent proteins are able to confer altered properties when recombined, provided the novel sequence folds properly. Although they may well be useful, additional point mutations were not required to achieve functional diversity.

D.2 Structural features of chimeric heme domains

The chimeric heme domains that retain the ability to fold and/or function did not result from swapping recognizable structural domains or distinct secondary structural elements. Instead, as illustrated in Figure 4.6A, the swapped fragments encompass non-trivial structural elements that would be difficult to identify without using an algorithm like SCHEMA, which takes into account sequence identity when calculating disruption. Using structural compactness alone to identify modules (for example, using the centripetal definition of Go [58,59]) does not identify most of our swapped elements as exchangeable. A great majority of the crossovers in the folded chimeras occur within these compact elements, rather than at their boundaries.

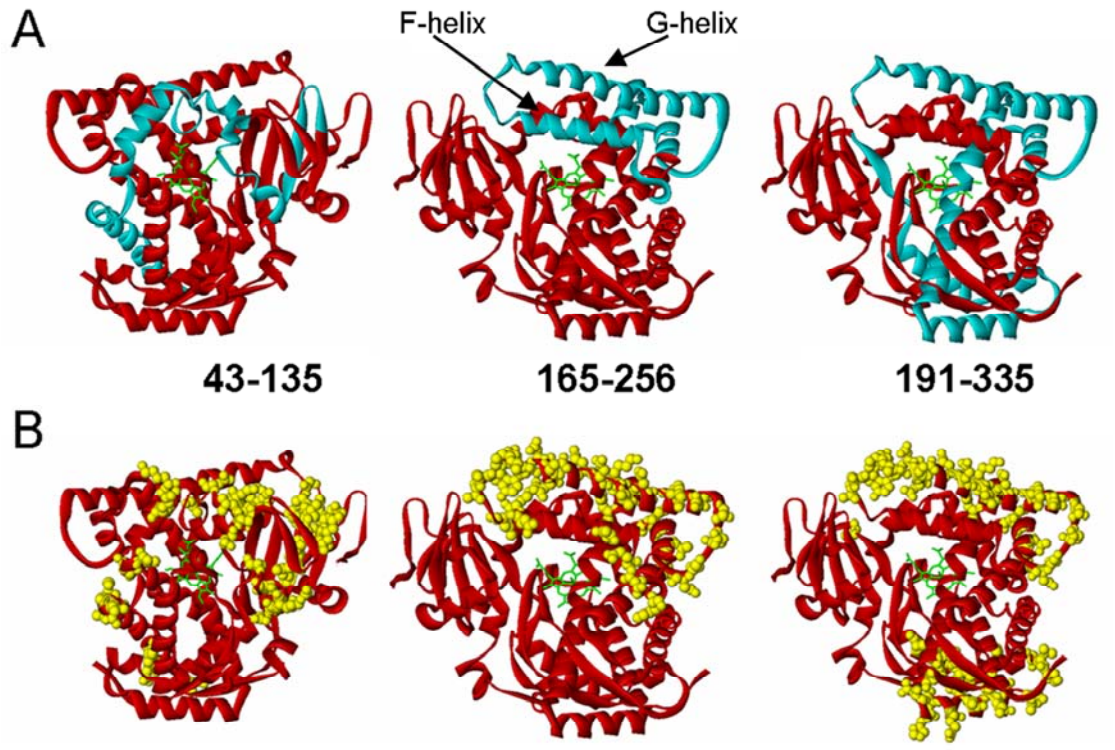

Figure 4.6. Structural models of chimeric cytochrome P450 heme domains. The numbers shown for each chimera represent the residues fromA1 that have been swapped for those from A2. A) Residues from A1 (red) and A2 (cyan) are mapped onto the structure of the CYP102A1 heme domain [30]. Arrows indicate the F and G-helices. B) Most of the effective mutations in the chimeras (shown in yellow) are located on the surface of the protein.

For proteins that share 63% amino acid sequence identity, most non-shared amino acids are on the protein surface. Not surprisingly, therefore, most of the sequence changes in the chimeras are found on the exterior of the protein (Figure 4.6B). Such mutations are less disruptive, on average, than changes in the core. A1 and A2 differ at six of the 21 residues postulated to contact a fatty-acid substrate [60]. However, it is unclear to what extent, if at all, sequence changes at these sites contribute to altered functional properties,

since no single change or combination of them is responsible for a particular activity. This suggests that mutations outside the active site effectively modulate substrate specificities and activities, as has been observed in previous random mutagenesis studies [61-63]. The 'novel' activity on allyloxybenzene and altered substrate-specificity profiles cannot be attributed to any specific residue alterations since chimeras exhibiting similar changes in activity arose by swapping distinct polypeptides in different places in the enzyme (Figure 4.4). Clearly, there are multiple ways to evolve functionally similar enzymes through recombination of homologous proteins at structurally-related residues.

We nonetheless point out one structural anecdote. The two P450 heme-domain chimeras with the highest peroxidase and peroxygenase rates (169-197) and broadest substrate specificity (165-256) have swapped a region of amino acids comprising the F helix. It has been shown that the F and G helices (Figure 4.6A) move approximately 6 Å upon substrate binding [32-34], and mutations affecting catalytic activity have been observed there in other protein engineering studies [63]. The new, favorable combination of the F helix from A2 and the G helix from A1 in the heme-domain chimera 169-197 and the complete substitution of the F and G helices in A1 with that from A2 in the heme-domain chimera 165-256 indicate a key role of this region in determining P450 catalytic properties.

D.3 Structure-guided design of chimeric enzyme libraries

Libraries generated by recombination of homologous proteins are rich in folded proteins if the parent proteins are highly similar [51,64] or if appropriate structural information is

incorporated in the library design [20,21]. It is not known, however, how functional diversity depends on the level of sequence diversity in such libraries, and whether recombination of less-similar sequences provides any advantage in the search for improved or novel functions. We hope this study begins to address this question. Figure 4.1 shows how chimera function is related to calculated disruption and effective mutations. Among the folded chimeras, those with substrate activity profiles similar to the parents typically cluster together with lower average mutation ($\langle m \rangle = 22$) than those with altered profiles ($\langle m \rangle = 34$). Thus chimeras with higher levels of mutation, provided they fold, may be more likely to have altered properties, while those with lower levels of mutation tend to be more similar to the parents. Theoretical models predict that recombination facilitates fitness changes [65,66], but there is still ittle information on how recombination and mutation level relates to functional evolution.

The probability of retaining function in the P450 chimeras decreases as calculated disruption (E = total number of residue-residue contacts broken upon recombination) increases. P450 chimeras with as many as 50, 61, and 70 effective mutations were still able to properly incorporate a heme cofactor, particularly with chimeric sequences characterized by low calculated disruption (typically $E \leq 30$) (Table 4.1). We found very similar results in a recent study of more than 16,000 chimeric lactamases [21]. Thus we believe that E is a useful measure of the likelihood a chimeric protein will retain its structure.

Taking together the experimental results and disruption calculations, we have in effect 'calibrated' this cytochrome P450 pair with respect to recombination. For example, we can now predict that a large fraction of all possible double-crossover chimeras of A1 and A2 will fold properly, because most are characterized by values of E < 30-35 (Figure 4.1). Once a particular set of crossover positions has been selected, however, only a limited number of chimeric sequences can be made (for two parents, this is $2^3 = 8$ sequences, including the parental ones). In generating large libraries that incorporate a larger number of crossovers, reducing disruption becomes an important design criterion. Figure 4.7 shows an *in silico* analysis of 5,000 different libraries in which 10 crossovers were allowed between A1 and A2, with the crossover positions chosen at random. Each library contains $2^{11} = 2{,}048$ different chimeric sequences. For each library, we calculated i) the fraction that is predicted to fold (using F_{30} = fraction of sequences with $E \leq 30$) and ii) the average level of effective mutation in these folded chimeras ($<m>_{30}$). Using this method of random enumeration of libraries, we find that the choice of crossover points can dramatically affect these values and, in all likelihood, the distribution and nature of functional proteins in the library. A library of A1 and A2 chimeras may contain as little as 9% that fold properly; on average 42% will fold. In contrast, by constructing libraries *in silico* and using SCHEMA to guide the choice of crossover points, the percentage folded can in principle be as high as 75%, and a very high effective level of mutation can be retained (with >50 mutations on average per folded sequence). We propose that this latter library will contain more folded chimeras and be richer in novel functional proteins than libraries made at random. We expect even greater benefits of using SCHEMA when

recombining more parents or parental sequences with less sequence identity, provided their structures are highly similar overall.

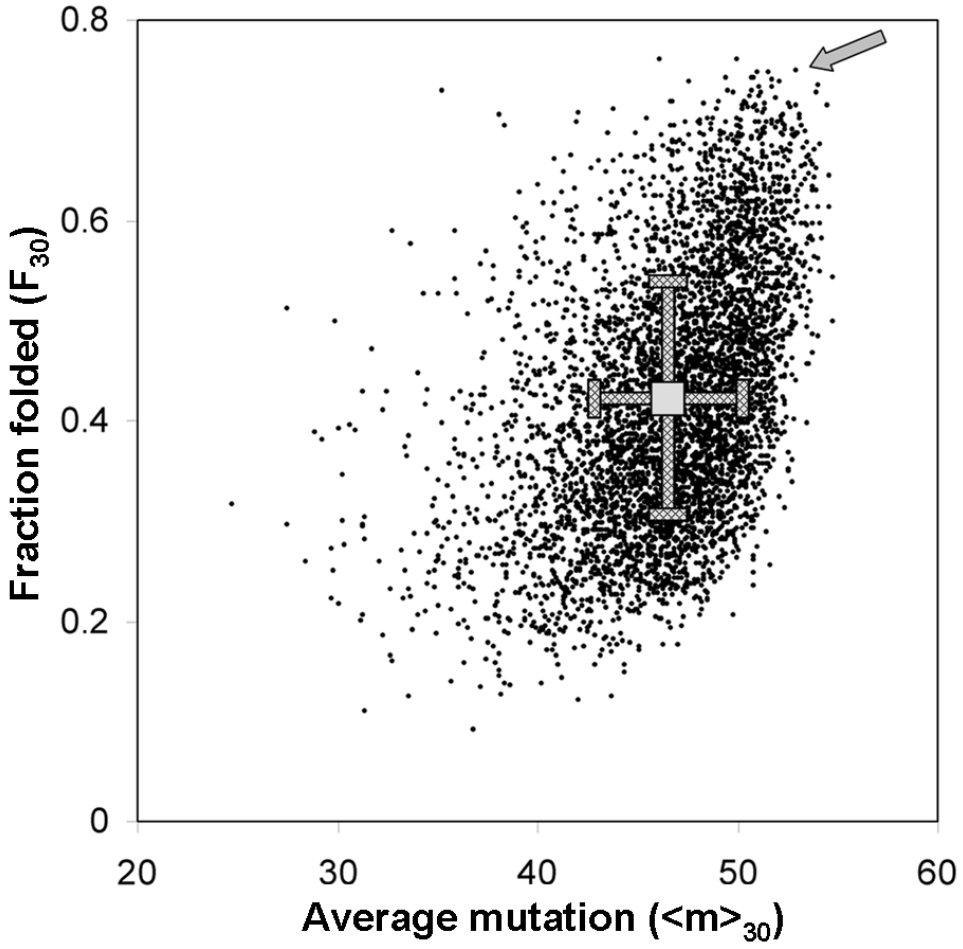

Figure 4.7. Theoretical library analysis. Five thousand libraries were generated *in silico,* in which 10 randomly-selected crossovers were allowed between A1 and A2. The fraction of each library that is predicted to be folded (F_{30}, those with $E \leq 30$) is plotted against the average level of effective mutation for the fraction that should fold ($<m>_{30}$). The average $<m>_{30}$ and F_{30} for the population is shown as a square with one standard deviation. The arrow points to a library with a F_{30} of 75% with $<m>_{30}$ greater than 50.

E. Conclusions

In nature, cytochromes P450 often protect organisms from toxic compounds [3,5] or help them adapt to new food sources [4,5]. Thus a scaffold that allows for rapid functional evolution could be beneficial. Such a scaffold is also desirable for protein engineering. Recent engineering efforts have demonstrated that P450s can acquire new or improved activities by point mutation [38,43,63]; here we show that recombination of homologous sequences should be able to generate significant functional diversity as well. We propose that SCHEMA can help identify appropriate crossover locations for large, combinatorial libraries [21], which can be generated using targeted recombination methods [35,67]. With appropriate high-throughput screening, we may then be able to discover new P450s with properties that nature has not yet needed or explored.

F. Experimental procedures

F.1 Materials

Enzymes for DNA manipulations were obtained from New England Biolabs (Beverly, MA) and Stratagene (La Jolla, CA). Synthetic oligonucleotides were obtained from Invitrogen (Carlsbad, CA) or the California Institute of Technology oligonucleotide facility. DNA purification kits were from Zymo Research (Orange, CA) and Qiagen (Valencia, CA). Other reagents and chemicals were from Fisher Scientific (Pittsburgh, PA), Becton Dickinson (Franklin Lake, NJ) and Sigma Chemical Co (St. Louis, MO).

F.2 Calculations

The number of contacts broken by recombination (*E*) was calculated as described using coordinates from the substrate-bound (1JPZ) and substrate-free structures of CYP102A1 (2HPD) [20,30,31]. Hydrogens, backbone nitrogens, backbone oxygens, and heme atoms were not included in the calculation. The sequences of A1 and A2 were aligned using ClustalW (Appendix A) [68], revealing the existence of a one-amino-acid insertion relative to A1, between Q229 and S230. This insertion was ignored in the calculations. A1 residues G227 and E228 were also ignored because they are unresolved in the substrate-bound structure (1JPZ). For calculation of *E* and mutation for all double crossover chimeras, we applied a minimum insert size of 10 residues. The error values reported for *E* and mutation represent one standard deviation.

The recombination libraries analyzed contained 10 randomly-chosen crossovers, each separated by a minimum of 10 residues. Using the substrate-bound structure of CYP102A1 (1JPZ) [30], we calculated the total number of contacts disrupted (*E*) and effective level of mutation (m) for all 2^{11} (2,048) chimeras in 5,000 libraries. We also computed the fraction of chimeras in each library with $E \leq 30$, denoted F_{30}, and the average effective level of mutation $\langle m \rangle_{30}$ in this low-disruption fraction.

F.3 Construction of chimeras

Selected chimeras were constructed using SOEing methods, as described previously [35]. Heme-domain chimeras contained residues 1-463 from CYP102A1 or the corresponding residues in CYP102A2 (1-466). Two primers consisting of a 5' sequence from one parent

(A) and a 3' sequence from the other (B) that encompass the crossover site were used to amplify the sequence to be inserted (B) with 25-30 bp overhangs from the other sequence (A). The PCR protocol was to heat the plasmids and primers at 95 ºC followed by 22 cycles of 95 ºC for 1 minute, 48 ºC for 1 minute, and 72 ºC for 2 minutes with a final extension at 72ºC for 10 minutes. These products acted as primers in a further PCR reaction along with forward and reverse primers external to the ends of the gene containing BamHI and EcoRI restriction sites, respectively, for cloning into the pCWori vector. The PCR protocol was 95 ºC for 1 minute, 46 ºC for 1 minute, and 72 ºC for 2 minutes for 22 cycles with a final extension at 72 ºC for 10 minutes. These two products were assembled in a two-step PCR reaction: 95 ºC for 1 minute followed by 14 cycles of 95 ºC for 1 minute, 46 ºC for 1 minute, and 72 ºC for 2 minutes. External primers were added followed by PCR 95ºC for 1 minute followed by 14 cycles of 95 ºC for 1 minute, 46 ºC for 1 minute, and 72 ºC for 2 minutes with a final extension of 72 ºC for 10 minutes. All PCR products were gel-purified using the Zymoclean-5 column from Zymo Research. High-fidelity Pfu Turbo and Pfu Ultra polymerases (Stratagene) were used for PCR. Final products were digested with BamHI and EcoRI and cloned into pCWori. Plasmids were transformed into a catalase-deficient strain of *E. coli*. Chimeras were sequenced at Laragen Inc. (Los Angeles, CA) and the California Institute of Technology sequencing facility (Pasadena, CA) to confirm the sequences, with the absence of point mutations. pCWori expression vectors, encoding heme-domain chimeras fused to the N-terminus of the CYP102A1 reductase domain (CRFP = chimera reductase fusion proteins), were constructed for five of the chimeras (43-135, 46-73, 118-184, 165-256, and 169-197) using a method similar to that described above.

F.4 Protein expression

Chimeric heme domains and CRFPs were expressed in catalase-deficient *E. coli* strain SN0037 [37] using the isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible pCWori vector [36]. Cultures grown in terrific broth (TB) were shaken at 250 rpm and 30 °C until they reached an OD_{600} of approximately 0.8. They were induced with 0.6 mM IPTG, supplemented with 25 μg/ml thiamine and 0.5 mM δ-aminolevulinic acid, and grown for 20 hours at 180 rpm and 25 °C. This procedure yields approximately 100 mg/L of P450 protein for A1 and A2. Cultures were pelleted at 5,500 g for 15 min, resuspended in 50 mM Tris (pH 8.2), and lysed by sonification. Centrifugation was used to clear the supernatant, which was used for further assays.

F.5 Folding assay

Carbon monoxide reduced difference spectroscopy was performed as reported [69]. Cell extracts were diluted into 800 μL of 100 mM Tris buffer (pH 8.2) in a cuvette at room temperature. A few mg of sodium hydrosulfite on the tip of a spatula were added, and a blank spectrum was determined from 400 to 500 nm. Carbon monoxide was bubbled in for 20 seconds at a rate of approximately one bubble per second. Two minutes were allowed to pass before a spectrum was taken. Spectra were determined at multiple times to ensure complete carbon monoxide binding and maximum absorbance. There were no increases beyond 5 minutes of incubation with carbon monoxide for any of the chimeras. P450 enzyme concentrations were quantified for further assays using an extinction coefficient of 91 $mol^{-1}cm^{-1}$ for the absorbance difference between 448 nm and 490 nm.

F.6 Peroxygenase activity

First-order rates of *p*-nitrophenolate accumulation were determined using 1 μM enzyme, 20 mM H_2O_2, 250 μM 12-pNCA and 0.5% dimethyl sulfoxide (DMSO) in 100 mM Tris-HCl (pH 8.2) at room temperature. Enzyme, substrate, buffer and DMSO were combined in a cuvette and zeroed at 410 nm. Reaction mixtures were allowed to incubate for 4 minutes and initiated by the addition of H_2O_2 to a final concentration of 20 mM. Initial rates were determined by monitoring the accumulation of *p*-nitrophenolate at 410 nm, and data from the first six seconds were used to determine initial rates. If no activity was observed at 20 mM H_2O_2, a second trial at 100 mM was done. No chimera inactive at 20 mM H_2O_2 showed activity at the higher concentration. The extinction coefficient of *p*-nitrophenolate is 13,200 $M^{-1}cm^{-1}$ [42]. All rates reported represent the average of three independent experiments with error bars corresponding to one standard deviation.

Catalytic activities on 2-phenoxyethanol and allyloxybenzene were determined using 2 μM enzyme, 20 mM H_20_2, 1% DMSO, and 1% acetone in 100 mM N-[2-hydroxyethyl]piperazine-N'-[3-propanesulfonic acid] (EPPS) (pH 8.2) at room temperature. Substrate concentrations for 2-phenoxyethanol (100 mM) and allyloxybenzene (50 mM) maintained saturation. Total reaction volumes were 400 μl for heme domain chimeras and were initiated by the addition of H_2O_2 and monitored for up to 90 minutes. Aliquots of the reaction were removed at time points within the linear region of the time course and mixed with an equal volume of a quench solution containing 4 M urea and 100 mM NaOH. 15 μl per 100 μl of 0.6% 4-AAP was added,

followed by mixing and addition of 15 µl per 100 µl of 0.6% potassium persulfate. Color was allowed to develop for 10 minutes before absorbance was read at 500 nm. The major products were determined by GC/MS to be the hemiacetal which decomposes to phenol. The extinction coefficient for the 4-AAP/phenol complex was determined to be 4,800 $M^{-1}cm^{-1}$.

F.7 Monooxygenase activity

CRFP monooxygenase activities were determined under identical conditions as the peroxygenase reactions, except H_2O_2 was replaced with 500 µM NADPH in all reactions.

F.8 Clustering analysis

Chimeras that retained the ability to fold were analyzed using hierarchical clustering analysis as performed by the Spotfire® software package (Spotfire, Somerville, MA). Chimeras were clustered based on their substrate-specificity profiles, i.e. whether or not they possessed measurable peroxygenase activity towards the substrates 12-pNCA, 2-phenoxyethanol, and allyloxybenzene. Therefore, activities were normalized to the presence of activity (1) or the lack of activity (0). UPGMA (Unweighted Pair Group Method using Arithmetic Averages) clustering was performed using Euclidean distance as a similarity metric and the average value as an ordering function.

F.9 Peroxidase activity

Peroxidase activities were measured by monitoring the accumulation of the radical cation of ABTS at 414 nm [45]. Enzyme (1 μM) was mixed with 10 mM ABTS in 200 mM phosphate buffer (pH 5.0) in a cuvette at room temperature. Samples were zeroed and reactions were initiated with the addition of H_2O_2 to a concentration of 20 mM. The absorbance at 414 nm was monitored for 5 minutes. Rates were determined from the initial slope of the time course (typically the first 30 seconds). An extinction coefficient of 36,000 $mol^{-1}cm^{-1}$ for ABTS was used.

F.10 Thermostability

Cell extracts were heated in a thermocycler for 10 minutes at various temperatures, followed by cooling to 4°C. Extracts were centrifuged for 5 minutes at 3500 x g to remove any precipitates. Carbon monoxide reduced difference spectroscopy was used to quantitate the amount of P450. The reduction of the carbon monoxide peak was monitored over a range of temperatures.

G. References

1. Ortiz de Montellano PR (1995) Cytochrome P450: Structure, Mechanism, and Biochemistry. New York: Plenum Press.
2. Gotoh O (1993) Evolution and Differentiation of P-450 Genes. In: Omura T, Ishimura Y, Fujii-Kuriyama Y, editors. Cytochrome P450. 2nd ed. Tokyo: Kodansha. pp. 255-272.
3. Gonzalez FJ, Nebert DW (1990) Evolution of the P450 gene superfamily: animal-plant 'warfare', molecular drive and human genetic differences in drug oxidation. Trends Genet. 6: 182-186.
4. Mauersberger S, Schunck WH, Muller HH (1981) The induction of cytochrome P-450 in Lodderomyces elongisporus. Z. Allg. Mikrobiol. 21: 313-321.
5. Porter TD, Coon MJ (1991) Cytochrome P-450. Multiplicity of isoforms, substrates, and catalytic and regulatory mechanisms. J. Biol. Chem. 266: 13469-13472.
6. Sono M, Roach MP, Coulter ED, Dawson JH (1996) Heme-Containing Oxygenases. Chem. Rev. 96: 2841-2888.
7. Eichhorn GL, Marzilli LG (1988) Heme proteins. New York, NY: Elsevier Science Publishing Co., Inc.
8. Cirino PC, Arnold FH (2002) Exploring the diversity of heme enzymes through directed evolution. In: Brakmann S, Johnsson K, editors. Directed Molecular Evolution of Proteins. Germany: Wiley-VCH. pp. 215-243.
9. Stemmer WP (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc. Natl. Acad. Sci. USA 91: 10747-10751.
10. Coco WM, Levinson WE, Crist MJ, Hektor HJ, Darzins A, et al. (2001) DNA shuffling method for generating highly recombined genes and evolved enzymes. Nat. Biotechnol. 19: 354-359.
11. Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. Nat. Biotechnol. 16: 258-261.
12. Volkov AA, Shao Z, Arnold FH (1999) Recombination and chimeragenesis by in vitro heteroduplex formation and in vivo repair. Nucleic Acids Res. 27: e18.
13. Kikuchi M, Ohnishi K, Harayama S (2000) An effective family shuffling method using single-stranded DNA. Gene 243: 133-137.
14. Bittker JA, Le BV, Liu JM, Liu DR (2004) Directed evolution of protein enzymes using nonhomologous random recombination. Proc Natl Acad Sci U S A 101: 7011-7016.
15. Ostermeier M, Nixon AE, Benkovic SJ (1999) Incremental truncation as a strategy in the engineering of novel biocatalysts. Bioorg Med Chem 7: 2139-2144.
16. Ostermeier M, Shim JH, Benkovic SJ (1999) A combinatorial approach to hybrid enzymes independent of DNA homology. Nat. Biotechnol. 17: 1205-1209.
17. Lutz S, Ostermeier M, Moore GL, Maranas CD, Benkovic SJ (2001) Creating multiple-crossover DNA libraries independent of sequence identity. Proc. Natl. Acad. Sci. USA 98: 11248-11253.
18. Kawarasaki Y, Griswold KE, Stevenson JD, Selzer T, Benkovic SJ, et al. (2003) Enhanced crossover SCRATCHY: construction and high-throughput screening of

a combinatorial library containing multiple non-homologous crossovers. Nucleic Acids Res 31: e126.
19. Sieber V, Martinez CA, Arnold FH (2001) Libraries of hybrid proteins from distantly related sequences. Nat. Biotechnol. 19: 456-460.
20. Voigt CA, Martinez C, Wang ZG, Mayo SL, Arnold FH (2002) Protein building blocks preserved by recombination. Nat. Struct. Biol. 9: 553-558.
21. Meyer MM, Silberg JJ, Voigt CA, Endelman JB, Mayo SL, et al. (2003) Library analysis of SCHEMA-guided protein recombination. Protein Sci. 12: 1686-1693.
22. Narhi LO, Kim BH, Stevenson PM, Fulco AJ (1983) Partial characterization of a barbiturate-induced cytochrome P-450-dependent fatty acid monooxygenase from Bacillus megaterium. Biochem. Biophys. Res. Commun. 116: 851-858.
23. Gustafsson MC, Roitel O, Marshall KR, Noble MA, Chapman SK, et al. (2004) Expression, purification, and characterization of bacillus subtilis cytochromes P450 CYP102A2 and CYP102A3: flavocytochrome homologues of P450 BM3 from Bacillus megaterium. Biochemistry 43: 5474-5487.
24. Munro AW, Leys DG, McLean KJ, Marshall KR, Ost TW, et al. (2002) P450 BM3: the very model of a modern flavocytochrome. Trends Biochem. Sci. 27: 250-257.
25. Li QS, Ogawa J, Schmid RD, Shimizu S (2001) Residue size at position 87 of cytochrome P450 BM-3 determines its stereoselectivity in propylbenzene and 3-chlorostyrene oxidation. FEBS Lett. 508: 249-252.
26. Cirino PC, Arnold FH (2002) Regioselectivity and activity of cytochrome P450 BM-3 and mutant F87A in reactions driven by hydrogen peroxide. Advanced Synthesis & Catalysis 344: 932-937.
27. Joern JM (2003) DNA shuffling. Methods Mol. Biol. 231: 85-89.
28. Abecassis V, Pompon D, Truan G (2000) High efficiency family shuffling based on multi-step PCR and in vivo DNA recombination in yeast: statistical and functional analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2. Nucleic. Acids Res. 28: E88.
29. Coco WM (2003) RACHITT. Methods Mol. Biol. 231: 111-127.
30. Haines DC, Tomchick DR, Machius M, Peterson JA (2001) Pivotal role of water in the mechanism of P450BM-3. Biochemistry 40: 13456-13465.
31. Ravichandran KG, Boddupalli SS, Hasermann CA, Peterson JA, Deisenhofer J (1993) Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. Science 261: 731-736.
32. Paulsen MD, Ornstein RL (1995) Dramatic differences in the motions of the mouth of open and closed cytochrome P450BM-3 by molecular dynamics simulations. Proteins 21: 237-243.
33. Modi S, Sutcliffe MJ, Primrose WU, Lian LY, Roberts GC (1996) The catalytic mechanism of cytochrome P450 BM3 involves a 6 A movement of the bound substrate on reduction. Nat. Struct. Biol. 3: 414-417.
34. Li H, Poulos TL (1997) The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. Nat. Struct. Biol. 4: 140-146.
35. Horton RM, Hunt HD, Ho SN, Pullen JK, Pease LR (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. Gene 77: 61-68.

36. Barnes HJ, Arlotto MP, Waterman MR (1991) Expression and enzymatic activity of recombinant cytochrome P450 17 alpha-hydroxylase in Escherichia coli. Proc. Natl. Acad. Sci. USA 88: 5597-5601.
37. Nakagawa S, Ishino S, Teshiba S (1996) Construction of catalase deficient Escherichia coli strains for the production of uricase. Biosci. Biotechnol. Biochem. 60: 415-420.
38. Cirino PC, Arnold FH (2003) A self-sufficient peroxide-driven hydroxylation biocatalyst. In press Angew. Chemie Intl. Edn.
39. Omura T, Sato R (1964) The carbon monoxide-binding pigment of liver microsomes. J. Biol. Chem. 239: 2370-2378.
40. Martinis SA, Blanke SR, Hager LP, Sligar SG, Hoa GH, et al. (1996) Probing the heme iron coordination structure of pressure-induced cytochrome P420cam. Biochemistry 35: 14530-14536.
41. Wells AV, Li P, Champion PM, Martinis SA, Sligar SG (1992) Resonance Raman investigations of Escherichia coli-expressed Pseudomonas putida cytochrome P450 and P420. Biochemistry 31: 4384-4393.
42. Schwaneberg U, Schmidt-Dannert C, Schmitt J, Schmid RD (1999) A continuous spectrophotometric assay for P450 BM-3, a fatty acid hydroxylating enzyme, and its mutant F87A. Anal. Biochem. 269: 359-366.
43. Li QS, Schwaneberg U, Fischer M, Schmitt J, Pleiss J, et al. (2001) Rational evolution of a medium chain-specific cytochrome P-450 BM-3 variant. Biochim. Biophys. Acta 1545: 114-121.
44. Mansuy D (1998) The great diversity of reactions catalyzed by cytochromes P450. Comp. Biochem. Physiol. C. Pharmacol. Toxicol. Endocrinol. 121: 5-14.
45. Childs RE, Bardsley WG (1975) The steady-state kinetics of peroxidase with 2,2'-azino-di-(3-ethyl-benzthiazoline-6-sulphonic acid) as chromogen. Biochem. J. 145: 93-103.
46. Beguin P (1999) Hybrid enzymes. Curr. Opin. Biotechnol. 10: 336-340.
47. Campbell RK, Bergert ER, Wang Y, Morris JC, Moyle WR (1997) Chimeric proteins can exceed the sum of their parts: implications for evolution and protein design. Nat. Biotechnol. 15: 439-443.
48. Crameri A, Raillard SA, Bermudez E, Stemmer WP (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. Nature 391: 288-291.
49. Kumamaru T, Suenaga H, Mitsuoka M, Watanabe T, Furukawa K (1998) Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase. Nat. Biotechnol. 16: 663-666.
50. Raillard S, Krebber A, Chen Y, Ness JE, Bermudez E, et al. (2001) Novel enzyme activities and functional plasticity revealed by recombining highly homologous enzymes. Chem. Biol. 8: 891-898.
51. Hansson LO, Bolton-Grob R, Massoud T, Mannervik B (1999) Evolution of differential substrate specificities in Mu class glutathione transferases probed by DNA shuffling. J. Mol. Biol. 287: 265-276.
52. Christians FC, Scapozza L, Crameri A, Folkers G, Stemmer WP (1999) Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling. Nat. Biotechnol. 17: 259-264.

53. Ramarao MK, Straub P, Kemper B (1995) Identification by in vitro mutagenesis of the interaction of two segments of C2MstC1, a chimera of cytochromes P450 2C2 and P450 2C1. J. Biol. Chem. 270: 1873-1880.
54. Chen J, Lu Z, Sakon J, Stites WE (2000) Increasing the thermostability of staphylococcal nuclease: implications for the origin of protein thermostability. J. Mol. Biol. 303: 125-130.
55. Matthews BW, Nicholson H, Becktel WJ (1987) Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. Proc. Natl. Acad. Sci. U.S.A. 84: 6663-6667.
56. Sevrioukaova IF, Li H, Zhang H, Peterson JA, Poulos TL (1999) Structure of a cytochrome P450-redox partner electron-transfer complex. Proc. Natl. Acad. Sci. USA 96: 1863-1868.
57. Pompon D, Nicolas A (1989) Protein engineering by cDNA recombination in yeasts: shuffling of mammalian cytochrome P-450 functions. Gene 83: 15-24.
58. Roy SW, Nosaka M, de Souza SJ, Gilbert W (1999) Centripetal modules and ancient introns. Gene 238: 85-91.
59. Go M, Nosaka M (1987) Protein architecture and the origin of introns. Cold Spring Harb. Symp. Quant. Biol. 52: 915-924.
60. Li H, Poulos TL (1999) Fatty acid metabolism, conformational change, and electron transfer in cytochrome P-450(BM-3). Biochim Biophys Acta 1441: 141-149.
61. Zha DX, Wilensek S, Hermes M, Jaeger KE, Reetz MT (2001) Complete reversal of enantioselectivity of an enzyme-catalyzed reaction by directed evolution. Chem. Commun. 24: 2664-2665.
62. Zhang JH, Dawes G, Stemmer WP (1997) Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. Proc. Natl. Acad. Sci. USA 94: 4504-4509.
63. Glieder A, Farinas ET, Arnold FH (2002) Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. Nat. Biotechnol. 20: 1135-1139.
64. Moore JC, Jin HM, Kuchner O, Arnold FH (1997) Strategies for the in vitro evolution of protein function: enzyme evolution by random recombination of improved sequences. J. Mol. Biol. 272: 336-347.
65. Cui Y, Wong WH, Bornberg-Bauer E, Chan HS (2002) Recombinatoric exploration of novel folded structures: a heteropolymer-based model of protein evolutionary landscapes. Proc. Natl. Acad. Sci. USA 99: 809-814.
66. Bogarad LD, Deem MW (1999) A hierarchical approach to protein molecular evolution. Proc. Natl. Acad. Sci. USA 96: 2591-2595.
67. Hiraga K, Arnold FH (2003) General method for sequence-independent site-directed chimeragenesis. J. Mol. Biol. 330: 287-296.
68. Thompson JD, Higgins DG, Gibson TJ (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: 4673-4680.
69. Schenkman JB, Jansson I (1998) Spectral Analyses of Cytochromes P450. In: Phillips IR, Shephard EA, editors. Cytochrome P450 Protocols. Totowa, New Jersey: Human Press, Inc. pp. 25-33.

Chapter 5

Structure-guided recombination creates an artificial family of cytochromes P450

Material from this chapter appears in: Otey, C.R., Landwehr, M., Endelman, J.B., Hiraga, K., Bloom, J.D., and Arnold F.H. (2006) **Structure-Guided Recombination Creates an Artificial Family of Cytochromes P450**. *PLos Biology* 4(5): e112. In addition, Kiowa Bauer provided laboratory assistance and this work was supported by NIH Grant R01 GM068664-01, ARO Contract DAAD19-03-D-0004, a NSEG Fellowship (to JBE), a JSPS Postdoctoral Fellowship (to KH), and an HHMI predoctoral fellowship (to JDB).

A. Abstract

Creating artificial protein families affords new opportunities to explore the determinants of structure and biological function free from many of the constraints of natural selection. We have created an artificial family comprising ~3,000 P450 heme proteins that correctly fold and incorporate a heme cofactor by recombining three cytochromes P450 at seven crossover locations chosen to minimize structural disruption. Members of this protein family differ from any known sequence at an average of 72 and by as many as 109 amino acids. Most (>73%) of the properly folded chimeric P450 heme proteins are catalytically active peroxygenases; some are more thermostable than the parent proteins. A multiple sequence alignment of 955 chimeras, including both folded and not, is a valuable resource for sequence-structure-function studies. Logistic regression analysis of the multiple sequence alignment identifies key structural contributions to cytochrome P450 heme incorporation and peroxygenase activity and suggests possible structural differences between parents CYP102A1 and CYP102A2.

B. Introduction

Our understanding of how protein sequence relates to structure and function is aided by comparisons of sequences related by evolution [1-3]. With only limited numbers of highly divergent sequences, however, such analyses are often uninformative. Furthermore, because the sequences have been culled by natural selection, relationships between sequence and physical or chemical properties not under direct selection are difficult or impossible to discern. We would like to create artificial protein families in order to probe the range of sequence and functional diversity that is compatible with a given structure, free from the constraint of having to function in the narrow context of the host organism. These artificial sequences would help us to identify connections to functions that may not be important biologically (e.g., high thermostability, new substrate specificity, or ability to fold into a particular structure, but not catalyze a particular reaction), but are critical for understanding the proteins themselves [4,5].

The products of millions of years of divergence and natural selection, protein families contain members that differ at large numbers of amino acids residues. Creating numerous diverse and folded sequences in the laboratory is challenging, due in part to the sparsity of proteins in sequence space. Among random sequences, estimates of the frequency of functional proteins range from 1 in 10^{11} [6] to as little as 1 in 10^{77} [7]. Randomly mutating a functional parent sequence improves the odds, but highly mutated sequences are still exceedingly unlikely to fold into recognizable proteins [8,9]. The methods by which novel proteins have been created, including selection from libraries of random [6] or patterned [10] sequences, evolution from existing sequences by iterative mutation or

recombination [11], and by structure-guided design [12] as well as computation-intensive protein design [13,14], either yield small numbers of characterized sequences or numerous sequences with low diversity (few sequence changes).

We are developing site-directed, homologous recombination guided by structure-based computation (SCHEMA) [15-17] to create libraries of protein sequences that are simultaneously highly mutated and have a high likelihood of folding into the parental structure. Mutations made by recombination of functional sequences are much more likely to be compatible with the particular protein fold than are random mutations [18]. SCHEMA calculations allow us to minimize the number of structural contacts that are disrupted when portions of the sequence are inherited from different parents, further increasing the probability that the chimeric proteins will fold. The validity of the SCHEMA disruption metric has been demonstrated in previous work [15-17]. SCHEMA, however, has not yet been used to design a library to maximize the number of sequences with low disruption and high mutation.

Here we report SCHEMA-guided recombination of three cytochromes P450 to create 6,561 chimeras, of which ~3,000 are properly folded P450 proteins. Cytochromes P450 comprise a superfamily of heme enzymes with myriad biological functions, including key roles in drug metabolism, breakdown of xenobiotics, and steroid and secondary metabolite biosynthesis [19]. More than 4,500 sequences of this ubiquitous enzyme are known [20]. Members of the artificial family of chimeric P450s reported here differ from any known protein by up to 109 amino acids, yet most retain significant catalytic activity.

Unlike natural protein families, this artificial family also includes sequences that do not fold or function. Inclusion of nonfunctional sequences enables us to apply powerful logistic regression tools [21] to the multiple sequence alignment (MSA) of the laboratory-generated proteins and determine which elements contribute to correct heme incorporation and retention of catalytic activity in the cytochrome P450 heme domain.

C. Results/Discussion

C.1 SCHEMA design and construction of a chimeric P450 library

We generated an artificial family of cytochromes P450 by recombining fragments of the genes encoding the heme-binding domains of three bacterial P450s, A1, A2, and A3, which share ~65% amino acid identity [22,23] (Figure 5.1). The parent proteins are 463–466 amino acids long and contain the single substitution F87A (A1) or F88A (A2 and A3), which increases the peroxygenase activities of these heme domains [24]. Calculations of the SCHEMA disruption that results when residue–residue contacts present in the parent structure are broken by recombination (see Materials and Methods) served to guide the placement of crossovers so as to maximize the number of highly mutated, folded proteins in the resulting library.

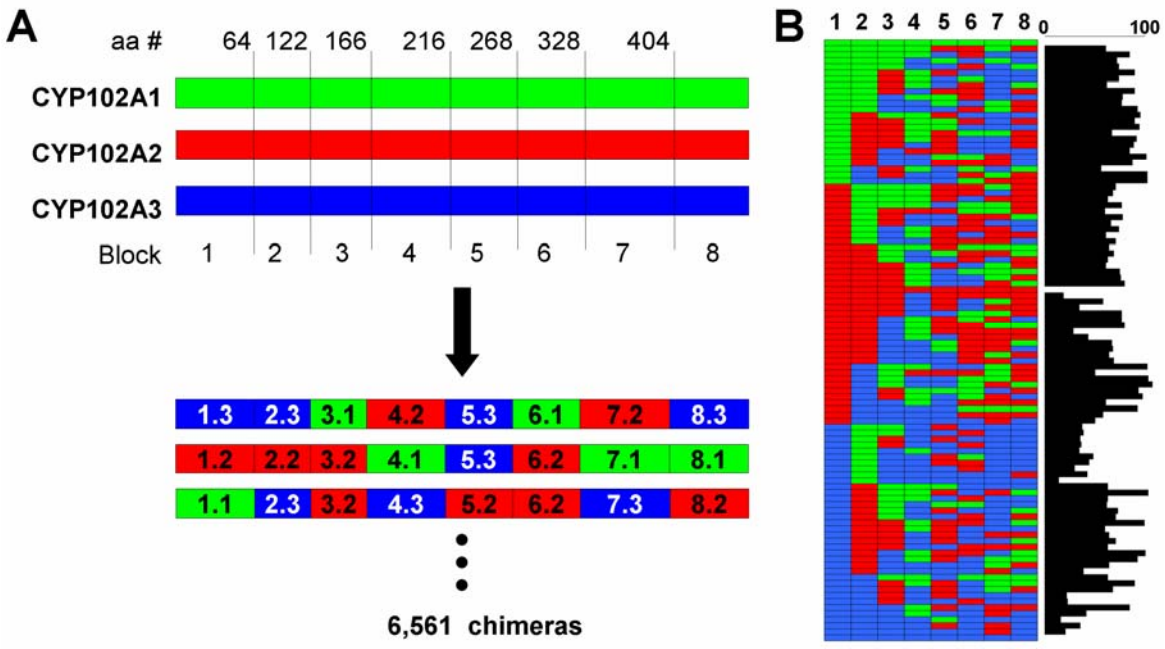

Figure 5.1. Diverse chimeras created by site-directed recombination. (A) Site-directed recombination of three bacterial cytochromes P450 showing crossover sites chosen to minimize the number of disrupted contacts (number is last residue of the sequence block according to CYP102A1 numbering). Blocks are assigned numbers 1 through 8 and three fragments are possible at each block. Three example chimeras are shown to illustrate the fragment nomenclature, e.g., fragment 1.3 is block 1 inherited from parent A3. (B) Sequences of three parents and 97 folded P450 chimeras and number of amino acid changes relative to the closest parent (bar on right).

To accomplish this, we used the structure of the heme domain from CYP102A1 [25] to computationally evaluate 5,000 libraries with seven crossovers, each of which contained 3^8 = 6,561 chimeric sequences (including the parents). Crossover sites were chosen randomly, with a minimum fragment size of 20 residues. To estimate the fraction of folded proteins in each library, we counted the number of structural contacts, E, disrupted in each chimeric sequence (see Materials and Methods) [15,17]. Based on data from the seventee A1–A2 chimeras individually constructed and studied in Chapter 4, we modeled the probability of folding as a step function which decreases from 1 to 0 at a threshold of

E = 30. Fraction folded was thus calculated as the number of chimeras in each library with $E \leq 30$ divided by the total number of chimeras (= 6,561). The average number of amino acid substitutions from the closest parent <m> for the folded proteins (those with $E \leq 30$) was also calculated as a measure of the library sequence diversity. The fraction folded and <m> for 5,000 randomly generated libraries can be seen in Figure 5.2.

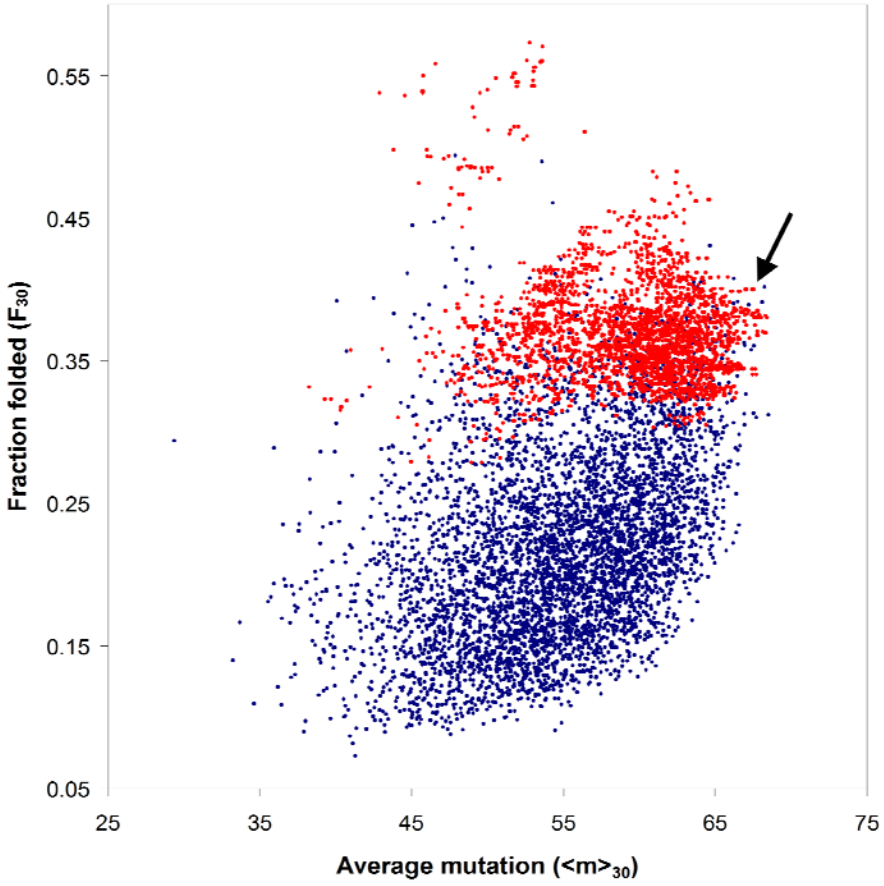

Figure 5.2. Choosing a library using random enumaration. Blue circles represent 5,000 randomly chosen seven-crossover libraries with minimum block sizes of twenty amino acids. The average fraction folded and average <m> for the randomly enumerated libraries are 22.0 ± 6.7% and 54.6 ± 6.0, respectively. Red circles represent complete enumeration of fourteen frequently occurring crossovers appearing in the randomly generated libraries with a fraction folded greater than 25%. The average fraction folded and average <m> for the completely enumerated libraries are 37.0 ± 3.6% and 59.6 ± 4.8, respectively. The arrow identifies the library chosen for construction (see text).

From the set of 5,000 randomly generated libraries, we selected only those with a fraction folded greater than 25% for further study. Within these, 14 crossover locations appeared most frequently (crossovers located after residues Ser65, Leu104, Ile122, Leu148, Tyr166, Gln204, Val216, Gly227, Thr268, Gln307, Ala328, His361, Pro376, and Gln403). Using these 14 crossover sites, we evaluated all 3,432 possible seven-crossover libraries and chose one with a high fraction folded (40%), high diversity ($\langle m \rangle = 68$ for the chimeras with $E \leq 30$, $\langle m \rangle = 76.4$ for the library as a whole), and crossovers distributed over the primary sequence (average number of residues per block = 59 ± 10). The final design has crossovers located after residues Glu64, Ile122, Tyr166, Val216, Thr268, Ala328, and Gln404, based on the numbering of the A1 sequence (Figure 5.1A).

Figure 5.2 shows how the libraries consisting of subsets of the 14 frequent crossovers (subset libraries) compare to those with crossovers chosen randomly. The average fraction folded and average diversity (average $\langle m \rangle$) are both significantly higher in the subset libraries than in the random libraries, demonstrating the enrichment of high quality libraries in the subset libraries (see Figure 5.2 legend).

Following design and construction of this library, Endelman et al. employed dynamic programming to quickly compute optimal site-directed libraries, termed RASPP (recombination as a shortest-path problem) [26,27]. Instead of fraction folded, RASPP uses average *E* to assess the frequency of viable protein sequences. A comparison of the random-enumeration to the RASPP curve shows that random enumeration of this small

number of libraries does not discover libraries with comparable <*E*> at a given value of <*m*> (Figure 5.3). Interestingly, the final library design chosen from the population of subset libraries falls along the RASPP curve. Thus by choosing frequently occurring crossover sites within a small set of randomly sampled libraries and completely enumerating those libraries we were able to achieve similar design results to RASPP. However, the computational time required for random enumeration followed by complete enumeration of the subset libraries is far greater than with RASPP.

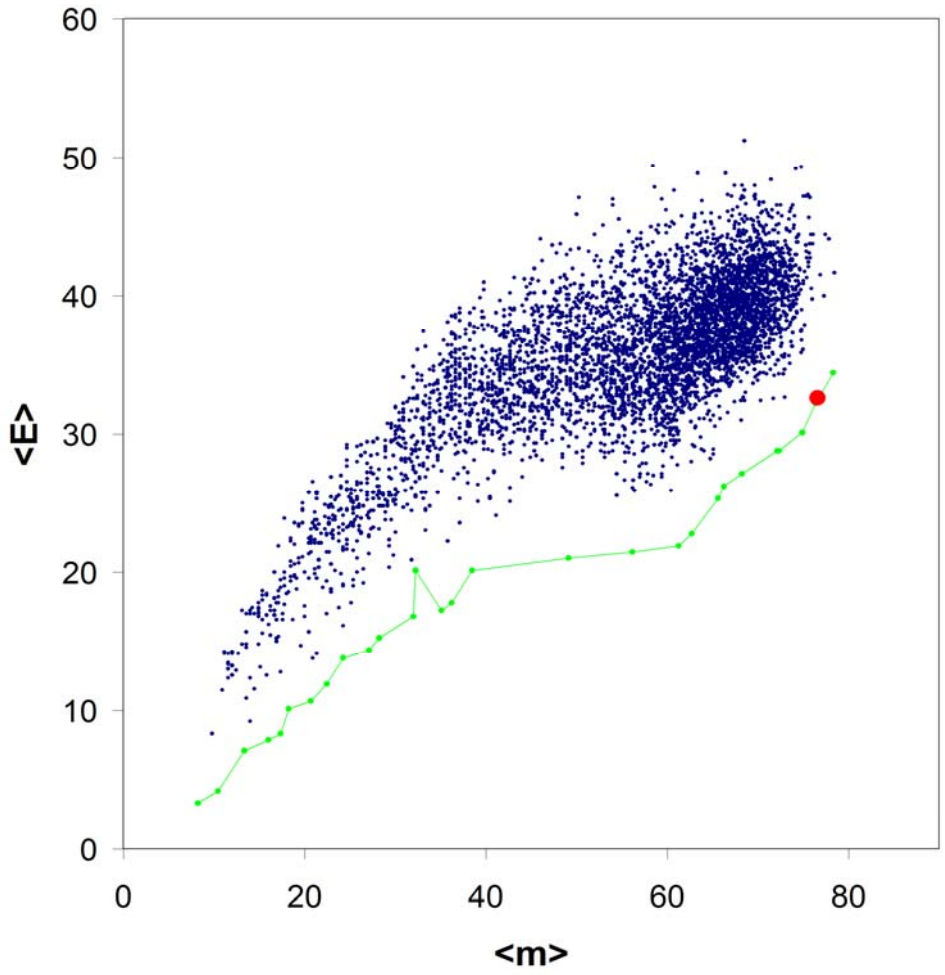

Figure 5.3. RASPP curve for seven crossover libraries between A1, A2 and A3. The blue circles represent 5,000 randomly generated libraries with a minimum mutational level per block of two. The green line represents the RASPP curve generated with the same mutational constraints as the random generation with a bin size of two. The large red point represents the <E> and <m> values for the library that was chosen here.

The individual structural elements identified by SCHEMA are not obvious based on secondary or domain structure (Figure 5.4 and 5.5A). For example, the crossovers between blocks 2–3, 4–5, 5–6, and 7–8 lie within the D, G, I and L helices, respectively [28]. Individual blocks, however, combine to form larger structural elements that coincide with protein domains determined from inspection of the A1 crystal structure

[28] and concerted motions evident in molecular dynamics simulations of the same protein [29] (Figure 5.5A). Blocks 1 and 7 comprise the independent "β domain," most of which is a five-stranded β-sheet. The two-stranded, anti-parallel β-sheet comes from block 7, while the remaining three β-strands are contributed by block 1. The library design divided this domain into the fewest possible pieces. The remaining blocks comprise the "α domain" [28], which on the basis of concerted protein motions has been divided further into α′ (corresponding to blocks 4 and 5) and α″ domains (blocks 6 and 8) [29]. These three domains reflect groups of residues that move together not only in molecular dynamic simulations but also between different conformations of A1, which undergoes a large conformational change upon substrate binding [30]. Considering the root-mean-square deviation (RMSD) between the substrate-bound (closed) and substrate-free (open) forms of A1 (Figure 5.5B) [31], five of seven crossovers are in regions which move 1.2 Å or less, significantly less than the average displacement of 2.2 Å, and capture the boundaries of the previously defined domains within six residues.

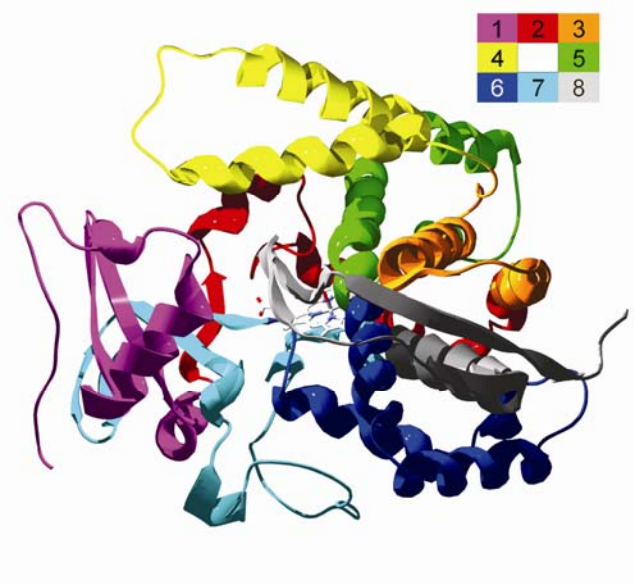

Figure 5.4. Structural model of heme-domain backbone structure showing positions of each block. Model is based on the crystal structure of CYP102A1 (2HPD) [28]. Blocks are color-coded as shown and heme is shown in CPK coloring.

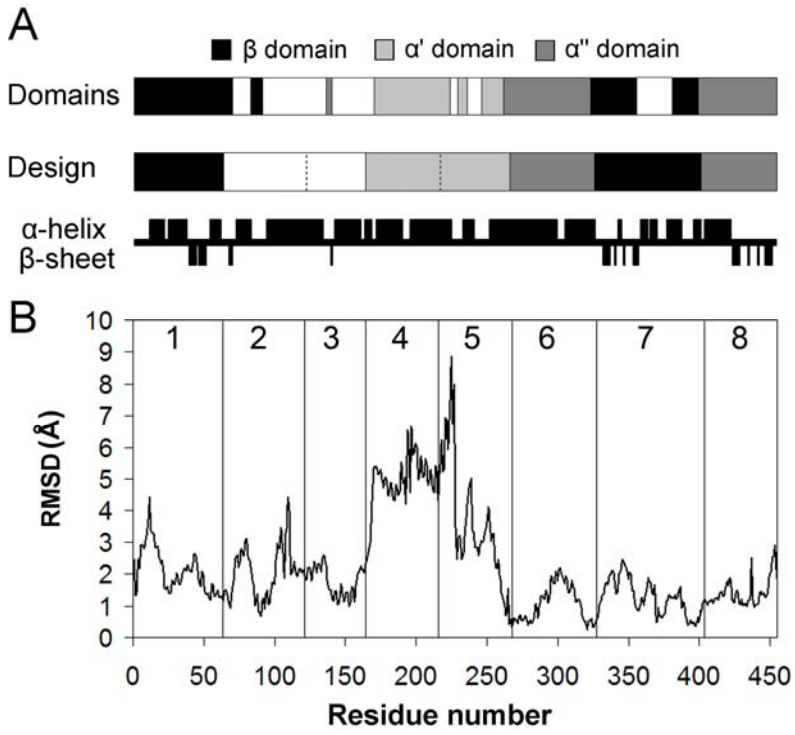

Figure 5.5. Comparison of library design to domains, dynamics, and secondary structure of CYP102A1. (A) Crossovers in the library designed using the SCHEMA energy function capture domain boundaries of CYP102A1 determined from molecular dynamics simulations [29]. Crossovers between blocks 2–3, 4–5, 5–6, and 7–8 lie within α-helices. (Secondary structure assignment is based on the CYP102A1 crystal structure [25]). (B) Plot of the RMSD between the backbone atoms of the substrate-bound (closed) and unbound (open) structures of CYP102A1. The RMSD was calculated by comparing molecule B of the substrate-free structure [31] and molecule A of the structure bound to palmitoleic acid [28] using Swiss PDB Viewer. Vertical lines designate crossover locations and blocks are numbered. Crossovers between blocks 1–2, 5–6, 6–7, and 7–8 occur at positions that move < 1.2 Å between the two structures. Crossover 3–4 is located next to a region of high identity and may be shifted towards the N-terminus by up to 14 residues and still produce the same chimeras. This shift allows it to occur at a position which moves < 1.2 Å.

The three gene fragments encoding each of the eight blocks were combinatorially assembled using the sequence-independent site-directed chimeragenesis (SISDC) [32] method developed specifically for this application to generate a gene library containing

6,561 different sequences (Figure 5.1A). These genes were expressed in *E. coli,* where high-throughput sequencing by DNA probe hybridization and functional assays determined the sequences and functions of the proteins they encoded.

C.2 Sequence analysis

Because the crossover locations are fixed, the complete sequence of a chimera (absent any point mutations, insertions, or deletions) can be obtained by determining which parent sequence is present at each block by DNA probe hybridization [33]. Out of 1,512 randomly selected colonies analyzed this way, 754 complete sequences were obtained. Of these, 628 were unique. The distribution of fragments in this sample revealed two main biases from the ideal incorporation of 33% of each parent at each block (Figure 5.6): at block 1, parent A1 is present in 10% of the chimeras, while parent A2 is present at block 4 in only 0.5%.

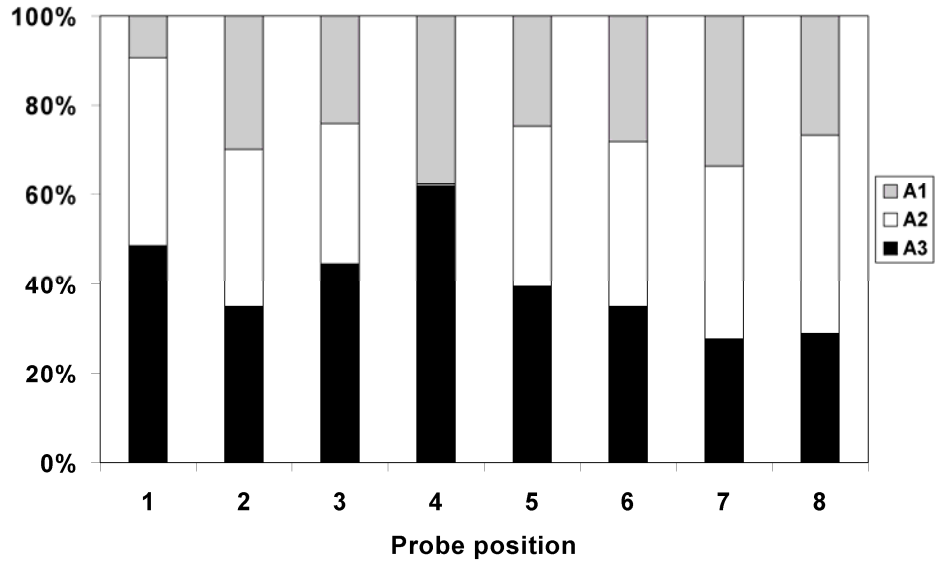

Figure 5.6. Fragment distribution at each block based on probe hybridization of genes from 754 unselected clones.

We completely sequenced 39 chimeras in order to assess the frequency of point mutations and of insertions, deletions, and remaining tag sequences (indels). Tag sequences were inserted at each crossover location for library construction by SISDC, and any remaining tag sequences result in a large insertion. In seven randomly chosen chimeras we found only one synonymous point mutation and no indels. We also sequenced 32 randomly chosen chimeras for which folding status had been determined. Twenty of these encoded folded P450s, while 12 encoded proteins that were not P450s. In the 20 folded P450 sequences, there were zero remaining tag indels and two point mutations. In the 12 not-folded sequences, one point mutation and one remaining tag sequence were found. From the overall point mutation frequency of 0.007% (in 51,568 nucleotides), we estimate that fewer than 10% of the chimeras in the library contain a point mutation. No indels or tag sequences were found in any of the folded P450 sequences, and fewer than 9% of the not folded chimeras contain indels or tags. Comparing the results from DNA sequencing and probe hybridization analysis, we found that probe hybridization identified the correct fragment at all eight blocks in 31 of 32 sequences. Thus the sequencing information from probe hybridization reflects the true sequences of the chimeras with errors in less than 10% of the chimeras, the majority of which are due to single point mutations.

C.3 Assignment of folding status

Using high-throughput CO difference spectroscopy (see Chapter 3), we assayed clones from the chimeric P450 library for the characteristic Soret peak at 450 nm. The presence of this peak indicates correct heme binding and thus a properly folded P450 heme

protein. Of the 628 unique full-length sequences, 293 (47%) encoded folded P450s. Additional sequencing of folded P450s yielded an expanded dataset containing 955 unique sequences (including the three parents), of which 620 correctly incorporate heme and 335 do not (Appenix B). Thirty-eight of these 335 not-folded sequences gave a peak at 420 nm, characteristic of improperly incorporated heme and a nonfunctional enzyme [34,35]. The remaining not-folded sequences lack a compatible heme-binding site and likely do not fold into a well-defined structure.

The folded sequences are highly mosaic and differ from their parents by 72.5 amino acids on average, with as many as 109 amino acid substitutions from the nearest parent sequence (Figure 5.1B and Appendix B). The average number of disruptions (<*E*>) is lower in chimeras that bind heme (29.5) versus those that do not (34.8). The average number of mutations in the heme-binding chimeras is also lower, 72.9 versus 77.5. The compositions of chimeras can be easily visualized using ternary diagrams (Figure 5.7). For example, the sequence biases against single A1 and A2 fragments in the library construction generates fewer chimeras whose compositions are very close to A1 or A2 (Figure 5.7A). It is clear from this plot, however, that the overall compositions of folded and not-folded chimeras are not markedly different and are well distributed over the accessible composition space.

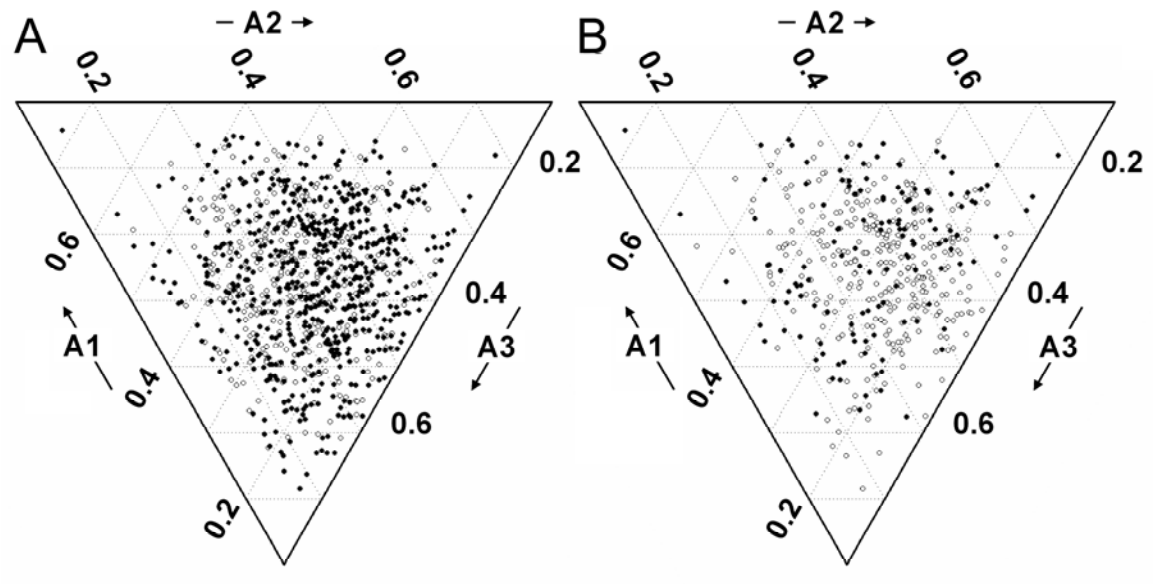

Figure 5.7. Ternary diagrams showing the distribution of chimera amino acid compositions. (A) Compositions of 955 folded (closed circles) and not-folded (open circles) chimeric sequences. Each data point represents the relative amino acid identity between a chimera and each parental sequence not including positions conserved between all three parents. This distance was calculated by determining the number of amino acids a chimera shares with each parent and dividing by their sum. The three relative identities add up to one. Since each parent shares some sequence identity with the other two, they do not lie at the corners of the diagram. (B) Compositions of 441 chimeras tested for activity on 12-pNCA: active chimeras (closed circles) and not active (open circles). Chimeras composed mostly of A3 and chimeras near the center tend to be inactive on 12-pNCA.

C.4 Catalytic activities of folded P450 chimeras

We estimated the fraction of chimeras that are functional by assaying 320 folded P450 chimeras for peroxygenase activity on 2-phenoxyethanol, a substrate accepted by all three parents. Reaction on this substrate yields phenol (Figure 5.8), which is detectable in high throughput(see Chapter 3). The three parent P450s naturally occur as fusion proteins to an FAD- and FMN-containing NADP reductase [22]. These monooxygenases use NADPH and molecular oxygen to hydroxylate fatty acids [23]. The parent heme

domains, by virtue of the single amino acid substitutions F87A in A1 and F88A in A2 and A3, also function as peroxygenases, catalyzing oxygen insertion in the presence of hydrogen peroxide [24]. Chimeras that produced at least 25% of the total product formed in the assay by the most active parent (A1) were considered active. Of the 320 folded chimeras assayed, 72% were found to be active on 2-phenoxyethanol.

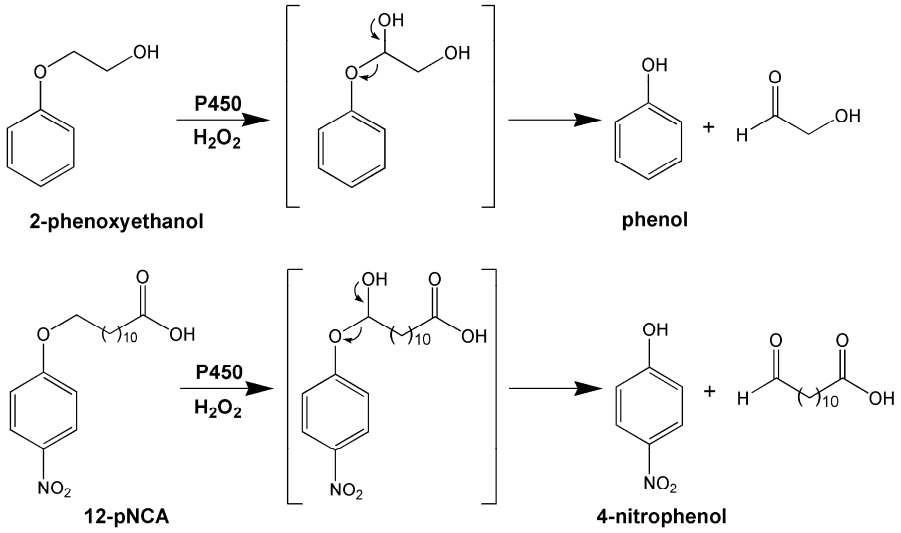

Figure 5.8. Substrates and major products of P450 peroxygenase reactions with 2-phenoxyethanol and 12-pNCA. In both cases, hydroxylation yields a hemiacetal which decomposes to phenolic products detectable in high-throughput assays.

We also assayed all the 955 chimeras for which the sequences and folding status were determined for activity on the fatty acid analog p-nitrophenoxydodecanoic acid (12-pNCA, Figure 5.8). The parent A1 and A2 heme domains are active on 12-pNCA, while A3 is not. Chimeras with 25% of the total product formed by A1 during the assay were considered active. None of the chimeras that did not fold properly showed activity. We then determined activity status for folded P450s whose concentration was at least 500

nM, in order to remove false negatives based on low expression or other experimental errors. Of the folded chimeras, 441 met this constraint, of which 134 (30%) were active on 12-pNCA (Appendix B). The average number of disruptions is lower for chimeras active on 12-pNCA versus those that are not (<*E*> = 26.3 versus 31.4). Mutations are similarly lower in active chimeras (<*m*> = 70.9 versus 76.9). The average number of crossovers in the active chimeras was 4.6 compared to 4.4 for those that were not active.

A ternary diagram showing the 441 chimeras tested for activity on 12-pNCA (Figure 5.8B) demonstrates that the sampled sequences are distributed similarly to the larger dataset (Figure 5.7A). Parent A3 is inactive on 12-pNCA, and there are only a few chimeras with a high fraction of sequence from A3 that exhibit this activity. Additionally, there is a lower density of active chimeras near the center, where the chimeric sequences have the greatest divergence from the parents.

Fewer chimeras showed activity on 12-pNCA than on 2-phenoxyethanol, which we attribute to the fact that one parent, A3, is not active towards 12-pNCA, while all three parents are active on 2-phenoxyethanol. Overall, 73% of the folded chimeras assayed exhibited peroxygenase activity on at least one of these two substrates. Thus, at least 35% of the 6,561 sequences in the library are folded and functional, corresponding to 2,300 new P450 enzymes, not including any that are active on substrates not tested. This functional fraction is roughly three times higher than reported in a study in which more closely related cytochromes P450 (>71% amino acid identity) were recombined using a

DNA shuffling methodology that leads to crossovers at regions of high sequence identity [36].

C.5 Thermostabilities of folded P450 chimeras

To examine how recombination affects protein stability, we measured the melting temperatures of the parent P450s and 14 chimeras (all of which denature irreversibly at high temperature) by monitoring the disappearance of the P450 Soret peak with increasing temperature. A range of T_m's (42 °C–62 °C) was observed in this small sample (Table 5.1). The most stable chimera differs from its closest parent by 84 amino acid substitutions, yet its melting temperature is 7 °C higher than the most stable parent. It is also higher than that of a variant of the A1 heme domain previously stabilized by sequential random mutagenesis and screening [37]. If a chimera is able to bind heme, then on average its stability appears not to be compromised relative to the parent proteins. The ability of the blocks to assemble into more thermostable proteins when removed from their natural context supports the modular nature of these elements and likely reflects some intrinsic stability of the individual blocks, due to the large number of structural contacts preserved by the library design.

Table 5.1. Thermostabilities of parent and chimeric heme domains.

Sequence[*]	**T_m (°C)** [†]
A1	55
A2	44
A3	49
23113312	43
23133121	45
32312231	51
22312333	62
32312332	52
32312333	56
21333223	54
12112333	49
12313331	49
11213231	53
22313232	52
21113212	49
22213222	48
32213333	47

[*] Sequence listed as parent sequence (A1 = 1, A2 =2, A3 = 3) at each block

[†] T_m's are the average of three measurements. Standard deviations is less than 1°C.

C.6 Logistic regression analysis of the multiple sequence alignments

Small sets of chimeric P450s have been constructed previously for investigations of sequence-structure-function relationships [38,39]. The MSA of natural protein families are also widely used for this purpose. Comprised of sequences largely uncoupled from natural selection, including sequences that encode nonnatural functions (such as not folding or not functioning), the artificial protein family described here offers a unique opportunity to elucidate key sequence and structural contributions to P450 folding and function. By analyzing the MSAs of the chimeric P450s we can identify how different

blocks and their parental identities influence folding and heme binding or catalytic activity. Because this dataset also includes sequences that encode not-folded and not-functional proteins, we can use logistic regression analysis (LRA), an analog of linear regression suitable for the type of binary data presented here, to analyze the MSAs. Other, more commonly used methods such as contingency table [40,41] and statistical coupling [1,42] are unable to utilize the additional information provided by the sequences that do not fold or function.

Underlying our LRA of the folded/not-folded dataset is the idea that individual fragments and interactions between fragment pairs contribute to whether a chimera will fold and bind heme. LRA fits an energy model containing intra- and inter-fragment terms; the magnitude of each term reflects how strongly that variable affects the likelihood of folding, with negative values increasing the likelihood and positive values decreasing it [21]. If energy is below a threshold, a chimera is assumed to be folded; otherwise it is not. In order to avoid overfitting the data, *p*-value testing is used to determine which fragments make a significant contribution to predicting chimera folding status.

We applied LRA to the MSA of the entire set of 955 chimeric P450s in Appendix B to determine which blocks contribute to folding and correct heme binding. The resulting energy model includes blocks and block pairs that are significant with the likelihood ratio test and cross-validation (see Materials and Methods). This analysis revealed that blocks 1, 5, and 7 by themselves and the interaction between blocks 1 and 7 (abbreviated 1–7) contribute significantly to whether a chimeric P450 folds and binds heme (Figure 5.9).

All other blocks and block pairs are apparently to a large extent interchangeable with respect to whether a chimera folds properly.

A

Block	Parent A1	Parent A2	Parent A3
1	0.3	-1.1	0.8
5	1.8	-1.1	-0.7
7	-0.1	1.2	-1.1

B

	7.1	7.2	7.3
1.1	-1.6	1.5	0.1
1.2	0.3	-1.3	1.0
1.3	1.3	-0.2	-1.1

C

Block	Parent A1	Parent A2	Parent A3
2	-1.4	-1.5	2.9
4	-1.3	0.1	1.2

D

	8.1	8.2	8.3
1.1	-1.3	1.7	-0.4
1.2	1.2	-2.4	1.2
1.3	0.2	0.6	-0.8

Figure 5.9. Logistic regression analysis of multiple sequence alignments identified blocks and block pairs that contribute to whether a chimera folds and binds heme and whether it exhibits activity on 12-pNCA. (A) Intra-fragment terms in the energy model from LRA of folded/not-folded sequences indicate that blocks 1, 5, and 7 make significant contributions to folding and incorporation of heme. Negative energies increase the likelihood of folding and correctly binding heme while positive ones decrease it. (B) The single significant inter-fragment interaction from LRA of folded/not-folded sequences comes from pair 1–7 and includes the nine energy terms for pair 1–7, which can be divided into three groups. The on-diagonal elements (filled black) are the most stabilizing. The three terms filled gray have roughly average energy. The three white elements are destabilizing relative to the others. (C) Significant intra-fragment terms from LRA of the MSA of active/not-active sequences indicate that blocks 2 and 4 have significant effects on peroxygenase activity. (D) The single significant inter-fragment interaction between blocks 1 and 8, showing the nine terms, divided into similar groups as in part B.

As shown in Figure 5.9A, the intra-fragment terms for fragments 1.2 and 7.3 have lower energy relative to the other parents, which means the sequence changes in these fragments are more favorable for heme binding. Blocks 1 and 7 are in fact expected to be important, because they contain the most residues, the greatest number of intra-fragment contacts (Figure 5.10), and block 1 has the highest average number of sequence changes,

whereas block 7 has the third most (Figure 5.10). In contrast, block 5 has the third fewest intra-fragment contacts and the second fewest average number of sequence changes (Figure 5.10**).** At this block, fragment 5.1 is the least favored of all the fragments for folding and heme binding (Figure 5.9A). Parent A1 contains a deletion relative to A2 and A3 in block 5, which may contribute to this behavior. We suspect that some of the importance of block 5 is due to the dynamic nature of cytochromes P450, similar to what has been observed in multiple sequence analyses of other protein families [2]. The F, G, and H helices (in blocks 4 and 5) undergo displacements of more than 5 Å between the substrate-bound and substrate-free forms of A1 [31], and block 5 moves an average of 3.6 Å (Figure 5.11A). This portion of the enzyme acts as a "hinge" by which the F and G helices close down upon the substrate. Because none of the residues in block 5 that contact the heme differ among the three parents, the importance must stem from how variable amino acids in block 5 affect dynamics or interact with conserved residues.

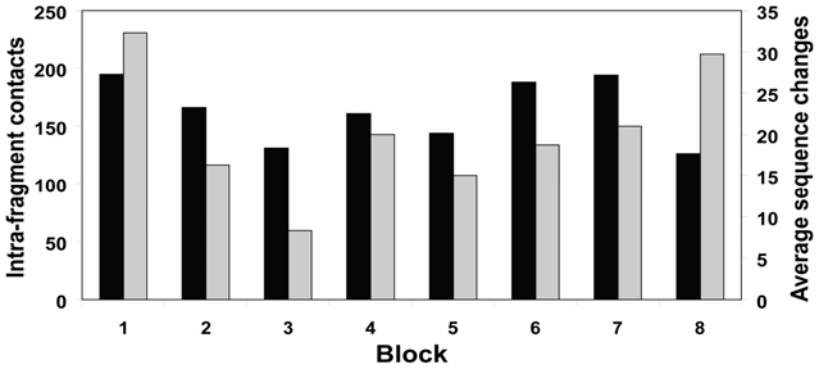

Figure 5.10. Black bars, intra-fragment contacts within each block, as defined by the SCHEMA distance of 4.5 Å [17]. Gray bars, the average number of sequence changes between each parent.

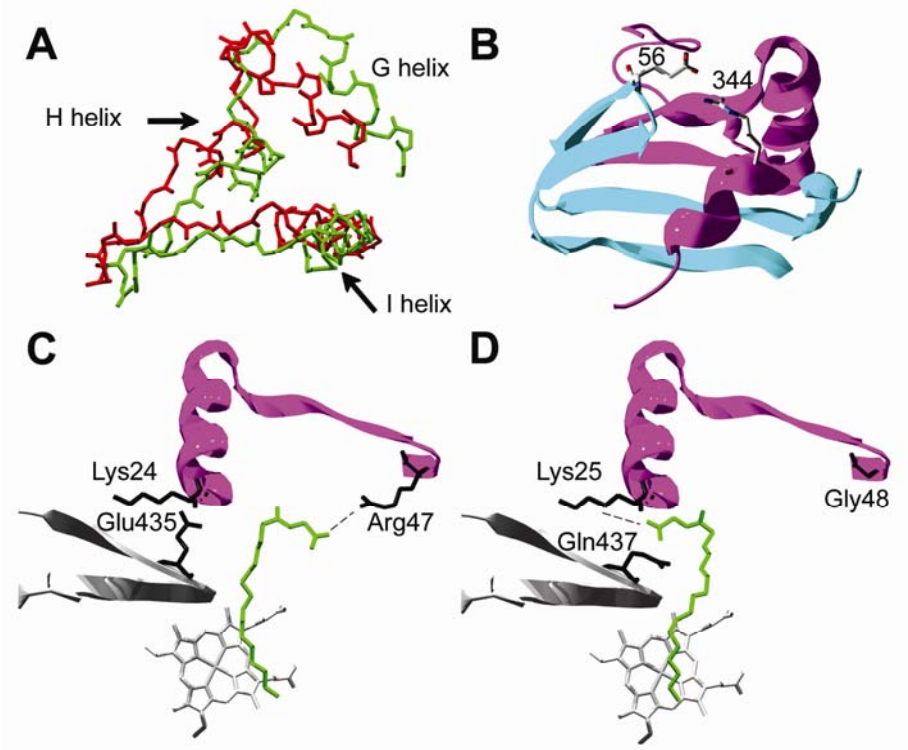

Figure 5.11. Structural elements that contribute significantly to proper folding and incorporation of heme and model of substrate binding in CYP102A1 and CYP102A2. (A) Movement of block 5 between open (red) and closed (green) structural forms based on alignment of heme cofactor. The average displacement over the whole block is 3.6 Å. (B) Residues that could contribute to positively and negatively interacting fragments at blocks 1 and 7. Residue 56 (shown as arginine) is an arginine, glutamate, and glutamine; and residue 344 (shown as glutamate) is a glutamate, lysine, and glutamate in A1, A2, and A3, respectively. The fragment pairs that result in unfavorable charge–charge interactions for these closely spaced side chains are unfavorable overall for folding and heme incorporation. (C) In CYP102A1 the carboxylate group of the fatty acid substrate (in green) interacts with arginine 47 from block 1 (dashed line). Residue 435, from block 8, and residue 24 may form a salt bridge. Portions of blocks 1 and 8 are shown in purple and grey, respectively. (D) Proposed model for CYP102A2 showing an alternative binding configuration for the fatty acid substrate. Residue 437 (in block 8) is a glutamine in A2. Thus in A2, lysine 25 is free to interact with the substrate carboxylate group (dashed line). Structure shown is 1FAG [28]. Amino acid residues are in black and heme is grey.

Block pair 1–7 was the only pair revealed by LRA as significant for folding and incorporation of heme. Blocks 1 and 7 interact extensively to form the β-domain (Figure

5.11B) and experience the largest average number of broken contacts when the blocks are inherited from different parents. As expected, chimeras that inherit blocks 1 and 7 from the same parent are more likely to fold and bind heme (Figure 5.9B). This result supports the core hypothesis of SCHEMA and other penalizing energy functions [43] which assign the best possible score to these wild-type interactions.

Inspection of the sequences of the parents in these two blocks revealed an electrostatic interaction that could contribute to the pattern of energies in Figure 5.9B. Residues 56 (block 1) and 344 (block 7) are 2.8 Å apart in the A1 crystal structure (Figure 5.11B). At position 56, parent A1 contains a positively charged arginine, A2 has a negatively charged glutamate, and A3 has a neutral glutamine. Residue 344 is a glutamate in A1 and A3, but lysine in A2. Thus the interaction 1.1–7.2 pairs arginine and lysine, while 1.2–7.3 pairs glutamate and glutamate, both of which are repulsive.

We repeated the logistic regression analysis to determine which blocks affect activity on 12-pNCA, independent of heme binding, by applying LRA to the subset of 441 folded chimeras for which presence or absence of activity on 12-pNCA had been determined (Appendix A). This analysis revealed that blocks 2 and 4 by themselves and block pair 1–8 contribute to whether a folded chimera is catalytically active on this substrate. At blocks 2 and 4, the fragments derived from parent A3 are detrimental to activity (Figure 5.9C). These sequence elements likely account for A3's lack of activity on this substrate, since sequence from this parent at other blocks has little affect on 12-pNCA activity in the chimeras. The importance of block pair 1–8 may reflect a difference between A1 and

A2 with respect to substrate binding: when A1 or A2 is present at either block 1 or 8, activity is strongly dependent on whether the other block comes from the same parent (Figure 5.9D). This indicates that there are one or more interactions between blocks 1 and 8 that must be preserved in order for the enzyme to be active on 12-pNCA.

C.7 Residues contributing to peroxygenase activity on 12-pNCA

We sought to determine what interactions(s) might be responsible for the importance of the 1–8 pair, using the sequence differences in parents A1 and A2 for guidance. One obvious difference occurs at the position corresponding to Arg47 in fragment 1.1, which is located at the opening of the active site and is thought to interact with the carboxylate group of fatty acid substrates [31]. Substitutions of this residue in the A1 holoenzyme significantly reduce catalytic activity [44,45]. In A2, the equivalent residue is Gly48, a residue that favors the binding of polycyclic aromatic hydrocarbons when present in the A1 holoenzyme [46]. We tested the importance of R47 to peroxygenase activity by swapping the residues at position 47/48 in A1 and A2, i.e., making the single mutation R47G in A1 and G48R in A2. The R47G mutation in A1 reduced the initial rate nearly 25 fold (from 65.9 ± 8.5 to 2.7 ± 0.5 nmol product/nmol P450/min), making it comparable to the activity of A2. On the other hand, the G48R mutation in A2 had no effect on rate. This suggested to us that G48 in A2 does not interact with the substrate carboxylate group, as the equivalent residue appears to do in A1.

We postulated that the different mode of substrate binding could be facilitated by a positively charged residue elsewhere in the A2 sequence. Only a small portion of block 8,

consisting of halves of two β-strands (residues 434 to 439), is located near the active site (Figure 5.11C). Examination of the parental sequence alignment in this region (Appendix A), however, revealed no lysines or arginines unique to fragment 8.2. Because fragments 8.1 and 8.3 are equally incompatible with 1.2 according to the LRA, we looked for a residue between 434 and 439 that was shared by A1 and A3 but not A2. Residue 435 in A1 (437 in A2 and A3), which is a glutamate in A1 and A3 and a glutamine in A2, met these criteria.

We then swapped these residues by making the E435Q mutation in A1 and the Q437E mutation in A2. The E435Q mutation in A1 reduced catalytic rate by 8 fold, whereas the Q437E mutation completely abolished the activity of A2 (Table 5.2). Having shown this residue to be important to activity in both parents, we next chose eight inactive chimeras containing unfavorable 1–8 block combinations to determine whether swapping these positions could "rescue" the activity. We introduced the Q437E mutation into four chimeras with fragments 1.1 and 8.2 and the E435Q mutation into four with fragments 1.2 and 8.1 (Table 5.2). This single substitution was able to confer activity in two of the eight chimeras.

Table 5.2. Peroxygenase activities of site-directed mutants of parents CYP102A1 and CYP102A2 and selected chimeric heme domains on 12-pNCA.

Sequence	Wild type[1]	Glu435Gln or Gln437Glu[2]
CYP102A1	65.9 ± 8.5	8.9 ± 1.7
CYP102A2	2.3 ± 0.5	n.d.
11332212	n.d.	n.d.
11331312	n.d.	0.8 ± 0.3
12232232	n.d.	n.d.
13233212	n.d.	n.d.
21113211	n.d.	n.d.
23213211	n.d.	n.d.
22131221	n.d.	n.d.
22233211	n.d.	0.9 ± 0.1

[1]All rates are reported in nmol product/nmol P450/minute. Activities < 0.1 were not detectable (n.d.). Wild type indicates heme-domain sequence with F87A (A1) or F88A (A2) mutation.
[2]The Glu435Gln mutation was made when block 8 contained fragment 8.1. The Gln437Glu mutation was made in fragment 8.2.

Thus the LRA analysis in combination with mutation studies uncovered a residue (Glu435/Gln437) previously unknown to be important for catalytic activity and suggests a different substrate binding mode in CYP102A2. One structural explanation for these results is illustrated in Figure 5.11C and 5.11D. Since A2 lacks a positive charge at position 48 and has no unique positively charged residues in the small portion of block 8 near the active site (or block 8 altogether), we hypothesized that another sequence change may have caused a positively charged residue to be made available elsewhere. Glu435 in A1 appears to participate in a salt bridge with Lys24, which is roughly 4 Å away in the crystal structure. The equivalent residue 25 is a lysine in A2 and a glutamine in A3. The

lack of a salt bridge partner near Lys25 in A2 could free Lys25 to interact with the carboxylate tail of the fatty acid (Figure 5.11D). In support of this, a single substitution of Gln437 to Glu rescued the activity of a chimera containing A2 sequence at block 8, but A1 sequence at block 1. Conversely, switching Glu435 to Gln in a chimera containing A1 sequence at block 8 but A2 sequence at block 1 was also able to rescue the activity. Of course, this single switch was unable to rescue activity in six more folded, but inactive chimeras, which indicates that additional interactions are also important (such as the contributions from residues in blocks 2 and 4).

C.8 SCHEMA-guided recombination creates a library rich in properly folded, highly mutated sequences

The approach used here to identify optimal recombination sites differs from the SCHEMA profile described previously [15]. Evaluating libraries with randomly sampled crossovers, as was done here, and a recently developed global optimization of recombination sites [26] are both preferred over the SCHEMA profile, which neglects important structural interactions between amino acids distant in the primary sequence. Based on this design, three cytochromes P450 were divided into “building blocks” and combinatorially reassembled to yield a library in which 47% of the members fold and correctly bind heme. This folded fraction is slightly larger than the prediction of 40% from the design. The full library therefore contains an estimated 3,000 unique chimeric P450s, many of which are highly mutated compared to the parent P450s.

It is interesting to estimate the extent to which SCHEMA recombination has enriched the library relative to a library having the same distribution of mutation levels, but made using random mutagenesis. The fraction of folded proteins in a random library can be estimated using the protein's "neutrality," or probability that a random amino acid substitution will not disrupt folding. Neutrality v has been calculated for other proteins and ranges from 0.38 to 0.56 [8]. Using 0.6 as a conservative estimate for P450 neutrality, the fraction of folded P450s having a mutation distribution equaling that of the chimeras (ff_r) is given by

$$ff_r = \frac{1}{N}\sum_{m=1}^{109} N_m \times v^m , \qquad (3)$$

were $v = 0.60$, N = total number of mutants (628, equal to the unique set of randomly sampled chimeras), m = number of amino acid changes, and N_m = number of mutants with a given value of m. This yields a fraction folded $ff_r = 6.3 \times 10^{-5}$. The fraction of folded chimeras in the library is 0.47, giving an enrichment of $0.47/ff_r = 7.5 \times 10^3$. Thus, by this conservative estimate, SCHEMA-guided recombination has increased the frequency of folded chimeras by nearly four orders of magnitude.

D. Conclusions

Protein families generated in the laboratory can be used to identify regions of the sequence and structure that are important for folding and function. This approach may be especially valuable for proteins with few naturally occurring family members. Datasets such as this one, containing hundreds of proteins for which functional information can be

determined in high-throughput assays, will be invaluable for developing and validating structure prediction tools and for protein sequence-structure-function analysis. Finally, rich in sequence diversity as well as the ability to fold properly, these proteins may be sources of novel functions for laboratory protein evolution.

E. Materials and Methods

E.1 Calculation of SCHEMA disruption

The parent heme-domain sequences of A1, A2, and A3 were aligned using ClustalW [47] (Appendix A). The number of broken contacts in a chimera *E* [15,17] is

$$E = \sum_{i} \sum_{j>i} C_{ij} \Delta_{ij} , \qquad (4)$$

where the C_{ij} are elements of the contact matrix which depend solely on the protein structure. Specifically, $C_{ij} = 1$ if residues *i* and *j* are within 4.5 Å in the structure of A1 bound to N-palmitoylglycine (1JPZ) [25]; otherwise $C_{ij} = 0$. The SCHEMA delta function Δ_{ij} uses only the parental sequence alignment: $\Delta_{ij} = 0$ if the amino acids found in the chimera at positions *i* and *j* are also found together in any single parent at the same positions. Otherwise, the *i–j* contact is considered broken, and $\Delta_{ij} = 1$.

E.2 Library construction

The heme domains of A1 and A2 are described in Chapter 4. The heme domain (first 1,401 nucleotides) of the A3 gene (a gift from Claes von Wachenfeldt, Lund University) was subcloned into the BamHI/EcoRI sites of the pCWori expression vector [48], and the

mutation corresponding to F88A was introduced. The chimeric library was constructed following the sequence-independent site-directed chimeragenesis (SISDC) method [32], using the type IIb restriction endonuclease *Bsa*XI. This required removal of a *Bsa*XI site in parent A1 using a synonymous mutation at Gly368 (GGA to GGT). N- (blocks 1-4) and C-terminal (blocks 5-8) half-length parental genes were constructed separately with a tag sequence containing a *Bsa*XI recognition site at each of the three crossover locations (Table 5.3). The half-length genes were cloned into the pCR®-Blunt II-TOPO plasmid (Invitrogen) and their DNA sequences were confirmed by sequencing (Laragen, Los Angeles, CA). Digestion of the half-length parental genes with *BsaX*I to remove the tag sequences yielded DNA fragments with unique 3 bp overhangs at each crossover site. After column purification to remove the small (30 bp) tags, the fragments from three parents were mixed in equimolar amounts and ligated to make two half-length chimeric libraries. Digestion with *Nhe*I (restriction site designed into tag sequence) and *BsaXI* after ligation removed genes with residual tags from the final library. Each half-length library was PCR amplified, cloned back into the TOPO vector and transformed into E*scherichia coli* TOP10 cells (Invitrogen).

Table 5.3. Primers used for site-directed recombination of cytochromes P450 heme domains of A1, A2 and A3.

Tag	Site	Tag Sequence (5' to 3')	Primer (5' to 3')	
N	N-term	-	A123-Nf	GGAGGTACCAGATCT*ATCGATGCTTAGGAGGTCATATG*
Tag 1	64	gaaTGAAGTACT**AC**CTGTC**CTCC**GCTAGCTgaa	A1-64f	GTACTACCTGTCCTCCGCTAGCT*gaaTCACGCTTTGATAAAAACTTAAGT*
			A1-64r	AGCGGAGGACAGGTAGTACTTCA*ttcATCGCATGCTTCTTTAATTAGAC*
			A2-64f	GTACTACCTGTCCTCCGCTAGCT*gaaGAACGGTTTGATAAAAGCATTG*
			A2-64r	AGCGGAGGACAGGTAGTACTTCA*ttcATCACAAACCTCTTTCACCA*
			A3-64f	GTACTACCTGTCCTCCGCTAGCT*gaaAAACGCTTTGACAAGAACCT*
			A3-64r	AGCGGAGGACAGGTAGTACTTCA*ttcATCACACACTTCAGCCAC*
Tag 2	122	atcTAAAGGCCT**AC**ATACT**CTCC**GCTAGCAatc	A1-122f	GGCCTACATACTCTCCGCTAGCA*atcGCCGTGCAGCTTGTTC*
			A1-122r	AGCGGAGAGTATGTAGGCCTTTA*gatATCGACCATCATCGCATGAT*
			A2-122f	GGCCTACATACTCTCCGCTAGCA*atcGCTGTTCAGCTCATTCAAAAATG*
			A2-122r	AGCGGAGAGTATGTAGGCCTTTA*gatATCGACCATTTTCTCATGATAGTC*
			A3-122f	GGCCTACATACTCTCCGCTAGCA*atcGCAACCCAGCTGATTC*
			A3-122r	AGCGGAGAGTATGTAGGCCTTTA*gatATCCAGCATCATAGAATGATAGC*
Tag 3	166	tacTGATCGCGAA**AC**AATTA**CTCC**GCTAGCTtac	A1-166f	CGCGAACAATTACTCCGCTAGCT*tacCGAGATCAGCCTCATCC*
			A1-166r	AGCGGAGTAATTGTTCGCGATCA*gtaAAAGCTGTTAAAGCGATAGTTAAA*
			A2-166f	CGCGAACAATTACTCCGCTAGCT*tacAGAGAAACGCCCCACC*
			A2-166r	AGCGGAGTAATTGTTCGCGATCA*gtaGTAACTGTTAAAGCGGTAGTTAAA*
			A3-166f	CGCGAACAATTACTCCGCTAGCT*tacCGTGATTCACAGCATCC*
			A3-166r	AGCGGAGTAATTGTTCGCGATCA*gtaAAAGCTGTTGAATCGATAGTTAAA*
	216/217	-	A1-216f	CGAGGTACCgtcgac*AAAATTATTGCAGATCGCAAAGC*
			A1-216r	CCGCTCGAGgtcgac*TAGGTCGTTCATCACCTTG*
			A2-216f	CGAGGTACCgtcgac*AGCATTATTGCAGAG*
			A2-216r	CCGCTCGAGgtcgac*TAACGAAAACATCGTTTG*
			A3-216f	CGAGGTACCgtcgac*AGAATGATAGCGGAGCGAAAG*
			A3-216r	CCGCTCGAGgtcgac*CAGGGAGTTCATGACTTC*
Tag 4	268	acaTGAACGCGT**AC**TTATT**CTCC**GCTAGCTaca	A1-268f	CGCGTACTTATTCTCCGCTAGCT*acaACAAGTGGTCTTTTATCATTTGC*
			A1-268r	AGCGGAGAATAAGTACGCGTTCA*tgtTTCGTGTCCCGCAATTAAG*
			A2-268f	CGCGTACTTATTCTCCGCTAGCT*acaACGAGCGGCCTGCTTT*
			A2-268r	AGCGGAGAATAAGTACGCGTTCA*tgtTTCATGGCCGGCAATC*

			A3-268f	CGCGTACTTATTCTCCGCTAGCT*acaACAAGCGGGTTGCTATC*
			A3-268r	AGCGGAGAATAAGTACGCGTTCA*tgtCTCATGTCCAGCAATTAAAAATG*
Tag 5	328	gctTAAGTATAC**AC**CATCC**CTCC**GCTAGCAgct	A1-328f	TATACACCATCCCTCCGCTAGCA*gctCCTGCGTTTTCCCTATATG*
			A1-328r	AGCGGAGGGATGGTGTATACTTA*agcAGTTGGCCATAAGCGC*
			A2-328f	TATACACCATCCCTCCGCTAGCA*gctCCGGCTTTCAGCCTTTATC*
			A2-328r	AGCGGAGGGATGGTGTATACTTA*agcTGTCGGCCATAAGCGC*
			A3-328f	TATACACCATCCCTCCGCTAGCA*gctCCGGCTTTTTCTCTATATGC*
			A3-328r	AGCGGAGGGATGGTGTATACTTA*agcTGTTGGATACAGTCTGAGG*
Tag 6	404	cagTGATTCGAA**AC**CATTA**CTCC**GCTAGCTcag	A1-404f	TCGAAACCATTACTCCGCTAGCT*cagTTCGCTCTTCATGAAGC*
			A1-404r	AGCGGAGTAATGGTTTCGAATCA*ctgCTGACCGATACACGCAC*
			A2-404f	TCGAAACCATTACTCCGCTAGCT*cagTTTGCCCTTCATGAAGC*
			A2-404r	AGCGGAGTAATGGTTTCGAATCA*ctgCATGCCGATACAGGCC*
			A3-404f	TCGAAACCATTACTCCGCTAGCT*cagTTTGCTCTTCAAGAAGCG*
			A3-404r	AGCGGAGTAATGGTTTCGAATCA*ctgCATGCCAATACAAGCGC*
C	C-term	-	A1-Cr	CAACAATTG*TTAAGTGCTAGGTGAAGGAATACC*
			A2-Cr	CAACAATTG*TTATTCAGCTGCCTGGACGTC*
			A3-Cr	CAACAATTG*TTATGCCTGTTCTTTTCTCTGTAC*

Bold letters: *BsaX*I recognition site; underline: *Nhe*I recognition site; Lower-case letters: consensus sequence (sticky ends by *BsaX*I digestion); Italic letters: complementary to parental gene sequence; Shadowed letters: *Bgl*II or *Mfe*I recognition site of N-terminal or C-terminal primer, respectively.

The N- and C-terminal half-libraries were digested with *Bgl*II/*Sal*I, and *Sal*I/*Mfe*I, respectively, and cloned into the *BamH*I/*EcoR*I sites of the pCWori vector. The *Sal*I recognition site corresponds to crossover site 216-217 and was introduced into all three P450 genes synonymously. The N- and C- terminal libraries were ligated at this site to yield the full-length library, which was transformed into the catalase-deficient *E. coli* strain SN0037 [49].

E.3 Probe hybridization analysis

Probe hybridization was performed as described [33,50]. Twenty-four unique oligonucleotide probes were designed to specifically recognize each parent fragment with a T_m ~63°C (Table 5.4). Probes were labeled using the terminal transferase reaction with either fluorescein (Amersham) or digoxigenin (Roche) following the manufacturer's protocol. Anti-fluorescein (Amersham) or anti-digoxigenin (Roche) antibodies conjugated to alkaline phosphatase were used for chemiluminescent detection according to the manufacturer's protocol and recorded with a Gel-Doc system from BIORAD or BIOMAX films from Kodak.

Table 5.4. Sequences of probes for hybridization analysis.

Block	Parent	Sequence (5' to 3')
1	A1	GAC AAT TAA AGA AAT GCC TCA GCC AAA AAC GTT TGG
	A2	GAA GGA AAC AAG CCC GAT TCC TCA GCC
	A3	GAA ACA GGC AAG CGC AAT ACC TCA GCC
2	A1	GCG CAT AAT ATC TTA CTT CCA AGC TTC AGT CAG C
	A2	GCG CAC AAC ATT CTG ATG CCG ACG TTC
	A3	CC CAC CGC ATT TTG CTG CCG AGT TTT AG
3	A1	GAG CGT CTA AAT GCA GAT GAG CAT ATT GAA GTA CCG
	A2	GCA AGG CTC AAC CCG AAT GAA GCA GTC
	A3	G AGC CGG TTA AAC CCC AAT GAA GAA ATT GAT GTA G
4	A1	CCA GAC GAC CCA GCT TAT GAT GAA AAC AAG C
	A2	GTT CAA GAT AAG CTT ATG GTC AGA ACA AAG CGG C
	A3	CTG CAA GAT AAA ATG ATG GTG AAA ACG AAG CTG CAG
5	A1	GCA AGC GGT GAA CAA AGC GAT GAT TTA TTA ACG C
	A2	GCG AAT GGA GAC CAG GAT GAA AAA GAT TTG CTC
	A3	GCG AAT CCG GAT GAA AAC ATT AAG GAT CTC TTG TC
6	A1	GCA GCA CGA GTT CTA GTA GAT CCT GTT CC
	A2	GTC GAT CGG GTG CTG ACG GAT GCA G
	A3	GCG GAT CGC GTG TTA ACG GAT GAC AC
7	A1	CCA AGT GCG ATT CCG CAG CAT GCG
	A2	CAG GAC CAA GTG CCT CAT CAT GCG TAC
	A3	CCT TCA AGT ATC CCT CAC CAT GCG TAT AAG C
8	A1	ATC GAA AAA AAT TCC GCT TGG CGG TAT TCC TTC AC
	A2	GCC GTC ATC AGG AAG CCA TTC ATG CAG A
	A3	CGC GAA AAA CAG CGG CAA TCA ATG TAC AGA G

E.4 High-throughput carbon monoxide binding assay

Clones grown in 96-well plates were replicated into 500 µl of Luria-Bertani (LB) medium with 100 µg/ml ampicillin in 2 ml deep-well plates (Falcon, manufacturer location) and grown in a humidified shaker (Kuhner ISF-1-W, manufacturer location) for 20 h at 210 rpm, 30 °C and 80% relative humidity. Samples (150 µl) of these saturated cultures were transferred to 850 µl of terrific broth (TB) medium supplemented with 117

μg/ml ampicillin, 30 μg/ml thiamine, 0.6 mM δ-aminolevulinic acid, and 0.7 mM IPTG. These were grown for 24 h at 210 rpm, 25 °C and 80% relative humidity and harvested by centrifugation at 4 °C, 4,900 × g. Cell pellets were stored frozen at −20 °C until they were resuspended in 300 μl of lysis buffer (100 mM Tris [pH 8.2] with 0.5 mg/ml lysozyme and 2 units/ml DNAse) using a pipetting robot (Beckman Multimek 96, manufacturer location). Plates were incubated at room temperature for 1 h, followed by centrifugation at 4,900 × g for 10 min at 4 °C to clear the lysate.

CO binding assays were carried out as described in Chapter 2 with minor modification. Lysate (160 μl) was added to wells of a 96-well plate followed by addition of 40 μl of freshly-prepared 0.1 M sodium hydrosulfite (in 1.3 M phosphate buffer, pH 8.0). Using a Spectra Max Plus 384 plate reader (Molecular Devices), blank spectra were taken every 10 nm from 400 to 500 nm, as were readings for 450 and 490 nm. Plates were put in a vacuum oven and the atmosphere was removed (to ~350 mbar) and replaced with CO. Plates were allowed to incubate for 5 to 10 minutes, at which point the spectra and absorbencies at 450 and 490 nm were read. Assays were performed in duplicate.

Chimeras were assigned a folding status (folded P450 = 1, not folded = 0) based on the following. To be folded, the spectrum must have a distinct Soret peak within an absorbance range of -0.01 to 0.01 at 450 nm. In addition, the minimum CO difference value ($A_{450} - A_{490}$) must be ≥ 0.0015. The background, as assayed using cells with a null vector (pCWori with no insert), was $4.9 \times 10^{-4} \pm 4.2 \times 10^{-4}$, which set 1.5×10^{-3} as the minimum CO difference value giving greater than 95% confidence (greater than two

standard deviations). If the sequence was not assigned folded status, it was deemed 'not folded' = 0.

E.5 Functional assays

Chimeras were assayed for peroxygenase activity on 12-pNCA in 96-well plate format as described [51]. Reactions were carried out in a volume of 200 µl with 250 µM 12-pNCA and 20 mM H_2O_2 in 100 mM Tris (pH 8.2) at room temperature and monitored at 410 nm for 30 min for accumulation of 4-nitrophenol. Chimeras in wells with total product formation greater than 25% of the average of four control wells with the A1 heme domain after 30 min were considered active (corresponding to > 5 µM product).

Activity on 2-phenoxyethanol was assayed in 96-well plates using the 4-aminoantipyrine assay (4-AAP), which detects phenol-like compounds (see Chapter 3). Reactions were carried out in 120 µl with 1% DMSO, 1% acetone, 100 mM 2-phenoxyethanol and 20 mM H_2O_2 in 100 mM EPPS (pH 8.2). Reactions were mixed and left at room temperature without shaking for 2 h then quenched with 120 µl of 0.1 M NaOH and 4 M urea. Thirty-six µl of 0.6% 4-AAP was added, the 96-well plate reader was zeroed at 500 nm, and 36 µl of 0.6% potassium persulfate was added. After 20 min the A_{500} was read. Chimeras in wells with an A_{500} greater than 25% of the average of four control wells with the A1 heme domain were considered active, corresponding to > 20 µM product.

E.6 Thermostability

Thermostabilities (as described by T_m, the temperature at which half of the protein is unfolded) were measured using CO difference spectroscopy to monitor the disappearance of the Soret band with increasing temperature as described as described in Chapter 3 .

E.7 Logistic regression analysis

The significance of each block (intra-fragment) and block pair (inter-fragment) was calculated relative to a reference model with all eight blocks using the likelihood ratio test [21]. In the case of heme binding, this identified six potentially significant variables which were collected into a second-round reference model and reevaluated using the likelihood ratio test (Table 5.5). Blocks 1, 5, 7, and block pair 1–7 remained highly significant in the second round, whereas pairs 1–5 and 5–8 dropped in significance to $p > 10^{-3}$, a threshold established previously [21]. Cross-validation tests (data not shown) provide further evidence that only the variables 1, 5, 7, and 1–7 are significant. The same analysis was done for activity on 12-pNCA and determined blocks 2, 4 and 1–8 are significant.

Table 5.5. Significance of the top six variables identified by logistic regression analysis of cytochrome P450 chimeras.

Variable	p-value[1]
1	1×10^{-12}
5	3×10^{-8}
7	2×10^{-17}
1-7	3×10^{-12}
1-5	0.01
5-8	0.02

[1]p-values were calculated by removing each variable from the model and applying the likelihood ratio test (see Methods).

E.8 Construction and analysis of site-directed mutants

Single mutations were made in the A1 and A2 genes and in the genes of the eight chimeras seen in Table 5.2. The R47G and G48R mutations were made using the codon from the alternate parent, Arg (CGT) and Gly (GGC), respectively. The E435Q and Q437E mutations were made in the same fashion with the codons Glu (GAA) and Gln (CAA) being swapped. Mutants were constructed using PCR overlap extension mutagenesis [52], cloned into the BamHI/EcoRI site of pCWori and transformed into catalase-deficient *E. coli.* P450 chimeras and parents were cultured in 200 ml of TB medium and the initial rates on 12-pNCA were measured with 1 μM enzyme, 250 μM 12-pNCA, 1% DMSO, 20 mM H_2O_2 in 100 mM EPPS (pH 8.2), as described Chapter 4.

F. References

1. Lockless SW, Ranganathan R (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. Science 286: 295-299.
2. Saraf MC, Moore GL, Maranas CD (2003) Using multiple sequence correlation analysis to characterize functionally important protein regions. Protein Eng. 16: 397-406.
3. Larson SM, Di Nardo AA, Davidson AR (2000) Analysis of covariation in an SH3 domain sequence alignment: applications in tertiary contact prediction and the design of compensating hydrophobic core substitutions. J. Mol. Biol. 303: 433-446.
4. Moffet DA, Hecht MH (2001) De novo proteins from combinatorial libraries. Chem. Rev. 101: 3191-3203.
5. Arnold FH, Wintrode PL, Miyazaki K, Gershenson A (2001) How enzymes adapt: lessons from directed evolution. Trends Biochem. Sci. 26: 100-106.
6. Keefe AD, Szostak JW (2001) Functional proteins from a random-sequence library. Nature 410: 715-718.
7. Axe DD (2004) Estimating the prevalence of protein sequences adopting functional enzyme folds. J. Mol. Biol. 341: 1295-1315.
8. Bloom JD, Silberg JJ, Wilke CO, Drummond DA, Adami C, et al. (2005) Thermodynamic prediction of protein neutrality. Proc. Natl. Acad. Sci. U.S.A. 102: 606-611.
9. Guo HH, Choe J, Loeb LA (2004) Protein tolerance to random amino acid change. Proc. Natl. Acad. Sci. U.S.A. 101: 9205-9210.
10. Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH (1993) Protein design by binary patterning of polar and nonpolar amino acids. Science 262: 1680-1685.
11. Farinas ET, Bulter T, Arnold FH (2001) Directed enzyme evolution. Curr. Opin. Biotechnol. 12: 545-551.
12. Patel PH, Loeb LA (2000) DNA polymerase active site is highly mutable: evolutionary consequences. Proc. Natl. Acad. Sci. U.S.A. 97: 5095-5100.
13. Bolon DN, Voigt CA, Mayo SL (2002) De novo design of biocatalysts. Curr. Opin. Chem. Biol. 6: 125-129.
14. Dwyer MA, Looger LL, Hellinga HW (2004) Computational design of a biologically active enzyme. Science 304: 1967-1971.
15. Voigt CA, Martinez C, Wang ZG, Mayo SL, Arnold FH (2002) Protein building blocks preserved by recombination. Nat. Struct. Biol. 9: 553-558.
16. Meyer MM, Silberg JJ, Voigt CA, Endelman JB, Mayo SL, et al. (2003) Library analysis of SCHEMA-guided protein recombination. Protein Sci. 12: 1686-1693.
17. Silberg JJ, Endelman JB, Arnold FH (2004) SCHEMA-guided protein recombination. Methods Enzymol. 388: 35-42.
18. Drummond DA, Silberg JJ, Meyer MM, Wilke CO, Arnold FH (2005) On the conservative nature of intragenic recombination. Proc. Natl. Acad. Sci. U.S.A. 102: 5380-5385.
19. Lewis DFV (2001) Guide to cytochromes P450: structure and function. London; New York: Taylor & Francis.
20. Nelson D (2005) Homepage <http://drnelson.utmem.edu/CytochromeP450.html>.

21. Endelman JB, Bloom JD, Otey CR, Landwehr M, Arnold FH (2005) Inferring interactions from an alignment of folded and unfolded protein sequences. arXiv: q-bio.BM/0505018.
22. Munro AW, Leys DG, McLean KJ, Marshall KR, Ost TW, et al. (2002) P450 BM3: the very model of a modern flavocytochrome. Trends Biochem. Sci. 27: 250-257.
23. Gustafsson MC, Roitel O, Marshall KR, Noble MA, Chapman SK, et al. (2004) Expression, purification, and characterization of bacillus subtilis cytochromes P450 CYP102A2 and CYP102A3: flavocytochrome homologues of P450 BM3 from Bacillus megaterium. Biochemistry 43: 5474-5487.
24. Cirino PC, Arnold FH (2002) Regioselectivity and activity of cytochrome P450 BM-3 and mutant F87A in reactions driven by hydrogen peroxide. Adv. Synth. Catal. 344: 932-937.
25. Haines DC, Tomchick DR, Machius M, Peterson JA (2001) Pivotal role of water in the mechanism of P450BM-3. Biochemistry 40: 13456-13465.
26. Endelman JB, Silberg JJ, Wang ZG, Arnold FH (2004) Site-directed protein recombination as a shortest-path problem. Protein Eng. Des. Sel. 17: 589-594.
27. Endelman JB (2005) Design and analysis of combinatorial protein libraries created by site-directed recombination. Pasadena: California Institute of Technology. 122 p.
28. Ravichandran KG, Boddupalli SS, Hasermann CA, Peterson JA, Deisenhofer J (1993) Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. Science 261: 731-736.
29. Arnold GE, Ornstein RL (1997) Molecular dynamics study of time-correlated protein domain motions and molecular flexibility: cytochrome P450BM-3. Biophys. J. 73: 1147-1159.
30. Li H, Poulos TL (1999) Fatty acid metabolism, conformational change, and electron transfer in cytochrome P-450(BM-3). Biochim. Biophys. Acta. 1441: 141-149.
31. Li H, Poulos TL (1997) The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. Nat. Struct. Biol. 4: 140-146.
32. Hiraga K, Arnold FH (2003) General method for sequence-independent site-directed chimeragenesis. J. Mol. Biol. 330: 287-296.
33. Meinhold P, Joern JM, Silberg JJ (2003) Analysis of shuffled libraries by oligonucleotide probe hybridization. Methods Mol. Biol. 231: 177-187.
34. Wells AV, Li P, Champion PM, Martinis SA, Sligar SG (1992) Resonance Raman investigations of Escherichia coli-expressed Pseudomonas putida cytochrome P450 and P420. Biochemistry 31: 4384-4393.
35. Martinis SA, Blanke SR, Hager LP, Sligar SG, Hoa GH, et al. (1996) Probing the heme iron coordination structure of pressure-induced cytochrome P420cam. Biochemistry 35: 14530-14536.
36. Abecassis V, Pompon D, Truan G (2000) High efficiency family shuffling based on multi-step PCR and in vivo DNA recombination in yeast: statistical and functional analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2. Nucleic Acids Res. 28: E88.
37. Salazar O, Cirino PC, Arnold FH (2003) Thermostabilization of a cytochrome p450 peroxygenase. Chembiochem 4: 891-893.

38. Brock BJ, Waterman MR (2000) The use of random chimeragenesis to study structure/function properties of rat and human P450c17. Arch. Biochem. Biophys. 373: 401-408.
39. Kronbach T, Larabee TM, Johnson EF (1989) Hybrid cytochromes P-450 identify a substrate binding domain in P-450IIC5 and P-450IIC4. Proc. Natl. Acad. Sci. U.S.A. 86: 8262-8265.
40. Kass I, Horovitz A (2002) Mapping pathways of allosteric communication in GroEL by analysis of correlated mutations. Proteins 48: 611-617.
41. Fodor AA, Aldrich RW (2004) Influence of conservation on calculations of amino acid covariance in multiple sequence alignments. Proteins 56: 211-221.
42. Suel GM, Lockless SW, Wall MA, Ranganathan R (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat. Struct. Biol. 10: 59-69.
43. Saraf MC, Horswill AR, Benkovic SJ, Maranas CD (2004) FamClash: a method for ranking the activity of engineered enzymes. Proc. Natl. Acad. Sci. U.S.A. 101: 4142-4147.
44. Noble MA, Miles CS, Chapman SK, Lysek DA, MacKay AC, et al. (1999) Roles of key active-site residues in flavocytochrome P450 BM3. Biochem. J. 339 (Pt 2): 371-379.
45. Graham-Lorence S, Truan G, Peterson JA, Falck JR, Wei S, et al. (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: 1127-1135.
46. Carmichael AB, Wong LL (2001) Protein engineering of Bacillus megaterium CYP102.The oxidation of polycyclic aromatic hydrocarbons. Eur. J. Biochem. 268: 3117-3125.
47. Thompson JD, Higgins DG, Gibson TJ (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: 4673-4680.
48. Barnes HJ, Arlotto MP, Waterman MR (1991) Expression and enzymatic activity of recombinant cytochrome P450 17 alpha-hydroxylase in Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. 88: 5597-5601.
49. Nakagawa S, Ishino S, Teshiba S (1996) Construction of catalase deficient Escherichia coli strains for the production of uricase. Biosci. Biotechnol. Biochem. 60: 415-420.
50. Joern JM, Meinhold P, Arnold FH (2002) Analysis of shuffled gene libraries. J. Mol. Biol. 316: 643-656.
51. Cirino PC, Arnold FH (2003) A self-sufficient peroxide-driven hydroxylation biocatalyst. Angew Chem Int Ed Engl 42: 3299-3301.
52. Higuchi R, Krummel B, Saiki RK (1988) A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucleic Acids Res. 16: 7351-7367.

Chapter 6

Functional analysis of an artificial family of P450s

This work was contributed to directly by Marco Landwehr. I would also like to thank Yougen Li and Lisa Hochrein for experimental assistance. This work was supported by NIH Grant R01 GM068664-01.

A. Abstract

The limited number of protein folds found in Nature requires that one fold serve many functional roles for its host. In fact, certain protein folds appear to be favored as they dominate known protein structure space. These protein superfamilies have diversified over millions of years to yield numerous diverse sequences capable of variety of functions. We have created an artificial protein family containing thousands of new, diverse P450s. Here, we analyze the functional diversity present in this family on a variety of substrates. We show that chimeras are able to exceed the parental peroxygenase activities as well as acquire new ones on a set of eight substrates. K-means clustering analysis of a set of 37 P450s shoed they cluster into five groups based on their activity profiles. Additionally, a large data set of 330 chimeras is assembled for further analysis.

B. Introduction

The amino acid sequence of a protein determines its structure, but that structure does not restrict its biological function. This is evident in Nature were similar protein structures perform a variety of biological functions. In fact, the number of protein folds is thought to be limited to only a few thousand [1,2] demonstrating that protein structure changes slowly over evolutionary time while amino acid sequences changs more quickly. The utilization of a similar protein structure for multiple functions is well demonstrated by the $(\beta\alpha)_8$ (or TIM) barrel fold which has been recruited to fulfill numerous biological roles such as glycosidases, synthetases as well as binding a wide range of substrate [3,4]. The ubiquitous thioredoxins also share a common protein structure and use sulfur redox chemistry to carry out transferase, peroxidase and isomerase reactions to name a few[5]. Numerous examples of these protein superfamilies exist [6].

Cytochromes P450 comprise a superfamily of heme enzymes with myriad biological functions, including key roles in drug metabolism, breakdown of xenobiotics, and steroid and secondary metabolite biosynthesis [7]. P450s are postulated to have first functioned in the detoxification of oxygen or peroxide [8,9]. From this ancestral P450, evolution has created the plethora of sequences that exist today. Cytochromes P450 tend to share high levels of structural similarity despite their low levels of sequence identity [10-12]. This high level of sequence diversity allows for the various functions present in this enzyme scaffold.

In the previous chapter, we described the creation of an artificial family of cytochromes P450 by recombining three bacterial P450s, A1, A2 and A3 using structure-guided recombination guided by SCHEMA. SCHEMA is an energy function that an allows us to minimize the number of structural contacts that are disrupted when portions of the sequence are inherited from different parents, further increasing the probability that the chimeric proteins will fold. This library contained roughly 3,000 new P450 sequences which had 72 amino sequence changes relative to the closest natural P450 on average. Analysis of these new sequences identified key structural contributions to correct heme binding and peroxygenase activity. Here we begin to analyze the functional diversity present within these artificially created P450s.

C. Results

C.1 Characterization of functional diversity

Using the artificial family of cytochromes P450 described in Chapter 5, we randomly selected 640 chimeras that were able to correctly bind heme as characterized by a Soret band at approximately 450 nm using high-throughput CO difference spectroscopy (see Chapter 3). This set of chimeras was assayed in high-throughput on a total of 47 different substrates (Appendix C and 12-pNCA) that range in size from single aromatic rings to larger substrates with multiple rings and numerous side groups. Of these 47, chimeras showed detectable activity using the 4-AAP assay on 18 substrates. A final set of eight substrates was selected based on chemical diversity and having activity levels high enough for repeatabilty in high throughput. For example, 11-phenoxyundecanoic acid was not used since its structure is very similar to 12-pNCA. Molecules such as

zoxazolamine, chlorzoxazone, tolbutamide and allyloxybenzene were not used do to poor reproducability from low activity or high background signals in high throughput. The final set of eight substrates is shown in Figure 7.1.

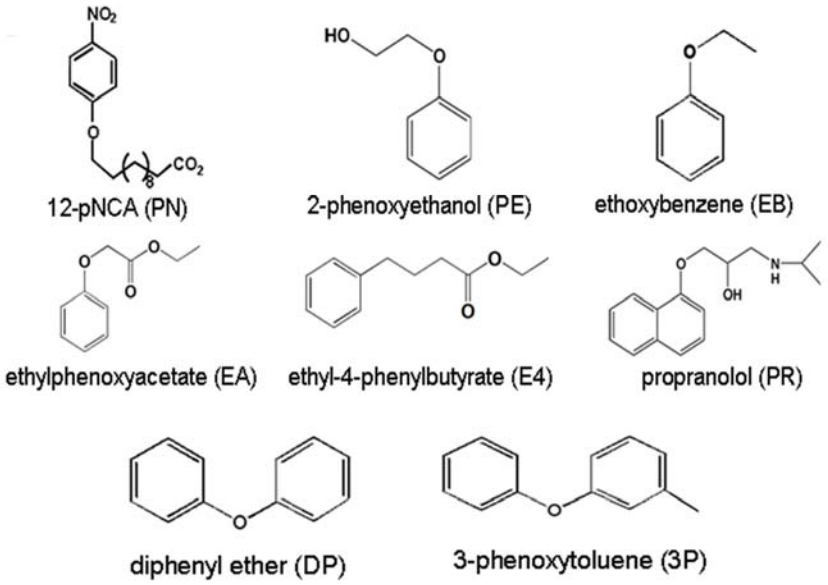

Figure 6.1 Set of eight substrates for characterization of chimeric P450s. Abbreviations are shown in parentheses.

We first determined the activity profile for each of the three parent P450s on these eight substrates (Figure 6.2). Two μM P450 was combined with substrate (Table 6.1) and 20 mM H_2O_2 and allowed to react for 2 hours. After 2 hours the reactions were quenched and the hydroxylated aromatic products were detected using the 4-AAP assay (Chapter 3). Activity on 12-pNCA was assayed by monitoring the accumulation of 4-nitrophenol at 410 nm due to hydroxylation forming a hemiacetal [13]. The parents showed detectable activity on six of the eight substrates. As expected from work in previous chapters, A1 was most or equally active on the six substrates compared to the other two parents. A1 and A2 were active on the same six substrates while A3 was only active on four. None of the parents showed activity on diphenyl ether or 3-phenoxytoluene.

Table 6.1 Substrate concentrations used in bioconversions.

Substrate	Concentration (mM)
12-pNCA	0.25
2-phenoxyethanol	100
ethoxybenzene	50
ethylphenoxyacetate	10
ethyl-4-phenylbutyrate	5
propranolol	4
diphenyl ether	10
3-phenoxytoluene	10

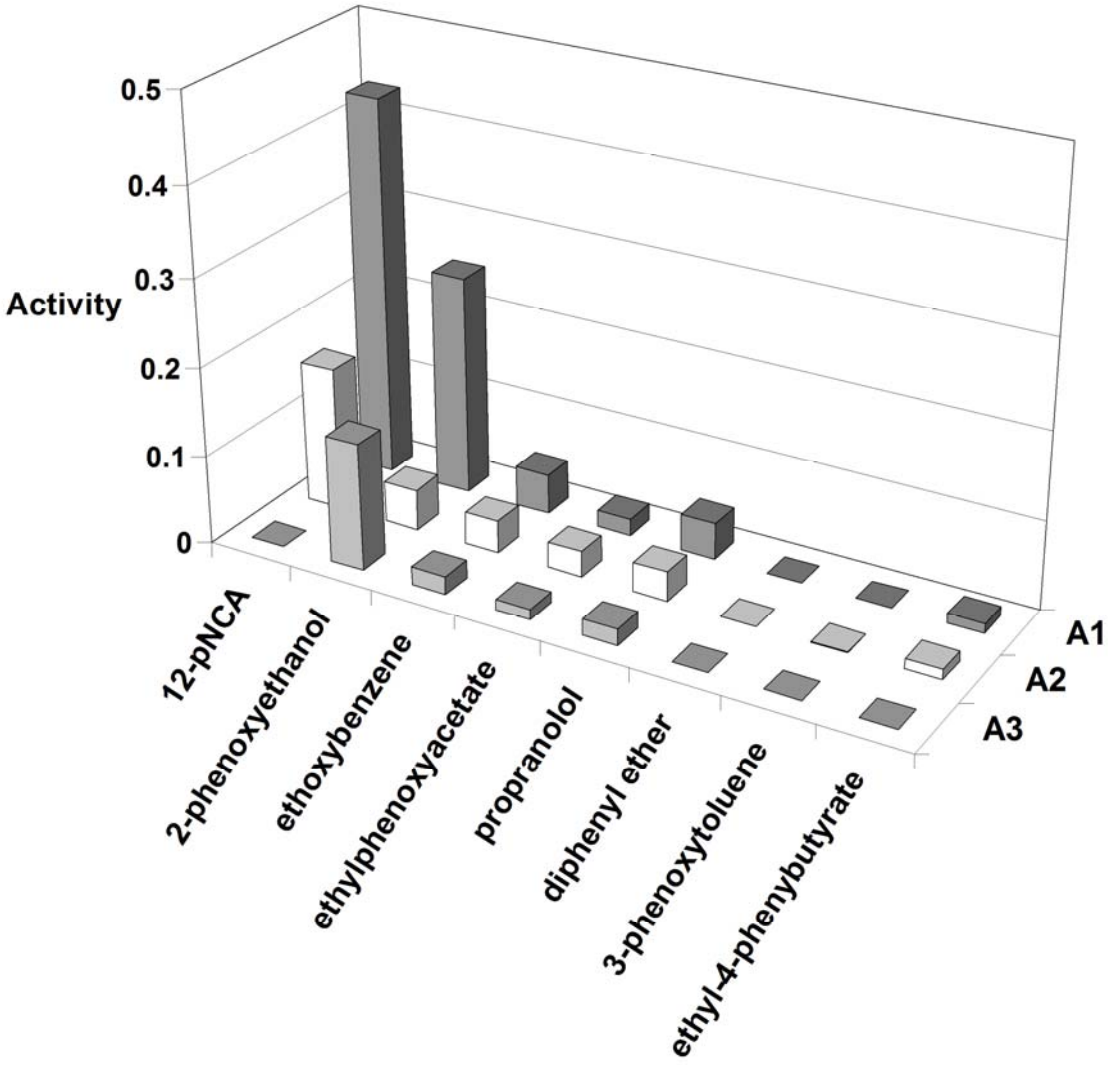

Figure 6.2 Peroxygenase activity of parental P450s A1, A2 and A3 on eight substrates using the 4-AAP assay and 12-pNCA. Activity was measured as the absorbance at 500 nm minus the background absorbance when the reaction was pre-quenched. One standard deviation for any of the measurements was less than 30%.

This analysis was then extended to thirty-four chimeric P450s from the set of 640 mentioned above. These folded P450 chimeras were assayed in the same way as the parents A1, A2 and A3. We found numerous chimeras with activity on the substrates not accepted by the parents (Figure 6.3). We also identified chimeras with greater activity than any of the parents on all of the substrates on which the parents had activity. Chimeras were able to exeed total activity of the most active parent, A1, on all eight substrates. Chimera 112122112 (see naming convention from Chapter 5) is more active than A1 on both 12-pNCA and propranolol. Chimera 21313311 is more active on 2-phenoxyethanol, ethyl-4-phenylbutyrate, ethoxybenzene and ethylphenoxyacetate. Chimeras also acquire activiy on diphenyl ether and 3-phenoxytoluene, substrates for which the parents showed no activity. Chmiera 32313233 is the most active chimera on diphenyl ether while 21313311 is most active on 3-phenoxytoluene. Chimera 32313231 is an example of a chimera that acquired activity on diphenyl ether and 3-phenoxytoluene without a large increase on all other substrates such as 2-phenoxyethanol. Chimeras 23312123 and 23132111 (not shown) are folded P450s that show no catalytic activity on any of the substrates tested.

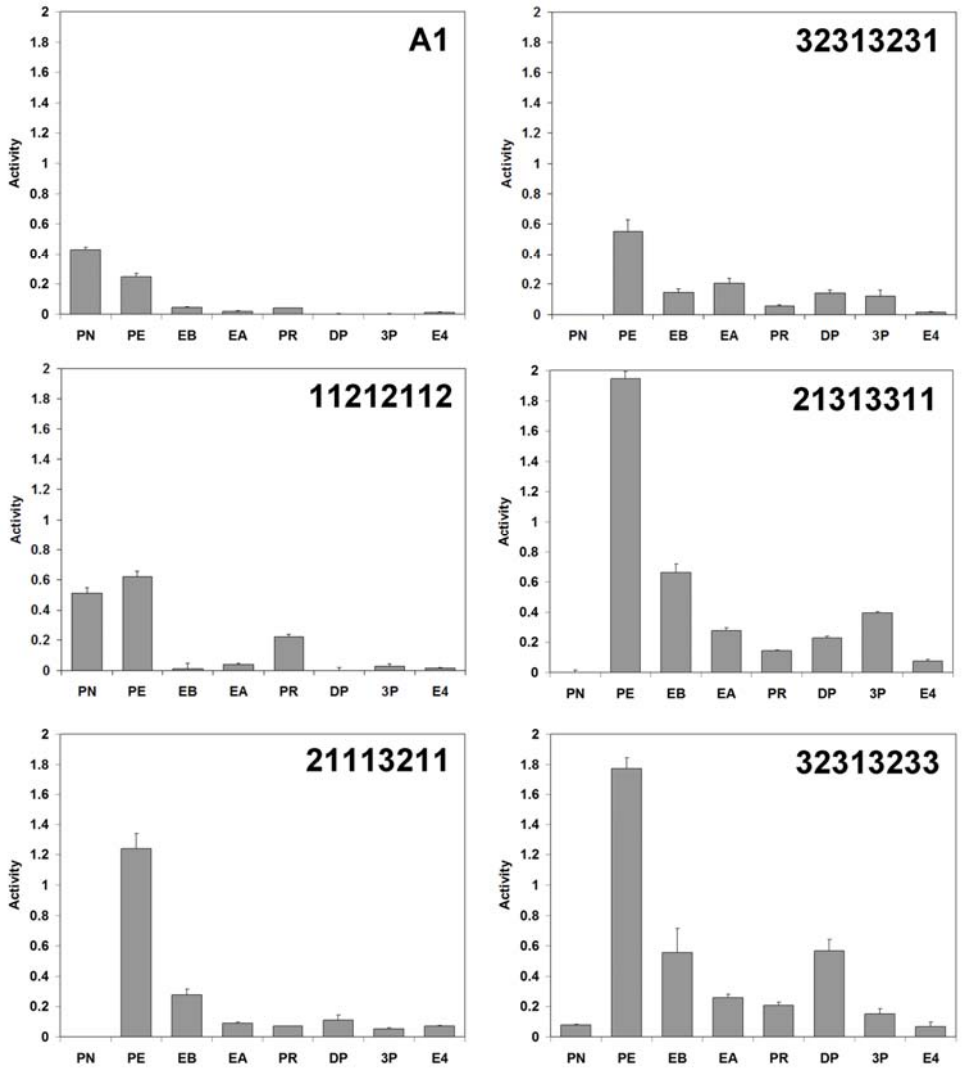

Figure 6.3 Select chimeras showing altered activity profiles and enhanced activities.

C.2 K-means clustering

K-means clustering can be used to determine how the expression of genes or enzymes with different activity profiles group together [14,15]. The activities of the 34 chimeras and three parent P450s were normalized to the maximum value for each substrate to avoid overweighting more active substrates. We used Cluster 3.0 [16] to perform K-means clustering on chimeras based on their activity profiles. The total number of groups

was increased until groups began containing only one member, which occurred with six groups. Silhouette values can be calculated to determine optimal group numbers based on their distance from other groups but that was not done here. We selected the five activity groups shown in Figure 6.4. Group A is characterized by high activity on 12-pNCA and propranolol. Group B is active on 2-phenoxyethanol, ethylphenoxyacetate, diphenyl ether and 3-phenoxytoluene. Group C is active on 2-phenoxyethanol and ethyl-4-phenylbutyrate. Group D has generally low activity with some members being active on 2-phenoxyethanol, ethylphenoxyacetate and propranolol. These groups demonstrate some of the functional diversity present in this artificial protein family of cytochromes P450.

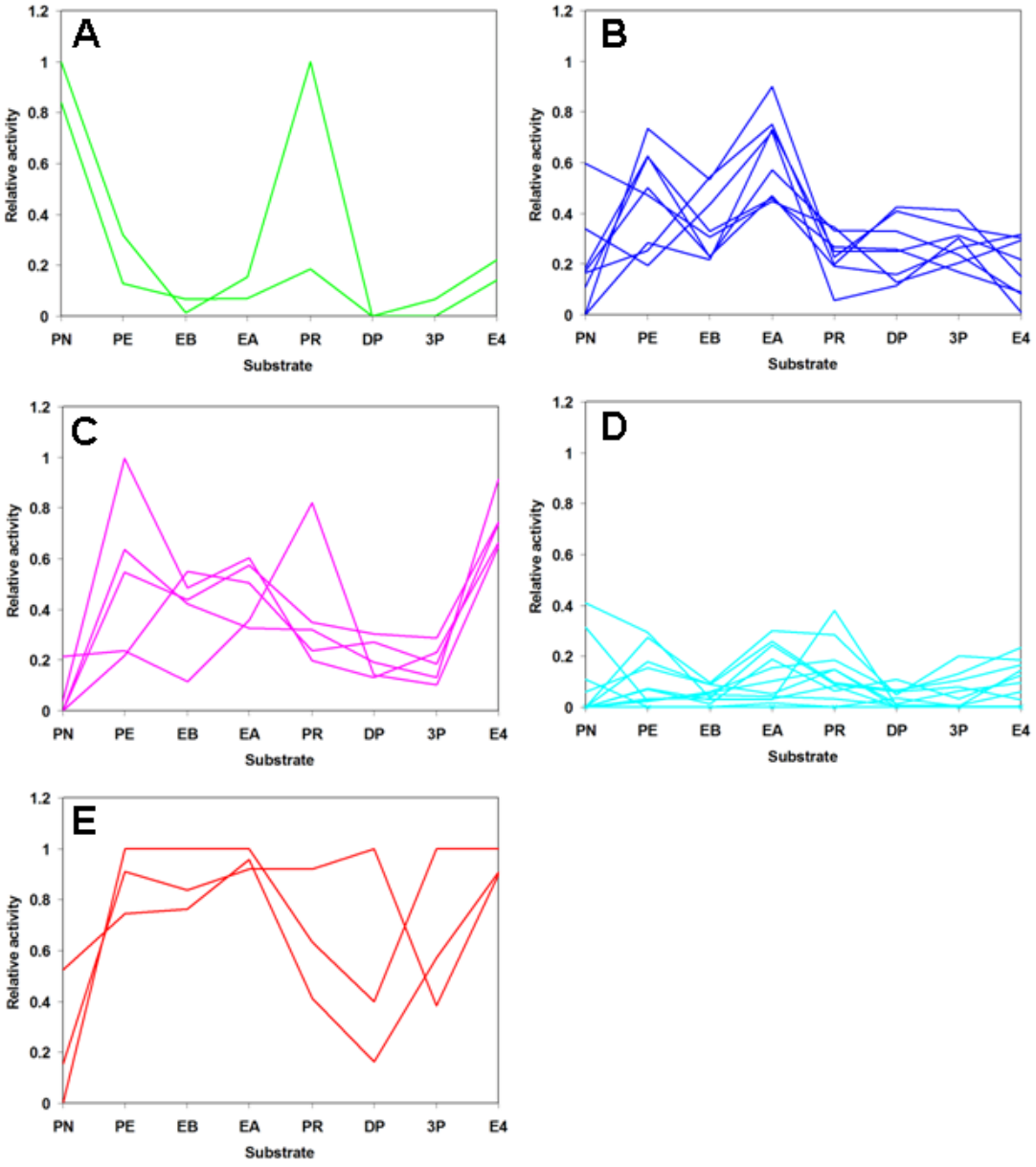

Figure 6.4 K-means clustering of the acitivy profiles 37 P450 peroxygenase on eight substrates cluster them into five groups. Relative activity represents the activity on each substrate normalized to the highest activity of all the P450s tested using the 4-AAP assay. The chimeras are assigned to each group as follows: A) A1, 11212112; B) 21113212, 12113221, 11113311, 32313231, 22213132, 32111333, 32312333, 32312231; C) 12212212, 21113211, 22233211, 21313112, 21313111; D) A1, A2, 22113232, 23312123, 23132233, 22312132, 22132231, 11332212, 23213311, 23132111, 21132112 22113111, 23312112; E) 21113312, 32313233, 21313311.

C.3 High-throughput activity measurements

To assemble a larger data set of hundreds of chimera activity profiles we measured the activity of 640 folded chimeras in high-throughput. The sequences of many of these are known from DNA probe hybridization described in Chapter 5 [17]. Measurements on each substrate were done at least in triplicate, and if a chimera had less than 500 nM P450 in a given assay it was removed to eliminate false negatives due to low expression or experimental errors. If this resulted in fewer than two measurements for a chimera on a substrate, this chimera was discarded. After removing chimeras with too little data or without sequences, 330 unique chimeric P450 sequences remained with data on all eight substrates. For each of these chimeras there are activity measurements and CO difference measurement which reflect the enzyme concentration. Control wells containing the parents A1, A2 and A3 as well as a null vector (pCWori with truncated CYP102A1) were included on each microtiter plate. The null vector wells serve to determine the background absorbance level.

To demonstrate the validity of the high-throughput data, activity values were determined in two ways. In method A, we normalized the raw activity absorbance to the average of control wells containing A1 (Equation 5):

$$activity(A) = \frac{ABS_{chimera} - \langle ABS_{null} \rangle}{\langle ABS_{CYP102A1} \rangle} \qquad (5),$$

were ABS = activity absorbance, null = null vector, and EXPR = CO difference value ($Abs_{450} - Abs_{490}$). The final activity value was the average of two greatest values. The mean percent standard deviation was calculated for each substrate by taking the average

of the standard deviation for the pairs of measurements (Table 6.2). In method B, the average background measurement (from the null vector) was subtracted from the activity value for absorbance and normalized to the average of control wells containing A1 as well as the CO difference value for that chimera:

$$activity(B) = \frac{ABS_{chimera} - <ABS_{null}>}{(EXPR_{chimera})(<ABS_{CYP102\,A1}>)} \qquad (6).$$

This effectively removes the background absorbance and controls for expression. However, it also leads to a much higher mean standard deviation across each substrate set (Table 6.2). This is due to the chimeras with absorbencies near the background level giving artificially high standard deviations. To remedy this, those with activities less than 10% of the most active chimera were removed and the mean percent standard deviation was recalculated (Table 6.2). This reduced the mean standard deviation to a level similar to that of Method A and eliminates chimeras with low activities.

Table 6.2 Mean percent standard deviations for each substrate across data set of 330 active chimeras.

	Mean percent standard deviation (%)		
Substrate	Method A[a]	Method B[b]	Method B with removal[c]
12-pNCA	31.1	105.1	22.0
2-phenoxyethanol	25.4	27.8	22.4
ethoxybenzene	23.9	48.0	22.2
ethylphenoxyacetate	20.0	158.6	33.8
propranolol	40.3	117.5	42.9
diphenyl ether	26.8	162.2	27.6
3-phenoxytoluene	18.5	208.6	40.0
ethyl-4-phenylbutyrate	13.9	170.1	37.6

[a]Method A uses the raw activity value and normalizes it the average A1 activity value. See Results.

[b]The average background measurement (from the null vector) was subtracted from the activity value for absorbance and normalized to the average of control wells containing A1 as well as the CO difference value for that chimera. See Results.

[c]Data from method be with values less than 10% of most active chimera removed. See Results.

D. Discussion

Many natural enzymes are able to perform a variety of functions using the same protein fold. Here we show that by diversifying protein sequence using site-directed recombination and retaining the ability to fold into a cytochrome P450, we can diversify the activity of hundreds of new sequences on a set of eight substrates (Figure 6.1). Some chimeras are more active than any of the parent sequences, A1, A2 and A3, and others are even able to accept diphenyl ether and 3-phenoxytoluene, substrates that the parents do not accept. Additionally, K-means clustering analysis shows that chimeras may be clustered into five groups based on their activity profiles. This initial analysis shows that

functional diversity exists in our artificial family of cytochromes P450 generated using site-directed protein recombination in the absence of external mutations.

This functional information can also be used to derive information about the parental enzymes. This is similar to ‘enzyme fingerprinting’ where understanding the different substrates an enzyme accepts can assist in developing surrogate substrates, examining enzyme specificity and understanding enzyme mechanisms [18-21]. The larger data set discussed here will be very useful to these ends. The data set appears to be accurate due to its acceptable level of mean percent standard deviation and may be further improved by removal of chimeras with high variance or through the use of additional processing methods. Analysis of these data may reveal a number of things. First, they can be analyzed to pinpoint which sequence blocks contribute to specific activities. This would best be done using linear regression analysis similar to that done previously [22]. Data on a large number of these chimeras can be used to learn which regions of the sequence contribute to altered activity profiles similar to how logistic regression was used to learn about what contributed to the overall structure in Chapter 4. This can be achieved using hierarchical or k-means clustering or linear regression [15,23-25]. Methods such as principal component analysis can be used to extract how different substrates may covary together and serve as surrogates for one another.

The new and diverse functions present in this library are not limited to the substrates tested here. For example, in the next chapter we discuss how they are able to accept the substrates of human P450s and produce authentic human metabolites. As new substrates

for which a cytochrome P450 is needed, an artificial family such as this can be used in conjunction with natural enzymes to search for the desired function. This methodology may lead to these enzymes use in drug discovery or chemical synthesis [26].

E. Materials and Methods

E.1 Materials

All chemicals were purchased from Sigma-Aldrich, Inc. (St. Louis, MO) or MP Biomedicals, Inc. (Solon, OH). Solvents were from EM Sciences (Gibbstown, NJ).

E.2 Testing of individual chimeras

Cleared cell lysates of individual chimeras were prepared as in Chapter 3. Enzyme concentrations were determined using CO binding difference spectroscopy with an extinction coefficient of 91 mM^{-1} cm^{-1} [27]. 4-AAP reactions were carried out in 120 µl with 2 µM P450, 1% dimethyl sulfoxide, 1% acetone and substrate concentrations according to Table 7.1 in 100 mM EPPS (pH 8.2). Reactions were initiated by the addition of H_2O_2 to a final concentration of 20 mM. After two hours at room temperature, reactions were quenched by addition of 120 µl of 4 M urea and 100 mM NaOH. Additionally, identical reactions were carried out in which the reaction was quenched prior to the addition of H_2O_2 in order to establish the background level in the presence of no reaction. To these quenched reactions, 36 µl of 0.6 % 4-AAP was added and the spectrophotometer was zeroed. Thirty-six µl of 0.6% potassium persulfate was added and the A_{500} was read after ten minutes. Each measurement was done in triplicate and the total activity was determined by subtracting the pre-quenched point from the two hour

point. Assays on 12-pNCA were done with 1 μM P450 with 2% DMSO in 100 mM EPPS (pH 8.2). Before addition of H_2O_2 the spectrophotometer was zeroed at 410 nm. H_2O_2 was then added to a final concentration of 20 mM and after 30 minutes the A_{410} was read.

E.3 High-throughput functional assays

Chimeras were cultured and prepared in 96-well plate format as described in Chapter 5. 12-pNCA assays were conducted as previously. For the other substrates, 80 μl of cell lysate was added to a 96-well plate and brought to a final volume of 120 μl containing substrate (Table 7.1) with 1% dimethyl sulfoxide (DMSO) and 1% acetone. Reactions were initiated with addition of H_2O_2 to a final concentration and after two hours they were quenched with 120 μl quench (0.1 M NaOH and 4 M urea). To these quenched reactions, 36 μl of 0.6 % 4-AAP was added and the spectrophotometer was zeroed at 500 nm. Thirty-six μl of 0.6% potassium persulfate was added and the absorbance at 500 nm was read after ten minutes. The concentration of P450 was also monitored in each assay using high-throughput CO-difference spectroscopy (see Chapter 3).

E.4 K-means clustering analysis

K-means clustering was done using Cluster 3.0 [16]. Activity on each substrate for the set of 34 chimeras and three parents was normalized to the highest activity. We used 20,000 runs to determine optimal groups.

F. References

1. Chothia C (1992) Proteins. One thousand families for the molecular biologist. Nature 357: 543-534.
2. Orengo CA, Jones DT, Thornton JM (1994) Protein superfamilies and domain superfolds. Nature 372: 631-634.
3. Gerlt JA, Raushel FM (2003) Evolution of function in (beta/alpha)8-barrel enzymes. Curr. Opin. Chem. Biol. 7: 252-264.
4. Nagano N, Orengo CA, Thornton JM (2002) One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. J. Mol. Biol. 321: 741-765.
5. Arner ES, Holmgren A (2000) Physiological functions of thioredoxin and thioredoxin reductase. Eur. J. Biochem. 267: 6102-6109.
6. Todd AE, Orengo CA, Thornton JM (2001) Evolution of function in protein superfamilies, from a structural perspective. J. Mol. Biol. 307: 1113-1143.
7. Lewis DFV (2001) Guide to cytochromes P450: structure and function. London; New York: Taylor & Francis.
8. Gotoh O (1993) Evolution and Differentiation of P-450 Genes. In: Omura T, Ishimura Y, Fujii-Kuriyama Y, editors. Cytochrome P450. 2nd ed. Tokyo: Kodansha. pp. 255-272.
9. Wickramasinghe RH, Villee CA (1975) Early role during chemical evolution for cytochrome P450 in oxygen detoxification. Nature 256: 509-511.
10. Hasemann CA, Kurumbail RG, Boddupalli SS, Peterson JA, Deisenhofer J (1995) Structure and function of cytochromes P450: a comparative analysis of three crystal structures. Structure 3: 41-62.
11. Mestres J (2005) Structure conservation in cytochromes P450. Proteins 58: 596-609.
12. Peterson JA, Graham SE (1998) A close family resemblance: the importance of structure in understanding cytochromes P450. Structure 6: 1079-1085.
13. Schwaneberg U, Schmidt-Dannert C, Schmitt J, Schmid RD (1999) A continuous spectrophotometric assay for P450 BM-3, a fatty acid hydroxylating enzyme, and its mutant F87A. Anal. Biochem. 269: 359-366.
14. Qin L, Rueda L, Ali A, Ngom A (2005) Spot detection and image segmentation in DNA microarray data. Appl. Bioinformatics 4: 1-11.
15. Kurtovic S, Runarsdottir A, Emren LO, Larsson AK, Mannervik B (2007) Multivariate-activity mining for molecular quasi-species in a glutathione transferase mutant library. Protein Eng. Des. Sel. 20: 243-256.
16. de Hoon M, Imoto S, Miyano S (2002) Cluster 3.0 (http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/software.htm#ctv). Tokyo.
17. Joern JM, Meinhold P, Arnold FH (2002) Analysis of shuffled gene libraries. J. Mol. Biol. 316: 643-656.
18. Grognux J, Reymond JL (2004) Classifying enzymes from selectivity fingerprints. Chembiochem 5: 826-831.
19. Srinivasan R, Huang X, Ng SL, Yao SQ (2006) Activity-based fingerprinting of proteases. Chembiochem 7: 32-36.
20. Goddard JP, Reymond JL (2004) Enzyme activity fingerprinting with substrate cocktails. J. Am. Chem. Soc. 126: 11116-11117.

21. Greenbaum DC, Arnold WD, Lu F, Hayrapetian L, Baruch A, et al. (2002) Small molecule affinity fingerprinting. A tool for enzyme family subclassification, target identification, and inhibitor design. Chem. Biol. 9: 1085-1094.
22. Li Y, Drummond DA, Sawayama AM, Snow CD, Bloom JD, et al. (2007) A diverse family of thermostable cytochrome P450s created by recombination of stabilizing fragments. Nat. Biotechnol. 25: 1051-1056.
23. Landwehr M, Carbone M, Otey CR, Li Y, Arnold FH (2007) Diversification of catalytic function in a synthetic family of chimeric cytochrome p450s. Chem. Biol. 14: 269-278.
24. Joern JM (2003) Engineering dioxygenases by laboratory evolution: a comparision of evolutionary search strategies. Pasadena: California Institute of Technology. 233 p.
25. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25: 25-29.
26. Guengerich FP (2002) Cytochrome P450 enzymes in the generation of commercial products. Nat. Rev. Drug Discov. 1: 359-366.
27. Omura T, Sato R (1964) The carbon monoxide-binding pigment of liver microsomes. II. Solubilization, purification, and properties. J. Biol. Chem. 239: 2379-2385.

Chapter 7

Chimeric P450s produce authentic human metabolites

I would like to acknowledge James Ross for direct contributions to this work. Additionally I would like to thank Lisa Hochrein for assistance with HPLC and Marco Landwehr for enzyme preparation. This work was supported by NIH Grant R01 GM068664-01.

A. Abstract

One of the main functions of P450s in humans is the processing and breakdown of drug compounds and other xenobiotics. The products of these reactions can have various biological properties including being more active drugs, toxic compounds or prepared for degradation. Here we show that chimeric bacterial P450s are capable of accepting human drug compounds as substrates and are able to produce authentic human metabolites of propranolol, chlorzoxazone, zoxazolamine. Chimeric P450s also accepted two additional substrates from human P450s, the drug tolbutamide and the carcinogen 4-aminobiphenyl. The product of the tolbutamide reaction was not any of the known human metabolites and the product of the 4-aminobiphenyl reaction was the same as that produced by horse radish peroxides (HRP). This demonstrates the utility of chimeric P450s created by site-directed recombination in the hydroxylation of valuable drug-like compounds.

B. Introduction

Cytochromes P450 from humans drove early research in the field due to their importance in drug metabolism and steroid biosynthesis [1]. Not long after these initial studies, it was discovered that P450s were plentiful in lower organisms as well and these P450s had very similar mechanistic properties to their human counterparts. These P450s have since proven extremely valuable as model systems for understanding the details of the enzymes structure and mechanism [2]. Bacterial P450s, for example, are easier to work with in that they are soluble, highly-active on their natural substrates and easily expressed at high levels. In contrast, human P450s are generally membrane associated, less stable and suffer from low activity. One property that is missing, however, is similar substrate specificities whereby bacterial P450s could accept the substrates of the human enzymes. The ability of a bacterial variant to produce the same chemicals as a human P450 has great value in drug discovery and testing [3].

Members of the artificial family of cytochromes P450 discussed in previous chapters were tested for their activity on five substrates of human P450s, four of which are drug molecules. We show that bacterial P450 chimeras are able to produce many of the same products as the human enzymes and in some cases not by any of the parental enzymes, A1, A2 and A3. Additionally, P450 chimeras are potentially able to produce novel metabolites not produced by the human P450s as well as form a conjugate product using a peroxidase mechanism like that in horse radish peroxidase (HRP).

C. Results

C.1 4-AAP analysis of drug-like compounds

The chimeras individually assayed in Chapter 6 were also assayed on chlorzoxazone, tolbutamide, zoxazolamine, propranolol and 4-aminobiphenyl. The 4-AAP assay was used since it should detect any products in which the aromatic rings of these substrates were hydroxylated. Chlorzoxazone, zoxazolamine, 4-aminobiphenyl, propranolol and tolbutamide are substrates of the mammalian P450s CYP2E1, CYP1A2, CYP2D6 and CYP2C9 respectively [4]. The main human metabolite of chlorzoxazone made by CYP2E1 has been shown to be 6-hydroxychlorzoxazone [5] and the metabolites of tolbutamide made by CYP2C9 have been shown to be 4-hydroxytolbutamide and carboxytolbutamide [6]. Zoxazolamine is converted to 6-hydroxyzoxazolamine [4] and 4-aminobiphenyl is converted to n-hydroxyaminobiphenl by CYP1A2 [7]. Propranolol is converted to 4'-hydroxypropranolol (4'OHP), 5'-hydroxypropranolol (5'OHP) and desisopropylpropranolol by CYP2D6 (DIP) [8].

The four chimeric P450s which showed the highest levels of activity on these substrates in the 4-AAP assay were used for further study. These reactions were driven by 20 mM H_2O_2 using the peroxygenase mechanism [9]. The activities of these four chimeras and the three parental enzymes measured with the 4-AAP assay can be seen in Figure 7.1. Most notably, the parental P450s A1 and A2 show low activity on propranolol and chlorzoxazone and no detectable activity on tolbutamide or zoxazolamine. Parent A3 only shows activity on propranolol. Interestingly, the chimeras do not just gain activity on all substrates but appear to prefer specific substrates. For example, chimera 22213132

has high activity on chlorzoxazone but low activity on the other three substrates. In contrast, chimera 11212112 has no detectable activity on chlorzoxazone but high activity on propranolol.

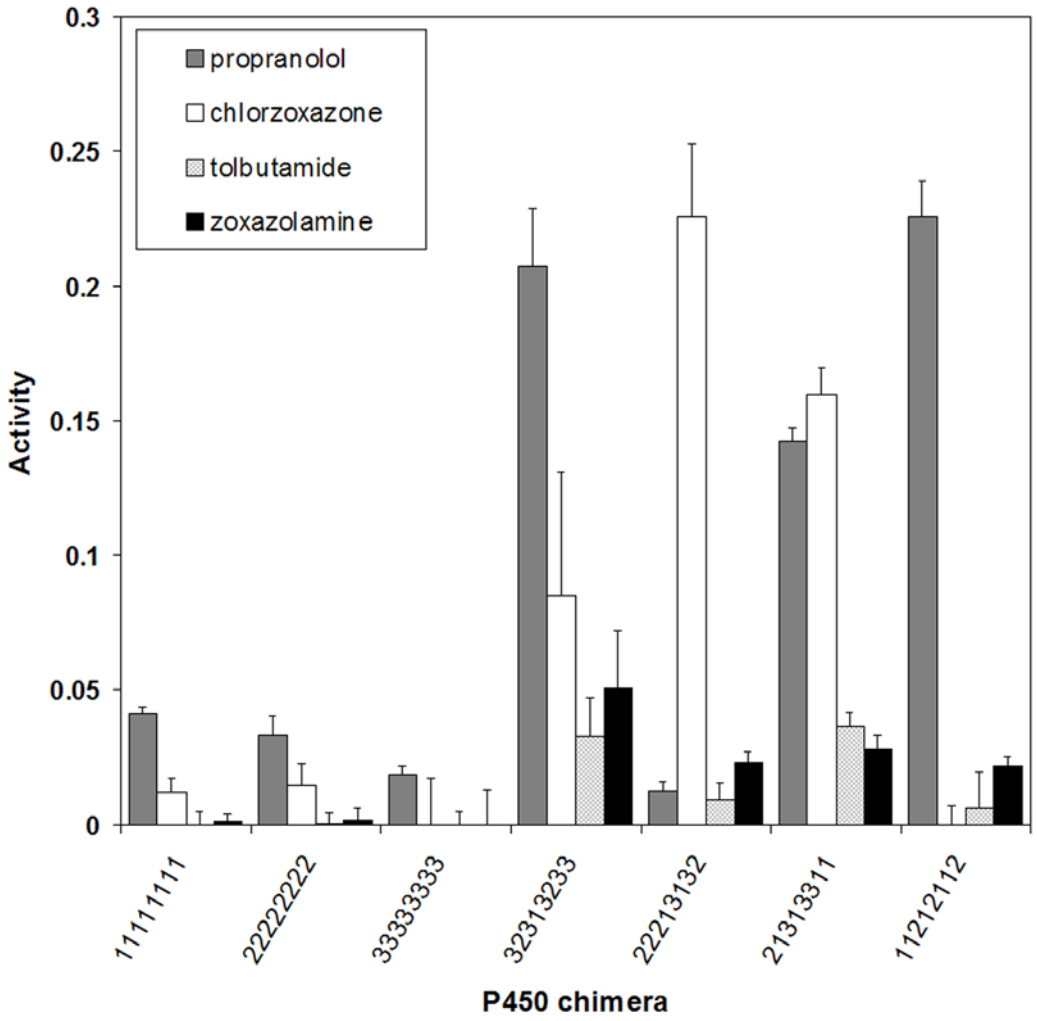

Figure 7.1 Total activity of P450 chimeras and parental enzymes A1, A2 and A3 on drug-like compounds (P450 chimera sequence represented in block-fragment format, see Chapter 5). Activity is measured as the absorbance at 500 nm minus the absorbance of control wells. Error bars are one standard deviation.

C.2 HPLC analysis of reactions with drug compounds

The most active chimera on each substrate was used for HPLC analysis along with the three parental enzymes, A1, A2 and A3. P450s were combined with substrate and 20 mM H_2O_2 and allowed to react for two hours. Control reactions were done in which substrate was omitted and in which peroxide was omitted. The reaction lacking substrate tests for any side products being created by reaction of P450 with molecules present besides the specific substrate. The reaction without peroxide tests for product formation not originating from non-peroxygenase reactions. After sample preparation (see Methods), the reactions were analyzed on HPLC.

Chimera 22213132 was used to examine the chlorzoxazone reaction. Figure 7.2 shows this chimera produced 6-hydroxychlorzoxazone as identified by retention time and matching absorbance spectra (see Appendix D for spectra of all substrates and products). Comparing peak areas normalized to the internal standard, zoxazolamine, we found that chimera 22213132 produced ten times as much 6-hydroxychlorzoxazone as the most active parent, A2. This is similar to the fifteen times more product produced in the 4-AAP assay.

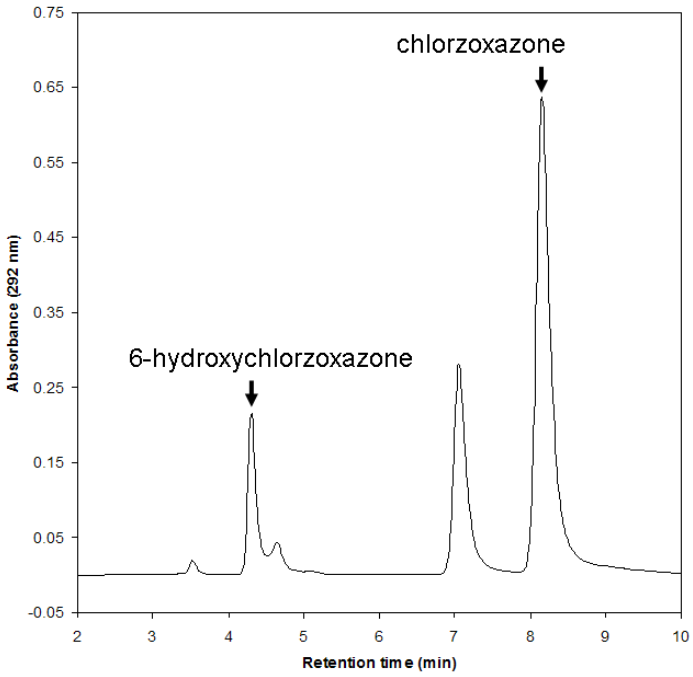

Figure 7.2 HPLC chromatogram of chlorzoxazone and the metabolite 6-hydroxychlrozazone produced by chimera 22213132. The retention times for chlorzoxazone and 6-hydroxychlorzoxazone were 8.0 and 4.3 min, respectively.

Chimera 11212112 was used for analysis of propranolol product formation. We found that it produced the three major metabolites 4'OHP, 5'OHP and DIP (Figure 7.3). Chimera 32313233 was used to analyze the products formed from zoxazolamine and Figure 7.4 shows that 6-hydroxyzoxazolamine was produced. The products of the reaction of 21313311 with tolbutamide were also analyzed. Standards for the two known metabolites 4-hydroxytolbutamide and carboxytolbutamide had retention times of 3.3 and 3.9 minutes (Figure 7.5). These peaks were not present in either control reaction and did not match the retention time or absorbance spectra for carboxytolbutamide or 4-hydroxytolbutamide. Therefore the specific product of the tolbutamide reaction is neither of these metabolites and remains unknown.

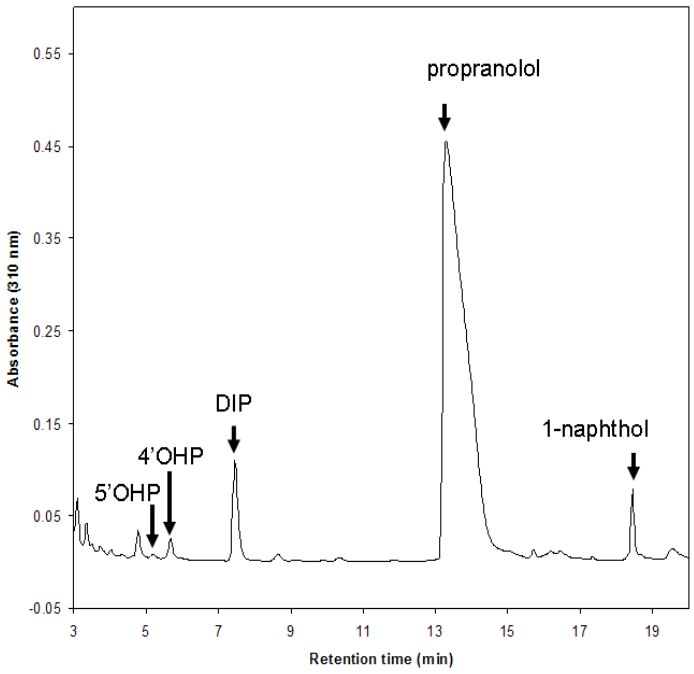

Figure 7.3 HPLC chromataogram of propranolol and its metabolites 5'OHP, 4'OHP, DIP and the side product 1-naphthol produced by chimera 11212112. The retention times for 5'OHP, 4'OHP, DIP, propranolol and 1-naphthol were 5.2 min, 5.7 min, 7.4 min, 12.9 min, 18.5 min, respectively.

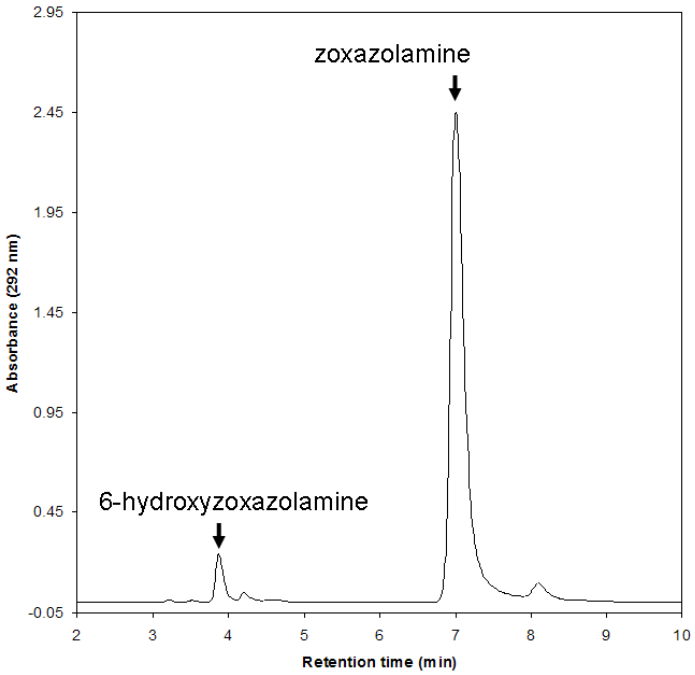

Figure 7.4 HPLC chromatogram of zoxazolamine and the metabolite 6-hydroxyzoxazolamine by chimera 32313233. The retention times for zoxazolamine and 6-hydroxyzoxazolamine were 7.1 and 3.9 minutes, respectively.

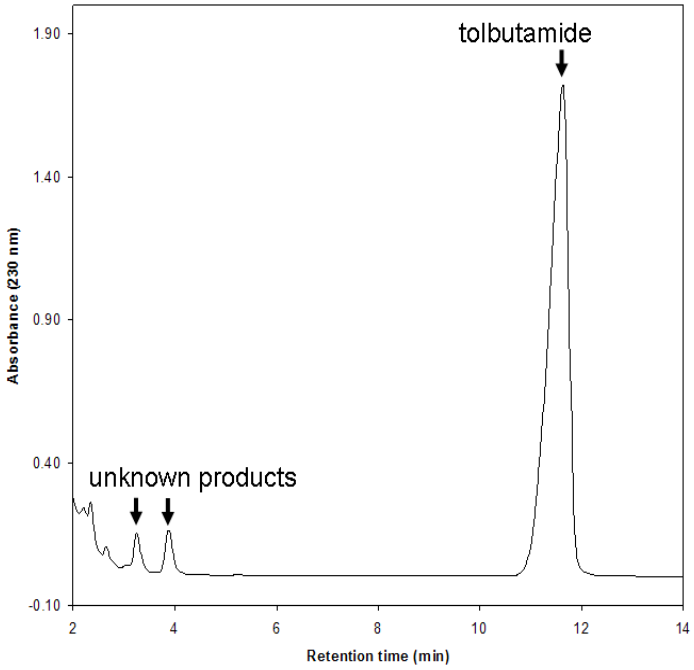

Figure 7.5 HPLC chromatogram of tolbutamide and two unidentified products produced by chimera 21313311. The retention times for the two unidentified products were 3.3 and 3.9 minutes. The retention time for tolbutamide was 11.7 min.

C.3 Reactions and analysis of 4-aminobiphenyl

Reactions containing chimeric P450s, H_2O_2 and 4-aminobiphenyl produced a reddish color and showed an increase in absorbance at 360 nm prior to addition of 4-AAP. Addition of 4-AAP did not increase absorbance near 500 nm suggesting there was no phenolic compound. Since 4-AAP was not required for the appearance of the reddish color, we suspected that a peroxidase reaction may be responsible for the color change via a conjugation reaction. HRP has been shown to react with 4-aminobiphenyl in the presence of H_2O_2 [10]. In this reaction, 4-aminobiphenyl serves as a reducing cosubstrate for HRP followed by conjugation of two 4-aminobiphenyl molecules forming 4,4'-azobis(biphenyl) [10]. Since cytochrome P450s also perform peroxidase chemistry we

tested to see if this was the reaction that occurred with the chimeric P450 21313311. The HRP reactions and P450 reactions were run side by side and analyzed using HPLC (Figure 7.6). They both turned a reddish color, showed an absorbance increase near 360 nm and the HPLC chromatogram had the same product peak with matching absorbance spectra at the same retention time. This supports that the product of the P450 chimera reaction with 4-aminobiphenyl was 4,4'-azobis(biphenyl).

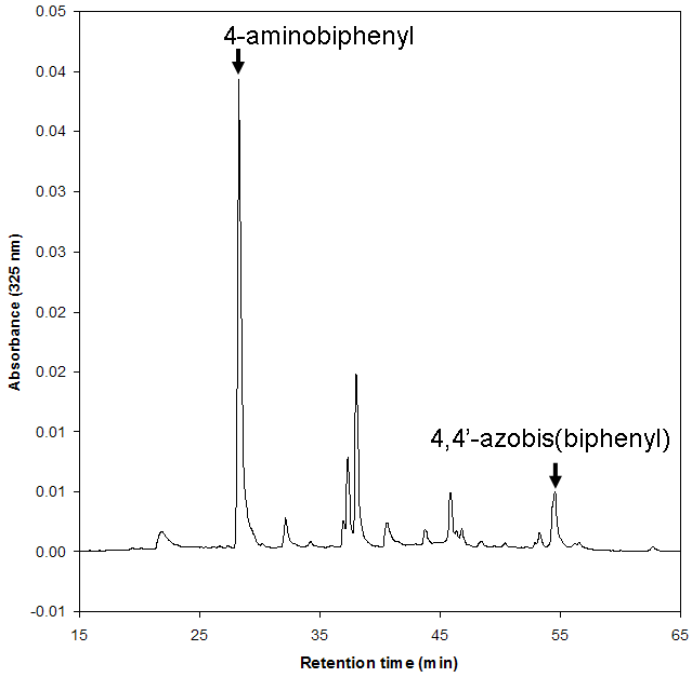

Figure 7.6 HPLC chromataogram of 4-aminobiphenyl and the conjugation product 4,4'-azobis(biphenyl) produced by 21313311. The retention times for 4-aminobiphenyl and 4,4'-azobis(biphenyl) were 28.5 and 54.8 min, respectively.

D. Discussion

The production of the drug metabolites produced by human P450s is of great interest since these chemicals may cause negative affects in a patient or have enhanced properties beyond that of the original drug. In fact, the FDA now requires study of all relevant drug

metabolites making their production critical to the drug approval process. The use of bacterial P450s able to accept drug-like compounds could be of great use in production of these metabolites. Here we show that bacterial P450s and their chimeric variants are able to produce authentic human metabolites as well as other chemical products. Figure 7.7 summarizes the findings showing the products of zoxazolamine, propranolol and chlorzoxazone were the authentic human metabolites. Chlorzoxazone and zoxazolamine are centrally acting muscle relaxants [11] while propranolol is a beta-adrenergic blocker [12]. 4-aminobiphenyl is a carcinogen found in tobacco smoke and serves as a carcinogen in human cancer [13]. The P450 chimeras did not appear to produce the human metabolite n-hydroxyaminobiphenl [7]. Instead, a peroxidase mechanism was utilized to conjugate two molecules of 4-aminobiphenyl to produce 4,4'-azobis(biphenyl) (Figure 7.7). Peroxidase chemistry in P450 chimeras was also observed in Chapter 4 on the substrate ABTS.

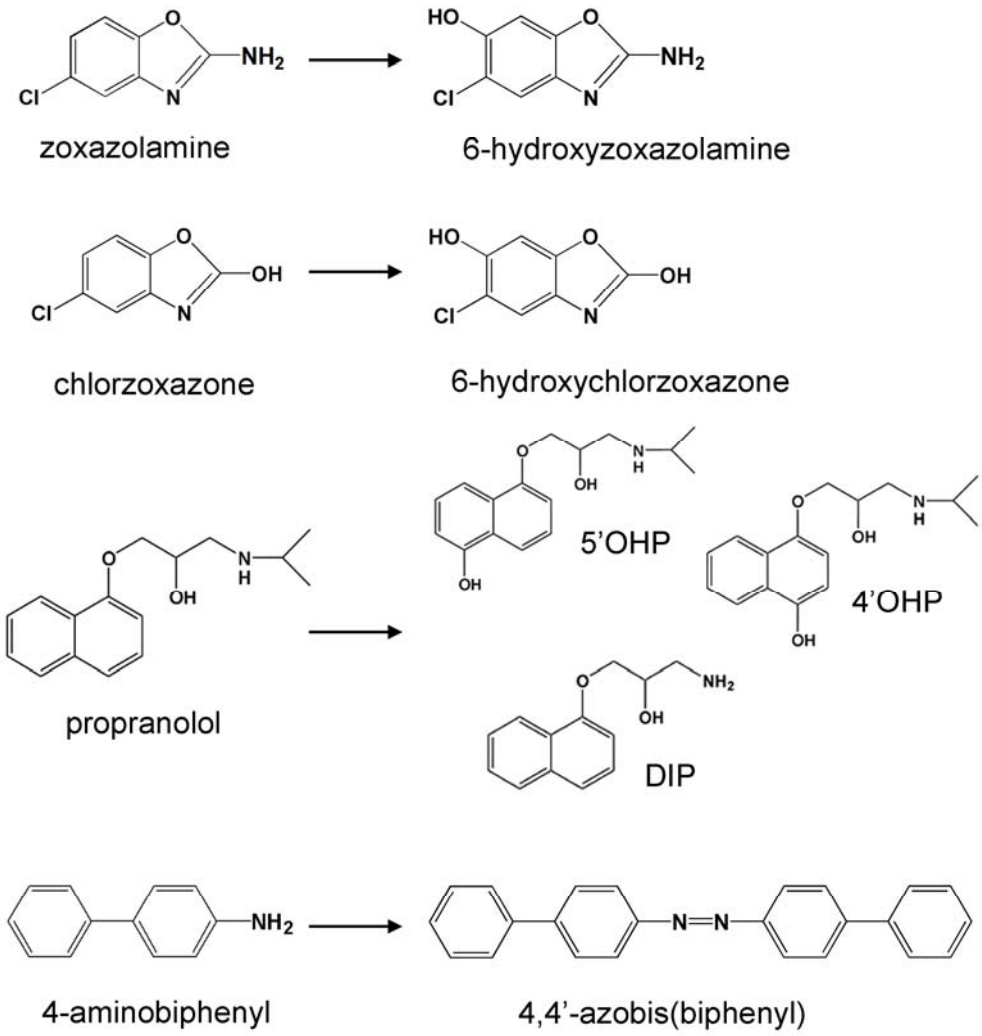

Figure 7.7 Products identified by HPLC that are produced by chimeric P450s.

The product of tolbutamide, an antidiabetic [14], was undetermined since it was not either of the known human metabolites. This was not entirely unexpected however since the 4-AAP assay detects aromatic ring hydroxylation and neither carboxytolbutamide nor 4-hydroxytolbutamide are hydroxylated on the aromatic ring. It is therefore likely that the product formed by the chimeric P450s is hydroxylated on the single aromatic ring in the substrate (Figure 7.8). Production of novel metabolites not made by the human enzymes such as this has potential for creating new drug molecules with enhanced efficacy from

existing chemical scaffolds. For example, new protein kinase inhibitors based on the indigoid scaffold [15] or novel metabolites of warfarin and an angiotensin II antagonist [16,17] have been generated. Finally, since the products formed by the chimeric P450s were not always the human metabolites it seems that merely acquire the ability to bind the substrate is enough for a specific activity. The orientation of the molecule in the active site is important for selectivity. The acceptance of drug-like compounds by engineered P450s opens up the potential for the production of known human metabolites as well as novel metabolites with possibly enhanced properties.

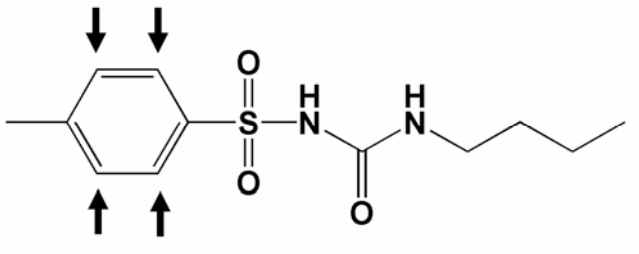

Figure 7.8 Potential hydroxylation sites for the products of tolbutamide produced by chimeric P450s.

E. Materials and Methods

E.1 Materials

All chemicals were purchased from Sigma-Aldrich, Inc. (St. Louis, MO), MP Biomedicals, Inc. (Solon, OH) or BD Biosciences (San Jose, CA). Solvents were from EM Sciences (Gibbstown, NJ). HRP was from Amersham Biosciences (Piscataway, NJ).

E.2 Protein expression and preparation

All P450 BM3 heme domains were cloned, transformed and expressed in the catalase-deficient strain of *E. coli* SN0037 [18] using the IPTG-inducible pCWori vector [19]. For

protein expression, 50 mL of Terrific Broth (TB) supplemented with 100 μg/mL ampicillin and 25 μg/mL thiamine were inoculated with 500 μL of an overnight culture and incubated for 5 hours at 30 ^{o}C with shaking. P450 expression was induced by adding 0.5 mM IPTG and the heme precursor δ-ALA to a final concentration of 1 mM. The cultures were grown for another 18 hours and the cells were pelleted and stored at -20 ^{o}C. Cell pellets were resuspended in 100 mM Epps pH 8.2, sonicated for 3 x 45 seconds and centrifuged at 10,000 g for 30 min. The supernatant was passed through a PD10 size exclusion column and P450 concentration was measured by CO difference spectroscopy [20].

E.3 Testing of individual chimeras

Cleared cell lysates of individual chimeras were prepared as in Chapter 4. Enzyme concentrations were determined using CO binding difference spectroscopy with an extinction coefficient of 91 mM^{-1} cm^{-1} [21]. 4-AAP reactions were carried out in 120 μl with 2 μM P450, 1% dimethyl sulfoxide, 1% acetone in 100 mM EPPS (pH 8.2). Substrate concentrations were 5 mM for chlorzoxazone, zoxazolamine, 4-aminobiphenyl and propranolol and 10 mM for tolbutamide. Reactions were initiated by the addition of H_2O_2 to a final concentration of 20 mM. After two hours at room temperature, reactions were quenched by addition of 120 μl of 4 M urea and 100 mM NaOH. Additionally, identical reactions were carried out in which the reaction was quenched prior to the addition of H_2O_2 in order to establish the background level. To these quenched reactions, 36 μl of 0.6 % 4-AAP was added and the spectrophotometer was zeroed at 500 nm. Thirty-six μl of 0.6% potassium persulfate was added and the A_{500} was read after ten

minutes. Each measurement was done in triplicate and the total activity was determined by subtracting the pre-quenched point from the two hour point. 4-aminobiphenyl conditions were carried out in an identical fashion except no 4-AAP or potassium persulfate was added. Instead, the reactions were quenched and allowed to incubate for 30 minutes in order to partially dissolve precipitates. The A_{500} was then read.

E.4 Bioconversions for HPLC analysis and sample preparation

Bioconversions for HPLC analysis of chimeric P450s were prepared as above except were done in 600 µl volumes and with 5 µM P450. Propranolol analysis was similar to that described previously [22]. Perchloric acid (7%) was added to the reaction mixture at a volume of 10 µL/100 µL of reaction, followed by addition of 2.5 mg/100 µL ascorbic acid. The reaction mixture was vortexed and centrifuged. The chlorzoxazone and zoxazolamine bioconversions were adjusted to pH 7 using H_3PO_4 and extracted three times with equivolumes of CH_2Cl_2. After removal of solvent and resuspension in 300 µl of 65% solvent A (1% triethylamine (v/v) and 0.8% acetic acid (v/v) in H_2O) and 35% solvent B (acetonitrile), an internal standard was added to a final concentration of 100 µM. Zoxazolamine was used as an internal standard for chlorzoxazone and vice versa for zoxazolamine reactions. 4-aminobiphenyl reactions were run directly on the HPLC. Tolbutamide bioconversions were acidified with 60 µl 7% perchloric acid and 15 mg ascorbic acid followed by vortexing and centrifugation.

E.5 HRP reaction with 4-aminobiphenyl

Reactions contained five units of HRP, 5 mM aminobiphenyl in 100 mM EPPS (pH 8.2) and were initiated by the addition of H_2O_2 to a final concentration of 1 mM. The reactions were allowed to proceed for 30 minutes at which point they were run directly on HPLC for analysis.

E.6 HPLC analysis

All bioconversions were analyzed on an Alliance HPLC system (Waters, Milford, MA) equipped with a photodiode array detector. Metabolites were identified by retention times and absorption spectra (see Appendix D for substrate and product spectra). Propranolol, chlorzoxazone, zoxazolamine and 4-aminobiphenyl were separated on a Microsorb-MV phenyl column (250 x 4.6 mm, particle diameter 5 µm, Varian, Palo Alto, CA) with varying HPLC conditions using solvent A (1% triethylamine (v/v) and 0.8% acetic acid in H_2O) and solvent B (acetonitrile). Fifty µl of the propranolol preparation was analyzed using the following elution protocol: 0-10 min., A:B 75:25; 10-20 min., linear gradient to A:B 25:75; 20-22 min., A:B 75:25; 22-27 min., linear gradient to A:B 75:25. Retention times for 5'-hydroxypropranolol, 4'-hydroxypropranolol, desisopropylpropranolol, propranolol and 1-naphthol were 5.2 min, 5.7 min, 7.4 min, 12.9 min, 18.5 min, respectively. Twenty µl of the chlorzoxazone and zoxazolamine extracts were run isocratically for 10 minutes with solvent A:B 65:35 at 1 ml/min. Retention times for chlorzoxazone and 6-hydroxychlorzoxazone were 8.0 and 4.3 min, respectively. Retention times for zoxazolamine and 6-hydroxyzoxazolamine were 7.1 and 3.9 min, respectively. Twenty µl of the 4-aminobiphenyl bioconversions were run directly on

HPLC and eluted with the following conditions: 0-10 min, A:B 75:25; 10-50 min, linear gradient to A:B 25:75; 50-70 min, A:B 25:75; 70-80 min, linear gradient to A:B 75:25. The retention times for 4-aminobiphenyl and 4,4'-azobis(biphenyl) were 28.5 and 54.8 min, respectively. Twenty µl of the tolbutamide extracts were run on a Phenomenex C8 (2) column (150 x 4.6mm, particle diameter 5 µm, Phenomenex, Torrance, CA) using 45% acetonitrile and 55% 0.325% H_3PO_4 at 1.8 ml/min for 15 minutes. The retention times for tolbutamide, carboxytolbutamide and 4-hydroxytolbutamide were 11.7, 3.8 and 2.4 min.

F. References

1. Lewis DF (2004) 57 varieties: the human cytochromes P450. Pharmacogenomics 5: 305-318.
2. Guengerich FP, Gillam EM, Shimada T (1996) New applications of bacterial systems to problems in toxicology. Crit. Rev. Toxicol. 26: 551-583.
3. Guengerich FP (2002) Cytochrome P450 enzymes in the generation of commercial products. Nat. Rev. Drug Discov. 1: 359-366.
4. Lewis DFV (2001) Guide to cytochromes P450: structure and function. London; New York: Taylor & Francis.
5. Peter R, Bocker R, Beaune PH, Iwasaki M, Guengerich FP, et al. (1990) Hydroxylation of chlorzoxazone as a specific probe for human liver cytochrome P-450IIE1. Chem. Res. Toxicol. 3: 566-573.
6. Ho JW, Moody DE (1993) Determination of tolbutamide hydroxylation in rat liver microsomes by high-performance liquid chromatography: effect of psychoactive drugs on in vitro activity. Life Sci. 52: 21-28.
7. Ozawa S, Katoh T, Inatomi H, Imai H, Kuroda Y, et al. (2002) Association of genotypes of carcinogen-activating enzymes, phenol sulfotransferase SULT1A1 (ST1A3) and arylamine N-acetyltransferase NAT2, with urothelial cancer in a Japanese population. Int. J. Cancer. 102: 418-421.
8. Masubuchi Y, Hosokawa S, Horie T, Suzuki T, Ohmori S, et al. (1994) Cytochrome P450 isozymes involved in propranolol metabolism in human liver microsomes. The role of CYP2D6 as ring-hydroxylase and CYP1A2 as N-desisopropylase. Drug Metab. Dispos. 22: 909-915.
9. Cirino PC, Arnold FH (2003) A self-sufficient peroxide-driven hydroxylation biocatalyst. Angew. Chem. Int. Ed. Engl. 42: 3299-3301.
10. Hughes MF, Smith BJ, Eling TE (1992) The oxidation of 4-aminobiphenyl by horseradish peroxidase. Chem. Res. Toxicol. 5: 340-345.
11. Syme CA, Gerlach AC, Singh AK, Devor DC (2000) Pharmacological activation of cloned intermediate- and small-conductance Ca(2+)-activated K(+) channels. Am. J. Physiol. Cell Physiol. 278: C570-C581.
12. Shand DG (1975) Drug therapy: Propranolol. N. Engl. J. Med. 293: 280-285.
13. Flamini G, Romano G, Curigliano G, Chiominto A, Capelli G, et al. (1998) 4-Aminobiphenyl-DNA adducts in laryngeal tissue and smoking habits: an immunohistochemical study. Carcinogenesis 19: 353-357.
14. Tucker HA (1965) Oral antidiabetic therapy, 1956-1965: with particular reference to tolbutamide. Springfield, Ill.: Thomas.
15. Guengerich FP, Sorrells JL, Schmitt S, Krauser JA, Aryal P, et al. (2004) Generation of new protein kinase inhibitors utilizing cytochrome p450 mutant enzymes for indigoid synthesis. J. Med. Chem. 47: 3236-3241.
16. Cannell RJ, Knaggs AR, Dawson MJ, Manchee GR, Eddershaw PJ, et al. (1995) Microbial biotransformation of the angiotensin II antagonist GR117289 by Streptomyces rimosus to identify a mammalian metabolite. Drug Metab. Dispos. 23: 724-729.
17. Cannell RJ, Rashid T, Ismail IM, Sidebottom PJ, Knaggs AR, et al. (1997) Novel metabolites of warfarin produced by Beauveria bassiana and Streptomyces rimosus: a novel application of hplc-nmr. Xenobiotica 27: 147-157.

18. Nakagawa S, Ishino S, Teshiba S (1996) Construction of catalase deficient Escherichia coli strains for the production of uricase. Biosci. Biotechnol. Biochem. 60: 415-420.
19. Barnes HJ, Arlotto MP, Waterman MR (1991) Expression and enzymatic activity of recombinant cytochrome P450 17 alpha-hydroxylase in Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. 88: 5597-5601.
20. Schenkman JB, Jansson I (1998) Spectral Analyses of Cytochromes P450. In: Phillips IR, Shephard EA, editors. Cytochrome P450 Protocols. Totowa, New Jersey: Human Press, Inc. pp. 25-33.
21. Omura T, Sato R (1964) The carbon monoxide-binding pigment of liver microsomes. II. Solubilization, purification, and properties. J. Biol. Chem. 239: 2379-2385.
22. Upthagrove AL, Nelson WL (2001) Importance of amine pKa and distribution coefficient in the metabolism of fluorinated propranolol derivatives. Preparation, identification of metabolite regioisomers, and metabolism by CYP2D6. Drug Metab. Dispos. 29: 1377-1388.

Chapter 8

Preparation of human metabolites of propranolol using laboratory-evolved bacterial cytochromes P450

Material from this chapter appears in Otey CR, Bandara G, Lalonde J, Takahashi K, Arnold FH (2006) **Preparation of human metabolites of propranolol using laboratory-evolved bacterial cytochromes P450**. *Biotechnol. Bioeng.* 93: 494-499, and is reprinted with permission from John Wiley & Sons, Inc. Additional thanks to Professor Wendel Nelson for generously providing the 4' and 5'-hydroxy propranolol standards and James Ross for laboratory assistance. This work was supported by National Institutes of Health Grant R01 GM068664-01.

A. Abstract

Testing the toxicities and biological activities of the human metabolites of drugs is important for the development of safe and effective pharmaceuticals. Producing these metabolites using human cytochrome P450s is difficult, however, because the human enzymes are poorly stable and slow. We have used directed evolution to generate variants of P450 BM3 from *Bacillus megaterium* that function via the 'peroxide shunt' pathway, using hydrogen peroxide in place of the reductase domain, oxygen and NADPH. Here, we report further evolution of the P450 BM3 heme domain to optimize production of the authentic human metabolites of propranolol by this approach. This biocatalytic system offers a versatile, cost-effective, and scaleable route to the synthesis of drug metabolites.

B. Introduction

The cytochromes P450 comprise a ubiquitous superfamily of heme-containing enzymes which perform a variety of oxidative reactions on a wide range of substrates [1,2]. In humans, P450s are responsible for the disposition of many xenobiotics, including pharmaceuticals [3]. During initial clearance, most drugs are metabolized in the liver by P450s [4,5]. Some of the metabolites are biologically active themselves, and understanding their effects is crucial in evaluating a drug's efficacy, toxicity and pharmacokinetics [6]. Identification of so-called "active metabolites," as in the case of terfenadine [7], early on in the drug development process could result in the early identification of safer, more well tolerated therapies. Such studies, however, can require large quantities of the pure metabolites, and these may be difficult to synthesize. An alternative to chemical synthesis is to use P450s to generate the metabolites of drugs or drug candidates. Hepatic microsomes are a source of human P450s, but their limited availability and highly variable expression levels make their use in preparative-scale metabolite synthesis impractical. Some human enzymes can also be obtained by expression in recombinant hosts. As membrane-bound, multi-protein systems, however, they commonly misfold and aggregate or are not expressed in active form. Metabolite preparation has been demonstrated using human P450s expressed in *E. coli* and in insect cells [8-10], but these systems are costly and have low productivities due to limited stabilities and slow reaction rates (usually <5 min^{-1} [11]).

The bacterial cytochrome P450 BM3, a fatty acid hydroxylase from *Bacillus megaterium*, is expressed at high levels in *E. coli* and exhibits very fast reaction rates (thousands per

minute) on favored substrates [12]. This enzyme has been engineered to accept a variety of new substrates, on which its variants catalyze regio- and enantio-selective oxidations [13-19]. In previous work, we used directed evolution, with iterations of random mutagenesis and high-throughput screening, to generate variants of the CYP102A1 heme domain which efficiently hydroxylate fatty acids using hydrogen peroxide in place of the reductase domain, oxygen and the costly NADPH cofactor (P450 'peroxygenases') [20]. Here, we demonstrate that an engineered P450 CYP102A1 heme domain peroxygenase can be used to produce the authentic human metabolites of propranolol, a multi-function beta-adrenergic blocker used to treat hypertension, arrhythmia, angina, migraine headaches, overactive thyroids and anxiety [21]. Furthermore, this activity can be improved by directed evolution to produce more of the biologically active metabolites.

C. Results

C.1 Bioconversion of propranolol by P450 A1-9C1

In previous work, we created variants of A1 with high peroxygenase activity and thermostability by directed evolution, using multiple rounds of mutagenesis and screening [20]. In each round, a mutant library created by error-prone PCR random mutagenesis was screened for H_2O_2-driven hydroxylation of 12-*p*-nitrophenoxycarboxylic acid (12-pNCA) [22]. The most active mutants were recombined to obtain the parents for subsequent rounds of evolution. A1 variant 21B3 obtained in this way was further evolved during four rounds of error-prone PCR and screening to generate variant 9C1, which has 13 amino acid substitutions relative to the original

sequence, wild-type P450 CYP102A1 with the F87A mutation (Table 8.1) and catalyzes ~1000 turnovers on 12-pNCA.

Table 8.1. Amino acid substitutions in cytochrome P450 A1 variants.

Position (A1-9C1[a])	DE10	D6H10	2C11
K24[b]			R
R47[b,c]			H
A74[c]	V		V
L75[c]		H	
V78[c]		E	
A82[c]	L	P	L
A87[c]	G		G

[a] The 9C1 parent has the following amino acid substitutions compared to wild-type CYP102A1: F87A, I58V, H100R, I102T, F107L, A135S, M145A, N239H, S274T, L324I, I366V, K434E, E442K, V446I.
[b] Identified by random mutagenesis
[c] Targeted active-site residues

In humans, CYP2D6 and CYP1A2 are the main P450s responsible for the metabolism of propranolol [23], forming the three major products shown in Figure 8.1. The ring-hydroxylated products, 4'OHP and 5'OHP, are of particular importance due to their pharmacological activities [24,25]. Upon testing different peroxygenase variants of A1 from our previous work, we found that variant 9C1 produces a very similar product profile in a bioconversion carried out using *E. coli* cell extracts overexpressing the P450 heme domain. A representative chromatogram showing the products from a propranolol bioconversion by A1-H 9C1 is shown in Figure 8.2.

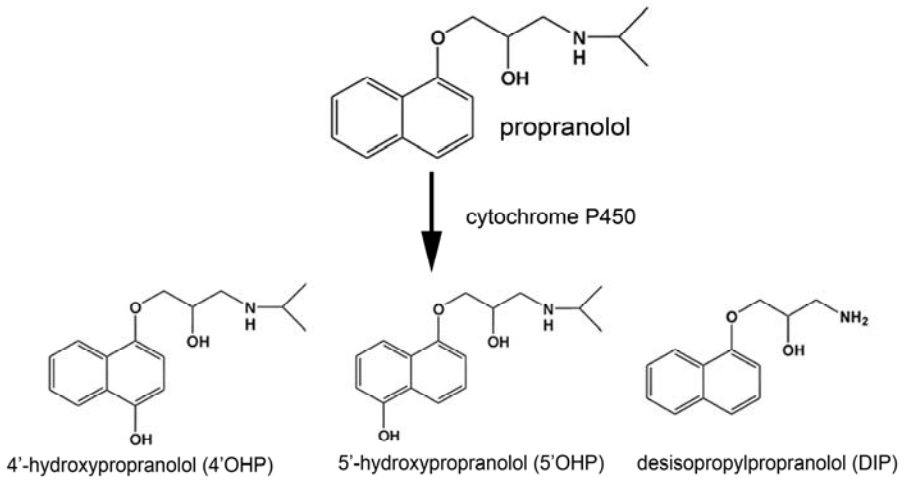

Figure 8.1. Major metabolites of propranolol formed by human cytochromes P450.

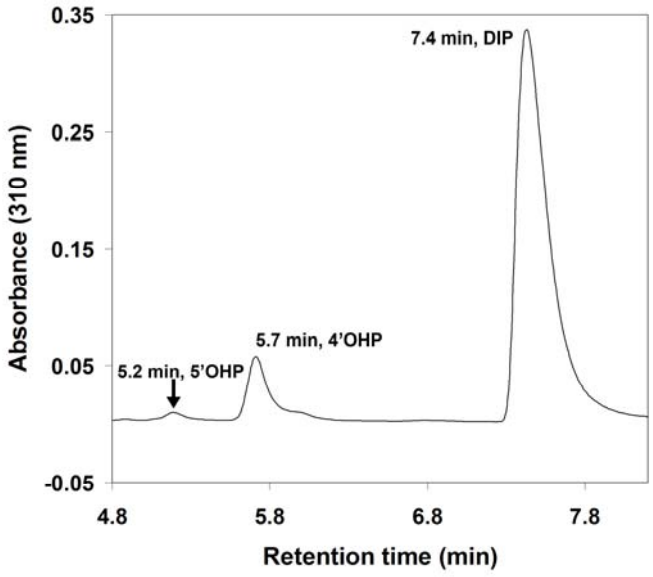

Figure 8.2. HPLC chromatogram showing the relevant human metabolites of propranolol produced in a bioconversion of (5 mM) propranolol with A1-9C1. Products from left to right are 5'-hydroxypropranolol (5.2 min.), 4'-hydroxypropranolol (5.7 min.), desisopropylpropranolol (7.4 min.). Propranolol (12.9 min.) and 1-naphthol (18.5 min.) are not shown.

Since A1-9C1 was evolved for activity on the fatty acid analog, 12-pNCA [20], we did not expect it to function optimally on the chemically different propranolol. Initial reactions containing 5 μM P450 and 5 mM propranolol driven by a single addition of H_2O_2 to a final concentration of 1 mM converted 8% of the substrate to products. To test whether H_2O_2 or P450 was limiting in this reaction we followed the amount of P450 present and product produced every 30 minutes for 6.5 hours (Figure 8.3). H_2O_2 was added either once or multiple times and in the presence or absence of the substrate propranolol. We found that roughly 75% of the initial P450 was still present after 3 hours when H_2O_2 was added only once. Production could be more than doubled to 18% conversion by multiple additions of 1 mM H_2O_2 every 30 minutes which equates to ~180 turnovers per P450 active site. Under these conditions, the reaction was complete in 180 minutes, at which point no more product was made, and no P450 remained (as determined by carbon monoxide difference spectroscopy). Interestingly, in the absence of the substrate propranolol, P450 degraded much more rapidly suggesting that the substrate is protecting the enzyme. The precise mechanism is unknown but it likely serves to reduce the frequency of interactions between the heme and oxygen radicals resulting from H_2O_2 which can cause heme alkylation [26-28] or oxidize or form radicals at individual amino acids [29-34], both of which inactivate the enzyme.

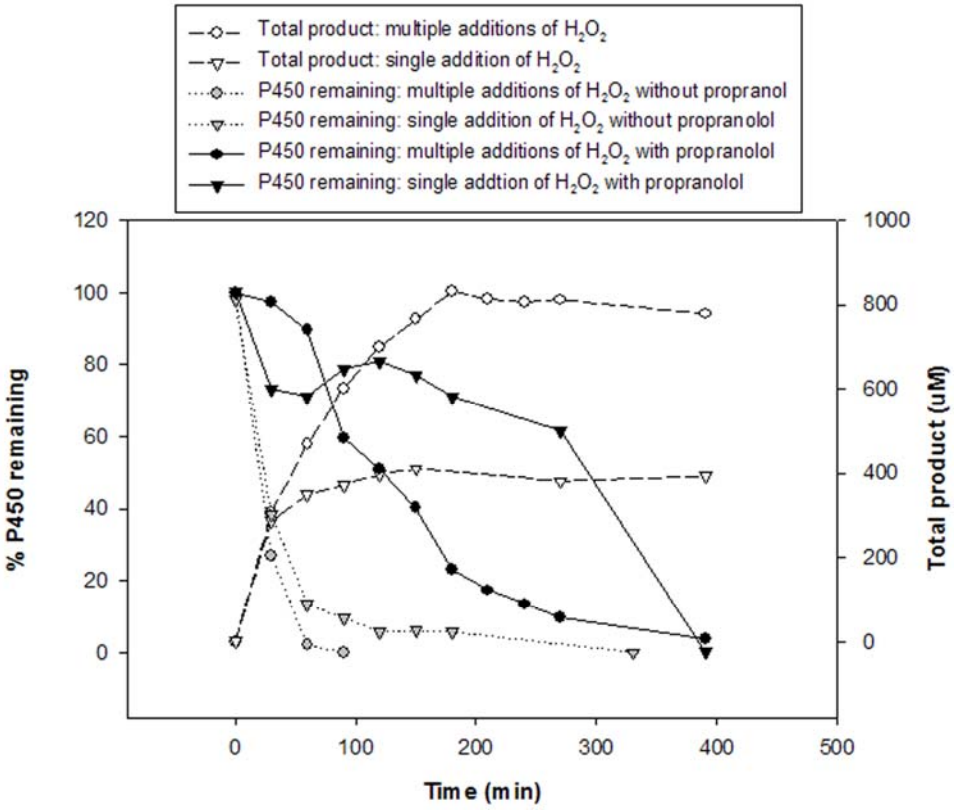

Figure 8.3. Productivity and lifetime of A1-9C1 under varying reaction conditions. Product formation was monitored by HPLC and remaining P450 by CO difference spectroscopy. H_2O_2 was added either once (single) or multiple times to a concentration of 1 mM. The concentrations of P450 and propranolol were 5 μM and 5 mM, respectively.

C.2 Directed evolution to enhance activity of A1-9C1 on propranolol

To test whether the bacterial P450 heme domains can be engineered to increase production of the hydroxylated products 4'OHP and 5'OHP and total activity towards drug compounds such as propranolol, we carried out additional directed evolution, this time with mutations targeted specifically to the active site. Using A1-9C1 as the parent, a library was created by directing mutations to seven active-site residues (A74, L75, V78,

F81, A82, F87 and T88) which are within 5 Å of the substrate, based on the crystal structure of P450 BM3 bound to N-palmitoylglycine [35]. The mutagenic oligonucleotides were doped such that two of the seven residues were mutated in each sequence, on average. This gene library was transformed into catalase-deficient *E. coli* [36] to allow screening for activity in the presence of hydrogen peroxide. Propranolol hydroxylation activity was monitored using the 4-AAP assay (see Chapter 3), which detects the products of aromatic hydroxylation, 4'OHP, 5'OHP, and 1-naphthol. It does not detect the dealkylation product DIP.

Mutants DE10 and D6H10 were selected on the basis of increased formation of detectable products compared to parent A1-9C1. Sequencing showed that each mutant contained three amino acid substitutions relative to A1-9C1 (Table 8.1). HPLC analysis of the reaction mixture showed that DE10 and D6H10 convert 13% and 9% of the propranolol, respectively, to products. Although DE10 and D6H10 lead to less conversion than 9C1, they produce more of the metabolites detected by the 4-AAP assay. As shown in Figure 8.4, DE10 produces twice as much 4'OHP, while D6H10 produces 50% more 5'OHP. These variants make 33% less DIP.

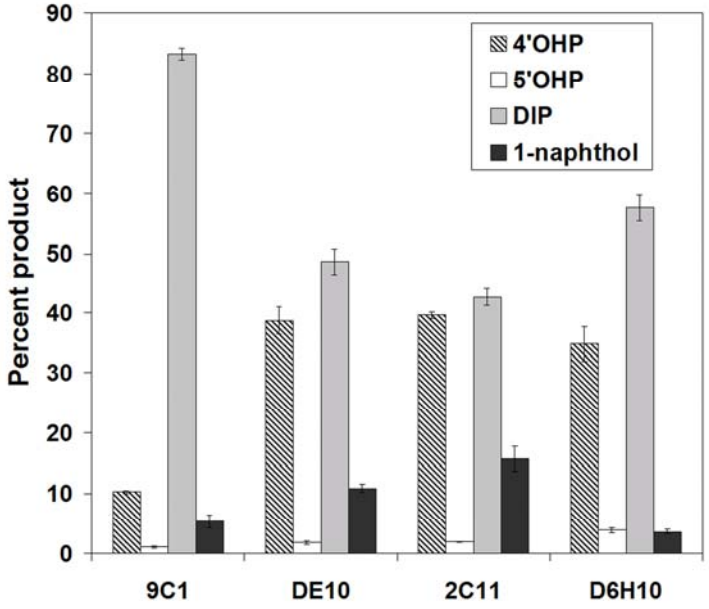

Figure 8.4. Propranolol metabolites produced by A1-9C1 and variants obtained by further directed evolution. Results are from bioconversions containing 5 μM P450 and 5 mM propranolol, driven by H_2O_2. Error bars represent one standard deviation from three experiments.

We next generated a second mutant library using error-prone PCR to introduce random mutations throughout the DE10 gene and screened the resulting clones for enhanced product formation. From this library we isolated 2C11. This mutant generates a product profile similar to DE10 but converts more (20%) of the propranolol to the hydroxylated products (Figure 8.4). Relative to A1-9C1, it produces 50% more 5'OHP, 200% more 4'OHP, 100% more 1-naphthol and 40% of the DIP. The four products 4'OHP, 5'OHP, and 1-naphthol and DIP account for ~75-80% of the propranolol converted. Other, minor products were also visible in the HPLC traces, but their identities are unknown. Mutant 2C11 contains an additional two amino acid substitutions compared to DE10 (Table 8.1).

D. Discussion

We have demonstrated that a bacterial P450 heme domain can be used to prepare the authentic human metabolites of propranolol in a reaction driven by hydrogen peroxide. Furthermore, the activity can be enhanced and product selectivity can be optimized by directed evolution. In the first round of evolution, mutations were targeted to active-site residues believed to contact the substrate during catalysis. In wild-type P450 CYP102A1, phenylalanine 87 sequesters the ω-end of the fatty acid, which is hydroxylated at adjacent positions [37]. Substitution of this residue for alanine increases the activity of wild-type BM3 on polycyclic aromatics [38] and affects the regioselectivity of fatty acid hydroxylation [39]. It is also known that the residue size at position 87 plays a critical role in H_2O_2 dependent substrate hydroxylation: replacement of phenylalanine 87 with smaller residues such as alanine or glycine increases peroxygenase activity [39,40]. It is thought that the additional space provided by these mutations allows more water molecules to remain at the active site pocket after substrate binding, promoting H_2O_2 access. We therefore introduced the F87A mutation into the wild-type parent CYP102A1 during the initial directed evolution efforts to find an efficient peroxygenase. This F87A mutation in the parent 9C1 used for the present study was exchanged for the even smaller, glycine residue after the first round of evolution.

Mutations in the active site have been shown to alter the stereo- and regioselectivity of P450 BM3 [13,14,17,19]. Saturation mutagenesis of active-site residues and screening has generated a range of CYP102A1 variants with novel activities, ranging from ethane hydroxylation [41] to enantioselective epoxidation of terminal alkenes [15] and

enantioselective hydroxylation of protected carboxylic acids [18], demonstrating the ease with which this enzyme's substrate specificity and product selectivity can be modified. The active-site mutations described here produce similar results, altering the regioselectivity of A1 in the peroxygenase reaction on propranolol. Activity is shifted activity away from the dealkylation reaction and towards aromatic hydroxylation (Figure 8.4).

The specific effects of each mutation are difficult to rationalize since the substrate likely binds in more than one conformation, as indicated by its multiple products. The mutations in the active site, however, influence the binding geometries available to the substrate and thereby alter the regioselectivity of oxidation. In variants DE10 and 2C11, active-site residues A82 and A74 are both changed to larger hydrophobic amino acids, leucine and valine, respectively, which presumably reduce the volume of the active site. A82L has previously been shown to alter the regioselectivity of substrate hydroxylation [17]. In contrast, A87G increases the volume of the active site directly above the heme. In 2C11, substitutions R47H and K24R introduced during random mutagenesis increase total activity while retaining a product profile similar to DE10 (Figure 8.4). K24R, which is not an active-site residue, is conservative and may be neutral. R47H, in contrast, is located at the mouth of the active site. Mutation of this residue to leucine has been shown to increase activity towards polycyclic aromatic hydrocarbons [42]. D6H10 has three mutations which lower total activity relative to parent A1-9C1 but increase production of 5'OHP. In this variant, the valine at position 78 is mutated to glutamic acid, introducing a

charged residue into the hydrophobic active site and likely resulting in weaker binding of the substrate.

In human liver microsomes, ring hydroxylation of propranolol is mediated by CYP2D6, whereas desisopropylation is mediated mainly by CYP1A2 and to a lesser extent by CYP2D6 [23]. Of these metabolites, the ring-hydroxylated products are of particular importance. 4'OHP has been demonstrated to be equipotent to propranolol as a β-receptor antagonist [24] and is believed to contribute to the bioactivity of propranolol in humans [43]. Other monohydroxylated products have also been shown to produce β-blockade and direct vasodilation in dogs [25]. Chemical synthesis for each individual metabolite involves multiple steps that require organic solvents and expensive chemical catalysts such as palladium [25,44]. The variants of bacterial P450 A1 described here produce all the major products of human propranolol metabolism, but, importantly, the evolved mutants make more of the key ring-hydroxylated products 4'OHP and 5'OHP. The variants here are also more productive than the human P450s. CYP2D6 has rates of < 5 per min and capable of less than 10 turnovers per P450 [45], much lower than the mutant 2C11 which catalyzes 180 total turnovers.

The heme domain of CYP102A1 is expressed at high levels (~100 mg/L) in a typical shake culture, and this can be further increased in a high density fermentation. Using crude cell lysate containing evolved heme domains capable of 180 turnovers when provided with H_2O_2, one could generate >70 mg of products with the enzyme from one liter of batch culture. This is similar to the quantities produced in *S. frugiperda* and *E.*

coli cells using human P450s [9,10], but the preparation is simpler and far less costly. In addition, the productivity can likely be increased by further optimizing the reaction conditions or protein expression. By simply serially adding H_2O_2 in small amounts rather than a large, single addition to high concentration, we were able to double the productivity (Figure 8.3). Recently it has been shown that the addition of other proteins, such as bovine serum albumin [46], increases the productivity of CYP102A1 and other parameters such as enzyme: substrate ratio, pH, buffer, ionic strength, temperature, etc. could be modified to further increase productivity.

An advantage of using the CYP102A1 heme domain, which utilizes H_2O_2 to drive these reactions, is its self-sufficiency. Catalysis does not depend on the addition or regeneration of NADPH, the cofactor required for most P450 reactions. Using the peroxygenase activity significantly decreases the cost and complexity of the preparation of these metabolites. With further optimization of the catalyst and reaction conditions, we anticipate that this system can be used to generate larger quantities of valuable drug metabolites. Additionally, laboratory-evolved P450s may be used for combinatorial biocatalysis of drug scaffolds to further diversify and increase the efficacy of these molecules. Using bacterial P450s to create the authentic human metabolites of drugs provides an easily scaleable and improvable system for the biosynthesis of these valuable compounds.

E. Methods

E.1. Materials

All chemicals were purchased from Sigma-Aldrich, Inc. (St. Louis, MO). Solvents were from EM Sciences (Gibbstown, NJ). Enzymes for DNA manipulation were from New England Biolabs (Beverly, MA). DNA purification kits were from Zymo Research (Orange, CA) and Qiagen (Valencia, CA). T4 DNA ligase was from Invitrogen (Carlsbad, CA). Propranolol standards 4'-hydroxypropranolol (4'OHP) and 5'-hydroxypropranolol (5'OHP) were generously provided by Professor Wendel Nelson of the University of Washington, Seattle. (S)-*N*-desisopropylpropranolol was purchased from ABX Advanced Biochemical Compounds (Radeberg, Germany).

E.2. Protein expression

All P450 BM3 heme domains were cloned, transformed and expressed in the catalase-deficient strain of *E. coli* SN0037 [36] using the IPTG-inducible pCWori vector [47]. For protein expression, 50 mL of Terrific Broth (TB) supplemented with 100 μg/mL ampicillin and 25 μg/mL thiamine were inoculated with 500 μL of an overnight culture and incubated for 5 hours at 30 ^{o}C with shaking. P450 expression was induced by adding 0.5 mM IPTG and the heme precursor δ-ALA to a final concentration of 1 mM. The cultures were grown for another 18 hours and the cells were pelleted and stored at -20 ^{o}C.

E.3. Bioconversion of propranolol

Cell pellets were resuspended in 100 mM EPPS pH 8.2, sonicated for 3 x 45 seconds and centrifuged at 10,000 g for 30 min. The supernatant was passed through a PD10 size

exclusion column and P450 concentration was measured by CO difference spectroscopy [48].

Bioconversions were done in 400 μL of 100 mM Epps pH 8.2 containing 5 μM P450 and 5 mM propranolol. Reactions were initiated by adding H_2O_2 to a final concentration of 1 mM, and the equivalent amount every 30 minutes thereafter. After 180 minutes at room temperature, the reaction products were analyzed using the 4-AAP assay and HPLC. Control reactions performed without the addition of enzyme or peroxide produced no detectable products.

E.4. HPLC analysis

Product analysis was similar to that described previously [49]. Perchloric acid (7%) was added to the reaction mixture at a volume of 10 μL/100 μL reaction, followed by addition of 2.5 mg/100 μL ascorbic acid. Samples were vortexed, centrifuged, and 50 μL of the supernatant were analyzed using an Alliance HPLC system (Waters, Milford, MA) equipped with a photodiode array detector. Separation was achieved on a Microsorb-MV phenyl column (250 x 4.6 mm, particle diameter 5 μm, Varian, Palo Alto, CA). Conditions with solvent A (1% triethylamine (v/v) and 0.8% acetic acid in H_2O) and solvent B (acetonitrile) used to elute propranolol and products were: 0-10 min., A:B 75:25; 10-20 min., linear gradient to A:B 25:75; 20-22 min., A:B 75:25; 22-27 min., linear gradient to A:B 75:25. Metabolites were identified by retention times and absorption spectra and were quantified by comparison to peak areas of known amounts of synthetic standards.

E.5. Saturation mutagenesis

Using A1-9C1 as the parent sequence, a mutant library was generated in which active site residues A74, L75, V78, P81, A82, F87 and T88 were randomly mutated. The primers were designed so that each clone would contain, on average, two amino acid substitutions [50]. The sequence of the forward primer was 5'-CTCGCTTTGATAAAAA*CTTAAG*TCAA**RSRSXX**AAATTT**RXW**CGTGAT**XXXRS**

WGGAGACGGGTTG**RSSWSW**AGCTGGACG-3', in which R=85%G, 5% of C, A and T; S=85%C, 5% of G, C and T; W=85% A, 5% of G, C and T; X=85% T, 5% of G, C and A. The bold letter shows the doped nucleotides, and italic sequence is a *BfrI* restriction site. The reverse primer sequence was 5'-CGTACTATGGTTTGCTTTGACGC-3'. Silent mutation A321G was introduced to the parent 9C1 gene to disrupt a *BrfI* restriction site for subsequent cloning purposes. Polymerase chain reaction was performed with *Pfu* Turbo DNA polymerase using these primers and the mutated A1-9C1 gene (with A321G) as the template. PCR conditions were 94 ^{o}C (1 min.), {94 ^{o}C (30 s) -> 63 ^{o}C (30 s) -> 72 ^{o}C (2 min.)} x 30 cycles, 72 ^{o}C (10 min.). Following agarose gel electrophoresis, the band corresponding to the P450 gene was excised from the gel and purified. The purified DNA fragment was treated with EcoRI and BfrI and ligated into the pCWori plasmid. The ligated DNA was transformed into catalase-deficient *E. coli*, and the resulting clones were screened as described below.

E.6. Random mutagenesis

Error-prone PCR was performed on the BM3-H DE10 gene with the GeneMorph PCR Mutagenesis kit (Stratagene, La Jolla, CA) according to the manufacturers' protocol, with 5-50 ng of target DNA. The reaction mixture was treated with DpnI for template degradation followed by agarose gel purification. The purified DNA fragment was digested with EcoRI and BamHI, gel purified, and ligated into the pCWori vector. The ligation product was used to transform the catalase-deficient *E. coli* strain, and the transformants were plated on LB agar medium supplemented with 100 μg/mL of ampicillin.

E.7. Screening for increased activity on propranolol

Single colonies were picked and inoculated into 500 μL of Luria-Bertani (LB) medium supplemented with 100 μg/mL ampicillin, in 96 well deep-well plates. The plates were incubated overnight at 30 ^{o}C with shaking and 80% humidity. 150 μL of the pre-cultures were inoculated into 850 μL of TB supplemented with 100 μg/mL ampicillin, 25 μg/mL thiamine, 0.5 mM IPTG and 1 mM δ-ALA, and grown under the same conditions for another 20 hours. The plates were then centrifuged to pellet the cells, supernatants discarded, and frozen overnight at -20 ^{o}C. Cells were resuspended in 300 μL of 100 mM Epps pH 8.2 containing 0.5 mg/mL lysozyme and 2 units/mL DNaseI. After 60 minutes at 37 ^{o}C, the lysates were centrifuged, and the supernatant was used for the activity assay. To 80 μL of the cleared supernatant, 30 μL of Epps pH 8.2 containing propranolol was added followed by 10 μL of 12 mM H_2O_2. Final concentrations in the reaction mixture were 1 mM propranolol, 1% DMSO, 1% acetone and 1 mM H_2O_2. After 90 minutes at

room temperature, the 4-AAP assay was used to quantify the products of propranolol bioconversions.

F. References

1. Sono M, Roach MP, Coulter ED, Dawson JH (1996) Heme-Containing Oxygenases. Chem. Rev. 96: 2841-2888.
2. Lewis DFV (2001) Guide to Cytochromes P450: structure and function. London ; New York: Taylor & Francis.
3. Guengerich FP (2003) Cytochromes P450, drugs, and diseases. Mol. Interv. 3: 194-204.
4. Lewis DF (2004) 57 varieties: the human cytochromes P450. Pharmacogenomics 5: 305-318.
5. Guengerich FP (2001) Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. Chem. Res. Toxicol. 14: 611-650.
6. Johnson MD, Zuo H, Lee K, P. TJ, Rae JM, et al. (2004) Pharmacological characterization of 4-hydroxy-N-desmethyl tamoxifen, a novel metabolite of tamoxifen. Breast Cancer Res. Treat. 207: 1-9.
7. Markham A, Wagstaff AJ (1998) Fexofenadine. Drugs 55: 269-274.
8. Parikh A, Gillam EM, Guengerich FP (1997) Drug metabolism by Escherichia coli expressing human cytochromes P450. Nat. Biotechnol. 15: 784-788.
9. Rushmore TH, Reider PJ, Slaughter D, Assang C, Shou M (2000) Bioreactor systems in drug metabolism: synthesis of cytochrome P450-generated metabolites. Metab. Eng. 2: 115-125.
10. Vail RB, Homann MJ, Hanna I, Zaks A (2005) Preparative synthesis of drug metabolites using human cytochrome P450s 3A4, 2C9 and 1A2 with NADPH-P450 reductase expressed in Escherichia coli. J. Ind Microbiol. Biotechnol. 32: 67-74.
11. Guengerich FP, Gillam EM, Shimada T (1996) New applications of bacterial systems to problems in toxicology. Crit. Rev. Toxicol. 26: 551-583.
12. Noble MA, Miles CS, Chapman SK, Lysek DA, MacKay AC, et al. (1999) Roles of key active-site residues in flavocytochrome P450 BM3. Biochem. J. 339 (Pt 2): 371-379.
13. Graham-Lorence S, Truan G, Peterson JA, Falck JR, Wei S, et al. (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: 1127-1135.
14. Li QS, Ogawa J, Schmid RD, Shimizu S (2001) Residue size at position 87 of cytochrome P450 BM-3 determines its stereoselectivity in propylbenzene and 3-chlorostyrene oxidation. FEBS Lett. 508: 249-252.
15. Kubo T, Peters MW, Meinhold P, Arnold FH (2005) Enantioselective epoxidation of terminal alkenes to (R)- and (S)-epoxides by engineered cytochromes P450 BM-3. Chemistry, Eur. J. 12: 1216-1220.
16. Meinhold P, Peters MW, Hartwick A, Hernandez AR, Arnold FH (2006) Engineering cytochrome P450 BM3 for terminal alkane hydroxylation. Adv. Synth. Catal. 348: 763-772.
17. Peters MW, Meinhold P, Glieder A, Arnold FH (2003) Regio- and enantioselective alkane hydroxylation with engineered cytochromes P450 BM-3. J. Am. Chem. Soc. 125: 13442-13450.

18. Munzer DF, Meinhold P, Peters MW, Feichtenhofer S, Griengl H, et al. (2005) Stereoselective hydroxylation of an achiral cyclopentanecarboxylic acid derivative using engineered P450s BM-3. Chem. Commun.: 2597-2599.
19. Landwehr M, Hochrein L, Otey CR, Kasrayan A, Backvall JE, et al. (2006) Enantioselective alpha-hydroxylation of 2-arylacetic acid derivatives and buspirone catalyzed by engineered cytochrome P450 BM-3. J. Am. Chem. Soc. 128: 6058-6059.
20. Cirino PC, Arnold FH (2003) A self-sufficient peroxide-driven hydroxylation biocatalyst. Angew Chem. Int. Ed. Engl. 42: 3299-3301.
21. Shand DG (1975) Drug therapy: Propranolol. N. Engl. J. Med. 293: 280-285.
22. Schwaneberg U, Schmidt-Dannert C, Schmitt J, Schmid RD (1999) A continuous spectrophotometric assay for P450 BM-3, a fatty acid hydroxylating enzyme, and its mutant F87A. Anal. Biochem. 269: 359-366.
23. Masubuchi Y, Hosokawa S, Horie T, Suzuki T, Ohmori S, et al. (1994) Cytochrome P450 isozymes involved in propranolol metabolism in human liver microsomes. The role of CYP2D6 as ring-hydroxylase and CYP1A2 as N-desisopropylase. Drug Metab. Dispos. 22: 909-915.
24. Fitzgerald JD, O'Donnell SR (1971) Pharmacology of 4-hydroxypropranolol, a metabolite of propranolol. Br. J. Pharmacol. 43: 222-235.
25. Oatis JE, Jr., Russell MP, Knapp DR, Walle T (1981) Ring-hydroxylated propranolol: synthesis and beta-receptor antagonist and vasodilating activities of the seven isomers. J. Med. Chem. 24: 309-314.
26. Boersma MG, Primus JL, Koerts J, Veeger C, Rietjens IM (2000) Heme-(hydro)peroxide mediated O- and N-dealkylation. A study with microperoxidase. Eur. J. Biochem. 267: 6673-6678.
27. He K, He YA, Szklarz GD, Halpert JR, Correia MA (1996) Secobarbital-mediated inactivation of cytochrome P450 2B1 and its active site mutants. Partitioning between heme and protein alkylation and epoxidation. J. Biol. Chem. 271: 25864-25872.
28. Raner GM, Hatchell JA, Dixon MU, Joy TL, Haddy AE, et al. (2002) Regioselective peroxo-dependent heme alkylation in P450(BM3)-F87G by aromatic aldehydes: effects of alkylation on cataysis. Biochemistry 41: 9601-9610.
29. Vogt W (1995) Oxidation of methionyl residues in proteins: tools, targets, and reversal. Free Radic Biol Med 18: 93-105.
30. Schechter Y, Burstein Y, Patchornik A (1975) Proceedings: Selective oxidation of methionine residues in proteins. Isr J Med Sci 11: 1171.
31. Estell DA, Graycar TP, Wells JA (1985) Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. J. Biol. Chem. 260: 6518-6521.
32. Schunemann V, Lendzian F, Jung C, Contzen J, Barra AL, et al. (2004) Tyrosine radical formation in the reaction of wild type and mutant cytochrome P450cam with peroxy acids: a multifrequency EPR study of intermediates on the millisecond time scale. J Biol Chem 279: 10919-10930.
33. Giulivi C, Cadenas E (1998) Heme protein radicals: formation, fate, and biological consequences. Free Radic. Biol. Med. 24: 269-279.
34. Villegas JA, Mauk AG, Vazquez-Duhalt R (2000) A cytochrome c variant resistant to heme degradation by hydrogen peroxide. Chem Biol 7: 237-244.

35. Haines DC, Tomchick DR, Machius M, Peterson JA (2001) Pivotal role of water in the mechanism of P450BM-3. Biochemistry 40: 13456-13465.
36. Nakagawa S, Ishino S, Teshiba S (1996) Construction of catalase deficient Escherichia coli strains for the production of uricase. Biosci. Biotechnol. Biochem. 60: 415-420.
37. Li H, Poulos TL (1997) The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. Nat. Struct. Biol. 4: 140-146.
38. Li QS, Ogawa J, Schmid RD, Shimizu S (2001) Engineering cytochrome P450 BM-3 for oxidation of polycyclic aromatic hydrocarbons. Appl. Environ. Microbiol. 67: 5735-5739.
39. Cirino PC, Arnold FH (2002) Regioselectivity and activity of cytochrome P450 BM-3 and mutant F87A in reactions driven by hydrogen peroxide. Adv. Synth. Catal. 344: 932-937.
40. Li QS, Ogawa J, Shimizu S (2001) Critical role of the residue size at position 87 in H2O2- dependent substrate hydroxylation activity and H2O2 inactivation of cytochrome P450BM-3. Biochem. Biophys. Res. Commun. 280: 1258-1261.
41. Meinhold P, Peters MW, Chen MMY, Takahashi K, Arnold FH (2005) Direct conversion of ethane to ethanol by engineered cytochrome P450 BM3. Chembiochem 6: 1765-1768.
42. Carmichael AB, Wong LL (2001) Protein engineering of Bacillus megaterium CYP102.The oxidation of polycyclic aromatic hydrocarbons. Eur. J. Biochem. 268: 3117-3125.
43. Coltart DJ, Shand DG (1970) Plasma propranolol levels in the quantitative assessment of beta-adrenergic blockade in man. Br. Med. J. 3: 731-734.
44. Oatis JE, Jr., Walle T, Daniell HB, Gaffney TE, Knapp DR (1985) Synthesis of 4'-hydroxypropranolol sulfate, a major non-beta-blocking propranolol metabolite in man. J. Med. Chem. 28: 822-824.
45. Bichara N, Ching MS, Blake CL, Ghabrial H, Smallwood RA (1996) Propranolol hydroxylation and N-desisopropylation by cytochrome P4502D6: studies using the yeast-expressed enzyme and NADPH/O2 and cumene hydroperoxide-supported reactions. Drug Metab. Dispos. 24: 112-118.
46. Maurer SC, Kuhnel K, Kaysser LA, Eiben S, Schmid RD (2005) Catalytic hydroxylation in biphasic systems using CYP102A1 mutants. Adv. Synth. Catal. 347: 1090-1098.
47. Barnes HJ, Arlotto MP, Waterman MR (1991) Expression and enzymatic activity of recombinant cytochrome P450 17 alpha-hydroxylase in Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. 88: 5597-5601.
48. Schenkman JB, Jansson I (1998) Spectral Analyses of Cytochromes P450. In: Shephard EA, editor. Cytochrome P450 Protocols. Totowa, New Jersey: Human Press, Inc. pp. 25-33.
49. Upthagrove AL, Nelson WL (2001) Importance of amine pKa and distribution coefficient in the metabolism of fluorinated propranolol derivatives. Preparation, identification of metabolite regioisomers, and metabolism by CYP2D6. Drug Metab. Dispos. 29: 1377-1388.

50. Suzuki M, Christians FC, Kim B, Skandalis A, Black ME, et al. (1996) Tolerance of different proteins for amino acid diversity. Mol. Divers. 2: 111-118.

Appendix A

Nucleotide and amino acid sequences

This appendix contains a ClustalW alignment of the heme domains of CYP102A1, CYP102A2 and CYP102A3 amino acid sequences with the F87A mutation (in CYP102A1) and F88A mutation (in CYP102A2 and CYP102A3). Following these are the different amino acid and nucleotide sequences of the genes and proteins used in this thesis. The accession numbers for the genes and gene products discussed are CYP102A1 (J04832 and CAB12544) and CYP102A3 (U93874) (http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi).

Table A.1. ClustalW amino acid sequence alignment of the heme domain of CYP102A1 (residues 1-463), CYP102A2 and CYP102A3 (residues 1-466). Numbering is according to the CYP102A1 sequence.

```
          1                                                          60
CYP102A1  .TIKEMPQPKTFGELKNLPLLNTDKPVQALMKIADELGEIFKFEAPGRVTRYLSSQRLIKE
CYP102A2  KETSPIPQPKTFGPLGNLPLIDKDKPTLSLIKLAEEQGPIFQIHTPAGTTIVVSGHELVKE
CYP102A3  KQASAIPQPKTYGPLKNLPHLEKEQLSQSLWRIADELGPIFRFDFPGVSSVFVSGHNLVAE

          61                                                         121
CYP102A1  ACDESRFDKNLSQALKFVRDFAGDGLATSWTHEKNWKKAHNILLPSFSQQAMKGYHAMMVD
CYP102A2  VCDEERFDKSIEGALEKVRAFSGDGLATSWTHEPNWRKAHNILMPTFSQRAMKDYHEKMVD
CYP102A3  VCDEKRFDKNLGKGLQKVREFGGDGLATSWTHEPNWQKAHRILLPSFSQKAMKGYHSMMLD

          122                                                        183
CYP102A1  IAVQLVQKWERLNADEHIEVPEDMTRLTLDTIGLCGFNYRFNSFYRDQPHPFITSMVRALD
CYP102A2  IAVQLIQKWARLNPNEAVDVPGDMTRLTLDTIGLCGFNYRFNSYYRETPHPFINSMVRALD
CYP102A3  IATQLIQKWSRLNPNEEIDVADDMTRLTLDTIGLCGFNYRFNSFYRDSQHPFITSMLRALK

          184                                                        245
CYP102A1  EAMNKLQRANPDDPAYDENKRQFQEDIKVMNDLVDKIIADRKASGEQ.SDDLLTHMLNGKD
CYP102A2  EAMHQMQRLDVQDKLMVRTKRQFRYDIQTMFSLVDSIIAERRANGDQDEKDLLARMLNVED
CYP102A3  EAMNQSKRLGLQDKMMVKTKLQFQKDIEVMNSLVDRMIAERKANPDENIKDLLSLMLYAKD

          246                                                        307
CYP102A1  PETGEPLDDENIRYQIITFLIAGHETTSGLLSFALYFLVKNPHVLQKAAEEAARVLVDPVP
CYP102A2  PETGEKLDDENIRFQIITFLIAGHETTSGLLSFATYFLLKHPDKLKKAYEEVDRVLTDAAP
CYP102A3  PVTGETLDDENIRYQIITFLIAGHETTSGLLSFAIYCLLTHPEKLKKAQEEADRVLTDDTP


          308                                                        369
CYP102A1  SYKQVKQLKYVGMVLNEALRLWPTAPAFSLYAKEDTVLGGEYPLEKGDELMVLIPQLHRDK
CYP102A2  TYKQVLELTYIRMILNESLRLWPTAPAFSLYPKEDTVIGGKFPITTNDRISVLIPQLHRDR
CYP102A3  EYKQIQQLKYIRMVLNETLRLYPTAPAFSLYAKEDTVLGGEYPISKGQPVTVLIPKLHRDQ

          370                                                        431
CYP102A1  TIWGDDVEEFRPERFENPSAIPQHAFKPFGNGQRACIGQQFALHEATLVLGMMLKHFDFED
CYP102A2  DAWGKDAEEFRPERFEHQDQVPHHAYKPFGNGQRACIGMQFALHEATLVLGMILKYFTLID
CYP102A3  NAWGPDAEDFRPERFEDPSSIPHHAYKPFGNGQRACIGMQFALQEATMVLGLVLKHFELIN

          432                                 463
CYP102A1  HTNYELDIKETLTLKPEGFVVKAKSKKIPLGGIPSPST.
CYP102A2  HENYELDIKQTLTLKPGDFHISVQSRHQEAIHADVQAAE
CYP102A3  HTGYELKIKEALTIKPDDFKITVKPRKTAAINVQRKEQA
```

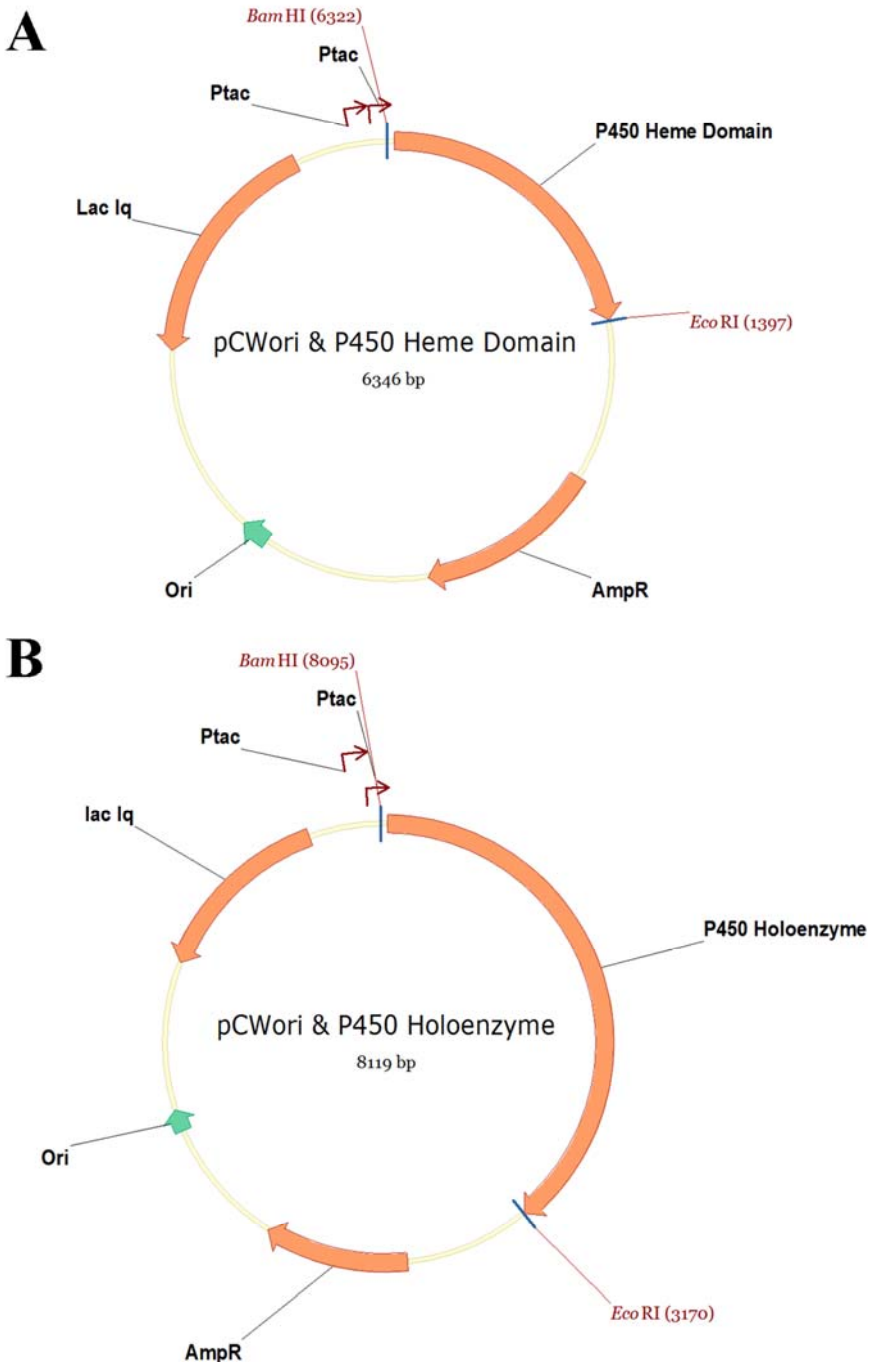

Figure A.1. Plasmid map of pCWori [1] with A) P450 heme domain or B) P450 holoenzyme (heme domain plus reductase domain). Expression is under the control of a double tac promoter (Ptac)[2] and has a β-lactamase as an ampicillin resistance marker (AmpR). The lacIq gene (lac Iq) which expresses the Lac I repressor and origin of replication (Ori). BamHI and EcoRI are typical cloning sites.

Nucleotide sequences

CYP102A1 Heme Domain – F87A – library

```
   1 ATGACAATTA AAGAAATGCC TCAGCCAAAA ACGTTTGGAG AGCTTAAAAA TTTACCGTTA
  61 TTAAACACAG ATAAACCGGT TCAAGCTTTG ATGAAAATTG CGGATGAATT AGGAGAAATC
 121 TTTAAATTCG AGGCGCCTGG TCGTGTAACG CGCTACTTAT CAAGTCAGCG TCTAATTAAA
 181 GAAGCATGCG ATGAATCACG CTTTGATAAA AACTTAAGTC AAGCGCTTAA ATTTGTACGT
 241 GATTTTGCAG GAGACGGGTT GGCCACAAGC TGGACGCATG AAAAAAATTG GAAAAAAGCG
 301 CATAATATCT TACTTCCAAG CTTCAGTCAG CAGGCAATGA AAGGCTATCA TGCGATGATG
 361 GTCGATATCG CCGTGCAGCT TGTTCAAAAG TGGGAGCGTC TAAATGCAGA TGAGCATATT
 421 GAAGTACCGG AAGACATGAC ACGTTTAACG CTTGATACGA TTGGTCTTTG CGGCTTTAAC
 481 TATCGCTTTA ACAGCTTTTA CCGAGATCAG CCTCATCCAT TTATTACAAG TATGGTCCGT
 541 GCACTGGATG AAGCAATGAA CAAGCTGCAG CGAGCAAATC CAGACGACCC AGCTTATGAT
 601 GAAAACAAGC GCCAGTTTCA AGAAGATATC AAGGTGATGA ACGACCTAGT CGACAAAATT
 661 ATTGCAGATC GCAAAGCAAG CGGTGAACAA AGCGATGATT TATTAACGCA CATGCTAAAC
 721 GGAAAAGATC CAGAAACGGG TGAGCCGCTT GATGACGAGA ACATTCGCTA TCAAATTATT
 781 ACATTCTTAA TTGCGGGACA CGAAACAACA AGTGGTCTTT TATCATTTGC GCTGTATTTC
 841 TTAGTGAAAA ATCCACATGT ATTACAAAAA GCAGCAGAAG AAGCAGCACG AGTTCTAGTA
 901 GATCCTGTTC CAAGCTACAA ACAAGTCAAA CAGCTTAAAT ATGTCGGCAT GGTCTTAAAC
 961 GAAGCGCTGC GCTTATGGCC AACTGCTCCT GCGTTTTCCC TATATGCAAA AGAAGATACG
1021 GTGCTTGGAG GAGAATATCC TTTAGAAAAA GGCGACGAAC TAATGGTTCT GATTCCTCAG
1081 CTTCACCGTG ATAAAACAAT TTGGGGTGAC GATGTGGAAG AGTTCCGTCC AGAGCGTTTT
1141 GAAAATCCAA GTGCGATTCC GCAGCATGCG TTTAAACCGT TTGGAAACGG TCAGCGTGCG
1201 TGTATCGGTC AGCAGTTCGC TCTTCATGAA GCAACGCTGG TACTTGGTAT GATGCTAAAA
1261 CACTTTGACT TTGAAGATCA TACAAACTAC GAGCTGGATA TTAAAGAAAC TTTAACGTTA
1321 AAACCTGAAG GCTTTGTGGT AAAAGCAAAA TCGAAAAAAA TTCCGCTTGG CGGTATTCCT
1381 TCACCTAGCA CT
```

CYP102A2 Heme Domain – F88A – library

```
   1 ATGAAGGAAA CAAGCCCGAT TCCTCAGCCG AAGACGTTTG GGCCGCTCGG CAATTTGCCT
  61 TTAATTGATA AAGACAAACC GACGCTTTCG CTGATCAAAC TGGCGGAAGA ACAGGGCCCG
 121 ATTTTTCAAA TCCATACACC CGCGGGCACG ACCATTGTAG TGTCCGGCCA TGAATTGGTG
 181 AAAGAGGTTT GTGATGAAGA ACGGTTTGAT AAAAGCATTG AAGGCGCCTT GGAAAAGGTT
 241 CGCGCATTTT CCGGTGACGG ATTGGCCACT AGTTGGACGC ATGAGCCTAA CTGGAGAAAA
 301 GCGCACAACA TTCTGATGCC GACGTTCAGC CAGCGGGCCA TGAAGGACTA TCATGAGAAA
 361 ATGGTCGATA TCGCTGTTCA GCTCATTCAA AAATGGGCAA GGCTCAACCC GAATGAAGCA
 421 GTCGATGTCC CGGGAGATAT GACCCGGCTG ACGCTCGACA CCATTGGGCT ATGCGGGTTT
 481 AACTACCGCT TTAACAGTTA CTACAGAGAA ACGCCCCACC CGTTTATCAA CAGCATGGTG
 541 CGGGCGCTTG ATGAAGCGAT GCATCAAATG CAGCGGCTTG ATGTTCAAGA TAAGCTTATG
 601 GTCAGAACAA AGCGGCAATT CCGCTATGAT ATTCAAACGA TGTTTTCGTT AGTCGACAGC
 661 ATTATTGCAG AGCGCAGGGC GAATGGAGAC CAGGATGAAA AAGATTTGCT CGCCCGCATG
 721 CTGAATGTGG AAGATCCGGA AACTGGTGAA AAGCTCGACG ACGAAAATAT CCGCTTTCAG
 781 ATCATCACGT TTTTGATTGC CGGCCATGAA ACAACGAGCG GCCTGCTTTC CTTTGCGACT
 841 TACTTTTTAT TGAAGCATCC TGACAAACTG AAAAAGGCGT ATGAAGAGGT CGATCGGGTG
 901 CTGACGGATG CAGCGCCGAC CTATAAACAA GTGCTGGAGC TTACATACAT ACGGATGATT
 961 TTAAATGAAT CACTGCGCTT ATGGCCGACA GCTCCGGCTT TCAGCCTTTA TCCAAAAGAA
1021 GACACAGTCA TTGGCGGAAA ATTTCCGATC ACGACGAATG ACAGAATTTC TGTGCTGATT
1081 CCGCAGCTTC ATCGTGATCG AGACGCTTGG GGAAAGGACG CAGAAGAGTT CCGGCCGGAA
1141 CGGTTTGAGC ATCAGGACCA AGTGCCTCAT CATGCGTACA AACCATTCGG AAATGGACAA
1201 CGGGCCTGTA TCGGCATGCA GTTTGCCCTT CATGAAGCCA CACTTGTGTT AGGCATGATT
1261 CTAAAATATT TCACATTGAT TGATCATGAG AATTATGAGC TTGATATCAA ACAAACCTTA
1321 ACACTTAAGC CGGGCGATTT TCACATCAGT GTTCAAAGCC GTCATCAGGA AGCCATTCAT
1381 GCAGACGTCC AGGCAGCTGA A
```

CYP102A3 Heme Domain – F88A - library

```
   1 ATGAAACAGG CAAGCGCAAT ACCTCAGCCC AAAACATACG GACCTTTAAA AAATCTTCCG
  61 CATCTGGAAA AAGAACAGCT TTCTCAATCC TTATGGCGGA TAGCTGATGA ATTGGGACCG
 121 ATTTTCCGTT TTGATTTTCC GGGAGTATCC AGTGTTTTTG TGTCCGGCCA CAATCTTGTG
 181 GCTGAAGTGT GTGATGAAAA ACGCTTTGAC AAGAACCTTG GCAAAGGCTT GCAAAAGGTG
 241 CGTGAGTTCG GGGGAGATGG CTTAGCCACT AGTTGGACGC ACGAACCGAA CTGGCAAAAA
 301 GCCCACCGCA TTTTGCTGCC GAGTTTTAGT CAAAAAGCGA TGAAAGGCTA TCATTCTATG
```

```
 361 ATGCTGGATA TCGCAACCCA GCTGATTCAA AAGTGGAGCC GGTTAAACCC CAATGAAGAA
 421 ATTGATGTAG CGGACGATAT GACACGTCTG ACGCTTGATA CGATTGGGTT ATGCGGGTTT
 481 AACTATCGAT TCAACAGCTT TTACCGTGAT TCACAGCATC CGTTTATCAC CAGTATGCTC
 541 CGTGCCTTAA AAGAGGCGAT GAATCAATCG AAAAGACTGG GCCTGCAAGA TAAAATGATG
 601 GTGAAAACGA AGCTGCAGTT CCAAAAGGAT ATAGAAGTCA TGAACTCCCT GGTCGACAGA
 661 ATGATAGCGG AGCGAAAGGC GAATCCGGAT GAAAACATTA AGGATCTCTT GTCTCTCATG
 721 CTTTATGCCA AAGATCCAGT AACGGGTGAA ACGCTGGATG ACGAAAACAT TCGATACCAA
 781 ATCATCACAT TTTTAATTGC TGGACATGAG ACAACAAGCG GGTTGCTATC CTTTGCGATT
 841 TATTGTCTGC TTACACATCC GGAAAAACTG AAAAAAGCTC AGGAGGAAGC GGATCGCGTG
 901 TTAACGGATG ACACGCCTGA ATATAAACAA ATCCAGCAGC TCAAATACAT TCGGATGGTT
 961 TTAAATGAAA CCCTCAGACT GTATCCAACA GCTCCGGCTT TTTCTCTATA TGCGAAGGAG
1021 GATACTGTTC TTGGCGGGGA ATATCCGATC AGCAAAGGGC AGCCAGTCAC TGTTTTAATT
1081 CCAAAACTGC ACCGGGATCA AAACGCTTGG GGACCGGATG CGGAAGATTT CCGTCCGGAA
1141 CGGTTTGAAG ATCCTTCAAG TATCCCTCAC CATGCGTATA AGCCGTTTGG AAACGGACAG
1201 CGCGCTTGTA TTGGCATGCA GTTTGCTCTT CAAGAAGCGA CAATGGTTCT CGGTCTTGTA
1261 TTAAAGCATT TTGAATTGAT AAACCATACT GGCTACGAAC TAAAAATCAA AGAAGCATTA
1321 ACGATCAAGC CGGATGATTT TAAAATTACT GTGAAACCGC GAAAAACAGC GGCAATCAAT
1381 GTACAGAGAA AAGAACAGGC A
```

CYP102A1 Holoenzyme

```
   1 ATGACAATTA AAGAAATGCC TCAGCCAAAA ACGTTTGGAG AGCTTAAAAA TTTACCGTTA
  61 TTAAACACAG ATAAACCGGT TCAAGCTTTG ATGAAAATTG CGGATGAATT AGGAGAAATC
 121 TTTAAATTCG AGGCGCCTGG TCGTGTAACG CGCTACTTAT CAAGTCAGCG TCTAATTAAA
 181 GAAGCATGCG ATGAATCACG CTTTGATAAA AACTTAAGTC AAGCGCTTAA ATTTGTACGT
 241 GATTTTGCAG GAGACGGGTT ATTTACAAGC TGGACGCATG AAAAAAATTG GAAAAAAGCG
 301 CATAATATCT TACTTCCAAG CTTCAGTCAG CAGGCAATGA AAGGCTATCA TGCGATGATG
 361 GTCGATATCG CCGTGCAGCT TGTTCAAAAG TGGGAGCGTC TAAATGCAGA TGAGCATATT
 421 GAAGTACCGG AAGACATGAC ACGTTTAACG CTTGATACAA TTGGTCTTTG CGGCTTTAAC
 481 TATCGCTTTA ACAGCTTTTA CCGAGATCAG CCTCATCCAT TTATTACAAG TATGGTCCGT
 541 GCACTGGATG AAGCAATGAA CAAGCTGCAG CGAGCAAATC CAGACGACCC AGCTTATGAT
 601 GAAAACAAGC GCCAGTTTCA AGAAGATATC AAGGTGATGA ACGACCTAGT AGATAAAATT
 661 ATTGCAGATC GCAAAGCAAG CGGTGAACAA AGCGATGATT TATTAACGCA TATGCTAAAC
 721 GGAAAAGATC CAGAAACGGG TGAGCCGCTT GATGACGAGA ACATTCGCTA TCAAATTATT
 781 ACATTCTTAA TTGCGGGACA CGAAACAACA AGTGGTCTTT TATCATTTGC GCTGTATTTC
 841 TTAGTGAAAA ATCCACATGT ATTACAAAAA GCAGCAGAAG AAGCAGCACG AGTTCTAGTA
 901 GATCCTGTTC CAAGCTACAA ACAAGTCAAA CAGCTTAAAT ATGTCGGCAT GGTCTTAAAC
 961 GAAGCGCTGC GCTTATGGCC AACTGCTCCT GCGTTTTCCC TATATGCAAA AGAAGATACG
1021 GTGCTTGGAG GAGAATATCC TTTAGAAAAA GGCGACGAAC TAATGGTTCT GATTCCTCAG
1081 CTTCACCGTG ATAAAACAAT TTGGGGAGAC GATGTGGAAG AGTTCCGTCC AGAGCGTTTT
1141 GAAAATCCAA GTGCGATTCC GCAGCATGCG TTTAAACCGT TTGGAAACGG TCAGCGTGCG
1201 TGTATCGGTC AGCAGTTCGC TCTTCATGAA GCAACGCTGG TACTTGGTAT GATGCTAAAA
1261 CACTTTGACT TTGAAGATCA TACAAACTAC GAGCTCGATA TTAAAGAAAC TTTAACGTTA
1321 AAACCTGAAG GCTTTGTGGT AAAAGCAAAA TCGAAAAAAA TTCCGCTTGG CGGTATTCCT
1381 TCACCTAGCA CTGAACAGTC TGCTAAAAAA GTACGCAAAA AGGCAGAAAA CGCTCATAAT
1441 ACGCCGCTGC TTGTGCTATA CGGTTCAAAT ATGGGAACAG CTGAAGGAAC GGCGCGTGAT
1501 TTAGCAGATA TTGCAATGAG CAAAGGATTT GCACCGCAGG TCGCAACGCT TGATTCACAC
1561 GCCGGAAATC TTCCGCGCGA AGGAGCTGTA TTAATTGTAA CGGCGTCTTA TAACGGTCAT
1621 CCGCCTGATA ACGCAAAGCA ATTTGTCGAC TGGTTAGACC AAGCGTCTGC TGATGAAGTA
1681 AAAGGCGTTC GCTACTCCGT ATTTGGATGC GGCGATAAAA ACTGGGCTAC TACGTATCAA
1741 AAAGTGCCTG CTTTTATCGA TGAAACGCTT GCCGCTAAAG GGGCAGAAAA CATCGCTGAC
1801 CGCGGTGAAG CAGATGCAAG CGACGACTTT GAAGGCACAT ATGAAGAATG GCGTGAACAT
1861 ATGTGGAGTG ACGTAGCAGC CTACTTTAAC CTCGACATTG AAAACAGTGA AGATAATAAA
1921 TCTACTCTTT CACTTCAATT TGTCGACAGC GCCGCGGATA TGCCGCTTGC GAAAATGCAC
1981 GGTGCGTTTT CAACGAACGT CGTAGCAAGC AAAGAACTTC AACAGCCAGG CAGTGCACGA
2041 AGCACGCGAC ATCTTGAAAT TGAACTTCCA AAAGAAGCTT CTTATCAAGA AGGAGATCAT
2101 TTAGGTGTTA TTCCTCGCAA CTATGAAGGA ATAGTAAACC GTGTAACAGC AAGGTTCGGC
2161 CTAGATGCAT CACAGCAAAT CCGTCTGGAA GCAGAAGAAG AAAAATTAGC TCATTTGCCA
2221 CTCGCTAAAA CAGTATCCGT AGAAGAGCTT CTGCAATACG TGGAGCTTCA AGATCCTGTT
2281 ACGCGCACGC AGCTTCGCGC AATGGCTGCT AAAACGGTCT GCCCGCCGCA TAAAGTAGAG
2341 CTTGAAGCCT TGCTTGAAAA GCAAGCCTAC AAAGAACAAG TGCTGGCAAA ACGTTTAACA
2401 ATGCTTGAAC TGCTTGAAAA ATACCCGGCG TGTGAAATGA AATTCAGCGA ATTTATCGCC
2461 CTTCTGCCAA GCATACGCCC GCGCTATTAC TCGATTTCTT CATCACCTCG TGTCGATGAA
```

```
2521 AAACAAGCAA GCATCACGGT CAGCGTTGTC TCAGGAGAAG CGTGGAGCGG ATATGGAGAA
2581 TATAAAGGAA TTGCGTCGAA CTATCTTGCC GAGCTGCAAG AAGGAGATAC GATTACGTGC
2641 TTTATTTCCA CACCGCAGTC AGAATTTACG CTGCCAAAAG ACCCTGAAAC GCCGCTTATC
2701 ATGGTCGGAC CGGGAACAGG CGTCGCGCCG TTTAGAGGCT TTGTGCAGGC GCGCAAACAG
2761 CTAAAAGAAC AAGGACAGTC ACTTGGAGAA GCACATTTAT ACTTCGGCTG CCGTTCACCT
2821 CATGAAGACT ATCTGTATCA AGAAGAGCTT GAAAACGCCC AAAGCGAAGG CATCATTACG
2881 CTTCATACCG CTTTTTCTCG CATGCCAAAT CAGCCGAAAA CATACGTTCA GCACGTAATG
2941 GAACAAGACG GCAAGAAATT GATTGAACTT CTTGATCAAG GAGCGCACTT CTATATTTGC
3001 GGAGACGGAA GCCAAATGGC ACCTGCCGTT GAAGCAACGC TTATGAAAAG CTATGCTGAC
3061 GTTCACCAAG TGAGTGAAGC AGACGCTCGC TTATGGCTGC AGCAGCTAGA AGAAAAAGGC
3121 CGATACGCAA AAGACGTGTG GGCTGGGCAT CATCATCATC ATCAT
```

CYP102A2 Holoenzyme

```
   1 ATGAAGGAAA CAAGCCCGAT TCCTCAGCCG AAGACGTTTG GGCCGCTCGG CAATTTGCCT
  61 TTAATTGATA AAGACAAACC GACGCTTTCG CTGATCAAAC TGGCGGAAGA ACAGGGCCCG
 121 ATTTTTCAAA TCCATACACC CGCGGGCACG ACCATTGTAG TGTCCGGCCA TGAATTGGTG
 181 AAAGAGGTTT GTGATGAAGA ACGGTTTGAT AAAAGCATTG AAGGCGCCTT GGAAAAGGTT
 241 CGCGCATTTT CCGGTGACGG ATTGTTTACG AGCTGGACGC ATGAGCCTAA CTGGAGAAAA
 301 GCGCACAACA TTCTGATGCC GACGTTCAGC CAGCGGGCCA TGAAGGACTA TCATGAGAAA
 361 ATGGTCGATA TCGCTGTTCA GCTCATTCAA AAATGGGCAA GGCTCAACCC GAATGAAGCA
 421 GTCGATGTCC CGGGAGATAT GACCCGGCTG ACGCTCGACA CCATTGGGCT ATGCGGGTTT
 481 AACTACCGCT TTAACAGTTA CTACAGAGAA ACGCCCCACC CGTTTATCAA CAGCATGGTG
 541 CGGGCGCTTG ATGAAGCGAT GCATCAAATG CAGCGGCTTG ATGTTCAAGA TAAGCTTATG
 601 GTCAGAACAA AGCGGCAATT CCGCTATGAT ATTCAAACGA TGTTTTCGTT AGTCGACAGC
 661 ATTATTGCAG AGCGCAGGGC GAATGGAGAC CAGGATGAAA AAGATTTGCT CGCCCGCATG
 721 CTGAATGTGG AAGATCCGGA AACTGGTGAA AAGCTCGACG ACGAAAATAT CCGCTTTCAG
 781 ATCATCACGT TTTTGATTGC CGGCCATGAA ACAACGAGCG GCCTGCTTTC CTTTGCGACT
 841 TACTTTTTAT TGAAGCATCC TGACAAACTG AAAAAGGCGT ATGAAGAGGT CGATCGGGTG
 901 CTGACGGATG CAGCGCCGAC CTATAAACAA GTGCTGGAGC TTACATACAT ACGGATGATT
 961 TTAAATGAAT CACTGCGCTT ATGGCCGACA GCTCCGGCTT TCAGCCTTTA TCCAAAAGAA
1021 GACACAGTCA TTGGCGGAAA ATTTCCGATC ACGACGAATG ACAGAATTTC TGTGCTGATT
1081 CCGCAGCTTC ATCGTGATCG AGACGCTTGG GGAAAGGACG CAGAAGAGTT CCGGCCGGAA
1141 CGGTTTGAGC ATCAGGACCA AGTGCCTCAT CATGCGTACA AACCATTCGG AAATGGACAA
1201 CGGGCCTGTA TCGGCATGCA GTTTGCCCTT CATGAAGCCA CACTTGTGTT AGGCATGATT
1261 CTAAAATATT TCACATTGAT TGATCATGAG AATTATGAGC TTGATATCAA ACAAACCTTA
1321 ACACTTAAGC CGGGCGATTT TCACATCAGT GTTCAAAGCC GTCATCAGGA AGCCATTCAT
1381 GCAGACGTCC AGGCAGCTGA AAAAGCCGCG CCTGATGAGC AAAAGGAGAA AACGGAAGCA
1441 AAGGGTGCAT CGGTCATCGG TCTTAACAAC CGCCCGCTTC TCGTGCTGTA CGGCTCAGAT
1501 ACCGGCACCG CAGAAGGCGT CGCCCGGGAG CTTGCTGATA CTGCCAGTCT TCACGGCGTA
1561 AGGACAAAGA CAGCACCTCT GAACGACCGG ATTGGAAAGC TGCCGAAAGA GGGAGCGGTT
1621 GTCATTGTGA CCTCGTCTTA TAATGGAAAG CCGCCAAGCA ATGCCGGACA ATTCGTGCAG
1681 TGGCTTCAAG AAATCAAACC GGGTGAGCTT GAGGGCGTCC ATTACGCGGT ATTTGGCTGC
1741 GGCGACCACA ACTGGGCGAG CACGTATCAA TACGTGCCGA GATTCATTGA TGAGCAGCTT
1801 GCGGAGAAAG GCGCGACTCG GTTTTCTGCG CGCGGGGAAG GGGATGTGAG CGGTGATTTT
1861 GAAGGGCAGC TTGACGAGTG GAAAAAAAGC ATGTGGGCGG ATGCCATCAA AGCATTCGGA
1921 CTTGAGCTTA ATGAAAACGC TGATAAGGAA CGAAGCACGC TGAGCCTTCA GTTTGTCAGA
1981 GGGCTGGGCG AGTCTCCGCT CGCTAGATCG TACGAAGCCT CTCACGCATC CATTGCCGAA
2041 AATCGTGAAC TCCAGTCCGC AGACAGCGAT CGAAGCACTC GCCATATCGA AATTGCATTG
2101 CCGCCGGATG TTGAATATCA AGAGGGCGAC CATCTTGGCG TATTGCCAAA AAACAGCCAA
2161 ACCAATGTCA GCCGGATTCT TCACAGATTC GGTCTGAAGG GAACCGACCA AGTGACATTG
2221 TCGGCAAGCG GCCGCAGTGC GGGGCATCTG CCATTGGGCC GTCCTGTCAG CCTGCATGAT
2281 CTTCTCAGCT ACAGCGTCGA GGTGCAGGAA GCAGCCACAA GAGCGCAAAT ACGTGAACTG
2341 GCGTCATTTA CAGTGTGTCC GCCGCATAGG CGCGAATTAG AAGAACTGTC AGCAGAGGGT
2401 GTTTATCAGG AGCAAATATT GAAAAAACGA ATTTCCATGC TGGATCTGCT TGAAAAGTAT
2461 GAAGCGTGTG ACATGCCGTT TGAACGATTT TTAGAGCTTT TACGGCCGTT AAAACCGAGA
2521 TACTATTCGA TTTCAAGCTC TCCAAGAGTG AATCCGCGGC AAGCATCGAT CACAGTCGGT
2581 GTCGTGCGCG GCCCGGCGTG GAGCGGCCGT GGCGAATACA GGGGTGTGGC ATCAAATGAT
2641 TTAGCTGAGC GTCAAGCCGG TGATGATGTC GTGATGTTTA TCCGCACACC GGAATCCCGG
2701 TTTCAGCTTC CGAAAGACCC TGAAACGCCA ATTATTATGG TCGGGCCAGG CACGGGAGTC
2761 GCGCCATTTC GCGGTTTCCT TCAAGCCCGC GATGTTTTAA AGCGGGAGGG CAAAACGCTC
2821 GGTGAGGCTC ATCTCTATTT TGGATGCAGG AACGATCGGG ATTTTATTTA CCGAGATGAG
2881 CTTGAGCGGT TTGAAAAAGA CGGAATCGTC ACTGTCCACA CAGCCTTTTC CCGAAAAGAG
```

```
2941 GGCATGCCGA AAACATATGT CCAGCATCTC ATGGCTGACC AAGCAGATAC ATTAATATCA
3001 ATCCTTGACC GCGGTGGCAG GCTTTATGTA TGCGGTGATG GCAGCAAAAT GGCCCCGGAT
3061 GTGGAGGCGG CACTTCAAAA AGCGTATCAG GCTGTCCATG GAACCGGGGA ACAAGAAGCG
3121 CAAAACTGGC TGAGACATCT GCAGGATACC GGTATGTACG CTAAGGATGT CTGGGCAGGG
3181 ATA
```

CYP102A3 Holoenzyme

```
   1 ATGAAACAGG CAAGCGCAAT ACCTCAGCCC AAAACATACG GACCTTTAAA AAATCTTCCG
  61 CATCTGGAAA AAGAACAGCT TTCTCAATCC TTATGGCGGA TAGCTGATGA ATTGGGACCG
 121 ATTTTCCGTT TTGATTTTCC GGGAGTATCC AGTGTTTTTG TGTCCGGCCA CAATCTTGTG
 181 GCTGAAGTGT GTGATGAAAA ACGCTTTGAC AAGAACCTTG GCAAAGGCTT GCAAAAGGTG
 241 CGTGAGTTCG GGGGAGATGG CTTATTTACA AGCTGGACGC ACGAACCGAA CTGGCAAAAA
 301 GCCCACCGCA TTTTGCTGCC GAGTTTTAGT CAAAAAGCGA TGAAAGGCTA TCATTCTATG
 361 ATGCTGGATA TCGCAACCCA GCTGATTCAA AAGTGGAGCC GGTTAAACCC CAATGAAGAA
 421 ATTGATGTAG CGGACGATAT GACACGTCTG ACGCTTGATA CGATTGGGTT ATGCGGGTTT
 481 AACTATCGAT TCAACAGCTT TTACCGTGAT TCACAGCATC CGTTTATCAC CAGTATGCTC
 541 CGTGCCTTAA AAGAGGCGAT GAATCAATCG AAAAGACTGG GCCTGCAAGA TAAAATGATG
 601 GTGAAAACGA AGCTGCAGTT CCAAAAGGAT ATAGAAGTCA TGAACTCCCT GGTTGATAGA
 661 ATGATAGCGG AGCGAAAGGC GAATCCGGAT GAAAACATTA AGGATCTCTT GTCTCTCATG
 721 CTTTATGCCA AAGATCCAGT AACGGGTGAA ACGCTGGATG ACGAAAACAT TCGATACCAA
 781 ATCATCACAT TTTTAATTGC TGGACATGAG ACAACAAGCG GGTTGCTATC CTTTGCGATT
 841 TATTGTCTGC TTACACATCC GGAAAAACTG AAAAAAGCTC AGGAGGAAGC GGATCGCGTG
 901 TTAACGGATG ACACGCCTGA ATATAAACAA ATCCAGCAGC TCAAATACAT TCGGATGGTT
 961 TTAAATGAAA CCCTCAGACT GTATCCAACA GCTCCGGCTT TTTCTCTATA TGCGAAGGAG
1021 GATACTGTTT TAGGCGGGGA ATATCCGATC AGCAAAGGGC AGCCAGTCAC TGTTTTAATT
1081 CCAAAACTGC ACCGGGATCA AAACGCTTGG GGACCGGATG CGGAAGATTT CCGTCCGGAA
1141 CGGTTTGAAG ATCCTTCAAG TATCCCTCAC CATGCGTATA AGCCGTTTGG AAACGGACAG
1201 CGCGCTTGTA TTGGCATGCA GTTTGCTCTT CAAGAAGCGA CAATGGTTCT CGGTCTTGTA
1261 TTAAAGCATT TTGAATTGAT AAACCATACT GGCTACGAAC TAAAAATCAA AGAAGCATTA
1321 ACGATCAAGC CGGATGATTT TAAAATTACT GTGAAACCGC GAAAAACAGC GGCAATCAAT
1381 GTACAGAGAA AAGAACAGGC AGACATCAAA GCAGAAACAA AGCCAAAAGA AACCAAACCT
1441 AAACACGGCA CACCTTTACT TGTTCTTTTT GGTTCAAATC TTGGGACAGC TGAGGGAATA
1501 GCCGGTGAAC TGGCTGCTCA AGGCCGCCAG ATGGGCTTTA CAGCTGAAAC GGCTCCGCTT
1561 GATGATTATA TCGGCAAGCT CCCTGAAGAA GGGGCAGTCG TCATTGTAAC GGCTTCTTAT
1621 AATGGGGCGC CGCCTGATAA TGCTGCCGGA TTTGTAGAGT GGCTGAAAGA GCTTGAGGAA
1681 GGCCAATTGA AAGGTGTTTC CTATGCGGTA TTCGGCTGCG GAAACCGGAG CTGGGCCAGC
1741 ACGTATCAGC GGATTCCCCG CCTGATTGAT GACATGATGA AAGCAAAGGG GGCATCGCGT
1801 TTAACAGCGA TTGGGGAAGG TGACGCCGCC GATGATTTTG AAAGCCACCG CGAGTCTTGG
1861 GAAAACCGCT TCTGGAAGGA AACGATGGAC GCATTTGATA TTAACGAAAT AGCCCAGAAA
1921 GAAGACAGGC CTTCATTATC GATTACTTTT CTCAGTGAAG CGACGGAAAC GCCGGTTGCT
1981 AAAGCATATG GCGCGTTTGA AGGGATTGTG TTAGAGAATC GAGAACTCCA GACAGCTGCC
2041 AGCACGCGTT CAACCCGCCA TATTGAATTG GAAATTCCGG CTGGTAAAAC ATATAAAGAA
2101 GGCGATCATA TCGGAATCCT GCCAAAGAAC AGCAGGGAGC TTGTTCAGCG GGTTCTCAGC
2161 CGATTCGGTT TGCAGTCCAA TCATGTGATA AAAGTAAGCG GAAGCGCTCA TATGGCTCAT
2221 CTGCCGATGG ATCGGCCAAT CAAAGTAGTG GATTTATTGT CGTCCTATGT AGAGCTGCAG
2281 GAACCGGCAT CAAGGCTTCA GCTTCGGGAG CTGGCCTCTT ATACAGTTTG TCCGCCGCAT
2341 CAAAAAGAGC TGGAACAGCT CGTTTCAGAT GATGGCATTT ACAAAGAGCA GGTACTTGCA
2401 AAACGTCTTA CCATGCTTGA TTTTTTAGAG GATTATCCTG CTTGCGAAAT GCCGTTTGAA
2461 CGGTTTTTAG CACTTTTGCC ATCACTAAAA CCGAGATACT ATTCCATTTC AAGCTCACCG
2521 AAAGTTCATG CAAATATCGT GAGCATGACG GTAGGAGTTG TGAAAGCCTC AGCATGGAGC
2581 GGCCGAGGTG AATACCGGGG TGTCGCCTCT AATTATTTAG CAGAATTGAA TACAGGTGAT
2641 GCAGCAGCTT GCTTCATTCG TACGCCGCAG TCCGGATTTC AGATGCCGAA TGATCCTGAA
2701 ACGCCTATGA TTATGGTCGG GCCGGGCACA GGAATTGCGC CATTCAGAGG CTTTATTCAG
2761 GCAAGATCGG TTTTGAAGAA GGAAGGAAGC ACCCTTGGTG AAGCACTTTT ATACTTCGGC
2821 TGCCGCCGCC CGGACCATGA CGACCTTTAC AGAGAAGAGC TGGATCAAGC GGAACAGGAC
2881 GGTTTGGTCA CAATCCGCCG ATGCTACTCG CGCGTCGAAA ACGAACCAAA AGGATATGTC
2941 CAGCACTTGC TCAAGCAAGA TACGCAGAAA TTGATGACAC TCATTGAAAA AGGGGCTCAT
3001 ATTTACGTAT GCGGTGATGG ATCGCAAATG GCTCCTGATG TAGAGAGAAC TTTGCGATTG
3061 GCATATGAAG CTGAAAAAGC AGCAAGTCAG GAAGAATCAG CTGTATGGCT GCAAAAGCTG
3121 CAAGATCAAA GACGTTATGT GAAAGACGTT TGGACAGGAA TG
```

CYP102A1 Holoenzyme F87A

```
   1 ATGACAATTA AAGAAATGCC TCAGCCAAAA ACGTTTGGAG AGCTTAAAAA TTTACCGTTA
  61 TTAAACACAG ATAAACCGGT TCAAGCTTTG ATGAAAATTG CGGATGAATT AGGAGAAATC
 121 TTTAAATTCG AGGCGCCTGG TCGTGTAACG CGCTACTTAT CAAGTCAGCG TCTAATTAAA
 181 GAAGCATGCG ATGAATCACG CTTTGATAAA AACTTAAGTC AAGCGCTTAA ATTTGTACGT
 241 GATTTTGCAG GAGACGGGTT GGCCACAAGC TGGACGCATG AAAAAAATTG GAAAAAAGCG
 301 CATAATATCT TACTTCCAAG CTTCAGTCAG CAGGCAATGA AAGGCTATCA TGCGATGATG
 361 GTCGATATCG CCGTGCAGCT TGTTCAAAAG TGGGAGCGTC TAAATGCAGA TGAGCATATT
 421 GAAGTACCGG AAGACATGAC ACGTTTAACG CTTGATACAA TTGGTCTTTG CGGCTTTAAC
 481 TATCGCTTTA ACAGCTTTTA CCGAGATCAG CCTCATCCAT TTATTACAAG TATGGTCCGT
 541 GCACTGGATG AAGCAATGAA CAAGCTGCAG CGAGCAAATC CAGACGACCC AGCTTATGAT
 601 GAAAACAAGC GCCAGTTTCA AGAAGATATC AAGGTGATGA ACGACCTAGT AGATAAAATT
 661 ATTGCAGATC GCAAAGCAAG CGGTGAACAA AGCGATGATT TATTAACGCA TATGCTAAAC
 721 GGAAAAGATC CAGAAACGGG TGAGCCGCTT GATGACGAGA ACATTCGCTA TCAAATTATT
 781 ACATTCTTAA TTGCGGGACA CGAAACAACA AGTGGTCTTT TATCATTTGC GCTGTATTTC
 841 TTAGTGAAAA ATCCACATGT ATTACAAAAA GCAGCAGAAG AAGCAGCACG AGTTCTAGTA
 901 GATCCTGTTC CAAGCTACAA ACAAGTCAAA CAGCTTAAAT ATGTCGGCAT GGTCTTAAAC
 961 GAAGCGCTGC GCTTATGGCC AACTGCTCCT GCGTTTTCCC TATATGCAAA AGAAGATACG
1021 GTGCTTGGAG GAGAATATCC TTTAGAAAAA GGCGACGAAC TAATGGTTCT GATTCCTCAG
1081 CTTCACCGTG ATAAAACAAT TTGGGGAGAC GATGTGGAAG AGTTCCGTCC AGAGCGTTTT
1141 GAAAATCCAA GTGCGATTCC GCAGCATGCG TTTAAACCGT TTGGAAACGG TCAGCGTGCG
1201 TGTATCGGTC AGCAGTTCGC TCTTCATGAA GCAACGCTGG TACTTGGTAT GATGCTAAAA
1261 CACTTTGACT TTGAAGATCA TACAAACTAC GAGCTGGATA TTAAAGAAAC TTTAACGTTA
1321 AAACCTGAAG GCTTTGTGGT AAAAGCAAAA TCGAAAAAAA TTCCGCTTGG CGGTATTCCT
1381 TCACCTAGCA CTGAACAGTC TGCTAAAAAA GTACGCAAAA AGGCAGAAAA CGCTCATAAT
1441 ACGCCGCTGC TTGTGCTATA CGGTTCAAAT ATGGGAACAG CTGAAGGAAC GGCGCGTGAT
1501 TTAGCAGATA TTGCAATGAG CAAAGGATTT GCACCGCAGG TCGCAACGCT TGATTCACAC
1561 GCCGGAAATC TTCCGCGCGA AGGAGCTGTA TTAATTGTAA CGGCGTCTTA TAACGGTCAT
1621 CCGCCTGATA ACGCAAAGCA ATTTGTCGAC TGGTTAGACC AAGCGTCTGC TGATGAAGTA
1681 AAAGGCGTTC GCTACTCCGT ATTTGGATGC GGCGATAAAA ACTGGGCTAC TACGTATCAA
1741 AAAGTGCCTG CTTTTATCGA TGAAACGCTT GCCGCTAAAG GGGCAGAAAA CATCGCTGAC
1801 CGCGGTGAAG CAGATGCAAG CGACGACTTT GAAGGCACAT ATGAAGAATG GCGTGAACAT
1861 ATGTGGAGTG ACGTAGCAGC CTACTTTAAC CTCGACATTG AAAACAGTGA AGATAATAAA
1921 TCTACTCTTT CACTTCAATT TGTCGACAGC GCCGCGGATA TGCCGCTTGC GAAAATGCAC
1981 GGTGCGTTTT CAACGAACGT CGTAGCAAGC AAAGAACTTC AACAGCCAGG CAGTGCACGA
2041 AGCACGCGAC ATCTTGAAAT TGAACTTCCA AAAGAAGCTT CTTATCAAGA AGGAGATCAT
2101 TTAGGTGTTA TTCCTCGCAA CTATGAAGGA ATAGTAAACC GTGTAACAGC AAGGTTCGGC
2161 CTAGATGCAT CACAGCAAAT CCGTCTGGAA GCAGAAGAAG AAAAATTAGC TCATTTGCCA
2221 CTCGCTAAAA CAGTATCCGT AGAAGAGCTT CTGCAATACG TGGAGCTTCA AGATCCTGTT
2281 ACGCGCACGC AGCTTCGCGC AATGGCTGCT AAAACGGTCT GCCCGCCGCA TAAAGTAGAG
2341 CTTGAAGCCT TGCTTGAAAA GCAAGCCTAC AAAGAACAAG TGCTGGCAAA ACGTTTAACA
2401 ATGCTTGAAC TGCTTGAAAA ATACCCGGCG TGTGAAATGA AATTCAGCGA ATTTATCGCC
2461 CTTCTGCCAA GCATACGCCC GCGCTATTAC TCGATTTCTT CATCACCTCG TGTCGATGAA
2521 AAACAAGCAA GCATCACGGT CAGCGTTGTC TCAGGAGAAG CGTGGAGCGG ATATGGAGAA
2581 TATAAAGGAA TTGCGTCGAA CTATCTTGCC GAGCTGCAAG AAGGAGATAC GATTACGTGC
2641 TTTATTTCCA CACCGCAGTC AGAATTTACG CTGCCAAAAG ACCCTGAAAC GCCGCTTATC
2701 ATGGTCGGAC CGGGAACAGG CGTCGCGCCG TTTAGAGGCT TTGTGCAGGC GCGCAAACAG
2761 CTAAAAGAAC AAGGACAGTC ACTTGGAGAA GCACATTTAT ACTTCGGCTG CCGTTCACCT
2821 CATGAAGACT ATCTGTATCA AGAAGAGCTT GAAAACGCCC AAAGCGAAGG CATCATTACG
2881 CTTCATACCG CTTTTTCTCG CATGCCAAAT CAGCCGAAAA CATACGTTCA GCACGTAATG
2941 GAACAAGACG GCAAGAAATT GATTGAACTT CTTGATCAAG GAGCGCACTT CTATATTTGC
3001 GGAGACGGAA GCCAAATGGC ACCTGCCGTT GAAGCAACGC TTATGAAAAG CTATGCTGAC
3061 GTTCACCAAG TGAGTGAAGC AGACGCTCGC TTATGGCTGC AGCAGCTAGA AGAAAAAGGC
3121 CGATACGCAA AAGACGTGTG GGCTGGGCAT CATCATCATC ATCAT
```

CYP102A2 Holoenzyme F88A

```
   1 ATGAAGGAAA CAAGCCCGAT TCCTCAGCCG AAGACGTTTG GGCCGCTCGG CAATTTGCCT
  61 TTAATTGATA AAGACAAACC GACGCTTTCG CTGATCAAAC TGGCGGAAGA ACAGGGCCCG
 121 ATTTTTCAAA TCCATACACC CGCGGGCACG ACCATTGTAG TGTCCGGCCA TGAATTGGTG
 181 AAAGAGGTTT GTGATGAAGA ACGGTTTGAT AAAAGCATTG AAGGCGCCTT GGAAAAGGTT
 241 CGCGCATTTT CCGGTGACGG ATTGGCCACT AGTTGGACGC ATGAGCCTAA CTGGAGAAAA
 301 GCGCACAACA TTCTGATGCC GACGTTCAGC CAGCGGGCCA TGAAGGACTA TCATGAGAAA
```

```
 361 ATGGTCGATA TCGCTGTTCA GCTCATTCAA AAATGGGCAA GGCTCAACCC GAATGAAGCA
 421 GTCGATGTCC CGGGAGATAT GACCCGGCTG ACGCTCGACA CCATTGGGCT ATGCGGGTTT
 481 AACTACCGCT TTAACAGTTA CTACAGAGAA ACGCCCCACC CGTTTATCAA CAGCATGGTG
 541 CGGGCGCTTG ATGAAGCGAT GCATCAAATG CAGCGGCTTG ATGTTCAAGA TAAGCTTATG
 601 GTCAGAACAA AGCGGCAATT CCGCTATGAT ATTCAAACGA TGTTTTCGTT AGTCGACAGC
 661 ATTATTGCAG AGCGCAGGGC GAATGGAGAC CAGGATGAAA AAGATTTGCT CGCCCGCATG
 721 CTGAATGTGG AAGATCCGGA AACTGGTGAA AAGCTCGACG ACGAAAATAT CCGCTTTCAG
 781 ATCATCACGT TTTTGATTGC CGGCCATGAA ACAACGAGCG GCCTGCTTTC CTTTGCGACT
 841 TACTTTTTAT TGAAGCATCC TGACAAACTG AAAAAGGCGT ATGAAGAGGT CGATCGGGTG
 901 CTGACGGATG CAGCGCCGAC CTATAAACAA GTGCTGGAGC TTACATACAT ACGGATGATT
 961 TTAAATGAAT CACTGCGCTT ATGGCCGACA GCTCCGGCTT TCAGCCTTTA TCCAAAAGAA
1021 GACACAGTCA TTGGCGGAAA ATTTCCGATC ACGACGAATG ACAGAATTTC TGTGCTGATT
1081 CCGCAGCTTC ATCGTGATCG AGACGCTTGG GGAAAGGACG CAGAAGAGTT CCGGCCGGAA
1141 CGGTTTGAGC ATCAGGACCA AGTGCCTCAT CATGCGTACA AACCATTCGG AAATGGACAA
1201 CGGGCCTGTA TCGGCATGCA GTTTGCCCTT CATGAAGCCA CACTTGTGTT AGGCATGATT
1261 CTAAAATATT TCACATTGAT TGATCATGAG AATTATGAGC TTGATATCAA ACAAACCTTA
1321 ACACTTAAGC CGGGCGATTT TCACATCAGT GTTCAAAGCC GTCATCAGGA AGCCATTCAT
1381 GCAGACGTCC AGGCAGCTGA AAAAGCCGCG CCTGATGAGC AAAAGGAGAA AACGGAAGCA
1441 AAGGGTGCAT CGGTCATCGG TCTTAACAAC CGCCCGCTTC TCGTGCTGTA CGGCTCAGAT
1501 ACCGGCACCG CAGAAGGCGT CGCCCGGGAG CTTGCTGATA CTGCCAGTCT TCACGGCGTA
1561 AGGACAAAGA CAGCACCTCT GAACGACCGG ATTGGAAAGC TGCCGAAAGA GGGAGCGGTT
1621 GTCATTGTGA CCTCGTCTTA TAATGGAAAG CCGCCAAGCA ATGCCGGACA ATTCGTGCAG
1681 TGGCTTCAAG AAATCAAACC GGGTGAGCTT GAGGGCGTCC ATTACGCGGT ATTTGGCTGC
1741 GGCGACCACA ACTGGGCGAG CACGTATCAA TACGTGCCGA GATTCATTGA TGAGCAGCTT
1801 GCGGAGAAAG GCGCGACTCG GTTTTCTGCG CGCGGGGAAG GGGATGTGAG CGGTGATTTT
1861 GAAGGGCAGC TTGACGAGTG GAAAAAAAGC ATGTGGGCGG ATGCCATCAA AGCATTCGGA
1921 CTTGAGCTTA ATGAAAACGC TGATAAGGAA CGAAGCACGC TGAGCCTTCA GTTTGTCAGA
1981 GGGCTGGGCG AGTCTCCGCT CGCTAGATCG TACGAAGCCT CTCACGCATC CATTGCCGAA
2041 AATCGTGAAC TCCAGTCCGC AGACAGCGAT CGAAGCACTC GCCATATCGA AATTGCATTG
2101 CCGCCGGATG TTGAATATCA AGAGGGCGAC CATCTTGGCG TATTGCCAAA AAACAGCCAA
2161 ACCAATGTCA GCCGGATTCT TCACAGATTC GGTCTGAAGG GAACCGACCA AGTGACATTG
2221 TCGGCAAGCG GCCGCAGTGC GGGGCATCTG CCATTGGGCC GTCCTGTCAG CCTGCATGAT
2281 CTTCTCAGCT ACAGCGTCGA GGTGCAGGAA GCAGCCACAA GAGCGCAAAT ACGTGAACTG
2341 GCGTCATTTA CAGTGTGTCC GCCGCATAGG CGCGAATTAG AAGAACTGTC AGCAGAGGGT
2401 GTTTATCAGG AGCAAATATT GAAAAAACGA ATTTCCATGC TGGATCTGCT TGAAAAGTAT
2461 GAAGCGTGTG ACATGCCGTT TGAACGATTT TTAGAGCTTT TACGGCCGTT AAAACCGAGA
2521 TACTATTCGA TTTCAAGCTC TCCAAGAGTG AATCCGCGGC AAGCATCGAT CACAGTCGGT
2581 GTCGTGCGCG GCCCGGCGTG GAGCGGCCGT GGCGAATACA GGGGTGTGGC ATCAAATGAT
2641 TTAGCTGAGC GTCAAGCCGG TGATGATGTC GTGATGTTTA TCCGCACACC GGAATCCCGG
2701 TTTCAGCTTC CGAAAGACCC TGAAACGCCA ATTATTATGG TCGGGCCAGG CACGGGAGTC
2761 GCGCCATTTC GCGGTTTCCT TCAAGCCCGC GATGTTTTAA AGCGGGAGGG CAAAACGCTC
2821 GGTGAGGCTC ATCTCTATTT TGGATGCAGG AACGATCGGG ATTTTATTTA CCGAGATGAG
2881 CTTGAGCGGT TTGAAAAAGA CGGAATCGTC ACTGTCCACA CAGCCTTTTC CCGAAAAGAG
2941 GGCATGCCGA AAACATATGT CCAGCATCTC ATGGCTGACC AAGCAGATAC ATTAATATCA
3001 ATCCTTGACC GCGGTGGCAG GCTTTATGTA TGCGGTGATG GCAGCAAAAT GGCCCCGGAT
3061 GTGGAGGCGG CACTTCAAAA AGCGTATCAG GCTGTCCATG GAACCGGGGA ACAAGAAGCG
3121 CAAAACTGGC TGAGACATCT GCAGGATACC GGTATGTACG CTAAGGATGT CTGGGCAGGG
3181 ATA
```

CYP102A3 Holoenzyme F88A

```
   1 ATGAAACAGG CAAGCGCAAT ACCTCAGCCC AAAACATACG GACCTTTAAA AAATCTTCCG
  61 CATCTGGAAA AAGAACAGCT TTCTCAATCC TTATGGCGGA TAGCTGATGA ATTGGGACCG
 121 ATTTTCCGTT TTGATTTTCC GGGAGTATCC AGTGTTTTTG TGTCCGGCCA CAATCTTGTG
 181 GCTGAAGTGT GTGATGAAAA ACGCTTTGAC AAGAACCTTG GCAAAGGCTT GCAAAAGGTG
 241 CGTGAGTTCG GGGGAGATGG CTTAGCCACT AGTTGGACGC ACGAACCGAA CTGGCAAAAA
 301 GCCCACCGCA TTTTGCTGCC GAGTTTTAGT CAAAAAGCGA TGAAAGGCTA TCATTCTATG
 361 ATGCTGGATA TCGCAACCCA GCTGATTCAA AAGTGGAGCC GGTTAAACCC CAATGAAGAA
 421 ATTGATGTAG CGGACGATAT GACACGTCTG ACGCTTGATA CGATTGGGTT ATGCGGGTTT
 481 AACTATCGAT TCAACAGCTT TTACCGTGAT TCACAGCATC CGTTTATCAC CAGTATGCTC
 541 CGTGCCTTAA AAGAGGCGAT GAATCAATCG AAAAGACTGG GCCTGCAAGA TAAAATGATG
 601 GTGAAAACGA AGCTGCAGTT CCAAAAGGAT ATAGAAGTCA TGAACTCCCT GGTTGATAGA
 661 ATGATAGCGG AGCGAAAGGC GAATCCGGAT GAAAACATTA AGGATCTCTT GTCTCTCATG
 721 CTTTATGCCA AAGATCCAGT AACGGGTGAA ACGCTGGATG ACGAAAACAT TCGATACCAA
```

```
 781 ATCATCACAT TTTTAATTGC TGGACATGAG ACAACAAGCG GGTTGCTATC CTTTGCGATT
 841 TATTGTCTGC TTACACATCC GGAAAAACTG AAAAAAGCTC AGGAGGAAGC GGATCGCGTG
 901 TTAACGGATG ACACGCCTGA ATATAAACAA ATCCAGCAGC TCAAATACAT TCGGATGGTT
 961 TTAAATGAAA CCCTCAGACT GTATCCAACA GCTCCGGCTT TTTCTCTATA TGCGAAGGAG
1021 GATACTGTTC TTGGCGGGGA ATATCCGATC AGCAAAGGGC AGCCAGTCAC TGTTTTAATT
1081 CCAAAACTGC ACCGGGATCA AAACGCTTGG GGACCGGATG CGGAAGATTT CCGTCCGGAA
1141 CGGTTTGAAG ATCCTTCAAG TATCCCTCAC CATGCGTATA AGCCGTTTGG AAACGGACAG
1201 CGCGCTTGTA TTGGCATGCA GTTTGCTCTT CAAGAAGCGA CAATGGTTCT CGGTCTTGTA
1261 TTAAAGCATT TTGAATTGAT AAACCATACT GGCTACGAAC TAAAAATCAA AGAAGCATTA
1321 ACGATCAAGC CGGATGATTT TAAAATTACT GTGAAACCGC GAAAAACAGC GGCAATCAAT
1381 GTACAGAGAA AAGAACAGGC AGACATCAAA GCAGAAACAA AGCCAAAAGA AACCAAACCT
1441 AAACACGGCA CACCTTTACT TGTTCTTTTT GGTTCAAATC TTGGGACAGC TGAGGGAATA
1501 GCCGGTGAAC TGGCTGCTCA AGGCCGCCAG ATGGGCTTTA CAGCTGAAAC GGCTCCGCTT
1561 GATGATTATA TCGGCAAGCT CCCTGAAGAA GGGGCAGTCG TCATTGTAAC GGCTTCTTAT
1621 AATGGGGCGC CGCCTGATAA TGCTGCCGGA TTTGTAGAGT GGCTGAAAGA GCTTGAGGAA
1681 GGCCAATTGA AAGGTGTTTC CTATGCGGTA TTCGGCTGCG GAAACCGGAG CTGGGCCAGC
1741 ACGTATCAGC GGATTCCCCG CCTGATTGAT GACATGATGA AAGCAAAGGG GGCATCGCGT
1801 TTAACAGCGA TTGGGGAAGG TGACGCCGCC GATGATTTTG AAAGCCACCG CGAGTCTTGG
1861 GAAAACCGCT TCTGGAAGGA AACGATGGAC GCATTTGATA TTAACGAAAT AGCCCAGAAA
1921 GAAGACAGGC CTTCATTATC GATTACTTTT CTCAGTGAAG CGACGGAAAC GCCGGTTGCT
1981 AAAGCATATG GCGCGTTTGA AGGGATTGTG TTAGAGAATC GAGAACTCCA GACAGCTGCC
2041 AGCACGCGTT CAACCCGCCA TATTGAATTG GAAATTCCGG CTGGTAAAAC ATATAAAGAA
2101 GGCGATCATA TCGGAATCCT GCCAAAGAAC AGCAGGGAGC TTGTTCAGCG GGTTCTCAGC
2161 CGATTCGGTT TGCAGTCCAA TCATGTGATA AAAGTAAGCG GAAGCGCTCA TATGGCTCAT
2221 CTGCCGATGG ATCGGCCAAT CAAAGTAGTG GATTTATTGT CGTCCTATGT AGAGCTGCAG
2281 GAACCGGCAT CAAGGCTTCA GCTTCGGGAG CTGGCCTCTT ATACAGTTTG TCCGCCGCAT
2341 CAAAAAGAGC TGGAACAGCT CGTTTCAGAT GATGGCATTT ACAAAGAGCA GGTACTTGCA
2401 AAACGTCTTA CCATGCTTGA TTTTTTAGAG GATTATCCTG CTTGCGAAAT GCCGTTTGAA
2461 CGGTTTTTAG CACTTTTGCC ATCACTAAAA CCGAGATACT ATTCCATTTC AAGCTCACCG
2521 AAAGTTCATG CAAATATCGT GAGCATGACG GTAGGAGTTG TGAAAGCCTC AGCATGGAGC
2581 GGCCGAGGTG AATACCGGGG TGTCGCCTCT AATTATTTAG CAGAATTGAA TACAGGTGAT
2641 GCAGCAGCTT GCTTCATTCG TACGCCGCAG TCCGGATTTC AGATGCCGAA TGATCCTGAA
2701 ACGCCTATGA TTATGGTCGG GCCGGGCACA GGAATTGCGC CATTCAGAGG CTTTATTCAG
2761 GCAAGATCGG TTTTGAAGAA GGAAGGAAGC ACCCTTGGTG AAGCACTTTT ATACTTCGGC
2821 TGCCGCCGCC CGGACCATGA CGACCTTTAC AGAGAAGAGC TGGATCAAGC GGAACAGGAC
2881 GGTTTGGTCA CAATCCGCCG ATGCTACTCG CGCGTCGAAA ACGAACCAAA AGGATATGTC
2941 CAGCACTTGC TCAAGCAAGA TACGCAGAAA TTGATGACAC TCATTGAAAA AGGGGCTCAT
3001 ATTTACGTAT GCGGTGATGG ATCGCAAATG GCTCCTGATG TAGAGAGAAC TTTGCGATTG
3061 GCATATGAAG CTGAAAAAGC AGCAAGTCAG GAAGAATCAG CTGTATGGCT GCAAAAGCTG
3121 CAAGATCAAA GACGTTATGT GAAAGACGTT TGGACAGGAA TG
```

Amino acid sequences

Methionine in sequence is cleaved and is not included in standard numbering of CYP102A1.

CYP102A1 Heme F87A

```
   1 TIKEMPQPKT FGELKNLPLL NTDKPVQALM KIADELGEIF KFEAPGRVTR YLSSQRLIKE
  61 ACDESRFDKN LSQALKFVRD FAGDGLATSW THEKNWKKAH NILLPSFSQQ AMKGYHAMMV
 121 DIAVQLVQKW ERLNADEHIE VPEDMTRLTL DTIGLCGFNY RFNSFYRDQP HPFITSMVRA
 181 LDEAMNKLQR ANPDDPAYDE NKRQFQEDIK VMNDLVDKII ADRKASGEQS DDLLTHMLNG
 241 KDPETGEPLD DENIRYQIIT FLIAGHETTS GLLSFALYFL VKNPHVLQKA AEEAARVLVD
 301 PVPSYKQVKQ LKYVGMVLNE ALRLWPTAPA FSLYAKEDTV LGGEYPLEKG DELMVLIPQL
 361 HRDKTIWGDD VEEFRPERFE NPSAIPQHAF KPFGNGQRAC IGQQFALHEA TLVLGMMLKH
 421 FDFEDHTNYE LDIKETLTLK PEGFVVKAKS KKIPLGGIPS PST
```

CYP102A2 Heme F88A

```
   1 KETSPIPQP KTFGPLGNLP LIDKDKPTLS LIKLAEEQGP IFQIHTPAGT TIVVSGHELVK
  61 EVCDEERFD KSIEGALEKV RAFSGDGLAT SWTHEPNWRK AHNILMPTFS QRAMKDYHEKM
 121 VDIAVQLIQ KWARLNPNEA VDVPGDMTRL TLDTIGLCGF NYRFNSYYRE TPHPFINSMVR
 181 ALDEAMHQM QRLDVQDKLM VRTKRQFRYD IQTMFSLVDS IIAERRANGD QDEKDLLARML
 241 NVEDPETGE KLDDENIRFQ IITFLIAGHE TTSGLLSFAT YFLLKHPDKL KKAYEEVDRVL
 301 TDAAPTYKQ VLELTYIRMI LNESLRLWPT APAFSLYPKE DTVIGGKFPI TTNDRISVLIP
 361 QLHRDRDAW GKDAEEFRPE RFEHQDQVPH HAYKPFGNGQ RACIGMQFAL HEATLVLGMIL
 421 KYFTLIDHE NYELDIKQTL TLKPGDFHIS VQSRHQEAIH ADVQAAE
```

CYP102A3 Heme F88A

```
   1 KQASAIPQPK TYGPLKNLPH LEKEQLSQSL WRIADELGPI FRFDFPGVSS VFVSGHNLVA
  61 EVCDEKRFDK NLGKGLQKVR EFGGDGLATS WTHEPNWQKA HRILLPSFSQ KAMKGYHSMM
 121 LDIATQLIQK WSRLNPNEEI DVADDMTRLT LDTIGLCGFN YRFNSFYRDS QHPFITSMLR
 181 ALKEAMNQSK RLGLQDKMMV KTKLQFQKDI EVMNSLVDRM IAERKANPDE NIKDLLSLML
 241 YAKDPVTGET LDDENIRYQI ITFLIAGHET TSGLLSFAIY CLLTHPEKLK KAQEEADRVL
 301 TDDTPEYKQI QQLKYIRMVL NETLRLYPTA PAFSLYAKED TVLGGEYPIS KGQPVTVLIP
 361 KLHRDQNAWG PDAEDFRPER FEDPSSIPHH AYKPFGNGQR ACIGMQFALQ EATMVLGLVL
 421 KHFELINHTG YELKIKEALT IKPDDFKITV KPRKTAAINV QRKEQA
```

CYP102A1 Holoenzyme F87A

```
   1 TIKEMPQPKT FGELKNLPLL NTDKPVQALM KIADELGEIF KFEAPGRVTR YLSSQRLIKE
  61 ACDESRFDKN LSQALKFVRD FAGDGLATSW THEKNWKKAH NILLPSFSQQ AMKGYHAMMV
 121 DIAVQLVQKW ERLNADEHIE VPEDMTRLTL DTIGLCGFNY RFNSFYRDQP HPFITSMVRA
 181 LDEAMNKLQR ANPDDPAYDE NKRQFQEDIK VMNDLVDKII ADRKASGEQS DDLLTHMLNG
 241 KDPETGEPLD DENIRYQIIT FLIAGHETTS GLLSFALYFL VKNPHVLQKA AEEAARVLVD
 301 PVPSYKQVKQ LKYVGMVLNE ALRLWPTAPA FSLYAKEDTV LGGEYPLEKG DELMVLIPQL
 361 HRDKTIWGDD VEEFRPERFE NPSAIPQHAF KPFGNGQRAC IGQQFALHEA TLVLGMMLKH
 421 FDFEDHTNYE LDIKETLTLK PEGFVVKAKS KKIPLGGIPS PSTEQSAKKV RKKAENAHNT
 481 PLLVLYGSNM GTAEGTARDL ADIAMSKGFA PQVATLDSHA GNLPREGAVL IVTASYNGHP
 541 PDNAKQFVDW LDQASADEVK GVRYSVFGCG DKNWATTYQK VPAFIDETLA AKGAENIADR
 601 GEADASDDFE GTYEEWREHM WSDVAAYFNL DIENSEDNKS TLSLQFVDSA ADMPLAKMHG
 661 AFSTNVVASK ELQQPGSARS TRHLEIELPK EASYQEGDHL GVIPRNYEGI VNRVTARFGL
 721 DASQQIRLEA EEEKLAHLPL AKTVSVEELL QYVELQDPVT RTQLRAMAAK TVCPPHKVEL
 781 EALLEKQAYK EQVLAKRLTM LELLEKYPAC EMKFSEFIAL LPSIRPRYYS ISSSPRVDEK
 841 QASITVSVVS GEAWSGYGEY KGIASNYLAE LQEGDTITCF ISTPQSEFTL PKDPETPLIM
 901 VGPGTGVAPF RGFVQARKQL KEQGQSLGEA HLYFGCRSPH EDYLYQEELE NAQSEGIITL
 961 HTAFSRMPNQ PKTYVQHVME QDGKKLIELL DQGAHFYICG DGSQMAPAVE ATLMKSYADV
1021 HQVSEADARL WLQQLEEKGR YAKDVWA
```

CYP102A2 Holoenzyme F88A

```
   1 KETSPIPQPK TFGPLGNLPL IDKDKPTLSL IKLAEEQGPI FQIHTPAGTT IVVSGHELVK
  61 EVCDEERFDK SIEGALEKVR AFSGDGLATS WTHEPNWRKA HNILMPTFSQ RAMKDYHEKM
 121 VDIAVQLIQK WARLNPNEAV DVPGDMTRLT LDTIGLCGFN YRFNSYYRET PHPFINSMVR
 181 ALDEAMHQMQ RLDVQDKLMV RTKRQFRYDI QTMFSLVDSI IAERRANGDQ DEKDLLARML
 241 NVEDPETGEK LDDENIRFQI ITFLIAGHET TSGLLSFATY FLLKHPDKLK KAYEEVDRVL
 301 TDAAPTYKQV LELTYIRMIL NESLRLWPTA PAFSLYPKED TVIGGKFPIT TNDRISVLIP
 361 QLHRDRDAWG KDAEEFRPER FEHQDQVPHH AYKPFGNGQR ACIGMQFALH EATLVLGMIL
 421 KYFTLIDHEN YELDIKQTLT LKPGDFHISV QSRHQEAIHA DVQAAEKAAP DEQKEKTEAK
 481 GASVIGLNNR PLLVLYGSDT GTAEGVAREL ADTASLHGVR TKTAPLNDRI GKLPKEGAVV
 541 IVTSSYNGKP PSNAGQFVQW LQEIKPGELE GVHYAVFGCG DHNWASTYQY VPRFIDEQLA
 601 EKGATRFSAR GEGDVSGDFE GQLDEWKKSM WADAIKAFGL ELNENADKER STLSLQFVRG
 661 LGESPLARSY EASHASIAEN RELQSADSDR STRHIEIALP PDVEYQEGDH LGVLPKNSQT
 721 NVSRILHRFG LKGTDQVTLS ASGRSAGHLP LGRPVSLHDL LSYSVEVQEA ATRAQIRELA
 781 SFTVCPPHRR ELEELSAEGV YQEQILKKRI SMLDLLEKYE ACDMPFERFL ELLRPLKPRY
 841 YSISSSPRVN PRQASITVGV VRGPAWSGRG EYRGVASNDL AERQAGDDVV MFIRTPESRF
 901 QLPKDPETPI IMVGPGTGVA PFRGFLQARD VLKREGKTLG EAHLYFGCRN DRDFIYRDEL
 961 ERFEKDGIVT VHTAFSRKEG MPKTYVQHLM ADQADTLISI LDRGGRLYVC GDGSKMAPDV
1021 EAALQKAYQA VHGTGEQEAQ NWLRHLQDTG MYAKDVWAG
```

CYP102A3 Holoenzyme F88A

```
   1 KQASAIPQPK TYGPLKNLPH LEKEQLSQSL WRIADELGPI FRFDFPGVSS VFVSGHNLVA
  61 EVCDEKRFDK NLGKGLQKVR EFGGDGLATS WTHEPNWQKA HRILLPSFSQ KAMKGYHSMM
 121 LDIATQLIQK WSRLNPNEEI DVADDMTRLT LDTIGLCGFN YRFNSFYRDS QHPFITSMLR
 181 ALKEAMNQSK RLGLQDKMMV KTKLQFQKDI EVMNSLVDRM IAERKANPDE NIKDLLSLML
 241 YAKDPVTGET LDDENIRYQI ITFLIAGHET TSGLLSFAIY CLLTHPEKLK KAQEEADRVL
 301 TDDTPEYKQI QQLKYIRMVL NETLRLYPTA PAFSLYAKED TVLGGEYPIS KGQPVTVLIP
 361 KLHRDQNAWG PDAEDFRPER FEDPSSIPHH AYKPFGNGQR ACIGMQFALQ EATMVLGLVL
 421 KHFELINHTG YELKIKEALT IKPDDFKITV KPRKTAAINV QRKEQADIKA ETKPKETKPK
 481 HGTPLLVLFG SNLGTAEGIA GELAAQGRQM GFTAETAPLD DYIGKLPEEG AVVIVTASYN
 541 GAPPDNAAGF VEWLKELEEG QLKGVSYAVF GCGNRSWAST YQRIPRLIDD MMKAKGASRL
 601 TAIGEGDAAD DFESHRESWE NRFWKETMDA FDINEIAQKE DRPSLSITFL SEATETPVAK
 661 AYGAFEGIVL ENRELQTAAS TRSTRHIELE IPAGKTYKEG DHIGILPKNS RELVQRVLSR
 721 FGLQSNHVIK VSGSAHMAHL PMDRPIKVVD LLSSYVELQE PASRLQLREL ASYTVCPPHQ
 781 KELEQLVSDD GIYKEQVLAK RLTMLDFLED YPACEMPFER FLALLPSLKP RYYSISSSPK
 841 VHANIVSMTV GVVKASAWSG RGEYRGVASN YLAELNTGDA AACFIRTPQS GFQMPNDPET
 901 PMIMVGPGTG IAPFRGFIQA RSVLKKEGST LGEALLYFGC RRPDHDDLYR EELDQAEQDG
 961 LVTIRRCYSR VENEPKGYVQ HLLKQDTQKL MTLIEKGAHI YVCGDGSQMA PDVERTLRLA
1021 YEAEKAASQE ESAVWLQKLQ DQRRYVKDVW TGM
```

References

1. Barnes HJ, Arlotto MP, Waterman MR (1991) Expression and enzymatic activity of recombinant cytochrome P450 17 alpha-hydroxylase in Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. 88: 5597-5601.
2. Amann E, Brosius J, Ptashne M (1983) Vectors bearing a hybrid trp-lac promoter useful for regulated expression of cloned genes in Escherichia coli. Gene 25: 167-178.

Appendix B

Data set for Chapter 5

Table B.1. 955 chimeric and parent P450 heme domain sequences with their folding state, 12-pNCA activity state and number of sequence changes.

Sequence[1]	Folding State[2]	12-pNCA activity[3]	m[4]
11111111	1	1	0
11112212	1	1	62
11113233	1	--	86
11113311	1	1	36
11131313	1	--	73
11133231	1	--	75
11212112	1	1	51
11212333	1	1	91
11213133	1	--	75
11213231	1	1	63
11232111	1	--	44
11232232	1	--	91
11232333	1	--	70
11311233	1	1	79
11312233	1	1	93
11313233	1	1	85
11313333	1	1	69
11331312	1	0	78
11331333	1	0	64
11332212	1	0	92
11332233	1	0	79
11332333	1	--	63
11333212	1	0	94
12112333	1	1	97
12113221	1	1	77
12211232	1	1	91
12211333	1	1	96
12212112	1	--	68
12212211	1	1	60
12212212	1	1	80
12212332	1	1	93
12213212	1	1	90
12232111	1	1	61
12232112	1	--	89
12232232	1	0	74
12232233	1	0	90
12232332	1	0	90
12233212	1	0	92
12313331	1	1	80
31233233	1	--	37
31233333	1	--	21
31311231	1	1	81
31311233	1	0	67
31311332	1	1	80
31312113	1	1	88
31312133	1	1	70
31312212	1	0	100
31312222	1	--	76
31312231	1	--	95
31312233	1	1	66
31312323	1	1	71
31312332	1	1	79
31312333	1	1	50
31313111	1	--	59
31313131	1	1	77
31313132	1	1	84
31313133	1	--	55
31313223	1	--	72
31313232	1	1	80
31313233	1	--	51
31313333	1	1	35
31331331	1	--	62
31331333	1	0	30
31332131	1	0	81
31332133	1	--	49
31332232	1	0	74
31332233	1	0	45
31332312	1	--	76
31332322	1	--	79
31332323	1	--	50
31332333	1	--	29
31333233	1	0	30
31333322	1	--	64
31333332	1	0	43
31333333	1	0	14
32111333	1	1	64
32112212	1	--	85
32112321	1	--	105

12322333	1	1	86	32113131	1	0	85
12331123	1	--	103	32113232	1	1	93
12331333	1	1	68	32113233	1	1	64
12332223	1	--	89	32131133	1	--	63
12332233	1	1	83	32132232	1	0	79
12332333	1	--	67	32132233	1	--	58
12333331	1	--	84	32132331	1	0	74
12333333	1	1	52	32133111	1	--	88
13113311	1	0	50	32133232	1	--	72
13213131	1	--	57	32133233	1	--	43
13222212	1	--	76	32133331	1	--	59
13233212	1	0	104	32211323	1	0	83
13332333	1	0	49	32212133	1	0	81
13333122	1	--	104	32212231	1	0	101
13333131	1	--	78	32212232	1	0	73
13333211	1	0	80	32212233	1	1	77
21111321	1	0	78	32212321	1	0	96
21111323	1	0	107	32212323	1	1	82
21111333	1	0	90	32212332	1	0	89
21112122	1	1	68	32212333	1	1	61
21112123	1	0	97	32213123	1	--	87
21112132	1	1	89	32213132	1	--	95
21112212	1	1	72	32213231	1	0	94
21112222	1	1	48	32213333	1	1	46
21112232	1	1	69	32232131	1	--	92
21112311	1	0	68	32232322	1	--	65
21112312	1	1	88	32232331	1	--	72
21112322	1	1	64	32232333	1	--	40
21112332	1	1	85	32233222	1	--	64
21113111	1	0	50	32233332	1	--	54
21113112	1	1	78	32311131	1	--	78
21113122	1	1	83	32311323	1	1	76
21113133	1	1	94	32312212	1	--	83
21113211	1	0	70	32312231	1	1	102
21113212	1	1	87	32312233	1	--	70
21113221	1	--	91	32312311	1	--	94
21113223	1	0	92	32312322	1	1	75
21113312	1	1	98	32312323	1	--	75
21113322	1	1	79	32312331	1	--	86
21131121	1	0	79	32312332	1	--	83
21132112	1	1	89	32312333	1	1	54
21132113	1	0	101	32313133	1	--	59
21132212	1	1	69	32313231	1	0	87
21132222	1	0	45	32313232	1	0	84
21132311	1	0	89	32313233	1	1	55
21132313	1	--	86	32313313	1	--	57
21132321	1	0	89	32313332	1	--	68
21132323	1	0	89	32313333	1	1	39
21133112	1	1	99	32332133	1	0	53
21133113	1	0	91	32332223	1	--	70

21133131	1	--	89	32332231	1	0	81
21133211	1	0	91	32332232	1	0	77
21133222	1	1	60	32332322	1	0	72
21133223	1	--	89	32332323	1	0	54
21133232	1	1	81	32332331	1	0	65
21133233	1	0	69	32332332	1	0	62
21133312	1	1	100	32332333	1	1	33
21133313	1	0	71	32333223	1	--	55
21133321	1	--	104	32333232	1	1	63
21133322	1	--	76	32333233	1	--	34
21133331	1	0	85	32333312	1	--	65
21133332	1	0	82	32333323	1	0	39
21211223	1	0	82	32333333	1	--	18
21211321	1	--	87	33113111	1	--	64
21212111	1	0	57	33113211	1	0	84
21212112	1	1	83	33113212	1	--	93
21212122	1	--	59	33113233	1	--	46
21212123	1	0	88	33131333	1	--	25
21212133	1	1	107	33133131	1	--	61
21212212	1	1	63	33133333	1	--	9
21212213	1	0	92	33212213	1	1	77
21212231	1	0	88	33212311	1	--	91
21212321	1	0	83	33212333	1	0	43
21212332	1	1	76	33213211	1	--	93
21212333	1	0	87	33213232	1	--	73
21213121	1	0	83	33213333	1	0	28
21213212	1	1	78	33232233	1	--	38
21213223	1	0	83	33232312	1	--	69
21213231	1	0	97	33232333	1	--	22
21213321	1	0	98	33233131	1	--	59
21213332	1	1	91	33233233	1	--	23
21222112	1	0	61	33233333	1	--	7
21231233	1	0	83	33311231	1	--	85
21232112	1	0	80	33312133	1	0	56
21232122	1	--	56	33312322	1	--	86
21232132	1	0	77	33312333	1	0	36
21232212	1	0	60	33313233	1	0	37
21232222	1	0	36	33313323	1	--	42
21232231	1	0	85	33313333	1	0	21
21232232	1	0	57	33331232	1	--	61
21232233	1	0	82	33331233	1	--	32
21232321	1	0	80	33331333	1	--	16
21232322	1	0	52	33332131	1	--	67
21232323	1	0	81	33332133	1	--	35
21232332	1	0	73	33332232	1	--	60
21233111	1	0	80	33332233	1	--	31
21233132	1	0	92	33332323	1	--	36
21233212	1	1	75	33332333	1	0	15
21233221	1	--	79	33333231	1	0	48
21233233	1	0	67	33333232	1	--	45

21233312	1	--	91	33333233	1	--	16
21233321	1	--	95	33333323	1	--	21
21311122	1	1	80	33333333	1	0	0
21311223	1	0	89	11313223	0*	--	101
21311331	1	0	81	11333323	0*	--	69
21312111	1	0	57	12212223	0*	--	85
21312112	1	1	85	12331221	0*	--	91
21312122	1	1	66	12332123	0*	--	108
21312123	1	0	95	13132333	0*	--	58
21312211	1	0	77	13311311	0*	--	43
21312222	1	1	46	13331123	0*	--	91
21312223	1	0	75	13331333	0*	--	50
21312321	1	0	90	13332332	0*	--	78
21312322	1	1	62	13333123	0*	--	75
21312323	1	0	91	13333223	0*	--	71
21313111	1	0	59	13333233	0*	--	50
21313112	1	1	87	13333323	0*	--	55
21313122	1	1	81	21113222	0*	--	63
21313221	1	0	89	21133123	0*	--	94
21313231	1	1	97	21233322	0*	--	67
21313311	1	0	79	21312121	0*	--	81
21313312	1	1	101	21331112	0*	--	92
21313313	1	0	83	22232213	0*	--	72
21313322	1	1	77	23231121	0*	--	99
21331223	1	0	86	23232221	0*	--	65
21331332	1	0	89	23313322	0*	--	78
21331333	1	0	60	23333122	0*	--	79
21332111	1	0	78	31131323	0*	--	60
21332112	1	0	87	31212132	0*	--	103
21332113	1	0	97	31233111	0*	--	80
21332122	1	0	63	32111211	0*	--	71
21332131	1	0	96	32212122	0*	--	72
21332212	1	0	67	32232222	0*	--	49
21332221	1	0	71	32333322	0*	--	68
21332223	1	0	72	33112112	0*	--	90
21332231	1	0	92	33133323	0*	--	30
21332233	1	0	75	33232322	0*	--	72
21332312	1	0	83	33331223	0*	--	53
21332322	1	0	59	33331321	0*	--	69
21332323	1	--	80	33332221	0*	--	84
21332331	1	0	91	33333321	0*	--	53
21332332	1	0	80	11112123	0	--	70
21332333	1	--	59	11113223	0	--	92
21333111	1	0	80	11132223	0	--	108
21333122	1	0	78	11132232	0	--	100
21333131	1	0	96	11132323	0	--	93
21333132	1	--	93	11231232	0	--	96
21333211	1	0	100	11232323	0	--	91
21333212	1	1	82	11331123	0	--	86
21333221	1	0	86	11332221	0	--	88

21333223	1	0	81	11333122	0	--	98
21333233	1	0	60	12133223	0	--	98
21333312	1	0	91	12211222	0	--	70
21333321	1	0	97	12231231	0	--	85
21333333	1	1	44	12233112	0	--	91
22111223	1	0	74	12233323	0	--	80
22111332	1	1	82	12311333	0	--	89
22112111	1	--	65	12331211	0	--	67
22112131	1	0	83	13132223	0	--	95
22112211	1	0	83	13132322	0	--	96
22112223	1	0	60	13133323	0	--	64
22112321	1	--	75	13212122	0	--	89
22112323	1	0	76	13212321	0	--	81
22112331	1	0	96	13231332	0	--	86
22113111	1	0	67	13232123	0	--	97
22113211	1	--	87	13232311	0	--	78
22113223	1	0	75	13232323	0	--	77
22113232	1	1	67	13233133	0	--	61
22113233	1	0	94	13233233	0	--	57
22113313	1	--	96	13233322	0	--	91
22113323	1	0	91	13332223	0	--	86
22113332	1	1	83	13333222	0	--	93
22131221	1	0	70	21111112	0	--	62
22132112	1	1	72	21111212	0	--	82
22132113	1	0	101	21111312	0	--	82
22132212	1	1	52	21111322	0	--	78
22132231	1	0	77	21131111	0	--	55
22132233	1	0	78	21131212	0	--	83
22132312	1	--	68	21131321	0	--	99
22132323	1	0	73	21132121	0	--	93
22132331	1	0	93	21133212	0	--	84
22133112	1	1	87	21211113	0	--	75
22133211	1	0	95	21211122	0	--	73
22133212	1	1	67	21211211	0	--	63
22133232	1	1	64	21211222	0	--	53
22133312	1	0	83	21233112	0	--	95
22133322	1	1	59	21311111	0	--	43
22133323	1	1	78	21311311	0	--	63
22212111	1	0	74	21311333	0	--	81
22212123	1	0	71	21312212	0	--	70
22212131	1	0	91	21331111	0	--	64
22212212	1	1	46	21331131	0	--	82
22212232	1	1	43	21331312	0	--	97
22212312	1	--	62	21331313	0	--	78
22212321	1	0	66	21333121	0	--	104
22212322	1	1	38	21333133	0	--	64
22213111	1	--	76	21333331	0	--	76
22213112	1	0	81	21333332	0	--	73
22213132	1	1	78	22111231	0	--	89
22213212	1	1	61	22113132	0	--	87

22213222	1	1	37	22121331	0	--	88
22213223	1	0	66	22131111	0	--	72
22213233	1	0	87	22131112	0	--	86
22213312	1	0	77	22131132	0	--	83
22213321	1	--	81	22131133	0	--	93
22222121	1	0	48	22131222	0	--	42
22222222	1	1	0	22131321	0	--	86
22231221	1	0	61	22211132	0	--	77
22231223	1	--	62	22211222	0	--	36
22232112	1	--	63	22211331	0	--	98
22232121	1	0	67	22211332	0	--	73
22232122	1	0	39	22231113	0	--	106
22232123	1	0	68	22231122	0	--	53
22232212	1	1	43	22231211	0	--	85
22232222	1	1	19	22231212	0	--	57
22232223	1	0	48	22231232	0	--	54
22232232	1	1	40	22231311	0	--	101
22232233	1	0	69	22231312	0	--	73
22232311	1	0	87	22231333	0	--	71
22232312	1	0	59	22233331	0	--	87
22232322	1	0	35	22233333	0	--	55
22232323	1	--	64	22311121	0	--	84
22232331	1	0	84	22311212	0	--	67
22232333	1	0	70	22311231	0	--	92
22233112	1	1	78	22311332	0	--	80
22233211	1	0	86	22313333	0	--	69
22233212	1	1	58	22331121	0	--	88
22233221	1	0	62	22331133	0	--	84
22233222	1	--	34	22331211	0	--	92
22233223	1	--	63	22331212	0	--	64
22233312	1	--	74	22331222	0	--	40
22233323	1	0	76	22331321	0	--	84
22233332	1	--	71	22333111	0	--	97
22311123	1	0	92	23111112	0	--	76
22312111	1	0	74	23111212	0	--	87
22312123	1	0	78	23112123	0	--	98
22312132	1	1	70	23112222	0	--	49
22312133	1	0	99	23131332	0	--	84
22312211	1	0	81	23132322	0	--	62
22312221	1	--	57	23133312	0	--	86
22312222	1	1	29	23211121	0	--	81
22312223	1	0	58	23211131	0	--	75
22312231	1	0	78	23211222	0	--	54
22312232	1	1	50	23211311	0	--	77
22312311	1	0	94	23211332	0	--	91
22312312	1	--	69	23212312	0	--	80
22312322	1	1	45	23213231	0	--	104
22312332	1	0	66	23231212	0	--	75
22312333	1	0	84	23231323	0	--	74
22313122	1	1	64	23232223	0	--	66

22313212	1	0	68	23233322	0	--	68
22313221	1	0	72	23311112	0	--	85
22313222	1	--	44	23311221	0	--	89
22313232	1	1	65	23311222	0	--	61
22313233	1	0	85	23311313	0	--	85
22313323	1	1	89	23313121	0	--	97
22313331	1	0	101	23313231	0	--	99
22313332	1	0	81	23331112	0	--	102
22323313	1	0	85	23331212	0	--	82
22331123	1	0	89	23331232	0	--	79
22331221	1	--	68	23332322	0	--	60
22331223	1	0	69	31112121	0	--	72
22331323	1	0	85	31113321	0	--	94
22331332	1	0	77	31131312	0	--	86
22332112	1	1	70	31132221	0	--	103
22332113	1	0	99	31132223	0	--	75
22332121	1	0	74	31132311	0	--	88
22332123	1	0	75	31132312	0	--	85
22332132	1	1	67	31133112	0	--	90
22332211	1	0	78	31133123	0	--	64
22332221	1	0	54	31133212	0	--	86
22332222	1	1	26	31211122	0	--	95
22332223	1	0	55	31211132	0	--	89
22332232	1	--	47	31211211	0	--	63
22332233	1	0	76	31211312	0	--	91
22332312	1	0	66	31212113	0	--	89
22332321	1	--	70	31212211	0	--	77
22332322	1	0	42	31213122	0	--	104
22332332	1	0	63	31213223	0	--	79
22333112	1	0	85	31231211	0	--	84
22333122	1	0	61	31231311	0	--	84
22333131	1	0	100	31231323	0	--	58
22333132	1	0	82	31232322	0	--	82
22333133	1	0	68	31233122	0	--	91
22333211	1	0	93	31233133	0	--	41
22333212	1	1	65	31233222	0	--	81
22333221	1	0	69	31311112	0	--	71
22333222	1	--	41	31311122	0	--	95
22333223	1	0	70	31311212	0	--	91
22333231	1	0	90	31311312	0	--	91
22333311	1	0	98	31312221	0	--	101
22333313	1	--	66	31313123	0	--	76
22333321	1	0	85	31313222	0	--	91
22333323	1	--	69	31313321	0	--	88
22333332	1	0	77	31331221	0	--	99
23112213	1	0	102	31331222	0	--	87
23112221	1	0	77	31331223	0	--	67
23112223	1	0	78	31331332	0	--	59
23112233	1	0	91	31332112	0	--	96
23112323	1	0	94	31332132	0	--	78

23112333	1	0	75	31332221	0	--	98
23113111	1	0	64	31333112	0	--	81
23113112	1	0	92	31333222	0	--	80
23113121	1	0	88	31333223	0	--	51
23113131	1	0	82	31333232	0	--	59
23113212	1	--	88	31333311	0	--	64
23113311	1	0	84	32111112	0	--	79
23113312	1	0	104	32111121	0	--	75
23113323	1	0	81	32111123	0	--	105
23113332	1	0	89	32111311	0	--	71
23122212	1	--	51	32112232	0	--	82
23131323	1	--	76	32112311	0	--	85
23132111	1	0	83	32113112	0	--	95
23132121	1	--	94	32131212	0	--	96
23132212	1	--	70	32131311	0	--	92
23132221	1	0	74	32132211	0	--	106
23132231	1	0	95	32132212	0	--	82
23132232	1	0	67	32132221	0	--	86
23132233	1	0	70	32133113	0	--	65
23132311	1	0	103	32133122	0	--	93
23132323	1	--	75	32133212	0	--	90
23133112	1	--	105	32133223	0	--	64
23133113	1	0	77	32133311	0	--	77
23133121	1	0	109	32133312	0	--	74
23133233	1	0	55	32133321	0	--	80
23133311	1	0	89	32133323	0	--	48
23133321	1	--	92	32211111	0	--	60
23133331	1	0	71	32211112	0	--	88
23133333	1	--	39	32211133	0	--	82
23211132	1	--	95	32211212	0	--	90
23212112	1	--	84	32212111	0	--	74
23212211	1	0	91	32212311	0	--	94
23212212	1	0	64	32213311	0	--	96
23212221	1	0	68	32213312	0	--	93
23212222	1	0	40	32231122	0	--	83
23212231	1	0	89	32231222	0	--	63
23212332	1	--	77	32231332	0	--	70
23212333	1	0	73	32232111	0	--	95
23213112	1	0	99	32232133	0	--	60
23213121	1	0	97	32232212	0	--	73
23213123	1	--	99	32232213	0	--	74
23213211	1	0	93	32232221	0	--	77
23213212	1	1	79	32233112	0	--	92
23213223	1	--	84	32233122	0	--	84
23213232	1	0	76	32233123	0	--	66
23213311	1	0	93	32233232	0	--	70
23213322	1	--	71	32233233	0	--	41
23213333	1	0	58	32311132	0	--	104
23231233	1	0	69	32311212	0	--	97
23232113	1	--	90	32311221	0	--	101

23232131	1	0	104	32311322	0	--	89
23232211	1	0	89	32313122	0	--	94
23232212	1	--	61	32313212	0	--	98
23232233	1	0	68	32313222	0	--	74
23232311	1	--	102	32331111	0	--	81
23232323	1	--	73	32331112	0	--	101
23233212	1	0	76	32331122	0	--	90
23233221	1	--	80	32331132	0	--	83
23233231	1	0	85	32331212	0	--	94
23233232	1	--	73	32331221	0	--	98
23233312	1	--	84	32331222	0	--	70
23233333	1	--	37	32331311	0	--	84
23311233	1	--	83	32331313	0	--	52
23311323	1	0	88	32332111	0	--	95
23312111	1	0	71	32332112	0	--	100
23312112	1	0	91	32332123	0	--	74
23312121	1	0	95	32332211	0	--	99
23312122	1	--	67	32332212	0	--	80
23312123	1	0	96	32332222	0	--	56
23312131	1	0	89	32333122	0	--	88
23312223	1	--	76	32333212	0	--	81
23312311	1	0	91	32333311	0	--	68
23312312	1	0	87	33111122	0	--	100
23312323	1	0	87	33111212	0	--	96
23313111	1	0	73	33111222	0	--	93
23313133	1	--	71	33111312	0	--	93
23313212	1	0	86	33113121	0	--	88
23313222	1	0	62	33113223	0	--	67
23313232	1	0	83	33131122	0	--	95
23313233	1	0	67	33131123	0	--	66
23313323	1	0	72	33131332	0	--	54
23313333	1	0	51	33132222	0	--	76
23331233	1	0	62	33132223	0	--	61
23331323	1	--	67	33132322	0	--	74
23332112	1	0	88	33133121	0	--	82
23332221	1	0	72	33133223	0	--	46
23332222	1	--	44	33133233	0	--	25
23332223	1	0	73	33133321	0	--	62
23332231	1	0	93	33133332	0	--	38
23332311	1	0	95	33211112	0	--	85
23332323	1	0	66	33211211	0	--	77
23332331	1	0	77	33211312	0	--	91
23333111	1	0	94	33211321	0	--	97
23333123	1	--	71	33212222	0	--	70
23333131	1	0	82	33212312	0	--	90
23333211	1	--	96	33212313	0	--	61
23333212	1	0	83	33213112	0	--	95
23333213	1	0	64	33231212	0	--	86
23333222	1	--	59	33231221	0	--	92
23333223	1	0	67	33231312	0	--	70

23333232	1	0	75	33231333	0	--	23
23333233	1	0	46	33232112	0	--	89
23333312	1	0	77	33232122	0	--	87
23333323	1	--	51	33232123	0	--	63
31111233	1	--	76	33232222	0	--	67
31112231	1	0	86	33232223	0	--	59
31112333	1	1	59	33232323	0	--	43
31113131	1	1	68	33233112	0	--	74
31113132	1	1	93	33233221	0	--	76
31113222	1	--	93	33233222	0	--	73
31113323	1	--	65	33233223	0	--	44
31113331	1	1	76	33233323	0	--	28
31113332	1	--	73	33311122	0	--	107
31131233	1	--	55	33311223	0	--	74
31132231	1	0	86	33311311	0	--	77
31132232	1	0	83	33311312	0	--	84
31132333	1	--	38	33311322	0	--	87
31133233	1	--	39	33311332	0	--	66
31133331	1	--	55	33312233	0	--	52
31211131	1	0	61	33312312	0	--	83
31211232	1	0	103	33312323	0	--	57
31212112	1	--	85	33313122	0	--	91
31212212	1	0	93	33313223	0	--	58
31212232	1	0	90	33313311	0	--	71
31212321	1	0	101	33331122	0	--	86
31212323	1	0	78	33331133	0	--	36
31212331	1	1	89	33331311	0	--	66
31212332	1	0	86	33331331	0	--	48
31212333	1	1	57	33332112	0	--	82
31213232	1	--	87	33332121	0	--	88
31213233	1	1	58	33332123	0	--	56
31213323	1	--	63	33332211	0	--	81
31213331	1	1	74	33332223	0	--	52
31213332	1	1	71	33333133	0	--	20
31232231	1	--	84	33333222	0	--	66
31232312	1	--	83	33333223	0	--	37
31232332	1	0	65	33333311	0	--	50
31232333	1	0	36	33333332	0	--	29
31233221	1	--	90				

[1]Sequence listed as parent sequence (A1 = 1, A2 =2, A3 = 3) at each block.

[2]Folding state is 1 for P450s and 0 for no P450; * signifies significant P420 peak.

[3]Chimeras with at least 25% of total product formed by A1 in 30 minutes are active = 1, not active = 0, -- = did not meet minimum P450 concentration requirement of 500 nM.

[4]m is number of sequence changes relative to the closest parent sequence.

Appendix C

Substrates tested with the 4-AAP assay

This table lists the substrates tested for activity with members of the chimeric library described in Chapter 5, 6 and 7. An **x** in activiy column designates the appearance of a red color over background as well an increase in absorbance near 500 nm.

Table C.1. Substrates used for activity tests with chimeric P450s with the 4-AAP assay.

Substrate	Activity	Notes
1,2-methylenedioxybenzene		
1,2-methylenedioxybenzene		
11-phenoxyundecanoic acid	x	fatty acid analog
1-phenyldecane	x	
1-phenyloctane	x	
2,3-(methylenedioxy)benzaldehyde		
2,6-dichlorobenzonitrile		
2-phenoxyethanol	x	bactericide
3-phenoxytoluene	x	
4-aminobiphenyl	x	carcinogenic
4-chromanone		
5α-cholestane		cholesterol derivative
9-amino-1,2,3,4-tetrahydroacridine		drug, cholinesterase inhibitor, a.k.a. Tacrine, Cognex
allyloxybenzene	x	a.k.a. allyl phenyl ether, background with 4-AAP
anthracene		
benzene		carcinogenic
benzofuran-2-carboxylic acid		
benzyl ether		flavoring agent
benzylacetone		
chlorzoxazone	x	drug, muscle relaxant

cinnamaldehyde		flavoring agent
cinnamic acid		
coumarone		a.k.a. Benzofuran, high-background with 4-AAP
diclofenac		drug-like, anti inflammatory
diphenyl ether	X	a.k.a. 2-methylbenzofuran
diphenyl ether		
DL-mandelic acid		
DL-α-phenylglycine		
dopamine		neurotransmitter
ethyl-4-phenylbutyrate	X	
ethylphenoxyacetate	X	
isosafrol		
melatonin		neurotransmitter
methoxychlor		insecticide
methylphenoxyacetate	X	
naphthalene		holoenzyme
naproxen		drug, non-steroidal anti-inflammatory, a.k.a. Aleve
p-acetophenetidide		a.k.a phenacetin
phenethyl alcohol		holoenzyme
ethoxybenzene	X	a.k.a. phenetole
propranolol	X	drug, beta blocker
pseudocumene		
serotonin		neurotransmitter
tolbutamide	X	drug, stimulates insulin secretion
warfarin		drug, anticoagulent
zoxazolamine	X	drug, muscle relaxant, a.k.a. 2-amino-5-chlorobenzoxazole

Appendix D

Spectra of drug compounds and their products

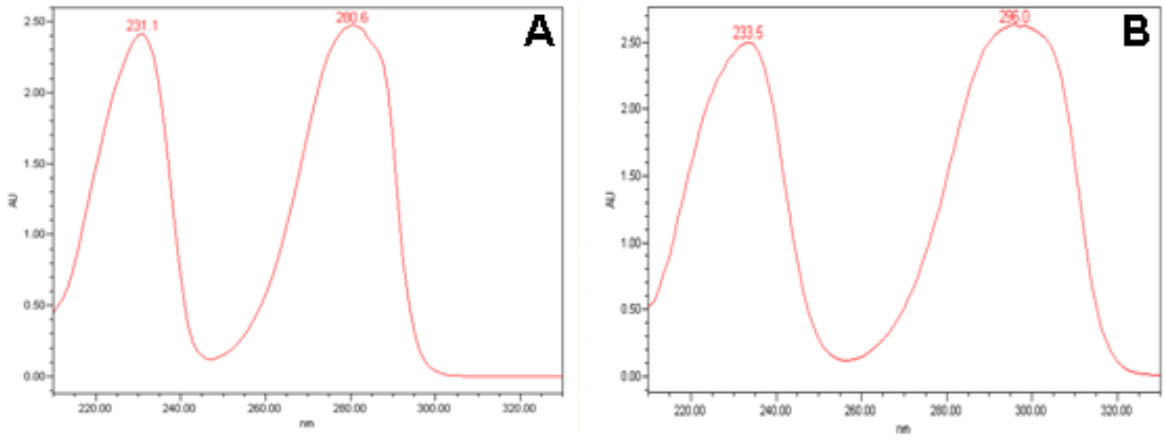

Figure D.1. UV-Vis spectra of chlorzoxazone (A) and 6-hydroxychlorzoxazone (B).

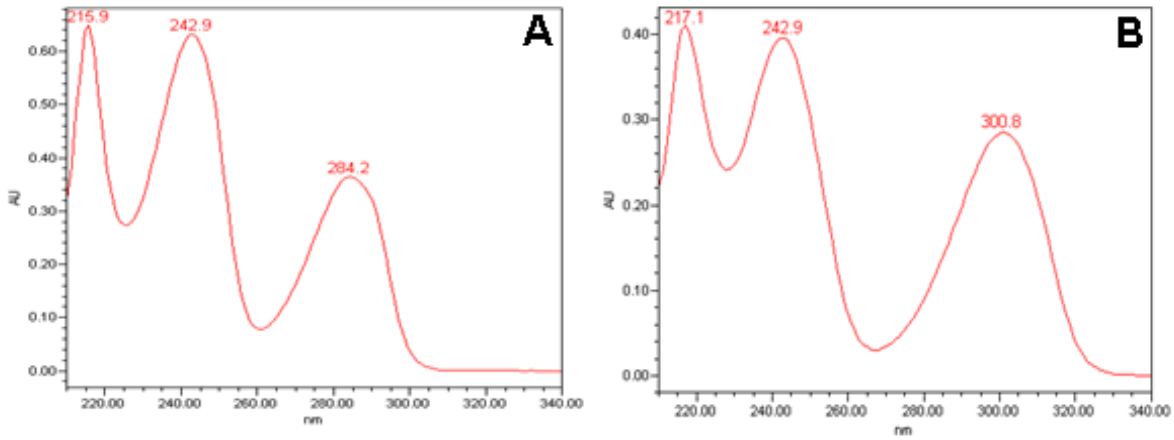

Figure D.2. UV-Vis spectra of zoxazolamine (A) and 6-hydroxyzoxazolamine (B).

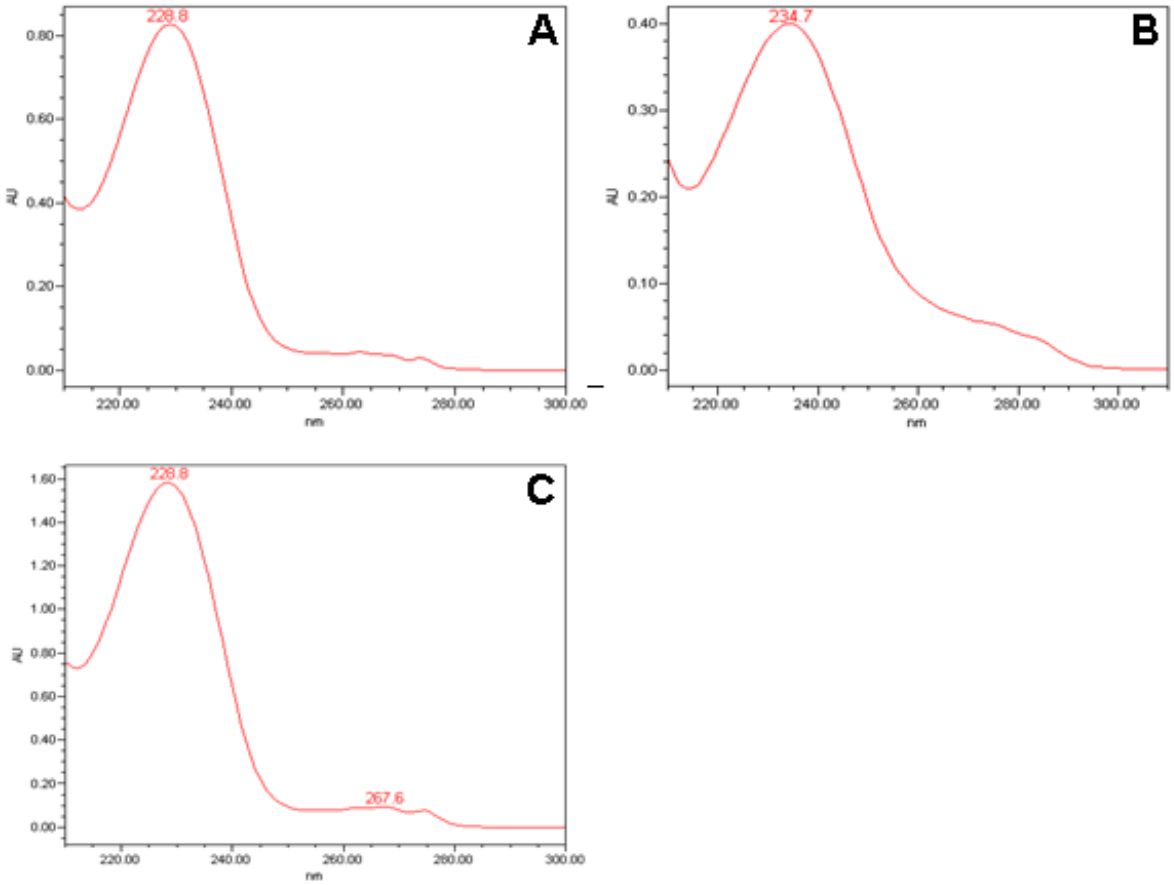

Figure D.3. UV-Vis spectra of tolbutamide (A), carboxytolbutamide (B) and 4-hydroxytolbutamide (C).

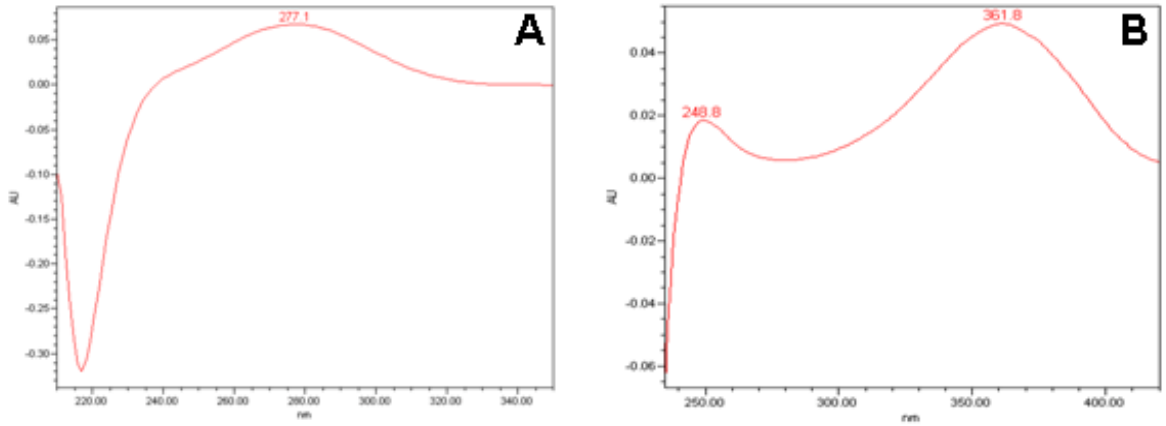

Figure D.4. UV-Vis spectra of 4-aminobiphenyl (A) and 4,4'-azobis(biphenyl) (B).

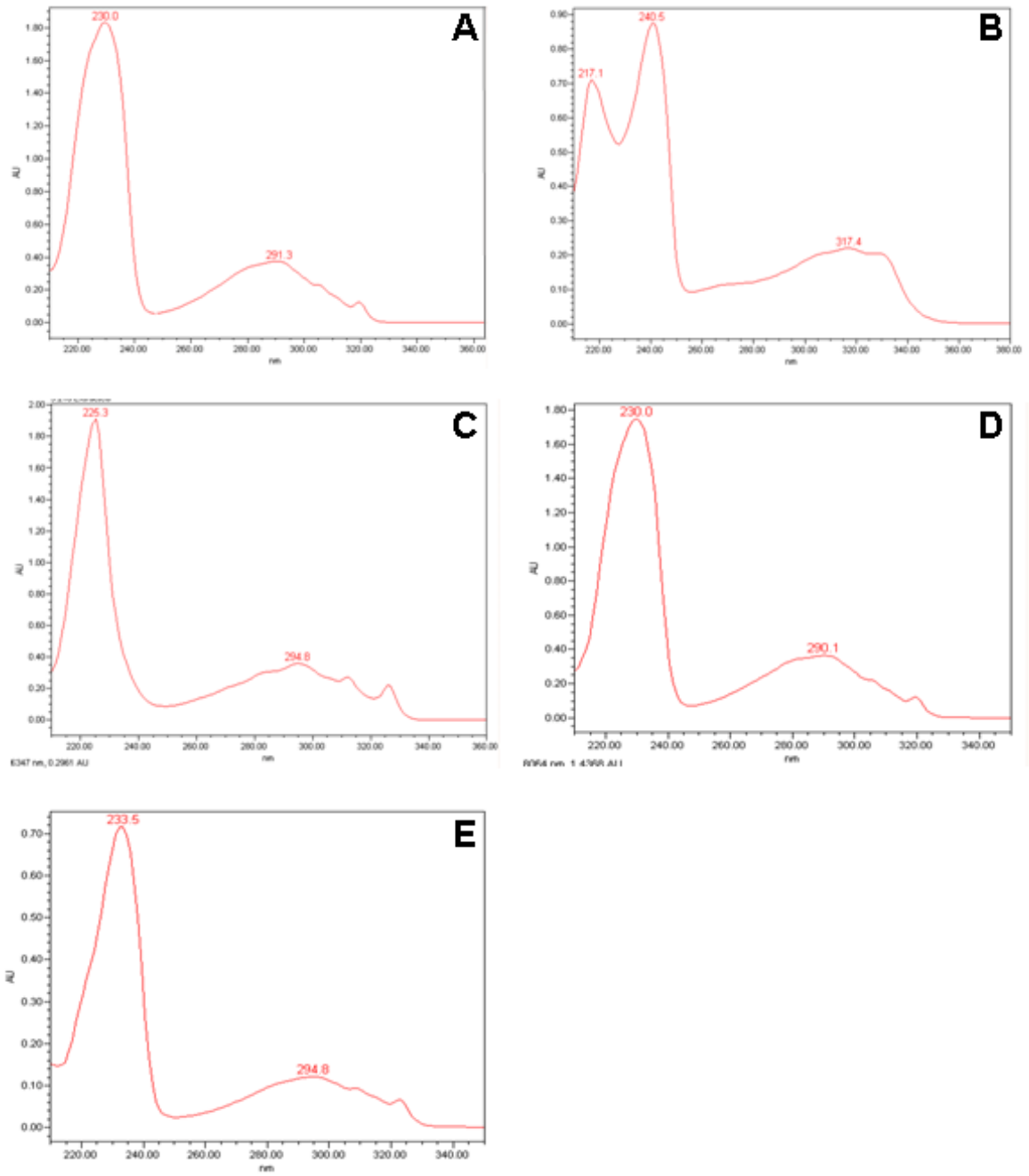

Figure D.5. UV-Vis spectra of propranolol (A), 4'-hydroxypropranolol (B), 5'-hydroxypropranolol (C), desisopropylpropranolol (D), 1-nphthol (E).

Appendix E

Creating chimeras of more distantly related P450s

To begin to explore the methodology of recombining more distantly related P450s, two chimeras between CYP102A1 and CYP101 (15% amino acid identity) were constructed. Chimeras made from P450s with low sequence identity (< 30%) have been successfully constructed before but have only contained single crossovers near the N- or C-termini of the enzyme [1,2]. In order to create large libraries between distantly related P450s, one must address the main issue of decreased accuracy in the sequence alignment due to lower sequence identity and a larger number of insertions and deletions. A sequence alignment alone, such as ClustalW [3], would likely yield little success since the aligned sequences may not structurally correlate. To remedy this, a structure-based alignment such as 3D-Coffee [4] or combinatorial extension [5] can be employed to correctly align elements of the core structure.

Two chimeras were designed by first aligning their structures using combinatorial extension and generating a structure-based sequence alignment (Figure E.1). This was done using PDB's 1JPZ [7] and 1AKD [8] for CYP102A1 and CYP101, respectively. Using this alignment, two structural elements were swapped to create two chimeras. The first chimera, named 172-224, has the F and G helices from CYP101 swapped into CYP102A1 (Figure E.2A). This resulted in 31 mutations and a deletion of 16 amino acids relative to CYP102A1. Chimera 234-266 has a direct replacement of the I-helix in CYP102A1 with the I helix from CYP101 resulting in 24 mutations (Figure E.2B).

A.

```
CYP102A1    ALMKIADELGEIFKFEAPGRVTRYLSSQRLIKEACDESRFDKN--LSQALKFVRDFAGDG     85
CYP101      AWAVLQESNVPDLVWTRCNGGHWIATRGQLIREAYEDYRHFSSECPFIPREAGEA--YDF     99

CYP102A1    LATSWTHEKNWKKAHNILLPSFSQQAMKGYHAMMVDIAVQLVQKWERLNADEHIEVPE-D    144
CYP101      IPTSMDPPE-QRQFRALANQVVGMPVVDKLENRIQELACSLIESLRPQ---GQCNFTEDY    155

CYP102A1    MTRLTLDTIGLCGFNYRFNSFYRDQPHPFITSMVRALDEAMNKLQRANPDDPAYDENKRQ    204
CYP101      AEPFPIRIFMLL-AGLPEED-----IPHLKYLTDQMTRPD---------------GSMT    193

CYP102A1    FQEDIKVMNDLVDKIIADRKASGEQSDDLLTHMLNGKDPETGEPLDDENIRYQIITFLIA    264
CYP101      FAEAKEALYDYLIPIIEQRR--QKPGTDAISIVANGQVN--GRPITSDEAKRMCGLLLVG    249

CYP102A1    GHETTSGLLSFALYFLVKNPHVLQKAAEEAARVLVDPVPSYKQVKQLKYVGMVLNEALRL    324
CYP101      GLDTVVNFLSFSMEFLAKSPEHRQELIQRP----------------ERIPAACEELLRR    292

CYP102A1    WPTAPAFSLYAKEDTVLGGEYPLEKGDELMVLIPQLHRDKTIWGDDVEEFRPERFENPSA    384
CYP101      FSLVADGRILTSDYEFH--GVQLKKGDQILLPQMLSGLDERENAC-PMHVDFSRQ----K    345

CYP102A1    IPQHAFKPFGNGQRACIGQQFALHEATLVLGMMLKHFDFEDHTN-YELDIKETLTLKPEG    443
CYP101      V---SHTTFGHGSHLCLGQHLARREIIVTLKEWLTRIPDFSIAPGAQIQHKSGIVSGVQA    402

CYP102A1    FVVKAK---SKKI                                                   453
CYP101      LPLVWDPATTKAV                                                   415
```

B.

```
CYP102A1    ALMKIADELGEIFKFEAPGRVTRYLSSQRLIKEACDESRFDKNLSQALKFVRDFAGDGLF     87
CYP101      AWAVLQESNVPDLVWTRCNGGHWIATRGQLIREAYEDYR---HFSSECPFIPREAG----     94

CYP102A1    TSWTHEKNWKKAHNILLPSFSQQAMKGYHAMMVDIA-VQLVQKWERLNADEHIEVPEDMT    146
CYP101      ----------EAYDFIPTSMDPPEQRQFRALANQVVGMPVVDKLENRIQELACSLIESLR    144

CYP102A1    RLTLDTIGLCGFNYRFNSFYRDQPHPFITSMVRALDEAMNKLQRANPDDPAYDENKRQFQ    206
CYP101      -----PQGQCNFTEDYAEPFPIR----IFMLLAGLPEEDIPHLKYLTDQMTRPDGSMTFA    195

CYP102A1    EDIKVMNDLVDKIIADRKASGEQSDDLLTHMLNGKDPETGEPLDDENIRYQIITFLIAGH    266
CYP101      EAKEALYDYLIPIIEQRRQ--KPGTDAISIVANG--QVNGRPITSDEAKRMCGLLLVGGL    251

CYP102A1    ETTSGLLSFALYFLVKNPHVLQKAAEEAARVLVDPVPSYKQVKQLKYVGMVLNEALRLWP    326
CYP101      DTVVNFLSFSMEFLAKSPEHRQELIQRPERIPAA----------------CEELLRRFS    294

CYP102A1    TAPAFSLYAKEDTVLGGEYPLEKGDELMVLIPQLHRDKTIWGDDVEEFRPERFENPSAIP    386
CYP101      LVADGRILTSDYEFHG--VQLKKGD--QILLPQMLSG----LDERENACPMHVDFS--RQ    344

CYP102A1    QHAFKPFGNGQRACIGQQFALHEATLVLGMMLKHF-DFEDHTNYELDIKETLT--LKPEG    443
CYP101      KVSHTTFGHGSHLCLGQHLARREIIVTLKEWLTRIPDFSIAPGAQIQHKSGIVSGVQALP    404

CYP102A1    FVVKAKSKKI-                                                     453
CYP101      LVWDPATTKAV                                                     415
```

Figure E.1. Sequence alignments of CYP102A1 and CYP101 based on structure (A) using combinatorial extension [5] and sequence (B) using ClustalW [3].

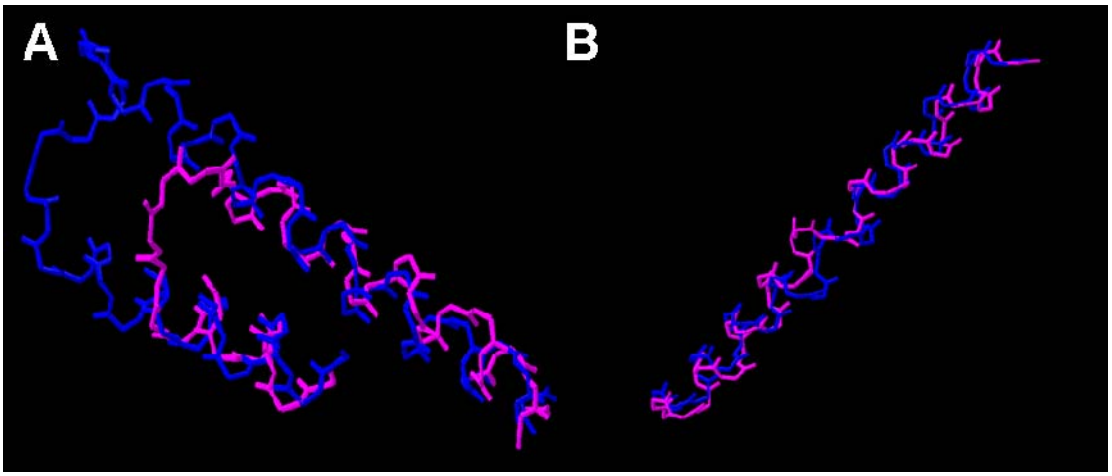

Figure E.2 Structural alignments of sequence elements swapped in CYP102A1-CYP101 chimeras. (A) Structural alignment of the F and G helices from CYP102A1 (blue) and CYP101 (purple) in 172-224. (B) Structural alignment of the I helices from CYP102A1 and CYP101 in 234-266.

Chimeras were assembled using SOEing from the heme domain of CYP102A1 with the F87A mutation and the wild-type sequence of CYP101 [6]. They were then expressed and analyzed as discussed in Chapter 3. Carbon monoxide difference spectroscopy showed that chimera 172-224 was able to correctly bind its heme cofactor and thus correctly fold into a cytochrome P450. Chimera 234-266, however, did not. Since it folded into a P450, 172-224 was tested for activity on 12-pNCA but showed no production of p-nitrophenol in the presence of 1 and 20 mM H_2O_2. Residue 87 was mutated back to phenylalanine, the residue present in the wild-type CYP102A1 seqence, and tested for activity on 12-pNCA. It was still inactive. Other tests could include testing for activity on camphor, the 'natural' substrate of CYP101 or the use of a reductase domain and NADPH in place of H_2O_2 to supply reducing equivalents.

Figure E.3 shows the alignments of the F and G helices that were swapped between CYP102A1 and CYP101 in chimera 172-224. Comparison of the structural and

sequence-based alignments shows a very different correspondence between amino acids. Figure E.2A shows that the insertion was correctly aligned in the structural alignment and that CYP101 has shortened F and G helices relative to CYP102A1. The sequence alignment using ClustalW does not identify this alignment and would likely result in the swapping of structurally incompatible elements. Further experiments would need to be performed to confirm this. Chimera 234-266 was likely unsuccessful since the I helix makes up the core of the P450 structure and the sequence changes are highly disruptive. Obviously more work is needed here but these methodologies could be extended to choose swappable elements for the creation of a large library in conjunction with computational algorithms to minimize structural disruption (i.e. SCHEMA [9]).

Structural alignment

```
CYP102A1    FITSMVRALDEAMNKLQRANPDDPAYDENKRQFQEDIKVMNDLVDKIIADRK
CYP101      LKYLTDQMTRPD----------------GSMTFAEAKEALYDYLIPIIEQRR
```

Sequence alignment

```
CYP102A1    FITSMVRALDEAMNKLQRANPDDPAYDENKRQFQEDIKVMNDLVDKIIADRK
CYP101      -IFMLLAGLPEEDIPHLKYLTDQMTRPDGSMTFAEAKEALYDYLIPIIEQRR
```

Figure E.3. The structural alignment compared to the sequence alignment for the F and G helices swapped in chimera 172-224.

References

1. Sieber V, Martinez CA, Arnold FH (2001) Libraries of hybrid proteins from distantly related sequences. Nat. Biotechnol. 19: 456-460.
2. Sukumaran S, Atkins WM, Shanker R (2002) Engineering cytochrome P-450s: chimeric enzymes. Appl. Biochem. Biotechnol. 102-103: 291-302.
3. Thompson JD, Higgins DG, Gibson TJ (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: 4673-4680.
4. O'Sullivan O, Suhre K, Abergel C, Higgins DG, Notredame C (2004) 3DCoffee: combining protein sequences and structures within multiple sequence alignments. J. Mol. Biol. 340: 385-395.
5. Shindyalov IN, Bourne PE (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng. 11: 739-747.
6. Horton RM, Hunt HD, Ho SN, Pullen JK, Pease LR (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. Gene 77: 61-68.
7. Haines DC, Tomchick DR, Machius M, Peterson JA (2001) Pivotal role of water in the mechanism of P450BM-3. Biochemistry 40: 13456-13465.
8. Schlichting I, Jung C, Schulze H (1997) Crystal structure of cytochrome P-450cam complexed with the (1S)-camphor enantiomer. FEBS Lett. 415: 253-257.
9. Voigt CA, Martinez C, Wang ZG, Mayo SL, Arnold FH (2002) Protein building blocks preserved by recombination. Nat. Struct. Biol. 9: 553-558.

Appendix F

Miscellaneous laboratory methods

This appendix includes laboratory methods used in this research that are unpublished since they were not part of the primary methodologies. They were frequently requested by other lab members and are thus included here.

Preparation of electrocompetent cells

Materials

1. 1 L of LB media with appropriate antibiotics (Fernbock)
2. 10 mL of LB media (in a 100 or 250 ml flask)
3. 1.75 L ice-cold 10% glycerol (autoclaved and stored at 4°C)
4. Two large centrifuge bottles (autoclaved and stored at 4°C)
5. Fifty 1.7 ml microcentrifuge tubes (autoclaved and stored at 4°C)

Methods

Keep all materials as close to 4 ºC as possible throughout the preparation. Resuspend cells gently with pipette. Do not vortex.

1. Inoculate a 10 ml culture of LB with your cell strain. Grow overnight at 30ºC.
2. Subculture the 10 ml into 1 liter of fresh LB. Grow at 30ºC until the OD reaches 0.8 to 0.9 (generally 3-4 hours).
3. Put on ice or in refrigerator for 10 minutes.
4. Spin down cells at 5,000 xg for 10 minutes. You will lose some of the cells here (~10%) if you only use two bottles.
5. Resuspend in 1 growth volume of ice-cold 10% glycerol.
6. Spin down at 5,000 xg for 10 minutes.
7. Resuspend in ½ growth volume of ice-cold 10% glycerol.
8. Spin down at 5,000 xg for 10 minutes.
9. Resuspend in 1/500 growth volume of ice-cold 10% glycerol.
10. Flash freeze in 100 µL aliquots in cold eppendorf tubes. Makes about 30 tubes.
11. Store at -80ºC.

Stripping blots for probe hybridization

Blots refer to nitrocellulose paper with bound DNA used for DNA probe hybridization (southern blots). They were used here for high-throughput sequencing.

Materials

1. Old blots
2. 400 ml 0.1% SDS per blot to be stripped
3. Hybridization oven set at 80°C
4. TE buffer, pH 8.0 (10 mM Tris (pH 8.0), 1 mM EDTA (pH 8.0))

Methods

1. Boil 0.1% SDS.
2. Add boiling 0.1% SDS to blot in hybridization bottle (100 ml per bottle works well).
3. Put hybridization bottle in hybridization oven (set at 80°C) for 10 minutes.
4. Repeat a total of four times.
5. Wash each blot once in 20 ml TE buffer, pH 8.0.
6. Store blots in small volume of TE buffer pH 8.0 (enough to cover blot).

Storage of cells in 96-well microtiter plates

This method is useful for storing libraries or large numbers of individual clones for long periods of time.

Materials

1. Sterilized 96-well plates
2. LB media with appropriate antibiotics
3. 50% glycerol (sterilized)
4. 96-well plates with the clones that are to be stored

Methods

1. Label sterilized 96-well plates. It is suggested that the plates be labeled on the edge and the top or bottom for easy identification and in case a label is lost or destroyed.
2. In a sterile hood, fill each well of the 96-well microtiter plate with 120 µl of LB media with the appropriate antibiotics.
3. Using a 96-well replicator, transfer small amounts of culture from an existing 96-well plate with saturated cultures of the clones to be stored.
4. Wrap plate with parafilm and grow for 14 to 20 hours.
5. Remove from shaker and add 60 µl of 50% glycerol to each well of the 96-well microtiter plate in a sterile hood.
6. Shake at 100 rpm for 10 minutes at room temperature.
7. Store at -80°C.