Chemistry & Biology, Vol. 11, 309–318, March, 2004, ©2004 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.chembiol.2004.02.018

Functional Evolution and Structural Conservation in Chimeric Cytochromes P450: Calibrating a Structure-Guided Approach

Christopher R. Otey,[1] Jonathan J. Silberg,[2] Christopher A. Voigt,[4] Jeffrey B. Endelman,[3] Geethani Bandara,[2] and Frances H. Arnold[1,2,3,*]
[1]Biochemistry and Molecular Biophysics Option
[2]Division of Chemistry and Chemical Engineering
[3]Bioengineering Option
California Institute of Technology
Mail Code 210-41
Pasadena, California 91125
[4]Department of Biopharmaceutical Sciences
University of California, San Francisco
San Francisco, California 94143

Summary

Recombination generates chimeric proteins whose ability to fold depends on minimizing structural perturbations that result when portions of the sequence are inherited from different parents. These chimeric sequences can display functional properties characteristic of the parents or acquire entirely new functions. Seventeen chimeras were generated from two CYP102 members of the functionally diverse cytochrome P450 family. Chimeras predicted to have limited structural disruption, as defined by the SCHEMA algorithm, displayed CO binding spectra characteristic of folded P450s. Even this small population exhibited significant functional diversity: chimeras displayed altered substrate specificities, a wide range in thermostabilities, up to a 40-fold increase in peroxidase activity, and ability to hydroxylate a substrate toward 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.

Introduction

The cytochrome P450 superfamily of enzymes exhibits an impressive range of chemical activities and biological roles. Nature has exploited 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, 15], these approaches generate few crossovers and large numbers of inactive sequences, due to insertions, deletions, and 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 [16, 17]. 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 [17]. 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 CYP102A1 (commonly referred to as cytochrome P450 BM-3) and CYP102A2, which are 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 will help to guide much larger efforts to explore the functional variation that is possible within the P450 scaffold.

Results

Chimera Design

For this work, we constructed chimeras of CYP102A1 [18] and CYP102A2, homologs from *Bacillus megaterium* and *Bacillus subtilis*, respectively. Their 64% nucleotide identity places them below the limits for effective recombination using annealing-based (DNA-shuffling) meth-

*Correspondence: frances@cheme.caltech.edu

ods. (Our previous attempts to shuffle these genes generated primarily the parental sequences, with few chimeras.) These soluble fusion proteins, consisting of a catalytic heme domain and an FAD- and FMN-containing NADPH reductase [19], 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 [20, 21]; 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 (referred to herein as CYP102A1-h and CYP102A2-h). 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 domains can also be fused to one of the parental reductase domains to regenerate a chimeric holoenzyme (see below).

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 CYP102A1-h and CYP102A2-h 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 1. Fourteen chimeras were individually designed and constructed to encompass a broad range of *E* (2 to 42) and *m* (11 to 70) (Table 1). Crossovers were placed in regions of low and high sequence identity. For three of the chimeras, we generated both "mirror" sequences, i.e., chimeras that derive sequences from opposite parents at every position. The other 11 sequences consisted of CYP102A1-h with an internal fragment derived from CYP102A2-h.

E values for the different chimeras were computed using the high-resolution structure for CYP102A1 with palmitoglycine bound in the active site [22]. Because previous studies have shown that substrate binding causes a large conformational change in CYP102A1 [23–25], we also calculated *E* using the substrate-free CYP102A1 structure [26]. As shown in Table 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*.

Folding of Chimeric Heme Domains

The chimeras were constructed using SOEing [27], cloned into the IPTG-inducible pCWori vector [28], and sequenced to confirm the absence of point mutations. All proteins were overexpressed in a catalase-free strain of *E. coli* [29], which allows the peroxygenase activity of the heme domains to be monitored directly in cell extracts [30]. We used carbon monoxide difference spectroscopy to assess the level of structural disruption

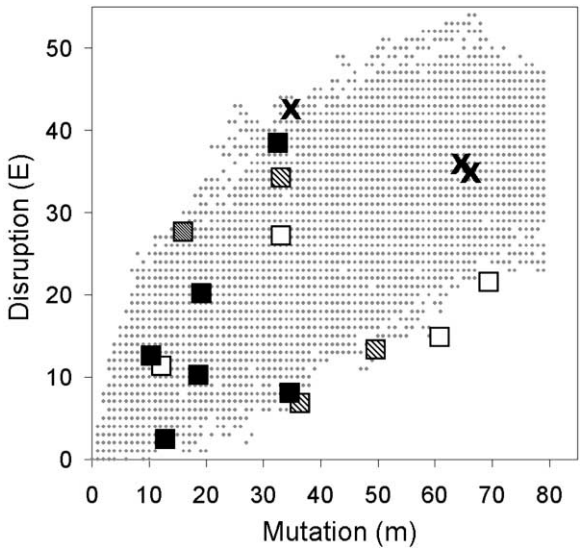

Figure 1. Effective Mutation and Disruption for Cytochrome P450 Heme Domain Chimeras

Disruption values for all double-crossover chimeras with a minimum insert size of 10 amino acids between CYP102A1-h and CYP102A2-h were determined using the structure for the CYP102A1 heme domain. Disruption (*E*) 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-phenoxy-ethanol, and allyloxy-benzene: squares represent chimeras that retain the ability to fold, and ×'s indicate those that did not. Symbols denote 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 allyloxy-benzene (▧).

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 [31]. 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 [32, 33]. Fourteen chimeras displayed detectable Soret bands: thirteen appeared at 450 nm and one at 420 nm (Figure 2). 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 1).

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 CYP102A1 and CYP102A2 to yield *p*-nitrophenolate [34]. 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 [35]). Activities on 2-phenoxy-ethanol and allyloxy-benzene were also determined, using the 4-aminoantipyrine (4-AAP) assay, which is sensitive to phenols and catechols [36]. This assay yields a detectable product if hydroxylation occurs at the *ortho* or *meta* positions of the aromatic ring or when hydroxyl-

Table 1. Properties of Designed CYP102A1-CYP102A2 Chimeric P450s

Protein[a]	*E* (Substrate Bound)[b]	*E* (Substrate-Free)[b]	*m*[c]	Folded (λ max)[d]	T_m (°C)	Peroxidase Activity[e]
CYP102A1-h	–	–	–	yes (448)	55	2.6 ± 0.1
CYP102A2-h	–	–	–	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] CYP102A1-h and CYP102A2-h 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 CYP102A2-h inserted into CYP102A1-h according to the numbering of CYP102A1. "M" indicates mirror chimeras where CYP102A1-h is inserted into CYP102A2-h.
[b] SCHEMA-calculated disruption (see Experimental Procedures) based on substrate bound (1JPZ) and substrate-free structures (2HPD). 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.).

ation 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-phenoxy-ethanol; neither showed measurable peroxygenase activity toward allyloxy-benzene.

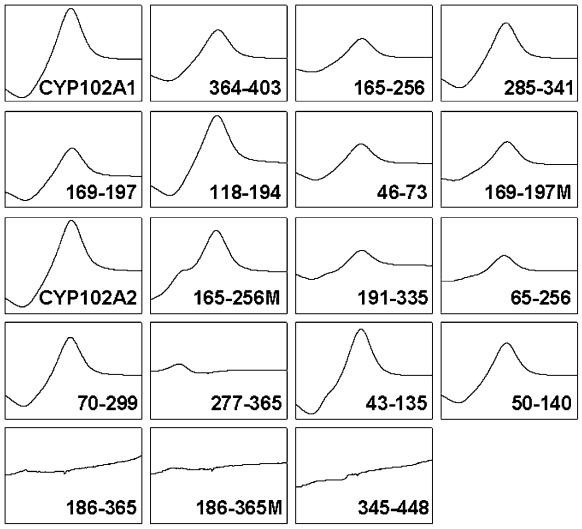

Figure 2. Reduced CO Difference Spectra for CYP102A1-h, CYP102A2-h, and Their Heme Domain Chimeras

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 3× with a range of −0.17 to 0.2, and the last row is magnified 40× 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.

Figure 3 compares the activities of the folded chimeric heme domains to those of the parent enzymes. Unfolded chimeras showed no activity toward 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 toward all substrates tested. Several chimeric heme domains were more active than the best parent, CYP102A1-h, on one or more substrates. Chimera 169–197 was the most active toward 12-pNCA, 46–73 had the highest activity on 2-phenoxy-ethanol, and 43–135 and 165–256 were the most active on allyloxy-benzene. The chimeras also showed different specificities, falling roughly into three groups: (1) chimeras with little or no detectable activity toward any substrate; (2) parent-like chimeras, active on 12-pNCA and 2-phenoxy-ethanol; and (3) chimeras with altered substrate specificities relative to the parents, due to loss of activity toward 12-pNCA and/or acquisition of activity on allyloxy-benzene. 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-phenoxy-ethanol and allyloxy-benzene, but not 12-pNCA. The members of each pair of mirror chimeras had equivalent folding properties but were not functionally equivalent.

Peroxidase Activities of Chimeric Heme Domains

P450s can reduce peroxide to water (and a proton) using a mechanism similar to that of peroxidases [37], although the intrinsic rate for P450s is orders of magnitude

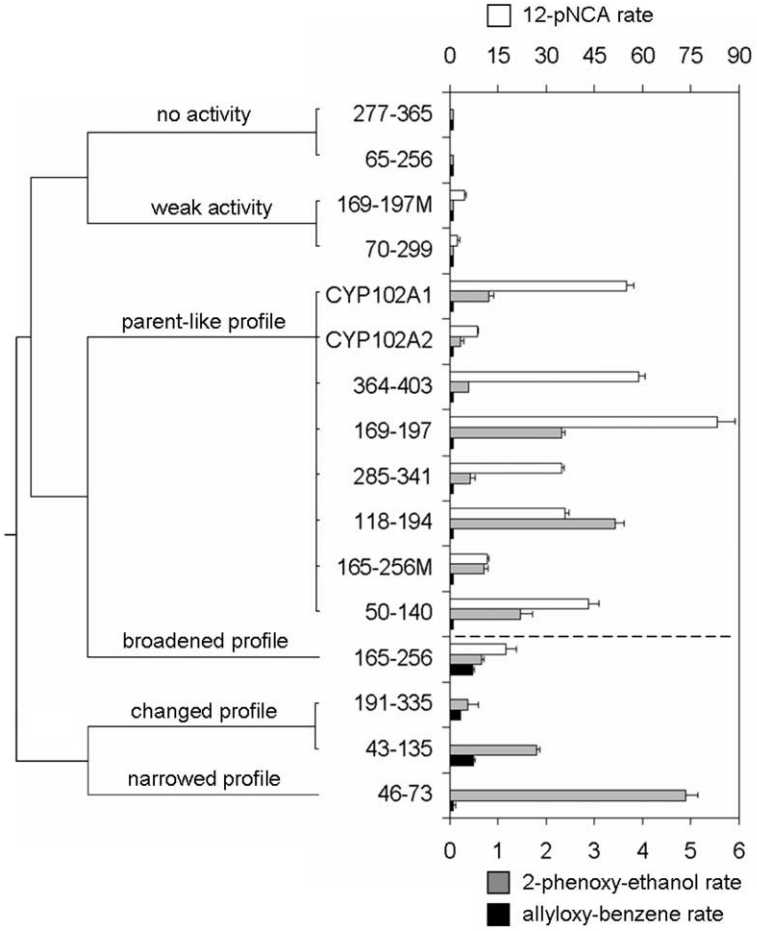

Figure 3. Substrate Activity Profiles of CYP102A1-h, CYP102A2-h, and the Folded Chimeric Heme Domains

Chimeras were assayed for peroxygenase activity on 12-pNCA, 2-phenoxy-ethanol, and allyloxy-benzene, 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-phenoxy-ethanol), 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 allyloxy-benzene). 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/min. 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-phenoxy-ethanol, and allyloxy-benzene, respectively. Chimeras lacking detectable peaks in the CO difference spectra showed no activity on the substrates assayed.

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 [38]. CYP102A1-h and CYP102A2-h both show low, but detectable, peroxidase activity (Table 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.

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 1), with none more stable than CYP102A1-h ($T_m = 55°C$). More than half of the folded chimeric heme domains were more thermostable than CYP102A2-h ($T_m = 44°C$); 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.

Folding and Monooxygenase Activities of Chimera-Reductase Fusion Proteins

We fused five of the functional chimeras (43–135, 46–73, 118–194, 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-phenoxy-ethanol, and allyloxy-benzene 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 2). Monooxygenase activities of the fusion proteins were comparable to the peroxygenase activities of the respective heme domains for 12-pNCA and 2-phenoxy-ethanol. The fusion protein monooxygenases were roughly an order of magnitude more active toward allyloxy-benzene than were the heme domain peroxygenases.

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

Table 2. Monooxygenase Activities of Holoenzymes CYP102A1, CYP102A2 and Chimera-Reductase Fusion Proteins, CRFPs, on Three Substrates

Protein[a]	12-pNCA[b]	2-Phenoxy-Ethanol[c]	Allyloxy-Benzene[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/min. Activity less than 0.1 was not detectable (N.D.).
[c] Reported in nmol product/nmol P450/min. Activity less than 0.2 was not detectable (N.D.).

to CYP102A1-h and CYP102A2-h (see Figure 3); 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 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 allyloxy-benzene 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.

Discussion

Activities and Specificities of Recombined P450s

While a chimeric protein often equals the sum of its parts [39], it is also possible for a chimera to exceed its parents and find amino acid combinations that allow new properties to emerge [40–46]. Creating these beneficial amino acid combinations from different parental sequences that are "prescreened" by nature is one goal of protein engineering by recombination [41]. We find that recombination is an effective way to alter the function of bacterial cytochrome P450s—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 3). 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-phenoxy-ethanol than the parent most active on that substrate. Three chimeric heme domains hydroxylated allyloxy-benzene, an activity not detectable in either of the parent heme domains.

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 [30]. A chimeric heme domain can also be reconstituted into a holoenzyme by addition of a CYP102 reductase domain. Four of the five such CRFPs that were constructed in fact functioned as monooxygenases (Table 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 CYP102A1 peroxygenase regioselectivity [21].

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 [47]. 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 [41–45] and with mammalian P450s [46, 48]. 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 [43]. 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 [41], the high levels of point mutation made it impos-

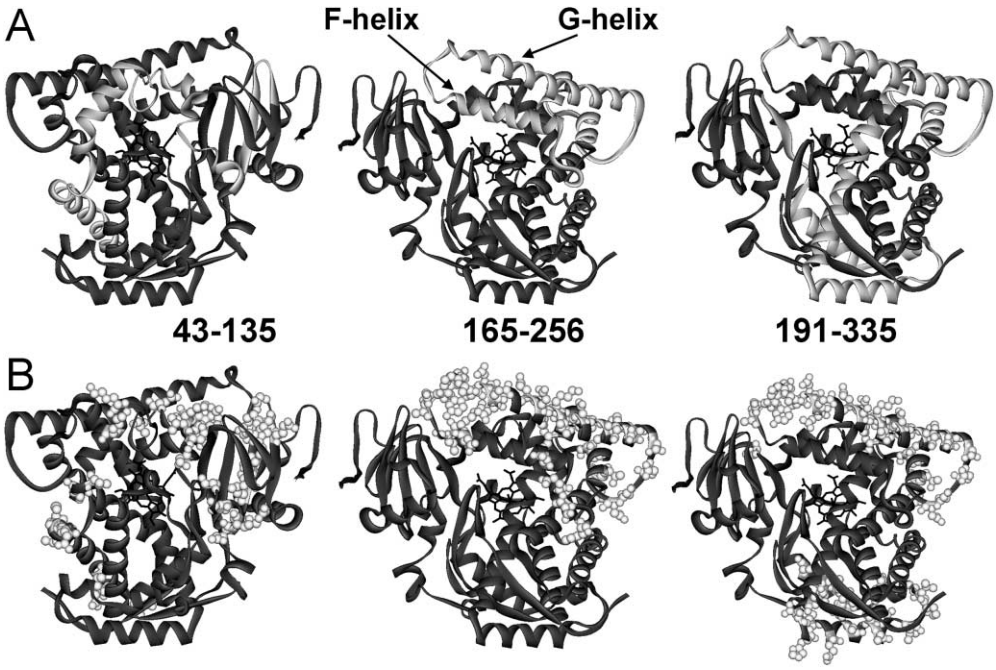

Figure 4. Structural Models of Chimeric Cytochrome P450 Heme Domains

The numbers shown for each chimera represent the residues from CYP102A1 that have been swapped for those from CYP102A2.
(A) Residues from CYP102A1 (dark gray) and CYP102A2 (light gray) are mapped onto the structure of the CYP102A1 heme domain [22]. 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.

sible 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.

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 4A, the swapped fragments encompass nontrivial 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 [49, 50]), 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 (data not shown).

For proteins that share 63% amino acid sequence identity, most nonshared 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 4B). Such mutations are less disruptive, on average, than changes in the core. CYP102A1 and CYP102A2 differ at six of the 21 residues postulated to contact a fatty-acid substrate [51]. 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 [52–54]. The "novel" activity on allyloxy-benzene 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 3). 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 P450 heme domain chimeras with the highest peroxidase and peroxygenase rates (169–197) and broadest substrate specificity (165–256) have both swapped a region of amino acids comprising the F helix. It has been shown that the F and G helices (Figure 4A) move approximately 6 Å upon substrate binding [23–25], and mutations affecting catalytic activity have been observed there in other protein engineering studies [54]. The new, favorable combination of the F helix from CYP102A2 and the G helix from CYP102A1 in the heme domain chimera 169–197 and the complete substitution of the F and G helices in CYP102A1 with that from CYP102A2 in the heme domain chimera 165–256 indicate a key role of this region in determining P450 catalytic properties.

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 [44, 55] or if appropriate structural information is incorporated in the library design [16, 17]. 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 that our studies will begin to address this question. Figure 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 ($<m>$ = 22) than those with altered profiles ($<m>$ = 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 [56, 57], but we still have little information on how recombination 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 1). We found very similar results in a recent study of more than 16,000 chimeric lactamases [17]. 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 CYP102A1-h and CYP102A2-h will fold properly, because most are characterized by values of $E < 30$–35 (Figure 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 multiple crossovers, reducing disruption becomes an important design criterion. Figure 5 shows an in silico analysis of 5000 different libraries in which 10 crossovers were allowed between CYP102A1-h and CYP102A2-h, with the crossover positions chosen at random. Each library contains $2^{11} = 2048$ different chimeric sequences. For each library, we calculated (1) the fraction that is predicted to fold (using F_{30} = fraction of sequences with $E \leq 30$) and (2) the average level of effective mutation in these folded chimeras ($<m>_{30}$). 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 CYP102A1-h and CYP102A2-h 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.

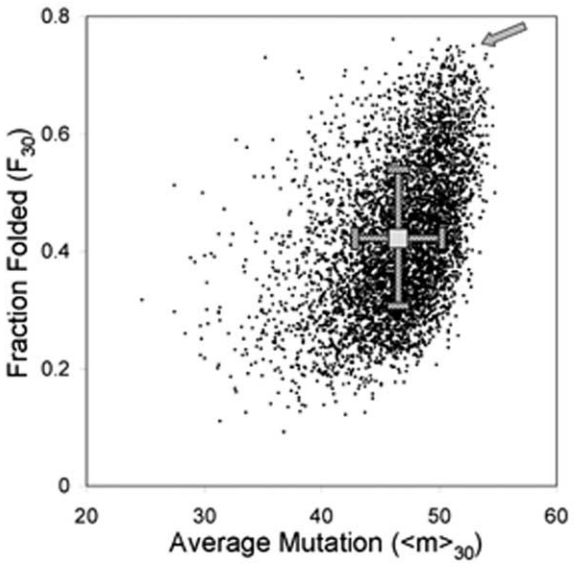

Figure 5. Theoretical Library Analysis

5000 libraries were generated in silico, in which 10 randomly selected crossovers were allowed between CYP102A1-h and CYP102A2-h. 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.

Significance

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 [30, 35, 54]; 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 [17], which can be generated using targeted recombination methods [27, 58]. With appropriate high throughput screening, we may then be able to discover new P450s with properties that nature has not yet needed or explored.

Experimental Procedures

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).

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) [16, 22, 26]. Hydrogens, backbone nitrogens, backbone oxygens, and heme atoms were not included in the calculation. The sequences of CYP102A1 and CYP102A2 were aligned using ClustalW [59], revealing the existence of a 1 amino acid insertion relative to CYP102A1, between Q229 and S230. This insertion was ignored in the calculations. CYP102A1 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) [22], we calculated the total number of contacts disrupted (E) and effective level of mutation (m) for all 2^{11} (2048) chimeras in 5000 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 $<m>_{30}$ in this low-disruption fraction.

Construction of Chimeras

Selected chimeras were constructed using SOEing methods, as described previously [27]. 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°C followed by 22 cycles of 95°C for 1 min, 48°C for 1 min, and 72°C for 2 min with a final extension at 72°C for 10 min. 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°C for 1 min, 46°C for 1 min, and 72° for 2 min for 22 cycles with a final extension at 72°C for 10 min. These two products were assembled in a two-step PCR reaction: 95°C for 1 min followed by 14 cycles of 95°C for 1 min, 46°C for 1 min, and 72°C for 2 min. External primers were added, followed by PCR 95°C for 1 min, then followed by 14 cycles of 95°C for 1 min, 46°C for 1 min, and 72°C for 2 min, with a final extension of 72°C for 10 min. 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.

Protein Expression

Chimeric heme domains and CRFPs were expressed in catalase-deficient *E. coli* (strain SN0037) [29] using the isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible pCWori vector [28]. 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 hr at 180 rpm and 25°C. This procedure yields approximately 80 mg/l of P450 protein for CYP102A1 and CYP102A2. Cultures were pelleted at 5500 × 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.

Folding Assay

Carbon monoxide-reduced difference spectroscopy was performed as reported [60]. 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 s 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 min 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.

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 min 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 6 s 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 $M^{-1}cm^{-1}$ [34]. All rates reported represent the average of three independent experiments, with error bars corresponding to one standard deviation.

Catalytic activities on 2-phenoxy-ethanol and allyloxy-benzene were determined using 2 μM enzyme, 20 mM H_2O_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-phenoxy-ethanol (100 mM) and allyloxy-benzene (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 min. 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 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 20 min 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 4800 $M^{-1}cm^{-1}$.

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.

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 toward the substrates 12-pNCA, 2-phenoxy-ethanol, and allyloxy-benzene. 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.

Peroxidase Activity

Peroxidase activities were measured by monitoring the accumulation of the radical cation of ABTS at 414 nm [38]. 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 min. Rates were determined from the initial slope of the time course (typically the first 30 s). An extinction coefficient of 36,000 $mol^{-1}cm^{-1}$ for ABTS was used.

Thermostability

Cell extracts were heated in a thermocycler for 10 min at various temperatures, followed by cooling to 4°C. Extracts were centrifuged for 5 min at 3500 $\times$ 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.

Acknowledgments

The authors thank 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.).

Received: June 13, 2003
Revised: November 12, 2003
Accepted: December 2, 2003
Published: March 18, 2004

References

1. Ortiz de Montellano, P.R. (1995). Cytochrome P450: Structure, Mechanism, and Biochemistry (New York: Plenum Press).
2. Gotoh, O. (1993). Evolution and differentiation of P-450 genes. In Cytochrome P450, 2nd Edition, T. Omura, Y. Ishimura, and Y. Fujii-Kuriyama, eds. (Tokyo: Kodansha), pp. 255–272.
3. Gonzalez, F.J., and Nebert, D.W. (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, W.H., and Muller, H.H. (1981). The induction of cytochrome P-450 in *Lodderomyces elongisporus*. Z. Allg. Mikrobiol. *21*, 313–321.
5. Porter, T.D., and Coon, M.J. (1991). Cytochrome P-450. Multiplicity of isoforms, substrates, and catalytic and regulatory mechanisms. J. Biol. Chem. *266*, 13469–13472.
6. Sono, M., Roach, M.P., Coulter, E.D., and Dawson, J.H. (1996). Heme-containing oxygenases. Chem. Rev. *96*, 2841–2888.
7. Eichhorn, G.L., and Marzilli, L.G. (1988). Heme Proteins (New York: Elsevier Science Publishing Co., Inc.).
8. Cirino, P.C., and Arnold, F.H. (2002). Exploring the diversity of heme enzymes through directed evolution. In Directed Molecular Evolution of Proteins, S. Brakmann and K. Johnsson, eds. (Germany: Wiley-VCH), pp. 215–243.
9. Stemmer, W.P. (1994). DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc. Natl. Acad. Sci. USA *91*, 10747–10751.
10. Coco, W.M., Levinson, W.E., Crist, M.J., Hektor, H.J., Darzins, A., Pienkos, P.T., Squires, C.H., and Monticello, D.J. (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, J.A., and Arnold, F.H. (1998). Molecular evolution by staggered extension process (StEP) in vitro recombination. Nat. Biotechnol. *16*, 258–261.
12. Volkov, A.A., Shao, Z., and Arnold, F.H. (1999). Recombination and chimeragenesis by in vitro heteroduplex formation and in vivo repair. Nucleic Acids Res. *27*, e18.
13. Kikuchi, M., Ohnishi, K., and Harayama, S. (2000). An effective family shuffling method using single-stranded DNA. Gene *243*, 133–137.
14. Lutz, S., Ostermeier, M., Moore, G.L., Maranas, C.D., and Benkovic, S.J. (2001). Creating multiple-crossover DNA libraries independent of sequence identity. Proc. Natl. Acad. Sci. USA *98*, 11248–11253.
15. Sieber, V., Martinez, C.A., and Arnold, F.H. (2001). Libraries of hybrid proteins from distantly related sequences. Nat. Biotechnol. *19*, 456–460.
16. Voigt, C.A., Martinez, C., Wang, Z.G., Mayo, S.L., and Arnold, F.H. (2002). Protein building blocks preserved by recombination. Nat. Struct. Biol. *9*, 553–558.
17. Meyer, M.M., Silberg, J.J., Voigt, C.A., Endelman, J.B., Mayo, S.L., Wang, Z.G., and Arnold, F.H. (2003). Library analysis of SCHEMA-guided protein recombination. Protein Sci. *12*, 1686–1693.
18. Narhi, L.O., Kim, B.H., Stevenson, P.M., and Fulco, A.J. (1983). Partial characterization of a barbiturate-induced cytochrome P-450-dependent fatty acid monooxygenase from *Bacillus megaterium*. Biochem. Biophys. Res. Commun. *116*, 851–858.
19. Munro, A.W., Leys, D.G., McLean, K.J., Marshall, K.R., Ost, T.W., Daff, S., Miles, C.S., Chapman, S.K., Lysek, D.A., Moser, C.C., et al. (2002). P450 BM3: the very model of a modern flavocytochrome. Trends Biochem. Sci. *27*, 250–257.
20. Li, Q.S., Ogawa, J., Schmid, R.D., and 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.
21. Cirino, P.C., and Arnold, F.H. (2002). Regioselectivity and activity of cytochrome P450 BM-3 and mutant F87A in reactions driven by hydrogen peroxide. Adv. Synth. Catalysis *344*, 932–937.
22. Haines, D.C., Tomchick, D.R., Machius, M., and Peterson, J.A. (2001). Pivotal role of water in the mechanism of P450BM-3. Biochemistry *40*, 13456–13465.
23. Paulsen, M.D., and Ornstein, R.L. (1995). Dramatic differences in the motions of the mouth of open and closed cytochrome P450BM-3 by molecular dynamics simulations. Proteins *21*, 237–243.
24. Modi, S., Sutcliffe, M.J., Primrose, W.U., Lian, L.Y., and Roberts, G.C. (1996). The catalytic mechanism of cytochrome P450 BM3 involves a 6 Å movement of the bound substrate on reduction. Nat. Struct. Biol. *3*, 414–417.
25. Li, H., and Poulos, T.L. (1997). The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. Nat. Struct. Biol. *4*, 140–146.
26. Ravichandran, K.G., Boddupalli, S.S., Hasermann, C.A., Peterson, J.A., and Deisenhofer, J. (1993). Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. Science *261*, 731–736.
27. Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K., and Pease, L.R. (1989). Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. Gene *77*, 61–68.
28. Barnes, H.J., Arlotto, M.P., and Waterman, M.R. (1991). Expression and enzymatic activity of recombinant cytochrome P450 17 alpha-hydroxylase in *Escherichia coli*. Proc. Natl. Acad. Sci. USA *88*, 5597–5601.
29. Nakagawa, S., Ishino, S., and Teshiba, S. (1996). Construction of catalase deficient *Escherichia coli* strains for the production of uricase. Biosci. Biotechnol. Biochem. *60*, 415–420.
30. Cirino, P.C., and Arnold, F.H. (2003). A self-sufficient peroxide-driven hydroxylation biocatalyst. Angew. Chemie. Intl. Edn. Engl. *42*, 3299–3301.
31. Omura, T., and Sato, R. (1964). The carbon monoxide-binding pigment of liver microsomes. J. Biol. Chem. *239*, 2370–2378.
32. Martinis, S.A., Blanke, S.R., Hager, L.P., Sligar, S.G., Hoa, G.H., Rux, J.J., and Dawson, J.H. (1996). Probing the heme iron coordination structure of pressure-induced cytochrome P420cam. Biochemistry *35*, 14530–14536.
33. Wells, A.V., Li, P., Champion, P.M., Martinis, S.A., and Sligar,

S.G. (1992). Resonance Raman investigations of *Escherichia coli*-expressed *Pseudomonas putida* cytochrome P450 and P420. Biochemistry *31*, 4384–4393.

34. Schwaneberg, U., Schmidt-Dannert, C., Schmitt, J., and Schmid, R.D. (1999). A continuous spectrophotometric assay for P450 BM-3, a fatty acid hydroxylating enzyme, and its mutant F87A. Anal. Biochem. *269*, 359–366.
35. Li, Q.S., Schwaneberg, U., Fischer, M., Schmitt, J., Pleiss, J., Lutz-Wahl, S., and Schmid, R.D. (2001). Rational evolution of a medium chain-specific cytochrome P-450 BM-3 variant. Biochim. Biophys. Acta *1545*, 114–121.
36. Otey, C.R., and Joern, J.M. (2003). High-throughput screen for aromatic hydroxylation. In Directed Enzyme Evolution: Screening and Selection Methods, F.H. Arnold and G. Georgiou, eds. (Totowa, NJ: Humana Press), pp. 141–148.
37. Mansuy, D. (1998). The great diversity of reactions catalyzed by cytochromes P450. Comp. Biochem. Physiol. C Pharmacol. Toxicol. Endocrinol. *121*, 5–14.
38. Childs, R.E., and Bardsley, W.G. (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.
39. Beguin, P. (1999). Hybrid enzymes. Curr. Opin. Biotechnol. *10*, 336–340.
40. Campbell, R.K., Bergert, E.R., Wang, Y., Morris, J.C., and Moyle, W.R. (1997). Chimeric proteins can exceed the sum of their parts: implications for evolution and protein design. Nat. Biotechnol. *15*, 439–443.
41. Crameri, A., Raillard, S.A., Bermudez, E., and Stemmer, W.P. (1998). DNA shuffling of a family of genes from diverse species accelerates directed evolution. Nature *391*, 288–291.
42. Kumamaru, T., Suenaga, H., Mitsuoka, M., Watanabe, T., and Furukawa, K. (1998). Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase. Nat. Biotechnol. *16*, 663–666.
43. Raillard, S., Krebber, A., Chen, Y., Ness, J.E., Bermudez, E., Trinidad, R., Fullem, R., Davis, C., Welch, M., Seffernick, J., et al. (2001). Novel enzyme activities and functional plasticity revealed by recombining highly homologous enzymes. Chem. Biol. *8*, 891–898.
44. Hansson, L.O., Bolton-Grob, R., Massoud, T., and Mannervik, B. (1999). Evolution of differential substrate specificities in Mu class glutathione transferases probed by DNA shuffling. J. Mol. Biol. *287*, 265–276.
45. Christians, F.C., Scapozza, L., Crameri, A., Folkers, G., and Stemmer, W.P. (1999). Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling. Nat. Biotechnol. *17*, 259–264.
46. Ramarao, M.K., Straub, P., and 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.
47. Sevrioukaova, I.F., Li, H., Zhang, H., Peterson, J.A., and Poulos, T.L. (1999). Structure of a cytochrome P450-redox partner electron-transfer complex. Proc. Natl. Acad. Sci. USA *96*, 1863–1868.
48. Pompon, D., and Nicolas, A. (1989). Protein engineering by cDNA recombination in yeasts: shuffling of mammalian cytochrome P-450 functions. Gene *83*, 15–24.
49. Roy, S.W., Nosaka, M., de Souza, S.J., and Gilbert, W. (1999). Centripetal modules and ancient introns. Gene *238*, 85–91.
50. Go, M., and Nosaka, M. (1987). Protein architecture and the origin of introns. Cold Spring Harb. Symp. Quant. Biol. *52*, 915–924.
51. Li, H., and Poulos, T.L. (1999). Fatty acid metabolism, conformational change, and electron transfer in cytochrome P-450(BM-3). Biochim. Biophys. Acta *1441*, 141–149.
52. Zha, D.X., Wilensek, S., Hermes, M., Jaeger, K.E., and Reetz, M.T. (2001). Complete reversal of enantioselectivity of an enzyme-catalyzed reaction by directed evolution. Chem. Commun. *24*, 2664–2665.
53. Zhang, J.H., Dawes, G., and Stemmer, W.P. (1997). Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. Proc. Natl. Acad. Sci. USA *94*, 4504–4509.
54. Glieder, A., Farinas, E.T., and Arnold, F.H. (2002). Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. Nat. Biotechnol. *20*, 1135–1139.
55. Moore, J.C., Jin, H.M., Kuchner, O., and Arnold, F.H. (1997). Strategies for the in vitro evolution of protein function: enzyme evolution by random recombination of improved sequences. J. Mol. Biol. *272*, 336–347.
56. Cui, Y., Wong, W.H., Bornberg-Bauer, E., and Chan, H.S. (2002). Recombinatoric exploration of novel folded structures: a heteropolymer-based model of protein evolutionary landscapes. Proc. Natl. Acad. Sci. USA *99*, 809–814.
57. Bogarad, L.D., and Deem, M.W. (1999). A hierarchical approach to protein molecular evolution. Proc. Natl. Acad. Sci. USA *96*, 2591–2595.
58. Hiraga, K., and Arnold, F.H. (2003). General method for sequence-independent site-directed chimeragenesis. J. Mol. Biol. *330*, 287–296.
59. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.
60. Schenkman, J.B., and Jansson, I. (1998). Spectral analyses of cytochromes P450. In Cytochrome P450 Protocols, Volume *107*, I.R. Phillips and E.A. Shephard, eds. (Totowa, NJ: Humana Press, Inc.), pp. 25–33.