Libraries of hybrid proteins from distantly related sequences

Volker Sieber, Carlos A. Martinez, and Frances H. Arnold*

We introduce a method for sequence homology–independent protein recombination (SHIPREC) that can create libraries of single-crossover hybrids of unrelated or distantly related proteins. The method maintains the proper sequence alignment between the parents and introduces crossovers mainly at structurally related sites distributed over the aligned sequences. We used SHIPREC to create a library of interspecies hybrids of a membrane-associated human cytochrome P450 (1A2) and the heme domain of a soluble bacterial P450 (BM3). By fusing the hybrid gene library to the gene for chloramphenicol acetyl transferase (CAT), we were able to select for soluble and properly folded protein variants. Screening for 1A2 activity (deethylation of 7-ethoxyresorufin) identified two functional P450 hybrids that were more soluble in the bacterial cytoplasm than the wild-type 1A2 enzyme.

Recombination is an important mechanism for the acquisition of novel function in proteins. Various methods for using recombination in the design of protein libraries for laboratory evolution have been described[1–6]. The majority of these "DNA-shuffling" methods can recombine only closely related sequences (more than ~70% identity) and generate crossovers only in regions of the highest sequence identity. However, many proteins having closely similar three-dimensional structures show low or even no discernible sequence similarity. Rational design[7,8], computational[9], and combinatorial[6] approaches have nonetheless shown that functional proteins can be obtained by recombination of such distantly related sequences.

Ostermeier *et al.* recently reported a method for constructing hybrid gene libraries that is independent of sequence homology[6]. Two parent genes are truncated incrementally and the fragments are fused pairwise, creating libraries of single-crossover hybrid sequences. However, because random-length fragments are fused to other random-length fragments, the gene length is not conserved and recombination occurs mainly at sites that are not structurally related. Only a tiny fraction of the crossovers connect the two parent genes at sites where the sequences align. Thus this approach is most useful when crossover is limited to a small section of the gene and in combination with a powerful selection system.

Here we present a method for sequence homology–independent protein recombination (SHIPREC) that produces libraries of hybrid sequences in which the crossovers retain proper sequence alignment and therefore occur predominantly at structurally related sites. We have used this method to produce hybrids of a membrane-associated human cytochrome P450 (1A2)[10] and a soluble bacterial P450 (the heme domain of cytochrome P450 BM3 from *Bacillus megaterium*)[11]. These proteins share only 16% amino acid sequence identity. We generated a SHIPREC library of sequences with single crossovers all along the aligned genes. An efficient selection for hybrid genes encoding soluble, folded proteins was followed by screening of the selected cells for P450 enzyme activity. In this way we were able to find sequences encoding human P450 enzymes that were more soluble in the bacterial cytoplasm.

Results and discussion

The recombination method. The goal is to create libraries of hybrid genes from distantly related sequences. To maximize the fraction of hybrid genes encoding functional proteins, the crossovers should occur at positions that are in similar structural environments. However, without sufficient sequence similarity to direct homologous recombination, known DNA-shuffling methods will not work. We therefore developed a method that uses sequence length rather than sequence similarity to achieve crossovers at positions likely to be structurally related.

The SHIPREC method is outlined in Figure 1. The key step is the fusion of two parent genes to form a gene dimer, which is then digested with DNase I to form an ensemble of random-length fragments. Fragments of a length corresponding to either of the parent sequences are isolated. After S1 treatment to produce blunt ends, the remaining ends of the two genes are fused by circularization. The gene that was at the 5′ position in the dimer will now be at the 3′ position and will donate the C terminus of the hybrid protein. The fragment size selection ensures that the two amino acids that meet at the crossover are in similar positions in the parent structures.

Library construction. A dimer of the genes for 1A2 and BM3 was constructed with a short linker sequence that contains unique restriction sites (Fig. 1). This dimer of ~3 kilobases was digested with DNase I in the presence of Mn^{2+} to favor double-strand breaks over nicks[1,12] until a smear from 3 kilobases down to 0 appeared on an agarose gel. The DNA ends were converted from staggered to blunt ends by treatment with S1 nuclease. It was important to use S1 nuclease in this step; the use of DNA polymerases (T4, Vent) to repair nicks and gaps and create blunt ends, reminiscent of methods for creating libraries of circularly permutated sequences[13,14], yielded fragments that could not be resolved well in later steps. Fragments in the range of 1,450–1,550 base pairs were isolated by preparative agarose gel electrophoresis, circularized by intramolecular blunt-end ligation, and linearized by restriction digestion using an enzyme that has a site in the linker region between the two genes. To improve the yield in subsequent steps, we amplified the fragment pool by PCR using an N-terminal primer of BM3 and a C-terminal primer of 1A2.

Division of Chemistry and Chemical Engineering 210-41, California Institute of Technology, Pasadena CA 91105.
**Corresponding author (frances@cheme.caltech.edu).*

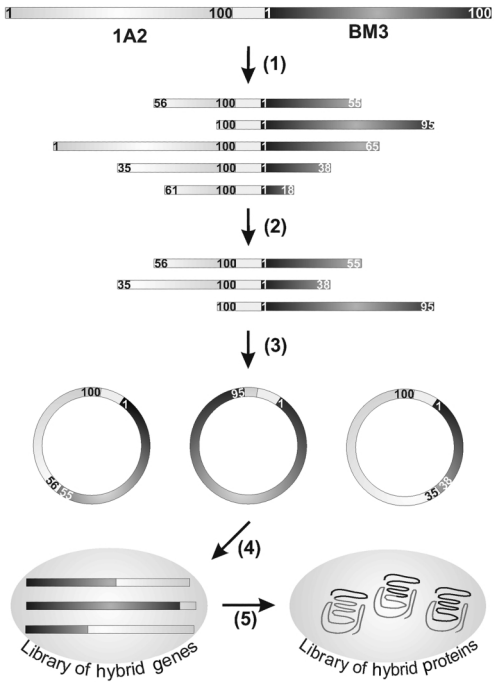

Figure 1. SHIPREC procedure. A gene dimer that consists of (from 5′ to 3′) the gene of protein 1 (1A2), a linker sequence containing useful restriction sites, and the gene of protein 2 (BM3) is constructed. (1) This dimer is fragmented (e.g., by digestion with DNase I in the presence of Mn^{2+}) and treated to produce blunt ends (e.g., using S1 nuclease or T4 polymerase). (2) Fragments of the length of a single gene (plus the length of the linker sequence) are separated from the pool (e.g., by agarose gel electrophoresis). (3) Single-gene length fragments are circularized by intramolecular blunt-end ligation. (4) Linearization of the circular DNA pieces by restriction digestion in the linker region creates a library of chimeric genes with members having an N-terminal part from protein 2 and a C-terminal part from protein 1, the crossovers being distributed over the entire length of the gene. (5) The chimeric genes are cloned into an expression vector directly or after amplification by PCR using one terminal primer from each of the two parents. The numbering shown in the genes represents percentage of gene length.

Library analysis. The amplified library was cloned into pBluescript KS(II)+ and analyzed by colony PCR on 44 randomly chosen colonies. The length variation was determined by PCR with primers that anneal to the termini of the hybrid genes. About 20% of the fragments were significantly smaller than 1.5 kilobases (mostly <500 base pairs). The average length of the remaining fragments was 1.4 kilobases with a standard deviation of 0.1 kilobases. (Note that the fragments purified from the agarose gel contain an additional 40 base-pair linker.)

The locations of the crossovers in the full-length variants were determined by colony PCR with different combinations of internal and terminal primers and found to be distributed over the gene, with 43% in the first third and ~28% each in the central and last thirds. Three randomly chosen clones were sequenced and found to consist of the first 612, 810, and 1,182 base pairs of BM3 followed by the last 816, 547, and 283 base pairs of 1A2.

Library pre-selection. Although sequence alignment is retained, we still expect that a significant fraction of the hybrid genes code for unfolded or unstable proteins. For one thing, two thirds of the library should have a frameshift at the crossover position, leading to massive mutation and premature truncation of the proteins. We wished to eliminate these variants and at the same time isolate hybrid proteins that are more soluble than P450 1A2. We therefore used a general pre-selection to eliminate unfolded and insoluble variants and greatly reduce the number that had to be screened for enzyme activity.

For this pre-selection, the library variants were fused to CAT. Bacteria containing P450 variants with a frameshift will express CAT in the wrong frame and not be resistant to chloramphenicol (cam). In addition, Maxwell *et al.*[15] recently showed that cells expressing fusions of an insoluble protein to CAT had decreased resistance to cam compared with fusions with soluble proteins. We reasoned that the localization of 1A2 in the membrane would impede the efficient trimerization required for the function of the fused CAT[15,16]. In contrast, fusion to cytosolic proteins (BM3 and any folded, soluble hybrid) should allow CAT to be fully active. CAT fusion selection would therefore eliminate incomplete variants and also select more soluble ones.

To test this approach we constructed several derivatives of 1A2 and BM3 fused to CAT and analyzed their influence on the resistance of XL1-Blue cells to different levels of cam (Table 1). We found it essential to remove the start codon of *cat* to avoid reinitiation of translation in reading frame (RF) variants. Cells with the negative controls (empty, *1A2*) did not grow on cam concentrations ≥10 μg/ml, whereas cells with the positive control (*catctrl*) grew without detectable differences at cam concentrations as high as 1 mg/ml. Cells with the 1A2-cat fusion showed decreased growth even at 10 μg/ml and did not grow above 40 μg/ml cam concentration, whereas those with the BM3 fusions grew very well

Table 1. Resistance toward chloramphenicol conferred by different *cat* fusions when expressed in XL1-Blue cells[a]

Variant[b]	Description[c]	Liquid[d]	Agar40[e,f]	Agar200[e]
Empty	No insert	–	–	–
1A2	Full-length 1A2	–	–	–
1A2cat	Full-length 1A2, fused to *cat* gene	–	+	–
BM3cat	Full-length BM3, fused to *cat* gene	++++	+++	++
1A2rfcat	Full-length 1A2 with shift in RF at AA 478, fused to *cat* gene	–	–	–
BM3rfcat	Full-length BM3 with shift in RF at AA 485, fused to *cat* gene	+++	+++	+
1A2delcat	1A2 with deletion of amino acids 60–478, fused to *cat* gene	++++	++	–
BM3delcat	BM3 with deletion of amino acids 60–390, fused to *cat* gene	+++	++	++
BM3rfdel-cat	BM3 with deletion of amino acids 60–390, including shift in RF, fused to *cat* gene	++	+++	–
catctrl	*cat* gene	++++	+++	++

[a]Resistance was measured by growth rate.
[b]Name of variant as it appears in the text in italics.
[c]Gene that replaces *1A2* in pCW1A2.
[d]Cells were grown in TB^{++} medium with different concentrations of chloramphenicol (cam). OD was measured after 12 h and compared to OD of the cam-free control. Shown are the concentrations of cam at which the growth was still >20% of the control: ++++, ≥160 μg/ml; +++, 80 μg/ml; ++, 40 μg/ml; +, 20 μg/ml; –, <20 μg/ml. [e]Cells were grown for several days at 30°C on TB^{++} agar with 40 or 200 μg/ml cam. Times that colonies appeared are given: +++, ≤24 h; ++, 36 h; +, 48 h; –, no colonies after 60 h. [f]Conditions chosen for the pre-selection (see Experimental Protocol for details).

even at the highest concentration of the antibiotic. The amount of folded protein in the cells, determined by the absorption at 450 nm after reduction and CO binding[17], is similar for 1A2-CAT and BM3-CAT and cannot be the cause of this difference. Instead, it indicates that the position or flexibility of the protein (free in cytoplasm or rigid in membrane) indeed affects the activity of CAT and that selection for soluble, folded proteins is feasible.

Cells with *1A2rfcat*, a frameshift mutant, did not grow in the presence of cam at all. But those with *BM3rfdelcat* grew unexpectedly well. Upstream from *cat*, within the BM3 gene, there is an ATG codon that could be acting as an initiation codon because it keeps the correct frame of the expressed cat. This, however, should not significantly affect the selection of the hybrid library because the hybrids have the C terminus of 1A2, not BM3. There is also a difference between the two deletion variants *1A2delcat* and *BM3delcat* in that the latter confers a higher resistance than the former. Neither is expected to form a folded protein. Due to its higher hydrophobicity, 1A2del, when overexpressed, is more prone than BM3del to form inclusion bodies, pulling CAT into inclusion bodies as well[15]. The ability of the pre-selection to eliminate 1A2 reading frame variants is very helpful, and if the selection pressure is adjusted carefully, small increases in solubility should be observable.

For the library pre-selection, we chose to grow the cells on agar plates with 40 μg/ml cam. The hybrid DNA library was cloned into pCW1A2rfcat, an inactive variant, chosen so that unrestricted vector contamination will not appear as false positives and also because these contaminants are efficiently removed by the cat selection. Approximately 250,000 clones were obtained after plating on Luria–Bertani broth containing ampicillin (LB Amp). Cells were resuspended in LB, and dilutions were plated on TB-expression agar (see Experimental Protocol), including 40 μg/ml cam, and on LB as a reference. The number of colonies appearing on the selection agar was ~20% of those on LB.

CO-binding screen for folded P450s. To test the efficiency of the selection, 116 variants from the selected library were analyzed for the production of folded P450 protein using the increase in absorbance upon reduction and CO binding[17]. Approximately 80% showed a strong peak at 450 nm, indicating that the hybrid variants were properly folded. PCR analysis and sequencing of several clones showed that all these folded variants consisted of mostly BM3 with a crossover to 1A2 at the far C terminus. All the folded variants had the junction in the correct reading frame.

Screen for active P450 chimeras. Approximately 2,000 variants from the selected library were grown in 96-well plates and screened for the deethylation of 7-ethoxyresorufin, an activity that is exhibited by 1A2 (ref. 18) but not BM3. Twelve variants that showed the greatest increase in fluorescence were rescreened. The DNA was purified from cells in the microplate and used to transform *Escherichia coli* XL1-Blue. Expression and activity of each variant were tested after growth in 5 ml cultures and in parallel in microtiter plates. Two variants, termed RC1 and RC2, showed the characteristic P450 absorption at 450 nm as well as enzyme activity (development of fluorescence); the remaining 10 were false positives. RC1 and RC2 were sequenced to determine the position of the crossover (Fig. 2). In RC1 the first 15 amino acids of 1A2 were replaced by the 14 N-terminal amino acids of BM3. RC1, therefore, is nearly a full-length 1A2 with a more hydrophilic N terminus. RC2 contains the first 44 nucleotides from BM3 but with a deletion of one A in the A-quintuplet at nucleotides 27–31. This results in a frameshift at amino acid 11 of BM3. The crossover 12 nucleotides further downstream restores the correct reading frame at amino acid 25 of 1A2.

Characterization of RC1 and RC2. Cytochrome P450 1A2 and chimeras RC1 and RC2 were subcloned to remove the cat fusions from the C terminus and expressed in *E. coli* DH5α. Cells were lysed, the cytosolic fraction was separated from the membrane by ultracentrifugation, and the concentrations of folded P450 in each fraction and in the whole cells were determined using the absorption at 450 nm after reduction and CO binding (see Table 2).

When lysates of *E. coli* expressing the chimeric proteins and the parent 1A2 sequence were prepared using 50 mM KP_i buffer, it was found that both hybrids expressed well and at levels close to 1A2. Whereas almost no 1A2 could be found in the soluble, cytosolic fraction, RC1 and RC2 appear there at concentrations of 8.8 nM and 30.4 nM, respectively. This corresponds to partitioning into the cytosol of ~14% for RC2 and 5% for RC1, compared with < 1.5% for 1A2. Western blot analysis of the different fractions using polyclonal antibodies against 1A2 gave the same result: the cytosolic fractions gave a signal for RC2, whereas no protein could be detected for 1A2.

Table 2. Distribution and activity of P450 enzymes in different *E. coli* cell compartments

	1A2				RC1	RC2			
[Phosphate buffer]	10 mM	50 mM	100 mM	500 mM	50 mM	10 mM	50 mM	100 mM	500 mM
Cytosolic P450 (%)[a]	2.5 ± 0.8	1.5 ± 0.9	4.9 ± 0.5	23 ± 1.7	5.0 ± 1.2	5.1 ± 0.5	14 ± 1.4	34 ± 3.3	97 ± 2.1
NADH oxidase activity in cytosol (%)[b]	2.1 ± 1.1	2.7 ± 0.8	2.7 ± 0.7	4.1 ± 0.3	5.0 ± 0.8	2.7 ± 1.3	10 ± 0.8	10 ± 0.3	12 ± 0.5

	1A2	RC1	RC2
[P450][c]	88 ± 0.3 nM	177 ± 0.8 nM	217 ± 0.1 nM
P450 activity in cytosolic fraction[d]	1	1.5	8
P450 activity in membrane fraction[d]	1	0.66	0.63
P450 Specific activity in membrane[e]	1	0.76	0.52

[a]Data represent average percentage ± s.d. of the total P450. Three measurements per sample were taken in each case. All samples were obtained from single lysates of 1 L cultures.
[b]NADH oxidase activity was measured from the initial decrease in absorbance of NADH with the addition of 10 μl of cytosolic fraction.
[c]Total concentration of cytochrome P450 estimated from CO-reduced difference spectra using an extinction coefficient of 91 nM/cm. The change in absorbance was obtained by subtracting the absorbance at 450 nM from the one at 490 nM. Thus, [P450] = $\Delta A \times 91$ nM for measurements made on a 1 cm path cuvette. Samples were prepared by fractionation with 50 mM KP_i buffer.
[d]Total activity was measured from the initial increase in fluorescence of 7-ethoxyresorufin with the addition of 2.0 mM $NADP^+$, 5.0 mM glucose-6-phosphate, 0.6 units/ml glucose-6-phosphate dehydrogenase, and 30 μg/ml oxidoreductase, relative to 1A2wt for samples obtained by fractionation at 50 mM KP_i buffer.
[e]In the absence of reliable concentration data for 1A2wt released in the cytosol after fractionation at 50 mM KP_i buffer, specific activity of the soluble enzymes was not calculated.

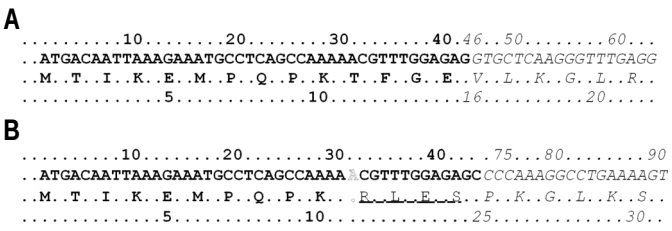

Figure 2. Nucleotide and amino acid sequences in the region of the crossover of functional hybrid cytochrome P450s RC1 and RC2. BM3 sequences are shown in bold and 1A2 sequences are shown in italics, both with the amino acid numbering of the respective wild type. (A) RC1; (B) RC2. The deleted A base is shown in outline, and the resulting new amino acid sequence is underlined.

RC1 and RC2 from the membrane fraction could also be solubilized with lower levels of detergents than 1A2. Both RC1 and RC2 were completely solubilized in 0.5% sodium cholate and 0.05% Triton X-100, whereas 1A2 was still bound to the membrane.

When lysates of *E. coli* expressing chimera RC2 were prepared at moderate (100 mM KP_i buffer) or high (500 mM) ionic strengths, the P450 partitioned readily into the cytosol (Table 2). At medium ionic strength, RC2 was solubilized to 34%, compared to only 4% for the mammalian CYP1A2. High concentrations of phosphate buffer afforded almost 100% of RC2 in the cytoplasm, compared to 23% for the mammalian parent. Even under these latter conditions, only ~10% of the total NADH oxidase activity is found in the cytosol, demonstrating a significant difference in the properties of RC2 relative to other membrane proteins. The lower solubility of RC1 compared to RC2 is due to the amino acid sequence VLKGLRPR still present in RC1. This combination of hydrophobic and basic residues probably sticks to the surface of the membrane, with the result that RC1 distributes less into the cytosol.

The P450 activities in each compartment were analyzed using an NADPH regeneration system and P450 oxidoreductase from rat in microsomes. RC1 and RC2 have slightly lower specific activity than 1A2. Dong *et al.*[19] reported a similar result for variants of 1A2 with deletions in the N terminus, and postulated less favorable interaction with the oxidoreductase. Indeed, we found that increasing the concentration of the reductase led to a greater increase in the activity of RC1 and RC2 than for 1A2 (data not shown).

All the soluble, folded hybrid P450s have the crossover close to one of the termini, where there is no structural similarity between 1A2 and BM3. SHIPREC efficiently produces the sequences with crossovers farther into the protein, but these apparently do not fold and bind heme. Earlier experiments to produce single-crossover hybrids of two very closely related cytochrome P450s 1A2 and 1A1 using homologous recombination in yeast also generated only variants with crossovers very close to the N or the C terminus[20]. Single crossovers in the interior of this protein are apparently too disruptive. The generation of useful hybrid cytochrome P450s will undoubtedly require multiple crossovers or, alternatively, mutational fine-tuning to counter the disruption introduced by recombination. Libraries containing multiple crossovers can be generated by iterative SHIPREC, ideally coupled with the *cat*-fusion selection to maintain a high concentration of folded, soluble hybrids.

Conclusions. Recombination by DNA shuffling has become an essential tool for creating high-quality sequence diversity for directed evolution. Unfortunately, methods for DNA shuffling usually cannot recombine sequences that are <70–80% identical. Most homologous sequences, however, show much less sequence similarity, and these sequences are valuable, untapped sources of diversity for directed evolution[21]. With SHIPREC we introduce an approach to overcome this limitation of current methods. SHIPREC can produce combinatorial libraries of hybrid proteins with single crossovers from proteins of low, and even no, sequence identity.

The *cat*-fusion pre-selection is a useful and generally applicable tool for eliminating variants that do not have a continuous reading frame or otherwise lead to nonfolded proteins: 80% of the hybrid P450s from the selected library bound heme and presumably were properly folded. To distinguish between soluble and membrane-bound proteins, careful adjustment of the selection pressure was necessary in order to retain partially soluble variants.

Experimental protocol

Construction of gene dimer. The gene of BM3 was kindly provided by Dr. Ulrich Schwaneberg. The *1A2* gene was kindly provided by Prof. F.P. Guengerich in pCWori$^+$ (pCW1A2), which contains the gene for rat NADPH-oxidoreductase in the same cistron[22]. The genes of BM3 and of 1A2 were amplified by PCR and cloned separately into pBluescript KS(II)+ by using *Xho*I + *Bam*HI to obtain pB-B' (BM3) and *Xba*I + *Sac*I to obtain pB-1' (1A2). During the PCR, two restriction sites on the interior of BM3 (*Nde*I and *Mfe*I) were removed by introducing synonymous mutations using overlap extension PCR. Both constructs were verified by sequencing. The gene of BM3 was then excised from pB-B' by *Xho*I + *Bam*HI and inserted into appropriately restricted pB-1' to yield pB-1B.

Library preparation. A 100 µg aliquot of pB-1B was digested with *Asn*I, *Ssp*I, *Xho*I, and *Sac*I, concentrated by EtOH precipitation, and used for the DNase I digestion. Next, 25 µg DNA in a volume of 300 µl of 25 mM Tris-HCl pH 7.5, 10 mM $MnCl_2$, 0.05 mg/ml BSA, were digested with 1,250 mU DNase I (Sigma, St. Louis, MO; 1 mg/ml correspond to 500 mU/µl) for 15 min at 26°C. The reaction was stopped by addition of 900 µl QX1-buffer (QiaexII kit, Qiagen, Valencia, CA). Using the QiaexII kit the DNA was desalted and the DNase removed. The DNA was then digested with 34 units S1 nuclease (MBI Fermentas, Hanover, MD) in 35 µl S1 buffer for 40 min at 22°C. The reaction was stopped by addition of ethylenediamine tetraacetic acid (EDTA) to 5 mM and incubation at 70° C for 10 min. After separation by agarose gel electrophoresis, fragments of ~1,500 bases were purified using the QiaexII kit. The fragments were circularized by ligation in 27 µl 100 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 25 µg/ml BSA, 3 mM dithiothreitol (DTT), 50 µM ATP, and 1,400 units T4 DNA Ligase (NEB, Beverly, MA) for 16 h at room temperature. The DNA concentration was kept below 3 ng/µl to favor the intramolecular reaction[12]. Five microliters of each ligation mix were digested with 0.3 units *Xba*I in 20 µl total volume to linearize the circular DNA in the linker region between the gene fragments of *1A2* and *BM3*. With a 1:10 dilution this preparation was used as template for a PCR using Vent-polymerase (NEB) and the primer pair 1a2d (5′-CGCTCTAGAGGTACCCCAATTGATG-GAGAAGCGCCGC-3′, anneals to 3′ end of the *1A2* gene) and bm3u (5′-CGACGGATCCGGAAGGAAGGGCCCATAT GACAATTAAAGAAAT-GCCTCAG-3′, anneals to 5′ end of the *BM3* gene) to amplify the libraries. The restriction digestion preceding the PCR, though in theory unnecessary, proved important in obtaining good PCR products.

Library analysis. The libraries were cloned into pBluescript KS(II)+ using *Bam*HI and *Xba*I restriction sites, and XL1-Blue cells were transformed and plated on Xgal/isopropyl-β-D-thiogalactoside (IPTG). White colonies were used as templates for colony PCR with the primer pair bm3u and 1a2d. The length variation of the fragments was analyzed by agarose gel electrophoresis and the gel analysis software package Quantity One (Bio-Rad, Hercules, CA). The same colonies were used as templates for colony PCR with primers that bind inside the gene of either BM3 or 1A2 to estimate the crossover positions.

Construction of *cat*-fusion variants: The *cat* gene was amplified from pACYC184 (NEB) by PCR and inserted into pCW1A2 using *Mfe*I and *Xba*I to produce pCW1A2'cat. This leaves a linker peptide sequence of WPGSPA between the gene of *1A2* and *cat*. The gene of *1A2* was replaced in this construct by the PCR-amplified gene of *BM3* using *Nde*I and *Mfe*I to yield pCWBM3'cat. pCW1A2rf'cat was obtained by linearizing pCW1A2'cat with *Sal*I, overhang fill-in and religation, pCW1A2del'cat by restricting pCW1A2'cat with *Sal*I and *Eco47*III, overhang fill-in and religation. pCWBM3rf'cat was obtained by linearizing pCWBM3'cat with *Mfe*I, overhang fill-in and religation, pCWBM3del'cat by restricting pCWBM3'cat with *Sph*I, overhang removal and religation, and pCWBM3rfdel'cat by restricting pCWBM3'cat with *Sph*I and

religation. The *cat* start codon was removed from these variants using the Quickchange mutagenesis protocol (Stratagene, La Jolla, CA) to change ATG (Met) to AGC (Ser) and introduce a restriction site for *Nhe*I. As a positive control, pCWcatctrl was constructed by restricting pCW1A2'cat with *Nde*I and *Ppu*MI. As a negative control, pCWempty was constructed by removing the gene for CAT as well as for the oxidoreductase from pCWcatctrl using *Bam*HI and *Hin*dIII and religating the vector after creating blunt ends.

Chloramphenicol (cam) resistance analysis. XL1-Blue cells transformed by the different variants were grown on TB medium supplemented with trace elements[22] and 1 mM thiamine (termed TB^+) at 30°C or in LB medium containing 0.01% glucose until an optical density (OD) of ~1 was reached. For liquid growth, cells were diluted to an OD of 0.02 into TB^+ medium supplemented by 0.5 mM δ-aminolevulenic acid[23] and 1 mM IPTG (termed TB^{++}) and 0 to 1 mg/ml cam and grown for 12 h at 30°C, and the OD analyzed. For growth on agar plates, cells were diluted and spread onto TB^{++} agar with 0–200 μg/ml cam. Plates were incubated at 30°C for several days and the appearance of colonies was observed.

Pre-selection and screening: The hybrid library was PCR amplified using the primer pair bm3u and 1a2d and cloned into pCW1A2rfcat (no *cat* start codon) using restriction sites for *Bam*HI and *Mfe*I after 1A2rf had been removed. XL1-Blue cells were transformed by electroporation to obtain 2.5×10^5 colonies on LB agar. Cells were scraped from the plates, resuspended in LB medium, and plated on to TB^{++} agar plates containing 40 μg/ml cam and in parallel on LB Amp. After 36 h at 30°C ~2,000 colonies were picked from the TB plates and used to inoculate 25 μl TB^+ medium in 96-well fluorescence microtiter plates. Cells were grown at 30°C at 270 r.p.m. for 30h, then 100 μl TB^{++} were added to induce expression. After additional growth for 20 h, 125 μl of 5 μM 7-ethoxyresorufin in 25 mM Tris-HCl, pH 7.4, 100 mM KCl and 10 mM $MgCl_2$ were added and the fluorescence emission was measured immediately and after 3 h at 595 nm after excitation at 550 nm. The plates were incubated at 37°C for 2 h and the fluorescence was measured again. From wells that showed the largest increase in fluorescence, cells were plated on LB Amp agar, from which colonies were rescreened using the same protocol.

Expression of hybrid proteins. RC1 and RC2 were subcloned, using *Bam*HI and *Mfe*I to remove the *cat*-fusions from the C termini, by reintroducing the native stop codons. The proteins were expressed in DH5α cells as described[24].

Subcellular fractionation. Membranes and soluble fractions were prepared as described[24]. For detergent treatment, the membrane fractions were analyzed for the total protein content with the BCA assay (Pierce, Rockford, IL) and BSA as standard. Membranes were diluted into 0.5 % sodium cholate and 0, 0.01, 0.05, or 0.2% Triton X-100 so that the concentration of total protein was 2 mg/ml. The samples were incubated at 4°C for 2 h at 50 r.p.m. and centrifuged at >150,000 *g* (Airfuge; Beckman Coulter, Fullerton, CA). The pellet was resuspended in 0.25 M sucrose.

Activity assay. Whole cells were incubated at 37°C with 2.5 μM 7-ethoxyresorufin in 25 mM Tris-HCl, pH 7.4, 100 mM KCl and 10 mM $MgCl_2$ and the fluorescence emission at 595 nm after excitation at 550 nm was measured at different time points. The slope of the initial increase in fluorescence was used as a measure of activity.

CO-binding assay. P450 content was estimated from CO reduced difference spectra, as described[17] using an extinction coefficient of 91 nM^{-1} cm^{-1}.

Western blot analysis. Protein samples were separated by sodium dodecyl sulfate (SDS) gel electrophoresis (Nupage 10% Bis-Tris; Novex, San Diego, CA). Western blotting was performed using polyclonal antibodies against human P450 1A1/1A2 as primary antibodies (Gentest, 1:500 dilution) and anti-goat antibodies coupled to horseradish peroxidase (HRP; Sigma, 1:2000 dilution) as secondary antibodies and detected with the ECL system (Amersham Pharmacia, Piscataway, NJ).

Acknowledgments

V.S. was supported by a research fellowship from the Deutsche Forschungsgemeinschaft (DGF). This work was funded by Maxygen Inc.

Received 21 November 2000; accepted 31 January 2001

1. Stemmer, W.P. DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91**, 10747–10751 (1994).
2. Zhao, H., Giver, L., Shao, Z., Affholter, J.A. & Arnold, F.H. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol.* **16**, 258–261 (1998).
3. Shao, Z., Zhao, H., Giver, L. & Arnold, F.H. Random-priming in vitro recombination: an effective tool for directed evolution. *Nucleic Acids Res.* **26**, 681–683 (1998).
4. Volkov, A.A., Shao, Z. & Arnold, F.H. Recombination and chimeragenesis by in vitro heteroduplex formation and in vivo repair. *Nucleic Acids Res.* **27**, e18 (1999).
5. Cherry, J.R. *et al*. Directed evolution of a fungal peroxidase. *Nat. Biotechnol.* **17**, 379–384 (1999).
6. Ostermeier, M., Shim, J.H. & Benkovic, S.J. A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.* **17**, 1205–1209 (1999).
7. Mitra, B. *et al*. A novel structural basis for membrane association of a protein: construction of a chimeric soluble mutant of (S)-mandelate dehydrogenase from *Pseudomonas putida*. *Biochemistry* **32**, 12959–12967 (1993).
8. Shimoji, M., Yin, H., Higgins, L. & Jones, J.P. Design of a novel P450: a functional bacterial–human cytochrome P450 chimera. *Biochemistry* **37**, 8848–8852 (1998).
9. Bogarad, L.D. & Deem, M.W. A hierarchical approach to protein molecular evolution. *Proc. Natl. Acad. Sci. USA* **96**, 2591–2595 (1999).
10. Quattrochi, L.C., Okino, S.T., Pendurthi, U.R. & Tukey, R.H. Cloning and isolation of human cytochrome P-450 cDNAs homologous to dioxin-inducible rabbit mRNAs encoding P-450-4 and P-450-6. *DNA* **4**, 395–400 (1985).
11. Narhi, L.O., Kim, B.H., Stevenson, P.M. & Fulco, A.J. Partial characterization of a barbiturate-induced cytochrome P-450- dependent fatty acid monooxygenase from *Bacillus megaterium*. *Biochem. Biophys. Res. Commun.* **116**, 851–858 (1983).
12. Sambrook, J., Fritsch, E.F. & Maniatis, T. *Molecular cloning: a laboratory manual*. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; 1989).
13. Graf, R. & Schachman, H.K. Random circular permutation of genes and expressed polypeptide chains: application of the method to the catalytic chains of aspartate transcarbamoylase. *Proc. Natl. Acad. Sci. USA* **93**, 11591–11596 (1996).
14. Baird, G.S., Zacharias, D.A. & Tsien, R.Y. Circular permutation and receptor insertion within green fluorescence proteins. *Proc. Natl. Acad. Sci. USA* **96**, 11241–11246 (1999).
15. Maxwell, K.L., Mittermaier, A.K., Forman-Kay, J.D. & Davidson, A.R. A simple in vivo assay for increased protein solubility. *Protein Sci.* **8**, 1908–1911 (1999).
16. Shaw, W.V. & Leslie, A.G. Chloramphenicol acetyltransferase. *Ann. Rev. Biophys. Chem.* **20**, 363–386 (1991).
17. Schenkman, J.B. & Janson, I. Spectral analysis of cytochromes P450. In *Cytochrome P450 protocols*, Vol. 107. (eds Phillips, I.R. & Shephard, E.A.) 25–33 (Humana Press Inc., Totowa, N.J.; 1998).
18. Chang, T.K.H. & Waxman, D.J. Catalytic assays for human cytochrome P450. In *Cytochrome P450 protocols*, Vol. 107. (eds Phillips, I.R. & Shephard, E.A.) 85–93 (Humana Press Inc., Totowa, N.J; 1998).
19. Dong, M.S., Yamazaki, H., Guo, Z. & Guengerich, F.P. 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 (1996).
20. Bellamine, A., Gautier, J.C., Urban, P. & Pompon, D. 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 (1994).
21. Ostermeier, M. & Benkovic, S.J. Evolution of protein function by domain swapping. *Adv. Protein Chem.* **55**, 29–77 (2000).
22. Parikh, A., Gillam, E.M. & Guengerich, F.P. Drug metabolism by *Escherichia coli* expressing human cytochromes P450. *Nat. Biotechnol.* **15**, 784–788 (1997).
23. Woodard, S.I. & Dailey, H.A. Regulation of heme biosynthesis in *Escherichia coli*. *Arch. Biochem. Biophys.* **316**, 110–115 (1995).
24. Jenkins, C.M., Pikuleva, I. & Waterman, M.R. Expression of eukaryotic cytochromes P450 in *E. coli*. In *Cytochrome P450 protocols*, Vol. 107. (eds Phillips, I.R. & Shephard, E.A.) 181–193 (Humana Press Inc., Totowa, N.J.; 1998).