PEDS Advance Access published February 14, 2012

Protein Engineering, Design & Selection pp. 1–8, 2012
doi:10.1093/protein/gzs004

Comparison of random mutagenesis and semi-rational designed libraries for improved cytochrome P450 BM3-catalyzed hydroxylation of small alkanes

Mike M.Y.Chen[1], Christopher D.Snow[2], Christina L.Vizcarra[1,3], Stephen L.Mayo[1,4] and Frances H.Arnold[1,5]

[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA, [2]Department of Chemical and Biological Engineering, Colorado State University, 1370 Campus Delivery, Fort Collins, CO 80523, USA, [3]Department of Chemistry and Biochemistry, University of California Los Angeles, Los Angeles, CA 90095, USA and [4]Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

[5]To whom correspondence should be addressed.
E-mail: frances@cheme.caltech.edu

Received January 9, 2012; revised January 9, 2012; accepted January 10, 2012

Edited by Alan Berry

Three semi-rational approaches, combinatorial site-saturation mutagenesis (CSSM) using a reduced amino acid set and two libraries based on C^{orbit} and CRAM computational design algorithms targeting up to 10 active site residues, were used to engineer cytochrome P450 BM3 to demethylate dimethyl ether and hydroxylate propane and ethane. These small libraries (343–1028 variants) were all enriched with respect to the fraction functional and maximal activities compared with a random mutagenesis library and individual site-saturation libraries targeting the same residues. Despite high average amino acid substitution levels of 2.6, 5 and 7.5, the CSSM, C^{orbit} and CRAM libraries had at least 75% of library members properly folded. Propane- and ethane-hydroxylating P450 BM3 variants were identified using all three mutagenesis approaches, with as few as two amino acid substitutions. The library designed using the CRAM algorithm, which sought to reduce the size of the binding pocket, produced both a higher number of active variants and variants supporting the greatest number of catalytic turnovers. The most active variant E32 supports 16 800 propane turnovers at 36% coupling, which rivals the activity of variants obtained after 10–12 rounds of directed evolution using random and site-saturation mutagenesis. None of the variants in this study achieved the complete re-specialization for propane hydroxylation (including 93% coupling) previously obtained via multiple rounds of mutagenesis and screening. However, these semi-rational approaches allowed for large jumps in sequence space to variants with the desired functions.

Keywords: alkane hydroxylation/C–H activation/directed evolution/P450 enzymes/semi-rational library design

Introduction

Cytochrome P450s (P450s) are a superfamily of heme-thiolate monooxygenases involved in the biosynthesis of secondary metabolites, degradation of carbon sources and detoxification of xenobiotics (de Montellano, 1986). Using molecular oxygen and electrons from a nicotinamide adenine dinucleotide (NAD(P)H) cofactor, these enzymes catalyze a variety of oxidative transformations on a wide array of substrates that range from small hydrocarbons to complex terpenes (Morant *et al.*, 2003; Isin and Guengerich, 2007; de Montellano, 2010). The breadth of chemical reactivity and substrate range of P450s has generated interest in their use as biocatalysts for industrial synthesis, generation of metabolites for toxicology studies and bioremediation. CYP102A1 (P450 BM3) isolated from *Bacillus megaterium* is one of the most promising P450s for utilization in biotechnology applications as it is soluble, readily over-expressed in industrial hosts and catalytically self-sufficient with its hydroxylase and diflavin reductase domains fused in a single polypeptide chain (Chefson and Auclair, 2006; Urlacher and Eiben, 2006). Over the last decade, protein engineering efforts aimed at expanding P450 BM3's substrate range beyond its preferred substrates of C_{12}–C_{18} fatty acids have produced variants with increased activity for hydroxylation of small alkanes, regioselective hydroxylation of indoles, arenes and polycyclic aromatic hydrocarbons, and selective demethylation of permethylated monosaccharides, among other activities (Ost *et al.*, 2000; Li *et al.*, 2001; Meinhold *et al.*, 2006; Fasan *et al.*, 2007; Lewis *et al.*, 2009).

In a few instances, engineered variants have been used as practical catalysts for preparative scale synthesis (Lewis *et al.*, 2010; Zhang *et al.*, 2011). However, most evolved variants exhibit low total turnover numbers (TTN) and poor coupling between cofactor consumption and product formation, typically 5–30%, which hampers their utility as biocatalysts. Low coupling efficiency reflects a poor fit between substrate and the enzyme active site, which disrupts the usually tightly regulated electron transfer and protonation reactions of the P450 catalytic cycle. Complete re-specialization of P450 BM3 to yield a variant with wild-type-like affinity and coupling for a novel substrate has only been demonstrated for propane (Fasan *et al.*, 2008). The resulting variant, $P450_{PMO}$, acquired 22 mutations from 15 rounds of directed evolution. The differences in molecular size and chemical functionality between propane and the P450 BM3-preferred fatty acid substrates certainly contributed to the extensive mutagenesis and screening effort required to produce $P450_{PMO}$. However, the lack of comparable results for other non-natural substrates suggests that although identification of P450 BM3 variants with modified

substrate specificity at modest activity levels can be achieved readily with a small number of mutations, producing variants that can be used as practical catalysts may require many rounds of labor-intensive mutagenesis and screening.

To reduce screening effort during directed evolution, a number of semi-rational 'library design' approaches such as SCOPE (O'Maille *et al.*, 2002), combinatorial site-saturation mutagenesis (CSSM) (Widersten *et al.*, 2000; Geddie and Matsumura, 2004) and others (Damborsky and Brezovsky, 2009; Lassila, 2010) have been described with the aim of generating smaller and/or functionally enriched libraries. These approaches rely on structural and sequence information to design libraries with mutations targeted to specific regions such as the enzyme active site. Such methods may be particularly useful for the engineering of P450s, as many of the residues involved in substrate recognition do not participate directly in catalysis (Pylypenko and Schlichting, 2004). In this study, we sought to determine the extent of re-specialization of function that could be obtained through semi-rational mutagenesis and compare the functional richness of libraries generated by these approaches to error-prone polymerase chain reaction (EP-PCR) libraries and site-saturation mutagenesis (SSM) libraries. We applied three semi-rational mutagenesis approaches, CSSM and two structure-based computational library design algorithms, C^{orbit} (Treynor *et al.*, 2007) and CRAM, to engineer P450 BM3 for propane hydroxylation, an activity not exhibited by the wild-type enzyme. Efficient P450-catalyzed small alkane hydroxylation could enable gas-to-liquid conversion under mild conditions. Our previous effort to evolve P450 BM3 to accept propane as its preferred substrate relied solely on random mutagenesis with EP-PCR and SSM, in conjunction with high-throughput screening (Fasan *et al.*, 2008). That earlier study showed that a step-wise path of accumulating mutations over multiple rounds of evolution existed for re-specialization of P450 BM3 for activity on dimethyl ether (DME), propane and ethane. Here, we determined that a considerable degree of specialization can be obtained through semi-rational library design, circumventing several generations of step-wise evolution.

Materials and methods

Materials

All chemicals were purchased from Sigma-Aldrich (St Louis, USA). Enzymes for DNA manipulation were obtained from New England Biolabs (Ipswich, USA), Roche (Indianapolis, USA) and Finnzymes (Woburn, USA). $NADP^+$ and NADPH were purchased from Codexis, Inc. (Redwood City, USA) and recombinant isocitrate dehydrogenase was purchased from OYC, Inc. (Azusawa Japan). Alcohol oxidase and horseradish peroxidase were purchased from Sigma-Aldrich. For cloning and expression of P450 BM3 variants, the bacterial strain *Escherichia coli* DH5α was used along with the plasmid pCWori as reported previously (Peters *et al.*, 2003). Oligonucleotides were purchased from Invitrogen (Carlsbad, USA) and Integrated DNA Technology (Coralville, USA). Sequencing reactions were carried out by Laragen (Culver City, USA).

Expression and purification of P450 BM3 variants

Escherichia coli DH5α cells transformed with plasmids containing P450 BM3 variants were cultivated in TB medium supplemented with 100 μg/ml of ampicillin. The cultures were grown at 37°C to an optical density of 1–1.5 at 600 nm and cooled to 25°C before protein expression was induced by the addition of IPTG (1 mM) and δ-aminolevulinic acid hydrochloride (0.5 mM). Following 16–20 h of expression, the cells were harvested by centrifugation and stored at −20°C. An established single-step purification was used for all variants (Schwaneberg *et al.*, 1999). Cell pellets were resuspended in 25 mM Tris–HCl (pH 8.0) and lysed by sonication. Cell debris was removed by centrifugation for 30 min at 20 000 *g*, and the crude cell extraction was loaded on a pre-equilibrated Toyopearl® Super Q-650 M column for ion exchange chromatography. After washing with five column volumes of both 25 mM Tris–HCl (pH 8.0) and 25 mM Tris–HCl with 0.15 M NaCl, the protein was eluted with 25 mM Tris–HCl (pH 8.0) with 0.34 M NaCl. The purified protein solution was buffer exchanged with potassium phosphate buffer (100 mM, pH 8.0) and stored at −80°C. The P450 enzyme concentrations were quantified by CO-binding difference spectra as described in Omura and Sato (1964), using 91/mM/cm as the extinction coefficient.

Cloning and mutagenesis library construction

Mutagenesis libraries were constructed of the heme domain of P450 BM3 (aa1-433) using primers listed in Supplementary Table SI. All library constructs were reinserted into pCWori through the flanking restriction sites *BamHI* and *EcoRI*. A random mutagenesis library of the P450 BM3 heme domain was created by EP-PCR (Cadwell and Joyce, 1994) using 150 μM Mn^{2+} to induce mutations with *Taq* DNA polymerase and primers BamHI-fwd and SacI-rev. An average error rate of 2.1 amino acids/protein was determined by sequencing 10 randomly selected clones. SSM libraries were constructed at selected residues using overlap extension PCR (Kunkel *et al.*, 1987) with complementary sense and antisense primers bearing the NNK triplet. This method was similarly used to construct the targeted mutagenesis libraries designed by the CSSM, C^{orbit} and CRAM algorithms with mutagenic primers containing degenerate codons and mixtures of such primers. For example, a mixture of three primers containing codons: GYN (encodes A and V), WTS (encodes L, I, M and F) and TGG (encodes W) at a ratio of 2 : 4 : 1 was used to achieve the desired set of allowed amino acid substitutions (L, I, F, V, A, M and W) of the CSSM libraries. Likewise, for the construction of CRAM and C^{orbit} libraries, primers with codon degeneracies listed in Supplementary Table SIII were used to generate the desired amino acid substitutions.

Library designs

See supporting information for library design details.

High-throughput screening

Single colonies grown on LB-agar plates supplemented with 100 μg/ml of ampicillin were picked to inoculate 300 μl of LB pre-culture grown in 1-ml 96-well plates. After 16–20 h of incubation at 37°C, 250 rpm and 80% relative humidity, 50 μl of the pre-culture was used to inoculate expression cultures of 650 μl TB grown in 2-ml 96-well plates. After 5 h growth at 37°C, protein expression was induced with the addition of IPTG and δ-aminolevulinic acid hydrochloride at final concentrations of 0.5 and 0.25 mM, respectively. The

cultures were then cooled to 25°C for the duration of protein expression, 20–24 h. The cells were pelleted and stored at −20°C until lysis, which was completed by resuspending the pellets in 350 µl of 0.1 M phosphate buffer (pH 8.0) with 0.5 mg/ml lysozyme, 2 unit/ml DNaseI and 10 mM $MgCl_2$ and incubated at 37°C for 60 min. Crude cell extracts obtained after centrifugation were transferred to microtiter plates to be used for all screening assays.

P450 folding and concentration were determined using a CO-binding difference spectroscopy assay developed for 96-well microtiter plates (Otey and Joern, 2003). Colorimetric quantification of the P450 demethylation of DME was conducted with Purpald® to detect the resulting aldehydes (Peters *et al.*, 2003). In a 96-well microtiter plate, 120 µl of 0.1 M, pH 8.0 potassium phosphate buffer presaturated with DME was added to 30 µl of cell lysate, and the reaction was initiated with the addition of 50 µl of 4 mM NADPH. After 15–30 min, 50 µl of 168 mM Purpald® in 2 M NaOH was added to quench the reaction and produce the purple adduct. The reaction yield was quantified by absorbance at 550 nm after an additional 15 min.

Alkane hydroxylation reactions

P450 ethane and propane hydroxylation reactions were performed in a pressurizable 96-well reactor (Accelrys, USA). Reactions with 0.5 ml total volume typically contained 20–40 µl of filtered (2 µm) crude cell extract or purified protein at concentrations of 50–200 nM, along with an NADPH regeneration system consisting of 20 mM DL-isocitric acid trisodium salt, 400 µM $NADP^+$ and 0.5 U/ml isocitrate dehydrogenase in alkane saturated 0.1 M potassium phosphate buffer. The reactor headspace was pressurized to 30 psi with alkane and the reactions were stirred at least 20 h at 4°C. The reactions were quenched with the addition of 15 µl of 5 M HCl and neutralized with 75 µl of 1 M potassium phosphate buffer, pH 8.0. After the removal of precipitated cell debris and denatured protein by centrifugation, 1-pentanol was added to the reaction mixture as an internal standard and alcohol yield was determined by gas chromatography with a flame ionization detector.

Gas chromatography analysis

Analysis for hydroxylation products of P450 reactions with small alkanes was performed on a Hewlett Packard 5890 Series II Plus gas chromatograph with a flame ionization detector and fitted with an HP-7673 autosampler system. Analysis of ethanol and propanol products was performed on a Supelco SPB-1 column (60 m length, 0.32 mm ID, 0.25 µm film thickness). A typical temperature program used for separating the alcohol products was 250°C injector, 300°C detector, 50°C oven for 3 min, then 10°C/min gradient to 200°C, 25°C/min gradient to 250°C and then 250°C for 3 min.

Results

We constructed mutant libraries of wild-type P450 BM3 heme domain using EP-PCR, SSM, CSSM and two structure-based computational design algorithms, C^{orbit} and CRAM, with library compositions listed in Table I. The residues targeted for mutagenesis, A74, L75, V78, A82, F87, L181, A184, L188, A328 and A330, were selected using the crystal structure of the P450 BM3 heme domain bound with *N*-palmitoyl glycine, PDB:1JPZ (Haines *et al.*, 2001) (see Fig. 1). These 10 residues were included based on their close proximity (<4.5 Å) to the bound substrate and were chosen over adjacent candidates due to favorable side chain orientation that point directly into the enzyme active site. Six of these 10 residues (A74, V78, A82, A184, L188 and A328) were also mutated in $P450_{PMO}$ (Fasan *et al.*, 2008) and thus were known sites of beneficial mutations for small alkane hydroxylation. Site-saturation libraries were constructed at each of these 10 positions using primers with NNK codon degeneracy to determine the range of activities that could be obtained by single amino acid substitutions.

Table I. Targeted residues and allowed amino acids for the semi-rational libraries of P450 BM3

Residue	CSSM	CRAM	C^{orbit}
A74	–	LW	AV
L75	–	LF	LF
V78	LIVMFAW	IF	VL
A82	LIVMFAW	VL	AS
F87	–	FA	FA
L181	–	LW	LF
A184	–	AV	AT
L188	–	LW	LW
A328	LIVMFAW	VF	AF
A330	–	LW	AV

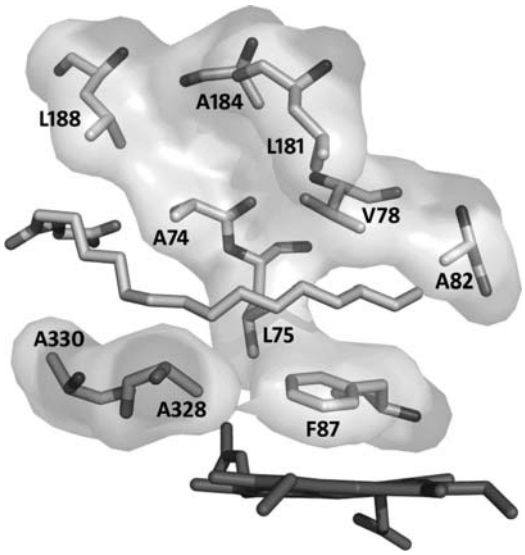

Fig. 1. The 10 residues targeted for SSM and structure-based computational library design were chosen based on their proximity to the bound *N*-palmitoyl glycine substrate in the 1JPZ structure of P450 BM3 (Haines *et al.*, 2001). Heme is shown in dark gray, *N*-palmitoyl glycine is shown in light gray.

Three different approaches were used to generate mutant libraries that explore combinations of mutations at these 10 positions. The CSSM library focused mutations to three residues, V78, A82 and A328, previously found to be particularly important for shifting P450 BM3's substrate specificity toward smaller substrates (Peters *et al.*, 2003). Mutating all three positions with NNK codon degeneracy would yield a library of $32^3 = 32\,768$ members, which is too large to screen. To reduce the library to a more manageable size

Table II. Protein folding and DME demethylation activity of P450 BM3 mutant libraries

Library properties	EP-PCR	Site saturation libraries				CSSM	CRAM	C^{orbit}
		Average	V78	A82	A328			
$<M_{AA}>$	2.1	0.9	0.9	0.9	0.9	2.6	7.5	5
Library size (number of unique sequences)	1166[a]	20	20	20	20	343	1024	1024
Fraction of folded variants[b]	0.52	0.89	0.90	0.96	0.62	0.94	0.75	0.84
Fraction of active variants[c]	0.36	0.10	0.26	0.21	0.36	0.54	0.34	0.31
Average DME activity	0.06	0.08	0.05	0.09	0.11	0.22	0.33	0.19
Maximum DME activity	0.52	0.34	0.17	0.21	0.34	0.81	2.6	1.9

$<M_{AA}>$, average mutation level.
[a]Estimate of the number of unique sequences sampled from 1408 clones of the EP-PCR library made using PEDEL (Patrick *et al.*, 2003).
[b]Fraction of folded variants determined by CO-binding spectroscopy, corrected for stop codon presence.
[c]Fraction of variants active on DME was determined in cell-free extract, corrected for background Purpald® oxidation.

($7^3 = 343$ members), the allowed amino acids were restricted to L, I, M, V, F, A and W, following the intuition that introducing amino acids with large hydrophobic side chains should reduce active site volume and promote binding of small hydrophobic substrates. To generate libraries with even higher levels of mutation, two structure-based computational design algorithms, C^{orbit} (Hayes *et al.*, 2002) and CRAM, were used to introduce mutations at all 10 targeted residues allowing for two possible amino acids at each position (see Table I). The C^{orbit} algorithm, which was used previously to alter the green fluorescent protein fluorescence spectrum (Treynor *et al.*, 2007), models protein stability as a surrogate for function. Mutations were selected based on the frequency of their appearance in sequences predicted to maintain the P450 BM3 substrate-bound conformation in the absence of substrate. The second approach, which we called CRAM, was developed with the principal objective of reducing the P450 BM3 active site volume by incorporating the largest tolerated amino acid. The amino acid tolerance at each target position was evaluated as a single amino acid substitution based on the P450 BM3 substrate-bound conformation (1JPZ) with the substrate removed. The potential for van der Waal clashes introduced by each amino acid substitution was scored using the repulsive van der Waal energy determined by the Rosetta design software after side-chain optimization (Rohl *et al.*, 2004). The mutations were selected to maximize the side chain size of the amino acid substitution while minimizing the potential for introducing steric clashes with the protein backbone and the side chains adjacent to the 10 designed residues. As the CRAM algorithm evaluates mutations based on single amino acid substitutions, it does not take into account potential van der Waals clashes between designed mutations. However, we hypothesized that relaxation of the protein backbone would relieve the energetic strain of side chain/side chain clashes in the substrate-binding pocket.

Library characterization for DME demethylation and protein folding

Each mutant library was characterized for DME demethylation and protein folding using high-throughput assays; the results are summarized in Table II. DME demethylation activity was quantified colorimetrically using Purpald® dye, which reacts with the formaldehyde product of the P450 reaction to form a purple adduct with a UV/Vis peak maximum at 550 nm (Peters *et al.*, 2003). The folded state of each variant was determined from the CO-binding difference spectra of the cell-free extract. All libraries were screened with at least 90% library coverage, with the exception of the EP-PCR library. The screening of 1408 variants of the EP-PCR library with an error rate of 2.1 amino acid substitutions/protein, corresponds to sampling 1166 unique sequences (Programme for estimating diversity in error-prone PCR libraries (PEDEL); Patrick *et al.*, 2003). A surprisingly large fraction of the EP-PCR library variants, 36%, was found to be active on DME, and 52% were folded. In comparison, only 8% of variants generated by saturation mutagenesis at the 10 individual active site residues were observed to demethylate DME, even though 89% of these variants were folded. Variants active on DME were found only in SSM libraries with mutations at V78, A82, A328 and A330. The high mutational tolerance of these active site positions is also reflected in the high fraction of folded variants of the CSSM, C^{orbit} and CRAM libraries, which despite having average amino acid substitution rates of 2.6, 5 and 7.5, contained 91, 84 and 75% folded variants, respectively. The CSSM library had the largest fraction of variants active for DME demethylation, 54%, whereas both computationally designed libraries had 31–34% active variants.

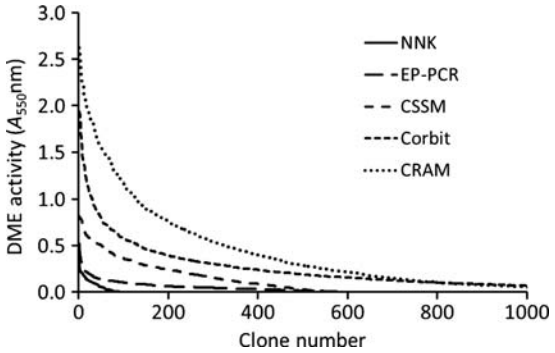

Fig. 2. Profile of DME demethylation activity for the various libraries. Activities of individual variants are plotted in descending order. All variants from the 10 SSM libraries are shown, as are the 1000 most active variants of the EP-PCR, CSSM, C^{orbit} and CRAM libraries.

Although all the mutagenesis strategies generated variants with DME demethylation activity, the activity levels varied (see Fig. 2). The functional richness of the EP-PCR library and SSM libraries are similar, with average DME activities of 0.06 and 0.08 (A_{550nm}) and maximum DME activities of 0.54 and 0.34, respectively. The averages reflect the overall

low activity of variants in these libraries. In contrast, the variants of the CSSM libraries exhibit DME activities up to 0.81, with an average activity of 0.22, more than twice those of the EP-PCR and single-site saturation libraries. The C^{orbit} and CRAM libraries contain variants with even higher DME activity, with maximum activities of 1.9 and 2.6 A_{550nm}. However, as the majority of variants in these two libraries has only mediocre activity, the library averages of 0.19 and 0.33 are similar to the CSSM library average.

Propane and ethane hydroxylation

Top-performing variants were purified and characterized for propane and ethane hydroxylation activity (see Table III for a representative sample and supporting information for complete sequence and activity information). From the 10 SSM libraries, 12 variants were identified that supported propane TTN from 120 to 2200. Mutations at V78 (T, C, S) and A82 (E, Q) located in the B' helix yielded variants with moderate propane activity, 120–370 TTN. More active variants, with >1000 propane TTN, were obtained with mutations at residues A328 (I, P, L, V) and A330 (L, P, V), which are located in the loop between the J and K helices. The best single active-site variant, A328V, supports 2200 propane TTN with a product formation rate of 7.1/ min and 8.1% coupling of cofactor consumption to product formation. Variants from the EP-PCR library supported 130–3300 propane TTN. The best one, WT-F162L, supports 3300 propane TTN with a product formation rate of 19/min and 15% coupling. The F162L mutation is in the linker between the E and F helices, outside the active site. Although this residue was not mutated in $P450_{PMO}$, several positions in the adjacent F-helix were mutated, which indicates that this region is important for substrate specificity.

Nine variants supporting 380–4200 propane TTN were identified from the CSSM library. Four of the nine contained mutations at all three targeted residues, the remaining variants had mutations at two residues. V78L was the only mutation found at position 78, whereas more substitutions were beneficial at A82 (L, W, M and V) and A328 (L, V and F). The best variant, WT-A82L-A328V, supports 4200 propane TTN with a product formation rate of 40/min and 44% coupling. In addition, two of the nine propane-active variants, WT-A82L-A328L and WT-A82L-A328V, also hydroxylated ethane, with 140 and 200 TTN, respectively. The CRAM and C^{orbit} libraries, which contained variants with much higher DME demethylation activity in the initial screen, produced 37 variants supporting at least 3500 propane TTN; 16 of these hydroxylated ethane with TTN of at least 300. A significantly higher number of active variants for propane and ethane hydroxylation was found in the CRAM library, 25 and 13, respectively, compared with the C^{orbit} library, which produced only 12 propane-hydroxylating variants and three ethane-hydroxylating variants. The most active CRAM variant WT-A74W-V78I-A82L-A184V-L188W-A328F-A330W (E32) supported 16 800 propane TTN and 1200 ethane TTN. The most active C^{orbit} variant WT-A74V-L181F-A328F (OD2) supported 11 600 propane TTN and 660 ethane TTN. The coupling of co-factor consumption to propanol formation for most of these variants ranged from 36 to 52%, with the best variant, E31, exhibiting 68% coupling.

Table III. *In vitro* small alkane hydroxylation activity of representative P450 BM3 variants

Variants	Library	Mutations	Propane TTN[a]	Coupling[b] (%)	Ethane TTN[a]
A328V	SSM	A328V	2300	8.0	0
4F9	EP-PCR	F162L	3300	18	0
4E10	CSSM	A82L, A328V	4200	44	200
OD2	C^{orbit}	A74V, L181F, A328F	11 600	42	660
E32	CRAM	A74W, V78I, A82L, A184V, L188W, A328F, A330W	16 800	36	1200

[a]TTN determined as nanomole product/nanomole enzyme. Alkane reactions contained 25–250 nM protein, 100 mM potassium phosphate buffer, pH 8.0, saturated with alkane and an NADPH regeneration system containing 100 μM $NADP^+$, 2 U/ml isocitrate dehydrogenase and 10 mM isocitrate. Standard errors are within 15% of average.
[b]Coupling is the ratio of product formation rate to NADPH consumption rate (%).

The 25 propane-hydroxylating CRAM library variants form a concise data set for sequence analysis. The distribution of amino acids for all 25 variants, shown in Fig. 3a, displays strong biases at 7 of the 10 targeted positions. Tryptophan appears at residue 74 and 188 in >72% of the sequences. Likewise, strong preferences exist for L at positions 75 (84%), 82 (76%) and 181 (88%), I at 78 (76%) and F at 87 (96%). Of the remaining residues, a weaker preference exists for W at 330 (68%) and V at 184 (60%), and nearly equal representations of F and V were observed at position 328. In most instances, the representation of the less preferred amino acid decreases with increasing propane hydroxylation activity (see Fig. 3c and e). For variants supporting <7500 propane TTN, a higher fraction of the less-preferred amino acids is found at positions 74, 75, 78, 82 and 188. Proceeding to variants supporting higher propane TTN, the less-preferred amino acids decrease, culminating in nearly absolute preference for W at position 74, L at positions 75, 82 and 181, I at position 78 and F at position 87 in variants supporting >10 000 propane TTNs. These results indicate that the screening process was able to find a narrow section of the total allowed sequence space containing the best solutions for propane hydroxylation.

The amino acid consensus for ethane-hydroxylating variants of the CRAM library shares many of the same preferences (see Fig. 3b). The biggest difference occurs at residue 188, where the ethane-hydroxylating variants strongly prefer tryptophan (92%) and the propane-hydroxylating variants have a weaker preference for tryptophan (72%). Interestingly, the sequence consensus of the ethane-hydroxylating variants most strongly resembles that of propane-hydroxylating variants with intermediate activity, those supporting 7500–10 000 propane TTN, rather than the most active propane-hydroxylating variants. This implies that even for similar substrates such as propane and ethane, the optimal active site configurations occupy distinct and non-overlapping regions of sequence space. Therefore, regions of sequence space with the best correlation of activities are occupied by generalist enzymes with intermediate activity for both substrates.

Discussion

Variants with DME demethylation and propane hydroxylation activity were identified through all the mutagenesis

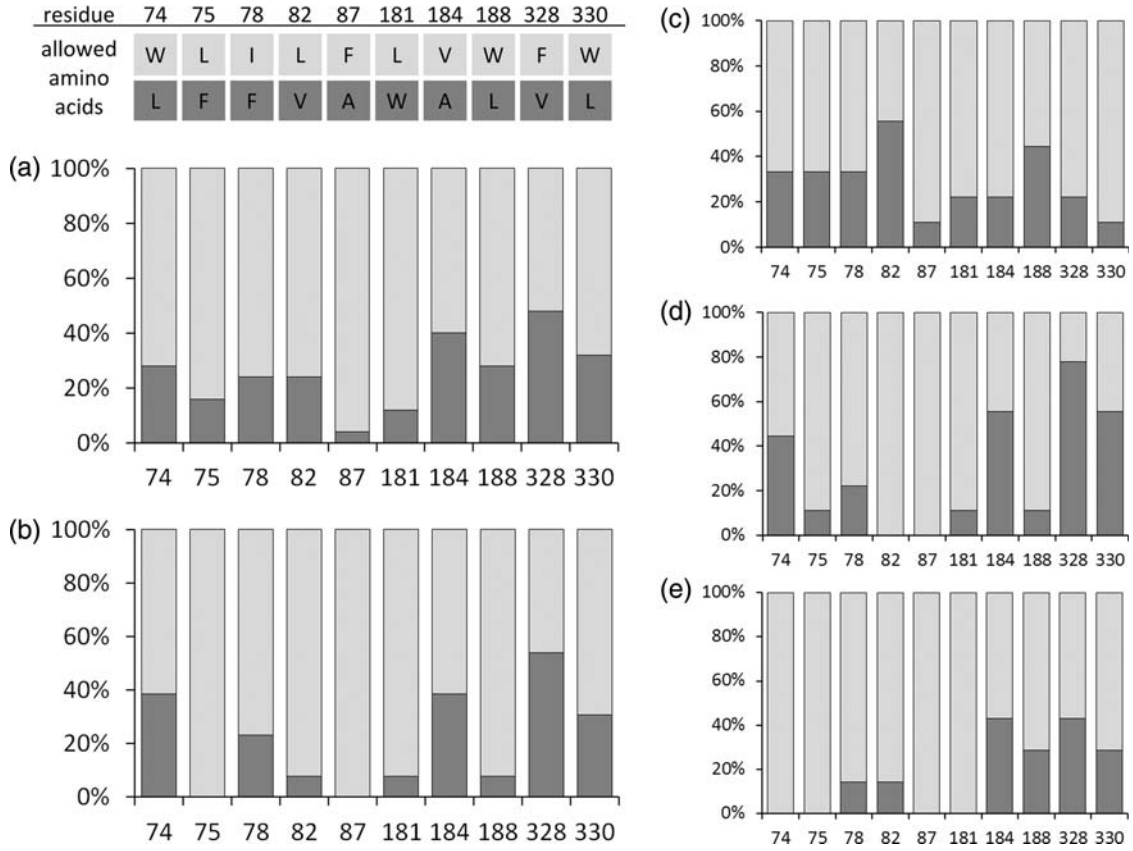

Fig. 3. Amino acid distributions at the 10 targeted positions of CRAM library variants that support (**a**) propane hydroxylation, (**b**) ethane hydroxylation, (**c**) <7500 propane TTN, (**d**) between 7500 and 10 000 propane TTN and (**e**) >10 000 propane TTN.

approaches investigated, with as few as one mutation. The surprising ease with which these two functions are acquired contrasts with the results of our early directed evolution study (Glieder *et al.*, 2002). One major difference between this study and the previous effort is this study's use of DME demethylation as a surrogate for activity on propane directly on libraries generated from wild-type P450 BM3. In the previous study, P450 BM3 mutant libraries were initially screened for activity on *para*-nitrophenoxy octyl ether, an octane surrogate, in a circuitous route to obtaining propane-hydroxylating variants. The large differences in effort and number of accumulated mutations between these two studies is another example for the importance of the screen in directed evolution, reaffirming the first rule of directed evolution: 'You get what you screen for' (You and Arnold, 1996).

As a substrate, propane shares many similarities with hexane, the smallest alkane known to be hydroxylated by wild-type P450 BM3. They are both hydrophobic with poor water solubility and possess sub-terminal alkane C–H bonds of bond strength comparable with P450 BM3's preferred fatty acid substrates (99–100 kcal/mol). The only difference between propane and known P450 BM3 substrates is its smaller molecular size, which would result in a lower binding affinity. Poorly bound substrates cannot displace the distal water ligand to initiate catalysis (Denisov *et al.*, 2005), which results in weak activation of the P450 catalytic cycle. For wild-type P450 BM3, propane binding would be the sole trigger for catalysis, as the enzyme exhibits very low (2/min) resting state oxidase activity (Kadkhodayan *et al.*, 1995), indicating that substrate-independent activation of the catalytic cycle occurs rarely. However, P450 BM3 variants can exhibit much higher resting state oxidase activity, thereby reducing the requirement of propane binding to induce catalysis. In fact, variants generated in the $P450_{PMO}$ lineage and many of the variants found in this study have substrate-free cofactor consumption rates that are orders of magnitude higher than the wild-type enzyme. The existence of this alternative pathway for propane hydroxylation, which can be achieved without appreciable propane-induced activation of the catalytic cycle (Fasan *et al.*, 2008), could account for the large number of propane-hydroxylating variants that were identified.

These targeted mutagenesis libraries also revealed the high mutational tolerance of the P450 BM3 active site, which goes against the general observation that mutations in the core of a protein are on average more destabilizing than mutations of solvent-exposed residues (Reidhaarolson and Sauer, 1988). Residues in the P450 BM3 active site may have an atypical degree of flexibility compared with typical core residues, as the active site of P450 BM3 undergoes significant motion between the 'closed' substrate-bound state and its 'open' resting state (Arnold and Ornstein, 1997). Local backbone flexibility could allow mutations with greater probability than highly constrained position buried within a rigid protein. The active site environment between these two states also differs significantly in terms of solvent accessibility (Schlichting *et al.*, 2000), which may allow these residues to tolerate polar or even charged amino acid substitutions. Finally, many of the P450 BM3 active site residues that participate in substrate recognition are not directly

involved in enzyme catalysis, which also contributes to the tolerance of these residues for substitution. The malleable nature of the active site in combination with the capacity to decouple substrate binding from catalysis could account for the ease with which P450 BM3 variants are able to acquire activity for non-native substrates.

In comparing the functional richness between the libraries generated by the various mutagenesis methods investigated, it was surprising to find that the EP-PCR library appears to generate more variants that catalyze DME demethylation than the combined efforts of the SSM libraries (see Fig. 2). The functional richness of a random mutagenesis library has been generally characterized as poor, as the majority of variants with point mutations are expected to exhibit a parent-like phenotype. Although the number of clones sampled from these libraries is similar (1408 for EP-PCR vs. 910 for SSM), the actual genotypic diversity is quite different: 200 unique variants for the SSM libraries vs. 1166 expected unique variants for the EP-PCR library. This six-fold difference in the number of sequences sampled could account for the higher number of DME-demethylating variants identified in the EP-PCR library. Of the 10 positions that were subjected to SSM, only four of the positions, V78, A82, A328 and A330, yielded variants that were active on DME. A different selection of SSM targets could yield more active variants, but the opposite result is also possible. The ease with which EP-PCR can generate sequence diversity differentiates these two approaches. For activities that can be altered by mutations dispersed across the protein sequence, EP-PCR will be advantageous compared with SSM.

The CSSM, C^{orbit} and CRAM libraries with higher average mutation levels generated variants with much higher levels of DME demethylation activity compared with variants from the SSM and EP-PCR libraries. This is unsurprising in that a variant with more beneficial mutations is likely to have greater activity. However, because most mutations are neutral or deleterious, the fraction of total sequence space with functional sequences decreases precipitously as a function of mutational distance. As a consequence, random mutagenesis at a high mutation rate generally results in libraries with lower functional richness due to the exponential decrease of protein folding/function with mutation (Drummond *et al.*, 2005). Designed libraries avoid this problem by targeting mutations to residues that are highly tolerant to mutation and by favoring amino acid substitutions that are compatible with the target substrates. Restricting amino acid substitutions to hydrophobic residues with larger side chains enriched these libraries for functional variants compared with the SSM libraries that replaced the targeted residues with all 19 amino acids. Although the CSSM library generated the highest fraction of variants active on DME, the maximum activities lagged behind CRAM and C^{orbit} libraries. This is likely the result of the limited volume reduction that was accessible to the CSSM library, which only allowed mutations at three residues compared with the 10 that could be mutated in the CRAM and C^{orbit} libraries. The C^{orbit} design was able to generate a library with a higher fraction of folded variants (84%) compared with the CRAM library (75%). However, it is difficult to assess the relative success of the C^{orbit} design at preserving the folded structure as the mutation level of the C^{orbit} library, 5 mutations/protein, is 2.5 mutations/protein lower than that of the CRAM library. Although the C^{orbit} library contained more folded variants than the CRAM library, this advantage was not reflected in the fraction functional for DME demethylation, which was similar (31 vs. 34%). The high neutrality of the targeted residues devalues the need to preserve folding, which was the core criterion of the C^{orbit} design.

The single mutations at V78, A82, A328 and A330 that resulted in variants with DME demethylation activity generally introduced amino acids with bulkier side chains into the active site, which follows the intuition that reducing active site volume would promote activity on a smaller substrate. Because none of the amino acids found in $P450_{PMO}$ (F78, G82 and F328), were found at these positions, one can conclude that they are not beneficial individually. The presence of proline mutations at A328 and A330 suggests that altering the orientation of the loop containing these residues can generate DME activity. Of the seven mutations found using random mutagenesis, only I260V, located on the I-helix, is near the active site. Mutations F162L and I153V found in the more active variants are in the region between the E and F helices. The remaining mutations, E4D, T235M, D232V and Q359R, are at surface-exposed residues, which is typical of mutations found using random mutagenesis (Arnold, 1998), whose effects are difficult to rationalize. Of the mutations identified from the CSSM variants, V78L, A82 (L, W, M and V) and A328 (L, V and F), only A328F was previously found with $P450_{PMO}$. As this mutation was not found to be beneficial individually, its presence in $P450_{PMO}$ and these reduced CSSM variants suggests there are synergistic effects between this residue and neighboring amino acids. The effect of the V78L mutation on propane TTN also showed a dependence on the surrounding mutations: when V78L was introduced to WT-A328L or WT-A82L-A328L, the resulting variants lost 48% of the parental propane activity (see Supplementary Table SII). However, when V78L was mutated in WT-A82W-A328F, the resulting variant had 6-fold improved propane activity. This illustrates the ruggedness of active site fitness landscape, in which the effects of mutations are highly dependent on the identity of neighboring amino acids.

The most effective libraries for DME demethylation and propane hydroxylation were the CRAM and C^{orbit} structure-based computationally designed libraries. These two libraries globally redesigned the P450 BM3 active site by mutating 10 residues to two possible amino acids. From these libraries, we obtained variants with propane and ethane activity rivaling those achieved by variants of the $P450_{PMO}$ lineage. The best variant from the CRAM library (E32) supported 16 800 propane TTN and 1200 ethane TTN, which represents nearly 50% of $P450_{PMO}$'s activity on these substrates. The best variant from the C^{orbit} library (OD2) supported 11 600 propane TTN and 660 ethane TTN, or 34 and 27% of $P450_{PMO}$'s activity on these substrates. As a comparison, the $P450_{PMO}$ activity levels were obtained after 10–12 rounds of mutagenesis and screening. Of the 37 variants isolated from these two designed libraries, all allowed amino acids were found at least once. The consensus sequence of propane-hydroxylating CRAM variants, WT-L74W-V78I-A82L-A184V-L188W-A328F-A330W, is actually the sequence of the most active variant E32, which suggests that the screening process was able to find an optimal solution within the allowed sequence space (2^{10}). However, assigning

significance to these amino acid preferences in the context of the total possible sequence space of these 10 residues (10^{20}) is problematic, as each mutation was compared against only one other amino acid within a limited set of surrounding mutations. Of the six positions where the $P450_{PMO}$ residue was an allowed choice by the library design (75, 78, 87, 181, 184 and 328), five positions converged on that amino acid. This convergence on $P450_{PMO}$ mutations is not surprising as the $P450_{PMO}$ active site is a sequence solution for the screened activities, and it may suggest that the $P450_{PMO}$ active site sequence is a very accessible solution from wild-type P450 BM3.

These libraries demonstrate that it is possible to jump from P450 BM3 to variants with moderate propane hydroxylation activity (~10 000 TTN). None of the variants, however, reached the level of specialization that was previously obtained with $P450_{PMO}$, in either propane TTN or coupling of cofactor consumption. In the evolution of P450 BM3 to $P450_{PMO}$, the specialization for propane hydroxylation did not occur evenly through the 16 rounds of mutagenesis. In fact, the variants of the lineage can be categorized into three groups by their substrate specificity for linear alkanes as (i) preferring longer chain alkanes, (ii) having equal preference for chain lengths C_3–C_{10} and (iii) preferring shorter chains the length of propane (Fasan *et al.*, 2008). These three groups of variants represent a transition from a specialized fatty acid hydroxylase to generalist P450s with broad alkane substrate acceptance followed by a second transition to a specialized propane monooxygenase. This last transition occurs in the final four rounds of mutagenesis where the largest improvements in cofactor coupling (44–93%) and propane TTN (10 550–33 400) occurred. In addition, mutations acquired in these final rounds of mutagenesis appeared in the reductase domain as well as in the heme domain. This suggests that mutations outside the active site or even the heme domain may be necessary for functional optimization. The ranges of propane TTN (3500–16 800) and coupling of cofactor consumption (36–68%) for variants in the designed libraries are those of the generalist intermediates found preceding the propane specialization phase of the $P450_{PMO}$ evolution. This suggests that semi-rational library design can be an effective strategy to move toward generalist enzymes, but that functional specialization still requires optimization through several rounds of random mutagenesis and screening.

Supplementary data

Supplementary data are available at *PEDS* online.

Acknowledgements

This work was supported by the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Science, US Department of Energy Grant DE-FG02-06ER15762 and the DARPA Protein Design Processes Grant. C.D.S. was supported by a Jane Coffin Childs postdoctoral fellowship. The authors thank Dr Nathan Dalleska for assistance with gas chromatography.

References

Arnold,F.H. (1998) *Acc. Chem. Res.*, **31**, 125–131.
Arnold,G.E. and Ornstein,R.L. (1997) *Biophys. J.*, **73**, 1147–1159.
Cadwell,R.C. and Joyce,G.F. (1994) *PCR-Methods Appl.*, **3**, S136–S140.
Chefson,A. and Auclair,K. (2006) *Mol. Biosyst.*, **2**, 462–469.
Damborsky,J. and Brezovsky,J. (2009) *Curr. Opin. Chem. Biol.*, **13**, 26–34.
de Montellano,P.R.O. (1986) *Cytochrome P450*, 1st edn, Plenum Publishing Corp., New York.
de Montellano,P.R.O. (2010) *Chem. Rev.*, **110**, 932–948.
Denisov,I.G., Makris,T.M., Sligar,S.G. and Schlichting,I. (2005) *Chem. Rev.*, **105**, 2253–2277.
Drummond,D.A., Iverson,B.L., Georgiou,G. and Arnold,F.H. (2005) *J. Mol. Biol.*, **350**, 806–816.
Fasan,R., Chen,M.M., Crook,N.C. and Arnold,F.H. (2007) *Angew. Chem. Int. Ed.*, **46**, 8414–8418.
Fasan,R., Meharenna,Y.T., Snow,C.D., Poulos,T.L. and Arnold,F.H. (2008) *J. Mol. Biol.*, **383**, 1069–1080.
Geddie,M. and Matsumura,I. (2004) *J. Mol. Biol.*, **279**, 26462–26468.
Glieder,A., Farinas,E.T. and Arnold,F.H. (2002) *Nat. Biotechnol.*, **20**, 1135–1139.
Haines,D.C., Tomchick,D.R., Machius,M. and Peterson,J.A. (2001) *Biochemistry*, **40**, 13456–13465.
Hayes,R.J., Bentzien,J., Ary,M.L., Hwang,M.Y., Jacinto,J.M., Vielmetter,J., Kundu,A. and Dahiyat,B.I. (2002) *Proc. Natl Acad. Sci. USA*, **99**, 15926–15931.
Isin,E.M. and Guengerich,F.P. (2007) *Biochim. Biophys. Acta-Gen Subj.*, **1770**, 314–329.
Kadkhodayan,S., Coulter,E.D., Maryniak,D.M., Bryson,T.A. and Dawson,J.H. (1995) *J. Biol. Chem.*, **270**, 28042–28048.
Kunkel,T.A., Roberts,J.D. and Zakour,R.A. (1987) *Methods Enzymol.*, **154**, 367–382.
Lassila,J.K. (2010) *Curr. Opin. Chem. Biol.*, **14**, 676–682.
Lewis,J.C., Bastian,S., Bennett,C.S., Fu,Y., Mitsuda,Y., Chen,M.M., Greenberg,W.A., Wong,C.H. and Arnold,F.H. (2009) *Proc. Natl Acad. Sci. USA*, **106**, 16550–16555.
Lewis,J.C., Mantovani,S.M., Fu,Y., Snow,C.D., Komor,R.S., Wong,C.H. and Arnold,F.H. (2010) *Chembiochem*, **11**, 2502–2505.
Li,Q.S., Ogawa,J., Schmid,R.D. and Shimizu,S. (2001) *Appl. Environ. Microbiol.*, **67**, 5735–5739.
Meinhold,P., Peters,M.W., Hartwick,A., Hernandez,A.R. and Arnold,F.H. (2006) *Adv. Synth. Catal.*, **348**, 763–772.
Morant,M., Bak,S., Moller,B.L. and Werck-Reichhart,D. (2003) *Curr. Opin. Biotechnol.*, **14**, 151–162.
O'Maille,P.E., Bakhtina,M. and Tsai,M.D. (2002) *J. Mol. Biol.*, **321**, 677–691.
Omura,T. and Sato,R. (1964) *J. Biol. Chem.*, **239**, 2370–2374.
Ost,T.W.B., Miles,C.S., Murdoch,J., Cheung,Y.F., Reid,G.A., Chapman,S.K. and Munro,A.W. (2000) *FEBS Lett.*, **486**, 173–177.
Otey,C. and Joern,J.M. (2003) In Arnold,F.H. and Georgiou,G. (eds), *Methods in Molecular Biology*. Humana Press, Totowa, pp. 141–148.
Patrick,W.M., Firth,A.E. and Blackburn,J.M. (2003) *Protein Eng.*, **16**, 451–457.
Peters,M.W., Meinhold,P., Glieder,A. and Arnold,F.H. (2003) *J. Am. Chem. Soc.*, **125**, 13442–13450.
Pylypenko,O. and Schlichting,I. (2004) *Annu. Rev. Biochem.*, **73**, 991–1018.
Reidhaarolson,J.F. and Sauer,R.T. (1988) *Science*, **241**, 53–57.
Rohl,C.A., Strauss,C.E.M., Misura,K.M.S. and Baker,D. (2004) *Numerical Computer Methods, Partt D*, Academic Press Inc, San Diego, p. 66.
Schlichting,I., Berendzen,J., Chu,K., *et al.* (2000) *Science*, **287**, 1615–1622.
Schwaneberg,U., Sprauer,A., Schmidt-Dannert,C. and Schmid,R.D. (1999) *J. Chromatogr. A*, **848**, 149–159.
Treynor,T.P., Vizcarra,C.L., Nedelcu,D. and Mayo,S.L. (2007) *Proc. Natl Acad. Sci. USA*, **104**, 48–53.
Urlacher,V.B. and Eiben,S. (2006) *Trends Biotechnol.*, **24**, 324–330.
Widersten,M., Hansson,L.O., Tronstad,L. and Mannervik,B. (2000) *Methods Enzymol.*, **328**, 389–404.
You,L. and Arnold,F.H. (1996) *Protein Eng.*, **9**, 77–83.
Zhang,K.D., El Damaty,S. and Fasan,R. (2011) *J. Am. Chem. Soc.*, **133**, 3242–3245.