doi:10.1016/j.jmb.2008.06.060 J. Mol. Biol. (2008) 383, 1069–1080

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Evolutionary History of a Specialized P450 Propane Monooxygenase

Rudi Fasan[1], Yergalem T. Meharenna[2], Christopher D. Snow[1], Thomas L. Poulos[2] and Frances H. Arnold[1]*

[1]*Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 E. California Blvd., MC 210-41, Pasadena, CA 91125, USA*

[2]*Department of Molecular Biology & Biochemistry, Chemistry, and Pharmaceutical Sciences, University of California, Irvine, Irvine, CA 92697, USA*

Received 22 April 2008; received in revised form 13 June 2008; accepted 19 June 2008
Available online 28 June 2008

Edited by F. Schmid

The evolutionary pressures that shaped the specificity and catalytic efficiency of enzymes can only be speculated. While directed evolution experiments show that new functions can be acquired under positive selection with few mutations, the role of negative selection in eliminating undesired activities and achieving high specificity remains unclear. Here we examine intermediates along the 'lineage' from a naturally occurring C_{12}–C_{20} fatty acid hydroxylase ($P450_{BM3}$) to a laboratory-evolved P450 propane monooxygenase ($P450_{PMO}$) having 20 heme domain substitutions compared to $P450_{BM3}$. Biochemical, crystallographic, and computational analyses show that a minimal perturbation of the $P450_{BM3}$ fold and substrate-binding pocket accompanies a significant broadening of enzyme substrate range and the emergence of propane activity. In contrast, refinement of the enzyme catalytic efficiency for propane oxidation (~9000-fold increase in k_{cat}/K_m) involves profound reshaping and partitioning of the substrate access pathway. Remodeling of the substrate-recognition mechanisms ultimately results in remarkable narrowing of the substrate profile around propane and enables the acquisition of a basal iodomethane dehalogenase activity as yet unknown in natural alkane monooxygenases. A highly destabilizing L188P substitution in a region of the enzyme that undergoes a large conformational change during catalysis plays an important role in adaptation to the gaseous alkane. This work demonstrates that positive selection alone is sufficient to completely respecialize the cytochrome P450 for function on a nonnative substrate.

Published by Elsevier Ltd.

Keywords: alkane oxidation; protein evolution; directed evolution; propane monooxygenase; cytochrome P450 BM3

Introduction

Organisms can adapt to changing environments by acquiring alternative biosynthetic pathways or detoxification mechanisms upon discovery and subsequent optimization of new enzyme catalysts. Despite fundamental contributions from genetic and structural analyses,[1,2] our understanding of the forces driving the diversification and specialization of enzymes to occupy new functional niches remains limited.

Directed evolution experiments have demonstrated that a small number of mutations can alter enzyme specificity and enhance activity toward poor substrates.[3–8] Catalytic promiscuity and the ease with which these side activities can be improved provide a mechanism for enzyme diversification in nature.[9,10] Although there are cases in which having a broad substrate range is linked to the physiological role of the enzyme (e.g., the ability of certain P450s to metabolize a wide array of xenobiotics),[11–14] many enzymes show exceptional power to discriminate among similar substrates. How this specialization occurs is largely unknown. In the laboratory, newly specialized enzyme variants have been obtained by applying positive and

**Corresponding author.* E-mail address: frances@cheme.caltech.edu.

Abbreviations used: PMO, propane monooxygenase; MMO, methane monooxygenase; TTN, total turnover number; NPG, *N*-palmitoylglycine; sMMO, soluble non-heme diiron methane monooxygenase; GC–ECD, gas chromatography–electron capture detector; FID, flame ionization detector; CRS, cofactor regeneration system.

0022-2836/$ - see front matter. Published by Elsevier Ltd.

Table 1. Summary of directed evolution steps and amino acid substitutions accumulated in $P450_{BM3}$ variants

Variant	Target substrate	Mutagenesis method	Heme domain substitutions[a]
139-3[b]	Octane	Random/recombination	V78A, H138Y, T175I, V178I, A184V, H236Q, E252G, R255S, A290V, A295T, L353V
J[c]	Propane	DNA shuffling	Y138H, I178V, F205C, S226R, T295A
35E11[d]	Propane	Site saturation–recombination	R47C, A78F, A82S, K94I, P142S, A328F
ETS8[e]	Stability/propane	Site directed	L52I, I366V
19A12[e]	Propane	Random	L188P
11-3[e]	Propane	Site saturation	A74S
1-3[e]	Propane	Recombination	V184A
$P450_{PMO}$	Propane	Saturation–recombination/site directed	A74E, S82G, A184V, G443A

[a] Substitutions are given with respect to the preceding variant.
[b] From Ref. 29.
[c] From Ref. 31.
[d] From Ref. 30. This and subsequent variants carry also two mutations in the reductase domain (E464G, I710T). While important for ethane oxidation, these mutations have only a small effect on propane oxidation.
[e] From Ref. 27.

negative selection pressures that reinforce the desired phenotype while removing unfavorable side activities.[15,16] Negative selection may guide natural enzyme evolution in those cases where a promiscuous activity has a negative impact on viability, leading, for example, to accumulation of toxic intermediates or the waste of cellular resources. However, negative selection may not be required to achieve high specificity. It has been postulated that 'generalist' enzymes may have to trade off high activity on any one substrate for the ability to recognize and react with a wide range of substrates.[17] 'Specialist' enzymes could thus be the natural result of maximizing activity on one particular substrate and not require negative selection to remove other activities. Specialization under positive selection alone, however, has not yet been demonstrated experimentally.

A large family of heme-containing monooxygenases comprising more than 7000 known members†, the cytochrome P450s provide an extraordinary example of how a conserved structural fold diversified over millions of years to catalyze oxygenation reactions on widely different substrates.[18–20] In addition to their key roles in the breakdown of xenobiotics, cytochrome P450s catalyze important steps in the biosynthesis of flavonoids,[21] alkaloids,[22] terpenes,[23] sterols,[24] and other natural products. Unlike the xenobiotic-oxidizing P450s, these latter, anabolic enzymes tend to display narrow substrate preference.[25,26] We have recently reported the engineering of a proficient P450 propane monooxygenase ($P450_{PMO}$R2) that catalyzes the subterminal hydroxylation of propane with *in vivo* activities comparable to those of naturally occurring alkane monooxygenases from alkanotrophic organisms.[27] This enzyme was obtained through multiple rounds of laboratory evolution starting from $P450_{BM3}$, a long-chain (C_{12}–C_{20}) fatty acid hydroxylase from *Bacillus megaterium* that has no measurable activity on propane.[28] Propane activity was first identified in a $P450_{BM3}$ variant (139-3) that was selected for improved activity toward octane.[29] Further evolution of this variant was directed to improving activity on propane.[27,30,31]

The entire 'fossil record'—sequences and biochemical properties of all mutants in the lineage, selection pressure, and even the mutational process—is knowable for laboratory evolution experiments.[32] Here, we examine intermediates along the pathway to the P450 propane monooxygenase to investigate how the applied evolutionary pressure affected the enzyme's structure and biochemical properties and to elucidate the mechanisms by which this novel, nonnative activity emerged and was refined.

Results

The 'evolutionary lineage' connecting $P450_{BM3}$ to $P450_{PMO}$ was reconstructed through eight variants. The series begins with the octane-hydroxylating variant 139-3 and includes subsequent variants that show at least a twofold increase in propane total turnover number (TTN) and/or >10% improvement in coupling efficiency for propane oxidation compared to their predecessors.[27,30,31] $P450_{PMO}$R2 carries a reductase domain mutation (D698G) beneficial to propane oxidation.[27] To facilitate comparison with wild-type $P450_{BM3}$ and other variants, we generated $P450_{PMO}$, where Gly698 was reverted to Asp by site-directed mutagenesis. $P450_{PMO}$ is highly active, supporting 33,400 turnovers on propane (*versus* 45,800 for $P450_{PMO}$R2) with a coupling efficiency of 93% (*versus* 98% for $P450_{PMO}$R2). The mutagenesis/screening methods used for the isolation of the selected variants together with their amino acid substitutions are summarized in Table 1.

Biochemical analysis of $P450_{PMO}$ and its evolutionary precursors

Changes in the catalytic parameters and stabilities of the variants are reported in Fig. 1. Kinetic constants (K_m, k_{cat}) for propane oxidation were determined from Michaelis–Menten plots of initial

† http://drnelson.utmem.edu/P450.statsfile.html

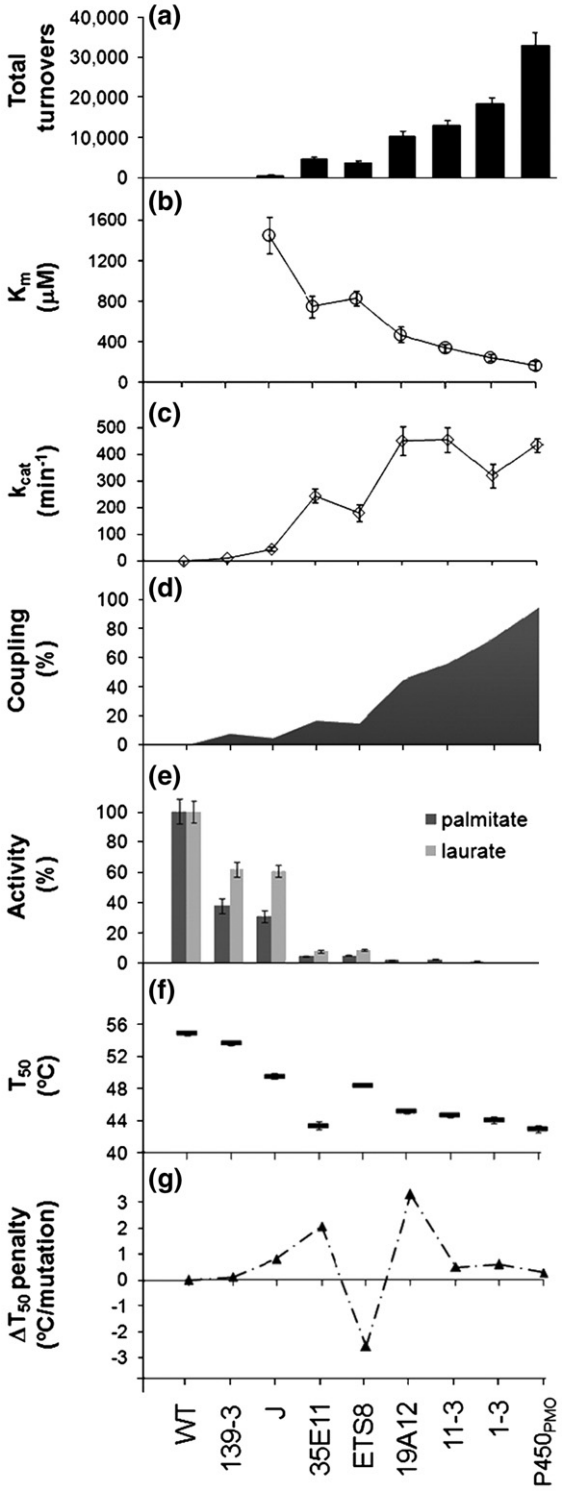

Fig. 1. Functional and kinetic properties of $P450_{PMO}$ evolutionary intermediates. (a) Total turnover numbers for propane hydroxylation (moles propanol per mole P450). (b) K_m values and (c) catalytic rate constants (k_{cat}) for propane oxidation. (d) Coupling efficiency as given by propanol formation rate/NADPH oxidation rate. Errors are within 10%. (e) Relative activity with palmitate and laurate as compared to $P450_{BM3}$. (f) Enzyme thermostability, reported as T_{50} values from heat inactivation of heme domain after 10 min incubation. (g) Average stability penalty per amino acid substitution for each step in the lineage.

propanol formation rates at different propane concentrations. Since saturation of 139-3 and J could not be achieved within the limits of propane solubility in water (≈1.6 mM), the kinetic parameters determined for these variants are approximate. Compared to 139-3 (wild-type $P450_{BM3}$ has no detectable propane activity), total turnovers supported by $P450_{PMO}$ increased 170-fold (Fig. 1a). This increase was accompanied by a 180-fold decrease in K_m (from ~30 mM in 139-3 to 170 μM in $P450_{PMO}$; Fig. 1b) and by a 45-fold increase in k_{cat} (from ~10 min^{-1} in 139-3 to 450 min^{-1}in $P450_{PMO}$; Fig. 1c). Overall, $P450_{PMO}$ shows a catalytic efficiency (k_{cat}/K_m) on propane of 4.4×10^4 M^{-1} s^{-1}, which is approximately 4 orders of magnitude greater than that of the first propane oxidizer, 139-3 (~5 M^{-1} s^{-1}). Despite the much reduced size and apolar nature of propane, $P450_{PMO}$'s K_m for propane oxidation is comparable to that of $P450_{BM3}$ for laurate oxidation (130 μM),[33] while its catalytic efficiency is only four times lower (cf. 1.9×10^5 M^{-1} s^{-1}).[33]

The coupling efficiency for propane oxidation improved from only 5–8% in the earliest generations to 93% in $P450_{PMO}$ (Fig. 1d). The largest change in coupling efficiency (15%→44%) was caused by the single L188P mutation between ETS8 and 19A12. Coupling of product formation to NADPH oxidation approaches 100% only with ideal substrates for which no competitive side reactions occur during the catalytic cycle.[34] Thus, the increasing coupling efficiency along the variant series suggests that the active site is becoming more optimized for propane binding. This conclusion is supported by the kinetic data and is consistent with the large decrease in K_m.

How the evolution of propane activity affects the native activity was evaluated by measuring variants' activity on the long-chain fatty acids palmitate (C16) and laurate (C12) and comparing those to that of wild-type $P450_{BM3}$ (Fig. 1e). $P450_{PMO}$ has no detectable activity on laurate and only marginal activity on palmitate (<150 TTN, or less than 0.5% that of $P450_{BM3}$). Palmitate is a significantly better substrate for $P450_{BM3}$ than laurate in terms of catalytic efficiency (6.0×10^7 M^{-1} s^{-1} *versus* 1.9×10^5 M^{-1} s^{-1}),[35] which probably explains why laurate activity was lost completely during $P450_{PMO}$ evolution. As Fig. 1e shows, native function was only mildly affected in the first two intermediates, 139-3 and J [J displays 31% (palmitate) and 61% (laurate) activity relative to wild type]. A more significant decrease was observed with 35E11 (<10% relative activity). The mutations accumulated in 35E11 include R47C (Table 1), which removed a side chain important for recognition of the carboxylate moiety in fatty acid substrates.[36] We reverted Cys47 to Arg in both 35E11 and $P450_{PMO}$ to examine its contribution to the loss of fatty acid activity. In 35E11, C47R led to moderate recovery of activity (laurate, 8–24% relative activity; palmitate, 5–18% relative activity), whereas $P450_{PMO}$-C47R still had no measurable activity on laurate and negligible activity on palmitate (~180 TTN). These results indicate that refinement of propane activity came at the cost of the

original function regardless of the presence of specific fatty-acid-binding determinants. In our experiments, no detectable activity on propane or fatty acids corresponds to a catalytic efficiency of less than 1 M^{-1} s^{-1}. Therefore, from $P450_{BM3}$ to $P450_{PMO}$ the overall change in specificity is at least a factor of $(4.4\times10^4/1)/(1/1.9\times10^5)\sim10^{10}$.

To establish the impact of the accumulated mutations on enzyme stability, T_{50} values were determined from heat-inactivation curves of the heme-dependent CO-binding spectral shift (Fig. 1f). These analyses show a gradual but continual decrease in the heme domain stability along the evolutionary pathway, reflecting the fact that stability was not under selection. The average impact of each mutational event on stability varies considerably though, as shown by the plot of per-mutation stability penalties (Fig. 1g). With 35E11, the much reduced stability compared to wild type ($\Delta T_{50}=-11.6$ °C) prevented further evolution and necessitated the stabilization step that led to ETS8, at which point a new beneficial mutation (L188P) could be found.[27] To investigate whether L188P exerts the same effect in the background of the less stable precursor, we introduced L188 in 35E11, where it caused a perturbation in the heme environment as indicated by the appearance of a 420-nm peak (~40%) in the CO-difference spectrum of the purified enzyme (data not shown). Furthermore, in contrast to the ~300% increase in $TTN_{propane}$ as a result of Leu188 mutation to proline during ETS8→19A12 transition, L188P generated only a small improvement (<40%) in 35E11. Altogether, these results highlight the role of the stabilizing mutations in enabling the fold to accommodate the beneficial but destabilizing L188P substitution. Saturation mutagenesis of position 188 in 19A12 confirmed proline as the best possible amino acid substitution (data not shown).

Substrate specificity profiles

Specificities of the evolved P450 variants were investigated using the alkane series from C_2 (ethane) to C_{10} (decane). Total turnovers—a cumulative measure of both catalytic and coupling efficiency—were determined for each substrate–P450 pair under identical reaction conditions (Supplementary Table S1). To enable comparison, TTNs were normalized to the highest value in the series (maximal turnover number). The resulting profiles (Fig. 2a) reveal a remarkable shift in substrate preference along the evolutionary path. The activity of $P450_{BM3}$ on the C_2–C_{10} alkanes was marginal (maximal turnover number = 160) and restricted to C_n >8 alkanes. In the earliest steps, substrate preference centered on the medium-chain alkanes (C_8–C_7), extending only later (35E11) toward shorter alkanes. A change in substrate preference appeared with 19A12. Subsequent variants 11-3, 1-3, and $P450_{PMO}$ showed a progressive decrease in TTN with alkane length. In $P450_{PMO}$, specificity along the alkane series has become remarkably restricted around propane: TTN drops by 55% for butane and by 95% for pentane.

Initial rates of product formation were compared for C_3 to C_9 alkanes (Fig. 2b). In the earliest intermediates, the preference for longer-chain alkanes is indicated by high $v_o(C_9)/v_o(C_3)$ ratios (51 for 139-3; 14 for J). This ratio decreased to ~1 in 35E11 and then switched in favor of the smaller alkane in subsequent generations. In $P450_{PMO}$, the product formation rate was highest on propane and decreased along the alkane series [$v_o(C_9)/v_o(C_3)=0.039$]. Overall, the ratio of the rates on propane *versus* nonane changed 1300-fold from 139-3 to $P450_{PMO}$.

Upon substrate binding and displacement of the water ligand from the heme iron in the low-spin configuration, ferric heme converts to a high-spin configuration, which appears as a blue shift of the Soret band from ~41 to ~392 nm.[37] How the variants interact with alkanes was further investigated by testing the heme iron spin shifts in the presence of propane, pentane, and octane (Fig. 2c). High-spin content was estimated based on reference spectra of $P450_{BM3}$ in complex with increasing amounts of palmitate. Octane binding induced between 35% and 50% spin shift in 139-3 and later-generation variants. Significant propane- and pentane-induced spin shifts (>25%) were observed only in the last three variants studied. Interestingly, 19A12 and its successors showed partial high-spin content (5–10%) even in the absence of substrate, suggesting that the mutations displace water in the active site or interfere with hydrogen bonding to the axial water ligand.[38]

In a further probe of the active site, we also tested $P450_{PMO}$ and its precursors on three structurally diverse terpenes: guaiol, valencene, and limonene. Oxidation products were quantified by GC (Fig. 3). For all variants, the TTN changed inversely with substrate size (guaiol < valencene ≪ limonene). Early intermediates 139-3 and particularly J showed the highest activities with all three, demonstrating substrate promiscuity that extends beyond alkanes. A large drop in activity was observed in the J→35E11 transition, followed by a gradual but steady decrease as activity on propane increased. $P450_{PMO}$ behaved similarly to $P450_{BM3}$, showing negligible activity on guaiol and valencene (TTN <150) and very weak activity on limonene (TTN = 580).

Halomethane dehalogenase activity

$P450_{PMO}$ supports thousands of turnovers on ethane (TTN ~2450) despite the higher C–H bond strength (101.0 kcal mol^{-1} for ethane *versus* 98.6 kcal mol^{-1} for the subterminal C–H of propane). To determine whether $P450_{PMO}$ can insert oxygen into C–H bonds stronger than those of ethane, $P450_{PMO}$ activity was tested on halomethanes (CH_3Cl, CH_3Br, and CH_3I), which have molecular sizes and C–H bond energies intermediate between those of methane and propane[39] (Supplementary Fig. S2). Reactions with halomethanes were analyzed for formation of formaldehyde (Fig. 4a) due to the

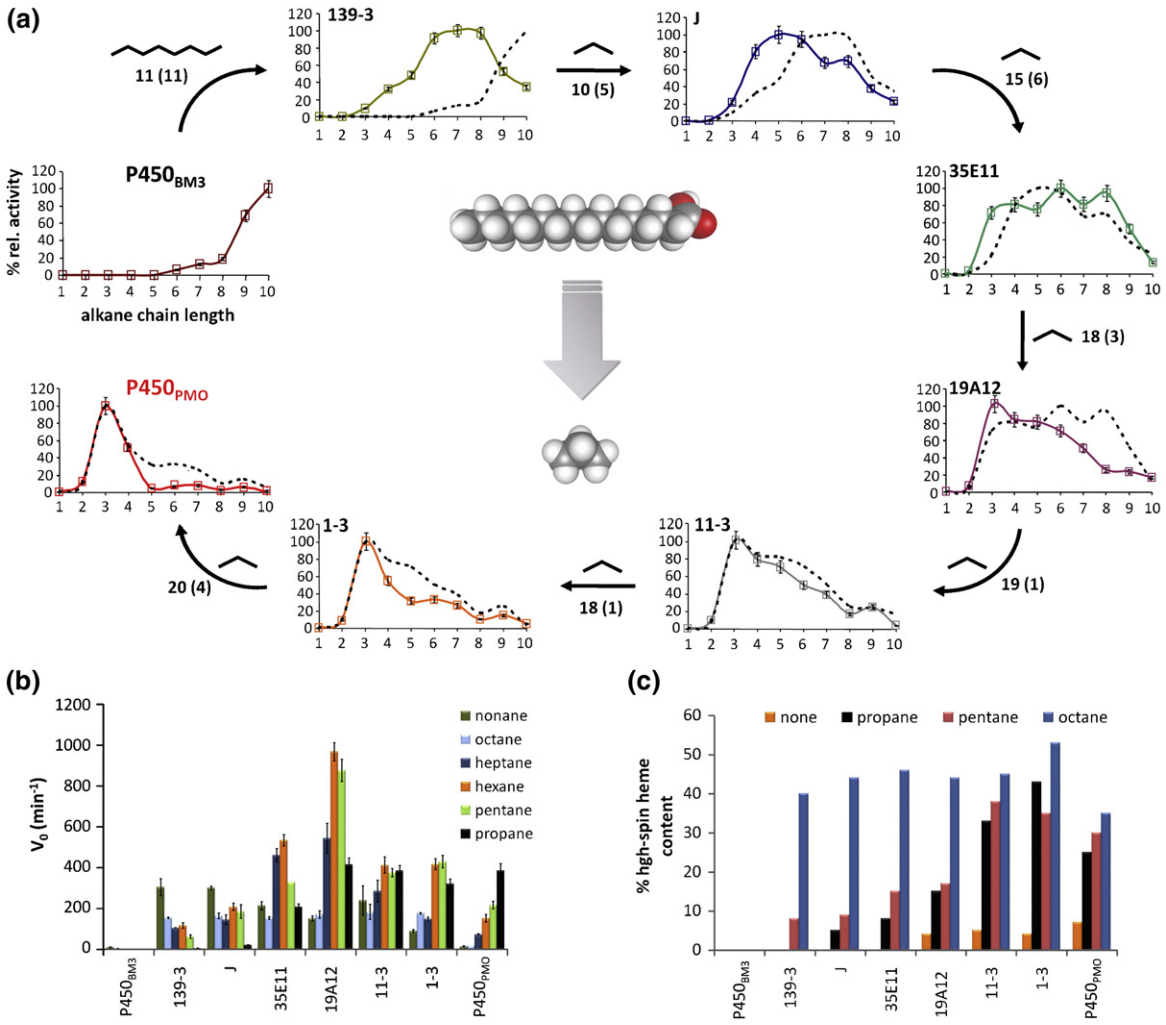

Fig. 2. Evolution of a P450 propane monooxygenase from a long-chain fatty acid hydroxylase. (a) Substrate specificity profiles of $P450_{PMO}$ and its evolutionary intermediates on the C_2–C_{10} alkane series. Maximal turnover numbers: $P450_{BM3}$ = 160; 139-3 = 2100; J = 3450; 35E11 = 6350; 19A12 = 10,500; 11-3 = 13,200; 1-3 = 19,200; $P450_{PMO}$ = 33,400. The dashed line indicates the profile of the preceding variant. Numbers indicate the number of heme domain mutations with respect to $P450_{BM3}$ and the preceding variant (brackets). The chemical structures next to the arrows refer to the selective pressure (octane or propane). The preferred substrates of $P450_{BM3}$ (palmitate) and $P450_{PMO}$ (propane) are displayed as space-filling models. (b) Initial oxidation rates with different alkanes, as measured by total nanomoles of hydroxylated alkane per nanomole enzyme per minute. (c) Alkane-induced shifts of heme spin state with different substrates, as determined from 350 to 500 nm spectra.

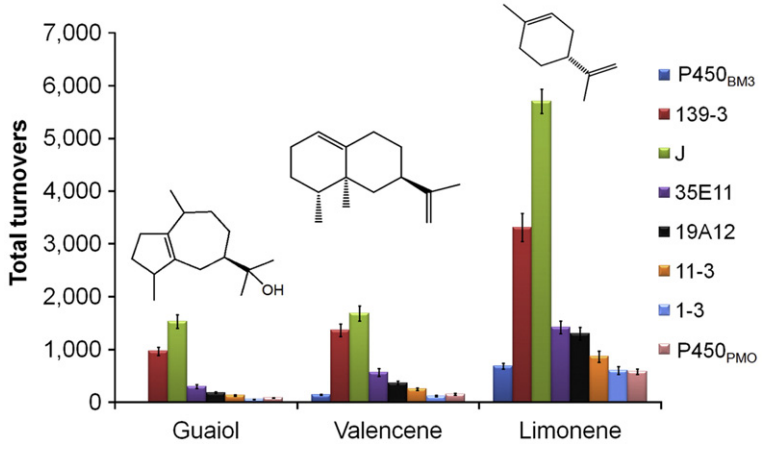

Fig. 3. Activities of $P450_{PMO}$ evolutionary intermediates on terpene substrates, as determined by GC analysis of the oxidation products.

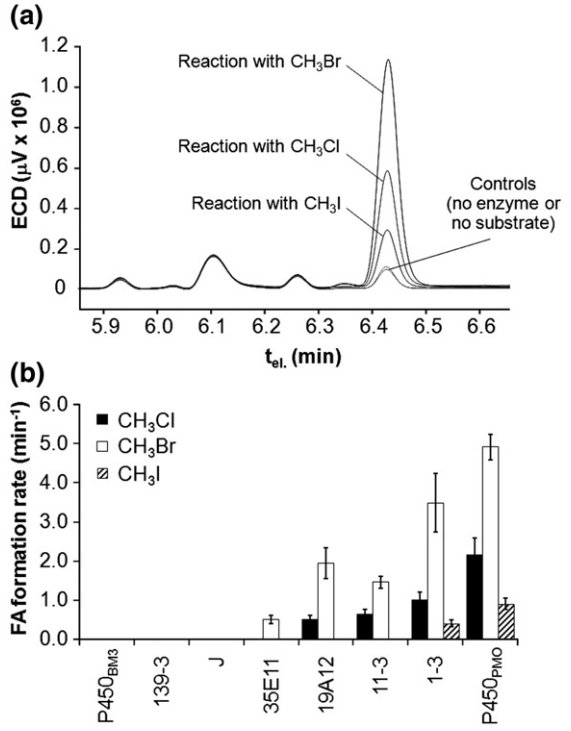

Fig. 4. Halomethane dehalogenase activities of $P450_{PMO}$ evolutionary intermediates. (a) GC traces of $P450_{PMO}$ reactions with CH_3Cl, CH_3Br and CH_3I after formaldehyde derivatization. (b) Halomethane dehalogenase activities as measured by the rate of formaldehyde formation.

expected α-elimination of HX upon hydroxylation. Measurable bromomethane activity first appeared in 35E11 and increased through the series at a level that is 1–5% of the propane activity (Fig. 4b). Despite the ~1 kcal mol^{-1} weaker C–H bond, chloromethane activity emerged only later (19A12) and was detected at lower levels (0.5–1% of propane activity), reflecting the preference of these variants for propane-sized substrates. Unlike CH_3Cl and CH_3Br, iodomethane is not oxidized by mammalian P450s.[40] Interestingly, $P450_{PMO}$ and its precursor 1–3 exhibited detectable activity on iodomethane, albeit with low turnover rates (0.5–1 min^{-1}). These results provide the first demonstration that a 103.0 kcal mol^{-1} sp^3 C–H bond is energetically accessible to oxygen insertion by a P450 enzyme.

Structural investigation of the $P450_{PMO}$ lineage

Crystallization experiments were undertaken to investigate structural changes along the lineage. While $P450_{PMO}$ and other variants failed to yield diffraction-quality crystals, early intermediate 139-3 could be crystallized in the presence of *N*-palmitoylglycine (NPG). The structure of the complex was solved at 2.6 Å resolution with $R = 0.21$ and $R_{free} = 0.26$ (see Table 2 for data collection and refinement statistics). The asymmetric unit contains two chains, A and B. The structure of $P450_{BM3}$ in complex with NPG is available.[41] Superposition of the 139-3 and $P450_{BM3}$ complexes revealed no gross changes in the enzyme fold (Fig. 5a, average C^{α} RMSD = 0.5 Å). The largest deviations (>1 Å) were in the N-terminal region (Thr1–Leu20), the loop in β-sheet β1 (Pro45–Arg47), the region between helix H and I (Pro243–Glu247), and the 3_{10} helix between Glu380 and Ser383 (Fig. 5a and Supplementary Fig. S3). The NPG ligand was found to adopt a more extended conformation in complex with 139-3, and a slight movement of β-sheet β1 toward helix F caused a restriction at the entrance of the substrate access channel (Fig. 5a). The relative positions of the residues critical for fatty acid interaction (Arg47, Tyr51) were unaltered. V78A and A184V are in the active site and are at short distance from one another (Fig. 5b). V78A is responsible for expansion of the hydrophobic pocket near the heme, whose volume increased by 29 $Å^3$ (Fig. 6a). In 139-3, Val184 forms a much tighter (3.3 Å *versus* 4.2 Å) contact with Leu437 across the substrate channel. Assuming octane orients similarly to NPG in the active site, Leu437 would establish closer van der Waals contacts near the terminal carbon of octane and enable more efficient exclusion of the shorter substrate from the solvent. This also translates into a change of the substrate orientation in the active site as reflected by the regioselectivity for octane oxidation in 139-3, which hydroxylates primarily at the 2-position (55% 2-octanol, 20% 3-octanol, 18% 4-octanol, and 7% ketone) *versus* the

Table 2. Data collection and refinement statistics

Data collection	
Maximum resolution (Å)	2.6
Unit cell dimensions	
a (Å)	62.05
b (Å)	127.82
c (Å)	183.94
Space group	$P2_12_12_1$
Total observations	154,017
Unique reflections	42,077
Redundancy	3.66
Completeness[a]	97.9 (92.6)
R_{sym}[a,b]	0.0078 (0.784)
$\langle I \rangle / \sigma$[a]	8.5 (1.3
Refinement	
R-factor[c]	0.21
R_{free}[d]	0.26
No. of protein atoms	7334
No. of heteroatoms	130
No. of water molecules	341
RMSD from ideal	
Bond length (Å)	0.007
Bond angle (deg)	1.4
PDB entry	3CBD

Backbone geometry analysis revealed only one residue (Leu^{437}) in a disallowed region of Ramachandran plot, which is nevertheless very well defined with clear density.

[a] Values in parentheses are for the highest-resolution shell.

[b] $R_{sym} = \sum |I - \langle I \rangle| / \sum I$, where I is the observed intensity and $\langle I \rangle$ the averaged intensity of multiple symmetry-related observations of the reflection.

[c] *R*-factor = $\sum ||F_o| - |F_c|| / \sum |F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively.

[d] R_{free} was calculated with the 5% of reflections set aside randomly throughout the refinement.

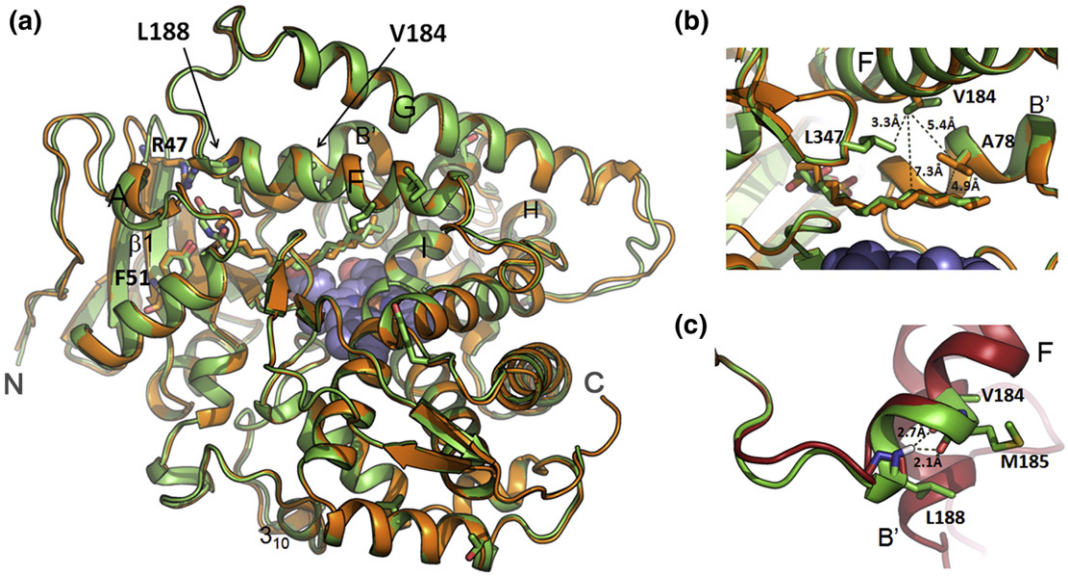

Fig. 5. Structure of variant 139-3. (a) Superposition of NPG-bound structures of $P450_{BM3}$ (orange, PDB code 1JPZ[41]) and evolutionary intermediate 139-3 (green, this work; PDB code 3CBD). The mutated residues, fatty acid-binding Arg47 and Tyr51, and Leu188 are displayed as stick models. The heme is displayed as a space-filling model. N- and C-termini and selected secondary-structure elements are indicated with capital letters. (b) Contacts of V78A and A184V substitutions with Leu347 and the bound ligand in 139-3 (green). (c) Superposition of 139-3 structure (green) and 19A12 model (red) illustrating the helix-capping role of Leu188 and the disruption of helical conformation as a consequence of the L188P mutation in 19A12.

wild-type enzyme, which prefers the 3- and 4-positions (5% 2-octanol, 49% 3-octanol, 43% 4-octanol, and 3% ketone).

Conservative structural models of 35E11, 19A12, 1-3, and $P450_{PMO}$ were calculated based on the 139-3 crystallographic model (chain B) using the Rosetta[42] structure prediction algorithm. In 35E11, three mutations (V78F, A82S, A328F) cause a major change in the substrate channel. Specifically, Phe328 (in concert with Leu437 and Ala74) constricts the

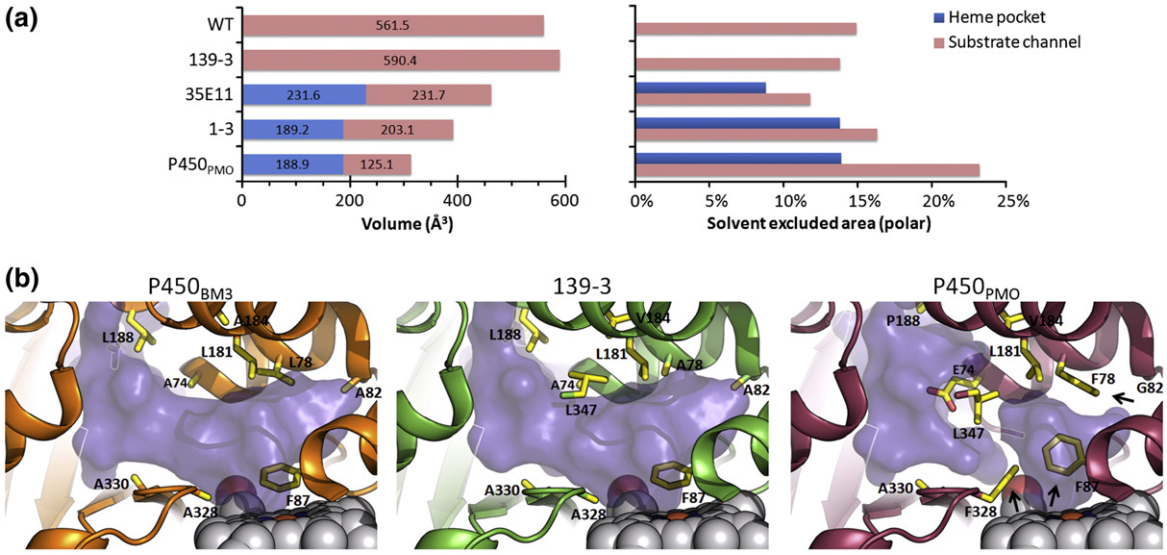

Fig. 6. Substrate access pathway. (a) Volumetric restriction and compartmentalization of the substrate access pathway during the transition of $P450_{BM3}$ to $P450_{PMO}$ (left). The A328F mutation in 35E11 interrupts the contiguity of the path by defining an inner cavity (at the heme pocket) from the remaining of the access path (indicated as 'substrate channel'). The computed fraction of polar surface area (right) shows an increasing difference in polarity between the inner and outer compartments along the lineage. (b) Surface representations of the substrate access pathway in $P450_{BM3}$ structure, 139-3 structure, and $P450_{PMO}$ structural model. The cavity above the heme first expands to accommodate octane (in 139-3). Upon selection for propane activity, this space becomes narrowly confined by a phenylalanine triad (black arrows) and segregated from the outer substrate channel through the interaction of Leu347 with opposing Glu74.

channel so that the contiguous substrate access pathway found in $P450_{BM3}$ and 139-3 (Fig. 6b) appears compartmentalized into an inner heme pocket (HP) and a more solvent-accessible substrate channel (SC) (as mapped using a probe radius of 1.4 Å). The total volume of these cavities is dramatically reduced compared to that of 139-3 ($\Delta V_{SC+HP} = -127$ Å^3; Fig. 6a). Experimental data support these calculations. The >70% drop in activity toward the terpenes suggests a much reduced accessibility of 35E11's active site compared to those of the preceding variants (139-3, J), while the increased regioselectivity with C_7–C_9 alkanes (>90% 2-alcohol) is indicative of a significant decrease in these substrates' conformational freedom within the active site.

In contrast to the other mutations, L188P was associated with a very unfavorable energy. The structure of 19A12 was therefore calculated allowing backbone motion for residues 185 to 191 and updating the conformation of this region using Loopy. In the resulting model, Pro188 causes partial unwinding of helix F C-terminus and removes Leu188 steric bulk, reducing the distance between helix F and adjacent helix B′ (Fig. 5c). These changes appear to have significant impact on the properties of the enzyme (see Discussion). In $P450_{PMO}$, the active-site mutations A184V, S74E, and S82G further restrict the substrate channel (Fig. 6a). This enables a more productive interaction of the enzyme with the substrate as indicated by the higher high-spin content upon propane binding and increased coupling in propane oxidation. The G443A mutation was found to affect the propane oxidation rate (twofold) but has little effect on coupling (91%) and K_m for propane (170 μM).

Discussion

Propane is one-fifth the size of a typical cytochrome $P450_{BM3}$ substrate and lacks the carboxylate moiety and long aliphatic chain through which the fatty acid interacts with the enzyme (Fig. 2a).[19,36] By studying the intermediates along the pathway to $P450_{PMO}$, we were able to observe how the enzyme has reelaborated its substrate-recognition strategies to achieve catalytic proficiency on this gaseous alkane. The 139-3 crystal structure shows that broadening of the substrate range of $P450_{BM3}$ to include octane and other medium-chain alkanes occurred with minimal perturbation of the overall fold. An enlarged hydrophobic pocket near the heme (Fig. 6a and b) appears to provide an energetically more favorable environment for binding to octane while also enabling the emergence of propane activity. These changes interfered only slightly with the native enzyme activity, which decreased by a factor of ~2. The latter reflects a previously observed feature of enzyme evolution: improvement of side, or 'promiscuous,' activities under positive selection pressure often comes with only marginal cost to the native activity.[17] The substrate range of variants 139-3 and J appears to be broad and includes other hydrophobic substrates (i.e., terpenes) that differ considerably from those used for activity screening (octane and propane, respectively). This generalist character has been observed in other early products of directed evolution such as, for example, in engineered variants of a beta-glucuronidase.[4]

In the next few generations, the prevailing strategy for improvement of propane activity involved compartmentalization and restriction of the active-site cavity (cf. 35E11). This facilitated the interaction of the enzyme with the smaller substrate, as evidenced by (a) a sevenfold increase in $TTN_{propane}$, (b) appearance of the heme spin shift (8%) upon propane binding, and (c) a decrease in K_m to the submillimolar range. The substrate preference remained centered on C_6–C_8 alkanes, but a significant decrease (>90%) in the native function was observed. The 'turning point' in substrate specificity in the next step (35E11 → 19A12) was induced by a single amino acid substitution, L188P (ETS8's profile overlaps with that of 35E11; data not shown). In the 139-3 structure, Leu188 lies at the C-terminus of helix F that, together with helix G, forms a lid that undergoes conformational change upon substrate binding (Fig. 5a).[36] L188P removes the intrastrand H bonds established by the Leu188 amide NH and is associated with a perturbation of the surrounding helical region (Fig. 5c) and a large destabilizing effect ($\Delta T_{50} = -3.3$ °C *versus* mean value of −0.6 °C/mutation, Fig. 1g). The stability cost of this mutation is offset by the advantage—the largest on a per-residue basis—it provides in propane oxidation, reflected in a threefold improvement in TTN and coupling efficiency, a twofold increase in turnover rate constant k_{cat}, and a twofold reduction in K_m (830 μM in ETS8 *versus* 475 μM in 19A12). By bringing helix F and helix B′ in closer proximity, this mutation may force the enzyme to adopt a conformation similar to the closed, catalytically more favorable one.[36]

After 19A12, a gradual decrease in activity towards longer alkanes accompanied the increased activity on propane, even though no negative selection was applied during the evolution of these variants to specifically favor this trend. Mapping of the substrate access pathway revealed a twisted path connecting the exterior to the heme (Fig. 6b). This configuration apparently provides an effective solution to channeling propane to the active site. It is also likely to hamper diffusion of longer-chain alkanes to and from the active site, explaining the observed difference in oxidation rates and TTNs observed within the alkane series. A comparison can be made with the soluble non-heme diiron methane monooxygenase (sMMO), an evolutionarily unrelated enzyme found in methanotrophic bacteria.[43] Studies addressed the question of how sMMO can discriminate methane from hundreds of other potential substrates with weaker C–H bonds.[44,45] The α subunit of sMMO hydroxylase domain was found to direct methane to the active-site pocket through a sequence of five buried cavities.[45] The

remarkable restriction ($\Delta V = -248$ Å^3; Fig. 6a) and compartmentalization of the substrate access pathway along the $P450_{PMO}$ lineage suggest a similar adaptive strategy to dealing with the problem of recognizing a small gaseous substrate. This can be rationalized in light of the negligible binding energy contribution ($-\Delta\Delta G$ <3–4 kcal/mol) provided by neutral ligands with fewer than five non-hydrogen atoms[46] and the fact that propane presumably has insufficient binding energy to induce the lid closure and accompanying conformational changes that characterize the interaction of $P450_{BM3}$ with long-chain fatty acids.[36,47] In contrast to the highly hydrophobic 'chambers' in sMMO, however, the substrate access channel in $P450_{PMO}$ has a more pronounced polar character. In particular, a progressive partitioning between a hydrophobic heme pocket and a relatively more polar substrate channel was observed along the variant series (Fig. 6a), a trend that culminated with a charged residue (Glu74) interposing between the two (Fig. 6b). Noteworthy in this regard was the progressive increase in polarity and bulkiness of residue 74 over the final evolutionary steps: Ala (19A12)→Ser (1–3)→Glu ($P450_{PMO}$).

While optimal for oxidation of the gaseous substrate, the acquired changes are incompatible with the native and other activities. In $P450_{PMO}$, there remains little trace of the original $P450_{BM3}$ function. Furthermore, in contrast to the earliest intermediates, which exhibit broad substrate ranges, $P450_{PMO}$ appears to have completely respecialized, displaying very narrow substrate tolerance within the C_2–C_{10} alkane series and little activity toward other, unrelated substrates (Figs. 2 and 3). Thus, the respecialization of $P450_{PMO}$ is induced by the same forces that pushed for optimization of the function under selective pressure and appears to be a consequence of acquisition of catalytic proficiency with propane.

Altogether, our results have several important implications. First, these data show that far from being 'evolutionary dead ends,' specialized enzymes can serve as starting points for divergent trajectories that lead to other, specialized members of the same enzyme family. Also, we provide the first experimental evidence that positive selection alone is sufficient to induce complete respecialization of a P450 enzyme, which suggests that negative selection need not have played an important role in the evolution of specialized members of this and possibly other enzyme families. Finally, the emergence of iodomethane dehalogenase activity in $P450_{PMO}$ is noteworthy, as mammalian P450s and sMMO do not exhibit it.[40,48] Iodomethane is known to be degraded by a very small subset of methylotrophic species, using enzymes such as corrinoid-dependent methyltransferases.[49] The propane-adapted features of $P450_{PMO}$'s active site enable the emergence of a basal activity towards this substrate, thus providing an alternative enzymatic strategy for the degradation of this compound and further expanding the range of known P450 activities.

Materials and Methods

Kinetic, biochemical, and biophysical analyses

Enzymes were expressed and purified as previously described.[27] $P450_{PMO}$, $P450_{PMO}$(C47R), 35E11(C47R), and 35E11(L188P) were obtained by site-directed mutagenesis (see Supplementary Table S2 for primer sequences). Michaelis–Menten plots for propane oxidation (e.g., Supplementary Fig. S1) were obtained varying propane concentration by dilution of propane-saturated solutions (≈1.6 mM). Reactions (5 mL) were carried out at 25 °C in 100 mM KPi buffer (pH 8.0) using 100–400 nM enzyme and 1 mM NADPH. After 20 s, reactions were stopped with 200 μL concentrated H_2SO_4 and analyzed by gas chromatography–electron capture detector (GC–ECD) as previously described.[27] Initial rates (20 s) for alkane oxidation were measured using 1 mM substrate (2% ethanol), 50 nM P450, and 1 mM NADPH. Reactions were stopped with 100 μL 2 N HCl, extracted with CH_2Cl_2, and analyzed by GC–flame ionization detector (GC-FID). Coupling efficiency corresponds to the ratio between propanol formation rate and NADPH oxidation rate in the presence of propane and was measured as previously described.[27] Half-denaturation temperatures (T_{50}) were determined as follows. Samples of purified enzyme (~3 μM) were incubated for 10 min at different temperatures (from 20 to 60 °C) in a PCR thermocycler. After centrifugation, 160-μL protein solutions were mixed with 40 μL 0.1 M sodium hydrosulfite on a microtiter plate and incubated for 10 min with carbon monoxide in a sealed chamber. T_{50} values were calculated from heat-inactivation curves of CO-binding difference spectra. Alkane-induced spin shifts were recorded on a Shimadzu UV–VIS spectrophotometer using P450 (2 μM) and 1 mM octane, pentane, or propane (2% ethanol). High-spin content was estimated by comparison with spectra of $P450_{BM3}$ and palmitate mixtures. Reactions with alkanes from pentane to decane were carried out at 3-mL scale using KPi buffer containing 2% ethanol, 50 nM enzyme, 1.6 mM alkane, and a cofactor regeneration system (CRS: 10 U/mL isocitrate dehydrogenase, 7 mg/mL isocitrate, 0.15 mg/mL $NADP^+$). Reaction mixtures were sealed, stirred for 24 h (4 °C), extracted with CH_2Cl_2 and analyzed by GC–FID. Reactions with ethane, propane, and butane were carried out as described for C_5–C_{10} alkanes using alkane-saturated buffers. All kinetic and biochemical measurements were performed in triplicate. For C_5–C_{10} alkane reactions, GC analyses were carried out using a Shimadzu GC-17A gas chromatograph, an Agilent HP5 column (30 m × 0.32 mm × 0.1 μm film), 1 μL injection, FID detector, and the following program: 250 °C inlet, 50 °C (3 min), 5 °C/min to 100 °C, 25 °C/min to 230 °C, and 230 °C (3 min). For butane and propane reactions (TTN), GC analyses were carried out using a Hewlett-Packard 5890 Series II Plus gas chromatograph, a Supelco SPB-1 column (60 m × 0.52 mm × 0.5 μm film), 0.5 μL injection, FID detector, and the following program: 250 °C inlet, 80 °C (2 min), 10 °C/min to 110 °C, 25 °C/min to 275 °C, and 275 °C (2.5 min).

Analysis of fatty acid and terpene activity

Reactions with laurate and palmitate were carried out at 2-mL scale using 2 mM fatty acid, 100–500 nM enzyme, and CRS. After stirring for 24 h, solutions were acidified with HCl, extracted with hexane, and dried over Na_2SO_4. *N*-Methyl-*N*-trimethylsilylheptafluorobutyramide 1:10

(v/v) was added and the mixture was analyzed by GC–FID using a Shimadzu GC-17A gas chromatograph, an Agilent HP5 column, 1 μL injection, and the following program: 300 °C inlet, 100 °C (2 min), 15 °C/min to 150 °C, 10 °C/min to 250 °C, and 250 °C (4 min). Reactions with terpene substrates were carried out at 2-mL scale using 1 mM guaiol, 1 mM valencene, or 5 mM limonene, 0.5 μM enzyme, and CRS. After the mixture was stirred for 24 h, reaction products were extracted with 200 μL $CHCl_3$ and analyzed by GC–FID using a Shimadzu GC-17A gas chromatograph, an Agilent HP5 column, 1 μL injection, FID detector, and the following program: 300 °C inlet, 70 °C, 10 °C/min to 210 °C, 50 °C/min to 260 °C, and 260 °C (2 min). All activity measurements were performed in triplicate.

Reactions with halomethanes

Reactions were carried out at 1-mL scale using halomethane-saturated 100 mM KPi (pH 8.0), 50 to 500 nM enzyme, and CRS. Buffers were prepared with HPLC-grade water and autoclaved twice prior to use. After 30 min reaction, 1-mL aliquots were mixed with 50 μL *o*-(2,3,4,5,6-pentafluorobenzyl)-hydroxylamine hydrochloride (Sigma) in formaldehyde-free water (17 mg/mL) and incubated for 15 min at 25 °C. Mixtures were mixed with 10 μL 25 mM 1-bromohexane (internal standard) and extracted with 500 μL hexane. The organic phase was diluted 1/100 (v/v) and analyzed by GC–ECD using a Shimadzu GC-17A gas chromatograph, an Agilent HP1 column (30 m ×0.53 mm×2.65 μm film), 1 μL injection, and the following program: 70 °C inlet, 45 °C (1 min), 15 °C/min to 200 °C, 50 °C/min to 220 °C, and 220 °C (3 min). A calibration curve was constructed from freshly prepared solutions of formaldehyde. All measurements were performed in triplicate.

Protein crystallization

For crystallization, the heme domain of 139-3 (1–455 residues; 139-3H) was cloned into the EcoRI/KpnI cassette of vector pProEX-1 using 5′-GATCCGGAATTCATGACAATTAAAGAAATG-3′ and 5′-GCTAAAAAAGTATAGGTACCAAGCCTTGCC-3′ as forward and reverse primers, respectively. Recombinant *Escherichia coli* DH5α cells were grown in terrific broth medium and induced with 0.6 mM IPTG at OD_{600} 0.7. Flasks were transferred to 25 °C and harvested after 20 h. Cells were lysed in 50 mM sodium phosphate (pH 7.4), 100 mM NaCl, 1 mM PMSF, and 10 mM β-mercaptoethanol (buffer A). Cell lysate was loaded onto a Ni–NTA agarose column preequilibrated with buffer A. After the column was washed with 50 mM $NaHPO_4$ (pH 7.4), 100 mM NaCl, 10 mM β-mercaptoethanol, and 10% glycerol (buffer B), bound protein was eluted with buffer B in the presence of 100 mM imidazole. After concentration, the protein was loaded onto a Sephacryl S200 gel-filtration column reequilibrated with buffer A. Purified 139-3H was mixed with NPG and crystallized by hanging-drop vapor-diffusion method from a protein solution at 40 mg/mL and a solution containing 100 mM Mops (pH 7.0), 50 mM LiBr, and 20% PEG (polyethylene glycol) 20,000. Crystals were soaked in 2 M sodium malonate as cryoprotectant prior to data collection.

Data collection and structure determination

X-ray diffraction data were collected on beam line 5.0.1 at Advanced Light Source (Berkeley, CA) using an ADSC Quantum 4 CCD detector. Optimization of data collection was guided by the STRATEGY function of MOSFLM.[50] All data were reduced using DENZO and scaling and rejections were performed with SCALEPACK.[51] The structure was solved using molecular replacement as implemented in AMORE in CCP4 suite[52] and the structure of $P450_{BM3}$ complexed with palmitoleic acid (PDB ID code 1FAG)[36] as search model. Two solutions in the asymmetric unit were found using data between 43.0 and 3.0 Å. Phases were improved and extended to 2.65 Å over 50 iterations of density modification using DM[53] as well as twofold noncrystallographic symmetry averaging. The structure was further refined using the graphical model building program O[54] and CNS[55] to a final resolution of 2.65 Å. Backbone geometry was checked in PROCHECK.[56] Data collection and refinement statistics are provided in Table 2.

Model calculations and substrate access pathway analysis

Structural models were calculated with Rosetta[42] using the structure of intermediate 139-3 (chain B) as reference model. The positions of mutated side chains were optimized (including off-rotamer continuous optimization) using full-atom Rosetta energy function. To accommodate L188P, the backbone between residues 185 and 191 was updated using Loopy.[57] Substrate access pathways were analyzed with Quickhull[58] using Cys400 and C^{α} atoms of residues 72–75, 78, 82, 87, 88, 177, 182, 185, 259, 263–268, 328–332, 357, and 435–439 as reference points and constructing a convex hull around these points. Within the hull, a 'cast' was formed by placing dummy atoms at all 0.5 Å grid points at least 3 Å from a protein atom. Substrate pathways were visualized displaying the surface of the cast with PyMOL. Solvent-excluded volumes were computed using MSMS[59] with a probe radius of 1.4 Å and a density of 3.0/$Å^2$. Reported volume and surface area values refer to volume occluded and the area buried, respectively. Based on MSMS per atom surface area breakdown, polar and nonpolar surface areas were measured using OPLSaa charges and grouped atoms with $|q| < 0.2$ into the nonpolar category.

Protein Data Bank accession number

Coordinates and structure factors have been deposited in the Protein Data Bank‡ with accession number 3CBD.

Acknowledgements

We are grateful to Huiying Li for assistance in collecting diffraction data. This work was supported by a Swiss National Science Foundation postdoctoral fellowship to R.F., a Jane Coffin Childs postdoctoral fellowship to C.S.F., NIH Grant GM32688 to T.L.P., and DOE grant (DE-FG02-06ER15762) and U.S. Army Research Office grant (AROICB DAAD19-03-D-0004) to F.H.A. Graphics were prepared using PyMOL Molecular Graphics System§.

‡ www.pdb.org
§ http://www.pymol.org

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2008.06.060

References

1. Tatusov, R. L., Koonin, E. V. & Lipman, D. J. (1997). A genomic perspective on protein families. *Science*, **278**, 631–637.
2. Laskowski, R. A. & Thornton, J. M. (2008). Understanding the molecular machinery of genetics through 3D structures. *Nat. Rev. Genet.* **9**, 141–151.
3. Olsen, M. J., Stephens, D., Griffiths, D., Daugherty, P., Georgiou, G. & Iverson, B. L. (2000). Function-based isolation of novel enzymes from a large library. *Nat. Biotechnol.* **18**, 1071–1074.
4. Matsumura, I. & Ellington, A. D. (2001). In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. *J. Mol. Biol.* **305**, 331–339.
5. Zaccolo, M. & Gherardi, E. (1999). The effect of high-frequency random mutagenesis on in vitro protein evolution: a study on TEM-1 beta-lactamase. *J. Mol. Biol.* **285**, 775–783.
6. Zhang, J. H., Dawes, G. & 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.
7. Yano, T., Oue, S. & Kagamiyama, H. (1998). Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proc. Natl Acad. Sci. USA*, **95**, 5511–5515.
8. Yang, J., Fu, X., Liao, J., Liu, L. & Thorson, J. S. (2005). Structure-based engineering of *E. coli* galactokinase as a first step toward in vivo glycorandomization. *Chem. Biol.* **12**, 657–664.
9. Jensen, R. A. (1974). Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **30**, 409–425.
10. Aharoni, A., Gaidukov, L., Khersonsky, O., Mc, Q. G. S., Roodveldt, C. & Tawfik, D. S. (2005). The 'evolvability' of promiscuous protein functions. *Nat. Genet.* **37**, 73–76.
11. Williams, P. A., Cosme, J., Sridhar, V., Johnson, E. F. & McRee, D. E. (2000). Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol. Cell*, **5**, 121–131.
12. Gillam, E. M. & Guengerich, F. P. (2001). Exploiting the versatility of human cytochrome P450 enzymes: the promise of blue roses from biotechnology. *IUBMB Life*, **52**, 271–277.
13. Muralidhara, B. K., Negi, S., Chin, C. C., Braun, W. & Halpert, J. R. (2006). Conformational flexibility of mammalian cytochrome P450 2B4 in binding imidazole inhibitors with different ring chemistry and side chains. Solution thermodynamics and molecular modeling. *J. Biol. Chem.* **281**, 8051–8061.
14. Ekroos, M. & Sjogren, T. (2006). Structural basis for ligand promiscuity in cytochrome P450 3A4. *Proc. Natl Acad. Sci. USA*, **103**, 13682–13687.
15. Varadarajan, N., Rodriguez, S., Hwang, B. Y., Georgiou, G. & Iverson, B. L. (2008). Highly active and selective endopeptidases with programmed substrate specificities. *Nat. Chem. Biol.* **4**, 290–294.
16. O'Loughlin, T. L., Greene, D. N. & Matsumura, I. (2006). Diversification and specialization of HIV protease function during in vitro evolution. *Mol. Biol. Evol.* **23**, 764–772.
17. Khersonsky, O., Roodveldt, C. & Tawfik, D. S. (2006). Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.* **10**, 498–508.
18. Ortiz de Montellano, P. R. (2005). *Cytochrome P450: Structure, Mechanism, and Biochemistry.* Kluver Academic/Plenum Publishers, New York.
19. Pylypenko, O. & Schlichting, I. (2004). Structural aspects of ligand binding to and electron transfer in bacterial and fungal P450s. *Annu. Rev. Biochem.* **73**, 991–1018.
20. Denisov, I. G., Makris, T. M., Sligar, S. G. & Schlichting, I. (2005). Structure and chemistry of cytochrome P450. *Chem. Rev.* **105**, 2253–2277.
21. Lepiniec, L., Debeaujon, I., Routaboul, J. M., Baudry, A., Pourcel, L., Nesi, N. & Caboche, M. (2006). Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant Biol.* **57**, 405–430.
22. Facchini, P. J. (2001). Alkaloid biosynthesis in plants: biochemistry, cell biology, molecular regulation, and metabolic engineering applications. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 29–66.
23. Gang, D. R. (2005). Evolution of flavors and scents. *Annu. Rev. Plant Biol.* **56**, 301–325.
24. Edwards, P. A. & Ericsson, J. (1999). Sterols and isoprenoids: signaling molecules derived from the cholesterol biosynthetic pathway. *Annu. Rev. Biochem.* **68**, 157–185.
25. Wittstock, U. & Halkier, B. A. (2000). Cytochrome P450 CYP79A2 from *Arabidopsis thaliana* catalyzes the conversion of L-phenylalanine to phenylacetaldoxime in the biosynthesis of benzylglucosinolate. *J. Biol. Chem.* **275**, 14659–14666.
26. Larbat, R., Kellner, S., Specker, S., Hehn, A., Gontier, E., Hans, J. *et al.* (2007). Molecular cloning and functional characterization of psoralen synthase, the first committed monooxygenase of furanocoumarin biosynthesis. *J. Biol. Chem.* **282**, 542–554.
27. Fasan, R., Chen, M. M., Crook, N. C. & Arnold, F. H. (2007). Engineered alkane-hydroxylating cytochrome P450(BM3) exhibiting nativelike catalytic properties. *Angew Chem. Int. Ed. Engl.* **46**, 8414–8418.
28. Narhi, L. O. & Fulco, A. J. (1987). Identification and characterization of two functional domains in cytochrome P-450BM-3, a catalytically self-sufficient monooxygenase induced by barbiturates in *Bacillus megaterium*. *J. Biol. Chem.* **262**, 6683–6690.
29. Glieder, A., Farinas, E. T. & Arnold, F. H. (2002). Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. *Nat. Biotechnol.* **20**, 1135–1139.
30. Meinhold, P., Peters, M. W., Chen, M. M., Takahashi, K. & Arnold, F. H. (2005). Direct conversion of ethane to ethanol by engineered cytochrome P450 BM3. *ChemBioChem*, **6**, 1765–1768.
31. Peters, M. W., Meinhold, P., Glieder, A. & Arnold, F. H. (2003). Regio- and enantioselective alkane hydroxylation with engineered cytochromes P450 BM-3. *J. Am. Chem. Soc.* **125**, 13442–13450.
32. Adami, C. (2006). Digital genetics: unravelling the genetic basis of evolution. *Nat. Rev. Genet.* **7**, 109–118.
33. Noble, M. A., Miles, C. S., Chapman, S. K., Lysek, D. A., Mackay, A. C., Reid, G. A. *et al.* (1999). Roles of active site residues in flavocytochrome P450 BM3. *Biochem. J.* **339**, 371–379.
34. Kadkhodayan, S., Coulter, E. D., Maryniak, D. M., Bryson, T. A. & Dawson, J. H. (1995). Uncoupling oxygen transfer and electron transfer in the oxygenation of camphor analogues by cytochrome P450-

CAM. Direct observation of an intermolecular isotope effect for substrate C–H activation. *J. Biol. Chem.* **270**, 28042–28048.
35. Oliver, C. F., Modi, S., Primrose, W. U., Lian, L. & Roberts, G. C. K. (1997). Engineering the substrate specificity of *Bacillus megaterium* cytochrome P-450 BM3: hydroxylation of alkyl trimethylammonium compounds. *Biochem. J.* **327**, 537–544.
36. Li, H. & 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.
37. Tsai, T., Yu, C. A., Gunsalus, I. C., Peisach, J., Blumberg, W., Orme-Johnson, W. H. & Beinart, H. (1970). Spin-state changes in cytochrome P450cam on binding of specific substrates. *Proc. Natl Acad. Sci. USA*, **66**, 1157–1163.
38. Raag, R. & Poulos, T. L. (1989). The structural basis for substrate-induced changes in redox potential and spin equilibrium in cytochrome P-450CAM. *Biochemistry*, **28**, 917–922.
39. Cherkasov, A. & Jonsson, M. (2000). A new method for estimation of homolytic C–H bond dissociation enthalpies. *J. Chem. Inf. Comput. Sci.* **40**, 1222–1226.
40. Chamberlain, M. P., Lock, E. A. & Reed, C. J. (1998). Investigations of the pathways of toxicity of methyl iodide in the rat nasal cavity. *Toxicology*, **129**, 169–181.
41. Haines, D. C., Tomchick, D. R., Machius, M. & Peterson, J. A. (2001). Pivotal role of water in the mechanism of P450BM-3. *Biochemistry*, **40**, 13456–13465.
42. Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A. J., Read, R. J. & Baker, D. (2007). High-resolution structure prediction and the crystallographic phase problem. *Nature*, **450**, 259–264.
43. Lieberman, R. L. & Rosenzweig, A. C. (2004). Biological methane oxidation: regulation, biochemistry, and active site structure of particulate methane monooxygenase. *Crit. Rev. Biochem. Mol. Biol.* **39**, 147–164.
44. Zhang, J., Zheng, H., Groce, S. L. & Lipscomb, J. D. (2006). Basis for specificity in methane monooxygenase and related non-heme iron-containing biological oxidation catalysts. *J. Mol. Catal. A: Chem*, **251**, 54–65.
45. Sazinsky, M. H. & Lippard, S. J. (2006). Correlating structure with function in bacterial multicomponent monooxygenases and related diiron proteins. *Acc. Chem. Res.* **39**, 558–566.
46. Kuntz, I. D., Chen, K., Sharp, K. A. & Kollman, P. A. (1999). The maximal affinity of ligands. *Proc. Natl Acad. Sci. USA*, **96**, 9997–10002.
47. Murataliev, M. B. & Feyereisen, R. (1996). Functional interactions in cytochrome P450BM3. Fatty acid substrate binding alters electron-transfer properties of the flavoprotein domain. *Biochemistry*, **35**, 15029–15037.
48. Smith, T. J. & Dalton, H. (2004). Biocatalysis by methane monooxygenase and its implications for the petroleum industry. In *Petroleum Biotechnology, Developments and Perspectives* (Vazquez-Duhalt, R. & Quintero-Ramirez, R., eds), pp. 177–192, 2004 edit. Elsevier Science, Amsterdam, Netherlands.
49. McDonald, I. R., Warner, K. L., McAnulla, C., Woodall, C. A., Oremland, R. S. & Murrell, J. C. (2002). A review of bacterial methyl halide degradation: biochemistry, genetics and molecular ecology. *Environ. Microbiol.* **4**, 193–203.
50. Powell, H. R. (1999). The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **55**, 1690–1695.
51. Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
52. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A*, **50**, 157–163.
53. Cowtan, K. (1994). An automated procedure for phase improvement by density modification. *Joint CCP4 & ESF-EACBM Newslett.* **J31**, 34–38.
54. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **47**, 110–119.
55. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S. *et al.* (1998). Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **54**, 905–921.
56. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
57. Xiang, Z., Soto, C. S. & Honig, B. (2002). Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction. *Proc. Natl. Acad. Sci. USA*, **99**, 7432–7437.
58. Barber, C. B., Dobkin, D. P. & Huhdanpaa, H. T. (1996). The Quickhull algorithm for convex hulls. *ACM Trans. Math. Software*, **22**, 469–483.
59. Sanner, M. F., Olson, A. J. & Spehner, J. C. (1996). Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers*, **38**, 305–320.