DOI: 10.1002/chem.200900643

A Panel of Cytochrome P450 BM3 Variants to Produce Drug Metabolites and Diversify Lead Compounds

Andrew M. Sawayama,[a] Michael M. Y. Chen,[a] Palaniappan Kulanthaivel,[b] Ming-Shang Kuo,[b] Horst Hemmerle,[b] and Frances H. Arnold*[a]

Abstract: Herein we demonstrate that a small panel of variants of cytochrome P450 BM3 from *Bacillus megaterium* covers the breadth of reactivity of human P450s by producing 12 of 13 mammalian metabolites for two marketed drugs, verapamil and astemizole, and one research compound. The most active enzymes support preparation of individual metabolites for preclinical bioactivity and toxicology evaluations. Underscoring their potential utility in drug lead diversification, engineered P450 BM3 variants also produce novel metabolites by catalyzing reactions at carbon centers beyond those targeted by animal and human P450s. Production of a specific metabolite can be improved by directed evolution of the enzyme catalyst. Some variants are more active on the more hydrophobic parent drug than on its metabolites, which limits production of multiply-hydroxylated species, a preference that appears to depend on the evolutionary history of the P450 variant.

Keywords: C–H activation · cytochrome P450 · drug development · drug metabolism · oxidation

Introduction

Selective C–H oxidation represents one of the great challenges for which synthetic chemists find only very substrate-specific solutions.[1–4] Breakthroughs in selective C–H oxidation methodology could benefit drug discovery, among other fields, by enabling rapid and parallel analogue construction for a specific molecular scaffold. Such improvements in efficiency would greatly increase the number and variety of compounds that could be produced, and raise the likelihood of identifying effective therapeutic agents.

Many synthetic strategies for C–H oxidation rely on a reactive intermediate that plays upon subtle differences in C–H bond strength (1–5 kcal mol^{-1}) to achieve regioselectivity.[5] Owing to the large number of C–H bonds in most bioactive chemicals, identifying a reagent that can react at one C–H bond in preference to all others can be difficult, or even impossible. Nature solves the selectivity problem by incorporating discrete molecular recognition elements into enzyme catalysts so that they can use specific enzyme–substrate interactions to impart reactivity to a specific C–H bond.

Cytochromes P450 (CYPs) are a large superfamily of heme-containing C–H oxidation enzymes. In humans, CYPs play key roles in drug metabolism and clearance. An ability to predict how potential pharmaceuticals will be metabolized will better equip us to identify derivatives with improved biological activity, solubility, toxicity, stability or bioavailability.[6] In fact, FDA guidelines indicate that uniquely human metabolites and metabolites present at disproportionately higher levels in humans as compared to the animal species used during standard toxicology testing may require safety assessment before beginning large-scale clinical trials.[7]

Preparation of these metabolites at sufficient scale for evaluation is not trivial and often requires a de novo synthesis for *each* metabolite. Biosynthetic methods that employ purified human CYPs or crude liver microsomes are not much better, because human CYPs are poorly stable, membrane-bound, multiprotein systems that exhibit low reaction rates. As an alternative to using human CYPs as biocatalysts for metabolite production, we and others have focused on soluble, bacterial P450 BM3 (also known as CYP102A1)[8] as a C–H oxidation platform.[9–11] Derived from *Bacillus megaterium*, P450 BM3 has properties that greatly facilitate

[a] Dr. A. M. Sawayama, M. M. Y. Chen, Prof. F. H. Arnold
Division of Chemistry and Chemical Engineering 210-41
California Institute of Technology, Pasadena, CA 91125-4100 (USA)
Fax: (+1) 626-528-8743
E-mail: frances@cheme.caltech.edu

[b] Dr. P. Kulanthaivel, Dr. M.-S. Kuo, Dr. H. Hemmerle
Eli Lilly & Company, Indianapolis, IN 46285 (USA)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.200900643.

its engineering and use in synthesis: it can be expressed at high levels in *E. coli* (~12% dry cell mass), and, unlike nearly all other CYPs, its hydroxylase, reductase and electron-transfer domains are all in one contiguous polypeptide chain. This last feature might contribute to its relatively high activity (>1000 turnovers per min) on its preferred fatty acid substrates.[8] Like most CYPs implicated in anabolic pathways, P450 BM3 is substrate specific, and hydroxylates a C-20 fatty acid over a C-12 fatty acid with more than 200-fold higher efficiency.[12]

We chose three structurally diverse drug compounds with known patterns of mammalian CYP-dependent clearance to evaluate whether P450 BM3 variants can catalyze similar C–H oxidations. Verapamil is a calcium channel blocker used in the treatment of hypertension and arrhythmia.[13] Astemizole is a potent H_1-histamine receptor antagonist used for treatment of common sinus allergy symptoms.[14] The third compound, LY294002, is an antiproliferative agent that inhibits phosphatidylinositol 3-kinase, a target with potential for treatment of malignancies.[15] We report here that a small collection of P450 BM3 variants can produce nearly all the known human (or rat) metabolites for each of the three drugs. Within each set of active enzyme variants, we identified several that produce selected metabolites in yields and activities suitable for preparative scale synthesis. We also identified variants that generate metabolites *not* produced by rat liver microsome controls or known to be human metabolites; the results demonstrate the ability of this C–H oxidation platform to target a range of carbon centers.

Results

Selection of P450 variants: From our extensive collections we selected 120 P450 BM3 variants that had previously demonstrated activity on substrates not in the wild-type enzyme's repertoire. These variants were constructed by using a variety of commonly implemented genetic diversification techniques, including error-prone PCR and targeted mutagenesis of active site residues. We also included variants derived from structure-guided recombination of BM3 and its homologues, CYP102A2 and CYP102A3, to produce chimeric enzymes.[16] These enzymes had been selected based on their activities towards a variety of substrates, including straight-chain alkanes, cycloalkanes, alkyl ethers,[17–19] aromatic compounds,[20] derivatives of fatty acids,[21] and drugs.[10,22]

We also distinguished between enzyme variants that function as monooxygenases and use NADPH to reduce the catalytic Fe^{III} center, and peroxygenases that contain only the heme domain and use H_2O_2 in place of O_2 and NADPH. Each has its own catalytic and operational advantages and disadvantages that have been described elsewhere.[21] Though we had evidence that the presence of a reductase domain impacts activity favorably,[20] we were unsure what effect, if any, this domain would have on regioselectivity. The variants, their sequences relative to wild type, and the criteria for their prior selection from mutant libraries are described in Table S1 in the Supporting Information.

Reactions with human and bacterial CYPs: Verapamil and astemizole undergo extensive biotransformation by the major human CYPs to produce demethylated and dealkylated products[23] by hydroxylation of activated carbon centers adjacent to heteroatoms (summarized in Tables 1A, 2A). Experiments conducted with rat liver microsomes in vitro indicate that LY294002 metabolism is confined to single and double alkyl hydroxylations of the morpholine ring (Table 3), a pattern also observed in other drugs possessing this moiety.

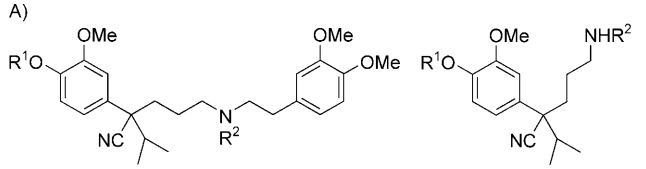
Reactions of the 120 P450 BM3 variants with each of the three drugs were assessed by HPLC. Metabolite products were evident in the HPLC traces for a significant number of variants (43 exhibited some activity on verapamil, 42 on astemizole and 18 on LY294002); the metabolites and their distributions were subsequently characterized by using LCMS and MS/MS (Tables S3, S4 and S5 in the Supporting Information). Metabolite structure assignments were aided by comparison of MS and MS/MS spectra with spectra of known metabolites obtained in our laboratories using rat microsomal systems.

Verapamil: Table 1 summarizes the performance of selected enzymes on verapamil (all 43 active enzymes and their product distributions are listed in Table S3 in the Supporting Information). For example, entry 4 displays the activity of chimera 22313333 on verapamil. It converts 34% of the starting material into an assortment of products, the distribution of which is 41% **2**, 15% **5**, 20% **6**, 9% **7** and 15% **10**. Variants 2C11 and 9C1 furnished the widest array of products, with 6 and 8 individual metabolites, respectively. In particular, 9C1 produced several metabolites that had undergone two hydroxylation events (**1**, **4** and **8**). Variant 22313333 (entry 4) was the only enzyme capable of producing double hydroxylation product **5**, while 32313233 (entry 6) best produced norverapamil (**3**). Interestingly, the addition of reductase as a fusion to the 32313233 heme domain (entry 7) rendered the enzyme unable to produce new metabolite **7**. Variant 7-11D (entry 13) differs by no more than two mutations from the variants in each of entries 9–12, but its product distribution more closely parallels chimera 21312332. The 9-10A family of enzymes was the most active, with conversions exceeding 30% (entries 9–12), and produced metabolite **7** with excellent selectivity.

The enzymes that best mimicked human CYP reactivity on verapamil were derived from two variant families. Three had been isolated by directed evolution for activity on propranolol,[22] while the remaining four were chimeras.[16,24] The BM3 variants that best produced new metabolites were derived from the alkane-hydroxylating 9-10A variant.[18,19] These enzymes catalyzed a new regioselective demethylation reaction at position R^4 and a new benzylic oxidation at position R^3. Even without optimization of the reaction con-

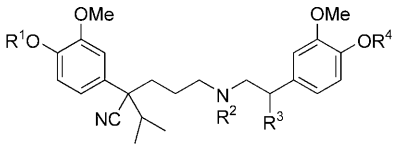
Table 1. Conversion of verapamil to: A) its most abundant[a] human metabolites, and B) new metabolites by different P450 BM3 variants.

A)



	M_w	R^1	R^2		M_w	R^1	R^2
Verapamil	454	Me	Me	**4**	276	H	Me
1	426	H	H	**5**	276	Me	H
2	440	H	Me	**6**	290	Me	Me
3	440	Me	H				

B)



	M_w	R^1	R^2	R^3	R^4	other
7	440	Me	Me	H	H	
8	456	Me	Me	OH	H	
9	470	Me	Me	OH	Me	
10	470	Me	Me	H	Me	OH

	Variant[b]	% Conversion	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	*2C11*	25		8	28			20	8		8	28
2	*9C1*[c]	31	3		29	3		6	13	3	10	13
3	*D6H10*	78			26	8		31			3	24
4	**22313333**	34		41			15	20	9			15
5	**22313231**	43		46				33	5			16
6	**32313233**	24			50			21	13			17
7	**32313233-R1**	25			40			44				16
8	**21312332**	6			17			33	17			33
9	9-10A A78F	30		10					83			
10	9-10A A82L	51						2	94			
11	9-10A F87L	49						6	94			
12	12-10C	34			6			6	79	3		
13	7-11D	21			28			29	14			29

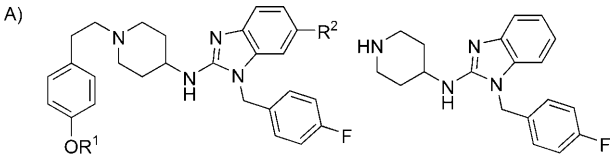
[a] Defined as >1% abundance following oral ^{14}C verapamil administration in humans.[23] [b] Variants shown in italics were selected for activity on propranolol,[22] variants in bold text are chimeras,[16] variants in normal type were selected for activity on alkanes.[18,19] Chimeras are written according to fragment composition: 32313233-R1, for example, represents a protein that inherits the first fragment from parent CYP102A3, the second from CYP102A2, the third from CYP102A3, and so on. R1 connotes a fusion to the reductase domain from parent A1. Chimera fusions were used as monooxygenases; chimera heme domains were used as peroxygenases. [c] Because not all P450 BM3 oxidation products could be identified, product distribution totals can be less than 100%.

ditions, some enzymes were highly active on verapamil; D6H10 (entry 3), for example, transformed verapamil into metabolites at 78% conversion and a total turnover number (TTN) greater than 1500.

The enzymes possess contrasting degrees of regioselectivity. Some were unselective and produced a spectrum of metabolites (e.g., propranolol-evolved enzymes 2C11, 9C1,

Table 2. Conversion of astemizole to: A) its most abundant[a] human metabolites, and B) new metabolites by different P450 BM3 variants.

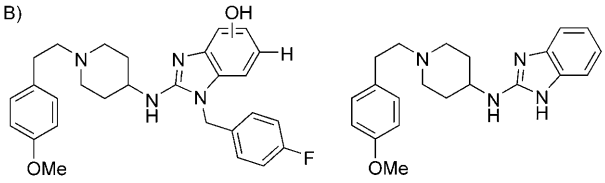
A)



	M_w	R^1	R^2
Astemizole	458	Me	H
11	444	H	H
12	460	H	OH
13	474	Me	OH

	M_w
14	324

B)



	M_w	other
15	474	
16	490	OH

	M_w
17	350

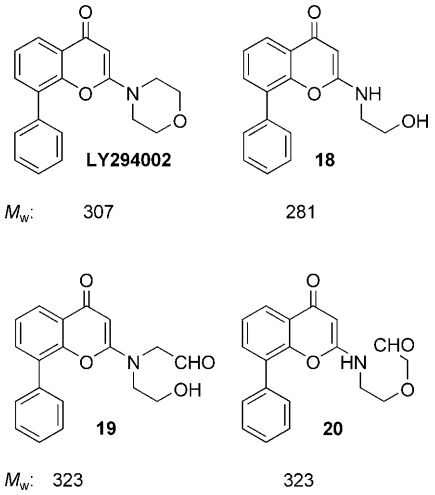
	Variant[b]	% Conversion	**11**	**12**	**13**	**14**	**15**	**16**	**17**
1	*DE10*	9				56			
2	**21313311**	10	20			40			40
3	**22313333**	9	22			56			22
4	9-10A A78F	21	48		38			14	
5	9-10A A78T	27	37	7	44			4	
6	9-10A A82S	30	20	7	70			3	
7	9-10A F87L	36		3	88			6	
8	41-5B	5	100						
9	**32313233**	11				45	27		27
10	**32313233-R1**	16				13	69		
11	**32312333-R1**	78				4	67		9
12	9-10A A78S	15	13	7			80		
13	9-10A A82L	49	24				61	4	
14	9-10A A82I	32	31	6	63				
15	9-10A F87A	31			23	16	45		

[a] Defined as >1% abundance following oral ^{14}C astemizole administration in humans.[25] [b] See Table 1 for further explanation of variant nomenclature.

D6H10). By contrast, chimera 22313231 produced metabolite **2**, chimera 32313233-R1 made dealkylated compound **6** (entries 5 and 7), and many variants produced new metabolites **7** and **10** (entries 8–13) with sufficient selectivity (≥ 30%) for larger scale production without further optimization. To demonstrate that useful quantities of metabolites can be produced, we used purified 9-10A F87L to produce metabolite **7** (9.4 mg) from verapamil (25 mg) in 39% yield (1560 TTN, 0.025 mol% catalyst).

Astemizole: Activity of the enzyme panel on astemizole resulted in the seven metabolites described in Table 2 (all 42 active enzymes and their product distributions are listed in Table S4 in the Supporting Information). Variants DE10,

Table 3. Conversion of LY294002 by different P450 BM3 variants into the most abundant[a] metabolites produced by rat liver microsomes.



	Variant[b]	% Conversion	18	19	20
1	9-10A L75I	10		50	
2	9-10A A78S	7		71	
3	9-10A F87V	6			100
4	9-10A T260S	11		9	
5	68-8F	9		11	

[a] Defined as >1% abundance following in vitro reaction of LY294002 with rat liver microsomes. [b] Variants in normal type were selected for activity on alkanes.[18,19]

21313311, 22313333, 32313233 (entries 1–3 and 9) produced dealkylated metabolite **14** in preference to other compounds. The identity of the residue at position 78 had a strong impact on the product distribution within the 9-10A backbone; mutations A78F, A78T and A78S (entries 4, 5 and 12) produced metabolites **11**, **13** and **15** as the most abundant respective products. Furthermore, 41-5B (entry 8), a 9-10A family member that contains the A78F mutation (in addition to A82G and A328V) also favors demethylated **11**. Variants 9-10A A82S and 9-10A F87L both produced metabolite **13** in good yield and selectivity (entries 6 and 7). Two chimeric monooxygenases, 32313233-R1 and 32312333-R1 (entries 10 and 11) were effective in generating new aromatic hydroxylation product **15**.

The 9-10A-derived monooxygenases were the most adept at producing human metabolites of astemizole (entries 4–8). However, these enzymes were unable to dealkylate astemizole and produce metabolite **14**. Oxidation of this more hindered C–H bond was best accomplished by the propranolol-active variants and chimeric peroxygenases (entries 1 and 2, 3, respectively). A new benzylic site not hydroxylated by human CYPs was targeted by six chimeric BM3 enzymes to produce **17**. A second benzimidazole site was hydroxylated (as demonstrated by metabolites **15** and **16**), which is different from that observed in human metabolites **12** and **13**. Discrimination of these sterically and electronically identical sp^2 C–H bonds is virtually impossible by using traditional transition metal catalysts and emphasizes the power of molecular recognition as a regiocontrolling element.

Small changes in the substrate channel can affect the regioselectivity of aromatic C–H oxidation. For instance, the presence and position of a single methyl group in Thr78 vs. Ser78 (entry 5 vs. 12) and Ile82 vs. Leu82 (entry 14 vs. 13) were sufficient to bias oxidation at C6 instead of at another position on the benzimidazole ring. Product **16** has undergone two oxidations. Though this metabolite was observed in reactions with ten variants, it was produced in greater than 10% abundance by only one (entry 4). New metabolite **15** is a candidate for scale-up and could be produced at high conversion (>25% selectivity) by 15 variants of different lineages (Table S4 in the Supporting Information). Even without optimization of reaction conditions, several enzymes showed good activity towards astemizole. Variant 32312333-R1, for example, converted 78% astemizole into metabolites (70 TTN).

LY294002: The most abundant metabolites produced by rat liver microsomes are detailed in Table 3. Both single hydroxylation products **19** and **20** were identified in BM3 variant reactions with very good regioselectivity. Aminoalcohol **18** is the sole metabolite of all three drugs not observed in the reactions. Derivative **18** requires two oxidations and might not appear simply due to the low conversion in these 96-well plate reactions.

The P450s catalyzing the single hydroxylations were monooxygenases derived from 9-10A (Table 3). Though several variants could produce metabolite **19**, only 9-10A F87V oxidized the other position of the morpholine ring and made derivative **20** (entry 3). The remaining 12 metabolite-producing variants were also monooxygenases of both 9-10A and chimeric origin, and produced a metabolite with $M_w = 238$; this indicates morpholine loss and addition of water to the bis-aryl backbone (Table S5 in the Supporting Information). In all cases, LY294002 conversion was very low (<15%) and correlates with the low rat liver microsome conversion for LY294002 (40%) relative to microsome activity on verapamil (85%) and astemizole (85%). LY294002 better evades both mammalian and bacterial P450-catalyzed C–H oxidation than verapamil and astemizole under the same conditions.

Activities on singly-hydroxylated metabolites: Of the 12 mammalian metabolites produced by the P450 BM3 variants, eight were made with sufficiently high selectivity to enable preparative scale production without further optimization (**2**, **3**, **6**, **11**, **13**, **14**, **19** and **20**). Of the seven new metabolites to which structures could be assigned, four were made with sufficiently high regioselectivity to enable preparative scale production (**7**, **10**, **15** and **17**). All 11 of the highly produced metabolites arose from single oxidations. Seven of the eight remaining metabolites are the products of two or more oxidations and are present as minor components in mixtures with the singly-hydroxylated products. Of the 69 individual reactions that underwent appreciable conversion (≥10% drug consumed) only five generated products of multiple C–H oxidations in ≥10% abundance.

The remainder contained single-oxidation products, including mixtures of singly-oxidized species. The lack of bis-hydroxylated products could reflect a regiochemical effect in which the second C–H site is less reactive. In this case, one of the two possible single-hydroxylation products would be favored. Sixteen variants produced only one single-hydroxylation product; reaction mixtures from ten additional variants contained two (metabolites were present in 1:1 to 3:1 ratios in 6/11 reactions). It is also possible that the relative lack of bis-hydroxylation products reflects discrimination on the part of the enzymes in which the BM3 binds and hydroxylates the parent drug in preference to its metabolites. Because BM3-catalyzed C–H oxidation increases polarity by either unmasking heteroatoms or formally substituting hydrogen with a hydroxyl group, production of poly-hydroxylated metabolites could be disfavored, for example, if the enzyme prefers a more hydrophobic substrate.

To assess enzyme activity on the metabolites produced by a single oxidation, all 120 enzymes were incubated with purified demethylated metabolites norverapamil (**3**) and desmethylastemizole[26] (**11**). Product distributions and extents of reaction were determined by HPLC. In every case in which an enzyme was active on both parent drug and metabolite, regioselectivities were unchanged; this indicates that the metabolite and parent drug bind in similar orientations. Comparison of each enzyme's activity on the parent drug and the associated metabolite showed that enzymes from the alkane-evolved 9-10A lineage tended to be more active on astemizole and verapamil than on their metabolites (true of 49 out of 55 active enzymes; Table 4). Because

Table 4. Preference for more hydrophobic parent drug over its metabolite depends on evolutionary history of the variant.

P450 Variant family	Number of P450s more active on parent drug	Number of P450s more active on metabolite
Alkane-selected lineage	49/55	6/55
Propranolol-selected and chimera lineages	18/30	12/30

neither the *N*-methyl group of verapamil nor the *O*-methyl group of astemizole is the preferred site for C–H oxidation by the alkane-evolved variants, this bias could reflect a preference for the more hydrophobic substrate. In contrast, when this pairwise reaction matrix is analyzed across the propranolol-evolved and chimeric P450 BM3 lineages, there is no statistical difference from random at the 95% confidence interval ($0.43 \leq p = 0.60 \leq 0.77$).

Directed evolution can improve metabolite production: None of the 120 members of this catalyst panel had been selected for activity on any of the three drugs. Thus, any initial activity represents a promiscuous activity; such side activities are often easy to improve by directed evolution.[27,28] Of the 103 enzymes that reacted with these substrates, 9-10A F87L possessed the best combination of activity and regioselectivity (Table 2, entry 7); its reaction with astemizole, which produces metabolite **13** with 88% selectivity, is representative of a reaction for which one would like to improve metabolite production. We speculated that a high-throughput screen for improved overall aromatic hydroxylation activity would yield at least a few variants that retained this regioselectivity while producing more metabolite. We screened 2000 variants made by error-prone PCR of the 9-10A F87L gene using a colorimetric screen for products of aromatic hydroxylation[29] and identified three new sequences with improved metabolite production (Table 5). The new variants improved the conversion to 51–52% while preserving high regioselectivity (~80% for metabolite **13**).

Table 5. Production of astemizole metabolites by 9-10A F87L variants.

	Variant	% Conversion	% Selectivity **13**
1	9-10A F87L	34	88
2	E4D D68G	51	80
3	C205R D338G	51	80
4	E4D H92Q	52	75

Discussion

A small, 120-member panel of P450 BM3 variants captured nearly all of the mammalian P450 scope of reactivity by producing 12 of 13 known metabolites. In their ability to mimic human CYPs, the P450 BM3 variants demonstrated considerable versatility, activating C–H bonds of varying strength (90–105 kcal mol^{-1}) and steric encumbrance (sp^2 vs. sp^3 and 1° vs. 2° carbon centers). We were able to assign structures to seven new metabolites (Tables 1B, 2B) all of which have undergone oxidation at new carbon centers.

Cytochrome P450 enzymes are versatile catalysts the biological activities of which are encoded in a wide range of primary sequences. The 26 CYPs for which the crystal structures have been solved possess as little as 15% sequence identity,[30] but share a highly conserved fold. Emphasizing the versatility of the P450 fold, P450cam and P450cin catalyze the oxidation of isosteres camphor and cineole. Although the substrates are nearly identical, the two enzymes differ not only in sequence (27% identity), but also in the structure of their active sites, most notably a complete lack of the B′ helix in P450cin.[31,32] Whereas large changes in sequence are possible, only small perturbations are necessary to produce significant changes in function. For example, CYP2A4 and CYP2A5 differ by only 11 amino acids, yet catalyze hydroxylations on structurally dissimilar coumarin and testosterone substrates.[33]

Only 57 human enzymes[34] are responsible for known CYP-dependent drug metabolism, and a single enzyme, CYP3A4, accounts for >50% of the burden for xenobiotic CYP-mediated clearance.[35] Because CYPs can be broadly or narrowly specific, we and others have speculated that it should be possible to take advantage of the high native activity of P450 BM3 and use mutation to either relax or shift its substrate specificity in order to generate useful C–H oxi-

dation catalysts.[9,36–39] For example, BM3 variant 9-10A, which is 13 mutations away from the wild type, exhibits broad activity across short- and medium-chain alkanes—activity that is low or completely absent in its wild-type parent.[18,19] Furthermore, 9-10A could be respecialized to hydroxylate propane, preferring it over alkanes that differ by a single methylene group.[40] We also showed previously that a variant of 9-10A was able to hydroxylate drug-like compounds efficiently and selectively.[10] Here we wanted to determine whether variants of 9-10A and other BM3-derived enzymes could cover or even exceed the broad substrate range of mammalian CYPs. Within any catalyst panel, both extremes of regioselectivity can be useful: enzymes that already possess the desired selectivity can be used to produce individual metabolites, whereas less selective enzymes can be used to survey metabolite possibilities. Both can serve as starting sequences for directed evolution to enhance activity or tune selectivity. That a systematic and general evolutionary algorithm can be used for catalyst improvement is a particularly appealing aspect of DNA-encoded reagents. Complementary optimization studies used in traditional synthesis methods usually rely on chemical intuition to improve a catalyst and require a good understanding of the catalytic mechanism.

When we examined conversion of the parent drug versus its demethylated metabolite, we noted that one enzyme family consistently converted more of the parent drug than the demethylated and more polar metabolite. These enzymes tend to catalyze single hydroxylations, while enzymes evolved for activity on propranolol and the chimeric variants often catalyze bis-hydroxylations. Structure–activity relationships of this type should help in the future to select enzymes that are most likely to react with as yet untested substrates and could also help predict product profiles.

Conclusion

This panel should enable rapid identification and production of relevant quantities of the human metabolites of drug candidates for pharmacological and toxicological evaluations in preclinical species.[41,42] Although we have highlighted the potential of these enzymes to accelerate preparation of metabolites for pharmacological and toxicological testing, this enzyme panel is likely also to be useful further upstream in the drug-development process as general reagents for lead diversification. Reagents that rely on molecular recognition will always be restricted in their scope of use. However, because the functionality of small molecules is not evenly distributed across all possible molecular architectures,[43] it should be worthwhile to engineer P450-derived reagents that are active on privileged scaffolds that reside in these densely functional regions of structure space. The plurality of C–H sites targeted by this small P450 BM3 variant set—including and extending human P450 metabolism—augurs well for the development of a truly general panel of C–H oxidation catalysts.

Experimental Section

General: All chemicals were purchased from Sigma–Aldrich, Inc. or provided by Eli Lilly directly. Solvents were purchased from EM Sciences. Lysis enzymes were purchased from Sigma. Absorbance measurements were conducted by using a SpectraMax 384 Plus plate reader. HPLC separations were performed by using a Supelco Discovery C18 column (2.1 × 150 mm, 3μ) on a Waters 2690 Separation module in conjunction with a Waters 996 PDA detector. LCMS and MS/MS spectra were obtained by using the ThermoFinnigan LCQ classic at the shared Caltech MS facility.

Protein expression: LB agar plates supplemented with ampicillin (100 $\mu g\,mL^{-1}$) were streaked with the catalase-deficient strain of *E. coli* SN0037[44] (for peroxygenase) or DH5α *E. coli* (for monooxygenase) containing a desired P450 BM3 variant in the isopropyl β-D-thiogalactopyranoside (IPTG)-inducible pCWori vector.[45] These were grown at 37 °C for 12 h before single clones were picked and added in quadruplicate to 1 mL 96-well plates containing LB medium (400 μL) supplemented with ampicillin (100 $\mu g\,mL^{-1}$). After being shaken at 80 % humidity, 30 °C for 24 h to grow the precultures to saturation, an aliquot (50 μL) was used to inoculate 2 mL 96-well plates containing TB medium (900 μL) supplemented with ampicillin (100 $\mu g\,mL^{-1}$). After being shaken at 80 % humidity, 30 °C for 5 h, P450 expression was induced by addition of IPTG (500 mM) and the heme precursor δ-aminolevulinic acid (δ-ALA) to a final concentration of 1 mM. The cultures were grown for another 24 h before the cells were centrifuged and stored at −20 °C.

Activity-based screening: All three drugs and two metabolites were screened against the cell lysate of the panel of 120 variants. Lysate was prepared by resuspending cell pellets with a buffer (600 μL) containing $MgCl_2$ (10 mM), lysozyme (0.5 $mg\,mL^{-1}$) and DNAse I (8 $U\,mL^{-1}$). For the holoenzyme reactions phosphate buffer (0.1 M, pH 8) was used and for the peroxygenase reactions EPPS buffer (0.1 M, pH 8.2) was used. The lysis reactions were incubated at 37 °C for 1 h and then centrifuged; the supernatant was used in three assays.

CO binding: Heme proteins absorb light at 450 nm corresponding to the CO stretch frequency when the Fe-heme is bound to CO. This Soret band can be used to quantify the amount of folded protein.[46] Lysate (100 μL) and sodium dithionite (100 μL, 140 mM in 1 M buffer, pH 8) were added to a 96-well flat-bottom screening plate. These were preread by using a plate reader at 450 and 490 nm before being incubated in a 1 atm CO chamber for 15 min. The plate was then read again at 450 and 490 nm.

Enzymatic activity: In a 96-well (2 mL) plate, the following mixtures were prepared for each of the activity-based assays. Holoenzyme: lysate (60 μL), phosphate (110 μL, 0.1 M, pH 8), drug/metabolite (10 μL, 5 mM), NADPH (20 μL, 20 mM); peroxygenase: lysate (50 μL), EPPS (100 μL, 0.1 M, pH 8.2), drug/metabolite (10 μL, 20 mM), hydrogen peroxide (40 μL, 5 mM).

Upon ultimate addition of NADPH or H_2O_2, the plates were briefly shaken and incubated for 2 h. After this time, acetonitrile (200 μL) was added to quench the reactions. The reactions were centrifuged and the supernatants were used for subsequent analysis with HPLC and LCMS.

HPLC: Supernatant (25 μL) was analyzed by HPLC. Conditions with solvent A (0.2 % formic acid *v*/*v*, in H_2O) and solvent B (acetonitrile) used to elute the products of metabolism were: 0–3 min, A/B 90:10; 3–25 min, linear gradient to A/B 30:70; 25–30 min, linear gradient to A/B 10:90.

LCMS, MS/MS: Identical conditions to the HPLC method detailed above were used for the LC portion of the analysis. The MS was operated in positive ESI mode. MS/MS spectra were acquired in a data-dependent manner for the most intense ions.

Preparation of metabolite 7: Glucose-6-phosphate (80 mM), glucose-6-phosphate dehydrogenase (2 $U\,mL^{-1}$), superoxide dismutase (100 $U\,mL^{-1}$), 9-10A F87L (250 nM), verapamil (24.6 mg, 1 mM) and NADP (5 mM) were added to an Erlenmeyer flask (250 mL) containing potassium phosphate buffer (50 mL, 100 mM, pH 8.0). These reagents were stirred vigorously at room temperature while metabolite production was monitored every hour for 4 h by HPLC. The reaction was quenched by precipitating the enzymes with acetonitrile (50 mL) and by stirring for

an additional 30 min followed by filtration. The resulting solution contained 9.4 mg metabolite **7** (39%).

Directed evolution of 9-10A F87L for activity on astemizole: Error-prone PCR was used to create the variant library. Cells containing the 9-10A F87L template in a pCWori plasmid were grown and plasmid DNA isolated. Error-prone PCR reactions (50 μL) contained water (38 μL), 10× Applied Biosystems PCR buffer (5 μL), forward primer 5'-GGA AAC AGG ATC CAT CGA TGC-3' (1 μL, 10 μM), reverse primer 5'-GTG AAG GAA TAC CGC CAA GC-3' (1 μL, 10 μM) and template plasmid (0.5 μL, 200 ng μL^{-1}), Taq polymerase (1 μL, 5 U μL^{-1}), dNTPs (1 μL, 10 μM) and $MnCl_2$ (2.5 μL, 1 mM). PCR conditions were 95 °C for 5 min, followed by 30 cycles of 30 s a 95 °C, 60 s at 52 °C and 120 s at 72 °C. The PCR products were treated with DpnI to remove template DNA followed by digestion with BamHI and SacI. Following agarose gel purification, the oligonucleotide library was ligated into the pCWori vector and then electroporated into DH5α *E. coli* cells. The transformants were plated on LB agar medium supplemented with ampicillin (100 μg mL^{-1}) and grown at 37 °C over 16 h. Protein was expressed, as described in the "Protein expression" portion of the Experimental Section. Variant hits were identified by first screening with a high-throughput colorimetric assay for aromatic hydroxylation.[29] Cells were lysed and prepared for reaction as described above. Upon ultimate addition of NADPH or H_2O_2, the plates were briefly shaken and incubated for 2 h. At this time, urea (60 μL, 8 M in 200 mM NaOH), 4-aminoantipyrine (18 μL, 1.2%) and potassium peroxodisulfate (18 μL, 1.2%) were added. Absorption at 510 nm was measured after 10 min. Reactions containing variant enzymes more active than the 9-10A F87L parent were repeated and characterized by HPLC, LCMS and MS/MS as described above; validated hits were sequenced.

Acknowledgements

This work was supported in part by NIH grant GM068664, the Jacobs Institute for Molecular Engineering and Medicine at Caltech, Eli Lilly and Co., and an NSF predoctoral fellowship to MMYC. We thank Dr. Terry Lindstrom for a critical reading of the manuscript.

[1] R. G. Bergman, *Nature* **2007**, *446*, 391–393.
[2] R. H. Crabtree, *J. Chem. Soc. Dalton Trans.* **2001**, 2437–2450.
[3] K. Godula, D. Sames, *Science* **2006**, *312*, 67–72.
[4] J. A. Labinger, J. E. Bercaw, *Nature* **2002**, *417*, 507–514.
[5] M. S. Chen, M. C. White, *Science* **2007**, *318*, 783–787.
[6] M. D. Johnson, H. Zuo, K. H. Lee, J. P. Trebley, J. M. Rae, R. V. Weatherman, Z. Desta, D. A. Flockhart, T. C. Skaar, *Breast Cancer Res. Treat.* **2004**, *85*, 151–159.
[7] FDA, FDA Guidance of Industry Safety Testing of Drug Metabolites Homepage: http://www.questpharm.com/FDA.html
[8] L. O. Narhi, A. J. Fulco, *J. Biol. Chem.* **1986**, *261*, 7160–7169.
[9] B. M. A. van Vugt-Lussenburg, E. Stjernschantz, J. Lastdrager, C. Oostenbrink, N. P. E. Vermeulen, J. N. M. Commandeur, *J. Med. Chem.* **2007**, *50*, 455–461.
[10] M. Landwehr, L. Hochrein, C. R. Otey, A. Kasrayan, J. E. Bäckvall, F. H. Arnold, *J. Am. Chem. Soc.* **2006**, *128*, 6058–6059.
[11] M. Budde, M. Morr, R. D. Schmid, V. B. Urlacher, *ChemBioChem* **2006**, *7*, 789–794.
[12] T. W. B. Ost, C. S. Miles, A. W. Munro, J. Murdoch, G. A. Reid, S. K. Chapman, *Biochemistry* **2001**, *40*, 13421–13429.
[13] D. Mctavish, E. M. Sorkin, *Drugs* **1989**, *38*, 19–76.
[14] D. M. Richards, R. N. Brogden, R. C. Heel, T. M. Speight, G. S. Avery, *Drugs* **1984**, *28*, 38–61.
[15] C. J. Vlahos, W. F. Matter, K. Y. Hui, R. F. Brown, *J. Biol. Chem.* **1994**, *269*, 5241–5248.
[16] C. R. Otey, M. Landwehr, J. B. Endelman, K. Hiraga, J. D. Bloom, F. H. Arnold, *PLoS Biol.* **2006**, *4*, 789–798.
[17] A. Glieder, E. T. Farinas, F. H. Arnold, *Nat. Biotechnol.* **2002**, *20*, 1135–1139.
[18] M. W. Peters, P. Meinhold, A. Glieder, F. H. Arnold, *J. Am. Chem. Soc.* **2003**, *125*, 13442–13450.
[19] P. Meinhold, M. W. Peters, A. Hartwick, A. R. Hernandez, F. H. Arnold, *Adv. Synth. Catal.* **2006**, *348*, 763–772.
[20] M. Landwehr, M. Carbone, C. R. Otey, Y. G. Li, F. H. Arnold, *Chem. Biol.* **2007**, *14*, 269–278.
[21] P. C. Cirino, F. H. Arnold, *Angew. Chem.* **2003**, *115*, 3421–3423; *Angew. Chem. Int. Ed.* **2003**, *42*, 3299–3301.
[22] C. R. Otey, G. Bandara, J. Lalonde, K. Takahashi, F. H. Arnold, *Biotechnol. Bioeng.* **2006**, *93*, 494–499.
[23] M. Eichelbaum, M. Ende, G. Remberg, M. Schomerus, H. J. Dengler, *Drug Metab. Dispos.* **1979**, *7*, 145–148.
[24] Y. G. Li, D. A. Drummond, A. M. Sawayama, C. D. Snow, J. D. Bloom, F. H. Arnold, *Nat. Biotechnol.* **2007**, *25*, 1051–1056.
[25] W. Meuldermans, J. Hendrickx, W. Lauwers, R. Hurkmans, E. Swysen, J. Heykants, *Drug Dev. Res.* **1986**, *8*, 37–51.
[26] F. Janssens, J. Torremans, M. Janssen, R. A. Stokbroekx, M. Luyckx, P. A. J. Janssen, *J. Med. Chem.* **1985**, *28*, 1934–1943.
[27] H. Joo, Z. L. Lin, F. H. Arnold, *Nature* **1999**, *399*, 670–673.
[28] J. C. Moore, F. H. Arnold, *Nat. Biotechnol.* **1996**, *14*, 458–467.
[29] C. R. Otey, J. M. Joern in *Methods in Molecular Biology, Vol. 230* (Eds.: F. H. Arnold, G. Georgiou), Humana Press, Totowa, **2003**, pp. 141–148.
[30] J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **1994**, *22*, 4673–4680.
[31] Y. T. Meharenna, H. Y. Li, D. B. Hawkes, A. G. Pearson, J. De Voss, T. L. Poulos, *Biochemistry* **2004**, *43*, 9487–9494.
[32] T. L. Poulos, B. C. Finzel, I. C. Gunsalus, G. C. Wagner, J. Kraut, *J. Biol. Chem.* **1985**, *260*, 6122–6130.
[33] R. L. P. Lindberg, M. Negishi, *Nature* **1989**, *339*, 632–634.
[34] D. F. V. Lewis, *Pharmacogenomics* **2004**, *5*, 305–318.
[35] F. P. Guengerich, *Annu. Rev. Pharmacol. Toxicol.* **1999**, *39*, 1–17.
[36] D. Appel, S. Lutz-Wahl, P. Fischer, U. Schwaneberg, R. D. Schmid, *J. Biotechnol.* **2001**, *88*, 167–171.
[37] T. Kubo, M. W. Peters, P. Meinhold, F. H. Arnold, *Chem. Eur. J.* **2006**, *12*, 1216–1220.
[38] T. W. B. Ost, C. S. Miles, J. Murdoch, Y. F. Cheung, G. A. Reid, S. K. Chapman, A. W. Munro, *FEBS Lett.* **2000**, *486*, 173–177.
[39] W. T. Sulistyaningdyah, J. Ogawa, Q. S. Li, C. Maeda, Y. Yano, R. D. Schmid, S. Shimizu, *Appl. Microbiol. Biotechnol.* **2005**, *67*, 556–562.
[40] R. Fasan, M. M. Chen, N. C. Crook, F. H. Arnold, *Angew. Chem.* **2007**, *119*, 8566–8570; *Angew. Chem. Int. Ed.* **2007**, *46*, 8414–8418.
[41] F. M. Szczebara, C. Chandelier, C. Villeret, A. Masurel, S. Bourot, C. Duport, S. Blanchard, A. Groisillier, E. Testet, P. Costaglioli, G. Cauet, E. Degryse, D. Balbuena, J. Winter, T. Achstetter, R. Spagnoli, D. Pompon, B. Dumas, *Nat. Biotechnol.* **2003**, *21*, 143–149.
[42] J. B. van Beilen, R. Holtackers, D. Luscher, U. Bauer, B. Witholt, W. A. Duetz, *Appl. Environ. Microbiol.* **2005**, *71*, 1737–1744.
[43] G. W. Bemis, M. A. Murcko, *J. Med. Chem.* **1996**, *39*, 2887–2893.
[44] S. Nakagawa, S. Ishino, S. Teshiba, *Biosci. Biotechnol. Biochem.* **1996**, *60*, 415–420.
[45] H. J. Barnes, M. P. Arlotto, M. R. Waterman, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 5597–5601.
[46] T. Omura, R. Sato, *J. Biol. Chem.* **1964**, *239*, 2370–2378.

Received: March 10, 2009
Revised: August 11, 2009
Published online: September 22, 2009

