DOI: 10.1002/cbic.201000565

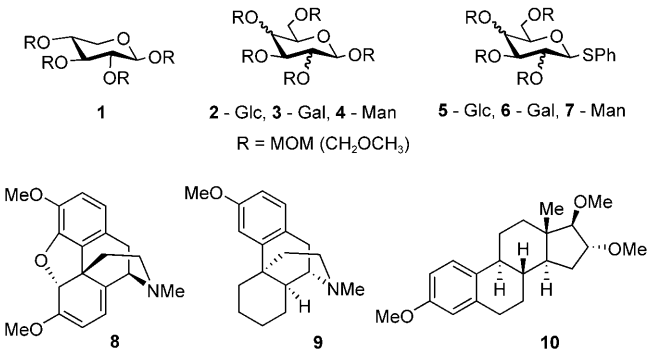
Combinatorial Alanine Substitution Enables Rapid Optimization of Cytochrome $P450_{BM3}$ for Selective Hydroxylation of Large Substrates

Jared C. Lewis,[d] Simone M. Mantovani,[e] Yu Fu,[b] Christopher D. Snow,[a] Russell S. Komor,[a] Chi-Huey Wong,[b, c] and Frances H. Arnold*[a]

Selective hydroxylation of C–H bonds in organic compounds provides an efficient means to access valuable drug metabolites, natural product derivatives, and other fine chemicals.[1] While chemical methods to accomplish this transformation have improved, these generally require the presence of directing groups or electronic properties inherent to certain substrate classes for the desired transformations to occur at all or with useful regioselectivity.[2] Enzymes are capable of avoiding this limitation by employing potent H-abstraction mechanisms with selectivity imposed by specific substrate binding.[3] Furthermore, systematic optimization of these catalysts by directed evolution has been demonstrated extensively and constitutes a powerful advantage of these systems over small molecule catalysts.[4]

Members of the cytochrome P450 monooxygenase superfamily are remarkable examples of such catalysts.[5] These enzymes utilize a cysteine-bound heme cofactor to catalyze a wide range of oxidative transformations including hydroxylation and epoxidation. Cytochrome $P450_{BM3}$ (BM3) from *Bacillus megaterium* possesses a number of features that make it particularly attractive for applications in chemical synthesis.[6] For example, the heme domain in which hydroxylation occurs and the diflavin reductase domains (FMN and FAD) that contribute electrons for oxygen activation in the heme domain are fused in a single polypeptide chain; this improves the rate and operational simplicity of BM3-catalyzed reactions. Indeed, BM3 catalyzes the subterminal hydroxylation of C_{12}–C_{18} fatty acids, its natural substrates, at rates of thousands of turnovers per minute, making it one of the most active hydroxylases known. BM3 is soluble, readily over-expressed in a variety of heterologous hosts, and requires only atmospheric oxygen and a supply of nicotinamide adenine dinucleotide phosphate (NADPH) for hydroxylation activity. These properties have led our groups and others to expand the substrate scope of this enzyme with the goal of creating efficient enzyme and whole-cell catalysts for a variety of oxidative transformations.[7]

The development of BM3 variants that catalyze regioselective demethylation or demethoxymethylation of methylated or methoxymethylated monosaccharides was recently reported by our groups.[8] While a chemoenzymatic method employing these enzymes provided a convenient means to access otherwise difficult-to-synthesize monosaccharide derivatives, it also highlighted some limitations of existing BM3 variants. For example, while MOM-protected pentoses were compatible with these enzymes, MOM-protected hexoses were deprotected to only a minor extent (Scheme 1). This same limitation was



Scheme 1. Structures of compounds utilized in enzyme library design and screening: methoxymethyl (MOM)-protected xylose, hexoses, and thioglycosides (**1**–**7**); alkaloids thebaine and dextromethorphan (**8** and **9**); trimethyl estriol (**10**).

encountered with a number of additional bulky compounds including various opiate alkaloids and steroid derivatives. Attempts to identify enzymes compatible with these substrates from random mutant libraries generated by error prone PCR were complicated by the low activity of all BM3 variants examined. We hypothesized that the shape and volume of these substrates exceeded the capacity of the enzyme active site, and that sufficient expansion of the active site to obtain significant improvements in activity might require more extensive mutation than can be readily accomplished with error prone PCR.[9]

An alternative approach involving extensive replacement of bulky active site residues with alanine was, therefore, explored as a means to obtain enzymes with activity against the aforementioned compounds. This involved first selecting a thermostable parent capable of tolerating extensive mutagenesis,[10]

[a] Dr. C. D. Snow, R. S. Komor, Prof. F. H. Arnold
Division of Chemistry and Chemical Engineering
California Institute of Technology, Pasadena, CA 91125 (USA)
Fax: (+1) 626-568-8743
E-mail: arnold@cheme.caltech.edu

[b] Y. Fu, Prof. C.-H. Wong
Department of Chemistry, The Scripps Research Institute
La Jolla, CA, 92037 (USA)

[c] Prof. C.-H. Wong
Genomics Research Center, Academia Sinica
Nankang, Taipei 115 (Taiwan)

[d] Dr. J. C. Lewis
Current address: Department of Chemistry
University of Chicago, Chicago, IL 60637 (USA)

[e] S. M. Mantovani
Current Address: Institute of Chemistry, University of Campinas
P.O. Box 6154, 13083-970 Campinas SP (Brazil)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.201000565.

WILEY ONLINE LIBRARY
© 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

and identifying active site residues likely to clash with large substrates based on computational models of these compounds docked in the enzyme active site. Because it is difficult to predict which of the chosen residues might actually impact catalysis,[11] a combinatorial library design that includes all permutations of alanine substitutions was utilized. Library members were then screened for activity with a panel of bulky substrates that possess methylated heteroatoms by detecting formaldehyde released during P450-catalyzed removal of Me or MOM groups.[8] These substrates thus serve as efficient probes for enzyme activity and for compatibility of valuable natural product scaffolds with BM3 variants.

BM3 variant 9-10A F87V TS, a thermostable form of an enzyme (9-10A F87V) previously found to have activity against α-1,2,3,4-tetramethoxymethyl xylose (**1**),[8] was selected as a parent for library creation. Structures for the most stable conformers of **1** and previously unreactive monosaccharide substrates, including 1,2,3,4,6-pentamethoxymethyl glucose (**2**), galactose (**3**), and mannose (**4**), were generated by using the program Omega.[12] The resulting conformers were then placed into the active site of a model of 9-10A F87V TS such that their terminal methyl C–H bonds were oriented according to the transition state geometry for cytochrome P450-catalyzed H-abstraction proposed by Rydberg et al.[13] The transition state ensemble was expanded by varying rotational degrees of freedom and reduced by eliminating substrate poses that clashed with the enzyme (within 2.5 Å of a backbone or β carbon). Inspection of the final ensemble of prospective transition state conformations led to the identification of eight residues for replacement with alanine: K69, L75, M177, L181, T260, I263, T268, and L437. Combinatorial substitution was accomplished by cloning fragments of the parent gene containing these residues by using degenerate primers or primer mixtures encoding either the original residue or alanine at each site (see the Supporting Information). These fragments were assembled to generate a 2^8 (256) member library containing the desired alanine substitutions. *E. coli* were transformed with the library mixture and 767 single colonies (ca. 3× theoretical library size) were picked and used to inoculate media in 96-well deep-well plates. Following protein expression and cell lysis, CO binding analysis revealed that 65% of the enzymes were properly folded.[14] Sequences from 1% of the library indicated unbiased incorporation of alanine at all of the desired sites, and an average alanine substitution per sequence of 3.9.

Several compounds, including MOM-protected thioglycosides **5**, **6**, and **7**, thebaine (**8**), dextromethorphan (**9**),[15] and trimethyl estriol (**10**; Scheme 1), were selected to probe the substrate scope of the library. Each of these is relatively large, possesses several methylated heteroatoms, and belongs to a privileged class of compounds (i.e., monosaccharides, alkaloids, and steroids), for which novel hydroxylation catalysts could be highly valuable (vide infra). High-throughput screening was carried out in 96-well microtiter plate format[8] by addition of NADPH to solutions of the appropriate substrate and cell lysate. Following incubation at room temperature, the reactions were quenched with a basic solution of Purpald, which reacts with the formaldehyde produced from P450-catalyzed heteroatom demethylation to generate a purple dye. The parent enzyme, 9-10A F87V TS, displayed only weak activity with these substrates, and provided a signal of 20% over background in the best cases (data not shown). However, many variants with high activity against several substrates were immediately evident based on visual inspection, and the formaldehyde concentration (and thus the extent of demethylation) was quantified by using a plate reader ($A_{550\,nm}$).

Variants with 4.1–7.9-fold improvement in demethoxymethylation activity toward each of the MOM-protected hexose derivatives were identified (Table 1). Furthermore, GC or HPLC analysis of the crude mixtures indicated that these reactions proceeded in moderate to high conversion while still maintaining high regioselectivity (vide infra). On the other hand, few variants with improved activity against the smaller pentose substrate, **1**, were obtained from this library; this is consistent with the expanded active site providing an advantage only for the reaction of larger substrates (see the Supporting Information). This library also contained variants with marked improvements in activity with alkaloid and steroid substrates **8**, **9**, and **10**, despite the fact that such structures were not used in the library design. For most substrates, the best variants possessed at least two alanine substitutions, and in general, positions 75, 177, 181, 260, 263, and 437 proved advantageous, while alanine substitution at 69, 263, and 268 did not.

Given that these variants possessed sufficient activity with substrates **2**–**10** to enable their detection with the aforemen-

Table 1. Sequence–activity relationship for alanine substitution.

Substrate	Variant	Alanine substitution (+) at residue								Fold improvement[a]
		69	75	177	181	260	263	268	437	
thioglycosides (**5**–**7**)	2A1	–	+	–	+	–	–	–	+	4.4
	4H9	–	–	–	+	+	–	–	+	4.1
	8C7	–	+	–	+	–	–	–	–	7.9
alkaloids (**8**, **9**)	4H5	–	+	+	+	–	–	–	–	2.7
	4H9	–	–	–	+	+	–	–	+	3.9
	7A1	–	+	+	+	+	–	–	–	2.8
	8C7	–	+	–	+	–	–	–	–	2.7
	8F11	–	–	–	–	–	–	–	+	3.3
steroid (**10**)	8F11	–	–	–	–	–	–	–	+	n.a.[b]

[a] Ratio of A_{550} measurement for reaction of each variant to that of the parent (9-10A F87V TS). Ratio for substrate with maximum improvement shown. [b] Improvement identified by visual inspection; substrate insolubility complicated plate reader measurement.

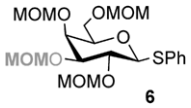
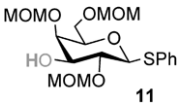
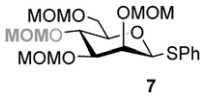
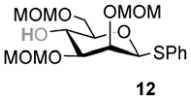
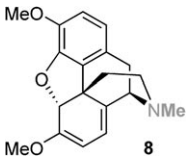
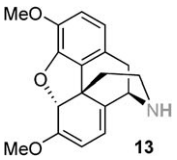
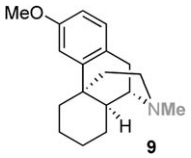
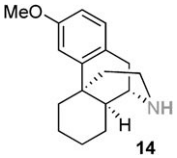
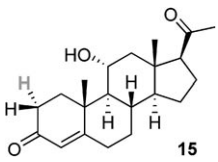
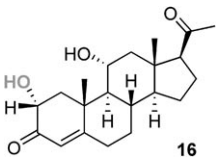
 © 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

tioned high-throughput screen, further enzyme optimization was possible by using directed evolution.[4] Steroid hydroxylation is a particularly valuable reaction due to the biological activity of these compounds and their common occurrence as metabolites.[16] Indeed, native P450s have been used for many years in the preparation of metabolites for both analytical studies and pharmaceutical manufacture.[17] Because only a single variant, 8F11, exhibited activity with the steroid derivative trimethyl estriol, a library of variants of this enzyme was generated by using error prone PCR and screened for improved demethylation of trimethyl estriol (see the Supporting Information). This led to the identification of a variant with four mutations, none of which was located in the enzyme active site, with 1.6-fold improvement in activity over the 8F11. This enzyme, F1, was found to have moderate activity with additional steroids, including 11-α-hydroxyprogesterone (vide infra) and testosterone acetate (data not shown). This result indicated that heteroatom-methylated substrates could be used as convenient probes for BM3 activity on related compounds lacking these handles for high-throughput analysis.

To demonstrate the synthesis utility of the enzymes obtained from combinatorial alanine substitution and error prone PCR, preparative scale bioconversions were conducted (Table 2). Reaction conditions previously developed in our laboratory were utilized;[8] yields and selectivities of the reactions translated well to the preparative scale. For example, site-selective deprotection of MOM-protected hexoses **6** and **7** proceeded in 75 and 70% yield, respectively (Table 2, entries 1 and 2). BM3-catalyzed deprotection of such substrates significantly expands the utility of our previously reported chemoenzymatic monosaccharide elaboration procedure[8] due to the mild conditions required for chemical deprotection of MOM groups and the potential use of the thiophenyl substituent as a leaving group in subsequent glycosylation reactions. Demethylation and N-substitution of opiate alkaloids is commonly used to vary the properties of these compounds,[18] and BM3 catalysis provides a mild and operationally simple method to accomplish the required demethylation step (Table 2, entries 3 and 4). Finally, selective hydroxylation of multifunctional molecules represents a great challenge in synthetic chemistry.[1] While P450-catalyzed hydroxylation of steroids has been demonstrated, most of these reactions require the use of whole-cell biocatalysts[16] or multicomponent enzyme systems.[19] On the other hand, BM3 variant F1 enabled regio- and diastereoselective steroid hydroxylation by using a single enzyme and thus provided a convenient platform from which additional catalysts for this valuable transformation can be developed (Table 2, entry 5).

Together, these results demonstrate the utility of combinatorial alanine substitution for generation of BM3 variants with activity against bulky, synthetically useful substrates. We have demonstrated that the resulting enzymes have novel activities that can be further optimized by directed evolution. Monosaccharides, alkaloids, and steroids were all viable substrates de-

Table 2. Substrate scope and reaction selectivity.

	Substrate	BM3 variant	Product	Conv. [%][a]	Select. [%][b]	Yield [%][c]
1	 **6**	2A1	 **11**	80	90	75
2	 **7**	8C7	 **12**	93	80	70
3	 **8**	8C7	 **13**	88	72	60
4	 **9**	4H5	 **14**	54	98	50
5	 **15**	F1	 **16**	28	82	20

[a] Conversion of starting material determined by HPLC or GC analysis of crude reaction mixture. [b] Percentage of desired product relative to additional products determined by HPLC or GC analysis of crude reaction extracts. [c] Isolated yield of the pure product.

 © 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

spite their large size. This approach could be useful for improving activity of other enzymes for which substrate size appears to limit the scope of their reactivity.

Experimental Section

For details see the Supporting Information.

Acknowledgements

J.C.L. is supported by a U.S. National Institutes of Health Pathways to Independence Award (1K99M087551-01A1). S.M.M. is supported by the Fundação Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES; 1756-09-5). This work was supported by the U.S. National Institutes of Health (2R01 M068664-05A1), the U.S. Department of Energy, Office of Basic Science, grant DE-FG02-06ER15762, and King Abdullah University of Science and Technology (KAUST), Award No. KUS-F1-028-03.

Keywords: alkaloids · biocatalysis · cytochromes · monosaccharides · steroids

[1] D. E. Torres Pazmiño, M. Winkler, A. Glieder, M. W. Fraaije, *J. Biotechnol.* **2010**, *146*, 9 – 24.
[2] J. C. Lewis, R. G. Bergman, J. A. Ellman, *Acc. Chem. Res.* **2008**, *41*, 1013 – 1025.
[3] P. A. Frey, *Chem. Rev.* **1990**, *90*, 1343 – 1357.
[4] F. H. Arnold, *Acc. Chem. Res.* **1998**, *31*, 125 – 131.
[5] *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Ed.: P. R. Ortiz de Montellano), Kluwer, New York, **2005**.
[6] A. J. Warman, O. Roitel, R. Neeli, H. M. Girvan, H. E. Seward, S. A. Murray, K. J. McLean, M. G. Joyce, H. Toogood, R. A. Holt, D. Leys, N. S. Scrutton, A. W. Munro, *Biochem. Soc. Trans.* **2005**, *33*, 747 – 753.
[7] J. C. Lewis, F. H. Arnold, *Chimia* **2009**, *63*, 309 – 312.
[8] J. C. Lewis, S. Bastian, C. S. Bennett, Y. Fu, Y. Mitsuda, M. M. Chen, W. A. Greenberg, C. H. Wong, F. H. Arnold, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 16550 – 16555.
[9] T. P. Treynor, C. L. Vizcarra, D. Nedelcu, S. L. Mayo, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 48 – 53.
[10] J. D. Bloom, S. T. Labthavikul, C. R. Otey, F. H. Arnold, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 5869 – 5874.
[11] C. A. Tracewell, F. H. Arnold, *Curr. Opin. Chem. Biol.* **2009**, *13*, 3 – 9.
[12] J. Boström, J. R. Greenwood, J. Gottfries, *J. Mol. Graphics Modell.* **2003**, *21*, 449 – 462.
[13] P. Rydberg, L. Olsen, P. O. Norrby, U. Ryde, *J. Chem. Theory Comput.* **2007**, *3*, 1765 – 1773.
[14] T. Omura, R. Sato, *J. Biol. Chem.* **1964**, *239*, 2370 – 2378.
[15] 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.
[16] T. Furuya, D. Shibata, K. Kino, *Steroids* **2009**, *74*, 906 – 912.
[17] *Steroid Hydroxylation: Microbial Steroid Biotransformations Using Cytochrome P450 Enzymes, M. Bureik, R. Bernhardt in Modern Biooxidation: Enzymes, Reactions, and Applications* (Eds.: R. D. Schmid, V. B. Urlacher), Wiley-VCH, **2007**, pp. 155 – 176.
[18] K. McCamley, J. A. Ripper, R. D. Singer, P. J. Scammells, *J. Org. Chem.* **2003**, *68*, 9847 – 9850.
[19] D. Zehentgruber, F. Hannemann, S. Bleif, R. Bernhardt, S. Lutz, *ChemBioChem* **2010**, *11*, 713 – 721.

Received: September 21, 2010
Published online on November 24, 2010

 © 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim