Solvay Conferences in Chemistry: Expanding the Protein Universe
Brussels, October 2013

EXPANDING THE ENZYME UNIVERSE THROUGH A MARRIAGE OF CHEMISTRY AND EVOLUTION

FRANCES H. ARNOLD
Division of Chemistry and Chemical Engineering, California Institute of Technology 210-41, 1200 E. California Blvd., Pasadena CA 91125, USA

Present state of research on expanding enzyme catalysis beyond nature

For more than twenty years this laboratory has used directed evolution to modify enzymes. It is now widely accepted that directed evolution can change substrate specificity or reaction selectivity in desired ways, even if it sometimes remains difficult in practice. It is no longer surprising that enzymes adapt readily by accumulating beneficial mutations. And why should it be, since this is how nature tailors them for myriad biological roles? More difficult to grasp is how nature discovers new enzyme functions, particularly new catalytic activities. We know that the biological world's diverse catalytic repertoire is the product of evolution by natural selection, but we have little understanding of how nature's tinkering generates new functions. Sometimes we are (un)lucky enough to catch them in the act—e.g. the acquisition of antibiotic resistance or the ability to degrade man-made toxins. But for the vast majority of activities, the fossil record is nonexistent or too sparse to tell the molecular story. This leaves us without much guidance for evolving new enzymes in the laboratory. Consequently we are forced to contaminate our evolution experiments with knowledge—e.g. computational design [1,2]—in order to jumpstart the discovery process.

If every bad catalyst could become a good one by directed evolution, part of the problem would be solved. We would be able to take any catalytic antibody or computationally-designed enzyme (or bovine serum albumin) and convert it into a great catalyst with multiple rounds of random mutagenesis and screening. We have learned the hard way, however, that not every bad catalyst lies at the base of a tall fitness peak, at least one that can be scaled by a random uphill walk. We therefore want to know what features make a protein with a new catalytic activity the potential mother of a whole new enzyme family. And, what are good ways to find new enzyme activities in the first place?

Recent research contributions to creating new enzymes

My laboratory has been directing the evolution of a remarkable enzyme, a bacterial cytochrome P450, for at least a dozen years. This particular P450, from *Bacillus megaterium*, is quite specific—it catalyzes subterminal hydroxylation of fatty acids. As a family, however, the P450s are wonderfully diverse, catalyzing a wide range of reactions

on an even wider range of substrates. Nature has demonstrated the 'evolvability' of this bit of protein that binds an iron-heme (along with its various electron-transfer partners), and we spent several happy years creating new versions of the bacterial enzyme that could mimic known functions of other (e.g. human) P450 family members [3,4]. We could also extend its function beyond what was known to be catalyzed by P450s, making, for example, versions that hydroxylate gaseous alkanes (propane and ethane), transformations thought to lie in the functional realm of the methane monooxygenase family [5].

The versatile iron-heme prosthetic group has been adopted by numerous transport, signaling, and catalytic proteins. Within just the P450 family the range of reactions catalyzed is impressive: hydroxylation, epoxidation, sulfoxidation, peroxidation, N- and O-dealkylation, and more. We found that we could access all of these reactivities, starting from wildtype or variants of our bacterial fatty acid monooxygenase and using directed evolution to increase the initially-low activities. We learned that even a single P450 is highly evolvable and lies at the base of many different fitness peaks. The P450 is also catalytically promiscuous—the bacterial enzyme or its close variants could catalyze low levels of many different reactions. Presumably natural selection made the same discovery many times, setting old P450s to new tasks in many new contexts.

More recently we decided to explore new reactions, not known in nature. Reactions whose mechanisms share features of the P450 machinery but have not been discovered in nature's many evolution experiments, because the context was not there. Understanding that natural selection can create new enzymes from promiscuous activities when presented with the opportunity to occupy a new niche (for example, a new substrate becomes available), we tested a collection of P450s for promiscuous activity in a few carbene and nitrene insertion reactions that are isoelectronic to the well-established formal 'oxene' transfer reaction of ferric-P450 enzymes with iodosylbenzene. The similarity is such that some of these reactions were explored in the 1980's by 'biomimetic' chemists, notably Breslow and Dawson [6] and others [7]. The poor turnover numbers they reported for P450s, however, discouraged further work, and no more was said for 30 years. But we knew from experience that when a P450 is a bad catalyst, it can often easily become a good one.

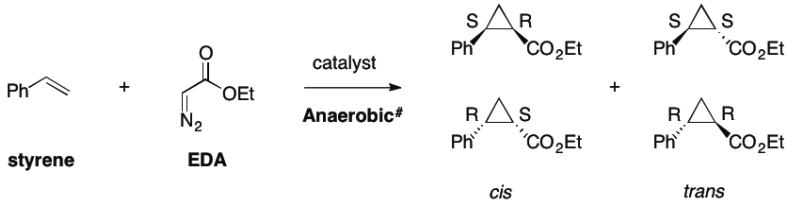
The wildtype P450 enzyme catalyzed just a few turnovers for cyclopropanation of styrene with EDA, fewer even than free hemin (Table 1) and other heme proteins [8]. However, it was the only catalyst to exhibit enantioselectivity, indicating the reaction took place in an active site that could exert some control on selectivity. We quickly identified variants in our collection, such as H2A10 and CIS (Table 1), that exhibited high selectivity for the *cis* diastereomer (opposite from that of hemin, which makes mostly the *trans* cyclopropane product), high activity, and high enantioselectivity.

The big breakthrough came, however, when we replaced the cysteine residue that ligates the heme iron with Ser [9]. Our goal was to have the reaction proceed *in vivo*, where it would have to rely on endogenous NADPH rather than the sodium dithionite used to reduce the heme Fe(III) to Fe(II) *in vitro*. This generated a very active

cyclopropanation enzyme that functions extremely well *in vivo*: the 67,800 turnovers for the P411-CIS enzyme (Table 1) is, we believe, the highest activity ever reported for this reaction with any catalyst. The Cys-Ser mutation abolished all monooxygenase activity and caused the typical peak at 450 nm in the CO-difference spectrum to shift to 411 nm. Thus we call this new catalyst a P411. Much more active than P450-CIS, the P411-CIS has a crystal structure nearly identical to that of P450-CIS. We have started to diversify this enzyme by directed evolution to expand its substrate range and selectivity.

Olefin cyclopropanation is not (yet!) a biologically relevant transformation, because P450s do not encounter the reactive diazoesters in their native environments. They nonetheless have this promiscuous activity, which can be captured by evolution when the opportunity arises. Our work shows that this promiscuous activity can be enhanced significantly with just a few mutations, something that can happen readily in a protein evolving under selective pressure, either natural or forced.

Table 1. Activities of different P450 variants for styrene cyclopropanation [9]. Yields are based on EDA. TTN = total turnover number.



Catalyst	[EDA] (mM)	[P450] (μM)	% yield	TTN	cis:trans	%ee$_{cis}$	%ee$_{trans}$
Hemin	10	20	15	73	6:94	1	0
P450$_{BM3}$	10	20	1	5	37:63	27	2
9-10A TS F87V	10	20	1	7	35:65	41	8
H2A10	10	20	33	167	60:40	95	78
P450$_{BM3\text{-}heme}$-CIS	8.5	15	32	212	77:23	94	91
P411$_{BM3\text{-}heme}$-CIS	8.5	15	51	342	93:7	99	51
P450$_{BM3}$-CIS*	8.5	3.7	42	950	22:78	60	22
P411$_{BM3}$-CIS*	8.5	1.3	55	3700	76:24	96	25
P411$_{BM3}$-CIS*	170	1.8	72	67800	90:10	99	43

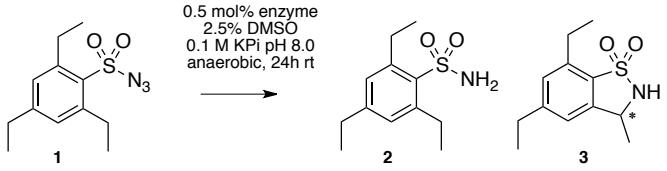
*conducted with intact *E. coli* cells.

#Conditions for reactions with purified P450s: 1 equiv $Na_2S_2O_4$, Ar atmosphere, 0.1 M KPi pH 8.0. Conditions for reactions with intact *E. coli* cells: 0.2 equiv glucose, Ar atmosphere, M9-N medium.

Enzymes that catalyze the concerted oxidative amination of C-H bonds are also apparently absent from nature's catalyst repertoire. Synthetic chemists, who are not limited to biologically accessible reagents and metals, have developed useful methods for C-H amination through a nitrenoid intermediate that also has no parallel in natural enzymes. Following up on studies performed in the 1980's [6], we investigated whether our P450 and P411 enzymes could catalyze intramolecular C-H amination of aryl sulfonylazides to form benzosultams (example shown in Table 2) [10]. Whereas wildtype P450 showed only weak activity, some P411 variants catalyzed several hundred TTN.

The purified enzymes as well as intact *E. coli* cells expressing the enzymes catalyze the amination reaction under anaerobic conditions (Table 2) [10]. P411-T268A and P411-CIS exhibited good activity and reasonable enantioselectivity (up to 89%); adding the T438S mutation to P411-CIS increased enantioselectivity (430 TTN, 86% *ee*). Optimization of expression conditions increased the productivity of whole-cell C-H amination, enabling conversions to **3** of up to 66% in small-scale reactions; higher yields have since been achieved at preparative scale.

Table 2. P450 and P411 enzymes catalyze direct C-H amination. Comparison of total turnover numbers (TTN) and enantioselectivities of intact *E. coli* cells expressing P450 and P411 variants, with azide **1** at 0.1 mol% catalyst loading, giving sulfonamide **2** and benzosultam **3** [10].



In vivo catalyst	[P450] or [P411] [(μM]	Yield 3 [%]	TTN[a]	ee [%][b]
Empty vector	0	0	0	n.d.
$P450_{BM3}$	6.6	0.5	5.1	n.d
$P450_{BM3}$-	5.8	7.8	26	84
$P411_{BM3}$	4.3	6.7	29	16
$P411_{BM3}$-	2.2	30	250	89
$P411_{BM3}$-CIS	1.4	46	680	60
$P411_{BM3}$-CIS- T438S	2.7	58	430	87

[a] TTN＝total turnover number to benzosultam **3**. [b] *(S-R)/(S+R). n.d.＝not determined.

Outlook for future enzymes

The tools are now in place to create enzymes that catalyze reactions not known in nature. We have reported olefin cyclopropanation and C-H amination catalyzed by P450s and P411s, and there is clearly opportunity for more useful reactions based on this system. The highly evolvable P450 is an excellent starting point, both for discovering new enzyme-catalyzed reactions and for diversifying them through directed evolution. But there are likely many more such enzymes waiting for the right substrates to come along. A tasteful mix of chemical intuition, computational design where appropriate, and evolution (to circumvent our near complete ignorance of the details of the sequence-function code for catalysis) will generate whole new families of genetically-encoded catalysts, greatly expanding the catalyst repertoire for biosynthesis and for organic synthesis. What is more, we will be able to observe the creation of new biological functions and follow the mechanisms by which they arise and are diversified and

optimized. The future will see nature's chemical universe expand to include more of the clever chemistry of man.

Acknowledgments

I have described the work of highly talented graduate students and postdocs: Pedro Coelho, Eric Brustad, John McIntosh, Jane Wang and Chris Farwell. Support was received from the Jacobs Institute for Molecular Engineering for Medicine at Caltech, and the Gordon and Betty Moore Foundation through Grant GBMF2809 to the Caltech Programmable Molecular Technology Initiative.

References

1. F. Richter, R. Blomberg, S. D. Khare, G. Kiss, A. P. Kuzin, A. J. T. Smith, J. Gallaher, Z. Pianowski, R. C. Helgeson, A. Grjasnow, R. Xiao, J. Seetharaman, M. Su, S. Vorobiev, S. Lew, F. Forouhar, G. J. Kornhaber, J. F. Hunt, G. T. Montelione, L. Tong, K. N. Houk, D. Hilvert, D. Baker, *J. Am. Chem. Soc.* **134**, 16197 (2012).
2. L. Giger, S. Caner, R. Obexer, P. Kast, D. Baker, N. Ban, D. Hilvert, *Nat. Chem. Biol.* **19**, 494 (2013).
3. M. Landwehr, L. Hochrein, C. R. Otey, A. Kasrayan, J.-E. Bäckvall, F. H. Arnold, *J. Am. Chem. Soc.* **128**, 6058 (2006)
4. A. M. Sawayama, M. M. Y. Chen, P. Kulanthaivel, M-S. Kuo, H. Hemmerle, F. H. Arnold, *Chemistry-A European Journal* **15**, 11723 (2009).
5. R. Fasan, M. M. Chen, N. C. Crook, F. H. Arnold, *Angew. Chem.* **46**, 8414-8418 (2007).
6. E. W. Svastis, J. H. Dawson, R. Breslow, S. H. Gellman, *J. Am. Chem. Soc.* **107**, 6427 (1985).
7. J. R. Wolf, C. G. Hamaker, J.-P. Djukic, T. Kodadek, L. K. Woo, *J. Am. Chem. Soc.* **117**, 9194 (1995).
8. P. S. Coelho, E. M. Brustad, A. Kannan, F. H. Arnold, *Science* **339**, 307 (2013).
9. P. S. Coelho, Z. J. Wang, M. E. Ener, S. A. Baril, A. Kannan, F. H. Arnold, E. M. Brustad, *Nat. Chem. Biol.* **9**, 485 (2013).
10. J. A. McIntosh, P. S. Coelho, C. C. Farwell, Z. J. Wang, J. C. Lewis, T. R. Brown, F. H. Arnold, *Angew. Chem. Int. Ed.* **52**, published online July 24, 2013.