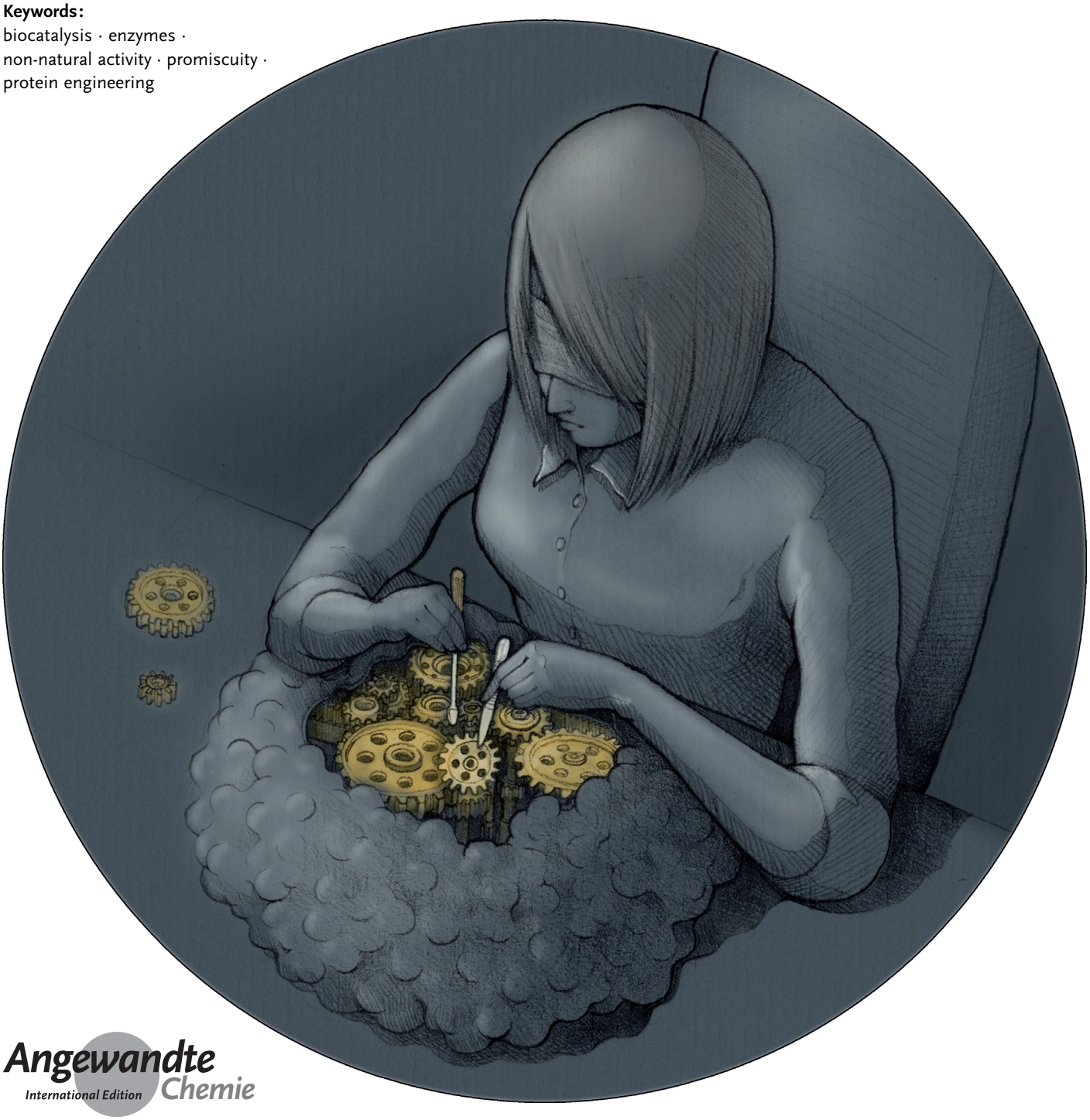
Biocatalysis | Special Issue 150 Years of BASF

DOI: 10.1002/anie.201409470

Expanding the Enzyme Universe: Accessing Non-Natural Reactions by Mechanism-Guided Directed Evolution

*Hans Renata, Z. Jane Wang, and Frances H. Arnold**

Keywords:
biocatalysis · enzymes · non-natural activity · promiscuity · protein engineering



Angewandte International Edition Chemie

High selectivity and exquisite control over the outcome of reactions entice chemists to use biocatalysts in organic synthesis. However, many useful reactions are not accessible because they are not in nature's known repertoire. In this Review, we outline an evolutionary approach to engineering enzymes to catalyze reactions not found in nature. We begin with examples of how nature has discovered new catalytic functions and how such evolutionary progression has been recapitulated in the laboratory starting from extant enzymes. We then examine non-native enzyme activities that have been exploited for chemical synthesis, with an emphasis on reactions that do not have natural counterparts. Non-natural activities can be improved by directed evolution, thus mimicking the process used by nature to create new catalysts. Finally, we describe the discovery of non-native catalytic functions that may provide future opportunities for the expansion of the enzyme universe.

From the Contents

1. Introduction

Replete with nature's solutions to the catalysis of chemical transformations, our burgeoning genomic databases beautifully illustrate how evolution generates chemical innovation in the form of new enzymes. The enormous biocatalytic diversity that we know today is the product of evolution from ancestral enzymes, the mechanisms of which are now being elucidated in unprecedented detail. Enzyme evolution is also alive and well and moving into the future: New enzymes continue to appear in response to opportunities to survive (often man-made) challenges (e.g. antibiotic resistance) or occupy new niches (e.g. catabolism of man-made compounds). Given nature's ability to innovate and our extremely limited ability to design new enzymes, we argue for the use of evolutionary strategies to create and tune enzymes fit for human applications.

Directed evolution is a powerful protein-engineering approach that has been applied with great success for nearly two decades to the fine-tuning of enzymes for chemical synthesis.[1] A simple strategy of accumulating mutations through iterative mutagenesis and screening for desired functions can effectively optimize properties of interest, such as activity on non-native substrates, enantioselectivity, product selectivity, and stability, and deftly circumvents our profound ignorance of how the enzyme sequence encodes these features. If chemists are to use enzymes in synthesis, these catalysts have to perform as well as or better than the alternatives. Often this bar is set very high, and no "rational-design" approach has been able to meet it on a regular basis; however, directed evolution performs well, given a good starting point.[2] As a result, enzymes are increasingly used in chemical synthesis, where they offer significant advantages for "green" processes,[3] the production of chemicals from renewable resources,[4] and the synthesis of complex natural products.

Directed evolution can be used to diversify existing enzymes by creating variants that function in non-native environments, accept non-native substrates, or exhibit non-native selectivities. But how do we create whole new enzymes, including enzymes that catalyze reactions not known in nature? Nature's catalyst reserve is vast and has not been fully mined; new enzymes will continue to be discovered. Chemists, however, are fond of a number of reactions for which there may well be no natural biocatalytic counterparts, either because nature has not discovered a need for them (our goals and requirements being different from those of a microbe or a tree) or because they require functional groups and reagents not normally found in the biological world. Our goal is to begin to address this gap between the enzymes we can find in nature and those we would like to have but may not exist. We believe we can use what we have learned about nature's approach to inventing new catalysts to jumpstart the evolution of new enzyme families in the laboratory. We first illustrate this evolutionary approach to catalyst discovery with some examples from nature's repertoire. In some cases, scientists have elucidated sequences and functional pathways that connect existing enzymes in order to demonstrate how natural function may have evolved from one to the other.

Can this knowledge help us step out into the unknown and create biocatalysts that have not yet been discovered in the natural world? The answer is an emphatic "Yes!", as illustrated by several powerful examples of enzymes engineered to catalyze reactions with no known natural counterparts. We end by describing a few non-natural activities that might afford a peek into future enzyme families.

[*] Dr. H. Renata, Dr. Z. J. Wang, Prof. Dr. F. H. Arnold
Division of Chemistry and Chemical Engineering
California Institute of Technology
1200 E. California Blvd. MC 210-41, Pasadena, CA 91125 (USA)
E-mail: fha@cheme.caltech.edu

2. Nature's Approach to Generating New Enzymes

Catalytic promiscuity refers to the ability of an enzyme to catalyze, in addition to its native function, reactions that target different functional groups on the substrate and proceed via different transition states and/or reactive intermediates. In 1976, Jensen proposed that ancient enzymes were characterized by broad substrate and reaction scope and that natural selection picked up and fine-tuned these different activities to generate contemporary enzymes with specific catalytic functions (Figure 1 A).[5] However, even today, enzymes are not as specific as often thought: Many can catalyze other transformations in their active sites and exhibit (usually low levels of) catalytic promiscuity.[6] Much evidence now suggests that this often serendipitous catalytic promiscuity is in fact vital to the evolution of new enzymes, as it provides a platform for the evolution of new functions by natural selection (Figure 1 B).[7]

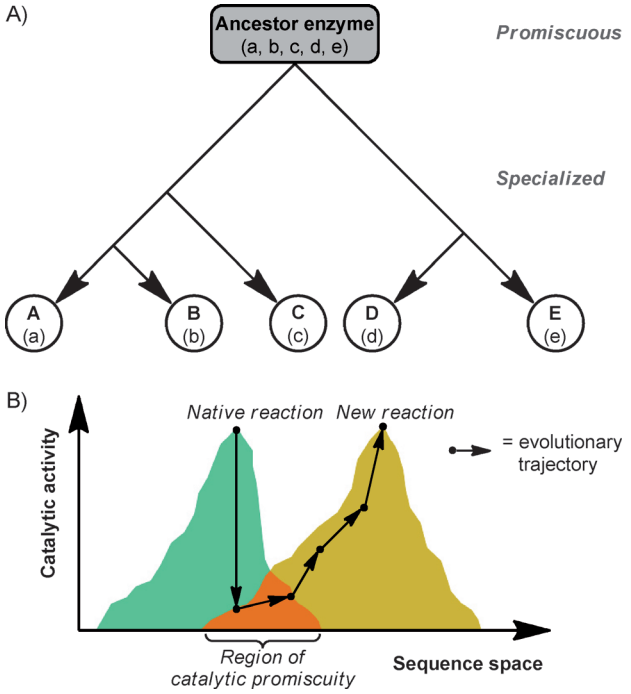


Figure 1. A) Divergence of an ancestral enzyme with broad catalytic capabilities (denoted a, b, c, d, e) to more specialized enzymes (denoted A, B, C, D, E) that catalyze primarily one reaction. B) Relationship between catalytic promiscuity and the evolution of a new function. A given protein sequence might catalyze multiple reactions. Under the right circumstances, a catalyst with a low level of a promiscuous activity can be improved by mutation and natural (or artificial) selection so that it becomes specialized for a new function. For further discussion, see Ref. [8].

2.1. Evolution of Atrazine Chlorohydrolase (AtzA)

The evolution of atrazine chlorohydrolase (AtzA) is one of the best case studies of how nature exploits catalytic promiscuity to create new enzymes. A potent herbicide introduced in the late 1950s, atrazine was initially found to be minimally biodegradable. Since 1993, however, atrazine has been observed to be degraded rapidly by soil microbes in diverse locales, a phenomenon attributed to the presence of the enzyme AtzA.[9] This enzyme catalyzes the hydrolysis of the C–Cl bond of atrazine (Figure 2 A) through a nucleophilic aromatic substitution reaction with a Fe^{2+}-activated water molecule.

The amino acid sequence of AtzA from *Pseudomonas* sp. ADP is 98% identical to that of melamine deaminase (TriA), an enzyme originally isolated from *Pseudomonas* sp. strain NRRL B-12227. TriA catalyzes the hydrolysis of the C–N bond of melamine, a non-natural compound that was originally classified as non-biodegradable in the 1930s, but was slightly degradable by the time atrazine was first introduced.[10] The two enzymes differ in only 9 out of 475 amino acids.[11] This extremely high level of identity and the fact that both enzymes can be found in at least one common bacterial species suggested that AtzA evolved from TriA or from a common ancestral enzyme similar to TriA (Figure 2 B), thus enabling the bacteria to capitalize on a new opportunity to use these synthetic compounds as nitrogen sources.[12]

Progression from melamine to triazine degradation indicates the evolution of a new catalytic function, from C–N bond cleavage (aminohydrolase, EC 3.5.4) to C–Cl cleavage (chlorohydrolyase, EC 3.8). Although other chlorohydrolases exist in nature, they typically use a carboxylate nucleophile instead of water activated by a divalent metal, as found for AtzA. Enzymes in the aminohydrolase (e.g. TriA) family

Hans Renata received his BA degree from Columbia University in 2008, for which he conducted research under the supervision of Professor Tristan H. Lambert. He earned his PhD in chemistry from The Scripps Research Institute in 2013 under the guidance of Professor Phil S. Baran. He is now carrying out postdoctoral studies in the research group of Professor Frances H. Arnold at the California Institute of Technology.

Z. Jane Wang received her BS degree in chemistry from the California Institute of Technology in 2007. She obtained her PhD in organic and organometallic chemistry with Professor F. Dean Toste and Professor R. Bergman at the University of California, Berkeley. Her research focused on the development of new methods for forming C–O and C–N bonds with coinage-metal complexes. She recently completed her postdoctoral studies at the California Institute of Technology with Professor Frances H. Arnold.

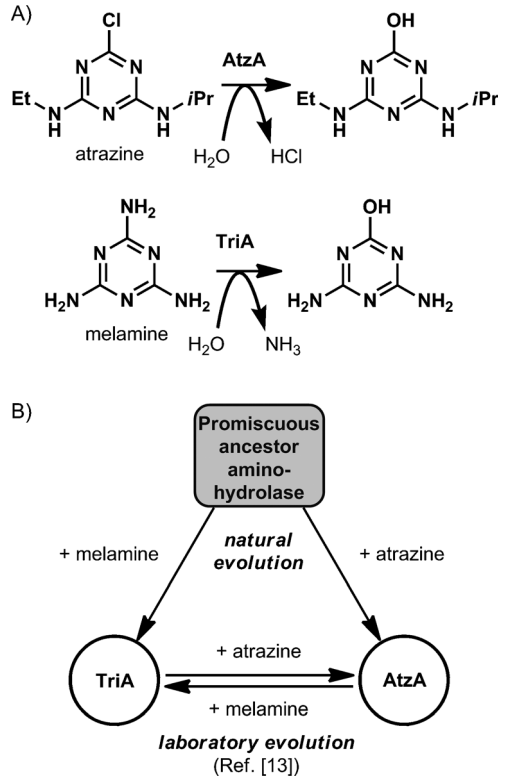


Figure 2. A) Chlorohydrolase activity of AtzA and aminohydrolase activity of TriA. B) TriA and AtzA (98% AA identity) are believed to be related through a common ancestor similar to TriA. These catalytic functions can be interconverted with a few amino acid mutations.[13]

typically utilize a divalent metal, such as the Zn^{2+} cation, to activate water for nucleophilic aromatic substitution. The active-site similarity between TriA and AtzA, as suggested by homology modeling, further corroborates the conjecture that AtzA evolved from an ancestral aminohydrolase (Figure 2B).

TriA and AtzA differ at 9 positions, but Scott and co-workers recently showed that two mutations are sufficient to convert TriA into an enzyme with atrazine chlorohydrolase activity comparable to that of AtzA.[13] Cys331Ser and Asp328Asn, mutations suggested by homology modeling to lie in the active site, improve the k_{cat}/K_M value for atrazine hydrolysis dramatically, from 60 to close to $10000\,M^{-1}s^{-1}$, while completely abolishing melamine-hydrolysis activity.

Frances H. Arnold is the Dickinson Professor of Chemical Engineering, Bioengineering and Biochemistry at the California Institute of Technology. After completing her PhD in 1985 with Henry Blanch at the University of California, Berkeley, she undertook postdoctoral research with Ignacio Tinoco (Berkeley) and John Richards (Caltech). Her research focuses on protein evolution and biocatalysis. She was elected to the US National Academies of Engineering (2000), Medicine (2004), and Sciences (2008), and was awarded the National Medal of Technology and the Draper Prize in Engineering in 2011.

Furthermore, Scott and co-workers showed that one can arrive at an atrazine chlorohydrolase by accumulating single beneficial mutations in a simple uphill walk, starting from TriA. The Cys331Ser mutation alone led to an almost 30-fold improvement in the k_{cat}/K_M value for atrazine hydrolysis. The introduction of Asp328Asn to the Cys331Ser mutant conferred a further sixfold improvement in catalytic efficiency.

These active-site mutations effect a dramatic shift from aminohydrolase to chlorohydrolase activity that can be rationalized after the fact. Scott and co-workers reasoned that the Asn/Ser dyad (positions 328 and 331 in AtzA) assists in the expulsion of a chloride anion through a hydrogen-bonding network, whereas the Asp/Cys dyad in TriA enables the release of ammonia through a proton-relay mechanism, whereby the thiol moiety of Cys331 donates a proton to the leaving NH_2^- group and is reprotonated by Asp328. Since NH_2^- is a poor leaving group ($pK_a = 34$), protonation by the more acidic Cys residue is necessary for the reaction to occur.

A number of functionally diverse enzyme superfamilies have been described.[14] Superfamily members share a structural fold and are believed to have diverged functionally from a common ancestor through a series of catalytically promiscuous intermediates. Most known member enzymes that catalyze different reactions, however, have accumulated many more sequence changes than the AtzA/TriA pair; large sequence distances make it more challenging to demonstrate simple evolutionary pathways among them or to pinpoint the functions of the ancestral enzymes. However, several research groups have taken on this challenge and used protein engineering and especially directed evolution to demonstrate how one function can become another in the context of extant enzymes.

2.2. Evolution of a Phosphotriesterase (PTE)

Another enzyme believed to have emerged very recently is phosphotriesterase (PTE), first identified in soil bacteria that can grow on synthetic organophosphate pesticides, such as parathion and paraoxon, as their sole phosphorus source (Figure 3A).[15] Given the recent introduction of parathion and paraoxon into the environment and the fact that PTE hydrolyzes them at near-diffusion-controlled rates, it is thought that PTE evolved recently from an ancestral enzyme with promiscuous organophosphate-hydrolysis activity. The ancestral enzyme is unknown, however, as no very close sequence homologue has been identified.

The PTE from *Pycnoclavella diminuta* has the $(\beta/\alpha)_8$-barrel fold and binuclear metal center common to many members of the amidohydrolase (AHS) superfamily, which hydrolyze different classes of substrates.[16] Afriat et al. proposed that the promiscuous lactonase activity of this PTE could be a vestige of its ancestral source, a clue to its ancestral function.[17] They showed that the three microbial enzymes that are the closest known homologues to *P. diminuta* PTE (ca. 30% sequence identity) are in fact highly active lactonases and also possess varying levels of promiscuous organophosphate-hydrolysis activity. These "phosphotriesterase-like lactonases" (PLLs) are especially active towards *N*-

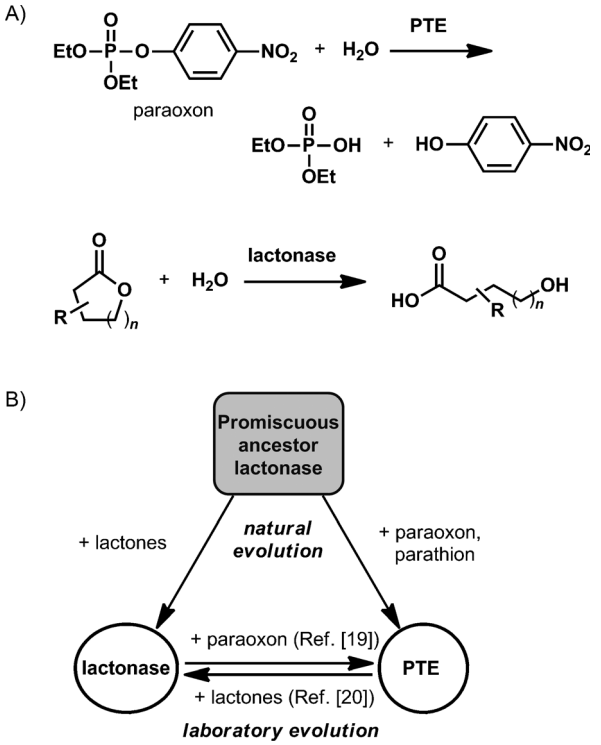


Figure 3. A) Hydrolysis reactions catalyzed by PTE and lactonase. B) Putative evolutionary relationship between lactonase and PTE and their interconversion in the laboratory.[19,20]

acyl homoserine lactones, which play a vital role in bacterial quorum sensing.[18] Afriat et al. argued that PTE may have evolved from the weak promiscuous activity of a bacterial PLL.

Raushel and co-workers recently conducted an experiment to try to recapitulate the conversion of a PLL with low-level organophosphate-degrading activity into a PTE.[19] By using a combination of rational and random mutagenesis, they were able to convert the PLL from *Deinococcus radiodurans* (*Dr*PLL) into an enzyme exhibiting PTE activity that is five orders of magnitude higher than that of the wild-type *Dr*PLL and just one order of magnitude lower than that of wild-type PTE. The most efficient enzyme contained only seven mutations from wild-type *Dr*PLL, three of which were sufficient to increase the PTE activity by two orders of magnitude. Tawfik and co-workers showed that the converse is also possible when they engineered a recombinant variant of PTE from *P. diminuta* to be a bifunctional PTE/PLL by active-site-loop remodeling.[20]

Similar to AtzA, PTE is believed to have arisen in response to the introduction of synthetic compounds that were potential new nutrient sources (Figure 3B). Unlike AtzA and TriA, however, PTE has low sequence identity to its closest known homologue, which is thought to be a lactonase. Despite the fact that the known PTE and PLL enzymes differ at hundreds of amino acid positions and in the structure and arrangement of the active-site loops through deletions and/or insertions, their functions overlap and could be interconverted by directed evolution and rational design. The ability to carry out this interconversion in the laboratory demonstrates the ease with which a promiscuous PLL could become a PTE in nature.[19] This example also demonstrates how readily a residual ancestral activity can be enhanced by the accumulation of beneficial mutations.

2.3. Catalytic Promiscuity in the MBL Superfamily

In a comprehensive study of how catalytic functions overlap in the metallo-β-lactamase (MBL) superfamily of αβ/βα proteins, Tokuriki and co-workers examined the activities of 24 enzymes from 15 distinct subfamilies.[21] They found that many members of the MBL superfamily, despite their low sequence identity (ca. 5–35%), catalyze at low levels the distinct reactions of distant family members in addition to their own (Figure 4). Echoing a common theme in enzyme evolution that active-site architecture within a superfamily is often at least partially conserved,[22] most members of this superfamily retain the binuclear active-site center for the activation of a water molecule by a divalent metal. Thus,

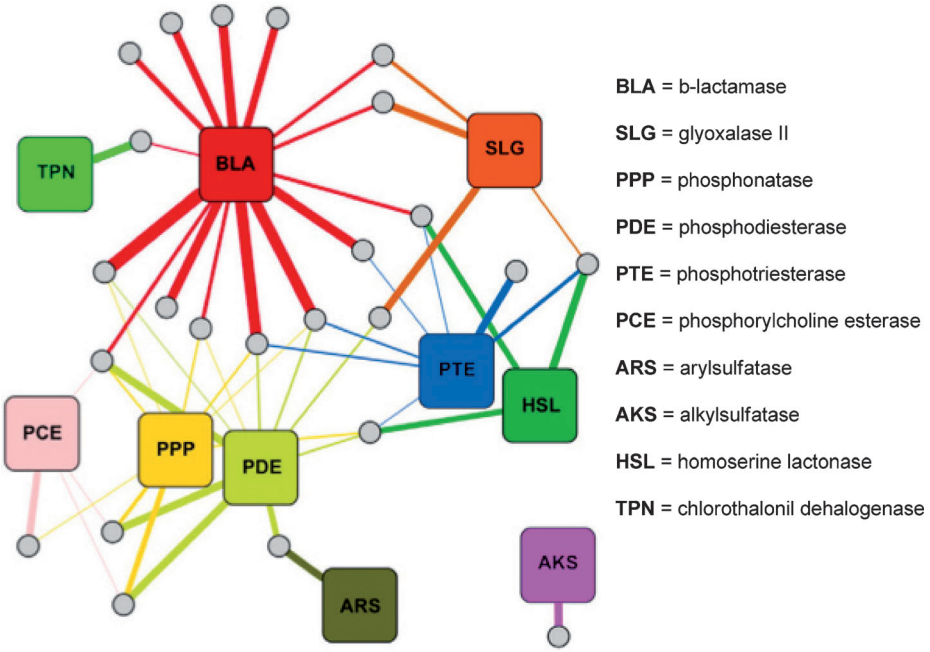


Figure 4. Members of the MBL superfamily are functionally highly interconnected, as illustrated by Tokuriki and co-workers.[21] The different reactions catalyzed by members of the superfamily are connected to one another through promiscuous enzymes (gray circles) that catalyze two or more reactions. Reproduced from Ref. [21].

observed promiscuous activities are likely to arise from the shared active-site features of the superfamily members. The MBL superfamily also includes a PTE family that probably evolved from lactonases, in parallel with the PTEs in the AHS superfamily and in what appears to be an example of the convergent evolution of a new function.[23]

Figure 4 illustrates the highly interconnected network of overlapping functions that Tokuriki and co-workers sampled in this superfamily. These enzymes presumably evolved from a common ancestor through a series of promiscuous intermediates. Promiscuity is still prevalent in the family, and given the level of functional overlap that still exists, it is reasonable to assume that it is possible move within this network from one catalytic function to others by directed evolution. In fact, Park et al. showed the feasibility of converting a glyoxalase II from this family into an enzyme with high β-lactamase activity by directed evolution.[24]

2.4. Evolution of a Hydroxylase from a Desaturase

Oleate desaturases and hydroxylases are integral membrane diiron enzymes that catalyze the modification of oleic acid (**1**) to form the corresponding dehydrogenated and oxygenated products, linoleic acid (**2**) and ricinoleic acid (**3**; Figure 5A). They typically exist in higher plants and are closely related members of a functionally diverse non-heme diiron enzyme family. The two reactions are also mechanistically related in that they are both initiated by a hydrogen-atom-abstraction step.[25] They diverge in the subsequent step, in which desaturation occurs through the abstraction of another hydrogen atom and hydroxylation proceeds through a radical-rebound/oxygen-transfer event.

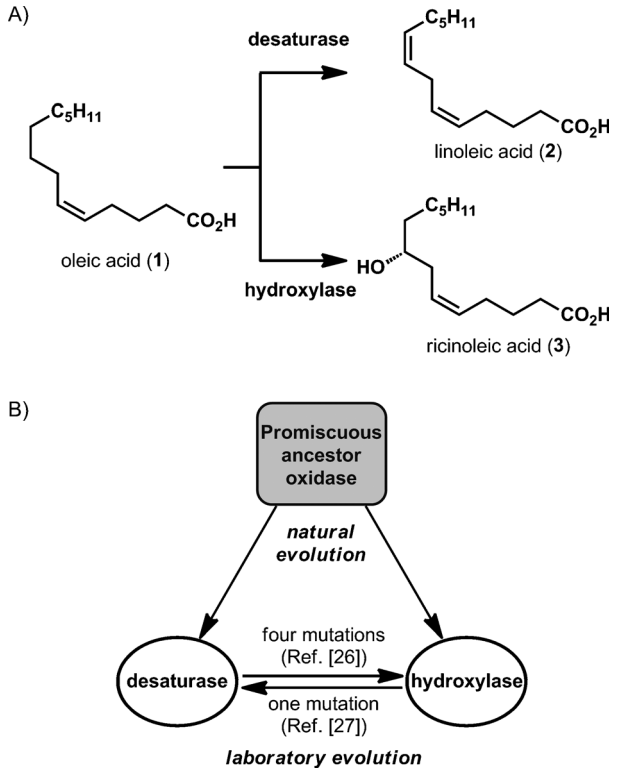


Figure 5. A) Desaturation and hydroxylation reactions of oleic acid catalyzed by FAD2 (a desaturase) and LFAH12 (a hydroxylase). B) Four mutations significantly increase the hydroxylase activity of *A. thaliana* FAD2 desaturase, and a single mutation significantly increases the desaturase activity of *L. fendleri* oleate hydroxylase LFAH12.[26,27]

When expressed in yeast, the hydroxylase from *Lesquerella fendleri*, LFAH12, was found to have appreciable desaturase activity in addition to its native hydroxylase activity (with the production of di-unsaturated fatty acids and hydroxylated fatty acids in an approximately 1:1 ratio).[26,27] In contrast, the desaturase from *Arabidopsis thaliana*, FAD2, was shown to catalyze desaturation almost exclusively, with only a very minor amount of hydroxylation products detected (a hydroxylation/desaturation product ratio of 0.006:1). These two enzymes are close relatives, with approximately 81% sequence identity, and sequence comparison of the two with a few other hydroxylases and desaturases led to the identification of several residues that are highly conserved in the desaturases but diverged in the hydroxylases.[26] On the basis of these results, seven residues from FAD2 were introduced into the corresponding positions in LFAH12. The resulting variant showed predominantly desaturase activity, and further mutation analysis demonstrated that as few as six mutations could transform LFAH12 into a desaturase (Figure 5B). Conversely, four mutations were found to be sufficient to convert FAD2 into a hydroxylase. Further studies by Broadwater et al.[27] showed that a single mutation was sufficient for a comparable boost in hydroxylase activity in FAD2.

This desaturase–hydroxylase example demonstrates how readily related enzymes with overlapping activities can be interconverted, similar to what was observed for the atrazine chlorohydrolase and phosphotriesterase examples. It was not necessary to identify and enlist the ancestral enzyme for the evolution of the new functions: The new enzymes were obtained in the laboratory by starting from the extant relative. As these examples suggest, a few mutations can be sufficient to convert an existing enzyme into a new enzyme with distinct, but mechanistically related activity.[22]

We have chosen just a few examples to illustrate how enzymes that catalyze different reactions can diverge from a common ancestor, especially when the activities overlap at least a little. Such divergence occurs in nature, and the laboratory experiments demonstrate just how readily it can happen. Although nature may not use the same routes or starting points, the laboratory experiments show how a new enzyme can appear and evolve as opportunities for a selective advantage arise. Now let us discuss the use of this strategy to make enzymes that catalyze reactions that have not yet been discovered in nature.

3. Using Mechanistic Similarities and Directed Evolution To Expand the Enzyme Universe

Many contemporary enzymes have versatile active sites and exhibit promiscuous activities at some level as a property of their mechanisms.[28] Thus, as the above examples show,

related enzymes that diverged by natural evolution for different functions can be interconverted and optimized, often with just a few mutations in an uphill evolutionary walk. Because catalytic promiscuity is common, the plethora of biocatalysts that exists today provides equal or perhaps even greater opportunity for innovation than during early evolution. Just as nature uses that rich source of starting materials to create new catalysts, so could we.

Let us suppose for a moment that an atrazine chlorohydrolase had not yet been discovered in nature. Could its appearance have been anticipated, and could such an enzyme have been created in the laboratory before it was found in nature? Given the similarity in substrate structure (*s*-triazine heterocyclic core) and reaction type (hydrolysis), one might logically test an aminohydrolase, such as TriA, as a starting point for directed evolution. As atrazine contains a better leaving group (Cl^- versus NH_2^-), its hydrolysis could be expected to be more facile than that of melamine. Indeed, TriA possesses a low level of atrazine chlorohydrolase activity ($k_{cat}/K_M = 60\ s^{-1}M^{-1}$). From this point, one could couple random or site-saturation mutagenesis with a high-throughput spectrometric assay[29] for the hydrolysis product to identify variants with increasing levels of activity. As Scott and co-workers showed, the accumulation of just two mutations improved the chlorohydrolase activity of TriA almost 200-fold.[13]

Given that a new, mechanistically related catalytic function can be imparted with just a few mutations to an enzyme that already possesses a low level of that function, an evolution-inspired approach to new catalyst discovery relies on the ability to identify an appropriate starting point, that is, an existing enzyme able to take on a new function. To find that enzyme, we can look for mechanistic similarities between an existing activity and a desired transformation for which no enzyme is known. Moreover, because low levels of catalytic activity exhibited by a promiscuous enzyme can be improved in many cases by engineering the protein sequence, the expectation is reasonable that a non-natural, but mechanistically related promiscuous activity can similarly improve.

In fact, several research groups have used this approach and capitalized on the catalytic promiscuity of enzymes and the similarity between the native and desired reaction mechanisms to create new enzymes. In some cases, however, the starting enzyme did not exhibit the desired promiscuous activity, and researchers had to rely on their chemical intuition that the desired function should be possible and could be obtained with one or a few mutations. In this section, we present some examples of novel non-natural functions that have been discovered by the use of this mechanism-based approach and then improved to useful levels by protein engineering and directed evolution.

3.1. Epoxide-Ring Opening with a Halohydrin Dehalogenase

Wild-type halohydrin dehalogenases (HHDH) catalyze the formation of epoxides from the corresponding chloro- and bromohydrins.[30] X-ray crystal structures of halohydrin dehalogenase from *Agrobacterium radiobacter* AD1 suggested the presence of a binding site for both the epoxide and the halide anion,[31] thus raising the possibility of using pseudohalides of varying sizes in the reverse epoxide-opening reaction. Indeed, Janssen and co-workers found that this enzyme accepts a wide range of non-natural nucleophiles, such as azide, nitrite, cyanate, and thiocyanate,[32] and used it for the kinetic resolution of various epoxides to give the ring-opened products with high enantioselectivity (Figure 6A). High selectivity for opening at the terminal position was also observed. In the case of aryl epoxide substrates, this finding stood in stark contrast to the noncatalyzed ring-opening reaction, in which nucleophilic attack usually occurs at the benzylic position.

In 2007, scientists at Codexis reported the use of HHDH in the asymmetric synthesis of ethyl (*R*)-4-cyano-3-hydroxybutyrate (**6**),[33] a valuable intermediate in the production of atorvastatin, a cholesterol-lowering drug. They were able to enhance the low activity of the wild-type enzyme for the cyanation of **5** by directed evolution to obtain enzyme variants that increased the volumetric productivity of the process by a factor of about 4000 (Figure 6B).[34] This catalyst enables the production of **6** with over 99.9% *ee* with a substrate loading of $130\ g\,L^{-1}$.

The promiscuous cyanation activity of HHDH was discovered through enzyme-structure analysis and by analogy to related epoxide-ring opening with non-natural pseudohalide nucleophiles. The low cyanation activity of wild-type HHDH could then be improved tremendously by directed evolution, through the accumulation of beneficial mutations in an uphill walk to the new function.

3.2. Synthesis of Thiooligosaccharides with an Engineered Glycosidase

A good strategy for introducing a new activity can be to divert a reactive intermediate to an alternative reaction pathway, as nature did with the hydroxylase and desaturase enzymes. A nice example comes from an early study by Withers and co-workers. By examining the catalytic mechanism of retaining β-glycosidases and rationally modifying the key catalytic residue(s), Withers and co-workers were able to divert a reactive intermediate in the hydrolysis of glycosidic bonds and redirect it to the synthesis of thiooligosaccharides.[35]

Thiooligosaccharides are of interest as carbohydrate mimics that possess a more stable and hydrolysis-resistant glycosidic bond.[36] These compounds are challenging to synthesize, because their preparation typically involves the manipulation of protecting groups and requires a high degree of stereocontrol at the anomeric position. A few enzymes are known that catalyze the formation of the C–S bond of naturally occurring thioglycosides, but there are only very few reports of their use in the preparation of thiooligosaccharides.[37]

Glycosidases catalyze the hydrolytic cleavage of glycosidic bonds and are mainly responsible for the degradation of carbohydrate-based biomass. At low water concentration, these enzymes are also capable of catalyzing glycoside

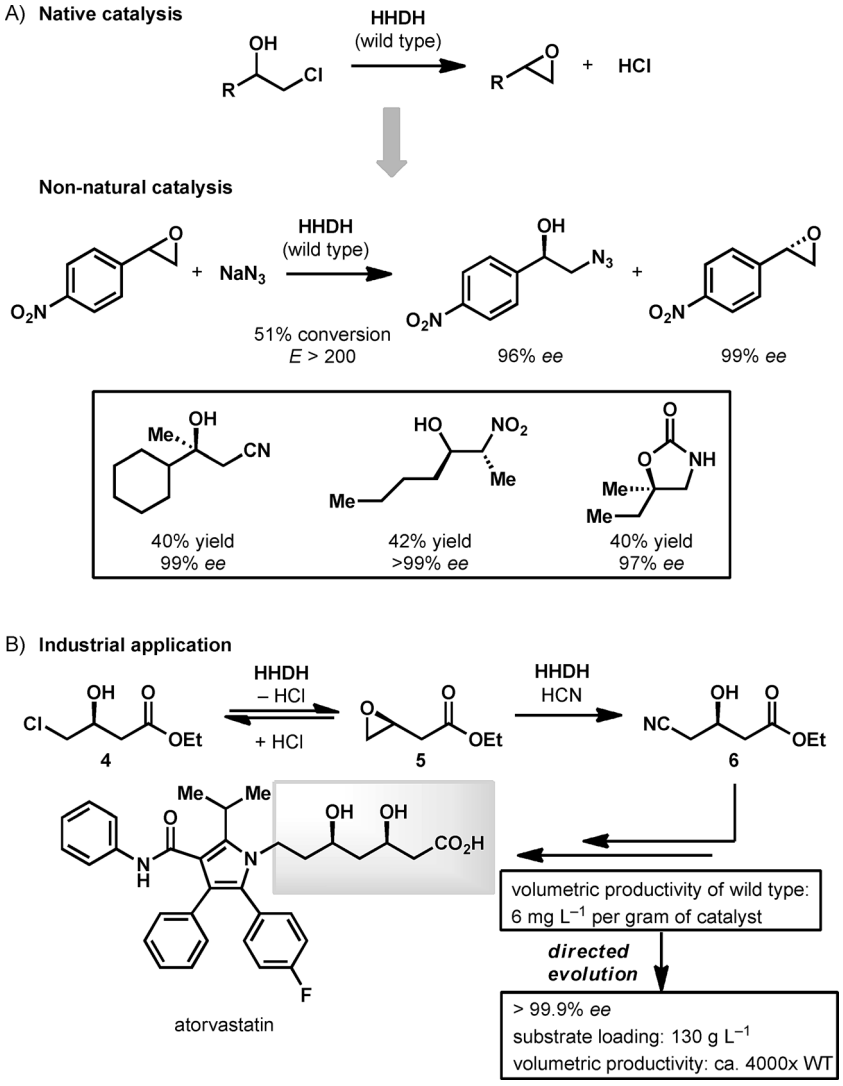


Figure 6. A) Promiscuous epoxide-ring-opening activity of wild-type HHDH.[32] B) Application of HHDH in the synthesis of the atorvastatin side chain and improvement of volumetric productivity by directed evolution.[33]

exchange. For retaining β-glycosidases, the enzymatic hydrolysis and glycoside-exchange reaction occurs through a double-substitution mechanism in which a catalytic nucleophilic residue first displaces the departing aglycon group and a catalytic acid/base residue activates the incoming water or glycoside nucleophile to form the new anomeric bond (Figure 7A).[38]

For the creation of a new activity, it is helpful if the starting enzyme already exhibits it at some level as a promiscuous activity, but this property is not necessarily required. In such a case, however, the catalytic mechanism must allow acquisition of the new activity without too much fine-tuning. By substituting the catalytic acid/base glutamine residue with an inert alanine residue (Figure 7B), Withers and co-workers redesigned the active site of β-glucosidase from *Agrobacterium* sp. (Abg) and β-mannosidase (Man2A) from *Cellulomonas fimi* in such a way that they would only catalyze the glycosylation of activated dinitrophenyl (DNP) glycoside donors and deoxythiosugars as acceptors. The former do not require acid activation since the leaving group is good enough to be displaced by the catalytic nucleophilic residue, and the latter contain a highly nucleophilic thiol moiety that does not require base catalysis for the formation of the glycosidic linkage.

Both glycosidase mutants were shown to catalyze the reaction of DNP glucose and DNP mannose with glycoside acceptors **7** and **8** in good to excellent yield (Figure 8). Notably, no protecting groups were required on the free hydroxy groups of the glycoside donors and acceptors. In further studies, Withers and co-workers showed that mutations at the catalytic acid/base residue could boost the thioglycoligase activity of these retaining β-glycosidases.[39]

3.3. Carbene- and Nitrene-Transfer Reactions with Cytochrome P450s

Metalloporphyrin complexes, long used as synthetic models for cytochrome P450 enzymes, also catalyze chemical reactions that have no natural counterparts. For example, the reaction of metalloporphyrins, including iron porphyrins, and diazo compounds to generate metal-carbenoid reactive intermediates for cyclopropanation reactions is well-documented (Figure 9A).[40] Whereas carbenes are generally too reactive for characterization, some of these metal-carbenoid species proved to be stable enough for isolation and X-ray crystallographic analysis.[41] These isolated metal carbenoids participate in cyclopropanation reactions, thus providing evidence that metalloporphyrin-catalyzed cyclopropanation reactions proceed through the metal-carbenoid intermediates.

Metalloporphyrins are also known to form reactive nitrenoids in the presence of activated species, such as azides and iminoiodinanes. Breslow and Gellman first showed that meso-tetraphenylporphyrin iron(III) chloride could catalyze intra- and intermolecular nitrene transfer when treated with iminoiodinanes,[42] presumably via a metal-nitrenoid species. A follow-up report by Dawson and co-workers further established that a rabbit-liver cytochrome P450 could catalyze the same nitrene-transfer reaction, albeit with a very low turnover.[43] Given the similarity between a carbene/nitrene and oxene (the free oxygen atom equivalent of the reactive species in P450 monooxygenation reactions) in terms of their electronic configuration and electrophilicity (Figure 9B), we hypothesized that cytochrome P450s may exhibit promiscuous activity for carbene- and nitrene-transfer reactions and that such activity could be improved by protein engineering. The addition of ethyl diazoacetate to styrene in

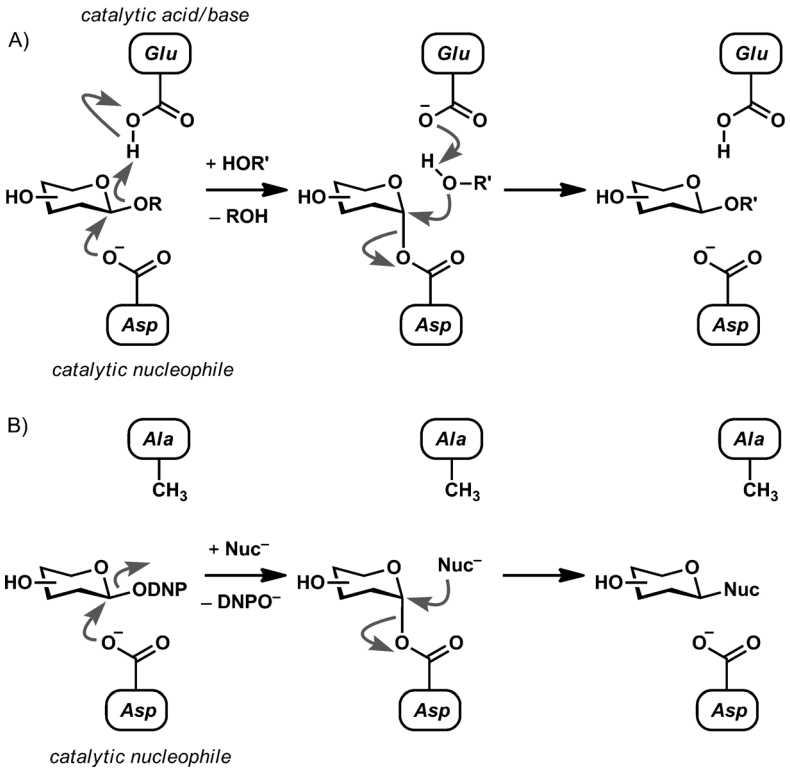


Figure 7. A) Mechanism of wild-type retaining β-glycosidases featuring catalytic acid/base and catalytic nucleophile residues; R′ = H or another sugar at low water concentration.[38] B) Mechanism of engineered thioglycoligases from which a catalytic acid/base residue has been removed. DNP = dinitrophenyl, Nuc = deoxythiosugar nucleophile as an acceptor.[35]

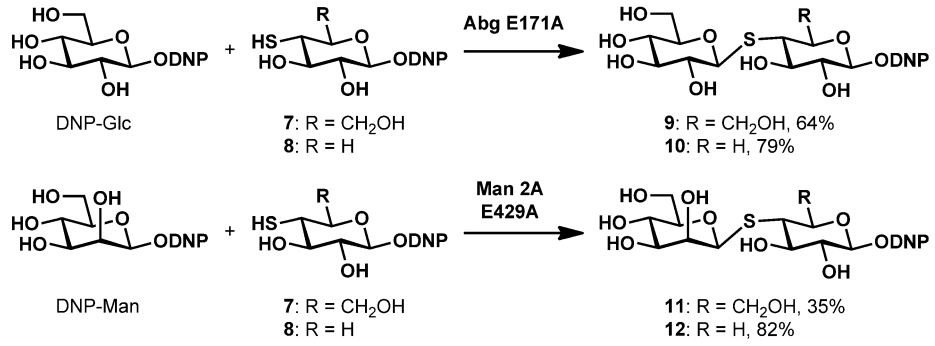


Figure 8. Use of engineered thioglycoligases for the synthesis of thiodisaccharides.[35] The reported yields are for the thiodisaccharides obtained after peracetylation. Wild-type enzymes do not catalyze this reaction.

the presence of wild-type P450 BM3 from *Bacillus megaterium* led to trace amounts of the corresponding cyclopropane product.[44] Further studies established that various hemeproteins, and even free hemin, catalyze olefin cyclopropanation in water.[45] Of particular interest, however, several wild-type P450s exhibited diastereoselectivity different from that of free hemin and showed some enantiomeric induction as well.[44,45] The unusual selectivity of the P450s suggested that cyclopropanation was taking place in the enzyme active site and that the active-site geometry exercised control over the stereochemical outcome of the reaction. We thus felt that the P450s were suitable starting points for the engineering and evolution of a new family of enzymes that could activate diazo compounds for carbene transfer to organic molecules.[44,46–50]

Mutations dramatically increase the non-natural cyclopropanation activity of P450 BM3. Substitution of the distal threonine residue (Thr268), a key residue in the native catalytic cycle for monooxygenation, with alanine improved the turnover number more than 60-fold. Further tailoring of the active site led to variant $P450_{BM3}$-CIS T438S, which catalyzed the cyclopropanation of styrene in high yield with excellent diastereoselectivity and enantioselectivity (Figure 10 A). Mutation at the cysteine axial ligand led to the greatest improvement in cyclopropanation activity. Mutation of the cysteine residue at position 400 to serine (Cys400Ser) in P450 BM3 shifted the characteristic 450 nm peak in the CO difference spectrum to 411 nm, hence the designation "P411" for the serine-ligated catalysts. The serine mutation also allowed the iron–heme complex to be reduced under cellular conditions to the Fe^{2+} active catalyst, thereby enabling cyclopropanation to be carried out with whole cells expressing these proteins. Styrene cyclopropanation was catalyzed on a gram scale to 67 000 turnovers in 72 % yield by a P411 enzyme, which is competitive with some of the most active reported rhodium catalysts.[46]

Enzyme-catalyzed cyclopropanation has been applied to the formal synthesis of levomilnacipran, a serotonin- and norepinephrine-reuptake inhibitor approved for the treatment of clinical depression.[47] A variant of P450 BM3 containing only five amino acid mutations, including the mutation of the proximal cysteine residue to histidine, catalyzed the cyclopropanation of *N*,*N*-diethyl-2-phenylacrylamide (**13**) to 86 % yield with 92 % enantioselectivity on a preparative scale (Figure 10 B). The laboratory-evolved catalyst BM3-Hstar promoted cyclopropanation in the presence of oxygen and exhibited an initial rate of reaction close to that reported for monooxygenation by wild-type P450 BM3 (> 1000 turnovers per minute). Examination of a panel of 2-phenylacrylamide derivatives revealed that BM3-Hstar was quite a general cyclopropanation catalyst and could be used on substrates with varied steric and electronic properties.[48]

Carbene insertion into aryl N–H bonds, another reaction catalyzed by iron porphyrins, can also be catalyzed by variants of P450 BM3.[49] The P411 variant H2-5-F10 catalyzed the

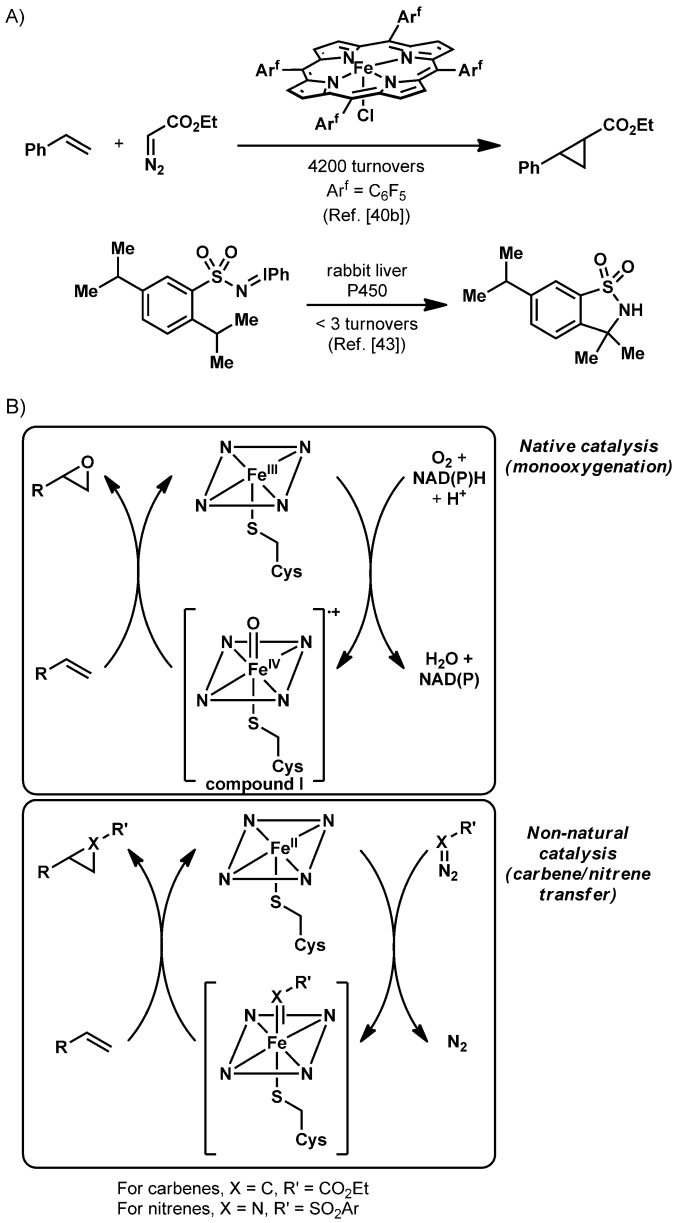


Figure 9. A) Precedent for carbene and nitrene reactivity with iron porphyrins.[40b,43] B) Comparison of the monooxygenation activity and non-natural carbene/nitrene-transfer reactivity of P450-BM3. Top box: The reaction of Fe^{III} of P450 BM3 with O_2 and NAD(P)H generates compound I, the active species in monooxygenation. Bottom box: The reaction of reduced Fe^{II} with an activated species, followed by extrusion of N_2, generates a putative carbenoid or nitrenoid species.

insertion into aryl N–H bonds in up to 83% yield and 354 turnovers in vitro (Figure 10C). Free Fe-protoporphyrin IX can catalyze aniline N–H insertion reactions to produce a mixture of single- and double-insertion products. In contrast, the enzyme provided the single-insertion product selectively, thus highlighting the important role that the protein-binding pocket plays in controlling selectivity.

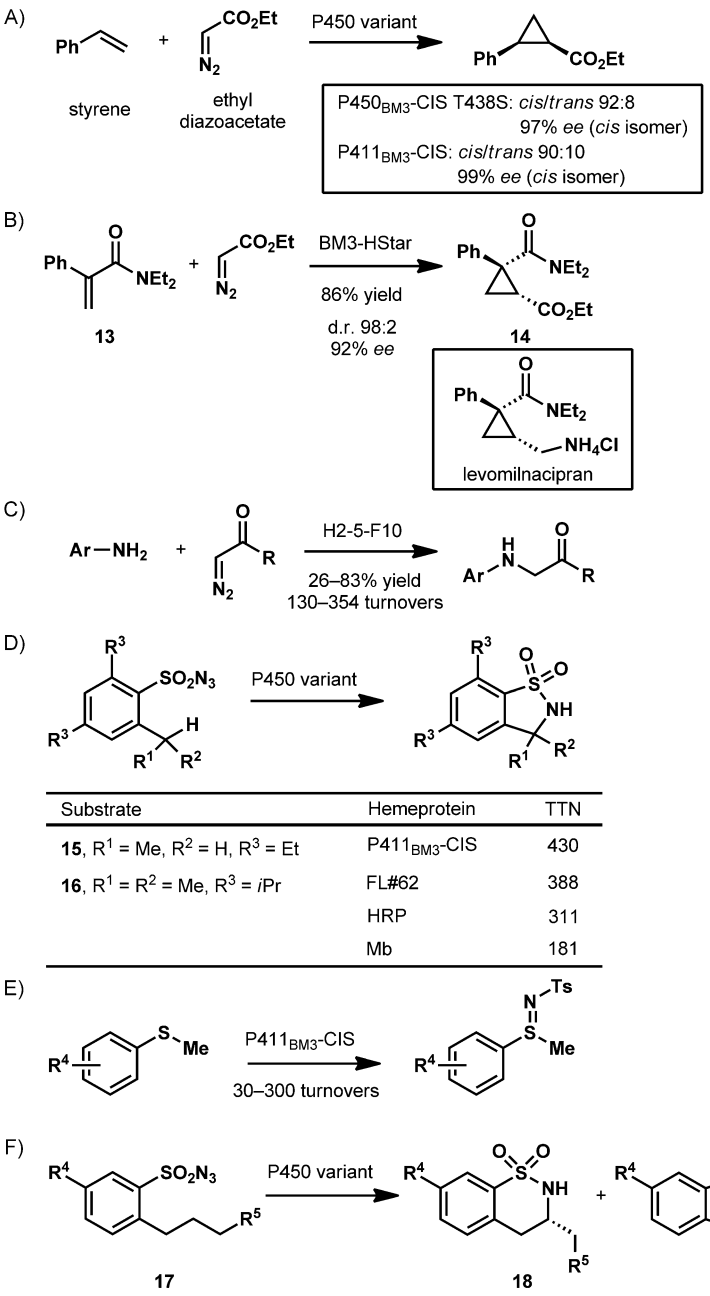
Following up on the early observation of Dawson and co-workers,[43] McIntosh et al. greatly improved the catalytic performance of P450 BM3 for intramolecular C–H amination with sulfonyl azides as the nitrene precursor.[50a] For azide **15**, mutations at key residues, such as Thr268 on the I helix (Thr268Ala) and the cysteine proximal ligand for the heme prosthetic group (Cys400Ser), were found to be crucial for improving the total turnover number (TTN) for amination (Figure 10D). In particular, variant $P411_{BM3}$-CIS catalyzed the amination reaction to give the product with up to 87% *ee* in 430 turnovers. The same variant was also found to catalyze nitrene transfer from *p*-toluenesulfonyl azide to a series of thioethers to generate the corresponding sulfimides in 30–300 turnovers (Figure 10E).[50b] Since free hemin does not catalyze this sulfimidation reaction at all, the ligation state of the heme cofactor in the protein and/or the protein itself play a key role in modulating the reactivity of the nitrenoid species.

An independent report by Fasan and co-workers revealed that the intramolecular C–H amination of sulfonyl azide **16** could be effected by a different variant of P450 BM3 (FL#62) that does not contain the Thr268Ala and Cys400Ser mutations.[51a] Their investigation of the scope of intramolecular C–H amination with FL#62 showed that the biocatalyst tolerates a range of substituents on the aryl ring of the substrate. A follow-up report[51b] further showed that the C–H amination of azide **16** could be catalyzed with good turnover numbers with either myoglobin (Mb) or horseradish peroxidase (HRP). Whereas wild-type Mb showed no detectable enantioselectivity in the amination of **16**, the introduction of mutations His64Val and Val68Ala led to appreciable enantiomeric induction (60% *ee* with **16**). These results suggest that other hemeproteins are also viable platforms for the discovery of new catalysts for non-natural reactions.

Hyster et al. very recently showed that the regioselectivity of this enzyme-catalyzed C–H amination can be tuned by mutations (Figure 10F).[50c] The variant $P411_{BM3}$-CIS-T438S-I263F catalyzed the C–H amination of substrates **17a–c** at the homobenzylic position with excellent regio- and enantioselectivity. The $P411_{BM3}$-T268A-F87A variant, in contrast, showed a strong preference for C–H amination at the benzylic position, also with high regio- and enantioselectivity. Thus, tailoring of the active site can alter the conformation of the reactive intermediates to the extent that the catalyst can override the thermodynamic bias towards reaction at the benzylic position (the bond-dissociation energy for the benzylic C–H bond is weaker by more than 10 kcal mol^{-1} than that of the nonbenzylic C–H bond).

3.4. Redirecting Cyclization with Terpene Synthases

Squalene–hopene cyclase (SHC) catalyzes the cationic polycyclization of squalene to the pentacyclic products hopene and hopanol (Figure 11), a reaction that Hauer and co-workers noted is highly reminiscent of polycyclizations catalyzed by chiral Brønsted acids.[52] The crystal structure of SHC from *Alicyclobacillus acidocaldarius* (*Aac*SHC) was disclosed in 1997,[53] but its promiscuity was known as early as



Substrate	Hemeprotein	TTN
15, R^1 = Me, R^2 = H, R^3 = Et	$P411_{BM3}$-CIS	430
16, R^1 = R^2 = Me, R^3 = *i*Pr	FL#62	388
	HRP	311
	Mb	181

P450 variant	Substrate	TTN	**18**/**19**	*ee* (**18**) [%]	*ee* (**19**) [%]
$P411_{BM3}$-CIS-T438S-I263F	**17a**, R^4 = *n*Pr, R^5 = H	361	97:3	99	–
	17b, R^4 = *n*Bu, R^5 = Me	178	90:10	99	–
	17c, R^4 = CO_2Et, R^5 = H	192	95:5	97	–
$P411_{BM3}$-T268A-F87A	**17a**, R^4 = *n*Pr, R^5 = H	187	30:70	–	99
	17b, R^4 = *n*Bu, R^5 = Me	128	3:97	–	99
	17c, R^4 = CO_2Et, R^5 = H	130	5:95	–	99

Figure 10. P450-catalyzed non-natural carbene- and nitrene-transfer reactions: A) styrene cyclopropanation;[44,46] B) cyclopropanation of *N*,*N*-diethyl-2-phenylacrylamide en route to levomilnacipran;[47] C) N—H insertion reaction;[49] D) intramolecular C—H amination;[50a,51] E) intermolecular sulfimidation;[50b] F) regioselective C—H amination by different P450 variants.[50c]

1986, when Neumann and Simon showed that homofarnesol could be cyclized by SHC to ambroxan, a valuable fragrance compound.[54] This reaction can be regarded as both a substrate- and catalytic-promiscuity feature of the enzyme, because attack by an internal nucleophile terminates the cyclization reaction.

Hauer and co-workers further investigated the inherent promiscuity of this enzyme with a range of terpene-like substrates to construct novel carbocyclic skeletons (Figure 12 A). By varying the terminator groups for the cyclization reactions, they were able to perform the SHC-catalyzed cyclization of various non-natural substrates in low to moderate yield.[54,55] The scope of the SHC-catalyzed cyclization reaction has also been studied by Hoshino and co-workers.[56]

SHC also possesses weak activity for the Prins cyclization of citronellal to form isopulegol.[57] Siedenburg et al. conducted site-saturation mutagenesis on three residues in the active site of SHC from *Zymomonas mobilis* (ZMO1548) to improve the production of isomers of isopulegol from racemic citronellal (Figure 12 B).[58] Two amino acid positions were identified as important for increased isopulegol formation. The Trp555Tyr mutant gave more than 70 % total conversion into isomeric mixtures of isopulegol, as compared to approximately 30 % observed with the wild-type enzyme. The Phe486Cys variant provided more than 50 % total conversion, with slightly improved diastereoselectivity.

In a very recent study,[59] Hauer and co-workers engineered an SHC from *A. acidocaldarius* (*Aac*SHC) to improve the catalytic activity for several different modes of Brønsted acid catalyzed cyclization, including the Prins cyclization of (*S*)-citronellal (Figure 13). By screening a library of enzyme variants made by mutating several amino acids in proximity to the catalytic Asp376 residue, they discovered variants with greatly improved activities for various cyclization reactions. The Tyr420Trp–Gly600Phe mutant catalyzed the cyclization of 6,7-epoxygeraniol to cyclohexanoid **22** with 78 % conversion, an approximately 140-fold improvement over the wild-type enzyme. Similarly, the Ile261Ala mutant catalyzed the Prins cyclization of (*S*)-citronellal to *iso*-isopulegol with 11 % conversion, an approximately 20-fold improvement over the wild-type enzyme. It will be interesting

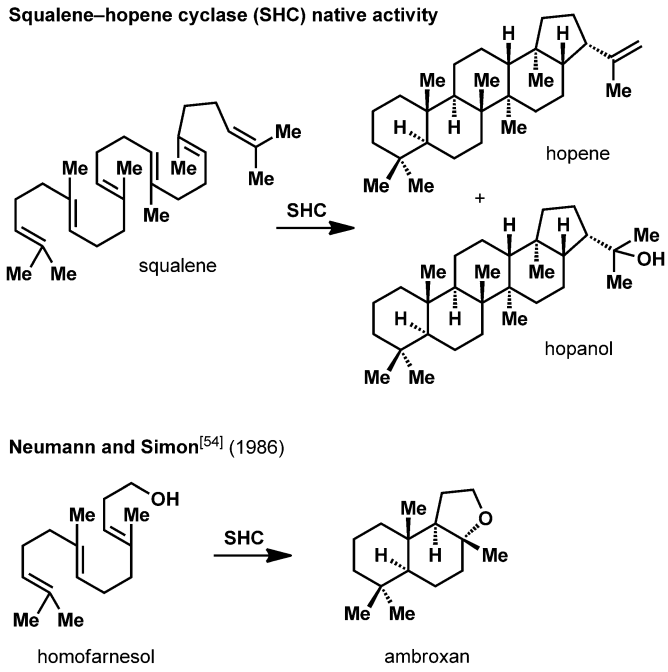


Figure 11. SHC-catalyzed cyclization of squalene and promiscuous activity of SHC on homofarnesol.[54]

to see if this non-natural cyclization activity of SHC or other terpene synthases can be expanded further to include other electrophiles as initiators.[60]

4. New Opportunities on the Horizon?

At present, biocatalytic transformations constitute a small but growing subset of industrially relevant chemical processes. For the potential of biocatalysis for sustainable chemistry to be fully realized, it will be important to expand the range of enzyme-catalyzed transformations to include reactions not yet discovered in nature. The examples we have presented illustrate an evolutionary approach that mimics some features of how nature creates new catalysts. Our feeling is that we have barely scratched the surface in terms of the possibilities available, as promiscuous activity among enzymes is widespread and can be improved by protein engineering and especially by directed evolution. Chemical intuition will help us find new opportunities.

Several recent discoveries of novel, promiscuous, and non-natural enzyme activities highlight the potential for future biocatalyst development. Still in their infancy, these next examples are important proofs-of-concept with potential for future applications. In cases in which wild-type enzymes catalyze non-natural reactions, it is entirely possible that protein engineering and directed evolution could boost their activity, fine-tune their selectivity, and make them more synthetically useful.

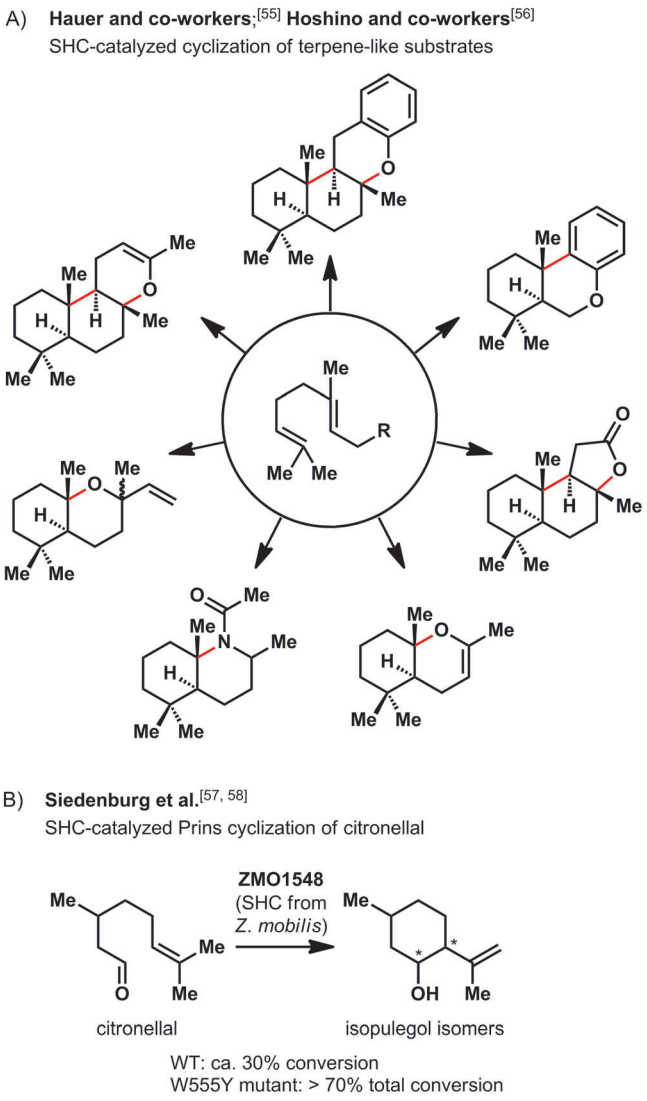


Figure 12. A) SHC-catalyzed cyclization of terpene-like substrates (bonds formed during the reactions are highlighted in red).[54–56] B) SHC-catalyzed Prins cyclization of citronellal for the production of isopulegol, a precursor to menthol, and activity improvement through mutation.[57,58]

4.1. C–H Functionalization by SyrB2 Halogenase

Nature utilized a similar bifurcation of reaction pathways to that described for the hydroxylase/desaturase example for halogenation and hydroxylation with iron(II)- and α-ketoglutarate-dependent (Fe/αKG) enzymes. An Fe/αKG enzyme from the syringomycin biosynthetic pathway of *Pseudomonas syringae* B301D, SyrB2 catalyzes the halogenation of the terminal methyl group of L-threonine appended to the carrier protein SyrB1.[61] This enzyme is related to Fe/αKG hydroxylases, which employ an almost identical reaction mechanism. Both reactions proceed through initial hydrogen-atom abstraction from the substrate by an Fe^{IV}–oxo intermediate to form a carbon-centered radical (Figure 14).[62] The subsequent step determines the product outcome: Homolytic coupling with a hydroxy ligand results in the hydroxylated

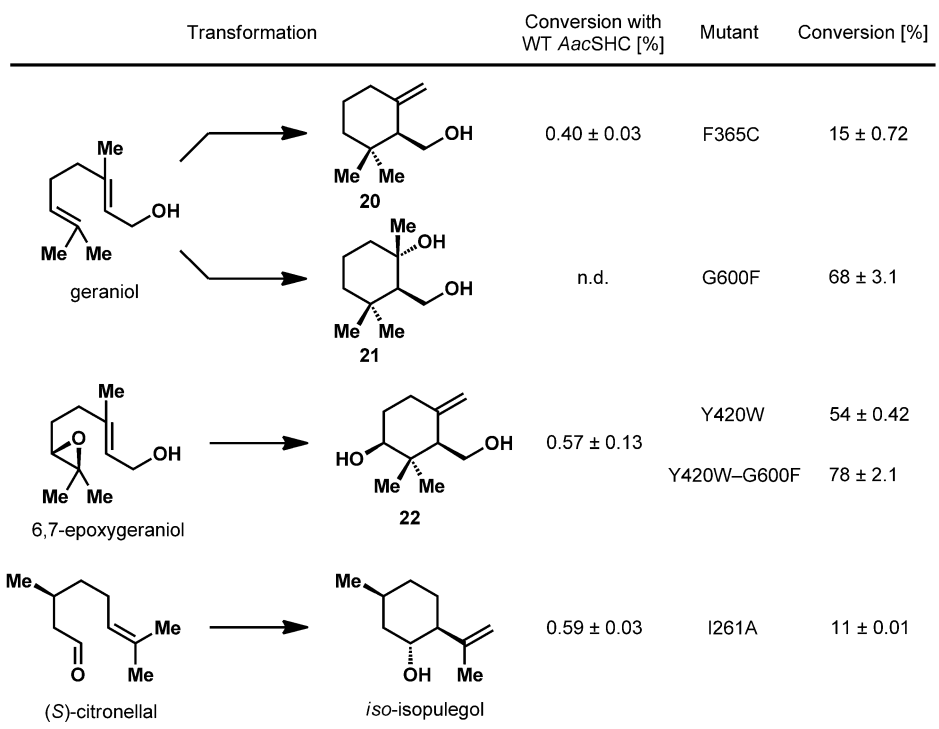


Figure 13. Cyclization reactions of various substrates by the use of functional-group initiators, such as an epoxide and an aldehyde, with *Aac*SHC and identification of enzyme variants with improved cyclization activity.[59]

product (path a), whereas coupling with a halogen ligand results in the chlorinated or brominated product (path b). One fundamental difference between the two enzymes lies in the presence of a coordinating Asp/Glu residue in Fe/αKG hydroxylases, whereas a noncoordinating Ala occupies this position in SyrB2 halogenase. As a result, the Fe center in the latter has an additional coordination site, which is occupied by a halide anion that is eventually incorporated in the product. Given such similarities in structure and mechanism, Walsh and co-workers proposed that Fe/αKG halogenase evolved from Fe/αKG hydroxylase.[63]

With threonine as the substrate, SyrB2 is remarkably selective for the production of 4-chloro-L-threonine, thus suggesting that hydroxyl-radical rebound from intermediate **A** (Figure 14A) is not competitive with halogen-radical rebound. Substitution of the noncoordinating Ala residue with Glu in SyrB2 led to the formation of a hydroxylase-like metal center, but this mutant was shown to be an inefficient hydroxylase.[64] Furthermore, the use of norvaline, a five-carbon-atom amino acid, as a substrate for wild-type SyrB2 led almost exclusively to the hydroxylation product (Figure 14B). These results suggest that a complex interplay between the substrate and the protein active site determines the selectivity between hydroxylation and halogenation. To further reinforce this notion, the simple substitution of coordinating Asp for Ala on prolyl 4-hydroxylase, an Fe/αKG enzyme, led to an inactive enzyme.[65] Thus, despite the presumed evolutionary relationship between the two enzymes, the simplistic notion of creating a vacant coordination site for halide binding to convert a hydroxylase into a halogenase is only successful in very special cases.

Many Fe/αKG hydroxylases have been discovered that act on different types of substrates, and some of these enzymes do not require the substrates to be appended to a carrier protein. Fe/αKG halogenases, however, are relatively rare in nature. The conversion of hydroxylases into halogenases would allow the rapid diversification of secondary metabolites produced by Fe/αKG enzymes. These metabolites include valuable β-lactam antibiotics and modified amino acids and nucleobases.[66]

The chemistry of SyrB2 can also be diversified to include non-natural functions. Matthews et al. recently demonstrated that in the presence of N_3^- or NO_2^-, wild-type SyrB2 could catalyze the radical azidation and nitration of substrates (L-2-aminobutyrate, L-threonine, and L-norvaline) bound to SyrB1 in modest yield under single-turnover conditions.[67] Such reactivity is reminiscent of radical-based C–N coupling in synthetic chemistry, where literature precedent suggests that many nitrogen-containing species, such as azides and nitrite salts, are viable partners in radical coupling reactions.[68] At this point, however, the enzyme-catalyzed reaction requires that the substrates be appended to the carrier protein and occurs in modest yield even under single-turnover conditions. Furthermore, given the apparent complexity of substrate positioning in the active site to determine the outcome of Fe/αKG enzyme-catalyzed reactions, the optimization of these new nitration and azidation reactions may be challenging.

4.2. Synthesis of Unnatural Amino Acids with O-Acetylserine Sulfhydrylase and Tryptophan Synthase

Natural and unnatural amino acids are important constituents of many active pharmaceutical ingredients (APIs); it has been estimated that they comprise 18% of the building blocks used in the pharmaceutical and agrochemical industries.[69] Natural L-amino acids are produced mainly by fermentation and extraction from raw feedstocks. Unnatural amino acids (UAAs), however, are commonly produced by chemical synthesis, as there exists no biosynthetic pathway for the introduction of unnatural side chains. Numerous synthetic methods have been developed for the synthesis of UAAs, but commercial production typically relies on asymmetric hydrogenation or the resolution of racemic mixtures.[70] These processes often require manipulation of protecting groups (which have to be removed) and catalysts that have to be designed de novo for new targets. The development of

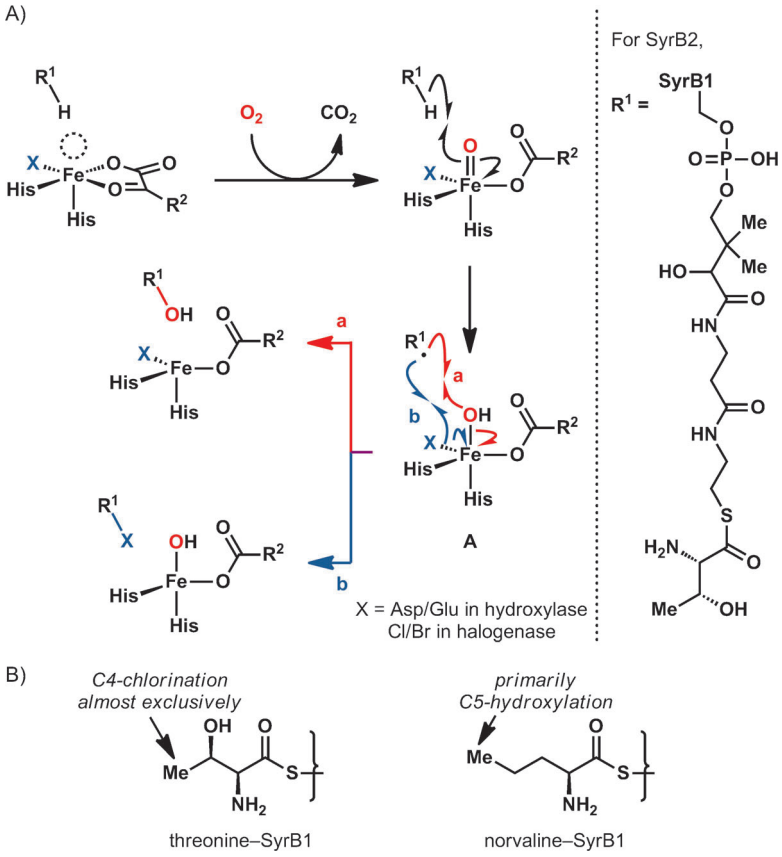


Figure 14. A) Mechanisms of Fe/αKG hydroxylase (path a, in red) and Fe/αKG halogenase (path b, in blue) in which a common reactive intermediate, **A**, is diverted into two reaction pathways depending on the ligand environment around the Fe center.[62] B) Divergent outcome of SyrB2-catalyzed reactions of threonine–SyrB1 and norvaline–SyrB1 indicating the complex interplay between the protein fold and substrate positioning in determining the outcome of the SyrB2-catalyzed reaction.

biosynthetic pathways for UAAs could potentially streamline their production.

O-Acetylserine sulfhydrylase (OASS, this term is used interchangeably with cysteine synthase) is a pyridoxal phosphate (PLP)-dependent enzyme that catalyzes the final step of cysteine biosynthesis.[71] After aldimine formation between *O*-acetylserine and PLP, loss of the acetate group of *O*-acetylserine gives an aminoacrylate intermediate, which then reacts with H_2S to form L-cysteine (Figure 15). In the late 1960s, Giovanelli and Mudd[72a] and Thompson and Moore[72b] independently established that this class of enzymes also catalyzed the synthesis of S-substituted cysteine derivatives, thus hinting at the relaxed substrate specificity of these enzymes. Although rare, heterocyclic β-substituted alanines do occur naturally in plants.[73] Elucidation of the biosynthetic pathways of these nonproteinogenic amino acids showed that they arose from the condensation of *O*-acetyl-L-serine with the appropriate nucleophiles.

Hypothesizing that this pathway shares a common reactive intermediate and reaction mechanism with OASS, Ikegami and co-workers showed in a series of studies that the OASS from higher plants could indeed catalyze the synthesis of β-(pyrazol-1-yl)-L-alanine, L-quisqualic acid, L-mimosine, and several other nonproteinogenic amino acids in low yield.[74] As Maier showed in a follow-up study,[75] the overall metabolic pathway could be engineered to improve the low titers. However, no improvement of the production yield of UAAs by engineering of the OASS enzyme has been disclosed. X-ray crystal structures of CysK1 and CysM, both cysteine synthases, indicated the presence of a substrate tunnel that probably acts as a passageway for incoming nucleophiles.[76] Engineering of this substrate tunnel to accommodate nucleophiles of different sizes could provide an alternative avenue to improve the production of UAAs.

Tryptophan synthase catalyzes the formation of tryptophan through a mechanism that bears a strong resemblance to that of OASS. In the β subunit of the enzyme, the condensation of serine with PLP is followed by dehydration to afford an aminoacrylate intermediate, which then undergoes a conjugate addition with indole.[77] Just like OASS, tryptophan synthase possesses a hydrophobic tunnel for the passage of indole. The wild-type enzyme has been shown to catalyze the production of various tryptophan analogues through the use of the corresponding heterocyclic nucleophiles, such as thienopyrroles, azaindoles, and indazole (Figure 16).[78] At present, however, optimization of this enzyme is still required to make the process practical. As with OASS, directed evolution and protein engineering could render this enzyme more useful for the production of UAAs.

5. Conclusions

Over the last fifty years, chemists have invented creative synthetic disconnections that are not found in nature. We believe that at least some of these reactions could be imported into biological systems. The challenge for engineering enzymes to catalyze non-natural chemistry is that there is not another enzyme to provide the inspiration or guide the engineering. That leap has to come from luck (accidental discovery of an interesting promiscuous activity), laborious screening of enzymes for non-natural functions, or, better, from chemical intuition/design based on known synthetic transformations. The carbene- and nitrene-transfer reactions catalyzed by engineered cytochrome P450s are a good example of how new enzymes can be generated by an approach that mimics nature and is based on chemical knowledge: Studies with transition-metal catalysts and mechanistic similarities provided the inspiration, the promiscuity of the natural P450 enzymes provided a starting point, and protein engineering/evolution provided the means to tune reactivity and selectivity. Similarly, the various synthetic methods developed for C–N coupling (for which there is no

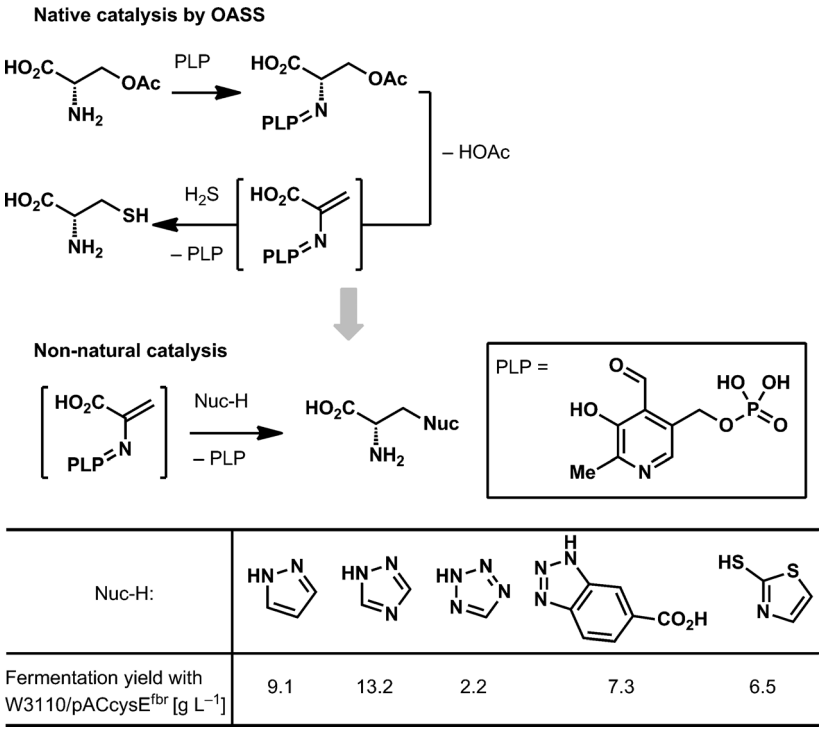


Figure 15. Use of *O*-acetylserine sulfhydrylase in the synthesis of unnatural β-substituted alanine derivatives and fermentation yields with *Escherichia coli* strain W3110/pACcysE^{fbr}.[75]

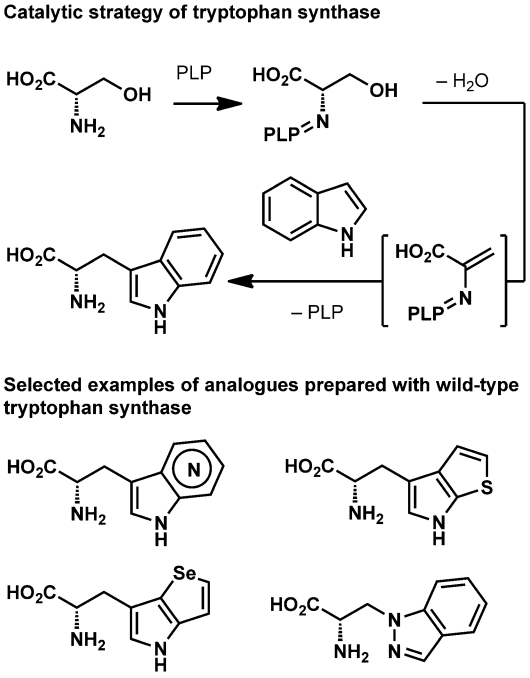


Figure 16. Mechanism of tryptophan synthase and some representative UAAs prepared with wild-type tryptophan synthase.[77,78]

natural counterpart) served as a motivation to investigate an enzymatic equivalent in SyrB2 halogenase.

But can every poorly active enzyme be engineered or evolved for high activity? Clearly not, or all weakly active computationally designed enzymes,[79] catalytic antibodies, or bovine serum albumin for that matter could be evolved in the laboratory to be highly efficient. Because we do not yet fully understand the features of an effective scaffold or starting point for the design or evolution of catalysis,[80] we find it prudent to start with enzyme scaffolds that have proven ability to evolve (i.e. have diverged naturally to catalyze different reactions). Experience shows that enzymes such as the cytochrome P450s or members of other functionally diverse superfamilies that have already diverged to catalyze many different reactions are also readily evolved in the laboratory, at least when the reactions share mechanistic features.[81] This divergence includes the evolution of catalytic activities that have no natural counterpart, such as olefin cyclopropanation with diazo compounds or cyanation through epoxide opening. If one particular enzyme does not exhibit a desired promiscuous activity, other family members or even close variants may, as was shown for the MBL family[21] and for the P450s.[45]

We can now begin to sketch out some general guidelines for engineering enzymes to catalyze reactions not known in nature: 1) For a given function, look for the most important feature(s) that enables the transformation, for example, an iron carbenoid for cyclopropanation, the acid-catalyzed formation of an oxonium ion in the Prins reaction, or something as simple as the presence of a Michael acceptor for UAA synthesis. 2) Establish this feature or key intermediate in an enzyme on the basis of known reactivity or mechanistic analogy, for example, carbene is isoelectronic to oxene, a key intermediate in monooxygenation, or Brønsted acid activation is a key step in many terpene synthase cyclizations. 3) Evaluate variants of the enzyme or closely related enzymes for the desired promiscuous activity. 4) Use directed evolution to improve the nonnatural activity or tune selectivity. Computational approaches may be able to assist this discovery process, possibly for the evaluation of suitable enzyme starting points in silico.[82]

Of course, the evolution and engineering of existing enzymes is not the only possible approach to the creation of new enzymes. Good progress has been made with artificial metalloenzymes and de novo enzyme design, but significant challenges remain, especially for the creation of synthetically useful catalysts and catalysts that function inside cells. We believe that a key advantage of the evolutionary approach reviewed herein is that the starting point for the creation of a new enzyme is an existing enzyme that is functionally expressed in a microbial host and that can be improved by directed evolution, or at least by genetic modification. The ability to evolve the new function in the laboratory, starting from an evolvable scaffold (an existing active site), greatly increases the chance that synthetically useful catalysts will emerge. The evolutionary approach we have described is limited to systems for which suitable starting enzymes exist. However, when such enzymes exist, progress to synthetically

useful catalysts can be very rapid.[33,47] For advances in the fields of artificial metalloenzymes and de novo enzyme design, we direct readers to the excellent reviews in Refs. [83] and [84].

Advances in mechanistic enzymology have allowed us to develop a greater understanding of the chemical basis of enzyme catalysis. This knowledge will help us select mechanisms and intermediates to "hijack" enzymes for non-natural catalysis. Protein engineering and enzyme evolution are also progressing rapidly. Armed with these tools, we believe that exciting times are ahead for bridging the gap between nature's chemical repertoire and the synthetic world.

Our research is supported by the Department of the Navy, Office of Naval Research (grant N00014-11-10205), the Jacobs Institute for Molecular Engineering for Medicine at Caltech, the Gordon and Betty Moore Foundation through grant GBMF2809 to the Caltech Programmable Molecular Technology Initiative, and the National Science Foundation, Office of Chemical, Bioengineering, Environmental and Transport Systems SusChEM Initiative (grant CBET-1403077). Z.J.W. was supported by a Ruth L. Kirschstein Fellowship from the National Institutes of Health (F32EB015846). We thank Prof. Dan S. Tawfik, Prof. Bernhard Hauer, Dr. John A. McIntosh, Dr. Andrew R. Buller, Dr. Sheel C. Dodani, and Dr. Christopher K. Prier for helpful suggestions and discussions. The content of this Review is solely the responsibility of the authors and does not represent the official views of the funding agencies.

Received: September 25, 2014
Published online: February 3, 2015

[1] a) J. C. Moore, F. H. Arnold, *Nat. Biotechnol.* **1996**, *14*, 458 – 467; b) Z. Shao, F. H. Arnold, *Curr. Opin. Struct. Biol.* **1996**, *6*, 513 – 518; c) F. H. Arnold, *Chem. Eng. Sci.* **1996**, *51*, 5091 – 5102; d) F. H. Arnold, J. C. Moore, *Adv. Biochem. Eng./Biotechnol.* **1997**, *58*, 1 – 14; e) M. T. Reetz, A. Zonta, k. Schimossek, K. Liebeton, K.-E. Jaeger, *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 2830 – 2832; *Angew. Chem.* **1997**, *109*, 2961 – 2963; f) M. T. Reetz, *J. Am. Chem. Soc.* **2013**, *135*, 12480 – 12496.

[2] a) C. K. Savile, J. M. Janey, E. C. Mundorff, J. C. Moore, S. Tam, W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz, J. Brands, P. N. Devine, G. W. Huisman, G. J. Hughes, *Science* **2010**, *329*, 305 – 309; b) J. Liang, J. Lalonde, B. Borup, V. Mitchell, E. Mundorff, N. Trinh, D. A. Kochrekar, R. N. Cherat, G. G. Pai, *Org. Process Res. Dev.* **2010**, *14*, 193 – 198; c) U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, K. Robins, *Nature* **2012**, *485*, 185 – 194; d) M. Alexeeva, A. Enright, M. J. Dawson, M. Mahmoudian, N. J. Turner, *Angew. Chem. Int. Ed.* **2002**, *41*, 3177 – 3180; *Angew. Chem.* **2002**, *114*, 3309 – 3312; e) N. J. Turner, *Nat. Chem. Biol.* **2009**, *5*, 567 – 573.

[3] R. Wohlgemuth, *Curr. Opin. Biotechnol.* **2010**, *21*, 713 – 724.

[4] a) E. J. Steen, Y. Kang, G. Bokinsky, Z. Hu, A. Schirmer, A. McClure, S. B. del Cardayre, J. D. Keasling, *Nature* **2010**, *463*, 559 – 562; b) P. P. Peralta-Yahya, F. Zhang, S. B. del Cardayre, J. D. Keasling, *Nature* **2012**, *488*, 320 – 328; c) S. Atsumi, T. Hanai, J. C. Liao, *Nature* **2008**, *451*, 86 – 89; d) S. Y. Lee, J. H. Park, S. H. Jang, L. K. Nielsen, J. Kim, K. S. Jung, *Biotechnol. Bioeng.* **2008**, *101*, 209 – 228.

[5] R. A. Jensen, *Annu. Rev. Microbiol.* **1976**, *30*, 409 – 425.

[6] a) P. J. O'Brien, D. Herschlag, *Chem. Biol.* **1999**, *6*, R91 – R105; b) S. D. Copley, *Curr. Opin. Chem. Biol.* **2003**, *7*, 265 – 272.

[7] O. Khersonsky, D. S. Tawfik, *Annu. Rev. Biochem.* **2010**, *79*, 471 – 505.

[8] D. S. Tawfik, *Nat. Chem. Biol.* **2010**, *6*, 692 – 696.

[9] J. L. Seffernick, L. P. Wackett, *Biochemistry* **2001**, *40*, 12747 – 12753.

[10] A. M. Cook, H. Grossenbacher, R. Hütter, *Experientia* **1983**, *39*, 1191 – 1198.

[11] J. L. Seffernick, M. L. De Souza, M. J. Sadowsky, L. P. Wackett, *J. Bacteriol.* **2001**, *183*, 2405 – 2410.

[12] N. Udiković-Kolić, C. Scott, F. Martin-Laurent, *Appl. Microbiol. Biotechnol.* **2012**, *96*, 1175 – 1189.

[13] a) C. Scott, C. J. Jackson, C. W. Coppin, R. G. Mourant, M. E. Hilton, T. D. Sutherland, R. J. Russell, J. G. Oakeshott, *Appl. Environ. Microbiol.* **2009**, *75*, 2184 – 2191; b) S. Noor, M. C. Taylor, R. J. Russell, L. S. Jermiin, C. J. Jackson, J. G. Oakeshott, C. Scott, *PLoS One* **2012**, *7*, e39822.

[14] A. E. Todd, C. A. Orengo, J. M. Thornton, *J. Mol. Biol.* **2001**, *307*, 1113 – 1143.

[15] C. M. Serdar, D. T. Gibson, D. M. Munnecke, J. H. Lancaster, *Appl. Environ. Microbiol.* **1982**, *44*, 246 – 249.

[16] C. M. Seibert, F. M. Raushel, *Biochemistry* **2005**, *44*, 6383 – 6391.

[17] L. Afriat, C. Roodveldt, G. Manco, D. S. Tawfik, *Biochemistry* **2006**, *45*, 13677 – 13686.

[18] M. B. Miller, B. L. Bassler, *Annu. Rev. Microbiol.* **2001**, *55*, 165 – 199.

[19] M. M. Meier, C. Rajendran, C. Malisi, N. G. Fox, C. Xu, S. Schlee, D. P. Barondeau, B. Höcker, R. Sterner, F. M. Raushel, *J. Am. Chem. Soc.* **2013**, *135*, 11670 – 11677.

[20] L. Afriat-Jurnou, C. J. Jackson, D. S. Tawfik, *Biochemistry* **2012**, *51*, 6047 – 6055.

[21] F. Baier, N. Tokuriki, *J. Mol. Biol.* **2014**, *426*, 2442 – 2456.

[22] a) P. C. Babbitt, J. A. Gerlt, *Adv. Protein Chem.* **2001**, *55*, 1 – 28; b) J. A. Gerlt, P. C. Babbitt, *Curr. Opin. Chem. Biol.* **2009**, *13*, 10 – 18.

[23] M. Elias, D. S. Tawfik, *J. Biol. Chem.* **2012**, *287*, 11 – 20.

[24] H.-S. Park, S.-H. Nam, J. K. Lee, C. N. Yoon, B. Mannervik, S. J. Benkovic, H.-S. Kim, *Science* **2006**, *311*, 535 – 538.

[25] J. Shanklin, J. E. Guy, G. Mishra, Y. Lindqvist, *J. Biol. Chem.* **2009**, *284*, 18559 – 18563.

[26] P. Broun, J. Shanklin, E. Whittle, C. Somerville, *Science* **1998**, *282*, 1315 – 1317.

[27] J. A. Broadwater, E. Whittle, J. Shanklin, *J. Biol. Chem.* **2002**, *277*, 15613 – 15620.

[28] M. Ben-David, M. Elias, J. J. Filippi, E. Duñach, I. Silman, J. L. Sussman, D. S. Tawfik, *J. Mol. Biol.* **2012**, *418*, 181 – 196.

[29] M. L. De Souza, M. J. Sadowsky, L. P. Wackett, *J. Bacteriol.* **1996**, *178*, 4894 – 4900.

[30] J. E. T. van Hylckama Vlieg, L. Tang, J. H. L. Spelberg, T. Smilda, G. J. Poelarends, T. Bosma, A. E. J. van Merode, M. W. Fraaije, D. B. Janssen, *J. Bacteriol.* **2001**, *183*, 5058 – 5066.

[31] a) R. M. de Jong, J. J. W. Tiesinga, H. J. Rozeboom, K. H. Kalk, L. Tang, D. B. Janssen, B. W. Dijkstra, *EMBO J.* **2003**, *22*, 4933 – 4944; b) R. M. de Jong, J. J. W. Tiesinga, A. Villa, L. Tang, D. B. Janssen, B. W. Dijkstra, *J. Am. Chem. Soc.* **2005**, *127*, 13338 – 13343.

[32] a) J. H. L. Spelberg, J. E. T. van Hylckama Vlieg, L. Tang, D. B. Janssen, R. M. Kellogg, *Org. Lett.* **2001**, *3*, 41 – 43; b) G. Hasnaoui-Dijoux, M. M. Elenkov, J. H. L. Spelberg, B. Hauer, D. B. Janssen, *ChemBioChem* **2008**, *9*, 1048 – 1051.

[33] R. J. Fox, S. C. Davis, E. C. Mundorff, L. M. Newman, V. Gavrilovic, S. K. Ma, L. M. Chung, C. Ching, S. Tam, S. Muley, J. Grate, J. Gruber, J. C. Whitman, R. A. Sheldon, G. W. Huisman, *Nat. Biotechnol.* **2007**, *25*, 338 – 344.

[34] a) R. J. Fox, *J. Theor. Biol.* **2005**, *234*, 187 – 199; b) R. J. Fox, *Protein Eng.* **2003**, *16*, 589 – 597.

[35] M. Jahn, J. Marles, R. A. J. Warren, S. G. Withers, *Angew. Chem. Int. Ed.* **2003**, *42*, 352 – 354; *Angew. Chem.* **2003**, *115*, 366 – 368.
[36] Z. J. Witczak, *Curr. Med. Chem.* **1999**, *6*, 165 – 178.
[37] a) G. Vic, D. H. G. Crout, *Tetrahedron: Asymmetry* **1994**, *5*, 2513 – 2516; b) J. B. Kempton, S. G. Withers, *Biochemistry* **1992**, *31*, 9961 – 9969.
[38] D. E. Koshland, *Biol. Rev.* **1953**, *28*, 416 – 436.
[39] J. Müllegger, M. Jahn, H.-M. Chen, R. A. J. Warren, S. G. Withers, *Protein Eng. Des. Sel.* **2005**, *18*, 33 – 40.
[40] a) J. L. Maxwell, K. C. Brown, D. W. Bartley, T. Kodadek, *Science* **1992**, *256*, 1544 – 1547; b) J. R. Wolf, C. G. Hamaker, J.-P. Djukic, T. Kodadek, L. K. Woo, *J. Am. Chem. Soc.* **1995**, *117*, 9194 – 9199.
[41] a) Y. Li, J.-S. Huang, Z.-Y. Zhou, C.-M. Che, X.-Z. You, *J. Am. Chem. Soc.* **2002**, *124*, 13185 – 13193; b) Z. Guo, X. Guan, J.-S. Huang, W.-M. Tsui, Z. Lin, C.-M. Che, *Chem. Eur. J.* **2013**, *19*, 11320 – 11331.
[42] R. Breslow, S. H. Gellman, *J. Am. Chem. Soc.* **1983**, *105*, 6728 – 6729.
[43] E. W. Svastits, J. H. Dawson, R. Breslow, S. H. Gellman, *J. Am. Chem. Soc.* **1985**, *107*, 6427 – 6428.
[44] P. S. Coelho, E. M. Brustad, A. Kannan, F. H. Arnold, *Science* **2013**, *339*, 307 – 310.
[45] T. Heel, J. A. Mcintosh, S. C. Dodani, J. T. Meyerowitz, F. H. Arnold, *ChemBioChem* **2014**, *15*, 2556 – 2562.
[46] P. S. Coelho, Z. J. Wang, M. E. Ener, S. A. Baril, A. Kannan, F. H. Arnold, E. M. Brustad, *Nat. Chem. Biol.* **2013**, *9*, 485 – 487.
[47] Z. J. Wang, H. Renata, N. E. Peck, C. C. Farwell, P. S. Coelho, F. H. Arnold, *Angew. Chem. Int. Ed.* **2014**, *53*, 6810 – 6813; *Angew. Chem.* **2014**, *126*, 6928 – 6931.
[48] H. Renata, Z. J. Wang, R. Z. Kitto, F. H. Arnold, *Catal. Sci. Technol.* **2014**, *4*, 3640 – 3643.
[49] Z. J. Wang, N. E. Peck, H. Renata, F. H. Arnold, *Chem. Sci.* **2014**, *5*, 598 – 601.
[50] a) 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.* **2013**, *52*, 9309 – 9312; *Angew. Chem.* **2013**, *125*, 9479 – 9482; b) C. C. Farwell, J. A. McIntosh, T. K. Hyster, Z. J. Wang, F. H. Arnold, *J. Am. Chem. Soc.* **2014**, *136*, 8766 – 8771; c) T. K. Hyster, C. C. Farwell, A. R. Buller, J. A. McIntosh, F. H. Arnold, *J. Am. Chem. Soc.* **2014**, *136*, 15505 – 15508.
[51] a) R. Singh, M. Bordeaux, R. Fasan, *ACS Catal.* **2014**, *4*, 546 – 552; b) M. Bordeaux, R. Singh, R. Fasan, *Bioorg. Med. Chem.* **2014**, *22*, 5697 – 5704.
[52] S. C. Hammer, P.-O. Syrén, M. Seitz, B. M. Nestl, B. Hauer, *Curr. Opin. Chem. Biol.* **2013**, *17*, 293 – 300.
[53] a) K.-U. Wendt, A. Lenhart, G. E. Schulz, C. Feil, K. Poralla, *Protein Sci.* **1997**, *6*, 722 – 724; b) K. U. Wendt, *Science* **1997**, *277*, 1811 – 1815.
[54] S. Neumann, H. Simon, *Biol. Chem. Hoppe-Seyler* **1986**, *367*, 723 – 730.
[55] a) M. Seitz, P.-O. Syren, L. Steiner, J. Klebensberger, B. M. Nestl, B. Hauer, *ChemBioChem* **2013**, *14*, 436 – 439; b) M. Seitz, J. Klebensberger, S. Siebenhaller, M. Breuer, G. Siedenburg, D. Jendrossek, B. Hauer, *J. Mol. Catal. B* **2012**, *84*, 72 – 77.
[56] Y. Yonemura, T. Ohyama, T. Hoshino, *Org. Biomol. Chem.* **2012**, *10*, 440 – 446.
[57] G. Siedenburg, D. Jendrossek, M. Breuer, B. Juhl, J. Pleiss, M. Seitz, J. Klebensberger, B. Hauer, *Appl. Environ. Microbiol.* **2012**, *78*, 1055 – 1062.
[58] G. Siedenburg, M. Breuer, D. Jendrossek, *Appl. Microbiol. Biotechnol.* **2013**, *97*, 1571 – 1580.
[59] S. C. Hammer, A. Marjanovic, J. M. Dominicus, B. M. Nestl, B. Hauer, *Nat. Chem. Biol.* **2014**, in press, DOI: 10.1038/nchembio.1719.
[60] a) W. S. Johnson, R. B. Kinnel, *J. Am. Chem. Soc.* **1966**, *88*, 3861 – 3862; b) J. A. Bender, A. E. Blize, C. C. Browder, S. Giese, F. G. West, *J. Org. Chem.* **1998**, *63*, 2430 – 2431; c) R. R. Knowles, S. Lin, E. N. Jacobsen, *J. Am. Chem. Soc.* **2010**, *132*, 5030 – 5032; d) W.-W. Qiu, K. Surendra, L. Yin, E. J. Corey, *Org. Lett.* **2011**, *13*, 5893 – 5895; e) M. A. Schafroth, D. Sarlah, S. Krautwald, E. M. Carreira, *J. Am. Chem. Soc.* **2012**, *134*, 20276 – 20278.
[61] F. H. Vaillancourt, J. Yin, C. T. Walsh, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 10111 – 10116.
[62] S. D. Wong, M. Srnec, M. L. Matthews, L. V. Liu, Y. Kwak, K. Park, C. B. Bell III, E. E. Alp, J. Zhao, Y. Yoda, S. Kitao, M. Seto, C. Krebs, J. M. Bollinger, E. I. Solomon, *Nature* **2013**, *499*, 320 – 323.
[63] F. H. Vaillancourt, E. Yeh, D. A. Vosburg, S. Garneau-Tsodikova, C. T. Walsh, *Chem. Rev.* **2006**, *106*, 3364 – 3378.
[64] M. L. Matthews, C. S. Neumann, L. A. Miles, T. L. Grove, S. J. Booker, C. Krebs, C. T. Walsh, J. M. Bollinger, Jr., *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 17723 – 17728.
[65] K. L. Gorres, K. H. Pua, R. T. Raines, *PLoS One* **2009**, *4*, e7635.
[66] M. Costas, M. P. Mehn, M. P. Jensen, L. Que, Jr., *Chem. Rev.* **2004**, *104*, 939 – 986.
[67] M. L. Matthews, W.-C. Chang, A. P. Layne, L. A. Miles, C. Krebs, J. M. Bollinger, *Nat. Chem. Biol.* **2014**, *10*, 209 – 215.
[68] a) E. K. Leggans, T. J. Barker, K. K. Duncan, D. L. Boger, *Org. Lett.* **2012**, *14*, 1428 – 1431; b) J. Waser, H. Nambu, E. M. Carreira, *J. Am. Chem. Soc.* **2005**, *127*, 8294 – 8295.
[69] C. Boswell, *Chem. Market Reporter* June 14 (1999), pp. 9 – 11.
[70] K. Drauz, *Chimia* **1997**, *51*, 310 – 314.
[71] C.-H. Tai, P. F. Cook, *Acc. Chem. Res.* **2001**, *34*, 49 – 59.
[72] a) J. Giovanelli, S. H. Mudd, *Biochem. Biophys. Res. Commun.* **1968**, *31*, 275 – 280; b) J. F. Thompson, D. P. Moore, *Biochem. Biophys. Res. Commun.* **1968**, *31*, 281 – 285.
[73] F. Ikegami, I. Murakoshi, *Phytochemistry* **1994**, *35*, 1089 – 1104.
[74] a) I. Murakoshi, M. Kaneko, C. Koide, F. Ikegami, *Phytochemistry* **1986**, *25*, 2759 – 2763; b) F. Ikegami, M. Mizuno, M. Kihara, I. Murakoshi, *Phytochemistry* **1990**, *29*, 3461 – 3465; c) I. Murakoshi, F. Ikegami, M. Kaneko, *Phytochemistry* **1985**, *24*, 1907 – 1911.
[75] T. H. P. Maier, *Nat. Biotechnol.* **2003**, *21*, 422 – 427.
[76] a) R. Schnell, W. Oehlmann, M. Singh, G. Schneider, *J. Biol. Chem.* **2007**, *282*, 23473 – 23481; b) M. T. Claus, G. E. Zocher, T. H. P. Maier, G. E. Schulz, *Biochemistry* **2005**, *44*, 8620 – 8626.
[77] M. F. Dunn, D. Niks, H. Ngo, T. R. M. Barends, I. Schlichting, *Trends Biochem. Sci.* **2008**, *33*, 254 – 264.
[78] R. S. Phillips, *Tetrahedron: Asymmetry* **2004**, *15*, 2787 – 2792.
[79] a) L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Rothlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas III, D. Hilvert, K. N. Houk, B. L. Stoddard, D. Baker, *Science* **2008**, *319*, 1387 – 1391; b) E. A. Althoff, L. Wang, L. Jiang, L. Giger, J. K. Lassila, Z. Wang, M. Smith, S. Hari, P. Kast, D. Herschlag, D. Hilvert, D. Baker, *Protein Sci.* **2012**, *21*, 717 – 726; c) G. Kiss, N. Çelebi-Ölçüm, R. Moretti, K. N. Houk, D. Baker, *Angew. Chem. Int. Ed.* **2013**, *52*, 5700 – 5725; *Angew. Chem.* **2013**, *125*, 5810 – 5836.
[80] D. Baker, *Protein Sci.* **2010**, *19*, 1817 – 1819.
[81] J. A. McIntosh, C. C. Farwell, F. H. Arnold, *Curr. Opin. Chem. Biol.* **2014**, *19*, 126 – 134.
[82] S. Chakraborty, B. J. Rao, *PLoS One* **2012**, *7*, e32011.
[83] a) M. Dürrenberger, T. R. Ward, *Curr. Opin. Chem. Biol.* **2014**, *19*, 99 – 106; b) J. Bos, G. Roelfes, *Curr. Opin. Chem. Biol.* **2014**, *19*, 135 – 143.
[84] H. Kries, R. Blomberg, D. Hilvert, *Curr. Opin. Chem. Biol.* **2013**, *17*, 221 – 228.