Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Angew Chem Int Ed Engl. Author manuscript; available in PMC 2016 March 09.

Published in final edited form as:

Angew Chem Int Ed Engl. 2015 March 9; 54(11): 3351–3367. doi:10.1002/anie.201409470.

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

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

Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 E. California Blvd. MC 210-41, Pasadena, CA 91125 (USA)

Frances H. Arnold: fha@cheme.caltech.edu

Abstract

High selectivities and exquisite control over reaction outcomes entice chemists to use biocatalysts in organic synthesis. However, many useful reactions are not accessible because they are not in nature's known repertoire. We will use this review to 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 progressions have been recapitulated in the laboratory starting from extant enzymes. We then examine non-native enzyme activities that have been discovered and exploited for chemical synthesis, emphasizing reactions that do not have natural counterparts. The new functions have mechanistic parallels to the native reaction mechanisms that often manifest as catalytic promiscuity and the ability to convert from one function to the other with minimal mutation. We present examples of how non-natural activities have been improved by directed evolution, mimicking the process used by nature to create new catalysts. Examples of new enzyme functions include epoxide opening reactions with non-natural nucleophiles catalyzed by a laboratory-evolved halohydrin dehalogenase, cyclopropanation and other carbene transfer reactions catalyzed by cytochrome P450 variants, and non-natural modes of cyclization by a modified terpene synthase. Lastly, we describe discoveries of non-native catalytic functions that may provide future opportunities for expanding the enzyme universe.

Keywords

biocatalysis; enzyme catalyst; non-natural function; promiscuity; protein engineering

1. Introduction

Replete with nature's solutions to catalyzing chemical transformations, our burgeoning genomic databases beautifully illustrate how evolution generates chemical innovation in the form of new enzymes. Today's enormous biocatalytic diversity 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:

Correspondence to: Frances H. Arnold, fha@cheme.caltech.edu.

The content of this paper is solely the responsibility of the authors and does not represent the official views of the funding agencies.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

new enzymes continue to appear in response to (often man-made) opportunities to survive challenges (*e.g.* antibiotic resistance) or occupy new niches (*e.g.* catabolize man-made compounds). Given nature's ability to innovate, and our extremely limited ability to design new enzymes, we argue for using 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 fine-tune enzymes for chemical synthesis.[1] A simple strategy of accumulating mutations via iterative mutagenesis and screening for desired functions can effectively optimize properties of interest—activity on non-native substrates, enantioselectivity, product selectivity, stability and more—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 is a high bar that no 'rational design' approach has been able to meet on a regular basis, but one where 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], production of chemicals from renewable resources,[4] and synthesis of complex natural products.

Directed evolution can be used to diversify existing enzymes, 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 of nature's approach to inventing new catalysts in order to jumpstart the evolution of new enzyme families in the laboratory. We illustrate this evolutionary approach to catalyst discovery by starting with some examples from nature's repertoire. In some cases, scientists have elucidated sequence and functional pathways that connect existing enzymes in order to demonstrate how natural evolution may have proceeded from one function 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? That the answer is an emphatic "Yes!" is illustrated with 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.

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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

through 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 1A).[5] But even today's 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, providing a platform for evolution of new functions by natural selection (Figure 1B).[7]

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 2A) 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, another 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 at only nine 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 2B), enabling the bacteria to capitalize on a new opportunity to use these synthetic compounds as nitrogen sources.[12]

Progression from melamine to triazine degradation represents 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. In contrast, enzymes in the aminohydrolase (*e.g*. TriA) family typically utilize a divalent metal such as 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 to an enzyme with atrazine chlorohydrolase activity comparable to AtzA.[13] Cys331Ser and Asp328Asn, mutations suggested by homology modeling to lie in the active site, improve the k_{cat}/K_M for atrazine hydrolysis dramatically, from 60 $M^{-1}s^{-1}$ to close to 10,000 $M^{-1}s^{-1}$, while completely abolishing melamine hydrolysis activity. Furthermore, Scott and co-workers showed that one can arrive at an atrazine chlorohydrolase by accumulating single beneficial mutations in a simple uphill

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

walk, starting from TriA. The Cys331Ser mutation alone improved the k_{cat}/K_M for atrazine hydrolysis by almost 30-fold. Introduction of Asp328Asn to the Cys331Ser mutant conferred a further six-fold 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 coworkers reasoned that the Asn/Ser dyad (positions 328 and 331 in AtzA) assists in the expulsion of 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. But several research groups have taken on this challenge, using 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 its 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 having promiscuous organophosphate hydrolysis activity. The ancestral enzyme is unknown, however, as no very close sequence homolog has been identified.

The PTE from *P. diminuta* has the $(\beta/\alpha)_8$-barrel fold and binuclear metal center common to many members of the amidohydrolase (AHS) superfamily that hydrolyze different classes of substrates.[16] Afriat *et al*. proposed that this PTE's promiscuous lactonase activity could be a vestige of its ancestral source, a clue to its ancestral function.[17] They showed that three microbial enzymes that are the closest known homologs to *P. diminuta* PTE (~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*-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 to a PTE.[19] Using a

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

combination of rational and random mutagenesis, they were able to convert the PLL from *Deinococcus radiodurans* (*Dr*PLL) to an enzyme exhibiting PTE activity that is five orders of magnitude higher than the wild-type *Dr*PLL and just one order of magnitude lower than 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 are potential new nutrient sources (Figure 3B). Unlike AtzA and TriA, however, PTE has low sequence identity to its closest known homolog, 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. That this could be done 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 accumulating 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 MBL superfamily members, despite their low sequence identities (*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 divalent metal activation of a water molecule. Thus, 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 likely evolved from lactonases, in parallel with the PTEs in the AHS superfamily and in what appears to be an example of 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 one could 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 to an enzyme with high β-lactamase activity using directed evolution.[24]

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

2.4. Evolution of a hydroxylase from a desaturase

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

When expressed in yeast, the hydroxylase from *L. fendleri*, LFAH12, was found to have appreciable desaturase activity in addition to its native hydroxylase activity (producing *ca.* 1:1 di-unsaturated fatty acids:hydroxylated fatty acid).[26,27] In contrast, the desaturase from *A. thaliana*, FAD2, was shown to catalyze desaturation almost exclusively, with only very minor hydroxylation products detected (hydroxylation:desaturation product ratio of 0.006). These two enzymes are close relatives, with ~81% sequence identity, and a 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] Based on this, 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 sufficient to convert FAD2 into a hydroxylase. Further studies by Broadwater *et al.*[27] showed that a single mutation was sufficient to achieve a comparable boost in hydroxylase activity in FAD2.

This desaturase-hydroxylase example demonstrates how easily 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 evolution of the new functions—the new enzymes were obtained in the laboratory starting from the extant relative. And, as these examples suggest, a few mutations can be sufficient to convert an existing enzyme to a new one with distinct, but mechanistically-related activity. [22]

We have chosen just a few examples to illustrate how enzymes catalyzing different reactions can diverge from a common ancestor, especially when the activities overlap at least a little. Nature does it, and the laboratory experiments demonstrate just how easily it can happen. Although nature may not have taken 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 using this strategy to make enzymes that catalyze reactions not already discovered in nature.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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, today's plethora of biocatalysts provides equal or perhaps even greater opportunities for innovation than in 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 melamine. Indeed, TriA possesses a low level of atrazine chlorohydrolase activity ($k_{cat}/K_M = 60\ s^{-1}M^{-1}$). From there, 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, accumulating 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 being able 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. And, because low levels of catalytic activity exhibited by a promiscuous enzyme can be improved in many cases by engineering the protein sequence, there is the reasonable expectation that a non-natural, but mechanistically-related promiscuous activity can similarly improve.

In fact, several laboratories have used this approach and capitalized on the catalytic promiscuity of enzymes and the similarity between the native and desired reaction mechanisms in order 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 will present some examples of novel non-natural functions that have been discovered using this mechanism-based approach and then improved to useful levels by protein engineering and especially directed evolution.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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 structures of halohydrin dehalogenase from *Agrobacterium radiobacter* AD1 suggested the presence of a binding site for both the epoxide and halide anion,[31] thus raising the possibility of using pseudohalides of varying sizes in the reverse epoxide opening reaction. Indeed, Janssen and coworkers found that this enzyme accepts a wide range of non-natural nucleophiles such as azide, nitrite, cyanate and thiocyanate[32] and used it for kinetic resolution of various epoxides to give the ring-opened products with high enantioselectivity (Figure 6A). In addition, high selectivity for opening at the terminal position was observed. In the case of aryl epoxide substrates, this finding stood in stark contrast to the non-catalyzed ring opening reaction, where 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 ~4000-fold (Figure 6B).[34] This catalyst enables production of **6** at >99.9%ee with a substrate loading of 130 g/L.

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

3.2. Synthesis of S-oligosaccharides 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 early work of Withers. By examining the catalytic strategy 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 synthesis of *S*-oligosaccharides.[35]

S-oligosaccharides 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 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 *S*-oligosaccharides.[37]

Glycosidases catalyze the hydrolytic cleavage of glycosidic bonds and are mainly responsible for the degradation of carbohydrate-based biomass. In low water concentration, these enzymes are also capable of catalyzing glycoside exchange. For retaining β-glycosidases, the enzymatic hydrolysis and glycoside exchange reaction occurs through a double substitution mechanism where a catalytic nucleophile residue first displaces the

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

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]

To create a new activity, it is helpful if the starting enzyme already exhibits it at some level as a promiscuous activity, but this 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 (Figure 7B), Withers and co-workers re-designed the active site of β-glucosidase from *Agrobacterium sp*. (Abg) and β-mannosidase (Man2A) from *Cellulomonas fimi* such that they would only catalyze the glycosylation of activated dinitrophenyl (DNP) glycoside donors and deoxythio sugars as acceptors. The former does not require acid activation since it is a good enough leaving group to be displaced by the catalytic nucleophile residue, and the latter contains 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 yields (Figure 8). Notably, no protecting groups were required on the free hydroxyl groups of the glycoside donors and acceptors. In further work, Withers and co-workers showed that mutations at the catalytic acid/base residue could boost the thioglycoligase activities 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 metallocarbenoid reactive intermediates for cyclopropanation reactions is well-documented in the synthetic literature (Figure 9A).[40] Whereas carbenes are generally too reactive for characterization, some of these metallocarbenoid species proved to be stable enough for isolation and X-ray crystallography analysis.[41] These isolated metallocarbenoids participate in cyclopropanation reactions, providing evidence that metalloporphyrin-catalyzed cyclopropanations proceed through the metallocarbenoid 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 reacted with iminoiodinanes,[42] presumably via a metallonitrenoid 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 in very low turnover.[43] Given the similarity between carbene/nitrene and oxene–the reactive species in P450 monooxygenation reactions–in electronic configuration and thus reactivity (Figure 9B), our laboratory hypothesized that cytochrome P450s may exhibit promiscuous activity for carbene and nitrene transfer reactions and that such activities could be improved by protein engineering. Addition of ethyl diazoacetate to styrene in the presence of wild-type P450-BM3 from *B. megaterium* led to trace amounts of the corresponding cyclopropane product.[44] Further work established that various hemeproteins, and even free hemin,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

catalyze olefin cyclopropanation in water.[45] Of particular interest, however, several wild-type P450s exhibited diastereoselectivity different from free hemin and showed some enantioinduction 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 engineering and evolution of a new family of enzymes that can activate diazo compounds for carbenoid transfers to organic molecules.[44,46–50]

Mutations dramatically increase the non-natural cyclopropanation activity of P450-BM3. Substitution of the distal threonine (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 that catalyzed the cyclopropanation of styrene in excellent yield, diastereoselectivity, and enantioselectivity (Figure 10A). Mutation at the cysteine axial ligand led to the greatest improvement in cyclopropanation activity. Mutating the cysteine at position 400 to serine (Cys400Ser) in P450-BM3 shifted the characteristic 450 nm peak in CO-difference spectrum to 411 nm, hence the “P411” name for the Ser-ligated catalysts. The Ser mutation also allowed the iron heme to be reduced under cellular conditions to the Fe^{2+} active catalyst, thereby enabling cyclopropanation with whole cells expressing these proteins. Styrene cyclopropanation was catalyzed on gram scale to 67,000 turnovers in 72% yield by a P411, 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 mutation of the proximal cysteine residue to histidine, catalyzed the cyclopropanation of *N*,*N*-diethyl-2-phenylacrylamide (**13**) to 86% yield and 92% enantioselectivity on preparative scale (Figure 10B). The laboratory-evolved catalyst BM3-Hstar performed cyclopropanation in the presence of oxygen and exhibited an initial rate of reaction close to what has been 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] P411 variant H2-5-F10 performed the 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, producing a mixture of single and double insertion products. In contrast, the enzyme provided the single insertion product selectively, highlighting the important role that the protein binding pocket plays in controlling selectivity.

Following up on the early observation of Dawson and coworkers,[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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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 for amination (Figure 10D). In particular, variant P411$_{BM3}$-CIS catalyzed the amination reaction in up to 87% enantioselectivity and 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 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 substrate scope of intramolecular C-H amination with FL#62 showed that the biocatalyst tolerates a range of substituents on the aryl ring. A follow-up report[51b] further showed that C-H amination on azide **16** could be catalyzed in good turnovers using either myoglobin (Mb) or horseradish peroxidase (HRP). Whereas wild-type Mb showed no detectable enantioselectivity in amination with **16**, introduction of mutations His64Val and Val68Ala led to appreciable enantioinduction (60% ee with **16**). These results suggest that other hemoproteins are also viable platforms for discovering 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] Variant P411$_{BM3}$-CIS-T438S-I263F catalyzed C-H amination of substrates **17a–c** at the homo-benzylic position with excellent regio- and enantioselectivity. The P411$_{BM3}$-T268A-F87A variant, in contrast, showed strong preference for C-H amination at the benzylic position, also with great regio- and enantioselectivity. Thus tailoring 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 is weaker by more than 10 kcal/mol relative to non-benzylic C-H).

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 chiral Brønsted acid-catalyzed polycyclizations.[52] The crystal structure of SHC from *Alicyclobacillus acidocaldarius* (*Aac*SHC) was disclosed in 1997,[53] but its promiscuity was known as early as 1986, when Neuman *et al*. 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 12A). By varying the terminator groups for the cyclization reactions, they were able to perform SHC-catalyzed cyclization of various non-natural substrates in low to moderate yield.[54,55] The

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

substrate scope for SHC-catalyzed cyclization reaction has also been studied by Hoshino and co-workers.[56]

SHC also possesses some weak activity for 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 *Z. mobilis* (ZMO1548) to improve production of isomers of isopulegol starting from racemic citronellal (Figure 12B).[58] Two amino acid positions were identified as important for increased isopulegol formation. The Trp555Tyr mutant gave more than 70% total conversion to isomeric mixtures of isopulegol, versus *ca*. 30% observed with the wild-type enzyme. The Phe486Cys variant provided more than 50% total conversion, with slightly improved product 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 cyclizations, including the Prins cyclization of (*S*)-citronellal (Figure 13). 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** in 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 in 11% conversion, an approximately 20-fold improvement over the wild-type enzyme. It will be interesting 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. In order to fully realize the potential of biocatalysis for sustainable chemistry, it will be important to expand the range of enzyme-catalyzed transformations to include ones 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 of possibilities, as promiscuous activity among enzymes is widespread and can be improved by protein engineering and especially directed evolution. Chemical intuition will help us know where to look for 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. Where wild-type enzymes catalyze non-natural reactions, it is entirely possible that protein engineering and directed evolution could boost activity, fine-tune selectivity, and make them more synthetically useful.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

4.1. C-H functionalization chemistry of SyrB2 halogenase

Reminiscent of the hydroxylase/desaturase example, nature utilized a similar bifurcation of reaction pathways for halogenation and hydroxylation with Fe(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 via an 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 hydroxyl ligand results in the hydroxylated product (path a) whereas coupling with a halogen ligand results in 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 non-coordinating 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 coworkers proposed that Fe/αKG halogenase evolved from Fe/αKG hydroxylase.[63]

Using threonine as substrate, SyrB2 is remarkably selective for the production of 4-chloro-L-threonine, suggesting that hydroxyl radical rebound from intermediate **A** (Figure 14A) is not competitive with halogen radical rebound. Substitution of the non-coordinating Ala 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 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 substrate and protein active site determines the selectivity between hydroxylation and halogenation. To further reinforce this notion, simple substitution of coordinating Asp to 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 to halogenase only works in very special cases.

Many Fe/αKG hydroxylases have been discovered that act on different types of substrates, and some of them do not require the substrates to be appended to a carrier protein. Fe/αKG halogenases, however, are relatively rare in nature. Conversion of hydroxylases to halogenases would allow 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 radical azidation and nitration of substrates (L-2-aminobutyrate, L-threonine, and L-norvaline) bound to SyrB1 with modest yields under single turnover conditions.[67] Such reactivity is reminiscent of radical-based C-N coupling in synthetic chemistry, where

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

literature precedents suggest 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 gives modest yields even under single turnover conditions. In addition, given the apparent complexity of substrate positioning in the active site in determining the outcome of Fe/αKG-catalyzed reactions, optimization of these new nitration and azidation reactions may be challenging.

4.2. Unnatural amino acid synthesis 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 via 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 resolution of racemic mixtures.[70] These processes often require manipulation of protecting groups (that have to be removed) and catalysts that have to be designed *de novo* for new targets. Development of 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, the acetate group of *O*-acetylserine is extruded to give an amino acrylate intermediate which then reacts with H_2S to form L-cysteine (Figure 15). In the late sixties, Mudd and Thompson independently established[72] that this class of enzymes also catalyzed the synthesis of *S*-substituted cysteine derivatives, hinting at 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 non-proteinogenic 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 in a series of publications showed that the OASS from higher plants could indeed catalyze the syntheses of β-(pyrazol-1-yl)-L-alanine, L-quisqualic acid, L-mimosine and several other non-proteinogenic amino acids in low yields.[74] As Maier showed in a follow-up work,[75] the overall metabolic pathway could be engineered to improve the low titers. However, no report has disclosed improving the production yield of UAAs by engineering the OASS enzyme. X-ray structures of CysK1 and CysM, both cysteine synthases, indicated the presence of a substrate tunnel that likely acts as a passageway for incoming nucleophiles.[76] Engineering this substrate tunnel to accommodate nucleophiles of different size could provide an alternative avenue to improve the production of UAAs.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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, condensation of serine with PLP is followed by a 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 analogs 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 efforts 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 carbenoid and nitrenoid transfer reactions catalyzed by engineered cytochrome P450s are a good example of how new enzymes can be generated using an approach that mimics nature and is based on chemical knowledge: work 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 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? This is clearly not the case. If it were, 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 like the cytochrome P450s or members of other functionally diverse superfamilies which 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 includes evolution of catalytic activities that have no natural counterpart, such as olefin cyclopropanation with diazo compounds or cyanation via 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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

feature(s) that enables the transformation, *e.g*. iron-carbenoid for cyclopropanation, acid-catalyzed formation of 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 based on known reactivity or mechanistic analogy, *e.g*. 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 non-natural activity or tune selectivity. Computational approaches may be able to assist this discovery process, possibly in evaluating suitable enzyme starting points *in silico*.[82]

Of course, evolution and engineering of existing enzymes is not the only possible approach to creating new enzymes. There has been good progress with artificial metalloenzymes and *de novo* enzyme design, but significant challenges remain, especially for creating synthetically useful catalysts and ones that function inside of cells. This, we believe, is a key advantage of the evolutionary approach we review here: the starting point for creating 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 already 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. But where that is the case, 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 the readers to 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' for non-natural catalysis. Protein engineering and enzyme evolution, too, are 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.

Acknowledgments

Our research is supported by the Department of 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 would like to 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.

References

1. a) Moore JC, Arnold FH. Nat Biotechnol. 1996; 14:458–467. [PubMed: 9630920] b) Shao Z, Arnold FH. Curr Opin Struct Biol. 1996; 6:513–518. [PubMed: 8794158] c) Arnold FH. Chem Eng Sci. 1996; 51:5091–5102.d) Arnold FH, Moore JC. Adv Biochem Eng Biotechnol. 1997; 58:1–14. [PubMed: 9103909] e) Reetz MT, Zonta A, Schimossek K, Liebeton K, Jaeger K-E. Angew Chem Int Ed. 1997; 36:2830–2832.f) Reetz MT. J Am Chem Soc. 2013; 135:12480–12496. [PubMed: 23930719]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2. a) Savile CK, Janey JM, Mundorff EC, Moore JC, Tam S, Jarvis WR, Colbeck JC, Krebber A, Fleitz FJ, Brands J, Devine PN, Huisman GW, Hughes GJ. Science. 2010; 329:305–308. [PubMed: 20558668] b) Liang J, Lalonde J, Borup B, Mitchell V, Mundorff E, Trinh N, Kochrekar DA, Cherat RN, Ganesh Pai G. Org Proc Res Dev. 2010; 14:193–198.c) Bornscheuer UT, Huisman GW, Kazlauskas RJ, Lutz S, Moore JC, Robins K. Nature. 2012; 485:185–194. [PubMed: 22575958] d) Alexeeva M, Enright A, Dawson MJ, Mahmoudian M, Turner NJ. Angew Chem Int Ed. 2002; 41:3177–3180.e) Turner NJ. Nat Chem Biol. 2009; 5:567–573. [PubMed: 19620998]
3. Wohlgemuth R. Curr Opin Biotechnol. 2010; 21:713–724. [PubMed: 21030244]
4. a) Steen EJ, Kang Y, Bokinsky G, Hu Z, Schirmer A, McClure A, del Cardayre SB, Keasling JD. Nature. 2010; 463:559–563. [PubMed: 20111002] b) Peralta-Yahya PP, Zhang F, del Cardayre SB, Keasling JD. Nature. 2012; 488:320–328. [PubMed: 22895337] c) Atsumi S, Hanai T, Liao JC. Nature. 2008; 451:86–89. [PubMed: 18172501] d) Lee SY, Park JH, Jang SH, Nielsen LK, Kim J, Jung KS. Biotechnol Bioeng. 2008; 101:209–228. [PubMed: 18727018]
5. Jansen RA. Annu Rev Microbiol. 1976; 30:409–425. [PubMed: 791073]
6. a) O'Brien PJ, Herschlag D. Chem Biol. 1999; 6:R91–R105. [PubMed: 10099128] b) Copley SD. Curr Opin Chem Biol. 2003; 7:265–272. [PubMed: 12714060]
7. Khersonsky O, Tawfik DS. Annu Rev Biochem. 2010; 79:471–505. [PubMed: 20235827]
8. Tawfik DS. Nat Chem Biol. 2010; 6:692–696. [PubMed: 20852602]
9. Seffernick JL, Wackett LP. Biochemistry. 2001; 40:12747–12753. [PubMed: 11669610]
10. Cook AM, Grossenbacher H, Hutter R. Experentia. 1983; 39:1191–1198.
11. Seffernick JL, De Souza ML, Sadowsky MJ, Wackett LP. J Bacteriol. 2001; 183:2405–2410. [PubMed: 11274097]
12. Udiković-Kolić N, Scott C, Martin-Laurent F. Appl Microbiol Biotechnol. 2012; 96:1175–1189. [PubMed: 23076592]
13. a) Scott C, Jackson CJ, Coppin CW, Mourant RG, Hilton ME, Sutherland TD, Russell RJ, Oakeshott JG. Appl Environ Microbiol. 2009; 75:2184–2191. [PubMed: 19201959] b) Noor S, Taylor MC, Russell RJ, Jermiin LS, Jackson CJ, Oakeshott JG, Scott C. PLoS ONE. 2012; 7:e39822. [PubMed: 22768133]
14. Todd AE, Orengo CA, Thornton JM. J Mol Biol. 2001; 307:1113–1143. [PubMed: 11286560]
15. Serdar CM, Gibson DT, Munnecke DM, Lancaster JH. Appl Environ Microbiol. 1982; 44:246–249. [PubMed: 16346063]
16. Seibert CM, Raushel FM. Biochemistry. 2005; 44:6383–6391. [PubMed: 15850372]
17. Afriat L, Roodveldt C, Manco G, Tawfik DS. Biochemistry. 2006; 45:13677–13686. [PubMed: 17105187]
18. Miller MB, Bassler BL. Annu Rev Microbiol. 2001; 55:165–199. [PubMed: 11544353]
19. Meier MM, Rajendran C, Malisi C, Fox NG, Xu C, Schlee S, Barondeau DP, Höcker B, Sterner R, Raushel FM. J Am Chem Soc. 2013; 135:11670–11677. [PubMed: 23837603]
20. Avriat-Jurnou L, Jackson CJ, Tawfik DS. Biochemistry. 2012; 51:6047–6055. [PubMed: 22809311]
21. Baier F, Tokuriki N. J Mol Biol. 2014; 426:2442–2456. [PubMed: 24769192]
22. a) Babbitt PC, Gerlt JA. Adv Prot Chem. 2001; 55:1–28.b) Gerlt JA, Babbitt PC. Curr Opin Chem Biol. 2009; 13:10–18. [PubMed: 19237310]
23. Elias M, Tawfik DS. J Biol Chem. 2012; 287:11–20. [PubMed: 22069329]
24. Park HS, Nam SH, Lee JK, Yoon CN, Mannervik B, Benkovic SJ, Kim HS. Science. 2006; 311:535–538. [PubMed: 16439663]
25. Shanklin J, Guy JE, Mishra G, Lindqvist Y. J Biol Chem. 2009; 284:18559–18563. [PubMed: 19363032]
26. Broun P, Shanklin J, Whittle E, Somerville C. Science. 1998; 282:1315–1317. [PubMed: 9812895]
27. Broadwater JA, Whittle E, Shanklin J. J Biol Chem. 2002; 277:15613–15620. [PubMed: 11864983]
28. Ben-David M, Elias M, Filippi JJ, Duñach E, Silman I, Sussman JL, Tawfik DS. J Mol Biol. 2012; 418:181–196. [PubMed: 22387469]

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

29. De Souza ML, Sadowsky MJ, Wackett LP. J Bacteriol. 1996; 178:4894–4900. [PubMed: 8759853]
30. Van Hylckama Vlieg JET, Tang L, Lutje Spelberg JH, Smilda T, Poelarends GJ, Bosma T, van Merode AEJ, Fraaije MW, Janssen DB. J Bacteriol. 2001; 183:5058–5066. [PubMed: 11489858]
31. a) de Jong RM, Tiesinga JJW, Rozeboom HJ, Kalk KH, Tang L, Janssen DB, Dijkstra BW. EMBO J. 2003; 22:4933–4944. [PubMed: 14517233] b) de Jong RM, Tiesinga JJW, Villa A, Tang L, Janssen DB, Dijkstra BW. J Am Chem Soc. 2005; 127:13 338–13 343.
32. a) Spelberg JHL, van Hylckama Vlieg JET, Tang L, Janssen DB, Kellogg RM. Org Lett. 2001; 3:41–43. [PubMed: 11429866] b) Hasnaoui-Dijoux G, Elenkov MM, Spelberg JHL, Hauer B, Janssen DB. Chem Bio Chem. 2008; 9:1048–1051.
33. Fox RJ, Davis SC, Mundorff EC, Newman LM, Gavrilovic V, Ma SK, Chung LM, Ching C, Tam S, Muley S, Grate J, Gruber J, Whitman JC, Sheldon RA, Huisman GW. Nat Biotechnol. 2007; 25:338–344. [PubMed: 17322872]
34. a) Fox RJ. J Theor Biol. 2005; 234:187–199. [PubMed: 15757678] b) Fox RJ. Protein Eng. 2003; 16:589–597. [PubMed: 12968076]
35. Jahn M, Marles J, Warren RAJ, Withers SG. Angew Chem Int Ed. 2003; 42:352–354.
36. Witczak ZJ. Curr Med Chem. 1999; 6:165–178. [PubMed: 10189230]
37. a) Vic G, Crout DHG. Tetrahedron:Asymmetry. 1994; 5:2513–2516.b) Kempton JB, Withers SG. Biochemistry. 1992; 31:9961–9969. [PubMed: 1390780]
38. Koshland DE. Biol Rev. 1953; 28:416–436.
39. Mülleggger J, Jahn M, Chen H-M, Warren RAJ, Withers SG. Prot Eng Des Sel. 2005; 18:33–40.
40. a) Maxwell JL, Brown KC, Bartley DW, Kodadek T. Science. 1992; 256:1544–1547. [PubMed: 17836323] b) Wolf JR, Hamaker CG, Djukic JP, Kodadek T, Woo LK. J Am Chem Soc. 1995; 117:9194–9199.
41. a) Li Y, Huang J-S, Zhou ZY, Che CM, You XZ. J Am Chem Soc. 2002; 124:13185–13193. [PubMed: 12405847] b) Guo Z, Guan X, Huang JS, Tsui WM, Lin Z, Che CM. Chem Eur J. 2013; 19:11320–11331. [PubMed: 23818419]
42. Breslow R, Gellman SH. J Am Chem Soc. 1983; 105:6728–6729.
43. Swastits EW, Dawson JH, Breslow R, Gellman SH. J Am Chem Soc. 1985; 107:6427–6428.
44. Coelho PS, Brustad EM, Kannan A, Arnold FH. Science. 2013; 339:307–310. [PubMed: 23258409]
45. Heel T, Mcintosh JA, Dodani SC, Meyerowitz JT, Arnold FH. ChemBioChem. 2014; 15:2259–2267. [PubMed: 25182183]
46. Coelho PS, Wang ZJ, Ener ME, Baril SA, Kannan A, Arnold FH, Brustad EM. Nat Chem Biol. 2013; 9:485–487. [PubMed: 23792734]
47. Wang ZJ, Renata H, Peck NE, Farwell CC, Coelho PS, Arnold FH. Angew Chem Int Ed. 2014; 126:6928–6931.
48. Renata H, Wang ZJ, Kitto RZ, Arnold FH. Cat Sci Tech. 2014; 4:3640–3643.
49. Wang ZJ, Peck NE, Renata H, Arnold FH. Chem Sci. 2013; 5:598–601. [PubMed: 24490022]
50. a) McIntosh JA, Coelho PS, Farwell CC, Wang ZJ, Lewis JC, Brown TR, Arnold FH. Angew Chem Int Ed. 2013; 52:9309–9312.b) Farwell CC, McIntosh JA, Hyster TK, Wang ZJ, Arnold FH. J Am Chem Soc. 2014; 136:8766–8771. [PubMed: 24901646] c) Hyster TK, Farwell CC, Buller AR, McIntosh JA, Arnold FH. J Am Chem Soc. accepted. 10.1021/ja509308v
51. a) Singh R, Bordeaux M, Fasan R. ACS Catal. 2014; 4:546–552. [PubMed: 24634794] b) Bordeaux M, Singh R, Fasan R. Bioorg Med Chem. 2014; 22:5697–5704. [PubMed: 24890656]
52. Hammer SC, Syrén PO, Seitz M, Nestl BM, Hauer B. Curr Opin Chem Biol. 2013; 17:293–300. [PubMed: 23485581]
53. a) Wendt KU, Lenhart A, Schulz GE, Feil C, Poralla K. Protein Sci. 1997; 6:722–724. [PubMed: 9070455] b) Wendt KU. Science. 1997; 277:1811–1815. [PubMed: 9295270]
54. Neumann S, Simon H. Biol Chem Hoppe-Seyler. 1986; 367:723–730. [PubMed: 3768141]
55. a) Seitz M, Syren PO, Steiner L, Klebensberger J, Nestl BM, Hauer B. ChemBioChem. 2013; 14:436–439. [PubMed: 23418022] b) Seitz M, Klebensberger J, Siebenhaller S, Breuer M, Siedenburg G, Jendrossek D, Hauer B. J Mol Catal B Enzym. 2012; 84:72–77.

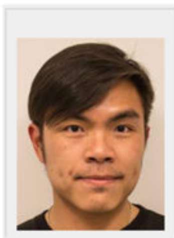
Author Manuscript Author Manuscript Author Manuscript Author Manuscript

56. Yonemura Y, Ohyama T, Hoshino T. Org Biomol Chem. 2012; 10:440–446. [PubMed: 22068606]
57. Siedenburg G, Jendrossek D, Breuer M, Juhl B, Pleiss J, Seitz M, Klebensberger J, Hauer B. Appl Environ Microbiol. 2012; 78:1055–1062. [PubMed: 22156419]
58. Siedenburg G, Breuer M, Jendrossek D. Appl Microbiol Biotechnol. 2013; 97:1571–1580. [PubMed: 22526778]
59. Hammer SC, Marjanovic A, Dominicus JM, Nestl BM, Hauer B. Nat Chem Bio. 2014 accepted. 10.1038/nchembio.1719
60. a) Johnson WS, Kinnel RB. J Am Chem Soc. 1966; 88:3861–3862. [PubMed: 5916378] b) Bender JA, Blize AE, Browder CC, Giese S, West FG. J Org Chem. 1998; 63:2430–2431. [PubMed: 11672099] c) Knowles RR, Lin S, Jacobsen EN. J Am Chem Soc. 2010; 132:5030–5032. [PubMed: 20369901] d) Qiu WW, Surendra K, Yin L, Corey EJ. Org Lett. 2011; 13:5893–5895. [PubMed: 21992657] e) Schafroth MA, Sarlah D, Krautwald S, Carreira EM. J Am Chem Soc. 2012; 134:20276–20278. [PubMed: 23193947]
61. Vaillancourt FH, Yin J, Walsh CT. Proc Natl Acad Sci USA. 2005; 102:10111–10116. [PubMed: 16002467]
62. Wong SD, Srnec M, Matthews ML, Liu LV, Kwak Y, Park K, Bell CB III, Alp EE, Zhao J, Yoda Y, Kitao S, Seto M, Krebs C, Bollinger JM, Solomon EI. Nature. 2013; 499:320–323. [PubMed: 23868262]
63. Vaillancourt FH, Yeh E, Vosburg DA, Garneau-Tsodikova S, Walsh CT. Chem Rev. 2006; 106:3364–3378. [PubMed: 16895332]
64. Matthews ML, Neumann CS, Miles LA, Grove TL, Booker SJ, Krebs C, Walsh CT, Bollinger JM Jr. Proc Natl Acad Sci USA. 2009; 106:17723–17728. [PubMed: 19815524]
65. Gorres KL, Pua KH, Raines RT. PLoS ONE. 2009; 4:e7635. [PubMed: 19890397]
66. Costas M, Mehn MP, Jensen MP, Que L Jr. Chem Rev. 2004; 104:939–986. [PubMed: 14871146]
67. Matthews ML, Chang W-C, Layne AP, Miles LA, Krebs C, Bollinger JM. 2014; 10:209–217.
68. a) Leggans EK, Barker TJ, Duncan KK, Boger DL. Org Lett. 2012; 14:1428–1431. [PubMed: 22369097] b) Waser J, Nambu H, Carreira EM. J Am Chem Soc. 2005; 127:8294–8295. [PubMed: 15941257]
69. Boswell C. Chem Market Reporter. Jun 14.1999 :9–11.
70. Drauz K. Chimia. 1997; 51:310–315.
71. Tai CH, Cook PF. Acc Chem Res. 2001; 34:49–59. [PubMed: 11170356]
72. a) Giovanelli J, Mudd SH. Biochem Biophys Res Comm. 1968; 31:275–280. [PubMed: 5656077] b) Thompson JF, Moore DP. Biochem Biophys Res Comm. 1968; 31:281–285. [PubMed: 5656078]
73. Ikegami F, Murakoshi I. Phytochemistry. 1994; 35:1089–1104.
74. a) Murakoshi I, Kaneko M, Koide C, Ikegami F. Phytochemistry. 1986; 25:2759–2763.b) Ikegami F, Mizuno M, Kihara M, Murakoshi I. Phytochemistry. 1990; 29:3461–3465.c) Murakoshi I, Ikegami F, Kaneko M. Phytochemistry. 1985; 24:1907–1911.
75. Maier THP. Nat Biotechnol. 2003; 21:422–427. [PubMed: 12640465]
76. a) Schnell R, Oehlmann W, Singh M, Schneider G. J Biol Chem. 2007; 282:23473–23481. [PubMed: 17567578] b) Claus MT, Zocher GE, Maier THP, Schulz GE. Biochemistry. 2005; 44:8620–8626. [PubMed: 15952768]
77. Dunn MF, Niks D, Ngo H, Barends TRM, Schlichting I. Trends Biochem Sci. 2008; 33:254–264. [PubMed: 18486479]
78. Phillips RS. Tetrahedron: Asymmetry. 2004; 15:2787–2792.
79. a) Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF III, Hilvert D, Houk KN, Stoddard BL, Baker D. Science. 2008; 319:1387–1391. [PubMed: 18323453] b) Althoff EA, Wang L, Jiang L, Giger L, Lassila JK, Wang Z, Smith M, Hari S, Kast P, Herschlag D, Hilvert D, Baker D. Prot Sci. 2012; 21:717–726.c) Kiss G, Çelebi-Ölçüm N, Moretti R, Houk KN, Baker D. Angew Chem Int Ed. 2013; 52:5700–5725.
80. Baker D. Prot Sci. 2010; 19:1817–1819.
81. McIntosh JA, Farwell CC, Arnold FH. Curr Opin Chem Biol. 2014; 19:126–134. [PubMed: 24658056]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

82. Chakraborty S, Rao BJ. PLoS ONE. 2012; 7:e32011. [PubMed: 22359655]
83. a) Dürrenberger M, Ward TR. Curr Opin Chem Biol. 2014; 19:99–106. [PubMed: 24608081] b) Bos J, Roelfes G. Curr Opin Chem Biol. 2014; 19:135–143. [PubMed: 24608083]
84. Kries H, Blomberg R, Hilvert D. Curr Opin Chem Biol. 2013; 17:221–228. [PubMed: 23498973]

Biographies



Hans Renata received his B. A. from Columbia University in 2008, conducting 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 conducting postdoctoral studies in the laboratories of Professor Frances H. Arnold at the California Institute of Technology.

Z. Jane Wang received her B. S. 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 work focused on developing new methods for forming C-O and C-N bonds with coinage metal complexes. Jane recently completed her postdoctoral studies at the California Institute of Technology with Professor Frances H. Arnold.

Frances H. Arnold is the Dickinson Professor of Chemical Engineering, Bioengineering and Biochemistry at the California Institute of Technology. Her research focuses on protein evolution and biocatalysis. She was elected to the U.S. 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.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

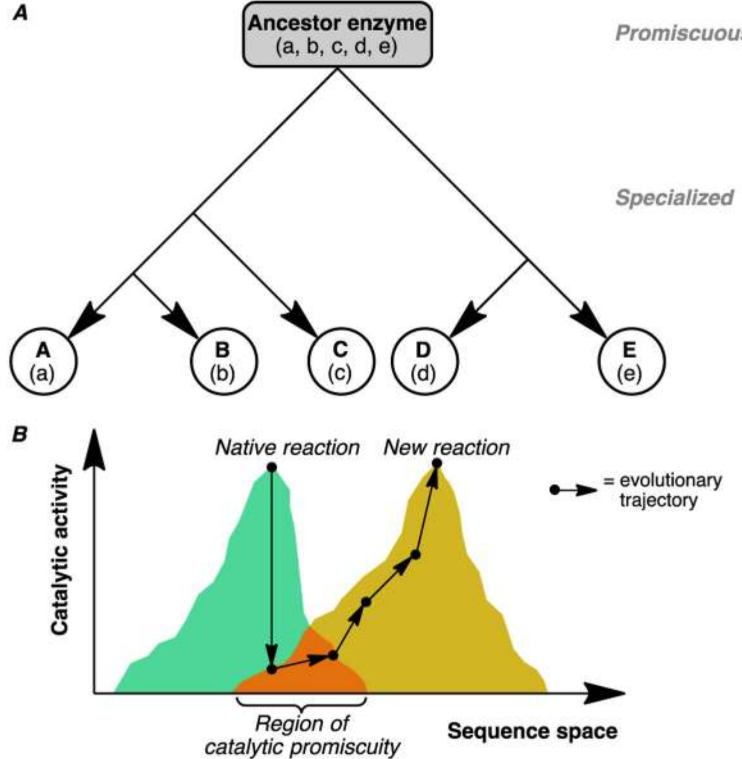


Figure 1.

(**A**) Divergence of an ancestral enzyme having 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 evolution of new function. A given protein sequence might catalyze multiple reactions. In 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 more discussion, see ref. 8.

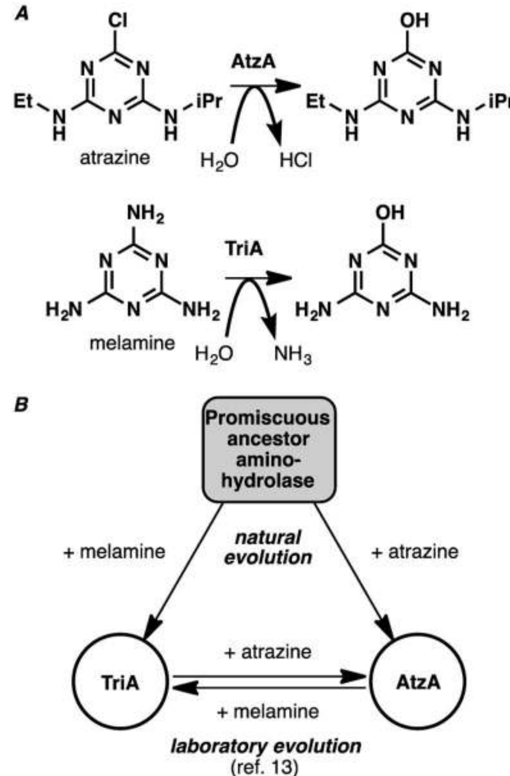


Figure 2.
(**A**) Chlorohydrolase activity of AtzA and aminohydrolase activity of TriA; (**B**) TriA and AtzA (98% AA identity) are believed to be related through a common ancestor similar to TriA. These catalytic functions can be interconverted with a few amino acid mutations.[13]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

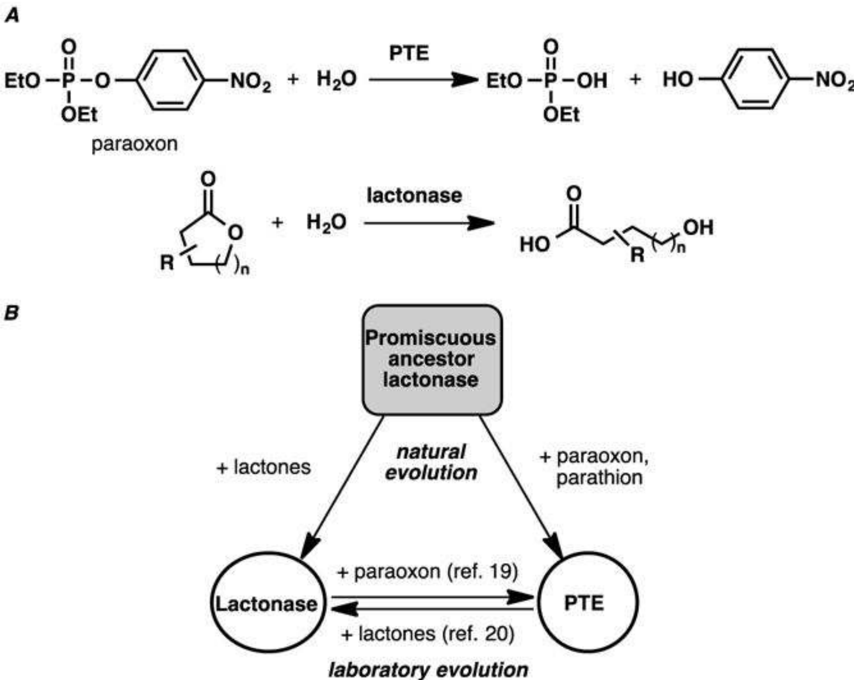


Figure 3.

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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

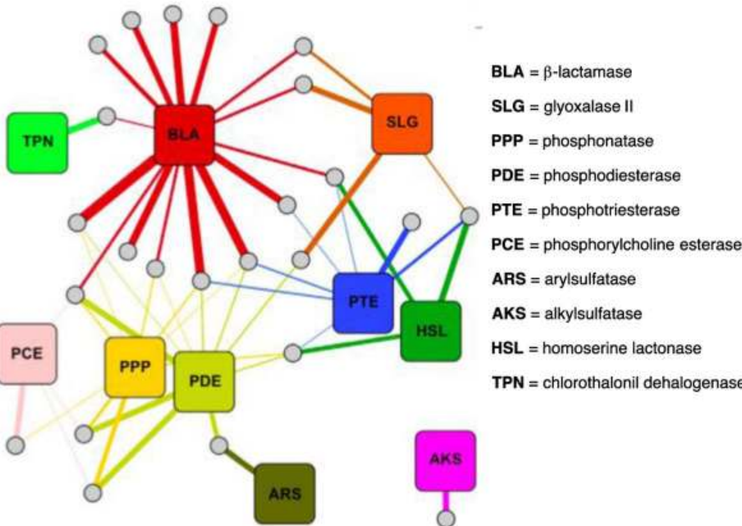
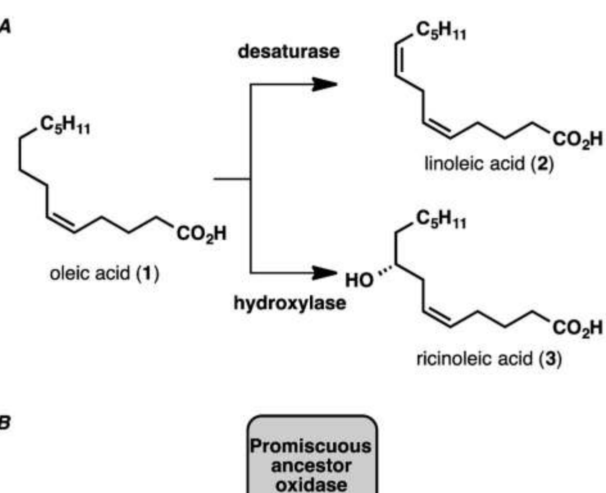


Figure 4.

Members of 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 via promiscuous enzymes (gray circles) that catalyze two or more reactions. Figure is reproduced from reference 21.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



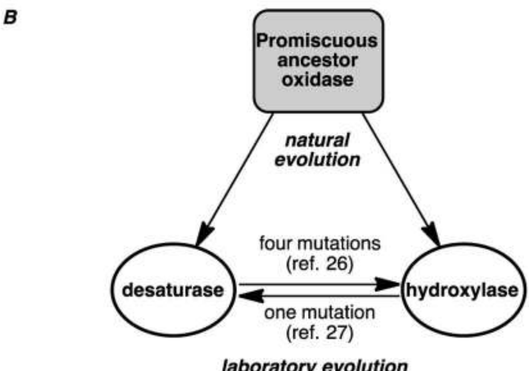


Figure 5.

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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

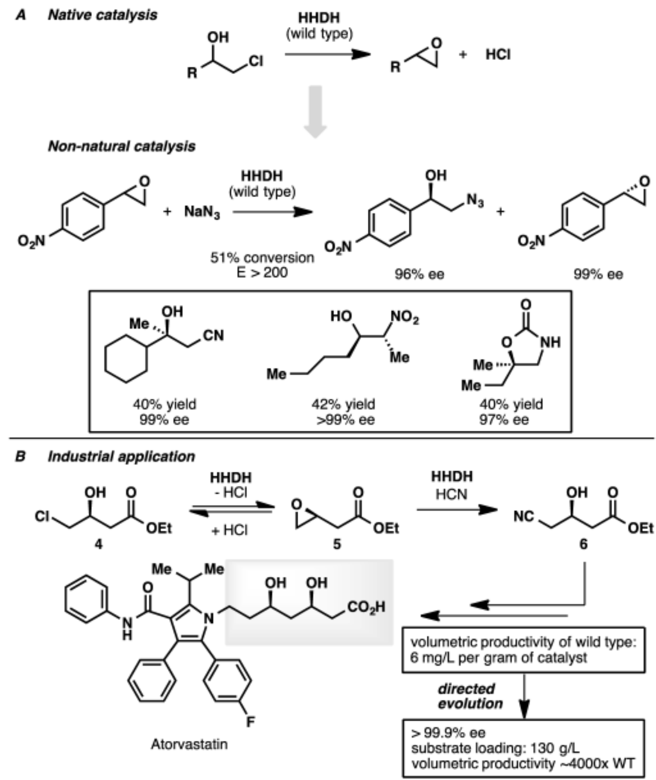


Figure 6.

(**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 using directed evolution.[33]

Figure 7.

(**A**) Mechanism of wild-type retaining β-glycosidase featuring catalytic acid/base and catalytic nucleophile residues where R′ = H or other sugar in low water concentration;[38] (**B**) Engineered thioglycoligase via removal of catalytic acid/base residue, DNP = dinitrophenyl, Nuc = deoxythio sugar nucleophile as acceptor.[35]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Figure 8.

Use of engineered thioglycoligase for the synthesis of thiodisaccharides.[35] Reported yields were after peracetylation of the thiodisaccharides. Wild-type enzymes do not catalyze this reaction.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

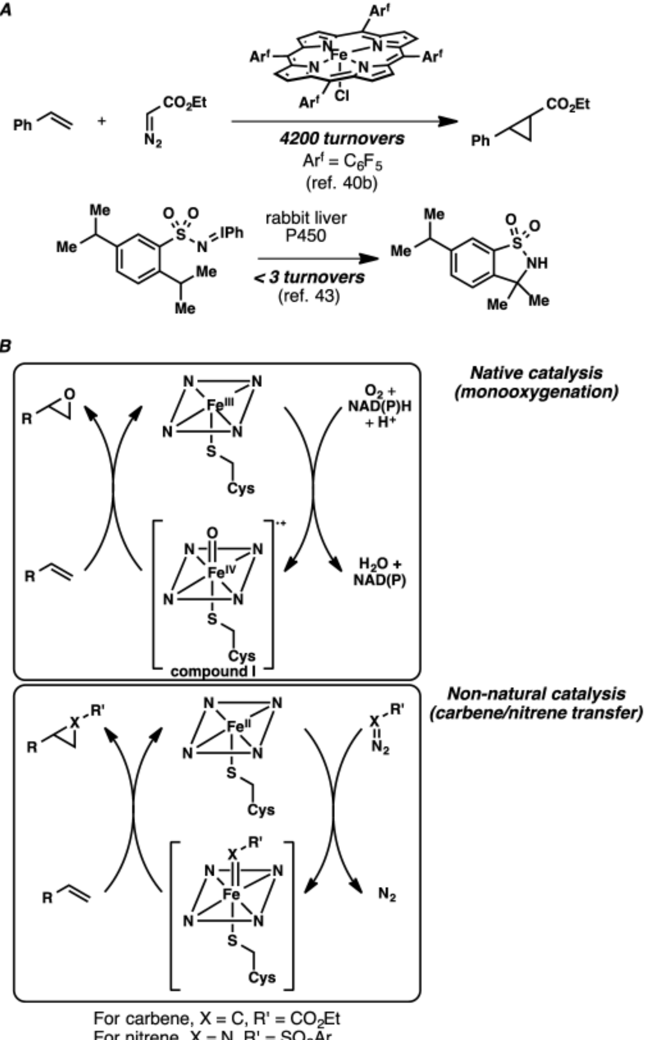
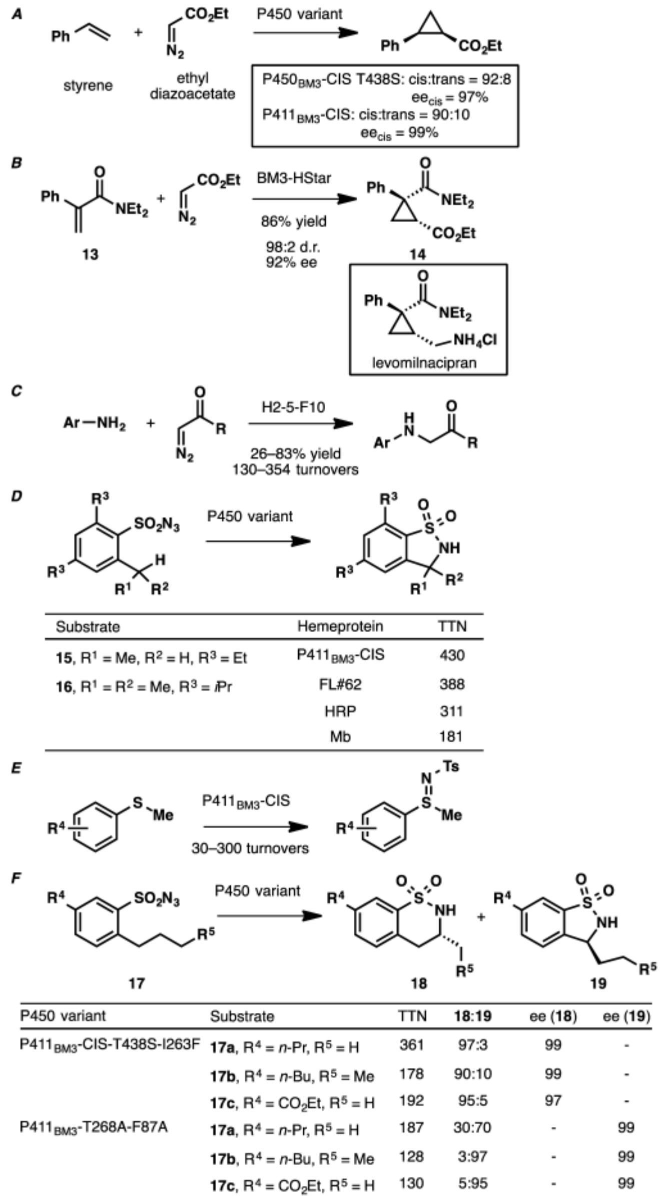


Figure 9.

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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



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]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 11.
Cyclization of squalene catalyzed by SHC and promiscuous activity of SHC on homofarnesol.[54]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Figure 12.

(**A**) SHC-catalyzed cyclization of terpene-like substrates, red color indicates bond(s) formed during the reaction;[54–56] (**B**) SHC-catalyzed Prins cyclization of citronellal for production of isopulegol, a precursor to menthol, and activity improvement via mutation.[57,58]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Transformation	Conversion with WT *Aac*SHC(%)	Mutant	Conversion (%)
geraniol → **20**	0.40 ± 0.03	F365C	15 ± 0.72
geraniol → **21**	n.d.	G600F	68 ± 3.1
6,7-epoxygeraniol → **22**	0.57 ± 0.13	Y420W	54 ± 0.42
		Y420W-G600F	78 ± 2.1
(*S*)-citronellal → *iso*-isopulegol	0.59 ± 0.03	I261A	11 ± 0.01

Figure 13.

Cyclization reactions of various substrates utilizing functional group initiators such as epoxide and aldehyde with *Aac*SHC and identification of enzyme variants with improved cyclization activities.[59]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

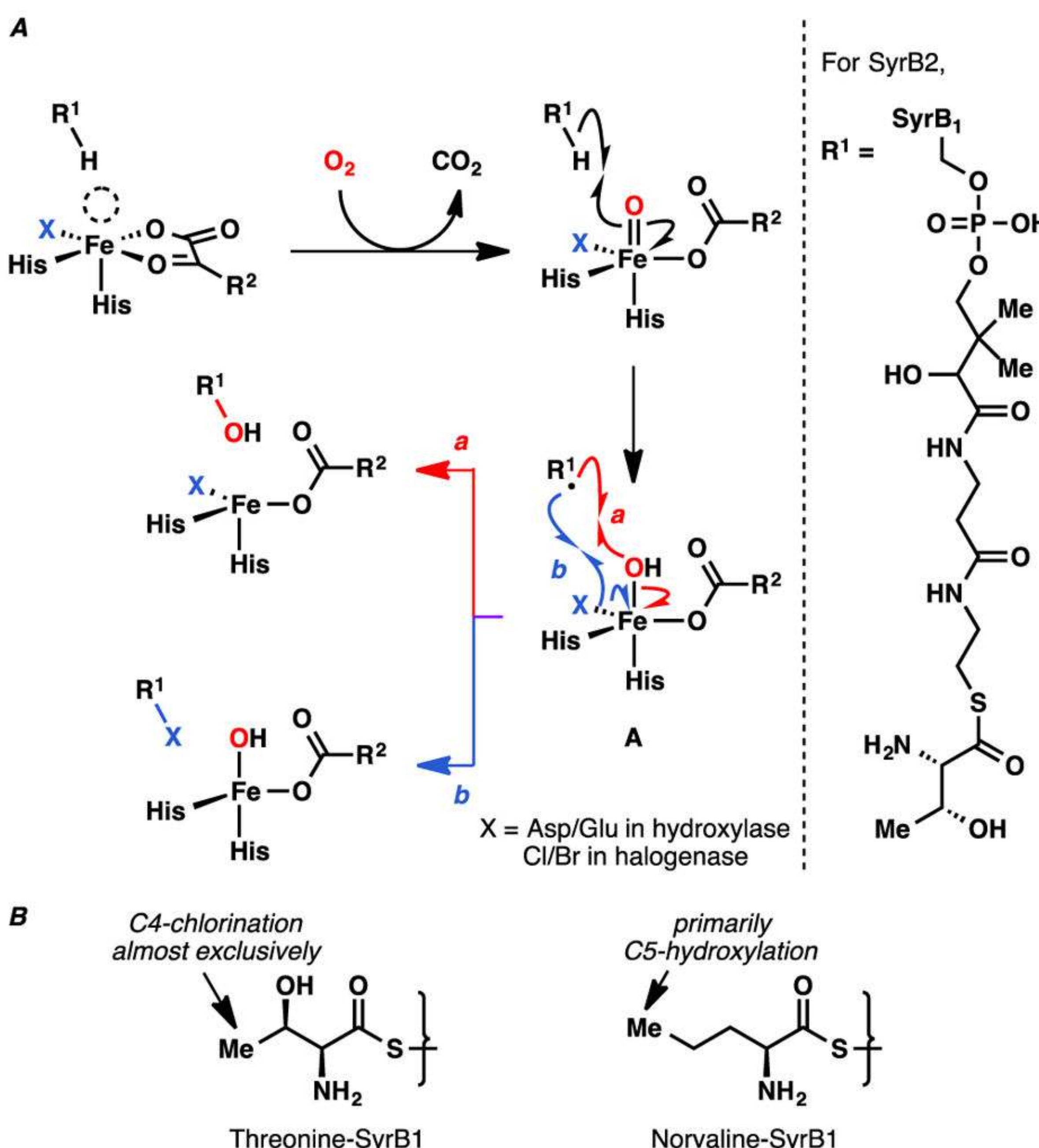


Figure 14.

(**A**) Mechanisms of Fe/αKG hydroxylase (path *a*, in red) and Fe/αKG halogenase (path *b*, in blue) where 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 SyrB2-catalyzed reaction.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Native catalysis by OASS

PLP

- HOAc

H_2S

- PLP

Non-natural catalysis

Nuc-H

- PLP

PLP =

Nuc-H:					
Fermentation yield with W3110/pACcysE^{fbr} (g/L)	9.1	13.2	2.2	7.3	6.5

Figure 15.

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

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

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