5. Directed evolution of artificial metalloenzymes: bridging synthetic chemistry and biology

Ruijie K. Zhang[1], David K. Romney[1], S. B. Jennifer Kan[1], Frances H. Arnold[1]

[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA.

Abstract: Directed evolution is a powerful algorithm for engineering proteins to have novel and useful properties. However, we do not yet fully understand the characteristics of an evolvable system. In this chapter, we present examples where directed evolution has been used to enhance the performance of metalloenzymes, focusing first on 'classical' cases such as improving enzyme stability or expanding the scope of natural reactivity. We then discuss how directed evolution has been extended to artificial systems, in which a metalloprotein catalyzes reactions using abiological reagents or in which the protein utilizes a non-natural cofactor for catalysis. These examples demonstrate that directed evolution can also be applied to artificial systems to improve catalytic properties, such as activity and enantioselectivity, and to favor a different product than that favored by small-molecule catalysts. Future work will help define the extent to which artificial metalloenzymes can be altered and optimized by directed evolution and the best approaches for doing so.

Keywords: directed evolution, fitness landscape, non-natural reactivity, artificial metalloenzyme

5.1. Evolution enables chemical innovation

Nature is expert at taking one enzyme framework and repurposing it to perform a multitude of chemical transformations. For example, the P450 superfamily consists of structurally similar heme-containing proteins that catalyze C–H oxygenation, alkene epoxidation, oxidative cyclization (1), aryl-aryl coupling (2), and nitration (3), among others (4). While an enzyme may be optimized to catalyze a specific process, it often exhibits activity for different processes. Evolution can take advantage of these 'promiscuous' reactivities in response to a changing environment (5-7) to create and optimize enzymes. Researchers are using similar concepts to make, diversify, and optimize new artificial metalloenzymes.

In directed evolution, which is the laboratory mimic of adaptive evolution, beneficial mutations identified by screening libraries are accumulated, one or a few at a time, to generate desired function(s) or performance. This technique has been used to confer useful properties such as stability under harsh conditions and improved activity on non-native substrates. More recently, scientists have found that synthetically useful reactions not found in nature can be catalyzed by metalloproteins, designed or natural, and that these non-natural activities can also be improved

by directed evolution. New activities can arise when natural metalloproteins are exposed to abiological reagents or when an artificial metallocofactor is introduced into a protein scaffold. Because the protein (ideally) plays a role in activating or otherwise controlling the cofactor and in determining substrate and reaction selectivity in these systems, both types of 'artificial' metalloenzymes can be optimized and diversified by directed evolution. This chapter will introduce directed evolution in the context of these new metalloenzymes. We describe the approaches that have been used to improve different enzyme systems in order to illustrate opportunities and requirements for using directed evolution.

5.1.1. Strategies for directed evolution

In a directed evolution experiment, the goal is to imbue a protein with a new property, such as high selectivity, activity with a non-native substrate, or stability in a different temperature or pH range. The process involves four key steps: 1) identification of a starting protein (that may have only low levels of the desired property), 2) diversification of the starting protein through mutagenesis of the corresponding DNA sequence, 3) functional screening of the variants to identify improved variants, and repeating steps 2 and 3 until sufficient levels of the desired property are achieved (step 4) (Figure 5.1). With advances in molecular biology and analytical tools, there are many methods available for executing each step, the choice of which will affect the evolutionary trajectory and end point.

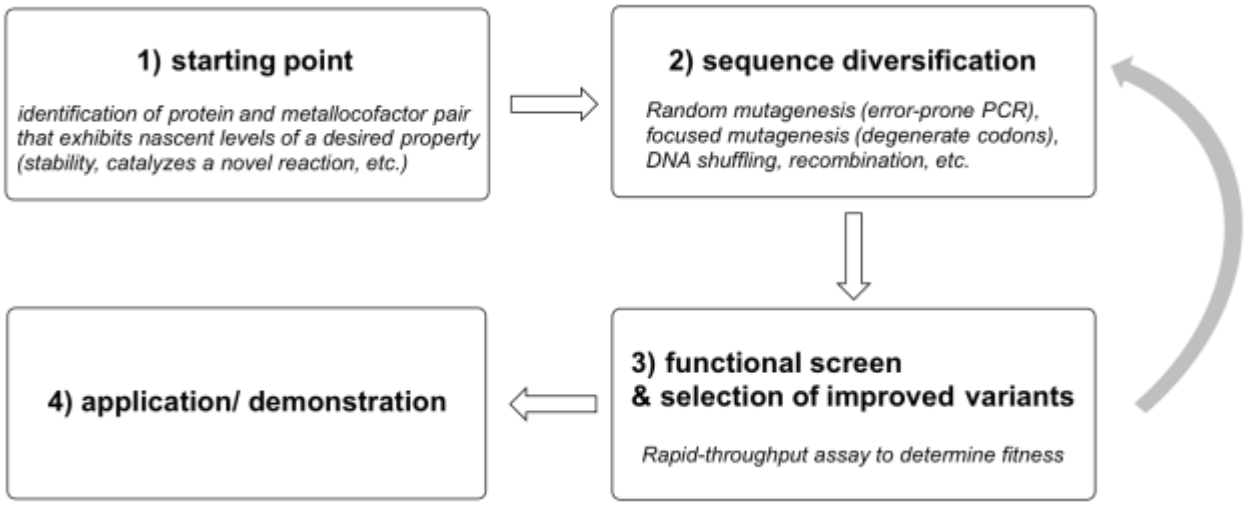


Figure 5.1. Overview of directed evolution.

Screening is usually the most experimentally challenging step in directed evolution of synthetically useful enzymes, and screening capacity will often determine how the sequence diversification is done. Because the goal of the experiment is to find beneficial mutations, the variant library should have a frequency of beneficial mutations that matches the screening

capability. Of course, the beneficial mutation frequency is not known at the beginning, but one can make educated guesses. In random mutagenesis, mutations are made throughout the protein. This approach is advantageous because it recognizes that the effects of mutations on protein structure, dynamics, and catalysis are not predictable *a priori*. For instance, mutations distant from enzyme active sites can affect catalysis. However, even for a small protein, the number of possible ways to make even a single mutation is so large that a rapid-throughput screen, such as a colorimetric assay, is usually required to sample enough sequences to find a beneficial mutation. (Whole protein random mutagenesis at 1-3 mutations per gene usually requires screening hundreds to thousands of mutant proteins per generation.) Unfortunately, many desired properties are not easily assayed in high throughput.

An alternative is to introduce mutations in a focused manner, guided by structural, mechanistic, or computational information. The assumption is that beneficial mutations will appear at the chosen positions (e.g., residues in the active site) with a higher frequency than at positions chosen at random (8). With the use of degenerate codons, focused libraries of variants can be created in which one or more specific residues are mutated to all or selected subsets of the 20 proteinogenic amino acids. This approach can significantly reduce library size and focus on amino acids believed to be key. However, success rests on making the right choice of residues to target. The fact remains that the effect of a mutation is often unpredictable. Thus, sites hypothesized to be important may in fact exert little influence on the desired function or property, or may not even tolerate mutation.

Other methods for sequence diversification include recombination, wherein genes or gene fragments of related proteins are shuffled and reassembled, thus enabling beneficial combinations of amino acids to appear in a single protein (9). Furthermore, beneficial mutations that are found separately can be recombined in a combinatorial fashion, or hybrids of evolutionarily related proteins can be made. Frequently, a combination of diversification methods is used, especially if multiple rounds of directed evolution are necessary (10).

5.1.2. Directed evolution as an uphill walk in the protein fitness landscape

The concept of directed evolution as a walk through sequence space that passes through functional proteins was born out of a paper by John Maynard Smith in 1970 (11). Smith asserted that in order for evolution by natural selection to be successful, there must be a continuous trail of functional proteins that can be traversed in single mutational steps without passing through nonfunctional proteins. He illustrated the idea with a word game, using conversion of the word ‘WORD’ to ‘GENE’ one letter at a time, with the requirement that all intermediate words be meaningful. A nonsense word, analogous to a nonfunctional protein, would be eliminated by selection before a second change could occur. His solution to the game, WORD → WORE → GORE → GONE → GENE, interconverts the two words through a series of single letter changes, each of which produces a meaningful word.

Directed evolution takes the game one step further and searches for improvements. With Smith's analogy, we can assign each word (and the protein it represents) a fitness, defined as the number of letters it matches with 'GENE', and think of this transformation from 'WORD' to 'GENE' as an adaptive uphill walk to a fitness peak. Laboratory protein evolution can then be envisioned as a walk on a high-dimensional fitness landscape in sequence space, where fitness is defined by the experimenter. Iterative rounds of mutagenesis and screening lead the experimenter to higher fitness levels, provided that the landscape is smooth (Figure 5.2) (10a). We do not know the structure of a protein fitness landscape; indeed, we expect the structure of the landscape to depend on the specific protein and problem. Past directed evolution successes and the success of natural evolution, however, suggest that enzyme landscapes are smooth for some properties, including catalysis and stability (*vide infra*), in at least some sequence dimensions. Because researchers have only recently begun to explore mutation of artificial metalloenzymes, we do not yet know whether their fitness landscapes enable similar optimization by iterative mutation and screening. However, early experiments indicate this will be the case.

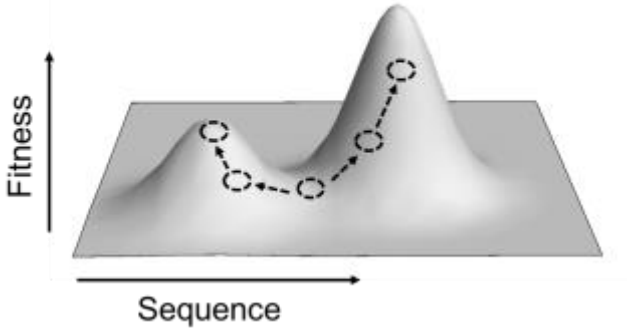


Figure 5.2. Directed evolution takes a protein along its fitness landscape, where fitness is a metric defined by the experimenter. Sequence diversification samples the nearby sequences, and screening identifies fitness improvements. Two possible evolutionary trajectories (green and black) from a single starting point (red) illustrate that there may exist multiple local maxima or solutions to any given optimization, and the results may depend on the path taken. Sequence space is of very high dimensionality; a simplified fitness landscape is presented here.

5.2. Directed evolution applied to natural metalloenzymes

Although natural metalloenzymes catalyze useful and chemically challenging reactions such as regioselective C–H oxygenation, utility of these biocatalysts for chemical synthesis or industrial applications may be limited. This may be due to restricted substrate scope or instability under process conditions. Directed evolution can address these issues by altering the activity profile of an enzyme or changing the range of optimal operating conditions. Directed evolution of natural metalloenzymes is well-known; we discuss only a few examples to illustrate how well this approach has worked to enhance biocatalyst performance.

5.2.1. Enhancing the stability of a carbonic anhydrase

Carbonic anhydrase is a Zn-dependent metalloenzyme that catalyzes the reversible hydration of carbon dioxide into bicarbonate and a proton (Figure 5.3a). It is a potentially useful catalyst for carbon capture and sequestration (CCS), a process to reduce anthropogenic CO_2 released from fossil fuel emissions (12). Current CCS processes effect CO_2 removal from emissions by absorption into an aqueous amine solvent, such as N-methyldiethanolamine (MDEA), followed by regeneration of the solvent by CO_2 stripping at elevated temperature (>87 °C). The CO_2-depleted gas can then be released into the atmosphere, while the pure CO_2 is compressed for geological sequestration (13). Introducing carbonic anhydrase to the CCS process to accelerate the capture of CO_2 is of great interest for environmental sustainability. However, naturally occurring carbonic anhydrases are limited by their inability to tolerate the harsh alkaline solvent and high process temperatures. Approaches that have been investigated to address this include carbonic anhydrase mimics (14), enzyme immobilization (15), and directed evolution to improve enzyme stability (16).

Lalonde and co-workers at Codexis used direction evolution to generate a highly stable carbonic anhydrase capable of withstanding CO_2 capture conditions (16). The starting enzyme, from *Desulfovibrio vulgaris* (*Dv*CA), was chosen because it was functional in 4.2 M MDEA, having a half-life of 15 minutes at 50 °C. Over nine rounds of evolution (Figure 5.3b), variants were challenged with increasingly harsh conditions (temperature from 42 °C to 107 °C, MDEA concentration from 3.0 M to 4.2 M), after which residual activity for the reverse reaction (carbonic acid to water and carbon dioxide) was assessed using a high-throughput assay based on phenolphthalein absorbance. The diversification strategy involved a combination of saturation mutagenesis, where chosen sites were mutated to all twenty proteinogenic amino acids, and recombination of beneficial mutations. Statistical analysis of screening results was performed using Protein Sequence Activity Relationship (ProSAR) analysis (17) to predict sets of beneficial mutations in the context of each parent protein during the evolution; this allowed for decreased screening effort. Although the use of statistical tools focused the evolutionary search, thousands of variants were still tested in each round, for a total of 27,000 variants screened (16).

Nine rounds of evolution converted the parent enzyme, which had negligible activity after exposure to 4.2 M MDEA at 60 °C, to a variant that retained activity after heating to 107 °C. Mutations at 36 positions (representing 15% of the enzyme) were accumulated during the evolution; most are located at a tetrameric interface based on alignment with the closest crystallized homolog. Exactly how the mutations enhance protein stability is not known, and this optimization solution could not be predicted *a priori*. Finally, the utility of an ultrastable variant was demonstrated in a pilot scale CO_2 capture system. Comparison of pilot scale CO_2 capture demonstrations performed in the presence or absence of the ultrastable enzyme showed that the rate of CO_2 capture was enhanced 25-fold using the evolved variant (16).

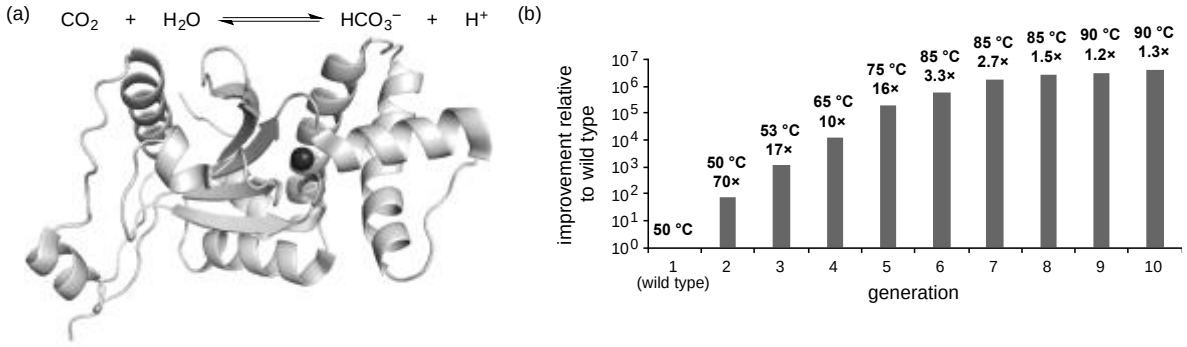


Figure 5.3. Directed evolution of a carbonic anhydrase stable to CO_2 capture process conditions (16). **(a)** Carbonic anhydrase catalyzes the reversible hydration of CO_2 to bicarbonate and a proton. The enzyme utilizes a catalytic zinc atom, displayed as a sphere. Only a single subunit of the tetrameric protein is shown (PDB 2A5V). **(b)** Evolution of an ultrastable carbonic anhydrase. Half-lives of the variants at the indicated temperature (black) were determined by measuring CO_2 absorption in a reactor. The fold improvement over the previous round is shown above each bar.

5.2.2. Expanding the scope of P450-catalyzed oxidation reactions

The cytochrome P450 superfamily has exploited a versatile protein framework to create enzymes with many different functions. For instance, P450 monooxygenases are responsible for the late-stage hydroxylation of diverse compounds, including terpenes, alkaloids, and steroids, in biosynthesis (18). A good indication of a protein family's evolvability in the laboratory is its functional diversity in nature (19), and indeed P450s are among the most studied metalloenzymes for directed evolution.

In one example, the cytochrome P450 from *Bacillus megaterium* ($P450_{BM3}$), whose native substrates are believed to be long-chain fatty acids, was optimized for regioselective oxidation of permethylated monosaccharides (20). Starting from a compiled library of eleven hundred $P450_{BM3}$ variants, some of which were engineered for different applications, Lewis *et al.* identified an enzyme with good activity (~800 turnovers) for the oxidative demethylation of galactose derivative **5.1**, but poor regioselectivity (1:1 **5.2** to **5.3**, Figure 5.4a). The large initial search space was enabled by their use of a colorimetric high-throughput assay that detected formaldehyde formation. Using a combination of random mutagenesis, site-directed mutagenesis, and recombination of beneficial mutations, they improved the enzyme for regioselectivity (Figure 5.4b). In four rounds, they obtained a variant that provided high regioselectivity for the conversion of **5.1** to **5.2** (96% conversion, single isomer) (20).

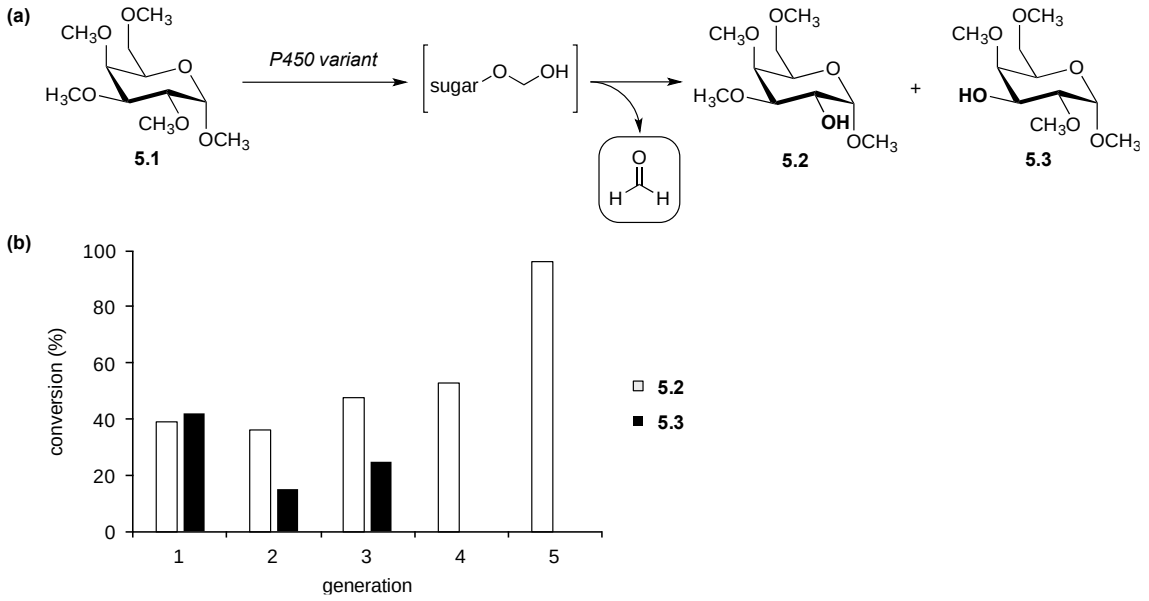


Figure 5.4. Engineering a cytochrome P450 for regioselective oxidation of permethylated monosaccharides (20). (**a**) Model reaction for P450-catalyzed demethylation. (**b**) Evolutionary trajectory of a regioselective catalyst.

Reetz and co-workers engineered two $P450_{BM3}$ variants for regiodivergent hydroxylation of a non-native steroid substrate, testosterone (21). Although testosterone is not accepted as a substrate by wild type $P450_{BM3}$, the authors observed that a single mutation, F87A, activated the enzyme towards this substrate. The F87A variant showed low activity (21% conversion) and gave a 52:45 mixture of 2β- and 15β-hydroxylated products (compounds **5.4** and **5.5**, respectively), with 6β- and 16β-hydroxytestosterone accounting for the remaining 3%. Initial studies focused on mutation of eight amino acid residues in the active site. These residues were grouped by location and, using degenerate codons, diversified simultaneously within the same group. Using this strategy, 8,700 transformants were initially screened by automated high-performance liquid chromatography (HPLC). Several variants demonstrated high 2β-selectivity with good conversion; notably, $P450_{BM3}$(F87A, A330W) gave 97% selectivity for the 2β-hydroxytestosterone product with 79% conversion (Figure 5.5). Variants with good 15β-selectivity were also identified, the best variant, $P450_{BM3}$(V78L, A82F, F87A), giving 91% 15β-selectivity with 86% conversion.

Further mutagenesis was performed to increase 15β-selectivity. A variant with intermediate 15β-selectivity was chosen as the new starting point, for which sites V78 and A82 were randomized simultaneously. After this second round of screening, improved variant $P450_{BM3}$(R47Y, T49F, V78L, A82M, F87A) was identified, giving 96% selectivity for 15β-hydroxytestosterone with 85% conversion (Figure 5.5).

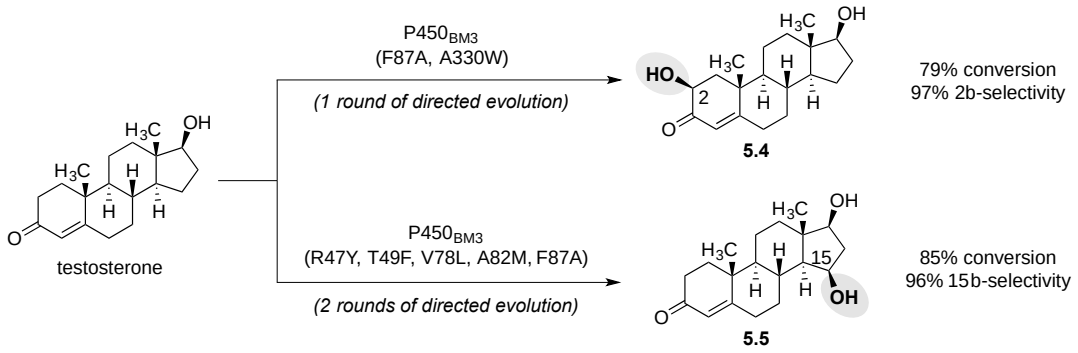


Figure 5.5. Regiodivergent hydroxylation of testosterone by engineered $P450_{BM3}$ variants (21).

The experimental solutions uncovered for these complex systems were not predictable *a priori*. Only recently has research begun to apply computational methods to results obtained by directed evolution in an effort to understand the molecular basis of the changes introduced by evolution. In the example of regiodivergent hydroxylation of testosterone discussed above, molecular dynamics simulations and substrate docking experiments were performed on the initial variant (F87A), a 2β-selective variant, and a 15β-selective variant to investigate the basis of the observed regioselectivity. The results of these studies suggest that in $P450_{BM3}$(F87A), testosterone can adopt two energetically equivalent orientations, one of which would lead to the 2β-hydroxylated product and the other to the 15β-hydroxylated product. The computational results for the 2β-selective and the 15β-selective variants show that each variant has a distinct binding pocket that allows only one active orientation of testosterone (21).

Additional applications of directed evolution to expand the scope of P450 oxidation reactions have led to the identification of numerous variants with impressive activities (22). These activities include hydroxylation of simple molecules, such as short-chain alkanes (23), as well as complex molecules, such as anti-malarial therapeutic artemisinin (24) and natural product parthenolide (25), with defined regio- and stereoselectivities. Evolution for chemoselectivity has also been investigated; in one report, an engineered variant displayed up to 90% chemoselectivity in favor of allylic hydroxylation versus epoxidation (26). As collections of diverse cytochrome P450 variants expand, more applications for these malleable protein scaffolds will undoubtedly arise.

5.3. Directed evolution of hemoproteins for abiological catalysis

Although metalloproteins catalyze an impressive set of reactions, there are many synthetically useful metal-catalyzed transformations that are *not* found in nature – the question becomes: how do we identify proteins capable of performing or acquiring some of these non-natural activities? And, if a protein catalyzes a non-natural reaction, can it be improved or diversified using directed evolution? One approach to creating new enzymes relies on using the catalytic

promiscuity of natural enzymes to find suitable starting points for directed evolution. Metalloproteins containing synthetically versatile cofactors are a good place to start – in mechanistically guided efforts, researchers challenge these proteins with synthetic reagents known to be capable of reacting with the cofactor. This approach has been especially successful with hemoproteins, which have served as starting points for a number of synthetically useful carbene and nitrene transfer reactions in recent years (27). After identifying a starting point, the next question is whether the fitness landscape for these abiological reactions is locally smooth and contains higher peaks, such that the new reactivity can be optimized through directed evolution. We will show that the answer appears to be yes, natural metalloproteins can acquire new, non-natural activities, and their performance can be enhanced by directed evolution.

5.3.1. Non-natural carbene transfer reactions with engineered P450$_{BM3}$ variants

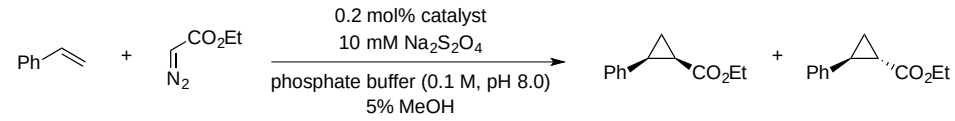
Iron-porphyrin complexes were originally synthesized as models to study the reaction mechanism of cytochrome P450s (28). Since then, these complexes have been shown to catalyze reactions not found in nature. For example, the reaction of an iron-porphyrin complex with a diazo compound and an alkene produces a cyclopropane product, with the reaction thought to proceed through a reactive iron-carbenoid intermediate (29a). Other examples of carbene transfer reactions mediated by iron porphyrins include carbenoid insertion into N–H bonds (29b) and formation of phosphonium ylides (29c-d).

Given this body of work and the similarity between the iron-carbenoid and iron-oxenoid reactive intermediate in the natural P450 hydroxylation mechanism, Coelho *et al.* hypothesized that a cytochrome P450 could perform abiological carbene transfer reactions. Importantly, they further hypothesized that a promiscuous, abiological activity such as olefin cyclopropanation could be enhanced by directed evolution (30). It was first observed that styrene and ethyl diazoacetate in the presence of wild-type P450$_{BM3}$ and sodium dithionite (a reductant) formed a small amount of cyclopropane product **5.6** (1% yield, 5 TON) under anaerobic conditions (Table 5-1). Hemin (Table 5-1, entry 1) and other hemoproteins (cytochrome c, myoglobin, and horseradish peroxidase) also produced **5.6** under the reaction conditions. However, only wild-type P450$_{BM3}$ gave a different product selectivity profile (*cis*:*trans* = 37:63, Table 5-1, entry 2) than that of hemin (>90% *trans*), suggesting that the reaction catalyzed by P450$_{BM3}$ occurred in the active site and therefore could be modulated by the protein scaffold.

In the initial search for a more active and *cis*-selective cyclopropanation catalyst, P450$_{BM3}$ variants from existing libraries (which had been evolved for monooxygenation of diverse substrates) were tested. These efforts identified the highly activating T268A mutation. Introduction of this mutation to variant 9-10A-TS(F87V) (12 mutations from wild type) yielded variant P450$_{BM3}$-CIS, with over 25-fold increase in activity and good selectivity for the *cis* diastereomer (Table 5-1, *cf.* entries 3 and 4). Further screening of active site variants for improvements in activity and *cis*-selectivity resulted in the addition of mutation T438S, which led to 92:8 *cis* to *trans* selectivity, as well as a 1.5-fold increase in TON (Table 5, entry 5). These

results demonstrated that a metalloenzyme can be improved by directed evolution for a completely non-natural reaction, with significant activity or selectivity enhancements in just a few rounds of mutation and screening.

Table 5-1. Improvement of $P450_{BM3}$ for a non-natural cyclopropanation reaction (30).



styrene ethyl diazoacetate *cis*-**5.6** *trans*-**5.6**

Entry	Catalyst	% yield	TON	cis:trans	$\%ee_{cis}$	$\%ee_{trans}$
1	Hemin	15	73	6:94	1	0
2	$P450_{BM3}$	1	5	37:63	27	2
3	9-10A-TS(F87V)	1	7	35:65	41	8
4	$P450_{BM3}$-CIS	40	199	71:29	94	91
5	$P450_{BM3}$-CIS(T438S)	59	293	92:8	97	66

Directed Evolution ?

While axial ligation of the iron center by a cysteine residue is crucial for monooxygenation activity, additional work on engineering $P450_{BM3}$ for carbene-transfer activity led to the discovery that mutation of the axial cysteine residue to serine (C400S) (31) and histidine (C400H) (32) improves carbene transfer and enables catalysis in whole cells. Introduction of the C400S mutation increases the reduction potential of the iron-heme complex by over 100 mV, allowing for the efficient reduction of the resting Fe(III) state to the active Fe(II) state in the cell. Additionally, introduction of the C400S mutation changes the characteristic CO-bound Soret peak from 450 to 411 nm, resulting in a novel 'P411' family of enzymes that catalyze a variety of non-natural reactions.

Variants bearing the C400S or C400H mutation have served as starting points for directed evolution of carbene- and nitrene-transfer catalysts. In one example, $P450_{BM3}$(C400H) variants were applied to the cyclopropanation of acrylamide **5.7** to form the key intermediate (**5.8**) in a formal synthesis of the antidepressant levomilnacipran (Figure 5.6a) (32). The initial variant investigated for this reaction, $P450_{BM3}$(T268A, C400H) provided the desired product in good yield and diastereoselectivity, but only modest enantioselectivity (42% *ee*). Screening for improvements in enantioselectivity, two rounds of site-saturation mutagenesis at active-site residues followed by recombination of beneficial mutations yielded $P450_{BM3}$-HStar (Figure 5.6b). Using whole *Escherichia coli (E. coli)* cells expressing $P450_{BM3}$-HStar, a preparative scale reaction provided the levomilnacipran precursor **5.8** in 86% isolated yield, with 98:2 diastereoselectivity and 92% *ee* for the desired stereoisomer. The variant also tolerated different

substitution patterns of the amide and the arene moieties, giving high yields and good stereoselectivities for a diverse panel of acrylamide substrates (33). This demonstrates that directed evolution can generate a catalyst with broad applicability to a particular substrate class.

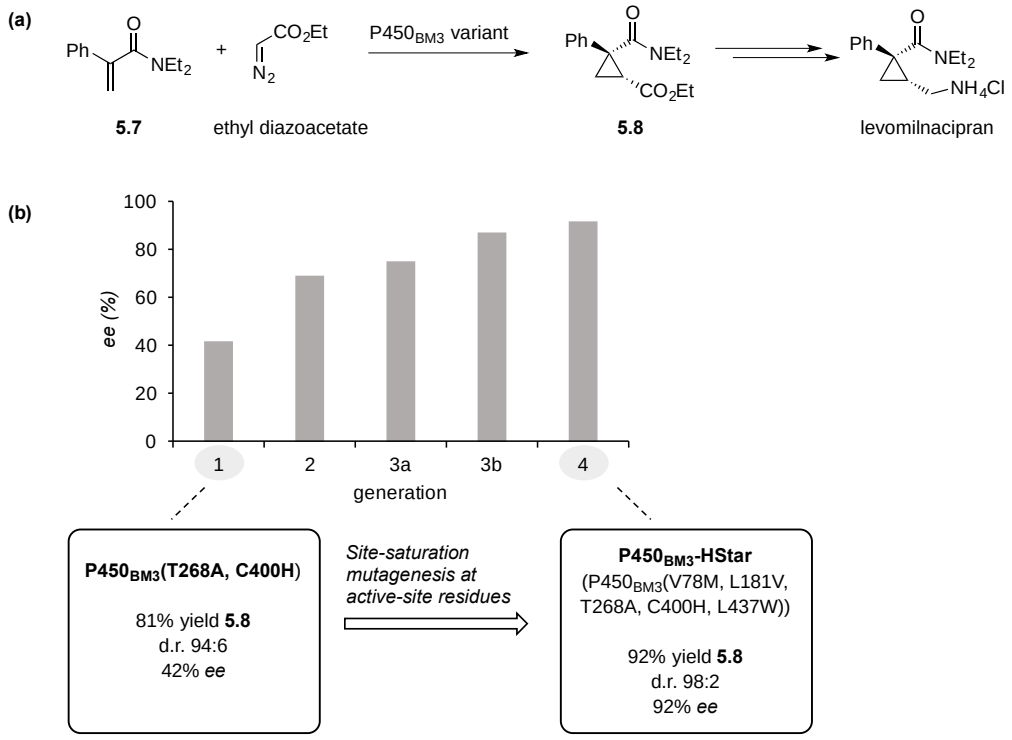


Figure 5.6. Developing a biocatalytic route to antidepressant levomilnacipran (32). **(a)** $P450_{BM3}$ variants catalyze the cyclopropanation of acrylamide **5.7** with ethyl diazoacetate to form levomilnacipran core **5.8**. **(b)** Evolution of a $P450_{BM3}$ variant for improved enantioselectivity for the cyclopropanation of **5.7**.

5.3.2. Non-natural nitrene transfer reactions with engineered $P450_{BM3}$ variants

Metalloporphyrins containing Fe, Co, Mn, or Ru have also been explored for nitrene-transfer reactions, including C–H amination and aziridination of alkenes (34). Given that $P450_{BM3}$ variants could readily adopt the ability to perform carbene transfer, it was hypothesized that their activity could be extended to nitrene transfer as well.

In an early report, trace levels of C–H amination products were detected when rabbit liver microsomal P450s were incubated with iminoiodinane reagents (2 turnovers) (35). Inspired by this report, McIntosh *et al.* investigated $P450_{BM3}$ for intramolecular C–H amination activity with sulfonyl azides (36). Though wild-type $P450_{BM3}$ could perform benzylic C–H amination at low levels (<6 turnovers), this activity could be greatly improved through mutations to the protein scaffold (up to 430 turnovers). In an independent report, Fasan and co-workers identified $P450_{BM3}$ variants with different mutations that also catalyzed amination of sulfonyl azide substrates (37).

Impressively, the $P450_{BM3}$ protein scaffold could be engineered to provide strong control over substrate orientation during a non-natural C–H amination reaction, such that it could override the kinetic bias of the reaction (38). Hyster *et al.* demonstrated this in a system where, within the same substrate, a stronger C–H bond could be aminated in the presence of a weaker C–H bond. They systematically varied the active-site residues and identified two $P450_{BM3}$ variants, both containing the C400S and T268A mutations, with divergent regioselectivity for benzylic or homobenzylic amination within the same substrate (Figure 5.7).

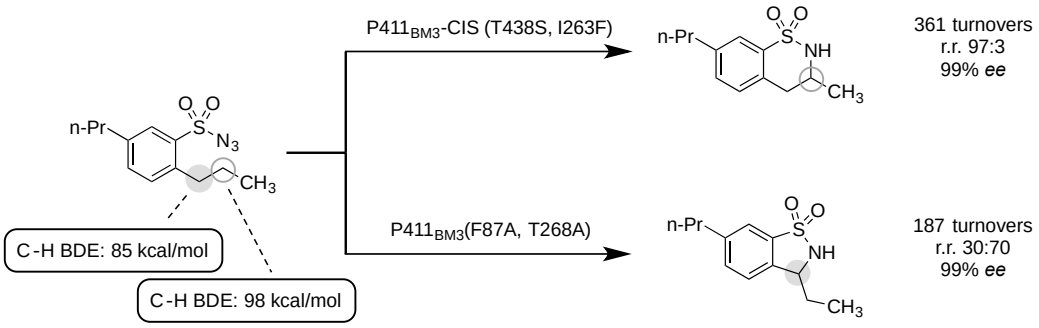


Figure 5.7. Regiodivergent C–H amination catalyzed by engineered cytochrome $P450_{BM3}$ variants (38).

Farwell *et al.* have shown that tosyl azide is a suitable nitrenoid precursor for intermolecular nitrene transfer to sulfides, to form sulfimides (39), and alkenes, to form aziridines (40), under catalysis by $P450_{BM3}$ variants. In the case of aziridination, sequential site-saturation mutagenesis and screening was used to evolve a more active and much more selective variant for the aziridination of 4-methylstyrene with tosyl azide (Figure 5.8). The resulting aziridination catalyst tolerates an array of substituted styrene substrates. Notably, the hemin cofactor is a poor catalyst for both of these transformations under the reaction conditions (<1 TON), highlighting the key role of the protein in introducing novel reactivity.

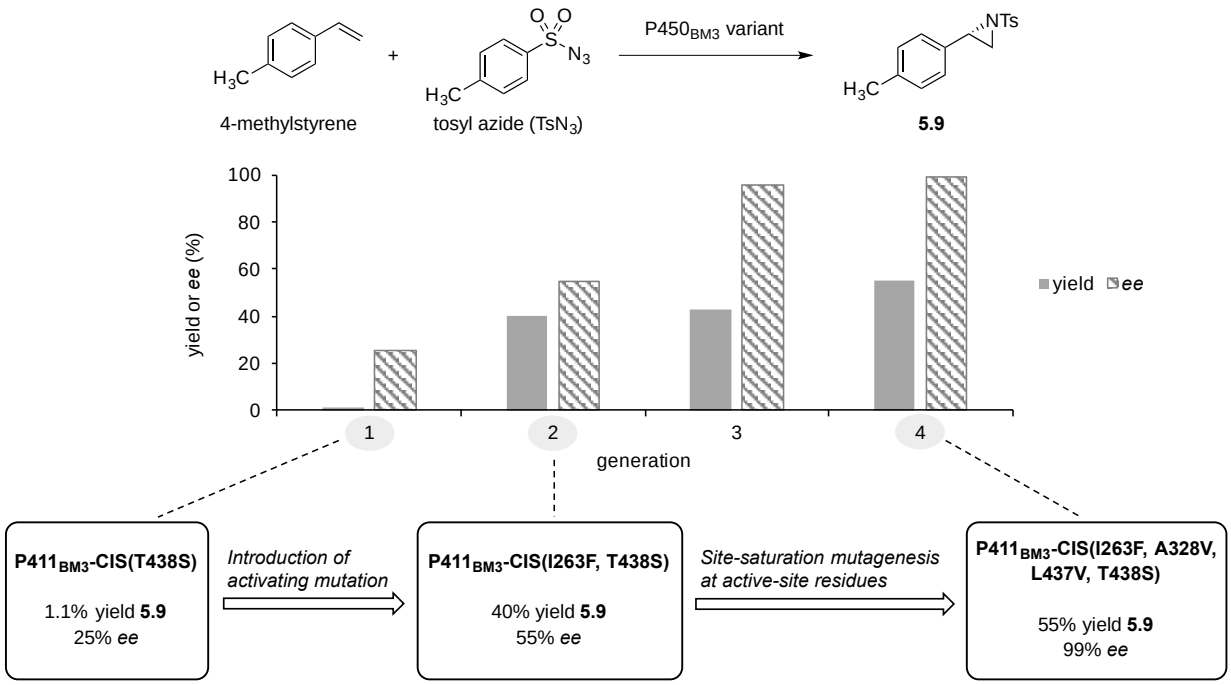


Figure 5.8. Evolution of a highly selective aziridination catalyst derived from $P450_{BM3}$ (40).

Prier *et al.* recently reported P450-catalyzed sulfimidation of allylic sulfides followed by a spontaneous sigmatropic rearrangement to access valuable chiral allylic amines (Figure 5.9a) (41). The initial report of P450-catalyzed nitrene transfer to sulfides focused on aryl-methyl sulfides, such as thioanisole (39). However, the best variant identified in this report, $P411_{BM3}$-CIS(T438S) (15 mutations relative to wild-type $P450_{BM3}$), performed poorly with larger aryl-alkyl substrates and was essentially inactive with sulfide **5.10** (<0.5% yield), the desired substrate for accessing allylic amine **5.11**. A 'substrate walk' approach was taken to evolve a variant for reaction on sulfide **5.10** (Figure 5.9b).

In the first stage, directed evolution involved screening for nitrene transfer to ethyl phenyl sulfide (**5.12**), which forms sulfimide **5.13**. $P411_{BM3}$-CIS(I263F, T438S) was chosen as the parent protein for making site-saturation libraries at active-site residues. The mutations that gave highest activity for substrate **5.12** were then combined in a new variant (P-3), which gave sulfimide **5.13** in >50% yield.

Importantly, P-3 was also significantly improved for nitrene transfer to sulfide **5.10**, validating the evolution strategy. Two additional rounds of evolution were then pursued using P-3 as parent; screening for activity on sulfide **5.10** gave a final variant (P-5) with two additional active-site mutations (A82I and A268G). When (*Z*)-**5.10** (>15:1 *Z*:*E*) was used as the substrate, the evolved variant was capable of promoting the reaction with 77% yield (2,200 turnovers) and 68% *ee*. Further work on this system suggested that the moderate enantiomeric excess is a result of imperfect stereofidelity in the rearrangement step, with the enzyme-catalyzed imidation occurring with high enantioselectivity. Finally, with variant P-5, the reaction of tosyl azide and

sulfide (*Z*)-**5.10** (>15:1 *Z*:*E*) was demonstrated on semi-preparative scale, affording amine **5.11** in 71% isolated yield (6,100 turnovers).

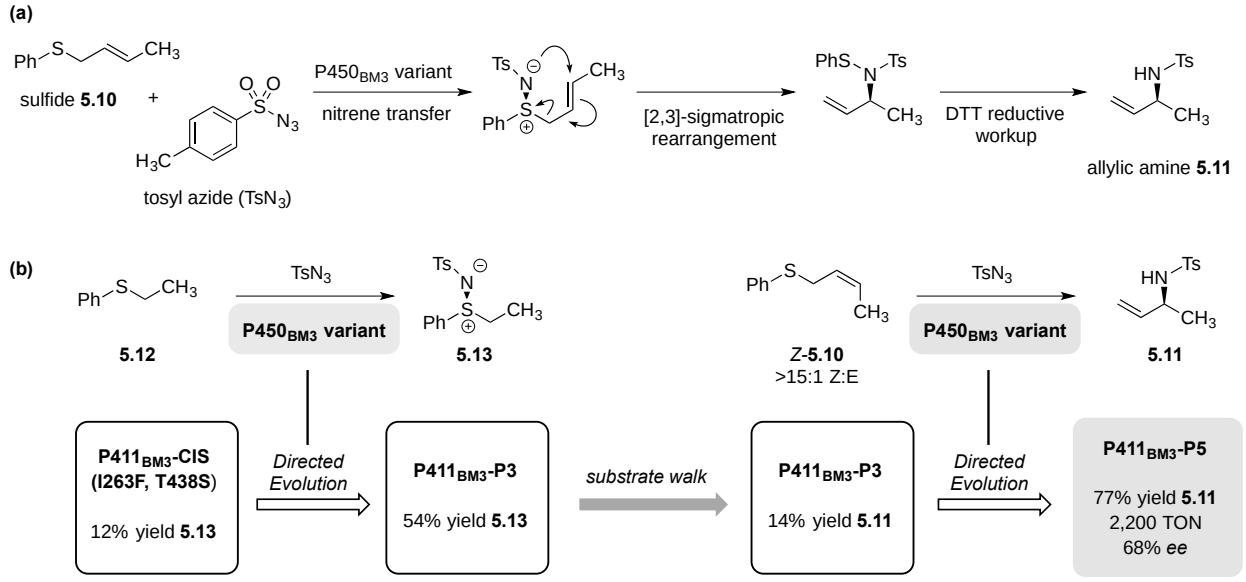


Figure 5.9. Evolution of a protein catalyst for formal asymmetric allylic amination (41). **(a)** Accessing chiral allylic amines via a sulfimidation/sigmatropic rearrangement sequence. DTT = dithiothreitol. **(b)** A substrate walk approach to evolve $P450_{BM3}$-based catalysts for nitrene transfer to an allyl sulfide. TON = turnover number.

5.3.3. Engineering cytochrome *c* for non-natural catalysis

Promiscuous carbene- and nitrene-transfer activity, and the potential to improve this activity through mutation of the protein sequence, is not limited to cytochrome $P450_{BM3}$, or even to metalloenzymes. Kan *et al.* recently discovered that cytochrome *c* from *Rhodothermus marinus* (*Rma* cyt *c*), a protein which normally serves as an electron transporter and has no known natural enzymatic function (42), possesses low-level activity for carbon–silicon bond formation (Figure 5.10a) (43). Enzymes that catalyze carbon–silicon bond formation are unknown in nature, despite the natural abundance of both elements. Interestingly, wild-type *Rma* cyt *c* catalyzed this reaction *via* carbene insertion into a Si–H bond with 10-fold higher activity than the free heme and excellent enantioselectivity (97% *ee*). This indicates that the heme-binding pocket of *Rma* cyt *c* offers a well-defined protein environment for the reaction to take place, even though this protein does not naturally possess an "active site" for substrate binding.

The carbon–silicon bond-forming ability of *Rma* cyt *c* was improved over three rounds of site-saturation mutagenesis. The axial methionine M100, which is coordinatively labile and common in cytochrome *c* proteins, was targeted first, as it was hypothesized that the side chain must be displaced upon iron-carbenoid formation. The M100D mutation stood out as highly activating,

affording the desired organosilicon product as a single enantiomer in 550 TTN, a 12-fold increase compared to the wild-type protein. Saturation mutagenesis at two additional amino acid residues close to the heme iron led to the discovery of triple-mutant V75T M100D M103E, which performed carbon–silicon bond formation with >1500 turnovers and >99% *ee* and is more efficient than any synthetic catalysts for this class of transformation (Figure 5.10a). The beneficial mutations increased the initial rate as well as the TTN: the triple mutant is seven times faster than the wild-type protein, with a turnover frequency of 46 min^{-1}. These mutations presumably also enhance the orientation and binding of the silicon substrate: when 4-(dimethylsilyl)aniline, a substrate that could undergo both Si–H and N–H insertions, was used to probe the bond-forming preference of *Rma* cyt *c* variants in the evolutionary lineage, the triple mutant showed 30-fold improved carbon–silicon bond-forming preference over wild-type cyt *c* (Figure 5.10b).

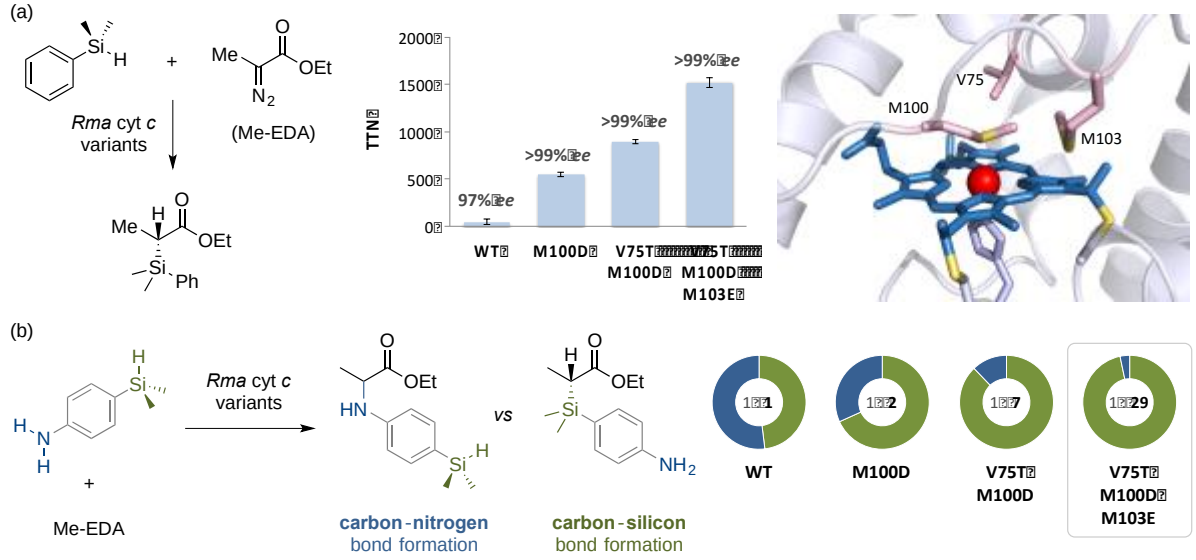


Figure 5.10. *Rma* cyt *c*-catalyzed carbon–silicon bond formation *via* carbene insertion into Si–H bond (43). **(a)** Directed evolution of *Rma* cyt *c* for carbon–silicon bond formation. Amino acid residues M100, V75 and M103 shown in the "active site" structure of wild-type *Rma* cyt *c* (PDB 3CP5) were subjected to sequential site-saturation mutagenesis. **(b)** Chemoselectivity for carbene Si–H insertion over N–H insertion increased markedly during directed evolution of *Rma* cyt *c*. EDA = ethyl diazoacetate. TTN = total turnover number. WT = wild-type.

5.3.4. Engineering myoglobin for non-natural catalysis

Fasan and co-workers have reported that mutation of residues located on the distal side of the heme in sperm whale myoglobin (Mb) can have a dramatic effect on cyclopropanation activity and selectivity (44), among other carbene- and nitrene-transfer reactions (45–49). Myoglobin is an oxygen binding protein with no known catalytic function *in vivo*. Fasan and co-workers

identified Mb as an attractive carbene-transfer catalyst due to its small size (17 kDa) and robustness to mutation.

Wild-type Mb lacks a binding cleft, leading to a similar activity and product selectivity profile as free cofactor (hemin) in the cyclopropanation of styrene with ethyl diazoacetate (Table 5-2, *cf.* entries 1 and 2). Fasan and co-workers hypothesized that amino acid residues on the periphery of the heme-cofactor may affect the transition state of the cyclopropanation reaction and therefore modulate the diastereo- and enantioselectivity of the reaction. To test this hypothesis, they prepared several myoglobin variants containing single amino-acid substitutions at chosen residues in proximity to the distal face of the heme (Figure 5.11). Combining beneficial mutations allowed them to transform myoglobin to a highly diastereo- and enantioselective catalyst capable of effecting cyclopropanation of terminal styrenes with ethyl diazoacetate to afford *trans*-cyclopropanes in high yield (Table 5-2, entries 3–5). Further engineering efforts enabled the development of myoglobin variants for synthesis of the chiral *trans*-cyclopropane cores of several drugs (tranylcypromine, tasimelteon, ticagrelor, and a TRPV1 inhibitor) (44b). However, no myoglobin variants selective for kinetically disfavored *cis*-cyclopropanation have been reported.

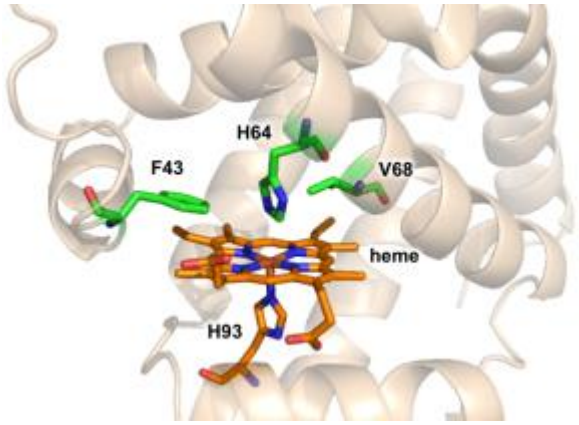
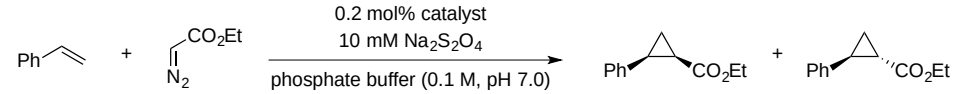


Figure 5.11. Heme cofactor in sperm whale myoglobin is solvent exposed (PDB 1A6K). Mutation of amino acid residues F43, H64, and V68, which are close to the heme cofactor, affect diastereo- and enantioselectivity of a non-natural cyclopropanation reaction.

Table 5-2. Stereoselective cyclopropanation with myoglobin variants (44a).



styrene ethyl diazoacetate *cis*-**5.6** *trans*-**5.6**

Entry	Catalyst	% Conv.	TON	cis:trans	%ee$_{cis}$	%ee$_{trans}$
1	Hemin	29	145	13:87	0	0
2	Mb	36	180	7:93	0	6
3	Mb(H64V)	73	365	4:96	–1	2

4	Mb(V68A)	56	280	2:98	–1	68
5	Mb(H64V, V68A)	>99	500	>0.5:99.5	–6	>99

5.3.5. Directed evolution of myoglobin-derived catalysts created through metal-ion replacement

Hartwig and co-workers recently used the myoglobin scaffold for non-natural catalysis by replacing the native Fe of the cofactor with alternative metals (50). This approach is advantageous because it introduces the reactivity of abiological metals and relies on a metalloprotein scaffold that has already been shown to be evolvable. To design the catalysts, eight myoglobin variants with different axial ligands (H93X) were expressed as apoproteins, then reconstituted with protoporphyrin IX (PPIX) complexes containing Fe, Co, Cu, Mn, Rh, Ir, Ru, or Ag. These were then tested as carbene-transfer catalysts. It was found that proteins bearing Ir(CH_3) porphyrins were superior catalysts for C–H insertion reactions to make cyclic products such as ether **5.14** (Figure 5.12a).

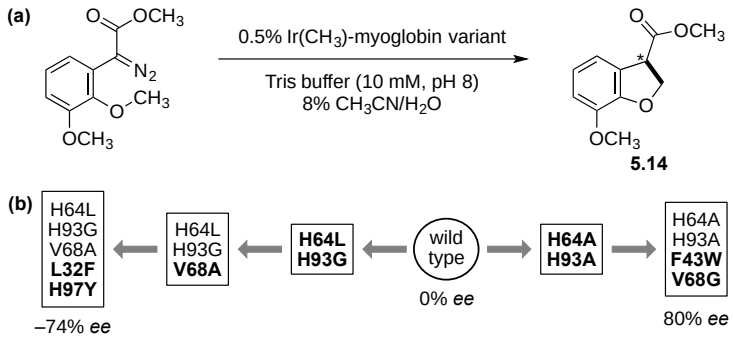


Figure 5.12. Carbene transfer catalyzed by metal-substituted myoglobin variants (50). **(a)** Representative C–H insertion reaction. **(b)** Evolutionary trajectory resulting in variants with divergent enantioselectivity for product **5.14**.

Since the Ir in the cofactor already has an axial ligand (CH_3), the histidine that normally coordinates to the Fe (H93) was mutated to either A or G (Figure 5.12b). To create space for lipophilic substrates to bind, the residue directly above the metal center (H64) was mutated to A, V, L, or I. In order to promote binding with hydrophobic substrates, residues F43 and V68 were mutated to hydrophobic or uncharged residues (V, A, G, F, Y, S, or T). This library of 225 variants was screened against a panel of substrates for reactions like the one depicted in Figure 5.12a. The 22 most active variants were then subjected to another round of mutagenesis, in which L32, F33, H97, and I99 were modified to hydrophobic or aromatic residues (V, L, I, F, H, W, or Y). This culminated in two catalysts that could deliver either enantiomer of **5.14** in moderate enantioselectivity (Figure 5.12b).

5.4. Metalloenzymes with artificial cofactors or metal-binding sites

While variants of natural metalloproteins such as hemoproteins are adept at catalyzing non-natural reactions, scientists have sought to expand the catalytic repertoire even further by creating enzymes with completely artificial metallocofactors. One approach is to introduce amino acids that can coordinate metal ions through their side-chains (e.g., histidine, cysteine, lysine). Several properly positioned coordinating residues will create a new metal-binding site that can compete for locally present metal ions. A second approach is to functionalize an organometallic complex with a pendant moiety that can bind to a protein scaffold. While both approaches have been used to create a number of novel metalloproteins, there are only a few examples of such proteins being optimized for catalytic activity through directed evolution (*vide infra*).

5.4.1. Artificial hydrolase with biotic metal ions in *de novo* binding sites

As part of their ongoing efforts to create supramolecular protein assemblies, Tezcan and co-workers modified cytochrome cb_{562} to contain binding domains (variant AB3) that allow it to spontaneously self-assemble through hydrophobic interactions into a homo-tetrameric complex (51). During expression, the newly synthesized proteins are translocated to the periplasm by means of an N-terminal signal sequence, which leads to the formation of disulfide bonds between cysteines on opposing protein faces, thus stabilizing the tetramer.

To impart catalytic activity, the protein-protein interfaces were designed to have triads of histidine and glutamate residues that could sequester Zn(II) ions present in the periplasm. The aim was to form three-coordinate Zn(II) centers, which would be competent at catalyzing hydrolysis (Figure 5.13a), in analogy to naturally occurring hydrolases such as metallo-β-lactamases. Although the initial construct, which contains eight zinc ions (Zn_8:$AB3_4$, Figure 5.13b), exhibited no hydrolase activity, a single mutation (K104A) led to a variant with detectable activity against *para*-nitrophenyl acetate (*p*-NPA) and the antibiotic ampicillin (Figure 5.13c).

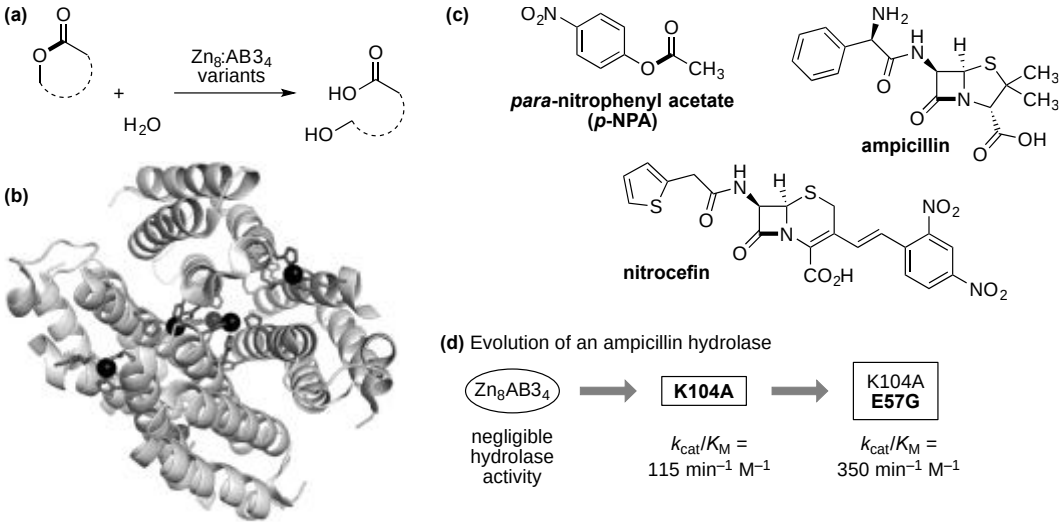


Figure 5.13. Artificial Zn-binding protein of Tezcan and co-workers (51). **(a)** Ester hydrolysis catalyzed by protein variants. **(b)** Tetrameric complex that served as parent for directed evolution (PDB 4U9D). **(c)** Ester substrates described in this study. **(d)** Progression of evolution for ampicillin-hydrolase activity.

This protein construct was an excellent candidate for directed evolution because, like the hemoproteins described in Section 5.3, it could be synthesized entirely *in vivo* and tested without purification from the bacterial cells. Furthermore, its hydrolase activity with ampicillin was sufficient to enable *E. coli* to survive on ampicillin-containing media, thus allowing the authors to use cell survival as a metric for ampicillin-hydrolysis activity. Site-saturation mutagenesis was performed at position 104 as well as three other positions around the active zinc ions: E57, D60, and Y105. At position 104, no residue was superior to alanine, but the additional mutations E57G or Y105T were found to improve cell survival (Figure 5.13d).

When Zn_8:$AB3_4$(K104A) and two double mutants (E57G K104A and K104A Y105T) were tested *in vitro*, only Zn_8:$AB3_4$(E57G, K104A) exhibited Michaelis-Menten kinetics with ampicillin (k_{cat} of 3.5 min^{-1}, K_M of 10 mM). Nonetheless, both double mutants showed higher selectivity for ampicillin *versus* *p*-NPA compared to the single mutant: 3-fold for Zn_8:$AB3_4$(E57G, K104A) and 4-fold for Zn_8:$AB3_4$(K104A, Y105T). The variant Zn_8:$AB3_4$(E57G, K104A) also exhibited activity with β-lactam nitrocefin, albeit with a 160-fold lower k_{cat}/K_M. The authors concluded from these observations that the active sites of the double mutants had been adapted specifically for ampicillin.

5.4.2 Artificial hydrogenases derived from streptavidin

Synthetic chemists have created a vast array of organometallic catalysts. In principle, any of these catalysts can be conjugated to a protein to create a hybrid catalyst whose properties may be tuned through directed evolution. In practice, however, there are fundamental design elements that must be carefully considered if such catalysts are to achieve high activity and selectivity.

Firstly, the metal center must be surrounded by the protein such that the incoming substrates are forced to interact with the protein environment. If the metal center is too exposed to the solvent, the protein may be unable to exert a significant influence on the reaction. Secondly, the protein pocket should hold the organometallic complex in a well-defined orientation, or it may be difficult to select for a single transition state and by extension, a single product. However, this second condition is an evolvable property, so long as the first condition is met.

One conjugation strategy explored for construction of artificial metalloenzymes is to append a biotin linker (Figure 5.14a) to an organometallic complex, then incorporate it into a biotin-binding protein, typically streptavidin (Figure 5.14b). However, there have been few attempts to optimize catalysts of this type through directed evolution. In one notable case, Reetz *et al.* used complex **5.15** to construct a metalloenzyme for the hydrogenation of acrylate **5.16** to alanine derivative **5.17** (52). Wild-type streptavidin (Sav) gave only low enantioselectivity (23% *ee*) in this reaction; thus, the authors' aim was to improve the enantioselectivity using directed evolution. However, these efforts were hindered by numerous technical difficulties, including the low activity of the catalyst system and poor expression levels of the Sav variants. As a result, screening required large expression cultures that made the exploration of broad sequence space, such as by random mutagenesis, infeasible. The screening process was further complicated by the need to titrate the Sav variants to avoid adding excess **5.15**, which would catalyze the hydrogenation unselectively. As a result, the authors selected a narrow set of positions for randomization, comprising residues near the Rh center (distance of 4 to 6 Å) and more distant residues that were expected to influence the protein structure (e.g., residues responsible for secondary structures). This resulted in a relatively small library, formed by independent site-saturation mutagenesis at eleven positions on the protein. This was further reduced to four when it was found that mutation of the other seven positions led to insoluble protein. The authors first identified Sav(S112G) as a variant with modestly improved enantioselectivity (35% *ee*), which served as a parent for site-saturation mutagenesis of N49 (Figure 5.14d). This second library revealed the double mutant Sav(S112G, N49V) as further improved. Finally, the authors randomized position 112 once again; this revealed that the single mutant Sav(N49V) was actually more selective than the double mutant, giving 65% *ee*. Throughout the evolution, there appeared to be no correlation between enantioselectivity and rate. The authors also attempted to generate a catalyst for the antipode of product **5.17**, but were only able to achieve –7% *ee*.

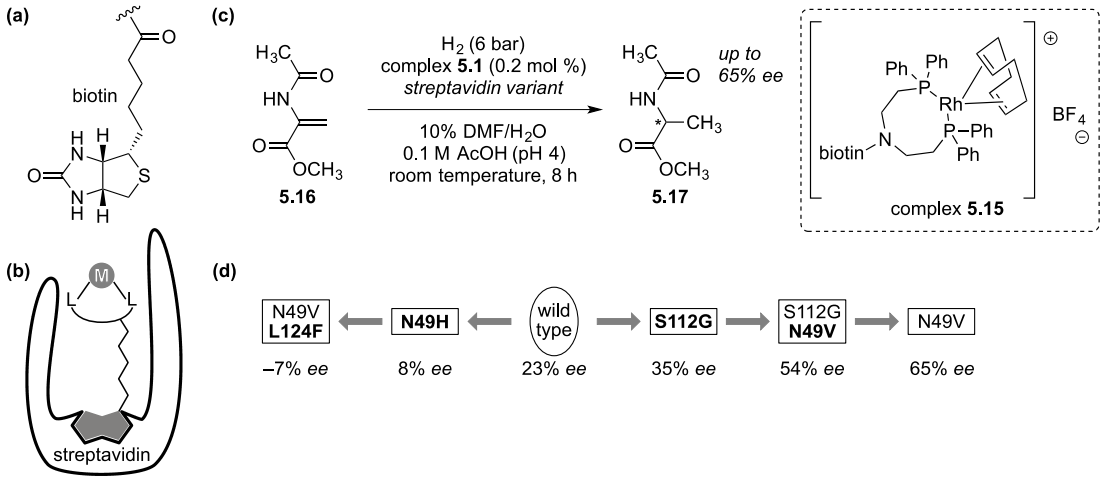


Figure 5.14. Enantioselective hydrogenation with Rh-streptavidin metalloenzymes. **(a)** Biotin linker for attachment of organometallic catalysts. **(b)** Schematic of catalyst design. **(c)** Hydrogenation of alkenes reported by Reetz and co-workers (52). **(d)** Directed evolution to improve enantioselectivity.

Directed evolution was more successful with the metalloenzyme derived from complex **5.18** (Figure 5.15a) (53). After examining a crystal structure of wild-type Sav with complex **5.18a**, Ward and co-workers concluded that residues K121 and L124 were in positions that could interact both with the metal complex and with incoming substrates. They prepared site-saturation libraries at each position for three parents: wild-type Sav and the two single mutants, Sav(S112A) and Sav(S112K). Screening was facilitated by immobilization of the variants on biotinylated Sepharose, which allowed excess **5.18** to be washed away. With complex **5.18b** as the cofactor, a single round of mutagenesis and selection revealed Sav(L124V) as a highly enantioselective catalyst for the reduction of ketone **5.19** to alcohol (*R*)-**5.20** (91% *ee*). For ketone **5.21**, however, two mutations (S112A and K121T) were required to give alcohol (*R*)-**5.22** in high enantioselectivity (88% *ee*, Figure 5.15b). Additionally, the authors found that complex **5.18a** in conjunction with the double mutant Sav(S112A, K121N) gave promising selectivity for the enantiomer (*S*)-**5.22** (–72% *ee*). This was not optimized further.

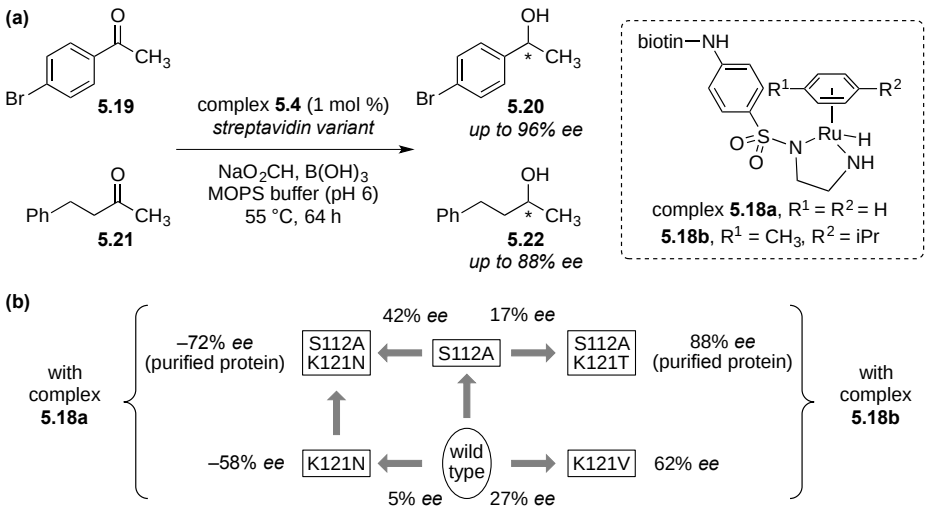


Figure 5.15. Reduction of ketones to alcohols with Ru-streptavidin metalloenzymes (53). **(a)** Reactions tested by Ward and co-workers. **(b)** Enantiodivergent catalysts obtained through directed evolution.

These studies are similar to the Ir-myoglobin carbene transferase (Section 5.3.5) in that, for all three cases, a known organometallic catalyst was introduced into a protein scaffold with the hypothesis that the protein would impart selectivity to the reaction. Notably, there appeared to be little correlation between enantioselectivity and activity: while the final variants achieved moderate to excellent levels of *ee*, the activities of the enzymes were not substantially different from the unbound cofactors. As a result, the directed evolution was complicated by the need to inactivate or extract the unbound cofactor. This contrasts with the $P450_{BM3}$ systems (Sections 5.3.1 and 5.3.2), in which whole cells can be used, due to the substantially higher activity of the protein catalysts over the cellular background.

5.4.3 Cross coupling with a Pd-streptavidin conjugate

In another study, Ward and co-workers combined Sav with Pd complex **5.23** to catalyze the Suzuki coupling of arenes **5.24** and **5.25** (Figure 5.16a) (54). At room temperature, wild-type Sav affords biaryl **5.26** with only moderate *ee* (58%), thus the goal was once again to apply directed evolution to improve the enantioselectivity (Figure 5.16b). Screening of site-saturation libraries at S112 and K121 revealed Sav(K121E) as the most improved variant (76% *ee* for product **5.26**). This was used as the parent for another site-saturation library at S112, which led to the double mutant Sav(S112Y, K121E). This variant afforded biaryl **5.26** in 80% *ee*, which improved to 90% *ee* upon lowering the reaction temperature to 4 °C.

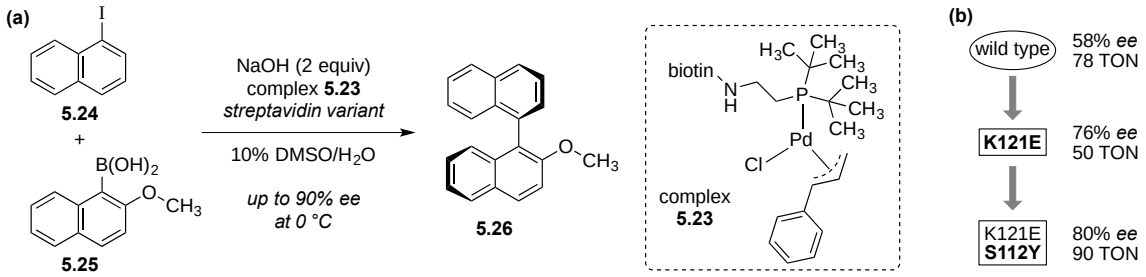


Figure 5.16. Suzuki cross-coupling catalyzed by a Pd-streptavidin metalloenzyme (54). **(a)** Reaction tested by Ward and co-workers. **(b)** Optimization of enantioselectivity.

5.4.4. Alkene metathesis catalyzed by a Ru-streptavidin conjugate

Recently, Ward and co-workers reported a Sav-based metalloenzyme for alkene metathesis (Figure 5.17a) that relies on Ru complex **5.27** (Figure 5.17b) (55). Similar to the approach of Tezcan *et al.*, the Sav variant was expressed with an N-terminal signal sequence that caused translocation of the protein to the periplasm, thus enabling the enzyme to assemble and function in the presence of whole cells. This greatly simplified the screening protocol compared to previous Sav systems because any unbound cofactor was inhibited by cellular components, thus obviating the need for catalyst purification. This, in turn, permitted the investigation of twenty site-saturation libraries, comprising the residues closest to the Ru center. Fluorescent lactone **5.28** (Figure 5.17c) was chosen as the product for screening. Fourteen of the residues could be mutated with positive effect, thus mutagenesis was repeated on these sites over five rounds of mutagenesis and screening. This resulted in a variant with five mutations (V47A, N49K, T114Q, A119G, and K121R) that increased the activity per cell by about 5-fold compared to wild-type Sav. When tested *in vitro*, however, the quintuple mutant showed only modest improvement in V_{max} and K_M (1.3- and 1.5-fold, respectively), leading to only a 2-fold increase in k_{cat}/K_M and a 2.6-fold increase in turnover number. The reason for this is unclear, but one possibility is that the directed evolution improved the cell specific activity by means unrelated to the metathesis, such as by increasing the protein-expression level or the periplasmic-localization efficiency.

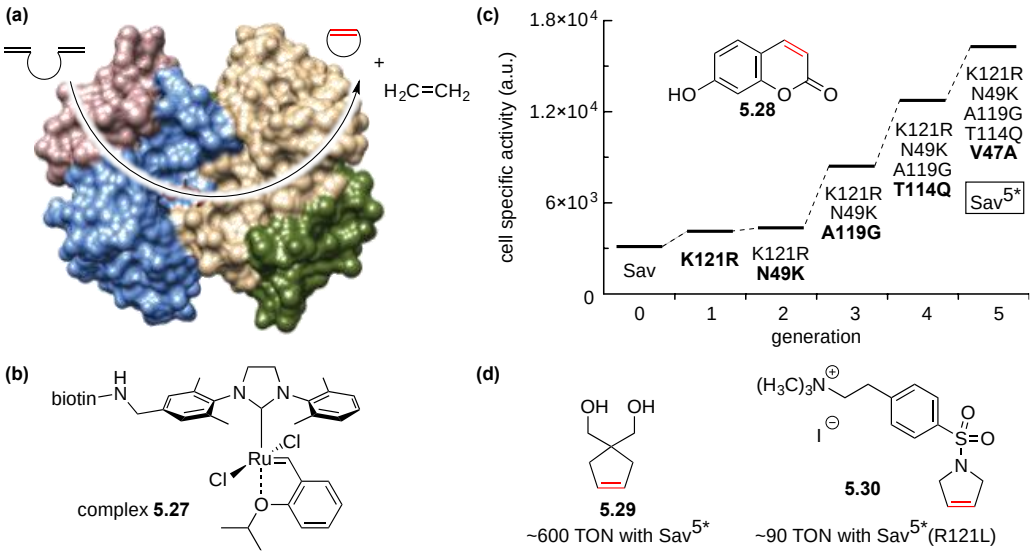


Figure 5.17. Alkene metathesis catalyst derived from streptavidin (55). **(a)** Model of complex **5.27** in streptavidin (reproduced from ref 55). **(b)** Ru-based metathesis catalyst modified for protein conjugation. **(c)** Directed evolution to improve activity for synthesis of lactone **5.28**. **(d)** Additional products described in this study.

The quintuple mutant also gave improvement in TON for product **5.29** (1.5-fold), but performed poorly with cationic product **5.30** (Figure 5.17d). Based on the hypothesis that productivity of **5.30** was decreased by Coulombic repulsion between the corresponding substrate and R121, residue 121 was subjected to another round of mutagenesis, revealing R121L as the optimal variant (1.6-fold increase in TON compared to wild-type Sav).

5.4.5. Carbene transfer with conjugate of rhodium and proline oligopeptidase

Lewis and co-workers have also reported a design strategy that relies on the strain-promoted azide-alkyne cycloaddition to attach a rhodium complex covalently to the interior of a protein (Figure 5.18a) (56). Proline oligopeptidase from *Pyrococcus furiosus* (*Pf*POP) was chosen due to its hyperthermostability and large interior cavity, which was expected to envelop the Rh complex, thus enhancing the ability of the protein to influence activity. In the first stage of protein engineering, the catalytic serine (S477) was mutated to *para*-azidophenylalanine (Z), then four other positions (E104, F146, K199, and D202) were mutated to alanine to allow for complex **5.31** to enter the active site and engage in the cycloaddition. After optimization of conditions, this enzyme (POP-A_4) was able to catalyze the cyclopropanation of styrene **5.32** with diazoester **5.33** to give compound **5.34** in 25% conversion and 38% *ee*. This was accompanied by a substantial amount of side product (0.6:1 **5.34** to **5.35**), which results from competing hydrolysis of the Rh-carbenoid (Figure 5.18b).

In the next phase, ten positions around the non-catalytic Rh were individually mutated to histidine, with the hypothesis that the imidazole side-chain would coordinate to the Rh and minimize movement of the cofactor within the active site. From this library, L328H was found to improve both conversion (61%) and *ee* (85%). Finally, phenylalanine residues were introduced combinatorially at positions predicted to project toward the active rhodium (S64, L97, G99, and G594). From this library, variants POP-A_4(L328H, L97F, G99F) and POP-A_4(L328H, G99F, G594F) improved the *ee* to >90% (Figure 5.18c). While the rate of consumption of diazoester **5.33** progressively decreased in the course of the evolution, the specificity for substrate **5.32** over water increased, such that the overall conversion to **5.34** improved to 73–74%.

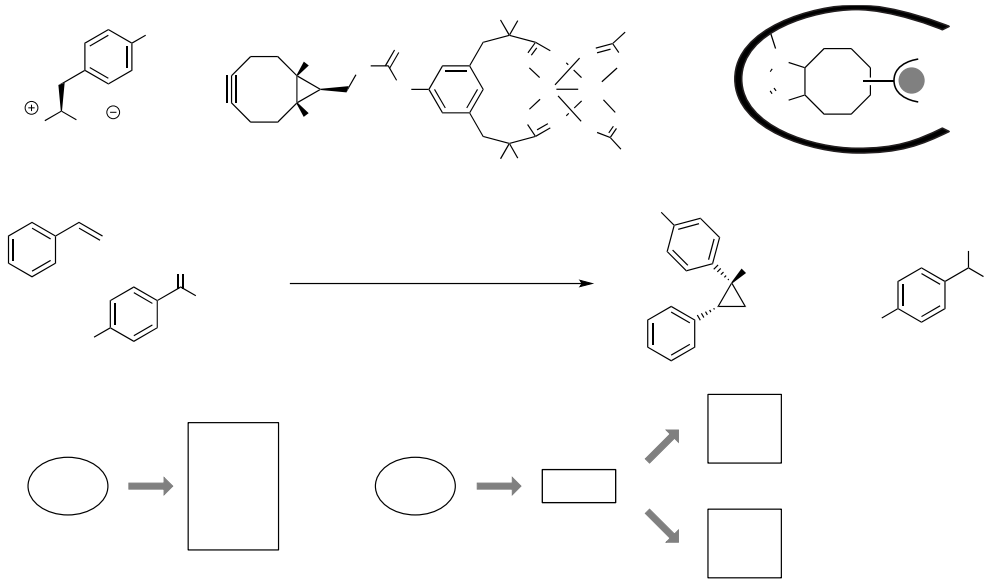


Figure 5.18. A platform for Rh-protein conjugates based on proline peptidase (56). **(a)** Strategy for formation of artificial metalloenzyme. **(b)** Carbene transfer catalyzed by enzyme variants. **(c)** Optimization of activity and enantioselectivity.

As with the Ir-myoglobin carbene transferase (Section 5.3.5) and the streptavidin hydrogenases (Sections 5.4.2 and 5.4.3), this system relies on an organometallic motif (here, dirhodium tetracarboxylate) that is known to catalyze the transformation of interest (here, cyclopropanation). In contrast to these systems, however, the evolution of the Rh-POP system led to concomitant increases in enantioselectivity and conversion. However, the activity increase relative to the free cofactor was still insufficient to obviate the need for removal of unbound cofactor. Thus, directed evolution of this system remains more complicated operationally than with the systems that work in whole cells (Sections 5.3.1–5.3.4).

5.5. Conclusion

Directed evolution, in which beneficial mutations are accumulated over multiple generations to enhance performance, is a powerful algorithm for transforming artificial metalloenzymes into synthetically useful catalysts by enhancing low levels of catalytic activity and selectivity. Through directed evolution, hemoproteins such as $P450_{BM3}$, cytochrome *c* and myoglobin can be repurposed to catalyze non-natural reactions such as carbene and nitrene transfer (Section 5.3). Moreover, natural protein scaffolds can be modified to function as hosts to artificial metallocofactors, as demonstrated in the Zn-binding hydrolase derived from cytochrome cb_{562} (Section 5.4.1); transition metal-conjugated streptavidins for hydrogenation (Section 5.4.2), cross coupling (Section 5.4.3), and olefin metathesis (Section 5.4.4); as well as rhodium complex-linked proline oligopeptidase for carbene transfer reactions (Section 5.4.5).

These examples demonstrate that a wide array of catalysts with a protein scaffold can be improved *via* mutagenesis and screening. However, not all catalysts will eventually produce the high levels of activity and selectivity that are hallmarks of enzymes because the fitness landscapes of artificial metalloenzymes may simply not have significant, accessible fitness peaks. If the local fitness landscape is shallow, or not sufficiently smooth, then finding beneficial mutations may require traversing or sampling a much larger sequence space. Information from X-ray crystal structures and homology models can increase the frequency of beneficial mutations by providing clues as to which residues will exert the largest influence on catalysis. Success is not guaranteed, however, since proteins are dynamic macromolecules and their interactions with abiological substrates and cofactors will likely induce conformational changes that are not predictable. So far, directed evolution of artificial metalloenzymes has relied exclusively on focused libraries guided by structural information, and therefore the evolvability of these scaffolds remains mostly unexplored. Further optimization of artificial metalloenzymes for non-natural reactions will continue to elucidate the requirements for a system that can be evolved effectively. In the meantime, however, we can opine on the necessary characteristics of an evolvable system.

First, it is important that the mature proteins can be assembled and tested with a minimum of post-expression processing, or else the sequence space that can be sampled will be minuscule. In this regard, artificial systems built from natural metalloproteins (Section 5.3), such as $P450_{BM3}$, cytochrome *c* and myoglobin, are ideal. The hydrolase of Tezcan (Section 5.4.1) provides an example of a protein with an artificial metal-binding site that can also assemble entirely during expression, but it is unclear whether such a meticulously designed system will be generalizable to more synthetically useful reactions. The alkene metathase of Ward (Section 5.4.4) and carbene transferase of Lewis (5.4.5) serve as a compromise, wherein protein-cofactor combinations that mediate synthetically desirable transformations can assemble in the presence of whole cells with a minimum of post-expression manipulation.

Second, the protein scaffold must permit optimal exposure of the cofactor to the solvent. This concern is especially relevant for non-natural chemistry because many synthetically valuable

transformations have low water tolerance. For example, in the carbene-transfer catalyst of Lewis (Section 5.4.5), the reactive Rh-carbenoid intermediate undergoes a competing hydrolysis reaction. This was significantly tempered by introduction of hydrophobic residues near the cofactor, suggesting that decreasing water accessibility to the cofactor played a major role in catalyst optimization.

The issue of cofactor placement within a protein scaffold is especially important because it will affect the ability of the protein to bind substrates and, by extension, impart rate acceleration. One way to achieve this is to use a scaffold that is a natural enzyme, such as $P450_{BM3}$, which already possesses a pre-constructed active site evolved to bind small molecules. It remains unclear how well such interactions can be achieved in non-enzyme protein scaffolds. While the streptavidin system has obtained excellent levels of enantioselectivity, the improvements in activity and substrate binding have been relatively minor. In this regard, the examples presented with cytochrome *c* and myoglobin, whose cofactors are much more solvent-exposed than that of a P450, are notable because they demonstrate that even a minimal protein scaffold can be engineered for high levels of activity and selectivity. Recent results with myoglobin (44–49) and other globin-like proteins (57) are promising.

A key question for abiological catalysis with proteins is: what can the protein impart that cannot be readily achieved with small-molecule catalysts? We believe this is a largely a substrate-binding problem. In addition to rate acceleration, such binding will enable precise control of transition-state orientation; this can impart selectivity for kinetically disfavored products, as in the *cis*-cyclopropanation of alkenes (30) or the amination of homobenzylic C–H bonds catalyzed by $P450_{BM3}$ variants (38). While the myoglobin systems can catalyze similar reactions, they have done so only for kinetically favored products, such as *trans*-cyclopropane (44), so far. The streptavidin system has also not yet demonstrated selectivity for kinetically disfavored products, such as *Z*-selectivity in alkene metathesis.

Furthermore, the ability to improve substrate binding through directed evolution may allow earth-abundant metals, like Fe and Co, to catalyze reactions that are typically associated with more precious transition metals, like Rh and Ir. For example, in the study of Ir-myoglobin catalysts for carbene insertion into C–H bonds (50), Hartwig and co-workers observed that Co-myoglobin complexes could also catalyze this reaction in >10 turnovers. Since this activity is measurable, one can imagine applying directed evolution to improve the activity to synthetically useful levels.

Synthetic chemistry has been revolutionized by the use of enzymes, which catalyze natural reactions with high activity and selectivity, often under mild conditions and without the need for protecting groups. We anticipate that directed evolution of artificial metalloenzymes will extend these advantages to non-natural reactions, including those currently inaccessible to small-molecule catalysts.

Acknowledgments

Research performed in the Arnold lab is supported by the National Science Foundation (Division of Molecular and Cellular Biosciences grant MCB-1513007; Office of Chemical, Bioengineering, Environmental and Transport Systems SusCHEM Initiative grant CBET-1403077), the Defense Advanced Research Projects Agency Biological Robustness in Complex Settings Contract (HR0011-15-C-0093), the Caltech Innovation Initiative (CI2) Program, the Jacobs Institute for Molecular Engineering and Medicine (Caltech), and the Gordon and Betty Moore Foundation (through grant GBMF2809 to the Caltech Programmable Molecular Technology Initiative). R. K. Z. is supported by a NSF graduate research fellowship (DGE-1144469) and acknowledges funding from the Donna and Benjamin M. Rosen Bioengineering Center; D. K. R. is supported by Ruth Kirschstein NIH Postdoctoral Fellowship (F32GM117635). Any opinions, conclusions, or recommendations expressed in this material are those of the authors and do not represent the views of any of the funding agencies.

References

1. He J, Muller M, Hertweck C. Formation of the aureothin tetrahydrofuran ring by a bifunctional cytochrome P450 monooxygenase. J Am Chem Soc. 2004 Dec; 126(51): 16742-16743
2. Schmartz PC, Wolfel K, Zerbe K, et al. Substituent effects on the phenol coupling reaction catalyzed by the vancomycin biosynthetic P450 enzyme OxyB. Angew Chem Int Ed Engl. 2012 Nov; 51(46): 11468-11472.
3. Barry SM, Kers JA, Johnson EG, et al. Cytochrome P450-catalyzed L-tryptophan nitration in thaxtomin phytotoxin biosynthesis. Nat Chem Biol. 2012 Oct; 8(10): 814-816.
4. McIntosh J, Farwell CC, Arnold FH. Expanding P450 catalytic reaction space through evolution and engineering. Curr Opin Chem Biol. 2014 Apr; 19: 126-134.
5. O'Brien PJ, Herschlag D. Catalytic promiscuity and the evolution of new enzymatic activities. Chem Biol. 1999 Apr; 6(4): R91-R105.
6. Copley SD. Enzymes with extra talents: moonlighting functions and catalytic promiscuity. Curr Opin Chem Biol. 2003 Apr; 7(2): 265-272.
7. Khersonsky O, Roodveldt C, Tawfik D. Enzyme promiscuity: evolutionary and mechanistic aspects. Curr Opin Chem Biol. 2006 Oct; 10(5): 498-508.
8. Morley KL, Kazlauskas RJ. Improving enzyme properties: when are closer mutations better? Trends Biotechnol. 2005 May; 23(5): 231-237.
9. Stemmer WPC. Rapid evolution of a protein *in vitro* by DNA shuffling. Nature. 1994 Aug; 370(4): 389-391.
10. (a) Romero PA, Arnold FH. Exploring protein fitness landscapes by directed evolution. Nat Rev Mol Cell Biol. 2009 Dec; 10(12): 866-876. (b) Jackel C, Kast P, Hilvert D. Protein design by directed evolution. Annu Rev Biophys. 2008; 37: 153-173. (c) Packer MS, Liu DR. Methods for the directed evolution of proteins. Nat Rev Genet. 2015 Jul; 16(7): 379-394.
11. Smith JM. Natural selection and the concept of a protein space. Nature. 1970 Feb; 225: 563-564.
12. Chu S. Carbon capture and sequestration. Science. 2009 Sep; 325(5948): 1599.
13. Rochelle GT. Amine scrubbing for CO_2 capture. Science. 2009 Sep; 325(5948): 1652-1654.
14. Floyd, 3rd WC, Baker SE, Valdez CA, et al. Evaluation of a carbonic anhydrase mimic for industrial carbon capture. Environ Sci Technol. 2013 Sep; 47(17): 10049-10055.

15. Zhang S, Lu H, Lu Y. Enhanced stability and chemical resistance of a new nanoscale biocatalyst for accelerating CO2 absorption into a carbonate solution. Environ Sci Technol. 2013 Dec; 47(23): 13882-13888.
16. Alvizo O, Nguyen LJ, Savile CK, et al. Directed evolution of an ultrastable carbonic anhydrase for highly efficient carbon capture from flue gas. Proc Natl Acad Sci USA. 2014 Nov; 111(46): 16436-16441.
17. Fox RJ, Davis SC, Mundorff EC, et al. Improving catalytic function by ProSAR-driven enzyme evolution. Nat Biotechnol. 2007 Mar; 25(3): 338-344.
18. (a) Munro AW, Girvan HM, Mason AE, et al. What makes a P450 tick? Trends Biochem Sci. 2013 Mar; 38(3): 140-150. (b) Coon MJ. Cytochrome P450: Nature's most versatile biological catalyst. Annu. Rev. Pharmacol. Toxicol. 2005; 45: 1-25.
19. O'Loughlin TL, Patrick WM, Matsumura I. Natural history as a predictor of protein evolvability. Protein Eng Des Sel. 2006 Nov; 19(10): 439-442.
20. Lewis JC, Bastian S, Bennett CS, et al. Chemoenzymatic elaboration of monosaccharides using engineered cytochrome $P450_{BM3}$ demethylases. Proc Natl Acad Sci USA. 2009 Sep; 106(39): 16550-16555.
21. Kille S, Zilly FE, Acevedo JP, Reetz MP. Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. Nat Chem. 2011 Aug; 3(9): 738-743.
22. For a recent review on this topic, see: Roiban GD, Reetz MT. Expanding the toolbox of organic chemists: directed evolution of P450 monooxygenases as catalysts in regio- and stereoselective oxidative hydroxylation. Chem Commun. 2015; 51: 2208-2224.
23. (a) Meinhold P, Peters MW, Chen MY, et al. Direct conversion of ethane to ethanol by engineered Cytochrome P450 BM3. ChemBioChem 2005 Oct; 6(10): 1765-1768. (b) Xu F, Bell SG, Lednik J, et al. The heme monooxygenase Cytochrome $P450_{cam}$ can be engineered to oxidize ethane to ethanol. Angew Chem Int Ed Engl. 2005 Jun; 44(26): 4029-4032. (c) Fasan R, Chen MM, Crook NC, Arnold FH. Engineered alkane-hydroxylating cytochrome $P450_{BM3}$ exhibiting nativelike catalytic properties. Angew Chem Int Ed Engl. 2007, 46(44): 8414-8418. (d) Fasan R, Meharenna YT, Snow CD, Poulos TL, Arnold FH. Evolutionary history of a specialized P450 propane monooxygenase. J Mol Biol. 2008 Nov; 383(5): 1069-1080.
24. Zhang K, Damaty SE, Fasan R. P450 fingerprinting method for rapid discovery of terpene hydroxylation P450 catalysts with diversified regioselectivity. J Am Chem Soc. 2011 Mar; 133(10): 3242-3245.
25. Kolev JN, O'Dwyer KM, Jordan CT, Fasan R. Discovery of potent parthenolid-based antileukemic agents enabled by late-stage P450-mediated C–H functionalization. ACS Chem Biol. 2014 Jan; 9(1): 164-173.
26. Neufeld K, Henβen B, Pietruszka J. Enantioselective allylic hydroxylation of ω-alkenoic acids and esters by P450 BM3 monooxygenase. Angew Chem Int Ed Engl. 2014 Nov; 53(48): 13253-13257.
27. For recent reviews on this topic, see: (a) Hyster TK, Ward TR. Genetic optimization of metalloenzyes: enhancing enzymes for non-natural reactions. Angew Chem Int Ed. 2016 Jun; 55(26): 7344-7357. (b) Prier CK, Arnold FH. Chemomimetic biocatalysis: exploiting the synthetic potential of cofactor-dependent enzymes to create new catalysts. J Am Chem Soc. 2015; 137(44): 13992-14006. (c) Renata H, Wang ZJ, Arnold FH. Expanding the enzyme universe: accessing non-natural reactions by mechanism-guided directed evolution. Angew Chem Int Ed. 2015 Mar; 54(11): 3351-3367.
28. Mansuy D. Cytochrome P-450 and synthetic models. Pure & Appl Chem. 1987; 59(6): 759-770.
29. (a) Robbins Wolf J, Hamaker CG, Djukic J-P, Kodadek T, Woo LK. Shape and stereoselective cyclopropanation of alkenes catalyzed by iron porphyrins. J Am Chem Soc. 1995 Sep; 117(36): 9194-9199. (b) Baumann LK, Mbuvi HM, Du G, Woo LK. Iron porphyrin catalyzed N–H insertion reactiosn with ethyl diazoacetate. Organometallics 2007 Jul; 26(16): 3995-4002. (c) Mirafzal GA, Cheng G, Woo KL. A new and efficient method for the selective olefination of aldehydes with ethyl diazoacetate catalyzed by an iron(II) porphyrin complex. J Am Chem Soc. 2002 Jan; 124(2): 176-177. (d) Chen Y, Huang L, Ranade MA, Zhang XP. Iron(III) and ruthenium(II) porphyrin complex-catalyzed selective olefination of aldehydes with ethyl diazoacetate. J Org Chem. 2003 May; 68(9): 3714-3717.

30. Coelho PS, Brustad EM, Kannan A, Arnold FH. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. Science. 2013 Jan; 339(6117): 307-310.
31. Coelho PS, Wang ZJ, Ener ME, Baril SA, Kannan A, Arnold FH, Brustad EM. A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins *in vivo*. Nat Chem Biol. 2013 Aug; 9(8): 485-487.
32. Wang ZJ, Renata H, Peck NE, Farwell CC, Coelho PS, Arnold FH. Improved cyclopropanation activity of histidine-ligated cytochrome P450 enables the enantioselective formal synthesis of levomilnacipran. Angew Chem Int Ed. 2014 Jun; 53(26): 6810-6813.
33. Renata H, Wang ZJ, Kitto RZ, Arnold FH. P450-catalyzed asymmetric cyclopropanation of electron-deficient olefins under aerobic conditions. Catal Sci Technol. 2014; 4: 3640-3643.
34. Roizen JL, Harvey ME, Du Bois J. Metal-catalyzed nitrogen-atom transfer methods for the oxidation of aliphatic C–H bonds. Acc Chem Res. 2012 Apr; 45(6): 911-922.
35. Svastits EW, Dawson JH, Breslow R, Gellman SH. Functionalized nitrogen atom transfer catalyzed by cytochrome P-450. J Am Chem Soc. 1985 Oct; 107(22): 6427-6428.
36. McIntosh JA, Coelho PS, Farwell CC, Wang ZJ, Lewis JC, Brown TR, Arnold FH. Enantioselective intramolecular C–H amination catalyzed by engineered cytochrome P450 enzymes in vitro and in vivo. Angew Chem Int Ed. 2013 Aug; 52(35): 9309-9312.
37. Singh R, Bordeaux M, Fasan R. P450-catalyzed intramolecular sp^3 C–H amination with arylsulfonyl azide substrates. ACS Catal. 2014 Feb; 4(2): 546-552.
38. Hyster TK, Farwell CC, Buller AR, McIntosh JA, Arnold FH. Enzyme-controlled nitrogen-atom transfer enables regiodivergent C–H amination. J Am Chem Soc. 2014 Nov; 136(44): 15505-15508.
39. Farwell CC, McIntosh JA, Hyster TK, Wang ZJ, Arnold FH. Enantioselective imidation of sulfides via enzyme-catalyzed intermolecular nitrogen-atom transfer. J Am Chem Soc. 2014 Jun; 136(24): 8766-8771.
40. Farwell CC, Zhang RK, McIntosh JA, Hyster TK, Arnold FH. Enantioselective enzyme-catalyzed aziridination enabled by active-site evolution of a cytochrome P450. ACS Cent Sci. 2015 May; 1(2): 89-93.
41. Prier CK, Hyster TK, Farwell CC, Huang A, Arnold FH. Asymmetric enzymatic synthesis of allylic amines: A sigmatropic rearrangement strategy. Angew Chem Int Ed. 2016 Apr; 55(15): 4711-4715.
42. Stelter M, Melo AMP, Pereira MM, Gomes CM, Hreggvidsson GO, Hjorleifsdottir S, Saraiva LM, Teixeira M, Archer M. A novel type of monoheme Cytochrome *c*: biochemical and structural characterization at 1.23 Å resolution of *Rhodothermus marinus* Cytochrome *c*. Biochemistry. 2008; 47(46): 11953-11963.
43. Kan SBJ, Lewis RD, Chen K, Arnold FH. Directed evolution of cytochrome c for carbon–silicon bond formation: Bringing silicon to life. Science. 2016 Nov; 354(6315): 1048-1051.
44. (a) Bordeaux M, Tyagi V, Fasan R. Highly diastereoselective and enantioselective olefin cyclopropanation using engineered myoglobin-based catalysts. Angew Chem Int Ed. 2015 Feb; 54(6): 1744-1748. (b) Bajaj P, Sreenilayam G, Tyagi V. Gram-scale synthesis of chiral cyclopropane-containing drugs and drug precursors with engineered myoglobin catalysts featuring complementary stereoselectivity. Angew Chem Int Ed. 2016; 55: 16110–16114.
45. Sreenilayam G, Fasan R. Myoglobin-catalyzed intermolecular carbene N–H insertion with arylamine substrates. Chem Commun. 2015 Jan; 51: 1532-1534.
46. Tyagi V, Bonn RB, Fasan R. Intermolecular carbene S–H insertion catalyzed by engineered myoglobin-based catalysts. Chem Sci. 2015 Apr; 6(4): 2488-2494.
47. Tyagi V, Fasan R. Myoglobin-catalyzed olefination of aldehydes. Angew Chem Int Ed Eng. 2016 Feb; 55(7): 2512-2516.
48. Bordeaux M, Singh R, Fasan R. Intramolecular C(sp^3)–H amination of arylsulfonyl azides with engineered and artificial myoglobin-based catalysts. Bioorganic & Medicinal Chemistry 2014 Oct; 22(20): 5697-5704.
49. Giovani S, Singh R, Fasan R. Efficient conversion of primary azides to aldehydes catalyzed by active site variants of myoglobin. Chem Sci. 2016; 7: 234-239.
50. Key HM, Dydio P, Clark DS, Hartwig JF. Abiological catalysis by artificial haem proteins containing noble metals in place of iron. Nature. 2016 Jun; 534(7608): 534-537.

51. Song WJ, Tezcan FA. A designed supramolecular protein assembly with in vivo enzymatic activity. Science. 2014 Dec; 346(6216):1525-8.
52. Reetz MT, Peyralans JJ-P, Maichele A, Fu Y, Maywald M. Directed evolution of hybrid enzymes: Evolving enantioselectivity of an achiral Rh-complex anchored to a protein. Chem. Commun. 2006 Nov; (41):4318-4320.
53. Creus M, Pordea A, Rossel T, Sardo A, Letondor C, Ivanova A, LeTrong I, Stenkamp RE, Ward TR. X-ray structure and designed evolution of an artificial transfer hydrogenase. Angew. Chem. Int. Ed. 2008 Jan; 47(8):1400-1404.
54. Chatterjee A, Mallin H, Klehr J, Vallapurackal J, Finke AD, Vera L, Marsh M, Ward TR. An enantioselective artificial Suzukiase based on the biotin–streptavidin technology. Chem. Sci. 2016 Jan; 7(1): 673-677.
55. Jeschek M, Reuter R, Heinisch T, Trindler C, Klehr J, Panke S, Ward TR. Directed Evolution of Artificial Metalloenzymes for *In Vivo* Metathesis. Nature. 2016 Sep; 537(7622): 661-665.
56. Srivastava P, Yang H, Ellis-Guardiola K, Lewis JC. Engineering a dirhodium artificial metalloenzyme for selective olefin cyclopropanation. Nat. Commun. 2015 Jul; 6, DOI: 10.1038/ncomms8789.
57. Hernandez KE, Renata H, Lewis RD, Kan SBJ, Zhang C, Forte J, Rozzell D, McIntosh J, Arnold FH. Highly stereoselective biocatalytic synthesis of key cyclopropane intermediate to ticagrelor. ACS Catal. 2016 Nov; 6(11): 7810-7813.