

1 **Combining chemistry and protein engineering for new-to-nature biocatalysis**

2
3 David C. Miller, Soumitra V. Athavale, and Frances H. Arnold*

4
5 *Division of Chemistry and Chemical Engineering, California Institute of Technology*

6 *Pasadena, California, 91125*

7 Email: frances@cheme.caltech.edu

8
9 ***Abstract:***

10 Biocatalysis, the application of enzymes to solve synthetic problems of human import,
11 has blossomed into a powerful technology for chemical innovation. In the past decade, a
12 threefold partnership, where nature provides blueprints for enzymatic catalysis, chemists
13 introduce innovative activity modes with abiological substrates, and protein engineers develop
14 new tools and algorithms to tune and improve enzymatic function, has unveiled the frontier of
15 *new-to-nature* enzyme catalysis. In this perspective, we highlight examples of interdisciplinary
16 studies which have helped to expand the scope of biocatalysis, including concepts of enzymatic
17 versatility explored through the lens of biomimicry, to achieve both activities and selectivities
18 that are not currently possible with chemocatalysis. We indicate how modern tools, such as
19 directed evolution, computational protein design and machine learning-based protein engineering
20 methods, have already impacted and will continue to influence enzyme engineering for new
21 abiological transformations. A sustained collaborative effort across disciplines is anticipated to
22 spur further advances in biocatalysis in the coming years.

23

24

25 Exploration of catalysts to facilitate and refine reactivity is a central enterprise of organic
26 chemistry. Over the past century, synthetic methods have been transformed by increasingly
27 sophisticated catalytic systems that enable kinetic and stereoelectronic control over intermediates
28 and reaction pathways. As a result, catalysis is used to synthesize medicines that cure once-lethal
29 diseases, agrochemicals that allow food production to sustain a growing population, and many
30 other essentials of modern society. However, these remarkable advances are only the latest
31 punctuation in Earth's chemical history. Billions of years before human chemists set out to
32 harness the practical entirety of the periodic table, let alone formulate the rudimentary principles
33 underlying modern chemistry, nature had already worked out a platform for executing catalysis
34 for the processes of life in the form of amino acid-based polymers – enzymes – whose intricate
35 and diverse three-dimensional structural space could be sampled by evolution to produce a
36 potentially endless variety of catalysts for implementing a diverse array of chemistry.

37 Although common to both human and natural chemistry, catalyst development is
38 approached differently by these parties. Enzymes are specialized macromolecules, evolved over
39 time to catalyze specific transformations with high efficiency under defined environmental
40 constraints. Assembled from a small set of building blocks and cofactors, they catalyze diverse
41 processes ranging from ester hydrolysis and radical methylations to photosynthesis and nitrogen
42 fixation. In contrast, human-developed catalysts capitalize on the entirety of the periodic table
43 and perform myriad types of reactions, including many unknown in nature. While versatile, these
44 catalysts are often unable to match the hallmark rates and selectivities of an enzyme. Biocatalysis
45 is emerging as an interdisciplinary field that seeks to bridge these capabilities. Protein engineers
46 have enabled us to steer enzymatic activities towards new functions to address synthetic
47 problems facing chemists.¹⁻³ By expanding the catalytic repertoire of nature through the

48 discovery of enzymes for abiological reactions, it will be possible to exploit the robust rates and
49 precise selectivities of enzymes for more transformations of human interest.^{4,5} Collaborations
50 between chemists and protein engineers, with help from nature, have already achieved
51 impressive successes, including establishing new chemical transformations. This Perspective
52 highlights a few interdisciplinary investigations underlying advances in the field and identifies
53 new directions that will accelerate the future development of enzymes for chemical synthesis.

54 Living systems have taught us much about synthesis through their diverse chemistry, and
55 the study of enzyme mechanisms has been a cornerstone for developing principles and theories
56 of chemical catalysis. Biomimetic catalysis, the development of human-made catalysts which
57 seek to imitate life's reactions and emulate enzymatic substrate activation, is the first of many
58 important collaborations between synthetic chemistry and enzymatic catalysis, with origins in the
59 1950s.^{6,7} While unable to recapitulate the fast rates and unparalleled selectivities available to
60 enzymes, biomimicry has continued to drive advances in synthetic chemistry, including the
61 development of catalytic reactions which nature is not yet known to perform. Examples of
62 applying the lessons of biology to catalysis include designing catalysts based on enzyme
63 cofactors (e.g., *N*-heterocyclic carbene catalysis),⁸ repurposing elementary steps involved in
64 biological redox processes (e.g., proton-coupled electron transfer),⁹ and capitalizing on
65 supramolecular catalysis to raise effective concentrations (e.g., β -cyclodextrin-based catalysts).¹⁰
66 Conversely, nature has proven to be an adept learner, and abiological reactions of biomimetic
67 catalysts have been a fruitful avenue to teach enzymes new functions.¹¹

68 The remarkable manner in which biology and chemistry have mutually inspired,
69 informed and advanced each other is apparent in a story of cytochrome P450 monooxygenases.
70 The astonishing ability of heme-containing P450 enzymes to selectively oxygenate specific C–H

71 bonds with molecular oxygen has captured the attention of the broader chemical community
72 since their discovery in the 1960s. Biomimetic chemists played a critical role in elucidating the
73 mechanism of these enzymes – pioneering studies by Groves using iron tetraphenylporphyrin
74 complexes as synthetic model systems elucidated the P450 “radical rebound” mechanism¹² –
75 while numerous other metalloporphyrin complexes were discovered to perform oxygenation
76 reactions.^{13,14} Inspired by these works and related studies for transition metal-catalyzed
77 amination, Breslow and Gellman demonstrated in 1982 that iron tetraphenylporphyrin complexes
78 could accept nitrene precursors and were competent to perform C–H amination reactions (Fig.
79 1a).^{15,16} They also speculated whether a cytochrome P450 itself could be induced to perform
80 amidation in place of its normal oxygenation. Three years later, in collaboration with Svastits
81 and Dawson, they demonstrated that rabbit liver microsomal cytochrome P450-LM3,4 could
82 accept nitrene precursors as substrates to perform C–H amination reactions (Fig. 1b).¹⁷ Although
83 the activity was far too low to be synthetically useful (2.2 turnovers observed for intramolecular
84 sulfamidation), this line of inquiry provided two critical insights: enzymes could perform
85 reactions outside of their natural purview, and chemists could use their knowledge to coax new
86 chemistry out of nature’s catalytic machinery.

87 The next step, elaboration of trace biological activity into robust catalysis, would require
88 protein engineering to train enzymes to perform new reactions with their hallmark rates and
89 selectivity. Here, nature has as much to teach us about her catalyst engineering process,
90 evolution, as she does about catalysis itself (*vide supra*). In his 1970 classic, ‘Natural Selection
91 and the Concept of a Protein Space’, John Maynard Smith provided a theoretical foundation for
92 how protein evolution by natural selection is operationally possible in a sequence space that is
93 both vast and mostly devoid of functional sequences.¹⁸ He argued that for evolution to occur,

94 among the set of possible single mutants of a functional protein there must exist at least one
95 daughter variant which is also functional (in practice, many such variants exist). The same
96 condition applies to the single-mutant neighborhood of the daughter variant, and so on. The
97 inductive logic implies that the sequence space of functional proteins forms a connected network
98 that can be accessed through single amino acid mutations. Advances in screening technology and
99 genetic engineering, notably error-prone polymerase chain reaction (PCR),¹⁹ in the 1980s
100 allowed ‘directed evolution’ to emerge as a tool for engineering proteins: Smith’s conjecture that
101 natural evolution required a minimum density of functional proteins within a single mutational
102 step was used by Arnold, who showed that enzymes could be evolved towards a user-defined
103 goal by accumulating mostly single amino acid changes in a random uphill walk.²⁰ Ever since the
104 first demonstrations in the early 1990s, directed evolution has been a reliable strategy to engineer
105 enzymes with improved properties such as thermostability, the ability to tolerate organic
106 cosolvents, catalytic efficiency, increased substrate tolerance, and many others.^{21,22}

107 But what about new reactivities? It was not immediately clear how such a conservative
108 process could generate enzymes that catalyze new reactions. Here the understanding that
109 enzymes selected by nature for specific transformations may nonetheless display other activities,
110 perhaps following mechanistically related pathways, played an important role. Appreciation of
111 this ‘catalytic promiscuity’ is often lost in the perception of enzymes as being highly specific for
112 single transformations.²³ While an enzyme’s biological function may be a product of natural
113 selection, its promiscuous activities—unless deleterious to the organism—are unrestrained by
114 selective pressures. Indeed, if conditions arise when a promiscuous activity becomes beneficial,
115 natural selection might produce a new enzyme with that function as its mainstay. In fact, new
116 enzymes appear, and they can do so in short time spans, for example, when a new food source or

117 the ability to degrade a new antibiotic or pesticide provides an opportunity to outpace neighbors.
118 This insight into how enzymes have evolved and diversified raised the tantalizing possibility that
119 human-imposed conditions, unnatural reactants, or artificial protein modifications which are
120 either inaccessible or irrelevant in nature might reveal promiscuous activities never seen in the
121 natural world (just as Breslow and coworkers discovered). Provided an enzyme capable of
122 performing a desired new-to-nature transformation can be identified, it should be possible to
123 direct its evolution, elaborating low-yielding, promiscuous enzymatic activity into robust and
124 efficient biocatalysis.

125 Because evolution on a human-friendly timescale can be applied to trace abiological
126 activities, nearly three decades after Breslow and Gellman's original inquiries into enzymatic
127 nitrene transfer, engineered enzymes now perform highly stereoselective C–H amination
128 reactions with hundreds to thousands of turnovers (Fig. 1c).^{24–27} The mechanistic analogy
129 between biomimetic iron-porphyrin catalysts and hemoproteins led to the discovery of other non-
130 natural reactivities supported by these proteins, such as carbene transfer, and even hemoproteins
131 with no natural enzyme functions (for example, globins and cytochromes *c*) have been
132 engineered for abiological carbene-transfer activity (Fig. 2a).^{28,29} Reactions discovered through
133 this manifold include some not observed with chemocatalytic approaches, such as the formation
134 of bicyclobutanes from alkynes.^{30,31}

135 Imaginative chemists have now started to discover other promiscuous activities a protein
136 may support when provided with non-natural reactants, external stimuli, or other abiological
137 components. In addition to hemoprotein-catalyzed carbene and nitrene transfer discussed above,
138 two other notable examples illustrate this creative exercise (Fig. 2). The wide class of flavin-
139 dependent proteins encompasses enzymes that perform myriad redox chemistry. Drawing upon

140 the rich literature of flavin photochemistry, Hyster demonstrated that flavin-dependent enzymes
141 can effect a range of new-to-nature photoredox radical transformations within the active site
142 (Figure 2b).³²⁻³⁴ Visible light excitation of the flavin cofactor allows radical generation from an
143 appropriate precursor (typically, alkyl halides) for subsequent radical reactions, in effect
144 converting an enzyme family known for two-electron reduction into biocatalysts for single-
145 electron transfer reactivity. Stereocontrol imposed by the active site allows asymmetric radical
146 transformations, including asymmetric hydrogen atom transfer – a longstanding challenge in
147 synthetic chemistry.³⁵ Most recently, Zhao and Hyster explored the mechanistic pliability of
148 these enzymes, demonstrating their ability to perform asymmetric hydroalkylation of styrenes
149 and other olefin acceptors.^{36,37} Hartwig has provided another example of how startling
150 promiscuous activities can lurk within the active sites of enzymes performing even the simplest
151 reactions. Carbonic anhydrase catalyzes the addition of water to carbon dioxide to generate
152 carbonic acid via an active site, zinc-hydroxide intermediate. Hartwig and coworkers
153 demonstrated that treating wild-type carbonic anhydrase with a silane can lead to an analogous
154 zinc-hydride intermediate (Figure 2c).³⁸ Although the generation of zinc-hydrides from silanes
155 and zinc salts in anhydrous conditions is known,^{39,40} the formation of such a strongly hydridic
156 species in an enzyme active site is striking. The intermediate could be harnessed for
157 stereoselective reduction of aryl-methyl ketones to secondary carbinols. High activity was often
158 observed in the above examples with just the wild-type (un-mutated) enzymes. Further
159 improvement in enzyme parameters and performance, as well as substrate scope, is possible with
160 directed evolution.⁴¹ The inherent mechanistic versatility of enzymes combined with protein
161 engineering provides immense potential for discovering new biocatalysts.

162 Several reports of new-to-nature biocatalytic transformations highlight the potential of
163 this collaborative effort among nature, synthetic chemistry, and protein engineering to overcome
164 longstanding synthetic challenges (Figure 3). While the development of enzymatic
165 cyclopropanation started with styrenyl alkenes as substrates, directed evolution has since
166 expanded the scope to unactivated olefins. Notably, catalyst-controlled stereodivergent synthesis
167 of all four diastereomeric cyclopropanation products could be achieved (Fig. 3a).²⁸ An
168 engineered cytochrome P450 that forms bicyclobutane products through two sequential carbene
169 addition reactions across alkyne substrates demonstrates the capacity of enzymes to access
170 highly strained carbocyclic cores with excellent efficiency (Figure 3b).³¹ Other engineered P450
171 enzymes catalyze unprecedented asymmetric nitrene C–H insertion reactions to access primary
172 amines at allylic and benzylic positions (Fig. 3c).²⁶ In this system, amination is believed to
173 proceed through an “unprotected” iron-nitrenoid active site intermediate which can be
174 considered as the nitrogen analogue of Compound I. In Hyster’s hydroalkylation reaction of α -
175 acyl radicals and olefins (Fig. 3d), substrate activation in the active site is achieved through a
176 rare quaternary charge-transfer complex between the radical precursor, olefin, flavin cofactor
177 and protein scaffold, which enables stereinduction over a challenging hydrogen-atom transfer
178 step.⁴² These results are corroborated by mechanistic work performed by Zhao on a similar
179 enzymatic hydroacylation reaction.³⁶ Finally, Lewis and colleagues reported the enantioselective
180 intramolecular halocyclization of carboxylic acids onto pendant olefins to form chiral,
181 halogenated γ -lactones (Fig. 3e).⁴³ The enzyme active site prevents racemization of chiral
182 bromonium ion intermediates by limiting halonium ion transfer between olefin substrates,
183 contributing to excellent stereinduction.

184 The extraordinary new chemistry exhibited in these systems is amplified by the
185 amenability of enzymes to operation in sequential reactions which generate tremendous
186 molecular complexity in a one-pot operation, just as they do inside a living cell. Recent examples
187 of biocatalytic cascade processes include Merck's nine-enzyme, three-step route for the synthesis
188 of the investigational HIV treatment islatravir⁴⁴ and Deska's one-pot total synthesis of
189 angiopterlactone B using a five-enzyme cascade.⁴⁵ Furthermore, new-to-nature transformations
190 can be used to develop artificial pathways for synthetic metabolism *in vivo*.⁴⁶ The ability to
191 perform exquisite chemistry in the context of both *in vitro* and *in vivo* cascades demonstrates the
192 unique synthetic opportunities available to biocatalysis, and suggests that the discovery of new-
193 to-nature reactions could serve as versatile modules for "plug-and-play" synthetic cascades. In
194 most of these examples, new activity is recognized using solely the genetically-encoded
195 enzymatic machinery. The growing field of artificial metalloenzymes, where new transition
196 metal cofactors are developed in order to imbue enzymes with novel activities,⁴⁷ promises to
197 further expand the reaction space of biocatalysts once robust methods for their assembly and
198 directed evolution are developed.

199 Advances in our ability to engineer and design proteins have, concurrently with the
200 chemical élan discussed above, facilitated the discovery of new abiological chemistry.
201 Computational modeling has been a useful tool towards this end, and protein modeling, with the
202 goal of designing new enzymes, has received considerable attention over the past two decades.⁴⁸
203 In 2008, Houk and Baker computationally designed an enzyme that catalyzed an abiological
204 Kemp elimination, and Tawfik evolved the minimally performing design to exhibit improved
205 activity.⁴⁹ Since this report, other enzymes have been computationally designed for reactions
206 with limited representation in the natural world, including Diels-Alderase⁵⁰ and Morita-Bayliss-

207 Hillmanases.⁵¹ Initial activities from *de novo* enzymes generally have been very low, reflecting
208 our limited understanding of the problem, but directed evolution can rescue the poor designs to
209 achieve rates and selectivities closer to those of natural enzymes.⁵² Computational design can
210 also guide the expansion of known enzymatic functions. For example, as part of their campaign
211 to develop a biocatalytic process route for the synthesis of sitagliptin, researchers at Merck
212 performed a computational redesign of a transaminase active site to accept the pro-sitagliptin
213 ketone, providing a critical starting point for a directed evolution campaign.⁵³ The wealth of
214 methods to reliably model proteins for both *de novo* generation of new enzymes and redesigning
215 extant ones are invaluable for biocatalysis.

216 Moving forward, new tools for protein engineering which leverage computational protein
217 design, next-generation sequencing and machine learning will accelerate our ability to engineer
218 enzymes with new activities. Directed evolution, agnostic by nature, frequently improves
219 enzyme performance by uncovering subtle mutations far away from the active site, which are
220 non-intuitive for chemists and protein engineers. However, the effects of such changes may be
221 predictable by computer algorithms that learn from data. Machine learning tools have enabled
222 protein structure to be predicted from a sequence with much greater reliability; correlating
223 enzyme sequences to their functions is the next big challenge. Nascent work in this field has
224 demonstrated that machine-learning directed evolution (MLDE) can achieve higher-fitness
225 variants more rapidly than the ‘greedy’ optimization strategies currently employed, reducing the
226 experimental effort needed to achieve a highly functional biocatalyst.⁵⁴ These approaches can
227 help develop enzymes with complementary selectivities: MLDE guided the divergent evolution
228 of a single enzyme for Si-H insertion to produce two biocatalysts with complementary
229 enantioselectivities, enabling access to both antipodes of the desired product.⁵⁵ The capacity to

230 leverage MLDE to engineer enzymes for complementary stereoiduction in parallel could
231 expedite the development of biocatalysts to access any desired stereoisomer. Such examples are
232 only the beginning of what these technologies can do for biocatalysis: the exploration of machine
233 learning to predict enzyme performance for reactions on non-natural substrates has already
234 started,⁵⁶ and we expect it to play a role in biocatalytic reaction discovery sooner rather than
235 later. The acceleration of directed evolution, combined with the ability to optimize enzymes with
236 complementary selectivities in parallel, dramatically amplifies the impact of discovering new-to-
237 nature biocatalytic transformations.

238 In conclusion, biocatalysis is a sustained and fruitful collaboration among synthetic
239 chemists, protein engineers, and nature. Just as nature has taught us chemistry and provided us
240 with the blueprint to improve her catalysts systematically, her machinery has proven highly
241 adaptable to acquiring new chemistry when presented with the right environment. The extent of
242 new chemistry that enzymes can perform remains an open question. We believe, however, that
243 scientists have just scratched the surface of chemical transformations that enzymes can achieve,
244 and anticipate that the boundless surprises from biology, the increasing power and sophistication
245 of protein engineering techniques, and continued chemical ingenuity will continue to power
246 innovations in biocatalysis.

247

248 *Acknowledgments*

249 This report was prepared as an account of work sponsored by an agency of the United
250 States Government. Neither the United States Government nor any agency thereof, nor any of
251 their employees, makes any warranty, express or implied, or assumes any legal liability or
252 responsibility for the accuracy, completeness, or usefulness of any information, apparatus,

253 product, or process disclosed, or represents that its use would not infringe on privately owned
254 rights. Reference herein to any specific commercial product, process, or service by trade name,
255 trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement,
256 recommendation, or favoring by the United States Government or any agency thereof. The views
257 and opinions of the authors expressed herein do not necessarily state or reflect those of the
258 United States Government or any agency thereof.

259 D.C.M. was supported by a Ruth Kirschstein NIH Postdoctoral Fellowship
260 (F32GM128247). This work was sponsored by the US Army Research Office and accomplished
261 under cooperative agreement W911NF-19-2-0026 for the Institute of Collaborative
262 Biotechnologies. This material is based upon work sponsored by the U.S. Department of Energy,
263 Office of Basic Energy Sciences, under Award Number DE-SC0021141. The authors wish to
264 acknowledge Dr. Edwin Alfonzo, Dr. Runze Mao and Kadina E. Johnston for helpful
265 discussions. All protein structures were prepared with PyMOL (The PyMOL Molecular Graphics
266 System, v.2.0 Schrödinger, LLC).

267

268 ***Competing Interests***

269 The authors declare no competing interests.

270

271

272 ***References***

- 273 1. Sun, H., Zhang, H., Ang, E. L. & Zhao, H. Biocatalysis for the synthesis of pharmaceuticals and
274 pharmaceutical intermediates. *Bioorg. Med. Chem.* **26**, 1275–1284 (2018).
- 275 2. Sheldon, R. A., Brady, D. & Bode, M. L. The Hitchhiker’s guide to biocatalysis: recent advances in
276 the use of enzymes in organic synthesis. *Chem. Sci.* **11**, 2587–2605 (2020).

- 277 3. Winkler, C. K., Schrittwieser, J. H. & Kroutil, W. Power of biocatalysis for organic synthesis. *ACS*
278 *Cent. Sci.* **7**, 55–71 (2021).
- 279 4. Renata, H., Wang, Z. J. & Arnold, F. H. Expanding the enzyme universe: accessing non-natural
280 reactions by mechanism-guided directed evolution. *Angew. Chem. Int. Ed.* **54**, 3351–3367 (2015).
- 281 5. Chen, K. & Arnold, F. H. Engineering new catalytic activities in enzymes. *Nat Catal.* **3**, 203–213
282 (2020).
- 283 6. Breslow, R. Biomimetic chemistry: Biology as an Inspiration. *J. Biol. Chem.* **284**, 1337–1342 (2009).
- 284 7. Swiegers, G. F. (ed.) *Bioinspiration and biomimicry in chemistry: reverse-engineering nature.*
285 (Wiley, 2012).
- 286 8. Flanigan, D. M., Romanov-Michailidis, F., White, N. A. & Rovis, T. Organocatalytic reactions
287 enabled by *N*-heterocyclic carbenes. *Chem. Rev.* **115**, 9307–9387 (2015).
- 288 9. Miller, D. C., Tarantino, K. T. & Knowles, R. R. Proton-coupled electron transfer in organic
289 synthesis: fundamentals, applications, and opportunities. *Top. Curr. Chem.* **374**, 30 (2016).
- 290 10. Breslow, R. & Dong, S. D. Biomimetic reactions catalyzed by cyclodextrins and their derivatives.
291 *Chem. Rev.* **98**, 1997–2012 (1998).
- 292 11. Prier, C. K. & Arnold, F. H. Chemomimetic biocatalysis: exploiting the synthetic potential of
293 cofactor-dependent enzymes to create new catalysts. *J. Am. Chem. Soc.* **137**, 13992–14006 (2015).
- 294 12. Huang, X. & Groves, J. T. Beyond ferryl-mediated hydroxylation: 40 years of the rebound
295 mechanism and C–H activation. *J. Biol. Inorg. Chem.* **22**, 185–207 (2017).
- 296 13. Kadish, K. M., Smith, K. M. & Guilard, R. *Bioinorganic and bioorganic chemistry.* (Academic Press,
297 2003).
- 298 14. Mansuy, D. A brief history of the contribution of metalloporphyrin models to cytochrome P450
299 chemistry and oxidation catalysis. *Comptes Rendus Chimie* **10**, 392–413 (2007).
- 300 15. Breslow, R. & Gellman, S. H. Tosylamidation of cyclohexane by a cytochrome P-450 model. *J.*
301 *Chem. Soc., Chem. Commun.* 1400–1401 (1982).

- 302 16. Breslow, R. & Gellman, S. H. Intramolecular nitrene carbon-hydrogen insertions mediated by
303 transition-metal complexes as nitrogen analogs of cytochrome P-450 reactions. *J. Am. Chem. Soc.*
304 **105**, 6728–6729 (1983).
- 305 17. Svastits, E. W., Dawson, J. H., Breslow, R. & Gellman, S. H. Functionalized nitrogen atom transfer
306 catalyzed by cytochrome P-450. *J. Am. Chem. Soc.* **107**, 6427–6428 (1985).
- 307 18. Maynard Smith, J. Natural selection and the concept of a protein space. *Nature* **225**, 563–564 (1970).
- 308 19. McCullum, E. O., Williams, B. A. R., Zhang, J. & Chaput, J. C. Random mutagenesis by error-prone
309 PCR. in *In Vitro Mutagenesis Protocols* (ed. Braman, J.) vol. 634 103–109 (Humana Press, 2010).
- 310 20. Arnold, F. H. Innovation by evolution: bringing new chemistry to life (Nobel Lecture). *Angew.*
311 *Chem. Int. Ed.* **58**, 14420–14426 (2019).
- 312 21. Kuchner, O. & Arnold, F. H. Directed evolution of enzyme catalysts. *Trends in Biotechnology* **15**,
313 523–530 (1997).
- 314 22. Zeymer, C. & Hilvert, D. Directed evolution of protein catalysts. *Annu. Rev. Biochem.* **87**, 131–157
315 (2018).
- 316 23. Khersonsky, O. & Tawfik, D. S. Enzyme promiscuity: a mechanistic and evolutionary perspective.
317 *Annu. Rev. Biochem.* **79**, 471–505 (2010).
- 318 24. McIntosh, J. A. *et al.* Enantioselective intramolecular C-H amination catalyzed by engineered
319 cytochrome P450 enzymes in vitro and in vivo. *Angew. Chem. Int. Ed.* **52**, 9309–9312 (2013).
- 320 25. Steck, V., Kolev, J. N., Ren, X. & Fasan, R. Mechanism-guided design and discovery of efficient
321 cytochrome P450-derived C-H amination biocatalysts. *J. Am. Chem. Soc.* **142**, 10343–10357 (2020).
- 322 26. Jia, Z.-J., Gao, S. & Arnold, F. H. Enzymatic primary amination of benzylic and allylic C(sp³)-H
323 bonds. *J. Am. Chem. Soc.* **142**, 10279–10283 (2020).
- 324 27. Athavale, S. *et al.* Biocatalytic, intermolecular C- H bond functionalization for the synthesis of
325 enantioenriched amides. *Angew. Chem. Int. Ed.* anie.202110873 (2021) doi:10.1002/anie.202110873.
- 326 28. Knight, A. M. *et al.* Diverse engineered heme proteins enable stereodivergent cyclopropanation of
327 unactivated alkenes. *ACS Cent. Sci.* **4**, 372–377 (2018).

- 328 29. Bordeaux, M., Tyagi, V. & Fasan, R. Highly diastereoselective and enantioselective olefin
329 cyclopropanation using engineered myoglobin-based catalysts. *Angew. Chem. Int. Ed.* **54**, 1744–1748
330 (2015).
- 331 30. Yang, Y. & Arnold, F. H. Navigating the unnatural reaction space: directed evolution of heme
332 proteins for selective carbene and nitrene transfer. *Acc. Chem. Res.* **54**, 1209–1225 (2021).
- 333 31. Chen, K., Huang, X., Kan, S. B. J., Zhang, R. K. & Arnold, F. H. Enzymatic construction of highly
334 strained carbocycles. *Science* **360**, 71–75 (2018).
- 335 32. Emmanuel, M. A., Greenberg, N. R., Oblinsky, D. G. & Hyster, T. K. Accessing non-natural
336 reactivity by irradiating nicotinamide-dependent enzymes with light. *Nature* **540**, 414–417 (2016).
- 337 33. Biegasiewicz, K. F. *et al.* Photoexcitation of flavoenzymes enables a stereoselective radical
338 cyclization. *Science* **364**, 1166–1169 (2019).
- 339 34. Grosheva, D. & Hyster, T. K. Light- driven flavin- based biocatalysis. in *Flavin- Based Catalysis*
340 (eds. Cibulka, R. & Fraaije, M.) 291–313 (Wiley, 2021). doi:10.1002/9783527830138.ch12.
- 341 35. Sandoval, B. A., Meichan, A. J. & Hyster, T. K. Enantioselective hydrogen atom transfer: discovery
342 of catalytic promiscuity in flavin-dependent α -ene-reductases. *J. Am. Chem. Soc.* **139**, 11313–11316
343 (2017).
- 344 36. Huang, X. *et al.* Photoenzymatic enantioselective intermolecular radical hydroalkylation. *Nature* **584**,
345 69–74 (2020).
- 346 37. Fu, H. *et al.* Ground-state electron transfer as an initiation mechanism for biocatalytic C–C bond
347 forming reactions. *J. Am. Chem. Soc.* **143**, 9622–9629 (2021).
- 348 38. Ji, P., Park, J., Gu, Y., Clark, D. S. & Hartwig, J. F. Abiotic reduction of ketones with silanes
349 catalysed by carbonic anhydrase through an enzymatic zinc hydride. *Nat. Chem.* **13**, 312–318 (2021).
- 350 39. Mukherjee, D., Ellern, A. & Sadow, A. D. Conversion of a zinc disilazide to a zinc hydride mediated
351 by LiCl. *J. Am. Chem. Soc.* **132**, 7582–7583 (2010).
- 352 40. Sattler, W. & Parkin, G. Zinc catalysts for on-demand hydrogen generation and carbon dioxide
353 functionalization. *J. Am. Chem. Soc.* **134**, 17462–17465 (2012).

- 354 41. Gao, X., Turek-Herman, J., Choi, Y. J., Cohen, R. & Hyster, T. Photoenzymatic synthesis of α -
355 tertiary amines by engineered flavin-dependent α -ene-reductases.
356 <https://chemrxiv.org/engage/chemrxiv/article-details/61429406b817b40e781e701c> (2021)
357 doi:10.33774/chemrxiv-2021-955hf.
- 358 42. Page, C. G. *et al.* Quaternary charge-transfer complex enables photoenzymatic intermolecular
359 hydroalkylation of olefins. *J. Am. Chem. Soc.* **143**, 97–102 (2021).
- 360 43. Mondal, D., Fisher, B. F., Jiang, Y. & Lewis, J. C. Flavin-dependent halogenases catalyze
361 enantioselective olefin halocyclization. *Nat Commun* **12**, 3268 (2021).
- 362 44. Huffman, M. A. *et al.* Design of an in vitro biocatalytic cascade for the manufacture of islatravir.
363 *Science* **366**, 1255–1259 (2019).
- 364 45. Kiefer, A., Liu, Y.-C., Gummerer, R., Jäger, C. & Deska, J. A fully biocatalytic approach to
365 angiopterlactone b based on a chemoinspired artificial in vitro metabolism.
366 [https://chemrxiv.org/articles/preprint/A_Fully_Biocatalytic_Approach_to_Angiopterlactone_B_Base
367 d_on_a_Chemoinspired_Artificial_in_Vitro_Metabolism/14679738/1](https://chemrxiv.org/articles/preprint/A_Fully_Biocatalytic_Approach_to_Angiopterlactone_B_Base_d_on_a_Chemoinspired_Artificial_in_Vitro_Metabolism/14679738/1) (2021)
368 doi:10.26434/chemrxiv.14679738.v1.
- 369 46. Cai, T. *et al.* Cell-free chemoenzymatic starch synthesis from carbon dioxide. *Science* **373**, 1523–
370 1527 (2021).
- 371 47. Schwizer, F. *et al.* Artificial Metalloenzymes: Reaction Scope and Optimization Strategies. *Chem.*
372 *Rev.* **118**, 142–231 (2018).
- 373 48. Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**,
374 320–327 (2016).
- 375 49. Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**,
376 190–195 (2008).
- 377 50. Siegel, J. B. *et al.* Computational design of an enzyme catalyst for a stereoselective bimolecular
378 Diels-Alder reaction. *Science* **329**, 309–313 (2010).

- 379 51. Bjelic, S. *et al.* Computational design of enone-binding proteins with catalytic activity for the
380 Morita–Baylis–Hillman reaction. *ACS Chem. Biol.* **8**, 749–757 (2013).
- 381 52. Blomberg, R. *et al.* Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature*
382 **503**, 418–421 (2013).
- 383 53. Savile, C. K. *et al.* Biocatalytic asymmetric synthesis of chiral amines from ketones applied to
384 sitagliptin manufacture. *Science* **329**, 305–309 (2010).
- 385 54. Wittmann, B. J., Yue, Y. & Arnold, F. H. Informed training set design enables efficient machine
386 learning-assisted directed protein evolution. *Cell Systems* S2405471221002866 (2021)
387 doi:10.1016/j.cels.2021.07.008.
- 388 55. Wu, Z., Kan, S. B. J., Lewis, R. D., Wittmann, B. J. & Arnold, F. H. Machine learning-assisted
389 directed protein evolution with combinatorial libraries. *Proc. Natl. Acad. Sci.* **116**, 8852–8858 (2019).
- 390 56. Goldman, S., Das, R., Yang, K. K. & Coley, C. W. Machine learning modeling of family wide
391 enzyme-substrate specificity screens. arXiv:2109.03900 (2021).

392

393 **Figure Captions**

394

395 **Figure 1. Biomimetic and enzymatic nitrene transfer for C-H insertion reactions.** (a) Biomimetic
396 intramolecular tosylamidation using Fe(TPP)Cl as a catalyst.¹⁶ (b) Enzyme-catalyzed intramolecular tosylamidation
397 using rabbit liver P450 LM3,4 as a biocatalyst.¹⁷ (c) Engineered P411_{CIS}-catalyzed asymmetric intramolecular
398 tosylamidation.²⁴ Bonds formed via nitrene C-H insertion are shown in red. e.e., enantiomeric excess; TPP,
399 tetraphenylporphyrin; TTN, total turnover number.

400

401 **Figure 2: Representative examples of cofactor adaptation for abiological reactions with new-to-nature**
402 **reactivity modes.** (a) The chemistry of metal porphyrins has been extended to hemoproteins for both carbene and
403 nitrene transfer when presented with carbene and nitrene precursors.²⁴⁻³¹ (b) Visible-light stimuli allows for

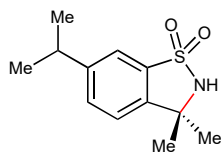
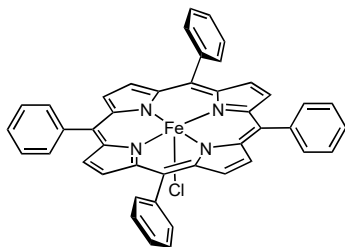
404 flavoproteins to engage in biocatalytic photoredox transformations of alkyl halides.³²⁻³⁴ (c) Carbonic anhydrase uses
405 silanes to generate an active zinc hydride intermediate for new reactions.³⁸ His, histidine.

406

407 **Figure 3: Select examples of new-to-nature enzyme catalysis.** (a) Hemoprotein-catalyzed, stereodivergent
408 cyclopropanation of unactivated olefin substrates.²⁸ (b) P411-catalyzed bicyclobutane formation through sequential
409 carbene additions across alkynes.³¹ (c) Asymmetric primary amination of benzylic C–H bonds catalyzed by P411s.²⁶
410 (d) Biocatalytic, asymmetric hydroalkylation of styrenes via photochemical activation of flavins.⁴² (e) Asymmetric
411 halolactonization catalyzed by halogenase enzymes.⁴³ GDH, glucose dehydrogenase; d.r., diastereomeric ratio; e.e.,
412 enantiomeric excess; LEDs, light-emitting diodes; OD, optical density; Piv, pivaloyl; RT, room temperature; Tf,
413 trifluoromethanesulfonyl; TTN, total turnover number; WCS, whole cell suspension.

a

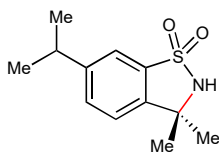
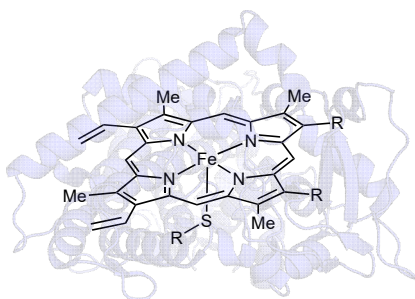
Breslow (1982)
Fe(TPP)Cl



77% yield
15 TTN

b

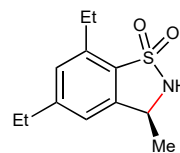
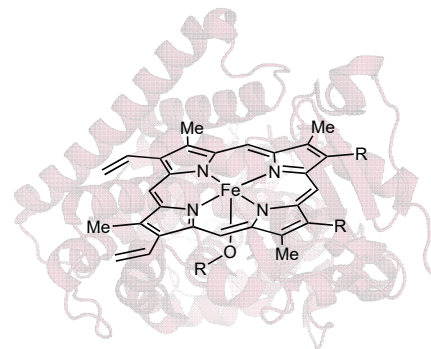
Svastits, Breslow (1985)
Rabbit Liver P450 LM3,4



Trace yield
2.2 TTN

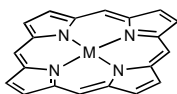
c

Arnold (2013)
Engineered P411_{CIS}

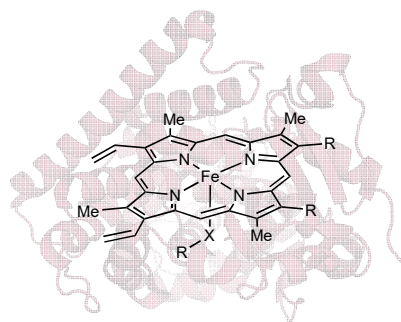


58% yield, 430 TTN
87% e.e.

a Metal porphyrins

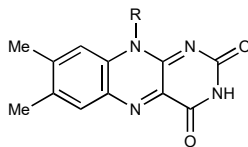


- Carbene transfer
- Nitrene transfer

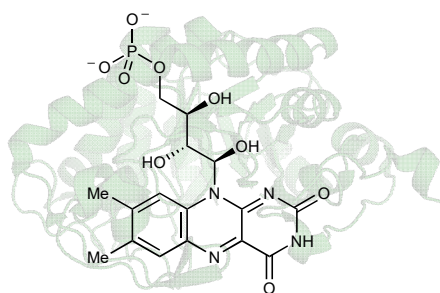


Hemoproteins
Biocatalytic carbene
& nitrene transfer

b Flavinoids

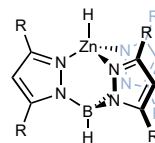


- Energy transfer
- Photoredox

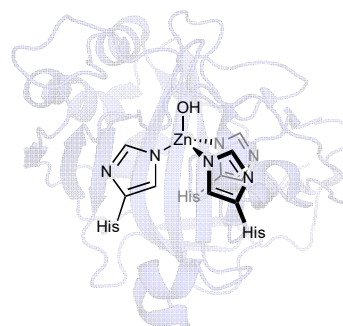


Ene-reductases
Biocatalytic
photoredox reactions

c Zinc hydrides

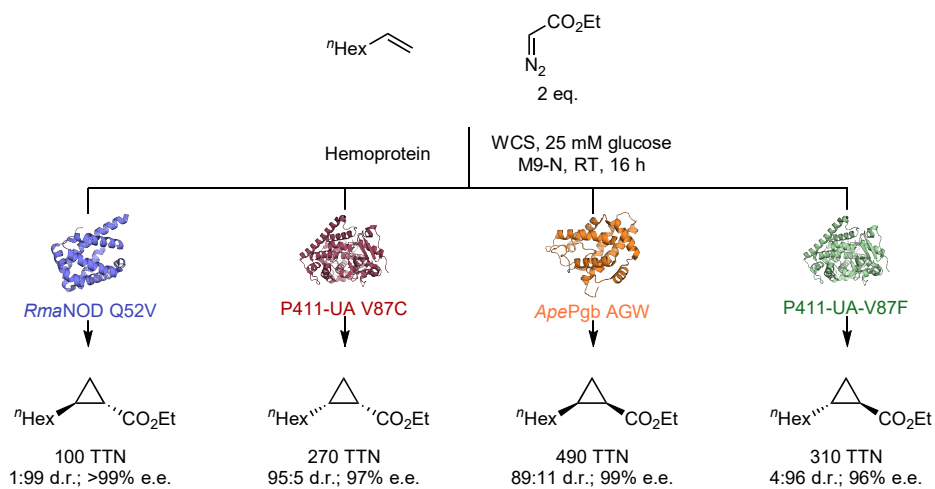


- Metal hydride transfer

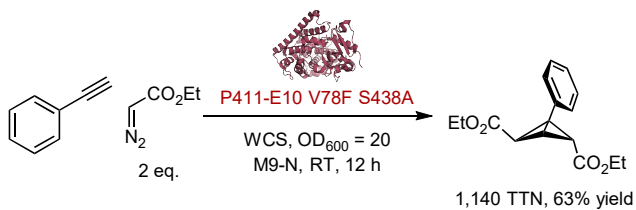


Carbonic anhydrases
Biocatalytic reactions
of metal hydride species

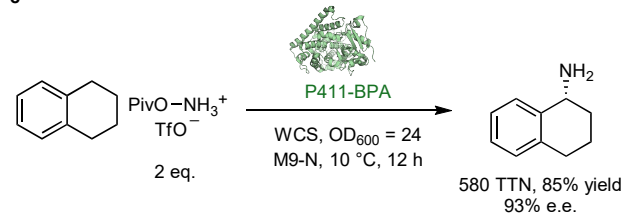
a



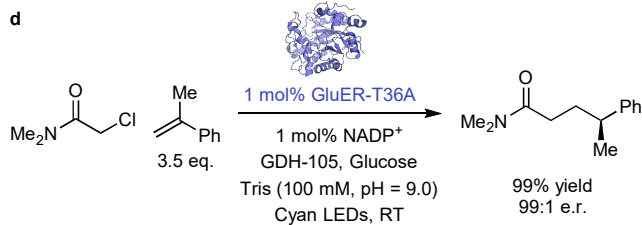
b



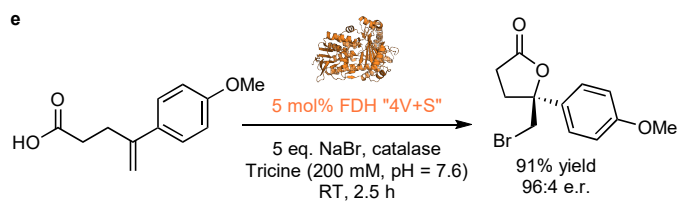
c

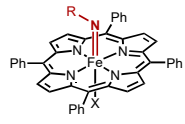


d

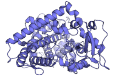


e



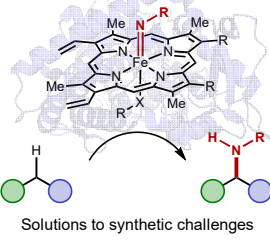


Inspiration from synthetic chemistry



Promiscuity and evolvability of enzymes

New-to-nature biocatalysis



Solutions to synthetic challenges