

Noncanonical heme ligands steer carbene transfer reactivity in an artificial metalloprotein

Moritz Pott^{1,†}, Matthias Tinzl^{1,†}, Takahiro Hayashi¹, Yusuke Ota¹, Daniel Dunkelmann¹, Peer R. E.

Mittl² and Donald Hilvert^{1,*}

¹ Laboratory of Organic Chemistry, ETH Zurich, 8093 Zurich, Switzerland

² Department of Biochemistry, University of Zürich, 8057 Zürich, Switzerland

ABSTRACT

Changing the primary metal coordination sphere is a powerful strategy for modulating metalloprotein properties. Taking advantage of this approach, we have replaced the proximal histidine ligand in myoglobin with the histidine analogues N_δ-methylhistidine (NMH), 5-thiazoylalanine (5ThzA), 4-thiazoylalanine (4ThzA) and 3-(3-thienyl)alanine (3ThiA) by amber stop codon suppression using engineered pyrrolysyl-tRNA synthetases, including two newly evolved enzymes. In addition to tuning the heme redox potential over a >200 mV range, these noncanonical ligands modulate the protein's promiscuous carbene transfer activity with ethyl diazoacetate. Myoglobin variants with increased reduction potentials (NMH and 5ThzA) proved superior for cyclopropanation and N-H insertion, especially under aerobic conditions, and could even promote these reactions in the absence of reducing agent. In contrast, the variants with the lowest E^0 values (4ThzA and 3ThiA) exhibit comparatively high S-H insertion activity even though the respective histidine surrogates do not coordinate the heme iron. Given the important functional roles played by histidine in many enzymes, these genetically encoded

histidine analogues represent valuable tools for probing mechanism and enabling new chemistries in metalloproteins.

MAIN TEXT

Functionally versatile heme proteins are involved in many biological processes. Examples include myoglobin and hemoglobin, responsible for oxygen transport and storage; the electron transport protein cytochrome *c*; catalase which relieves oxidative stress; and cytochrome P450 monooxygenases which oxidize xenobiotics and diverse metabolites.¹ Many of these proteins also promiscuously catalyze abiological reactions with high chemo- and stereoselectivity.² Carbene transfer reactions such as cyclopropanation,^{3,4} aldehyde olefination,⁵ X-H insertion (X= S, N, Si, B),^{6–10} and C-H activation¹¹ are notable in this regard. This novel reactivity has been enhanced and broadened by site directed mutagenesis,^{4–7} directed evolution,^{8–10} metal ion substitution,^{12–16} and replacement of heme by other cofactors.^{17–19}

The advent of genetic code expansion^{20,21} has created new opportunities for tuning heme properties, enabling recombinant replacement of the proximal iron ligand with noncanonical amino acids (ncAAs). Introducing N δ -methylhistidine (NMH) into ascorbate peroxidase, for example, disrupts what was thought to be a crucial Asp-His interaction but substantially increases turnover number without compromising catalytic efficiency.²² NMH also modulates the reactivity of compound II in cytochrome *c* peroxidase²³ and enhances the promiscuous peroxidase²⁴ and cyclopropanation²⁵ activities of myoglobin. Encouraged by the latter findings, we set out to explore the effect of NMH and three additional histidine isosteres—5-thiazoylalanine (5ThzA), 4-thiazoylalanine (4ThzA) and 3-(3-thienyl)alanine (3ThiA) (Figure 1)—on several representative carbene transfer reactions catalyzed by myoglobin.

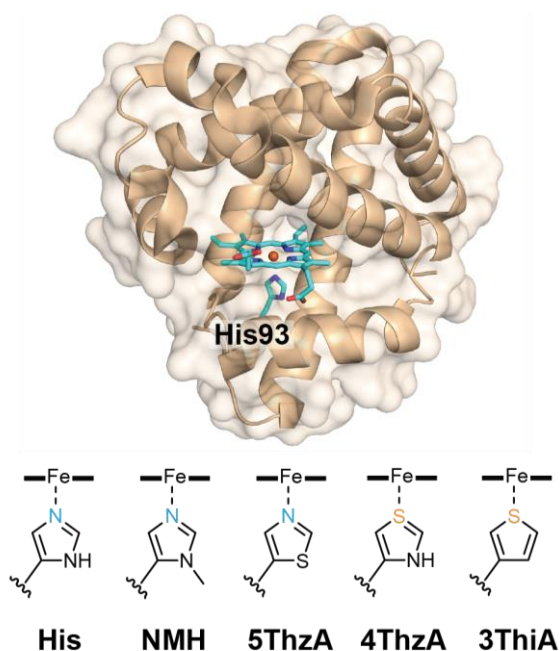


Figure1. Structure of myoglobin and the histidine analogues used to replace His93.

Genetic encoding of ncAAs requires a tRNA/aminoacyl-tRNA synthetase (aaRS) pair orthogonal to the *E. coli* translation machinery to enable efficient suppression of the amber stop codon (UAG) *in vivo*. *Methanosarcina barkeri* pyrrolyl-tRNA synthetase (PylRS) was previously evolved to encode a series of histidine analogues, including NMH,²⁶ so we tested the ability of this enzyme to accept the new ncAAs. We used a reporter assay based on expression of a superfolder GFP (sfGFP)-derived gene containing a single UAG codon for residue 133.^{26–28} 5ThzA afforded threefold higher fluorescence than NMH in this assay (Figure 2), indicating robust incorporation into the full length protein, but only low signal was observed for 3ThiA and 4ThzA.

Although 3ThiA insertion can be reportedly improved by using higher amino acid concentrations,³⁰ more efficient and specific PylRSs for 3ThiA and 4ThzA would be valuable. We therefore randomized nine residues that line the binding pocket of *M. barkeri* PylRS. Seven libraries, each containing different combinations of four mutagenized sites (Table S1), were created, combined, and subjected to three rounds of positive selection for chloramphenicol resistance and two rounds of negative

selection using a toxic barnase gene.³¹ This procedure afforded variants that efficiently accept 4ThzA (PylRS_{4ThzA}: A267Q/N311S/C313W) and 3ThiA (PylRS_{3ThiA}: N311Q/ Y349F) as substrates (Figure 2). Although low levels of sfGFP fluorescence were observed for both aaRSs in the amber suppression assay in the absence of ncAA, addition of the respective histidine isostere increased fluorescence by a factor of five (PylRS_{3ThiA}) to eight (PylRS_{4ThzA}), matching the incorporation efficiency seen for 5ThzA with PylRS_{NMH}. Notably, the new PylRS_{4ThzA} variant outperforms PylRS_{4ThzA_1}, an aaRS for 4ThzA reported while our work was in progress,²⁹ with respect to both efficiency and selectivity (Figure 2).

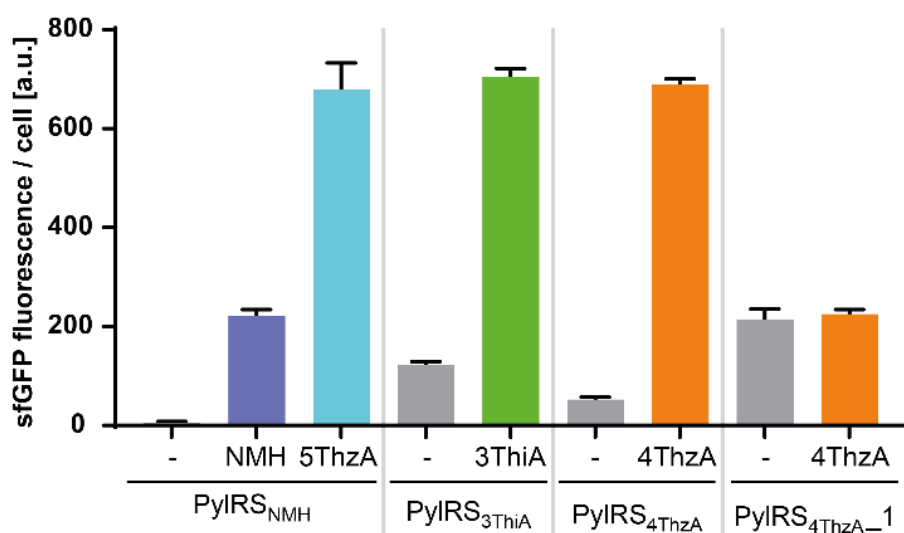


Figure 2. PylRS variants were benchmarked in an assay for amber suppression efficiency. sfGFP(D133AMBER) production levels in the absence (-) or presence of 2 mM ncAA were measured by cellular fluorescence. PylRS_{3ThiA} and PylRS_{4ThzA} were developed in this work, PylRS_{NMH}²⁶ and PylRS_{4ThzA_1}²⁹ were reported previously.

Having the required aaRSs in hand, we replaced the proximal His93 ligand in myoglobin with the four histidine isosteres which present either a nitrogen or sulfur atom in a position suitable for coordination to the heme iron. For these experiments, we used the myoglobin variant Mb* which contains two additional mutations, H64V and V68A, known to boost carbene transferase activity.⁴ The four proteins

were produced in excellent yield (>100 mg/L of culture). Incorporation of the ncAAs was confirmed by HR-MS (Table S2) and, for Mb*(5ThzA) and Mb*(3ThiA), by x-ray crystallography (Figure 3).

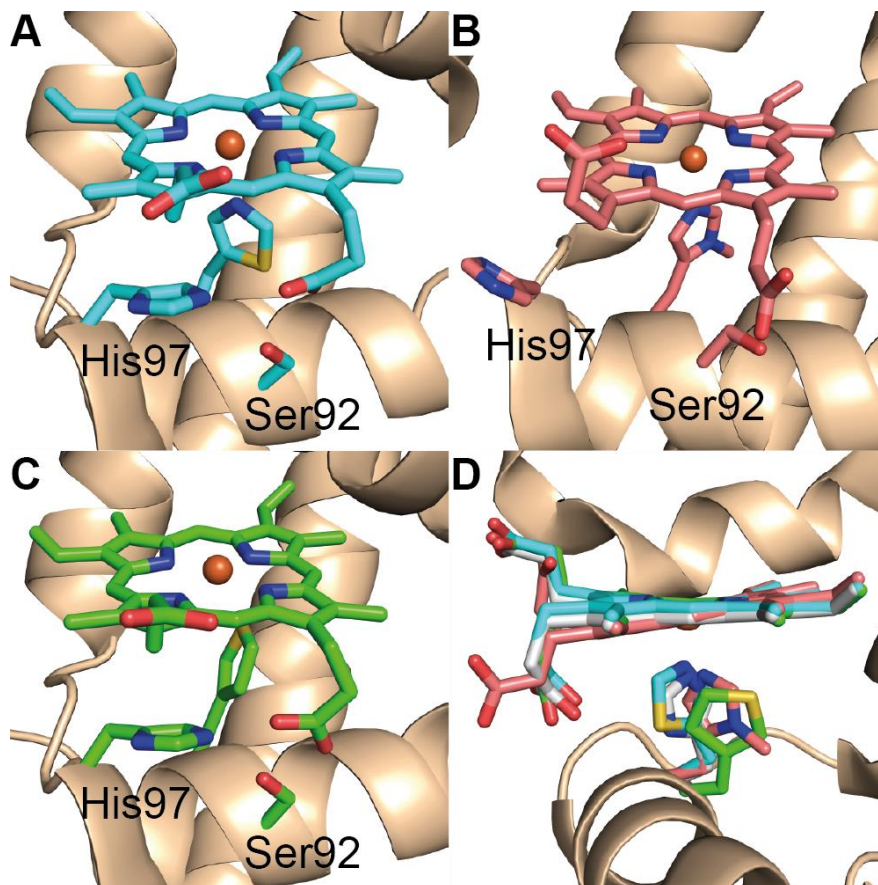


Figure 3. Crystal structures of noncanonical myoglobin variants: (A) Mb*(5ThzA), (B) Mb*(NMH),²⁵ and (C) Mb*(3ThiA). (D) Overlay of Mb*(gray), Mb*(5ThzA) (cyan), Mb*(NMH) (pink), and Mb*(3ThiA) (green). Oxygen, nitrogen and sulfur atoms are colored red, blue and yellow, respectively; iron is shown as an orange sphere.

The crystal structure of Mb*(5ThzA), determined to 1.23 Å, closely resembles Mb* (rmsd 0.10 Å). The 5ThzA thiazole side chain adopts the same orientation as His93 in Mb* and, as expected, coordinates the heme iron via its ϵ -nitrogen atom (Figures 3A, S1). The main difference between Mb*(5ThzA) and Mb* is a 90° rotation of the Ser92 side chain away from the thiazole sulfur atom (Figure S1). Despite this twist, Ser92 still maintains its hydrogen bond to one of the heme propionate groups.

5ThzA thus appears to be a better His93 mimic than NMH, which also coordinates the cofactor via Ne but adopts a different rotameric conformation, leading to a cascade of small changes in the active site, including a 180° rotation of the Ser92 side chain, tilting of the heme plane, and displacement of nearby His97 (Figures 3B, S2).²⁵

In contrast to 5ThzA, 3ThiA is not an effective His93 surrogate. In analogy to the methionine-iron bond in *c* type cytochromes, we had expected the thiophene sulfur atom to coordinate the heme iron, but this is not observed in the 1.96 Å crystal structure of Mb*(3ThiA) (Figure 3C, S2). Instead, the 3ThiA side chain is shifted away from the iron center, preventing a direct S-Fe interaction. As a result, C4 of the thiophene ring is the closest atom (3.63 Å) to the iron center. Because the steric bulk of the thiophene blocks binding of water or another ligand, Mb*(3ThiA) lacks a proximal ligand in its resting state.

Heme proteins have diagnostic UV/vis spectra. Both Mb* and Mb*(NMH) exhibit a Soret maximum at 407 nm in the Fe(III) state, which shifts to 432-434 nm in the reduced Fe(II) state.²⁵ Similarly, the Soret band for oxidized Mb*(5ThzA), Mb*(4ThzA) and Mb*(3ThiA) appears between 405 and 407 nm (Figure 4). The unexpectedly low 407/280 ratio observed for Mb*(4ThzA) and Mb*(3ThiA) suggests relatively poor cofactor binding, so these samples may contain some free heme in solution. Upon reduction of the new noncanonical variants with dithionite, the Soret band redshifts to 423-428 nm. Interestingly, reduced Mb*(5ThzA) has a substantially larger extinction coefficient than the other proteins.

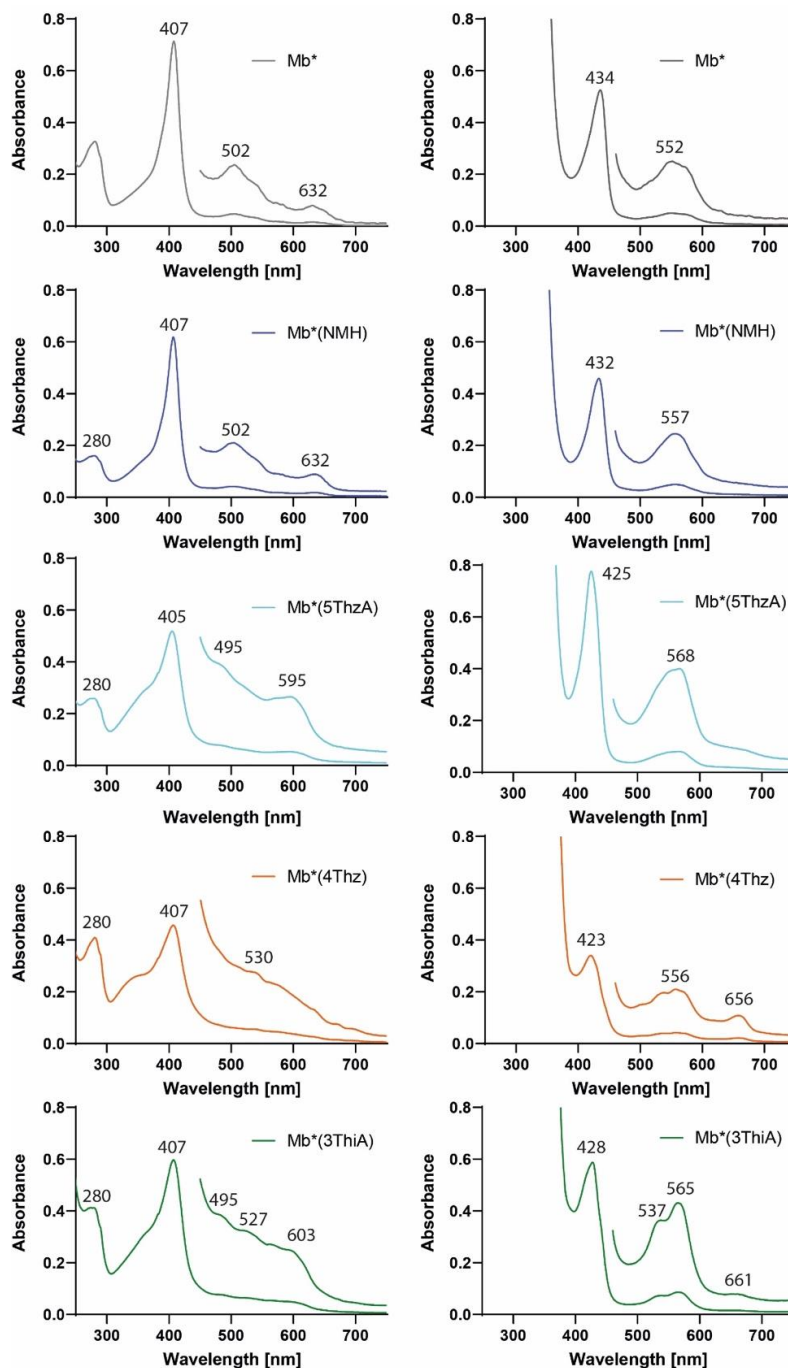


Figure 4. UV/vis spectra of the ferric (left) and ferrous (right) forms of myoglobin Mb* and its variants containing noncanonical histidine analogues in place of His93.

Changing the coordination sphere of metals can significantly alter their reduction potential.³² We therefore determined E^0 for the new Mb* variants using a photometric assay³³ (Figure S3). Mb*(NMH) ($E^0 = 77 \pm 6$ mV)²⁵ and Mb*(5ThzA) (118 ± 10 mV) have more positive reduction potentials than the

parent Mb* (30 ± 3 mV).²⁵ In both cases, this change can be largely attributed to disruption of the His93-Ser92 hydrogen bond and the associated increase in hydrophobic character of the heme proximal pocket.¹⁹ Because 5ThzA is less electron donating than NMH, as indicated by its lower pK_a , Mb*(5ThzA) has the most positive E° of the variants tested.³⁴ In contrast, the E° values for Mb*(3ThiA) (-83 ± 8 mV) and Mb*(4ThzA) (-91 ± 7 mV) are considerably lower, albeit still well above the reduction potential of cysteine-ligated myoglobin (-230 mV)³⁵ and free heme (-470 mV).³⁶ These low values likely reflect the inability of either 3ThiA or 4ThzA to coordinate to the heme iron.

The catalytic capabilities of the new myoglobin variants were tested in three model carbene transfer reactions: cyclopropanation, N-H insertion, and S-H insertion (Figure 5).

Previous work showed that Mb*(NMH) catalyzes efficient styrene cyclopropanation in the presence and absence of reductant, even under aerobic conditions.²⁵ Unlike Mb*, which promotes such transformations via an Fe(II)-carbenoid species, it appears to utilize an Fe(III) reaction manifold. The three new noncanonical proteins also catalyze styrene cyclopropanation in good yield under standard anaerobic conditions in the presence of the reducing agent sodium dithionite. Mb*(5ThzA) is comparable to Mb*(NMH) (both 71% yield) and superior to Mb* (61%), Mb*(4ThzA) (50%), Mb*(3ThiA) (49%) and free heme (29%) (Figure 5A). Like Mb*(NMH), Mb*(5ThzA) also catalyzes the reaction in the absence of reductant. The observation that the catalysts with the most positive reduction potentials perform better than Mb* and the other catalysts is consistent with computational studies indicating that more electron deficient iron-carbenoid complexes should speed up the concerted cyclopropanation reaction.³⁷ Notably, Mb*, Mb*(NMH) and Mb*(5ThzA) promote cyclopropanation with high enantioselectivity ($>99\%$ ee), whereas Mb*(3ThiA) (44% ee) and Mb*(4ThzA) (11% ee) exhibit only poor selectivity (Table S3). The loss in stereochemical fidelity in the latter cases likely reflects competition from heme either free in solution or non-specifically bound to solvent exposed residues.

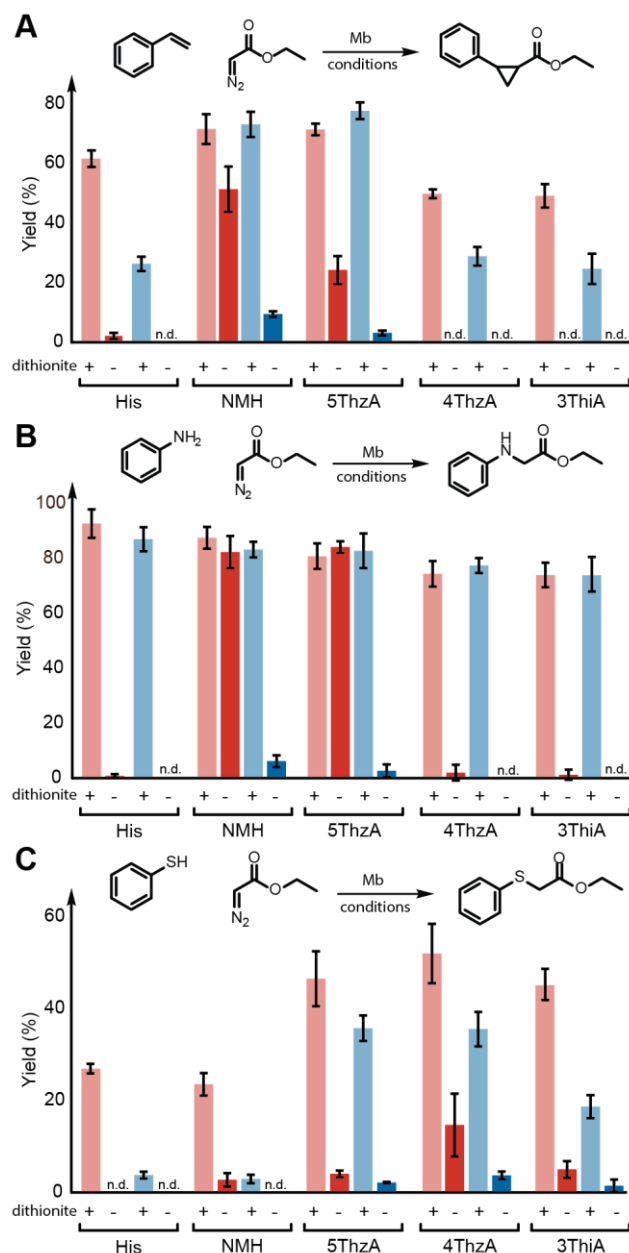


Figure 5. Reactivity of myoglobin variants in different carbene transfer reactions. (A) Cyclopropanation. (B) N-H insertion. (C) S-H insertion. The reactions were carried out both anaerobically (red bars) and in open reaction vessels (blue bars) in the presence (+) or absence (-) of 10mM dithionite. Reaction conditions: 50 mM potassium phosphate buffer pH=7.0, 10 μ M enzyme, 10 mM EDA, 10 mM aromatic compound, 1 h, r.t.

Heme-catalyzed N-H insertions are thought to proceed via nucleophilic attack on the iron porphyrin carbenoid rather than by a concerted mechanism.³⁸ We found that all four myoglobin variants efficiently catalyze the reaction of ethyl diazoacetate and aniline to give N-phenylglycine ethyl ester in

74-93% yield under anaerobic conditions in the presence of 10 mM dithionite (Figure 5B). Notably, Mb*(NMH) and Mb*(5ThzA) also yield comparable amounts of product (80%) in the absence of dithionite. This novel reactivity may reflect initial reduction of the protein by aniline or, as in the case of cyclopropanation,²⁵ a parallel Fe(III) reaction manifold.

The trends observed for the reaction between thiophenol and ethyl diazoacetate, an S-H insertion, differ from the other two carbene transfer reactions. In this case, Mb*(4ThzA) is the most effective catalyst, followed by Mb*(5ThzA) and Mb*(3ThiA) (Figure 5C). All three outperform Mb* and Mb*(NMH). The reactions with Mb*(4ThzA) and Mb*(5ThzA), and to a lesser extent Mb*(3ThiA), can be carried out in open reaction vessels. Mb*(4ThzA) is also active in the absence of reducing agent, albeit only weakly (15% yield). Computational studies suggest that hydrogen atom transfer (HAT) is the rate determining step for S-H insertion.³⁹ HAT is a radical process involving oxidation of the Fe(II) iron metal center (Scheme S1), which would explain why Mb variants with lower reduction potentials such as Mb*(3ThiA) and Mb*(4ThzA) outperform Mb* and Mb*(NMH). The reasons for the enhanced activity of Mb*(5ThzA) are less clear. One possibility is that recombination of the initially generated thiyl radical with the iron carbenoid intermediate after the rate-determining HAT step (Scheme S1) is faster in Mb*(5ThzA) than in other variants as it involves reduction of an Fe(III) to an Fe(II) species. Faster radical recombination may mitigate inactivating side reactions of the thiyl radical with the cofactor or the protein.

Our results illustrate how different ncAAs can modulate the structural and electronic properties of myoglobin-bound heme, giving rise to distinctive reactivity profiles. The excellent performance of 5ThzA, a new genetically encoded ncAA, in three mechanistically distinct carbene transfer reactions is particularly noteworthy. It is an excellent mimic of His93, minimally perturbing the protein binding pocket, and also affords the most positive redox potential of all the ligands tested, facilitating direct reaction with EDA in the absence of reductant. In contrast, 3ThiA and 4ThzA, which were previously proposed as isosteric histidine analogues, fail to coordinate the myoglobin heme iron via their heterocyclic sulfur atom and are hence poor His93 surrogates. Nevertheless, other metal centers may prove more

tolerant to the N→S substitution; 4ThzA could also be an attractive replacement for histidines that coordinate metal ions via N_δ rather than N_ε. In the future, genetically encoded ligand exchanges in heme and non-heme iron enzymes, copper binding enzymes, and many other metalloproteins can be expected to become as routine as altering the coordination sphere of synthetic transition metal complexes as a means of rationally tuning metal ion reactivity.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge.

Complete experimental procedures, Supplementary Figures and Tables

AUTHOR INFORMATION

Corresponding Author

*Donald Hilvert; Email: hilvert@org.chem.ethz.ch

Author Contributions

† M.P. and M.T. contributed equally

Notes

The authors declare no competing financial interests.

Data availability

Atomic coordinates of the Mb* variants and structure factors have been deposited in the RCSB Protein Data Bank (PDB) under accession codes 6Z4R (Mb*(3ThiA)) and 6Z4T (Mb*(5ThzA)). The experimental diffraction data are available at www.proteindiffraction.org.

ACKNOWLEDGMENTS

The authors are grateful to the Swiss National Science Foundation and the ETH Zurich for generous support of this work. We also thank Beat Blattmann and his team at the protein crystallization facility at the University of Zurich for setting up crystallization screens, the staff of the Swiss Light Source for supporting x-ray data collection as well as Fabian Glatz for assisting with SFC measurements. Plasmids pBK_PylRS, pBAR_PylT_Barnase(2,44AMBER) and pREP_PylT_CAT(111AMBER)_T7RP(8,114AMBER)_GFP, were generously provided by Jason Chin, MRC Laboratory of Molecular Biology.

REFERENCES

- (1) Poulos, T. L. Heme Enzyme Structure and Function. *Chem. Rev.* **2014**, *114*, 3919–3962.
- (2) Brandenberg, O. F.; Fasan, R.; Arnold, F. H. Exploiting and Engineering Hemoproteins for Abiological Carbene and Nitrene Transfer Reactions. *Curr. Opin. Biotechnol.* **2017**, *47*, 102–111.
- (3) Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. Olefin Cyclopropanation via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes. *Science*. **2013**, *339*, 307–310.
- (4) Bordeaux, M.; Tyagi, V.; Fasan, R. Highly Diastereoselective and Enantioselective Olefin Cyclopropanation Using Engineered Myoglobin-Based Catalysts. *Angew. Chem. Int. Ed.* **2015**, *54*, 1744–1748.
- (5) Tyagi, V.; Fasan, R. Myoglobin-Catalyzed Olefination of Aldehydes. *Angew. Chem. Int. Ed.* **2016**, *55*, 2512–2516.
- (6) Tyagi, V.; Bonn, R. B.; Fasan, R. Intermolecular Carbene S-H Insertion Catalysed by Engineered Myoglobin-Based Catalysts. *Chem. Sci.* **2015**, *6*, 2488–2494.
- (7) Sreenilayam, G.; Fasan, R. Myoglobin-Catalyzed Intermolecular Carbene N-H Insertion with Arylamine Substrates. *Chem. Commun.* **2015**, *51*, 1532–1534.

- (8) Kan, S. B. J.; Lewis, R. D.; Chen, K.; Arnold, F. H. Directed Evolution of Cytochrome c for Carbon–Silicon Bond Formation: Bringing Silicon to Life. *Science* **2016**, *354*, 1048–1051.
- (9) Huang, X.; Garcia-Borràs, M.; Miao, K.; Kan, S. B. J.; Zutshi, A.; Houk, K. N.; Arnold, F. H. A Biocatalytic Platform for Synthesis of Chiral α -Trifluoromethylated Organoborons. *ACS Cent. Sci.* **2019**, *5*, 270–276.
- (10) Kan, S. B. J.; Huang, X.; Gumulya, Y.; Chen, K.; Arnold, F. H. Genetically Programmed Chiral Organoborane Synthesis. *Nature* **2017**, *552*, 132–136.
- (11) Zhang, R. K.; Huang, X.; Arnold, F. H. Selective C–H Bond Functionalization with Engineered Heme Proteins: New Tools to Generate Complexity. *Curr. Opin. Chem. Biol.* **2019**, *49*, 67–75.
- (12) Cai, Y. B.; Yao, S. Y.; Hu, M.; Liu, X.; Zhang, J. L. Manganese Protoporphyrin IX Reconstituted Myoglobin Capable of Epoxidation of the C–C Bond with Oxone®. *Inorg. Chem. Front.* **2016**, *3*, 1236–1244.
- (13) Sreenilayam, G.; Moore, E. J.; Steck, V.; Fasan, R. Metal Substitution Modulates the Reactivity and Extends the Reaction Scope of Myoglobin Carbene Transfer Catalysts. *Adv. Synth. Catal.* **2017**, *359*, 2076–2089.
- (14) Oohora, K.; Miyazaki, Y.; Hayashi, T. Myoglobin Reconstituted with Ni Tetrahydrocorrin as a Methane Generating Model of Methyl-Coenzyme M Reductase. *Angew. Chem. Int. Ed.* **2019**, *58*, 13813–13817.
- (15) Key, H. M.; Dydio, P.; Clark, D. S.; Hartwig, J. F. Abiological Catalysis by Artificial Haem Proteins Containing Noble Metals in Place of Iron. *Nature* **2016**, *534*, 534–537.
- (16) Dydio, P.; Key, H. M.; Nazarenko, A.; Rha, J. Y.-E.; Seyedkazemi, V.; Clark, D. S.; Hartwig, J. F. An Artificial Metalloenzyme with the Kinetics of Native Enzymes. *Science* **2016**, *354*, 102–106.
- (17) Sreenilayam, G.; Moore, E. J.; Steck, V.; Fasan, R. Stereoselective Olefin Cyclopropanation under Aerobic Conditions with an Artificial Enzyme Incorporating an Iron-Chlorin E6 Cofactor.

- (18) Oohora, K.; Meichin, H.; Zhao, L.; Wolf, M. W.; Nakayama, A.; Hasegawa, J.; Lehnert, N.; Hayashi, T. Catalytic Cyclopropanation by Myoglobin Reconstituted with Iron Porphycene: Acceleration of Catalysis Due to Rapid Formation of the Carbene Species. *J. Am. Chem. Soc.* **2017**, *139*, 17265–17268.
- (19) Bhagi-Damodaran, A.; Petrik, I. D.; Marshall, N. M.; Robinson, H.; Lu, Y. Systematic Tuning of Heme Redox Potentials and Its Effects on O₂ Reduction Rates in a Designed Oxidase in Myoglobin. *J. Am. Chem. Soc.* **2014**, *136*, 11882–11885.
- (20) Chin, J. W. Expanding and Reprogramming the Genetic Code. *Nature* **2017**, *550*, 53–60.
- (21) Drienovská, I.; Roelfes, G. Expanding the Enzyme Universe with Genetically Encoded Unnatural Amino Acids. *Nat. Catal.* **2020**, *3*, 193–202.
- (22) Green, A. P.; Hayashi, T.; Mittl, P. R. E.; Hilvert, D. A Chemically Programmed Proximal Ligand Enhances the Catalytic Properties of a Heme Enzyme. *J. Am. Chem. Soc.* **2016**, *138*, 11344–11352.
- (23) Ortmayer, M.; Fisher, K.; Basran, J.; Wolde-Michael, E. M.; Heyes, D. J.; Levy, C.; Lovelock, S. L.; Anderson, J. L. R.; Raven, E. L.; Hay, S.; Rigby, S. E. J.; Green, A. P. Rewiring the “Push-Pull” Catalytic Machinery of a Heme Enzyme Using an Expanded Genetic Code. *ACS Catal.* **2020**, *10*, 2735–2746.
- (24) Pott, M.; Hayashi, T.; Mori, T.; Mittl, P. R. E.; Green, A. P.; Hilvert, D. A Noncanonical Proximal Heme Ligand Affords an Efficient Peroxidase in a Globin Fold. *J. Am. Chem. Soc.* **2018**, *140*, 1535–1543.
- (25) Hayashi, T.; Tinzl, M.; Mori, T.; Kregel, U.; Proppe, J.; Soetbeer, J.; Klose, D.; Jeschke, G.; Reiher, M.; Hilvert, D. Capture and Characterization of a Reactive Haem–Carbenoid Complex in an Artificial Metalloenzyme. *Nat. Catal.* **2018**, *1*, 578–584.
- (26) Xiao, H.; Peters, F. B.; Yang, P.; Reed, S.; Chittuluru, J. R.; Schultz, P. G. Genetic Incorporation

of Histidine Derivatives Using an Engineered Pyrrollysyl-TRNA Synthetase. *ACS Chem. Biol.* **2014**, *9*, 1092–1096.

- (27) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. An Enhanced System for Unnatural Amino Acid Mutagenesis in *E. Coli*. *J. Mol. Biol.* **2010**, *395*, 361–374.
- (28) Pédelacq, J. D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and Characterization of a Superfolder Green Fluorescent Protein. *Nat. Biotechnol.* **2006**, *24*, 79–88.
- (29) Sharma, V.; Wang, Y. S.; Liu, W. R. Probing the Catalytic Charge-Relay System in Alanine Racemase with Genetically Encoded Histidine Mimetics. *ACS Chem. Biol.* **2016**, *11*, 3305–3309.
- (30) Moore, E. J.; Fasan, R. Effect of Proximal Ligand Substitutions on the Carbene and Nitrene Transferase Activity of Myoglobin. *Tetrahedron* **2019**, *75*, 2357–2363.
- (31) Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. An Efficient System for the Evolution of Aminoacyl-TRNA Synthetase Specificity. *Nat. Biotechnol.* **2002**, *20*, 1044–1048.
- (32) Garner, D. K.; Vaughan, M. D.; Hwang, H. J.; Savelieff, M. G.; Berry, S. M.; Honek, J. F.; Lu, Y. Reduction Potential Tuning of the Blue Copper Center in Pseudomonas Aeruginosa Azurin by the Axial Methionine as Probed by Unnatural Amino Acids. *J. Am. Chem. Soc.* **2006**, *128*, 15608–15617.
- (33) Efimov, I.; Parkin, G.; Millett, E. S.; Glenday, J.; Chan, C. K.; Weedon, H.; Randhawa, H.; Basran, J.; Raven, E. L. A Simple Method for the Determination of Reduction Potentials in Heme Proteins. *FEBS Lett.* **2014**, *588*, 701–704.
- (34) Hosseinzadeh, P.; Lu, Y. Design and Fine-Tuning Redox Potentials of Metalloproteins Involved in Electron Transfer in Bioenergetics. *Biochim. Biophys. Acta* **2016**, *1857*, 557–581.
- (35) Adachi, S. ichi; Nagano, S.; Watanabe, Y.; Ishimori, K.; Morishima, I. Alteration of Human Myoglobin Proximal Histidine to Cysteine or Tyrosine by Site-Directed Mutagenesis: Characterization and Their Catalytic Activities. *Biochem. Biophys. Res. Commun.* **1991**, *180*, 138–144.

- (36) Sun, S. C.; Hsieh, B. C.; Chuang, M. C. Electropolymerised-Hemin-Catalysed Reduction and Analysis of Tartrazine and Sunset Yellow. *Electrochim. Acta* **2019**, *319*, 766–774.
- (37) Wei, Y.; Tinoco, A.; Steck, V.; Fasan, R.; Zhang, Y. Cyclopropanations via Heme Carbenes: Basic Mechanism and Effects of Carbene Substituent, Protein Axial Ligand, and Porphyrin Substitution. *J. Am. Chem. Soc.* **2017**, *140*, 1649–1662.
- (38) Sharon, D. A.; Mallick, D.; Wang, B.; Shaik, S. Computation Sheds Insight into Iron Porphyrin Carbenes' Electronic Structure, Formation, and N-H Insertion Reactivity. *J. Am. Chem. Soc.* **2016**, *138*, 9597–9610.
- (39) Chen, K.; Zhang, S. Q.; Brandenburg, O. F.; Hong, X.; Arnold, F. H. Alternate Heme Ligation Steers Activity and Selectivity in Engineered Cytochrome P450-Catalyzed Carbene-Transfer Reactions. *J. Am. Chem. Soc.* **2018**, *140*, 16402–16407.

Supporting Information

I. Materials and Methods

All chemicals and solvents were purchased from commercial suppliers and were used without further purification. H-His(3-Me)-OH (NMH) was obtained from Bachem (Bubendorf, Switzerland). All other ncAAs were either synthesized or bought from Ark Pharm (Illinois, USA). Product standards for GC-calibration curves were obtained from Enamine (rac-ethyl (1R, 2R)-2-phenylcyclopropane-1-carboxylate), Maybridge (ethyl 2-(phenylthio)acetate) and Acros (N-phenylglycine ethyl ester). LB medium and 2xYT were purchased from MP Biomedicals, LLC (California, USA) and Sigma-Aldrich (Missouri, USA). Hemin chloride and δ -aminolevulinic acid were purchased from Merck Millipore (Darmstadt, Germany) and Acros Organics (Geel, Belgium). Glucose oxidase was obtained from TCI (Tokyo, Japan). PEG1500, xanthine oxidase from bovine milk and bovine liver catalase were purchased from Sigma-Aldrich (Buchs, Switzerland). Oligonucleotides were synthesized by Microsynth AG (Balgach, Switzerland).

All UV-vis spectra were recorded on Perkin Elmer Spectrometers of the Lambda series (25, 35 or 45) at 20 °C. Depending on the enzyme concentrations, cuvettes with path lengths of either 1 cm, 2 mm or 1 mm were used. The scan speed was 960 nm/min.

GC measurements were performed on a Shimadzu instrument. Conditions were optimized to achieve baseline separation of all species.

Fast protein liquid chromatography (FPLC) was carried out on an NGC Quest 10 plus chromatography system (Bio-rad laboratories, California, USA) equipped with a multi-wavelength detector, BioFrac fraction collector and a Superdex 75 increase 10/300 GL (GE healthcare Life Sciences, Illinois, USA) size exclusion (SEC) column. Hydrophobic interaction chromatography (HIC) was performed on a HiPrep Butyl FF 16/10 column (GE healthcare Life Sciences, Illinois, USA).

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed on a PhastSystem electrophoresis apparatus using precast PhastGel homogeneous 20.0 (GE healthcare Life Sciences).

High-resolution mass spectra were recorded on an electrospray ionization quadrupole time-of-flight (ESI-Q-TOF) MaXis system (Bruker Daltonics, Bremen, Germany) coupled to an Agilent 1200 HPLC system (Agilent technology, California, USA).

II. General Procedures

(A) aaRS Library construction

Seven *M. barkeri* PylRS mutant libraries were created by cassette mutagenesis of four positions that line the amino acid binding pocket (Table S1). Libraries V-VII contained a fixed Y349F mutation that was reported to increase the rate of tRNA aminoacylation independent of the substrate in homologous PylRS of *M. mazei*.¹ For all other sites, full randomization to any amino acid was enabled by diversifying the genes encoding PylRS² or PylRS (Y349F) by overlap extension polymerase chain reaction using degenerate primers (Table S1, Table S4). Equimolar amounts of the respective fragments were generated by PCR and assembled to the full-length gene by amplification with the outermost primers o16 and o21 (Table S4). All PCR products were digested with NdeI and PstI restriction endonucleases according to the manufacturer's protocol (using a twofold excess of NdeI) and ligated with a threefold molar excess into the analogously digested acceptor vector-pBK²- using T4 ligase. Ligation mixtures were purified with the Clean and Concentrator-5 Kit (Zymo Research) and used to transform electrocompetent XL1-Blue cells. The transformants were recovered in 50 mL SOC medium at 37 °C for 1 h before cooling to 4 °C. Cells were subsequently pelleted by centrifugation at 2,137 g at 4 °C for 20 min and resuspended in 1.1 mL cold SOC medium. To determine library size, dilutions were plated on kanamycin-containing LB agar plates. The remaining culture was selected for antibiotic resistance by plating each library on seven LB agar plates supplemented with kanamycin and incubating them at 37 °C overnight. Resulting lawns of *E.coli* were scraped off the plates the next day and resuspended in LB medium. Suspensions of the same library were combined and plasmids isolated. The integrity of the library was confirmed by pool sequencing, as well as by sequencing of five individual clones from each library using primers o22 and o25. Purified plasmids were used to transform electrocompetent DH10β cells which had been previously transformed with the positive selection plasmid pREP_PylT_CAT(111AMBER)_T7RP(8,114AMBER)_GFP². The transformants were recovered in 50

mL SOC medium at 37 °C for 1 h before being cooled to 4 °C. Cells were subsequently pelleted by centrifugation at 2,137 g at 4 °C for 20 min and resuspended in 1.1 mL cold SOC medium. To determine library size, dilutions were plated on LB agar plates containing both kanamycin and tetracycline. The remaining culture was selected for antibiotic resistance by plating each library on seven LB agar plates supplemented with kanamycin and tetracycline and incubating them at 37 °C overnight. *E. coli* lawns were scraped off the plates the next day and suspended in LB medium containing kanamycin and tetracycline. Subsequently, suspensions of the same libraries were combined and the OD₆₀₀ measured to estimate the concentration of colony forming units (cfu). It was assumed that an OD₆₀₀ of 1 corresponds to 5.5x10⁸ cfu/mL. Aliquots of an estimated 2.1x10⁸ cfus (200-fold library coverage) were stored in LB medium supplemented with kanamycin, tetracycline and 30 % (v/v) glycerol at -80 °C until further use. An accurate number of cfus per aliquot was determined by thawing individual vials and plating dilutions onto LB agar plates containing kanamycin and tetracycline.

(B) PylRS selection

4ThzA: Aliquots of glycerol stocks of all seven PylRS libraries were thawed. Approximately 1.1x10⁷ cfu (10-fold library coverage) from each were pooled in 2 mL glycerol minimal media with leucine (GMML, 1 % (w/v) glycerol, 1x M9 salts, 1mM MgSO₄, 0.1 mM CaCl₂, 5 µM FeSO₄, 1x trace metals, 8.5 mM NaCl, 0.3 mM L-leucine) supplemented with 50 µg/mL kanamycin, 12.5 µg/mL tetracycline and 2 mM 4ThzA. As a control, an identical sample was prepared excluding the ncAA. Cells were grown in a shaking incubator at 230 rpm at 37 °C for 4 h before 500 µL were plated on GMML agar plates (GMML medium + 15 g/L agar-agar) containing 12.5 µg/mL tetracycline, 50 µg/mL kanamycin, 80 µg/mL chloramphenicol and 2 mM 4ThzA. In the control 4ThzA was not added. The resulting plates were incubated for 48 h at 37 °C in 100 % humidity, then for another 12 h at 30 °C and 100% humidity. After incubation, a growth advantage for cells grown on the plate containing 4ThzA compared to those grown on non-supplemented plates was evident. Cells were scraped off the plate, resuspended in cold LB

medium containing 50 µg/mL kanamycin and 12.5 µg/mL tetracycline and the plasmids were isolated. pBK and pREP plasmids were separated by 0.75 % (w/v) agarose gel electrophoresis. The pBK plasmids were extracted using the Gel DNA Recovery kit (Zymo Research) and subsequently purified using the Clean and Concentrator-5 kit. Purified plasmids were used to transform electrocompetent DH10β cells containing the negative selection plasmid pBAR_PylT_Barnase(2,44AMBER).² Transformants were rescued in 1 mL SOC medium for 1 h at 37 °C and shaken at 230 rpm before being plated on LB agar plates supplemented with 50 µg/mL kanamycin, 34 µg/mL chloramphenicol and 0.2 % (w/v) L-arabinose to induce the araBAD promoter controlling the barnase gene. Cells were incubated overnight at 37 °C and subsequently scraped off the plates and resuspended with cold LB medium containing kanamycin and chloramphenicol. Plasmids were isolated and pBK and pBAR plasmids were separated by 0.75 % (w/v) agarose gel electrophoresis. The pBK plasmids were extracted using the Gen DNA Recovery kit and subsequently purified using the Clean and Concentrator-5 kit. For the subsequent round of positive selection purified plasmids were used to transform electrocompetent DH10β cells carrying the positive selection plasmid. Transformed cells were rescued in 2 mL SOC medium and shaken at 230 rpm, 37 °C for 1 h before being chilled on ice and pelleted by centrifugation at 3500 g for 3 min at 4 °C. Cell pellets were resuspended in 2 mL GMML medium containing tetracycline and kanamycin. The resulting suspensions were split and each half was added to 1 mL GMML medium containing kanamycin, tetracycline, one half containing 2 mM 4ThzA (final concentration), the other half did not contain any ncAA. Cell cultures were incubated at 37 °C and 230rpm for 3.5 h before being plated on positive selection plates with and without 4ThzA. Subsequent rounds of positive and negative selection were carried out as detailed above. In total, three rounds of positive and two rounds of negative selection were conducted. For the last round of negative selection, the stringency was increased by growing cells on richer 2xYT agar plates. Individual clones from the last step of positive selection were further analyzed in a subsequent growth assay.

3ThiA: In analogy to the aforementioned procedure, PylRS libraries were subjected to a round of positive selection with a histidine analog, followed by negative selection. In the subsequent positive selection round, the PylRS variants were selected towards acceptance of 3ThiA as substrate for aminoacylation. Individual clones from the last step of positive selection were further analyzed by a growth assay.

(C) Growth assay

92 single clones from the last round of positive selection were grown in a 96-deepwell plate, each well containing 1 mL LB medium supplemented with 50 µg/mL kanamycin and 12.5 µg/mL tetracycline. Two wells that were not inoculated served as sterile controls and two additional wells were inoculated with DH10β cells harboring pBK_PylRS and the positive selection plasmid. The latter samples served as negative controls. Plates were incubated overnight at 37 °C in 100% humidity at 1,050 rpm. Cells were then washed twice with 1 mL PBS and subsequently diluted 1:50 in PBS. 5 µL of the dilutions were plated on LB agar plated supplemented with or without 2 mM ncAA, in addition to 50 µg/mL kanamycin, 12.5 µg/mL tetracycline and increasing concentrations of chloramphenicol (0, 50, 100, 150 µg/mL). Clones showing strongest growth on plates containing the ncAA and high chloramphenicol concentrations, but little growth on plates without ncAA were selected for sequencing using the primers o22 and o25. The assay was repeated with fresh transformants of variants of interest to confirm the initial results.

(D) Plasmid construction

pEVOL_PylT_PylRS: For efficient production of proteins containing a ncAA, the corresponding PylRS variants needed to be cloned into the pEVOL_PylT plasmid system.³ The plasmid has two copies of the gene one under the constitutive glnS and one under the inducible araBAD promoter. The first copy was inserted by amplifying the respective PylRS encoding gene from the pBK system using primers o26 and o27 (Table S4). The PCR products were digested with BglII and SalI restriction endonucleases according

to the manufacturer's protocol and ligated with a threefold molar excess into the analogously digested pEVOL_PylT acceptor vector using T4 ligase. Cloning success was confirmed by sequencing the resulting plasmid with o32 and pBADrev (Table S4). The second copy of the gene was inserted after it was digested with NdeI and PstI restriction endonucleases using a twofold excess of NdeI and ligated with a threefold molar excess into the analogously digested pEVOL_PylT acceptor vector obtained after the first insertion. Cloning success was confirmed by sequencing the resulting plasmid with primers o23 and o24 (Table S4).

sfGFP:

For construction of the pET29b(+)_sfGFP(133AMBER)-6His the Asp133 codon in the gene encoding sfGFP was mutated to an amber stop codon by overlap extension PCR. Two fragments corresponding to the N- and C- termini of the sfGFP gene were amplified from pACYC_6His-sfGFP⁴ using primer pairs o28/o30 and o29/o31 (Table S4). The two PCR products were purified and then combined in an equimolar ratio for the final PCR. The full-length fragment was assembled using the primer pair o28 and o29. The mutated fragment was digested by NdeI (twofold excess) and XhoI and ligated into a pET29b(+) vector digested with the same restriction enzymes, yielding pET29b(+)_sfGFP(133AMBER)-6His.

Myoglobin variants:

Construction of plasmids encoding sperm whale myoglobin variants (pET29b_Mb H64V V68A and pET29b_Mb H64V V68A H93TAG) was described previously.⁵

The DNA sequence of pET29b_Mb H64V V68A:

ATGATGGTTCTGTCTGAAGGTGAATGGCAGCTGGTTCTGCATGTTTGGGCTAAAGTTGAAG
CTGACGTCGCTGGTCATGGTCAGGACATCTTGATTCGACTGTTCAAATCTCATCCGGAAAC
TCTGGAAAAATTCGATCGTTTCAAACATCTGAAAACCTGAAGCTGAAATGAAAGCTTCTGAA
GATCTGAAAAAAGTGGGTGTTACCGCGTTAACTGCCCTAGGTGCTATCCTTAAGAAAAAAG
GGCATCATGAAGCTGAGCTCAAACCGCTTGCACAATCGCATGCTACTAAACATAAGATCCC
GATCAAATACCTGGAATTCATCTCTGAAGCGATCATCCATGTTCTGCATTCTAGACATCCA
GGTGACTTCGGTGCTGACGCTCAGGGTGCTATGAACAAAGCTCTGGAGCTGTTCCGTAAAG
ATATCGCTGCTAAGTACAAAGAACTGGGTACCAGGGTGGCTCGGGACATCATCACCATCA
CCATTGA

The amino acid sequence of Mb*:

MMVLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEMKASED
LKKVGVLTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGDFGA
DAQGAMNKALELFRKDIAAKYKELGYQGGSGHHHHHH

The DNA sequence of pET29b Mb H64V V68A H93TAG:

ATGATGGTTCTGTCTGAAGGTGAATGGCAGCTGGTTCTGCATGTTTGGGCTAAAGTTGAAG
CTGACGTCGCTGGTCATGGTCAGGACATCTTGATTCGACTGTTCAAATCTCATCCGGAAAC
TCTGGAAAAATTCGATCGTTTCAAACATCTGAAAACCTGAAGCTGAAATGAAAGCTTCTGAA
GATCTGAAAAAAGTGGGTGTTACCGCGTTAACTGCCCTAGGTGCTATCCTTAAGAAAAAAG
GGCATCATGAAGCTGAGCTCAAACCGCTTGCACAATCGTAGGCTACTAAACATAAGATCCC
GATCAAATACCTGGAATTCATCTCTGAAGCGATCATCCATGTTCTGCATTCTAGACATCCA
GGTGACTTCGGTGCTGACGCTCAGGGTGCTATGAACAAAGCTCTGGAGCTGTTCCGTAAAG

ATATCGCTGCTAAGTACAAAGAACTGGGTTACCAGGGTGGCTCGGGACATCATCACCATCA
CCATTGA

The amino acid sequence of Mb*(ncAA):

MMVLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEMKASED
LKKHGVTVLTALGAILKKKGHHEAELKPLAQS(ncAA)ATKHKIPIKYLEFISEAIIHVLHSRHPG
DFGADAQGAMNKALELFRKDIAAKYKELGYQGGSGHHHHHH

(E) Protein production and purification⁶. For expression of Mb*, pET29b_Mb H64V V68A was transformed into BL21(DE3) *E. coli* and the cells were plated on a LB agar plate containing 50 µg/ml kanamycin. A single colony of freshly transformed cells was cultured overnight in 3 ml of LB medium containing 50 µg/ml kanamycin. 1 ml of the culture was used to inoculate 100 mL of 2xYT medium supplemented with 1 mM δ -aminolevulinic acid (1M stock) and 50 µg/ml kanamycin. The culture was incubated for ~2 h at 37 °C at a shaking speed of 230 rpm. When the OD₆₀₀ of the culture reached 0.4 to 0.5, IPTG was added to a final concentration of 0.1 mM to induce expression of the Mb* protein. For expression of Mb*(ncAA), pET29b_Mb H64V V68A H93TAG and the correct pEVOL_PylT_PylRS vector were used to cotransform BL21(DE3)Gold *E. coli*. The cells were plated on an LB agar plate containing 50 µg/ml kanamycin and 34 µg/ml chloramphenicol. A single colony of freshly transformed cells was cultured overnight in 3 ml of LB medium containing 50 µg/ml kanamycin and 34 µg/ml chloramphenicol. 1 mL of the culture was used to inoculate 100 mL of 2xYT medium supplemented with 1 mM δ -aminolevulinic acid, 12 mM of ncAA (8 mM in case of 3ThiA), 50 µg/ml kanamycin and 34 µg/ml chloramphenicol. The culture was incubated for ~2.5 h at 37 °C at a shaking speed of 230 rpm. When the OD₆₀₀ of the culture reached 0.4 to 0.5, IPTG (1 M stock) and arabinose (10% stock in water) were added to a final concentration of 0.1 mM and 0.05%, respectively.

The induced cultures were incubated for ~24 h at 25 °C, and the cells were subsequently harvested by centrifugation at 8,000 g for 10 min. The pelleted bacterial cells were suspended in PBS buffer (Na₂HPO₄ 10 mM, KH₂PO₄ 1.8 mM, NaCl 137 mM, KCl 2.7 mM pH 7.4) and sonicated to disrupt the cells. To maximize heme occupancy, all proteins were reconstituted with hemin chloride. Briefly, the lysate solutions were rapidly mixed with 0.2 mM of hemin chloride (20 mM stock solution in 10 mM NaOH), and incubated at 4 °C for 10 min. The lysate was centrifuged at 11,500 g for 15 min and the supernatant was subjected to affinity chromatography using Ni-NTA Agarose (Qiagen, Helden, Germany). After two washes with buffer containing 50 mM Tris-HCl, 300 mM NaCl and 10 mM imidazole at pH 8.0, the proteins were eluted by 50 mM Tris-HCl, 300 mM NaCl at pH 8.0 containing 300 mM imidazole. The proteins were then further purified by size-exclusion (SEC) chromatography on Superdex 75 10/300 increase GL (GE healthcare Life Sciences). If necessary, an additional HIC step was added. Purified samples were analyzed by SDS-PAGE for purity and by HR-MS for correct incorporation of the ncAA (Table S2). If necessary, the protein solution was concentrated by Amicon Ultra-15 10 kDa cut-off centrifugal filter device (Merck Millipore, Massachusetts, USA). The concentrations of heme proteins were determined by UVvis absorbance measurements at the Soret band using extinction coefficients determined by the hemochrome assay (see below).

(F) GFP amber suppression assay

To assay the efficiency of amber suppression enabled by different PylRS variants E. coli BL21-Gold(DE3) cells were cotransformed with pET29b(+)_sfGFP(133AMBER)-6His and the respective pEVOL_PylT_PylRS plasmids. For each PylRS variant three single clones from the transformation were used to inoculate 1 mL LB medium supplemented with 34 µg/mL chloramphenicol and 50 µg/mL kanamycin in three separate wells of a 96-deepwell plate. Plates were incubated overnight at 37 °C in 100% humidity at 1,050 rpm. The overnight cultures were used to inoculate 2x 1mL LB medium containing 34 µg/mL chloramphenicol and 50 µg/mL kanamycin either supplemented with 2 mM ncAA

or, as negative control, unsupplemented antibiotics-containing LB. Cultures were incubated for approximately 2 h at 37 °C in 100% humidity and 1,050 rpm before gene expression was induced by the addition of IPTG (1mM final) and L-arabinose (0.02% final). Plates were further incubated overnight at 25 °C in 100% humidity at 1,050 rpm. On the next day, cells were washed twice with 1 mL PBS before being resuspended in the same volume. 200 µL of the suspension were transferred to a black 96-well microtiter plate, and cellular sfGFP fluorescence intensity was measured at 510 nm upon excitation at 485 nm. 200 µL of the same suspension were used to measure the OD₆₀₀ of the culture in a microtiter assay. sfGFP fluorescence was normalized to the OD₆₀₀ of the respective culture. Reported values are the average of three analyzed single clones and errors are given by the resulting standard deviation.

To determine the ideal concentration of ncAA for each PylRS variant during protein production, the GFP assay was repeated in 2xYT using increasing concentrations of the amino acid (0, 1, 2, 4, 8, 12 mM)

(G) Crystallization and structure determination

Myoglobin variants in PBS buffer were concentrated to approximately 45 mg/ml using an Amicon Ultra-15 10 kDa cut-off centrifugal filter device. The protein solutions were passed through a Zeba spin desalting column pre-equilibrated with pure water, and the final concentration was adjusted to 30 mg/ml. The proteins were crystallized by vapor diffusion in sitting drops at 20 °C using seeds of Mb*(NMH) crystals. Optimal crystallization conditions were found to be 0.1 M SPG buffer at pH 6.8-7.2 containing 28-32% (w/v) PEG1500. Crystals were cryo-protected by plunging into mother liquor supplemented with 20% (w/v) ethylene glycol prior to flash freezing in liquid nitrogen. Diffraction data were collected at 100 K using a wavelength of 1.000 Å at the Swiss Light Source (SLS). The structures were solved by molecular replacement using the Mb* structure (PDB ID: 6G5T) and program Phaser⁷ and were further refined using the program PHENIX⁸ Refinement statistics are summarized in Supplementary Table S5. Anomalous maps were also calculated using the program PHENIX.

(H) UV-vis measurements

All spectra were recorded anaerobically in 50 mM potassium phosphate buffer at pH 7.0. Enzyme and reactant concentrations were chosen as indicated in each experimental section.

(I) Hemochrome assay

In order to determine the heme occupancy of Mb* and Mb*NMH, a hemochrome assay was carried out following a slightly modified literature procedure.⁹ A solution of pyridine-NaOH was prepared by mixing 1 mL of pyridine with 0.4 mL 1 M NaOH and removing the excess aqueous phase. To a cuvette containing 700 μ L degassed protein solution in 50 mM potassium phosphate buffer pH 8.0, 250 μ L pyridine-NaOH and 50 μ L of dithionite solution (10 mg/mL in 50 mM potassium phosphate pH 8.0) were added anaerobically. The UV-vis spectra were recorded and the concentration was determined using an extinction coefficient, $\epsilon_{556-540}$, of 22,100 M⁻¹cm⁻¹.¹⁰ Determination of the concentrations was performed in triplicate.

To determine the extinction coefficients of the ferric myoglobin variants, the UV-vis spectrum of 1 mL protein solution in 50 mM potassium phosphate buffer pH 8.0 was recorded. By plotting the measured absorbance values of the Soret bands against protein concentration, the extinction coefficients, ϵ_{soret} , for Mb*(5Thz), Mb*(4ThzA) and Mb*(3ThiA) were determined to be 142,000 M⁻¹cm⁻¹, 188,000 M⁻¹cm⁻¹ and 126,000 M⁻¹cm⁻¹, respectively.

(J) Reduction potential determination

Reduction potentials were determined using a slightly modified literature procedure.¹¹ For reduction potential measurements, 50 mM potassium phosphate buffer pH 7.0 containing 10 mM glucose and 300

μ M xanthine was made anaerobic by bubbling nitrogen for 10 minutes through the buffer in a 1 mL cuvette sealed with a septum. Then toluidine blue or Nile blue, enzyme (2 μ M final concentration), 30 μ L catalase (stock 10 mg/mL, final 0.3 mg/mL) and 10 μ L of glucose oxidase (stock 5 mg/mL, final 50 μ g/mL) were added. To initiate the reaction, 2 μ L of a 25 μ M xanthine oxidase solution (50 nM final concentration) was added. The reactions were monitored by UV-vis for approximately 30 minutes. The reduction potential was determined by adding the standard reduction potential of toluidine blue ($E^\circ = +34$ mV) or Nile blue ($E^\circ = -116$ mV) to the value of the y-intercept obtained by fitting the data to the Nernst equation (Figure S3). When fitting the data, the first few points were ignored for Mb*(4ThzA) and Mb*(3ThiA) as it took awhile for the system to reach equilibrium. Due to the low reduction potential of the dye and the protein, the two variants and Nile blue were not completely reduced over the course of the assay. However, complete reduction could be achieved using dithionite as reductant.

(K) Carbene transfer reactions

Carbene transfer reactions were run on a 400 μ L scale in 2 mL sealed glass vials. The enzyme (final concentration 10 μ M) was added to degassed 50 mM potassium phosphate buffer pH 8.0 and the solution was made anaerobic by continuous headspace exchange with nitrogen gas for a minimum of 10 minutes. The reactions were initiated by adding dithionite (400 mM stock solution in 50 mM potassium phosphate buffer pH 8.0), substrate (styrene, aniline or thiophenol, 10 μ L of 400 mM stock in methanol) and EDA (10 μ L of 400 mM stock). For reactions under non-reducing conditions, dithionite was excluded. The mixtures were stirred for 1 h at 200 rpm using a magnetic stirring bar. For product quantification, 100 μ L of 5 mM 1,2-dimethoxybenzene in chloroform was added as standard and the product extracted with 900 μ L ethyl acetate. The product was quantified by GC-FID analysis using calibration curves prepared from authentic commercial standards. For analysis of ee, samples were analyzed by supercritical fluid chromatography (SFC) using a chiral column.

III.) Supplementary Figures

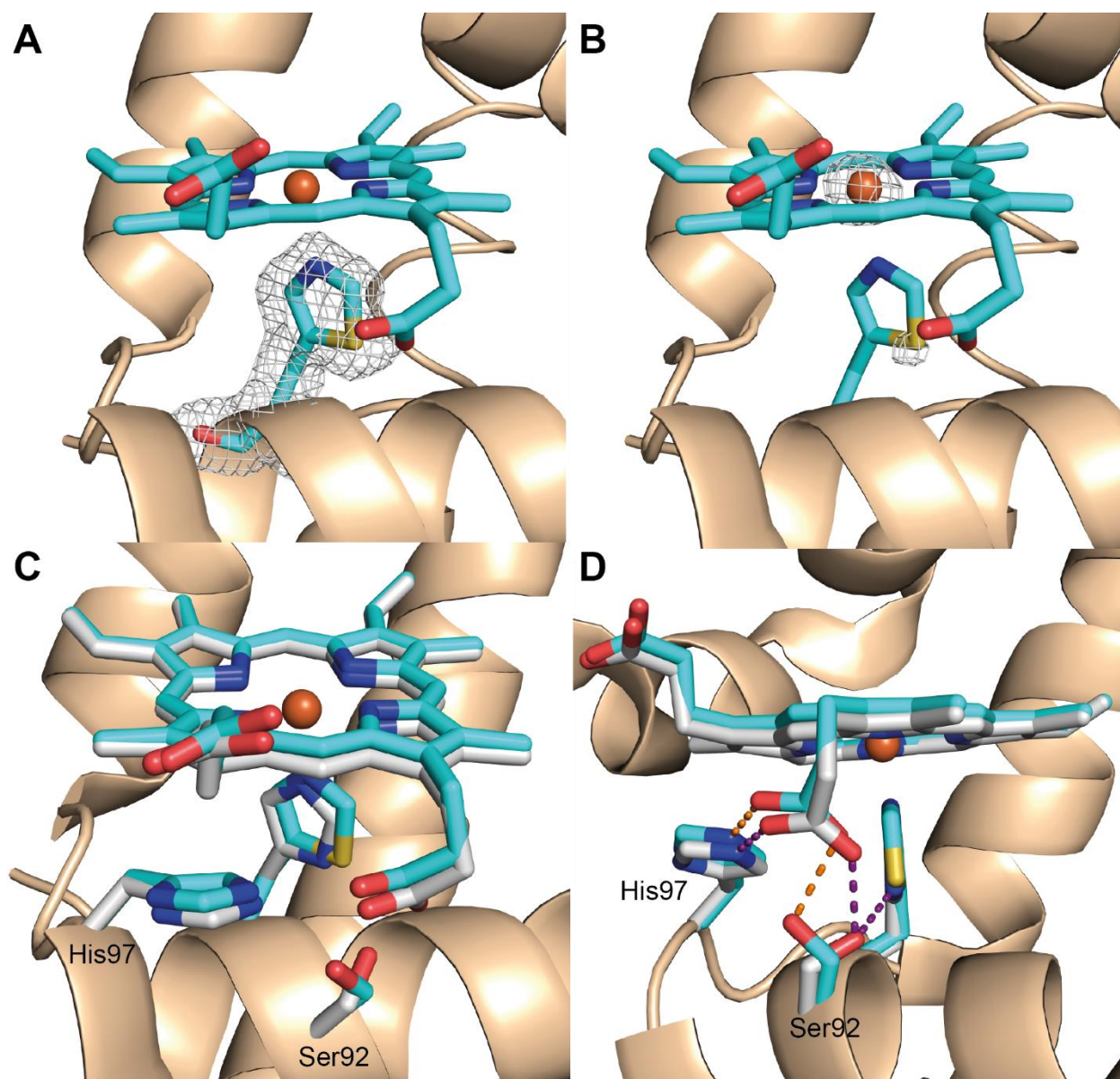


Figure S1. (A) 2Fo-Fc map of Mb*(5ThzA) contoured at 3σ showing the electron density for the ncAA in the proximal pocket. (B) Anomalous map of Mb*(5ThzA) contoured at 2σ . The anomalous signal of sulfur clearly shows the sidechain is oriented as indicated in the figure. (C) Overlay of Mb*(5ThzA) (cyan) and Mb* (gray). (D) The hydrogen bonding networks in the proximal pocket of Mb*(5ThzA) and (red) Mb* (purple). Although replacement of histidine with 5ThzA disrupts the hydrogen bond between

the proximal ligand and Ser92, the latter maintains its interaction with one of the heme propionic acid group via an altered rotameric conformation. Oxygen atoms are colored red, nitrogen atoms blue, sulfur atoms yellow; the iron ion is shown as an orange sphere.

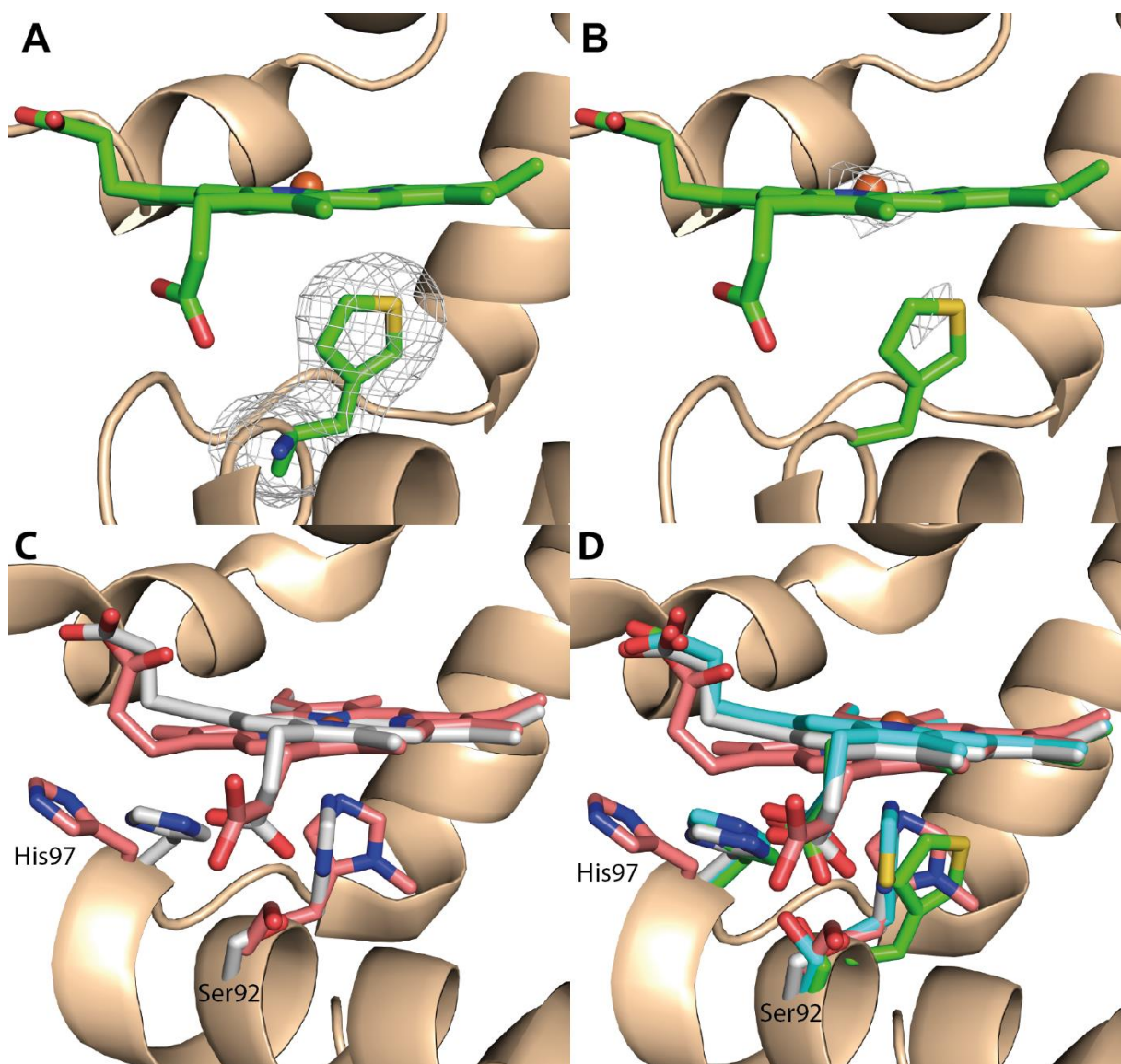


Figure S2. (A) 2Fo-Fc map of Mb*(3ThiA) contoured at 3σ showing the ncAA in the proximal pocket. (B) Anomalous map of Mb*(3ThiA) contoured at 2σ . While the position of the sulfur atom cannot be assigned unambiguously from the anomalous signal, the density in the 2Fo-Fc map suggests that it points away from the iron. (C) Overlay of Mb*(NMH) (pink) and Mb* (gray). Note the altered conformations of the proximal ligand (NMH vs. histidine), Ser92 and His97, which alter the tilt of the heme plane. (D) Overlay of Mb*(5ThzA) (cyan), Mb*(NMH) (pink), Mb*(3ThiA) (green) and Mb* (gray), shown from a different perspective than Figure 3D in the main text, highlighting the different

conformations of the noncanonical proximal ligands. Oxygen atoms are colored red, nitrogen atoms blue, and sulfur atoms yellow; the iron ion is shown as an orange sphere.

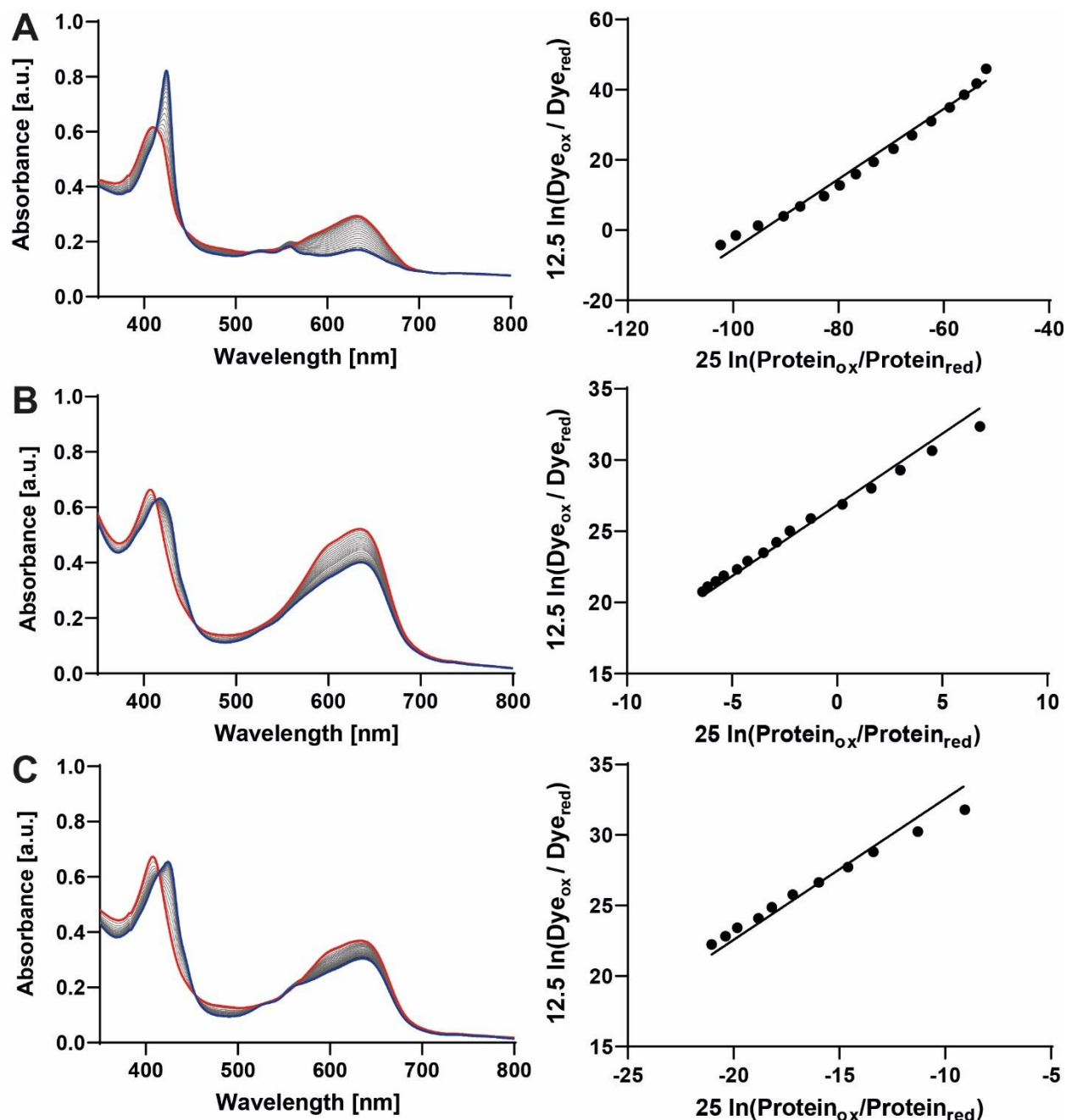
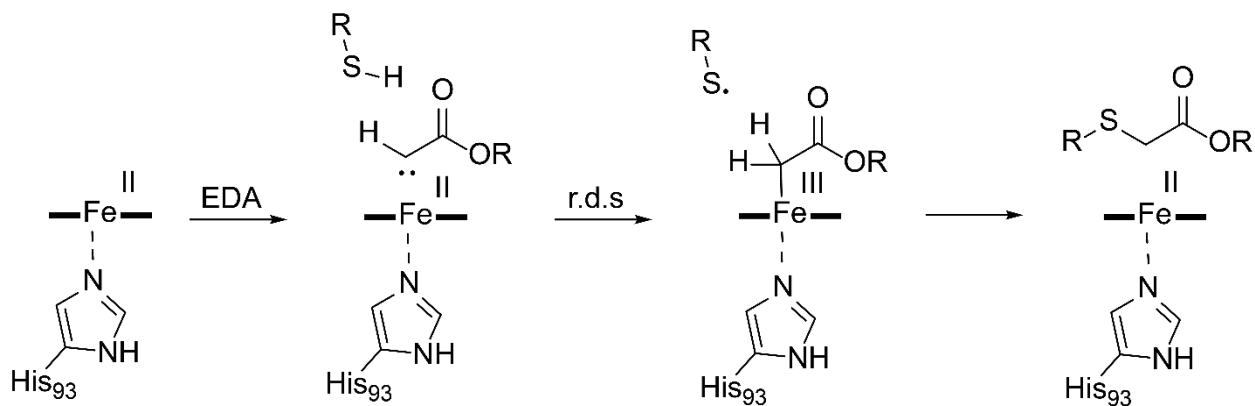


Figure S3: Reduction potential of myoglobin variants. Representative spectra (left) and Nernst plots (right) for reduction potential measurements of (A) Mb*(5ThzA), (B) Mb*(4ThzA), and (C) Mb*(3ThiA) are shown. Spectra of fully oxidized and reduced myoglobin variants are red and blue, respectively. For Mb*(4ThzA) and Mb*(3ThiA), the dye was not completely reduced.

IV. Supplementary Schemes



Scheme S1. Proposed mechanism of carbene S-H insertion reaction.¹² The carbene species is formed upon reaction of the Fe(II) iron with the carbene precursor EDA. Next, hydrogen atom transfer (HAT) from the thiol to the carbene species occurs, which is the rate determining step of the reaction. As a last step, the thiyl radical recombines with the carbenoid intermediate, regenerating the Fe(II) resting state.

IV. Supplementary Tables

Table S1: Focused libraries for parallel randomization of four residues in PylRS and PylRS(Y349F)

Library	Fixed	Position 1	Position 2	Position 3	Position 4	Primers
I		L270X	Y271X	L274X	C313X	o16+o17 o18+o19 o20+o21
II		A267X	Y271X	N311X	C313X	o16+o1 o2+o3 o4+o21
III		Y271X	L274X	C313X	Y349X	o16+o5 o6+o19 o20+o7 o8+o21
IV		A267X	N311X	V367X	G384X	o16+o1 o9+o3 o10+o11 o12+o13 o14+o21
V	Y349F	L270X	Y271X	L274X	C313X	o16+o17 o18+o19 o20+o21
VI	Y349F	A267X	Y271X	N311X	C313X	o16+o1 o2+o3 o4+o21
VII	Y349F	A267X	N311X	V367X	G384X	o16+o1 o9+o3 o10+o11 o12+o13 o14+o21

Table S2: HR-MS of myoglobin variants.

variant	Theoretical mass [Da]	Mass found [Da]
Mb*	18289.06	18288.69 ⁵
Mb*(NMH)	18303.07	18302.69 ⁵
Mb*(5ThzA)	18305.65	18305.44
Mb*(4ThzA)	18305.65	18305.21
Mb*(3ThiA)	18304.66	18304.48

Table S3: Diastereoselectivity and enantioselectivity of cyclopropanation reactions.

variant	Diastereomeric Ratio	Enantiomeric Excess [%]
Mb*	>99:1	>99
Mb*(NMH)	>99:1	>99
Mb*(5ThzA)	>99:1	>99
Mb*(4ThzA)	91:9	11
Mb*(3ThiA)	96:4	44

Table S4: List of Primers

Name	Sequence 5'→3'
o1	CAGCATCGGACGAGG
o2	CCTGCGTCCGATGCTGNNKCCGACGCTGNNKAATTATCTGCGTAAAC
o3	CACCATGGTGAATTCTTCCAGG
o4	CCTGGAAGAATTCACCATGGTGNNKTTTCNNKCAAATGGGTAGCGGCTGC A
o5	CAGCGTCGGGGCCAG
o6	CTGGCCCCGACGCTGNNKAACTACNNKCGCAAACCTGGATCGT
o7	AACCATGCAGCTATCGCCC
o8	GGGCGATAGCTGCATGGTTNNKGGTGATACCCTGGAT
o9	CCTGCGTCCGATGCTGNNKCCGACGCTGTATAACTAT
o10	CCTGGAAGAATTCACCATGGTGNNKTTCTGCCAAATGGGC
o11	AACCGCGCTGCTCAGTT
o12	AACTGAGCAGCGCGGTTNNKGGCCCGGTTAGCCTGGA
o13	GATCCACGGTTTATCAATGCC
o14	GGCATTGATAAACCGTGGATCNNKGCCGGTTTTGGCCTGGAA
o15	AAAATGCTCGAGTCCCGAGCCACC
o16	GAGGAATCCCATATGATGGATAAA
o17	GGTCGGGGCCAGCAT
o18*	ATGCTGGCCCCGACCNNKNNKAATTATNNKCGT... (end unknown)
o19	AAAGTTAACCATGGTGAATTCTTC
o20*	GAAGAATTCACCATGGTTAACTTTNNKCAGATGGGCAG... (end unknown)
o21	CGTTTGAAACTGCAGTTACAGGTT
o22	CGGCGGCTTTGTTGAATAA
o23	AAACTCAATATATTGCAGAG
o24	TCCCGCTTTAATATCATAC
o25	TATCCGGCCTACAAAAGCA
o26	GAGGAATCCAGATCTATGATGGATAAAAAACCGCTGG
o27	CGTTTGAAAGTCGACTTACAGGTTTCGTGCTAATGCCG
o28	GATATACATATGAGCAAAGGAGAAGAACTTTTC
o29	GGTGGTGCTCGAGGGATCCTTTGTAGAGTTCATCCATGCC
o30	GTCCGAGAATGTTTCCCTATTCTTTAAAATCAATAC
o31	GTATTGATTTTAAAGAATAGGGAAACATTCTCGGAC
o32	GGCGTCACACTTTGCTAT
pBADrev	GATTTAATCTGTATCAGG

*Primers were obtained from an old lab stock and the sequence of the last nucleotides could not be verified anymore

Table S5: Data collection and refinement statistics

	Mb*(5ThzA)	Mb*(3ThiA)
PDB ID	6Z4T	6Z4R
Number of crystal	1	1
Data collection		
X-ray wave length (Å)	1.000	1.000
Rotation range (°)	0 -180	0 -180
$\Delta \phi$ (°)	0.1	0.1
Exposure time (sec)	0.05	0.05
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a, b, c (Å)	40.0, 47.3, 76.9	40.1, 47.3, 76.2
Resolution (Å)	40.3-1.23	40.2-1.96
	(1.27-1.23)*	(2.03-1.96)
R_{merge}	5.3 (120)	22.4 (234)
$I / \sigma I$	16.6 (0.68)	11.1 (1.31)
CC (1/2) (%)	100 (28.5)	99.7 (60.3)
Completeness (%)	96.8 (75.6)	99.2 (98.0)
Redundancy	5.7 (2.2)	12.5 (12.5)
Refinement		
Resolution (Å)	40.3-1.23	40.2-1.96
No. reflections	41767	10834
$R_{\text{work}} / R_{\text{free}}$	15.1/18.8	21.9/26.7
No. atoms		
Protein	1280	1216
Ligand/ion	45	43
Water	193	63
B -factors (Å ²)		
Protein	18.4	33.3

Ligand	16.5	29.2
Water	39.4	37.2
R.m.s. deviations		
Bond lengths (Å)	0.006	0.014
Bond angles (°)	0.88	1.12
Ramachandran plot (%)		
Favored	98.7	97.3
Allowed	1.3	2.7
Outliers	0	0
No. TLS groups	-	5

V. SI References

- (1) Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. Multistep Engineering of Pyrrolysyl-TRNA Synthetase to Genetically Encode N ϵ -(o-Azidobenzoyloxycarbonyl) Lysine for Site-Specific Protein Modification. *Chem. Biol.* **2008**, *15*, 1187–1197.
- (2) Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. Genetically Encoding N ϵ -Acetyllysine in Recombinant Proteins. *Nat. Chem. Biol.* **2008**, *4*, 232–234.
- (3) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. An Enhanced System for Unnatural Amino Acid Mutagenesis in E. Coli. *J. Mol. Biol.* **2010**, *395*, 361–374.
- (4) Azuma, Y.; Bader, D. L. V.; Hilvert, D. Substrate Sorting by a Supercharged Nanoreactor. *J. Am. Chem. Soc.* **2018**, *140*, 860–863.
- (5) Hayashi, T.; Tinzl, M.; Mori, T.; Krengel, U.; Proppe, J.; Soetbeer, J.; Klose, D.; Jeschke, G.; Reiher, M.; Hilvert, D. Capture and Characterization of a Reactive Haem–Carbenoid Complex in an Artificial Metalloenzyme. *Nat. Catal.* **2018**, *1*, 578–584.
- (6) Green, A. P.; Hayashi, T.; Mittl, P. R. E.; Hilvert, D. A Chemically Programmed Proximal Ligand Enhances the Catalytic Properties of a Heme Enzyme. *J. Am. Chem. Soc.* **2016**, *138*, 11344–11352.
- (7) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40*, 658–674.
- (8) Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. Towards Automated Crystallographic Structure Refinement with Phenix.Refine. *Acta Crystallogr. Sect. D*

- (9) Berry, E. A.; Trumpower, B. L. Simultaneous Determination of Hemes a, b, and c from Pyridine Hemochrome Spectra. *Anal. Biochem.* **1987**, 161, 1–15.
- (10) Williams, J. N. A Method for the Simultaneous Quantitative Estimation of Cytochromes a, b, and c in Mitochondria. *Arch. Biochem. Biophys.* **1964**, 107, 537–543.
- (11) Efimov, I.; Parkin, G.; Millett, E. S.; Glenday, J.; Chan, C. K.; Weedon, H.; Randhawa, H.; Basran, J.; Raven, E. L. A Simple Method for the Determination of Reduction Potentials in Heme Proteins. *FEBS Lett.* **2014**, 588, 701–704.
- (12) Chen, K.; Zhang, S. Q.; Brandenberg, O. F.; Hong, X.; Arnold, F. H. Alternate Heme Ligation Steers Activity and Selectivity in Engineered Cytochrome P450-Catalyzed Carbene-Transfer Reactions. *J. Am. Chem. Soc.* **2018**, 140, 16402–16407.