Accepted Manuscript

Effect of proximal ligand substitutions on the carbene and nitrene transferase activity of myoglobin

Eric J. Moore, Rudi Fasan

Please cite this article as: Moore EJ, Fasan R, Effect of proximal ligand substitutions on the carbene and nitrene transferase activity of myoglobin, *Tetrahedron* (2019), doi: https://doi.org/10.1016/j.tet.2019.03.009.

Effect of proximal ligand substitutions on the carbene and nitrene transferase activity of myoglobin

Eric J. Moore[a] and Rudi Fasan[*,a]

Department of Chemistry, University of Rochester, Rochester, NY 14627, United States. Email: rfasan@ur.rochester.edu

_____

Abstract: Engineered myoglobins were recently shown to be effective catalysts for abiological carbene and nitrene transfer reactions. Here, we investigated the impact of substituting the conserved heme-coordinating histidine residue with both proteinogenic (Cys, Ser, Tyr, Asp) and non-proteinogenic Lewis basic amino acids (3-(3'-pyridyl)-alanine, *p*-aminophenylalanine, and β-(3-thienyl)-alanine), on the reactivity of this metalloprotein toward these abiotic transformations. These studies showed that mutation of the proximal histidine residue with both natural and non-natural amino acids result in stable myoglobin variants that can function as both carbene and nitrene transferases. In addition, substitution of the proximal histidine with an aspartate residue led to a myoglobin-based catalyst capable of promoting stereoselective olefin cyclopropanation under nonreducing conditions. Overall, these studies demonstrate that proximal ligand substitution provides a promising strategy to tune the reactivity of myoglobin-based carbene and nitrene transfer catalysts and provide a first, proof-of-principle demonstration of the viability of pyridine-, thiophene-, and aniline-based unnatural amino acids for metalloprotein engineering.

Heme-containing proteins have recently emerged as promising biocatalysts for promoting an array of abiological carbene and nitrene transfer reactions useful for the construction of new carbon-carbon[1-14] and carbon-nitrogen bonds,[15-20] respectively. In particular, our group has previously reported the versatility and synthetic value of engineered myoglobins for catalyzing stereoselective olefin cyclopropanation reactions in the presence of diazo-based carbene donors[3,4,21,11,22] and C—H amination reactions in the presence of azide-based nitrene donors,[17] among other abiological reactions.[23] The reactivity of myoglobin in these transformations derives from its hemin cofactor (iron-protoporphyrin IX), whose iron center is coordinated at the axial position by a conserved 'proximal' histidine residue (His93 in sperm whale myoglobin).[24] The latter is implicated in mediating binding of the cofactor to the protein as well as in modulating the oxygen binding properties of this metalloprotein.[24]

Modification of the metalloporphyrin cofactor environment in myoglobin (Mb) represents a promising approach toward tuning or enhancing the reactivity of myoglobin in the context of nitrene and carbene transfer reactions. In previous studies, we and others have demonstrated that substitution of the iron center with other metals (e.g., Mn, Co, Rh, Ru, Ir) led to Mb variants with altered nitrene[17] or carbene transferase activity,[25-27] also enabling reactions that cannot be achieved with the iron-containing counterparts.[25,26] In another study, the combination of metal substitution with a proximal ligand mutation (serine-ligated Co-Mb) enabled the development of a Mb-based catalyst capable of favoring olefin cyclopropanation over the Y—H carbene insertion reactions.[11] Furthermore, replacement of the protoporphyrin IX ligand with other porphyrin derivatives has yielded Mb-based cyclopropanation catalysts with distinct catalytic properties compared to wild-type Mb.[28,9] In the context of P450 enzymes, substitution of the axial cysteine ligand coordinating heme with serine was reported to favor

nitrene transfer reactions such as aziridination and intermolecular C—H amination[15,18,20] or favor cyclopropanation over S—H carbene insertion.[29] Substitution of the cysteine axial ligand with selenocysteine also affected the reactivity of P450s as monooxygenases.[30] More recently, substitution of the proximal histidine ligand in Mb with N^{δ}-methylhistidine (NMH) was found to increase its peroxidase activity[31] and the same modification in an engineered Mb-based carbene transferase[3] resulted in increased cyclopropanation activity under non-reducing conditions.[32] Despite this progress, the potential value of altering the iron first coordination sphere residue in engineered myglobins toward modulating their non-native carbene and nitrene transferase reactivity have remained largely undefined. In addition, the feasibility of exploiting non-canonical amino acids for this purpose beyond N-methyl-histidine has not been established. Here, we addressed these questions through the preparation of a panel of proximal ligand Mb variants (**Figure 1**) and their characterization in terms of their stability and spectroscopic properties and reactivity in representative carbene and nitrene transfer reactions.
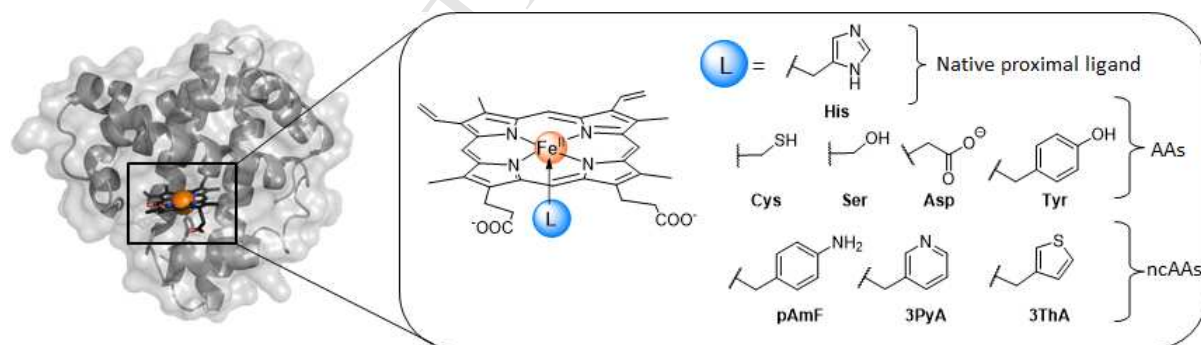

Figure 1. Overview of myoglobin structure and chemical structures of the natural and non-canonical amino acid residues introduced at the heme proximal axial ligand position (His93).

3

For these studies, a series of proximal ligand variants were prepared based on Mb(H64V,V68A), an engineered active site variant of sperm whale myoglobin that exhibits high activity and stereoselectivity in olefin cyclopropanation with ethyl α-diazoacetate (EDA)[3,4] and other acceptor-only diazo compounds,[21,22] along with intramolecular C—H amination activity in the presence of sulfonyl azides.[17] A such, this variant was deemed useful for assessing the impact of altering the heme coordination environment on both Mb-mediated carbene and nitrene transfer activity and on subtle properties such as diastereo- and enantioselectivity. Accordingly, a first set of Mb(H64V,V68A)-based variants were generated upon substituting the proximal histidine (His93) for another proteinogenic nucleophilic amino acid capable of metal coordination, namely, Ser, Cys, Tyr, and Asp. Cys and Tyr are involved in heme ligation in P450s[33] and catalase,[34] respectively, whereas Asp is implicated in metal coordination in various metalloenzymes.[35,36] To further extend the range of nucleophilic groups to be introduced at the proximal site of the hemoprotein beyond that allowed by proteinogenic amino acids, we also targeted the preparation of Mb(H64V,V68A) variants in which His93 is replaced with the non-canonical amino acid (ncAA) *p*-aminophenylalanine (*p*AmF),[37] 3-(3'-pyridyl)-alanine (3PyA),[38] and β-(3-thienyl)-alanine (3ThA)[38] (**Figure 1**). While sharing a similar size with histidine, these amino acids incorporate a side-chain group with significantly different basicity/nucleophilicty, as supported by the different pKa and/or Lewis basicity of aniline, pyridine and thiophene compared to imidazole (or N-methyl-imidazole).[39] To generate these variants, a Mb(H64V,V68A) construct containing an amber stop codon (TAG) at position 93 was expressed in the presence of the desired ncAA and orthogonal aminoacyl-tRNA synthetases previously developed for the genetic incorporation of these ncAAs[37,38] via the amber stop codon suppression technology.[40,41] Gratifyingly, all of these ncAA-containing Mb variants could be successfully

produced and isolated from *E. coli*. In addition, selective incorporation of the corresponding ncAA at the desired position (93) was confirmed by LC-MS and MS/MS analysis of peptide fragments obtained after proteolytic digestion of these proteins (**Figure S1**).

Analysis of the electronic absorption spectrum of the purified Mb(H64V,V68A) variants showed the presence of a prominent Soret band, indicating proper incorporation of the hemin cofactor in each case, including the ncAA-containing variants. At the same time, side-by-side comparison of the UV-vis spectra of the proximal ligand-substituted Mb variant in their ferric, ferrous and CO-bound forms show distinguishable features compared to the histidine-ligated Mb(H64V,V68A) (**Figure 2, Table S1**). Ferric Mb(H64V,V68A) shows a Soret band with λ_{max} at 409 nm and Q bands at 530 and 636 nm, characteristic of a hexacoordinated high-spin iron species.[24] Reduction to the ferrous state with sodium dithionite results in a red shift of the Soret band to 432 nm, which undergoes a blue-shift to 424 nm upon complexation with CO. As summarized in **Table S1**, differences in the absorption maxima for the Soret band and/or Q bands are observable for the proximal ligand-substituted variants when compared to each other and to Mb(H64V,V68A), for one or more forms of the hemoprotein (i.e., ferric, ferrous, CO-bound form). For example, the Soret bands of Mb(H64V,V68A,H93(3PylA)) have λ_{max} at 412 (ferric)/419(ferrous)/418(CO-bound) compared to 410/424/420 for Mb(H64V,V68A,H93pAmF), 406/421/420 for Mb(H64V,V68A,H93S), and 409/432/424 for Mb(H64V,V68A) (**Table S1**). Moreover, the spectroscopic features of these proximal ligand variants are distinct from those of Mb(H64V,V68A) variants in which His93 was substituted with a non-nucleophilic Phe residue (405/417/415)[25] or N-methyl-His, the latter being reported to exhibit a UV-vis spectrum indistinguishable from that of Mb(H64V,V68A).[32] Altogether, these spectroscopic analyses

5

indicated that the heme cofactor environment in Mb(H64V,V68A) is altered upon mutation of the proximal histidine residue with both the natural and unnatural amino acids.
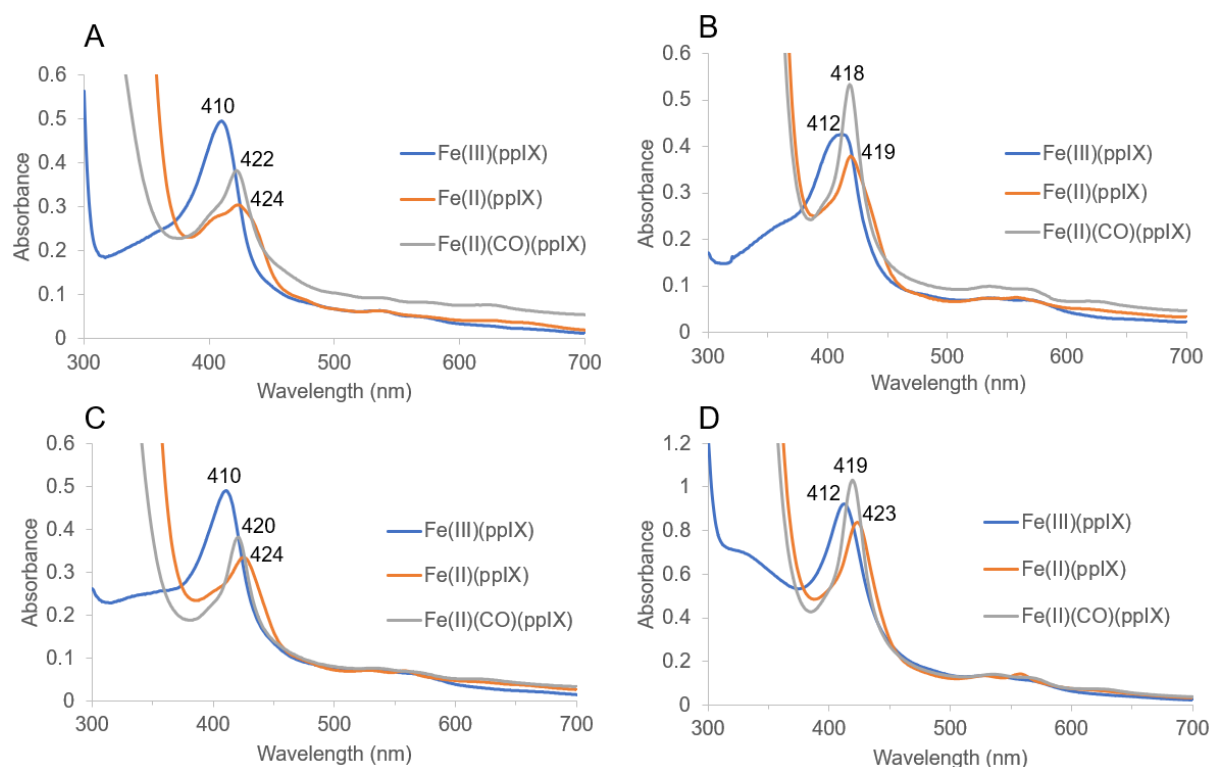

Figure 2: Electronic absorption spectra for representative proximal ligand Mb variants. (A) Mb(H64V,V68A,H93D), (B) Mb(H64V,V68A,H93(3PylA)), (C) Mb(H64V,V68A,H93(pAmF)) and (D) Mb(H64V,V68A,H93(3ThA)). Spectra correspond to the proteins in their ferric (blue line), ferrous (orange line) (solid line), and CO-bound form (grey line). The λ_{max} for the Soret band are indicated. See Table S1 for additional spectroscopic data on these and the other Mb variants.

To examine the impact of these mutations on protein stability, the Mb variants were further characterized with respect to their stability against thermal denaturation. Consistent with

previous observations,[42] Mb(H64V,V68A) was determined to have an apparent T_m of 66.0 °C using variable temperature circular dichroism spectroscopy (**Figure 3**, entry 1). In comparison, the proximal ligand variants were found to show a general reduction in thermostability, the majority exhibiting an approximately 10-12 °C reduction in T_m value. These results are not entirely surprising considering the structural role of the proximal histidine residue in Mb.[24] The largest change in thermostability was observed with the H93D substitution (T_m: 47.9 vs. 66.0 °C), whereas smallest decrease in thermostability was observed for the H93Y mutation (T_m: 60.9 vs. 66.0 °C). Notably, the mutations involving the non-canonical amino acids pAmF, 3PylA, and 3ThA were well tolerated by the myoglobin scaffold, resulting in folded proteins (Figure 1) with melting temperatures of 55-57 °C (**Figure 3**). Thus, despite the reduction in T_m compared to the parent protein, the stability of these and the other proximal ligand Mb variants remain comparable or higher than that of heme-dependent enzymes from mesophilic organisms ($T_m = 45 - 55$ °C).[43,44]
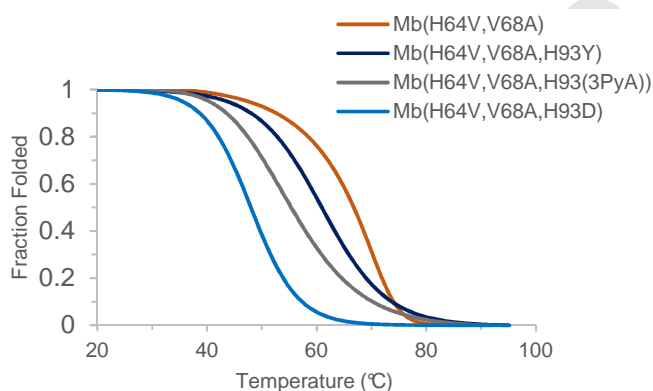
Entry	Variant	Apparent T_m (°C)
1	Mb(H64V,V68A)	66.0 ± 1.0
2	Mb(H64V,V68A,H93(3PyA))	55.3 ± 0.8
3	Mb(H64V,V68A,H93(pAmF))	55.6 ± 0.4
4	Mb(H64V,V68A,H93(3ThA))	57.3 ± 1.5
5	Mb(H64V,V68A,H93C)	58.5 ± 0.5
6	Mb(H64V,V68A,H93Y)	60.9 ± 0.8
7	Mb(H64V,V68A,H93S)	55.1 ± 0.6
8	Mb(H64V,V68A,H93D)	47.9 ± 2.0

Figure 3. Thermal denaturation curves for Mb(H64V,V68A) and selected proximal ligand mutants as determined by variable temperature circular dichroism spectroscopy. The Table reports the apparent T_m values for Mb(H64V,V68A) and its variants as determined by fitting raw

CD data to a sigmoidal equation via nonlinear regression analysis. Mean values and standard errors are derived from experiments performed in triplicate.

Encouraged by these results, we examined the carbene transfer activity of the proximal ligand Mb variants using the cyclopropanation of styrene (**1**) in the presence of ethyl α-diazoacetate (**2**) as a model reaction. Under 'standard' (i.e., anaerobic and reducing) conditions,[3] Mb(H64V,V68A) catalyzes the nearly quantitative conversion of the olefin to the cyclopropanation product **3** (97%, 1,940 TON) with high enantio- and diasteroselectivity (99% *de* and *ee*; **Table 1**, Entry 1). Under these conditions, nearly all of the proximal ligand-substituted variants, including those incorporating the non-canonical amino acids pAmF, 3PylA and 3ThA, were found to be viable cyclopropanation catalysts, supporting between 900 and 1,100 catalytic turnovers (TON) in this reaction (**Table 1**). As an exception, the serine-containing variant Mb(H64V,V68A,H93S) shows a significant reduction in cyclopropanation activity compared to the parent protein (25% vs. 98% conv.). Among the proximal ligand variants, Mb(H64V,V68A,H93D) emerged as the most efficient and selective cyclopropanation biocatalyst, producing **3** with nearly identical turnovers and diastereoselectivity (1,930 TON, 95% *de*) and only moderately decreased enantioselectivity (75% *ee*) compared to Mb(H64V,V68A) (**Table 1**, Entry 8).

We next performed the cyclopropanation reactions in the presence of oxygen ('aerobic/reducing' conditions) and in the absence of reductant ('anaerobic/nonreducing conditions'), the latter being beneficial for reducing the ferric protein to the catalytically active ferrous form.[3] Under aerobic conditions, all of the proximal ligand variants, except Mb(H64V,V68A,H93D), showed a ~ 2-fold reduction in cyclopropanation reactivity compared

8

to the anaerobic reaction, likely due to inhibition resulting from oxygen binding to the heme cofactor. In stark contrast, Mb(H64V,V68A,H93D) maintains high cyclopropanation activity also in the presence of oxygen, producing **3** in 70% conversion and 1,400 turnovers. Notably, Mb(H64V,V68A,H93D) was found to exhibit even higher cyclopropanation activity along with higher diastereo- and enantioselectivity under non-reducing conditions (78% conv., 1,590 TON, 99% *de*, 92% *ee*). This result is in contrast to Mb(H64V,V68A) and the other proximal ligand variants, which show significantly reduced (27% conv.) and negligible cyclopropanation activity (≤ 6% conv.), respectively, under the same reaction conditions. Thus, the introduction of Asp at the proximal ligand site of the hemoprotein has a unique effect toward enabling Mb to function as a carbene transferase under non-reducing conditions. Moreover, this mutation can recapitulate the effect previously observed for the same Mb variant upon substitution of His93 with N-methyl-histidine.[32] To investigate the involvement of ferric vs. ferrous species in Mb(H64V,V68A,H93D)-catalyzed cyclopropanation, the reaction under anaerobic/non-reducing conditions was also performed in the presence of CO-saturated buffer, which resulted in no product formation. As CO binds with high affinity only to ferrous heme, these results suggested that the ferrous form of Mb(H64V,V68A,H93D) is responsible for catalysis. However, since it is known that CO is able to slowly reduce ferric hemoproteins to their ferrous form,[45] further experiments were performed to rule out the possibility that ferric Mb(H64V,V68A,H93D) may be reduced by CO and then inhibited by it over the course of these reactions. Incubation of ferric Mb(H64V,V68A,H93D) with CO showed no formation of the ferrous CO-bound species up to 40-45 min, and formation of a only small amount of this species (<10%) at more extended incubation times (> 60 min) (**Figure S3**). On the other hand, analysis of ferric Mb(H64V,V68A,H93D)-catalyzed cyclopropanation of styrene with EDA after short reaction

9

times (i.e., 5 and 10 min, within which period no ferrous Mb(H64V,V68A,H93D) is formed

upon reduction by CO; **Figure S3**) showed formation of the cyclopropanation production in the

absence of CO, but not in the presence of it. Based on these results, we conclude that the H93D

mutation enables Mb(H64V,V68A,H93D) to be effectively reduced to the catalytically active

ferrous form by the diazo compound *in situ*,[46-48] resulting in efficient olefin cyclopropanation

without the need of an exogeneous reductant.

Table 1: Catalytic activity and selectivity of proximal ligand-substituted Mb variants in the

cyclopropanation of styrene with ethyl α-diazoacetate (EDA).[a]
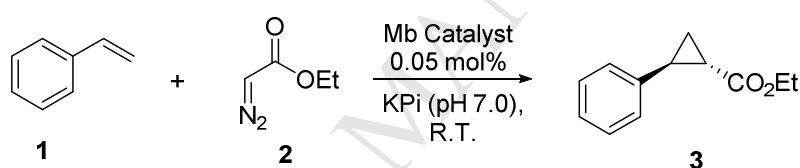
	Reaction Conditions:	*Anaerobic / Reducing*			*Aerobic / Reducing*			*Anaerobic / Non-reducing*		
Entry	Catalyst	TON	Conv.[b]	% *de* (*ee*)[c]	TON	Conv.[b]	% *de* (*ee*)[c]	TON	Conv.[b]	% *de* (*ee*)[c]
1	Mb(H64V,V68A)	1940	97%	>99 (>99)	1560	78%	97 (96)	545	27%	>99 (98)
2	Mb(H64V,V68A,H93(3PyA))	1045	52%	84 (8)	625	31%	85 (11)	105	5%	95 (62)[d]
3	Mb(H64V,V68A,H93(*p*AmF))	1115	56%	84 (5)	905	45%	84 (5)	50	3%	90 (36)[d]
4	Mb(H64V,V68A,H93(3ThA))	845	42%	83 (8)	590	29%	84 (10)	45	2%	93 (60)[d]
5	Mb(H64V,V68A,H93C)	905	45%	90 (27)	540	27%	90 (22)	125	6%	88 (29)
6	Mb(H64V,V68A,H93Y)	910	45%	90 (26)	475	24%	90 (31)	110	5%	93 (42)
7	Mb(H64V,V68A,H93S)	490	25%	90 (19)	270	13%	89 (18)	40	2%	86 (14)
8	Mb(H64V,V68A,H93D)	1930	96%	95 (75)	1400	70%	94 (68)	1590	79%	99 (92)

[a] *Reaction Conditions*: 0.01 M styrene, 0.02 M EDA, 5 µM Mb catalyst, 50 mM potassium phosphate buffer (pH 7) containing 5% ethanol, room temperature.
[b] As determined by GC using calibration curves with isolated **3**.
[c] For *trans*-(1*S*,2*S*) stereoisomer as determined by chiral GC.
[d] Diastereomeric and enantiomeric ratios may be overestimated due to low product formation in these reactions.

Mb(H64V,V68A) was previously shown to catalyze the cyclization of sulfonylazides to

give the corresponding sultam,[17] a reaction involving a nitrene $C(sp^3)$—H insertion via a heme-

nitrenoid intermediate.[16,19] To test the effect of the mutations on the nitrene transfer reactivity of this protein, a model C—H amination reaction with 2,4,6-triisopropyl-benzenesulfonyl azide (**4**) was used (**Table 2**). Under anaerobic and reducing conditions, Mb(H64V,V68A) catalyzes the conversion of **4** into the sultam **5** with 32% yield (GC) and 635 catalytic turnovers (**Table 2**, Entry 1). Interestingly, all of the proximal ligand variants proved to be functional and efficient nitrene transferases, supporting between 290 and 650 turnovers (**Table 2**). This catalytic activity compares well with that of synthetic organometallic catalysts previously reported for this transformation (50-100 TON).[49] Furthermore, the ncAA-containing variants Mb(H64V,V68A,H93(3PylA)) and Mb(H64V,V68A,H93pAmF) as well as Mb(H64V,V68A,H93D) were found to exhibit comparable efficiency as nitrene transferases as the histidine-ligated Mb(H64V,V68A) (635-650 vs. 635 TON, **Table 2**), thus indicating a significantly broader tolerance of this non-native reactivity to alteration of the heme coordinating residue compared to the carbene transferase reactivity (**Table 1**). In contrast to the results obtained with the cyclopropanation reaction (**Table 1**), a significant reduction in C—H amination activity was observed for all the Mb variants in the aerobic reactions, with Mb(H64V,V68A,H93pAmF) being the most efficient biocatalyst under these conditions. No C—H amination activity was observed instead in the absence of reductant with all the variants, indicating that, unlike for Mb(H64V,V68A,H93D) in the presence of the diazo reagent, the catalytically active ferrous form of these proteins is not produced in the presence of the azide reagent. Altogether, the results further highlight the differential impact of proximal ligand substitutions on Mb reactivity in the context of the two group transfer reactions.

Table 2. Catalytic activity of proximal ligand-substituted Mb variants for C-H amination with 2,4,6 triisopropylbenzenesulfonyl azide.[a]
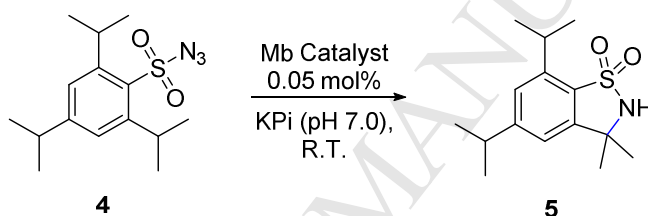
	Reaction Conditions:	Anaerobic + Reductant		Aerobic + Reductant	
Entry	Catalyst	TON	Conv.[b]	TON	Conv.[b]
1	Mb(H64V,V68A)	635	32%	76	4%
2	Mb(H64V,V68A,H93(3PyA))	650	33%	70	3%
3	Mb(H64V,V68A,H93pAmF)	640	32%	95	5%
4	Mb(H64V,V68A,H93(3ThA))	535	27%	72	4%
5	Mb(H64V,V68A,H93C)	290	14%	48	2%
6	Mb(H64V,V68A,H93Y)	475	24%	44	2%
7	Mb(H64V,V68A,H93S)	405	20%	24	1%
8	Mb(H64V,V68A,H93D)	635	32%	15	1%

[a] *Reaction conditions*: 0.01 M triisopropylbenzenesulfonyl azide, 5 µM Mb catalyst, 50 mM phosphate buffer (pH 7) containing 2.5% ethanol, room temperature.
[b] As determined by GC calibration curves using isolated **5**.

In summary, this work demonstrates how modification of the heme cofactor environment through substitution of the proximal axial ligand can be exploited to alter and modulate the carbene and nitrene transfer reactivity of myoglobin-based catalysts. Extending upon our

previous findings,[25,28] these results highlight the robustness and functional plasticity of the myoglobin scaffold toward tolerating structural alterations of the heme cofactor environment for tuning such non-native reactivities. Most notably, the present strategy proved effective for obtaining a Mb catalyst, Mb(H64V,V68A,H93D), that is capable of promoting efficient and stereoselective olefin cyclopropanation under nonreducing conditions. This reactivity, which was uniquely imparted by replacement of the proximal histidine ligand with an aspartate residue, results for the enhanced ability of this Mb variant to undergo conversion to the catalytically active, ferrous form directly upon reaction with the diazo substrate. Finally, the functionality of the ncAA-containing Mb variants as carbene and nitrene transferases demonstrate the utility of *p*-aminophenylalanine (*p*AmF), 3-(3'-pyridyl)-alanine (3PyA), and β-(3-thienyl)-alanine (3ThA) for re-engineering the first coordination sphere of this metalloprotein. Although these Mb variants did not outperform those generated by proximal ligand substitutions via proteinogenic amino acids in the context of the present reactions, these findings provide support to the feasibility of exploiting these pyridine-, thiophene-, and aniline-based unnatural amino acids for myoglobin engineering, and more broadly, for the engineering of metalloenzymes and metalloproteins.

Materials and Methods

General Information

All chemicals and reagents were purchased from commercial suppliers and used without any further purification unless otherwise stated. EDA was purchased from Sigma-Aldrich as 87%

m/v solution in dichloromethane. The unnatural amino acids were purchased from Chem-Impex. LC-MS grade trypsin was purchased from Pierce.

Cloning and Mutagenesis

The cloning of a pET22-based vector for the expression of Mb variant Mb(H64V,V68A) was described previously.[3] In this vector, the protein gene is under a IPTG-inducible T7 promoter. Additional mutations were incorporated into the Mb(H64V,V68A) template using SOE PCR and the primers listed in Table S2. The preparation of a pEVOL-based plasmid[50] for the expression of an engineered *M. jannaschii* tyrosyl-tRNA synthetase (TyrRS) and cognate amber suppressor tRNA (MjtRNA$_{CUA}^{Tyr}$) for the genetic incorporation of pAmF was described previously.[51] Using a similar procedure, a second pEVOL-based plasmid was prepared for encoding for a polyspecific AARS derived from *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase and the cognate amber suppressor tRNA (MbtRNA$_{CUA}^{Pyl}$) for the genetic incorporation of 3-pyridylalanine and 3-thienylalanine.[38] This gene was prepared by introducing the mutations L270I, Y271F, L274G, C313F, and Y349F into the MbPylRS gene via site-directed mutagenesis and the resulting gene was re-introduced into the pEVOL vector using the *Bgl* II/*Sal* I cassette. In these vectors, the AARS gene is under an arabinose-inducible araBAD promoters and the tRNA is under a constitutive (proK) promoter.

Growth media

Cell cultures were grown in enriched M9 medium, which was prepared as follows. For 1 L, 740 mL of deionized H_2O was added with 200 mL of M9 salts (5×) solution, 50 mL of 10% (w/v) yeast extract, 10 mL of glycerol, 1 mL of $MgSO_4$ (2 M), and 100 μL of $CaCl_2$ (1 M). The M9 salt (5×) solution was prepared by dissolving 34 g of Na_2HPO_4, 15 g of K_2HPO_4, 2.5 g of NaCl,

14

5 g of NH_4Cl in 1 L of deionized H_2O and sterilized by autoclaving. The glycerol and $MgSO_4$

solutions were autoclaved separately. The $CaCl_2$ stock solution was sterilized by filtration.

Enriched M9 agar plates were prepared by adding 17 g of agar to 1 L of enriched M9 media

containing all of the aforementioned components at the specified concentrations with the

exception of $CaCl_2$, which was added immediately prior to plating. To media and plates was

added ampicillin to a final concentration of 100 mg/L and chloramphenicol was added to a final

concentration of 34 mg/L.

Protein expression

The Mb variants containing all natural amino acids were expressed from pET22-based vectors in

E. coli BL21(DE3) as described previously.[3] The ncAA-containing Mb variants were expressed

from *E. coli* BL21(DE3) cells co-transformed with the pET22-based vector encoding for

Mb(H64V,V68A,H93(TAG)) and the appropriate pEVOL-based vector for the expression of

pAmF-RS (pAmF) or PylHRS (3PyA and 3ThA). After selection on M9 agar plates containing

ampicillin (100 mg L^{-1}) and chloramphenicol (34 mg L^{-1}), single colonies of the transformed

cells were used to inoculate 5 mL overnight cultures of enriched M9 medium (37°C, 200 rpm, 12

hours). The overnight cultures were used to inoculate 1L of enriched M9 medium. When OD_{600}

reached 0.4 – 0.6, cell cultures were condensed by centrifugation (4,000 rpm, 4°C, 20 min)

followed by resuspension of the cell pellet in 200 mL of fresh enriched M9 medium. Arabinose,

δ-aminolevulinic acid, and the desired amino acid were added to the expression media at final

concentrations of 4 mM, 0.3 mM, and 5 mM respectively followed by incubation for 30 minutes

with shaking (27°C, 200 rpm) after which protein expression was induced with 0.5 mM IPTG.

Following induction, cultures were shaken at 200 rpm and 27°C and harvested after 20 hours by

centrifugation (4000 rpm, 4°C, 20 minutes). Cell pellets were resuspended in 20 mL of Ni-NTA

Lysis Buffer (50 mM KPi, 250 mM NaCl, 10 mM histidine, pH 8.0) and protein extraction and purification were performed immediately.

Protein Purification and Characterization

Cell suspensions were lysed by sonication and clarified by centrifugation (14,000 rpm, 20 min, 4°C). The clarified lysate was transferred to a Ni-NTA column equilibrated with Ni-NTA Lysis Buffer (50 mM KPi, 250 mM NaCl, 10 mM histidine, pH 8.0). The resin was washed with 50 mL of Ni-NTA Lysis Buffer followed by 50 mL of Ni-NTA Wash Buffer (50 mM KPi, 250 mM NaCl, 20 mM histidine, pH 8.0). Proteins were eluted with Ni-NTA Elution Buffer (50 mM KPi, 250 mM NaCl, 250 mM histidine, pH 7.0). After elution from the Ni-NTA column, the protein was buffer exchanged against 50 mM KPi buffer (pH 7.0) using 10 kDa Centricon filters. The concentration of the Mb variants was determined using extinction coefficients as determined using the hemochrome assay[52] (Table S3). UV-vis absorbance spectra were recorded using a protein solution at 2 - 5 µM in 50 mM KPi (pH 7.0). For LC-MS analyses, the ncAA-containing amino acids were digested with trypsin in 50 µL reactions containing 2.5 µg trypsin, 10 mM $CaCl_2$, and 60 – 70 µg of myoglobin in 50 mM NH_4HCO_3 (pH 8.0). Digestion reactions incubated at 37 °C without shaking for 12 – 15 hours and were cleaned up using the C18 ZipTip protocol. Digested samples were given to the URMC Mass Spectrometry Resource Facility for LC-MS/MS analysis.

Cyclopropanation and C—H Amination Reactions

The cyclopropanation reactions with styrene and EDA and the C—H amination reactions with 2,4,6-triisopropyl-benzenesulfonyl azide were carried out as described previously.[3,17] Briefly, cyclopropanation reactions under anaerobic/reducing conditions were carried out at a 400 µL

scale using 5 µM myoglobin catalyst, 10 mM styrene, 20 mM EDA, and 10 mM dithionite in 50

mM KPi (pH 7.0). For analysis of C—H amination activity, reactions were typically carried out

at a 400 µL scale using 5 µM myoglobin catalyst, 10 mM 2,4,6 triisopropylbenzenesulfonyl

azide, and 10 mM dithionite in 50 mM KPi (pH 7.0). Under anaerobic conditions, reactions

performed in an anaerobic chamber and were stirred for 16 hours at room temperature and

aerobic reactions were performed in vessels open to the atmosphere. The reactions under non-

reducing conditions were performed in a similar manner without the addition of sodium

dithionite.

Product Analysis

The products of the cyclopropanation and C—H amination reactions were analyzed by chiral GC

and HPLC, respectively, using synthetically prepared (racemic) standards of **3**[4] and **5**[17] as

references. The cyclopropanation reactions were analyzed by adding 20 µL of internal standard

(benzodioxole, 50 mM in ethanol) to the reaction mixture followed by extraction with 400 µL of

dichloromethane and analysis by gas chromatography (GC). GC analyses were carried out using

a Shimadzu GC-2010 gas chromatograph equipped with an FID detector and a Chiral Cyclosil-B

column (30 m × 0.25 mm × 0.25 µm film). Separation method: 1 µL injection, injector temp: 200

°C, detector temp: 300 °C. Gradient: column temperature set at 120 °C for 3 min, then to 150 °C

at 0.8 °C/min, then to 245 °C at 25 °C/min. Total run time was 46.30 min. For the reactions with

2,4,6 triisopropylbenzenesulfonyl azide, the DCM extraction solution was evaporated, and the

residue was redissolved in 400 µL of methanol followed by HPLC analysis using a Shimadzu

LC-2010A instrument equipped with a UV–vis detector. Analytical conditions: Thermo Fisher

Scientific Hypersil GOLD C18 column (250 × 4.6 mm); UV detection: 254 nm; flow rate: 1.0

mL/ min: solvent A: 1% trifluoroacetic acid in H_2O; solvent B: 1% trifluoroacetic acid in

acetonitrile; gradient: 0−3 min: 30% B; 3−21 min: 30% B to 90% B; 21−24 min: 90% B; 24−24.5 min: 90% B to 30% B; 24.5−30 min: 30% B. Retention times: 17.35 min for **5**; and 24.54 min for **4.** Calibration curves for quantification of each product were constructed using authentic standards prepared synthetically. All measurements were performed at least in duplicate.

Analysis of thermostability

Circular dichroism (CD) spectroscopy and thermal denaturation analysis were performed using a JASCO J-1100 CD spectrophotometer equipped with variable temperature/wavelength denaturation software. Far UV CD spectra (250−190 nm) were obtained using 3 μM solutions of purified Mb variant in 50 mM potassium phosphate buffer (pH 7.0) and recorded at 20 °C at a scan rate of 50 nm/min with a bandwidth of 1 nm and an averaging time of 10 s per measurement. Thermal denaturation curves were measured by monitoring the change in molar ellipticity at 222 nm (θ_{222}) over a temperature range from 20 to 100 °C. The temperature increase was set to 0.5 °C per minute with an equilibration time of 10 s. Data integration time for the melt curve was set to 4 s with a bandwidth of 1 nm. Linear baselines for the folded (θ_f) and unfolded state (θ_u) were generated. Melting temperature (T_m) was determined using the low temperature ($\theta_f = m_f T + b_f$) and high temperature ($\theta_u = m_u T + b_u$) equations fitted to the experimental data before and after global unfolding, respectively. These data were normalized and converted to fraction of folded protein (F_f) vs. temperature plots, and the resulting curve was fitted to a sigmoidal equation (θ_{fit}) via nonlinear regression analysis in SigmaPlot from which apparent melting temperatures were derived. See Figure S for additional data.

ACKNOWLEDGEMENTS

18

References

1. Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. *Science* **2013,** *339*, 307-310.
2. Coelho, P. S.; Wang, Z. J.; Ener, M. E.; Baril, S. A.; Kannan, A.; Arnold, F. H.; Brustad, E. M. *Nat. Chem. Biol.* **2013,** *9*, 485-487.
3. Bordeaux, M.; Tyagi, V.; Fasan, R. *Angew. Chem. Int. Ed.* **2015,** *54*, 1744–1748.
4. Bajaj, P.; Sreenilayam, G.; Tyagi, V.; Fasan, R. *Angew. Chem. Int. Ed.* **2016,** *55*, 16110–16114.
5. Tyagi, V.; Fasan, R. *Angew. Chem. Int. Ed.* **2016,** *55*, 2512-2516
6. Tyagi, V.; Sreenilayam, G.; Bajaj, P.; Tinoco, A.; Fasan, R. *Angew. Chem. Int. Ed.* **2016,** *55*, 13562-13566.
7. Wang, Z. J.; Renata, H.; Peck, N. E.; Farwell, C. C.; Coelho, P. S.; Arnold, F. H. *Angew. Chem. Int. Ed.* **2014,** *53*, 6810-6813.
8. Weissenborn, M. J.; Low, S. A.; Borlinghaus, N.; Kuhn, M.; Kummer, S.; Rami, F.; Plietker, B.; Hauer, B. *Chemcatchem* **2016,** *8*, 1636-1640.
9. Ohora, K.; Meichin, H.; Zhao, L. M.; Wolf, M. W.; Nakayama, A.; Hasegawa, J.; Lehnert, N.; Hayashi, T. *J. Am. Chem. Soc.* **2017,** *139*, 17265-17268.
10. Knight, A. M.; Kan, S. B. J.; Lewis, R. D.; Brandenberg, O. F.; Chen, K.; Arnold, F. H. *ACS Central Sci.* **2018,** *4*, 372-377.
11. Moore, E. J.; Steck, V.; Bajaj, P.; Fasan, R. *J. Org. Chem.* **2018,** *83*, 7480-7490.
12. Chen, K.; Huang, X. Y.; Kan, S. B. J.; Zhang, R. K.; Arnold, F. H. *Science* **2018,** *360*, 71-+.
13. Brandenberg, O. F.; Prier, C. K.; Chen, K.; Knight, A. M.; Wu, Z.; Arnold, F. H. *ACS Catal.* **2018,** *8*, 2629-2634.
14. Villarino, L.; Splan, K. E.; Reddem, E.; Alonso-Cotchico, L.; Gutiérrez de Souza, C.; Lledós, A.; Maréchal, J.-D.; Thunnissen, A.-M. W. H.; Roelfes, G. *Angew. Chem. Int. Ed.* **2018,** *57*, 7785-7789.
15. McIntosh, J. A.; Coelho, P. S.; Farwell, C. C.; Wang, Z. J.; Lewis, J. C.; Brown, T. R.; Arnold, F. H. *Angew. Chem. Int. Ed.* **2013,** *52*, 9309-9312.
16. Singh, R.; Bordeaux, M.; Fasan, R. *ACS Catal.* **2014,** *4*, 546-552.
17. Bordeaux, M.; Singh, R.; Fasan, R. *Bioorg. Med. Chem.* **2014,** *22*, 5697-5704.
18. Hyster, T. K.; Farwell, C. C.; Buller, A. R.; McIntosh, J. A.; Arnold, F. H. *J. Am. Chem. Soc.* **2014,** *136*, 15505-15508.

19. Singh, R.; Kolev, J. N.; Sutera, P. A.; Fasan, R. *ACS Catal.* **2015,** *5*, 1685-1691.
20. Prier, C. K.; Zhang, R. J. K.; Buller, A. R.; Brinkmann-Chen, S.; Arnold, F. H. *Nature Chem.* **2017,** *9*, 629-634.
21. Tinoco, A.; Steck, V.; Tyagi, V.; Fasan, R. *J. Am. Chem. Soc.* **2017,** *139*, 5293-5296.
22. Chandgude, A. L.; Fasan, R. *Angew. Chem. Int. Ed.* **2018,** *57*, 15852-15856.
23. Brandenberg, O. F.; Fasan, R.; Arnold, F. H. *Curr. Opin. Biotech.* **2017,** *47*, 102-111.
24. Springer, B. A.; Sligar, S. G.; Olson, J. S.; Phillips, G. N. *Chem. Rev.* **1994,** *94*, 699-714.
25. Sreenilayam, G.; Moore, E. J.; Steck, V.; Fasan, R. *Adv. Synth. Cat.* **2017,** *359*, 2076–2089.
26. Key, H. M.; Dydio, P.; Clark, D. S.; Hartwig, J. F. *Nature* **2016,** *534*, 534-537.
27. Wolf, M. W.; Vargas, D. A.; Lehnert, N. *Inorg. Chem.* **2017,** *56*, 5623-5635.
28. Sreenilayam, G.; Moore, E. J.; Steck, V.; Fasan, R. *ACS Catal.* **2017,** *7*, 7629-7633.
29. Chen, K.; Zhang, S. Q.; Brandenberg, O. F.; Hong, X.; Arnold, F. H. *J. Am. Chem. Soc.* **2018,** *140*, 16402-16407.
30. Onderko, E. L.; Silakov, A.; Yosca, T. H.; Green, M. T. *Nat. Chem.* **2017,** *9*, 623-628.
31. Pott, M.; Hayashi, T.; Mori, T.; Mittl, P. R. E.; Green, A. P.; Hilvert, D. *J. Am. Chem. Soc.* **2018,** *140*, 1535-1543.
32. Hayashi, T.; Tinzl, M.; Mori, T.; Krengel, U.; Proppe, J.; Soetbeer, J.; Klose, D.; Jeschke, G.; Reiher, M.; Hilvert, D. *Nature Catal.* **2018,** *1*, 578-584.
33. Poulos, T. L. *Chem. Rev.* **2014,** *114*, 3919-3962.
34. Fita, I.; Rossmann, M. G. *Proc. Natl. Acad. Sci. USA* **1985,** *82*, 1604-1608.
35. Krebs, C.; Fujimori, D. G.; Walsh, C. T.; Bollinger, J. M. *Accounts Chem. Res.* **2007,** *40*, 484-492.
36. White, M. D.; Flashman, E. *Curr. Opin. Chem. Biol.* **2016,** *31*, 126-135.
37. Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. *Nature Biotechnol.* **2002,** *20*, 1044-1048.
38. Xiao, H.; Peters, F. B.; Yang, P. Y.; Reed, S.; Chittuluru, J. R.; Schultz, P. G. *Acs Chemical Biology* **2014,** *9*, 1092-1096.
39. Joule, J. A.; Mills, K.; Smith, G. F. *Heterocyclic Chemistry*; Springer; 3rd edition 1995.
40. Wang, L.; Schultz, P. G. *Angew. Chem. Int. Ed.* **2004,** *44*, 34-66.
41. Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G. *Chem. Sci.* **2015,** *6*, 50-69.
42. Moore, E. J.; Zorine, D.; Hansen, W. A.; Khare, S. D.; Fasan, R. *Proc. Natl. Acad. Sci. USA* **2017,** *114*, 12472-12477.
43. Anzenbacher, P.; Hudecek, J.; Struzinsky, R. *FEBS Lett.* **1982,** *149*, 208-210.
44. Prieto, G.; Suarez, M. J.; Gonzalez-Perez, A.; Ruso, J. M.; Sarmiento, F. *Phys. Chem. Chem. Phys.* **2004,** *6*, 816-821.
45. Bickar, D.; Bonaventura, C.; Bonaventura, J. *J. Biol. Chem.* **1984,** *259*, 777-783.
46. Wolf, J. R.; Hamaker, C. G.; Djukic, J. P.; Kodadek, T.; Woo, L. K. *J. Am. Chem. Soc.* **1995,** *117*, 9194-9199.
47. Lai, T. S.; Chan, F. Y.; So, P. K.; Ma, D. L.; Wong, K. Y.; Che, C. M. *Dalton T.* **2006,** 4845-4851.
48. Salomon, R. G.; Kochi, J. K. *J. Am. Chem. Soc.* **1973,** *95*, 3300-3310.
49. Ruppel, J. V.; Kamble, R. M.; Zhang, X. P. *Org. Lett.* **2007,** *9*, 4889-4892.
50. Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. *J. Mol. Biol.* **2010,** *395*, 361-74.
51. Frost, J. R.; Jacob, N. T.; Papa, L. J.; Owens, A. E.; Fasan, R. *ACS Chem. Biol.* **2015,** *10*, 1805-16.

52. Berry, E. A.; Trumpower, B. L. *Analyt. Biochem.* **1987,** *161*, 1-15.

Graphical Abstract