Enzyme stabilization via computationally guided protein stapling

Eric J. Moore[a], Dmitri Zorine[b,c], William A. Hansen[b,c,d], Sagar D. Khare[b,c,d,1], and Rudi Fasan[a,1]

[a]Department of Chemistry, University of Rochester, Rochester, NY 14627; [b]Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854; [c]Center for Integrative Proteomics Research, Rutgers University, Piscataway, NJ 08854; and [d]Quantitative Biomedicine Graduate Program, Rutgers University, Piscataway, NJ 08854

Thermostabilization represents a critical and often obligatory step toward enhancing the robustness of enzymes for organic synthesis and other applications. While directed evolution methods have provided valuable tools for this purpose, these protocols are laborious and time-consuming and typically require the accumulation of several mutations, potentially at the expense of catalytic function. Here, we report a minimally invasive strategy for enzyme stabilization that relies on the installation of genetically encoded, nonreducible covalent staples in a target protein scaffold using computational design. This methodology enables the rapid development of myoglobin-based cyclopropanation biocatalysts featuring dramatically enhanced thermostability ($\Delta T_m = +18.0\ °C$ and $\Delta T_{50} = +16.0\ °C$) as well as increased stability against chemical denaturation [ΔC_m (GndHCl) = 0.53 M], without altering their catalytic efficiency and stereoselectivity properties. In addition, the stabilized variants offer superior performance and selectivity compared with the parent enzyme in the presence of a high concentration of organic cosolvents, enabling the more efficient cyclopropanation of a water-insoluble substrate. This work introduces and validates an approach for protein stabilization which should be applicable to a variety of other proteins and enzymes.

protein thermostabilization | computational protein design | myoglobin | noncanonical amino acids | Rosetta macromolecular modeling

Enzymes play a major role in biotechnology and constitute attractive catalysts for the implementation of efficient, selective, and sustainable processes for the production of pharmaceuticals, fine chemicals, and biofuels (1, 2). The marginal stability of proteins, however, poses a fundamental challenge for the exploitation of enzymes for practical-scale syntheses and chemical manufacturing, which often require harsh reaction conditions such as elevated temperature and exposure to organic solvents (1). Due to these limitations, protein stabilization against thermal and chemical denaturation has represented a long-standing goal in enzyme design and engineering. In addition to higher robustness to operational conditions, increasing the thermostability of a protein can enhance its evolvability (i.e., its tolerance to mutagenesis for the acquisition of new or improved functions) (3).

Well-known experimental strategies for enzyme stabilization include directed evolution and consensus mutagenesis (4–6). While effective in many instances, these procedures remain laborious and time-consuming, typically requiring several rounds of mutagenesis and screening to obtain variants with significantly increased stability (i.e., $\Delta T_m > 5$–10 °C) (7). Notable examples of enzyme stabilization have been reported using computational design (8). In this area, methods have focused on optimizing native state interactions, for example through improving core packing (9–11), fragment contacts (12), combined structure- and phylogeny-guided energy optimization (13, 14), surface charge optimization (15, 16), and rigidification (17, 18). A complementary but comparatively less explored approach has focused on decreasing the configurational entropy of the unfolded state ensemble (19), primarily via the installation of nonnative disulfide linkages (20). Indeed, the introduction of nonnative disulfide bridges (20) has proven useful for increasing the thermostability of enzymes, although stability–activity trade-offs were also observed (21, 22). Disulfide cross-linking can additionally provide kinetic stabilization (23). Unfortunately, the redox instability of disulfide bridges make this approach unsuitable for many proteins and enzymes that are meant to function in reducing environments, such as the intracellular milieu and/or in the presence of reductants (24–26). Moreover, the reversibility and homotypic nature of disulfide bridges can complicate folding of variants with multiple cross-links due to disulfide scrambling (27). Thus, the development of computational strategies for protein stabilization by means of chemically stable covalent bonds remains an unmet challenge.

Here, we report the development of Rosetta-guided protein stapling (R-GPS), a method for enzyme/protein stabilization based on the computational design of genetically encodable, covalent "staples" in a protein of interest. This method utilizes the Rosetta enzyme design framework to identify optimal sites for intramolecular cross-linking of the protein scaffold via redox-stable and irreversible thioether bonds generated upon the reaction between a cysteine and a genetically encoded noncanonical amino acid containing a cysteine-reactive side-chain group (O-2-bromoethyl tyrosine, or O2beY) (28). The R-GPS approach was implemented and validated in the context of a myoglobin (Mb)-based cyclopropanation biocatalyst, resulting in the development of "cyclopropanases" with significantly increased stability against thermal and chemical denaturation as well as improved catalytic performance under harsh reaction conditions.

Results

Choice of Target Enzyme and Protein Stapling Chemistry. A recently reported Mb-derived cyclopropanation biocatalyst, Mb(H64V, V68A) (29), was selected as the test bed for the development of R-GPS. Compared with wild-type Mb, Mb(H64V,V68A) bears

Significance

The marginal stability of most natural proteins presents a challenge for the exploitation of natural and engineered enzymes in biotechnology and industrial biocatalysis. Protein stabilization can be time- and labor-intensive and often involves extensive amino acid substitutions, which may impair the activity and/or selectivity of the enzyme. Here, we describe a computational design method for enzyme stabilization that uses structure-based modeling to introduce covalent "staples" in a protein scaffold via a genetically encodable noncanonical amino acid. This method was applied to obtain stapled variants of a stereoselective cyclopropanation biocatalyst featuring greatly increased thermostability and robustness to high concentrations of organic cosolvents. This minimally invasive strategy for protein stabilization should be applicable to a variety of enzymes and proteins.

two active site mutations that confer enhanced catalytic activity as well as high diastereo- and enantioselectivity (>90–99% *de* and *ee*) in the cyclopropanation of styrenes and vinylarenes with ethyl α-diazoacetate as the carbene donor (29). On one hand, stabilization of Mb(H64V,V68A) was desirable for enhancing its robustness for synthetic applications, which include the synthesis of cyclopropane-containing drugs and other carbene transfer reactions (30, 31). On the other hand, the high stereoselectivity of Mb(H64V,V68A) was envisioned to provide a sensitive probe for evaluating the impact of the computationally designed covalent staples on subtle functional properties such as stereoinduction. Indeed, previous studies indicated that the stereoselectivity of Mb-based cyclopropanation catalysts is highly sensitive to small structural variations in their active sites (30).

The thioether bond-forming reaction between cysteine and the genetically encodable O-2-bromoethyl-tyrosine (O2beY) (28, 32) was selected as the protein cross-linking strategy (Fig. 1*A*). Based on these previous studies, the O2beY/Cys reaction was anticipated to offer several promising features for the purpose of protein stapling, namely (*i*) the ability to mediate the spontaneous formation of the thioether cross-link at the posttranslational level, (*ii*) high chemoselectivity of O2beY toward cysteine-mediated alkylation in the presence of other nucleophilic residues (e.g., Lys and His), and (*iii*) "spatially controlled" reactivity, whereby the O2beY and Cys undergo the nucleophilic substitution reaction only when located in close proximity.

Computational Design of Stapled Mb(H64V,V68A) Variants Using R-GPS. As a first design principle we reasoned that cross-linking residues that are distal in the primary sequence but proximal in space in the folded state would maximize thermostabilization by decreasing the chain entropic cost of folding (33–36). As a second criterion we envisioned the need to identify compatible backbone positions for placing the haloalkane (O2beY) and thiol (Cys) side-chain groups so that the cross-link is accommodated in a strain-free configuration while making energetically favorable interactions with its surrounding residues (33). Following these guiding principles, the N-terminal helices A and B and the C-terminal helices G and H of the target enzyme, Mb(H64V,V68A), were chosen as the target regions for the installation of the thioether bridges (Fig. 1*B*), as cross-links between these structural elements would bear the highest contact order. The RosettaMatch algorithm (37) was then applied to identify positions where the O2beY and Cys residues can be accommodated for formation of the thioether staple. Once geometrically feasible backbone

positions of the cross-linking residues were obtained (Fig. 1*B*), RosettaDesign (38) was performed to identify additional sequence changes to minimize steric clashes with the modeled staple. To assess the method, we selected for experimental characterization a diverse set of nine designs based on (*i*) the solvent accessibility of the staple in the protein structure, (*ii*) total number of amino acid substitutions, and (*iii*) Rosetta energies (Table 1). These designs were termed sMb1 through sMb9 and had differences in scores relative to the parent protein Mb(H64V,V68A) ranging from −6.7 (sMb1) to +16.6 (sMb9) Rosetta energy units (Reu), and intervening segment lengths ranging from 73 (sMb2) to 121 (sMb5) residues. Seven of designed variants feature a surface exposed thioether linkage, whereas for the remaining two (sMb6 and sMb8) the cross-link is buried (Fig. 1*C*). The number of designed amino acid substitutions, excluding the Cys and O2beY residues, ranged from zero (sMb2 and sMb5) to two (sMb3, sMb6, sMb7, sMb8, and sMb9).

Expression and Characterization of the sMb Designs. All of the designed Mb(H64V,V68A)-derived variants could be expressed in soluble form from *Escherichia coli* cells containing an orthogonal aminoacyl-tRNA synthetase/tRNA$_{CUA}$ pair for O2beY incorporation (28) via amber stop codon suppression (39). In addition, the constructs were all able to bind heme and fold properly, as evinced from the characteristic Soret band (~410 nm) in the corresponding UV-visible spectra (*SI Appendix*, Fig. S4).

SDS/PAGE analysis revealed an increase in electrophoretic mobility for five out of the nine protein constructs, namely sMb2, sMb3, sMb4, sMb5, and sMb7, compared with Mb(H64V,V68A) (Fig. 2*A* and *SI Appendix*, Fig. S5). This behavior is indicative of a more compact structure of the protein under denaturing and reducing conditions (SDS + DTT), which is consistent with the presence of the nonreducible thioether (O2beY/Cys) cross-link. These conclusions were corroborated by MALDI-TOF MS (Fig. 2*B* and *SI Appendix*, Fig. S6), which showed a single signal corresponding to expected mass of these proteins minus 82 Da, deriving from the loss of HBr as a result of the O2beY/Cys cross-linking reaction. For sMb4, two *m/z* signals consistent with the stapled and unstapled form of the protein indicated the O2beY/cysteine cross-linking reaction has occurred only partially. For the remaining designs, the only observable species in the MS spectra corresponded to the O2beY-containing protein lacking the thioether cross-link. Notably, the integrity of the O2beY residue in these unstapled constructs demonstrates the lack of undesirable reactivity toward hydrolysis or abundant thiol-containing intracellular metabolites such as glutathione. These results also show how placing the O2beY/Cys pair in close spatial proximity in the context of a folded protein is a necessary but not sufficient condition for productive stapling. The positional dependence of the stapling reaction could be rationalized based on analysis of the rotamers of Cys and O2beY residues (Fig. 1*B*) in models of the sMb constructs in their unstapled forms. In particular, a good agreement was found between the experimental results and the accessibility of near-attack conformations (NACs) compatible with a bimolecular nucleophilic substitution reaction between the thiol group of the cysteine and the alkyl bromide group in O2beY (see *SI Appendix* for further discussion).

Thermostability of the sMb Designs. The thermostability of the Mb variants was examined by measuring their melting temperatures (T_m) using CD (*SI Appendix*, Fig. S7). Mb(H64,V68A) was found to exhibit an apparent T_m of 66.0 °C (Fig. 2*C*), which is about 14 °C lower than that of wild-type Mb. Notably, all of the sMb constructs containing the thioether staple showed an increase in thermostability compared with the parent protein, with ΔT_m values ranging from 3.9 °C to 10 °C (Table 1). The largest thermostabilization effect was observed for sMb5 (ΔT_m: +10.0 °C), followed by sMb2 (ΔT_m: +7.2 °C). In contrast, the sMb variants lacking the cross-link show comparable (sMb1 and sMb9) or lower T_m values (sMb4, sMb6, and sMb8) compared with Mb(H64V,V68A) (ΔT_m: +1.5 to −16.8 °C; Table 1). To further examine the impact of the thioether staple, analogs of the two most stable sMb variants, sMb2 and sMb5, were prepared by substituting O2beY for O-propargyl-tyrosine
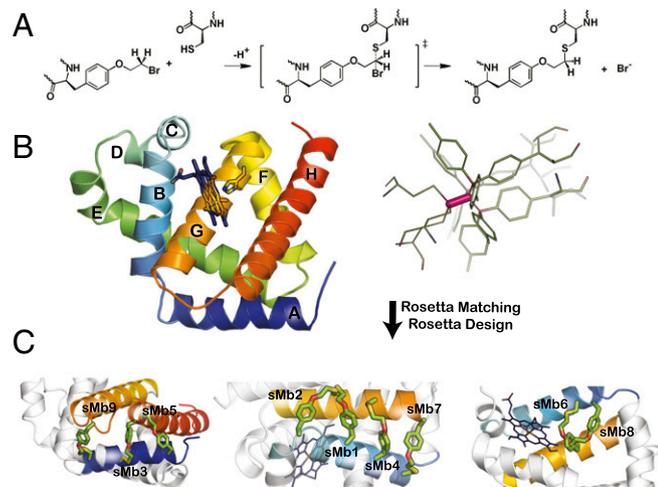
Fig. 1. Computational design approach. (*A*) Stapling reaction between the noncanonical amino acid O-2-bromoethyl tyrosine and cysteine resulting in a chemically stable thioether bond. (*B*) Structure of Mb (Protein Data Bank ID code 1JP9) highlighting helices A–H (*Left*). The active-site heme group and metal-coordinating histidine residue are shown as sticks. Conformational ensemble of the modeled thioether linkage used to find compatible locations for stapling (*Right*). (*C*) Models showing locations of covalent staples between A–H and B–G helices in designs sMb1 through sMb9.

Table 1. Computational and experimental values for Mb(H64V,V64A) and its variants

Variant	Mutations	ΔG^{comp}, Reu	Location staple*	Attack angle, deg[†]	Staple (MS)	T_m °C (ΔT_m)[‡]	T_{50}, °C (ΔT_{50})[§]	C_m (M) (ΔC_m)[¶]
Mb(H64V,V68A)	—	0	None	n.a.	n.a.	66.0 ± 1.0 (0.0)	62.3 ± 0.6 (0.0)	1.55 ± 0.02 (0.0)
sMb1	R31(O2beY), S35K, E109C	−6.73	B–G	102	No	67.5 ± 1.3 (+1.5)	59.1 ± 0.6 (−3.2)	n.d.
sMb2	H36(O2beY), E109C	−5.05	B/C–G	142	Yes	73.2 ± 0.2 (+7.2)	63.8 ± 0.3 (+1.5)	n.d.
sMb3	L9R, H12C, D122T, A127(O2beY)	−2.66	A–H	126	Yes	72.2 ± 1.4 (+6.2)	60 ± 3 (−2.3)	n.d.
sMb4	D27(O2beY), H113C, V114G	−0.76	B–G	98	Yes (partial)	56.1 ± 1.1 (−9.9)	50 ± 2 (−12.3)	n.d.
sMb5	G5(O2beY), D126C	−0.68	A–H	169	Yes	76.0 ± 2.0 (+10.0)	71.5 ± 0.9 (+9.2)	1.79 ± 0.02 (+0.24)
sMb6	G25C, I28A, L69A, I111(O2beY)	+1.70	B–G (core)	132	No	53.0 ± 1.0 (−13.0)	51.2 ± 0.3 (−11.1)	n.d.
sMb7	D20S, G23C, D27A, R118(O2beY)	+2.34	B-G	156	Yes	69.9 ± 1.4 (+3.9)	54 ± 1 (−8.3)	n.d.
sMb8	H24(O2beY), I28S, L69S, I111C	+3.05	B–G (core)	108	No	49.2 ± 1.0 (−16.8)	43.3 ± 0.6 (−19.0)	n.d.
sMb9	K16C, H119A, G121(O2beY), D122S	+16.6	A-G/H	112	No	64.3 ± 1.3 (−1.7)	52 ± 2 (−10.3)	n.d.
sMb10	G5(O2beY), H36(O2beY), E109C, D126C	n.a.	B/C-G + A–H	n.a.	Yes (double)	82.8 ± 0.8 (+16.8)	73 ± 2 (+10.7)	2.01 ± 0.05 (+0.46)
sMb11	G5(O2beY), H36(O2beY), F106A, E109C, D126C	n.a.	B/C–G + A–H	n.a.	Yes (double)	n.d.	61 ± 1 (−1.3)	n.d.
sMb12	G5(O2beY), H36(O2beY), E109C, D126C, G129E	n.a.	B/C–G + A–H	n.a.	n.d.	n.d.	66.6 ± 0.8 (+4.3)	n.d.
sMb13	G5(O2beY), H36(O2beY), E109C, H113E, D126C	n.a.	B/C–G + A–H	n.a.	Yes (double)	84.0 ± 0.2 (+18.0)	78.3 ± 0.8 (+16.0)	2.08 ± 0.03 (+0.53)

deg, degrees; n.a., not applicable; n.d., not determined.
*Corresponding to alpha helices A through H as labeled in Fig. 1.
[†]Angle for nucleophilic attack of Cys thiol group to O2beY side-chain alkyl bromide group (θ_{attack}) as described in *SI Appendix*, Fig. S1A.
[‡]Apparent melting temperatures as determined via circular dichroism (*SI Appendix*, Fig. S7).
[§]Half-maximal denaturation temperature of holoprotein (Soret band) after 10-min incubation.
[¶]Midpoint guanidinium chloride concentration from chemical denaturation curves measured by CD.

(OpgY) (40). OpgY is an isostere of O2beY but is unable to react with cysteine to form the thioether cross-link, as confirmed by SDS/PAGE (Fig. 2*A*) and MALDI-TOF MS analyses (*SI Appendix*, Fig. S6). The resulting OpgY-containing variants, sMb2(OpgY) and sMb5(OpgY), displayed significantly lower T_m values than their stapled counterparts (ΔT_m: –12.2 °C and –11.9 °C, respectively), confirming the stabilizing effect of the thioether cross-link, as encoded in computational design.

As a second measure of thermostability, half-maximal denaturation temperatures (T_{50}) were determined by monitoring heme loss (λ_{max}: 408 nm) upon incubation of the hemoproteins (10 min) at variable temperatures (Fig. 2*D*). This is a more stringent assay of thermal stability since it monitors the ability of the sMb variants to remain associated with the heme cofactor, which is essential for their activity as cyclopropanation biocatalysts. As thermal denaturation of holomyoglobin is irreversible, an increase in T_{50} is also a measure of kinetic stabilization upon cross-linking. In this assay, Mb(H64V,V68A) showed a T_{50} value of 62.3 °C (Table 1). Whereas sMb3 and sMb7 showed reduced T_{50} values compared with Mb(H64V,V64A) (Table 1), both sMb5 and sMb2 exhibited improved thermal stability compared with the parent protein (ΔT_{50}: +9.2 °C and +1.5 °C, respectively; Fig. 2*C*). These results demonstrate that the thermostabilization effect induced by the O2beY/Cys cross-links in both sMb2 and sMb5 is not restricted to the protein secondary structure, as determined by the CD melting curves, but extends to the heme-bound ("holo") forms of these proteins.

Design and Characterization of Doubly Stapled Mb Variants. Further stabilization of Mb(H64V,V68A) was then pursued by combining the covalent staples from the two most promising variants, namely sMb2 and sMb5, resulting in the sMb10 design. Upon production in *E. coli*, MS analysis confirmed the successful formation of both cross-links within this protein (Fig. 2*B*). Further characterization of sMb10 showed a nearly additive effect of the two thioether staples toward increasing thermostability of the hemoprotein in terms of both T_m and T_{50} (ΔT_m: +16.8 °C; ΔT_{50}: +10.7 °C; Table 1). Comparison of the CD spectra corresponding to sMb5, sMb10, and Mb(H64V,V68A) revealed no major changes in the protein secondary structure as a result of the presence of a single (sMb5) and double staple (sMb10) (*SI Appendix*, Fig. S8).

Evaluation and further stabilization of sMb10 was then pursued via structure-guided point mutagenesis. Investigation of the F106A substitution in sMb11 confirmed the energetic importance of the computed π–π stacking interaction between the O2beY residue in position 36 and the phenyl ring of Phe106 (see *SI Appendix* for further discussion). Next, we noticed that installation of the 36–109 cross-link (from sMb2) and 5–126 cross-link (from sMb5) replaces in each case a solvent-exposed negatively charged residue (i.e., Glu109 and Asp126, respectively) with the less-polar O2beY/Cys staple. Accordingly, two additional constructs (sMb12 and sMb13) were designed to incorporate "compensating" neutral-to-negatively charged amino acid substitutions (i.e., G129E and H113E, respectively) at a compatible position in proximity of each staple. Whereas sMb12 showed reduced thermostability (T_{50}) compared with sMb10, sMb13 was found to exhibit higher T_m (ΔT_m: +1.2 °C) as well as significantly increased stability against temperature-induced heme loss (ΔT_{50}: +5.3 °C) compared with sMb10. The presence of the two O2beY/Cys cross-links in sMb13 was confirmed by MS

Moore et al.
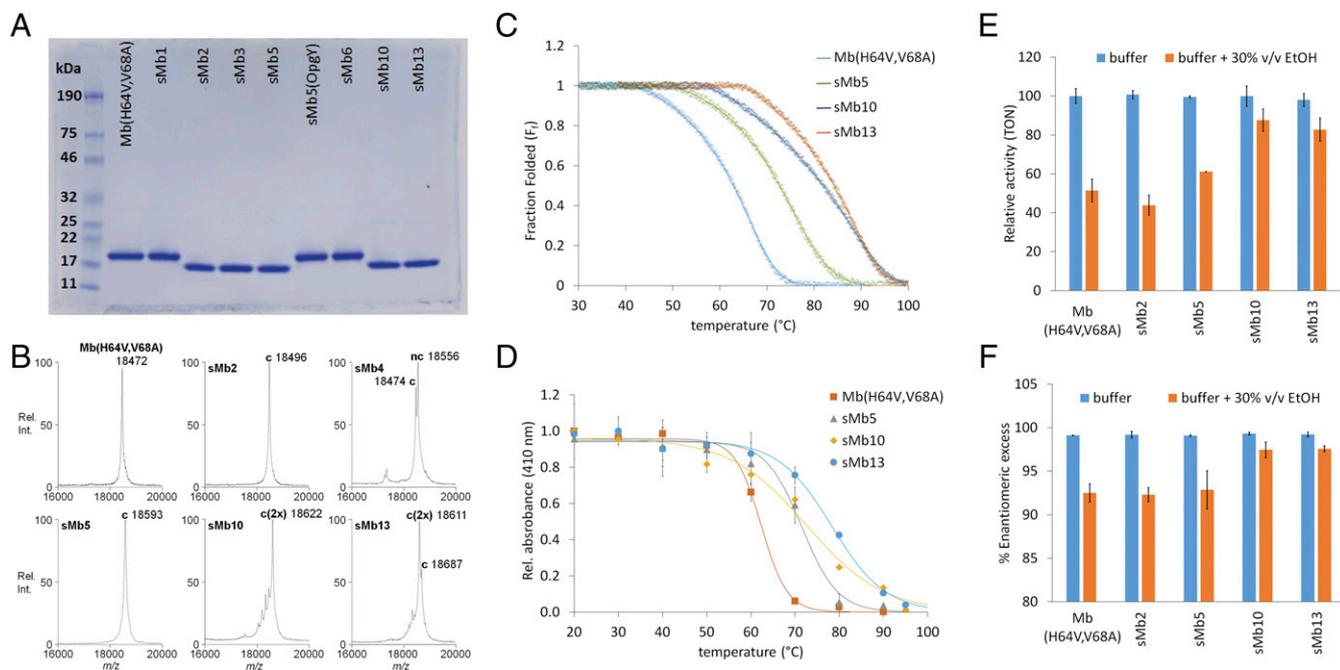
Fig. 2. Characterization of stapled Mb(H64V,V68A) variants (sMb variants). (*A* and *B*) SDS/PAGE gel (*A*) and MALDI-TOF MS spectrum (*B*) of Mb(H64V,V68A) and representative sMb variants. c, cross-linked; c(2×), doubly cross-linked; nc, not cross-linked. Calculated masses: Mb(H64V,V68A): 18,474 Da; sMb2 (c): 18,500 Da; sMb4 (c): 18,472 Da; sMb5 (c): 18,594 Da; sMb10 [c(2×)]: 18,621 Da; sMb13 [c(2×)]: 18,612 Da. (*C*) Thermal denaturation curves for Mb(H64V,V68A) and selected stapled variants as measured via CD at 220 nm (T_m determination). (*D*) Heat-induced inactivation curves (heme loss) for the same proteins as determined by decrease of Soret band signal (408 nm) after incubation (10 min) at variable temperatures (T_{50} determination). See *SI Appendix*, Figs. S5–S7 for additional data. (*E* and *F*) Catalytic activity (*E*) and enantioselectivity (*F*) of Mb(H64V,V68A) and stapled variants in styrene cyclopropanation reactions with EDA in buffer only and in the presence of 30% vol/vol ethanol. Relative activities refer to normalized catalytic turnovers (TON) relative to TON measured with Mb(H64V,V68A) in buffer only reactions. See *SI Appendix*, Figs. S10–S12 for related data with other organic cosolvents. Rel Int, relative intensity.

analysis, which also revealed the presence of a small fraction of singly stapled protein (Fig. 2*B*). The CD spectra of sMb13 and sMb10 were found to be superimposable, indicating that the "charge-compensating" mutation H113E had minimal impact on the protein structure (*SI Appendix*, Fig. S8). Thus, as a result of merely five solvent-exposed mutations in sMb13, the thermostability (T_m) of the original Mb(H64V,V64A) biocatalyst could be augmented by +18 °C. These modifications also translated into a significantly increased thermal stability of the heme-bound form of the protein (ΔT_{50}: +16 °C).

Catalytic Activity and Selectivity of Stapled Mb(H64V,V68A) Variants in Cyclopropanation Reactions. In the absence of selective pressure to maintain function, thermostabilization via protein engineering is often accompanied by a reduction in catalytic efficiency and/or selectivity of the target enzyme (41–43). To assess the impact of the thioether staples on the catalytic properties and stereoselectivity of the thermostabilized sMb variants these biocatalysts were tested in a model cyclopropanation reaction with styrene (**1**) and α-diazoacetate (EDA, **2**) (Table 2). Under the applied conditions [10 mM styrene, 20 mM EDA, 2 μM Mb variant (0.02 mol%)] Mb(H64V,V68A) was found to produce (1*S*,2*S*)-ethyl 2-phenylcyclopropanecarboxylate (**3a**) in 94% yield (4,710 turnovers or TON) and high diastereomeric (99.4% *de*) and enantiomeric excess (99.1% *ee*) (entry 1, Table 2). Under identical conditions, both sMb10 and sMb13 were able to catalyze the formation of **3a** with equally high catalytic activity (92–94% yields; 4,710–4,615 TON) and stereoselectivity (>99% *de* and *ee*; entries 4–5, Table 2). These data clearly indicated that thermostabilization induced by the stapling procedure had no negative impact on the carbene transfer reactivity of these hemoproteins, nor did it perturb the asymmetric environment provided by the distal heme pocket, which is critical for inducing the (1*S*,2*S*)-enantioselectivity during the cyclopropanation reaction (29).

Increased Stability Against Chemical Denaturation. Enhanced robustness to chemical denaturation and organic solvents is a highly desirable trait of enzymes to be employed for synthetic applications (44, 45). To examine this aspect, the most promising sMb constructs (sMb5, sMb10, and sMb13) were subjected to denaturation experiments in the presence of guanidinium chloride (Gnd·HCl) (*SI Appendix*, Fig. S9). Reflecting the trend emerging from the thermostability assays, a progressive stability increase in the presence of the chaotropic agent was observed going from the singly stapled sMb5 (C_m: 1.79 M) to the doubly stapled sMb10 and sMb13 (C_m: 2.01 M and 2.08 M, respectively), compared with the parent enzyme (C_m: 1.55 M).

Next, we investigated the effect of the staples toward improving the performance of the sMb variants in organic solvents. In the presence of 30% (vol/vol) ethanol, Mb(H64V,V68A) catalyzes the cyclopropanation of styrene with reduced activity (48% yield; 2,420 TON; 51% relative activity) and lower enantioselectivity (92% *ee*) compared to the same reaction in buffer (Fig. 2 *E* and *F*; entry 6 vs. 1, Table 2). Compared with Mb(H64V,V68A), both sMb10 and sMb13 better tolerate the presence of the organic cosolvent, affording the cyclopropanation product **3a** in higher yields (78–82%; 3,900–4,100 TON; entry 7, Table 2) and with higher diastereo- (1:400 vs. 1:130 diastereomeric ratio for *trans:cis*) and enantioselectivity (97% vs. 92% *ee*) (entry 7 vs. 6, Fig. 2*F* and Table 2). As stereoselectivity is highly sensitive to structural perturbation within the active site of these biocatalysts (30), these results suggest that stabilization of the Mb scaffold by covalent stapling minimizes the disruptive effects induced by the organic solvent. A similar trend was observed upon comparing the activity and selectivity of sMb5, sMb10, and sMb13 with those of Mb(H64V,V68A) in cyclopropanation reactions in the presence of high concentrations (30% vol/vol) of other organic solvents such as methanol, dimethylformamide (DMF), and DMSO (*SI Appendix*, Figs. S10–S12). In comparison, tetrahydrofuran (THF) was tolerated equally well by both the parent protein and the stapled variants (entry 9, Table 2). In

Table 2. Catalytic activity and selectivity of Mb(H64V,V68A) and its stapled variants for cyclopropanation of styrene (1a) and pentafluorostyrene (1b) with ethyl 2-diazo-acetate (2) in the absence and in the presence of organic cosolvents
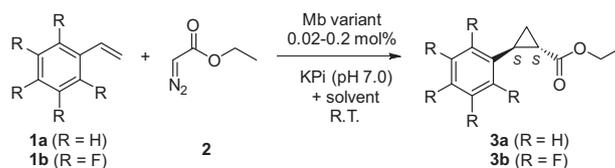
Entry	Variant	Product	Solvent	Yield, %	TON	% de	% ee
1	Mb(H64V,V68A)	3a	—	94	4,710 ± 180	99.4	99.1
2	sMb2	3a	—	94	4,740 ± 100	99.3	99.2
3	sMb5	3a	—	94	4,690 ± 25	99.1	99.1
4	sMb10	3a	—	94	4,710 ± 250	99.4	99.3
5	sMb13	3a	—	92	4,610 ± 140	99.2	99.2
6	Mb(H64V,V68A)	3a	30% EtOH	48	2,420 ± 280	98	92
7	sMb10	3a	30% EtOH	82	4,120 ± 270	99.3	97
8*	sMb10	3a	30% DMF	84 (64)	4,220 ± 140 (3,480)	99.8 (99)	99.6 (99)
9*	sMb13	3a	30% THF	81 (92)	4,360 ± 90 (4,660)	97 (99)	95 (94)
10*	sMb10	3a	30% DMSO	64 (60)	3,220 ± 280 (3,000)	99.8 (99.8)	99.6 (99.4)
11*	sMb10	3a	45% DMSO	39 (20)	1,990 ± 190 (1,020)	99.9 (99.9)	98.5 (98.5)
12	Mb(H64V,V68A)	3b	45% DMSO	54	270 ± 5	99.8	99.9
13	sMb10	3b	45% DMSO	80	395 ± 5	99.8	99.9
14	sMb13	3b	45% DMSO	77	380 ± 5	99.8	99.9

Reaction conditions for **3a**: 2 μM protein (0.02 mol%), 10 mM **1a**, 20 mM EDA (**2**), 10 mM sodium dithionite in 50 mM potassium phosphate (pH 7), room temperature. Reaction conditions for **3b**: 10 μM protein (0.2 mol%), 5 mM **1b**, 10 mM EDA (**2**), 10 mM sodium dithionite in 50 mM potassium phosphate (pH 7), room temperature.
*Values in brackets correspond to reactions performed in the presence of Mb(H64V,V68A) as the catalyst.

the presence of these organic cosolvents, both sMb10 and sMb13 are able to produce **3a** in high yields (64–88%; 3,200–4,440 TON) and high stereoselectivity (99.5–99.8% *de*; 99.2–99.6% *ee*) (entries 8–10, Table 2). Furthermore, sMb10 and sMb13 maintain good catalytic activity (1,080–1,990 TON) as well as excellent stereoselectivity (>99% *de*; 97–98.5% *ee*) even in the presence of a nearly 1:1 mixture (45% vol/vol) of buffer with DMF or DMSO (entry 11, Table 2 and *SI Appendix*, Figs. S11 and S12). These results are remarkable considering that most heme-dependent enzymes are readily inactivated by low concentrations (>5–10% vol/vol) of these organic solvents (46, 47). Given their enhanced performance at high DMSO concentrations, sMb10 and sMb13 could be applied to afford the cyclopropanation of a water-insoluble substrate, that is, pentafluorostyrene (**1b**), with higher efficiency than possible using the parent Mb variant (Table 2, entries 13 and 14 vs. 12). Altogether, these results demonstrate the value of R-GPS toward enhancing the robustness and performance of a biocatalyst in the presence of chemical denaturants and organic cosolvents.

Discussion

As demonstrated through the design and characterization of highly stabilized variants of an Mb-based cyclopropanase, R-GPS enables the rapid identification of optimal sites within a target protein scaffold for structural stabilization via "stapling" through nonnative covalent bridges. In its current implementation, R-GPS utilizes a bioorthogonal reaction between cysteine and the genetically encodable noncanonical amino acid O2beY to produce chemically stable and nonreducible thioether cross-links. Key advantages of this chemistry include its chemoselectivity, spatially controlled reactivity, and efficient and spontaneous formation of the cross-link at the posttranslational level and in the intracellular space. The latter feature allows for the production of the stapled proteins inside cells and eliminates the need for protein manipulation during or after purification (48). Whereas R-GPS has been validated using the O2beY/Cys cross-linking method, we expect this approach can be extended to other noncanonical amino acids and/or other chemistries useful for protein cross-linking (49–52).

Two computationally designed staples could be effectively combined in the final designs (sMb10 and sMb13) to achieve additive effects for stabilization against thermal and chemical denaturation. While the introduction of computationally designed nonnative disulfide bridges may result in misfolded variants due to disulfide scrambling (27), no evidence of incorrect cross-linking was noted for the sMb variants containing the two thioether staples. We attribute this result to the heterotypic nature of the O2beY/Cys thioether linkage, which reduces the number of wrong cross-links that can be formed compared with the homotypic disulfide bridge. In addition, the presence of specific geometric requirements for successful formation of the O2beY/Cys cross-link, as revealed by the present studies (*SI Appendix*, Fig. S1), likely contributes to further disfavor the formation of incorrect linkages. At the same time, a noticeable reduction in the protein yields was observed upon incorporation of multiple copies of O2beY via amber stop codon suppression [2 mg/L culture for sMb10 compared with 10–15 mg/L for sMb2/sMb5 and 30 mg/L for Mb(H64V,V68A)]. While this result is not surprising given the less-efficient incorporation of noncanonical amino acids (ncAAs) compared with natural amino acids, various strategies have recently enabled the expression of proteins containing multiple ncAAs (53–55) and these protocols should increase the accessibility of proteins containing multiple staples designed with the present methodology.

R-GPS was able to identify a small set of structural modifications (i.e., four and five mutations for sMb10 and sMb13, respectively) capable of conferring these carbene transferase enzymes significantly increased stability against thermal and chemical denaturation without impairing their catalytic activity and subtle functional properties such as stereoselectivity. The improved stability properties of these biocatalysts translated into improved performance and stereoselectivity in reactions conducted in high concentrations of organic solvents, also enabling the more efficient transformation of a water-insoluble olefin substrate. Since R-GPS involves a small set of amino acid substitutions, most of the protein residues remain available for stabilization based on optimization of interactions in the native state. Thus, it should be possible to combine R-GPS with other computational design (13, 14, 17) and/or experimental

methods (4–6) to achieve additive or synergistic effects in protein/enzyme stabilization. The minimally invasive approach to protein stabilization described here is expected to provide a valuable strategy for greatly enhancing the stability of enzymes and proteins without impacting their functional properties.

Materials and Methods

See *SI Appendix* for detailed descriptions of computational methods and experimental procedures and for additional data: NAC analyses (*SI Appendix*, Figs. S1 and S2 and Table S1); UV-visible and CD spectra (*SI Appendix*, Figs. S4 and S8), protein gels (*SI Appendix*, Fig. S5), MS spectra (*SI Appendix*, Fig. S6), thermal and chemical denaturation curves (*SI Appendix*, Figs. S7 and S9), and reaction characterization data (*SI Appendix*, Figs. S11–S13).

1. Bornscheuer UT, et al. (2012) Engineering the third wave of biocatalysis. *Nature* 485: 185–194.
2. Clouthier CM, Pelletier JN (2012) Expanding the organic toolbox: A guide to integrating biocatalysis in synthesis. *Chem Soc Rev* 41:1585–1605.
3. Bloom JD, Labthavikul ST, Otey CR, Arnold FH (2006) Protein stability promotes evolvability. *Proc Natl Acad Sci USA* 103:5869–5874.
4. Gershenson A, Arnold FH (2000) Enzyme stabilization by directed evolution. *Genet Eng (N Y)* 22:55–76.
5. Lehmann M, Wyss M (2001) Engineering proteins for thermostability: The use of sequence alignments versus rational design and directed evolution. *Curr Opin Biotechnol* 12:371–375.
6. Bommarius AS, Paye MF (2013) Stabilizing biocatalysts. *Chem Soc Rev* 42:6534–6565.
7. Tracewell CA, Arnold FH (2009) Directed enzyme evolution: Climbing fitness peaks one amino acid at a time. *Curr Opin Chem Biol* 13:3–9.
8. Magliery TJ (2015) Protein stability: Computation, sequence statistics, and new experimental methods. *Curr Opin Struct Biol* 33:161–168.
9. Malakauskas SM, Mayo SL (1998) Design, structure and stability of a hyperthermophilic protein variant. *Nat Struct Biol* 5:470–475.
10. Korkegian A, Black ME, Baker D, Stoddard BL (2005) Computational thermostabilization of an enzyme. *Science* 308:857–860.
11. Borgo B, Havranek JJ (2012) Automated selection of stabilizing mutations in designed and natural proteins. *Proc Natl Acad Sci USA* 109:1494–1499.
12. Li Y, et al. (2007) A diverse family of thermostable cytochrome P450s created by recombination of stabilizing fragments. *Nat Biotechnol* 25:1051–1056.
13. Goldenzweig A, et al. (2016) Automated structure- and sequence-based design of proteins for high bacterial expression and stability. *Mol Cell* 63:337–346.
14. Bednar D, et al. (2015) FireProt: Energy- and evolution-based computational design of thermostable multiple-point mutants. *PLOS Comput Biol* 11:e1004556.
15. Gribenko AV, et al. (2009) Rational stabilization of enzymes by computational redesign of surface charge-charge interactions. *Proc Natl Acad Sci USA* 106:2601–2606.
16. Bjørk A, Dalhus B, Mantzilas D, Sirevåg R, Eijsink VG (2004) Large improvement in the thermal stability of a tetrameric malate dehydrogenase by single point mutations at the dimer-dimer interface. *J Mol Biol* 341:1215–1226.
17. Wijma HJ, et al. (2014) Computationally designed libraries for rapid enzyme stabilization. *Protein Eng Des Sel* 27:49–58.
18. Van den Burg B, Vriend G, Veltman OR, Venema G, Eijsink VG (1998) Engineering an enzyme to resist boiling. *Proc Natl Acad Sci USA* 95:2056–2060.
19. Zou J, Song B, Simmerling C, Raleigh D (2016) Experimental and computational analysis of protein stabilization by Gly-to-d-Ala substitution: A convolution of native state and unfolded state effects. *J Am Chem Soc* 138:15682–15689.
20. Dombkowski AA, Sultana KZ, Craig DB (2014) Protein disulfide engineering. *FEBS Lett* 588:206–212.
21. Matsumura M, Signor G, Matthews BW (1989) Substantial increase of protein stability by multiple disulphide bonds. *Nature* 342:291–293.
22. Liu T, et al. (2016) Enhancing protein stability with extended disulfide bonds. *Proc Natl Acad Sci USA* 113:5910–5915.
23. Sanchez-Romero I, et al. (2013) Mechanism of protein kinetic stabilization by engineered disulfide crosslinks. *PLoS One* 8:e70013.
24. Tyagi V, Bonn RB, Fasan R (2015) Intermolecular carbene S-H insertion catalysed by engineered myoglobin-based catalysts. *Chem Sci (Camb)* 6:2488–2494.
25. Carninci P, et al. (1998) Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. *Proc Natl Acad Sci USA* 95:520–524.
26. Tyagi V, Fasan R (2016) Myoglobin-catalyzed olefination of aldehydes. *Angew Chem Int Ed Engl* 55:2512–2516.
27. Tanghe M, et al. (2017) Disulfide bridges as essential elements for the thermostability of lytic polysaccharide monooxygenase LPMO10C from Streptomyces coelicolor. *Protein Eng Des Sel* 30:401–408.
28. Bionda N, Cryan AL, Fasan R (2014) Bioinspired strategy for the ribosomal synthesis of thioether-bridged macrocyclic peptides in bacteria. *ACS Chem Biol* 9:2008–2013.
29. Bordeaux M, Tyagi V, Fasan R (2015) Highly diastereoselective and enantioselective olefin cyclopropanation using engineered myoglobin-based catalysts. *Angew Chem Int Ed Engl* 54:1744–1748.
30. Bajaj P, Sreenilayam G, Tyagi V, Fasan R (2016) Gram-scale synthesis of chiral cyclopropane-containing drugs and drug precursors with engineered myoglobin catalysts featuring complementary stereoselectivity. *Angew Chem Int Ed Engl* 55:16110–16114.
31. Tyagi V, Sreenilayam G, Bajaj P, Tinoco A, Fasan R (2016) Biocatalytic synthesis of allylic and allenyl sulfides through a myoglobin-catalyzed Doyle-Kirmse reaction. *Angew Chem Int Ed Engl* 55:13562–13566.
32. Bionda N, Fasan R (2015) Ribosomal synthesis of natural-product-like bicyclic peptides in Escherichia coli. *ChemBioChem* 16:2011–2016.
33. Betz SF (1993) Disulfide bonds and the stability of globular-proteins. *Protein Sci* 2:1551–1558.
34. Zhang T, Bertelsen E, Alber T (1994) Entropic effects of disulphide bonds on protein stability. *Nat Struct Biol* 1:434–438.
35. Flory PJ (1956) Theory of elastic mechanisms in fibrous proteins. *J Am Chem Soc* 78:5222–5234.
36. Chan HS, Dill KA (1989) Intrachain loops in polymers–Effects of excluded volume. *J Chem Phys* 90:492–509.
37. Zanghellini A, et al. (2006) New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci* 15:2785–2794.
38. Kuhlman B, Baker D (2000) Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci USA* 97:10383–10388, and erratum (2000) 97:13460.
39. Liu CC, Schultz PG (2010) Adding new chemistries to the genetic code. *Annu Rev Biochem* 79:413–444.
40. Deiters A, Schultz PG (2005) In vivo incorporation of an alkyne into proteins in Escherichia coli. *Bioorg Med Chem Lett* 15:1521–1524.
41. Nagatani RA, Gonzalez A, Shoichet BK, Brinen LS, Babbitt PC (2007) Stability for function trade-offs in the enolase superfamily "catalytic module". *Biochemistry* 46:6688–6695.
42. Hamamatsu N, et al. (2006) Directed evolution by accumulating tailored mutations: Thermostabilization of lactate oxidase with less trade-off with catalytic activity. *Protein Eng Des Sel* 19:483–489.
43. Zhong CQ, et al. (2009) Improvement of low-temperature caseinolytic activity of a thermophilic subtilase by directed evolution and site-directed mutagenesis. *Biotechnol Bioeng* 104:862–870.
44. Savile CK, et al. (2010) Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* 329:305–309.
45. Fox RJ, et al. (2007) Improving catalytic function by ProSAR-driven enzyme evolution. *Nat Biotechnol* 25:338–344.
46. Wong TS, Arnold FH, Schwaneberg U (2004) Laboratory evolution of cytochrome p450 BM-3 monooxygenase for organic cosolvents. *Biotechnol Bioeng* 85: 351–358.
47. Chauret N, Gauthier A, Nicoll-Griffith DA (1998) Effect of common organic solvents on in vitro cytochrome P450-mediated metabolic activities in human liver microsomes. *Drug Metab Dispos* 26:1–4.
48. Abdeljabbar DM, Piscotta FJ, Zhang S, James Link A (2014) Protein stapling via azide-alkyne ligation. *Chem Commun (Camb)* 50:14900–14903.
49. Furman JL, et al. (2014) A genetically encoded aza-Michael acceptor for covalent cross-linking of protein-receptor complexes. *J Am Chem Soc* 136:8411–8417.
50. Xiang Z, et al. (2014) Proximity-enabled protein crosslinking through genetically encoding haloalkane unnatural amino acids. *Angew Chem Int Ed Engl* 53:2190–2193.
51. Chen XH, et al. (2014) Genetically encoding an electrophilic amino acid for protein stapling and covalent binding to native receptors. *ACS Chem Biol* 9:1956–1961.
52. Xuan W, Shao S, Schultz PG (2017) Protein crosslinking by genetically encoded noncanonical amino acids with reactive aryl carbamate side chains. *Angew Chem Int Ed Engl* 56:5096–5100.
53. Guo J, Melançon CE, 3rd, Lee HS, Groff D, Schultz PG (2009) Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids. *Angew Chem Int Ed Engl* 48:9148–9151.
54. Johnson DBF, et al. (2011) RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat Chem Biol* 7:779–786.
55. Lajoie MJ, et al. (2013) Genomically recoded organisms expand biological functions. *Science* 342:357–360.

BIOPHYSICS AND COMPUTATIONAL BIOLOGY

CHEMISTRY