

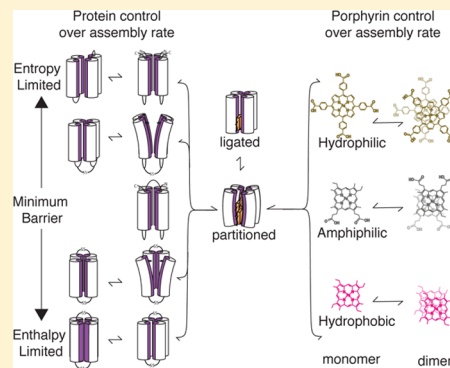
# Engineering the Assembly of Heme Cofactors in Man-Made Proteins

Lee A. Solomon, Goutham Kodali, Christopher C. Moser, and P. Leslie Dutton\*

The Johnson Research Foundation, Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, United States

## S Supporting Information

**ABSTRACT:** Timely ligation of one or more chemical cofactors at preselected locations in proteins is a critical preamble for catalysis in many natural enzymes, including the oxidoreductases and allied transport and signaling proteins. Likewise, ligation strategies must be directly addressed when designing oxidoreductase and molecular transport functions in man-made, first-principle protein constructs intended to operate in vitro or in vivo. As one of the most common catalytic cofactors in biology, we have chosen heme B, along with its chemical analogues, to determine the kinetics and barriers to cofactor incorporation and bishistidine ligation in a range of 4- $\alpha$ -helix proteins. We compare five elementary synthetic designs (maquettes) and the natural cytochrome  $b_{562}$  that differ in oligomeric forms, apo- and holo-tertiary structural stability; qualities that we show can either assist or hinder assembly. The cofactor itself also imposes an assembly barrier if amphiphilicity ranges toward too hydrophobic or hydrophilic. With progressive removal of identified barriers, we achieve maquette assembly rates as fast as native cytochrome  $b_{562}$ , paving the way to in vivo assembly of man-made hemoprotein maquettes and integration of artificial proteins into enzymatic pathways.



## INTRODUCTION

One ambition of synthetic biology is the creation of man-made enzymes with prescribed physical characteristics and performance specifications engineered for work in vitro and eventually in living cells. As with any practical machine, a natural enzyme molecule comprises a consortium of component parts reflecting production, assembly, and service engineering that supports the functionally active parts. The sharing of residues, structural motifs, and domains by these parts contributes to the complexity of natural proteins. This complexity is amplified by the unplanned modifications from repeated blind natural selection over evolutionary time, which confounds reliable extraction of essential elements of natural enzyme activity for translation into man-made reproductions. Our approach to the creation of man-made oxidoreductases and related proteins thus seeks to minimize complexity by combining first principles established for protein folding with those for engineering electron-transfer and oxidation–reduction systems.<sup>1</sup> We avoid the common practice of designing man-made proteins by mimicking natural proteins.

Our method to functionalize oxidoreductase maquettes endeavors to keep the cofactor assembly process much simpler than is evident in natural systems.<sup>2,3</sup> In the cell, the import and ligation of the large, complex Mg/Zn and Fe tetrapyrrole structures of chlorins and hemes, respectively, of photosynthesis, respiration, and oxidative and reductive metabolism employ elaborate engineering to regulate tetrapyrrole production and delivery into the apoprotein.<sup>4–9</sup> In the laboratory, the analogous in vitro equipping of man-made apoprotein designs with tetrapyrroles has to date been a matter of empirically

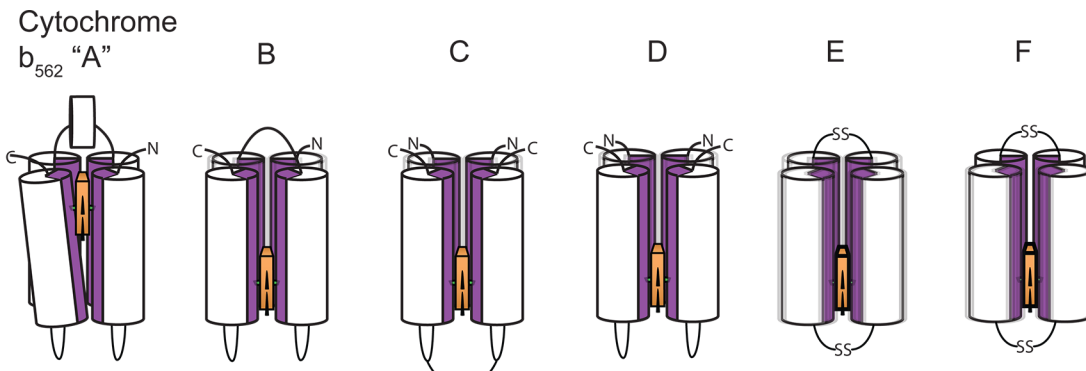
finding satisfactory conditions of mixing tetrapyrroles in organic solution with the apoprotein in aqueous media with minor attention to the essential details of the assembly process.<sup>10–12</sup> As maquettes, man-made 4- $\alpha$ -helical proteins designed with simple oxidoreductase functions that match key characteristics of their natural counterparts,<sup>13</sup> continue to advance to include multi-cofactor combinations and more sophisticated catalysis, it becomes important to learn how the assembly of tetrapyrrole and protein proceeds and to develop strategies to ensure competent integration of all cofactors into a designed host protein.

The staged, step-by-step developmental approach toward construction of heme maquettes displaying oxidoreductive activities has generated an extensive family of related 4- $\alpha$ -helical proteins.<sup>14</sup> Five of these are selected for their markedly different oligomeric and topological forms and the degree of structuring of their apo and holo heme-ligated states (Table 1). Together with the bacterial 4- $\alpha$ -helix hemoprotein cytochrome  $b_{562}$ ,<sup>15,16</sup> they offer a useful palette with which to determine the time scales and barriers governing tetrapyrrole partition from aqueous medium into the hydrophobic interior of the protein followed by positioning at a preselected location by bis-His ligation. In the present work, we begin comparing tetrapyrrole binding using the common natural cofactor Fe-protoporphyrin IX (heme B), which binds to all the selected proteins with similar nanomolar affinities (dissociation constants,  $K_D$ ).<sup>1,13,17,18</sup> The study is then extended beyond heme B to

Received: November 27, 2013

Published: February 4, 2014

Table 1. Topology and Physical Properties of the Six Heme Binding Protein Sequences<sup>a</sup>

	Cytochrome b <sub>562</sub> "A"	B	C	D	E	F
						
K <sub>D</sub> , nM (ΔG, kcal mol <sup>-1</sup> )	6 (-11.2)	<1 (-12.2)	<1 (-12.2)	<50 (-9.9)	15 (-10.6)	2 (-11.8)
Alpha Helix (%)	78	80	85	85	87	87
Helical Constraints	6	6	4	4	4	4
Syn/Anti Assembly	No	No	No	Apo-Only	Yes	Yes
Subunit Dissociation	No	No	No	Yes	Minor	Minor
Apo Structure	Partial	No	No	No	No	Yes
Holo Structure	Yes	Yes	Yes	Yes	No	No

<sup>a</sup>Sequences are shown in Supplementary Table 1. Heme dissociation values (K<sub>D</sub>) from previous works.<sup>1,13,17,18</sup> Absence of structure refers to NMR characterization of a molten globular state for the five maquettes and a random coil for the natural cytochrome b<sub>562</sub>.<sup>1,18,25–27</sup>

examine time scales of binding of a series of related heme analogues differing in peripheral substitutions and hydrophobicity to clarify the influence of cofactor physical chemistry on the assembly process.

## MATERIALS AND METHODS

**Supplies.** All chemicals were purchased from Sigma Aldrich unless otherwise noted.

**Heme Preparation.** Iron protoporphyrin IX (hemin) and the other porphyrins were purchased from Frontier Scientific. 2,6-Dinitriroporphyrin was synthesized as described.<sup>19</sup> Heme A was extracted and purified as described.<sup>20</sup> All porphyrins were dissolved in dimethyl sulfoxide (DMSO) with the concentration determined via weight, except for hemin whose concentration was determined via the hemochrome assay utilizing the extinction coefficient at 556 nm.<sup>21</sup> Each porphyrin was diluted into the appropriate aqueous buffer minutes before each experiment.

**Protein Preparation.** Proteins E and F were prepared at 0.1 mmol scale on a CEM Liberty microwave peptide synthesizer using standard fluorenylmethyloxycarbonyl (Fmoc)/*t*Bu protection protocols.<sup>22</sup> Amino acids were purchased from Nova Biochem. The side chain protecting groups were as follows: Cys (trt), Lys (Boc), His (Boc), and Asp/Glu (OtBu). After synthesis, the protein was cleaved from the resin by incubating it with a mixture of trifluoroacetic acid (TFA), ethanedithiol, anisole, and thioanisole in a 9:0.2:0.5:0.3 ratio for 2.5 h protected from light. After excess reagent removal via rotovap, protein was precipitated with methyl *t*-butyl ether and purified using a reverse-phase Waters HPLC equipped with a C18 column running a linear gradient of acetonitrile with 0.1% (v/v) TFA and water with 0.1% (v/v) TFA. The products were verified with MALDI mass spectroscopy using a sinapinic acid matrix.

The other proteins used in this paper were obtained as follows. The genes for the protein were ordered from DNA2.0 containing a His-tag, a linker with a TEV site, and the protein in PJ414 vector. DNA was transformed into BL-21DE3 strain *Escherichia coli* cells (New England Biolabs). These BL-21DE3 cells were grown in 4 L of media (12 g/L bacto-peptone, 24 g/L bacto-yeast extract, 4 mL/L glycerol, and 14 g/L KH<sub>2</sub>PO<sub>4</sub>, pH 7.5) until the OD<sub>600</sub> was at 0.6 au. Overexpression was

induced by addition of IPTG to a final concentration of 1 mM. The cells overexpressed the protein for 4.5 h and were spun down to a pellet. The pellet was then resuspended in lysis buffer (300 mM NaCl, 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 10 mM imidazole, pH 7) and sonicated at an amplitude of 90 (5 times, 20 s each time) in a Misonix sonicator (S-4000 with microtip attachment). The lysate was spun down for 1 h at 20 000g. The supernatant was then run through a GE-Histrap column. The histidine tag was removed by incubating the eluate with TEV protease (1 mM dithiothreitol, 1 unit TEV enzyme purchased from Invitrogen per 100 μL volume, overnight) and rerun on the Histrap column to separate cleaved from uncleaved. The flowthrough of this second column was dialyzed into *N*-cyclohexyl-2-aminoethanesulfonic acid (CHES) buffer (20 mM CHES, 150 mM KCl, pH 9).

The concentration for all protein variants was determined by the monitoring absorbance at 280 nm using an extinction coefficient of 22 500 M<sup>-1</sup> cm<sup>-1</sup>.

All proteins in this paper were purified by HPLC, and the masses were verified by MALDI as described above (Supplemental Figures 1–12).

**Stopped-Flow Spectroscopy.** Millisecond time scale measurements were made on an OLIS RSM 1000 stopped flow spectrophotometer, which takes a full visible spectrum every millisecond. The tetrapyrrole and protein were added to separate syringes and shot together, and the absorbance was monitored in a 2 cm flow cell. Temperature was controlled with a Fischer-Scientific IsoTemp 3031 water bath attached. Individual wavelengths were selected for further kinetic analysis. All experiments were performed in triplicate, and the data were averaged together for further analysis.

**Data Fitting.** Initial rates were determined by fitting the first 20 ms to a linear regression and dividing this slope by the extinction coefficient.

Pseudothermodynamic parameters of the transition state were derived using eq 1 for the temperature-dependent rates.

$$k = \kappa \frac{k_B T}{h} e^{-\Delta H^\ddagger / RT} e^{\Delta S^\ddagger / R} \quad (1)$$

where  $k$  is the rate of the reaction at a certain temperature,  $k_B$  is Boltzmann's constant,  $T$  is the temperature in Kelvin, and  $h$  is Planck's

constant;  $\kappa$  is the transmission coefficient, which is set to 1 for all reactions in this paper, indicating all reactions that reach the transition state proceed to completion.  $\Delta S^\ddagger$  is the entropy change between the transition state and the reactants,  $\Delta H^\ddagger$  is the enthalpy change between the transition state and the reactants, and  $R$  is the gas constant in units of  $\text{cal K}^{-1} \text{mol}^{-1}$ .<sup>23</sup> The linearized form (eq 2) is used in the Eyring plot of Figure 2:

$$\ln \frac{k}{T} = \frac{-\Delta H^\ddagger}{RT} + \ln \frac{k_B}{h} + \frac{\Delta S^\ddagger}{R} \quad (2)$$

**UV/Vis Spectroscopy.** A Varian Cary-500 spectrophotometer measured slower than millisecond kinetics. Data were collected between 200 and 800 nm. The spectrophotometer was blanked to the buffer used in the experiment.

**Partition Coefficients.** Partition coefficients were determined by first dissolving the chromophore in 1-octanol and taking a spectrum.<sup>24</sup> This solution was then mixed with an equal volume 50 mM CHES, 150 mM KCl buffer, pH 9, vortexed, and allowed to sit for 1 h at room temperature. After the incubation, spectra of the octanol layer were taken again. The concentration of porphyrin in the aqueous layer was calculated as the concentration in octanol before incubation minus the concentration in octanol after incubation. The partition coefficient ( $\log P$ ) was then determined as the log of the ratio of the concentration of porphyrin in the octanol layer to the aqueous layer.

## RESULTS

### Heme B Binding to the Selected Maquette Proteins.

There are many structural factors that can govern speed of cofactor assembly in proteins. In this work, we focus on the effect of structural elements at the higher tertiary and quaternary levels without changing the lower primary and secondary structural elements. Specifically, our maquette selections maintain the same positioning of bis-His heme ligation sites and the similarity of surrounding amino acids (see Supporting Information Table 1). The  $\alpha$ -helical secondary structure content of maquettes remains similar and comparable to  $b_{562}$  (see Table 1). Differences at the tertiary and quaternary levels for our set of six selected comparison sequences are illustrated in Table 1. The tertiary structure is altered by (a) helical constraint via connecting loops, (b) the ability or inability of helices to orient parallel versus antiparallel (syn vs anti conformations), and (c) the relative mobility or rigidity of interior amino acids packing between helices in the apo state, as assayed by NMR and X-ray crystallography.<sup>1,18,25–27</sup> To test for the effects of differences in quaternary structure, we vary the nature of the connection of the helical pair units, that is, whether or not they are linked to other helical pair units through covalent loops and/or disulfide bonds and whether helical pair units associate tightly or relatively loosely in the apo state.

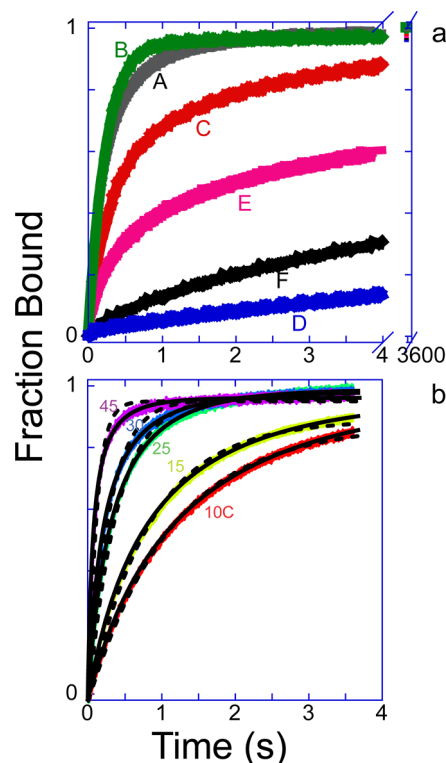
Specifically, maquettes B, C, and D have identical helical sequences but differ in the number and position of interhelix loops and the presence or absence of a disulfide link between loops. Maquettes D, E, and F are unlinked and capable of dissociating into separate helical pair subunits and associating in two distinct syn or anti forms based on the relative positions of the connecting loops to one another (Table 1) though remain as 4-helix units when in the apo state.<sup>28,29</sup> In this regard, maquette D has the greatest tendency to dissociate under the conditions used here.<sup>30</sup>

Table 1 shows that, despite the differences in tertiary and quaternary structure evident in the set of proteins, all sequences display  $K_D$  values for heme binding in the tight nanomolar range, which makes the binding effectively irreversible.<sup>1,31,32</sup> The heme binding reaction is monitored by a characteristic

spectral shift of the ferric heme Soret absorption from 395 nm in the aqueous phase to 412 nm on bis-His ligation. The heme binding process is straightforward and the heme-bound state stable beyond 20 h.

**Time Course of Heme B Binding to Maquettes and Cytochrome  $b_{562}$ .** Under all experimental conditions used, the assembly process can be regarded as a movement of the heme B from the aqueous phase into the apoprotein and ending after the heme has found and bis-His-ligated at the site.

Figure 1a presents time courses of heme B ligation to cytochrome  $b_{562}$  (A) and maquettes B–F under comparable



**Figure 1.** (a) Heme binding kinetics at 412 nm for proteins A–F. All data were collected at 30 °C in 20 mM CHES, 150 mM buffer, pH 9. Protein at 5  $\mu\text{M}$ , heme at 10  $\mu\text{M}$ . Data normalized to final overnight absorbance to ensure completion of the reaction. (b) Heme binding of maquette A at various temperatures fit to both a first-order (dashed lines) and second-order (solid lines) kinetic time course.

conditions of one heme B per ligation site. Rate constants at 30 °C (Figure 1) are summarized in Table 2. Heme B ligation to the six proteins covers a broad range of time, proceeding to completion following time courses characteristic of a second-order (Figure 1b), but not a first-order, reaction with equimolar reactants (Supplementary Figures 13 and 14).<sup>33</sup> The linear dependence of the initial rate of binding for 20  $\mu\text{M}$  excess heme combined with variable protein concentrations of 1, 2, 3, and 4  $\mu\text{M}$  confirmed a second-order process in all cases except for the untethered D, which was ambiguous (Supplementary Figure 15). Similarly, the linear dependence of the initial rate for 30  $\mu\text{M}$  excess maquette B protein with variable heme B concentrations of 6, 9, 12, and 18  $\mu\text{M}$  confirmed a second-order process (Supplementary Figure 16). The single-chain monomer maquette B, unstructured in the apo state and structured in the holo state,<sup>1</sup> displays the most rapid time course of heme ligation, proceeding to completion on a subsecond time scale with a second-order rate constant of 3.04



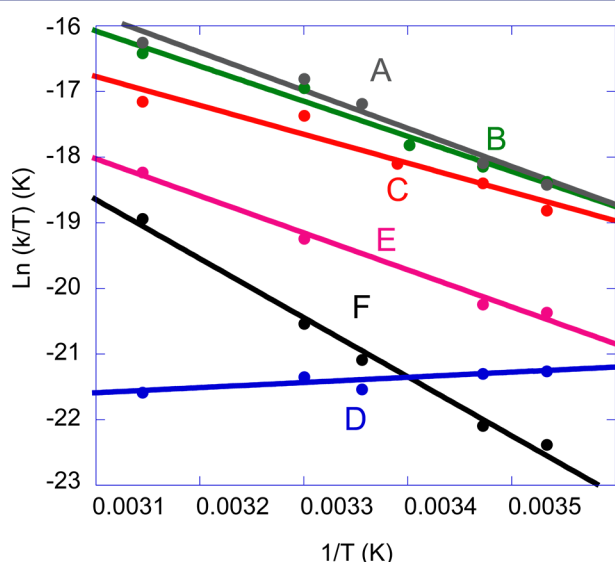
Table 2. Binding Rate Constants and Transition State Parameters of Sequences A–F<sup>a</sup>

protein	rate constant at 303 K (M <sup>-1</sup> s <sup>-1</sup> )	$\Delta H^\ddagger$ (kcal/mol)	$\Delta S^\ddagger$ (cal/mol K)	$-T\Delta S^\ddagger$ (kcal/mol)	$\Delta G^\ddagger$ (kcal/mol)
A	$3.04 \times 10^5 \pm 1.7 \times 10^4$	$11.53 \pm 1.1$	$-42.89 \pm 3.8$	<b>13.0</b>	24.5
B	$3.04 \times 10^5 \pm 6.4 \times 10^3$	$10.62 \pm 1.0$	$-46.22 \pm 3.4$	<b>14.0</b>	24.6
C	$1.73 \times 10^5 \pm 3.3 \times 10^3$	$8.73 \pm 1.3$	$-53.44 \pm 4.6$	<b>16.2</b>	24.9
D	$3.24 \times 10^3 \pm 5.5 \times 10^2$	$-1.56 \pm 0.6$	$-94.94 \pm 2.1$	<b>28.8</b>	27.2
E	$2.68 \times 10^4 \pm 2.2 \times 10^3$	$11.2 \pm 0.8$	$-48.3 \pm 2.8$	<b>14.6</b>	25.8
F	$7.3 \times 10^3 \pm 1.8 \times 10^3$	<b>17.87 <math>\pm</math> 1.0</b>	$-28.85 \pm 3.5$	8.74	26.6

<sup>a</sup>Reference temperature is 30 °C. The bold/italic numbers denote the predominant barrier for each protein.

$\times 10^5 \text{ M}^{-1} \text{ s}^{-1}$  (see Table 2 and Figure 1), the same rate as natural protein A. Rates for binding heme at different sites are equivalent for this maquette, as well (Supplemental Figure 17).

Compared to natural protein A, the temperature dependence for heme binding to single-chain maquette B (presented in Eyring plots of Figure 2) balances a slightly smaller enthalpic



**Figure 2.** Eyring plot of the temperature dependence of the initial rates of the various proteins studied. A (gray), B (green), C (red), and E (pink) share similar slopes and intercepts of their temperature dependence. F (black) has a higher transition state enthalpy of formation and D (blue) a higher transition state entropy of formation: 20 mM CHES, 150 mM KCl buffer, pH 9; protein at 5  $\mu\text{M}$ , heme at 10  $\mu\text{M}$ .

(10.6 vs 11.5 kcal/mol) with a slightly larger entropic (14.0 vs 13.0 kcal/mol at 30 °C) component. Despite markedly different primary sequences, natural sequence A and maquette B appear to present a common barrier height to heme entry, ligation, and establishment of a singular structure.

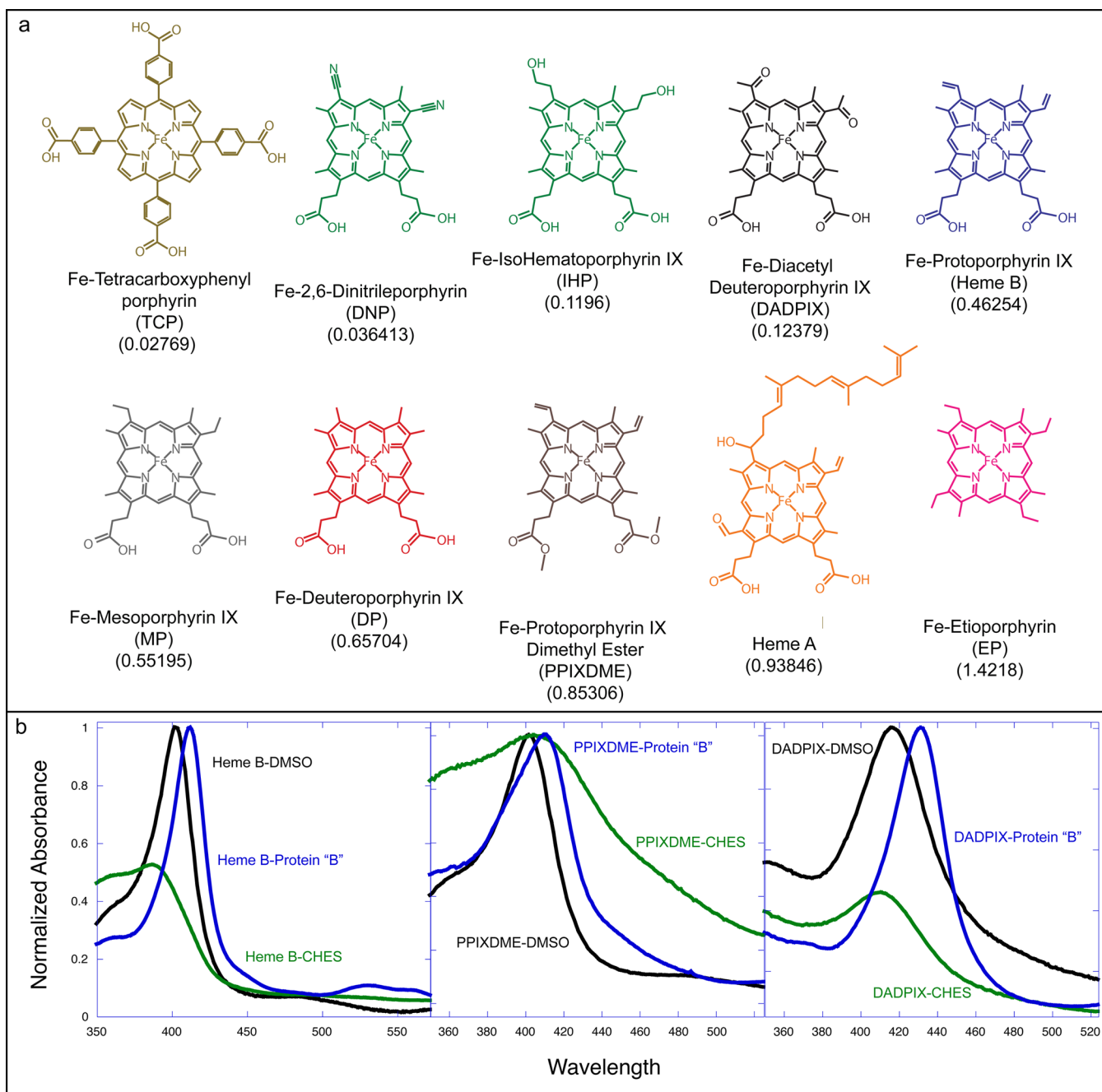
Dimeric maquette C shares the same sequence as monomeric maquette B and also assumes a singular structure in the holo state but has a different tertiary topology of helical tethering that involves removal of the loop connecting helices 2 and with the addition of a disulfide bond connecting the loops at the opposite end of the bundle (Table 1). The maquette C has a more exaggerated enthalpic and entropic counterbalancing (8.7 and 16.2 kcal/mol) with a net effect of only a 2-fold slowing of the second-order rate constant. The enhanced mobility at one end of the helical bundle in C appears to increase the entropic term while lowering the activation enthalpy. Individual site knockouts of protein C unveil peculiar kinetics (Supplemental Figure 17). The site farthest away from the tether (H42F)

assembles at a rate 3 times faster than the site nearest the tether (H7F). This could also be due to a combination of cofactor access to the individual sites and differences in the level of apo structure of maquette C.<sup>25</sup> This also manifests in the affinity, showing two separate  $K_D$  values.<sup>13</sup>

The simple removal of the disulfide tether in C to create the untethered helix–loop–helix homodimer D substantially lowers the second-order rate constant by 50-fold. The Eyring plot reveals a dramatically different rate-limiting barrier for the dominant slow phase. Figure 2 and Table 2 show that the heme B ligation process for D slows slightly with increasing temperature, yielding a modest enthalpic term of  $-1.54$  kcal/mol, with a dominating entropic term (28.8 kcal/mol). This suggests a distinctly different rate-limiting process that may be responding to the freedom of the apo-maquette helical pairs to pack parallel or antiparallel in a syn or anti conformation and/or the repelling effect of the high net charge on each of the homodimers ( $-8$  charge per one helix–loop–helix motif).

The untethered homodimer E with neutral net charge has a similar enthalpic but larger entropic contribution than natural sequence A. This slows the second-order rate constant  $\sim 10$ -fold. Although also net-neutral charge, the untethered homodimer F, the only sequence that is structured (X-ray<sup>29</sup> and NMR<sup>18</sup>) in the apo form, reverses the thermodynamic trend, raising the enthalpic term (17.9 kcal/mol) and lowering the entropic term (8.7 kcal/mol), for a net slowing of the binding rate by  $\sim 40$ -fold. These results separate D and F from the others as clearly rate-limited, respectively, by diffusion between heme B and the weakly coupled homodimers or a higher thermal activation barrier to access the heme into a singular structured apoprotein.

**Ligation of Fe-Porphyrin Variants.** In complementary trials, we kept the maquette sequence constant (B) but varied the Fe-porphyrin to test whether the transfer of cofactor from the aqueous phase into the maquette interior limits the rapid time course of ligation and to assess the effect of porphyrin aggregation in aqueous solution. Figure 3a presents a family of synthetic Fe-porphyrins with marked variance in their overall degree of hydrophobicity or hydrophilicity and in the pattern of polar and nonpolar groups around the porphyrin macrocycle (affinity data in Supplemental Table 2). We used  $[n\text{-octanol}]/[water]$  partition coefficients  $P^{24,34}$  as a guide to the effect of the porphyrin macrocycle substitution on the relationship between the aqueous phase and the hydrophobic interior of the maquette B (Figure 3a). Figure 3b displays typical spectra of three heme variants covering the experimental range of log  $P$  values at the concentrations of aqueous solutions used for the binding kinetic measurement. We assayed the solubility of the hemes and their tendency to form multimers in aqueous solution by UV–visible spectroscopy. In the left panel of Figure 3b, the visible spectrum of heme in pH 9 aqueous buffer (green) is dominated by the broad 350–400 nm Soret peak

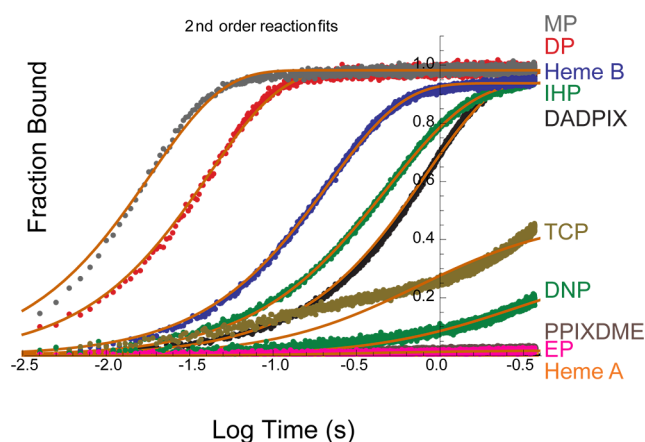


**Figure 3.** (a) Various protoporphyrin IX analogues used to depict the effect of changing the ring substituents on the assembly rate. Partition coefficient ( $\log P$ ) values in parentheses. (b) Spectra of three selected porphyrins in DMSO (black), 20 mM CHES, 150 mM KCl, pH 9, buffer (green), and bound to protein B (blue). Heme B (left), protoporphyrin IX dimethyl ester (middle), and diacetyl deuteroporphyrin IX (right).

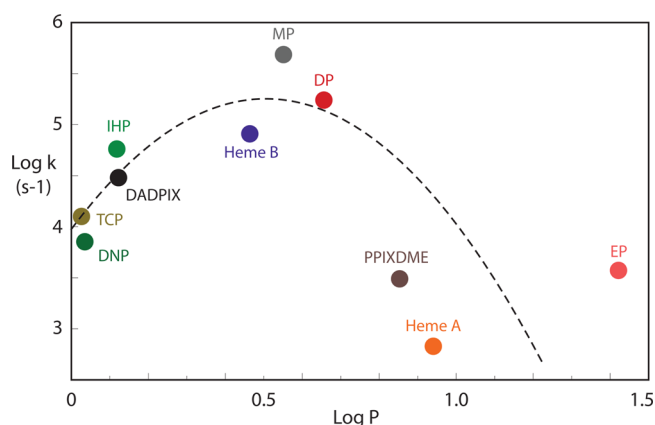
and the 605 nm peak characteristic of the  $\pi$ - $\pi$  dimer described by Asher et al.<sup>35</sup> The approximately 0.5  $\mu$ M aqueous dimer to monomer dissociation constant reported by this group<sup>36</sup> indicates that a substantial fraction of heme monomer is also present. There is no indication of the  $\mu$ -oxo heme dimer, reported decades ago by Brown et al., who also reported stronger heme dimer dissociation constants.<sup>37</sup> The left panel also shows that heme dissolved in DMSO (black) has the narrow Soret absorption band at 405 nm characteristic of heme in monomeric form.<sup>35</sup> The Soret absorption remains narrow and red shifts to 412 nm on bis-His ligation in the maquette (blue). Heme variants with higher  $\log P$ , such as protoporphyrin IX dimethyl ester, show the broad spectra of

multimer formation in aqueous solution (Figure 3b, middle), while hemes equipped with highly polar substituents and low  $\log P$  values (diacetyl deuteroporphyrin IX) exhibit simple monomeric solutions in water (Figure 3b, right).

Figure 4 shows time courses of binding of members of the family of Fe-porphyrins to maquette B, while Figure 5 summarizes the effect of  $\log P$  on the binding rate (linear time scale in Supplemental Figure 18). Compared to heme B, the slightly less polar deuteroporphyrin and mesoporphyrin bind to maquette B several-fold faster. On the other hand, making the Fe-porphyrin more polar by substituting the heme B vinyls with structurally comparable nitriles (2,6-dinitrileporphyrin) or adding polar hydroxyls (isohematoporphyrin) or



**Figure 4.** Assembly kinetics for the above series of Fe-porphyrins with the monomer protein fit to a second-order reaction (orange line). Rates were collected upon rapid mixing of protein B (24  $\mu\text{M}$ ) with porphyrin (5  $\mu\text{M}$ ) at 25  $^{\circ}\text{C}$  in 20 mM CHES, 150 mM KCl buffer, pH 9.



**Figure 5.** Comparison of protein B assembly rates with *n*-octanol/water log *P* values at pH 9 for the above series of Fe-porphyrins.

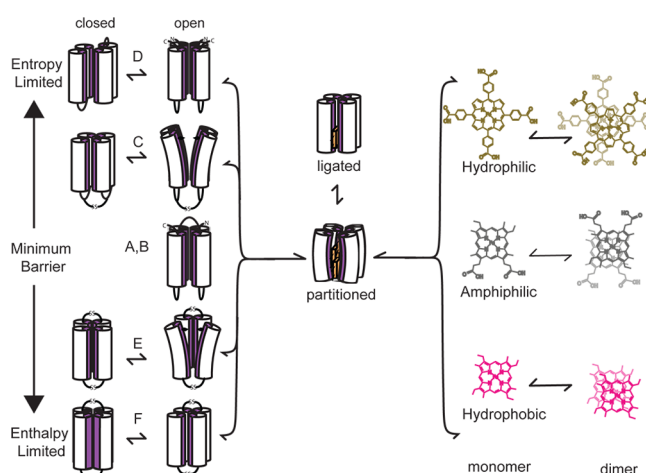
carbonyls (diacetyl deuteroporphyrin) slows the rate of binding. This suggests that the rate-limiting step of heme ligation to maquette B lies in the initial steps of partitioning into the hydrophobic interior of the maquette.

On the other hand, increasing the log *P* above a threshold value of about 0.75 by either removing the heme B propionates (etioporphyrin) or esterifying them (protoporphyrin IX dimethyl ester) leads to greater spectral evidence of porphyrin aggregation and significantly lowers the binding rates. This could reflect a lowering of the heme monomer concentration free in solution and/or a limiting rate of porphyrin disaggregation.

## DISCUSSION

Scheme 1 summarizes the hurdles that can stand in the way of Fe-porphyrin cofactor self-assembly into protein, starting in an aqueous medium and ending in a binding site in the interior of a protein that provides strong, effectively irreversible bis-histidine ligation for the heme iron. The first barrier is associated not with the protein but with the polarity balance and aggregation state of the Fe-porphyrin cofactor. Mesoporphyrin shows an optimal Fe-porphyrin amphiphilicity for rapid cofactor binding. Increased log *P*, indicating higher hydrophobic character, shifts the equilibrium toward the

## Scheme 1. Protein–Tetrapyrrole Binding Reaction<sup>a</sup>



<sup>a</sup>The various proteins are limited in their assembly by either entropic or enthalpic barriers related to protein mobility, while the porphyrin is limited by amphiphilicity. The entropic barrier of D is shown as flipping in this scheme, while F is shown as a tightly packed cartoon that opens up for assembly. Splaying open (E and C) is used as a pictorial representation of an open state capable of heme binding. A and B are not thought to be limited substantially by the protein structure. Porphyrins limit the assembly through their amphiphilic character.

aggregated Fe-porphyrin in solution, a form unable to ligate to the proteins. Decreased log *P*, indicating an increase in hydrophilic character, slows the rate by making partition of the unligated Fe-porphyrin in the hydrophobic core less favorable.

Heme association across our range of maquettes shows assembly on the 100 ms to tens of seconds time scale, much faster than the tens of minutes time scale reported by Kuzelova et al. for heme binding to natural proteins albumin and hemopexin.<sup>38</sup> Kuzelova interpreted the observed rates around 0.001 s<sup>−1</sup> to reflect rate-limiting heme dimer dissociation. Our distinctly faster assembly rates with all maquettes suggest that heme dimer dissociation is in fact much faster. While the assembly of most heme variants is well fit by a second-order time course, Fe-tetracarboxyphenyl porphyrin (TCP) displays unusual biphasic kinetics, with a burst phase extent of ~20% of unknown origin (Supplemental Figure 19). This may reflect some heterogeneity in the TCP dimer charge resulting from the p*K* values of 9.6 reported by Stong et al.<sup>39</sup>

We show two distinct ways protein structure can slow cofactor self-assembly. The first way is to impose a large entropic reorganization barrier, best demonstrated by protein D. Entropic barriers associated with minimal temperature dependence of the rate is a characteristic of structural rearrangements.<sup>40</sup> Structurally, this could reflect two separate untethered monomeric helix–loop–helix halves associating from solution to form the four-helix dimer that favorably buries the hydrophobic core of each monomer and enables heme binding. At extremely low nanomolar concentrations of protein, this might be the case, as heme binding is profoundly slowed. However, at micromolar concentrations, there is a conspicuous rapid burst phase of heme binding that indicates preformed dimers are available. On the other hand, the untethered monomers are free to associate in a helical bundle geometry (syn) that does not bring the histidines close enough together to form a bis-His heme ligation site. Similar geometry changes

are seen in the H10S24 protein maquette of Grosset et al.<sup>41</sup> Her work clearly shows a protein capable of reorganizing between syn and anti conformations on seconds time scale initiated by a change in heme redox states. In the case of maquette D, conformational change would impose a barrier wherein the maquette must search structural space in order to find one state suitable for cofactor assembly.

The addition of a simple tether, the disulfide bond linking loops in C, for example, can restrict the conformational freedom of the heme binding segments such that the two monomers cannot rearrange or diffuse away from one another, lowering the entropic barrier to moderate levels and speeding the rate of binding (Table 2). Protein C still retains residual conformational freedom associated with the smaller entropic barrier at the unlinked end of the bundle, while the increased structuring imposed by the tether adds to an enthalpic barrier.

The second way protein structure can slow cofactor self-assembly is to impose a large enthalpic barrier associated with apoprotein structuring. Protein F is highly structured in the apo state as seen in the crystal structure.<sup>29</sup> This structuring interferes with heme entry to the core as various packing interactions and internal hydrogen bonds must be broken. Such structural destabilization requires substantial amounts of thermal energy, thus giving protein F its strong temperature dependence.

These foregoing conclusions clearly demonstrate the power of maquettes to uncover the relevant physical chemical principles of protein design without complexity brought on by evolution. The stepwise redesign of maquettes from early multichain forms to a final single chain form that are sufficiently malleable in the apo state has increased the rate of cofactor self-assembly to a point comparable to natural proteins and limited by the physical chemical properties of the porphyrin itself. Like the maquette B, natural cytochromes  $b_{562}$  and  $b_5$  and oxygen transport proteins such as myoglobin are predominantly unstructured in the apo state, becoming substantially more structured when heme is bound.<sup>26,42</sup> In all cases, there is a small amount of structure in the apo state that restricts the conformational freedom of the unstructured regions. Apocytochrome  $b_{562}$  retains two intact helices, while the helices that bind heme are random coil. Moreover, in apocytochrome  $b_5$ , only a small region has any secondary structure, while the majority of the protein is random coil and unstructured. Though myoglobin is expressed as a single-chain protein, there is no published apo structure to reveal if there are minor structural elements restricting the conformational freedom. Despite this, NMR analysis of apomyoglobin has allowed Eliezer and Wright to construct a theoretical model of this protein without heme B showing, when combined with other literature describing its folding, that it too adheres to the principles outlined above: a conformationally restricted structure in which the binding site is malleable to allow access to heme B.<sup>43,44</sup>

The principles outlined in this article agree with the work of Shoemaker and Wolynes who describe the effect of molecular disorder on the binding of cofactors.<sup>45</sup> The authors from this work predict that disordered apo states have greater rates of binding in part due to their capture radius, the distance at which the protein will come into contact with a binding partner. This concept, referred to as the fly casting mechanism, can be easily applied to the maquettes in this paper owing to helical fraying and the molten globular nature of the apo state. However, difficulty in measuring structure in any localized

fashion, due to the repetition in maquette sequence and minimal change in CD signal between apo and holo states, makes proving this difficult.<sup>18,25</sup>

Beyond insights into natural proteins, the principles here can be readily applied to the design of cofactor binding proteins. Rate limitations can be uncovered through the use of temperature dependencies, and the structure can be addressed accordingly based on the information above.

Optimizing the ability of proteins to self-assemble with cofactors rapidly is as important to natural expression as increasing the affinity. Rapid cofactor binding can be of critical importance in the cell for cofactor modification and customization. It also minimizes the cytotoxicity of free heme B, which generates reactive oxygen species when unbound in solution.<sup>46–48</sup> Future man-made protein designs that seek to integrate with and exploit natural biochemical pathways in vivo will benefit by using these principles to diminish assembly barriers and effectively compete with host natural proteins for available cofactors described in this paper. Indeed, recent work utilizing a variant of maquette B has already been shown to co-opt the natural  $c$ -type machinery to covalently link heme B to an artificial protein.<sup>49</sup>

## ■ ASSOCIATED CONTENT

### § Supporting Information

Protein purification data as well as added kinetic analysis and protein sequences. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

[dutton@mail.med.upenn.edu](mailto:dutton@mail.med.upenn.edu)

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This research was supported in part by the U.S. National Institutes of Health, General Medical Institutes (RO1GM 41048), which funded molecular biology supplies, expression and purification from *E. coli* of the set of six different heme binding proteins that included five structurally distinct maquettes and our expressed reproduction of the natural cytochrome  $b_{562}$ ; also the basic biophysical and spectroscopic characterization of these six proteins and kinetic analysis of the heme B binding rates. The research was also supported in part by U.S. Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering (DE-FG02-05ER46223), which funded materials, design, synthesis, characterization, and development of the family of structural and polarity variants of Fe-tetrapyrroles and the characterization and analysis of their binding rates in the proteins. The authors would like to thank Jose Cerda for his help purifying heme A.

## ■ REFERENCES

- (1) Farid, T. A.; Kodali, G.; Solomon, L. A.; Lichtenstein, B. R.; Sheehan, M. M.; Fry, B. A.; Bialas, C.; Ennist, N. M.; Siedlecki, J. A.; Zhao, Z.; Stetz, M. A.; Valentine, K. G.; Anderson, J. L.; Wand, A. J.; Discher, B. M.; Moser, C. C.; Dutton, P. L. *Nat. Chem. Biol.* **2013**, *9*, 826.
- (2) Minagawa, J.; Takahashi, Y. *Photosynth. Res.* **2004**, *82*, 241.
- (3) Khalimonchuk, E.; Rodel, G. *Mitochondrion* **2005**, *5*, 363.



- (4) Kuras, R.; de Vitry, C.; Choquet, Y.; Girard-Bascou, J.; Culler, D.; Buschlen, S.; Merchant, S.; Wollman, F. A. *J. Biol. Chem.* **1997**, *272*, 32427.
- (5) Castelfranco, P. A.; Beale, S. I. *Annu. Rev. Plant Phys.* **1983**, *34*, 241.
- (6) Wang, Z. H.; Wang, Y. X.; Hegg, E. L. *J. Biol. Chem.* **2009**, *284*, 839.
- (7) Stevens, J. M.; Uchida, T.; Daltrop, O.; Ferguson, S. J. *Biochem. Soc. Trans.* **2005**, *33*, 792.
- (8) Canniffe, D. P.; Jackson, P. J.; Hollingshead, S.; Dickman, M. J.; Hunter, C. N. *Biochem. J.* **2013**, *450*, 397.
- (9) Hunter, C. N.; Tucker, J. D.; Niederman, R. A. *Photochem. Photobiol. Sci.* **2005**, *4*, 1023.
- (10) Robertson, D. E.; Farid, R. S.; Moser, C. C.; Urbauer, J. L.; Mulholland, S. E.; Pidikiti, R.; Lear, J. D.; Wand, A. J.; Degrado, W. F.; Dutton, P. L. *Nature* **1994**, *368*, 425.
- (11) Gibney, B. R.; Dutton, P. L. *Protein Sci.* **1999**, *8*, 1888.
- (12) Gibney, B. R.; Isogai, Y.; Rabanal, F.; Reddy, K. S.; Grosset, A. M.; Moser, C. C.; Dutton, P. L. *Biochemistry* **2000**, *39*, 11041.
- (13) Koder, R. L.; Anderson, J. L. R.; Solomon, L. A.; Reddy, K. S.; Moser, C. C.; Dutton, P. L. *Nature* **2009**, *458*, 305.
- (14) Lichtenstein, B. R.; Farid, T. A.; Kodali, G.; Solomon, L. A.; Anderson, J. L. R.; Sheehan, M. M.; Ennist, N. M.; Fry, B. A.; Chobot, S. E.; Bialas, C.; Mancini, J. A.; Armstrong, C. T.; Zhao, Z. Y.; Esipova, T. V.; Snell, D.; Vinogradov, S. A.; Discher, B. M.; Moser, C. C.; Dutton, P. L. *Biochem. Soc. Trans.* **2012**, *40*, S61.
- (15) Nikkila, H.; Gennis, R. B.; Sligar, S. G. *Eur. J. Biochem.* **1991**, *202*, 309.
- (16) Arnesano, F.; Banci, L.; Bertini, I.; Faraone-Mennella, J.; Rosato, A.; Barker, P. D.; Fersht, A. R. *Biochemistry* **1999**, *38*, 8657.
- (17) Feng, Y. Q.; Sligar, S. G. *Biochemistry* **1991**, *30*, 10150.
- (18) Gibney, B. R.; Huang, S. S.; Skalicky, J. J.; Fuentes, E. J.; Wand, A. J.; Dutton, P. L. *Biochemistry* **2001**, *40*, 10550.
- (19) Lichtenstein, B. R. Doctoral Thesis, University of Pennsylvania, 2010.
- (20) Takemori, S.; King, T. E. *J. Biol. Chem.* **1965**, *240*, 504.
- (21) Berry, E. A.; Trumpower, B. L. *Anal. Biochem.* **1987**, *161*, 1.
- (22) Bodansky, M. *Peptide Chemistry: A Practical Approach*, 2nd ed.; Springer-Verlag: New York, 1993.
- (23) Houston, P. L. *Chemical Kinetics and Reaction Dynamics*; 1st ed.; Dover Publications Inc.: Mineola, NY, 2001.
- (24) Leo, A.; Hansch, C.; Elkins, D. *Chem. Rev.* **1971**, *71*, S25.
- (25) Koder, R. L.; Valentine, K. G.; Cerda, J.; Noy, D.; Smith, K. M.; Wand, A. J.; Dutton, P. L. *J. Am. Chem. Soc.* **2006**, *128*, 14450.
- (26) Wand, A. J.; Feng, Y. Q.; Sligar, S. G. *J. Cell. Biochem.* **1993**, *286*.
- (27) Sturtevant, J. M.; Robinson, C. R.; Liu, Y. F.; Thomson, J. A.; Sligar, S. G. *Biochemistry* **1997**, *36*, 16141.
- (28) Gibney, B. R.; Rabanal, F.; Reddy, K. S.; Dutton, P. L. *Biochemistry* **1998**, *37*, 4635.
- (29) Huang, S. S.; Gibney, B. R.; Stayrook, S. E.; Dutton, P. L.; Lewis, M. J. *Mol. Biol.* **2003**, *326*, 1219.
- (30) Feng, Y. Q.; Wand, A. J.; Sligar, S. G. *Biochemistry* **1991**, *30*, 7711.
- (31) Zhang, L.; Anderson, J. L. R.; Ahmed, I.; Norman, J. A.; Negron, C.; Mutter, A. C.; Dutton, P. L.; Koder, R. L. *Biochemistry* **2011**, *50*, 10254.
- (32) Gibney, B. R.; Rabanal, F.; Reddy, K. S.; Dutton, P. L. *Biochemistry* **1998**, *37*, 4635.
- (33) Prince, R. C.; Bashford, C. L.; Takamiya, K. I.; van den Berg, W. H.; Dutton, P. L. *J. Biol. Chem.* **1978**, *253*, 4137.
- (34) Keske, J. M.; Bruce, J. M.; Dutton, P. L. *Z. Naturforsch. C* **1990**, *45*, 430.
- (35) Asher, C.; de Villiers, K. A.; Egan, T. J. *Inorg. Chem.* **2009**, *48*, 7994.
- (36) de Villiers, K. A.; Kaschula, C. H.; Egan, T. J.; Marques, H. M. J. *Biol. Inorg. Chem.* **2007**, *12*, 101.
- (37) Brown, S. B.; Dean, T. C.; Jones, P. *Biochem. J.* **1970**, *117*, 733.
- (38) Kuzelova, K.; Mrhalova, M.; Hrkal, Z. *Biochim. Biophys. Acta* **1997**, *1336*, 497.
- (39) Stong, J. D.; Hartzell, C. R. *Bioinorg. Chem.* **1976**, *5*, 219.
- (40) Oliveberg, M.; Tan, Y. J.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 8926.
- (41) Grosset, A. M.; Gibney, B. R.; Rabanal, F.; Moser, C. C.; Dutton, P. L. *Biochemistry* **2001**, *40*, 5474.
- (42) Davis, R. B., Jr.; Lecomte, J. T. *Biopolymers* **2008**, *90*, 556.
- (43) Eliezer, D.; Wright, P. E. *J. Mol. Biol.* **1996**, *263*, 531.
- (44) Culbertson, D. S.; Olson, J. S. *Biochemistry* **2010**, *49*, 6052.
- (45) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8868.
- (46) Prousek, J. *Pure Appl. Chem.* **2007**, *79*, 2325.
- (47) Allen, J. W. A.; Leach, N.; Ferguson, S. J. *Biochem. J.* **2005**, *389*, 587.
- (48) Verissimo, A. F.; Mohtar, M. A.; Daldal, F. *J. Biol. Chem.* **2013**, *288*, 6272.
- (49) Anderson, J. L. R.; Armstrong, C. T.; Kodali, G.; Lichtenstein, B. R.; Watkins, D. W.; Mancini, J.; Boyle, A. L.; Farid, T. A.; Crump, M. P.; Moser, C. C.; Dutton, P. L. *Chem. Sci.* **2014**, *5*, 507–514.