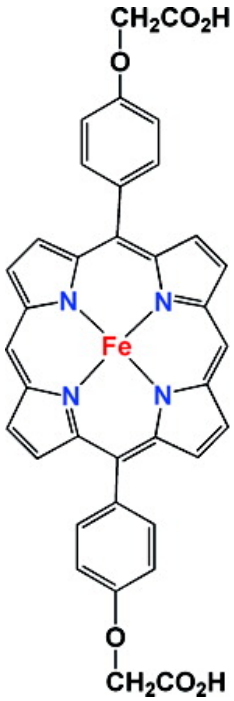
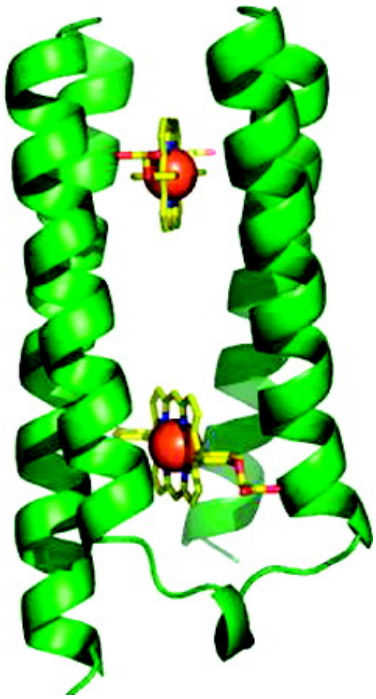
Article

De Novo Design of a Single-Chain Diphenylporphyrin Metalloprotein

Gretchen M. Bender, Andreas Lehmann, Hongling Zou, Hong Cheng, H. Christopher Fry, Don Engel, Michael J. Therien, J. Kent Blasie, Heinrich Roder, Jeffrey G. Saven, and William F. DeGrado

J. Am. Chem. Soc., **2007**, 129 (35), 10732-10740• DOI: 10.1021/ja071199j • Publication Date (Web): 10 August 2007





More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 4 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 08/10/2007

De Novo Design of a Single-Chain Diphenylporphyrin Metalloprotein

Gretchen M. Bender,[†] Andreas Lehmann,[‡] Hongling Zou,[‡] Hong Cheng,[§] H. Christopher Fry,[‡] Don Engel,[†] Michael J. Therien,[‡] J. Kent Blasie,[‡] Heinrich Roder,[§] Jeffrey G. Saven,[‡] and William F. DeGrado*[,†,‡]

Contribution from the Department of Biochemistry and Molecular Biophysics, Johnson Foundation, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania 19104, and Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111

Received February 19, 2007; E-mail: wdegrado@mail.med.upenn.edu

Abstract: We describe the computational design of a single-chain four-helix bundle that noncovalently self-assembles with fully synthetic non-natural porphyrin cofactors. With this strategy, both the electronic structure of the cofactor as well as its protein environment may be varied to explore and modulate the functional and photophysical properties of the assembly. Solution characterization (NMR, UV−vis) of the protein showed that it bound with high specificity to the desired cofactors, suggesting that a uniquely structured protein and well-defined site had indeed been created. This provides a genetically expressed single-chain protein scaffold that will allow highly facile, flexible, and asymmetric variations to enable selective incorporation of different cofactors, surface-immobilization, and introduction of spectroscopic probes.

Introduction

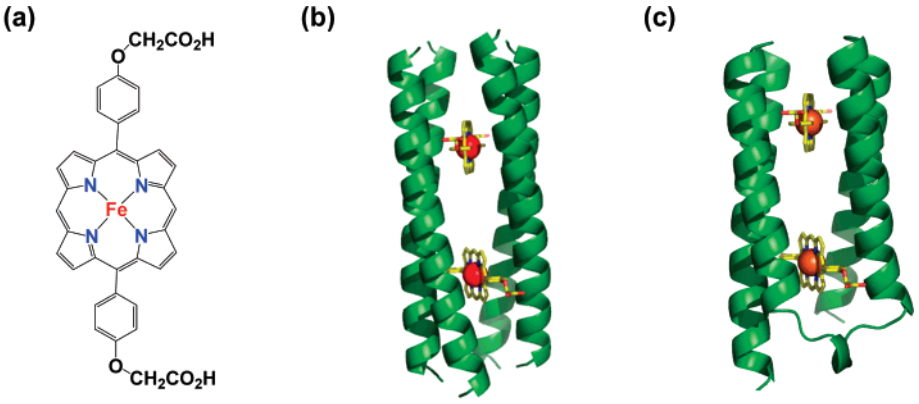
The design of optically and electronically responsive molecular assemblies capable of diverse photophysical processes is becoming increasingly feasible because of progress in the study of electron transfer and charge separation in both biological systems and synthetic small molecules.[1−4] Recently, studies aimed at combining the strengths of these two types of systems have been initiated through the computational design of four-helix bundles that noncovalently self-assemble with fully synthetic non-natural porphyrin cofactors. With this strategy, both the electronic structure of the cofactor and its protein environment may be varied to explore and modulate the functional and photophysical properties of the assembly. These systems are unlike previous de novo "maquettes"[3,5] in that these peptides were computationally designed to recognize a unique porphyrin cofactor. Thus, while maquettes bind a large number of porphyrins with low specificity, computationally designed proteins should have much higher specificity. Indeed, they bound with high specificity to their desired cofactors, suggesting that a uniquely structured protein and well-defined site had successfully been created. The initial assemblies had D_2 symmetry (PA_{TET}, Figure 1b),[6−8] comprising four identical helices that encapsulate two iron diphenylporphyrin molecules (FeDPP(III), Figure 1a).

Porphyrin-binding peptides have been extensively studied for the past 15 years, utilizing helix−loop−helix designs with disulfide bonds and template-assisted design strategies.[1−4,9−16] Until now, however, there have been no genetically expressed single-chain proteins that permit highly facile, flexible, and asymmetric variations of the protein to allow selective incorporation of different cofactors, surface immobilization, and introduction of spectroscopic probes. Additionally, a single-chain variant would allow for the modification of Cys residues, semisynthesis,[17] or genetic strategies for introducing unique

[†] School of Medicine, University of Pennsylvania.
[‡] Department of Chemistry, University of Pennsylvania.
[§] Fox Chase Cancer Center.

(1) Lombardi, A.; Nastri, F.; Pavone, V. *Chem. Rev.* **2001**, *101*, 3165−3189.
(2) Reedy, C. J.; Gibney, B. R. *Chem. Rev.* **2004**, *104*, 617−649.
(3) Kennedy, M. L.; Gibney, B. R. *Curr. Opin. Struct. Biol.* **2001**, *11*, 485−490.
(4) Xu, T.; Wu, S. P.; Miloradovic, I.; Therien, M. J.; Blasie, J. K. *Nano Lett.* **2006**, *6*, 2387−2394.
(5) Discher, B. M.; Koder, R. L.; Moser, C. C.; Dutton, P. L. *Curr. Opin. Chem. Biol.* **2003**, *7*, 741−748.
(6) Ghirlanda, G.; Osyczka, A.; Liu, W.; Antolovich, M.; Smith, K. P.; Dutton, P. L.; Wand, A. J.; DeGrado, W. F. *J. Am. Chem. Soc.* **2004**, *126*, 8141−8147.
(7) North, B.; Summa, C. M.; Ghirlanda, G.; DeGrado, W. F. *J. Mol. Biol.* **2001**, *311*, 1081−1090.
(8) Cochran, F. V.; Wu, S. P.; Wang, W.; Nanda, V.; Saven, J. G.; Therien, M. J.; DeGrado, W. F. *J. Am. Chem. Soc.* **2005**, *127*, 1346−1347.
(9) DeGrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F.; Lombardi, A. *Annu. Rev. Biochem.* **1999**, *68*, 779−819.
(10) Shifman, J. M.; Gibney, B. R.; Sharp, R. E.; Dutton, P. L. *Biochemistry* **2000**, *39*, 14813−14821.
(11) Rau, H. K.; Haehnel, W. *J. Am. Chem. Soc.* **1998**, *120*, 468−476.
(12) Mutter, M.; Tuchscherer, G. G.; Miller, C.; Altmann, K. H.; Carey, R. I.; Wyss, D. F.; Labhardt, A. M.; Rivier, J. E. *J. Am. Chem. Soc.* **1992**, *114*, 1463−1470.
(13) Liu, D. H.; Williamson, D. A.; Kennedy, M. L.; Williams, T. D.; Morton, M. M.; Benson, D. R. *J. Am. Chem. Soc.* **1999**, *121*, 11798−11812.
(14) Huffman, D. L.; Rosenblatt, M. M.; Suslick, K. S. *J. Am. Chem. Soc.* **1998**, *120*, 6183−6184.
(15) Rosenblatt, M. M.; Wang, J. Y.; Suslick, K. S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13140−13145.
(16) Arnold, P. A.; Benson, D. R.; Brink, D. J.; Hendrich, M. P.; Jas, G. S.; Kennedy, M. L.; Petasis, D. T.; Wang, M. X. *Inorg. Chem.* **1997**, *36*, 5306−5315.
(17) Muir, T. W. *Annu. Rev. Biochem.* **2003**, *72*, 249−289.

10.1021/ja071199j CCC: $37.00



(d)

```
                                a  d    a  d    a  d    a  d    a
PA_TET Sequence (1-34):      S LEEALQE AQQTAQE AQQALQK HQQAFQK FQKYG

PA_SC Sequence (1-31):       S PEEAMQE AQQTARE AEQAMQK HRQAYDKGD

PA_SC Sequence (32-60):        QQK ALQTAKE FQQAMQK HKQY    MNPQAISE

PA_SC Sequence (61-84):          S VQKTARY FEQAMQK HRQAYDKGD

PA_SC Sequence (85-108):       QQK ALQTAKE AQQAMQK HSQALRG
```

Figure 1. Model and sequence of porphyrin assemblers. (a) Iron diphenylporphyrin cofactor, (b) model of PA_{TET}[8], (c) model of PA_{SC}, (d) sequences of PA_{TET} and PA_{SC}, which have the helical regions underlined, and the heptad positions of the residues are lettered. Keystone residues are bolded.

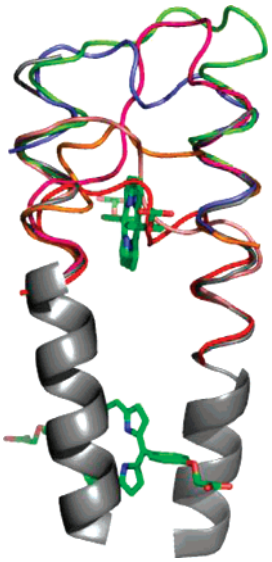


Figure 2. Model of loop choices for PA_{SC}. Each color corresponds to a section from the structures selected from the PDB: 2spc (light pink), lim8 (red), 1t06 (blue), 1fzn (magenta), 1g03 (orange), and lea3 (light green). The loop 1t06 (in blue) had the lowest deviation in rmsd from the PA_{TET} helices (shown in gray, they correspond to helices 2 and 3 of PA_{SC}) and was chosen for the PA_{SC} scaffold.

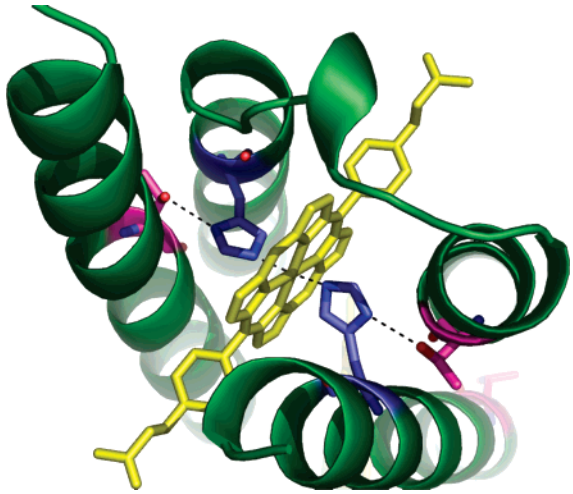


Figure 3. Keystone residues in PA_{SC}. Histidines are shown in blue, threonines are shown in pink, and the porphyrin ring of FeDPP(III) is shown in yellow; hydrogen bonds are shown as dotted lines from the Fe to N(ϵ) of His, and from the Thr O(γ) to N(δ) of His.

cofactors and probes.[18] To this end, we present a single-chain version, PA_{SC}, of the originally designed four-chain helical bundle (Figure 1c).

Although many porphyrin-binding helical bundles have been designed and extensively characterized, there have been significant problems associated with the determination of their three-dimensional structures, in part because many have conformationally fluctuating interiors.[19,20] Over the years, this problem has been alleviated somewhat through the synthesis of variants of the original "maquettes", resulting in better dispersion in the NMR spectra of the complexes.[21] Smaller peptide−porphyrin complexes with two helices per complex

(18) Wang, L.; Xie, J.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225−249.

(19) Gibney, B. R.; Rabanal, F.; Reddy, K. S.; Dutton, P. L. *Biochemistry* **1998**, *37*, 4635−4643.

(20) Betz, S. F.; Raleigh, D. P.; DeGrado, W. F. *Curr. Opin. Struct. Biol.* **1993**, *3*, 601−610.

(21) Huang, S. S.; Koder, R. L.; Lewis, M.; Wand, A. J.; Dutton, P. L. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5536−5541.

have also been structurally characterized.[21−22] The availability of a single-chain protein, such as PA_{SC}, would greatly expedite structural determination because of the ease of incorporation of isotopic labels for heteronuclear NMR studies.

Methods

Design of PA_{SC}. Requirements for efficient cofactor binding dictated the precise protein backbone geometry. A single low-energy structure was computed with a Monte Carlo simulated annealing protocol that considered the following constraints: (i) a metal−metal distance between 17 and 19 Å, (ii) optimal His N(ϵ) to Fe bonding interactions, (iii) second-shell hydrogen bonds between His N(δ) and Thr O(γ), (iv) minimal steric clashes, and (v) maintenance of D_2 symmetry. This process led to imidazole rings in near-perpendicular alignment as a result of the second-shell hydrogen bonds. These steps to obtain protein backbone and cofactor coordinates were repeated from Cochran et al.[8] and led to the same tetrameric four-helix bundle, which was then modified. All four 32-residue helices of the original PA_{TET} bundle were shortened by four residues at each end, except for the N-terminus of one helix, which would eventually provide the N-terminus of the entire single-chain four-helix bundle. The helices were shortened so as to obtain a compact (108-residue) single-chain peptide, PA_{SC}.

With regard to stabilizing the bundle, the loops take the place of hydrophobic interactions at each end of the bundle in the previous tetrameric complex.[8] The resulting bundle structure contained three shortened 24-residue helices and one 28-residue helix. The helices were connected through loops of natural protein structures taken from a protein database with reduced redundancy, PDB Select.[24] A program called STITCH was developed to identify loops that would be most suitable to connect a given pair of unconnected helices. STITCH enumerates helix−loop−helix superpositions and rank-orders these superpositions according to root mean square differences (rmsd) between C_α atoms of the helical residues that terminate the loops and the C_α atoms of the helices. Loop structures were selected having overlaps of at least five C_α positions at each loop−helix transition. Overlapping backbone fragments were deleted. Connection points were identified by visual inspection, looking for closest overlap of loop and helix backbone atoms. Residues near the connection points were relaxed using CHARMM[25] steepest descent optimization for 25 steps. Bond lengths and bond angles of the resulting backbone structure were checked using PROCHECK.[26]

Protein Design Steps. The identities of the four Fe-coordinating His residues and the four Thr residues to which these are hydrogen-bonded were held fixed to support the six-point coordination of the Fe(III) cations: T12, H23, T38, H49, T65, H76, T91, and H102. The identities of the remaining 100 out of 108 positions were determined by recursive calculations using the computational design algorithm SCADS.[27,28] SCADS (which provides site-dependent amino acid probabilities) has been used previously in a number of successful protein design studies.[8,28−29] All amino acid identities were allowed, except for cysteine and histidine, to eliminate the possibility of oxidative degradation and unwanted disulfide bridges and to avoid cofactor coordination at sites other than those in the positions designated above. After the first round of SCADS calculations, identities in 24 positions were fixed because they were the most probable amino acids with high probabilities of at least $p > 0.5$ and frequently $p > 0.8$: A9, A13, A16, M20, A26, K34, A35, A39, F42, A45, M46, K50, A66, F69, A72, M73, A79, K87, A88, A92, A95, A98, and M99. Of these, A16, M20, F42, M46, F69, M73, A95, and M99 form a complementary, asymmetric hydrophobic core between the two porphyrin macrocycles. Note that this nonsymmetrical packing illustrates a major advantage of a single-chain design as opposed to a tetramer because, as symmetry is not required, more complementary packing options become available. The hydrophobic core is shown in Figure 4.

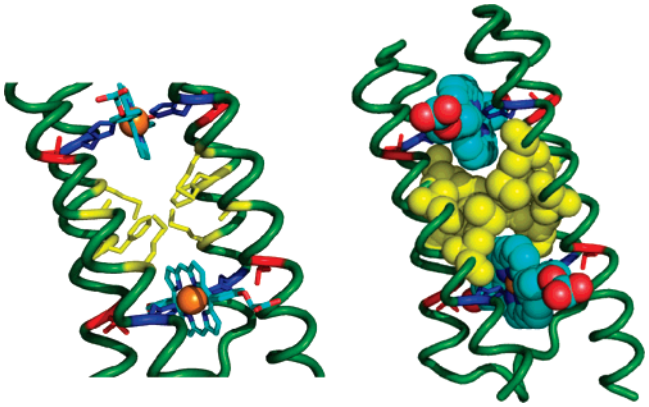


Figure 4. Asymmetric hydrophobic interior of PA_{SC}. Dark blue = His, yellow = Met and Phe, red = Ala; blue ribbon section are the histidines. The nonsymmetric hydrophobic core of PA_{SC} consists of residues A16, M20, F42, M46, F69, M73, A95, and M99. These residues were picked by SCADS as residues dominating the probability profiles at their respective positions.

In addition to these fixed residues, for the second-round of SCADS calculations nine interior positions 6, 27, 36, 53, 58, 62, 80, 89, and 106, each facing one of the cofactors, were limited to noncharged or hydrophobic amino acid identities. After this second round of SCADS calculations, decisions about the remaining 76 amino acid identities were made on the basis of the resulting SCADS probability profiles in conjunction with other considerations relevant to the design. Residues M6, Q11, Q18, R24, Q25, Q37, K40, Q44, Q51, Y52, I58, E60, V62, K64, Q71, R77, Q78, Q90, K93, Q100, S103, Q104, and L106 were chosen because their SCADS probabilities were the highest at their respective sites. That is, a total of 55 out of 108 amino acid identities were fixed on the basis of either the requirement for the hexacoordinated Fe(III) ions or on the highest SCADS probabilities. N-Terminal residues S1, P2, E3, E4, and the final G108 form known helix-capping motifs. Residues A5, Q7, E8, Q10, E15, Q21, K22, Q33, E41, Q47, K48, Q63, Q74, K75, Q86, E94, Q96, Q97, K101, and A105 were chosen because they were probable and because they also occupied the same positions relative to the cofactors in the earlier successful four-helix bundle design.[8] In particular, interhelical salt bridges E15-K48, K22-E41, K75-E94 were also introduced to stabilize the fold. Positively charged R14, R67, and R107 (together with K50 from the first round of SCADS calculations) were chosen near the negatively charged carboxylate groups of the cofactors. D28, K29, G30, D31, D81, K82, G83, and D84 were identified on the basis of a statistical analysis of α-helical hairpin turns[30,31] in which the authors analyzed the relative frequencies of amino acids in such hairpin turns and found that some amino acids are preferred. Similarly, in the larger loop, identities of the wild-type residues at positions P55, A57, S59, and S61, taken from 1T06.pdd, were kept. Q32, Q43, Q56, and Q85 were chosen over more probable charged amino acids to reduce the overall number of charged amino acids in the protein. L36, M53, and L89 are hydrophobic residues

(22) Huang, S. S.; Gibney, B. R.; Stayrook, S. E.; Dutton, P. L.; Lewis, M. *J. Mol. Biol.* **2003**, *326*, 1219−1225.
(23) Laue, T. M.; Shah, B. D.; Ridgeway, T. M.; Pelletier, S. L. Computer-aided interpretation of analytical sedimentation data for proteins. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science;* Harding, S. E., Rowe, A. J., Horton, J. C., Eds.; Royal Society of Chemistry: Cambridge, 1992; pp 90−125.
(24) Hobohm, U.; Sander, C. *Protein Sci.* **1994**, *3*, 522−524.
(25) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(26) Laskowski, R. A.; Macarthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283−291.
(27) Kono, H.; Saven, J. G. *J. Mol. Biol.* **2001**, *306*, 607−628.
(28) Calhoun, J. R.; Kono, H.; Lahr, S.; Wang, W.; DeGrado, W. F.; Saven, J. G. *J. Mol. Biol.* **2003**, *334*, 1101−1115.
(29) Nanda, V.; Rosenblatt, M. M.; Osyczka, A.; Kono, H.; Getahun, Z.; Dutton, P. L.; Saven, J. G.; DeGrado, W. F. *J. Am. Chem. Soc.* **2005**, *127*, 5804−5805.
(30) Lahr, S. J.; Engel, D. E.; Stayrook, S. E.; Maglio, O.; North, B.; Geremia, S.; Lombardi, A.; DeGrado, W. F. *J. Mol. Biol.* **2005**, *346*, 1441−1454.
(31) Engel, D. E.; DeGrado, W. F. *J. Mol. Biol.* **2004**, *337*, 1195−1205.

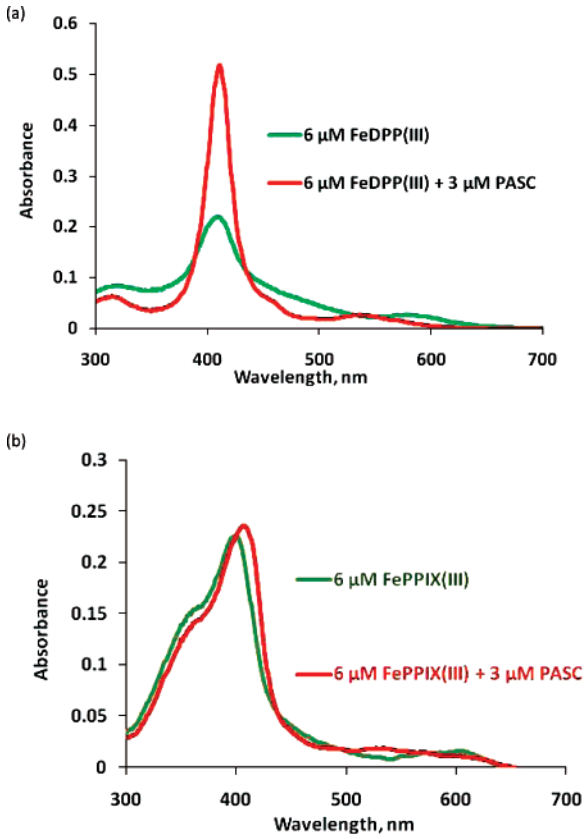


Figure 5. Cofactor binding selectivity. (a) PA_{SC} demonstrates cofactor selectivity for FeDPP(III) over a biological cofactor. (b) Iron protoporphyrin IX, FePPIX(III). Cofactor was added to the protein in 50 mM phosphate buffer and 100 mM NaCl, pH 7.5, as described in the Methods.

in interior positions facing the cofactors, which were preferred over more probable polar amino acids. Conversely, N54 and Y68 are polar residues facing the protein exterior and were selected over slightly more probable hydrophobic residues. Charged residues E17 and E70 were introduced to exploit potentially favorable intrahelical i, $i + 3$ interactions with R14 and R67, respectively. Finally, Y27 and Y80 were chosen over the more probable W because of their more hydrophilic character at these partially exposed sites.

PA_{SC}/FeDPP(III) Complex Preparation. The synthetic gene for PA_{SC} was obtained and the codon sequence optimized using software from DNA2.0. The sequence was transformed into the pET28a vector from the *Nco*I to *Bam*HI restriction sites (Novagen) and contains the IPTG-inducible T7 promoter. The protein was overexpressed in *Escherichia coli* and harvested by centrifugation, and the cells lysed by sonication. The supernatant was heat-denatured for 30 min at 70 °C and centrifuged, and the supernatant was purified to homogeneity by reverse-phase HPLC on a preparative C4 column (Vydac). The molecular weight was confirmed by MALDI mass spectrometry. The lyophilized protein was reconstituted in 50 mM phosphate and 100 mM NaCl at pH 7.5. As previously described, the chemically synthesized FeDPP(III) cofactor was added from a stock solution in DMSO.[8] Aggregation caused by this addition was minimized by briefly heating the complex to 70 °C for 15 min (above the T_m of the holoprotein). After being heated, the samples were equilibrated at room temperature for an hour and filtered though a 0.22-μm PVDF membrane centrifuge filter (Millipore) at 13 200 rpm in an Eppendorf 5415D microcentrifuge.

UV−Visible and Circular Dichroism (CD) Spectroscopy. To verify binding of the cofactor, complex formation was monitored at the Soret band at 410 nm. UV−visible data of the complex (as prepared above) was obtained on a Cary 300 Bio UV−visible spectrophotometer at room temperature in a 1-cm quartz cuvette. To evaluate the backbone and stability of the protein and the complex, CD spectroscopy was carried out at 25 °C on a Jasco 810 Spectropolarimeter in 0.1-cm cuvettes. Buffer conditions for CD experiments were 10 mM phosphate buffer, pH 7.5, and the protein concentration was 25−50 μM. For thermal denaturation experiments, the ellipticity at 222 nm was monitored as a function of temperature. Data were collected every 2 °C with a heating rate of 1 °C/min followed by a 4-min equilibration time. Melting temperature midpoints (T_m's) were estimated by plotting temperature (x axis) versus F_{fold} (y axis) in which the fraction folded was estimated from the observed ellipticity versus that for the folded and unfolded baselines. The apo states of the proteins were too unstable to resolve a folded baseline; therefore, the parameters for the holoprotein were used to approximate these values:

$$F_{fold} = ([\theta_{observed}] - [\theta_{unfolded}])/([\theta_{folded}] - [\theta_{unfolded}])$$

$$[\theta_{unfolded}] = [\theta_{unfolded}]_0 + T(\mathrm{d}[\theta_{unfolded}]/\mathrm{d}T)$$

$$[\theta_{folded}] = [\theta_{folded}]_0 + T(\mathrm{d}[\theta_{folded}]/\mathrm{d}T)$$

Determination of Aggregation State. Size Exclusion Chromatography (SEC). Gel filtration profiles were obtained using a Superdex 75 10/300GL column on an FPLC system (GE Healthcare AKTA FPLC System). To evaluate the oligomeric state, 100 μL of the complex was injected onto the column and eluted with a flow rate of 0.5 mL/min with a mobile phase of 50 mM phosphate buffer and 100 mM NaCl, pH 7.5. The monomeric species eluted at 12.3 mL, yielding an apparent molecular weight of 18 500 Da. MW_{app} was calculated from a standard curve done with mass standards blue dextran, V_o: aprotinin, 6500 Da; cytochrome C, 12 400 Da; carbonic anhydrase, 29 000 Da; and albumin, 66 000 Da.

Analytical Ultracentrifugation. For analytical ultracentrifugation, a 300 μL sample of 150 μM PA_{SC} with 300 μM FeDPP(III) in 50 mM phosphate buffer and 100 mM NaCl at pH 7.5 was prepared as previously mentioned. After being centrifuged for 10 min at 13 200 rpm in an Eppendorf 5415D microcentrifuge to remove insoluble material, an aliquot of 110 μL was used for sedimentation equilibrium experiments. These were performed at 25 °C using a Beckman XL-I analytical ultracentrifuge. The absorbance was monitored at 280 nm, and the sample was centrifuged at 25 000, 35 000, and 45 000 rpm (Supporting Information, Figure S1). The data were analyzed using a modified global fitting routine in IGOR Pro (Wavemetric). The data were well-described by assuming a single molecular weight species. The protein maintains its calculated partial specific volume when the monomeric molecular weight is held constant. The partial specific volume of 0.7137 was calculated using SEDNTERP.[23]

Stoichiometry. Binding stoichiometry was determined using the monomeric species isolated through SEC. The monomeric PA_{SC}/FeDPP(III) complex was monitored for formation by UV−vis spectroscopy at 410 nm. By holding the protein concentration fixed, various cofactor concentrations were evaluated using UV−visible spectroscopy to monitor the formation of the Soret band at 410 nm. The stoichiometry of the SEC fractions was determined by a hemochrome assay[32] for FeDPP(III) concentration and an HPLC assay for protein concentration (Figure S2). Binding stoichiometry was determined to be 2:1, as there was no significant increase in absorbance at 410 nm with excess cofactor (Figure S3).

NMR Analysis. Multiple preparations of 100 μM ^{15}N-labeled PA_{SC} with 200 μM FeDPP(III) were incubated at 70 °C for 15 min, allowed to equilibrate at room temperature, and filtered through a 0.45-μm Nylon syringe filter, and 2 mL was injected over a preparative Superdex 75 XK16/70 column (GE Healthcare AKTA FPLC System). The monomeric fractions were collected and concentrated using an Amicon Ultra-15 MWCO 10 000 filter (Millipore). The sample was buffer-exchanged

(32) Berry, E. A.; Trumpower, B. L. *Anal. Biochem.* **1987**, *161*, 1−15.

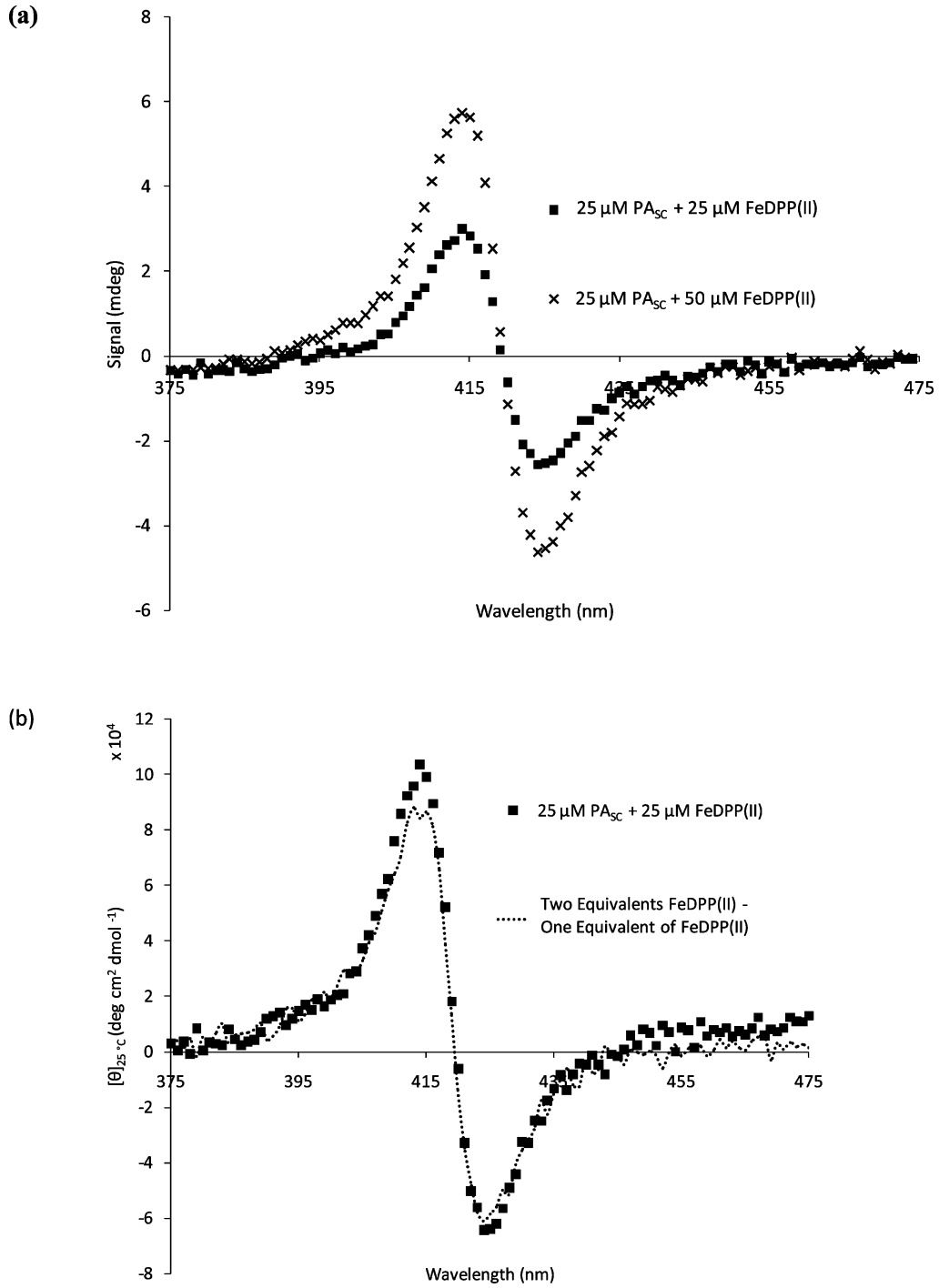


Figure 6. CD spectroscopy in the visible−UV region. (a) The addition of substoichiometric (25 μM) and stoichiometric (50 μM) amounts of FeDPP(II) to 25 μM PA$_{SC}$ at 25 °C in 10 mM phosphate buffer at pH 7.5. (b) The spectra demonstrate equality of binding sites within the PA$_{SC}$ scaffold.

into 50 mM phosphate buffer and 200 mM NaCl, pH 7.5, over a PD-10 column (GE Healthcare) and concentrated to 270 μL for the NMR experiment. The final sample was diluted to 300 μL with D_2O, and 2 μL of 150 mM 3-(trimethylsilyl)-1-propanesulfonic acid sodium salt stock (Fluka) in water was added. The final concentrations of cofactor and protein were determined by hemochrome and HPLC assays (Figure S2).

The $^1H-^{15}N$ heteronuclear single quantum correlation (HSQC) spectrum was recorded at 45 °C on a Bruker DMX-600 spectrometer equipped with a 5-mm *x*,*y*,*z*-shielded pulsed-field gradient triple-resonance probe. 2048 (^1H) × 128 (^{15}N) complex data points were collected for each dimension. The spectral widths used were 8992.806 Hz for ^1H and 1702.903 Hz for ^{15}N dimensions. The data were processed by using Felix 2002 (Felix NMR). The final matrix contained 2048 × 2048 data points along ^1H and ^{15}N dimensions.

Molecular Dynamics (MD) Simulations. The designed cofactor-bound PA$_{SC}$ was taken as the starting structure for the simulations. The protein was solvated within a 50 × 50 × 72 Å^3 water box, and periodic boundary conditions were employed in all cell dimensions. Simulations were performed using the NAMD 2.5 package[33] and the

(33) Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 283−312.

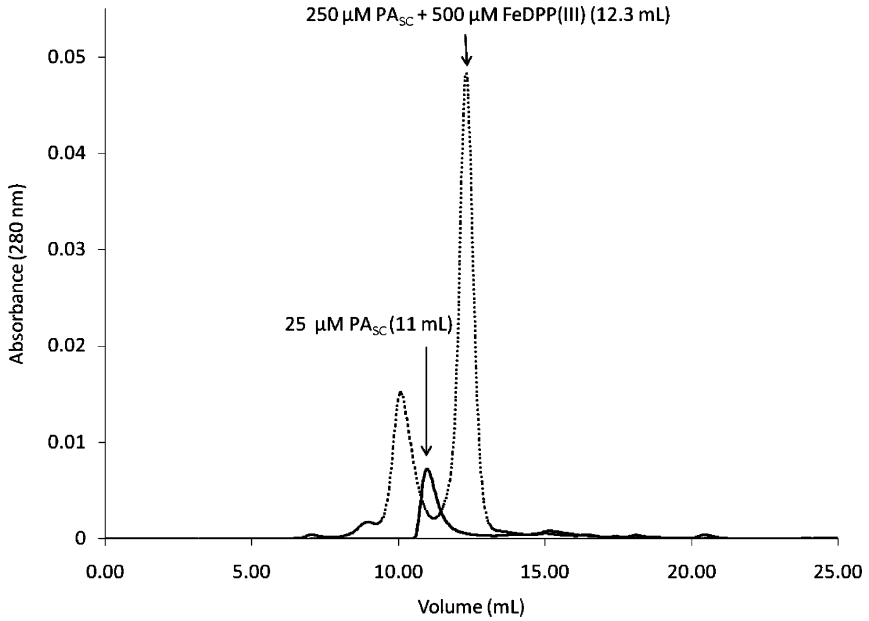


Figure 7. SEC of the PA_{SC} apo and holo complexes. The apoprotein eluted at a retention time (11 mL, black line) higher than that of the holoprotein (12.3 mL, dotted line), indicating that once the cofactor binds, the holoprotein is more compact.

CHARMM22 all-atom force field.[34] The parameterization of the six-coordinate FeDPP(III) macrocyle and substituents was derived from CHARM22 parameters for the iron heme and from functional groups of amino acid side chains.[34] A time step of 2 fs was used for bonded and nonbonded interactions, and a cutoff distance of 10 Å was maintained to calculate nonbonded interactions. Full electrostatic forces were evaluated at every other time step through application of the particle-mesh Ewald method.[35] All bonds between hydrogen atoms and heavy atoms were constrained using the SHAKE algorithm with a tolerance of 10^{-5} Å.[36]

Results

Design. The structure of the 108-residue single-chain protein, PA_{SC} (porphyrin assembler, single chain, Figure 1c) was designed, using the model for the corresponding homotetramer (Figure 1b, PA_{TET}) as a starting point. As described in the Methods section, interhelical loops with favorable conformations and sequences were introduced into the structure by grafting known interhelical backbone structures onto the PA_{TET} model. The precise conformations of the loops in the structure were obtained by (1) superimposing helical hairpins from proteins of known structure onto the helices of PA_{TET}, (2) selecting hairpins with a low rmsd for the superimposed helical residues, (3) inserting the loop at a position where the chains are particularly well-superimposed, and (4) selecting structures that did not interfere with cofactor binding (Figure 2).[30] As described in previous work,[8,28] eight "keystone" residues directly involved in cofactor ligation or second-shell interactions were predefined (four histidines and four threonines, Figure 3), while the remaining positions of the protein were specified using the computational methodology SCADS (Figure 1d). The resulting protein, PA_{SC}, exhibits complementary, asymmetric hydrophobic interactions within the interior that are difficult to realize in symmetric bundles (Figure 4). PA_{SC} was expressed from a synthetic gene (10−15 mg/L pure protein) in *E. coli*.

UV−Visible and CD Spectroscopy. UV−visible spectroscopy demonstrated that PA_{SC} selectively binds FeDPP(III) in the intended stoichiometry (2:1). Addition of protein to 2 equiv of FeDPP(III) resulted in a sharpening of the Soret band and a shift from 405 to 410 nm (Figure 5). Titration of PA_{SC} with FeDPP(III) indicates formation of a 2:1 cofactor to protein complex (Figure S3). Reduction of the PA_{SC}/FeDPP(III) complex with sodium dithionite resulted in a ferrous complex with spectral signatures typical of a bis-His ligated complex (λ_{max} = 525 nm).[6,37] The Soret peak for PA_{SC} is at 410 nm and shifts to 419 nm when reduced with solid sodium dithionite. For comparison, PA_{TET} showed a Soret peak at 408 nm, which shifts to 417 nm when reduced.[8] The PA_{SC}/FeDPP(III) complex gives an extinction coefficient of 177 000 M^{-1} cm^{-1}, whereas the PA_{TET} complex gives an extinction coefficient of 160 000 M^{-1} cm^{-1}, which are the same within the experimental error of the protein concentration determination.

Furthermore, the visible CD spectrum showed that each cofactor was bound in a unique chiral environment in PA_{SC}.[19] The UV−visible CD spectrum of the reduced species shows that both cofactors have been incorporated into the protein, with a change in the transition at 419 nm, consistent with the location of the peak of the ferrous species (Figure 6a). To evaluate the second binding site, the spectrum of 1 equiv each of protein and cofactor was subtracted from the spectrum of 1 equiv of protein and 2 equiv of cofactor (Figure 6b). This revealed that there is little difference in cofactor ligation at each site, because both sites produce a signal that is similar in magnitude and shape.

Size Exclusion Chromatography and Analytical Ultracentrifugation (AUC) of PA_{SC}/FeDPP(III). The association of PA_{SC} was examined by SEC, using a Superdex 75 analytical

(34) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(35) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(36) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(37) 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*, 425−431.

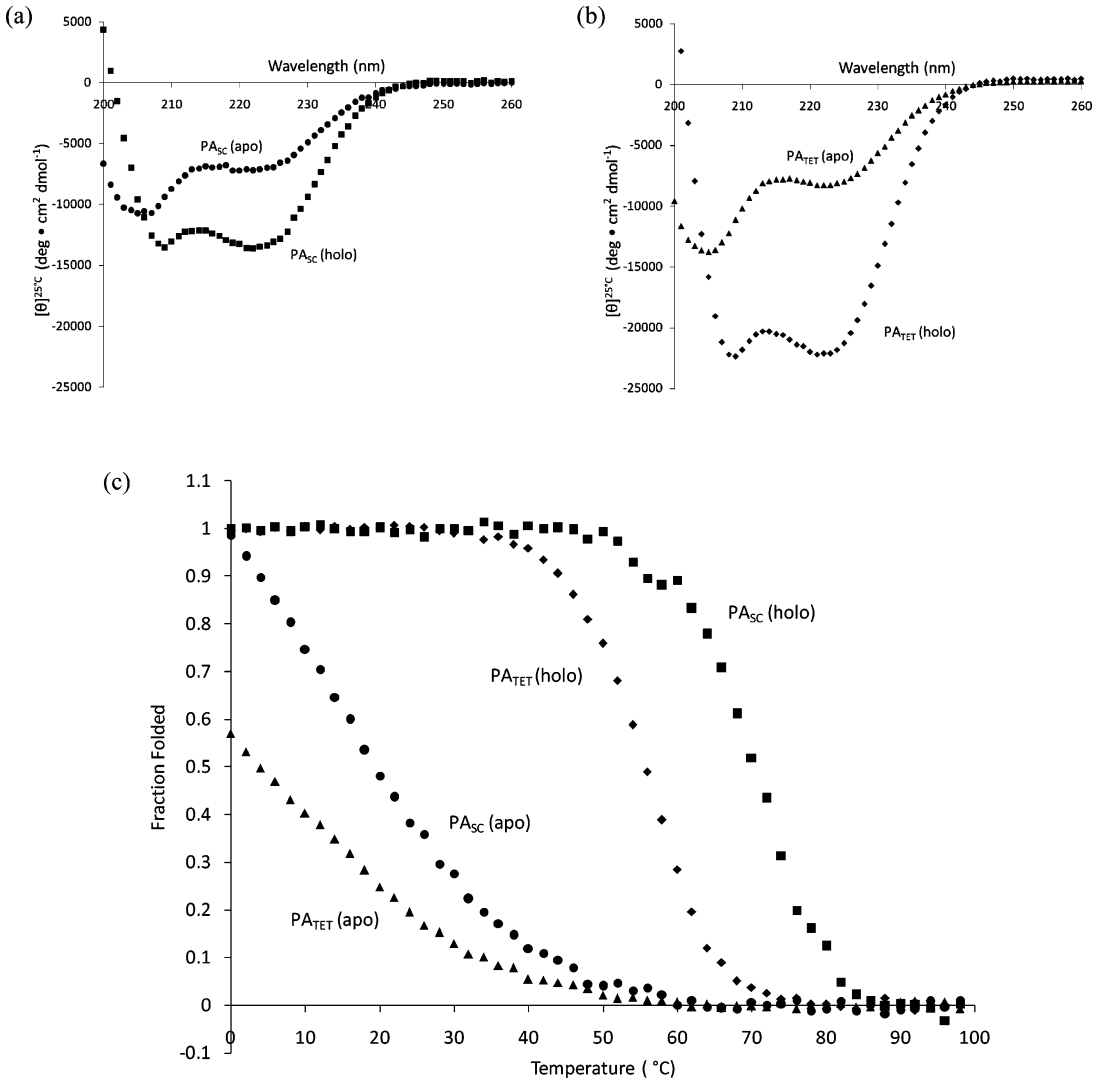


Figure 8. CD spectroscopy in the near-UV region. (a) The spectra of apo-PA$_{SC}$ (25 μM protein, ●) and holo-PA$_{SC}$ (25 μM protein + 50 μM FeDPP(III), ■) at 25 °C demonstrate the increase in helicity upon addition of FeDPP(III). (b) Spectra of apo-PA$_{TET}$ (50 μM protein, ▲) and holo-PA$_{TET}$ (50 μM protein + 25 μM FeDPP(III), ◆). (c) Fraction folded for PA$_{TET}$ (apo, holo) and PA$_{SC}$ (apo, holo) demonstrates that the T_m's (°C) for PA$_{SC}$ are higher than those for PA$_{TET}$, indicating increased stability.

column (Figure 7). In the absence of cofactor, the protein elutes in a symmetrical peak. Upon addition of 1 or 2 equiv of FeDPP-(III) per mole of PA$_{SC}$, the protein eluted at a smaller volume consistent with the compaction of the structure upon association with the cofactor. A second minor peak is also observed and is indicative of aggregation in the reconstitution assay. This peak can be minimized by reconstituting at a high dilution and annealing the sample at 70 °C. Once purified, the monomeric form is stable at millimolar concentrations for several weeks at room temperature. The monomeric peak, representing the PA$_{SC}$/FeDPP(III) complex, was further examined by AUC. AUC of the PA$_{SC}$/FeDPP(III) complex revealed that it sedimented as a single molecular species with molecular weight of 14 150 Da, in excellent agreement with that for the monomeric complex, 13 932 Da (Figure S1).

Far-UV CD and NMR Investigation of Backbone Conformation of PA$_{SC}$. Far-UV CD spectroscopy was used to examine the helical structure, stability, and the effects of porphyrin binding in the protein as well as that in PA$_{TET}$. The spectra of both apo-PA$_{SC}$ and apo-PA$_{TET}$ were nearly identical at room temperature (Figure 8a,b). The mean residue ellipticity for apo-PA$_{TET}$ ($[\theta]_{222nm,25°C}$) is reported to be -8300 deg cm^2 dmol^{-1}. Apo-PA$_{SC}$ is slightly less helical ($[\theta]_{222nm,25°C} = -7200$ deg cm^2 dmol^{-1}) because of the inclusion of the loop and also disorder near the N- and C-termini of PA$_{SC}$. This trend is also observed for the holoscaffolds; both show an increase in helical content upon addition of the cofactor and PA$_{SC}$ remains less helical than PA$_{TET}$ (Table 1).

Table 1. Summary of CD Data for PA$_{TET}$ and PA$_{SC}$, Both Apo- and Holoproteins

	$[\theta_{222nm}]^{0°C}$ [a]	T_m [b]	concn [c]
PA$_{TET}$ (holo) [d]	−23800	56	50
PA$_{TET}$ (apo)	−4200	4	50
PA$_{SC}$ (holo) [e]	−14500	68	25
PA$_{SC}$ (apo)	−3900	22	25

[a] In deg cm^2 dmol^{-1}. [b] Temperature (°C). [c] Concentration of protein scaffold (μM). [d] 25 μM FeDPP added to scaffold PA$_{TET}$. [e] 50 μM FeDPP added to scaffold PA$_{SC}$.

However, PA$_{SC}$ demonstrated higher stability than PA$_{TET}$, in both the apo and holo forms of the protein (Figure 8c). The addition of porphyrin resulted in a very large increase in the stability of PA$_{SC}$ (T_m values of the apo- and holoprotein were

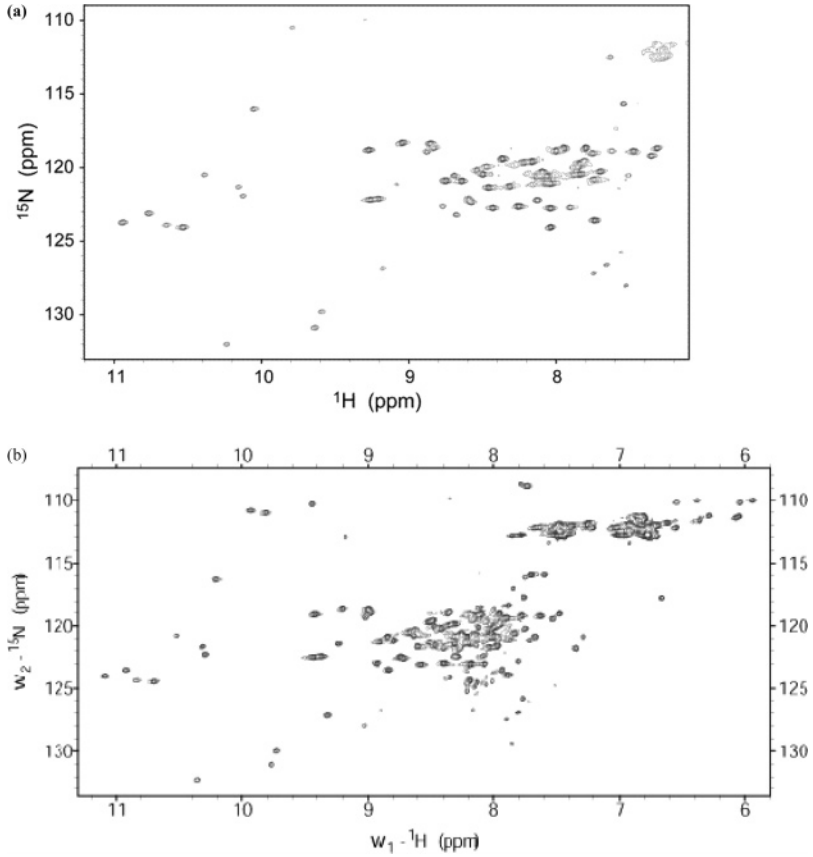


Figure 9. HSQC NMR of PA_{SC}. (a) The $^{1}H-^{15}N$ HSQC spectrum of 0.83 mM PA_{SC} and 1.6 mM FeDPP(III) in 50 mM phosphate buffer and 200 mM NaCl, pH 7.5, at 45 °C. (b) The $^{1}H-^{15}N$ HSQC spectrum of 1.1 mM PA_{SC} and 2.2 mM FeDPP(III) in 100 mM phosphate buffer and 150 mM NaCl, pH 6.0, at 45 °C.

22 and 68 °C, respectively, Table 1) and the steepness of the transition. Comparatively, the melting midpoints were determined to be 4 °C for apo-PA_{TET} and 56 °C for holo-PA_{TET}. This demonstrates the increased stability of the scaffold PA_{SC} because of inclusion of interhelical links as well as better interior packing. Given the increased stability and the fact that the scaffold demonstrated its designed behavior, it was important to establish whether the scaffold was a good structural candidate by evaluating the complex using NMR.

The NMR spectrum of PA_{SC} should provide valuable information concerning the degree to which the protein is uniquely folded, versus adopting a molten maquette-like conformation. Figure 9a shows that the ^{15}N proton HSQC spectrum of PA_{SC} with FeDPP(III) at pH 7.5 is indeed well-dispersed, and approximately 70 resonances are observed. The amide backbone resonances are spread over 7−11 ppm in the proton dimension and 110−132 ppm in the ^{15}N dimension, indicative of a well-defined, presumably uniquely folded structure. At pH 6.0, where the exchange with solvent is slower, we expect to observe more resonances that are better resolved (Figure 9b). The approximately 95 resonances are observed at pH 6.0, and they are dispersed over 6−11 ppm in the proton dimension and 108−134 ppm in the ^{15}N dimension, consistent with a natively folded structure.

Molecular Dynamics. MD calculations were performed to validate the retention of key design elements during an all-atom simulation of the molecular structure. If the sequence is well-suited to stabilize the starting three-dimensional model, the structure should equilibrate rapidly and then show only small local deviations from its equilibrium position. However, starting structures that are either unstable, far from their energy minimum, or dynamically averaging on the nanosecond time scale might be expected to continuously drift from their starting configuration during MD simulations.[38] The motions of PA_{SC} in a $50 \times 50 \times 72$ Å^3 periodic box of water were simulated for 7 ns. A plot of the rmsd of the observed structure (relative to the starting model) versus time shows the wild-type protein equilibrates within 200 ps and then remains essentially constant over the remaining 7 ns (Figure 10a). Examination of the contributions of the individual helices to the rmsd fluctuations showed that the first and fourth helices were most variable because of fraying of their N- and C-terminal sequences (Figure S4).

To determine the significance of the observed structural conservation of the structure over this time scale, we introduced four mutations at positions $i + 4$ from the His ligand. Previous work from the groups of Suslick and Benson[13−16] had shown that hydrophobic residues at these positions contribute significantly to the stability of heme−peptide complexes. Thus, Y27, M53, Y80, and L06 were mutated to alanine in a protein designated PA_{SC}-ALA, and the dynamics of this protein was simulated for 7 ns. Similar to the trajectory for PA_{SC}, PA_{SC}-ALA showed a rapid equilibration in approximately 200 ps, but then its structure continued to vary over the remainder of the simulation, never reaching a stable configuration (Figure 10c). This finding is reflected not only in the plots of rmsd versus

(38) Zou, H. L.; Strzalka, J.; Xu, T.; Tronin, A.; Blasie, J. K. *J. Phys. Chem. B* **2007**, *111*, 1823−1833.

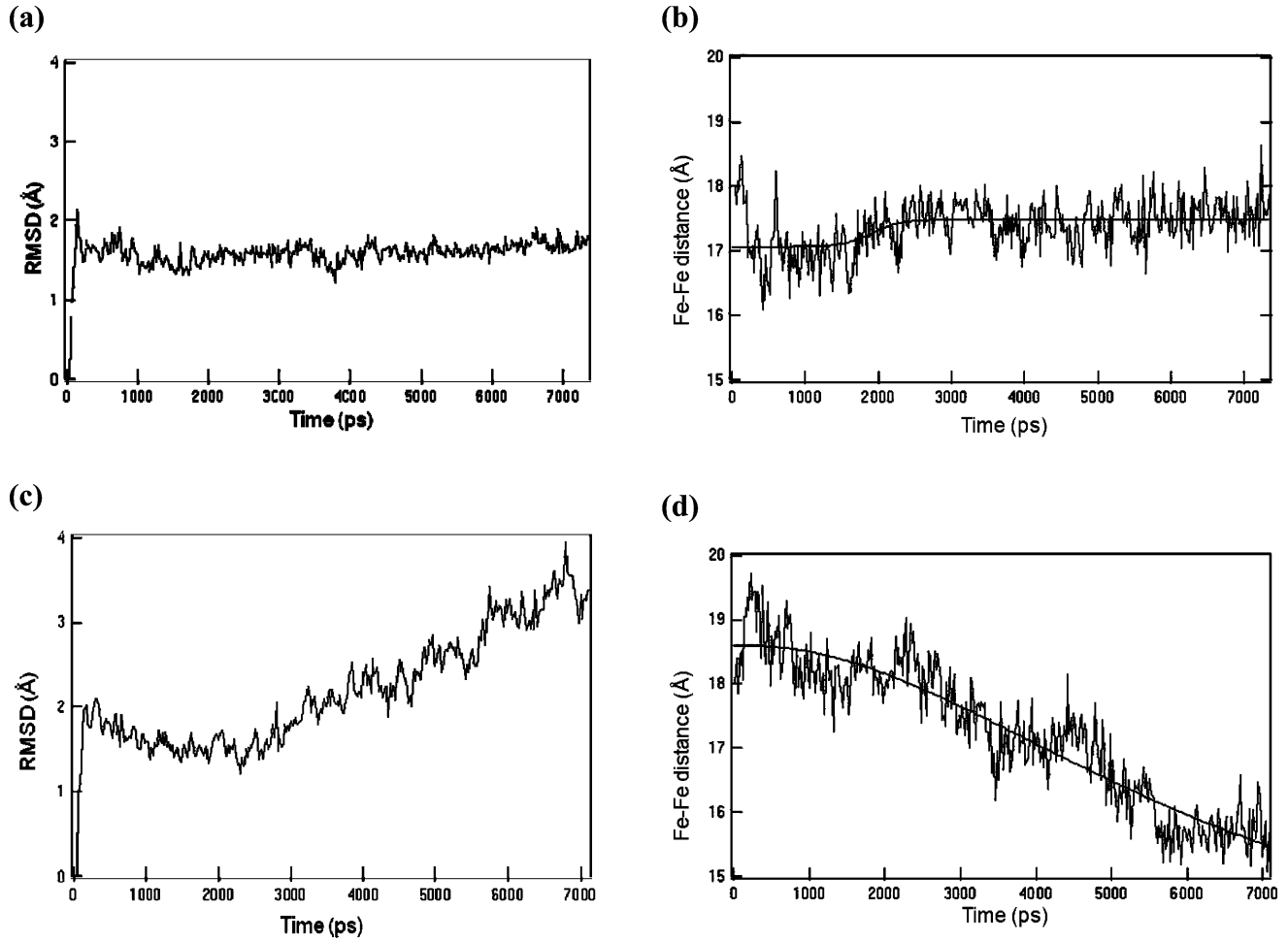


Figure 10. MD simulations. Comparison of trajectories of the backbone rmsd and the metal-to-metal distance versus time for the PA_{SC} scaffold. The wild-type system (PA_{SC}) backbone rmsd is shown in (a), whereas the mutant (PA_{SC}-ALA) backbone rmsd is in (c). The metal-to-metal distances for the wild-type system (b) are shown in contrast to those for the mutant system (d).

time but also from an examination of the distance between the metal ions in the two cofactors, which were essentially constant for PA_{SC} but became continuously smaller for PA_{SC}-ALA (Figure 10b,d). The computationally designed loops were also examined for stability, on the 7-ns time scale, and in PA_{SC} both the short as well as the long interhelical loops were highly stable on the nanosecond time scale (Table S1). Within the time of the simulation, the ϕ and ψ angles of the wild-type residues remained in the same basins of the Ramachandran plot (Table S1).

Conclusion

Here we demonstrate that modern methods of computational protein design were successfully used to yield a fully asymmetric protein that holds considerable promise for structural characterization as well as functional studies. Specifically, this genetically expressed scaffold allows for the facile incorporation of isotopic labels such as ^{15}N and ^{13}C necessary for multidimensional NMR experiments. Structural studies to understand the nature of the binding are underway and are made possible not only by the successful incorporation of the nonbiological cofactor, but also by the stability of PA_{SC} in solution. The selectivity of PA_{SC} for FeDPP(III) demonstrates the specificity of binding for a given cofactor. The design and incorporation of additional cofactors with various functionalities is now possible, and such designs are currently underway. In conclusion, de novo protein design is a promising and expanding approach to examine the delicate interplay between a cofactor's environment and its functional properties.

Acknowledgment. This work was supported by the DOE Grant DE-FG02-04ER46156 (J.K.B., W.F.D., J.G.S., M.J.T.) and the National Science Foundation Grant NSEC DMR-0425780. Partial support was also provided by the NIH Grants GM54616 to W.F.D.; GM61267 to J.G.S.; and GM071628 to M.J.T. We also thank James Lear for interpreting the AUC data. We thank the Spectroscopy Support Facility at the Fox Chase Cancer Center for access to NMR instumentation.

Supporting Information Available: AUC data, FeDPP(III) quantification, hemochrome assays, and MD simulation information. This material is available free of charge via the Internet at http://pubs.acs.org.

JA071199J