

De novo design of a hyperstable, non-natural protein-ligand complex with sub-Å accuracy

Nicholas F. Polizzi^{1,4}, Yibing Wu⁴, Thomas Lemmin⁴, Alison M. Maxwell⁴, Shao-Qing Zhang⁴,
Jeff Rawson², David N. Beratan^{1,2,3}, Michael J. Therien², William F. DeGrado⁴

¹Department of Biochemistry, ²Department of Chemistry, and ³Department of Physics, Duke
University, Durham, North Carolina 27710, USA

⁴Department of Pharmaceutical Chemistry, Cardiovascular Research Institute, University of
California, San Francisco, San Francisco, CA 94158, USA

Abstract

Protein catalysis requires atomic-level orchestration of side chains, substrates, and
cofactors, yet the ability to design a small-molecule-binding protein entirely from first principles
with a precisely predetermined structure has not been demonstrated. We designed a novel
protein, PS1, which binds a highly electron-deficient, non-natural porphyrin at temperatures up
to 100 °C. The high-resolution structure of holo-PS1 is in sub-Å agreement with the design.
The structure of apo-PS1 retains the remote core packing of the holo, predisposing a flexible
binding region for the desired ligand-binding geometry. Our results illustrate the unification of
core packing and binding site definition as a central principle of ligand-binding protein design.

Introduction

If we truly understand proteins, we should be able to design functional proteins
purposefully from scratch. While the de novo design of proteins has seen many successes¹⁻¹², no

1 small molecule ligand- or organic cofactor-binding protein has been designed entirely from first
2 principles to achieve i) a unique structure and ii) *a predetermined binding-site geometry with*
3 *sub-Å accuracy*. Such achievements are prerequisites for the design of proteins that control and
4 enable complex reaction trajectories, where the relative placements of cofactors, substrates, and
5 protein side chains must be established within the length scale of a chemical bond. Here, we
6 design a small molecule-binding protein based on the concept that the entire protein contributes
7 to establishing the binding geometry of a ligand¹³⁻¹⁶. Numerous mutational studies of natural
8 ligand-binding proteins have highlighted the counter-intuitive importance of distant amino acids
9 (10-20 Å from the binding site) on binding affinity, which work in concert with first-shell amino
10 acids surrounding the bound ligand¹³⁻¹⁶. We implement this concept for the first time in de novo
11 protein design. Hence, what are traditionally considered as separate sectors – the hydrophobic
12 core and ligand-binding site – we treat as an inseparable unit. We utilize flexible backbone
13 sequence design of a parametrically defined protein template to simultaneously pack the protein
14 interior both proximal to and remote from the ligand-binding site. Thus, tight interdigitation of
15 core side chains quite removed from the binding site structurally restrains the first- and second-
16 shell packing around the ligand. We apply this principle to the decades-old problem of structural
17 non-uniqueness in de novo-designed heme-binding proteins¹⁷. We designed a novel protein,
18 PS1, which binds a highly electron-deficient, non-natural porphyrin at temperatures up to 100
19 °C. The high-resolution structure of holo-PS1 is in sub-Å agreement with the design. The
20 structure of apo-PS1 retains the remote core packing of the holo, predisposing a flexible binding
21 region for the desired ligand-binding geometry. Our results illustrate the unification of core
22 packing and binding site definition as a fundamental principle of ligand-binding protein design.

Recent successes in the field of de novo design of coiled coils^{3,7} and metalloproteins^{4,8-10} are encouraging, but so far have not translated to more complex cofactors. In fact, attempts at computational design of novel small molecule ligand-binding proteins have been limited in number and generally focused on changing only the binding site of natural proteins, leaving the core of the protein intact^{18,19}. For example, the binding site of a natural scaffold was computationally redesigned to bind a hydrophobic organic ligand but required multiple rounds of mutagenesis and experimental selection using yeast display¹⁸. At the other extreme, de novo heme-binding helical bundle proteins have been designed entirely from first principles (reviewed in refs. 17, 20), but these “maquettes” have evaded structural determination, largely due to aggregation or their dynamical properties^{17,21,22}. With the exception of short, covalently linked peptide-heme complexes²³, the only structure of a de novo heme-binding protein was solved for an apo-protein, which showed a hydrophobically collapsed binding site with no space for binding heme^{21,24}. The lack of precise, predictive three-dimensional models of heme-binding maquettes, coupled with the failure to determine high-resolution structures, has limited their utility, although maquettes have elucidated electrostatic roles for tuning redox potentials of donors/acceptors in electron-transfer reactions²⁰. An iterative trial-and-error approach has been shown to incrementally improve NMR spectra of maquette proteins²⁵, and may ultimately lead to the determination of three-dimensional structures; however, a robust computational method is needed to deliver *precisely predetermined* structures with sub-Å accuracy.

Our own work has focused on the development of computational design of cofactor-binding proteins²⁶⁻²⁸ with atomic-level accuracy. We used a step-wise strategy in which we first employed a mathematical parameterization of an antiparallel coiled coil to construct a rigid binding site, then, in a separate calculation, introduced side chain packing constrained by this

1 rigid backbone²⁶⁻²⁸. This approach resulted in de novo porphyrin-binding proteins with the
2 desired tertiary structure and ligand-binding stoichiometry, but not of sufficient conformational
3 uniqueness to yield a high-resolution structure.

4 An extensive body of work with natural proteins¹³⁻¹⁶ has shown that side chain packing
5 quite distant from the binding site can propagate to significantly affect ligand binding, catalysis,
6 and allosteric regulation. Thus, the entire hydrophobic core – even residues 20 Å away from the
7 binding site – should be considered as an essential extension of the primary and secondary shell
8 interactions with the ligand. We noted that, unlike natural proteins (Fig. 1a), previous de novo
9 designed cofactor-binding proteins lack an extensive, well-defined apolar core. Instead, their
10 interior packing is dominated by interactions with one or more porphyrins or multi-functional
11 cofactors that span the length of the bundle (Fig. 1b). Where a cofactor-free core was included²⁹,
12 the core was not computationally designed, and high-resolution structures were not determined.
13 Here, we purposefully include a folded core remote from the ligand-binding site and optimize its
14 sequence and structure *in concert with the binding site* to ensure appropriate coupling (Fig. 1c).
15 As compared to earlier computational design of ligand-binding proteins^{11,18} our approach differs
16 by: 1) beginning with a mathematically parameterized backbone rather than a natural protein; 2)
17 applying flexible backbone design to the entire backbone as well as sequence design to all
18 interior and substrate-binding sites rather than just the first and second-shell contacting residues;
19 3) not relying on screening of large numbers of designs or genetic selections to achieve the
20 desired outcome.

22 Results

23 Protein design

The design of PS1 (Porphyrin-binding Sequence 1) began with the previously parameterized backbone from the de novo designed protein SCRPZ-2²⁸, a protein that bound an extended porphinato(metal)-polypyridyl(metal) cofactor (Fig. 1b). The backbone of SCRPZ-2 and its di-porphyrin-binding predecessors^{26,30} was designed with a simple equation defining a D_2 -symmetrical antiparallel coiled coil³¹. The parameters were adjusted to position a single His ligand to receive a second-shell hydrogen bond with Thr from a neighboring helix (see Fig. 2b). Side chains in the vicinity of the binding site were computationally designed to stabilize the asymmetric ligand environment while maintaining a rigid symmetrical backbone. Interhelical loops were then chosen following previously defined geometric principles^{26,28,32}. Although SCRPZ-2 bound to its desired cofactor, its NMR spectra was not as well dispersed as those for natural heme-containing proteins, and it lacked a cooperatively folded core.

We used the parameterized backbone of SCRPZ-2 as a starting point for design of a protein that binds a much smaller abiological porphyrin (CF₃)₄PZn ([5,10,15,20-tetrakis(trifluoromethyl)porphinato]Zn²⁺) (Fig. 2)³³, [a powerful photo-oxidant with an excited-state reduction potential similar to the ground-state reduction potential of the oxidized special pair of chlorophylls in photosystem II of green plants](#)³⁴. The reduced size of the (CF₃)₄PZn cofactor provided space for a hydrophobic core in what was formerly occupied by the large, bulky metal-polypyridyl group. We manually docked (CF₃)₄PZn in the porphyrin-binding site (Fig. 2b) and used Backrub within Rosetta³⁵ to sample small structural changes of the parameterized backbone; we then employed alternating loops of fixed backbone sequence design and backbone/sidechain minimization (see Methods). The models were assessed for packing of the porphyrin as well as the core. To isolate effects of introducing a well-defined hydrophobic core, we allowed sequence changes only in the protein interior and cofactor-binding site, keeping

the identities of most solvent-exposed and loop residues fixed from that of SCRPZ-2. The final sequence of PS1 shares no similarity with any known natural protein (BLAST *E* value < 0.06 against the non-redundant protein sequence database *nr*). Although the final backbone model of PS1 differed by only 1 Å root mean square deviation (RMSD) from the initial parameterized backbone of SCRPZ-2, fully 70% of the interior residues were changed from SCRPZ-2, and half of those retained were predicted to adopt different rotamers (Fig. S1).

Biophysical characterization of PS1

PS1 is monomeric (Fig. S2) and binds the water-insoluble cofactor, (CF₃)₄PZn, forming highly thermostable complexes (extrapolated *T*_m > 120 °C, Fig. 3c and Fig. S3) that are stable for over a year. The complex forms within seconds of adding (CF₃)₄PZn from organic solution to aqueous PS1, suggesting a small kinetic barrier for assembly (Fig. 3a). A tight dissociation constant of binding, *K*_D = 45 nM, was measured under conditions where the water-insoluble porphyrin was solubilized with 1% w/v octylglucopyranoside detergent (Fig. 3b). PS1 also binds the ferrous iron-derivative of the porphyrin, (CF₃)₄PFe (Fig. S4), despite the abysmal solubility in water of this cofactor. Loading of PS1 with (CF₃)₄PFe suggests that the protein could also be used as a platform for engineering ground-state redox chemistry, as (CF₃)₄PFe is an electron-deficient (porphinato)metal complex capable of molecular oxygen activation for alkane hydroxylation and alkene epoxidation³⁶.

Time-resolved transient absorption spectroscopy showed that protein/(CF₃)₄PZn interactions are preserved even at near-boiling temperatures where the protein retains its native structure (Fig. 3d). The excited-state spectra and dynamics of (CF₃)₄PZn within holo-PS1 at 21 and 100 °C are indistinguishable, which indicates that the protein does not detectably perturb the porphyrin molecular framework—intersystem crossing rates of electronically excited porphyrins

are known to be sensitive to temperature and environment³⁷. Furthermore, these data indicate that encapsulation of (CF₃)₄PZn in the binding site of PS1 shields the porphyrin from nucleophilic attack that would otherwise occur in water, especially at high temperatures, i.e. the protein safeguards the porphyrin against a wasteful, degradative, photochemical side reaction. Thus, PS1 effectively stabilizes an extraordinarily insoluble cofactor in aqueous solution, even at temperatures considered extreme for hyperthermophiles.

We also examined another high-scoring sequence (named PS2) of the design process, with a hydrophobic core unique from PS1, which was expressed, purified, and tested for binding to (CF₃)₄PZn (see Methods and supplementary information). Electronic absorption spectra of holo-PS2 shows narrow absorption bands similar to those evinced by holo-PS1 (Fig. S5), strongly suggesting that these designs analogously enfold the porphyrin in a unique binding environment, although structural characterization of PS2 has not yet been pursued.

Structural characterization of holo-PS1

An exceptionally well-resolved NMR structural ensemble of holo-PS1 (Figs. 4 and 5) was computed using 19 nuclear Overhauser effects (NOEs) per residue and nearly complete ¹D_{NH} residual dipolar coupling restraints. The backbone is in excellent agreement with the design (0.8 ± 0.1 Å helical backbone RMSD), and core residues each populate a single rotamer state, almost all in agreement with the design (Fig. 4a,d,e). While the PS1 design was selected based in part on its featuring an abundance of high-probability rotamer states of core residues, two low-probability rotamers were present in the designed core of PS1: one, Leu98, in the first-shell of the binding site, and the other, Leu19, in the remote folded core. Binding of the porphyrin forces Leu98 to adopt this low-probability rotamer, which is not present in the apo-protein (see below, Fig. 5e), whereas Leu19 adopts a more probable rotamer in both the holo- and apo-proteins.

Trp68, fit snugly between two CF₃ groups of the porphyrin, can also be seen to adopt its predicted rotamer upon binding of the porphyrin, driving a unique conformation of the cofactor within the binding site.

The location and orientation of the porphyrin within PS1 was determined by an exceptional number of porphyrin-protein NOEs (26 porphyrin-protein NOEs were used in the structural refinement, Fig. 4b). Most importantly, the observed orientation of the cofactor is exactly as designed, within the precision of the NMR structure (Fig. 4c). (CF₃)₄PZn was only displaced in its binding site relative to its predetermined orientation by an average translation (0.4 Å) half the size of a covalent C-H bond, and by a small average rotation (11°) within the porphyrin plane.

Ab initio folding predictions and NMR structure of apo-PS1

We wondered how the folded core of PS1 might contribute to favorable binding dynamics, as the ligand binding must compete with precipitation of the water-insoluble ligand. Ab initio folding³⁸ simulations of the apo-PS1 sequence predict a bipartite structure with a conformationally unique folded core, which closely resembles the core of holo-PS1, and a more flexible cofactor-binding region (Fig. 2c). Significantly, hydrophobic collapse in the binding region is avoided, because it contains a polar His and also is rich in small Ala and Gly side chains (Fig. 4d) to specifically associate with the face of the porphyrin ring, rather than the large hydrophobic residues used to stabilize hemes in maquettes. Thus, “negative design” in PS1 is implicitly achieved through the construction of a relatively polar cofactor-binding site, which creates a cofactor-shaped void in the apo-protein.

The NMR structure of apo-PS1 was also solved (Fig. 5), and the structural ensemble shows a folded core highly similar to that of holo-PS1. This finding indicates that the folded core

both predisposes and anchors the flexible binding region for productive binding of the ligand. The binding region is more dynamic in apo-PS1, which contains two clusters of structures, open and closed (Fig. S6). The open conformation likely facilitates binding of the large cofactor, but there is room for water to penetrate into the unoccupied binding site in both conformations.

Dynamics and structural comparisons of apo- vs holo-PS1

Solvent hydrogen-deuterium exchange (HDX) experiments and molecular dynamics simulations of apo-PS1 also show a gradient in conformational stability between the apolar core and the binding site of apo-PS1 (Fig. 5c, Figs. S7 and S8). The backbone surrounding the apolar core of both holo- and apo-PS1 is highly protected from exchange, an important characteristic of cooperatively folded native proteins. The protected region extends into the porphyrin-binding site in the holo-protein but not in the apo-structure (Fig. 5c). The increased protection in the binding site of holo-PS1 is seen at both solvent-exposed and interior positions, indicating increased conformational stability rather than steric restriction from the bound cofactor alone.

In both the apo- and the holo-structures, the interior side chains stack into four layers, beginning at the edge of the porphyrin-binding site and extending to the end of the bundle (Fig. 5, d-f). In the absence of cofactor to constrain and stabilize the tightly packed conformation of the holo-protein, the layers closest to the binding site explore more conformations, accessing rotamers not seen in holo-PS1 (Fig. 5e). By contrast, the packing of the more distal layers is identical in the apo- and holo-structures (Fig. 5f). Thus, the third- and fourth-shell layers, located up to 20 Å away from the binding site, are precisely preorganized to stabilize the conformation of the first-shell side chains when PS1 enfolds its cofactor. This finding is consistent with numerous studies on natural proteins¹³⁻¹⁶, which show that variation of residues

involved in core packing distant from an active site can have profound influences on binding and catalysis.

Discussion

The vast improvement in conformational specificity between PS1 and earlier designs illuminates the importance of considering hydrophobic core packing and the construction of ligand-binding sites as a joint optimization problem during computational design. Our previous studies indicate that the use of rigid backbones optimized for ligand-protein interactions alone are insufficient for conformational uniqueness without explicitly considering and designing a backbone that can also accommodate a well-defined apolar core. Similarly, attempts to radically change specificity of natural proteins by varying their binding sites, while treating the surrounding protein matrix as a rigid unit of fixed sequence, has required subsequent experimental optimization via extensive rounds of random mutagenesis and selection^{18,19,39}. The reliance on experimental methods such as directed evolution and genetic selections, while currently useful in many practical applications¹⁹, speaks to our incomplete understanding of protein structure and function, and the need to test and refine this knowledge through design. It is noteworthy that the first sequence designed and tested via our approach succeeded without need for experimental screening. Furthermore, another high-scoring protein design also bound the cofactor, suggesting a possible generality of the method within the helical bundle protein family. Future work will apply this principle to other families of tertiary structures. These studies bring chemists closer to the ultimate goal of the computational design of fully functional proteins with properties unprecedented in nature.

Figures

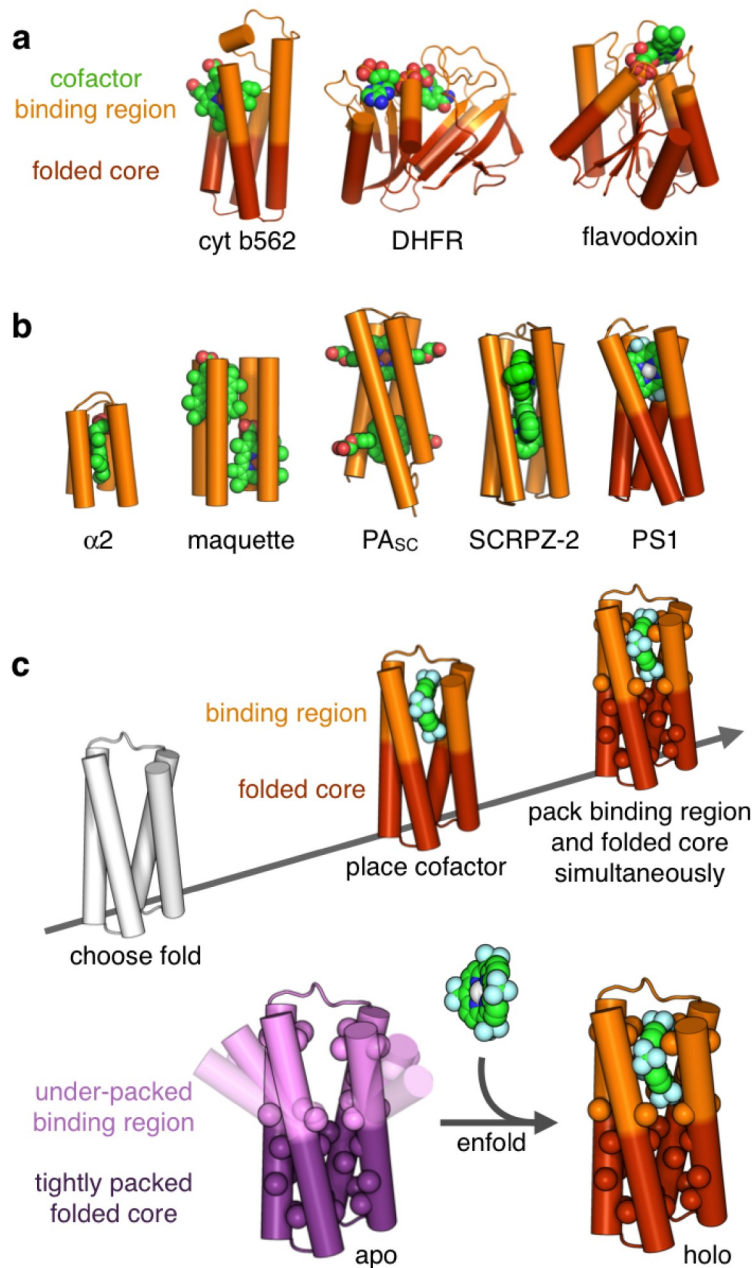
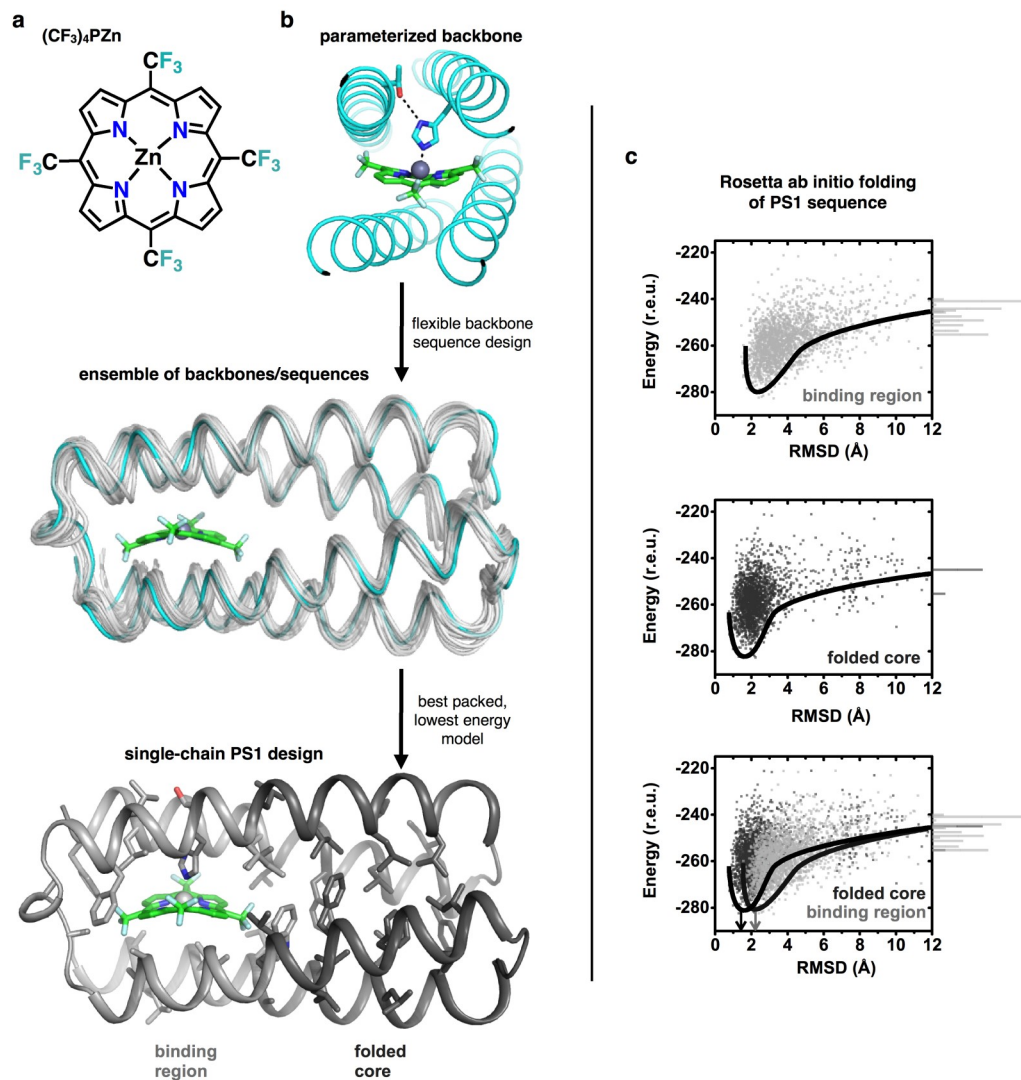
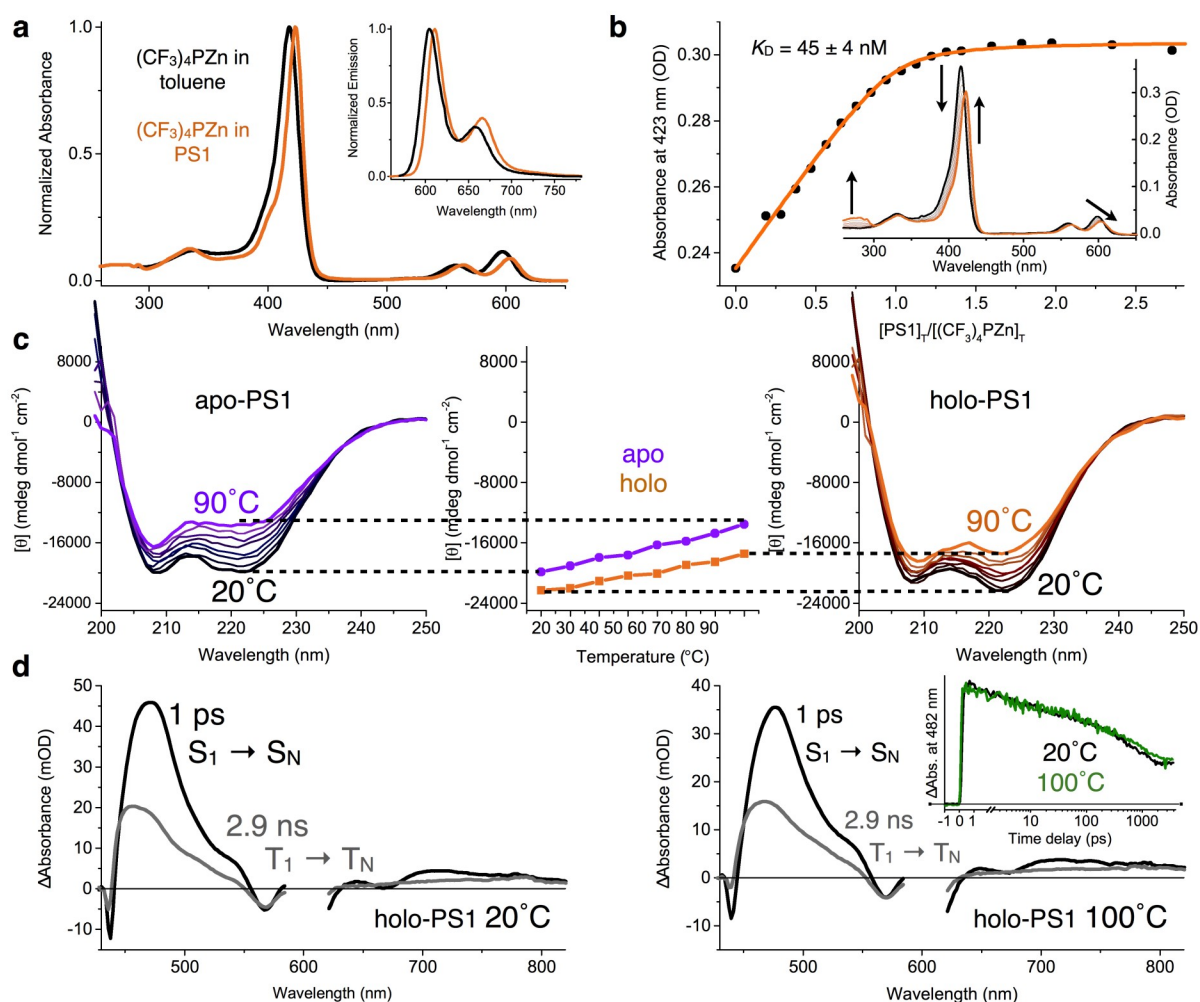


Figure 1 | The design strategy. **a**, Structures of natural cofactor-binding proteins show a folded core supporting a cofactor-binding region. **b**, Examples of previously designed tetra-helical porphyrin-binding proteins; all but PS1 (this work) lack a folded core (dark red). $\alpha 2$ protein is from ref 40; the remainder are described in the text. **c**, The design process starts with a parameterized backbone, which undergoes simultaneous optimization of packing of core residues (shown as spheres) in the binding region (light color) and folded core (dark color), with flexible

backbone. The resultant holo-protein (red) is tightly packed both in the binding region and in the folded core, whereas the apo-protein (purple) is tightly packed only in the folded core, which anchors the under-packed binding region to bind the cofactor.



- 1 chains and backbone to simultaneously conform to the porphyrin (CF₃)₄PZn (green). **c**, Ab initio
- 2 folding predictions of PS1 sequence. The Rosetta folding algorithm predicts a shallow folding



1 normalized emission spectrum of $(\text{CF}_3)_4\text{PZn}$ upon electronic excitation at 405 nm ($\text{OD} = 0.1$ at
2 excitation wavelength); buffer = 100 mM NaCl, 50 mM NaPi, pH 7.5. **b**, Determination of K_D
3 by apo-PS1 titration into a buffer solution (100 mM NaCl, 50 mM NaPi, pH 7.5) of $(\text{CF}_3)_4\text{PZn}$
4 with 1% w/v octyl- β -D-glucopyranoside. Inset shows spectral shifts upon porphyrin binding to
5 PS1. **c**, Circular dichroism (CD) spectra of apo- and holo-PS1 in 50 mM NaPi, 100 mM NaCl,
6 pH 7.5 as a function of temperature. The transitions appear reversible based on the fact that the
7 spectra are identical after cooling to room temperature. Units are in molar residue ellipticity.
8 Electronic absorbance spectra indicate holo-PS1 retains the porphyrin upon cooling. **d**, Pump-
9 probe transient absorption spectra of $(\text{CF}_3)_4\text{PZn}$ bound in the interior of holo-PS1 at 21 °C and
10 100 °C. The black spectrum shows characteristic $S_1 \rightarrow S_N$ absorptions of $(\text{CF}_3)_4\text{PZn}$, which
11 smoothly transitions into the gray spectrum showing characteristic $T_1 \rightarrow T_N$ absorptions of
12 $(\text{CF}_3)_4\text{PZn}$. Inset exemplifies identical transient dynamics (primarily intersystem crossing from
13 S_1 to T_1) at $\Delta\text{Abs.} = 482$ nm (scaled). Experimental conditions: solvent = 50 mM NaPi, 100 mM
14 NaCl, pH 7.5; excitation wavelength = 600 ± 5 nm; magic-angle polarization between pump and
15 probe pulses; pump-probe cross-correlation of ~ 250 fs.

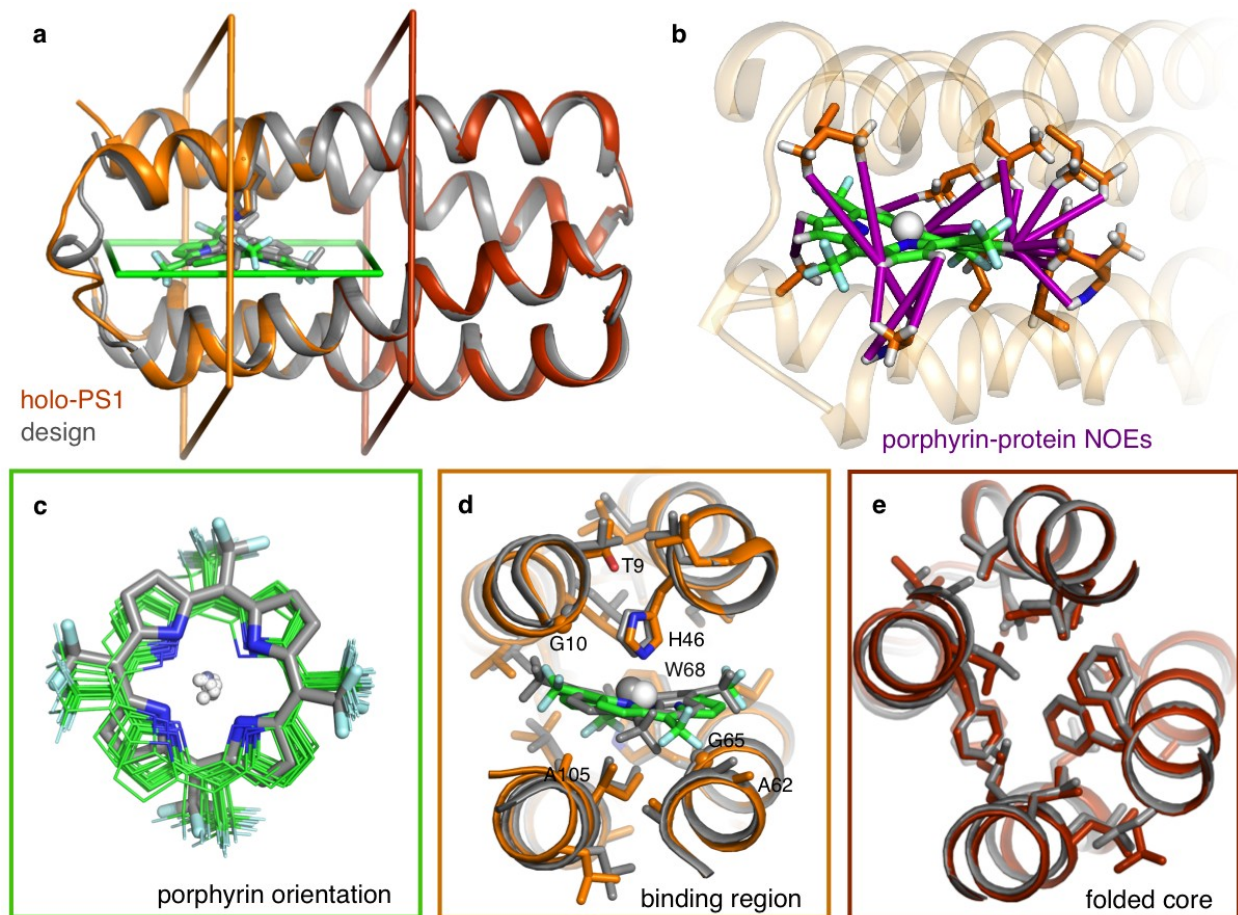


Figure 4 | The structure of holo-PS1 agrees closely with the design. **a**, The structure of holo-PS1 (orange) superimposed on the design (gray), with mean helical backbone RMSD of 0.8 ± 0.1 Å. The holo-PS1 model shown is the centroid of the NMR structural ensemble. **b**, 26 porphyrin-protein nuclear Overhauser effects (NOEs), drawn as purple sticks, experimentally determine the orientation of the porphyrin within the binding site of PS1. **c**, compares observed (green) vs. designed (gray) orientations. All hydrophobic and helical backbone heavy atoms within 4 Å of porphyrin heavy atoms in the design were used for alignment (0.9 ± 0.1 Å all-atom RMSD). **d,e**, shows ~ 10 Å slices of the holo-PS1 NMR centroid and design in the binding region and folded core, respectively.

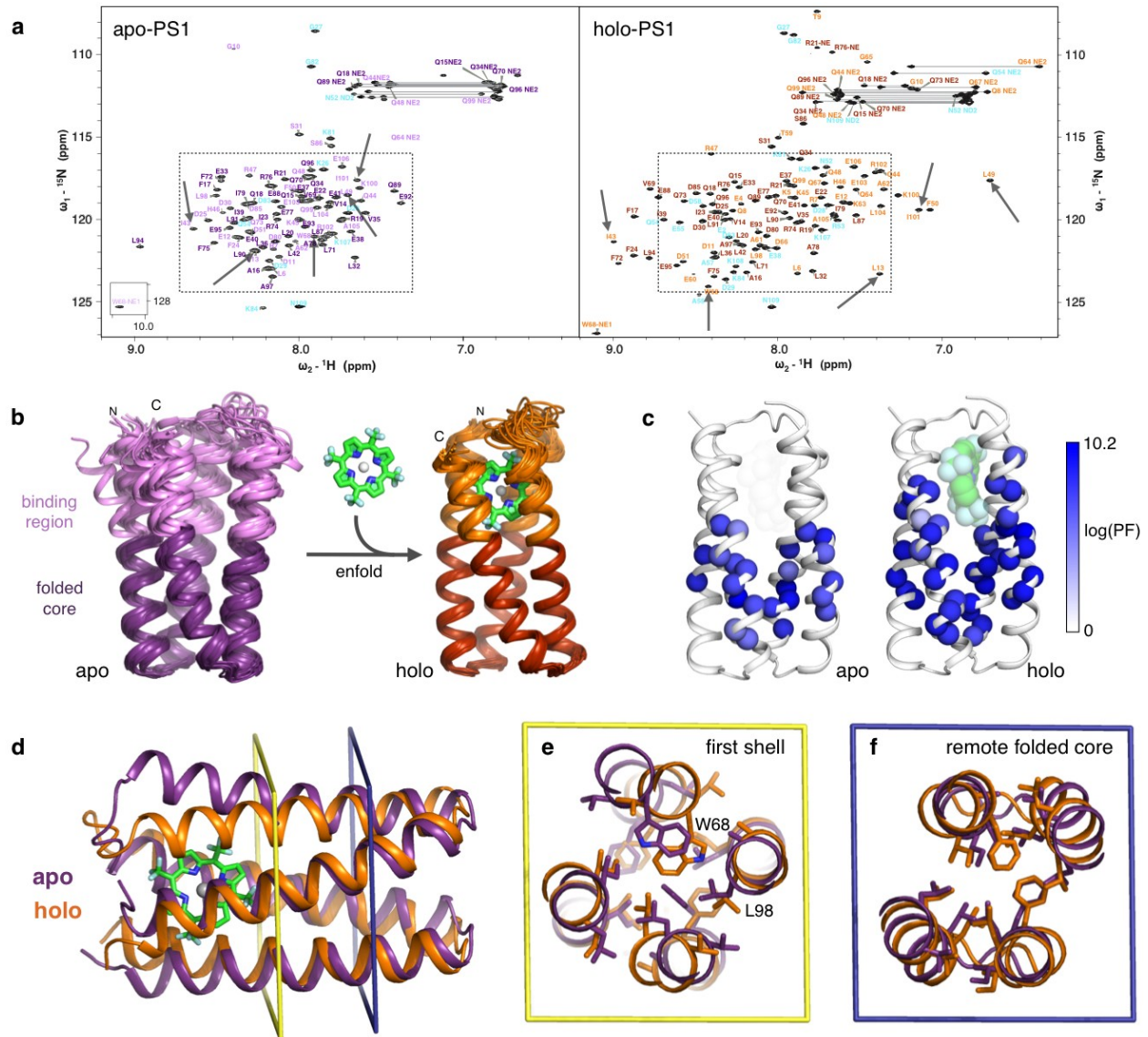


Figure 5 | Apo- and holo-PS1 share similar folded cores and differ in the binding region. a, 2D ^1H - ^{15}N HSQC spectra acquired for apo- and holo-PS1. Experimental conditions: 0.78 mM at 298K, 50 mM NaPi, 100 mM NaCl, pH 7.5, in 5% D_2O . Resonance assignments are indicated using the one-letter amino acid code. Signals arising from side chains (Asn HD2/ND2, Gln HE2/NE2, Arg HE/NE and Trp HE1/NE1) are also labeled. The residues belonging to the binding region and folded core are color-coded as in (b). Non-helical residues are labeled in cyan font face. The inset in the HSQC spectrum of apo-PS1 shows the chemical shift of the

indole proton of Trp68 near 10.2 ppm. A dashed box surrounds 90% of the backbone resonances of apo-PS1 and is also placed at the same position in the holo-PS1 spectrum. Arrows point to resonances of residues within the binding region that change dramatically upon binding of the cofactor. **b**, Solution NMR structures of apo-PS1 (purple) and holo-PS1 (orange). The structures were aligned to the backbone of the helical folded core of the lowest energy holo-PS1 model. Terminal residues 1, 108, and 109 are not shown for clarity. **c**, Hydrogen-deuterium exchange protection factors (PF) measured for apo- and holo-PS1, mapped onto the centroid structure of holo-PS1. Backbone amide nitrogens of residues with determined PFs are shown as spheres. Not shown: N of Trp68 indole side chain is protected in holo, but not apo. **d-f**, Backbone alignment of the holo- and apo-centroids at the folded core shows, **f**, agreement of side chain rotamer states far from the binding site and, **e**, differences in first-shell rotamers (e.g., Trp68, Leu98) accompanied by changes in backbone of the binding region. Centroids are from NMR structural ensembles clustered via RMSD of core side chain heavy atoms.

Methods

PS1 design process. Full methods and scripts regarding the design of PS1 can be found in the supplementary information. Briefly, the entire core of the D_2 -symmetrical parameterized backbone of SCRPZ-2 was redesigned to bind $(\text{CF}_3)_4\text{PZn}$ via a customized Rosetta script for flexible backbone sequence design. The flexible backbone design protocol was as follows: Distance and angle constraints between His and Zn were loaded, the model was repacked without mutations, the backbone was relaxed via Rosetta Backrub, three trials of a Monte Carlo flexible backbone design sub-protocol (see supplement) were performed, and models with native protein-like packing (i.e., a Rosetta PackStat score ≥ 0.58) were output. 170 designs were output

from 500 runs through the protocol (Fig. S1). We analyzed these 170 models for packing, radius of gyration, energy, and rotamer state probability within Matlab to select PS1 for expression. The design of PS2 proceeded in the same fashion.

Protein expression, purification, and biophysical characterization. Details regarding protein expression, purification, and biophysical characterization can be found in the supplement.

Briefly, genes for the proteins were ordered from GenScript, cloned into a pET-11a plasmid, and purified via a Ni column, followed by His-tag cleavage by TEV protease. The protein sequence of expressed, purified PS1 after His-tag cleavage is:

SEFEKLRQTGDELVQAFQRLREIFDKGDDDSLEQVLEEIEELIQKHRQLFDNRQEAAADTE
AAKQGDQWVQLFQRFREAIKGDKDSLEQLLEELEQALQKIRELAEKKN. The sequence for PS2 can be found in the supplement.

Porphyrin binding to apo-protein. A 2-fold excess of the cofactor $(CF_3)_4PZn$ was added from a 4 mM DMSO stock solution to a 50 mM NaPi, 100 mM NaCl, pH 7.5 buffer with apo-protein (Note that final DMSO concentrations were kept $< 1\%$). Buffer solution of apo-PS1 protein was heated for 5 minutes at 50 °C, $(CF_3)_4PZn$ was then added from DMSO stock, the resultant mixture was vortexed for 5 seconds, and placed back in the heat block at 50 °C for 15 minutes, with vortexing every 3 minutes. The protein/cofactor solution was then spun at 14000 x g in a Amicon Ultra-0.5 mL centrifuge filter for 10 min, three times, replacing the buffer to 0.5 mL after each 10 min spin. Finally, the protein solution was spun for 4 min at 12000 x g in an Amicon ultrafree-MC GV filter (UFC30GV0S). The holo-PS1 sample was then used for spectroscopic experiments immediately afterward, and diluted to an appropriate concentration if

necessary. Binding of $(\text{CF}_3)_4\text{PFe}$ was carried out in the same fashion, with the exception that the porphyrin was first dissolved in a stock of DMSO/ CHCl_3 .

Nuclear magnetic resonance spectroscopy. NMR spectra were recorded at 298 K on a 900 MHz Bruker Avance II spectrometer equipped with cryogenic probe for the holo-protein or on a Bruker 600 MHz spectrometer equipped with cryogenic probe for the apo-protein. Sequence specific backbone ($^1\text{H}^{\text{N}}$, ^{15}N , $^{13}\text{C}^{\alpha}$, ^{13}CO) and $^{13}\text{C}^{\beta}$ resonance assignments were obtained by using 3D HNCACB / CBCA(CO)NH and 3D HNCO / CO(CA)NH along with the program AUTOASSIGN⁴¹. $^1\text{H}^{\alpha}$ and $^1\text{H}^{\beta}$ assignments were extended by 3D HAHB(CO)NH experiment and more peripheral side chain chemical shifts were assigned with aliphatic 3D CCH-TOCSY (mixing time: 75 ms) and simultaneous 3D $^{15}\text{N}/^{13}\text{C}^{\text{aliphatic}}/^{13}\text{C}^{\text{aromatic}}$ -resolved [^1H , ^1H]-NOESY (mixing time: 120 ms). Overall assignments were obtained for 98.1% and 95.9% of the backbone (excluding the N-terminal NH_3^+) and $^{13}\text{C}^{\beta}$, and for 97% and 94.6% of the side chain chemical shifts (excluding Lys NH_3^+ , Arg NH_2 , OH, side chain ^{13}CO and aromatic $^{13}\text{C}^{\gamma}$) for the holo- and apo-proteins, respectively. All spectra were processed and analyzed with the programs NMRPIPE and XEASY, respectively^{42,43}. ^1H - ^1H upper distance limit constraints for structure calculations were extracted from NOESY. In addition, backbone dihedral angle constraints were derived from chemical shifts using the program TALOS for residues located in well-defined secondary structure elements⁴⁴. 2D constant-time [^{13}C , ^1H]-HSQC spectra were recorded as was described for the 5% fractionally ^{13}C -labeled samples to obtain stereo-specific assignments for isopropyl groups of Val and Leu⁴⁵. The $^1\text{D}_{\text{NH}}$ residual dipolar couplings (RDCs) were measured with 2D ^1H - ^{15}N IPAP-HSQC in samples aligned using Pfl phage (ASLA biotech). The program CYANA was used to assign long-range NOEs and calculate the structure^{46,47}. Backbone $^1\text{D}_{\text{NH}}$

RDCs were used as orientational constraints for the later stages of refinement with XPLOR-NIH⁴⁸. The final set of structures was further refined by restrained molecular dynamics in explicit water⁴⁸. NMR structure quality was assessed with the Protein Structure Validation Software Suite (PSVS)⁴⁹ (Table S4).

Hydrogen-deuterium exchange measurements. For the measurements of H/D exchange rates, a series of 2D ¹⁵N HSQC spectra were obtained on a 900 MHz Bruker Avance II spectrometer. The first spectra were recorded 9 minutes after the dilution of 100 µl of a high concentration sample in H₂O (2 mM for apo and 1.2 mM for holo) into 200 µl D₂O buffer. 15-min HSQC spectra were recorded successively in the first 12 hours, a 15-min spectrum in every hour in the second 12 hours, a 15-min spectrum in every two hours in the third 12 hours, and so on. The last points were 2730.6 and 4903.5 min for apo and holo, respectively. For the H/D exchange rate analysis, the peak height of each isolated peak was extracted by nmrDraw and fitted to one-phase exponential decay.

References

- 1 Roy, S. *et al.* A protein designed by binary patterning of polar and nonpolar amino acids displays native-like properties. *J. Am. Chem. Soc.* **119**, 5302-5306 (1997).
- 2 Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364-1368 (2003).
- 3 Nanda, V. & Koder, R. L. Designing artificial enzymes by intuition and computation. *Nat. Chem.* **2**, 15-24 (2010).

- 1 4 Peacock, A. F. A. Incorporating metals into de novo proteins. *Curr. Opin. Chem. Biol.*
2 **17**, 934-939 (2013).
- 3 5 Huang, P.-S. *et al.* High thermodynamic stability of parametrically designed helical
4 bundles. *Science* **346**, 481 (2014).
- 5 6 Thomson, A. R. *et al.* Computational design of water-soluble α -helical barrels. *Science*
6 **346**, 485 (2014).
- 7 7 Woolfson, D. N. *et al.* De novo protein design: How do we expand into the universe of
8 possible protein structures? *Curr. Opin. Struct. Biol.* **33**, 16-26 (2015).
- 9 8 Mocny, C. S. & Pecoraro, V. L. De novo protein design as a methodology for synthetic
10 bioinorganic chemistry. *Acc. Chem. Res.* **48**, 2388-2396 (2015).
- 11 9 Ulas, G., Lemmin, T., Wu, Y., Gassner, G. T. & DeGrado, W. F. Designed
12 metalloprotein stabilizes a semiquinone radical. *Nat. Chem.* **8**, 354-359 (2016).
- 13 10 Olson, T. L. *et al.* Design of dinuclear manganese cofactors for bacterial reaction centers.
14 *Biochim. Biophys. Acta: Bioenergetics* **1857**, 539-547 (2016).
- 15 11 Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design.
16 *Nature* **537**, 320-327 (2016).
- 17 12 Brunette, T. J. *et al.* Exploring the repeat protein universe through computational protein
18 design. *Nature* **528**, 580-584 (2015).

- 1 13 Bollen, Y. J. M., Westphal, A. H., Lindhoud, S., van Berkel, W. J. H. & van Mierlo, C. P.
2 M. Distant residues mediate picomolar binding affinity of a protein cofactor. *Nat. Comm.*
3 **3**, 1010 (2012).
- 4 14 Sela-Culang, I., Kunik, V. & Ofra, Y. The structural basis of antibody-antigen
5 recognition. *Front. Immunol.* **4** (2013).
- 6 15 van den Bedem, H., Bhabha, G., Yang, K., Wright, P. E. & Fraser, J. S. Automated
7 identification of functional dynamic contact networks from X-ray crystallography. *Nat.*
8 *Methods* **10**, 896-902 (2013).
- 9 16 Koulechova, D. A., Tripp, K. W., Horner, G. & Marqusee, S. When the scaffold cannot
10 be ignored: The role of the hydrophobic core in ligand binding and specificity. *J. Mol.*
11 *Biol.* **427**, 3316-3326 (2015).
- 12 17 Reedy, C. J. & Gibney, B. R. Heme protein assemblies. *Chem. Rev.* **104**, 617-650 (2004).
- 13 18 Tinberg, C. E. *et al.* Computational design of ligand-binding proteins with high affinity
14 and selectivity. *Nature* **501**, 212-216 (2013).
- 15 19 Prier, C. K. & Arnold, F. H. Chemomimetic biocatalysis: Exploiting the synthetic
16 potential of cofactor-dependent enzymes to create new catalysts. *J. Am. Chem. Soc.* **137**,
17 13992-14006 (2015).
- 18 20 Farid, T. A. *et al.* Elementary tetrahelical protein design for diverse oxidoreductase
19 functions. *Nat. Chem. Biol.* **9**, 826-833 (2013).

- 21 Skalicky, J. J. *et al.* Solution structure of a designed four- α -helix bundle maquette
scaffold. *J. Am. Chem. Soc.* **121**, 4941-4951 (1999).
- 22 Huang, S. S., Koder, R. L., Lewis, M., Wand, A. J. & Dutton, P. L. The HP-1 maquette:
From an apoprotein structure to a structured hemoprotein designed to promote redox-
coupled proton exchange. *Proc. Natl. Acad. Sci. USA* **101**, 5536-5541 (2004).
- 23 Lombardi, A., Nastri, F. & Pavone, V. Peptide-based heme-protein models. *Chem. Rev.*
101, 3165-3190 (2001).
- 24 Huang, S. S., Gibney, B. R., Stayrook, S. E., Leslie Dutton, P. & Lewis, M. X-ray
structure of a maquette scaffold. *J. Mol. Biol.* **326**, 1219-1225 (2003).
- 25 Gibney, B. R., Rabanal, F., Skalicky, J. J., Wand, A. J. & Dutton, P. L. Iterative protein
redesign. *J. Am. Chem. Soc.* **121**, 4952-4960 (1999).
- 26 Bender, G. M. *et al.* De novo design of a single-chain diphenylporphyrin metalloprotein.
J. Am. Chem. Soc. **129**, 10732-10740 (2007).
- 27 Fry, H. C., Lehmann, A., Saven, J. G., DeGrado, W. F. & Therien, M. J. Computational
design and elaboration of a de novo heterotetrameric alpha-helical protein that selectively
binds an emissive abiological (porphinato)zinc chromophore. *J. Am. Chem. Soc.* **132**,
3997-4005 (2010).
- 28 Fry, H. C. *et al.* Computational de novo design and characterization of a protein that
selectively binds a highly hyperpolarizable abiological chromophore. *J. Am. Chem. Soc.*
135, 13914-13926 (2013).

- 29 Solomon, L. A., Kodali, G., Moser, C. C. & Dutton, P. L. Engineering the assembly of
heme cofactors in man-made proteins. *J. Am. Chem. Soc.* **136**, 3192-3199 (2014).
- 30 Ghirlanda, G. *et al.* De novo design of a D_2 -symmetrical protein that reproduces the
diheme four-helix bundle in cytochrome bc_1 . *J. Am. Chem. Soc.* **126**, 8141-8147 (2004).
- 31 North, B., Summa, C. M., Ghirlanda, G. & DeGrado, W. F. D_n -symmetrical tertiary
templates for the design of tubular proteins. *J. Mol. Biol.* **311**, 1081-1090 (2001).
- 32 Lahr, S. J. *et al.* Analysis and design of turns in α -helical hairpins. *J. Mol. Biol.* **346**,
1441-1454 (2005).
- 33 Goll, J. G., Moore, K. T., Ghosh, A. & Therien, M. J. Synthesis, structure, electronic
spectroscopy, photophysics, electrochemistry, and x-ray photoelectron spectroscopy of
highly-electron-deficient [5,10,15,20-tetrakis(perfluoroalkyl)porphinato]zinc(II)
complexes and their free base derivatives. *J. Am. Chem. Soc.* **118**, 8344-8354 (1996).
- 34 Lubitz, W., Lendzian, F. & Bittl, R. Radicals, radical pairs and triplet states in
photosynthesis. *Acc. Chem. Res.* **35**, 313-320 (2002).
- 35 Kaufmann, K. W., Lemmon, G. H., DeLuca, S. L., Sheehan, J. H. & Meiler, J. Practically
useful: What the Rosetta protein modeling suite can do for you. *Biochemistry* **49**, 2987-
2998 (2010).
- 36 Moore, K. T., Horváth, I. T. & Therien, M. J. Mechanistic studies of (porphinato)iron-
catalyzed isobutane oxidation. Comparative studies of three classes of electron-deficient
porphyrin catalysts. *Inorg. Chem.* **39**, 3125-3139 (2000).

- 37 Gentemann, S. *et al.* Variations and temperature dependence of the excited state
properties of conformationally and electronically perturbed zinc and free base porphyrins.
J. Phys. Chem. B **101**, 1247-1254 (1997).
- 38 Bradley, P., Misura, K. M. S. & Baker, D. Toward high-resolution de novo structure
prediction for small proteins. *Science* **309**, 1868 (2005).
- 39 Tinberg, C. E. & Khare, S. D. in *Computational Design of Ligand Binding Proteins* (ed
Barry L. Stoddard) 155-171 (Springer New York, 2016).
- 40 Choma, C. T. *et al.* Design of a heme-binding four-helix bundle. *J. Am. Chem. Soc.* **116**,
856-865 (1994).
- 41 Zimmerman, D. E. *et al.* Automated analysis of protein NMR assignments using methods
from artificial intelligence. *J. Mol. Biol.* **269**, 592-610 (1997).
- 42 Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on
UNIX pipes. *J. Biomol. NMR* **6**, 277-293 (1995).
- 43 Bartels, C., Xia, T.-h., Billeter, M., Güntert, P. & Wüthrich, K. The program XEASY for
computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol.*
NMR **6**, 1-10 (1995).
- 44 Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a
database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289-302 (1999).
- 45 Neri, D., Szyperski, T., Otting, G., Senn, H. & Wuethrich, K. Stereospecific nuclear
magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-

1 binding domain of the 434 repressor by biosynthetically directed fractional carbon-13
2 labeling. *Biochemistry* **28**, 7510-7516 (1989).

3 46 Güntert, P., Mumenthaler, C. & Wüthrich, K. Torsion angle dynamics for NMR structure
4 calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283-298 (1997).

5 47 Herrmann, T., Güntert, P. & Wüthrich, K. Protein NMR structure determination with
6 automated NOE assignment using the new software CANDID and the torsion angle
7 dynamics algorithm DYANA. *J. Mol. Biol.* **319**, 209-227 (2002).

8 48 Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Marius Clore, G. The Xplor-NIH
9 NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65-73 (2003).

10 49 Bagaria, A., Jaravine, V., Huang, Y. J., Montelione, G. T. & Guntert, P. Protein structure
11 validation by generalized linear model root-mean-square deviation prediction. *Protein*
12 *Sci.* **21**, 229-238 (2012).

13
14 **Acknowledgements.** N.F.P., J.R., and M.J.T. acknowledge research support from the National
15 Science Foundation through Grant CHE-1413333. D.N.B. acknowledges the National Institutes
16 of Health (GM-071628 and GM-048043) for support of this work. W.F.D. acknowledges support
17 from the National Science Foundation (CHE-1413295), the National Institutes of Health (GM-
18 054616 and GM-071628), and the Materials Research Science and Engineering Centers program
19 of the NSF (DMR-1120901). Simulations were carried out in part with XSEDE resources
20 through grant MCB080011. **Coordinates and data files have been deposited to the Protein**
21 **Data Bank with accession codes 5TGW (apo-PS1) and 5TGY(holo-PS1) and to the BMRB**
22 **(chemical shifts) with codes 30185 (apo-PS1) and 30186 (holo-PS1).**

1
2 **Author Contributions.** N.F.P. and W.F.D. designed the protein. N.F.P., Y.W., A.M., S.-Q.Z.,
3 and J.R. performed experiments. T.L. performed molecular dynamics simulations. N.F.P, Y.W.,
4 W.F.D, and S.-Q.Z. performed data analysis. N.F.P., W.F.D., D.N.B, and M.J.T wrote the paper.

5
6 **Author Information.** Reprints and permissions information is available at
7 www.nature.com/reprints. Correspondence and requests for materials should be addressed to
8 David.Beratan@duke.edu, Michael.Therien@duke.edu, William.DeGrado@ucsf.edu. The
9 authors declare no competing financial interests.