HHS Public Access

Author manuscript

Nat Chem. Author manuscript; available in PMC 2018 March 20.

Published in final edited form as:

Nat Chem. 2017 December ; 9(12): 1157–1164. doi:10.1038/nchem.2846.

Strategy for designing hyperstable, non-natural protein-ligand complexes with sub-Å accuracy

Nicholas F. Polizzi[1,4], **Yibing Wu**[4], **Thomas Lemmin**[4], **Alison M. Maxwell**[4], **Shao-Qing Zhang**[4], **Jeff Rawson**[2], **David N. Beratan**[1,2,3], **Michael J. Therien**[2], and **William F. DeGrado**[4]

[1]Department of Biochemistry, Duke University, Durham, North Carolina 27710, USA

[2]Department of Chemistry, Duke University, Durham, North Carolina 27710, USA

[3]Department of Physics, Duke University, Durham, North Carolina 27710, USA

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

Abstract

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 successes[1–11], no small molecule ligand- or organic cofactor-binding protein has been designed entirely from first principles to achieve i) a unique structure and ii) a predetermined binding-site geometry with sub-Å accuracy. Such achievements are prerequisites for the design of proteins that control and enable complex reaction trajectories, where the relative placements of cofactors, substrates, and protein side chains must be established within the length scale of a chemical bond. Here, we develop and test a strategy for design of small molecule-binding proteins, based on the concept that the entire protein contributes to establishing the binding geometry of a ligand[12–15]. Hence, what are traditionally considered as separate sectors – the hydrophobic core and ligand-binding site – we treat as an inseparable unit. We utilize flexible backbone sequence design of a parametrically defined protein template to simultaneously pack the protein interior both proximal to and remote from the ligand-binding site. Thus, tight interdigitation of core side chains quite removed from the binding site structurally restrains the first- and second-shell packing around the ligand. We apply this principle to the decades-old problem of structural non-uniqueness in de novo-designed heme-binding proteins[16]. 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 reveal the unification of core packing and binding site definition as an essential principle of ligand-binding protein design.

Correspondence and requests for materials should be addressed to David.Beratan@duke.edu, Michael.Therien@duke.edu, William.DeGrado@ucsf.edu.

Author Contributions. N.F.P. and W.F.D. designed the protein. N.F.P., Y.W., A.M., S.-Q.Z., and J.R. performed experiments. T.L. performed molecular dynamics simulations. N.F.P, Y.W., 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.

The authors declare no competing financial interests.

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[17,18]. 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 display[17]. At the other extreme, de novo heme-binding helical bundle proteins have been designed entirely from first principles (reviewed in refs. [16, 19]), but these “maquettes” have evaded structural determination, largely due to aggregation or their dynamical properties[16,20,21]. With the exception of short, covalently linked peptide-heme complexes[22], 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[20,23]. The lack of precise, predictive three-dimensional models of heme-binding maquettes, coupled with the failure to determine high-resolution structures, has severely limited what could be learned or built upon in future endeavors. Meanwhile, successes in the field of de novo design of functional coiled coils[3,7] and metalloproteins[4,8–10] have so far not translated to more complex cofactors.

Our own work has focused on the development of computational design of cofactor-binding proteins[24–26] 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 rigid backbone[24–26]. This approach resulted in de novo porphyrin-binding proteins with the desired tertiary structure and ligand-binding stoichiometry, but not of sufficient conformational uniqueness to yield a high-resolution structure.

An extensive body of work with natural proteins[12–15] has shown that side chain packing quite distant from the binding site can propagate to significantly affect ligand binding, catalysis, and allosteric regulation. Thus, the entire hydrophobic core – even residues 20 Å away from the binding site – should be considered as an essential extension of the primary and secondary shell interactions with the ligand. We noted that, unlike natural proteins (Fig. 1a), previous de novo designed cofactor-binding proteins lack an extensive, well-defined apolar core. Instead, their interior packing is dominated by interactions with one or more porphyrins or multi-functional cofactors that span the length of the bundle (Fig. 1b). Where a cofactor-free core was included[27], the core was not computationally designed, and high-resolution structures were not determined. Here, we purposefully include a folded core remote from the ligand-binding site and optimize its sequence and structure *in concert with the binding site* to ensure appropriate coupling (Fig. 1c). As compared to earlier computational design of ligand-binding proteins[11,17] our approach differs by: 1) beginning with a mathematically parameterized backbone rather than a natural protein; 2) applying flexible backbone design to the entire backbone as well as sequence design to all interior and substrate-binding sites rather than just the first and second-shell contacting residues; 3) not relying on screening of large numbers of designs or genetic selections to achieve the desired outcome.

The design of PS1 (Porphyrin-binding Sequence 1) began with the previously parameterized backbone from the de novo designed protein SCRPZ-2[26], a protein that bound an extended

porphinato(metal)-polypyridyl(metal) cofactor (Fig. 1b). The backbone of SCRPZ-2 and its di-porphyrin-binding predecessors[24,28] was designed with a simple equation defining a D_2-symmetrical antiparallel coiled coil[29]. 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[24,26,30]. 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 $(CF_3)_4$PZn ([5,10,15,20-tetrakis(trifluoromethyl)porphinato]Zn^{2+}) (Fig. 2)[31], an example of electron-deficient (porphinato)metal complexes capable of molecular oxygen activation for alkane hydroxylation and alkene epoxidation[32]. The reduced size of the $(CF_3)_4$PZn cofactor provided space for a hydrophobic core in what was formerly occupied by the large, bulky metal-polypyridyl group. We manually docked $(CF_3)_4$PZn in the porphyrin-binding site (Fig. 2b) and used Backrub within Rosetta[33] 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 (Extended Data Fig. 1).

PS1 is monomeric (Extended Data Fig. 2) and it binds the water-insoluble cofactor, $(CF_3)_4$PZn, forming highly thermostable complexes (extrapolated T_m > 120 °C, Extended Data Fig. 4) that are stable for over a year. The complex forms within seconds of adding $(CF_3)_4$PZn from organic solution to aqueous PS1, suggesting a small kinetic barrier for assembly (Extended Data Fig. 3). Time-resolved transient absorption spectroscopy showed that protein-cofactor interactions are preserved even at near-boiling temperatures where the protein retains its native structure (Fig. 3e). The excited-state spectra and dynamics of $(CF_3)_4$PZn within holo-PS1 at 21 and 100 °C are indistinguishable, indicating that light absorption by the porphyrin populates nearly temperature-independent Franck-Condon surfaces. Thus, PS1 effectively stabilizes an extraordinarily insoluble cofactor in aqueous solution, even at temperatures considered extreme for hyperthermophiles.

An exceptionally well-resolved NMR structural ensemble of holo-PS1 (Fig. 3, Fig. 4a, and Extended Data Fig. 5) was computed using 19 nuclear Overhauser effects (NOEs) per residue, 26 cofactor-protein NOEs, 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. Most importantly, the observed orientation of the cofactor is exactly as designed, within the precision of the NMR structure (Fig. 3d). $(CF_3)_4$PZn 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.

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 folding[34] 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 (Extended Data Fig. 6). 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. 3b) 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. 4), 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 (Extended Data Fig. 7). 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.

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. 4b, Extended Data Figs. 8 and 9). 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. 4b). 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. 4, c–e). 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. 4d). By contrast, the packing of the more distal layers is identical in the apo- and holo-structures (Fig. 4e). 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 proteins[12–15], which show that variation of residues involved in core packing distant from an active site can have profound influences on binding and catalysis.

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[17,18,35]. The reliance on experimental methods such as directed evolution and genetic selections, while currently useful in many practical applications[18], 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 via our approach succeeded without need for experimental screening. The strategy reported here is not limited to idealized backbones and could also be immediately extended to redesign of natural proteins. These studies bring chemists closer to the ultimate goal of the computational design of fully functional proteins with properties unprecedented in nature.

Methods

Cofactor geometry optimization

The geometry of $(CF_3)_4$PZn was optimized via density functional theory using the B3-LYP functional and 6-31G* basis set implemented in Gaussian03. The starting geometry was obtained from the crystal structure of related meso-heptafluoropropyl(porphinato)Zn(II), with the fluoropropyl groups truncated to fluoromethyl[31]. Meso-heptafluoropropyl(porphinato)Zn(II) co-crystalized with an axially ligating pyridine; imidazole was computationally substituted for pyridine for the geometry optimization of $(CF_3)_4$PZn.

PS1 design process

The design of PS1 began with a D_2-symmetrical parameterized backbone of a 4-helix bundle (Tables S1 and S2)[29]. We have previously used this backbone parameterization to create a D_2-symmetrical diheme-binding tetrameric 4-helix bundle, PA_{TET}, which was composed of 4 copies of a 25 residue helix containing the requisite metal-coordinating His and second shell H-bonding Thr residues placed at *d* and *b* positions in a heptad repeat, respectively[28]. This tetramer bound two hemes with a bis-His ligation in a D_2-symmetrical bundle. Asymmetry of the sequence was later introduced in a single chain diporphyrin-binding design, PA_{SC} (Fig. 1B), where loops were selected to connect the helices via a structural bioinformatics approach[24,30]. The attachment of these loops cemented the Crick parameters of the helical backbone, which was later employed in another single-chain protein design,

SCRPZ-2, that bound an extended cofactor throughout the interior of the bundle (Fig. 1B)[26]. The design of PS1 utilized the His and Thr positioning of one porphyrin-binding region from these previous designs, with the remainder of the protein then designated as a cofactor-free folded core. Because SCRPZ-2 was soluble and expressed well in *E. coli*, we elected to retain its exterior-facing amino acids and loops within the PS1 design, while computationally designing the entire core (binding region and folded core simultaneously). In doing so, we also isolate effects on cofactor binding due solely to the creation of a folded core that uniquely predisposes the binding region for cofactor association, which is simultaneously optimized for sequence and side chain packing along with the binding region of the $(CF_3)_4$PZn porphyrin. A flexible backbone sequence design protocol was developed (see below) to fine-tune the parameterized backbone to $(CF_3)_4$PZn and to achieve optimal side chain packing for creation of the folded core and positioning of $(CF_3)_4$PZn within the binding region.

Flexible backbone sequence design

We wrote a RosettaScript for flexible backbone sequence design, implemented in Rosetta 3.5, that proceeds through a cycle of backbone/side-chain relaxations and fixed backbone design, with a filtering step based on core packing (RosettaScript provided below). Details of the process are provided in the subsections below.

Amino acids allowed to vary in the design—Because $(CF_3)_4$PZn could potentially act as a photo-oxidant, we disallowed any potentially oxidizable amino acids in the sequence (e.g., Tyr, Cys, Met, Trp, His) other than the single His and Trp residues described below. The initial residue identities of the bundle were chosen from a previous computationally designed 4-helix bundle SCRPZ-2 [26], with a few changes, e.g., surface-exposed Tyr residues of the SCRPZ-2 sequence were constrained to be polar or charged during the computational sequence design in Rosetta. The entire core (40 residues in total) of SCRPZ-2 was allowed to vary during the design process, except for His46 and Thr9, which are keystone interactions dedicated to Zn coordination of the porphyrin used in previous designs (see Fig. 2). (63% of the SCRPZ-2 sequence consists of exterior residues and loops, and these were held fixed during the design of PS1.) Ultimately, of the 40 residues that could vary (out of 108), 28 residues were changed and 12 were retained from the SCRPZ-2 sequence, such that 70% of the core was computationally mutated to establish a preferred orientation of the porphyrin cofactor, as well as an interdigitated folded core. This percentage of retained residues can be rationalized based on the expected results of choosing large space-filling amino acids (Phe, Leu, Ile, Val) at random, such that a residue that is Leu in the sequence has a 25% chance of retaining its identity as Leu. Table S3 and Extended Data Fig. 2b, as well as the residue file (resfile.txt) in the Supplementary Text, show precisely which residues were allowed to vary during the design process. Below and in the main text, we use residue numbering that is consistent with the expressed holo-protein, which contains an N-terminal Ser residue not present in the design, a remnant from a TEV protease cleavage site (see below).

Selection of residue 68 as Trp—We required a Trp residue in the protein interior as an absorption handle, as well as a fluorescent indicator of hydrophobic packing. Trp could be

used as a source of a photo- or chemically generated radical in future studies. To select the sequence position of the single Trp residue, we used the Rosetta Backrub program [38,39] to create an ensemble of backbones that were relaxed around the $(CF_3)_4$PZn cofactor, after the cofactor was docked in the porphyrin binding region of the SCRPZ-2 model, with an orientation described by CF_3 groups pointing down the long axis of the bundle. No sequence design was performed to generate this backbone ensemble. Next, we performed fixed backbone sequence design on each member of the backbone ensemble, allowing Trp at all core residues, to determine a probable location of Trp within the protein interior, based on the frequency of occurrence within the designed sequences. Based on this information, we constrained residue 68 to be Trp during the flexible backbone design process below.

Flexible backbone design protocol—Flexible backbone design utilized angle and distance constraints between the Zn and His to restrict the design space to those consistent with the DFT-optimized imidazole-Zn distance of 2.0 Å. We used an energy term (hack_aro = 1) that models quadrupolar interactions between aromatic side chains in every stage of the flexible backbone design protocol. We also employed an energy term (rg = 2) that penalizes bundles with a large radius of gyration (rg). We noticed a propensity within Rosetta to output bundles that received good packing scores (via Packstat or Rosetta Holes) but displayed helices separated by large distances (large rg). The packing algorithms could not differentiate between interior or exterior when the helix-helix interfaces were very wide, and often inappropriately gave good packing scores when the designed bundle was qualitatively poorly packed. The inclusion of the rg term, as well as employing Rosetta Backrub, ameliorated this issue.

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 below) were performed, and models with native protein-like packing (i.e., a Rosetta PackStat score ≥ 0.58) were output. The PackStat score was calculated 3 times per trial to account for its stochastic behavior. 170 designs were output from 500 runs through the protocol (Fig. S1). We analyzed these 170 models for packing, rg, energy, and rotamer state probability within Matlab to select PS1 for expression.

Flexible backbone design sub-protocol—The flexible backbone design sub-protocol consists of 3 Monte Carlo trials of (i) fixed backbone design with soft weights (decreased vdW interactions, i.e., soft_rep_design weights within Rosetta), (ii) sidechain minimization via MinMover, (iii) fixed backbone design with Score13 weights, where the electrostatic term (fa_pair) is replaced by hack_elec (hack_elec = 0.55), and the addition of extra rotamer sampling around χ_1 (ex1, level 3, i.e., sampled between 2 std of the mean chi angle value for each rotamer) and χ_2 (ex2, level 3) sidechain dihedrals, (iv) backbone minimization via MinMover, (v) repetition of step iii (due to propensities of Rosetta to design a particular sequence to a particular backbone). At the end of step (v), the model is filtered for native structure-like packing via PackStat (If 1 of 3 trials of PackStat score is > 0.58, the model passes the filter.). In all energy functions for flexible backbone design, hack_aro is set to 1 and rg is set to 2. The final, designed sequence (PS1) selected for protein expression was the

following 108 amino acids:
EFEKLRQTGDELVQAFQRLREIFDKGDDDSLEQVLEEIEELIQKHRQLFDNRQEAAD
TEA AKQGDQWVQLFQRFREAIDKGDKDSLEQLLEELEQALQKIRELAEKKN

Ab initio folding

Rosetta ab initio folding[34] was performed on the PS1 sequence in Rosetta 3.5. Cα RMSD of the folded core was scored against residues 14-23, 32-42, 69-79, and 87-97 of the design model. Cα RMSD of the binding region was scored against residues 5-13, 43-50, 61-68, and 98-105 of the design model.

Visualization of protein structures and image rendering

Protein models were visualized and rendered in the PyMol visualization program [40].

Protein expression and purification

The gene coding for the protein sequence of PS1 was ordered from GenScript, which was cloned into the IPTG-inducible pET-11a plasmid (cloning site NdeI-BamHI). The sequence also coded for an N-terminal 6xHis-tag followed by a TEV protease cleavage sequence, followed finally by the designed sequence. The cloned gene sequence is:
CATATGCATCACCATCACCATCACGAAAACCTGTATTTTCAGAGCGAATTCGAAAA
A
CTGCGTCAAACCGGCGACGAACTGGTGCAGGCATTTCAACGTCTGCGCGAAATTT
TC
GATAAAGGTGATGACGATAGTCTGGAACAGGTTCTGGAAGAAATTGAAGAACTGA
T
CCAGAAACATCGTCAACTGTTTGACAATCGCCAGGAAGCGGCCGATACGGAAGC
AG
CTAAACAGGGCGACCAATGGGTCCAGCTGTTTCAACGTTTCCGCGAAGCCATTGA
TA
AAGGTGACAAAGATAGCCTGGAACAGCTGCTGGAAGAACTGGAACAGGCGCTGC
A AAAAATCCGCGAACTGGCCGAAAAGAAAAACTAAGGATCC

The expressed protein sequence was finally: MHHHHHHENLYFQ/
SEFEKLRQTGDELVQAFQRLREIFDKGDDDSLEQVLEEIEELIQKH
RQLFDNRQEAADTEAAKQGDQWVQLFQRFREAIDKGDKDSLEQLLEELEQALQKI
REL AEKKN where the "/" defines the cleavage site of TEV protease. The plasmids were transfected into *E. coli* BL21(DE3) cells, which were grown in LB/ampicillin media (or, for NMR samples, M9 minimal media with isotope-labeled ammonia and glucose from Cambridge Isotopes) until OD @ 600 nm = 0.6. The cells were then induced with IPTG and allowed to grow for 4 more hours. Cells were then centrifuged and frozen. The frozen cell pellets were lysed in a French press in the Duke University Biology Department. The expressed, His-tagged PS1 protein was purified via a Ni NTA column (Invitrogen) and confirmed by gel electrophoresis. The buffer was exchanged to the Sigma-recommended TEV protease buffer (5 mM DTT, 50 mM Tris, 0.5 mM EDTA, pH 8.0), and the PS1/TEV solution (His-tagged TEV protease was ordered from Sigma.) was allowed to rock for 1 day at room temperature. The resulting His-tag-free PS1 protein was collected from the flow-

through of a Ni NTA column and concentrated in a stock of 50 mM NaPi, 100 mM NaCl, pH 7.5 buffer, with an approximate yield of 40 mg/L.

Clustering of apo-PS1 NMR models

We implemented a greedy clustering algorithm in Matlab to form clusters within the family of structures of apo-PS1 (Extended Data Fig. 7). A pairwise RMSD matrix of each apo-PS1 model was scored against residues 61-67 and 99-105. These residues, which lie on opposite helices, show the largest conformational variation within the apo-PS1 models. The clustering algorithm defines the centroid as the column of the RMSD matrix containing the largest number of RMSD values below a threshold of 1 Å. Components of this column below this threshold have their corresponding rows and columns removed from the RMSD matrix, and the clustering algorithm repeats again on this truncated RMSD matrix. Of the 20 NMR models, two clusters were found with > 4 members each. The cluster defining the closed conformation contained 13 members, and that of the open conformation contained 5 members.

Cofactor synthesis

The cofactor [5,10,15,20-tetrakis(trifluoromethyl)porphinato]zinc(II), abbreviated as $(CF_3)_4$PZn in the main text, was synthesized as previously reported[31], and was confirmed by NMR and electronic absorption spectra.

Cofactor reconstitution

A 2-fold excess of the cofactor $(CF_3)_4$PZn 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, $(CF_3)_4$PZn 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 × 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 × 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.

Analytical ultracentrifugation (AUC)

The oligomeric state of apo- and holo-PS1 were determined by analytical equilibrium sedimentation performed at 25 °C using a Beckman XL-I analytical ultracentrifuge. Ultracentrifugation was conducted at speeds of 25K, 30K, 35K, 40K and 45K r.p.m., and the radial gradient profiles were obtained by absorbance at 280 nm. A 200 μM solution of the apo- and a 100 μM solution of the holo-protein were prepared in 50 mM NaPi pH 7.5, 100 mM NaCl (apo) and 20 mM NaPi pH 7.5, 125 mM NaCl (holo). Data were globally fit to a single-species model of equilibrium sedimentation by a nonlinear least-squares method using IGOR Pro (Wavemetrics).

Size exclusion chromatography

Gel filtration profiles were obtained using a Superdex 75 5/150 column on an FPLC system (GE Healthcare AKTA). To evaluate the oligomeric state, 20 μL of 100 μM apo-PS1 or 37 μM holo-PS1 was injected onto the column and eluted with a 50 mM phosphate, 150 mM NaCl, pH 7.0 buffer mobile phase at a flow rate of 0.4 mL/min. The approximate molecular weight (MW_{app}) was calculated from a standard curve obtained with the GE LMW standard protein kit. From this curve, MW_{app} of the apo is 19.5 kD and that of holo is 17.9 kD. These 13 kD proteins elute at higher MW_{app} due to their large negative suface charge (q = −12). For apo-PS1, a small dimer peak elutes at MW_{app} of 44.1 kD, and a smaller tetramer (or pentamer) peak at 103.2 kD.

Circular dichroism (CD)

CD spectra were collected on a Jasco J-810 CD spectrometer in a 0.1 cm path length quartz cuvette, using temperature/wavelength mode. Spectra were collected from 20 to 95 °C with an interval of 5 °C and an increase rate of 1 °C/minute, over a wavelength range from 215 to 250 nm. Apo- and holo-PS1 were prepared at 20 μM in 50 mM NaPi pH 7.5, 100 mM NaCl buffer. Temperature melts of apo-PS1 were also performed at varying concentrations of Guanidine HCl denaturant (0M, 1M, 2M, 3M, 4M, 5M, 5.85, 7M).

Steady-state electronic absorption and emission spectroscopy

Electronic absorption spectra were collected using a Shimadzu UV-1700 UV-Vis spectrophotometer or Cary 5000 spectrophotometer. Steady-state emission spectra were obtained on FLS920P spectrophotometer (Edinburgh Instruments Ltd. Livingston, UK) in 1 cm quartz optical cells. The steady-state emission spectra were corrected using the correction factor generated by the manufacturer.

Pump-probe transient absorption spectroscopy

Ultrafast transient absorption spectra were obtained using standard pump-probe methods[41]. Optical pulses (≥ 120 fs) centered at 775 nm, were generated using a Ti:Sapphire laser (Clark-MXR, CPA-2001, Dexter, MI, USA), which consisted of a regenerative amplifier seeded by a mode-locked fiber oscillator. The output of the regenerative amplifier was split to feed an optical parametric amplifier (Light Conversion Ltd., TOPAS-C, Vilnius, Lithuania), which generates excitation pulses tunable in wavelength from the UV through the NIR region. The pump beam was chopped at half the laser repetition rate (~500 Hz). A fraction (<5%) of the output from the regenerative amplifier was passed through an optical delay line, and focused onto a 2 mm c-cut sapphire plate to generate a white light continuum, which was used as the probe beam. The polarization and attenuation of the pump and probe beams were controlled by half-wave plate and Rochon prism polarizer pairs. The polarization was set to the magic angle (54.7°) for these experiments. The pump beam was focused into the sample cell with an f = 20 cm lens, while the probe beam was focused with a concave mirror. The spot size diameter was 0.2–0.3 mm. The beam diameter was determined using the razor-blade method. The excitation pump power was measured using a power meter (Coherent, LabMax Top with PS19 head). After passing through the sample, the probe light was adjusted using a neutral density filter to avoid saturating the detector, and

focused onto the entrance slit of a computer-controlled image spectrometer (Acton Research Corporation, SpectraPro-150, Trenton, NJ, USA). A CCD array detector (1024 × 128 elements, Roper Scientific, Trenton, NJ, USA), interfaced to the spectrometer, recorded the spectrum of the probe light from the UV (~370 nm) to the NIR (~1100 nm), providing spectral resolution better than 0.5 nm. Pairs of consecutive spectra were measured with ($I_{on}(\lambda)$) and ($I_{off}(\lambda)$) to determine the difference spectrum, $\Delta A = \log(I_{off}(\lambda))/(I_{on}(\lambda))$. All these experiments utilized a custom-built 2 mm-path-length fused-silica sample cell; all transient optical studies were carried out at 21 ± 1 °C in HPLC grade solvents received from Sigma-Aldrich, unless otherwise noted. Elevated temperature experiments were performed in a custom-made temperature block of anodized aluminum, the temperature of which was controlled by heating rods and monitored by a pair of thermocouples wired to a PID through a solid-state relay. All transient spectra reported represent averages obtained over 3-5 scans, with each scan consisting of ~100-200 data points, with each point an average of 2000 frames. In these experiments, the delay line utilizes a computer-controlled delay stage. Delay times up to 4 ns were achieved using a Compumotor-6000 (Parker). The baseline noise level in these transient absorption experiments corresponded to ~ 0.2 mOD per second of signal accumulation. The time resolution is probe wavelength dependent; in these experiments, the FWHM of the instrument response function (IRF) varied between 140–200 fs (*e.g.*, at 680 nm, the IRF was 150 ± 6 fs). Following all pump-probe transient absorption experiments, electronic absorption spectra verified that the samples were robust.

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 ($^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, ^{13}CO) and $^{13}C^{\beta}$ resonance assignments were obtained by using 3D HNCACB/CBCA(CO)NH and 3D HNCO/CO(CA)NH along with the program AUTOASSIGN[42]. $^{1}H^{\alpha}$ and $^{1}H^{\beta}$ 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 $^{15}N/^{13}C^{aliphatic}/^{13}C^{aromatic}$-resolved [$^{1}H$,$^{1}H$]-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 $^{13}C^{\beta}$, 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 $^{13}C^{\gamma}$) for the holo- and apo-proteins, respectively. All spectra were processed and analyzed with the programs NMRPIPE and XEASY, respectively[43,44]. ^{1}H-^{1}H 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 elements[45]. 2D constant-time [^{13}C,^{1}H]-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 Leu[46]. The $^{1}D_{NH}$ residual dipolar couplings (RDCs) were measured with 2D ^{1}H-^{15}N IPAP-HSQC in samples aligned using Pf1 phage (ASLA biotech). The program CYANA was used to assign long-range NOEs and calculate the structure[47,48]. Backbone $^{1}D_{NH}$ RDCs were used as orientational constraints for the later stages of refinement with XPLOR-NIH[49]. The final set of structures was further refined by restrained molecular dynamics in explicit water[49]. NMR

structure quality was assessed with the Protein Structure Validation Software Suite (PSVS)[50] (Table S4).

Hydrogen-Deuterium Exchange Measurements

For the measurements of H/D exchange rates, a series of 2D ^{15}N 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 high concentration sample in H_2O (2 mM for apo and 1.2 mM for holo) into 200 μl D_2O 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.

Molecular dynamics simulations

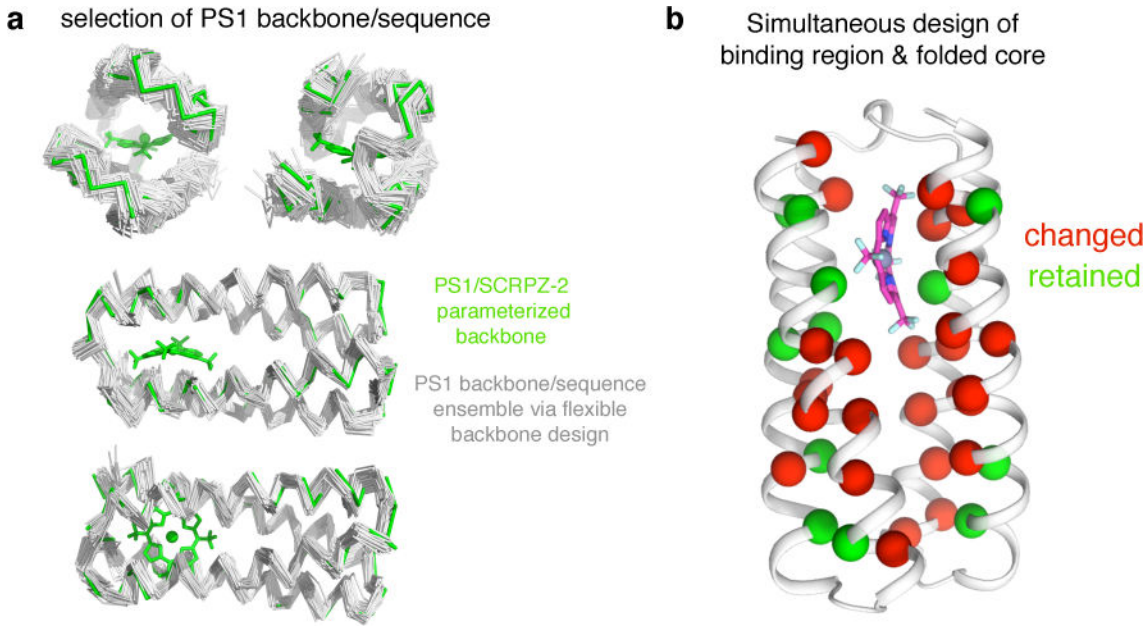
The lowest energy NMR structure of apo-PS1, which is the centroid of the closed conformation, was used as the starting conformation for the molecular dynamics simulation. The structure was solvated in a 17 Å padding water box, neutralized by the addition of 12 Na^+ counter ions. The AMBER force field 14SB was used for the parameterization of the protein. TIP3P water parameterization was used to describe the water molecules [51].

The molecular dynamics simulation was carried out using ACEMD[52]. The system was minimized for 2000 steps, followed by equilibration using the NPT ensemble for 10 ns at 1 atm using a time-step of 2 fs. We also used rigid bonds and a cutoff of 9 Å using PME for long-range electrostatics. Following the relaxation phase, the protein was allowed to move freely and simulated under the NVT ensemble using ACEMD's NVT ensemble with a Langevin thermostat. To achieve a time-step of 4 ps, we used damping at 0.1 ps-1 and a hydrogen mass repartitioning scheme. The simulation was carried out to 1 μs at 298 K.

SOCKET Server for assessment of knobs-into-holes packing

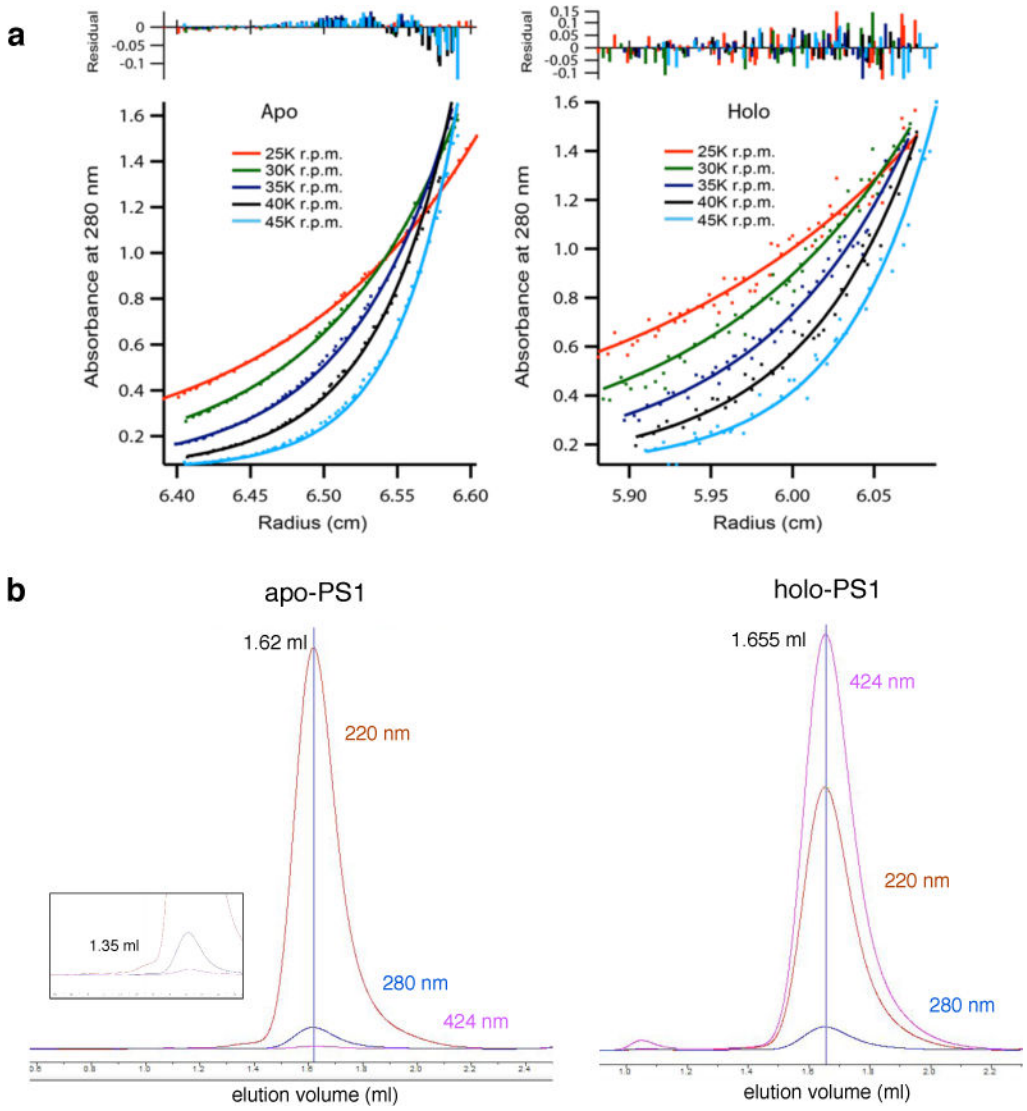
PDB files of the PS1 design model, holo-PS1 centroid, and apo-PS1 open/closed centroids were individually uploaded to and analyzed by the SOCKET server[53] for knobs-into-holes side chain packing (see Supplement). A helical residue was defined as a knob if its side chain was within 8 Å of 4 other side chains from residues on an adjacent helix (a hole). Output from the SOCKET server for each of these PDB files is displayed below showing the residues of each knob and hole. Note that the residue number of the PS1 design model is off register by 1 amino acid from the structural sequences, due to the presence of the N-terminal Ser residue from TEV cleavage of the expressed proteins.

Extended Data



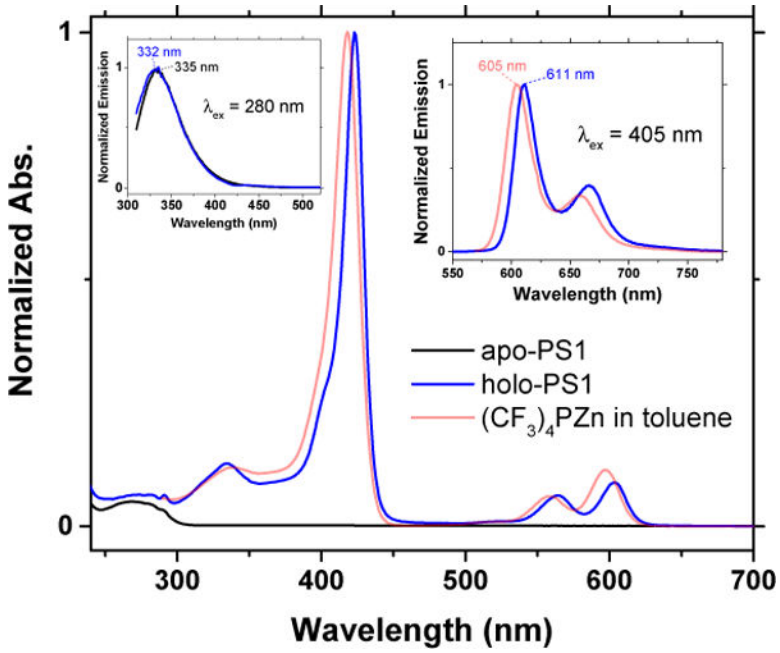
Extended Data Figure 1. PS1 design metrics

a, PS1 design ensemble resulting from flexible backbone sequence design. **b**, Residues (Cα atoms shown as blue and green spheres) within the PS1 design that were allowed to vary from the SCRPZ-2 sequence. 40 of the 108 residues were allowed to vary, and, of the 40 residues, 28 were mutated (blue) and 12 were retained (green) from the original SCRPZ-2 sequence as a result of the computational design process.



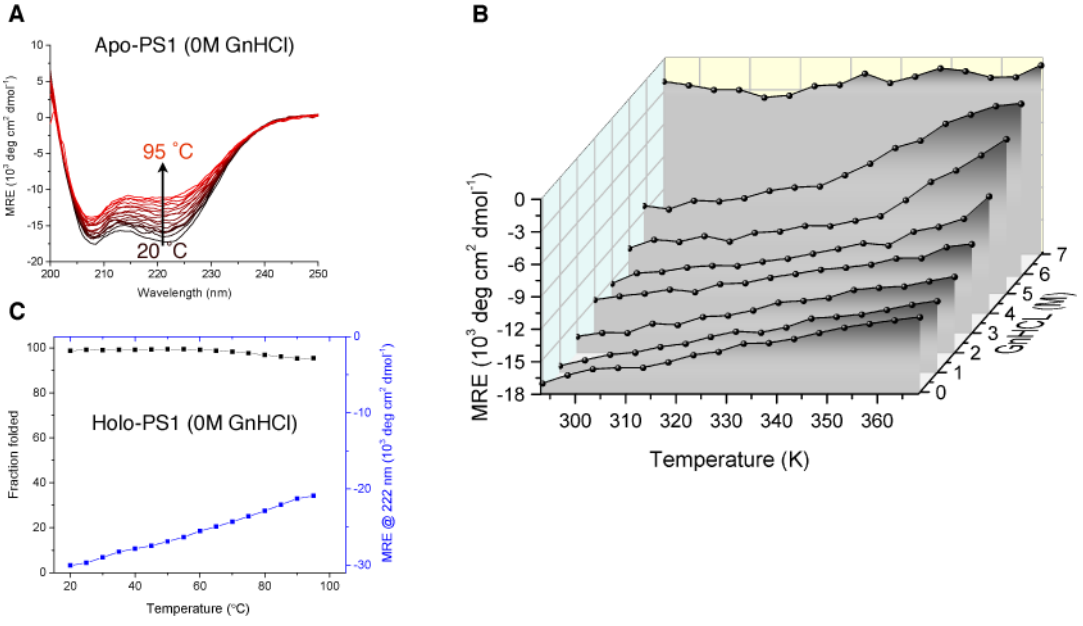
Extended Data Figure 2. Analytical ultracentrifugation and gel filtration analysis show that apo- and holo-PS1 are monomeric in solution

a, Analytical ultracentrifugation. Solutions of apo- and holo-PS1 were centrifuged at speeds ranging from 25,000 r.p.m. to 45,000 r.p.m. and monitored by absorbance at 280 nm. Parameters were globally fit to the data. Single-species fitting agrees well with the data over the entire range and yields the molecular weight of apo-PS1 15.81 ± 0.09 kD and holo-PS1 12.24 ± 0.91 kD, which agrees well with the 12.86 kD weight of PS1. At high concentration, the fit for apo-PS1 is not ideal, suggesting a small degree of aggregation. Partial specific volumes were estimated from SEDNTERP[37], for amino acid side chains. **b**, Analytical gel filtration analysis of apo- and holo-PS1. Detection wavelengths are labeled as the same color as their respective curves. Apo shows a small degree (< 5%) of dimerization (1.35 ml elution volume) relative to the monomer peak (1.62 ml elution volume). The small peak near 1.05 ml elution volume in holo-PS1 is unbound (excess), aggregated porphyrin eluting in the void volume of the superdex 75 5/150 column. Samples were run at concentrations of 100 μM and 37 μM for apo and holo, respectively, in 50 mM NaPi, 150 mM NaCl, pH 7.0 buffer.



Extended Data Figure 3. Electronic absorption and emission spectra of apo- and holo-PS1

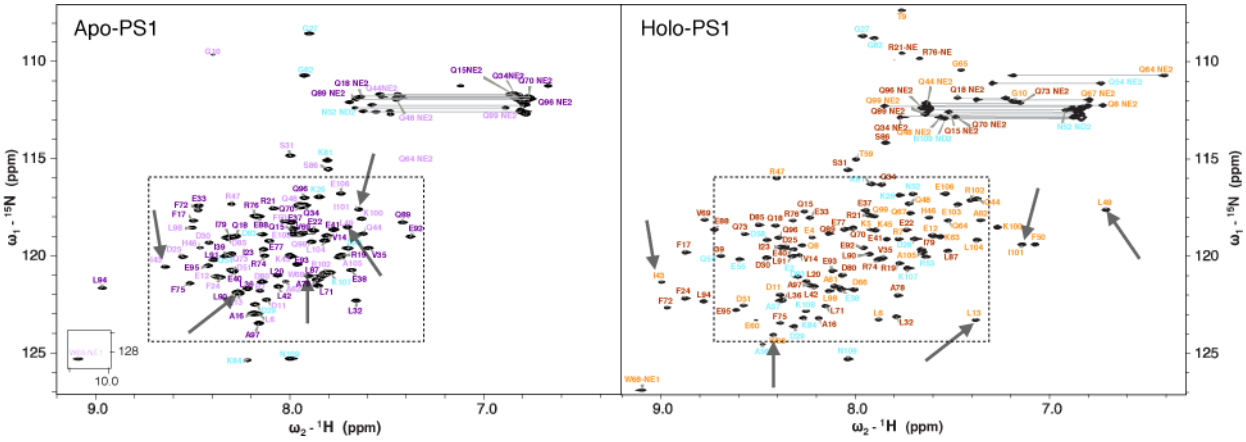
The insets show normalized emission spectra: (left) electronic excitation of Trp68 at 280 nm, monitoring Trp68 emission; (right) electronic excitation of $(CF_3)_4$PZn at 405 nm, monitoring porphyrin emission. OD = 0.1 at excitation wavelength, buffer = 100 mM NaCl, 50 mM NaPi, pH 7.5.



Extended Data Figure 4. Temperature and GnHCl induced unfolding of apo- and holo-PS1 show the designed protein is extremely thermostable

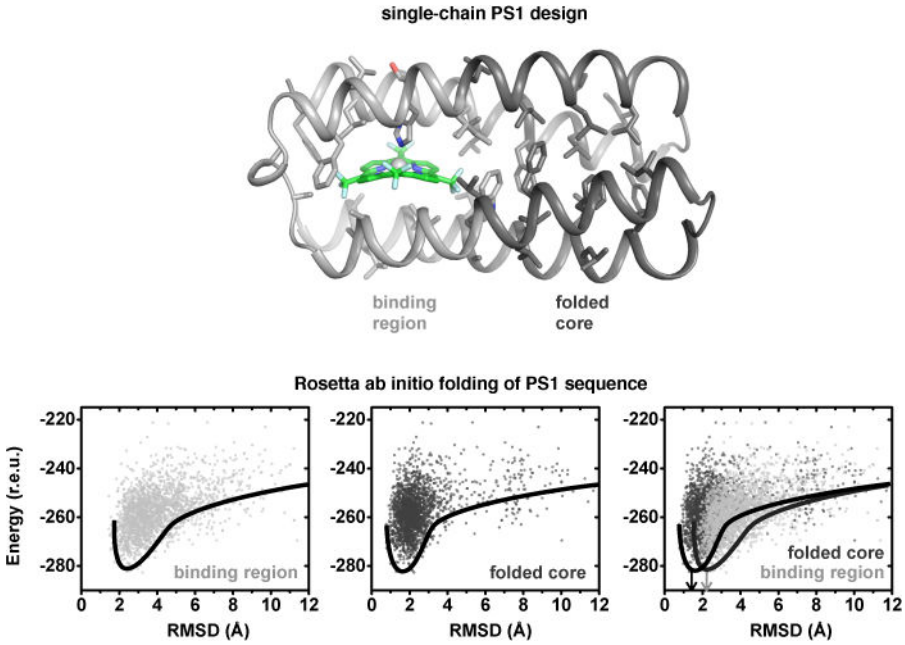
a, Circular dichroism (CD) spectra of apo-PS1 in 50 mM NaPi, 100 mM NaCl, pH 7.5 (No denaturant), as a function of temperature. **b**, CD spectra at 222 nm of apo-PS1 as a function

of temperature and denaturant (Guanidine HCl, GnHCl) concentration in 50 mM NaPi, 100 mM NaCl, pH 7.5 buffer. **c**, CD spectra (blue) at 222 nm of holo-PS1 as a function of temperature in 50 mM NaPi, 100 mM NaCl, pH 7.5 (No denaturant), and fraction folded (black), defined as [MRE@222nm (T)- unfolded@222nm (T)]/[folded@222nm (T)-unfolded@222nm (T)], where unfolded(T) and folded(T) are unfolded and folded baselines, respectively, and T is temperature. The transitions appear reversible based on the fact that the spectra are identical after cooling to room temperature. The midpoint for GnHCl-induced unfolding at 95 °C was approximately 4.5 M.



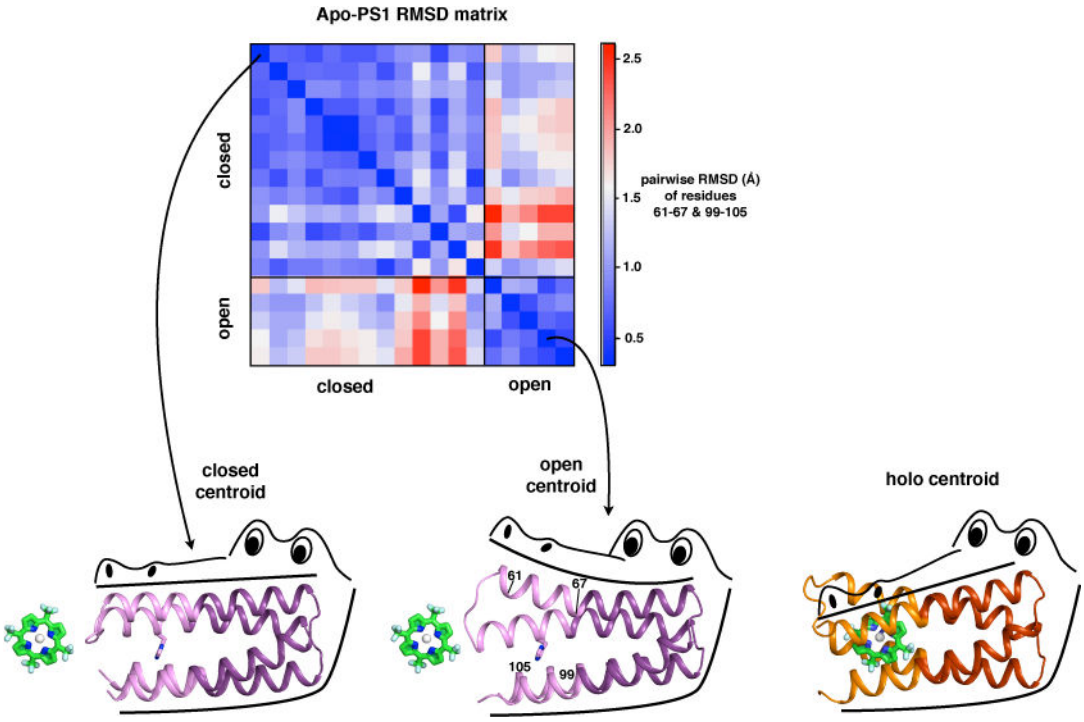
Extended Data Figure 5. 2D ^{1}H-^{15}N HSQC spectra acquired for apo- and holo-PS1

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 Fig. 4a. 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.



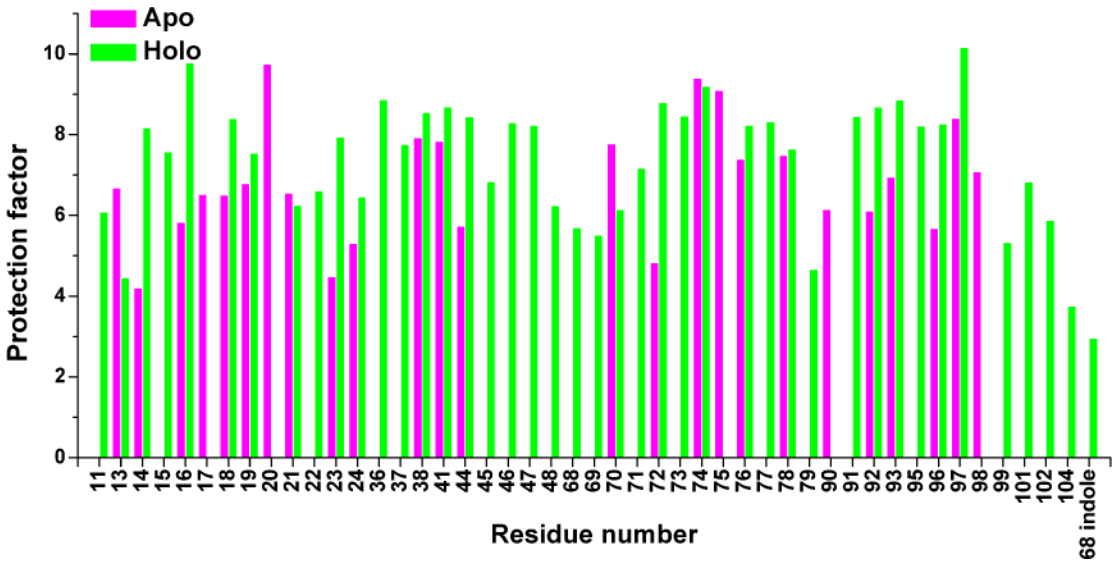
Extended Data Figure 6. Ab initio folding predictions of PS1 sequence

The Rosetta folding algorithm predicts a shallow folding funnel for the binding region (light gray) and a deep folding funnel shifted toward lower RMSD for the folded core (dark gray) of apo-PS1. The RMSD (root mean squared deviation) in Å is against the helical residues within these regions in the designed model. Energy is in Rosetta energy units (r.e.u.).

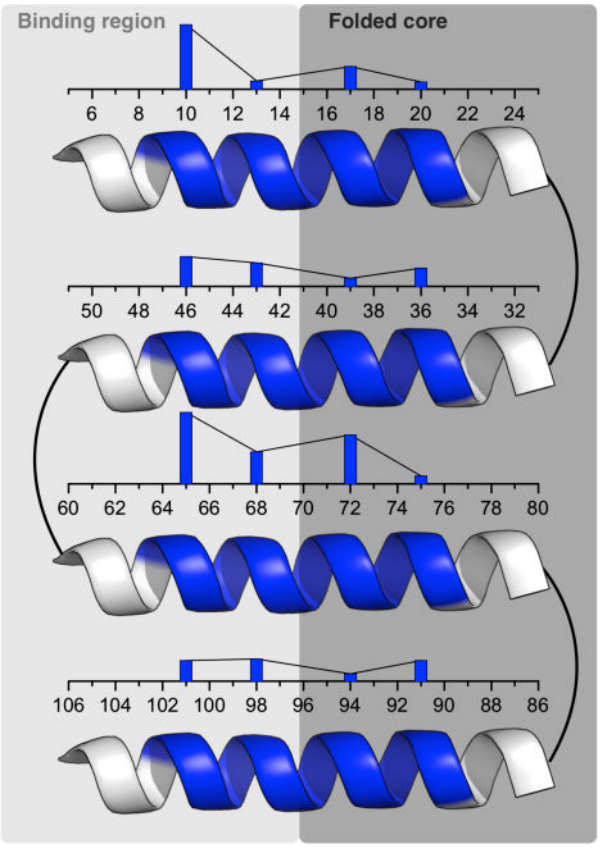


Extended Data Figure 7. The NMR structural ensemble of apo-PS1 contains two clusters of conformations, closed and open

Above, color mapping of the pairwise backbone RMSD matrix of each NMR ensemble member of apo-PS1. Apo models with high structural similarity in the region of residues 61-67 and 99-105 (labeled in the open structure shown below) are blue in the plot. Models that are structurally dissimilar (large RMSD) are red in the plot. Below, the model centroids representing the closed and open structures (models 1 and 18, respectively, in the deposited NMR structure). The porphyrin $(CF_3)_4$PZn is shown in green, and the holo centroid (orange) is also drawn for comparison.



Extended Data Figure 8. HDX protection factors for apo- and holo-PS1, as described in Table S5
Note that “68 indole” denotes the indole N of Trp68 side chain.



Extended Data Figure 9. Molecular dynamics simulations show the binding region of apo-PS1 is more accessible to solvent

Histogram of number of waters within 3.5 Å of any heavy atom of each buried amino acid side chain (an A or D position of the heptad repeat), from 1000 snapshots of a 1 μs trajectory of apo-PS1. All histograms are drawn to the same scale and show number of solvating waters normalized by side chain surface area. Binding region shown in light gray, and folded core in dark gray.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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

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. 1997; 119:5302–5306.
2. Kuhlman B, et al. Design of a novel globular protein fold with atomic-level accuracy. Science. 2003; 302:1364–1368. [PubMed: 14631033]
3. Nanda V, Koder RL. Designing artificial enzymes by intuition and computation. Nat Chem. 2010; 2:15–24. [PubMed: 21124375]
4. Peacock AFA. Incorporating metals into de novo proteins. Curr Opin Chem Biol. 2013; 17:934–939. [PubMed: 24183813]
5. Huang PS, et al. High thermodynamic stability of parametrically designed helical bundles. Science. 2014; 346:481. [PubMed: 25342806]
6. Thomson AR, et al. Computational design of water-soluble α-helical barrels. Science. 2014; 346:485. [PubMed: 25342807]
7. Woolfson DN, et al. De novo protein design: How do we expand into the universe of possible protein structures? Curr Opin Struct Biol. 2015; 33:16–26. [PubMed: 26093060]
8. Mocny CS, Pecoraro VL. De novo protein design as a methodology for synthetic bioinorganic chemistry. Acc Chem Res. 2015; 48:2388–2396. [PubMed: 26237119]
9. Ulas G, Lemmin T, Wu Y, Gassner GT, DeGrado WF. Designed metalloprotein stabilizes a semiquinone radical. Nat Chem. 2016; 8:354–359. [PubMed: 27001731]
10. Olson TL, et al. Design of dinuclear manganese cofactors for bacterial reaction centers. Biochim Biophys Acta: Bioenergetics. 2016; 1857:539–547.
11. Huang PS, Boyken SE, Baker D. The coming of age of de novo protein design. Nature. 2016; 537:320–327. [PubMed: 27629638]
12. Bollen YJM, Westphal AH, Lindhoud S, van Berkel WJH, van Mierlo CPM. Distant residues mediate picomolar binding affinity of a protein cofactor. Nat Comm. 2012; 3:1010.
13. Sela-Culang I, Kunik V, Ofran Y. The Structural Basis of Antibody-Antigen Recognition. Front Immunol. 2013; 4
14. van den Bedem H, Bhabha G, Yang K, Wright PE, Fraser JS. Automated identification of functional dynamic contact networks from X-ray crystallography. Nat Methods. 2013; 10:896–902. [PubMed: 23913260]
15. Koulechova DA, Tripp KW, Horner G, Marqusee S. When the Scaffold Cannot Be Ignored: The Role of the Hydrophobic Core in Ligand Binding and Specificity. J Mol Biol. 2015; 427:3316–3326. [PubMed: 26301601]
16. Reedy CJ, Gibney BR. Heme protein assemblies. Chem Rev. 2004; 104:617–650. [PubMed: 14871137]
17. Tinberg CE, et al. Computational design of ligand-binding proteins with high affinity and selectivity. Nature. 2013; 501:212–216. [PubMed: 24005320]
18. Prier CK, Arnold FH. Chemomimetic biocatalysis: Exploiting the synthetic potential of cofactor-dependent enzymes to create new catalysts. J Am Chem Soc. 2015; 137:13992–14006. [PubMed: 26502343]
19. Farid TA, et al. Elementary tetrahelical protein design for diverse oxidoreductase functions. Nat Chem Biol. 2013; 9:826–833. [PubMed: 24121554]
20. Skalicky JJ, et al. Solution structure of a designed four-α-helix bundle maquette scaffold. J Am Chem Soc. 1999; 121:4941–4951.
21. Huang SS, Koder RL, Lewis M, Wand AJ, Dutton PL. The HP-1 maquette: From an apoprotein structure to a structured hemoprotein designed to promote redox-coupled proton exchange. Proc Natl Acad Sci USA. 2004; 101:5536–5541. [PubMed: 15056758]
22. Lombardi A, Nastri F, Pavone V. Peptide-based heme–protein models. Chem Rev. 2001; 101:3165–3190. [PubMed: 11710067]
23. Huang SS, Gibney BR, Stayrook SE, Leslie Dutton P, Lewis M. X-ray structure of a maquette scaffold. J Mol Biol. 2003; 326:1219–1225. [PubMed: 12589764]

24. Bender GM, et al. De novo design of a single-chain diphenylporphyrin metalloprotein. J Am Chem Soc. 2007; 129:10732–10740. [PubMed: 17691729]
25. Fry HC, Lehmann A, Saven JG, DeGrado WF, Therien MJ. 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. 2010; 132:3997–4005. [PubMed: 20192195]
26. Fry HC, et al. Computational de novo design and characterization of a protein that selectively binds a highly hyperpolarizable abiological chromophore. J Am Chem Soc. 2013; 135:13914–13926. [PubMed: 23931685]
27. Solomon LA, Kodali G, Moser CC, Dutton PL. Engineering the assembly of heme cofactors in man-made proteins. J Am Chem Soc. 2014; 136:3192–3199. [PubMed: 24495285]
28. 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. 2004; 126:8141–8147. [PubMed: 15225055]
29. North B, Summa CM, Ghirlanda G, DeGrado WF. D_n-symmetrical tertiary templates for the design of tubular proteins. J Mol Biol. 2001; 311:1081–1090. [PubMed: 11531341]
30. Lahr SJ, et al. Analysis and design of turns in α-helical hairpins. J Mol Biol. 2005; 346:1441–1454. [PubMed: 15713492]
31. Goll JG, Moore KT, Ghosh A, Therien MJ. 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. 1996; 118:8344–8354.
32. Moore KT, Horváth IT, Therien MJ. Mechanistic studies of (porphinato)iron-catalyzed isobutane oxidation. Comparative studies of three classes of electron-deficient porphyrin catalysts. Inorg Chem. 2000; 39:3125–3139. [PubMed: 11196847]
33. Kaufmann KW, Lemmon GH, DeLuca SL, Sheehan JH, Meiler J. Practically useful: What the Rosetta protein modeling suite can do for you. Biochemistry. 2010; 49:2987–2998. [PubMed: 20235548]
34. Bradley P, Misura KMS, Baker D. Toward high-resolution de novo structure prediction for small proteins. Science. 2005; 309:1868. [PubMed: 16166519]
35. Tinberg, CE., Khare, SD. Computational Design of Ligand Binding Proteins. Stoddard, Barry L., editor. Springer; New York: 2016. p. 155-171.
36. Choma CT, et al. Design of a heme-binding four-helix bundle. J Am Chem Soc. 1994; 116:856–865.
37. Hayes, D., Laue, T., Philo, J. Program Sednterp: sedimentation interpretation program. Durham, NH: University of New Hampshire; 1995.
38. Davis IW, Arendall WB III, Richardson DC, Richardson JS. The backrub motion: How protein backbone shrugs when a sidechain dances. Structure. 2006; 14:265–274. [PubMed: 16472746]
39. Friedland GD, Lakomek NA, Griesinger C, Meiler J, Kortemme T. A correspondence between solution-state dynamics of an individual protein and the sequence and conformational diversity of its family. PLoS Comput Biol. 2009; 5:e1000393. [PubMed: 19478996]
40. Schrodinger LLC. The PyMOL Molecular Graphics System, Version 1.8. 2015
41. Rubtsov IV, Susumu K, Rubtsov GI, Therien MJ. Ultrafast singlet excited-state polarization in electronically asymmetric ethyne-bridged bis[(porphinato)zinc(II)] complexes. J Am Chem Soc. 2003; 125:2687–2696. [PubMed: 12603156]
42. Zimmerman DE, et al. Automated analysis of protein NMR assignments using methods from artificial intelligence. J Mol Biol. 1997; 269:592–610. [PubMed: 9217263]
43. Delaglio F, et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995; 6:277–293. [PubMed: 8520220]
44. 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. 1995; 6:1–10. [PubMed: 22911575]
45. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR. 1999; 13:289–302. [PubMed: 10212987]

46. 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-binding domain of the 434 repressor by biosynthetically directed fractional carbon-13 labeling. Biochemistry. 1989; 28:7510–7516. [PubMed: 2692701]

47. Güntert P, Mumenthaler C, Wüthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol. 1997; 273:283–298. [PubMed: 9367762]

48. Herrmann T, Güntert P, Wüthrich K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol. 2002; 319:209–227. [PubMed: 12051947]

49. Schwieters CD, Kuszewski JJ, Tjandra N, Marius Clore G. The Xplor-NIH NMR molecular structure determination package. J Magn Reson. 2003; 160:65–73. [PubMed: 12565051]

50. Bagaria A, Jaravine V, Huang YJ, Montelione GT, Guntert P. Protein structure validation by generalized linear model root-mean-square deviation prediction. Protein Sci. 2012; 21:229–238. [PubMed: 22113924]

51. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983; 79:926–935.

52. Harvey MJ, Giupponi G, Fabritiis GD. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. J Chem Theory Comput. 2009; 5:1632–1639. [PubMed: 26609855]

53. Walshaw J, Woolfson DN. SOCKET: a program for identifying and analysing coiled-coil motifs within protein structures. J Mol Biol. 2001; 307:1427–1450. [PubMed: 11292353]

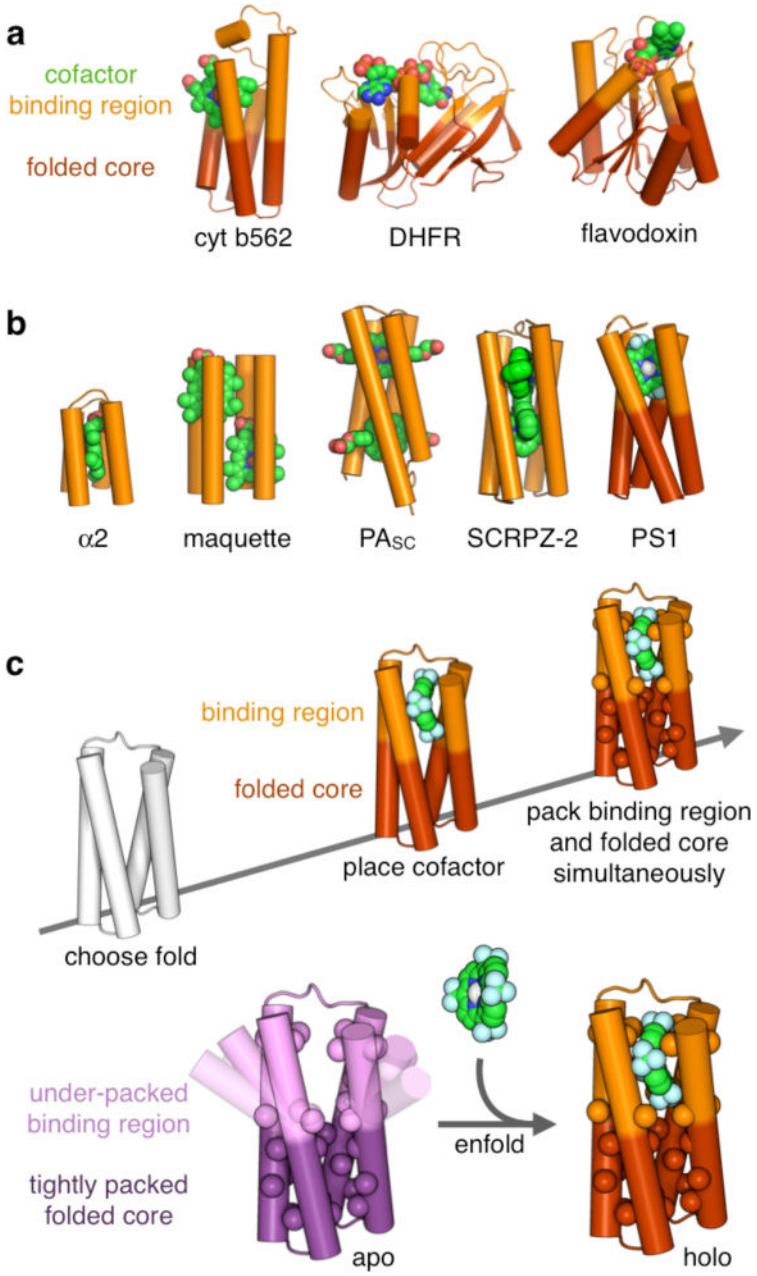


Figure 1. The design strategy

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). α2 protein is from ref [36]; 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. cyt b562, cytochrome b562 (pdb 256b), DHFR, dihydrofolate reductase (pdb 8dfr), flavodoxin (pdb 1czu).

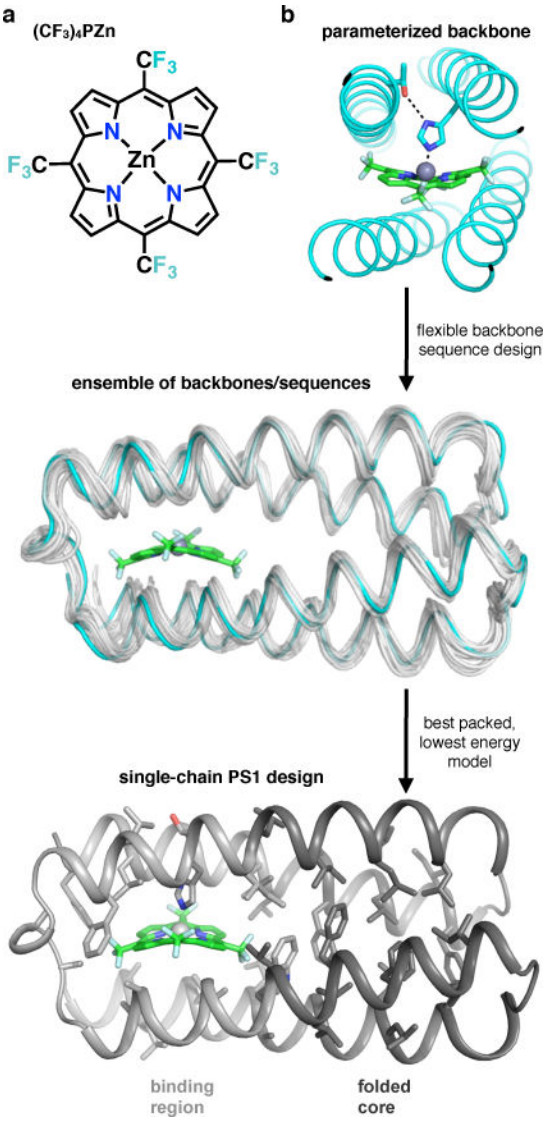


Figure 2. The computational design workflow for optimized core packing

a, The abiological porphyrin cofactor, $(CF_3)_4PZn$. **b**, The constrained, parameterized backbone of SCRPZ-2 (cyan) feeds into a flexible backbone design protocol that allows the interior side chains and backbone to simultaneously conform to the porphyrin $(CF_3)_4PZn$ (green).

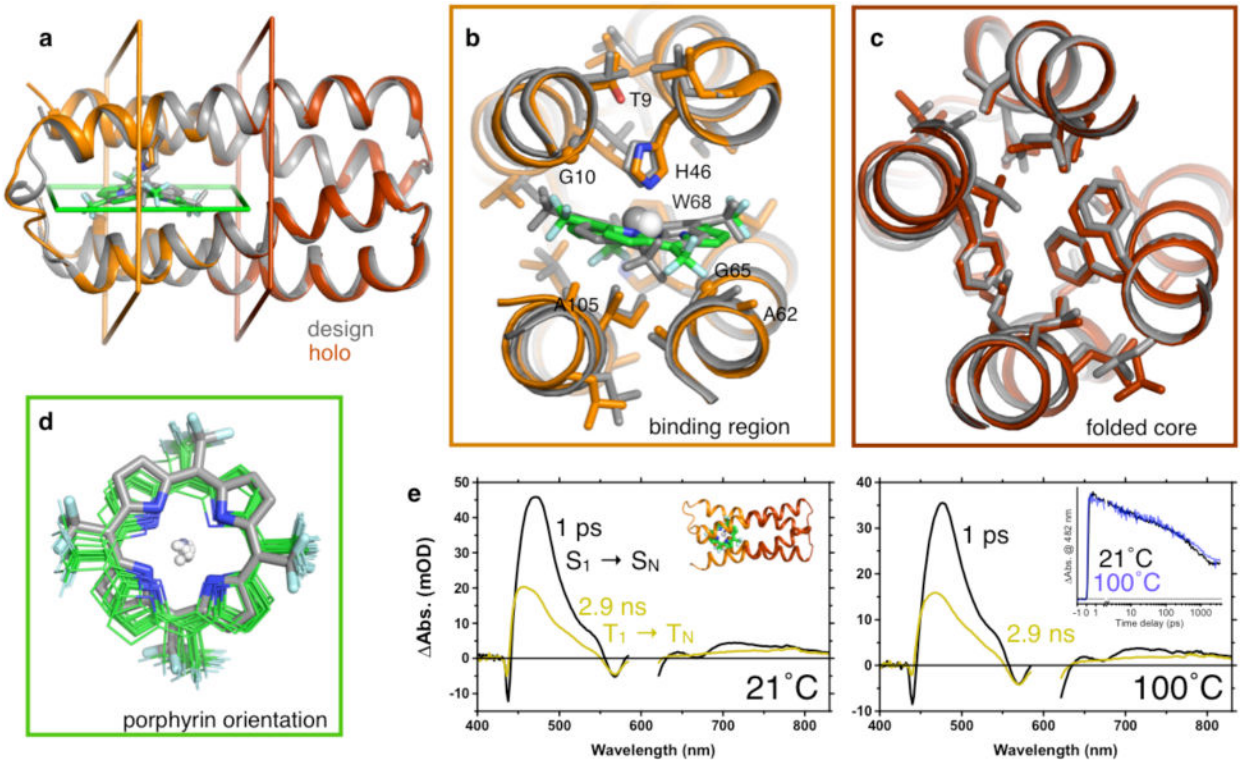


Figure 3. The structure of holo-PS1 agrees closely with the design

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**,**c**, shows ~10 Å slices of the holo-PS1 NMR centroid and design in the binding region and folded core, respectively. **d**, 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). **e**, Pump-probe transient absorption spectra of $(CF_3)_4$PZn bound in the interior of holo-PS1 at 21 °C and 100 °C. The black spectrum shows characteristic $S_1 \rightarrow S_N$ absorptions of $(CF_3)_4$PZn, which smoothly transitions into the gold spectrum showing characteristic $T_1 \rightarrow T_N$ absorptions of $(CF_3)_4$PZn. Inset exemplifies identical transient dynamics (primarily intersystem crossing from S_1 to T_1) at ΔAbs. = 482 nm (scaled). Experimental conditions: solvent = 50 mM NaPi, 100 mM NaCl, pH 7.5; excitation wavelength = 600 ± 5 nm; magic-angle polarization between pump and probe pulses; pump-probe cross-correlation of ~250 fs.

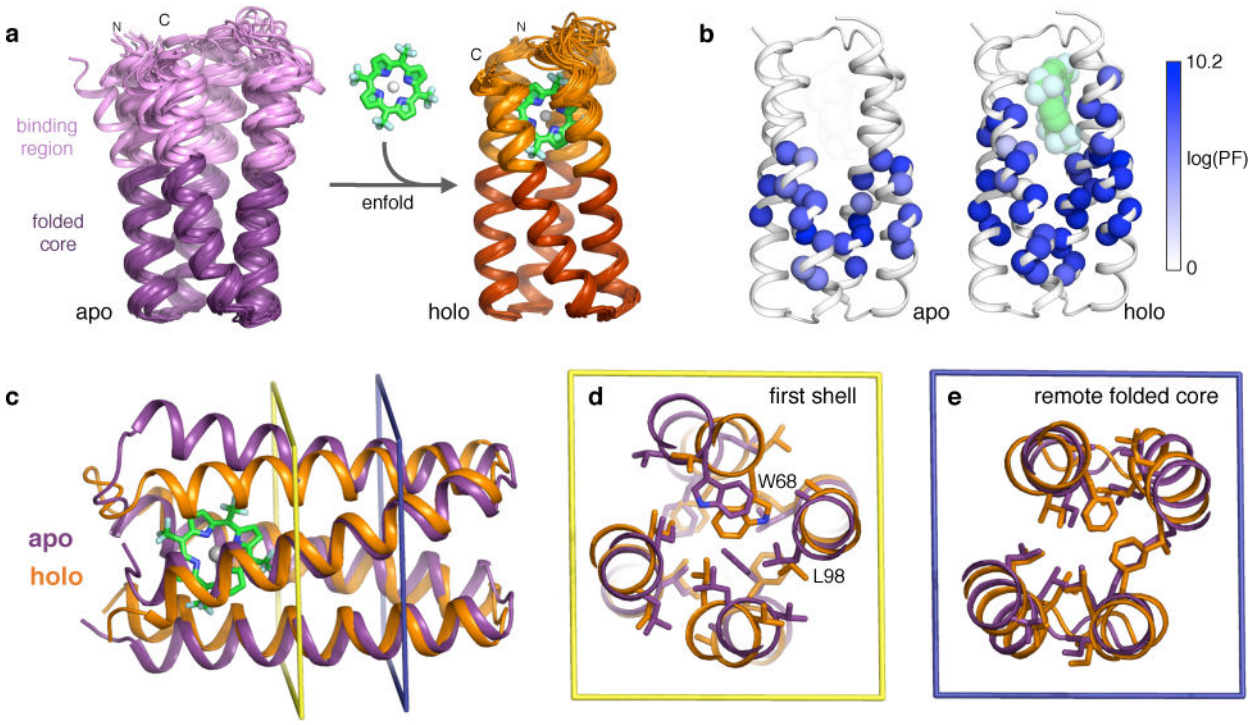


Figure 4. Apo- and holo-PS1 share similar folded cores and differ in the binding region

a, 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. **b**, 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 sidechain is protected in holo, but not apo. **c–e**, Backbone alignment of the holo- and apo-centroids at the folded core shows, **e**, agreement of side chain rotamer states far from the binding site and, **d**, 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.