Materials and Methods

Analysis of β-sheet curvature

To analyze β-sheet curvature in native protein structures we first identified for each amino acid position the secondary structure, as predicted with DSSP (*33*), and the ABEGO bin (*14*, *34*) corresponding to the Ramachandran plot region of its ϕ/ψ angles. Residues predicted as strand with "B" (beta region) and "A" (alpha region) ABEGO bins were defined as "regular" and "bulge" strand residues, respectively. In this work we have considered the "classic" bulge type, which is the most common in native β-sheets and adopts a α-helical conformation (strand residues with "A" ABEGO bin that are preceded and followed by the "B" ABEGO bin correspond to this bulge type). We only considered strands of more than four residues. It is convenient to describe the local curvature of a strand residue (*i*) with the segment of five consecutive residues (5-mer) centered in *i*, those between residue *i*-2 to *i*+2. Given the alternation of pleating and sidechain directionality in strands, we defined bending and intra-strand twist with strand residues sharing pleating and sidechain direction. We define the bending as the angle (α) formed by $C_\alpha(i\text{-}2)$-$C_\alpha(i)$-$C_\alpha(i\text{+}2)$, and intra-strand twist as the dihedral angle formed by $C_\beta(i\text{-}2)$-$C_\alpha(i\text{-}2)$-$C_\alpha(i\text{+}2)$-$C_\beta(i\text{+}2)$. For those 5-residue fragments including one bulge at position *b*, we accounted for the bulge offset in sidechain direction by calculating bending as the angle formed by $C_\alpha(b\text{-}2)$, the mid-way position between $C_\alpha(b)$ and $C_\alpha(b\text{+}1)$, and $C_\alpha(b\text{+}3)$; and intra-strand twist as the dihedral angle formed by $C_\beta(b\text{-}2)$-$C_\alpha(b\text{-}2)$-$C_\alpha(b\text{+}3)$-$C_\beta(b\text{+}3)$. We calculated the bend angle sign as a function of three vectors: (1) $\vec{c}$ as $\overrightarrow{c_1} + \overrightarrow{c_2}$, where $\overrightarrow{c_1}$ is the vector from $C_\alpha(i)$ to $C_\alpha(i\text{-}2)$ and $\overrightarrow{c_2}$ is the the vector from $C_\alpha(i)$ to $C_\alpha(i\text{+}2)$; (2) $\overrightarrow{s_1}$ as the vector from $C_\alpha(i\text{-}2)$ to $C_\alpha(i\text{+}2)$; (3) $\overrightarrow{s_{21}}$ as the vector from $C_\alpha(i)$ to the C_α of the paired residue. The bend angle sign is then calculated as $\hat{c} \cdot (\overrightarrow{s_1} \times \overrightarrow{s_{21}})$, as shown in Fig. 1A. It should be noted that under this definition the bend angle sign is unambiguous for edge strands, but for inner strands the sign changes depending on which of the two adjacent strands is considered to compute the $\overrightarrow{s_{21}}$ vector. Therefore, for comparing bend angles of inner strands the absolute value α is more appropriate. For regular and bulged strand segments we considered those with ABEGO strings "BBBBB" and "BBABBB",

respectively. To analyze bending and intra-strand twist from native β-sheets we collected 85721 regular strand 5-residue fragments and 2292 bulged strand 6-residue fragments from a non-redundant database of PDB structures obtained from the PISCES server (*35*) with sequence identity <30% and resolution ≤2 Å. We selected those 5-residue fragments only involved in antiparallel pairing and classified them as "edge" or "inner" segments depending on whether they have one or two pairing strands flanking them, respectively.

To identify the amino acid preferences of bent strands in uniform β-sheets we considered 5-residue fragments of edge strands and two neighboring strands. For bulged β-sheets, due to the offset in sidechain directionality, we considered 6-residue fragments of bulged (edge) strands and 5-residue fragments of the two neighboring strands. The frequency of each amino acid at each position was normalized by the frequency of the amino acid to be found in a strand.

Computational design process

Protein backbone construction

Protein backbones were generated by Monte Carlo fragment assembly using 9- and 3-residue fragments with the target secondary structure and torsion bins (ABEGO), using the Blueprint Builder mover (*12*) implemented in RosettaScripts (*36*). We restricted regular strand and bulge residues to the "B" and "A" ABEGO bins, respectively. These Rosetta folding simulations use a sequence-independent centroid representation of the protein, as well as a scoring function that includes a hydrogen bonding term for backbone atoms, a Van der Waals term to avoid steric clashes, an omega angle term to ensure planarity of the peptide bond, and a radius of gyration term to favor compact structures. Thousands of independent folding trajectories are performed and subsequently filtered.

When building backbones involving non-local contacts, adding a constraint term to the scoring function increases the efficiency of the folding simulations. Due to the non-local character of β-sheet contacts, we used distance and angle constraints to favor the ideal geometry of the backbone-backbone hydrogen bonds between the paired strand residues. For bulged strand pairs both the bulge and the residue following donate hydrogen bond to the same residue ("X") in the paired strand, but with different hydrogen

bond distances according to distributions from native protein structures. The hydrogen bond distances to residue X from the bulge and the residue following were constrained to 2.9 and 3.4 Å, respectively. Conveniently, once the register shift between paired strands and the strand pairing type (parallel or antiparallel) are defined, all pairings between strand residues and their corresponding constraints are determined. Additionally, when building flexible elements such as N- or C-terminal helices or loops with a particular hydrogen bond pattern, the use of constraints allows sampling structures closer to the target with more efficiency.

Strand fragments with low bending and twist are overrepresented in the fragment library derived from the PDB and, as a consequence, constraints are necessary to favor the construction of backbones with increased strand bending and twist. We used angle constraints between C-alpha atoms of residues with the same pleating at different separation levels, i.e. $C_\alpha(i\text{-}2n)$-$C_\alpha(i)$-$C_\alpha(i\text{+}2n)$ where i is the central residue and n is the separation level. Similarly, for twist we used dihedral constraints $C_\beta(i)$-$C_\alpha(i)$-$C_\alpha(i\text{+}2n)$-$C_\beta(i\text{+}2n)$. The separation level provides control on the degree of locality of strand curvature. In addition, the inter-strand twist for an antiparallel pairing can also be controlled with dihedral constraints for $C_\alpha(k\text{+}2)$-$C_\alpha(k)$-$C_\alpha(p_k)$-$C_\alpha(p_k\text{-}2)$, where p_k is the strand residue paired to residue k.

Stepwise backbone building

The introduction of constraints in the fragment sampling trajectory can rapidly increase the ruggedness of the energy landscape, leading to its frustration, i.e. the trajectory sticking at a local energy minimum. This is a general limitation of the fragment-based approach that we circumvented by building backbones stepwise and constraining non-local contacts. We divided the construction of the target folds in several steps. For instance, for Fold E: (1) central 4-stranded antiparallel β-sheet with a β-bulge in each edge-strand; (2) helix 3 and hairpin-interdomain connection; (3) helices 1 and 2 added at the N-terminus; (4) addition of C-terminal helix. We used constraints for building the β-sheet (strand pairings, bending and twist), the inter-domain connection and positioning helices 1 and 2. Helix 1 is a flexible element at the N-terminus that we constrained at interacting distance from the edge strand of the β-sheet. The loop

connecting helix 2 and the inter-domain connection was constrained to hydrogen bond the backbone of the edge strand. Helix 4 in Fold E designs was constrained to pack onto helix 3 at the entrance of the pocket.

We have used four criteria to filter protein backbones at each step:

1. *Target topology:* protein models are filtered according to the match between the blueprint and the detected secondary structure, ABEGO sequence and topology (strand and helix pairings) of the built model.
2. *Native-like backbones:* to favor native-like backbones, protein models are also filtered on the basis of backbone hydrogen bonding energy (lr_hb score), Cβ-average degree (average number of Cβ-Cβ contacts between residues within 10 Å) and balance between exposed and buried SASA to favor compact structures. Additionally, we checked for deviations between backbone fragments of the designed structures and native fragments (FragmentLookup filter), which is indicative of local backbone strain.
3. *Geometrical features defining target structure:* depending on the protein topology to be built, additional filters are considered to evaluate the geometry of secondary structure elements as well as their relative orientation, such as the strand twist/bending or the distance/angle between helix and strand.
4. *Canonical loops:* the conformations of loops connecting two secondary structure elements can be discretized by the sequence of their torsion bins (ABEGO). Previous works (*12*, *14*, *34*) have mined the PDB for information on the relationship between loop length and ABEGO, and the orientation and type of the secondary structure elements they bridge; we used this information to select the length and ABEGO bins of all loops. Only using the most frequent loop ABEGOs facilitates the design of their amino acid sequences, as explained below.

Sequence design

Thousands of backbones are subjected to RosettaDesign calculations (*24*, *37*) with the full-atom Talaris2013 (*38*, *39*) scoring function to favor amino acid identities and side-chain conformations with low-energy and tight packing. The design calculation

corresponds to cycles of fixed backbone design followed by backbone relaxation, and the designs were filtered based on three independent criteria:

- Low total energy
- Tight packing: RosettaHoles (*40*), shape complementarity between secondary structure elements, packstat and core side-chain average degree. Side-chain average degree is the average number of hydrophobic sidechain heavy-atom contacts within 4 Å. We developed this filter to improve the packing in the core of protein folds with large pockets, which are difficult to pack efficiently. This minimized the number of alanines in helices and valines in strands, while increasing the number of large hydrophobic sidechains.
- High sequence-structure compatibility: match between secondary structure of the designed structure and Psipred (*41*) secondary structure prediction from the designed amino acid sequence.

To achieve very low energy sequences with tight packing, for each backbone we ran multiple Generic Monte Carlo trajectories of the design protocol, optimizing simultaneously total energy and side-chain average degree, and subsequently applied all filters. The design calculations are performed using a restricted set of amino acids and rotamers for each position. The restrictions were such that hydrophobic amino acids were allowed in the core and polar amino acids in the surface. To improve the local sequence-structure compatibility in loops and β-bulges we restricted their amino acid identities to the subset of amino acids most frequently observed in similar fragments in the PDB. This was done by the creation of sequence profiles for loops that shared the same ABEGO bins and adjacent secondary structure elements. The top 5 most frequent amino acids in each position were the only ones allowed, unless there was a strong preference for a particular amino acid. Additionally, amino acids identities conflicting with the expected hydrophobicity pattern were excluded. The loop ABEGO classification in combination with the corresponding sequence profile allows the automatic identification of well-known local sequence-structure motifs, such as N-terminal helix capping residues (D, N, S and T) or prolines that restrict the ϕ/ψ of the residue immediately before. These sequence motifs are seldom identified by the score function, thus giving poorer local

sequence-structure compatibility. For β-bulges we built sequence profiles for positions *b*-1, *b*, *b*+1, *b*+2 and X; where *b* is the bulge position and X is the strand residue paired to the bulge. In general, positions *b* and *b*+1 were restricted to RKEQ, and *b*-1 and *b*+2 to ILVFY. To minimize the aggregation propensity, we incorporated polar residues at inward-pointing positions of edge strands and removed surface exposed hydrophobic residues. Due to the large size of the pockets of the target folds, efficient core packing was achieved by a high number of aromatic sidechains. As part of the protein core is solvent-exposed we preserved well-packed exposed aromatics that hydrogen bond polar residues at the surface (especially Trp-Glu and Tyr-Glu interactions).

Sequence-structure compatibility

The compatibility between sequence and backbone structure is assessed in three steps:

1) *Fragment quality assessment.* The designed model sequence is spliced in overlapping 9-residue fragments, and two hundred 9-residue fragments with the same sequence and secondary structure are picked from a PDB-derived fragment database for each position. The RMSDs between all picked 9-mer fragments and the corresponding 9-mer of the designed structure are calculated. Two metrics evaluating the overall structural similarity between the ensemble of picked fragments and the designed structure are calculated to rank designs based on fragment quality. First, the percentage of fragments with RMSD < 1.5 Å and, second, the RMSD of the best fragment at the worst position. The quality of these fragments tests compatibility of the sequence and backbone structure at the local level.

2) *Biased Forward Folding.* After verifying the fragment quality, the sequence-structure compatibility is assessed at the global level by characterizing the folding energy landscape with Rosetta *ab initio* folding simulations starting from an extended chain (*27*, *28*), on the Rosetta@home server. This is the most stringent computational test and those designs with funnel-shaped energy landscapes are selected for experimental characterization. In general, hundreds of designs pass the fragment quality filter and their folding energy landscape should be assessed. However, these simulations are too computationally demanding. The high contact order of the protein folds targeted in this work complicated the identification of designs with funnel-shaped energy landscapes and

required to screen by *ab initio* folding too many designs with good fragment quality. We developed a new method, *Biased Forward Folding*, to quickly assess the folding energy landscape and select the most promising candidates for unbiased *ab initio* structure prediction. The standard Rosetta *ab initio* structure prediction method starts with a fragment picking process in which at each residue position 9- and 3-residue fragments are selected from the fragment library on the basis of similarity in sequence and secondary structure prediction. The top scoring fragments are then subjected to a Monte Carlo assembly process using a low resolution scoring function and, in a second step, the lowest energy structures are relaxed with a high-resolution scoring function. The fragment assembly process performs the large scale conformational sampling, while the high-resolution relaxing step is limited to local backbone perturbations allowing sidechains to repack and find low energy structures. Therefore the selection of fragments and their assembly process are the two primary limiting factors in sampling conformations close to the designed structure and obtain funnel-shaped energy landscapes. We hypothesized that those picked fragments structurally similar to the designed structure fragments are the main contributors to sampling near the designed structure during *ab initio*. Biasing *ab initio* folding simulations using a small subset of fragments close in RMSD to the design structure is therefore expected to have predictive power of the funnel character of the energy landscape near the design structure. If under this bias, sampling trajectories do not reach the target structure it is very unlikely that the standard *ab initio* simulation will sample closer. With a smaller set of fragments the number of folding trajectories necessary to map the energy landscape available gets dramatically reduced. We selected the three lowest-rmsd fragments (9 and 3 residues long) picked at each position and ran a low number of *ab initio* folding trajectories (between 30 and 50). This allows screening 10-100 times more designs than with *ab initio* folding simulations.

3) *Ab initio structure prediction*. Those designs having funnel-shaped energy landscapes in Biased Forward Folding simulations are then subjected to standard *ab initio* structure prediction simulations on Rosetta@home. For an energy landscape obtained from Biased Forward Folding or *ab initio* structure prediction to be funnel shaped we required to get sampling below 2 Å RMSD to the relaxed structure and a large energy gap with

alternative structures to ensure that the designed structure is achievable and lower in energy to alternate states.

Computational design of homodimers

We used the Residue Pair Transform method (*42*) to generate docking configurations with C2 symmetry suitable for designing the homodimer interface. We restricted the docking process to configurations that exclude helices from the dimer interface and maximize the number of β-sheet contacts. The top 50 scoring docked configurations were subjected to interface design calculations. Those β-sheet residues at the convex face with the C_β atom within 10 Å of a C_β atom of the other subunit were selected for design. The possible amino acid identities at each design position were restricted based on the solvent accessible surface area (SASA). Designs were filtered based on buried SASA, shape complementarity and binding energy. Designs passing these criteria were subjected to asymmetric docking simulations and those with funnel-shaped energy landscapes were selected for experimental characterization.

Design of disulfide bonds

We used the *Disulfidize* mover implemented in RosettaScripts to screen for pairs of residue positions with proper geometry for disulfide bond formation. We favored disulfide bonds between residues distant in primary sequence (at least a 6-residue separation) and with a disulfide score < -1.0. To increase the likelihood of finding good geometries for disulfide bond we locally perturbed the backbone structure with small moves (*27*) using the *Small* mover in RosettaScripts.

Cavity-creating mutations

We redesigned residues close to the cone base and restricted the calculations to amino acid identities with smaller hydrophobic or polar sidechains.

Visualization of protein structures and image rendering

Images of protein structures were created with PyMOL (*43*) and Chimera (*44*).

Experimental characterization

Protein expression and purification

Genes encoding the designed protein sequences were obtained from Genscript and cloned into pET21_NESG (*45*, *46*) (with C-terminal 6xHis tag) or pET-28b+ (with N-terminal 6xHis tag and a thrombin cleavage site) expression vectors. Plasmids were transformed into chemically competent *Escherichia coli* BL21 Star (DE3) cells from Invitrogen. Starter cultures were grown at 37°C in Luria-Bertani (LB) medium overnight with antibiotic (50 μg/ml carbenicillin for pET21-NESG expression or 30 μg/ml kanamycin for pET-28b+ expression). For expression of non-labelled proteins, overnight cultures were used to inoculate 500 ml of LB medium supplemented with antibiotic. To express ^{15}N-labelled proteins for NMR spectroscopy, starter cultures were transferred to 40 mL of MJ9 minimal media (*47*) with antibiotic, were grown overnight and used to inoculate 500 ml of minimal media. After inoculation, cells were grown at 37 °C and 225 r.p.m until an optical density (OD_{600}) of 0.5-0.7 was reached. Protein expression was then induced with 1mM of isopropyl β-D-thiogalactopyranoside (IPTG) at 18 °C. After overnight expression, cells were collected by centrifugation (at 4 °C and 4400 r.p.m for 10 minutes) and resuspended in 25 ml of lysis buffer (20 mM imidazole and phosphate buffered saline, PBS). Resuspended cells were lysed by sonication or microfluidizer in the presence of lysozyme, DNAse and protease inhibitors. Lysates were centrifuged at 4 °C and 18,000 r.p.m. for 30 minutes; and the supernatant was filtered and loaded to a nickel affinity gravity column pre-equilibrated in lysis buffer for purification. The column was washed with three column volumes of PBS+30 mM imidazole and the purified protein was eluted with three column volumes of PBS+250 mM imidazole. The eluted protein solution was dialyzed against PBS buffer overnight. The expression of purified proteins was assessed by SDS-polyacrylamide gel electrophoresis and mass spectrometry; and protein concentrations were determined from the absorbance at 280 nm measured on a NanoDrop spectrophotometer (ThermoScientific) with extinction coefficients predicted from the amino acid sequences. Proteins were further purified by FPLC size-exclusion chromatography using a Superdex 75 10/300 GL (GE Healthcare) column.

Site-directed mutagenesis

Single-point mutations were obtained by QuikChange site-directed mutagenesis using 0.75 µl of the pET-28b+ constructs as templates, 1 µl of Phusion high-fidelity DNA polymerase (New England BioLabs), 10 µl of 5X Phusion buffer (New England BioLabs), 1.25 µl of a 10 mM deoxynucleotides (dNTP) solution mix and 1 µl of the designed forward and reverse primers solutions at 125 ng/µL. Primers were ordered from Integrated DNA Technologies. Full-length gene product was assembled by 1 cycle of PCR (95 °C 1.5 min), 18 cycles of PCR (95 °C 30 s, 55 °C 30 s, 72 °C 4 min) and 1 cycle of PCR (72 °C 6 min). Mutations were confirmed by sequencing.

Circular dichroism (CD)

Far-ultraviolet CD measurements were carried out with an AVIV spectrometer, model 420. Wavelength scans were measured from 260 to 195 nm at temperatures between 25 and 95 °C. Temperature melts monitored absorption signal at 220 nm in steps of 2 °C/min and 30 s of equilibration time. For wavelength scans and temperature melts a protein solution in PBS buffer (pH 7.4) of concentration 0.2-0.4 mg/ml was used in a 1 mm path-length cuvette.

Chemical denaturation experiments with guanidium chloride (GdmCl) were done with an automatic titrator using a protein concentration of 0.02-0.04 mg/ml and a 1 cm path-length cuvette with stir bar. PBS buffer (pH 7.4) was used for the cuvette solution and PBS+GdmCl for the titrant solution at the same protein concentration. GdmCl concentration was determined by refractive index. The denaturation process monitored absorption signal at 220 nm in steps of 0.2 M GdmCl with 1 min mixing time for each step and at 25 °C. The denaturation curves were fitted by non-linear regression to a two-state unfolding model to extract six parameters: slope and intercept for pre- and post-transition baselines, *m* value and the folding free energy (ΔG_{H2O}) (*48*, *49*). The deviation of the fitted *m* value from its expected value given protein size was computed using the empirical correlation between the number of protein residues and the protein *m* value for denaturation with GdmCl (*50*).

Size exclusion chromatography combined with multiple angle light scattering (SEC-MALS)

SEC-MALS experiments were performed using a Superdex 75 10/300 GL (GE Healthcare) column combined with a miniDAWN TREOS multi-angle static light scattering detector and an Optilab T-rEX refractometer (Wyatt Technology). One hundred microliter protein samples of 1-3 mg/ml were injected to the column equilibrated with PBS (pH 7.4) or TBS (pH 8.0) buffer at a flow rate of 0.5 ml/min. The collected data was analyzed with ASTRA software (Wyatt Technology) to estimate the molecular weight of the eluted species.

Nuclear magnetic resonance spectroscopy

^{15}N-HSQC screening

To evaluate whether the designed proteins fold into well-ordered structures ^{15}N-HSQC screening was carried out at 20 or 25 $^{\circ}$C using a 1.7 mm micro cryoprobe with automatic sample changer at 600 MHz. The spectra were generally recorded in multiple buffers, using standard protocols that have been published previously (*45*, *51*). The buffers and temperatures providing the best quality spectra were used for the analyses provided in this study.

NMR structure determination of dcs_A_3 and dcs_B_2

The selected designs (dcs_A_3, NESG target OR485; dcs_B_2, NESG target OR664) were expressed and purified by following the standard NESG protocols (*45*). Synthetic genes (Genscript) cloned into the pET21_NESG expression vector (*45*, *46*) were expressed in *E. coli* BL21 (DE3) pMGK cells as *U*-^{15}N, 5%^{13}C-enriched, and *U*-^{15}N, *U*-^{13}C-enriched proteins, using MJ9 minimal media (*47*), ^{13}C-glucose and $^{15}NH_3Cl$ as the sole sources of carbon and nitrogen, respectively. *U*-^{15}N, 5%^{13}C-labeled proteins were generated for stereo-specific assignments of isopropyl methyl groups of valines and leucines (*52*). Samples were determined to be homogeneous (>95%) by SDS-PAGE, and monomeric by size exclusion chromatography. The molecular weights of ^{13}C,^{15}N-enriched OR485 and ^{13}C,^{15}N-enriched OR664 were confirmed as 10.61 kDa and 14.31 kDa by MALDI-TOF, respectively, in good agreement with theoretical values (10.64 kDa

and 14.33 kDa, respectively). The yields were 20 mg and 15 mg per liter culture, respectively.

All NMR spectra were recorded at 25 °C using Bruker AVANCE NMR spectrometer systems with cryogenic NMR probes at 600 and 800 MHz. The NMR structures were determined using standard NMR structure determination protocols, as previously described (*53*). NMR structures were determined in a "blind" fashion; i.e. without knowledge of the design structure. Structure quality assessment was done using the Protein Structure Validation Software (PSVS) software suite (*54*, *55*). Chemical shifts data and final structure coordinates were deposited in the Biological Magnetic Resonance Bank and Protein Data Bank, respectively. (NESG ID, BMRB and PDB IDs: OR485, BMRB 30139, 5kph for dcs_A_3; and OR664, BMRB 30128, 5kpe for dcs_B_2). The refinement statistics for the final structures are summarized in Table S6.

Crystallization, data collection and structure determination

dcs_A_4 (NESG target OR486)

A DNA fragment encoding dcs_A_4 was synthesized and cloned into the bacterial expression vector pET21_NESG (*45*, *46*), with a short C-terminal purification tag "LEHHHHHH". The plasmid was then transformed into *E. coli.* BL21(DE3) cells (Strategene) and grown in LB media (1L) at 37 °C to OD_{600} of 0.8 units, and induced with 1 mM IPTG over night at 17 °C. The bacteria were pelleted by centrifugation, and resuspended in 1x PBS buffer by mild sonication to release the soluble target protein. After high-speed centrifugation, the supernatant was applied to a 5 ml His-tag affinity column (GE Healthcare), and eluted with a linear (50-500 mM) imidazole gradient. Further purification was carried out by size exclusion chromatography using a HighLoad 26/60 Superdex S75 column (GE Healthcare). The purified protein was over 95% pure based on SDS PAGE, and was also validated by MALDI-TOF mass spectrometry.

The purified dcs_A_4 (NESG target OR486) was concentrated to 10 mg/ml in 100 mM NaCl, 5 mM DTT, 0.02% NaN_3, 10 mM Tris-HCl at pH 7.5 and stored at -80 ˚C prior to crystallization. The initial crystallization screening was carried out at the high-throughput screening (HTS) facility at Hauptman-Woodward Institute (HWI) located in Buffalo, NY, where 1536 crystallization conditions were screened using the microbatch

method (*56*). Initial crystallization hits were further optimized manually to obtain diffraction quality crystals. The addition of detergents in this screen was key to improving the crystals' quality. Optimal conditions for crystallization were obtained at room temperature in 0.1 M NaH_2PO4, 0.1 M Na Acetate, pH 5.5 and 28% PEG 400. Diffraction of OR486 crystals was first tested using a home X-ray facility with a Rigaku RAXV ++ detector. The crystals were harvested directly from the drops and flash-frozen in liquid nitrogen. Diffraction data set to 2.44 Å was collected at the National Synchrotron Light Source, with beamline X4C, and the data were processed with HKL-2000 (HKL Research, Inc.). The structure was determined by molecular replacement using Phaser (*57*), with a preliminary NMR model of OR485 as initial search model. The refinement was carried out using Phenix (*58*, *59*), and model adjusting was done in Coot (*60*). The statistics for the final structure refinement and model geometry are summarized in Table S4.

dcs_C_1_ss, dcs_D_2, dcs_E_3, dcs_E_4, dcs_E_4_dim9 and dcs_E_4_dim9_cav3

To prepare protein samples for X-ray crystallography, the buffer of choice was 25 mM Tris, 300 mM NaCl, pH 8.0. Proteins were expressed from pET28b+ constructs to cleave the 6xHis tag with thrombin. Dialyzed proteins were incubated with thrombin (1:5000 dilution) overnight at room temperature and cleaved samples were loaded to a column of benzamidine resin pre-equilibrated in lysis buffer. Resin was resuspended and nutated for 30-60 minutes to remove thrombin from solution. Flow-through was collected and washed with 3-5 mL of lysis buffer. Protease inhibitor (phenylmethylsuphonyl fluoride, PMSF) was added to the eluted sample, which was then applied to a nickel affinity column pre-equilibrated in lysis buffer to remove the cleaved 6xHis tag from solution. Flow-through was collected and washed with 1-2 column volumes. Proteins were further purified by FPLC as described above and specific cleavage of the 6xHis tag was tested by mass spectrometry.

Purified proteins were concentrated to approximately 10-20 mg/ml for screening crystallization conditions. Commercially available crystallization screens were tested in 96-well sitting or hanging drops with different protein:precipitant ratios (1:1, 1:2 and 2:1) using a mosquito robot. When possible, initial crystal hits were grown in larger 24-well

hanging drops. Obtained crystals were flash-frozen in liquid nitrogen. X-ray diffraction data sets were collected at the Lawrence Berkeley National Laboratory (LBNL). Crystal structures were solved by molecular replacement with Phaser (*57*) using the design models as the initial search models. The structures were built and refined using Phenix (*58*, *59*) and Coot (*60*).

The crystallization conditions for the solved crystal structures are the following:

- **dcs_C_1_ss:**
  - Protein solution: 15 mg/ml, 25 mM Tris hydrochloride (pH 7) and 0.1 M sodium chloride
  - Reservoir solution: 0.1 M Tris hydrochloride, pH 8.5 and 25% PEG 3,350
  - 20% glycerol as a cryoprotection solution
- **dcs_D_2:**
  - Protein solution: 16 mg/ml, 25 mM Tris hydrochloride (pH 8) and 0.3 M sodium chloride
  - Reservoir solution: 0.1 M sodium MOPS/HEPES, pH 7.5, 12.5% PEG 1000, 12.5% PEG 3350 and 12.5% 2-methyl-2,4-pentanediol and 0.2 M of amino acids (sodium glutamate, DL-alanine, glycine, DL-lysine HCl and DL-serine).
  - No cryoprotection added
- **dcs_E_3:**
  - Protein solution: 11 mg/ml, 25 mM Tris hydrochloride (pH 8) and 0.1 M sodium chloride
  - Reservoir solution: 0.2 M ammonium citrate dibasic and 30% PEG 3350
  - No cryoprotection added
- **dcs_E_4:**
  - Protein solution: 27 mg/ml, 25 mM Tris hydrochloride (pH 8) and 0.3 M sodium chloride
  - Reservoir solution: 0.1 M bicine/Trizma base, pH 8.5, 10% PEG 20 000, 20% PEG MME 550 and 0.03 M of each ethylene glycol (diethyleneglycol, triethyleneglycol, tetraethyleneglycol and pentaethyleneglycol).
  - No cryoprotection added

- **dcs_E_4_dim9:**
  - Protein solution: 8 mg/ml, 25 mM Tris hydrochloride (pH 8) and 0.3 M sodium chloride
  - Reservoir solution: 0.1 M potassium thiocyanate, pH 8 and 30% PEG MME 2000
  - 32% PEG MME 2000 and 10% glycerol as a cryoprotection solution
- **dcs_E_4_dim9_cav3:**
  - Protein solution: 8 mg/ml, 30 mM Tris hydrochloride (pH 8) and 0.1 M sodium chloride
  - Reservoir solution: 0.1 M sodium MOPS/HEPES, pH 7.5, 10% PEG 20 000, 20% PEG MME 550 and 0.3 M of halides (sodium fluoride, sodium bromide and sodium iodide).
  - No cryoprotection added.

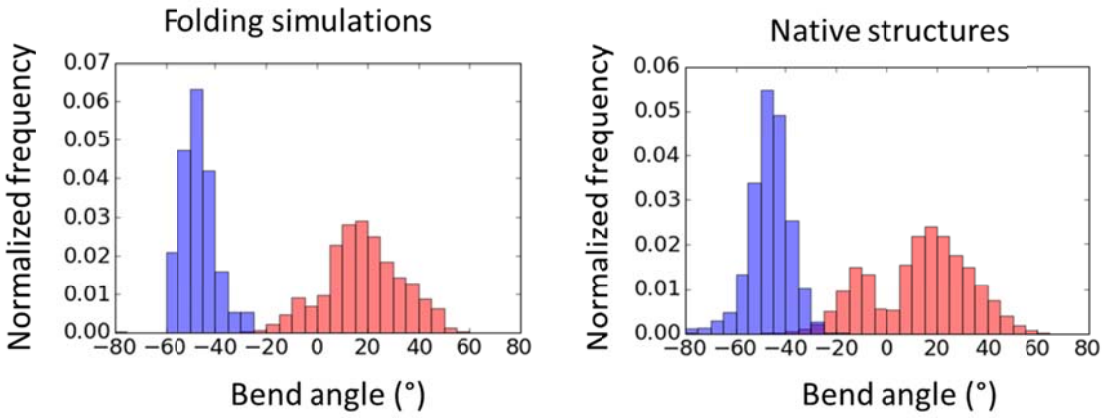

Fig. S1. Comparison of bend angle distributions from Rosetta folding simulations and native protein structures. Distributions for strand pairs formed by uniform and bulged strands are shown in red and blue colors respectively. Simulation distributions were obtained from two-stranded antiparallel β-sheets built by fragment assembly. Right panel shows the same distribution as in Fig. 1B for comparison. Both folding simulations and native structural analysis show that uniform and bulged strand pairs favor positive and negative bend angles respectively.

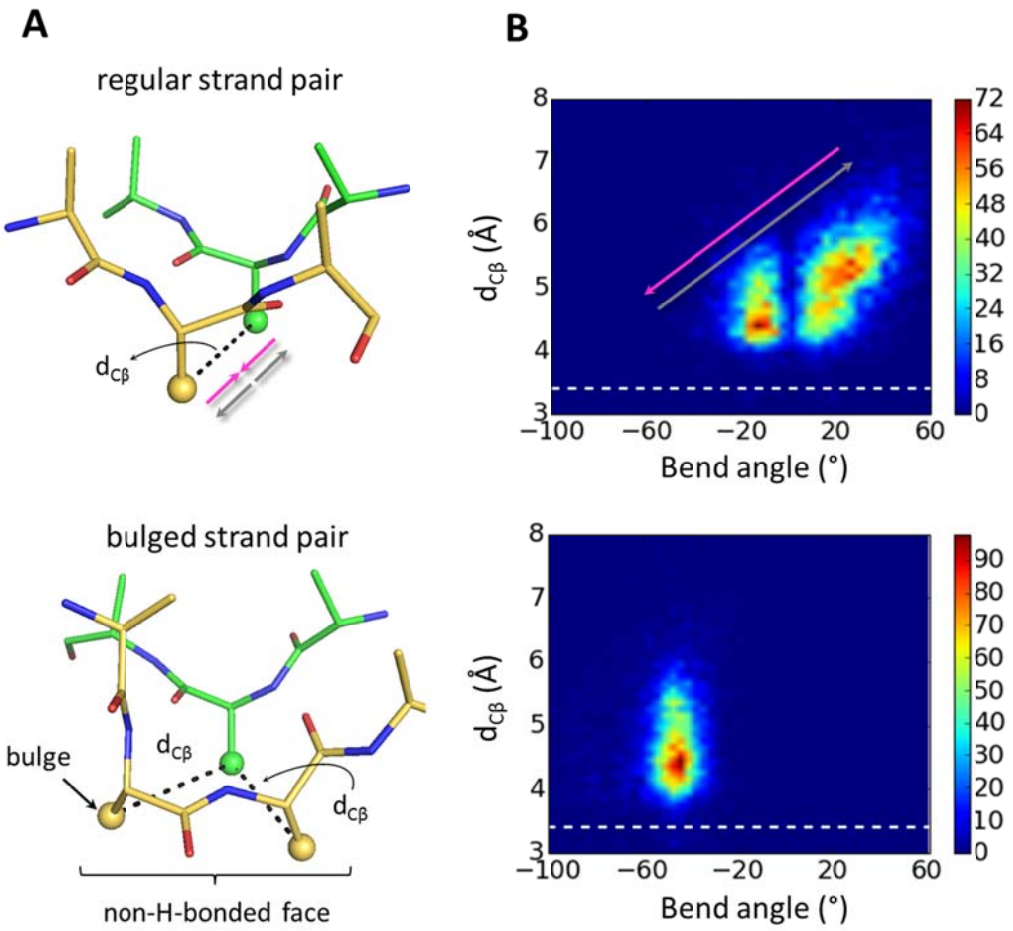

Fig. S2. Steric effects associated to the bend angle sign. (A) Local geometry of regular and bulged strand pairs indicating the $C_\beta \cdots C_\beta$ inter-atomic distance ($d_{C\beta}$) between paired residues (dashed lines). Gray and pink arrows show $d_{C\beta}$ changes correlated with positive and negative bend angles, as shown in panel B. For the bulged pair, due to the offset in sidechain directionality, the $d_{C\beta}$ is also considered for the C_β of the residue following the bulge. The different hydrogen bond pairing of bulges prevents strand pairing in one face of the strand as indicated. (B) Distribution of bend angle sign and $d_{C\beta}$ for the two strand pair types. The white dashed line at 3.4 Å shows the steric clash limit between two carbon atoms (sum of Van der Waals radii). For regular strand pairs, the increase of bend angle tends to increase $d_{C\beta}$. Bulged strand pairs achieve more negative bend angles than regular strand pairs without decreasing $d_{C\beta}$ further. While the local geometry of bulges minimizes steric effects favoring negative bend angles, it disallows positive bend angles by preventing hydrogen bond pairing in one face of the strand. The low frequency of perfectly flat regular strands (see gap close to 0°) is due to partial contribution of intra-strand twist to the bend angle calculation.

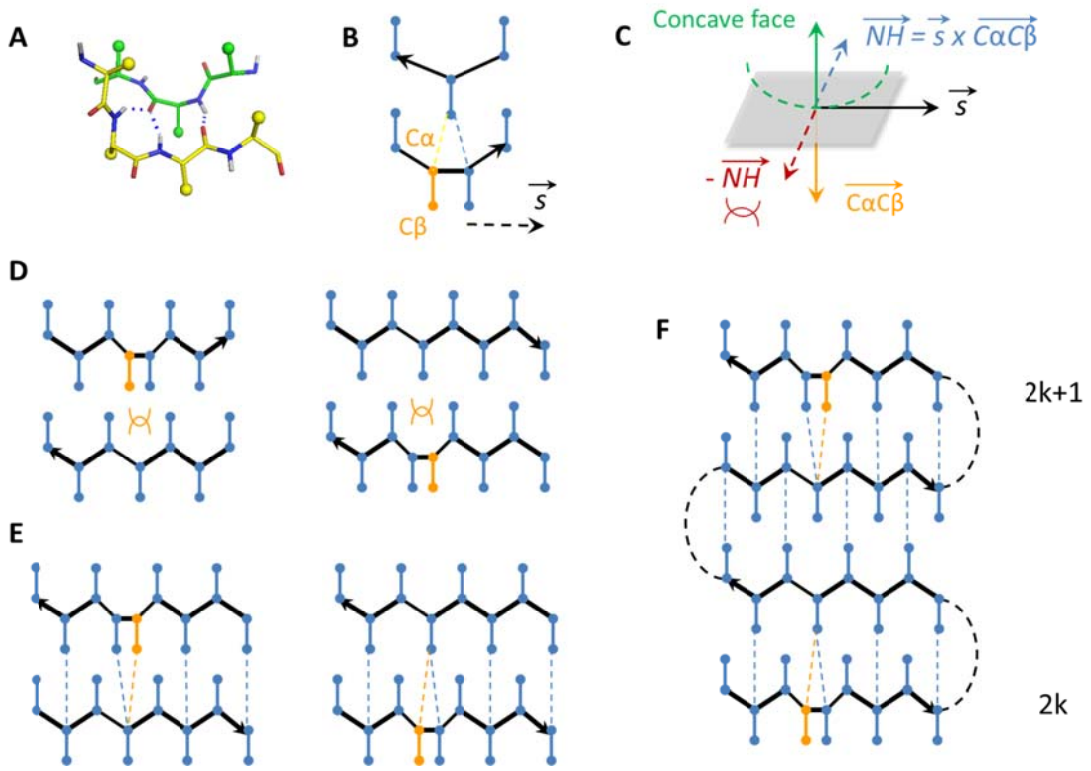

Fig. S3. Restrictions of β-bulges on antiparallel strand pairings. (A) Local geometry of a bulged strand pair and (B) its diagram representation. Bulges are highlighted in orange, regular strand residues are shown in blue and the vector $\vec{s}$ indicates the bulged strand direction. (C) Description of the hydrogen bonding orientation of the bulge with respect to the concave face of the bulge local bend. Blue and red arrows indicate the directions where hydrogen bond is allowed and disallowed respectively. (D) Diagram representation of incompatible strand pairings in the presence of a bulge. (E) Diagram representation of compatible strand pairings in the presence of a bulge. Antiparallel hydrogen bonding between paired residues is drawn with dashed lines. (F) Diagram of the strand pairing arrangement of a 4-stranded antiparallel β-sheet compatible with two bulges at the edge strands. Bulges must be located at even, 2k, and odd positions, 2k+1, from the following and previous hairpin connections, respectively.

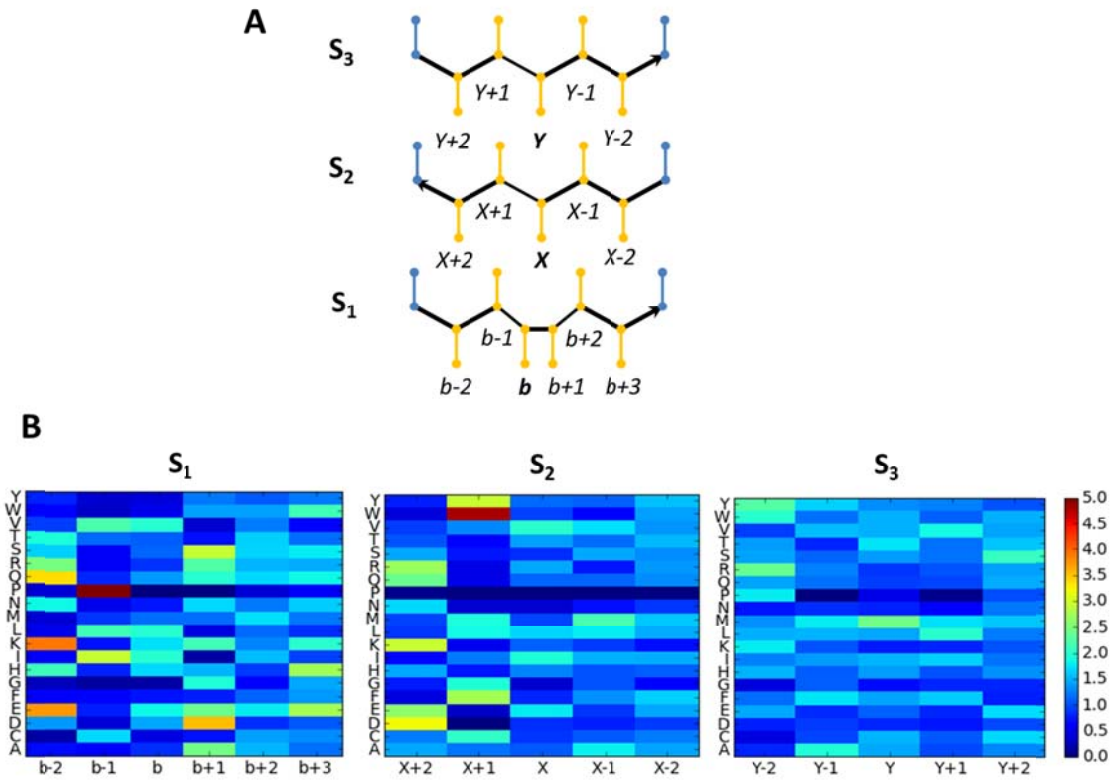

Fig. S4. Amino acid preferences in bulged β-sheets. (A) Diagram of a 3-stranded bulged β-sheet. Residues used for calculating sequence profiles are in orange and labeled; where *b* denotes the bulge position at strand S_1, X the residue paired to *b* and *b*+1 at strand S_2, and Y the residue paired to X at strand S_3. (B) Sequence profiles colored by the frequency of each amino acid relative to the frequency to be found in strands. Amino acids more favored at bulged strand (S_1) positions: *b*-2, polars; *b*-1, Pro or aliphatics; *b*, aliphatics or polars; *b*+1, glycine or polars; and *b*+3, polars. These preferences point to some degree of alternation in the pattern of polar and hydrophobic residues. At the inner strand (S_2): X+1, Gly and non β-branched amino acids (including aromatics); X+2, polars. Position X+1 favors Gly and non β-branched amino acids, which allow to increase bending in strand S_2 so as to follow the high bending of the strand S_1 bulge (residues without C_β or β-branching diminish steric repulsion between X+1 and *b*-1 in adjacent strands, thus favoring a highly negative bend angle; see fig. S2). At strand S_3 no significant preferences are observed.

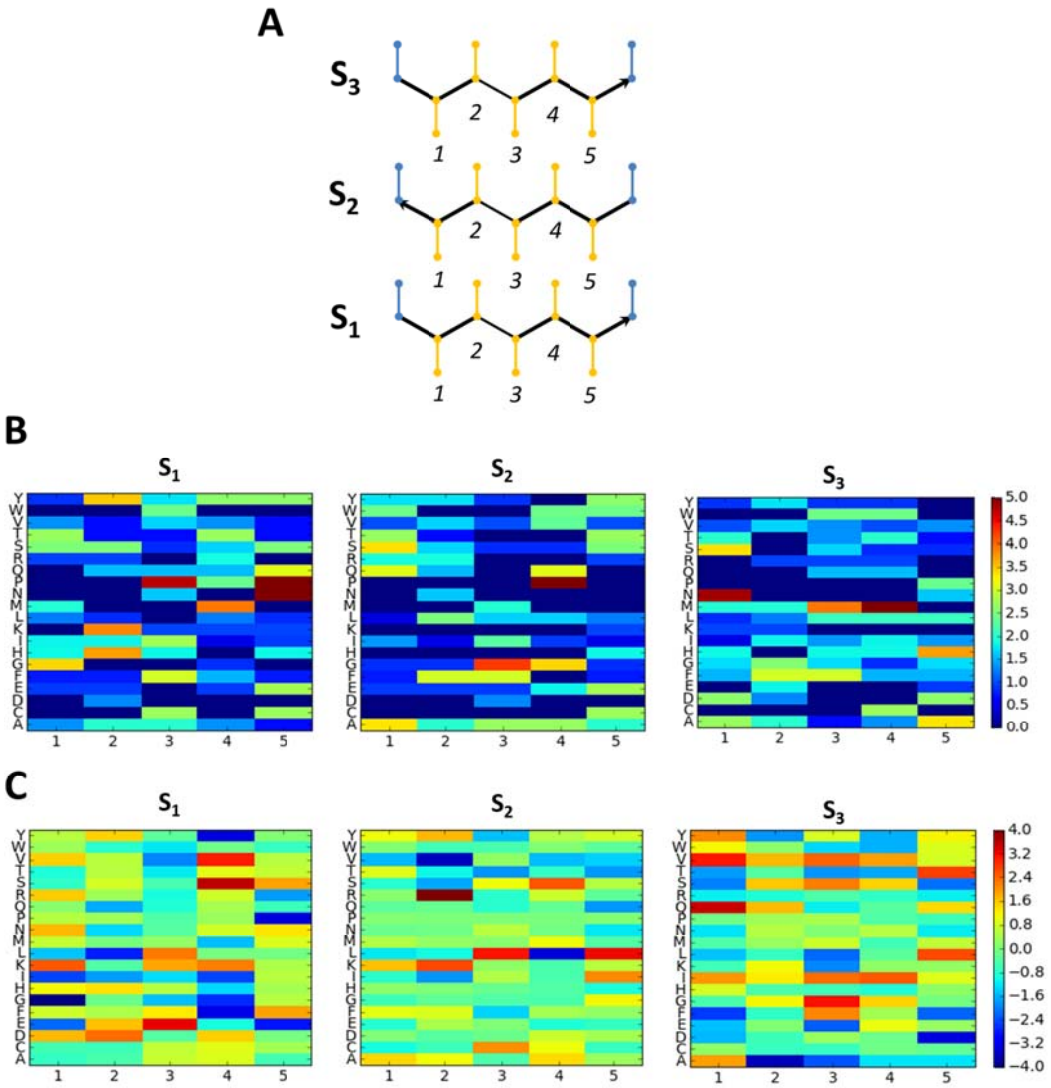

Fig. S5. Amino acid preferences in uniform β-sheets with different curvature. (A) Diagram of a uniform 3-stranded β-sheet; 5-residue strands segments used for calculating sequence profiles are in orange and labeled. (B) Inner strand with high positive bend angle (bend (S_1) > 0 and bend (S_2) ≥ 30°). Sequence profiles are colored by the frequency of each amino acid relative to the frequency to be found in strands. Central positions of the β-sheet favor G, F, P and M. (C) Positive *vs* negative bend angle. Sequence profiles are colored by the difference in amino acid frequency (%) between negative (-5° ≥ bend (S_1) ≥ -25°) and positive bend angle (25° ≥ bend (S_1) ≥ 5°). Amino acids favored and disfavored with negative bend angle are colored in red and blue respectively. For negative bend angle, β-branched amino acids with high strand propensity (V, I, T) are disfavored in the middle of edge strand segments (S_1), while L is strongly favored both in S_1 and S_2. In addition to the C_β-C_β steric interactions between paired residues (fig. S2), sidechain packing contributes to the stability of strand bend angle, therefore, leading to the complex sequence patterns observed.

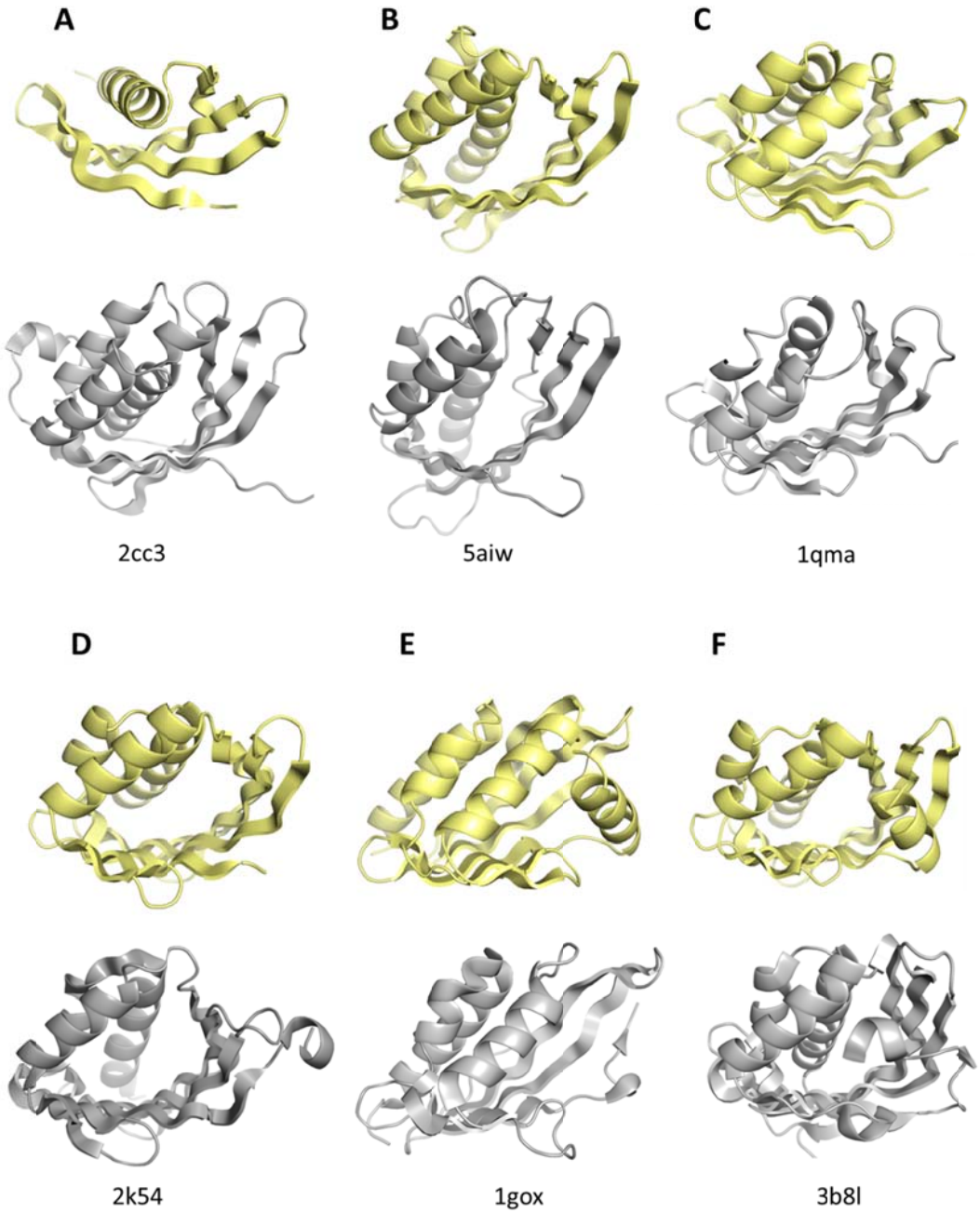

Fig. S6. Comparison of designed structures with native protein structures from the Protein Data Bank. A designed structure representative of each fold (A to F) is compared with the closest structural analog, as determined by a TM-align search (*31*, *61*). (A) TM-score 0.80, sequence id. 6.8%. (B) TM-score 0.78, sequence id. 9.3%. (C) TM-score 0.82, sequence id. 6.6%. (D) TM-score 0.86, sequence id. 19.2%. (E) TM-score 0.74, sequence id. 14.4%. (F) TM-score 0.79, sequence id. 6.9%. The top structural hits belong to the cystatin and NTF2-like superfamilies.

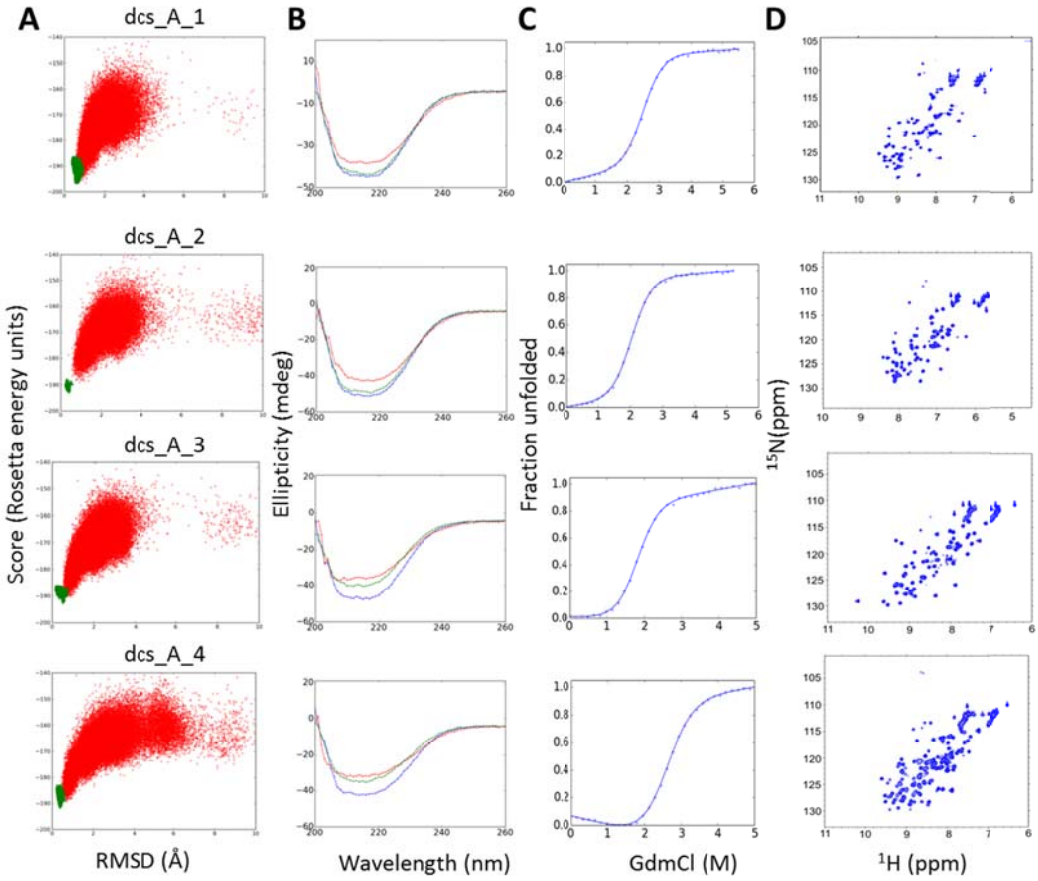

Fig. S7. Characterization of fold A designs. (A) Folding energy landscapes generated by *ab initio* structure prediction calculations. Each dot represents the lowest energy structure identified in an independent trajectory starting from an extended chain (red dots) or from the design model (green dots); *x*-axis shows the Cα-root mean squared deviation (RMSD) from the designed model; the *y*-axis shows the Rosetta all-atom energy. (B) Far-ultraviolet circular dichroism spectra (blue: 25 °C, red: 95 °C, green: 25 °C after cooling). (C) Chemical denaturation with GdmCl monitored with circular dichroism at 220 nm and 25 °C. (D) ^{1}H-^{15}N HSQC spectra obtained at 25 °C.

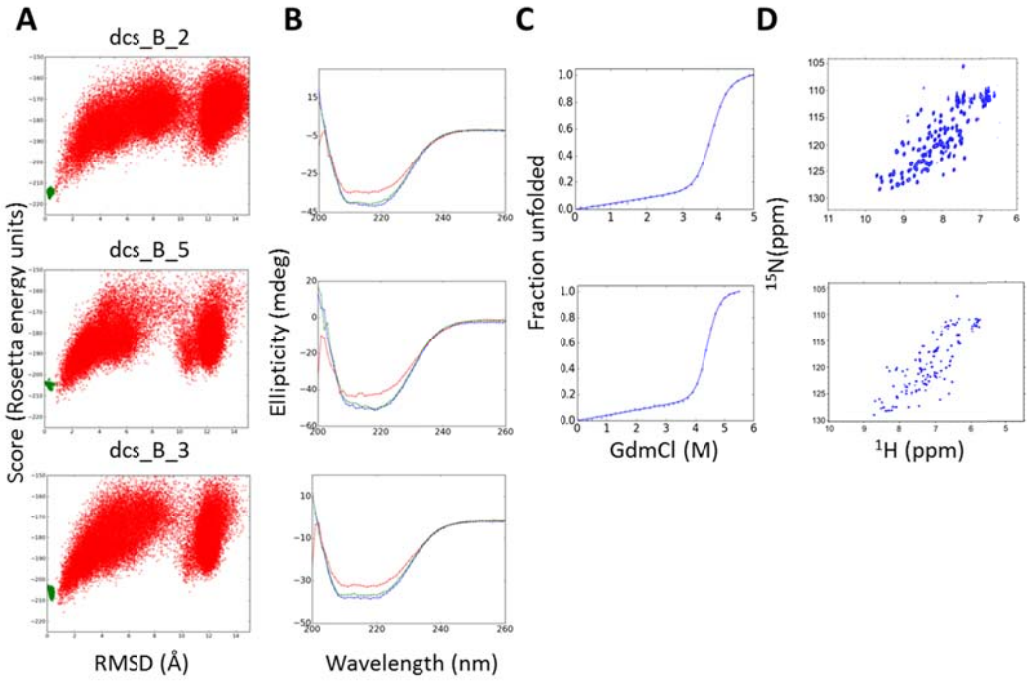

Fig. S8. Characterization of fold B designs. (A) Folding energy landscapes generated by *ab initio* structure prediction calculations. Each dot represents the lowest energy structure identified in an independent trajectory starting from an extended chain (red dots) or from the design model (green dots); *x*-axis shows the Cα-root mean squared deviation (RMSD) from the designed model; the *y*-axis shows the Rosetta all-atom energy. (B) Far-ultraviolet circular dichroism spectra (blue: 25 °C, red: 95 °C, green: 25 °C after cooling). (C) Chemical denaturation with GdmCl monitored with circular dichroism at 220 nm and 25 °C. (D) ^{1}H-^{15}N HSQC spectra obtained at 25 °C.

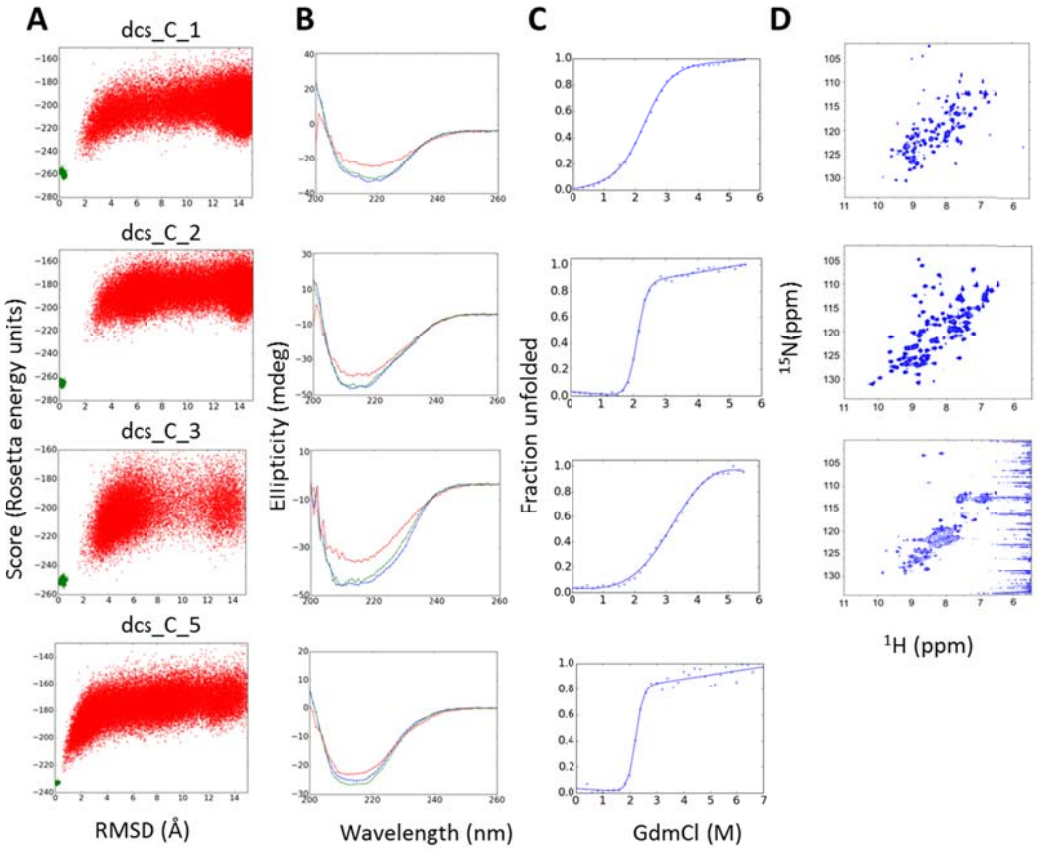

Fig. S9. Characterization of fold C designs. (A) Folding energy landscapes generated by *ab initio* structure prediction calculations. Each dot represents the lowest energy structure identified in an independent trajectory starting from an extended chain (red dots) or from the design model (green dots); *x*-axis shows the Cα-root mean squared deviation (RMSD) from the designed model; the *y*-axis shows the Rosetta all-atom energy. (B) Far-ultraviolet circular dichroism spectra (blue: 25 °C, red: 95 °C, green: 25 °C after cooling). (C) Chemical denaturation with GdmCl monitored with circular dichroism at 220 nm and 25 °C. (D) ^{1}H-^{15}N HSQC spectra obtained at 25 °C.

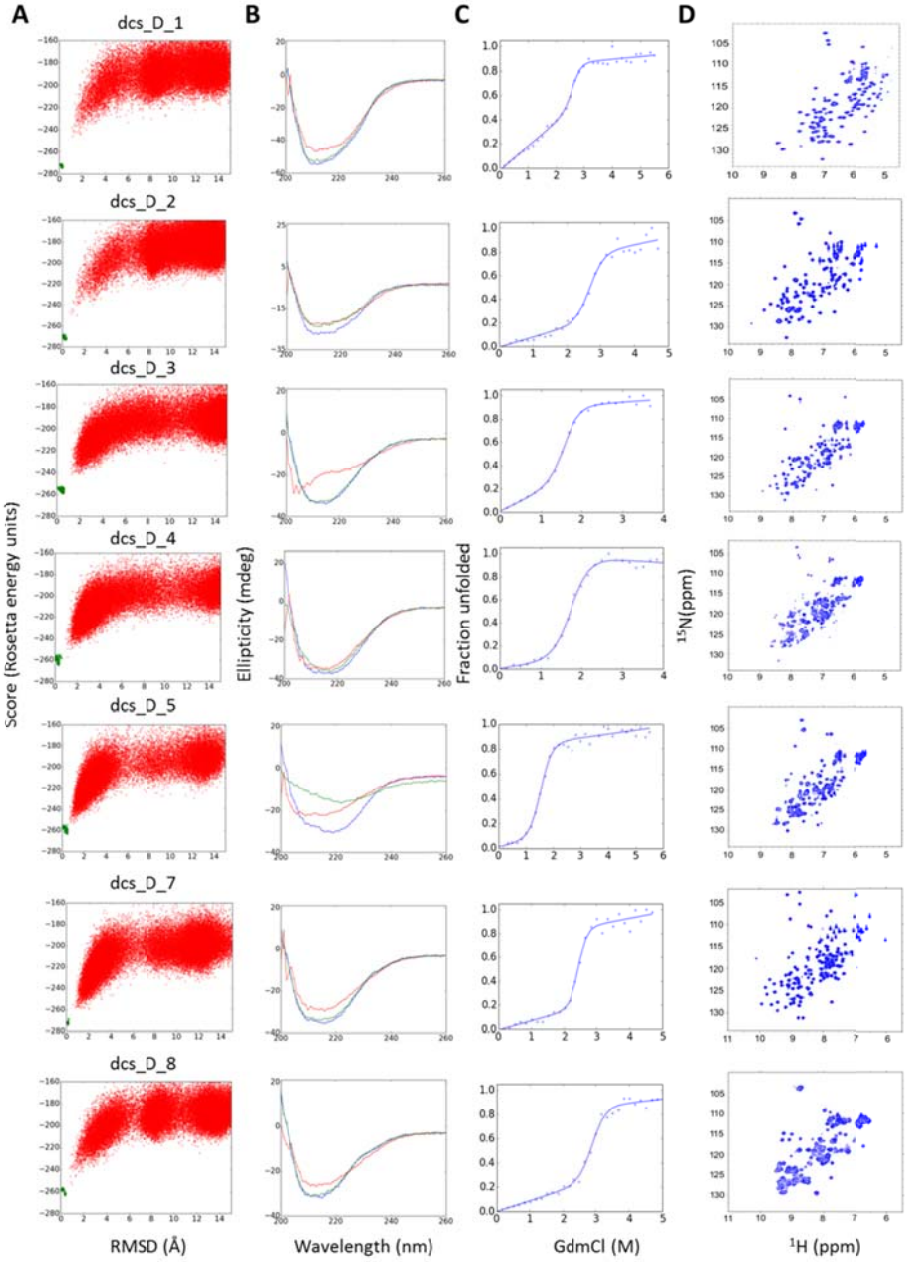

Fig. S10. Characterization of fold D designs. (A) Folding energy landscapes generated by *ab initio* structure prediction calculations. Each dot represents the lowest energy structure identified in an independent trajectory starting from an extended chain (red dots) or from the design model (green dots); *x*-axis shows the Cα-root mean squared deviation (RMSD) from the designed model; the *y*-axis shows the Rosetta all-atom energy. (B) Far-ultraviolet circular dichroism spectra (blue: 25 °C, red: 95 °C, green: 25 °C after cooling). These proteins are more sensitive to temperature than others from different folds due to the high solvent accessibility of the pocket. (C) Chemical denaturation with GdmCl monitored with circular dichroism at 220 nm and 25 °C. (D) ^{1}H-^{15}N HSQC spectra obtained at 25 °C.

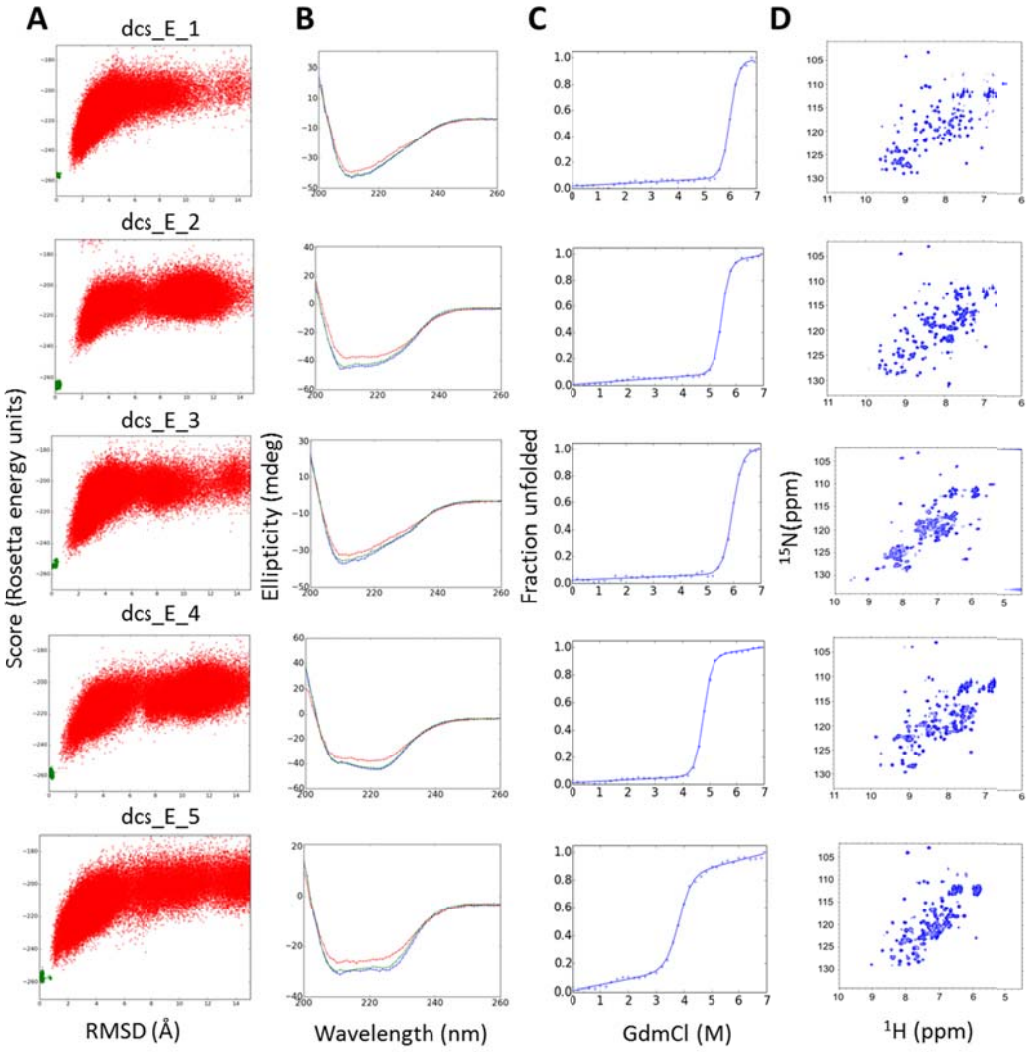

Fig. S11. Characterization of fold E designs. (A) Folding energy landscapes generated by *ab initio* structure prediction calculations. Each dot represents the lowest energy structure identified in an independent trajectory starting from an extended chain (red dots) or from the design model (green dots); *x*-axis shows the Cα-root mean squared deviation (RMSD) from the designed model; the *y*-axis shows the Rosetta all-atom energy. (B) Far-ultraviolet circular dichroism spectra (blue: 25 °C, red: 95 °C, green: 25 °C after cooling). (C) Chemical denaturation with GdmCl monitored with circular dichroism at 220 nm and 25 °C. (D) ^{1}H-^{15}N HSQC spectra obtained at 25 °C.

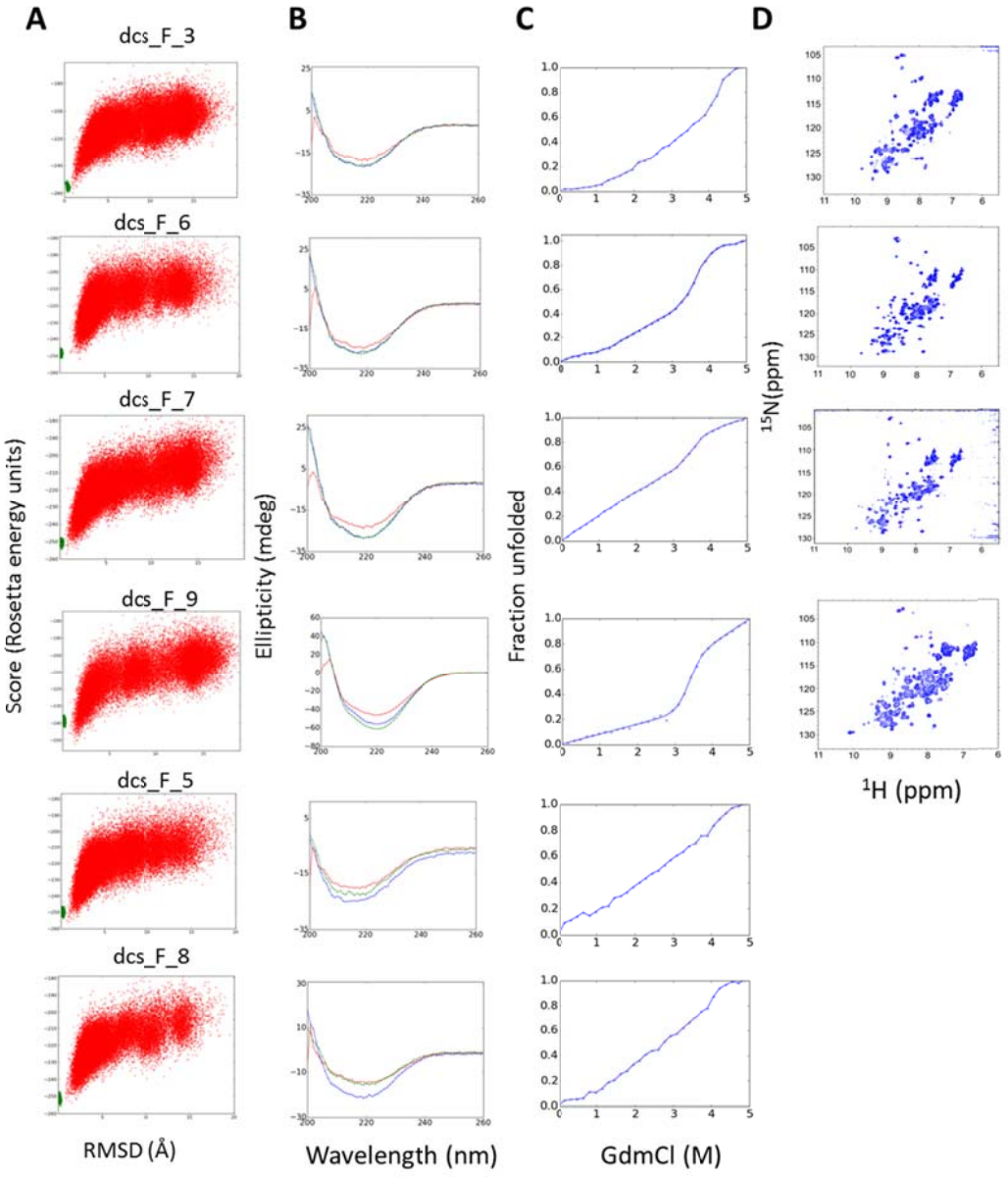

Fig. S12. Characterization of fold F designs. (A) Folding energy landscapes generated by *ab initio* structure prediction calculations. Each dot represents the lowest energy structure identified in an independent trajectory starting from an extended chain (red dots) or from the design model (green dots); *x*-axis shows the Cα-root mean squared deviation (RMSD) from the designed model; the *y*-axis shows the Rosetta all-atom energy. (B) Far-ultraviolet circular dichroism spectra (blue: 25 °C, red: 95 °C, green: 25 °C after cooling). (C) Chemical denaturation with GdmCl monitored with circular dichroism at 220 nm and 25 °C. The non-sigmoidal transitions suggest molten globule character for these proteins. (D) ^{1}H-^{15}N HSQC spectra obtained at 25 °C.

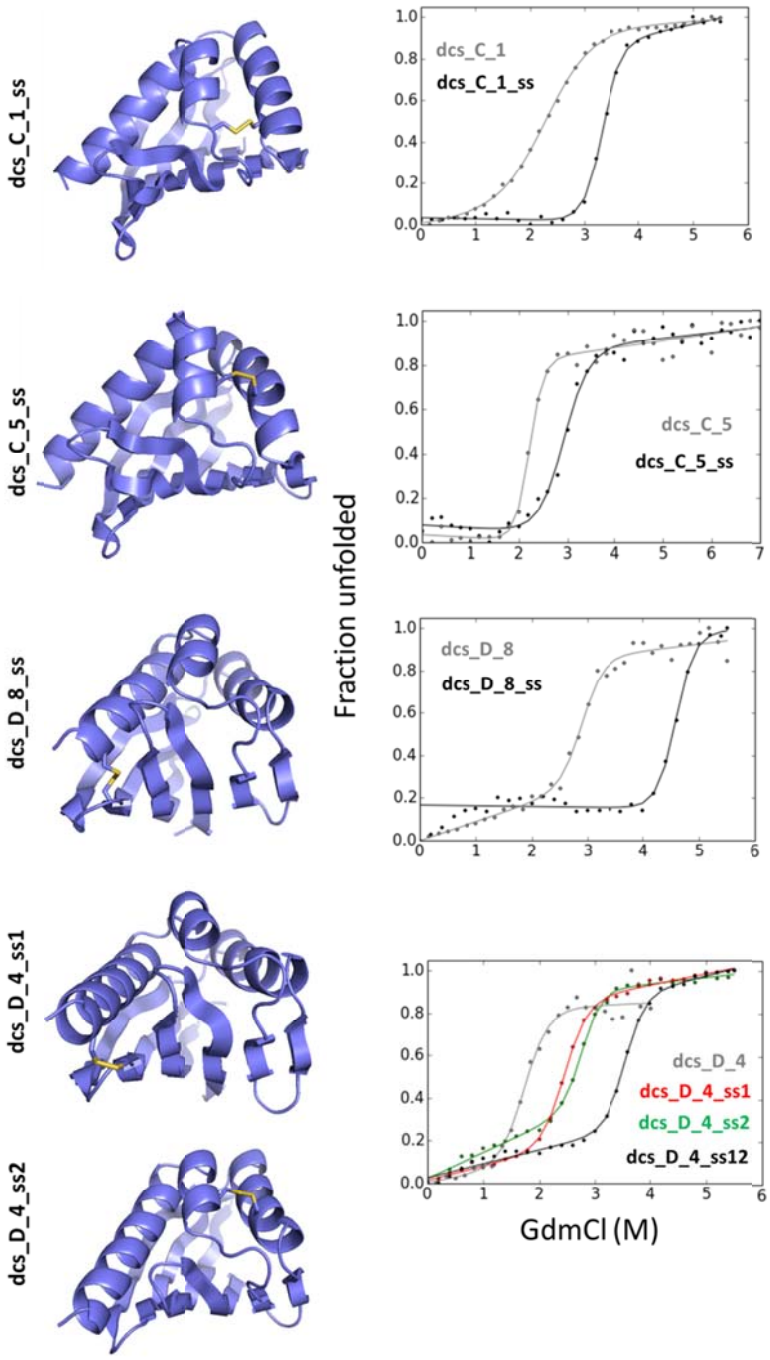

Fig. S13. Disulfides provide stability from the protein pocket periphery. (A) Designed models with disulfide bonds (yellow). Disulfides are located outside of the pocket area. (B) Chemical denaturation with GdmCl monitored with circular dichroism at 220 nm and 25 °C for each disulfide variant. For comparison, denaturation curves of the corresponding parent designs without disulfides are also shown. Design dcs_D_4_ss12 has the two disulfide bonds from dcs_D_4_ss1 and dcs_D_4_ss2 and exhibits cumulative stabilization from the two.

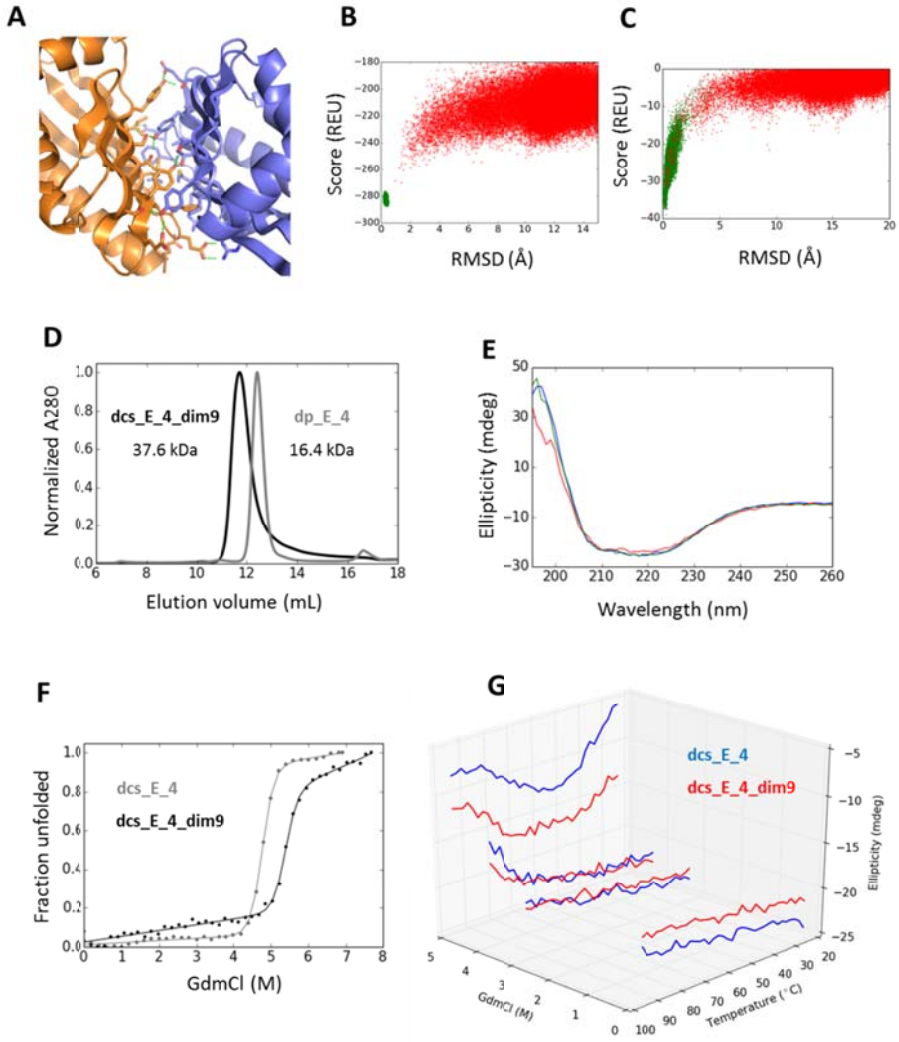

Fig. S14. Characterization of homodimer dcs_E_4_dim9 and comparison to its parent monomeric design dcs_E_4. (A) Designed interface interactions involving hydrogen bonding and aromatic stacking. Hydrogen bonds are highlighted in green dashed lines. (B) Folding energy landscape of the monomer subunit simulated with ab initio structure prediction. (C) Asymmetric docking simulations of dcs_E_4_dim9 predict stable formation of the designed homodimer interface. (D) Size-exclusion chromatograms monitoring UV absorbance at 280 nm. The shift in elution volume is consistent with dimer formation as assessed by multiple angle light scattering. (E) CD wavelength scans for dcs_E_4_dim9 at 25°C (blue), 95°C (red) and 25°C after cooling (green). (F) Chemical denaturation with GdmCl monitored with circular dichroism at 220 nm and 25°C. Continuous lines represent data fits to a two-state folding model. The higher C_m and lower folding free energy for dcs_E_4_dim9 indicate that the dimer interface provides additional stability ($\Delta\Delta G$ estimated in -1.4 kcal·mol^{-1}). (G) Chemical and thermal denaturation experiment on the monomer and dimer. At 4M GdmCl and ~90°C the monomer unfolds, whereas the dimer remains folded.

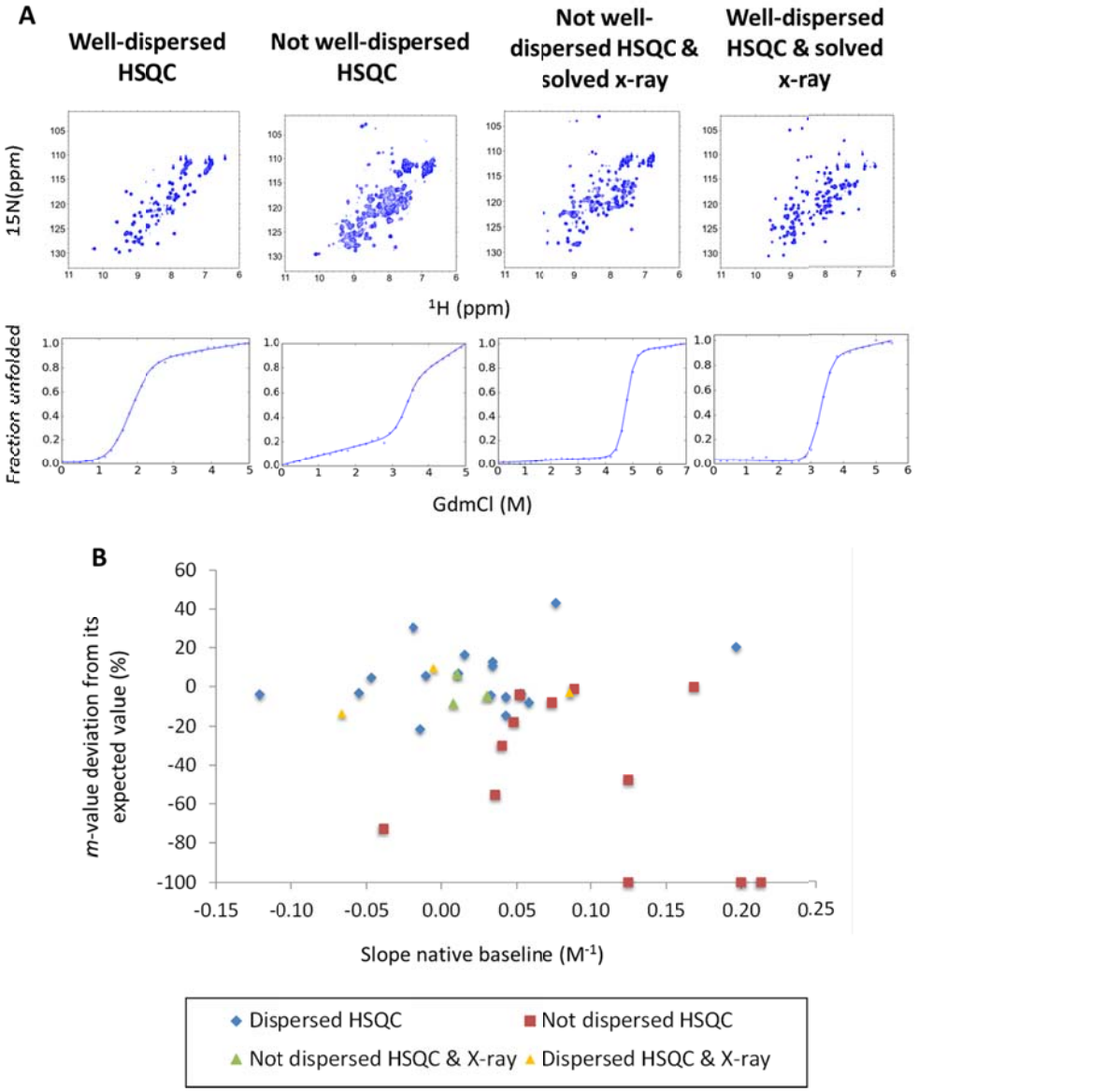

Fig. S15. Cooperative folding is determinant for structure elucidation. (A) Examples of designs differing in their folding cooperativity and HSQC quality. HSQCs are classified as well-dispersed, if have "good" or "excellent" scores, or not well-dispersed, "poor" or "promising" scores. Scores are given in Table S1. (B) The degree of folding cooperativity allows to distinguish those designs suitable for structure determination (well-dispersed HSQC spectra and/or crystallizable) from the rest. X-axis: slope of the native baseline obtained from the normalized unfolding transition curve. Y-axis: deviation of the fitted m-value with respect to its expected value based on protein size. This analysis includes 35 data points obtained from proteins characterized in this study (28 in total) and from other previously published *de novo* designed proteins (7 in total) (*12*, *14*).

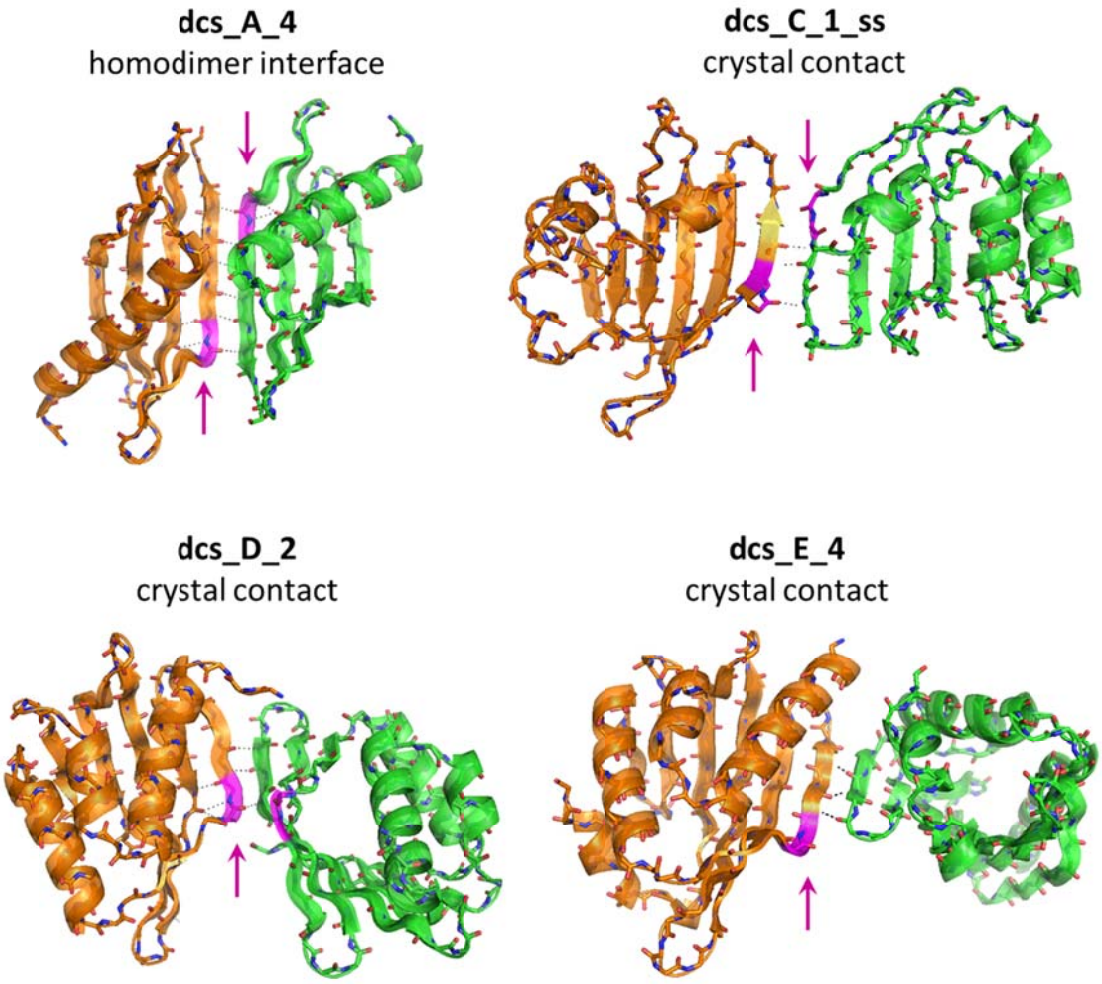

Fig. S16. Bulges limit strand pairing in solved crystal structures. The non-hydrogen bonding face of bulges (in magenta) restricts the hydrogen-bonded pairing between edge strands to regular segments, as observed in the homodimer interface of dcs_A_4 and in crystal contacts of dcs_C_1_ss, dcs_D_2 and dcs_E_4.

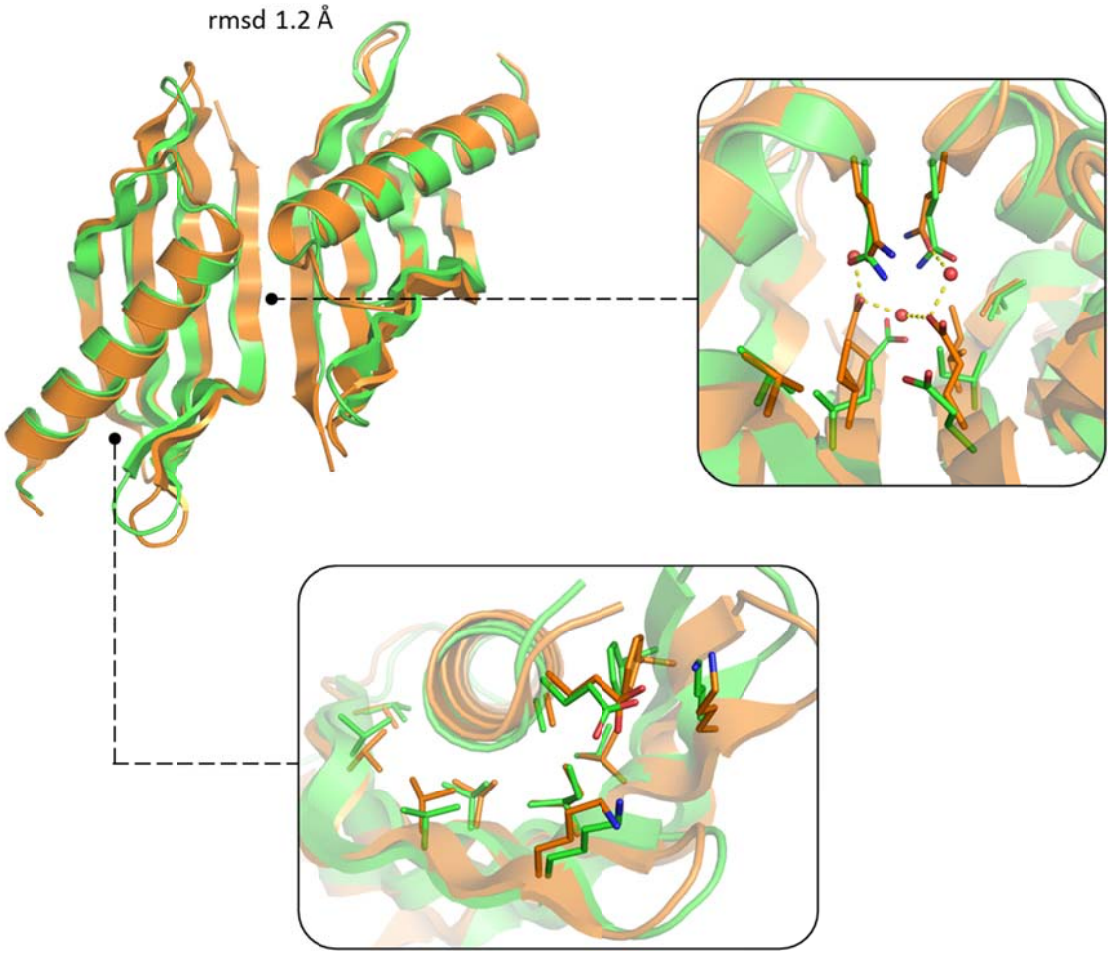

Fig. S17. Crystal structure of dcs_A_4 shows formation of a dimer interface. The monomeric design model is superimposed to each chain of the crystallized dimer for comparison. The experimental structure (2.4 Å resolution) and the design model are colored in orange and green, respectively; insets show comparisons of sidechain rotamers (right, homodimer interface; bottom, packing between β-sheet long arm and helix). Water mediated hydrogen bonds formed at the interface are shown in yellow dashed lines. The RMSD is calculated over all Cα atoms of each chain: rmsd(chain A) 1.22 Å and rmsd(chain B) 1.21 Å.

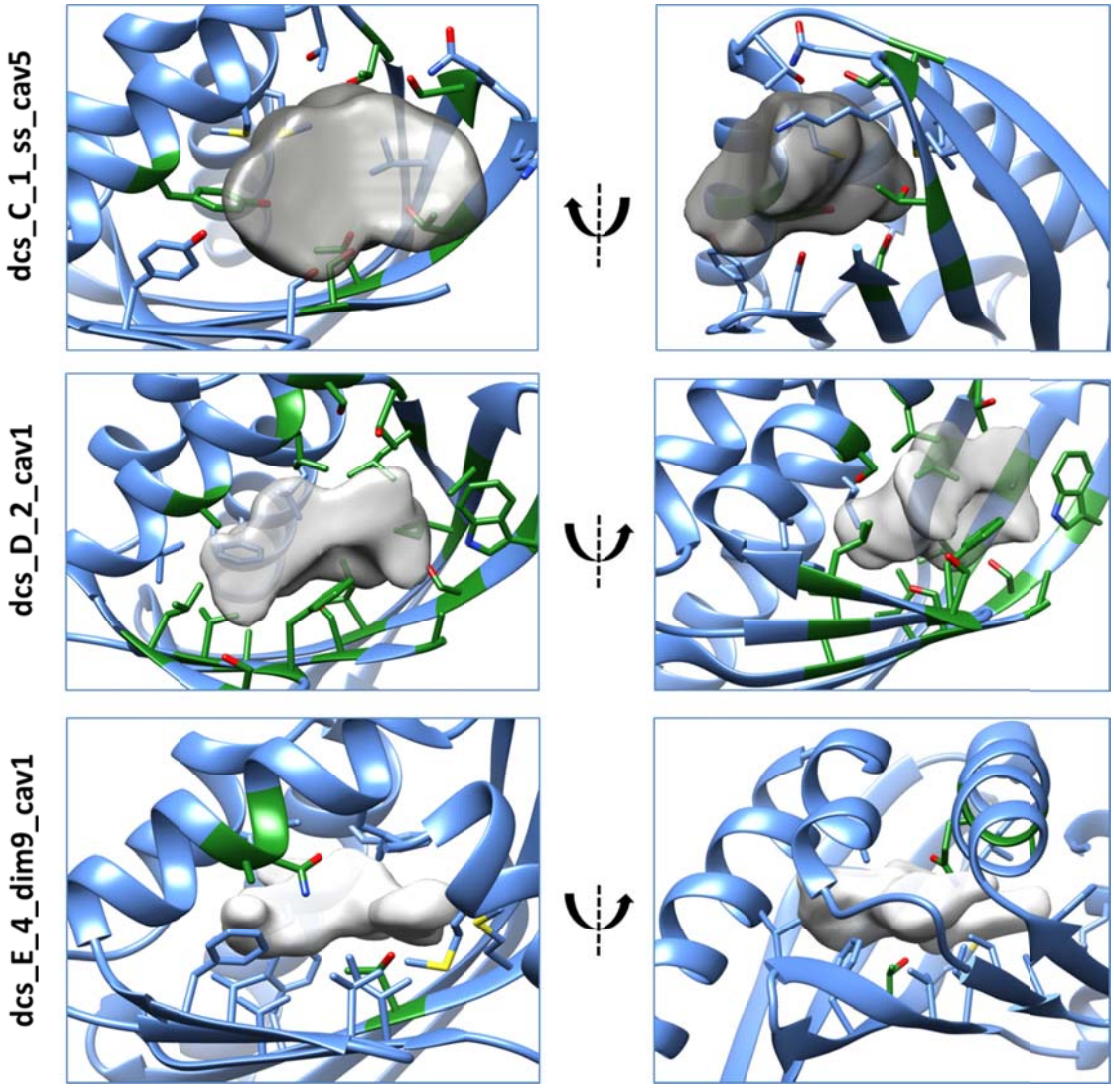

Fig. S18. Predicted cavities in stable cavity variants of dcs_C_1_ss, dcs_D_2 and dcs_E_4_dim9. Sidechains of residues lining the cavities are shown and the incorporated mutations are colored in green. Cavities were calculated with the 3V webserver (*62*) using different probe radii depending on the degree of cavity burial (outer probe radii from 4 to 6 Å and inner probe radii from 1 to 2 Å) and a grid size of 0.5 Å.

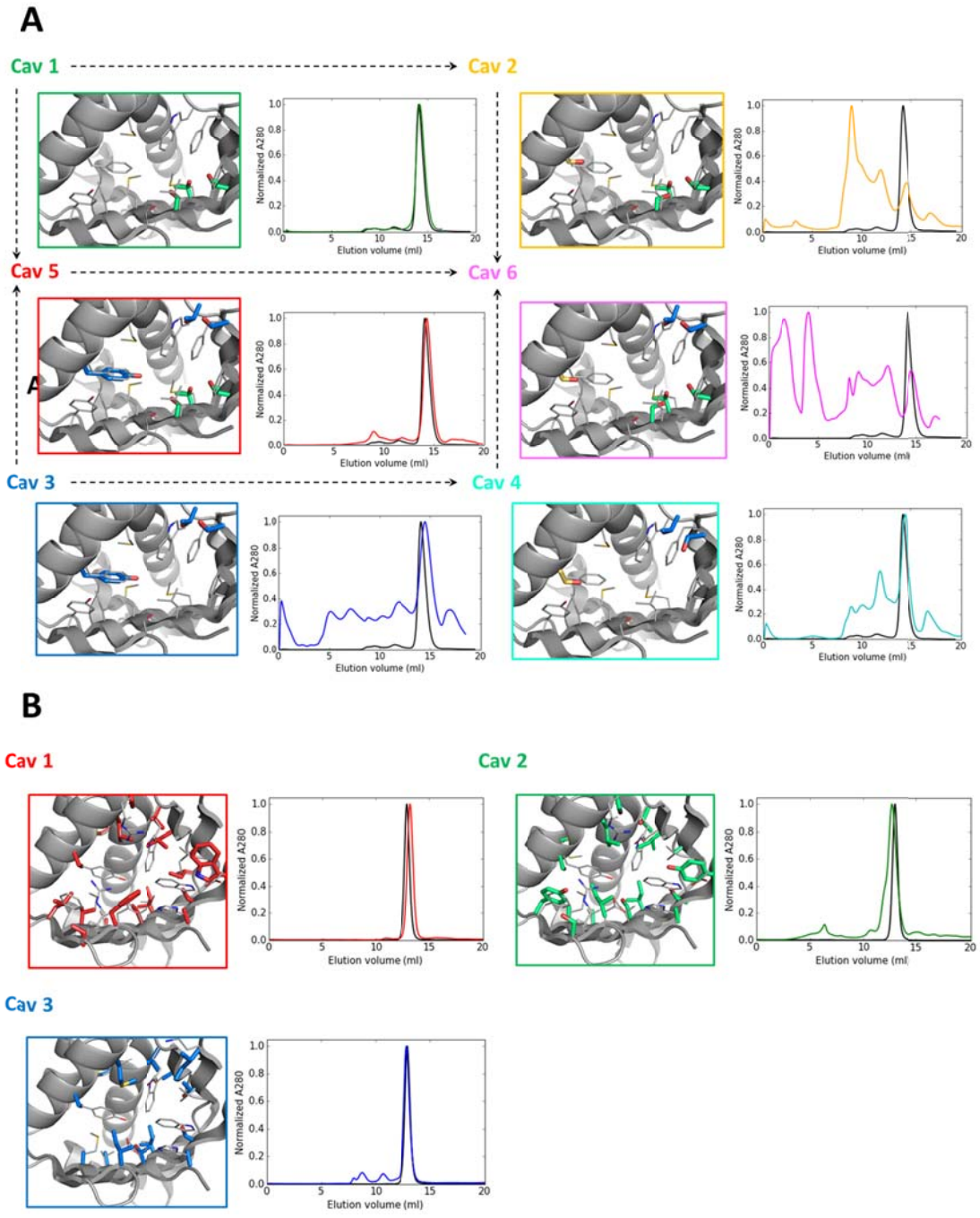

Fig. S19. The effect of cavity-creating mutations on the oligomeric state of dcs_C_1_ss (A) and dcs_D_2 (B). For each of the expressed mutants the left panel represents the amino acid substitutions overlaid with the corresponding parent design (in gray). The size-exclusion chromatogram of each mutant is compared with that from the parent design (in black) on the right panels. (A) Mutants result from combinations of three groups of mutations, whose sidechains are displayed in different colors: 84T, 86T and 104S (group 1, in green); 47Y, 57T and 79S (group 2, blue); 47S (group 3, in

yellow). Arrows connect the closest mutants. Some combinations of groups of mutations are found to be well tolerated (cav 5, groups1+2), while others form higher molecular weight species (cav 2 and cav6). (B) These three mutants of dcs_D_2 are purely monomeric and are the ones that accumulate the highest number of mutations.

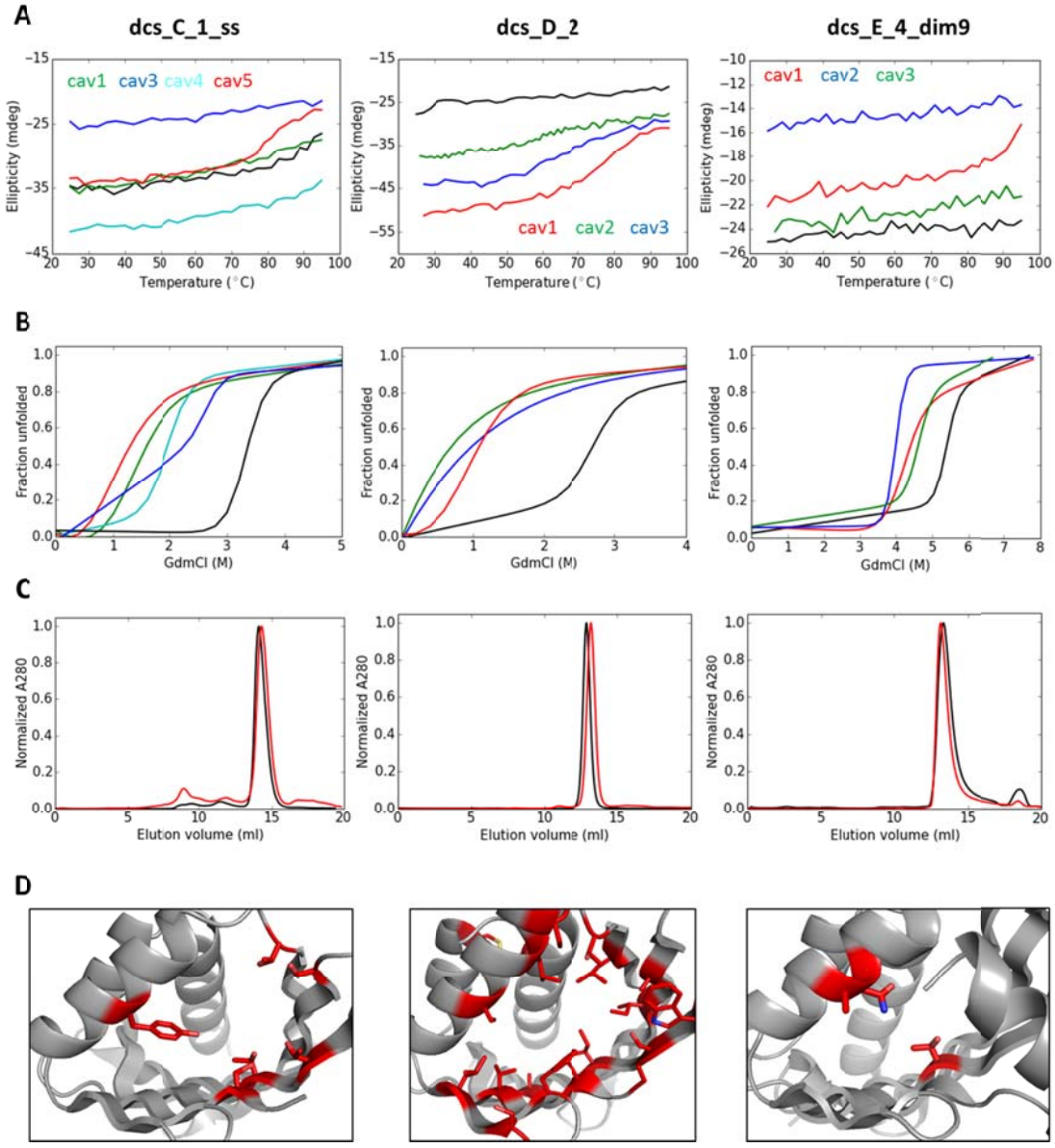

Fig. S20. Characterization of cavity mutations in dcs_C_1_ss, dcs_D_2 and dcs_E_4_dim9. (A) Temperature melting monitored with circular dichroism for the best expressed cavity-mutants and the corresponding parent designs (in black). Same colors identifying each mutant are used in the other panels. (B) Chemical denaturation with GdmCl monitored with circular dichroism at 220 nm and 25°C. Continuous lines represent data fits to a two-state folding model. (C) Size-exclusion chromatograms of a cavity mutant and its parent design (in black). (D) Design model representations coloring the incorporated mutations of each mutant.

.

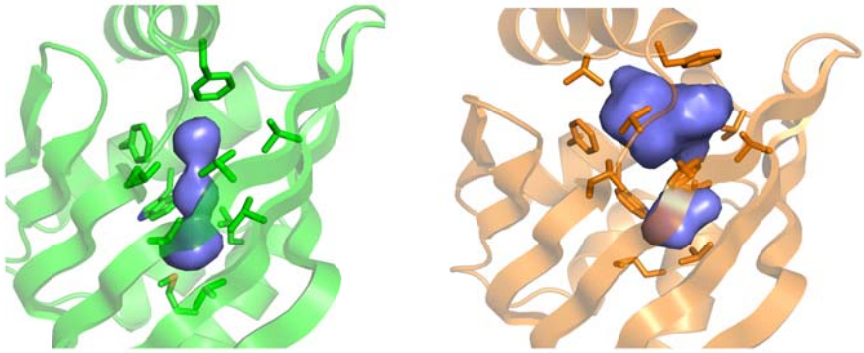

Fig. S21. Buried cavities in dcs_E_3 design. Both the design (left) and the crystal structure (right) have an internal hydrophobic cavity with a volume of 84 and 192 $Å^3$, respectively. The cavity expands in the crystal structure due to a slight reorientation of the C-terminal helix. Cavity volumes were calculated with 3V webserver (*62*) using an outer probe radius of 2.5 Å, an inner probe radius of 1.0 Å and a grid size of 0.5 Å.

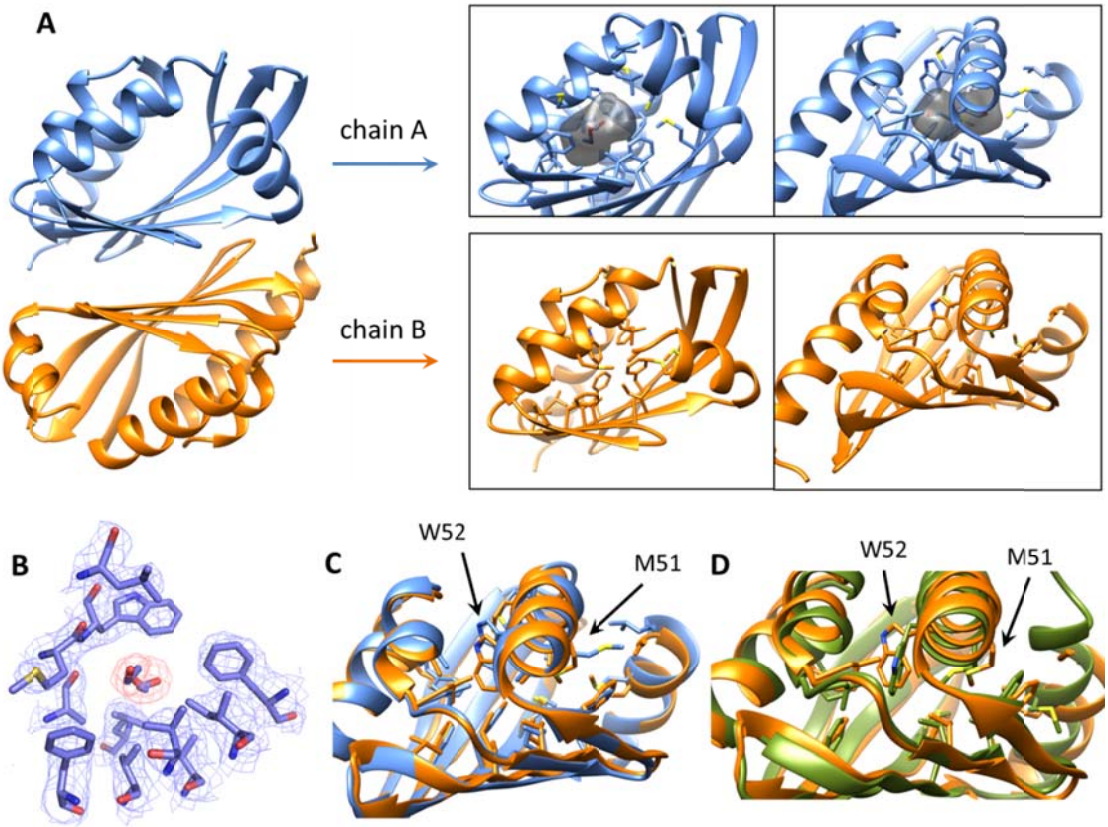

Fig. S22. Reorganization of two sidechains results in a cavity in the crystal structure of dcs_E_4_dim9. (A) Reorganization of M51 and W52 sidechains in chain A results in a cavity occupied by a diethylene glycol molecule from the crystallization solution. In chain B, the two sidechains have a different conformation that occludes the cavity by providing tighter hydrophobic packing. The cavity formed in chain A was calculated with 3V webserver (*62*) using an outer probe radius of 2.5 Å, an inner probe radius of 1.5 Å and a grid size of 0.5 Å. The calculated cavity volume is 191 $Å^3$. We hypothesized mutations at these positions could result in cavities with different shapes and sizes. (B) The internal cavity formed in chain A is occupied by a diethylene glycol molecule as shown by the electron density. (C) Superimposition of both chains highlights the differences in M51 and W52 sidechain conformations. (D) Superimposition of chain B from the crystal structure (orange) with one of the two symmetric chains of the design model (green) shows that the designed conformation of M51 closely matches that from the chain B crystal structure, which provides better hydrophobic packing.

Table S1. Summary of the experimental characterization of designs.

Design name	Expressed	Soluble	CD spectra (25 °C)	T_m(°C)	Two-state GdmCl unfolding §	Oligomeric state†	HSQC quality*
dcs_A_1	Y	Y	αβ	> 95°C	Y	M	2
dcs_A_2	Y	Y	αβ	> 95°C	Y	M	2
dcs_A_3	Y	Y	αβ	> 95°C	Y	M	2
dcs_A_4	Y	Y	αβ	> 95°C	Y	M	2
dcs_B_1	N						
dcs_B_2	Y	Y	αβ	> 95°C	Y	M	2
dcs_B_3	Low	Y	αβ	> 95°C	N/A	N/A	N/A
dcs_B_4	N						
dcs_B_5	Y	Y	αβ	> 95°C	Y	M	2
dcs_C_1	Y	Y	αβ	> 95°C	Y	M	3
dcs_C_2	Y	Y	αβ	> 95°C	Y	M	2
dcs_C_3	Low	Y	αβ	> 95°C	Y	M	4
dcs_C_4	N						
dcs_C_5	Y	Y	αβ	> 95°C	Y	M	N/A
dcs_D_1	Y	Y	αβ	> 95°C	Y	M	2
dcs_D_2	Y	Y	αβ	> 95°C	Y	M	2
dcs_D_3	Y	Y	αβ	~85°C	Y	M	3
dcs_D_4	Y	Y	αβ	> 95°C	Y	M	4
dcs_D_5	Y	Y	αβ	> 95°C	Y	M	3
dcs_D_6	Y	Y	αβ	> 95°C	Y	M	3
dcs_D_7	Y	Y	αβ	> 95°C	Y	M	1
dcs_D_8	Y	Y	αβ	> 95°C	Y	M	3
dcs_D_9	N						
dcs_E_1	Y	Y	αβ	> 95°C	Y	M	2
dcs_E_2	Y	Y	αβ	> 95°C	Y	M	2
dcs_E_3	Y	Y	αβ	> 95°C	Y	M	3
dcs_E_4	Y	Y	αβ	> 95°C	Y	M	3
dcs_E_5	Y	Y	αβ	> 95°C	Y	M	3
dcs_F_1	Y	Y	αβ	> 95°C	N/A	M	N/A
dcs_F_2	Y	Y	αβ	> 95°C	N/A	M	N/A
dcs_F_3	Y	Y	αβ	> 95°C	N	M	3
dcs_F_4	Y	Y	αβ	> 95°C	N/A	M	N/A
dcs_F_5	Y	Y	αβ	> 95°C	N	M	N/A
dcs_F_6	Y	Y	αβ	> 95°C	N	M	3
dcs_F_7	Y	Y	αβ	> 95°C	N	M	3
dcs_F_8	Y	Y	αβ	90°C	N	M	N/A
dcs_F_9	Y	Y	αβ	> 95°C	Y	M	3
Disulfide variants							
dcs_C_1_ss	Y	Y	αβ	> 95°C	Y	M	1
dcs_C_2_ss	Y	Y	αβ	> 95°C	Y	M	N/A

dcs_C_4_ss	N						
dcs_C_5_ss	Y	Y	αβ	> 95°C	Y	M	N/A
dcs_D_4_ss1	Y	Y	αβ	> 95°C	Y	M	N/A
dcs_D_4_ss2	Y	Y	αβ	> 95°C	Y	M	N/A
dcs_D_4_ss12	Y	Y	αβ	> 95°C	Y	M	N/A
dcs_D_8_ss	Y	Y	αβ	> 95°C	Y	M	N/A
Homodimers							
dcs_E_4_dim1	Y	Y	N/A	N/A	N/A	D	N/A
dcs_E_4_dim2	Y	Y	N/A	N/A	N/A	M/D‡	N/A
dcs_E_4_dim3	Y	Y	N/A	N/A	N/A	M/D‡	N/A
dcs_E_4_dim4	Y	Y	N/A	N/A	N/A	M/D‡	N/A
dcs_E_4_dim5	Y	Y	N/A	N/A	N/A	D	N/A
dcs_E_4_dim6	N						N/A
dcs_E_4_dim7	N						N/A
dcs_E_4_dim8	N						N/A
dcs_E_4_dim9	Y	Y	αβ	> 95°C	Y	D	4
Cavity mutants							
dcs_C_1_ss_cav1	Y	Y	αβ	~85°C	Y	M	N/A
dcs_C_1_ss_cav2	Y	Y	αβ	N/A	N/A	A	N/A
dcs_C_1_ss_cav3	Y	Y	αβ	> 95°C	Y	M	N/A
dcs_C_1_ss_cav4	Y	Y	αβ	> 95°C	Y	M	N/A
dcs_C_1_ss_cav5	Y	Y	αβ	> 95°C	Y	M	N/A
dcs_C_1_ss_cav6	Y	Y	αβ	N/A	N/A	A	N/A
dcs_D_2_cav1	Y	Y	αβ	75°C	Y	M	N/A
dcs_D_2_cav2	Y	Y	αβ	65°C	N	M	N/A
dcs_D_2_cav3	Y	Y	αβ	65°C	N	M	N/A
dcs_E_4_dim9_cav1	Y	Y	αβ	> 95°C	Y	D	N/A
dcs_E_4_dim9_cav2	Y	Y	αβ	> 95°C	Y	M	N/A
dcs_E_4_dim9_cav3	Y	Y	αβ	> 95°C	Y	D	N/A

§ The denaturation curve was sigmoidal and could be fitted to a two-state folding mechanism.

† Oligomeric state of the dominant species based on SEC-MALS (M, monomer ; D, dimer). A denotes dominant aggregate species

‡ The error in the molecular weight estimate is too high to determine whether the main peak corresponds to a monomer or dimer species.

* HSQC quality was ranked from 1 to 4 based on the peak dispersion and intensity (*63*): 1, excellent; 2, good; 3, promising; 4, poor.

Table S2. Designed protein sequences. The lowest E-value obtained from BLAST (*29*, *30*) searches (against the NCBI nr database of non-redundant protein sequences) is shown.

Design name	Amino acid sequence	E-value
dcs_A_1	KSDELQKRVVEYAKEVILRQKGDPTLDIQVKR VQTTGNTLRVELEIRTGNTTRQYQIEVEIRGDT FQVRRVQETGGS	>10
dcs_A_2	KDDELQKRVVEYAKEVLLRQKGDPTTDIQVKR VQTTGNTVRVELELRVGNETTQMQIEVEIQGD TFQVRRVQKTGGS	>10
dcs_A_3	PSEEEEKRQVKQVAKEKLLEQSPNSKVQVRRV QKQGNTIRVELELRTNGKKENYTVEVERQGNT WTVKRITRTVGS	>10
dcs_A_4	PSEEEEKRRAKQVAKEKILEQNPSSKVQVRRV QKQGNTIRVELEITENGKKTNITVEVEKQGNTF TVKRITETVGS	5.4
dcs_B_1	QDIVEAAKQAAIAIFQLWKNPTDPKAQKLLKKI LSPDLLKQMEKHARKLQKQGIHFEVKRVEVEK TGNTVQVTVEIEKTTGGTRQRRTYQMRFEVDG DTIRRVTVTEVGS	>10
dcs_B_2	QDIVEAAKQAAIAIFQLWKNPTDPEAQELLNKI LSPDVLDQVREHARELQKQGIHFEVKRVEVTT DGNTVNVTVELEETTGGTTTNTTYELRFEVDG DTIRRVTVTQNGS	0.81
dcs_B_3	QDIVEAAKQAAIAYFQLLKNPTDPEAQNLLNKI LSPDVLDQVKEHAKKLQKQGIHFEVKRVEVET TGNTVKVKVELEKETGGTRQRKRYTLRFEVDG DTIKRVTTTQTGSWS	2.3
dcs_B_4	QDIVEAAKQAVIAYFQLLKNPTDPDAQNLLRKI LSPDLLEQIKRHARQLQKQGIHFEVKRVEVETT GNTVKVTVEIEKKTGGTRTRKRYKLRFEVDGD TIKRVTVTQTGSWS	0.75
dcs_B_5	QDIVEAAKQAAIAYFQLLKNPTDPDAQNLLRKI LSPDVLEQIKRHARQLQKQGIHFEVKRVEVTTT GNTVQVTVEIEETTGGTTTQTTYKLRFEVDGD TIKRVTVTQTGSWS	>10
dcs_C_1	SEEAKIAIELFKEAMKDPERFKEMVSPDTRIESN GQEYRGSEEAKKFAEEMKKTHPWEVRVERYR SDGDRFEIELRVNFNGKTFRMEIRMRKVNGEF RIEEMRLHG	0.72
dcs_C_2	QPDEVKKIAQEWWERMMRNPRQIEELIDPNTR LRDGNTELTGREVQEYMKEWVTKVRFEVKEV TKEGNVYRVRLKVEENGKTKEMEIRLEDDNG RMRFKEIEIRG	>10
dcs_C_3	DKEEAKKLAELIERAYRNPDVAREVFSPNTRFE	0.43

	DNGRETHDVEEWMEEIKRQGRPVEVRVKEITR DGNEMRIRLRIRYNGEEYEMEIRFRHEDGQWK IEEMRWRG	
dcs_C_4	DDIEKMMKKFVQWMRDGNPEYVERMVSPNT KFRHNGQETKGSDIVREWMKKLLNMRVEVKR YRIKNGELELEIEFETGDRTSTVTFRLRLENGQ MHLEEMEFRN	1.0
dcs_C_5	SEDDVRREVQRVWEEIRNNPEALREYVDPNTH LHDGNQQYSGEEVQEYMRELVTRVEFRVRRV EKKGNTWKVEVEVRENGQEKEMHIEFEEDNG KFKFKRIEIRG	1.1
dcs_D_1	PEEEKMARLFIEAVEKGDPELMRKVISPDTRVE DNGREFTGDEVSEWVKEIQKRGEQWHLRRYT KEGNSWRFELQVDNNGQTEQWEVQIEVRNGRI KRVTVTHV	0.00002
dcs_D_2	PEEEKAARLFIEALEKGDPELMRKVISPDTRME DNGREFTGDEVVEYVKEIQKRGEQWHLRRYT KEGNSWRFEVQVDNNGQTEQWEVQIEVRNGR IKRVTITHV	0.00002
dcs_D_3	SPEKEESKLVEEFMKLMEQGDPEEMLKLISPDT RLEKDGEEYNGEEVRQYWEKEMREGTKFQVR EVTTQGNKVRIRVQVQQNGTTTQEQYEVEMR DGRIRRITVHTRG	0.026
dcs_D_4	SPEKEESKLVEEFMKLMEQGDPEEMKKLISPDT RLERDGEEYNGEEVRQFWEEEMRQGLKFQVR EVTTQGNKVRIRVQVQKNGTTTQVQFEVEMR DGRIRRITVHERG	0.003
dcs_D_5	SEEESKVAQEMMKMISKGDPDEIRKHMSPDTR VDFNGEEYSGEEVARMWEKERRKGRQYEVKR YQSKGNEVQFELEVQDNGKTETIQIRVRVENG RVKEVQITTH	>10
dcs_D_6	SEEEEKVAQEMMKAIQKGDPDEIRKYLSPDVR VKVNGEEYSGEEVVRYWEKERRKGRRWEVK RYQTDGNEVQFELQVEDNGKTEQYEIRVRVEN GRVKEIQITTH	0.087
dcs_D_7	SEEEERVAKEMMEAIQKGDPDEIRKYLSPDVR VKVNGEEYSGEEVVRYWEKEKRKGRRWEVK RYQTKGNEVQFELQVEDNGKTEQWEIRVRVE NGRVKEIQITQH	0.003
dcs_D_8	PEVVKVWKRIMEALQKGDPELLKKMISPDTRM EVNGQTFTGEEVVRYWEEEIRRGRQWTVKRY TEKGNEVEFEVEQQDGDETRTYRVQVRVRNG QVEEIQVTQV	0.53
dcs_D_9	SEHEKHARQIEKAWKKGNPEELKKVVSPDTR MDFNGEEYRGKERIEEMMRRKKERGVEITLER VQHKGNELQLRVQFTEGNQTKQYEFRFEFENG	0.022

	QVRRVEVREN	
dcs_E_1	SREEIRKVVEEMLRSLKQGSPEDISKYLSPDVR LEVGNYTFEGSEQVTKFWRMWTKFVDRVEVR KVQVDGNHVRVEMEVEWNGKRWTFEMEVEV RNGKIKRIRLQVDPEFKKVVQNIWNLL	0.007
dcs_E_2	TKDEVKKMVEILKKAFEEGDPEKIVSLLSPNVR LEMGNYTWEGSEQVEEFLRYLMEIVDRVEVRR IKVRPNHIEVEVEMEFNGKSFEVEWRFEIENGK VRRVEVRVTPEMKKIVEKVYRKA	0.23
dcs_E_3	SREEIRKVVEEMVRKLKQGSPEDISKYLSPDVR LEVGNYTFEGSEQVTKFWRMLTKFVDRVEVR KVQVDGNHVRVEVEVEWNGKKWTFEVEVEV RNGKIKRIRLQVDPEFKKVVQNIWNLL	0.002
dcs_E_4	TQEEVRKIMEKLKKAFKQGNPEQIVSLLSPDVR VKVGNQEFSGSEEAEKMWRKLMKFVDRVEVR RVKVDENRVEIEVEFEVNGQRYSMEFHFEVEN GKVRRVEIRISPTMKKLMKQILNYG	1.1
dcs_E_5	TKKEVEKMARTFKEAMNQGNPEQLTSKLSPD VRLRIGNQEFEGSEEVEKWLRRWFNLVDRVEV RRIKVEDNHVEVEVEVELNGKNVEIEFRFEIRN GKVERMEIRVTPDMKKFAEKINKYG	0.001
dcs_F_1	DENEKMKKMVRQFLELIEKEDPDEIRKLLSPDT RVTFNGRTFTGPEEFAKELQELRKQGIRFQFTE AEIQTDNGKLQIRVEVTLTVNGQEYRSEVTFTI RVENGVIKEVTIQFSPKLQEALKGGS	0.48
dcs_F_2	DENEKMKEAVRQFLELIEKEDPDEIRKLLDPNT RVTFNGKTFTGPEEFAKELQELRKQGIRFQFTV KEIQTDNGKLQIRVEVTLTVNGQEYRSEVTFTI RVENGVIKEVTIQFSPKLQEALKGGS	0.52
dcs_F_3	DENEKMKEMVREFLEIIEKRDPNEIRKLLDPNT RVTFDGRTYTGPEEFAKELQELEKQGIEFQFTIK EIQTDNGVLQIRVEVTLTVNGQEYRSEVTFTIR VENGVIKEVTIQFSPKLQEALKGGS	0.82
dcs_F_4	DENEKMKEMVREFLELIEKRDPEEMRKLLSPD TRVTFDGKTFTGPEEFAKELQELEKQGIEMQYT VKEIQTDNGVLQIRVEVTLTVNGQEYRSEVTFT IRVENGTIKEVTIQYSPKLQEALKGGS	0.18
dcs_F_5	DEDEKMKEIVKQFLELIKREDPEELRKLLSPDT RVTFNGRTYTGPEEFAKELQEMRKRGVRFQFTI KEVRTVNGVMKIRFEVQVTVNGVTYRSEVTIQ IRVENGVIKEVTIQFSPKLQEAIEGGS	0.067
dcs_F_6	DENEKMKEIVKQFLELIKREDPEELRKLLSPDT RVTFDGRTFTGPEEFAKELQEMRKRGVRFQFT EAEVQTDNGKLKIRFEVQVTVNGQTYRSEVTI QIRVENGVIKEVTIQFSPKLQEAIEGGS	0.007
dcs_F_7	DEDEKMKEIVKQFLELIKRRDPEELRKLLDPNT	6.5

	RVTFNGKTFTGPEEFAKELQELEKRGVEMQYTI KEVQTDNGKMKIRFEVQVTVNGQTYRSEVTIQ IRVENGVIKEVTIQYSPKLQEALEGGS	
dcs_F_8	DEDEKMKEIVKQFLELMKRRDPEEMRKLLDPN TRVTFNGKTFTGPEEFAKELQEMEKRGVEFQF TIKEVRTVNGVMKIRFEVQVTVNGVTYRSEVTI QIRVENGVIKEVTIQFSPKLQEAIEGGS	1.4
dcs_F_9	DPAEQAREIVRQFLELIQRRDPEELRRLLSPDTR VTFNGRTFTGPERFAEALQELERRGVEMQYTIQ EVQTENGRMSIRFEVQVTVNGQTYRSEVTIQIR VENGRIREVTIQYSPRLQEALEGGSGW	0.02
Disulfide variants		
dcs_C_1_ss	SEEAKIAIELFKEAMKDPERFKEMCSPDTRIESN GQEYRGSEECKKFAEEMKKTHPWEVRVERYR SDGDRFEIELRVNFNGKTFRMEIRMRKVNGEF RIEEMRLHG	0.84
dcs_C_2_ss	QPDEVKKIAQEWWERMMRNPRQIEELIDPNTR CRDGNTELTGRECQEYMKEWVTKVRFEVKEV TKEGNVYRVRLKVEENGKTKEMEIRLEDDNG RMRFKEIEIRG	>10
dcs_C_3_ss	DKEEAKKLCELIERAYRNPDVAREVFSPNTRFE DNGRETHDVEEWMEEIKRQGRPVECRVKEITR DGNEMRIRLRIRYNGEEYEMEIRFRHEDGQWK IEEMRWRG	1.2
dcs_C_5_ss	SEDDVRREVQRVWEEIRNNPEALCEYVDPNTH LHDGNQQYSGEEVCEYMRELVTRVEFRVRRV EKKGNTWKVEVEVRENGQEKEMHIEFEEDNG KFKFKRIEIRG	1.6
dcs_D_4_ss1	SPCKEESKLVEEFMKLMEQGDPEEMKKLISPDT RLERDGEEYNGEEVRQFWEEEMRQGLKFQVR EVTTQGCKVRIRVQVQKNGTTTQVQFEVEMR DGRIRRITVHERG	0.006
dcs_D_4_ss2	SPAKEESKLVEEFMKLMEQGDPEEMCKLISPDT RLERDGEEYNGEEVCQFWEEEMRQGLKFQVR EVTTQGAKVRIRVQVQKNGTTTQVQFEVEMR DGRIRRITVHERG	0.002
dcs_D_4_ss12	SPCKEESKLVEEFMKLMEQGDPEEMCKLISPDT RLERDGEEYNGEEVCQFWEEEMRQGLKFQVR EVTTQGCKVRIRVQVQKNGTTTQVQFEVEMR DGRIRRITVHERG	0.021
dcs_D_8_ss	PECVKVWKRIMEALQKGDPELLKKMISPDTRM EVNGQTFTGEEVVRYWEEEIRRGRQWTVKRY TEKGNECEFEVEQQDGDETRTYRVQVRVRNG QVEEIQVTQV	0.38
Homodimer designs		
dcs_E_4_dim1	TEEEVRKIMEKLKKAFKQGNPEQIVSLLSPDVR	0.061

	VQVGNQEFSGSEEAEKMWRKLMKFVDRVEVR RVSVFENVVVIEVEFEVNGQRYSMIFVFFVENG KVSMVIIYISPTMAKLMKQILNYG	
dcs_E_4_dim2	TREEVRKIMEKLKKAFKQGNPEQIVSLLSPDVV VVVGNQDFKGSEEAEKMWRKLMKFVDRVEV KKVQVYENIVIIEVEFEVNGQRYEMLFTFYVEN GKVKMVSIFISPTMKKLMKQILNYG	0.037
dcs_E_4_dim3	TEEEVRKIMEKLKKAFKQGNPEQIVSLLSPDVV VVVGNQSFSGSEEAEKMWRKLMKFVDRVEVR KVRVFENIVLIEVEFEVNGQRYSMFFTFYVENG KVAAVSIWISPTMKKLMKQILNYG	0.4
dcs_E_4_dim4	TAEEVRKIMEKLKKAFKQGNPEQIVSLLSPDVF VMVGNQSFSGSEEAEKMWRKLMKFVDRVEV KKVQVYENIVIIEVEFEVNGQRYAMLFTFYVEN GKVKAVSIFISPTMKKLMKQILNYG	1.2
dcs_E_4_dim5	TEEEVRKIMEKLKKAFKQGNPEQIVSLLSPDVA VQVGNQEFSGSEEAEKMWRKLMKFVDRVEVR DVRVAENIVVIFVEFEVNGQRYVMAFVFFVEN GKVSQVVIYISPTMKKLMKQILNYG	0.75
dcs_E_4_dim6	SREEIRKVVEEMLRSLKQGSPEDISKYLSPDVR LEVGNYTFEGSEQVTKFWRMWTKFVDRVEVK EVKVAGNYVIVVMSVEWNGKRWEATMIVTV RNGKIKRIILAVDEEFKKVVQNIWNLL	0.001
dcs_E_4_dim7	SREEIRKVVEEMLRSLKQGSPEDISKYLSPDVFL LVGNYTFEGSEQVTKFWRMWTKFVDRVEVRR VEVAGNAVVVLMEVEWNGKRWTFYMLVVVR NGKIKRIALAVDPEFSKVAQNIWNLL	0.16
dcs_E_4_dim8	TREEARKIMEKLKKAFKQGNPEQIVSLLSPDVR VVVGNQEFKGSEEAEKMWRKLMKFVDRVEV ARVRVDENMVVIAVEFEVNGQRYVMFFAFVV ENGKVKAVFIFISEEAMKLMKQILNYG	1.8
dcs_E_4_dim9	TEEEVRKIMEKLKKAFKQGNPEQIVSLLSPDVK VDVGNQSFSGSEEAEKMWRKLMKFVDRVEVR DVRVFENAVMIAVEFEVNGQRYKMIFTFYVEN GKVSMVSIYISPTMKKLMKQILNYG	0.007
Cavity mutants		
dcs_C_1_ss_cav1	SEEAKIAIELFKEAMKDPERFKEMCSPDTRIESN GQEYRGSEECKKFAEEMKKTHPWEVRVERYR SDGDRFEIELRVNFNGKTTRTEIRMRKVNGEFR IEEMRSHG	1.4
dcs_C_1_ss_cav2	SEEAKIAIELFKEAMKDPERFKEMCSPDTRIESN GQEYRGSEECKKSAEEMKKTHPWEVRVERYR SDGDRFEIELRVNFNGKTTRTEIRMRKVNGEFR IEEMRSHG	2.3
dcs_C_1_ss_cav3	SEEAKIAIELFKEAMKDPERFKEMCSPDTRIESN GQEYRGSEECKKYAEEMKKTHPTEVRVERYRS	7.0

	DGDRFEIELRVNSNGKTFRMEIRMRKVNGEFRIEEMRLHG	
dcs_C_1_ss_cav4	SEEAKIAIELFKEAMKDPERFKEMCSPDTRIESNGQEYRGSEECKKSAEEMKKTHPTEVRVERYRSDGDRFEIELRVNSNGKTFRMEIRMRKVNGEFRIEEMRLHG	1.7
dcs_C_1_ss_cav5	SEEAKIAIELFKEAMKDPERFKEMCSPDTRIESNGQEYRGSEECKKYAEEMKKTHPTEVRVERYRSDGDRFEIELRVNSNGKTTRTEIRMRKVNGEFRIEEMRSHG	2.5
dcs_C_1_ss_cav6	SEEAKIAIELFKEAMKDPERFKEMCSPDTRIESNGQEYRGSEECKKSAEEMKKTHPTEVRVERYRSDGDRFEIELRVNSNGKTTRTEIRMRKVNGEFRIEEMRSHG	0.85
dcs_D_2_cav1	PEEEKAARLFIECLEKGDPECMRKVISPDTRVEFNGSELTGDEVVESVKELQKSGTQLHLRRYTKEGNSWRFEIQADNNGQTWQSEIQIEVRNGRIKRATSTA	1.5
dcs_D_2_cav2	PEEEKAARLFIEALEKGDPELCRKVISPDTRAEINGSEYTGDEVVESCKELQKSGTQIHLRRYTKEGNSWRFEVQADNNGQTYQSEIQIEVRNGRIKRATSTA	2.7
dcs_D_2_cav3	PEEEKACRLFIEALEKGDPELMRKVISPDTRAEINGREFTGDEVVESVKEMQKRGVQAHLRRYTKEGNSCRFEVQTDINGQTEQSEIQIEVRNGRIKRATTTA	0.00005
dcs_E_4_dim9_cav1	TEEEVRKIMEKLKKAFKQGNPEQIVSLLSPDVKVDVGNQSFSGSEEAEKAQRKLMKFVDRVEVRDVRVFENAVMIAVEFEVNGQRYKMITTFYVENGKVSMVSIYISPTMKKLMKQILNYG	2.9
dcs_E_4_dim9_cav2	TEEEVRKIMEKLKKAFKQGNPEQIVSLLSPDVKVDVGNQSFSGSEEAEKAARKLMKFVDRVEVRDVRVFENAVMIAVEFEVNGQRYKVIVTFYVENGKVSMVSIYISPTMKKLMKQILNYG	0.72
dcs_E_4_dim9_cav3	TEEEVRKIMEKLKKAFKQGNPEQIVSLLSPDVKVDVGNQSFSGSEEAEKAARKLMKFVDRVEVRDVRVFENAVMIAVEFEVNGQRYKMIFTFYVENGKVSMVSIYISPTMKKLMKQILNYG	0.21

Table S3. Parameters fitted to GdmCl denaturation curves for designed proteins. Denaturation curves were measured for those proteins with soluble expression. N/A indicates data that is not available due to the lack of sigmoidal character in the denaturation curves (highly linear). In those cases the slope of the native baseline was calculated from a linear fit, which are indicated by an asterisk (*). For designs with disulfides, denaturation curves in the presence of the reducing agent Tris(2-carboxyethyl)phosphine (TCEP) are also shown.

Design name	slope native baseline (M^{-1})	m-value ($kcal \cdot mol^{-1} \cdot M^{-1}$)	m-value deviation (%)	ΔG ($kcal \cdot mol^{-1}$)	C_m (M)
dcs_A_1	0.053	2.0	-3.6	-5.0	2.5
dcs_A_2	0.033	2.0	-4.5	-4.0	2.1
dcs_A_3	-0.010	2.1	5.5	-3.8	1.8
dcs_A_4	-0.066	1.7	-13.4	-4.5	2.7
dcs_B_2	0.043	2.8	-14.6	-10.5	3.8
dcs_B_5	0.043	3.1	-5.3	-13.4	4.5
dcs_C_1	0.036	1.4	-55.5	-3.2	2.3
dcs_C_2	-0.019	4.0	30.2	-8.5	2.2
dcs_C_3	-0.039	0.8	-72.9	-2.8	3.3
dcs_C_5	-0.013	4.2	34.2	-9.2	2.3
dcs_D_1	0.197	3.7	20.3	-9.7	2.6
dcs_D_2	0.057	2.8	-8.0	-7.7	2.8
dcs_D_3	0.167	3.8	19.0	-6.0	1.6
dcs_D_4	0.047	2.7	-14.4	-4.8	1.8
dcs_D_5	0.052	3.0	-4.1	-4.5	1.6
dcs_D_6	0.064	2.6	-16.8	-6.1	2.4
dcs_D_7	0.068	4.4	42.1	-10.8	2.5
dcs_D_8	0.089	3.0	-1.2	-8.8	2.9
dcs_E_1	0.011	3.9	6.7	-23.1	6.0
dcs_E_2	0.016	4.2	16.3	-23.0	5.5
dcs_E_3	0.008	3.3	-8.4	-19.7	6.0
dcs_E_4	0.011	3.8	6.3	-18.4	4.8
dcs_E_5	0.040	2.5	-29.9	-9.7	3.9
dcs_F_3	0.213*	N/A	N/A	N/A	N/A
dcs_F_5	0.203*	N/A	N/A	N/A	N/A
dcs_F_6	0.125*	N/A	N/A	N/A	N/A
dcs_F_7	0.200*	N/A	N/A	N/A	N/A
dcs_F_8	0.221*	N/A	N/A	N/A	N/A
dcs_F_9	0.078	3.5	-4.7	-11.8	3.4
Disulfide variants					
dcs_C_1_ss	-0.005	3.5	9.6	-11.5	3.4
dcs_C_1_ss + TCEP	0.067	2.1	-33.5	-5.7	2.7
dcs_C_5_ss	-0.013	2.3	-25.9	-6.7	3.0
dcs_D_4_ss1	0.066	2.9	-8.5	-7.1	2.5
dcs_D_4_ss2	0.121	3.5	10.2	-9.7	2.8

dcs_D_4_ss12	0.065	3.3	3.0	-11.7	3.6
dcs_D_4_ss12 + TCEP	-0.032	1.4	-56.3	-2.5	3.2
dcs_D_8_ss	-0.003	3.6	17.1	-16.4	4.6
dcs_D_8_ss + TCEP					
Homodimer designs					
dcs_E_4_dim9	0.030	3.67	1.46	-19.77	5.44
Cavity mutants					
dcs_C_1_ss_cav1	-0.212	1.75	-44.74	-2.16	1.41
dcs_C_1_ss_cav3	0.223	4.0	26.49	-10.73	2.62
dcs_C_1_ss_cav4	0.061	3.02	-4.49	-5.94	1.98
dcs_C_1_ss_cav5	-0.547	1.5	-52.46	-1.10	1.06
dcs_E_4_dim9_cav1	-0.008	2.09	-42.24	-8.78	4.24
dcs_E_4_dim9_cav2	0.003	4.55	25.59	-18.13	4.03
dcs_E_4_dim9_cav3	0.029	3.02	-16.39	-13.19	4.66

Table S4. X-ray crystallography data collection and refinement statistics.

Design name	**dcs_A_4**	**dcs_D_2**	**dcs_C_1_ss**
PDB ID	**4R80**	**5L33**	**5TS4**
Data collection			
Space group	C2	P 21 21 21	C 2 2 21
Cell dimensions			
a, b, c (Å)	56.14, 70.62, 41.04	28.25, 34.36, 100.39	81.31, 101.54, 101.58
α, β, γ (°)	90, 113.16, 90	90, 90, 90	90, 90, 90
Wavelength (Å)	0.97916	0.97625	1.0
Resolution (Å)	2.44 (2.44-2.48)	2.0 (2.0-2.05)	3.0 (3.0-3.07)
R_{sym} or R_{merge} (%)	5.2 (6.0)	4.7 (23.3)	11 (101)
CC1/2	0.986	0.99 (0.98)	0.91 (0.64)
I/σI	33.9 (18.2)	24.6 (7.7)	16 (1.1)
Completeness (%)	93.1 (42.8)	92.3 (97.0)	98 (87)
Redundancy	6.7 (6.2)	7.8 (7.8)	8 (7)
Refinement			
Resolution (Å)	2.44	2.0	3.0
No. reflections	5311	6503	8365
R_{work} (%) / R_{free} (%)	21.8 / 25.6	17.2/20.1	27.4/31.6
No. atoms			
Protein	1216	918	2814
Water	59	69	23
B-factors (Å^2)			
Protein	27.6	31.1	108.2
Water	29.8	42.0	65.7
R.m.s. deviations			
Bond lengths (Å)	0.003	0.007	0.006
Bond angles (°)	0.584	0.875	0.719
Ramachandran statistics (%)			
Favored	99	99	98
Outliers	0	0	0
Rotamer outliers (%)	0	0	0.9

*Values in parentheses are for highest-resolution shell.

Table S5. X-ray crystallography data collection and refinement statistics.

Design name	**dcs_E_3**	**dcs_E_4**	**dcs_E_4_dim9**
PDB ID	**5TPJ**	**5TRV**	**5TPH**
Data collection			
Space group	P 41 21 2	P 42 21 2	P 1 21 1
Cell dimensions			
a, b, c (Å)	49.81, 49.81, 113.1	75.53, 75.53, 50.07	38.21, 32.79, 86.48
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 92.11, 90
Wavelength (Å)	0.99990	0.99990	0.97625
Resolution (Å)	3.10 (3.31-3.10)	2.91 (3.09-2.91)	2.47 (2.57-2.47)
R_{sym} or R_{merge} (%)	7.2 (26.8)	3.4 (20.5)	3.5 (18.4)
CC1/2	0.99 (0.99)	1.0 (0.98)	0.99 (0.98)
I/σI	21.7 (8.8)	72.8 (7.9)	22.1 (7.0)
Completeness (%)	100 (100)	97.9 (99.6)	99.9 (100)
Redundancy	11.7 (12.4)	4.5 (4.5)	3.7 (3.8)
Refinement			
Resolution (Å)	3.10	2.91	2.47
No. reflections	2881	3314	7943
R_{work} (%) / R_{free} (%)	22.1/ 26.5	24.9/29.7	22.3/25.7
No. atoms			
Protein	970	910	1813
Water	1	2	28
B-factors (Å^2)			
Protein	75.4	75.0	52.4
Water	31.9	102.6	55.6
R.m.s. deviations			
Bond lengths (Å)	0.007	0.002	0.003
Bond angles (°)	0.977	0.521	0.502
Ramachandran statistics (%)			
Favored	98	98	98
Outliers	1	0	0
Rotamer outliers (%)	3	0	0

*Values in parentheses are for highest-resolution shell.

Table S6. X-ray crystallography data collection and refinement statistics.

Design name	**dcs_E_4_dim9_cav3**
PDB ID	**5U35**
Data collection	
Space group	P 1 21 1
Cell dimensions	
a, b, c (Å)	38.01, 33.20, 86.59
α, β, γ (°)	90, 91.78, 90
Wavelength (Å)	0.99986
Resolution (Å)	1.8 (1.9-1.8)
R_{sym} or R_{merge} (%)	10.8 (112.2)
CC1/2	0.99 (0.50)
I/σI	7.8 (1.8)
Completeness (%)	99.9 (99.8)
Redundancy	4.3 (4.3)
Refinement	
Resolution (Å)	1.8
No. reflections	20357
R_{work} (%) / R_{free} (%)	20.1/24.7
No. atoms	
Protein	1926
Water	140
B-factors (Å^2)	
Protein	36.5
Water	44.5
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.762
Ramachandran statistics (%)	
Favored	99
Outliers	0
Rotamer outliers (%)	0

*Values in parentheses are for highest-resolution shell.

Table S7. NMR and refinement statistics for protein structures.

Design name	**dcs_A_3**	**dcs_B_2**
NESG ID	**OR485**	**OR664**
PDB ID	**5KPH**	**5KPE**
NMR distance and dihedral constraints		
Distance constraints		
Total NOE	2012	3395
Intra-residue	553	673
Inter-residue		
Sequential (|i-j| = 1)	505	865
Medium-range (|i-j| ≤ 4)	301	655
Long-range (|i-j| ≥ 5)	653	1202
Intermolecular		
Hydrogen bonds	76	56
Total dihedral angle restraints	139	186
Phi	35	93
Psi	35	93
Structure statistics		
Violations		
RMS of distance violation/constraint¶ (Å)	0.01	0.01
RMS of dihedral angle violation/constraint (°)	0.88	0.93
Max distance constraint violation (Å)	0.66	0.40
Max dihedral angle violation (°)	7.80	974
Average medoid r.m.s.d.** (Å)		
Heavy	0.5±0.15	0.5±0.19
Backbone	1.1±0.10	0.9±0.11
RPF Scores		
Recall	0.977	0.963
Precision	0.929	0.973
F-measure	0.952	0.968
DP-scores	0.786	0.886
Structure quality factors (raw/Z-score⌘)		
Procheck G-factor (phi / psi only)**	-0.42/1.34	-0.20/-0.47
Procheck G-factor (all dihedral angles)**	0-.19/-1.12	-0.14/-0.83
Verify3D	0.34/-1.93	0.407/0.16
ProsaII (-ve)	0.79/0.58	1.08/1.78
MolProbity clashscore	15.34/-1.11	13.11/-0.72
Ramachandran plot summary from Richardson's lab		
Most favored regions (%)	97.1	98.6
Allowed regions (%)	2.8	1.4
Disallowed regions (%)	0.1	0

***** Analyzed for the 20 lowest energy refined structures for each designed protein, which are deposited in the PDB: OR485 (5kph, residues 1-85), DI_7S, OR664 (5kpe, residues 1-120) using PDBSTAT (*64*) and PSVS 1.4 (*54*, *55*).

§ PEG and phage were used as alignment media 1 and 2.

¶ Calculated by using sum over r^{-6}.

⌘ With respect to mean and standard deviation for a set of 252 X-ray structures with sequence lengths < 500, resolution ≤ 1.80 Å, R-factor ≤ 0.25 and R-free ≤ 0.28; a positive value indicates a 'better' score.

** Calculated among 20 refined structures for ordered residues that have sum of phi and psi order parameters (*65*) S(phi)+S(psi)>1.8 (*54*). The ordered residues of OR485: 4-48, 50-75; OR664: 4-52 55-82, 84-108.

Example of RosettaScripts XML protocol used for the backbone generation of topologies with 6-stranded β-sheets and 3 helices (bbgen.xml):

```
<ROSETTASCRIPTS>
        <SCOREFXNS>
             # increased weight for the hbond_lr_bb score term rewards strand pair formation
              <SFXN1 weights=fldsgn_cen >
                        <Reweight scoretype=hbond_sr_bb weight=1.0 />
                        <Reweight scoretype=hbond_lr_bb weight=2.0 />
                        <Reweight scoretype=atom_pair_constraint weight=1.0 />
                        <Reweight scoretype=angle_constraint weight=1.0 />
                        <Reweight scoretype=dihedral_constraint weight=1.0 />
              </SFXN1>
              <SFXN2 weights=fldsgn_cen >
                        <Reweight scoretype=hbond_sr_bb weight=1.0 />
                        <Reweight scoretype=hbond_lr_bb weight=2.0 />
                        <Reweight scoretype=atom_pair_constraint weight=1.0 />
                        <Reweight scoretype=angle_constraint weight=1.0 />
                        <Reweight scoretype=dihedral_constraint weight=1.0 />
              </SFXN2>
              <SFXN3 weights=fldsgn_cen >
                        <Reweight scoretype=hbond_sr_bb weight=1.0 />
                        <Reweight scoretype=hbond_lr_bb weight=2.0 />
                        <Reweight scoretype=atom_pair_constraint weight=1.0 />
                        <Reweight scoretype=angle_constraint weight=1.0 />
                        <Reweight scoretype=dihedral_constraint weight=1.0 />
              </SFXN3>
              <SFXN4 weights=fldsgn_cen >
                        <Reweight scoretype=hbond_sr_bb weight=1.0 />
                        <Reweight scoretype=hbond_lr_bb weight=1.0 />
                        <Reweight scoretype=atom_pair_constraint weight=0.5 />
                        <Reweight scoretype=angle_constraint weight=0.25 />
                        <Reweight scoretype=dihedral_constraint weight=0.5 />
              </SFXN4>
        </SCOREFXNS>

        <FILTERS>
         # Step 1:  Build strands 4 and 5
        # Ensure the dssp secondary structure of the generated pose matches that of the blueprint file, and also that the ABEGO strings also match (regular strand residues have “B” abego and bulge residues “A” abego).
        <SecondaryStructure name=ss1  use_abego=1 blueprint="../bp1" cutoff=1.0 confidence=1/>
        # ensure strands to be paired and in which orientation (A: antiparallel, P: parallel)
          <SheetTopology name=st1 topology="1-2.A.99" blueprint="../bp1" confidence=1/>
        # filter strands with bending and intra-strand twist values out of specified bounds.
        <StrandCurvatureByLevels name=st1_curv StrandID=1 concavity_reference_residue="last" concavity_direction=1 bend_level=2 min_bend=20 max_bend=50 twist_level=2 min_twist=0 max_twist=180 confidence="1" />
          <StrandCurvatureByLevels name=st2_curv StrandID=2 concavity_reference_residue="first" concavity_direction=1 bend_level=2 min_bend=20 max_bend=50 twist_level=2 min_twist=0 max_twist=180 confidence="1" />
          <CompoundStatement name=secst1 >
            <AND filter_name=ss1 />
            <AND filter_name=st1 />
            <AND filter_name=st1_curv />
            <AND filter_name=st2_curv />
```

```
        </CompoundStatement>

         # Step 2:  Add strand 3
         <SecondaryStructure name=ss2  use_abego=1 blueprint="../bp2" cutoff=1.0 confidence=1/>
         <SheetTopology name=st2 topology="1-2.A.99;2-3.A.99" blueprint="../bp2" confidence=1 />
         <CompoundStatement name=secst2 >
           <AND filter_name=ss2 />
           <AND filter_name=st2 />
         </CompoundStatement>

         # Step 3:  Add strand 6
         <SecondaryStructure name=ss3  use_abego=1 blueprint="../bp3" cutoff=1.0 confidence=1/>
         <SheetTopology name=st3 topology="1-2.A.99;2-3.A.99;3-4.A.99" blueprint="../bp3"
confidence=1/>
         <CompoundStatement name=secst3 >
           <AND filter_name=ss3 />
           <AND filter_name=st3 />
         </CompoundStatement>

         # Step 4: Add helix 3 and strands 1 and 2
        <SheetTopology name=st4 topology="1-2.A.0;1-6.P.-5;3-4.A.99;4-5.A.99;5-6.A.99"
blueprint="../bp4.b" confidence="1"/>
        <SecondaryStructure name=ss4  use_abego=1 blueprint="../bp4" cutoff=1.0 confidence="1"/>
        <HelixBend name=hbend4 threshold=155.0 blueprint="../bp4.b" HelixID=2 confidence=1 />
         <CompoundStatement name=secst4 >
           <AND filter_name="st4" />
           <AND filter_name="dist4a" />
           <AND filter_name="hbend4" />
           <AND filter_name="st1hx2" />
        </CompoundStatement>

        </FILTERS>
        <TASKOPERATIONS>
        </TASKOPERATIONS>
        <MOVERS>
         # General movers
                <DumpPdb name="pdb1" fname="iter1.pdb" scorefxn="SFXN1" />
                <DumpPdb name="pdb2" fname="iter2.pdb" scorefxn="SFXN2" />
                <DumpPdb name="pdb3" fname="iter3.pdb" scorefxn="SFXN3" />
                <DumpPdb name="pdb4" fname="iter4.pdb" scorefxn="SFXN4" />
                <Dssp name=dssp/>

                <SwitchResidueTypeSetMover name=fullatom set=fa_standard/>
                <SwitchResidueTypeSetMover name=cent set=centroid/>
        <MakePolyX name="polyval" aa="VAL" />

                # Step 1:  Build strands 4 and 5
                <SetSecStructEnergies name=set_ssene1 scorefxn=SFXN1 blueprint="../bp1.b" />
        # Fragment assembly based on secondary structure and ABEGO bins from the blueprint file.
        Constraints specifying strand bending and hydrogen bond pairing are added to improve sampling.
                <BluePrintBDR name=bdr1 scorefxn=SFXN1 use_abego_bias=1 blueprint="../bp1.b"
        constraint_file="../cst1"/>
        <ConstraintSetMover name="addcst1" add_constraints="1" cst_file="../cst1"/>
                <MinMover name=min1 scorefxn=SFXN1 chi=1 bb=1
        type="dfpmin_armijo_nonmonotone_atol" tolerance=0.0001/>
                <ParsedProtocol name=cenmin1 >
```

```
            <Add mover_name=cent />
            <Add mover_name=addcst1 />
            <Add mover_name=min1 />
            <Add mover_name=fullatom />
          </ParsedProtocol>

          <ParsedProtocol name=bdr1ss >
            <Add mover_name=bdr1 />
            <Add mover_name=cenmin1 />
            <Add mover_name=dssp />
          </ParsedProtocol>
          <LoopOver name=loop1 mover_name=bdr1ss filter_name=secst1 drift=0 iterations=50
ms_whenfail=FAIL_DO_NOT_RETRY/>


          # Step 2:  Add strand 3
          <SetSecStructEnergies name=set_ssene2 scorefxn=SFXN2 blueprint="../bp2.b" />
          <BluePrintBDR name=bdr2 scorefxn=SFXN2 use_abego_bias=1 blueprint="../bp2.b"
constraint_file="../cst2"/>
          <ConstraintSetMover name="addcst2" add_constraints="1" cst_file="../cst2"/>
          <MinMover name=min2 scorefxn=SFXN2 chi=1 bb=1
type="dfpmin_armijo_nonmonotone_atol" tolerance=0.0001 />
          <ParsedProtocol name=cenmin2 >
            <Add mover_name=cent />
            <Add mover_name=addcst2 />
            <Add mover_name=min2 />
            <Add mover_name=fullatom />
          </ParsedProtocol>
          <ParsedProtocol name=bdr2ss >
            <Add mover_name=bdr2 />
            <Add mover_name=cenmin2 />
  <Add mover_name=dssp />
          </ParsedProtocol>
          <LoopOver name=loop2 mover_name=bdr2ss filter_name=secst2 drift=0 iterations=50
ms_whenfail=FAIL_DO_NOT_RETRY/>

          # Step 3:  Add strand 6
          <SetSecStructEnergies name=set_ssene3 scorefxn=SFXN3 blueprint="../bp3.b" />
          <BluePrintBDR name=bdr3 scorefxn=SFXN3 use_abego_bias=1 blueprint="../bp3.b"
constraint_file="../cst3" />
<ConstraintSetMover name="addcst3" add_constraints="1" cst_file="../cst3"/>
          <MinMover name=min3 scorefxn=SFXN3 chi=1 bb=1
type="dfpmin_armijo_nonmonotone_atol" tolerance=0.0001 />
          <ParsedProtocol name=cenmin3 >
            <Add mover_name=cent />
  <Add mover_name=addcst3 />
            <Add mover_name=min3 />
            <Add mover_name=fullatom />
          </ParsedProtocol>
          <ParsedProtocol name=bdr3ss >
            <Add mover_name=bdr3 />
            <Add mover_name=cenmin3 />
  <Add mover_name=dssp />
          </ParsedProtocol>
          <LoopOver name=loop3 mover_name=bdr3ss filter_name=secst3 drift=0 iterations=50
ms_whenfail=FAIL_DO_NOT_RETRY/>
```

```
        # Step 4: Add helix 3 and strands 1 and 2
        <SetSecStructEnergies name=set_ssene4 scorefxn=SFXN4 blueprint="../bp4.b"
hs_angle=180 hs_ortho_angle=65 hs_atr_dist=16.0 hs_atr_dist_wts=1.0 hs_angle_wts=1.0
hs_ortho_angle_wts=1.0 natbias_ss=1.0 natbias_hs=0.0 />
<ConstraintSetMover name="addcst4" add_constraints="1" cst_file="../cst4"/>
<MinMover name=min4 scorefxn=SFXN4 chi=1 bb=1 type="dfpmin_armijo_nonmonotone_atol"
tolerance=0.0001 />
        <ParsedProtocol name=cenmin4 >
          <Add mover_name=cent />
  <Add mover_name=addcst4 />
          <Add mover_name=min4 />
          <Add mover_name=fullatom />
        </ParsedProtocol>
        <BluePrintBDR name="bdr4" scorefxn="SFXN4" use_abego_bias="1"
blueprint="../bp4.b" constraint_file="../cst4"/>
        <ParsedProtocol name=bdr4ss >
          <Add mover_name=bdr4 />
  <Add mover_name=cenmin4 />
  <Add mover_name=dssp />
        </ParsedProtocol>
        <LoopOver name="loop4" mover_name="bdr4ss" filter_name="secst4" drift="0"
iterations="50" ms_whenfail="FAIL_DO_NOT_RETRY"/>

# Step 5: Add helix 1 and 2
# The MultipleOutputWrapper (MOW) is used to generate multiple poses with the same backbone
for those parts constructed until step4, but with backbone variability in those parts added in the
next steps. In this case step5 adds the two N-terminal helices what are flexible and in this way the
generation of the global topology becomes more cost-effective.
<MultipleOutputWrapper name="multi_step_5" max_output_poses=5>
<ROSETTASCRIPTS>
        <SCOREFXNS>
                <SFXN5 weights=fldsgn_cen >
                        <Reweight scoretype=hbond_sr_bb weight=2.0 />
                        <Reweight scoretype=hbond_lr_bb weight=2.0 />
                        <Reweight scoretype=atom_pair_constraint weight=0.5 />
                        <Reweight scoretype=angle_constraint weight=0.25 />
                        <Reweight scoretype=dihedral_constraint weight=0.5 />
        </SFXN5>
                <standardfxn weights=talaris2013.wts />
        </SCOREFXNS>
        <FILTERS>
                <HelixPairing name=hp23 dist=15 cross=50.0 helix_pairings="2-3.A"
blueprint="../bp5" output_type="dist"/>
                <HelixPairing name=hp12 dist=15 cross=20.0 helix_pairings="1-2.A"
blueprint="../bp5" output_type="dist"/>
                <HelixBend name=hbend5 threshold=155.0 blueprint="../bp5.b" HelixID=1
confidence=1 />
                # filter non-compact designs
                <SasaBalance name=sasa_balance5 ratio_sc=2.5 confidence=1 />
                # filter non-compact designs
                <AverageDegree name="avdeg5" threshold="15.0" distance_threshold="10.0"
confidence=1/>
                <CompoundStatement name=secst5 >
```

```
                                        <AND filter_name="hbend5" />
                                       <AND filter_name="hp23" />
                                       <AND filter_name="sasa_balance5" />
                                  <AND filter_name="avdeg5" />
                        </CompoundStatement>

                        # ensure the local backbone conformation is native-like
                                <FragmentLookupFilter name="faulty_fragments_all"
                                lookup_name="source_fragments_4_mer" store_path="/
                                lab/databases/VALL_clustered/backbone_profiler_database_06032014"
                                lookup_mode="first" chain="1" threshold="0" confidence="1" />

                        # Reporting filters used to rank the final generated backbones
                        <AverageDegree name="all_avdeg" threshold="15.0"
distance_threshold="10.0" confidence=0/>
                        <SasaBalance name="all_sasa_balance" ratio_sc=2.0 confidence=0 />
                        <ScoreType name="all_lr_hb" scorefxn="SFXN6" score_type="hbond_lr_bb"
threshold=0.0 confidence=0 />

                        <ScoreType name="score" scorefxn="standardfxn" score_type="total_score"
threshold=0.0 confidence="0" />
                        <ResidueCount name="nres" confidence="0" />
                        <CalculatorFilter name="score_res" confidence="0" equation="SCORE/NRES"
threshold="-1.9">
                                <SCORE name="SCORE" filter_name="score" />
                                <NRES name="NRES" filter_name="nres" />
                        </CalculatorFilter>

                </FILTERS>
                <TASKOPERATIONS>
                        <LimitAromaChi2 name=limitchi2 />
                        <LayerDesign name="layer_design" layer="all" use_sidechain_neighbors="1"
pore_radius="0.2" core="3.0" surface="1.8" repack_non_design="1" make_pymol_script="1">
                                <core>
                                        <all append="M"/>
                                 </core>
                        </LayerDesign>
                </TASKOPERATIONS>
                <MOVERS>
                        <SwitchResidueTypeSetMover name=fullatom set=fa_standard/>
                        <SwitchResidueTypeSetMover name=cent set=centroid/>
                        <DumpPdb name="pdb5" fname="iter5.pdb" scorefxn="SFXN5" />
                        <Dssp name=dssp/>

                        # Add helix 1 and 2
                         <ConstraintSetMover name="addcst5" add_constraints="1" cst_file="../cst5"/>
                         <MinMover name=min5 scorefxn=SFXN5 chi=1 bb=1
type="dfpmin_armijo_nonmonotone_atol" tolerance=0.0001>
                         <ParsedProtocol name=cenmin5 >
                                <Add mover_name=cent />
                                <Add mover_name=addcst5 />
                                <Add mover_name=min5 />
                                <Add mover_name=min5 />
                                <Add mover_name=fullatom />
                        </ParsedProtocol>
```

```
                        <BluePrintBDR name="bdr5" scorefxn="SFXN5" use_abego_bias="1"
        blueprint="../bp5.b" constraint_file="../cst5" />
                        <ParsedProtocol name=bdr5ss >
                                <Add mover_name=bdr5 />
                                <Add mover_name=cenmin5 />
                                <Add mover_name=dssp />
                        </ParsedProtocol>
                        <LoopOver name="loop5" mover_name="bdr5ss" filter_name="secst5"
        drift="0" iterations="50" ms_whenfail="FAIL_DO_NOT_RETRY"/>

                        # Step 6: Quick design
                        <FastDesign name="fdesign" task_operations="limitchi2, layer_design"
        scorefxn="standardfxn"  repeats="1" clear_designable_residues="1"  />
                </MOVERS>
                <PROTOCOLS>
                        <Add mover_name=loop5 />
                        <Add mover_name=design />
                        <Add filter_name=faulty_fragments_all />
                        <Add filter_name=all_avdeg />
                        <Add filter_name=all_sasa_balance />
                        <Add filter_name=all_lr_hb />
                        <Add filter_name=score_res />
                       </PROTOCOLS>
        </ROSETTASCRIPTS>
                </MultipleOutputWrapper>
        </MOVERS>

        <APPLY_TO_POSE>
        </APPLY_TO_POSE>

        <PROTOCOLS>
                 <Add mover_name=polyval />
                 <Add mover_name=set_ssene1 />
                 <Add mover_name=loop1 />
                 <Add mover_name=set_ssene2 />
                 <Add mover_name=loop2 />
                 <Add mover_name=set_ssene3 />
                 <Add mover_name=loop3 />
                 <Add mover_name=set_ssene4 />
                 <Add mover_name=loop4 />
        <Add mover_name="multi_step_5"/>
        </PROTOCOLS>
</ROSETTASCRIPTS>
```

Example of command line for backbone generation calculations:
rosetta_scripts.static.linuxgccrelease -database path_to_database -parser:protocol bbgen.xml -picking_old_max_score 1 -holes:dalphaball path_to_DAlphaBall/DAlphaBall.icc -nstruct 100

Example of RosettaScripts XML protocol used for sequence design (design.xml):

```
<ROSETTASCRIPTS>
        <SCOREFXNS>
                # modified talaris2013.wts with a high reference weight for Alanine (2.0) to avoid
overrepresentation, especially in helices.
                <SFXN1 weights=talaris2013_highAlanine.wts />
```

```
            <SFXN2 weights=talaris2013_highAlanine.wts />
            # use the actual talaris2013.wts for final reporting filters on the generated poses.
            <standardfxn weights=talaris2013.wts />
        </SCOREFXNS>

        <FILTERS>
            # secondary structure prediction based on the designed amino acid sequence. Necessary
to achieve good fragment quality afterwards.
            <SSPrediction name="sspred" confidence="1" cmd="/path/runpsipred_single"
use_probability="0" use_svm="0" threshold=0.75 blueprint="model.bp"/>
            <ScoreType name="rama" scorefxn="standardfxn" score_type="rama" threshold=0.0
confidence="0" />
            # filter designs with low packing
            <PackStat name=pack threshold=0.6 confidence=1/>
            # filter designs with low packing
            <Holes name=holes threshold=2.0 confidence=1/>
            # filter designs with low packing among hydrophobic residues in the core
            <SidechainAverageDegree name=sc_avdeg threshold=8.0 pho_pho=1 all_pho=0
confidence=1 task_operations="core_layer"/>
            # normalize total score by the number of residues.
            <ScoreType name="score" scorefxn="standardfxn" score_type="total_score"
threshold=0.0 confidence="1" />
            <ResidueCount name="nres" confidence="0" />
            <CalculatorFilter name="score_res" confidence="1" equation="SCORE/NRES"
threshold="-2.1">
                <SCORE name="SCORE" filter_name="score" />
                <NRES name="NRES" filter_name="nres" />
            </CalculatorFilter>

            # filter designs with low aromatic content
            <ResidueCount name=AroCount residue_types="PHE,TYR,TRP" min_residue_count=6
max_residue_count=20 confidence=1 />
            # filter designs with low shape complementarity for each of the helices.
        <SSShapeComplementarity name=sc_hx1 HelixID=1 helices=1 loops=0 confidence=1 verbose=1
        threshold=0.6/>
        <SSShapeComplementarity name=sc_hx2 HelixID=2 helices=1 loops=0 confidence=1 verbose=1
        threshold=0.6/>
            <SSShapeComplementarity name=sc_hx3 HelixID=3 helices=1 loops=0 confidence=1
verbose=1 threshold=0.6/>
            # Combined filter used in the generic montecarlo optimization of the designs. Increase
packing while minimizing the energy.
            <CombinedValue name=comb_filters confidence=0>
                <Add filter_name=sc_avdeg factor=-0.1/>
                <Add filter_name=score_res factor=1.0/>
            </CombinedValue>

            # set of filters that can be used in a LoopOver mover
            <CompoundStatement name=filt >
                <AND filter_name=score_res />
                <AND filter_name=pack />
                <AND filter_name=holes />
                <AND filter_name=sspred />
                <AND filter_name=sc_avdeg />
                <AND filter_name=AroCount />
            </CompoundStatement>
```

```
	</FILTERS>
	<TASKOPERATIONS>
		<LimitAromaChi2 name=limitchi2 />
		<ExtraRotamersGeneric name=ex1ex2 ex1=1 ex2=1 />
		# resfile with amino acid restrictions for bulges and loops.
		<ReadResfile name=resfile filename=%%resfile%%/>
		<LayerDesign name="layer_design" layer="all" use_sidechain_neighbors="1"
pore_radius="0.2" core="3.0" surface="1.8" repack_non_design="1" make_pymol_script="1">
			<core>
				<all append="M"/>
			</core>
		</LayerDesign>

		<LayerDesign name="core_layer" layer="core" use_sidechain_neighbors="1"
pore_radius="0.2" core="3.0" surface="1.8" repack_non_design="1" make_pymol_script="1">
			<core>
				<all append="M"/>
			</core>
		</LayerDesign>

		# layer definition to remove large hydrophobics exposed to solvent
		<LayerDesign name="layer_gen" layer="hydrophobes" use_sidechain_neighbors="0"
repack_non_design="0" pore_radius="2.0" make_pymol_script="0" core_E="20" surface_E="40"
core_H="20" surface_H="40" surface_L="40" core_L="10">
			<CombinedTasks name="hydrophobes">
				<all copy_layer="surface" />
				<SelectBySASA mode="sc" state="monomer" probe_radius="1.5"
core_asa="20" surface_asa="40" surface="1" />
				<OperateOnCertainResidues>
					<RestrictToRepackingRLT/>
					<NoResFilter>
						<ResidueName3Is name3="TRP,PHE,MET"/>
					</NoResFilter>
				</OperateOnCertainResidues>
			</CombinedTasks>
		</LayerDesign>
	</TASKOPERATIONS>
	<MOVERS>
		<Dssp name=dssp/>
		<FastDesign name="fdesign" task_operations="ex1ex2,resfile,limitchi2,layer_design"
scorefxn="SFXN1"  repeats="2" clear_designable_residues="0"  />
		<FastDesign name="rm_hydrophobes" task_operations="ex1ex2,resfile,layer_gen"
scorefxn="SFXN2"  repeats="1" clear_designable_residues="0" />
		<ParsedProtocol name=design >
			<Add mover_name=fdesign />
			<Add mover_name=rm_hydrophobes />
			<Add mover_name=dssp />
		</ParsedProtocol>
		<GenericMonteCarlo name=genericmc mover_name=design filter_name=comb_filters
trials=10 sample_type=low temperature=0.6 drift=1/>
	</MOVERS>
	<APPLY_TO_POSE>
	</APPLY_TO_POSE>
	<PROTOCOLS>
		<Add mover_name=genericmc />
		<Add filter_name=score_res />
```

```
                <Add filter_name=holes />
                <Add filter_name=pack />
                <Add filter_name=rama />
                <Add filter_name=sspred />
                <Add filter_name=AroCount />
                <Add filter_name=sc_avdeg />
                <Add filter_name=sc_hx1 />
                <Add filter_name=sc_hx2 />
                <Add filter_name=sc_hx3 />
        </PROTOCOLS>
</ ROSETTASCRIPTS >
```

Example of command line for sequence design calculations:
rosetta_scripts.static.linuxgccrelease -database path_to_database -parser:protocol design.xml -parser:script_vars resfile=resfile_name -holes:dalphaball path_to_DAlphaBall/DAlphaBall.icc -nstruct 100

References and Notes:

1. C. E. Tinberg, S. D. Khare, J. Dou, L. Doyle, J. W. Nelson, A. Schena, W. Jankowski, C. G. Kalodimos, K. Johnsson, B. L. Stoddard, D. Baker, Computational design of ligand-binding proteins with high affinity and selectivity. *Nature*. **501**, 212–216 (2013).
2. D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. Dechancie, J. Betker, J. L. Gallaher, E. a Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, D. Baker, Kemp elimination catalysts by computational enzyme design. *Nature*. **453**, 190–195 (2008).
3. L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Rothlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas, D. Hilvert, K. N. Houk, B. L. Stoddard, D. Baker, De novo computational design of retro-aldol enzymes. *Science*. **319**, 1387–1391 (2008).
4. J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, K. N. Houk, F. E. Michael, D. Baker, Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science*. **329**, 309–313 (2010).
5. S. Rajagopalan, C. Wang, K. Yu, A. P. Kuzin, F. Richter, S. Lew, A. E. Miklos, M. L. Matthews, J. Seetharaman, M. Su, J. F. Hunt, B. F. Cravatt, D. Baker, Design of activated serine-containing catalytic triads with atomic-level accuracy. *Nat. Chem. Biol.* **10**, 386–391 (2014).
6. F. Richter, R. Blomberg, S. D. Khare, G. Kiss, A. P. Kuzin, A. J. T. Smith, J. Gallaher, Z. Pianowski, R. C. Helgeson, A. Grjasnow, R. Xiao, J. Seetharaman, M. Su, S. Vorobiev, S. Lew, F. Forouhar, G. J. Kornhaber, J. F. Hunt, G. T. Montelione, L. Tong, K. N. Houk, D. Hilvert, D. Baker, Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. *J. Am. Chem. Soc.* **134**, 16197–16206 (2012).
7. L. Giger, S. Caner, R. Obexer, P. Kast, D. Baker, N. Ban, D. Hilvert, Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **9**, 494–498 (2013).

8. N. H. Joh, T. Wang, M. P. Bhate, R. Acharya, Y. Wu, M. Grabe, M. Hong, G. Grigoryan, W. F. DeGrado, De novo design of a transmembrane Zn^{2+}-transporting four-helix bundle. *Science*. **346**, 1520–1524 (2014).
9. A. R. Thomson, C. W. Wood, A. J. Burton, G. J. Bartlett, R. B. Sessions, R. L. Brady, D. N. Woolfson, Computational design of water-soluble α-helical barrels. *Science*. **346**, 485–488 (2014).
10. L. Doyle, J. Hallinan, J. Bolduc, F. Parmeggiani, D. Baker, B. L. Stoddard, P. Bradley, Rational design of α-helical tandem repeat proteins with closed architectures. *Nature*. **528**, 585–588 (2015).
11. A. J. Burton, A. Thomson, W. M. Dawson, R. Brady, D.N. Woolfson, Installing hydrolytic activity into a completely de novo protein framework. *Nat. Chem.* **8**, 837-844 (2016).
12. N. Koga, R. Tatsumi-Koga, G. Liu, R. Xiao, T. B. Acton, G. T. Montelione, D. Baker, Principles for designing ideal protein structures. *Nature*. **491**, 222–227 (2012).
13. P.-S. Huang, G. Oberdorfer, C. Xu, X. Y. Pei, B. L. Nannenga, J. M. Rogers, F. DiMaio, T. Gonen, B. Luisi, D. Baker, High thermodynamic stability of parametrically designed helical bundles. *Science*. **346**, 481–485 (2014).
14. Y. Lin, N. Koga, R. Tatsumi-Koga, G. Liu, A. F. Clouser, G. T. Montelione, D. Baker, Control over overall shape and size in de novo designed proteins. *Proc. Natl. Acad. Sci.* **112**, E5478–E5485 (2015).
15. T. Brunette, F. Parmeggiani, P.-S. Huang, G. Bhabha, D. C. Ekiert, S. E. Tsutakawa, G. L. Hura, J. A. Tainer, D. Baker, Exploring the repeat protein universe through computational protein design. *Nature*. **528**, 580–584 (2015).
16. P.-S. Huang, K. Feldmeier, F. Parmeggiani, D. A. Fernandez Velasco, B. Höcker, D. Baker, De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat. Chem. Biol.* **12**, 29–34 (2016).
17. T. M. Jacobs, B. Williams, T. Williams, X. Xu, A. Eletsky, J. F. Federizon, T. Szyperski, B. Kuhlman, Design of structurally distinct proteins using strategies inspired by evolution. *Science*. **352**, 687–690 (2016).
18. J. S. Richardson, E. D. Getzoff, D. C. Richardson, The beta bulge: a common

small unit of nonrepetitive protein structure. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 2574–2578 (1978).

19. A. W. Chan, E. G. Hutchinson, D. Harris, J. M. Thornton, Identification, classification, and analysis of beta-bulges in proteins. *Protein Sci*. **2**, 1574–1590 (1993).
20. C. Chothia, Coiling of beta-pleated sheets. *J. Mol. Biol.* **163**, 107–117 (1983).
21. F. R. Salemme, Structural properties of protein beta-sheets. *Prog. Biophys. Mol. Biol.* **42**, 95–133 (1983).
22. A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y. E. A. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker, P. Bradley, Rosetta3: An object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
23. Materials and methods are available as supplementary materials at the *Science* website.
24. B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, D. Baker, Design of a novel globular protein fold with atomic-level accuracy. **302**, 1364–1369 (2003).
25. P. Craveur, A. P. Joseph, J. Rebehmed, A. G. De Brevern, β-Bulges: Extensive structural analyses of β-sheets irregularities. *Protein Sci.* **22**, 1366–1378 (2013).
26. K. Fujiwara, S. Ebisawa, Y. Watanabe, H. Fujiwara, M. Ikeguchi, The origin of β-strand bending in globular proteins. *BMC Struct. Biol.* **15**, 21 (2015).
27. C. A. Rohl, C. E. M. Strauss, K. M. S. Misura, D. Baker, Protein Structure Prediction Using Rosetta. *Methods Enzymol.* **383** (2004), pages 66–93.
28. P. Bradley, K. M. S. Misura, D. Baker, Toward high-resolution de novo structure prediction for small proteins. *Science*. **309**, 1868–1871 (2005).
29. S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25** (1997), pages 3389–3402.

30. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: architecture and applications. *BMC Bioinformatics*. **10**, 421 (2009).
31. Y. Zhang, J. Skolnick, TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
32. J. S. Richardson, D. C. Richardson, Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 2754–2759 (2002).
33. W. Kabsch, C. Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. **22**, 2577–2637 (1983).
34. R. T. Wintjens, M. J. Rooman, S. J. Wodak, Automatic classification and analysis of alpha alpha-turn motifs in proteins. *J. Mol. Biol.* **255**, 235–253 (1996).
35. G. Wang, R. L. Dunbrack, PISCES: A protein sequence culling server. *Bioinformatics*. **19**, 1589–1591 (2003).
36. S. J. Fleishman, A. Leaver-Fay, J. E. Corn, E. M. Strauch, S. D. Khare, N. Koga, J. Ashworth, P. Murphy, F. Richter, G. Lemmon, J. Meiler, D. Baker, Rosettascripts: A scripting language interface to the Rosetta Macromolecular modeling suite. *PLoS One*. **6** (2011), doi:10.1371/journal.pone.0020161.
37. B. Kuhlman, D. Baker, Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci.* **97**, 10383–10388 (2000).
38. A. Leaver-Fay, M. J. O'Meara, M. Tyka, R. Jacak, Y. Song, E. H. Kellogg, J. Thompson, I. W. Davis, R. A. Pache, S. Lyskov, J. J. Gray, T. Kortemme, J. S. Richardson, J. J. Havranek, J. Snoeyink, D. Baker, B. Kuhlman, Scientific benchmarks for guiding macromolecular energy function improvement. *Methods Enzymol.* **523**, 109–143 (2013).
39. M. J. O'Meara, A. Leaver-Fay, M. D. Tyka, A. Stein, K. Houlihan, F. Dimaio, P. Bradley, T. Kortemme, D. Baker, J. Snoeyink, B. Kuhlman, Combined covalent-electrostatic model of hydrogen bonding improves structure prediction with Rosetta. *J. Chem. Theory Comput.* **11**, 609–622 (2015).
40. W. Sheffler, D. Baker, RosettaHoles2: A volumetric packing measure for protein

structure refinement and validation. *Protein Sci.* **19**, 1991–1995 (2010).

41. D. T. Jones, Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).
42. D. B. J. A. Fallas, G. Ueda, W. Sheffler, V. Nguyen, D. E. McNamara, B. Sankaran, J. H. Pereira, F. Parmeggiani, TJ Brunette, D. Cascio, T. R. Yeates, P. Zwart, Computational design of self-assembling cyclic protein homooligomers. *Nat. Chem.* (2016).
43. The PyMOL Molecular Graphics System, version 1.7.2. (Schrödinger, LLC, New York, 2016).
44. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
45. T. B. Acton, R. Xiao, S. Anderson, J. Aramini, W. A. Buchwald, C. Ciccosanti, K. Conover, J. Everett, K. Hamilton, Y. J. Huang, H. Janjua, G. Kornhaber, J. Lau, D. Y. Lee, G. Liu, M. Maglaqui, L. Ma, L. Mao, D. Patel, P. Rossi, S. Sahdev, R. Shastry, G. V. T. Swapna, Y. Tang, S. Tong, D. Wang, H. Wang, L. Zhao, G. T. Montelione, Preparation of protein samples for NMR structure, function, and small-molecule screening studies. *Methods Enzymol.* **493**, 21–60 (2011).
46. R. Xiao, S. Anderson, J. Aramini, R. Belote, W. A. Buchwald, C. Ciccosanti, K. Conover, J. K. Everett, K. Hamilton, Y. J. Huang, H. Janjua, M. Jiang, G. J. Kornhaber, D. Y. Lee, J. Y. Locke, L. C. Ma, M. Maglaqui, L. Mao, S. Mitra, D. Patel, P. Rossi, S. Sahdev, S. Sharma, R. Shastry, G. V. T. Swapna, S. N. Tong, D. Wang, H. Wang, L. Zhao, G. T. Montelione, T. B. Acton, The high-throughput protein sample production platform of the Northeast Structural Genomics Consortium. *J. Struct. Biol.* **172**, 21–33 (2010).
47. M. Jansson, Y.-C. Li, L. Jendeberg, S. Anderson, G. Montelione, B. Nilsson, High-level production of uniformly 15N-and 13C-enriched fusion proteins in Escherichia coli. *J. Biomol. NMR*. **7**, 131–141 (1996).
48. M. M. Santoro, D. W. Bolen, A test of the linear extrapolation of unfolding free energy changes over an extended denaturant concentration range. *Biochemistry*. **31**, 4901–4907 (1992).

49. J. M. Scholtz, G. R. Grimsley, C. N. Pace, in *Methods in enzymology* (2009; http://dx.doi.org/10.1016/S0076-6879(09)66023-7), vol. 466, pages 549–565.

50. C. D. Geierhaas, A. a Nickson, K. Lindorff-Larsen, J. Clarke, M. Vendruscolo, BPPred: A Web-based computational tool for predicting biophysical parameters of proteins. *Protein Sci.* **16**, 125–134 (2006).

51. P. Rossi, G. V. T. Swapna, Y. J. Huang, J. M. Aramini, C. Anklin, K. Conover, K. Hamilton, R. Xiao, T. B. Acton, A. Ertekin, J. K. Everett, G. T. Montelione, A microscale protein NMR sample screening pipeline. *J. Biomol. NMR*. **46** (2010), pages 11–22.

52. D. Neri, T. Szyperski, G. Otting, H. Senn, K. Wüthrich, 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 13C labeling. *Biochemistry*. **28**, 7510–7516 (1989).

53. G. H. Liu, Y. Shen, H. S. Atreya, D. Parish, Y. Shao, D. K. Sukumaran, R. Xiao, A. Yee, A. Lemak, A. Bhattacharya, T. A. Acton, C. H. Arrowsmith, G. T. Montelione, T. Szyperski, NMR data collection and analysis protocol for high-throughput protein structure determination. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10487–10492 (2005).

54. A. Bhattacharya, R. Tejero, G. T. Montelione, Evaluating protein structures determined by structural genomics consortia. *Proteins Struct. Funct. Genet.* **66**, 778–795 (2007).

55. Y. J. Huang, R. Powers, G. T. Montelione, Protein NMR recall, precision, and F-measure scores (RPF scores): Structure quality assessment measures based on information retrieval statistics. *J. Am. Chem. Soc.* **127**, 1665–1674 (2005).

56. J. R. Luft, R. J. Collins, N. A. Fehrman, A. M. Lauricella, C. K. Veatch, G. T. DeTitta, A deliberate approach to screening for initial crystallization conditions of biological macromolecules. *J. Struct. Biol.* **142**, 170–179 (2003).

57. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

58. P. H. Zwart, P. V Afonine, R. W. Grosse-Kunstleve, L.-W. Hung, T. R. Ioerger, A.

J. McCoy, E. McKee, N. W. Moriarty, R. J. Read, J. C. Sacchettini, N. K. Sauter, L. C. Storoni, T. C. Terwilliger, P. D. Adams, Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* **426**, 419–435 (2008).

59. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).

60. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

61. A. Roy, J. Yang, Y. Zhang, COFACTOR: An accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res.* **40** (2012), doi:10.1093/nar/gks372.

62. N. R. Voss, M. Gerstein, 3V: Cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.* **38** (2010), doi:10.1093/nar/gkq395.

63. D. A. Snyder, Y. Chen, N. G. Denissova, T. Acton, J. M. Aramini, M. Ciano, R. Karlin, J. Liu, P. Manor, P. A. Rajan, P. Rossi, G. V. T. Swapna, R. Xiao, B. Rost, J. Hunt, G. T. Montelione, Comparisons of NMR spectral quality and success in crystallization demonstrate that NMR and X-ray crystallography are complementary methods for small protein structure determination. *J. Am. Chem. Soc.* **127**, 16505–16511 (2005).

64. R. Tejero, D. Snyder, B. Mao, J. M. Aramini, G. T. Montelione, PDBStat: A universal restraint converter and restraint analysis software package for protein NMR. *J. Biomol. NMR.* **56**, 337–351 (2013).

65. S. G. Hyberts, M. S. Goldberg, T. F. Havel, G. Wagner, The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci.* **1**, 736–751 (1992).