JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

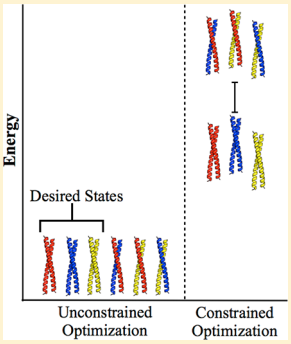
A Set of Computationally Designed Orthogonal Antiparallel Homodimers that Expands the Synthetic Coiled-Coil Toolkit

Christopher Negron[†] and Amy E. Keating*[,†,‡]

[†]Program in Computational and Systems Biology and [‡]Departments of Biology and Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 021393, United States

Supporting Information

ABSTRACT: Molecular engineering of protein assemblies, including the fabrication of nanostructures and synthetic signaling pathways, relies on the availability of modular parts that can be combined to give different structures and functions. Currently, a limited number of well-characterized protein interaction components are available. Coiled-coil interaction modules have been demonstrated to be useful for biomolecular design, and many parallel homodimers and heterodimers are available in the coiled-coil toolkit. In this work, we sought to design a set of orthogonal antiparallel homodimeric coiled coils using a computational approach. There are very few antiparallel homodimers described in the literature, and none have been measured for cross-reactivity. We tested the ability of the distance-dependent statistical potential DFIRE to predict orientation preferences for coiled-coil dimers of known structure. The DFIRE model was then combined with the CLASSY multistate protein design framework to engineer sets of three orthogonal antiparallel homodimeric coiled coils. Experimental measurements confirmed the successful design of three peptides that preferentially formed antiparallel homodimers that, furthermore, did not interact with one additional previously reported antiparallel homodimer. Two designed peptides that formed higher-order structures suggest how future design protocols could be improved. The successful designs represent a significant expansion of the existing protein-interaction toolbox for molecular engineers.



INTRODUCTION

Modular design is used for engineering complex devices in electronics, mechanics, nanotechnology and other fields. Recently, biologists have begun to exploit modular parts as a way to build novel synthetic biological systems.[1] Many types of parts are required to implement diverse structural, binding and catalytic functions. Here, we focus on the α-helical coiled coil, which is a protein-interaction domain highly suitable for inclusion in the growing molecular parts toolkit.[2,3] Coiled coils are prevalent in native proteins and are useful interaction motifs due to their capacity to encode complex interaction patterns in a short protein sequence.[4−6]

Coiled coils form a rod-like structure composed of α-helices that wrap around each other with a superhelical twist. Coiled-coil sequences have a characteristic motif commonly referred to as a heptad repeat, denoted as $[abcdefg]_n$. The *a* and *d* positions are dominated by hydrophobic residues, and are found at the core of the structure; we refer to *a* and *d* positions as core positions in this work. In coiled-coil dimers, *e* and *g* positions are typically occupied by charged residues and form the boundary between the core and the surface of the coiled coil. The *b*, *c*, and *f* positions are located on the surface and are most often polar or charged. In coiled-coil notation, a prime indicates a residue on an opposing chain. For example, positions *e* and *g*′ are proximal and can form interhelical salt bridges in parallel coiled-coil dimers, whereas *e*/*e*′ and *g*/*g*′ pairs can interact in antiparallel coiled coils.

The relationship between coiled-coil sequence and structure is incompletely understood, even after decades of study of native, mutant and de novo-designed coiled coils. This is partly due to the many topologies accessible to coiled-coil sequences. For example, coiled coils can fold into dimers, trimers, tetramers, and even higher-order oligomers. Additionally, oligomers can be homo- or heteroassemblies. Lastly, the orientations (parallel vs antiparallel) and axial alignments of the constituent helices can vary.[7,8] The general problem of predicting detailed coiled-coil structure from sequence has not been solved, although progress has been made developing methods to predict oligomerization state from sequence, and in particular to discriminate parallel dimers from parallel trimers.[9−14]

Coiled coils have been used in a wide range of applications. They have been applied to the design of artificial transcription factors and used to manipulate cell-signaling pathways.[15,16] They have also been used to build engineered crystals, and to modulate the charge-transfer properties of electronic devices.[17,18] In many of these studies, controlling the orientation of the helices in the coiled coil was important. For example, Shlizerman et al. modulated the conductance between two monolayers of gold using coiled-coil dimers and showed that parallel and antiparallel coiled coils differentially impacted the electronic properties of the system. Coiled coils of different

Received: July 31, 2014
Published: October 22, 2014

   dx.doi.org/10.1021/ja507847t | J. Am. Chem. Soc. 2014, 136, 16544−16556

orientations have net molecular dipoles of different magnitude and direction, and can thereby confer different electronic properties.[18]

Recently, an exciting strategy was developed to design polypeptide polyhedra based around coiled-coil dimers. Gradišar et al. used a set of parallel and antiparallel dimeric coiled coils as building blocks to engineer a nanoscale single-chain tetrahedron with coiled coils forming each edge.[19] The design strategy involved concatenating a series of 12 sequence segments coding for different coiled-coil helices into a single chain. The artificial protein sequence was designed such that folding of the chain, driven by pairing each coiled-coil helix with its appropriate intrachain partner helix, would generate a prespecified three-dimensional structure. A crucial aspect of the design strategy was the use of coiled-coil components that were orthogonal to one another, i.e., that had low potential to cross-interact. The designed tetrahedron was based on 4 parallel and 2 antiparallel coiled-coil dimers previously reported in the literature.[20−23] As part of their work, the authors computed the number and type of coiled coils that would be needed to build different polyhedra. Interestingly, they found that most polyhedra require orthogonal antiparallel and parallel dimers. For example, of the 6 polyhedra considered by the authors, only an octahedron could be built without using antiparallel dimers.

Despite the clear benefits of having reagents that allow manipulation of orientation in a molecular assembly, most designed coiled coils adopt a parallel orientation. Very few antiparallel coiled-coil dimers have been characterized or designed, and none have been tested for orthogonality. In contrast, dozens of native and synthetic parallel coiled coils have been tested for interactions and orthogonality.[6,23,24] There are currently two databases maintained for designed coiled coils, the SYNZIP database, and the *Pcomp* database.[2,3] Currently 96% of the SYNZIP sequences and ∼63% of the sequences in the *Pcomp* database form parallel dimers. Between these two databases, the biophysical properties of only one antiparallel coiled coil (a heterodimer) are reported.[2] Thus, designing sets of orthogonal antiparallel homodimers would expand the available coiled-coil parts in a meaningful way.

Because coiled-coil sequences can encode many different structures, negative design to destabilize undesired states is crucial when making peptides intended to assemble into a single topology.[25] Several negative design strategies have been used in the past that involve placing charged, beta-branched or polar asparagine residues such that they form unfavorable interactions in undesired states.[26−28] A recent study relied on all three of these strategies to design a parallel homodimer, homotrimer, and homotetramer.[3] The orientations of the helices were engineered by placing lysines at all *e* positions and glutamates at all *g* positions, which leads to electrostatic attraction in parallel assemblies but repulsion in antiparallel states. Oligomerization states were specified by the differential placement of beta-branched residues in core *a* and *d* heptad positions, a strategy first discovered by Harbury et al., and by the use of asparagine residues to specify dimer formation, which was originally reported by Lumb and Kim.[27,28] Including charged residues in core *a* or *d* positions has also been observed to destabilize nondimer states.[29]

Designing sets of orthogonal coiled-coil homodimers presents additional challenges related to encoding interaction specificity. This is due to the increased number of undesired, off-target states associated with forming hetero-oligomeric species. The number of possible hetero species increases dramatically as the number of designed orthogonal coiled coils grows, such that three orthogonal antiparallel homodimers have the potential to form six possible off-target parallel or antiparallel heterodimers; other undesired structures are also possible. To design sets of orthogonal antiparallel coiled-coil dimers, we therefore turned to computational methods to keep track of the numerous desired and undesired structures in this design problem.

Despite the many successes of structure-based approaches for modeling and designing protein−protein interactions, treating multiple states is difficult with these techniques.[30,31] The computational costs of modeling each structure can be large, and current optimization functions used with structure-based models do not provide efficient routines for optimizing one set of states while simultaneously destabilizing many off-target states. The multistate design framework CLASSY addresses these issues by carrying out design in protein sequence space, without the need to explicitly model all protein structures.[32,33] By using a transformation of structure-based models to sequence-based models, CLASSY addresses both the search and scoring problems of multistate design, and the method has previously been applied to design parallel coiled coils specific for binding to a target in preference to closely related off-target proteins.[32,34,35]

This paper describes our work applying CLASSY in conjunction with the DFIRE[36] statistical potential to the de novo design of sets of coiled coils consisting of three orthogonal antiparallel homodimers. We designed two sets of three proteins, and used biophysical techniques to determine the oligomerization state, helix orientation and thermal stability of structures formed by the designed sequences. Some designed peptides formed trimers or higher-order assemblies, but we identified 3 peptides (APH2, APH3, and APH4) that formed orthogonal antiparallel homodimers. In addition, we showed that these proteins homodimerize in preference to binding to APH, a previously reported antiparallel homodimer.[21] Thus, we provide evidence for four sequences that preferentially form antiparallel homodimers that can be used for protein engineering applications.

MATERIALS AND METHODS

Building and Scoring Structures with DFIRE*. As described in detail below, side chains were modeled on idealized coiled-coil backbones using Rosetta and scored using DIFIRE*, a modified version of the DFIRE statistical potential. To construct libraries of parallel and antiparallel backbones, a set of 214 canonical coiled coils (i.e., left-handed coiled coils with uninterrupted heptad registers, *abcdefg*) with 2 helices each longer then 20 residues were culled from the CC+ database as of August 18, 2010.[37] Within the parallel and antiparallel sets, examples were filtered to have ≤50% sequence identity. This set of structures is referred to as the filtered CC+ set. Seven geometrical parameters defined by Crick to describe a coiled coil were fit to each structure using the CCCP Structure Fitter.[38,39] This set of backbones was then filtered to give 25 parallel and 23 antiparallel backbones with parameters within one standard deviation of the average value for each parameter. Averages and standard deviations are reported in Table S1 (Supporting Information). Idealized versions of these 48 structures were generated using the CCCP Structure Generator.[39] Coiled-coil sequences to be scored were modeled on each idealized backbone using the fixed-backbone packing protocol of Rosetta 3.2.[40] The soft-potential flag and expansion of the first and second dihedral angles of the rotamer library were used, along with the side-chain minimization flag. All surface heptad positions (*b*, *c*, and *f*) were modeled as alanine. Structures were scored using a modified version of DFIRE, a distance-dependent pairwise statistical

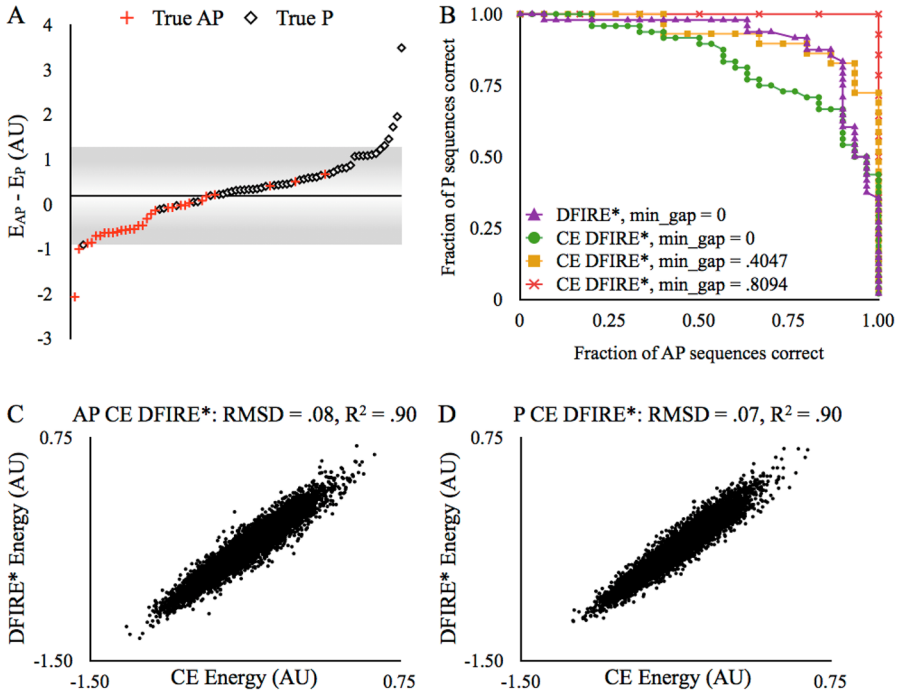


Figure 1. Predicting coiled-coil orientation preference and testing cluster-expanded DFIRE*. (A) E_{AP} and E_P are the antiparallel (AP) and parallel (P) DFIRE* energies for each orientation test set coiled coil. Antiparallel or parallel coiled coils (according to PDB structure) are plotted with red crosses or black diamonds, respectively. The line at $E_{AP} - E_P = 0.18$ AU gives optimal separation of parallel and antiparallel examples. Min_gap was used to remove examples with small DFIRE* orientation preferences (see text); shading indicates increasing min_gap from the line of optimal separation. (B) The fraction of antiparallel sequences predicted correctly vs the fraction of parallel sequences predicted correctly, as the cutoff value for $E_{AP} - E_P$ was changed, is plotted for DFIRE* and the CE model of DFIRE*. Curves for data sets with different values of min_gap are shown for the CE model of DFIRE*. (C, D) DFIRE* energies vs the CE model of DFIRE* energies for randomly generated dimer-like test structures in the antiparallel (C) and parallel (D) states.

potential based on the distance-scaled, finite ideal-gas reference state.[36] Two modifications were made to the published energy function. The cutoff distance, r_{cut}, was set to 5.8 Å, and interatomic energies were evaluated only between residues on opposite helices in the coiled coil. We refer to this modified version of DFIRE as DFIRE*. DFIRE* outperforms DFIRE on certain interaction prediction tests for parallel coiled coils (V. Potapov, personal communication). The lowest DFIRE* energy for each sequence over all 25 parallel or 23 antiparallel backbones was used as the parallel or antiparallel energy, respectively.

Deriving Cluster Expansion Models. Two cluster-expanded functions based on DFIRE* were derived to score the propensity of sequences to form antiparallel and parallel coiled coils. For an outline of the protocol, see Figure S1, and for an in-depth discussion of performing cluster-expansion calculations using CLEVER 1.0 see Negron et al.[33] In the present application, the cluster-expanded models express energy as a sum of terms corresponding to weights for single amino acids at *a*, *d*, *e*, and *g* heptad positions and pairs of amino acids at these positions. As in Grigoryan et al., only pairs of positions within the same or adjoining heptads were considered.[41] Weights were fit to training data using the CLEVER 1.0 package.[33,42] The training data consisted of DFIRE* energies for a central two-heptad unit within a six-heptad structure, calculated using the scoring protocol described in the previous section for 30 000 sequences. Another 8000 sequences, nonoverlapping with the training set, were generated in the same way to be used as a test set. Training sequences were 42 residues (six heptads) long and composed of a repeating two-heptad unit. Training sequences were generated randomly but with heptad-specific single-residue frequencies matching those of known coiled-coil dimers (both parallel and antiparallel). Antiparallel frequencies were obtained from antiparallel structures in the filtered CC+ set.[37] Parallel frequencies were obtained from the NPS database.[14] Once determined, cluster expansion (CE) weights can be used to score antiparallel and parallel coiled-coil dimers of arbitrary length.

Orientation Test Set. Examples of parallel and antiparallel coiled coils were obtained from the filtered CC+ set and further filtered to exclude those shorter than 28 residues and those that contained non-natural amino acids. For certain sequences, three residues at the terminal ends of the two chains were removed so that the two chains fully overlapped in both the parallel and antiparallel orientations; i.e., the coiled coils that were modeled were blunt-ended in both orientations. The final orientation test set contained 30 antiparallel complexes (composed of ~285 heptads) and 48 parallel complexes (composed of ~547 heptads). PDB IDs with chain and residue numbers for the orientation test set are given in Table S2.

CLASSY Peptide Design. A detailed description of how integer linear programming (ILP) can be applied as part of the CLASSY multistate design method is given in Negron et al.[33] In this work, the objective function for ILP was the total energy (E^T), given by the sum of the energies of three antiparallel homodimers ($E^T = E^1 + E^2 + E^3$). All energies were obtained from either the antiparallel or parallel cluster-expanded models. The ILP solver of the IBM ILOG CPLEX optimizer was used to minimize this objective function under a set of constraints.[43] The constraints included energy gaps to off-target dimer states (see Figure 2A,B), as well as constraints on the number of polar residues allowed at *a* and *d* heptad positions (maximum of 2 charged residues at *a*, and 1 Lys residue at *d* per design sequence). A constraint was included on the energy gap between every antiparallel homodimer and every off-target state (of those types considered in the calculation) that the constituent peptide could participate in. The constraints were of the form $E^{OT} - E^x > \Delta$, where E^{OT} represents the energy of a single off-target state, of which there were several as shown in Figure 2. E^x represents the energy of a single antiparallel homodimer, i.e., E^1, E^2, or E^3. Δ is a user-defined specificity gap, and different values of Δ were used as shown in Figure 2C,D. A solution, consisting of three sequences, was obtained for each Δ. Two sets of design calculations were done, one including glutamate as an option at *a* positions (sequence space 1) and one not allowing glutamate (sequence space

dx.doi.org/10.1021/ja507847t | J. Am. Chem. Soc. 2014, 136, 16544−16556

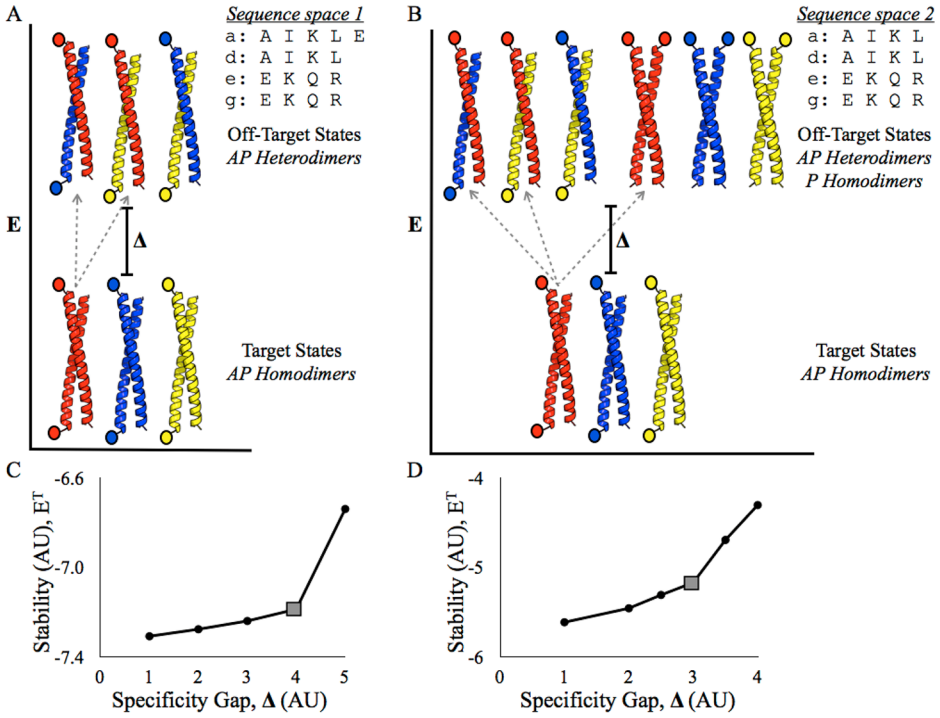


Figure 2. Computational design of orthogonal antiparallel homodimers. (A, B) Diagram of target and off-target states included in two design calculations. Colors represent distinct sequences, and colored circles indicate the N-terminus of each helix. An energetic constraint, Δ, was enforced between the energy of each target antiparallel homodimer state (E^1, E^2, E^3) and every off-target state that peptide could participate in (examples shown with gray dashed lines). The sequence space used for each design is indicated. Different numbers of off-target states were included for sequence space 1 (A) vs sequence space 2 (B). (C, D) The total energy $E^T = E^1 + E^2 + E^3$ vs Δ is plotted for sequence space 1 (C) and sequence space 2 (D). Each value of Δ led to a set of optimized sequences, and the gray squares mark the solutions chosen for experimental testing.

2). One solution was chosen manually for experimental testing from each calculation, based on predicted stabilities and specificities.

Cloning, Protein Expression, and Purification. Synthetic genes encoding computationally designed coiled-coil sequences, and control sequences, were constructed by PCR amplification from two 258-base pair oligonucleotides and one 157-base pair oligonucleotide (gblocks) purchased from Integrated DNA Technologies. DNA sequences were codon optimized for expression in *Escherichia coli* using DNAWorks.[44] Low-frequency *E. coli* codons selected by DNAWorks were manually switched with synonymous high-frequency codons.

Following amplification with primers to provide appropriate vector overlap, Gibson cloning (New England Biolabs) was used to clone synthetic genes into pENTR vectors. The products of the Gibson reactions were then recombined into pMAL (New England Biolabs) destination vectors using LR Clonase II (Invitrogen) in 2.5 μL reactions. pMAL encodes MBP followed by a TEV protease cleavage site (not used), a Gateway linker region, and a C-terminal His_6 tag. The LR Clonase II reaction inserted the synthetic gene between the Gateway linker region and the C-terminal His_6 site. The pMAL vectors were transformed into BL21 (DE3) cells (Agilent). BL21 cells were grown in liquid LB cultures (1 L) at 37 °C to an OD_{600} of ∼0.4−0.6. Protein expression was then induced with 1 mM IPTG for 4.5−5.5 h. Cells were pelleted, resuspended, and then lysed by sonication. MBP-fused proteins were purified from the supernatant using NiNTA (Qiagen) column purification under native conditions. The elution buffer contained 0.3 M imidazole, 20 mM Tris base, and 0.5 M NaCl at a pH of 7.91. The approximate sizes of MBP-fused proteins were confirmed using protein gels with size ladders.

A second set of constructs was made by amplifying from gblocks using primers encoding a cysteine either at the N-terminal or C-terminal end, as well as flanking *Bam*HI/*XhoI* restriction sites. The genes were cloned by means of the *Bam*HI/*XhoI* restriction sites into a modified version of the pDEST17 vector. This vector encodes an N-terminal His_6 tag as well as a GESKEYKKGSGS linker shown to improve the solubility of recombinant proteins.[34] Cysteine-containing constructs were expressed in RP3098 cells grown, induced and lysed as described above for BL21. However, these proteins were purified from the supernatant using NiNTA (Qiagen) under denaturing conditions. The elution buffer consisted of 60% acetonitrile (HPLC-grade) and 0.1% trifluoroacetic acid (TFA). Ni-affinity purification was followed by reverse-phase HPLC with a water/acetonitrile gradient in the presence of 0.1% TFA. Masses were confirmed by MALDI-TOF mass spectrometry.

Concentrations of all constructs were determined using the Edelhoch method, measuring UV absorbance of aromatic residues at 280 nm in 6 M guanidinium chloride.[45] Amino-acid sequences of all constructs are given in Table S3.

Sedimentation Equilibrium Analytical Ultracentrifugation. Proteins were dialyzed with three changes of reference buffer (40 mM Tris base, 150 mM NaCl, pH 7.91) over the course of 24 h. Sedimentation equilibrium runs were performed with a Beckman XL-I analytical ultracentrifuge using an An-50 Ti rotor at 20 °C. Proteins were spun at three speeds and at least two protein concentrations. Constructs fused to MBP were spun at concentrations ranging from 4 to 40 μM at 10 200, 16 300 and 20 400 rpm. These spins were monitored either using UV absorbance at 280 nm, or with interference optics when multiple MBP constructs were mixed. For protein constructs containing cysteine, 1 mM TCEP was added to the reference buffer prior to dialysis. These constructs were spun at concentrations of 20 and 40 μM at 28 000, 35 000 and 42 000 rpm and monitored using interference optics. For each speed, equilibrium was confirmed by negligible differences between the sample distributions in the cells over sequential scans. Data sets for each construct were globally fit to a model for a single ideal species using the program SEDPHAT.[46,47] Values for v-bar, solvent density, and viscosity were obtained from SEDNTERP.[48]

Disulfide-Exchange Experiments. Cysteine-containing proteins in varying states of oxidation/reduction (depending on construct) were placed in a redox buffer (500 μM reduced glutathione, 250 μM oxidized glutathione, 40 mM Tris base, 150 mM NaCl, pH 7.91) at 20

dx.doi.org/10.1021/ja507847t | J. Am. Chem. Soc. 2014, 136, 16544−16556

μM of each protein at room temperature. Redox reactions were quenched at different time points using a drop of 6 M hydrochloric acid. The products of the reactions were then run on an analytical Vydac C_{18} reverse-phase column with absorbance monitored at 220 nm using a linear water/acetonitrile gradient containing 0.1% TFA. Equilibrium was confirmed by monitoring changes in HPLC profiles as a function of time. Retention times for the reduced proteins and for the oxidized states for each of the 6 cysteine-containing proteins were assigned by HPLC analysis of the constructs in TBS (40 mM Tris base, 150 mM NaCl, pH 7.91) alone, in TBS with TCEP added for an incubation time of 30 min (to generate the fully reduced species), or in TBS solution left exposed to air and stirring overnight (to generate the fully oxidized species). Glutathione adduct peaks were assigned by the appearance, following incubation in redox buffer, of a peak with a retention time not consistent with the reduced or oxidized states of each of the six individual protein constructs. Antiparallel peaks were assigned by monitoring the appearance of a peak that was only observed after mixing two constructs that encoded the same coiled coil, but with cysteine residues at opposing ends.

Circular Dichroism (CD) Spectroscopy. CD spectra and thermal-denaturation curves were measured on an AVIV 400 CD spectrometer. Peptides were equilibrated in PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.4) containing 1 mM of dithiothreitol (DTT) at ~25 °C for at least 1.5 h prior to measurement. Measurements were made in a 1 mm quartz cuvette at a protein concentration of 20 μM using the N-terminal cysteine-containing constructs. CD spectra were measured at 25 °C. For each sample, three wavelength scans were measured and then averaged. For each wavelength scan, data were collected from 190 to 280 nm, in 1 nm steps, averaging for 5 s at each wavelength. Thermal denaturation curves were generated by monitoring θ_{222} using a 30 s averaging time, 3 min equilibration time, and temperature increments of 2.5 °C from 0 to 98 °C. Melting temperatures, T_m, were obtained by fitting the change of the CD signal over the change in temperature.[32,49] Fitting was performed using the nonlinear least squares method in Matlab 7.8. The fractional helicity of each design was estimated by substituting the experimentally measured θ_{222} into the equation $(\theta_{222} - 3000)/(-36\,000 - 3000)$.[50]

RESULTS

Benchmarking DFIRE* on Orientation-Preference Prediction. Computational design of orthogonal antiparallel homodimers requires an energy function capable of scoring antiparallel vs parallel dimers. To assess whether our design energy function could predict helix orientation for coiled-coil dimers of known structure, we implemented a test similar to that in Apgar et al.[51] We created a database of 30 antiparallel and 48 parallel dimer structures based on the CC+ database of Testa et al.;[37] we refer to this database as the orientation test set (see Materials and Methods). The orientation test set in this study differed from that used by Apgar et al. due to its higher stringency on length, ≥28 residues vs ≥18 residues.[51] This more stringent cutoff has the effect of removing examples of short coiled-coil sequences embedded in large structures, for which the helix orientation is less likely to be determined by the sequence of the coiled-coil region alone. Furthermore, sequence features of antiparallel coiled coils in the PDB are a function of their lengths; e.g., shorter coiled coils have a 16% higher frequency of hydrophobic residues at the *g* position (Table S4).

A modified version of DFIRE, DFIRE*, which includes only interchain energy terms, was used for scoring. The orientation test-set sequences were modeled in both parallel and antiparallel orientations using Rosetta and scored using DFIRE*, as described in the Materials and Methods. The DFIRE* energy gap between the antiparallel and parallel state for each sequence is plotted in Figure 1A. We report energies in arbitrary units (AU), as we have no information at this time about how predicted energies from this procedure correlate with experimental free energies. The ability of DFIRE* to predict orientation preference on the test set was measured using the area under the curve (AUC) when plotting the fraction of parallel test-set sequences predicted correctly vs the fraction of antiparallel sequences predicted correctly, as a function of the score cutoff used to discriminate parallel from antiparallel sequences. As seen in Figure 1B, DFIRE* predicts orientation preference in this test with an AUC value of 0.91 (random predictions would result in an AUC of 0.5).

Cluster Expansion of DFIRE*. Cluster expansion (CE) is a computational method for generating a sequence-based scoring function that approximates energies calculated using structure-based techniques.[41,42,52] Once generated, a CE model eliminates the need for computationally costly structure building in protein design. Two CE models were built to approximate DFIRE* energies for antiparallel and parallel coiled-coil dimers, as described in the Materials and Methods, and the models were used to score 8000 test sequences (Figure 1C,D). Both models showed good correlation with DFIRE*, R^2 = 0.90, indicating that the approximation of structure-based modeling with a sequence-based function introduced relatively little error within the sequence space explored.

We benchmarked the orientation prediction performance of the CE DFIRE* models using the orientation test set. Every pair of sequences in the set was scored with the antiparallel CE model and the parallel CE model. The energy difference between the two CE models was used to predict the orientation preference of each sequence. The AUC value using the CE approximation of DFIRE* was 0.84 (Figure 1B), demonstrating that the faster, yet more approximate model gave reduced performance, as expected. However, the AUC value significantly improved as coiled coils with small energy gaps were removed from the orientation test set. For 44 coiled coils with the largest predicted differences in CE energy between parallel and antiparallel orientation (greater than 0.4047), the prediction performance (0.93) was similar to the performance of DFIRE* on the entire orientation test set. For 20 examples with predicted energy gaps greater than 0.8094, prediction performance was perfect. This information was used to set energy gap requirements for off-target states during the sequence-design stage of CLASSY.

Computational Design of Orthogonal Antiparallel Homodimers using CLASSY. CLASSY is a protein-design method that uses integer linear programming (ILP) to optimize a protein sequence using a CE scoring function. Importantly, the method allows a user to impose numerous constraints on the designed sequence. These can include constraints on sequence composition or properties (e.g., total charge). In multistate design, it is convenient to impose a constraint on the energy of a designed sequence adopting an undesired structure, to disfavor formation of that structure.

In our application, the antiparallel and parallel CE models were combined with ILP to do CLASSY design of six-heptad antiparallel homodimers. Designed antiparallel coiled coil APH is also six heptads long, and a four-heptad variant of APH had low thermal stability.[21] On the basis of this, we reasoned that six heptads should provide ample space to include specificity elements while maintaining a folded structure. Only residues at *a*, *d*, *e*, and *g* positions were designed; these residues are thought to be most critical for interaction specificity.[53,54] The *b*,

Table 1. Sequences of APH and Candidate Antiparallel Homodimers

Design	fgabcde fgabcde fgabcde fgabcde fgabcde fgabcde fgabc[a]
APH	KQLE KELKQLE KELQAIE KQLAQLQ WKAQARK KKLAQLK KKLQA
APH_i	KEEKQIE KELKQIE KELQAIE WRLAQLR KRLQALR KRKAQKR E
APH_{ii} (APH2)	KRLKQLE KRLKQLR KRKQAKR WEEAQIE KELQAIE KQLAQIR E
APH_{iii} (APH3)	KRKKQKR KRAKQLR KRLQALE WQLAQIR KELQAAE KEEAQIE E
APH_{iv}	KEKKQLR KELKQLE KELQALR WRLAQIE KRLQAIR KRLAQKE E
APH_v	KRLKQKE KRKKQLR KRLQALR WQLAQIE KELQAAE KEAAQLR E
APH_{vi} (APH4)	KQLKQIE KRLKQIE KRLQAKE WEKAQLR KELQALR KKLAQLR E

[a]Some sequences have two names, as described in the text. [b]Indicates the heptad register.

c, and *f* surface positions were taken from APH, which is one of the few characterized antiparallel homodimers reported in the literature. The surface of APH mainly consists of patterned glutamine and alanine residues at *b* and *c* positions, and lysine residues at *f* positions. This surface design has been used for both parallel and antiparallel coiled coils, and is thought to play a minimal role in interaction specificity.[21,26]

We used the CE model of DFIRE* to design the globally best-scoring antiparallel homodimer in a sequence space without cysteine, proline, or glycine and found that the designed sequence was highly charged and contained no hydrophobic residues in any heptad position. This peptide would not be expected to fold into a coiled-coil structure. The unrealistic design sequence is not inconsistent with the good performance of DFIRE* and the CE model of DFIRE* on the orientation prediction test above. In the orientation test, each of two compared structures had the same sequence. In contrast, without constraints on sequence composition, optimization using the CE model of DFIRE* had the freedom to build a sequence entirely from charged pairs that have highly favorable CE weights. The 20 most favorable weights in the CE DFIRE* model are all core-to-edge (i.e., *a* or *d* to *e* or *g*), or core-to-core charge−charge residue interactions. The weight of the most stabilizing hydrophobic−hydrophobic interaction is 2-fold weaker than the most stabilizing charge−charge interaction. To use CE DFIRE* in protein design, we therefore imposed constraints on the number of polar residues allowed at core heptad positions (see Materials and Methods) and restricted the design calculations to subsets of sequence space, as described below.

Two separate sequence spaces, sequence space 1 and sequence space 2, were chosen to search for antiparallel homodimer sequences (Figure 2A,B). Both sequence spaces included residues known to influence coiled-coil structural specificity through mechanisms such as electrostatic attraction/repulsion and beta-branch residue packing/clashing.[26,27] Sequence space 1 differed from sequence space 2 by the addition of glutamate as a choice at *a* positions. Statistics from the coiled-coil databases we analyzed show a 3-fold frequency enrichment of glutamate in *a* sites of antiparallel dimers relative to parallel dimers (Table S5); this difference has also been noted by Straussman et al.[55]

To design three noninteracting coiled coils, we optimized the sum of the CE energies of three antiparallel homodimers using CLASSY. Constraints were added to allow no more than two hydrophilic residues at *a* positions and no more than one at *d* positions. This maintained the hydrophobicity of the design solutions at these positions close to that of known antiparallel dimers of lengths greater than four heptads. Constraints were also placed on the predicted energies of competing states. In particular, all design calculations treated all three possible antiparallel heterodimer states as undesired states. Without these constraints, the global energy minimum would correspond to three copies of the lowest-energy antiparallel homodimer. Constraints on the off-target states were imposed as an energy gap by requiring the energy of each antiparallel homodimer to be lower than the energy of each of the off-target states that sequence could participate in, by a fixed amount (Figure 2).

CLASSY design was done iteratively, by progressively increasing the energy gap that was imposed between the target antiparallel homodimers and off-target antiparallel heterodimer states. As the gap to off-target states increased, the total predicted stability of the three antiparallel homodimers decreased (Figure 2C,D). This type of stability-specificity trade-off has been observed previously in the case of parallel dimer design using CLASSY.[32] Two sets of solutions, one from each of the sequence spaces, were rationally chosen based on good stability-specificity trade-offs. The designs in sequence space 1 are referred to as APH_i, APH_{ii}, and APH_{iii}. The designs in sequence space 2 are referred to as $APH_{iv'}$, $APH_{v'}$, $APH_{vi'}$. For each set of designed sequences, parallel and antiparallel homo- and heterodimer states were scored with the original DFIRE* structure-based model to predict relative energies of target and off-target structures. For the antiparallel homodimers designed in sequence space 1, the predicted energies of all parallel and antiparallel off-target dimers were much higher than the predicted energies for the antiparallel homodimers. The smallest gap, of 0.77 AU, was between the antiparallel homodimer state of APH_{iii} and a parallel heterodimer consisting of APH_{iii} and APH_i (Figure S2A). The APH_i antiparallel homodimer gap to this state was 1.13 AU. At gaps of this magnitude, DFIRE* predicts the orientation preference of native sequences with an AUC = 1.0. Thus, no additional states were added to the optimization protocol for sequence space 1. For sequence space 2, we observed that one of the parallel homodimers was predicted to be lower in energy than the corresponding antiparallel homodimer (Figure S2B). Furthermore, other parallel homodimer states were closer in energy to the antiparallel homodimers than when design was done in sequence space 1. To address this, we added parallel homodimer states as off-target states in the optimization protocol used for sequence space 2, and chose a new set of solutions in that space. The final six designed sequences are

shown in Table 1, with APH_i, APH_{ii} and APH_{iii} resulting from design in sequence space 1, and APH_{iv}, APH_v and APH_{vi} from design in sequence space 2. The two sets of designed sequences were also scored for cross-reactivity using DFIRE*. Predicted energies for all parallel and antiparallel heterodimers that could be formed between sets were significantly larger than predicted energies for the antiparallel homodimer states, with the smallest energy gap of 0.61 AU between the antiparallel and parallel homodimer states of APH_{iv}.

Oligomerization States of Designs. The molecular weights of complexes formed by designed peptides APH_i–APH_{vi} were determined using sedimentation equilibrium analytical ultracentrifugation (see Materials and Methods). We anticipate that the APH coiled coils will be used as fusion proteins in many applications, so we did two sets of experiments: one in which the peptides were fused to maltose binding protein (MBP) and one in which they were not. The results are shown in Table 2. The data for two designed peptides, APH_{iii} and APH_{vi}, were consistent with these peptides forming homodimers. APH_i was determined to have a molecular weight greater than that expected for a dimer, and no further data were collected on this construct. Single-species fits to APH_{ii} and APH_{iv} gave molecular weights less than and greater than what was expected for a dimer, respectively. APH_{ii} and APH_{iv} were retested at higher concentrations to stabilize higher-order states. At 20 μM, APH_{ii} formed a homodimer, whereas APH_{iv} formed a homotrimer. Further experiments were carried out only on designs APH_{ii}, APH_{iii} and APH_{vi}, which we renamed APH2, APH3 and APH4, respectively (see Table 2).

Table 2. Molecular Weights Determined by Analytical Ultracentrifugation

protein	concentration (μM)	MW (global fit)/MW (calc.)[a]
APH_i	4, 8, 12	1.7
APH_{ii} (APH2)	4, 7.4, 11	0.76
APH_{ii}[b] (APH2)	20, 40	0.99
APH_{iii} (APH3)	4.5, 9, 14	0.94
APH_{iii}[b] (APH3)	20, 40	1.16
APH_{iv}	7.7, 15.3	1.23
APH_{iv}[b]	20	1.58
APH_{vi} (APH4)	4, 7.4, 12	0.96
APH_{vi}[b] (APH4)	20, 40	1.08

[a]MW(calc.) is the expected dimer mass of each designed coiled coil. [b]Data collected using interference optics, and a construct not fused to MBP.

Orientation and Orthogonality of Designs. To determine the helix orientation in complexes formed by APH2, APH3, and APH4, we performed disulfide-exchange experiments, and resolved the products of the reactions using HPLC (see Materials and Methods). Key peaks are labeled in Figure 3, which shows changes in the chromatograms over time. For all three designs, starting with a combination of oxidized parallel species and/or reduced peptides, only one oxidized peak was detected at the end of 5 h, corresponding to a disulfide-linked antiparallel homodimer. On the basis of the smallest detectable peak area, we estimated a minimum 10^5-fold preference for forming antiparallel complexes over parallel complexes for all designs.

The same constructs that were used to measure orientation preferences were used to determine whether the designs formed heterodimers. APH peptides were tested in a pairwise manner (Figure 4). Each design formed a disulfide cross-linked antiparallel homodimer over time, but we did not detect any disulfide bond formation between any pairs of designed peptides. Each design was additionally measured for cross reactivity with the antiparallel homodimer-forming peptide APH, in a pairwise manner (Figure 5). No design showed any detectable cross-reactivity with APH, in either orientation, extending the number of orthogonal antiparallel homodimers from three to four.

To determine whether mixtures of more than two APH coiled coils formed complexes other than the expected dimers, MBP fusions of all four APH peptides were mixed at 20 or 40 μM of each APH design and analyzed by sedimentation equilibrium ultracentrifugation (as done for individual MBP fusion proteins, see Materials and Methods). The ratio of the fitted mass to the dimer mass was 0.91, with good fit quality (representative data in Figure S3), indicating that dimers formed as expected and no higher-order species were present in a mixture of all four APH fusion proteins.

Helicity and Thermal Stability. We measured the circular dichroism (CD) spectra of the three designed peptides APH2, APH3, and APH4, using the N-terminal cysteine constructs in a reduced state. Each construct contained 65 residues, of which 43 corresponded to the designed coiled-coil sequence (Table S3). Our APH construct contained 66 residues, of which 44 corresponded to the APH sequence. The CD spectra of all three designs were characteristic of coiled coils, with distinct minima at 208 and 222 nm (Figure 6A). The mean residue ellipticity (MRE) of the designed peptides was similar to that of APH, which is longer by one residue in the coiled-coil region. Thermal denaturation experiments established that all designs unfolded cooperatively, which is a characteristic property of coiled coils (Figure 6B). The thermal stabilities (T_m) of the designs at 20 μM ranged from 47.4 °C for APH2, to 59.3 °C for APH4 and 78.3 °C for APH3, with APH3 being slightly less stable then APH, which had a T_m of 79.3 °C. All melts were reversible. Upon recooling, all peptides regained ≥95% of the original MRE, and fits of refolding curves gave melting temperatures within 1.5 °C of values obtained from the denaturing curves.

Estimating peptide helicity from the CD data using the method of Morrisett et al. indicated that linkers and tags appended to the designed coiled coil contributed helical signal, as shown in Figure S4.[50] However, most of this "extra" signal was lost gradually with temperature in the pretransition baseline, indicating these regions are not part of the cooperative unfolding event. Furthermore, these linker residues were not present in the MBP-fusion constructs used in the sedimentation equilibrium centrifugation experiments, consistent with them not being necessary for the specific interactions observed in those experiments. Finally, making APH in the construct that we used for CD experiments did not change its melting temperature from the value reported in the literature for only the coiled coil.[21]

■ DISCUSSION

An expanded toolkit of coiled-coil interaction parts would be of great utility in protein engineering. Many papers have reported the successful design of coiled-coil structures of diverse topologies, but apart from parallel dimers, the number of biochemically characterized complexes of any one type is limited.[3,6,56] Designing coiled coils de novo is complicated by

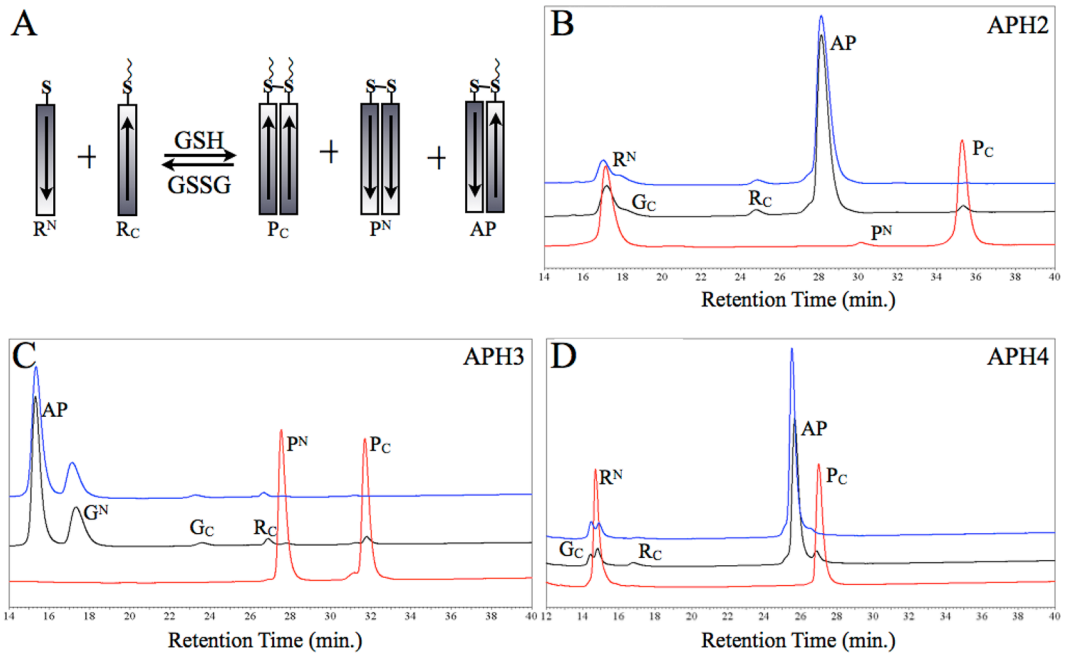


Figure 3. Designed peptides APH2, APH3, and APH4 adopt an antiparallel helix orientation. (A) Schematic of the assay. Arrows indicate helix direction from N to C terminus. The wavy line indicates two amino acids added to the designed sequence to change peptide retention times (APH2 = YY, APH3 = QW, APH4 = YY). S represents the sulfur atom in cysteine residue(s). (B, C, D) HPLC chromatograms show the results for the disulfide-exchange reactions upon mixing equimolar amounts of N-terminal and C-terminal cysteine variants of each design sequence (20 μM each). The reactions were quenched at 0 min (red), 15 min (black), or 5 h (blue). Peaks are labeled according to the scheme shown in panel A, with G indicating a glutathione adduct.

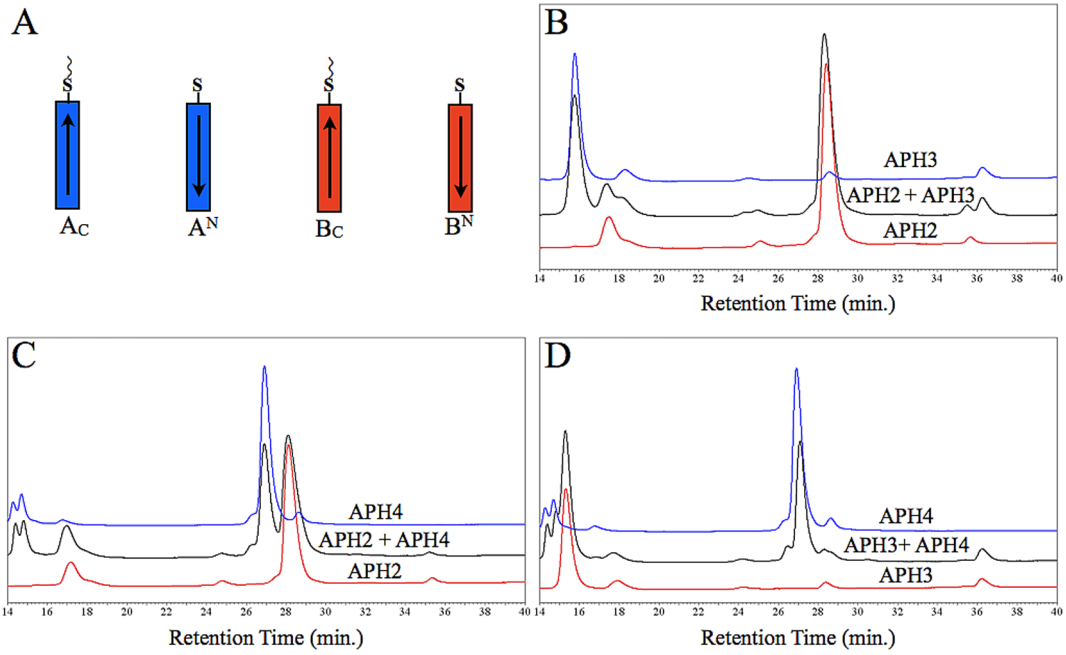


Figure 4. Designed peptides APH2, APH3 and APH4 do not form heterodimers. (A) Cartoon showing four cysteine-containing peptides, two for each of two designs, which were included in the disulfide-exchange cross-reactivity assay. (B, C, D) HPLC traces for all pairwise mixtures of designed peptides after equilibration for 15 min. The blue and red traces are for reactions with equimolar amounts of N- and C-terminal cysteine variants of a single designed peptide (20 μM each). The black trace is for a reaction with equimolar amounts of all four peptides in panel A (20 μM each). (B) APH2 + APH3, (C) APH2 + APH4, (D) APH3 + APH4.

the fact that different coiled-coil topologies have similar sequence requirements, and small sequence changes can alter coiled-coil structure. For these reasons, it is often necessary to explicitly consider competing states in the design process.[25,28,32]

Treating off-target states in computational protein design can be costly, particularly when there are many such states that must be modeled. One strategy is to incorporate a design element known to strongly destabilize a set of off-target topologies, to reduce the number of off-target states that must

dx.doi.org/10.1021/ja507847t | J. Am. Chem. Soc. 2014, 136, 16544−16556

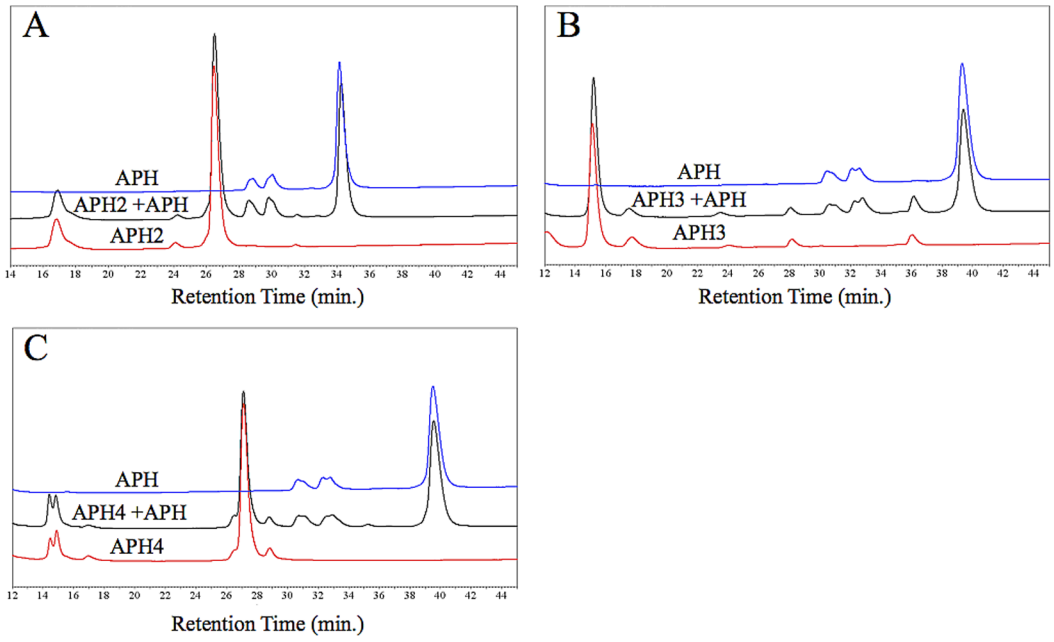


Figure 5. Designed peptides APH2, APH3, and APH4 do not heterodimerize with APH. (A, B, C) HPLC traces for all pairwise combinations of APH with the designed coiled coils, with experimental conditions as for Figure 4. The blue and red traces are for equimolar mixtures of N- and C-terminal cysteine variants of APH (blue) or APH2, APH3 or APH4 (red) (20 μM each). The black trace is for a mixture of four peptides, APH and the indicated design, each modified at the N- or C-terminus with a cysteine residue (20 μM each). (A) APH + APH2, (B) APH + APH3, (C) APH + APH4.

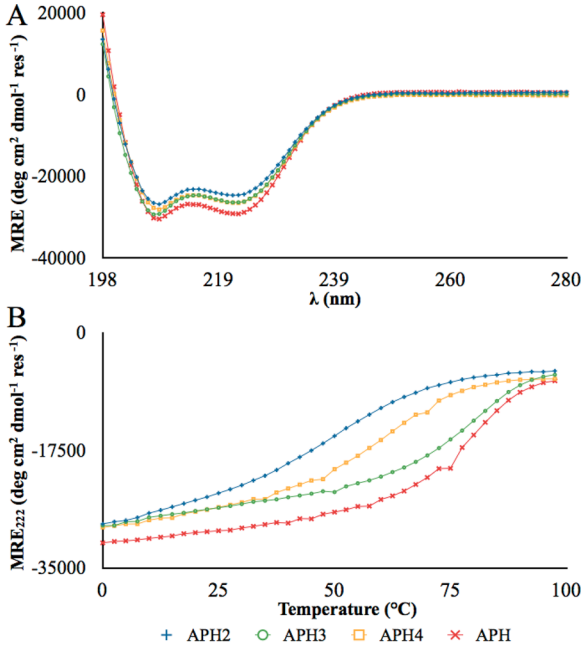


Figure 6. Circular dichroism spectra and thermal denaturation curves. (A) CD spectra and (B) thermal denaturation curves measured at 25 °C in PBS with 1 mM DTT. APH (red), APH2 (blue), APH3 (green) and APH4 (orange).

be modeled. For instance, Thomas et al. observed that the de novo design of parallel heterodimeric coiled coils composed entirely of isoleucine and leucine cores did not reliably destabilize higher-order states.[56] But the same design strategy in the background of a single asparagine−asparagine interaction, which was known from prior work to favor parallel dimer states over higher-order states, consistently gave dimeric assemblies.[28] Unfortunately, incorporating simple design elements that reliably destabilize all off-target topologies, in all sequence contexts, is not feasible. Exceptions have been reported for even the most thoroughly studied coiled-coil structural specificity determinants,[3,56] and for many coiled-coil topologies, the sequence-structure relationship is not well understood.

Of relevance for this work, there are few sequence features known to favor antiparallel over parallel helical alignments. Oakley et al. showed that, in analogy to the role of asparagines favoring dimers over higher-order states, paired asparagines can be introduced at opposing *a* and *d′* positions to favor an antiparallel helix alignment.[57] McClain et al. demonstrated that charge−charge interactions at *e* and *g* positions across the interface can impart an antiparallel vs parallel preference.[58] Gurnon et al. placed an isoleucine at a *d* heptad position and an alanine residue at an opposing *a′* heptad position to favor an antiparallel homodimer state over a parallel homodimer state in the designed sequence APH.[21] Further evidence supporting this interaction as an orientation specificity determinant was obtained via thiol-thioester exchange studies by Hadley et al.[59] Although simple rules do have some utility for design, Hadley et al. showed that residue−residue interactions in antiparallel coiled coils can also be highly context dependent, helping explain why rules extracted from studies of model systems do not satisfactorily explain the orientations of native coiled coils.[51,60]

Modeling off-target states explicitly and including them in the design process provides a broadly applicable mechanism for engineering specificity. In this work, we used explicit negative design to disfavor antiparallel heterodimer states by imposing energy gaps between antiparallel homo and heterodimers. Most of the sequence elements in our APH designs that disfavored antiparallel heterodimerization within a design set involved charged residues predicted to participate in repulsive interactions in heterodimer states. For example, all antiparallel

dx.doi.org/10.1021/ja507847t | J. Am. Chem. Soc. 2014, 136, 16544−16556

heterodimer states contained *a*-to-*e*′ and *d*-to-*g*′ charge−charge repulsions between lysine or arginine residues. Designs from sequence space 1 additionally contained *a*-to-*e*′ charge−charge repulsions between glutamate residues. These core-to-edge charge−charge repulsions were the most destabilizing weights available to the antiparallel CE DFIRE* model in the design sequence spaces chosen, with lysine at *d* to arginine at *g*′ being the most destabilizing.

The design strategies that led to destabilization of parallel homodimers differed in sequences spaces 1 and 2. In sequence space 1, we allowed glutamate at *a* positions, and all designed sequences included this element. In fact, we identified a motif consisting of two glutamate residues at *a* and *g*, and a lysine at *d*′ with an arginine at *e*′ on the opposing helix that was present in all of the sequence space 1 designs (Figure S5). Interactions between residues in this motif contain the first and fourth most favorable weights available in the CE DFIRE* model in sequence space 1, such that the motif is predicted to contribute strongly to antiparallel homodimer stability. Interestingly, in a parallel homodimer, the residues of this motif form unfavorable interactions sufficient to provide a large energy gap between parallel and antiparallel states. Certain unfavorable weights are shown in Figure S5, and this can be further demonstrated by modeling an artificial homodimer that includes the motif embedded in a poly-alanine sequence. Because of the symmetry of the homodimer, this results in two copies of the motif in the structure. Scoring parallel and antiparallel homodimeric structures with this sequence using DFIRE* revealed a significant preference of 1.64 energy units for the antiparallel state (poly alanine alone has a preference of 0.14 energy units for the antiparallel state using this model). Thus, in sequence space 1, charge networks predicted to stabilize the antiparallel state led to substantial destabilization of parallel homodimers, without explicit negative design. The situation was different in sequence space 2, which did not include glutamate residues at *a* positions. In this sequence space, designing antiparallel homodimers while disfavoring heterodimers did not automatically lead to large energy gaps to parallel homodimer states for all sequences (see Figure S2B); it was necessary to include parallel structures as off-target states in the optimization problem. Doing so led to sequences that placed more isoleucines at *d* heptad positions to favor antiparallel over parallel homodimers. For example, of the three sequences originally chosen in sequence space 2, two sequences had one isoleucine residue at a *d* position, while one sequence had no isoleucine residues at all. After placing constraints on the energies of the parallel homodimer states, all design sequences contained one or two isoleucine residues at *d* heptad positions. Each designed isoleucine at a *d* position introduced a *d*−*d*′ isoleucine pairing across the coiled-coil interface in the parallel homodimer state. As previously mentioned, this interaction destabilizes parallel dimers. The effect is captured in our models: isoleucine at *d*−*d*′ is the fourth most destabilizing weight for parallel dimers in sequence space 2.

Explicit consideration of off-target states requires enumerating and modeling the relevant competing states. We successfully used this strategy to destabilize antiparallel heterodimer states in sequence spaces 1 and 2, and to destabilize parallel homodimers when designing in sequence space 2. But we did not explicitly model formation of higher-order assemblies, and as a result, oligomers larger than dimers were formed by designs APH_i and APH_{iv}. Modeling higher-order coiled coils is challenging due to the many different topologies that are possible. Each helix pair can be antiparallel or parallel, heteroassemblies can form with different stoichiometries, and the geometry of helix associations can vary in subtle ways.[61,62] It is therefore difficult to include a comprehensive set of competing states and, even if such a set could be generated, the computational modeling costs for considering all possibilities explicitly would be high.

One approach to disfavoring higher-order states could be to include just a small number of trimer and tetramer topologies in the calculations. Adding representative off-target structures would minimally alter the computational complexity of the design framework, yet might lead to broader destabilization of additional higher-order states. Indeed, our study provided an example where specificity was obtained against states that were not explicitly modeled, possibly due to constraints on specificity against related states. The design solutions from sequence space 1 were predicted not to form heterodimers with design solutions from sequence space 2, despite these interactions not being explicitly constrained during optimization. We hypothesize that this occurred because the consideration of many off-target dimer states gave rise to interfaces with charge patterns low in symmetry, as well as hydrophobic cores with unique geometries due to the placements of beta-branched residues in the core. As a result, the probability of cross-reacting with another sequence to form dimers was low.

Considering just a few higher-order states may also have the effect of reducing or removing design features known to favor higher-order states generally. For example, isoleucines at *d* heptad positions are known to favor parallel trimer and tetramer states in preference to parallel dimer states.[3,27] Yet isoleucines at *d* heptad positions also favor antiparallel dimers over parallel dimers, and were included in many of our designs for this reason, as discussed above (also see Table 1). Interestingly, in native coiled coils isoleucines are approximately 4-fold more common in antiparallel dimers than in parallel dimers (Table S5). Isoleucines at *d* heptad positions that were included in the design to favor antiparallel dimers might have promoted the formation of higher-order assemblies, which were not treated in the model. A constraint to disfavor just a few trimers or tetramers might be sufficient to limit the use of this sequence element, or to drive inclusion of compensating elements that are poorly accommodated in higher-order assemblies.

A significant obstacle to including even a few higher-order states in design is the small amount of structural data available for coiled-coil trimers and tetramers of a specific toplogy.[37] Benchmarking the predictive power of models using experimental data is important for assessing performance, and is useful for setting meaningful energy cutoffs in design calculations. However, very few known structures of higher-order states of any specific topology passed our orientation test set filters of ≤50% sequence identity and >27 residues (0 antiparallel trimers, 6 antiparallel tetramers, and 9 parallel tetramers in the August 18, 2010 CC+ database). For these reasons, we did not benchmark DFIRE* on the problem of predicting oligomerization state, and we did not attempt to use it for this purpose.

We examined the structure-prediction power of other methods when applied to our antiparallel coiled coils. LOGICOIL is a computational predictor trained on coiled-coil sequences in the CC+ database to discriminate parallel dimers, antiparallel dimers, trimers and tetramers.[9] For APH2, APH3, and APH4, LOGICOIL assigned very similar scores for

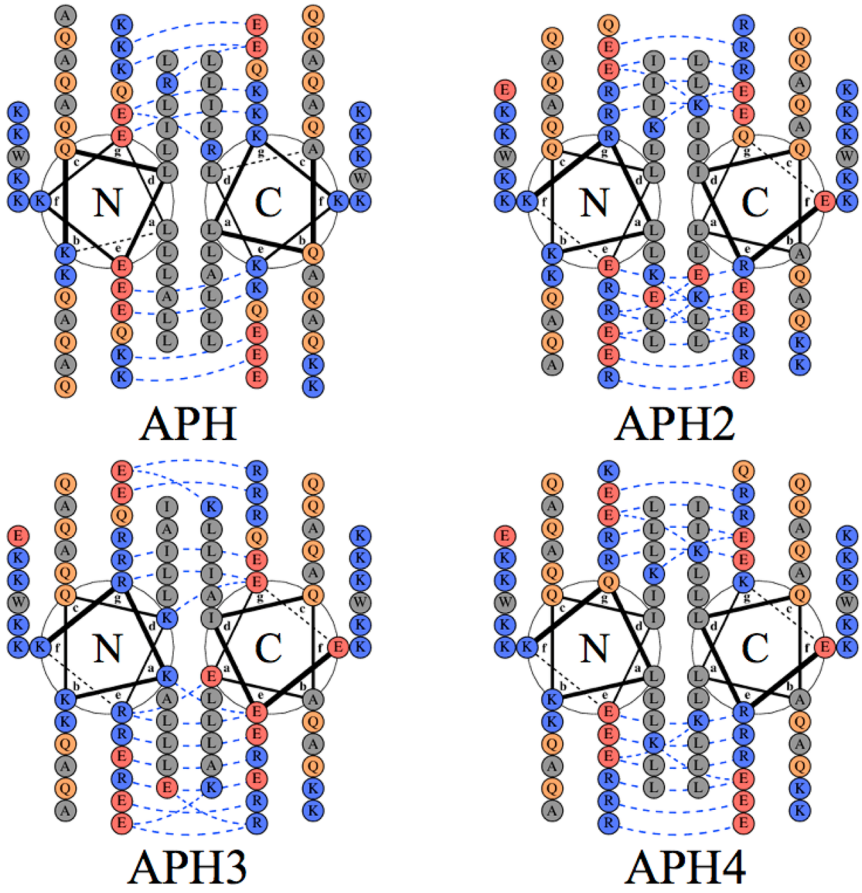


Figure 7. Helical-wheel diagrams of APH, APH2, APH3, and APH4 as antiparallel homodimers. Positively and negatively charged amino acids are shown in blue and red, respectively, with noncharged polar residues in orange and hydrophobic residues in gray. Potentially attractive salt bridges are shown as dashed lines. Sequences start at an *f* position and end at an *e* position. Diagrams were generated using DrawCoil 1.0, http://www.grigoryanlab.org/drawcoil.

each of the four topologies. The challenge the APH sequences present to LOGICOIL is not surprising. LOGICOIL makes predictions based on a single-chain sequence, without information about interchain interactions that are crucial design elements in the APH designs. Additionally, the APH sequences are de novo designed sequences, with intrachain pairwise frequencies that may not resemble those in native sequences. CCBuilder is a new web-based application that generates coiled-coil structures and can be used for predicting coiled-coil topology. CCBuilder can model parallel dimers, antiparallel dimers, parallel trimers, and parallel tetramers, and the program computes the stability of coiled-coil complexes using two energy functions: Rosetta and BUDE.[63] Both energy functions, used with default settings provided by the Web site, correctly predict that APH2, APH3, and APH4 favor the antiparallel dimer state over the parallel dimer state. When higher-order states are considered, both energy functions predict that APH2 and APH4 favor the trimer state, and Rosetta also predicts that APH3 will form a trimer. BUDE however correctly predicts that APH3 will favor the antiparallel dimer state. Better methods for oligomerization state prediction are needed and, if developed, could be incorporated into our design framework.

The rankings of the thermal stabilities (Figure 6) are not predicted well by DFIRE*. DFIRE* instead predicts that APH3 is the most stable complex, followed by APH2, APH4, and APH. These predicted ranking are consistent with the favorable weights that the CE of DFIRE* assigns between core and edge positions. But the relative thermal stabilities of the APH coiled coils appear to be related to the number of charged residues in the central two heptads of the designed coiled coils (Figure 7). APH2, which has the greatest number of charged residues in the central two heptads, is the least stable. In contrast, both APH and APH3 contain no charged residues in the central two heptads and are the most thermally stable. This is consistent with many studies showing coiled-coil destabilization by polar residues in the core.[59,60,66] It should be noted that CCBuilder accurately predicts the thermal stability rankings using Rosetta or BUDE scores, if all structures are scored as antiparallel dimers.

The new APH designs have many desirable properties for synthetic biology and materials science. First, the surface residues of all APH designs were engineered to be passive and may provide useful positions for adding novel functions or modulating stability.[52,64,65] The designed structures also provide users with a range of thermal stabilities, and it may be possible to tune the dimer stabilities using mutations of the surface residues, as needed. Finally, the designs are orthogonal to each other when used in pairwise or higher-order combinations. Proteins with this property have been sought for many applications in synthetic biology and are thought to be one of the limiting reagents slowing progress in this field.[19,67] It should also be noted that heterodimers involving the APH proteins could be included as off-target states in future design studies using the CLASSY framework, allowing for the

extension of this set. In conclusion, the antiparallel homodimer sequences represent a significant expansion to the coiled-coil toolkit, which is currently dominated by parallel dimers, and thus may find application in many molecular engineering projects.

■ ASSOCIATED CONTENT

Supporting Information

Additional figures, including a schematic of the cluster expansion protocol, DFIRE* energy gaps predicted for certain designs, representative sedimentation equilibrium centrifugation plots, predicted helicity of APH sequences from CD data, helical-wheel diagrams of the design motif in sequence space 1, and helical-wheel diagrams of each APH sequence showing how the helices would align in a symmetric parallel homodimer. Additional tables, including the average Crick parameters fit to crystal structures, details of the orientation test set, sequences of experimental constructs, and statistical analysis of dimeric coiled-coil sequences. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

keating@mit.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank members of the Keating lab, especially R. Rezaei Araghi for performing mass spectrometry analysis of peptides and J. B. Kaplan and K. Hauschild for experimental advice. This work used the MIT Bioinstrumentation Facility, and we are grateful to D. Pheasant for analytical ultracentrifugation support. Funding was from a National Science Foundation Graduate Research Fellowship awarded to C.N. and from NSF Award MCB-0950233 (supporting experimental applications) and NIH Award GM67681 (supporting computational design method development) to A.E.K. We used computer resources provided by National Science Foundation Award DBI-0821391.

■ REFERENCES

(1) Purnick, P. E.; Weiss, R. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 410−422.
(2) Thompson, K. E.; Bashor, C. J.; Lim, W. A.; Keating, A. E. *ACS Synth. Biol.* **2012**, *1*, 118−129.
(3) Fletcher, J. M.; Boyle, A. L.; Bruning, M.; Bartlett, G. J.; Vincent, T. L.; Zacci, N. R.; Armstrong, C. T.; Bromley, E. H. C.; Booth, P. J.; Brady, R. L.; Thomson, A. R.; Woolfson, D. N. *ACS Synth. Biol.* **2012**, *1*, 240−250.
(4) Wolf, E.; Kim, P. S.; Berger, B. *Protein Sci.* **1997**, *6*, 1179−1189.
(5) Rackham, O. J.; Madera, M.; Armstrong, C. T.; Vincent, T. L.; Woolfson, D. N.; Gough, J. *J. Mol. Biol.* **2010**, *403*, 480−493.
(6) Reinke, A. W.; Grant, R. A.; Keating, A. E. *J. Am. Chem. Soc.* **2010**, *132*, 6025−6031.
(7) Lupas, A. *Trends Biochem. Sci.* **1996**, *21*, 375−382.
(8) Woolfson, D. N. *Adv. Protein Chem.* **2005**, *70*, 79−112.
(9) Vincent, T. L.; Green, P. J.; Woolfson, D. N. *Bioinformatics* **2013**, *29*, 69−76.
(10) Woolfson, D. N.; Alber, T. *Protein Sci.* **1995**, *4*, 1596−1607.
(11) Wolf, E.; Kim, P. S.; Berger, B. *Protein Sci.* **1997**, *6*, 1179−1189.
(12) Armstrong, C. T.; Vincent, T. L.; Green, P. J.; Woolfson, D. N. *Bioinformatics* **2011**, *27*, 1908−1914.
(13) Mahrenholz, C. C.; Abfalter, I. G.; Bodenhofer, U.; Volkmer, R.; Hochreiter, S. *Mol. Cell. Proteomics* **2011**, *10*, M110.004994.
(14) Trigg, J.; Gutwin, K.; Keating, A. E.; Berger, B. *PLoS One* **2012**, *6*, e23519.
(15) Wolfe, S. A.; Grant, R. A.; Pabo, C. O. *Biochemistry* **2003**, *42*, 13401−13409.
(16) Bashor, C. J.; Helman, N. C.; Yan, S.; Lim, W. A. *Science* **2008**, *319*, 1539−1543.
(17) Lanci, C. J.; MacDermaid, C. M.; Kang, S.; Acharya, R.; North, B.; Yang, X.; Qiu, X. J.; DeGrado, W. F.; Saven, J. G. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 7304−7309.
(18) Shlizerman, C.; Atanassov, A.; Berkovich, I.; Ashkenasy, G.; Ashkenasy, N. *J. Am. Chem. Soc.* **2010**, *132*, 5070−5076.
(19) Gradišar, H.; Božič, S.; Doles, T.; Vengust, D.; Hafner-Bratkovič, I.; Mertelj, A.; Webb, B.; Šali, A.; Klavžar, S.; Jerala, R. *Nat. Chem. Biol.* **2013**, *9*, 362−366.
(20) Lumb, K. J.; Carr, C. M.; Kim, P. S. *Biochemistry* **1994**, *33*, 7361−7367.
(21) Gurnon, D. G.; Whitaker, J. A.; Oakley, M. G. *J. Am. Chem. Soc.* **2003**, *125*, 7518−7519.
(22) Taylor, C. M.; Keating, A. E. *Biochemistry* **2005**, *44*, 16246−16256.
(23) Gradišar, H.; Jerala, R. *J. Pept. Sci.* **2010**, *17*, 100−106.
(24) Reinke, A. W.; Baek, J.; Ashenberg, O.; Keating, A. E. *Science* **2013**, *340*, 730−734.
(25) Havranek, J. J.; Harbury, P. B. *Nat. Struct. Biol.* **2003**, *10*, 45−52.
(26) O'Shea, E. K.; Lumb, K. J.; Kim, P. S. *Curr. Biol.* **1993**, *3*, 658−667.
(27) Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T. *Science* **1993**, *262*, 1401−1407.
(28) Lumb, K. J.; Kim, P. S. *Biochemistry* **1995**, *34*, 8642−8648.
(29) McClain, D. L.; Gurnon, D. G.; Oakley, M. G. *J. Mol. Biol.* **2002**, *324*, 257−270.
(30) London, N.; Ambroggio, X. *J. Struct. Biol.* **2013**, *185*, 136−146.
(31) Davey, J. A.; Chica, R. A. *Protein Sci.* **2012**, *21*, 1241−1252.
(32) Grigoryan, G.; Reinke, A. W.; Keating, A. E. *Nature* **2009**, *458*, 859−864.
(33) Negron, C.; Keating, A. E. *Methods Enzymol.* **2013**, *523*, 171−190.
(34) Reinke, A. W.; Grigoryan, G.; Keating, A. E. *Biochemistry* **2010**, *49*, 1985−1997.
(35) Chen, T. S.; Reinke, A. W.; Keating, A. E. *J. Mol. Biol.* **2011**, *408*, 304−320.
(36) Yang, Y.; Zhou, Y. *Proteins* **2008**, *72*, 793−803.
(37) Testa, O. D.; Moutevelis, E.; Woolfson, D. N. *Nucleic Acids Res.* **2009**, *37*, 315−322.
(38) Crick, F. H. C. *Acta Crystallogr.* **1953**, *6*, 685−689.
(39) Grigoryan, G.; DeGrado, W. F. *J. Mol. Biol.* **2011**, *405*, 1079−1100.
(40) Kuhlman, B.; Baker, D. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 10383−10388.
(41) Grigoryan, G.; Zhou, F.; Lustig, S. R.; Ceder, G.; Morgan, D.; Keating, A. E. *PLoS Comput. Biol.* **2006**, *2*, 551−563.
(42) Hahn, S.; Ashenberg, O.; Grigoryan, G.; Keating, A. E. *J. Comput. Chem.* **2010**, *31*, 2900−2914.
(43) *IBM ILOG CPLEX Optimization Studio*, Version 12.5; IBM Corp.: Armonk, NY, 2012.
(44) Hoover, D. M.; Lubkowski, J. *Nucleic Acids Res.* **2002**, *30*, e43.
(45) Edelhoch, H. *Biochemistry* **1967**, *6*, 1948−1967.
(46) Schuck, P. *Anal. Biochem.* **2003**, *320*, 104−124.
(47) Vistica, J.; Dam, J.; Balbo, A.; Yikilmaz, E.; Mariuzza, R. A.; Roualt, T. A.; Schuck, P. *Anal. Biochem.* **2004**, *326*, 234−256.
(48) Laue, T. M.; Shah, B. D.; Ridgeway, T. M.; Pelletier, S. L. Computer-aided interpretation of analytical sedimentation data for proteins. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*; Harding, S. E., Rowe, A. J., Horton, J. C., Eds.; Royal Society of Chemistry: Cambridge, 1992; pp 90−125.
(49) John, D. M.; Weeks, K. M. *Protein Sci.* **2000**, *9*, 1416−1419.
(50) Morrisett, J. D.; David, J. S. K.; Pownhall, H. J.; Gotto, A. M., Jr. *Biochemistry* **1973**, *12*, 1290−1299.

(51) Apgar, J. R.; Gutwin, K. N.; Keating, A. E. *Proteins* **2008**, *72*, 1048−1065.
(52) Zhou, F.; Grigoryan, G.; Lustig, S. R.; Keating, A. E.; Ceder, G.; Morgan, D. *Phys. Rev. Lett.* **2005**, *95*, 148103.
(53) Li, Y.; Kaur, H.; Oakley, M. G. *Biochemistry* **2008**, *47*, 13564−13572.
(54) Mason, J. M.; Hagemann, U. B.; Arndt, K. M. *Biochemistry* **2009**, *48*, 10380−10388.
(55) Straussman, R.; Ben-Ya'acov, A.; Woolfson, D. N.; Ravid, S. *J. Mol. Biol.* **2007**, *366*, 1232−1242.
(56) Thomas, F.; Boyle, A. L.; Burton, A. J.; Woolfson, D. N. *J. Am. Chem. Soc.* **2013**, *135*, 5161−5166.
(57) Oakley, M. G.; Kim, P. S. *Biochemistry* **1998**, *37*, 12603−12610.
(58) McClain, D. L.; Woods, H. L.; Oakley, M. G. *J. Am. Chem. Soc.* **2001**, *123*, 3151−3152.
(59) Hadley, E. B.; Gellman, S. H. *J. Am. Chem. Soc.* **2006**, *128*, 16444−16445.
(60) Hadley, E. B.; Testa, O. D.; Woolfson, D. N.; Gellman, S. H. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 530−535.
(61) Deng, Y.; Zheng, Q.; Liu, J.; Cheng, C. S.; Kallenbach, N. R.; Lu, M. *Protein Sci.* **2007**, *16*, 323−328.
(62) Liu, J.; Zheng, Q.; Deng, Y.; Li, Q.; Kallenbach, N. R.; Lu, M. *Biochemistry* **2007**, *46*, 14951−14959.
(63) Wood, C. W.; Bruning, M.; Ibarra, A. A.; Bartlett, G. J.; Thomson, A. R.; Sessions, R. B.; Brady, R. L; Woolfson, D. N. *Bioinformatics* **2014**, *30*, 3029−3035.
(64) Dahiyat, B. I.; Gordon, B.; Mayo, S. L. *Protein Sci.* **1997**, *6*, 1333−1337.
(65) Kaplan, J. B.; Reinke, A. W.; Keating, A. E. *Protein Sci.* **2014**, *23*, 940−953.
(66) Acharya, A.; Rishi, V.; Vinson, C. *Biochemistry* **2006**, *45*, 11324−11332.
(67) Kapp, G. T.; Liu, S.; Stein, A.; Wong, D. T.; Reményic, A.; Yeh, B. J.; Fraser, J. S.; Taunton, J.; Lim, W. A.; Kortemme, T. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 5277−5282.