ACS SyntheticBiology

Research Article

pubs.acs.org/synthbio

A Basis Set of *de Novo* Coiled-Coil Peptide Oligomers for Rational Protein Design and Synthetic Biology

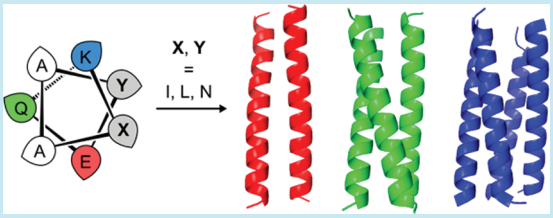
Jordan M. Fletcher,[†,#] Aimee L. Boyle,[†,#] Marc Bruning,[†,#] Gail J. Bartlett,[†] Thomas L. Vincent,[†,§] Nathan R. Zaccai,[‡,∥] Craig T. Armstrong,[†,‡] Elizabeth H. C. Bromley,[†,⊥] Paula J. Booth,[‡] R. Leo Brady,[‡] Andrew R. Thomson,*[,†] and Derek N. Woolfson*[,†,‡]

[†]School of Chemistry, University of Bristol, Cantocks Close, Bristol BS8 1TS, U.K.

[‡]School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol BS8 1TD, U.K.

Supporting Information

ABSTRACT: Protein engineering, chemical biology, and synthetic biology would benefit from toolkits of peptide and protein components that could be exchanged reliably between systems while maintaining their structural and functional integrity. Ideally, such components should be highly defined and predictable in all respects of sequence, structure, stability, interactions, and function. To establish one such toolkit, here we present a basis set of *de novo* designed α-helical coiled-coil peptides that adopt defined and well-characterized parallel dimeric, trimeric, and tetrameric states. The designs are based on sequence-to-structure relationships both from the literature and analysis of a database of known coiled-coil X-ray crystal structures. These give *foreground* sequences to specify the targeted oligomer state. A key feature of the design process is that sequence positions outside of these sites are considered non-essential for structural specificity; as such, they are referred to as the *background*, are kept non-descript, and are available for mutation as required later. Synthetic peptides were characterized in solution by circular-dichroism spectroscopy and analytical ultracentrifugation, and their structures were determined by X-ray crystallography. Intriguingly, a hitherto widely used empirical rule-of-thumb for coiled-coil dimer specification does not hold in the designed system. However, the desired oligomeric state is achieved by database-informed redesign of that particular foreground and confirmed experimentally. We envisage that the basis set will be of use in directing and controlling protein assembly, with potential applications in chemical and synthetic biology. To help with such endeavors, we introduce *Pcomp*, an on-line registry of peptide components for protein-design and synthetic-biology applications.



KEYWORDS: *coiled coil, oligomeric state, protein design, self-assembly, synthetic biology*

One of the underlying tenets of the emerging field of synthetic biology is to make the engineering of biological systems easier and more reliable.[1,2] This challenge can be approached at a number of levels relating to engineering whole genomes, clusters of genes and their products, and individual or small collections of biomolecules.[3,4] Whatever approach is taken, a key theme is *standardization* and ultimately the design and construction of standard parts or components that can be used to perform similar tasks in different contexts. At the protein level, the synthetic-biology approach differs from traditional protein engineering, which has largely focused on bespoke solutions to problems at hand and hence often produces non-transferable outcomes. An alternative synthetic-biology approach is to produce standard sets of polypeptide components, i.e., toolkits, that are designed and fully characterized once, but which can be used in a "plug-and-play" manner to solve problems across many different systems.[5,6] To succeed, the components of such toolkits should be fully characterized in all respects from sequence through to structure, function, and interactions. For this reason, *de novo* designed components of reduced complexity where the role of every residue is understood may offer certain advantages over more-complex sequences appropriated from nature.

Among the desired components are tunable protein−protein interaction motifs, which might be used as hubs, scaffolds, and nucleators for other biomolecular assemblies and functions both *in vitro* and *in vivo*. There are many natural protein domains that could be used as starting points for the construction of protein−protein interaction components.[7,8] However, another consideration is that components should be designable or at least engineerable, and this requires a good understanding of sequence-to-structure relationships for the targets. This requirement points to using rational *de novo* designed peptides. Sequence-to-structure relationships in proteins are immensely complex and are not fully understood. The ability to design folded proteins reliably and using a small number of rules-of-thumb is limited to a small number of specific examples, for instance, zinc fingers and collagens,[3,9−12] though other structures, including truly *de novo* structures, protein−protein interactions, and enzyme-like function have

Received: April 6, 2012
Published: May 4, 2012

been accessed using computational approaches.[13−15] α-Helical coiled-coil proteins, however, have considerable advantages in terms of designability.[16]

Coiled coils exhibit possibly the greatest geometric diversity of protein−protein interactions in terms of number, orientation, and constitution of interacting elements.[17,18] In addition, due to their abundance in nature, the large number of available X-ray crystal structures, and a plethora of experimental and theoretical studies over the past two decades, there are established rules-of-thumb that relate protein sequence and structure in coiled-coil domains.[16] These features make coiled coils attractive and tractable targets for rational protein design and in contributing to the construction of toolkits of peptide and protein components for synthetic biology.

In terms of sequence, coiled-coil domains are found on average in ∼3% of protein-encoding regions across all genomes.[19] They often exhibit a heptad repeat of hydrophobic (*H*) and polar (*P*) residues, *HPPHPPP*, usually denoted ***abcdefg***, and most coiled coils have 4 or more contiguous heptad repeats. In structural terms (Figure 1), these patterns

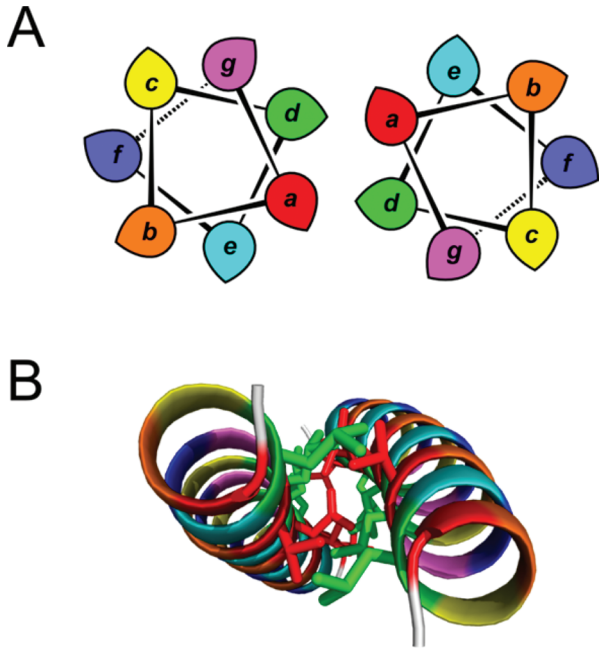


Figure 1. Coiled-coil assemblies. (A) Helical-wheel diagram showing the heptad repeat configured as a helix with 3.5 residues per turn, i.e., effectively incorporating the supercoil. Leaf shapes indicate the direction of the C_α−C_β vectors. Hydrophobic residues at the ***a*** and ***d*** positions provide most of the binding enthalpy, augmented by ion-pairing or hydrogen-bonding interactions between polar residues at opposing ***e*** and ***g*** positions. The ***b***, ***c***, and ***f*** positions are distant from the interface. (B) View down the long axis of a typical parallel, dimeric coiled coil (PDB ID 2ZTA[24]). Coloring of the heptad positions, ***abcdefg***, follows the CC+ standard for heptad positions (***a*** = red; ***b*** = orange; ***c*** = yellow; ***d*** = green; ***e*** = cyan; ***f*** = blue; ***g*** = magenta).[25] Structural images created using PyMol (http://www.pymol.org).

encode amphipathic α-helices, which assemble to bury the hydrophobic surface formed by the ***a*** and ***d*** residues. The slight mismatch between the periodicity of these hydrophobic residues and the α-helix—3.5 residues versus 3.6 residues/turn, respectively—leads to a left-handed supercoiling of the helices around each other and to characteristic and intimate knobs-into-holes side-chain packing between these helices.[20,21] Although this binding is primarily hydrophobic, the ***e*** and ***g*** positions flank the hydrophobic core and present possibilities for interhelix salt-bridging interactions. These apparently straightforward sequence-to-structure relationships mask considerable structural variety: classical coiled coils can have between 2 and 6 helices,[22] and more complex assemblies are possible.[18,23] The helices can be arranged in parallel, antiparallel, or mixed topologies in both homo- and heterotypic assemblies.[17,18,23] However, the precise energetic factors that discriminate between these various arrangements remain incompletely understood.

Though other states have been targeted successfully on an individual or bespoke basis, much of the work in coiled-coil *de novo* design has focused on dimers.[16,26] This work has culminated in the generation of partner-selective sets of orthogonal, that is, independently folding, heterodimeric coiled coils.[27−29] Undoubtedly, these will be of use in directing and cementing pairwise protein−protein interactions both *in vitro* and *in vivo* and in synthetic biology applications.[5,6] Here we focus on generating a set of parallel dimeric, trimeric, and tetrameric coiled-coil peptides, which constitute the major structural classes observed in known coiled-coil structures.[18] We regard these structures as the starting point for a toolkit of coiled coils, or a coiled-coil basis set. Our aims in this study were (1) to design these peptides *de novo* to facilitate understanding of sequence-to-structure relationships within the set; (2) to characterize the peptides as fully as possible both in solution and through X-ray crystallography; and (3) to deliver a robust set of peptides that can be manipulated precisely and predictably in future protein engineering and synthetic biology applications.[5,22,30]

RESULTS AND DISCUSSION

Rational Peptide Design. Residues at the ***g***, ***a***, ***d***, and ***e*** sites of the heptad repeat predominantly define the helix−helix interfaces in the vast majority of coiled-coil assemblies.[17,21,23] Therefore, we aimed initially to establish combinations of residues at these *foreground* sites to direct *de novo* peptide sequences into the different target oligomer states, namely, parallel dimer, trimer, and tetramer.

The seminal work of Harbury *et al.* on the leucine-zipper peptide (GNC4-p1) from the yeast transcriptional activator GCN4 provided the starting point for our designs.[31,32] In this study, combinations of isoleucine (Ile, I) and leucine (Leu, L) at the ***a*** and ***d*** positions of GCN4-p1 yield three different oligomer states: Ile at ***a*** plus Leu at ***d*** gives a dimer, Ile at both positions results in a trimer, and Leu at ***a*** plus Ile at ***d*** leads to a tetramer. Thus, for our starting set of sequences, we used these combinations at ***a*** and ***d*** in the peptides CC-pIL, CC-pII and CC-pLI, respectively (Table 1).

Next, we selected the amino acids at the core-flanking ***g*** and ***e*** sites. In natural coiled-coil sequences these are predominantly occupied by polar residues. More specifically, we found in parallel, homo-oligomers from the CC+ database[25] that glutamic acid (Glu, E), glutamine (Gln, Q), lysine (Lys, K), and less frequently arginine (Arg, R) dominate these positions (Supplementary Table S1). In addition, in these structures, there is the potential for g_h:e'_{h+1} salt-bridging interactions, where subscripts *h* and *h+1* signify consecutive heptads and the prime signifies a different peptide chain. On this basis, we chose Glu at ***g*** and Lys at ***e*** for all three starting sequences.

With the initial foreground ***g***, ***a***, ***d***, and ***e*** positions of the starting sequences set, the remaining *background* positions (***b***, ***c***,

dx.doi.org/10.1021/sb300028q | ACS Synth. Biol. 2012, 1, 240−250

Table 1. Sequences and Heptad Register for the Designed Coiled-Coil Peptides[a]

Systematic name	Sequence and heptad register	Basis-set name
	gabcdef gabcdef gabcdef gabcdef	
CC-pIL	Ac-GE**I**AA**L**KQ E**I**AA**L**KK E**I**AA**L**KW E**I**AA**L**KQ GYY-NH_2	-
CC-pIL-I17N	Ac-GE**I**AA**L**KQ E**I**AA**L**KK E**N**AA**L**KW E**I**AA**L**KQ GYY-NH_2	CC-Di
CC-pII	Ac-GE**I**AA**I**KQ E**I**AA**I**KK E**I**AA**I**KW E**I**AA**I**KQ GYG-NH_2	CC-Tri
CC-pII-I13N	Ac-GE**I**AA**I**KQ E**I**AA**N**KK E**I**AA**I**KW E**I**AA**I**KQ GYG-NH_2	CC-Tri-N13
CC-pLI	Ac-GE**L**AA**I**KQ E**L**AA**I**KK E**L**AA**I**KW E**L**AA**I**KQ GAG-NH_2	CC-Tet

[a]These were given unique mass and chromophoric signatures.

and ***f***) were kept the same in each peptide. These were designed to contribute to the overall stability of the assembly without favoring any particular oligomer state, i.e., to promote helicity only, and also to be exchangeable for other side chains in the future, as follows: for the designs presented here, ***b*** and ***c*** positions were made alanine as this is a small, helix-favoring and otherwise largely non-discriminating amino acid. The ***f*** positions were populated with combinations of Gln, Lys and tryptophan (Trp, W), to favor helicity and to improve solubility (Gln and Lys) and to provide a UV chromophore (Trp), respectively. Our analysis of CC+ also revealed that Ala, Gln, and Lys were among the most frequently observed amino acids at ***b***, ***c***, and ***f*** (Supplementary Tables S1 and S2). For some of our designed peptides, additional variants were made in which the Trp residues were replaced by 4-iodo-L-phenylalanine (iodo-Phe, Φ) to aid structure determination by X-ray crystallography.[33]

To provide sufficient stability, the sequences were made 4 heptads, i.e., 28 residues, long and were additionally capped with a Gly residue at each end. Finally, each sequence was distinguished by appending a two-residue sequence outside the helical region to give it a unique mass. The resulting sequences, which were named CC-pIL, CC-pII, and CC-pLI, respectively, are shown in Table 1, and a full set of sequences (including the iodo-Phe variants) are in Supplementary Table S3.

Solution-Phase Biophysical Characterization. The three sequences CC-pIL, CC-pII, and CC-pLI were synthesized by standard Fmoc solid-phase peptide synthesis, purified by reversed-phase HPLC, and confirmed by MALDI-TOF mass spectrometry (Supplementary Figure S1). We used circular dichroism (CD) spectroscopy to probe the nature and stability of secondary structure in the peptides. At 20 °C all three gave CD spectra consistent with near-complete α-helix formation (Figure 2A and Supplementary Figure S2), that is, with mean residue ellipticities at 222 nm (MRE_{222}) of ca. −34,000 to −40,000 deg cm^2 $dmol^{-1}$ res^{-1}, Table 2. Moreover, in each case the structures showed high thermal stability, with minimal denaturation, and no cooperative melting transition at temperatures up to 90 °C at 50 μM peptide concentrations, Figure 2B. N.B.: we interpret the slight, linear loss of MRE_{222} with increasing temperature as fraying of the termini of the helices.[34] Consistent with this, in the X-ray crystal structures reported below, we found increased B-factors for both the N- and C-terminal residues for all of our designed peptides (Supplementary Figure S3). In solution, reducing the peptide concentrations to 10 μM only revealed the beginnings of thermal unfolding curves above ∼70 °C (Supplementary Figure S4). Therefore, to access the full unfolding behavior of the peptides, we performed the thermal denaturation experiments in the presence of 3 M guanidinium chloride. These showed reversible, sigmoidal unfolding curves (Figure 2C, Supplementary Figure S4, and Table 2). Together, these CD data are consistent with CC-pIL, CC-pII, and CC-pLI folding to form thermally stable, cooperatively folded, and discrete helical bundles as designed.

To assess the size of the bundles formed, we turned to analytical ultracentrifugation (AUC). Sedimentation-equilibrium experiments for each peptide gave data that were readily interpreted by fits assuming single-ideal species, and more complex fitting to monomer−oligomer models was not necessary; in each case, the former returned oligomers with integer numbers of monomers. CC-pII and CC-pLI were trimeric and tetrameric, respectively, consistent with our design strategy and Harbury's experiments on GCN4-p1.[31,32] However, and to our surprise, CC-pIL was trimeric in solution. Consistent with these data, dynamic light scattering (DLS) measurements gave hydrodynamic diameters in the order CC-pLI > CC-pII ≈ CC-pIL (Supplementary Figure S6). These results challenge many first-order assumptions in the literature, including our own, that peptides with ***a*** = Ile and ***d*** = Leu should be dimeric.[3,16] This inspired further investigation.

X-ray Crystal Structures. To gain further structural insight into the assemblies of CC-pIL, CC-pII, and CC-pLI, X-ray crystal structures were determined using a combination of molecular replacement and single wavelength anomalous diffraction methods. The latter used experimental phases from iodine atoms incorporated *via* iodo-Phe in place of Trp (Supplementary Table S3). Full details of crystallization, structure determination, and refinement are given in Supplementary Tables S4 and S5.

The X-ray crystal structures revealed that each peptide formed parallel, blunt-ended assemblies, as per the design strategy, with oligomer states matching those observed in solution, i.e., CC-pIL and CC-pII were trimers, and CC-pLI was tetrameric. As judged by the program SOCKET,[21] all three assemblies had "knobs-into-holes" side-chain packing centered on a single hydrophobic seam of ***a*** and ***d*** residues (Figure 3, Supplementary Table S6) and, as determined by TWISTER,[35]

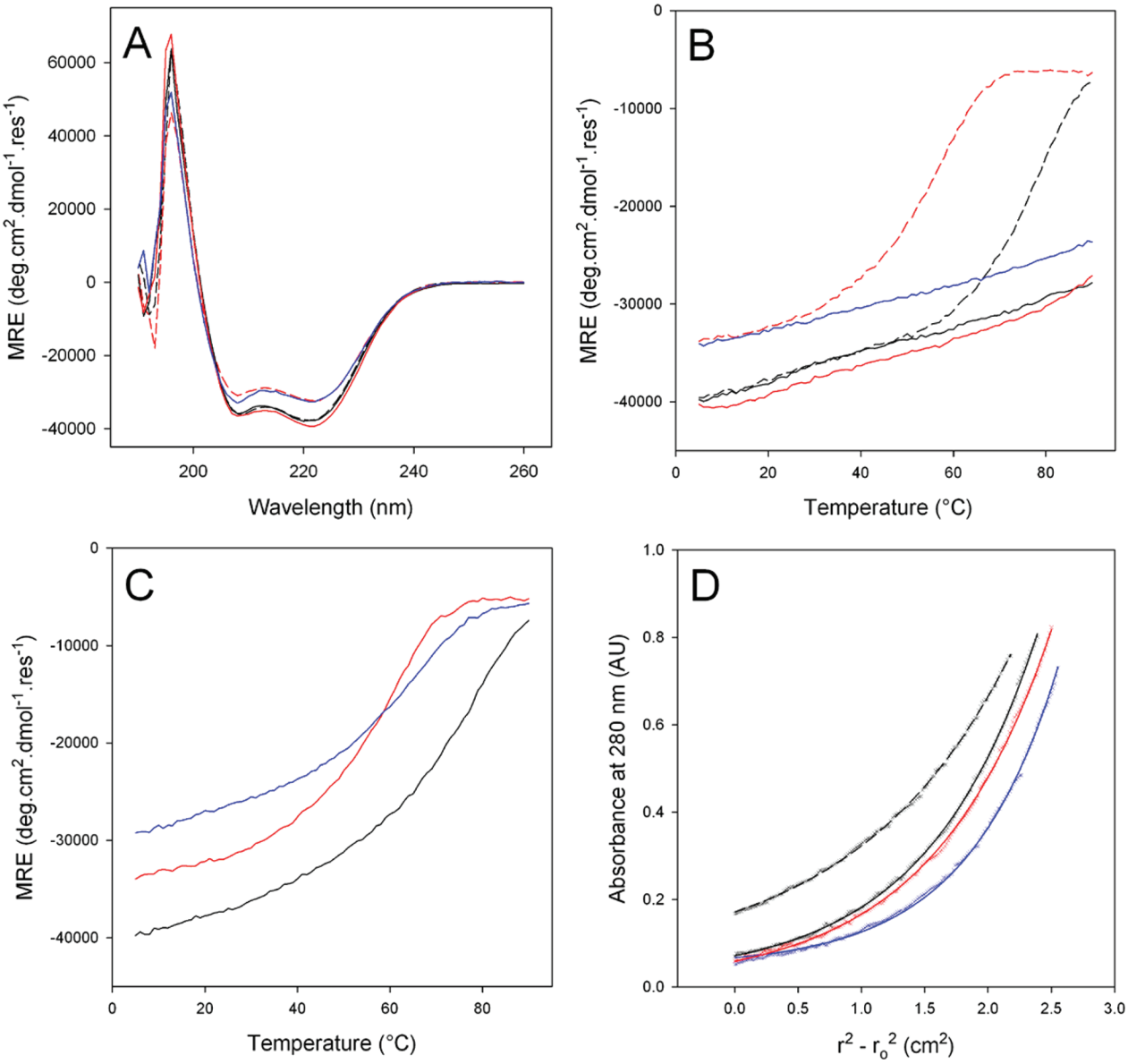


Figure 2. Solution-phase characterization of the designed coiled-coil peptides. (A) CD spectra recorded at 20 °C. (B) Thermal unfolding curves followed by CD spectroscopy. (C) Thermal denaturation curves in the presence of 3 M guanidine hydrochloride. (D) Representative AUC sedimentation equilibrium data (crosses) recorded at 20 °C and 46,000 rpm rotor speed in a Beckman AN60 Ti rotor and fits (lines). Key: CC-pIL, solid black lines; CC-pII, solid red lines; CC-pLI, solid blue lines; CC-pIL-I17N, broken black lines; and CC-pII-I13N, broken red lines. All samples were in phosphate-buffered saline (PBS) at pH 7.4. For panels A, B, and C, 50 μM peptide concentrations were used. For panel D concentrations were between 75 and 400 μM, giving an initial A_{280} of 0.4 AU. AUC data for CC-pII-I13N are given in Supplementary Figure S5.

Table 2. Biophysical Parameters for the Designed Coiled-Coil Peptides[a]

systematic name	basis-set name	MRE_{222} (deg cm^2 dmol^{-1}·res^{-1})	T_M (°C)	$K_{T,1/2}$ (M) (T = 20 °C); (T = 37 °C)	n solution	n X-ray crystallography (PDB ID)
CC-pIL		−38179	>90 (77.44)	n/d	3	3 (4DZN*)
CC-pIL-I17N	CC-Di	−37844	78.15	6.67×10^{-11}; 5.54×10^{-9}	2	2 (4DZM*)
CC-pII	CC-Tri	−39293	>90 (60.09)	n/d	3	3 (4DZL*)
CC-pII-I13N	CC-Tri-N13	−32232	56.47	1.74×10^{-8}; 9.42×10^{-7}	3	3 (4DZK)
CC-pLI	CC-Tet	−32808	>90 (63.18)	n/d	4	4 (3R4A)

[a]MRE values are for peptide at 50 μM peptide concentration in PBS at 20 °C. T_M values are for the mid-point of thermal denaturation at 50 μM peptide concentration in PBS, and values in parentheses indicate T_M in the presence of 3 M guanidine hydrochloride. $K_{1/2,T}$ is the concentration at which the complex is half folded, as judged by MRE_{222}, for the given temperature. n Refers to the number of helices in the assemblies. PDB Codes marked with an asterisk are for the 4-iodophenylalanine derivative.

had helical and superhelical parameters characteristic of classical coiled-coil assemblies (Supplementary Table S7).

In more detail, the core-packing angles made by the ***a*** and ***d*** "knob" residues into their respective "holes" for all three

dx.doi.org/10.1021/sb300028q | *ACS Synth. Biol.* 2012, 1, 240−250

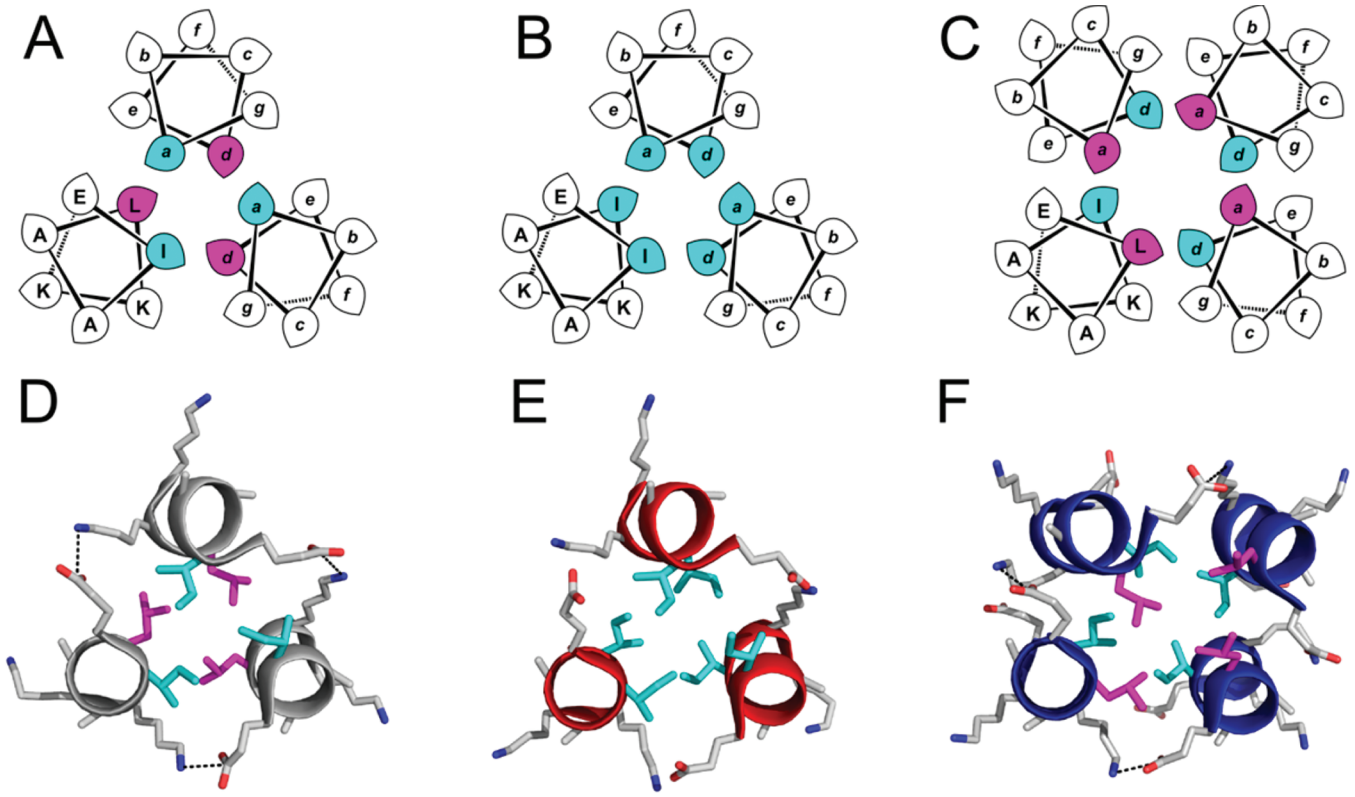


Figure 3. Structures formed by the designed coiled-coil peptides CC-pIL, CC-pII and CC-pLI. (A−C) Helical-wheel diagrams for trimeric and tetrameric coiled coils, showing sequences and heptad assignments. One-heptad slices through the X-ray crystal structures of CC-pIL (D, gray ribbon, PDB ID 4DZN), CC-pII (E, red, PDB ID 4DZL), and CC-pLI (F, blue, PDB ID 3R4A). Key features include the knobs-into-holes interactions formed by the isoleucine (cyan) and leucine residues (magenta) at the ***a***/***d*** positions and sub-4 Å salt-bridge interactions between Glu at ***g*** and Lys at the following ***e*** positions (shown as broken black lines). All images were created in PyMol (http://www.pymol.org), are viewed along the axis of the bottom-left helix from the N- to C-terminus, and cover one complete heptad register, ***gabcdef***.

structures fitted with the distributions measured for other classical coiled coils of known structure (Supplementary Figure S7). Specifically, the core residues in CC-pIL and CC-pII had "acute" packing arrangements typical of parallel trimers, and although the structure is less symmetrical, those in CC-pLI had alternating "perpendicular" and "parallel" packing at ***a*** and ***d***, respectively.

As CC-pII and CC-pLI conformed to our design expectations, we renamed these CC-Tri and CC-Tet, respectively, to signify that they form part of our designed coiled-coil basis set, Table 1. As stated above, CC-pIL unexpectedly formed a classical coiled-coil trimer with canonical knobs-into-holes interactions. This contrasts with the same ***a*** = Ile ***d*** = Leu sequence in the GCN4-p1 background, which forms a parallel dimer in solution.[31] N.B.: A crystal structure for this variant has not been reported. Therefore, we turned to the task of engineering dimer specificity into the CC-pIL foreground and to our initial goal of making a true Basis-set dimer, CC-Di.

Negative Design: Specifying Dimer in CC-pIL. The problem of specifying a particular oligomer state can be addressed by *negative design*, which entails energetically destabilizing any unwanted states relative to the target state, which may itself be destabilized to some extent by the change. Ideally, we wanted a single point mutation, or at least a small number of such mutations, to effect the switch from trimer to dimer in CC-pIL. In other words, we required an amino acid with a large oligomer-state discrimination factor (ODF) between parallel dimer and trimer when placed at a specific site in the sequence.

To inform our choice, we turned to the CC+ database of structurally verified coiled coils.[25] We collated all examples of canonical (i.e., heptad-based), parallel, homodimers and -trimers that were >21 residues long and had <50% sequence identity. This gave sets of 90 dimers and 43 trimers, with ∼500 and ∼220 heptad repeats in total, respectively (Supplementary Table S2). We focused on residues at the ***a*** and ***d*** sites, as mutations here were anticipated to have the largest effects on oligomer state, and selected only those that occurred at >5% in either of the ***a*** or ***d*** sites in the dimer or trimer sets; this was an arbitrarily chosen cutoff, which nonetheless covered 80% and 88% of the counts at the ***a*** and ***d*** sites in the dimer and trimer sets, respectively. This gave the following reduced set of preferred side chains at one or both of the sites in coiled-coil dimers and trimers: Ala, Ile, Lys, Leu, Asn (asparagine, N), Gln, and Val (valine, V). Next, we calculated the ODFs for these residues at both positions, Figure 4. We define ODF as the $\log_{10}$ of the ratio of the normalized percentages of the amino acid at the particular site in the dimer and the trimer (Supplementary Table S2). We find this so-called log-odds scoring useful as positive ODF values reveal preferences for dimer, and negative values are indicative of trimer.[36] It is important to stress, however, that this analysis assesses the impact of individual residues on oligomer state, and ignores any pairwise or higher-order side-chain interactions.[37] Despite this limitation, this method highlights first-order effects and directs straightforward experiments.

The ODF values are intriguing in themselves for a number of reasons. First, Leu, which is the most frequently found residue

dx.doi.org/10.1021/sb300028q | *ACS Synth. Biol.* 2012, 1, 240−250

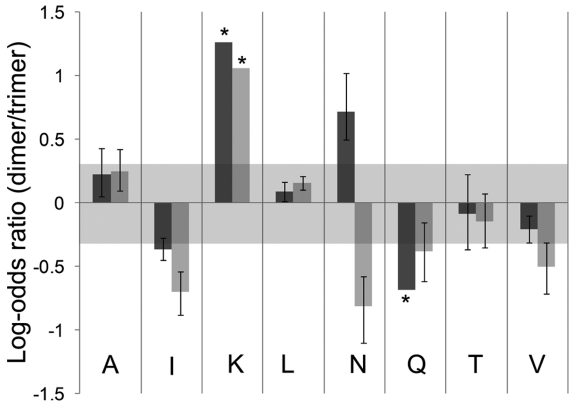


Figure 4. Oligomer-state discrimination factors (ODFs) for individual amino acids in parallel, homomeric coiled-coil dimers and trimers. ODFs were calculated as the $\log_{10}$ of the ratio of the normalized percentages of each amino acid at the specified positions: ***a*** (black bars) and ***d*** (gray bars), in the dimer and trimer data sets. The shaded region highlights ODFs in the range +0.3 and −0.3, i.e., preferences for dimer and trimer, respectively, of no more than twice the alternative oligomer state. Propensity is defined as the ratio of observed residue counts at a defined register position, divided by the expected count for the same residue at the same position, where the latter is calculated by a standard contingency table analysis, Supplementary Table S2. In this way, biases due to the different sizes and compositions of the data sets are accounted for, although we found the final relative values largely insensitive to the normalization procedure. N.B.: There were no examples of Lys at ***a*** or ***d*** in trimers; however, given the high occurrence of Lys at ***a*** in dimers (6%), we felt it important to assess the ODFs for this residue. Thus, to avoid infinite (or zero) ODFs, a dummy count of 1 was added to the raw count for every residue. Therefore, the absolute magnitude of the ODFs for Lys is dictated by this dummy count, and these values should only be qualitatively compared to other ODFs. Errors on the count data for the dimers and trimers were estimated using the normal distribution and propagated through the ODF calculation, with the error bars shown indicating 1 standard deviation. Starred data shown without error bars failed to meet the criteria for the normal approximation to the binomial distribution, as a result of one or more of the contributing count data elements being less than 5. The method was also extended to the ***e*** and ***g*** positions, for which the ODF values were smaller, reflecting the lower steric contribution of these residues to the binding surface, Supplementary Figure S8. Furthermore, as an aid to design, the ODF values for each position can be correlated to the coiled-coil propensity of a given residue (Supplementary Figure S9). Finally, there were insufficient data for natural parallel coiled-coil tetramers to extend the analysis to this state.

in both coiled-coil data sets, appears to be oligomer-state agnostic at both ***a*** and ***d***. The β-branched amino acids, Ile and Val, are non-discriminating at ***a*** but favor the trimer when placed at ***d***. This immediately explains why the ***a*** = Ile, ***d*** = Leu foreground of CC-pIL does not necessarily favor dimers, and considering the earlier studies on GCN4-p1,[31] it indicates that other factors within that background render GCN4-pIL a dimer. Nevertheless, of those selected for this study and Harbury's work on GCN4, this foreground appears to be one of the best starting points for making a dimer, because anything with Ile (or Val) at ***d*** strongly favors a trimer. Alternative foregrounds are ***a*** = Leu, ***d*** = Leu or ***a*** = Val, ***d*** = Leu, but considering the low overall specificity conferred by all of these ***a*** and ***d*** combinations the ***a*** = Ile, ***d*** = Leu foreground was retained, and dimer specificity was sought through point mutation of the CC-pIL peptide.

Returning to the question of what point mutation to make to switch CC-pIL from trimer to dimer, the choice was either to incorporate Asn at ***a*** or Lys at ***a*** or ***d***, as only these residue placements had ODF ≫ 0.3, i.e., strongly favoring dimers, Figure 4. We chose the former for several reasons. First, Asn occurs more frequently than Lys at the ***a*** and ***d*** sites of coiled coils overall; second, it has a higher coiled-coil propensity overall compared with Lys (Supplementary Figure S9); third, and interestingly, the other placement, Asn at ***d***, favors the trimer, which we return to below; and finally, it is known empirically that Asn at ***a*** is strongly associated with dimeric states.[38−40] To have the maximum impact on the structure and consistent with the conserved positioning of Asn at the ***a*** site of the third heptad repeat of leucine-zipper peptides including GCN4, we made the Ile-to-Asn mutation in the middle of the sequence at the ***a*** site of the third heptad repeat of CC-pIL to give CC-pIL-I17N, Table 1. N.B.: we have also found substitution at this position to be optimal for the CC-pIL system, as we will describe elsewhere (Fletcher *et al.*, unpublished data).

At low temperatures, the CC-pIL-I17N peptide was as helical as its parent, Figures 2A and B. However, the mutant was less stable to heat, showing a classical sigmoidal and reversible thermal unfolding transition with a midpoint of 78 °C at 50 μM peptide concentration, Figure 2B. This is consistent with the anticipated destabilizing effect of including a polar side chain within the otherwise hydrophobic core of the peptide assembly. Moreover, AUC analysis revealed that CC-pIL-I17N was dimeric in solution, Figure 2D, and the X-ray crystal structure confirmed the desired parallel dimer, Figure 5. Thus, we achieved CC-Di with two design iterations.

The X-ray crystal structure of CC-Di is noteworthy because the asparagine side chains are not involved in any hydrogen bonding contacts with each other, Figure 5A and B. Related to this, we have found that about two-thirds of the similarly buried Asn side chains in parallel coiled-coil dimers make interhelical hydrogen bonds (Supplementary Figure S10). This suggests that their influence in specifying oligomer state is more complicated than has often been argued[16,38−40] and that this is not necessarily mediated by well-defined interhelical interactions. Put another way, in the absence of any hydrogen-bonded interactions, burying the amide components of Asn side chains destabilizes all homo-oligomeric coiled coils, but the penalty is expected to be less severe for dimers, which partly bury only two such side chains, compared with higher-oligomer states that bury more. These arguments carry the caveat that there may be discrepancies between the fundamental dynamic nature of protein structures and the snap shots of these afforded through X-ray crystallography.

Second, and returning to the discussion of the $\boldsymbol{g}_h$:$\boldsymbol{e}'_{h+1}$ Glu:Lys salt-bridge interactions, in the parent, CC-pIL trimer, all possible 12 interactions are made, Figure 3D. However, in CC-Di, only 2 of the possible 8 interactions are made, and this is only in one of the two dimers observed in X-ray crystal structure (Figure 5A and B and Supplementarya Table S8); all of the other C_δ to N_ζ distances are ≥4.5 Å. Thus, in CC-pIL, which from the ODFs of Figure 4 have an oligomer-state-agnostic core, the balance may be tipped toward the trimer state because this allows closer contact of the $\boldsymbol{g}_h$ and $\boldsymbol{e}'_{h+1}$ residues and salt-bridge formation.

Revisiting GCN4-pIL. As mentioned above, an X-ray structure for GCN4-pIL has not been reported. However, solution-phase data indicate that the peptide is dimeric.[31] This

dx.doi.org/10.1021/sb300028q | *ACS Synth. Biol.* 2012, 1, 240−250

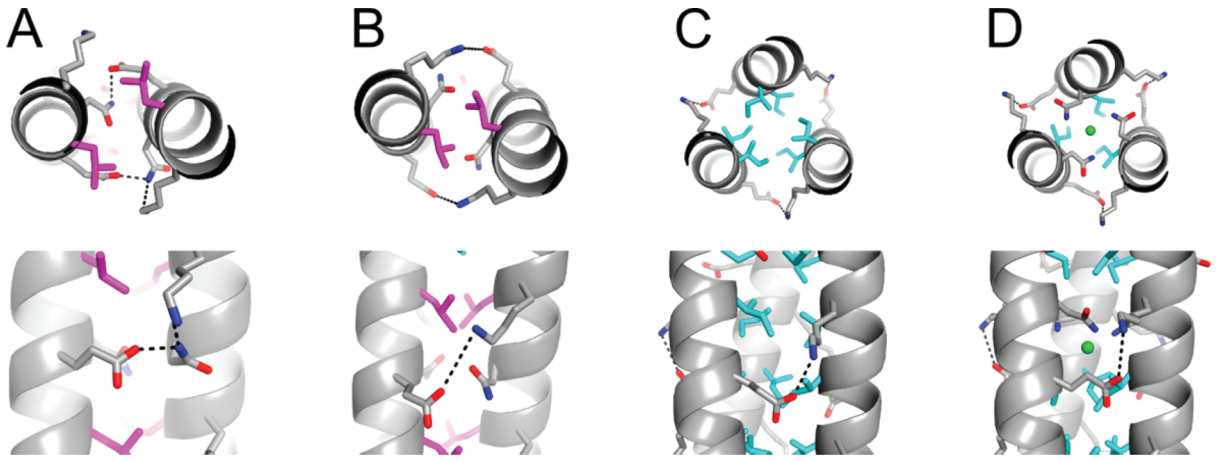


Figure 5. X-ray crystal structures of the basis-set dimer and trimer peptides. (A, B) CC-Di. (C) CC-Tri. (D) CC-Tri-N13. The two nonidentical dimer assemblies from the CC-pIL-I17N unit cell (PDB ID 4DZM) are shown in panels A and B for completeness. Putative hydrogen bonds, taking into account side-chain flexibility, are indicated with a dashed line, with distances ranging from 2.4 to 3.5 Å. For the polar side chains CPK coloring is used, while Leu and Ile side chains are colored magenta and cyan, respectively. Structural images were created using PyMol (http://www.pymol.org).

contrasts with our findings for ***a*** = Ile, ***d*** = Leu, albeit in a different background, and interestingly with those on another related design sequence, essentially $(IAALEQK)_4$, which also forms a parallel trimer in solution.[41] Given the importance of the GCN4 system in the development of the coiled-coil and protein-folding fields, its applications in protein engineering and design, and the necessity for robust design principles, we felt it important to try to reconcile these differences. Therefore, we synthesized GCN4-pIL (Supplementary Table S3) and characterized it in solution (Supplementary Figures S2 and S5). As expected the peptide was fully helical and stable in solution. However, we found that it sedimented with a molecular weight somewhere between that of dimer and trimer, depending on concentration. This ambiguity was also consistent with our inability to solve the crystal structure of this peptide: the space group of the crystal, and the asymmetric unit content were equally consistent with dimers or trimers in the crystal lattice. Together these observations suggest that the oligomerization state adopted by GCN4-pIL may be more condition-dependent than originally thought and that the energetic difference between dimeric and trimeric states for this particular sequence is small.

Negative Design: Modulating the Stabilities of CC-Tri and CC-Di. Interestingly, aside from Ile or Val at ***d***, the single-residue placement that most favors trimers is Asn at ***d***, Figure 4. Again, this has recently been corroborated experimentally.[42] Therefore, we explored the inclusion of a single Asn at a ***d*** position in CC-Tri: in this case, Ile-13 in the second heptad was mutated to give CC-Tri-N13, Table 1. As with CC-Di, this peptide was less stable than its parent, showing co-operative thermal denaturation with a T_M of 56.5 °C at 50 μM concentration. In this case, however, a change in oligomer state was not expected and, indeed, was not observed: the peptide was trimeric in solution by AUC, Table 2, and the X-ray crystal structure confirmed a parallel trimer, Figure 5D. The Asn side chains are accommodated without significant steric penalty but bind a chloride ion. Similar constellations have been observed in both natural and engineered coiled coils[42] and presumably help offset some but not all of the energetic penalty of including asparagine in the hydrophobic core. In summary, combined these studies lend weight to the idea that this residue placement helps further to specify oligomer state, ameliorates the stability of an all-hydrophobic core, and may possibly help in partner selection. All in all, and like specific positioning of Asn at ***a*** in dimers, it provides a further rule-of-thumb for tuning trimeric coiled-coil interactions.

Developing this last point, an important requirement for using the coiled-coil basis set in protein engineering and synthetic biology will be the ability to adjust complex stabilities to suit different applications. The high stabilities of CC-pIL, CC-Tri, and CC-Tet mitigate this; essentially, these peptides are oligomeric under all practically accessible concentrations. The inclusion of Asn at ***a*** and ***d*** of CC-Di and CC-Tri-N13, respectively, ameliorates their stabilities. Conventionally, the stabilities of biomolecular complexes are described in terms of a dissociation constant, K_D, i.e., the concentration at which the molecules are half associated and half dissociated. However, for states above dimer, this term is not simply that concentration but the concentration raised to the power $n - 1$, where n is the oligomeric state. Therefore to quantify oligomer-state stability more practically, we measured the midpoint of thermal unfolding (T_M) for CC-Di and CC-Tri-N13 over a range of concentrations (Supplementary Figure S11 and Table S9). From these data, the concentration at which the complex is half formed for any given temperature, $K_{1/2,T}$,[43] could be determined. For example, for CC-Di and CC-Tri-N13 at 37 °C, the values were in the nanomolar and micromolar regimes, respectively, Table 2. We argue that this parameter will be more useful to practicing protein engineers or synthetic biologists.

Conclusions, Implications, and Applications. The studies described herein provide the foundations for a toolkit or a basis set of *de novo* designed coiled-coil peptides of specified oligomer state, namely, parallel dimer (CC-Di), trimer (CC-Tri), and tetramer (CC-Tet). The designs incorporate previously known rules for oligomer-state selection that center on the hydrophobic core residues ***a*** and ***d*** placed within a designed and presumed oligomer-state-agnostic background. These studies confirm that ***a*** = ***d*** = Ile and ***a*** = Leu and ***d*** = Ile are good rules for specifying parallel trimeric and tetrameric coiled coils, respectively, as reported previously.[31,32] Moreover, our work demonstrates that these rules are robust in that they can be transferred from natural to fully designed backgrounds. However, and surprisingly, our initial design iteration for the targeted dimer, CC-pIL, reveals that a foregoing rule-of-thumb,

a = Ile and *d* = Leu, is not sufficient in itself to specify dimer; the peptide is trimeric both in solution and the crystal state. To solve this problem, based on the literature and analysis of a database of known coiled-coil structures, we made a single Ile-to-Asn substitution at a central *a* position, and the resulting peptide, CC-Di, is a parallel dimer.

The finding that the oligomeric state of a ∼30-residue peptide can be switched by a single-residue change indicates that the energetic difference between some coiled-coil states is small. This implies that the free-energy landscape of coiled-coil assemblies is complex and is unlikely to be typified by only few distinct minima. In other words, some effort has to be made to locate stable minima for desired states. Defining this energy landscape better will be non-trivial. We envisage that the coiled-coil basis set and variants within it will contribute to mapping this energy landscape.

These studies affirm that the set of CC-Di, CC-Tri, and CC-Tet provides a robust set of stand-alone peptides that are characterized to a very high level in terms of sequence, solution-phase behavior and 3D structure, thus meeting many of our initial requirements for a toolkit of peptide components for protein engineering and synthetic biology. Moreover, the basis set has the potential to be used to provide oligomerization domains for a variety of the chemical and synthetic biology studies. Indeed, we have already shown that CC-Tri acts as a hub to nucleate the folding and assembly of bacterial collagen sequences;[30] that CC-Tet presents a scaffold to explore larger oligomeric states;[22] and that CC-Di and CC-Tri can be used to study fundamental sequence-to-structure relationships in coiled coils (Fletcher *et al.*, unpublished).

Toward a Registry of Peptide Components (*Pcomp*) for Synthetic Biology. Finally, and as an aid to protein engineers and synthetic biologists, we have constructed a database, *Pcomp* (http://coiledcoils.chm.bris.ac.uk/pcomp/). This is a repository for structural and functional information for peptide components, with the coiled-coil basis set as its first occupants. The database is searchable by component structural type and houses data relating to component synthesis and biochemical, structural, and functional characterizations. In addition, *Pcomp* provides downloadable datasheets summarizing the full characterization of the components (Supplementary Figure S12). The overall aim of *Pcomp* is to provide all of the information necessary on robust synthetic components that might be used in a "plug-and-play" fashion to construct more complex systems, i.e., to perform a similar role to that of the Registry of Standard Biological Parts offered by the BioBricks foundation.[44] The database is freely accessible, and we welcome input and contributions from others.

METHODS

General. All Fmoc-amino acid derivatives, coupling agents, and peptide grade DMF were purchased from AGTC Bioproducts (Hessle, U.K.), with the exception of Fmoc-L-4-iodophenylalanine, which was purchased from GL Biochem (Shanghai, PRC). Trifluoroacetic acid and piperidine were purchased from Fisher Scientific (Loughborough, U.K.). Mass spectrometry was carried out on an Applied Biosystems 4700 Proteomics Analyzer MALDI instrument. All experiments were carried out in phosphate-buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, and 10 mM phosphate buffer at pH 7.4), unless otherwise stated. Crystallographic screens were purchased from Molecular Dimensions Ltd. (Suffolk, U.K.). X-ray diffraction data were collected at Diamond Light Source Ltd., Harwell Campus, Didcot, Oxfordshire, U.K.

Peptide Synthesis. All peptides were synthesized on a 0.1 mmol scale on Rink amide chemmatrix resin. Synthesis was conducted *via* standard Fmoc protocols on a CEM microwave-assisted peptide synthesizer, making use of HBTU activation. Prior to cleavage from the solid support the peptide was N-terminally acetylated using acetic anhydride/pyridine in DMF. Cleavage from the resin was effected with a 95:5:5 mix (10 mL) of trifluoroacetic acid, triisopropylsilane, and water. The resin was washed with additional trifluoroacetic acid (5 mL), and the combined acid extracts were concentrated to approximately 5 mL under a flow of nitrogen. The crude peptide was then precipitated by addition of cold diethyl ether (40 mL) to give an off-white precipitate that was isolated by centrifugation. The precipitate was redissolved in ca. 10 mL of a 1:1 mix of acetonitrile and water prior to freeze-drying to yield a fine white solid. Peptides were purified by reverse-phase HPLC using a Kromatek C18HQsil column (150 by 10 mm) running a linear gradient of acetonitrile and water, each containing 0.1% TFA. A typical gradient ran from 20% to 80% acetonitrile over 30 min.

Circular Dichroism Spectroscopy. CD spectra were obtained using a JASCO J-810 spectropolarimeter fitted with a Peltier temperature controller. Peptide concentrations were determined by UV absorption (ε_{280}(Trp) = 5690 mol^{-1} cm^{-1}; ε_{280}(Tyr) = 1280 mol^{-1} cm^{-1}). Peptide solutions were prepared in PBS and examined in 1 mm (50−100 μM samples), 5 mm (5−10 μM), and 10 mm (50−100 nM) quartz cuvettes. Thermal denaturation experiments were performed by ramping temperature from 5 to 90 °C at a rate of 10 °C h^{-1}. Full spectra were recorded at 5 °C intervals, and the CD at 222 nm was recorded at 1 °C intervals (1 nm interval, 1 nm bandwidth, 16 s response time). The midpoint of thermal denaturation (T_M) values were determined from the second derivative of the variable temperature slope.

Analytical Ultracentrifugation. Sedimentation equilibrium experiments were conducted at 20 °C in a Beckman-Optima XL-I analytical ultracentrifuge using an An-60 Ti rotor. Solutions were prepared in PBS with peptide concentrations in the range 75−400 μM and spun at speeds in the range 20,000−50,000 rpm. Data sets were initially fitted to a single, ideal species model using Ultrascan (http://www.ultrascan.uthscsa.edu/). The partial specific volume for each of the various peptides and the solvent density was calculated using Sednterp (http://www.jphilo.mailway.com/download.htm).

Dynamic Light Scattering. Measurements were made using a Malvern Zetasizer Nanoseries instrument. Samples were prepared at 100 μM peptide concentration in PBS. Samples were centrifuged prior to analysis to remove any large particulate material. Measurements were made at 20 °C using automated settings. The data were analyzed using the associated DTS Nano particle sizing software.

X-ray Crystallography: Crystallization. All peptides were crystallized using the sitting drop vapor diffusion method at 18 °C. For crystallization peptide solutions were prepared unbuffered in ultrapure water at the following concentrations: CC-pIL-I17N 8 mg mL^{-1}, CC-pIL 8 mg mL^{-1}, CC-pII-I13N 20 mg mL^{-1}, CC-pII 15 mg mL^{-1}, CC-pLI 10 mg mL^{-1}. For preparation of halide-derivative crystals iodo-phenylalanine containing peptides were used. Initial screens were generally carried out using a sparse matrix approach with a total number of 480 conditions of commercially available screens. Screens were prepared in 96-well MRC plates using the robotic Phoenix

Liquid Handling System (Art Robbins, Sunnyvale, CA, USA) with reservoir volumes of 50 μL and peptide drop volumes of 200 nL mixed with 200 nL of reservoir solution. Diffraction quality crystals were usually obtained within 1−4 weeks with optimization of initial crystal hits carried out as necessary. Final crystallization conditions for all peptides can be found in Supplementary Table S4.

X-ray Crystallography: Data Collection and Processing. Complete sets of diffraction data for all peptides were collected from loop mounted single crystals under cryo conditions at 100 K using a ADSC Q315 CCD detector at synchrotron beamlines IO2 or IO4 of Diamond Light Source Ltd., Harwell Campus, Didcot, Oxfordshire, U.K. Prior to loop mounting crystals were soaked in appropriate cryoprotectant solutions (usually 20−30% PEG400 or glycerol) where necessary. Data collection was carried out either at a wavelength of 0.98 Å for native data sets or at 1.7 Å for collection of anomalous data from iodide-derivatized crystals (Supplementary Table S5). Data indexing, integration and scaling were carried out for CC-pII-I13N and CC-pIL with MOSFLM[45] and SCALA,[46] for CC-pII with HKL-2000[47] or for CC-pIL-I17N with XDS.[48] Data collection statistics can be found in Supplementary Table S5.

X-ray Crystallography: Structure Solution, Model Building, and Refinement. Initial phases for CC-pII-I13N were obtained from molecular replacement (MR) phasing using PHASER[49] with a previously solved coiled-coil X-ray structure (PDB ID 1BB1) as the search model. For experimental phasing of CC-pIL-I17N halide sites were identified from anomalous data collected at 1.7 Å using SHELXD/E.[50] Similarly, for CC-pII SAD experimental phasing and chain tracing were successfully carried out with AutoSol in PHENIX.[51] In the case of CC-pIL attempts for SAD phasing failed using either SHELX or PHENIX. Since diffraction data for CC-pIL extended to 1.6 Å, a structure-solution strategy was employed that combined MR using PHASER with CC-pIL-I17N as search model. Molecular replacement was followed by *ab initio* phasing in ACORN[52] to remove model bias. During model building and refinement it became apparent that, probably due to radiation damage during data collection, not all iodide sites were fully occupied, indicating why experimental phasing using SAD data may have failed. After obtaining initial phases, further automated density modification and chain tracing using BUCCANEER[53]/PARROT[54] was carried out followed by ARP/WARP[55] or AutoBuild in PHENIX for automated model building, where applicable. Subsequent model completion was carried out by iterative manual modeling in COOT[56] and automated refinement in REFMAC5.[57] In the final REFMAC5 stages TLS refinement[58] was used for CC-pIL, CC-pIL-I17N, and CC-pII. TLS parameters were obtained from the TLSMD server[59] generally using one TLS group per chain. TLS refinement was run such that output PDB files contained full B-factors. Final models were checked with PROCHECK,[60] COOT, and MOLPROBITY[61] exhibiting good geometries with no outliers in the Ramachandran plot. A summary of the refinement and model building statistics together with PDB accession codes can be found in Supplementary Table S6. All figures for the crystallographic models were prepared using PyMOL (http://www.pymol.org). Structure solution of CC-pLI has been reported previously.[22]

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

This material is available free of charge *via* the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +44 (0) 117 95 46347. Fax: +44 (0) 117 929 8611. E-mail: Drew.Thomson@bristol.ac.uk; D.N.Woolfson@bristol.ac.uk.

Present Addresses

[§]Department of Genetic Medicine, Weill Cornell Medical College, 1305, York Avenue, New York, NY 10021, U.S.A.

[∥]Thomas C. Jenkins Department of Biophysics, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, U.S.A.

[⊥]Department of Physics, Durham University, South Road, Durham DH1 3LE, U.K.

Author Contributions

[#]These authors contributed equally to this work.

Author Contributions

All authors contributed to the overall basis-set concept and to the experimental design; J.M.F. and A.R.T. synthesized the peptides; J.M.F., A.L.B., and E.H.C.B. conducted the solution-phase biophysical experiments; M.B., N.R.Z., and R.L.B. solved the X-ray crystal structures; G.J.B., T.L.V., C.T.A., and D.N.W. performed the bioinformatic analyses; A.R.T. and D.N.W. wrote the majority of the main manuscript; P.J.B., R.L.B., and D.N.W. supervised the work; all authors assembled the Supporting Information.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Noah Linden and Richard Sessions for helpful discussions on thermodynamics. We are grateful to the BBSRC of the U.K. for grants to D.N.W. and P.J.B. (BB/G008833/1) and to R.L.B. and D.N.W. (BB/F007256/1). We thank the staff at Diamond Light Source for access to synchrotron radiation for the crystallographic analyses.

■ REFERENCES

(1) Purnick, P. E. M., and Weiss, R. (2009) The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol. 10*, 410−422.

(2) Benner, S. A., and Sismour, A. M. (2005) Synthetic biology. *Nat. Rev. Genet. 6*, 533−543.

(3) Bromley, E. H. C., Channon, K., Moutevelis, E., and Woolfson, D. N. (2008) Peptide and protein building blocks for synthetic biology: From programming biomolecules to self-organized biomolecular systems. *ACS Chem. Biol. 3*, 38−50.

(4) Channon, K., Bromley, E. H. C., and Woolfson, D. N. (2008) Synthetic biology through biomolecular design and engineering. *Curr. Opin. Struct. Biol. 18*, 491−498.

(5) Armstrong, C. T., Boyle, A. L., Bromley, E. H. C., Mahmoud, Z. N., Smith, L., Thomson, A. R., and Woolfson, D. N. (2009) Rational design of peptide-based building blocks for nanoscience and synthetic biology. *Faraday Discuss. 143*, 305−317.

(6) Thompson, K. E., Bashor, C. J., Lim, W. A., and Keating, A. E. (2012) SYNZIP protein interaction toolbox: *in vitro* and *in vivo* specification of heterospecific coiled-coil interaction domains. *ACS Synth. Biol. q*, 118−129.

(7) Stites, W. E. (1997) Protein-protein interactions: Interface structure, binding thermodynamics, and mutational analysis. *Chem. Rev. 97*, 1233−1250.
(8) Nooren, I. M. A., and Thornton, J. M. (2003) Diversity of protein-protein interactions. *EMBO J. 22*, 3486−3492.
(9) Krizek, B. A., Amann, B. T., Kilfoil, V. J., Merkle, D. L., and Berg, J. M. (1991) A consensus zinc finger peptide : design, high-affinity metal-binding, a pH-dependent structure, and a His to Cys sequence variant. *J. Am. Chem. Soc. 113*, 4518−4523.
(10) O'Leary, L. E. R., Fallas, J. A., and Hartgerink, J. D. (2011) Positive and negative design leads to compositional control in AAB collagen heterotrimers. *J. Am. Chem. Soc. 133*, 5432−5443.
(11) Berg, J. M., and Godwin, H. A. (1997) Lessons from zinc-binding peptides. *Annu. Rev. Biophys. Biomol. Struct. 26*, 357−371.
(12) Boyle, A. L., and Woolfson, D. N. (2011) De novo designed peptides for biological applications. *Chem. Soc. Rev. 40*, 4295−4306.
(13) Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science 302*, 1364−1368.
(14) Fleishman, S. J., Whitehead, T. A., Ekiert, D. C., Dreyfus, C., Corn, J. E., Strauch, E. M., Wilson, I. A., and Baker, D. (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science 332*, 816−821.
(15) Smith, B. A., and Hecht, M. H. (2011) Novel proteins: from fold to function. *Curr. Opin. Chem. Biol. 15*, 421−426.
(16) Woolfson, D. N. (2005) The design of coiled-coil structures and assemblies. *Adv. Protein Chem. 70*, 79−112.
(17) Lupas, A. N., and Gruber, M. (2005) The structure of alpha-helical coiled coils. *Adv. Protein Chem. 70*, 37−78.
(18) Moutevelis, E., and Woolfson, D. N. (2009) A Periodic table of coiled-coil protein structures. *J. Mol. Biol. 385*, 726−732.
(19) Rackham, O. J. L., Madera, M., Armstrong, C. T., Vincent, T. L., Woolfson, D. N., and Gough, J. (2010) The evolution and structure prediction of coiled coils across all genomes. *J. Mol. Biol. 403*, 480−493.
(20) Crick, F. H. C. (1953) The packing of alpha-helices - simple coiled-coils. *Acta Crystallogr. 6*, 689−697.
(21) Walshaw, J., and Woolfson, D. N. (2001) SOCKET: A program for identifying and analysing coiled-coil motifs within protein structures. *J. Mol. Biol. 307*, 1427−1450.
(22) Zaccai, N. R., Chi, B., Thomson, A. R., Boyle, A. L., Bartlett, G. J., Bruning, M., Linden, N., Sessions, R. B., Booth, P. J., Brady, R. L., and Woolfson, D. N. (2011) A de novo peptide hexamer with a mutable channel. *Nat. Chem. Biol. 7*, 935−941.
(23) Walshaw, J., and Woolfson, D. N. (2003) Extended knobs-into-holes packing in classical and complex coiled-coil assemblies. *J. Struct. Biol. 144*, 349−361.
(24) O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) X-Ray structure of the GCN4 leucine zipper, a 2-stranded, parallel coiled coil. *Science 254*, 539−544.
(25) Testa, O. D., Moutevelis, E., and Woolfson, D. N. (2009) CC plus: a relational database of coiled-coil structures. *Nucleic Acids Res. 37*, D315−D322.
(26) Schreiber, G., and Keating, A. E. (2011) Protein binding specificity versus promiscuity. *Curr. Opin. Struct. Biol. 21*, 50−61.
(27) Bromley, E. H. C., Sessions, R. B., Thomson, A. R., and Woolfson, D. N. (2009) Designed alpha-helical tectons for constructing multicomponent synthetic biological systems. *J. Am. Chem. Soc. 131*, 928−930.
(28) Gradisar, H., and Jerala, R. (2011) De novo design of orthogonal peptide pairs forming parallel coiled-coil heterodimers. *J. Pept. Sci. 17*, 100−106.
(29) Reinke, A. W., Grant, R. A., and Keating, A. E. (2010) A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. *J. Am. Chem. Soc. 132*, 6025−6031.
(30) Yoshizumi, A., Fletcher, J. M., Yu, Z. X., Persikov, A. V., Bartlett, G. J., Boyle, A. L., Vincent, T. L., Woolfson, D. N., and Brodsky, B. (2011) Designed coiled coils promote folding of a recombinant bacterial collagen. *J. Biol. Chem. 286*, 17512−17520.
(31) Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) A switch between 2-stranded, 3-stranded and 4-stranded coiled coils in Gcn4 leucine-zipper mutants. *Science 262*, 1401−1407.
(32) Harbury, P. B., Kim, P. S., and Alber, T. (1994) Crystal-structure of an isoleucine-zipper trimer. *Nature 371*, 80−83.
(33) Xie, J. M., Wang, L., Wu, N., Brock, A., Spraggon, G., and Schultz, P. G. (2004) The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nat. Biotechnol. 22*, 1297−1301.
(34) Dragan, A. I., and Privalov, P. L. (2002) Unfolding of a leucine zipper is not a simple two-state transition. *J. Mol. Biol. 321*, 891−908.
(35) Strelkov, S. V., and Burkhard, P. (2002) Analysis of alpha-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. *J. Struct. Biol. 137*, 54−64.
(36) Armstrong, C. T., Vincent, T. L., Green, P. J., and Woolfson, D. N. (2011) SCORER 2.0: an algorithm for distinguishing parallel dimeric and trimeric coiled-coil sequences. *Bioinformatics 27*, 1908−1914.
(37) Hadley, E. B., Testa, O. D., Woolfson, D. N., and Gellman, S. H. (2008) Preferred side-chain constellations at antiparallel coiled-coil interfaces. *Proc. Natl. Acad. Sci. U.S.A. 105*, 530−535.
(38) Lumb, K. J., and Kim, P. S. (1995) A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled-coil. *Biochemistry 34*, 8642−8648.
(39) Gonzalez, L., Woolfson, D. N., and Alber, T. (1996) Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nat. Struct. Biol. 3*, 1011−1018.
(40) Akey, D. L., Malashkevich, V. N., and Kim, P. S. (2001) Buried polar residues in coiled-coil interfaces. *Biochemistry 40*, 6352−6360.
(41) Zaytsev, D. V., Xie, F., Mukherjee, M., Bludin, A., Demeler, B., Breece, R. M., Tierney, D. L., and Ogawa, M. Y. (2010) Nanometer to millimeter scale peptide-porphyrin materials. *Biomacromolecules 11*, 2602−2609.
(42) Hartmann, M. D., Ridderbusch, O., Zeth, K., Albrecht, R., Testa, O., Woolfson, D. N., Sauer, G., Dunin-Horkawicz, S., Lupas, A. N., and Alvarez, B. H. (2009) A coiled-coil motif that sequesters ions to the hydrophobic core. *Proc. Natl. Acad. Sci. U.S.A. 106*, 16950−16955.
(43) Marky, L. A., and Breslauer, K. J. (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers 26*, 1601−1620.
(44) Smolke, C. D. (2009) Building outside of the box: iGEM and the BioBricks Foundation. *Nat. Biotechnol. 27*, 1099−1102.
(45) Powell, H. R. (1999) The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr., Sect. D 55*, 1690−1695.
(46) Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr., Sect. D 62*, 72−82.
(47) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol. 276*, 307−326.
(48) Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr., Sect. D 66*, 133−144.
(49) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr. 40*, 658−674.
(50) Sheldrick, G. M. (2008) A short history of SHELX. *Acta Crystallogr. Sect. A 64*, 112−122.
(51) Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D 66*, 213−221.
(52) Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Yao, J. X., and Zheng, C. D. (2000) A flexible and efficient procedure for the solution and phase refinement of protein structures. *Acta Crystallogr., Sect. D 56*, 1137−1147.

(53) Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr., Sect. D 62*, 1002−1011.
(54) Zhang, K. Y. J., Cowtan, K., and Main, P. (1997) Combining constraints for electron-density modification. *Methods Enzymol. 277*, 53−64.
(55) Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc. 3*, 1171−1179.
(56) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr., Sect. D 66*, 486−501.
(57) Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D 53*, 240−255.
(58) Painter, J., and Merritt, E. A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr., Sect. D 62*, 439−450.
(59) Painter, J., and Merritt, E. A. (2006) TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallogr. 39*, 109−111.
(60) Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) Procheck - a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr. 26*, 283−291.
(61) Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect. D 66*, 12−21.