Navigating the structural landscape of de novo α–helical bundles

Guto G. Rhys,Christopher W. Wood,Joseph L. Beesley,Nathan R. Zaccai,Antony J. Burton,R. Leo Brady,Andrew R. Thomson,Derek N. Woolfson

1School of Chemistry, University of Bristol, Cantock’s Close, Bristol BS8 1TS, United Kingdom
2School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol BS8 1TD, United Kingdom
3School of Chemistry, University of Glasgow, Glasgow G12 8QQ, United Kingdom
4Frick Chemistry Laboratory, Princeton University, Princeton, NJ 08544, USA
5BrisSynBio, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, United Kingdom

ABSTRACT: The association of amphipathic α helices in water leads to α-helical-bundle protein structures. However, the driving force for this—the hydrophobic effect—is not specific and does not define the number or the orientation of helices in the associated state. Rather, this is achieved through deeper sequence-to-structure relationships, which are increasingly being discerned. For example, for one structurally extreme but nevertheless ubiquitous class of bundle—the α-helical coiled coils—relationships have been established that discriminate between all-parallel dimers, trimers and tetramers. Association states above this are known, as are antiparallel and mixed arrangements of the helices. However, these alternative states are less-well understood. Here, we describe a synthetic-peptide system that switches between parallel hexamers and various up-down-up-down tetramers in response to single-amino-acid changes and solution conditions. The main accessible states of each peptide variant are characterized fully in solution and, in most cases, to high resolution with X-ray crystal structures. Analysis and inspection of these structures helps rationalize the different states formed. This navigation of the structural landscape of α-helical coiled coils above the dimers and trimers that dominate in nature has allowed us to design rationally a well-defined and hyperstable antiparallel coiled-coil tetramer (apCC-Tet). This robust de novo protein provides another scaffold for further structural and functional designs in protein engineering and synthetic biology.

INTRODUCTION

In nature, bundles of four α helices are common in protein structures and assemblies. These four-helix bundles perform a wide variety of functions including: acting as protein hormones and cytokines; providing scaffolds for metal- and co-factor-binding to facilitate storage, redox and enzymatic functions; cementing protein-protein interactions that direct DNA binding, membrane fusion and other processes; serving as building blocks for viral capsids; and spanning membranes to perform signal transduction and transport functions.

As such, four-helix bundles have become key targets and scaffolds for de novo protein design. Usually, these comprise amphipathic α helices. These helices assemble via their hydrophobic faces to form bundles with the hydrophobic side chains buried in consolidated cores. The associations can be of individual helices to form tetramers, of helical hairpins to give dimers, or intramolecularly within the same protein chain. However, and particularly for the intermolecular cases, alternative helical arrangements (e.g., all parallel, up-down antiparallel) and even other oligomeric states are possible. Therefore, the specification of helix-helix interactions that direct towards a specified structure and away from unwanted alternatives is critical for successful design; these two aspects are known as positive and negative design, respectively. Ultimately, what is required are clear sequence-to-structure relationships and/or robust computational methods to augment the amphipathic α-helical sequences and specify a target four-helix bundle. Here we explore the structural plasticity between all-parallel and up-down-up-down antiparallel four-helix coiled-coil structures and some of the other competing states.

Four-helix bundles are one of the earliest examples of a recognized protein-structure motif. Since then, structures of many four-helix bundles have enabled analyses of features that define the fold. Weber and Salemme expand on previous work, including Crick’s model for packing in α-helical coiled coils, to show that structural similarities between four-helix bundles are a consequence of basic physical properties of the component helices. Presnell and Cohen present further examples of four-helix bundles using computational screening and visual inspection of over 300 protein structures. This screening identifies 20 putative structures based on buried surface area, but when the 20 are examined visually, some are excluded as they “do not appear to be well packed”, with helix crossing angles diverging from extant four-helix bundles. These examples of true
four-helix bundles provide the basis for categorizing the fold by helix-crossing angles, with six categories emerging: square, splinter, x, undecorated, bicornate and splayed.

The square class contains the previously identified four-helix bundles, where all helices are aligned, while the other classes have packing more similar to previously identified α-helical globules.

The main differentiating factor between these structures is the number and regularity of interactions between helices.

Indeed, there is spectrum of side-chain-mediated helix-to-helix packing in four-helix bundles from few non-specific interactions to repeated regular interactions (Figure 1A). Short helices (< 14 residues) tend to be more free to associate in aligned or orthogonal configurations, resulting in a wide range of bundles. Whereas, longer helices (> 14 residues) favor aligned arrangements, packing optimally as extended bundles with regular and repeating interactions. Extended bundles can be categorized further based on the predominant mode of helix packing, from less-specific ridges-into-grooves (RIG) to intimate knobs-into-holes (KIH) interactions. These packing modes are characteristics of globular and coiled-coil domains, respectively.

Given their ubiquity and this level of understanding, four-helix bundles offer a fertile ground for the design and engineering of new proteins. DeGrado has pioneered this area initially, designing small synthetic peptides of leucine, glutamate and lysine to create self-assembling amphipathic helices, although these were later revealed to form larger globular bundles; then joining these together with proline- and arginine-based loops to generate hairpins that dimerize; and expressing single-chain four-helix bundles that are monomeric and hyperstable. Contemporaneously, but using a larger amino-acid palette, Richardson and co-workers designed a monomeric “native-like” four-helix protein. These bottom-up approaches have led to sophisticated de novo four-helix bundles with ion-transportation and cofactor-binding properties. For instance, Dutton and co-workers have also built four-helix bundle maquettes that incorporate heme to bind and transport oxygen. Similar scaffolds have been developed by others to introduce light-harvesting and enzyme-like functions.

Hecht and co-workers have taken a different approach to designing four-helix bundles. This pioneers the use of binary patterns of hydrophobic (h) and polar (p) residues to define amphipathic helices linked into single chains by intervening turns. Rather than specifying the sequences any further, however, redundant codons are used to place multiple different residues at the h and p sites; i.e., sequence libraries are generated to be compatible with the overall four-helix-bundle fold. Moreover, rather than actively selecting from or evolving these libraries, proteins that survive or operate in cells are picked out passively. In this way, the group have achieved stably folded and structured de novo four-helix bundles. In turn, these have been endowed with functions such as heme binding, abilities to substitute for deleted endogenous proteins, small-molecule binding, reducing copper toxicity, and, impressively, a de novo protein that catalyzes a life-sustaining reaction in a microorganism.

With some exceptions, these de novo four-helix bundles tend to have sequences that promote RIG packing of helices. Thus, while structural data is limited, it is likely that most de novo four-helix bundles have irregular side-chain packing, falling in the middle of the spectrum of Figure 1A. For more-intimate KIH packing and coiled-coil formation, two general features are required: (i) heptad repeats of h and p residues, hphpcpp (usually denoted abcddefg), or related repeats with

Figure 1. Four-helix bundles adopt a range of topologies with irregular to highly regular packing. From left to right, the structures are for a region of 3-isopropylmalate dehydrogenase (PDB ID code 1IPD) in the x class with only pairs of helices aligned; apolipoprotein E3 (PDB ID code 1LPE) in the square class, where all are helices aligned and interact via ridges-into-grooves packing; and CC-Tet (PDB ID code 3R4A) also in the square class but with knobs-into-holes helix-helix interactions. (B) The HAMP domain of TSR/AP1503, which forms an Alacoil (left and top right, PDB ID code 3Z6X), with the corresponding helical wheel showing helix packing (bottom right). (C) A designed hexameric coiled coil, CC-Hex (left and top right), also with a helical wheel of helix packing (bottom right).
Table 1. Sequences and summary of biophysical data for CC-Hex*-L24 point mutants.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Sequence and Register</th>
<th>Helix(^2) (%)</th>
<th>SV(^3) (mass/monomer mass)</th>
<th>SE(^4) (mass/monomer mass)</th>
<th>XRD(^5) oligomeric state</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC-Hex(^1)</td>
<td>Ac-GELKAIQAELKAIAKELKAIWELKAIAQGAG-NH(_2)</td>
<td>82</td>
<td>-</td>
<td>6.0</td>
<td>6 (pH 7.5)</td>
</tr>
<tr>
<td>CC-Hex*-L24D</td>
<td>Ac-GELKAIQAELKAIAKELKAIWEDKAIAGQ-NH(_2)</td>
<td>10</td>
<td>1.6 &amp; 4.6</td>
<td>2.1</td>
<td>4 (pH 8.5)</td>
</tr>
<tr>
<td>CC-Hex*-L24E</td>
<td>Ac-GELKAIQAELKAIAKELKAIWEEKAIAGQ-NH(_2)</td>
<td>15</td>
<td>2.0 &amp; 4.5</td>
<td>2.2</td>
<td>4 (pH 8.5) &amp; 6 (pH 6)</td>
</tr>
<tr>
<td>CC-Hex*-L24H</td>
<td>Ac-GELKAIQAELKAIAKELKAIWHEKAIAGQ-NH(_2)</td>
<td>55</td>
<td>4.2</td>
<td>4.9</td>
<td>4 (pH 6.5)</td>
</tr>
<tr>
<td>CC-Hex*-L24K</td>
<td>Ac-GELKAIQAELKAIAKELKAIWKEKAIAGQ-NH(_2)</td>
<td>41</td>
<td>4.3</td>
<td>4.7</td>
<td>4 (pH 8)</td>
</tr>
<tr>
<td>CC-Hex*-L24Dab</td>
<td>Ac-GELKAIQAELKAIAKELKAIWEKAIAGQ-NH(_2)</td>
<td>30</td>
<td>4.7</td>
<td>4.5</td>
<td>4 (pH 6.5)</td>
</tr>
<tr>
<td>CC-Hex*-L24Nle</td>
<td>Ac-GELKAIQAELKAIAKELKAIWEEKAIAGQ-NH(_2)</td>
<td>91</td>
<td>5.0 &amp; 12.0</td>
<td>6.2</td>
<td>6 (pH 6.5)</td>
</tr>
</tbody>
</table>

Complete biophysical analysis can be found in the supplementary information. Single letter codes for L-1,4-diaminobutyric acid and L-norleucine are \(\Gamma\) and \(\Delta\), respectively. Bold letters highlight point mutations in the CC-Hex* sequence. \(^1\)Data for CC-Hex from a previous publication. \(^2\)Fraction helix determined from the mean residue ellipticity at 222 nm in circular dichroism experiments. Conditions: 50 \(\mu\)M peptide concentration, phosphate-buffered saline (PBS) solution, pH 7.4 and 5 °C. \(^3\)Sedimentation-velocity analytical ultracentrifugation. Conditions: 150 \(\mu\)M peptide concentration, PBS solution, pH 7.4. \(^4\)Sedimentation-equilibrium analytical ultracentrifugation. Conditions: 70 \(\mu\)M peptide concentration, PBS solution, pH 7.4. \(^5\)X-ray crystallography oligomeric state of the biological unit. The hexamers are parallel \(\alpha\)-helical barrels and the tetramers are offset antiparallel \(\alpha\)-helical bundles.

3, 4 spacings of \(h\) residues; and (ii) specific combinations of predominantly aliphatic hydrophobic residues at the \(a\) and \(d\) sites. Designed four-helix bundles that are coiled coils lie far to the right of the spectrum of Figure 1A. Early designs in this area elucidated the rules for the formation of parallel four-helix bundles and also for parallel dimers and trimers. These rules have been embellished and used in rational \textit{de novo} design of many parallel coiled coils.

Previously, we have described the rational design and complete characterization through to X-ray protein crystal structures of a basis set of all-parallel dimeric, trimeric and tetrameric coiled coils. Serendipitously, a simple permutation to the sequence repeat of the tetramer, CC-Tet, produces an entirely new coiled-coil assembly, namely an all-parallel hexamer, CC-Hex. This has a central accessible channel. Therefore, it is an \(\alpha\)-helical barrel. The formation of the structure can be rationalized to the mutations of CC-Tet expanded the hydrophobic surfaces of the component helices allowing more of them to associate. Moreover, the introduction of complementary charged aspartic acid (Asp, D) and histidine (His, H) residues at the core position Leu24 in the otherwise hydrophobic central channel of CC-Hex to render two peptides—CC-Hex-D24 and CC-Hex-H24—that complement to form a parallel \(\alpha\_B\_2\) type heterohexamer. However, through a number of unpublished studies, we find that the CC-Hex scaffold is not completely robust and its oligomer state changes with other polar mutations. Subsequently, we have developed parametric computational design to deliver a series of \(\alpha\)-helical barrels, including pentamers, new hexamers and a heptamer. These \(\alpha\)-helical barrels are more robust to mutation and serve as platforms for rational design to introduce new functions.

Herein, we return to the mutants of CC-Hex and explore structural plasticity in the coiled-coil structural landscape. This must be navigated to deliver robust \textit{de novo} designs. Specifically, we describe a hitherto unexplored sequence with glutamic acid (Glu, E) at position 24. This assembles as a stable parallel hexamer at low pH, and as an up-down-up-down antiparallel tetramer, CDab, near neutral pH.

Moreover, substituting positively charged lysine (Lys, K) or 2,4-diaminobutyric acid (Dab) residues at position 24 gives the antiparallel tetramer. Thus, this study shows that the CC-Hex scaffold is far less mutable than previously believed. We use the X-ray crystal structures for the new state to guide the rational design of a robust and hyperstable anti-parallel tetramer, apCC-Tet, which we characterize fully. This required consideration of both the composition of the hydrophobic core and of interhelical salt-bridging in \textit{en bloc} mutations of the original CC-Hex sequence. apCC-Tet provides an additional \textit{de novo} protein fold that could be of use in protein design and engineering, materials science, and synthetic biology.

**RESULTS AND DISCUSSION**

Polar mutations at core sites of CC-Hex cause structural switches in solution.

To probe how robust CC-Hex-based sequences were to forming \(\alpha\)-helical barrels, we synthesized variants of CC-Hex at position 24. This is an \(\alpha\) position of the heptad repeat and, therefore, contributes to the cores of dimeric, trimeric and tetrameric coiled coils, and it points towards the lumens of \(\alpha\)-helical barrels. We tested additional charged side chains at the site, which we anticipated would be destabilizing. The variants included: negatively charged side chains...
Asp and Glu, to give L24D and L24E; positively charged side chains His, Lys and L-2,4-diaminobutyric acid (Dab), to give L24H, L24Dab and L24K; and the non-polar side chain norleucine (Nle), L24Nle (Table 1).

N.B. These and similar mutants described herein were made in a modified CC-Hex sequence, CC-Hex*, lacking the C-terminal AlaGly dipeptide, which was included originally as a mass tag. Hence the two nomenclatures used. For simplicity, hereafter mutations at position 24 in the CC-Hex* background are referred to as L24D, L24E etc.

Circular dichroism (CD) spectroscopy of L24D and L24E at pH 7.4 showed that the peptides had reduced helicities (Figure 2A) compared to the parent CC-Hex.56 Analytical ultracentrifugation (AUC) experiments indicated that both variants formed a mixture of dimeric and tetrameric species rather than hexamers (Figures S6.1,S6.2; Table 1). For a given concentration, L24E was more α-helical than L24D at neutral pH. Therefore, L24E was chosen for further biophysical characterization. Specifically, it was interrogated by CD spectroscopy and AUC over the pH range 3 – 7 (Figure 2B-D). At pH 3 and 4, the peptide was highly stable and did not fully unfold even by 90 °C. At pH 5 there was no significant loss of α helicity, but the peptide showed the start of a thermal unfolding transition. At pH 6 there was ≈30% drop in helicity and the peptide underwent full thermal denaturation. pH 7 saw a further drop in helicity and the peptide showed signs of cold denaturation. (N.B. This cold denaturation was typical of mutants later confirmed as tetramers (Figures S4.1–4.4,S4.6)). Sedimentation-equilibrium AUC experiments revealed that from pH 3 – 7.4 L24E switched from hexameric to dimeric species (Figure 2D), whereas sedimentation-velocity at pH 4 indicated multiple species (Figure S6.13). At pH 7.4, increasing the peptide concentration above 150 μM increased α helicity and shifted the equilibrium towards tetramers (Figures S5.1,S6.12).

To test if the tetramer resulted from charged residues at core sites generally, the positively charged variants, L24Dab, L24K and L24H (N.B. His has a pKa of 6.8) and the L24nLeu control were examined.67 Again, CD spectroscopy at pH 7.4 showed that L24H, L24K and L24Dab were significantly less α-helical than the parent (Figures S4.3–4.4,S4.6). By SE AUC, L24K, L24Dab and L24H had weights of ≈5 x monomer mass (Figures S6.3–6.4,S6.6). By contrast, L24Nle was highly folded and hexameric in solution at pH 7.4; although an additional higher-order species was observed by sedimentation-velocity experiments accounting for 35% of the sample (Figures S4.5,S6.5).

Thus, the CC-Hex background is not robust to charged polar mutations in its hydrophobic channel, in contrast to the more robust computationally designed α-helical barrels, including a hexamer.65 Moreover, at pH values where these residues are likely charged the structures are less helical, of lower thermal stability, and form lower oligomeric states.

X-ray crystal structures reveal a broader accessible structural landscape.

To add to the previously documented hexameric structures of CC-Hex-D24 and CC-Hex-H24 (PDB 1D codes 3R46 and 3R47), we obtained X-ray crystal structures for L24E, L24D, L24Dab, L24K, L24H and L24Nle. L24E crystallized in two forms that gave different structures: a hexameric blunted barrel and an antiparallel tetramer (Figure 3A,C, respectively). The crystallization conditions for the two states were markedly different. The tetramer was only observed at pH 8.5, while the hexamer crystallized from several conditions from pH 5 – 6.5, as well from unbuffered solutions where the pH was low due to residual trifluoroacetic acid from peptide synthesis (pH < 2). In the studies reported herein, L24D and L24H only crystallized as the antiparallel tetramer. However, as mentioned above, previously, we crystallized CC-Hex-D24 and CC-Hex-H24 from unbuffered solution and at pH 7.5, respectively.56 With the amine side chains L24K and L24Dab only crystallized as antiparallel tetramers, while L24Nle only crystallized as the parallel hexamer (Figure 3B); these structures were obtained in the pH range 6.5 – 8 (Table S2). For these three sequences, our inability to obtain other crystal forms does not necessarily mean that other states are inaccessible. Nevertheless, the solution-phase biophysical data corroborate the observed crystal structures (Table 1).

Comparing the low-pH hexameric structures of CC-Hex-D24 and CC-Hex*-L24E revealed the shorter aspartic acid could be accommodated within the core of a completely folded structure; whereas, glutamic acid could not be fully accommodated, as two of the six peptide chains frayed at the C terminus allowing the glutamates to extend into solvent (Figure 3A).

---

Table 2. Sequences and summary of biophysical data for CC-Hex* en bloc mutants and apCC-Tet.
<table>
<thead>
<tr>
<th>Peptide</th>
<th>Sequence and Register</th>
<th>Helix$^1$ (%)</th>
<th>SV$^2$ (mass/monomer mass)</th>
<th>SE$^3$ (mass/monomer mass)</th>
<th>XRD$^4$ oligomeric state</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC-Hex*-KgEb</td>
<td>Ac-KLEAIAQKLEAIKLEAIAKLEAIQG-NH$_2$</td>
<td>38</td>
<td>.5</td>
<td>.5</td>
<td>4</td>
</tr>
<tr>
<td>CC-Hex*-LL</td>
<td>Ac-GELKALAQELKALAKELKALAWELKALQQG-NH$_2$</td>
<td>61</td>
<td>4.6</td>
<td>4.4</td>
<td>4</td>
</tr>
<tr>
<td>CC-Hex*-*II</td>
<td>Ac-GEIKIAAQEKIAKEIKIAWEIKIAAQG-NH$_2$</td>
<td>69</td>
<td>6.0</td>
<td>5.6</td>
<td>6</td>
</tr>
<tr>
<td>CC-Hex*-LL*KgEb</td>
<td>Ac-GKLEAQLKLEAIKLEAIKLEAIQG-NH$_2$</td>
<td>62</td>
<td>3.95</td>
<td>4.05</td>
<td>4</td>
</tr>
<tr>
<td>apCC-Tet</td>
<td>Ac-GELELAQEGLEAIKLEAIKLEAIQG-NH$_2$</td>
<td>86</td>
<td>4.0</td>
<td>4.1</td>
<td>4</td>
</tr>
</tbody>
</table>

Complete biophysical analysis can be found in the supplementary information. $^1$Fraction helix determined from the mean residue ellipticity at 222 nm in circular dichroism experiments. Bold letters highlight point mutations in the CC-Hex* sequence. For clarity CC-Hex*-LL-KgEb is not highlighted. Conditions: 10 μM peptide concentration, phosphate buffered saline (PBS) solution, pH 7.4 and 5 °C. $^2$Sedimentation-velocity analytical ultracentrifugation. Conditions: 150 μM peptide concentration, PBS solution, pH 7.4. $^3$Sedimentation-equilibrium analytical ultracentrifugation. Conditions: 70 μM peptide concentration, PBS solution, pH 7.4. $^4$X-ray crystallography oligomeric state of the biological unit. CC-Hex*-*II forms a hexameric collapsed complex coiled coil and all others form tetrameric antiparallel α-helical bundles. **CC-Hex*-KgEb and CC-Hex*-LL-KgEb aggregate at 10 μM and 40 μM peptide concentration, respectively, when span at 3,000 rpm. Data was collected for CC-Hex*-LL-KgEb at 35 μM peptide concentration.

Turning to the structures determined near neutral pH, L24D and L24E plus those for L24Dab, L24H and L24K are closely similar antiparallel four-helix coiled coils with backbone and all-atom RMSDs for residues 1 – 23 across of the whole set of 0.44 Å ± 0.12 Å and 1.13 Å ± 0.44 Å, respectively (Figure 3C). They have staggered rather than blunt-end arrangements of the helices, and the helices are frayed or disordered after the polar 24th residue. As a result, the cores are exclusively hydrophobic. These cores have contributions from residues at a (Leu), d (Ile) and e (Ala) sites of the heptad repeats. This gives two distinct helix-helix interfaces: two “wide faces” centered on d = Ile (Figure 3C left) and two “narrow faces” centered on e = Ala (Figure 3C center). The Leu residues at a are directed towards central long axis of the bundle (Figure 3C right) reminiscent of complementary x-da layers observed in Alacoil (Figure 1B). In other words, the structures have oblate cross-sections. The two interfaces are flanked by pairs of the same amino acid from neighboring helices; i.e., b:b and g:g, respectively. As b = Lys and g = Glu, the narrow and wide faces present seams of positive and negative charge, respectively. Thus, the formation of the hydrophobic core overrides (i) the complete folding of each chain into α helices, and (ii) potential electrostatic repulsion between like residues at b and g.

Structural searches of PDBfold and CAME topsearch using the L24E tetramer identified the following: HAMP proteins (PDB ID code 3ZX6; Ca RMSD = 1.2 Å); variants of the GCN4-p1 peptides (PDB ID code 1W5J; Ca RMSD = 1.3 Å); the ROP proteins (PDB ID code 1QX8; Ca RMSD = 1.6 Å); αTet (PDB ID code 6C52; Ca RMSD = 2.1 Å), a non-aggregating variant of a de novo designed cross-α amyloid-like structure; and di-Zn(II)- DF3I (PDB ID code 2K1K; Ca RMSD = 2.3 Å), an artificial protein designed for phenol oxidase activity. While all structures are antiparallel tetrameric bundles, only the HAMP proteins contain regions with oblate cross-section due to regular Alacoil. The L24 variants are also structurally similar to a GCN4 leucine-zipper peptide mutant from Kallenbach and Lu (PDB ID code 2R2V), which crystallizes as an oblate antiparallel tetramer with wide interfaces centered on d = Val and Alacoil interfaces centered on e = Ala. Furthermore, the L24 variants resemble HexCoil-Ala (PDB ID code 35OR), a de novo designed antiparallel tetramer with a core of a, d and g residues where the Alacoil interfaces center on g = Ala. Interestingly, this different core results in topologies with opposite handedness.

Blunt-ended and fully folded tetramers can be redesigned from the CC-Hex* sequence.

![Figure 3. X-ray crystal structures of CC-Hex* point mutants.](image)

(A) The hexameric crystal form of L24E. (B) Slice through the structure of L24Nle viewed from the N termini. (C) The overlaid backbones of the tetrameric forms of L24E, L24D, L24Dab, L24K and L24H with the side chains of L24E displayed for reference. The chains are colored in rainbow spectrum from the N termini (blue) to the C termini (red).
The above structures of the polar mutants at position 24 of the CC-Hex* sequence demonstrate that another coiled-coil topology—i.e., an up-down-up-down antiparallel tetramer—is accessible to our basis-set of de novo designs. However, the tetramers described thus far are not fully folded. To explore how the fold might be optimized, we synthesized en bloc mutants of CC-Hex* at the a & d and g & b sites (Table 2). Core positions have a significant effect on specificity of oligomeric state and helix orientation of associated coiled-coil peptides. Interestingly, in the heptad repeat of CC-Hex charged Glu residues occupy core g positions. To test how residues at these positions alter specificity, we swapped the order of the charge residues at the g and b sites. In CC-Hex*, the g→f repeat is ELKAIAX. Therefore, the swap had the sequence KLEAIAX, which we refer to as CC-Hex*-KgEb. Unexpectedly, this peptide was only soluble up to 40 μM concentration in phosphate buffered saline (PBS). It was 38% α helical but could not be analyzed by AUC due to aggregation. Nevertheless, we were able to obtain an X-ray crystal structure of the charge-swapped variant in the original CC-Hex background (CC-Hex-KgEb) that revealed an antiparallel tetramer (Figure 4A). Unlike the L24 variants, this does not have shifted antiparallel helices and, consequently, it is a blunt-ended tetramer.

Thus, swapping the potential salt-bridge positions disrupts the coiled-coil assembly. Analysis of KIH interactions in the CC-Hex structure revealed Glu at g as knob residues, with the ethyl unit (Cβ and Cγ) packing into the interface and the carboxylic acid presented at the surface. By contrast, the Lys side chains at b extend fully into solution. Given this, it is surprising that the swap was not tolerated: it is reasonable to expect that the methylene groups of Lys should pack at g, and the Glu at b should be solvent accessible. We speculate that the entropic cost of reducing the conformational freedom of Lys side chains at g disfavors hexamer formation.

Next, we turned to the en bloc hydrophobic mutations at a and d positions of CC-Hex*. One of these mutants has already been reported: swapping the Leu at a and Ile at d results in the formation of a slipped hexameric barrel, CC-Hex*-IL (PDB ID code 4HBG). For the study presented herein, we made the two other permutants with either all Ile or all Leu at both a and d, i.e., CC-Hex*-II and CC-Hex*-LL, respectively. CC-Hex*-II was α-helical and completely unfolded upon heating, whereas CC-Hex*-LL was highly α-helical and hyperthermally stable (Figure S4.7). CC-Hex*-LL showed the start of a thermal unfolding transition at ≈75 °C and, on cooling, signs of thermal annealing. AUC experiments indicated that CC-Hex*-II was hexameric in solution, whereas CC-Hex*-LL formed species of lower mass (Table 2).

X-ray crystal structures of CC-Hex*-II and CC-Hex*-LL were determined (Figure 4B,C). The former revealed a collapsed hexameric coiled coil. This is a complex coiled coil with multiple unique helical environments in the homomeric assembly. Recently, we have described similar homomeric structures in which symmetry is broken. Albeit in a different e/g background, we argue that β-branched residues at the a/d sites promote barrel structures. Clearly, this is not the case for the new peptides described here: the introduction of additional Ile residues to give CC-Hex*-II results in a structure with a consolidated core. This demonstrates that residues peripheral to core sites also contribute to the final structure that is adopted. Previously, we solved an alternative structure for a variant of the CC-Hex*-II peptide, which is a parallel tetramer (PDB ID code 4H7R). The crystal structure for CC-Hex*-LL also revealed an antiparallel tetramer. As with CC-Hex-KgEb, the peptide chains are fully helical. However, the backbones of these two structures could not be aligned (Figure 4A,B): the narrow Alacoil interface of CC-Hex*-LL is offset longitudinally to a lesser extent than in CC-Hex-KgEb resulting in more-flushed termini.

N.B. All of the CC-Hex* variants reported here and structurally characterized by X-ray crystallography were confirmed to be coiled coils by SOCKET analysis (Figure S8.1). These results demonstrate further the complexity of the self-association landscape of CC-Hex-based peptides. And that alternative states can be accessed by changes in solution conditions and/or small changes to the sequence. Nonetheless, they also suggest ways in which structures within this landscape can be targeted, which we illustrate below.

**Designing apCC-Tet, an optimized antiparallel coiled-coil tetramer.**

From the above, we chose CC-Hex*-LL as the best starting point for the core in a rational design of a fully folded, stable antiparallel homotetramer. In addition, we swapped the charged residues at g and b positions, as this also favors antiparallel tetramer over hexamer. This gave CC-Hex*-LL-KgEb (Table 2). Compared with CC-Hex*-LL, CC-Hex*-LL-KgEb gave a sharp single peak in the c(s) distribution in sedimentation-velocity AUC. Sedimentation-equilibrium data and an X-ray crystal structure confirmed CC-Hex*-LL-KgEb as a tetramer. However, like CC-Hex*-KgEb, the new variant aggregated in PBS above 40 μM concentration. Aggregation of the tetramers could occur along the wide interface of these oblate structures: relative to the narrow-interface, the wide interfaces present a greater exposed hydrophobic surface. It is possible that peripheral lysine residues screen this
interface less well than glutamate. Thus, whilst swapping the potential salt bridges helps specify the tetrameric form relative to the hexamer this is at the cost of reduced solubility. Therefore, we sought an alternate and optimized pattern of charge on the exterior of the tetramer.

There are several ways to arrange Lys and Glu residues at $g$ and $b$ positions to disfavor parallel association of helices and favor the antiparallel alignment. For example, Glu could be placed at $g$ and $b$ of heptads 1 and 2, with Lys at these sites in heptads 3 and 4. This arrangement of Glu near to the $N$ terminus and Lys near the $C$ terminus is known to improve $\alpha$-helical stability. Retrospectively, we named this charge pattern in the CC-Hex*-LL background apCC-Tet.

apCC-Tet was highly $\alpha$ helical at 5 °C with a fraction helicity of 86%, an increase of 25% over CC-Hex*-LL (Figure 5A). apCC-Tet was thermally stable up to 95 °C, whereas CC-Hex*-LL showed the start of an unfolding transition at $\approx$75 °C (Figure 5B). This improvement in solution-phase properties of apCC-Tet was also evident in sedimentation-velocity AUC experiments, where apCC-Tet gave a sharper peak than CC-Hex*-LL (Figure 5C). For apCC-Tet, analyses of these data gave single discrete species with weights of 4.0 and 4.6 x monomer mass by sedimentation velocity and sedimentation equilibrium, respectively; whereas, CC-Hex*-LL returned non-integer oligomeric states of 4.6 and 4.4, respectively (Figure 5C,D). Crucially, apCC-Tet did not show any signs of aggregation at 150 μM peptide concentration during these experiments. Collectively, these data show that the arrangement of charged residues that should favor antiparallel helices also has improved the folding and thermal stability of apCC-Tet.

The X-ray crystal structure for apCC-Tet confirmed an antiparallel coiled-coil tetramer (Figures 5D-F, S8.1). The helix geometry and interfaces are similar to CC-Hex*-LL with backbone and all-atom RMSDs of 0.56 Å and 1.58 Å between the two structures. However, only 7 of the possible 16 pairs of Glu-Lys contacts in the apCC-Tet structure are within 4 Å to form interhelical salt bridges and the side-chain densities are diffuse (e.g. the average temperature factor for the Lys N$_{C}$ is 56.9 Å$^2$ in comparison the all-atom average is 40.5 Å$^2$). Therefore, despite improvements in the solution-phase behavior, the X-ray structure reveals that these are not due to altered backbone arrangements or core packing, nor are they from salt-bridge formation. We posit that the changes allow electrostatic steering in helix association and avoid uniformly charged surfaces on apCC-Tet.

To summarize this section, we have achieved the rational design of an antiparallel homomeric coiled-coil tetramer. In the register of the heptad repeat, this has a hydrophobic core of $a = d =$ Leu plus $e =$ Ala, with flanking Glu-Lys pairs at $g$ and $b$:b sites. This sequence pattern directs the assembly of a well-defined, discrete and hyperstable antiparallel tetramer with an up-down-up-down topology and oblate cross-section. This adds to a growing basis set of de novo coiled-coil modules for protein design and synthetic biology.

**CONCLUSION**

CC-Hex was the first reported X-ray crystal structure of a de novo $\alpha$-helical barrel. These barrels are a growing class of both natural and designed proteins that are robust and have potential in protein engineering and synthetic biology. Nonetheless, CC-Hex was discovered serendipitously: it was a variant of a de novo designed parallel coiled-coil...
tetramer, CC-Tet.62 Herein, we have described a series of point and en bloc changes to a modified CC-Hex sequence, CC-Hex*, which lacks the original C-terminal AlaGly mass tag.

These new variants reveal that the CC-Hex background is somewhat plastic with a variety of coiled-coil states being accessible from it. These include parallel and antiparallel tetramers, and open barrel and collapsed hexamers. Moreover, the structural plasticity is evident even for some of the charge-bearing point variants: charged side chains introduced that are charged are disfavored in the parallel barrel state and, therefore, switch to the offset antiparallel tetramer state. For self-associating systems, core positions strongly specify interface type and we identify Leu at a and d, in the CC-Hex background that preferentially forms antiparallel tetramers. These empirical explorations and observations have allowed us to design de novo a new antiparallel homotetrameric coiled coil, apCC-Tet. A combination of the established positive and negative design rules resulted in a robust and fully characterized structure.

This work highlights further the degrees of freedom available to self-associating peptide systems,59,60,78 and the associated challenge of designing specific oligomer states and topologies de novo.12,57,83–88 This problem is particularly acute for less-well-defined peptide-peptide interfaces where many near-isoenergetic states are accessible.12 The energy landscape for α-helical coiled-coil assemblies may well be more navigable than the general case of helical bundles. This is for three interrelated reasons: First, the hallmark knobs-into-holes packing of coiled coils dramatically reduces the number of helical arrangements possible. As a result, it is relatively straightforward to enumerate many coiled-coil backbones parametrically.86–91 Second, and directly related to this structural constraint, coiled-coil sequences have relatively simple repeat patterns of hydrophobic and polar residues. Third, and as a consequence of the first two points, only certain residues appear to be tolerated in the helix-helix interfaces of coiled coils.83–85 Thus, as coiled coils have low-complexity sequences, the protein-design problem is more sharply defined for these structures and sequences compared with less-regular structures.

That said, the coiled-coil energy landscape is still complex with multiple oligomer states, parallel/antiparallel/mixed arrangements of helices, and homo- and heterotypic assemblies all possible.56,92 Nonetheless, considerable progress has been made to discern sequence-to-structure relationships and to develop computational methods for coiled-coil design. These have led to robust rational and/or computational designs for parallel dimers through heptamers.56,62,64,81 The work presented herein adds to this effort: it illustrates how alternate states can be distinguished; and it provides guidelines for accessing de novo antiparallel structures, which have been less explored than parallel coiled coils.76–78,93

More subtly, certain polar residues are tolerated at the otherwise hydrophobic coiled-coil interfaces, and these influence partner and oligomer-state selection.83,94,95 This concept has recently been revisited.96,97 Here, we add to this showing that buried charged residues disfavor high-oligomer states for alternate antiparallel tetramers in the CC-Hex background. Furthermore, for certain sequences the two states—parallel hexamer and antiparallel tetramer—are sufficiently close in energy to effect switches between them simply by changing the pH of the solution. This presents exciting prospects for developing sequences to switch gross structural state in response to facile perturbations.75,96–102

Coiled-coil peptides that show multiple-defined states have been described by others. Lizatović et al. design a pH-responsive sequence that switches between a pentameric bundle and parallel hexameric bundle.98 Grigoryan et al. present the design of carbon-nanotube solubilizing peptides, which wrap around the surfaces of the nanotubes. In isolation, one of these peptides forms a tetramer, and another forms a dimer/hexamer mixture.69 Others are also discovering that seemingly benign sequence alterations can cause gross structural changes in self-associating coiled coils. Slovic et al. describe the redesign of the membrane-spanning peptide phospholamban to make it water soluble. Solution-phase biophysics shows that the full-length sequence remains pentameric. However, a truncated peptide (residues 21 – 52) is best modelled as a tetramer/pentamer equilibrium in AUC and a structure reveals an offset antiparallel tetramer.103 Finally, Spencer et al. have made variants of parent CC-Hex with phenylalanine in the core that forms a collapsed antiparallel hexameric bundle.105,106

The specification of orientation of peptides in self-associating systems is a standing challenge in protein design. Anti-parallel coiled coils remain a challenging design target with fewer examples of successful designs compared to parallel structures. We have delivered a robust de novo antiparallel tetramer that is characterized thoroughly. This provides another designed module or component adding to the growing basis set of de novo coiled coils with potential applications in the design and assembly of protein origami,107 peptide nanotubes,72 protein colocalization,108,109 and generally as tectons for generating complex self-assembling systems.110,111

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Supporting Figures and Tables (PDF)

Crystal-Structure Files

The crystal structures are all available from the Protein Data Bank using the following accession codes: CC-Hex*-L24D (6Q5H), CC-Hex*-L24E_tet (6Q5I), CC-Hex*-L24E_hex (6Q5J), CC-Hex*-L24K (6Q5K), CC-Hex*-L24H (6Q5L), CC-Hex*-L24DAB (6Q5M), CC-Hex*-L24Ne (6Q5N), CC-Hex*-LL (6Q5O), CC-Hex*-II (6Q5P), CC-Hex-KgEb (6Q5Q), CC-Hex*-LL-KgEb (6Q5R), apCC-Tet (6Q5S).

AUTHOR INFORMATION

Corresponding Author

* d.n.woolfson@bristol.ac.uk

Author Contributions

G.G.R., C.W.W., J.L.B., A.R.T. and D.N.W. conceived the project and designed the experiments. G.G.R., C.W.W., J.L.B. and A.J.B.
synthesized the peptides and conducted the solution-phase biophysics. G.G.R., C.W.W., J.L.B., N.R.Z., A.J.B. and R.L.B. determined the peptide X-ray crystal structures. G.G.R., C.W.W. and D.N.W. wrote the paper. All authors have read and contributed to the preparation of the manuscript.

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENT
G.G.R., C.W.W., J.L.B., A.R.T., A.J.B. and DNW are supported by a European Research Council Advanced Grant to DNW (340764). G.G.R. and A.J.B. thank the Bristol Chemical Synthesis Centre for Doctoral Training funded by the Engineering and Physical Sciences Research Council (EP/G036764/1). C.W.W. and D.N.W. thank the Biotechnology and Biological Sciences Research Council for funding the South West Biosciences Doctoral Training Partnership (BB/J014400/1) and a responsive-mode grant (BB/R00661X/1). We thank the University of Bristol School of Chemistry Mass Spectrometry Facility for access to the EPSRC-funded Bruker Ultraflex MALDI-TOF/TOF instrument (EP/K03927X/1). We thank the MX group at Diamond Light Source for their support. D.N.W. holds a Royal Society Wolfson Research Merit Award (WM140008). D.N.W. would like to dedicate this paper to Professor Ronald T Raines on the occasion of his 60th birthday.

REFERENCES


(18) Grayson, K. J.; Anderson, J. L. R. Designed for Life: Biocompatible de Novo Designed Proteins and


