



Thomson, A. R., Wood, C. W., Burton, A. J., Bartlett, G. J., Sessions, R. B., Brady, R. L., & Woolfson, D. N. (2014). Computational design of water-soluble α -helical barrels. *Science*, 346(6208), 485-488.
<https://doi.org/10.1126/science.1257452>

Peer reviewed version

Link to published version (if available):
[10.1126/science.1257452](https://doi.org/10.1126/science.1257452)

[Link to publication record in Explore Bristol Research](#)
PDF-document

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Submitted to Science for Consideration as a Report

Computational design of water-soluble α -helical barrels

Andrew R Thomson,^{1,*} Christopher W Wood,^{1,2,†} Antony J Burton,^{1,†} Gail J Bartlett,¹
Richard B Sessions,² R Leo Brady,² and Derek N Woolfson^{1,2,*}

¹School of Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS, UK

² School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol, BS8
1TD, UK

*To whom correspondence should be addressed: chzart@bristol.ac.uk; chdnw@bristol.ac.uk

†These authors contributed equally to this work.

Summary sentence: De novo protein design has been used to expand the repertoire of coiled coils to include barrel structures with 5 – 7 helices.

The design of protein sequences that fold into prescribed de novo structures is challenging. General solutions to this problem require geometric descriptions of protein folds, and methods to fit sequences to these. The α -helical coiled coils present a promising class of protein for this, and offer considerable scope for exploring hitherto unseen structures. For α -helical barrels, which have >4 helices and accessible central channels, many of the possible structures remain unobserved. Here we combine geometrical considerations, knowledge-based scoring, and atomistic modeling to facilitate the design of new channel-containing α -helical barrels. X-ray crystal structures of the resulting designs match predicted in silico models. Furthermore, the observed channels are chemically defined and have diameters related to oligomer state, which present routes to design protein function.

Defining protein sequences that fold into specified three-dimensional structures is called the ‘inverse protein-folding problem’ (1). Mostly, this has been applied to mimic existing folds (2). However, the design of structures not yet seen in nature can also be considered (3). Repeat proteins are of interest here as extrapolation from known structures can provide geometric parameters and sequence-to-structure relationships to guide design (4, 5), and proteins with cyclic symmetry offer possibilities for systematic variation of repeating elements to produce families of proteins (6). One example is the α -helical coiled coil (7, 8). Classical coiled coils comprise bundles of 2 – 4 α -helices, account for >98% of known coiled-coil structures (9, 10), and have well-understood sequence-to-structure relationships (7, 8). Unusually for proteins, the conformations of coiled-coil backbones are well described by a small number of parameters (11-14). Consequently, a relatively large number of successful coiled-coil structures have been designed (8), although with a few exceptions (8, 13) these have largely mimicked natural precedents.

The α -helical barrels present an intriguing subset of coiled coils to move beyond known structures (15). These have more-complex helical packing (15, 16), which results in the assembly of five or more α -helices into cylindrical bundles with central channels or pores (15). The few current examples include: natural parallel 5- and 10-helix structures, and antiparallel 10- and 12-helix bundles (17-20); a de novo parallel hexamer, achieved partly serendipitously (21); and a mutant leucine-zipper peptide that forms an unusual staggered parallel 7-helix arrangement (22). For these, there is a near-linear relationship between lumen size and oligomer state, Fig. 1A, which opens possibilities for designing channel proteins. However, due to the scarcity of α -helical barrels, and because these are usually parts of larger membrane-spanning proteins, it is difficult to derive rules to design new examples. To overcome this, we describe a geometrical and computational framework for designing α -helical barrels from first principles, and apply this to deliver discrete, water-soluble assemblies with 5 – 7 parallel and identical helices.

First, we required a means to map structural and sequence relationships between the new targets and the plentiful examples of classical coiled coils. The latter have heptad sequence repeats, $(hxxhxxx)_n$, where *h* and *x* are hydrophobic and polar residues, respectively; often labeled **abcdefg**, this places *h*-type residues at **a** and **d**. A resulting hydrophobic seam mediates helix association and packing, with the interface often buttressed by polar interactions between **e** and **g**, Figs. 1B&C. These are Type-N interfaces, and the residues at the “**gade**” positions determine oligomer state and partner selection (8, 23).

Fig. 1: Interfaces, packing and scoring in coiled-coil design. (A) Relationship between oligomer state and pore diameter for existing α -helical barrels (red) and the de novo structures described herein (blue); data are given as the mean \pm SD; the dotted line = linear regression ($R^2 = 0.86$). (B) Helical-wheel diagram representation of a Type N heterodimer. (C) Section through a coiled-coil heterodimer crystal structure (PDB ID: 1fos). (D) Complete helical-wheel diagram for a Type II α -helical barrel (pentamer shown with $M = 1$), illustrating the heterotypic interfaces between **cdga** and **deab**. Compared to classical parallel dimers (panel B), which make interhelical **gade** contacts **g** \rightarrow **e'**, **a** \rightarrow **a'**, **d** \rightarrow **d'** and **e** \rightarrow **g'**, the approximately equivalent primary contacts in helical α -barrels are

$c \rightarrow b'$, $d \rightarrow e'$, $g \rightarrow a'$ and $a \rightarrow d'$, respectively. In addition, three of the four geometric parameters required to describe α -helical barrels are shown: coiled-coil radius (r), oligomer state ($N = M + 4$), and helical offset (ω_1). Superhelical pitch, P , is not represented. (E) Section through a coiled-coil pentamer crystal structure (PDB ID: 1mz9). (F) Helical wheel for an isolated heterodimer-like interface in α -helical barrels. (G) Heterodimer-like interfaces in a single sequence with a Type II repeat, showing the shared a and d positions. (H) Calculation of fitness score. (I) Designed sequence classes following filtering. Numbers of interaction pairs/sequences are shown in brackets at each step for F-I. PyMol (www.pymol.org) was used to create panels C & E.

In an α -helical barrel each helix interacts with two neighbors via independent hydrophobic seams, Figs. 1D&E. There are three ways to achieve this within a heptad repeat: the two seams can share one residue (Type I interfaces); be adjacent (Type II), Fig. 1D; or be separated by an intervening residue (Type III) (15, 16). We hypothesized that Type II and III interfaces can lead to α -helical barrels, with the oligomer state determined by the angular offset between the two interfaces. Because differences in this contact angle between helices become smaller with increasing oligomeric state, Fig. S1A, we anticipated that controlling barrel size would be more tractable for smaller assemblies. Therefore, we concentrated on Type II interfaces and ' $hhxxhhx$ ' repeats (read $g \rightarrow f$), which should define oligomers with 5 – 7 helices.

Also with increasing oligomer state, the helix-helix interfaces in α -helical barrels become more like those of classical coiled-coil dimers, specifically heterodimeric interfaces, Figs. S1C and 1B&D. Therefore, we devised a scoring system to select sequences encoding two heterotypic hydrophobic seams. This treated each seam as one half of a heterodimer, with the seams sharing two residues, Figs. 1F&G. In terms of traditional heptads, these seams comprise residues at **deab** and **cdga**—which are each equivalent to **gade** positions in classical parallel dimers—and combine to give a **gabcde** repeat. We used the bZIP scoring function (24) to assess many **deab** plus **cdga** interfaces and identify potential heterotypic pairs. We considered all combinations of A, E, I, K, L, N, Q, R, S, and V residues, which are commonly found in parallel dimers (10). Because of the shared residues, the initial screen had 10^6 sequences. For each **deab** plus **cdga** pair a 'fitness score' was calculated by subtracting the highest homo-paired score from that for the hetero pair, Fig. 1H. The identified pairs were filtered further by the raw bZIP pairing score, to give 7578 hits.

Further screening removed sequences with anticipated destabilizing polar residues at a and d positions; selected those with potential $b \rightarrow c'$, $b \rightarrow g'$, $e \rightarrow c'$ or $e \rightarrow g'$ salt-bridging interactions; and excluded hydrophobic residues at peripheral b and c positions. This reduced the set to 370 sequences. Many of these resembled classical tetramers, with Leu at a , Ile at d , and polar residues at e and g . Thus, to select larger assemblies, we added a requirement for a hydrophobic residue at one or both of e and g , reducing the set to 188 sequences. There was some redundancy between Arg and Lys, so we retained only the more readily synthesized Lys-containing sequences. Of the initial 10^6 sequences, 76 met the full selection criteria. Of these, 22 repeats representing sequence diversity in the full set were chosen for further study, Fig. 1I. These were named after their **gabcde** repeat, and synthesized as four-heptad peptides, Table S1.

Next, we developed a software tool, CCBuilder (25), to construct *in silico* models of α -helical coiled coils and barrels. This uses Crick's equations to build coiled-coil backbones (11, 12), adds side chains using SCRWL (26) and PyRosetta (27), and assesses inter-helix packing through implementations of the BUDE forcefield (25, 28) and SOCKET (29). For the 22 sequences, models were generated for each oligomer state with 4 – 8 all-parallel helices. This employed a genetic algorithm to search and optimize structural space defined by three independent parameters: radius, pitch, and the rotational offset between helices, Fig. 1D and Table S2; *n.b.*, the number of residues per α -helical turn also varies as it is related to coiled-coil pitch, however, it was constrained within limits known for proteins (3.65 ± 0.07 (25)). We predicted the preferred association state for each sequence, produced an atomistic model for this, and estimated an energetic difference between it and alternative states.

The most commonly predicted oligomeric state was pentameric (12 sequences) consistent with the angular offset between the two hydrophobic seams in a Type II interface (103°) most closely matching the internal angle of a regular pentagon (108°), Fig. S1A. No sequences were predicted to form tetramers, while 7 were predicted to form hexamers, 2 to form heptamers, and 1 was predicted to form an octamer, Table S3.

We synthesized peptides for the 22 targets, Table S1 and Fig. S2. As judged by circular dichroism (CD) spectroscopy, all of these formed highly α -helical and thermally stable assemblies, Figs. 2A&B and S3. Two of the peptides showed low solubility and could not be characterized further. The oligomeric states of the remaining 20 soluble examples were assessed by analytical ultracentrifugation (AUC), Figs. 2C&D and S4, and Table S3, which indicated assemblies with 4 – 7 peptide chains consistent with the design rationale; specifically, there were 2 tetramers, 4 pentamers, 6 hexamers and a single heptamer, Table S3. In 7 cases a single state could not be identified by AUC. Where determined, the experimental oligomer states matched those predicted using CCBuilder in 8 out of 13 cases, Table S3.

Fig. 2 Solution-phase biophysical data for the designed peptides. (A) CD spectra at 5 °C for ILQKIE (red), SLKEIA (green), SIKEIA (green dashes) and ALKEIA (blue). (B) Thermal denaturation profiles monitored by the change in CD signal at 222 nm. (C) Sedimentation velocity $c(s)$ distribution fits at 20 °C for ILQKIE, SLKEIA and ALKEIA. (D) Representative sedimentation-equilibrium AUC data (dots) and fitted single-ideal species model curves recorded at 20 °C, 280 nm and 24,000 rpm for ILQKIE, SLKEIA and ALKEIA. **Color key:** Panels B – D same as for A. **Conditions:** 10 μ M (panels A and B), 150 μ M (panel C) and 70 μ M (panel D) peptide concentrations, phosphate-buffered saline (PBS), pH 7.4, except ALKEIA, panel C and D, tris-buffered saline, pH 7.5.

High-resolution X-ray crystal structures were determined for 4 of the peptides, Figs. 3 and S5 – S7, Tables 1 and S4. Structures were solved by molecular replacement with either part of, or intact CCBuilder predictions (backbone and C β atoms only) as initial search models. The structures revealed parallel, blunt-ended α -helical barrels, Figs. 3A&B, with knobs-into-holes packing confirmed by SOCKET (29), and Type II interfaces as designed, Fig. 3C. The experimental and *in silico* models were closely correlated

with low RMSDs for backbone and side-chain atoms, Table 1 and Fig. S5. The pore sizes of the assemblies are consistent with those expected for each oligomer state, Fig. 1A. To accord with our other de novo coiled coils (30), we renamed the sequences CC-Pent, CC-Hex2, CC-Hex3 and CC-Hept, Table 1.

Fig. 3. X-ray crystal structures for three de novo α -helical barrels. (A) and (B) From left to right, orthogonal views of ILQKIE (CC-Pent; red; PDB ID 4pn8), SLKEIA (CC-Hex2; green; PDB ID 4pn9) and ALKEIA (CC-Hept; blue; PDB ID 4pna). (C) Conserved packing of the **a** Leu (red) and **d** Ile (green) residues, and variation of the steric bulk of the **e** and **g** residues.

Repeat sequence (<i>gabcde</i>) systematic name	Predicted oligomeric state (N), pitch (Å), radius (Å), pore diameter (Å)	Oligomeric state by AUC	Crystal structure oligomeric state, pitch, radius, pore diameter.	Crystal structure resolution (Å), and PDB ID.	RMSD X-ray crystal structure – model (Å)
ILQKIE CC-pent	5, 190, 8.2, 5.1	5	5, 183, 8.6, 5.7	2.00, 4pn8	1.61 (0.67)
SLKEIA CC-Hex2	6, 305, 8.6, 5.0	6	6, 162, 9.5, 6.0	2.20, 4pn9	2.47 (1.77)
SIKEIA CC-Hex3	6, 135, 8.8, 5.0	6	6, 132, 9.7, 7.4	2.05, 4pnb	2.36 (1.26)
ALKEIA CC-Hept	7, 280, 9.5, 7.1	7	7, 481, 9.8, 7.6	2.10, 4pna	1.94 (1.17)

Table 1: Modeled and experimental oligomer-state and structural parameters for selected assemblies. Oligomeric state was taken from AUC sedimentation-equilibrium experiments. Pore diameters were measured with PoreWalker (31) and are given as the average pore diameter through the assembly. The coordinates of the experimental and model structures were fitted and the RMSD calculated using the McLachlan algorithm implemented by ProFit (32). RMSDs are for all non-hydrogen atoms, and in parentheses for C α coordinates only.

The crystal structure of CC-Pent is the first for a de novo designed pentameric α -helical coiled coil, and is one of four pentameric barrels from this study. CC-Hex2, CC-Hex3, and four other sequences are further examples of de novo hexameric coiled-coil barrels (21). CC-Hept is the first parallel, blunt-ended heptamer. The repeats for CC-Hex2 and CC-Hex3 are point mutants of each other; i.e., SLKEIAx and SIKEIAx, respectively. The former has significantly more favorable raw bZIP and fitness scores. The less-favorable score for CC-Hex3 likely reflects the β -branched residues at the **a** positions, Fig. S6, which are analogous here to **d** positions in Type-N, dimeric interfaces where Leu is favored (10, 23). Consistent with this, CC-Hex3 has the beginning of a sharp unfolding transition at high temperatures, whereas CC- is highly thermally stable, Fig. 2B. The sequence repeat of CC-Hept, ALKEIAx, has small Ala residues at both **g** and **e**, as seen in the staggered heptamer (22). These small residues appear critical to dictate the fold, and a heptamer may be the highest oligomeric state possible based on Type-II sequences. As mentioned above, one of our designs was predicted to form an octamer, and this also has **e** = **g** = Ala, Table S3. However, AUC measurements were ambiguous and the peptide did not crystallize. We posit that extending oligomer states past heptamer reliably will require Type-III repeats. However, this may

prove difficult to achieve because the energetic difference between oligomer states is predicted to diminish with increasing number of helices, Fig. S1A; and, therefore, single-chain constructs may be needed to direct specific topologies.

In summary, we have developed a geometrical and computational framework for designing α -helical barrels based on one type of complex coiled-coil packing, namely Type-II interfaces (15). This has produced *de novo* parallel pentameric, hexameric and heptameric structures, none of which are commonly observed in nature, and with a design success rate of ~36% (8/22). These expand the set of rationally designed coiled-coil assemblies from dimer through to heptamer (21, 30). The CC-Pent, CC-Hex2 and CC-Hept structures are based on a similar sequence framework of **a** = Leu plus **d** = Ile. Although this constellation appears to be a good general solution in terms of stability, it is not sufficient to confer oligomer-state specificity alone. That requires contributions from the remaining sites of the helical interfaces; i.e., the **g** and **e** sites, which complete the Type II pattern of **hhxxhhx**, and must possess progressively smaller side chains as oligomer state increases. These structures have accessible and chemically defined channels, opening tantalizing prospects for the rational design of channel- and pore-containing protein assemblies with defined internal chemistries and properties (15, 21, 33).

References

1. C. Pabo, Molecular Technology - Designing Proteins and Peptides. *Nature* **301**, 200 (1983).
2. N. Koga *et al.*, Principles for designing ideal protein structures. *Nature* **491**, 222 (2012).
3. B. Kuhlman *et al.*, Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364 (2003).
4. E. R. G. Main, A. R. Lowe, S. G. J. Mochrie, S. E. Jackson, L. Regan, A recurring theme in protein engineering: the design, stability and folding of repeat proteins. *Curr Opin Struc Biol* **15**, 464 (2005).
5. C. Reichen, S. Hansen, A. Pluckthun, Modular peptide binding: From a comparison of natural binders to designed armadillo repeat proteins. *J Struct Biol* **185**, 147 (2014).
6. K. O. Kopec, A. N. Lupas, beta-Propeller Blades as Ancestral Peptides in Protein Evolution. *PLoS One* **8**, (2013).
7. A. N. Lupas, M. Gruber, The structure of alpha-helical coiled coils. *Adv Protein Chem* **70**, 37 (2005).
8. D. N. Woolfson, The design of coiled-coil structures and assemblies. *Adv Protein Chem* **70**, 79 (2005).
9. E. Moutevelis, D. N. Woolfson, A Periodic Table of Coiled-Coil Protein Structures. *J Mol Biol* **385**, 726 (2009).
10. O. D. Testa, E. Moutevelis, D. N. Woolfson, CC plus : a relational database of coiled-coil structures. *Nucleic Acids Res* **37**, D315 (2009).
11. F. H. C. Crick, The Packing of Alpha-Helices - Simple Coiled-Coils. *Acta Crystallogr* **6**, 689 (1953).
12. G. Offer, R. Sessions, Computer Modeling of the Alpha-Helical Coiled-Coil - Packing of Side-Chains in the Inner-Core. *J Mol Biol* **249**, 967 (1995).
13. P. B. Harbury, J. J. Plecs, B. Tidor, T. Alber, P. S. Kim, High-resolution protein design with backbone freedom. *Science* **282**, 1462 (1998).
14. G. Grigoryan, W. F. DeGrado, Probing Designability via a Generalized Model of Helical Bundle Geometry. *J Mol Biol* **405**, 1079 (2011).
15. D. N. Woolfson, G. J. Bartlett, M. Bruning, A. R. Thomson, New currency for old rope: from coiled-coil assemblies to alpha-helical barrels. *Curr Opin Struc Biol* **22**, 432 (2012).

16. J. Walshaw, D. N. Woolfson, Open-and-shut cases in coiled-coil assembly: alpha-sheets and alpha-cylinders. *Protein Sci* **10**, 668 (2001).
17. K. V. A. Sharff, E. Koronakis, B. Luisi, C. Hughes, Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* **405**, 914 (2000).
18. V. N. Malashkevich, R. A. Kammerer, V. P. Efimov, T. Schulthess, J. Engel, The crystal structure of a five-stranded coiled coil in COMP: A prototype ion channel? *Science* **274**, 761 (1996).
19. D. Meusch *et al.*, Mechanism of Tc toxin action revealed in molecular detail. *Nature* **508**, 61 (2014).
20. L. Sun *et al.*, Icosahedral bacteriophage Phi X174 forms a tail for DNA transport during infection. *Nature* **505**, 432 (2014).
21. N. R. Zaccai *et al.*, A de novo peptide hexamer with a mutable channel. *Nat Chem Biol* **7**, 935 (2011).
22. J. Liu *et al.*, A seven-helix coiled coil. *P Natl Acad Sci USA* **103**, 15457 (2006).
23. P. B. Harbury, T. Zhang, P. S. Kim, T. Alber, A Switch between 2-Stranded, 3-Stranded and 4-Stranded Coiled Coils in Gcn4 Leucine-Zipper Mutants. *Science* **262**, 1401 (1993).
24. J. H. Fong, A. E. Keating, M. Singh, Predicting specificity in bZIP coiled-coil protein interactions. *Genome Biol* **5**, (2004).
25. C. W. Wood *et al.*, CCBuilder: an interactive web-based tool for building, designing and assessing coiled-coil protein assemblies. *Bioinformatics*, (2014). DOI: 10.1093/bioinformatics/btu502
26. G. G. Krivov, M. V. Shapovalov, R. L. Dunbrack, Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **77**, 778 (2009).
27. S. Chaudhury, S. Lyskov, J. J. Gray, PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689 (2010).
28. S. McIntosh-Smith, T. Wilson, A. A. Ibarra, J. Crisp, R. B. Sessions, Benchmarking Energy Efficiency, Power Costs and Carbon Emissions on Heterogeneous Systems. *Comput J* **55**, 192 (2012).
29. J. Walshaw, D. N. Woolfson, SOCKET: A program for identifying and analysing coiled-coil motifs within protein structures. *J Mol Biol* **307**, 1427 (2001).
30. J. M. Fletcher *et al.*, A Basis Set of de Novo Coiled-Coil Peptide Oligomers for Rational Protein Design and Synthetic Biology. *Acs Synth Biol* **1**, 240 (2012).
31. M. Pellegrini-Calace, T. Maiwald, J. M. Thornton, PoreWalker: A Novel Tool for the Identification and Characterization of Channels in Transmembrane Proteins from Their Three-Dimensional Structure. *Plos Comput Biol* **5**, (2009).
32. <http://www.bioinf.org.uk/software/profit/>
33. A. J. Burton *et al.*, Accessibility, Reactivity, and Selectivity of Side Chains within a Channel of de Novo Peptide Assembly. *J Am Chem Soc* **135**, 12524 (2013).

Acknowledgements: CWW and AJB are supported by the BBSRC South West Doctoral Training Partnership and the EPSRC Bristol Chemical Synthesis Centre for Doctoral Training, respectively; the work was funded by grants from the EPSRC (EP/J001430/1; DNW), BBSRC (BB/J008990/1; DNW and RLB) and the ERC (340764; DNW). DNW holds a Royal Society Wolfson Research Merit Award. We thank the Diamond Light Source for access to beamlines I0-3 & I0-4 (award MX8922). We thank Nathan Zaccai for advice on protein X-ray crystallography, and members of the Brady and Woolfson groups for general discussions. ART and DNW conceived the research programme and approach; ART designed the peptide sequences; ART, CWW and AJB synthesized the peptides and performed solution-phase characterizations; CWW, AJB and RLB solved the X-ray crystal structures; ART, CWW, GJB and RBS wrote the computational design and modeling algorithms; ART and DNW wrote the paper. All authors analyzed the data, and reviewed and contributed to the manuscript. Coordinates and structure factors have been deposited in the Protein Data Bank with the accession codes: CC-Pent, 4pn8; CC-Hex2, 4pn9;

CC-Hex3, 4pnb; CC-Hept, 4pna; CC-Pent-variant, 4pnd.

Supplementary Materials

Materials and Methods

Figs. S1 to S7

Tables S1 to S4

References (34 – 41)