De novo design of discrete, stable 3_{10}-helix peptide assemblies

Prasun Kumar[1,*], Neil G. Paterson[2], Jonathan Clayden[1] and Derek N. Woolfson[1,3,4,*]

[1]School of Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS, UK
[2]Diamond Light Source, Harwell Science and Innovation Campus, Didcot, OX11 0DE, Oxfordshire, UK
[3]School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol BS8 1TD, UK
[4]BrisSynBio, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK

*To whom correspondence should be sent: prasun.kumar@bristol.ac.uk; d.n.woolfson@bristol.ac.uk

The iconic α helix is pre-eminent in structural biology[1] and widely exploited in protein folding[2], design[3], and engineering[4]. Although other helical peptide conformations do exist near to the α-helical region of conformational space—namely, 3_{10} helices and π helices[5]—these occur much less frequently in protein structures. Less favourable internal energies and reduced tendencies to pack into higher-order structures mean that 3_{10} helices rarely extend past 6 residues in natural proteins, and that they tend not to form regular supersecondary, tertiary, or quaternary interactions. Here we show that despite their absence in nature, synthetic peptide assemblies can be built from 3_{10} helices. We report the rational design, solution-phase characterisation, and an X-ray crystal structure for water-soluble bundles of 3_{10} helices with consolidated hydrophobic cores. The design uses 6-residue repeats informed by analysing 3_{10}-helical conformations in known protein structures, and incorporates 2-aminoisobutyric acid (Aib) residues. Design iterations reveal a tipping point between α-helical and 3_{10}-helical folding, and identify features required for stabilising assemblies of 3_{10} helices. This work provides principles and rules to open opportunities for designing into this hitherto unexplored region of protein-structure space.

The challenge of designing beyond the iconic α helix

The α helix is one of the fundamental constants of biology: it is a cornerstone of structural biology[1], a model for protein-folding studies[2,6], a workhorse in protein design[3,7], and a scaffold for displaying functional moieties in protein engineering and biotechnology[4,8]. The α helix is highly defined conformationally: it has a narrow range of backbone-torsion or Ramachandran angles[9], which leads to tight helical parameters, Fig. 1; and stabilising intrahelical backbone hydrogen bonding between residues 'i' and 'i+4' along the polypeptide chain, $CO_i \rightarrow NH_{i+4}$. Put another way, energetically, the α helix sits in a narrow and deep free-energy well[10].

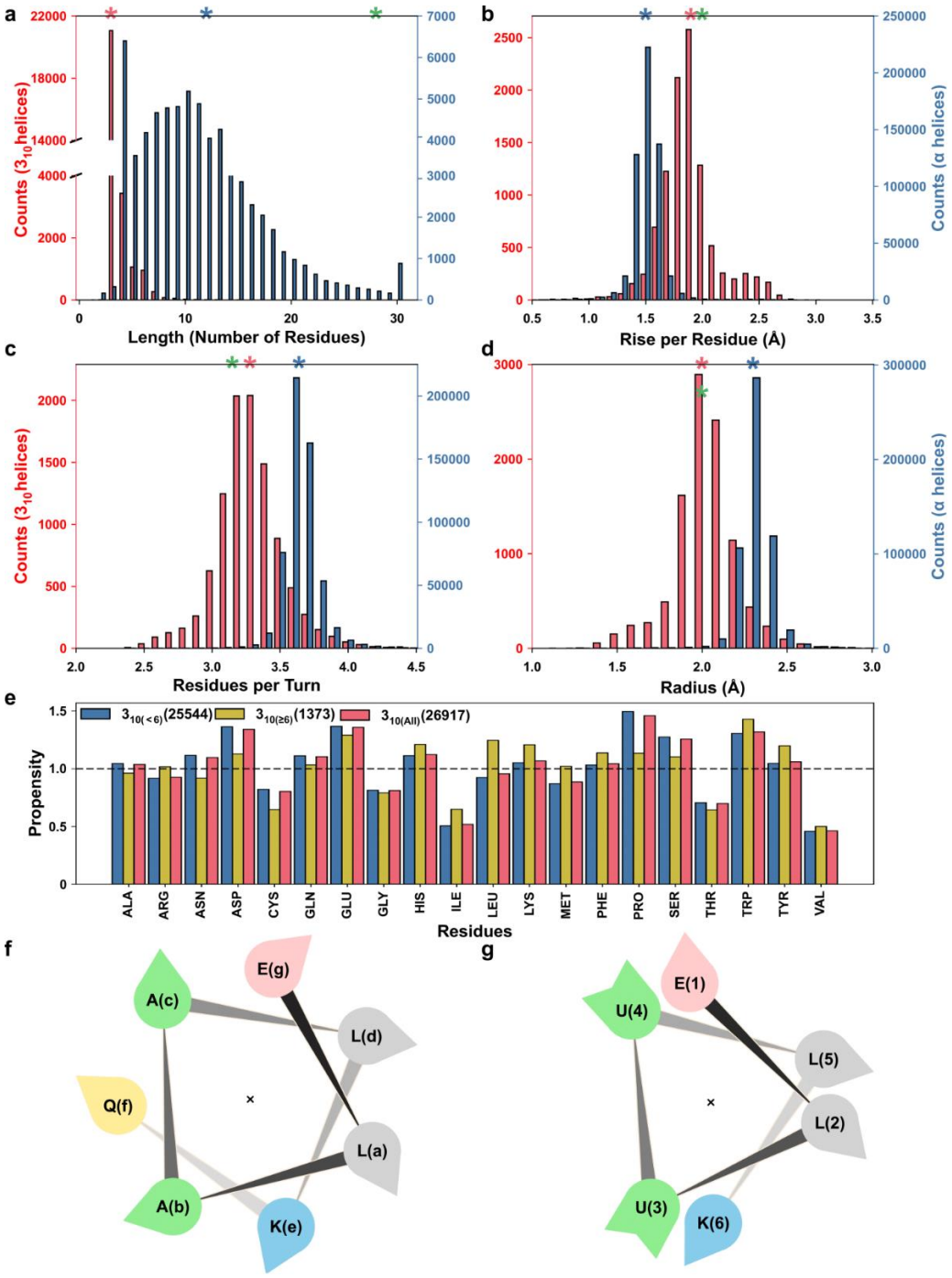

Fig. 1: Analysis of the PDB and design principles for 3_{10} helices. Helical parameters for 3_{10} helices (red) and α helices (blue) calculated from the PDB: (**a**) helix length; (**b**) rise per residue; (**c**) residues per turn; and (**d**) radius. Mean values are marked with red and blue asterisks. The green asterisks show the mean values from the new quaternary structure of 3_{10} helices reported here. For panel **A**, all a helices over 29 residues long are plotted as a single bar at 30 residues; the longest α helix found had 97 residues. (**e**) Propensity values of the 20 standard amino acids in 3_{10} helices, divided into two categories for helices shorter than 6 residues long ($3_{10\,(<6)}$; blue) or greater than 5 residues long ($3_{10\,(\geq6)}$; orange). (**f** and **g**) Helical-wheel diagrams for a *de novo* 7-residue sequence repeat on an α helix (**f**), and a *de novo* 6-residue repeat on a 3_{10} helix (**g**). Standard amino acids are represented by their one letter codes, and U denotes Aib.

By contrast, 3_{10} and π helices have different helical parameters leading to tighter and looser helical structures with $CO_i \rightarrow NH_{i+3}$ and $CO_i \rightarrow NH_{i+5}$ hydrogen-bonding patterns,

respectively. They lie on the rim of the α-helical free-energy well and are, thus, thermodynamically less stable. Consequently, it is not surprising that these potentially accessible polypeptide conformations occur less commonly in natural proteins. Indeed, we surveyed a non-redundant set of protein structures from the RCSB Protein Data Bank (PDB)[11] and found that while 33% of the >2 million residues were located within α helices, only 4% were within 3_{10} helices, and 0.5% within π helices, Fig. 1 and Supplementary Information Table S7.1. The 3_{10} and π helices were also much shorter than the α helices, Fig. 1a and Supplementary Information Fig. S1.1. Furthermore, we found no examples in proteins where consolidated packing arrangements—i.e., supersecondary, tertiary or quaternary interactions—are formed exclusively by these alternative helices. In comparison, such structures based on α helices are widely catalogued[12,13]. Evidently, nature appears not to use 3_{10} or π helices to construct higher-order peptide and protein assemblies in aqueous media. This raises the question of whether such objects can nonetheless exist, and the challenge of whether chemists might be able to make them.

One group of natural peptides do contain regions of 3_{10}-helical conformation; namely, the membrane-active fungal metabolites known as the peptaibols[14,15]. These short peptides are rich in α,α-disubstituted amino acids, in particular α-aminoisobutyric acid (Aib, U)[16-21], which favour tighter helical turns characteristic of 3_{10} helices[22,23]. However, the formation of 3_{10} helices appears to fall off with lengths >9 residues, where α-helix formation becomes favoured[24]. Nonetheless, long, synthetic polymers of Aib do form free-standing 3_{10} helices[17,18,25], and these have been used as functional synthetic models of membrane-penetrating proteins such as G-protein coupled receptors[26] and rhodopsin[27]. However, peptide solubility in water is lowered by increased content of hydrophobic Aib. Polar and charged groups can be added to increase water solubility of 3_{10}-helical peptides, but atomistic structures of these remain elusive[28,29].

Thus, whilst natural peptide and protein chains do access 3_{10}-helical conformations, these tend not to propagate beyond short stretches, and they do not facilitate higher-order tertiary and quaternary interactions. Furthermore, the folding and assembly of 3_{10} helices appears to be even more limited in aqueous solution. We hypothesised that if we could design a synthetic (*de novo*) sequence to form an amphipathic 3_{10} helix of sufficient length—*i.e.*, one with distinct hydrophobic and polar faces—it could be stabilised by helix-helix interactions to form a 3_{10}-helical bundle. Here, we present the design and high-resolution structure of such a bundle.

Bioinformatic analysis provides rules for designing 3_{10} helices

We took a rational design approach to generate repeat sequences aimed at forming long 3_{10} helices in aqueous solution. Our starting point was the set of protein structures harvested from the PDB and analysed in Fig. 1. Although 3_{10} helices are fewer in number and shorter in length than α helices, Fig. 1a, the data set contained sufficient examples to calculate propensities for each of the 20 proteinogenic residues to adopt this conformation compared with any other state, Fig. 1b and Supplementary Information Table S7.1. These data indicated that for 3_{10} helices of ≥6 residues, the residues with the strongest preferences for 3_{10}-helical secondary structures included glutamate (Glu, E), lysine (Lys, K), leucine (Leu, L) and tryptophan (Trp, W).

It is well understood that amphipathic α-helical peptides can be stabilised by association into coiled-coil bundles, and design principles for these are well established[3,7,30]. Therefore, we began our design process with a control α-helical bundle using the above amino-acid palette plus alanine (Ala, A). Amphipathic α helices can be encoded by placing hydrophobic residues 3 and 4 residues apart, as the average 3.5-residue spacing closely matches the 3.6 residues per turn of the helix, Fig. 1f. Based on past experience[31], we designed the 7-residue repeat, E-L-A-A-L-K-X, where X can be any amino acid. We repeated this 4 times to make peptide PK−1 (systematically named CC-TypeN-L_aL_d), Table 1.

Short name	Systematic peptide name	Sequence	AUC		XRD
			SV	SE	
PK-1	CC-TypeN-L_aL_d	G ELAALKQ ELAALKW ELAALKE ELAALKB G	3	3	1.41 Å, 7qdk
PK-2	not given	G ELUULKQ ELUULKW ELUULKE ELUULKY G	3	3	-
PK-3	not given	G ELAALK ELAALK ELAALK ELAALK WKG	1	1	-
PK-4	3_{10}HD	G ELUULK ELUULK ELUULK ELUULK UUU WKG	6	6	-
PK-5	D-3_{10}HD	G elUUlk elUUlk elUUlk elUUlk UUU bkG	8	7	2.34 Å, 7qdi
PK-6	not given	G EAUUAK EAUUAK EAUUAK EAUUAK UUU WKG	1	1	-
PK-7	not given	G EUUUUK EUUUUK EUUUUK EUUUUK UUU WKG	1	1	-
PK-8	not given	G EIUUIK EIUUIK EIUUIK EIUUIK UUU WKG	1	1	-
PK-9	not given	G EVUUVK EVUUVK EVUUVK EVUUVK UUU WKG	1	1	-
PK-10	not given	G ELUULK ELUULK UUU WKG	1	1	1.44 Å, 7qdj
PK-11	not given	G elUUlk elUUlk UUU wkG	1	1	
PK-12	not given	G ELUULK ELUULK ELUULK UUU WKG	4-6	4-6	-
PK-13	3_{10}HD-A	G ELUULE ELUULE ELUULE ELUULE UUU WKG	1	1	
PK-14	3_{10}HD-B	G KLUULK KLUULK KLUULK KLUULK UUU WKG	1	1	-

Table 1: De novo designed peptide sequences. Standard amino acids are represented by their one letter codes, and U denotes Aib and B 4-bromo-l-phenylalanine. D-Amino acids are in lowercase. For the analytical ultracentrifugation (AUC) sedimentation velocity (SV) and sedimentation equilibrium (SE), the numbers indicate the oligomeric states determined from these experiments. Resolution (Å) and PDB identifiers are given for structures solved by X-ray diffraction (XRD). Peptides were made by solid-phase peptide synthesis and confirmed by mass spectrometry, see Supplementary Information for details.

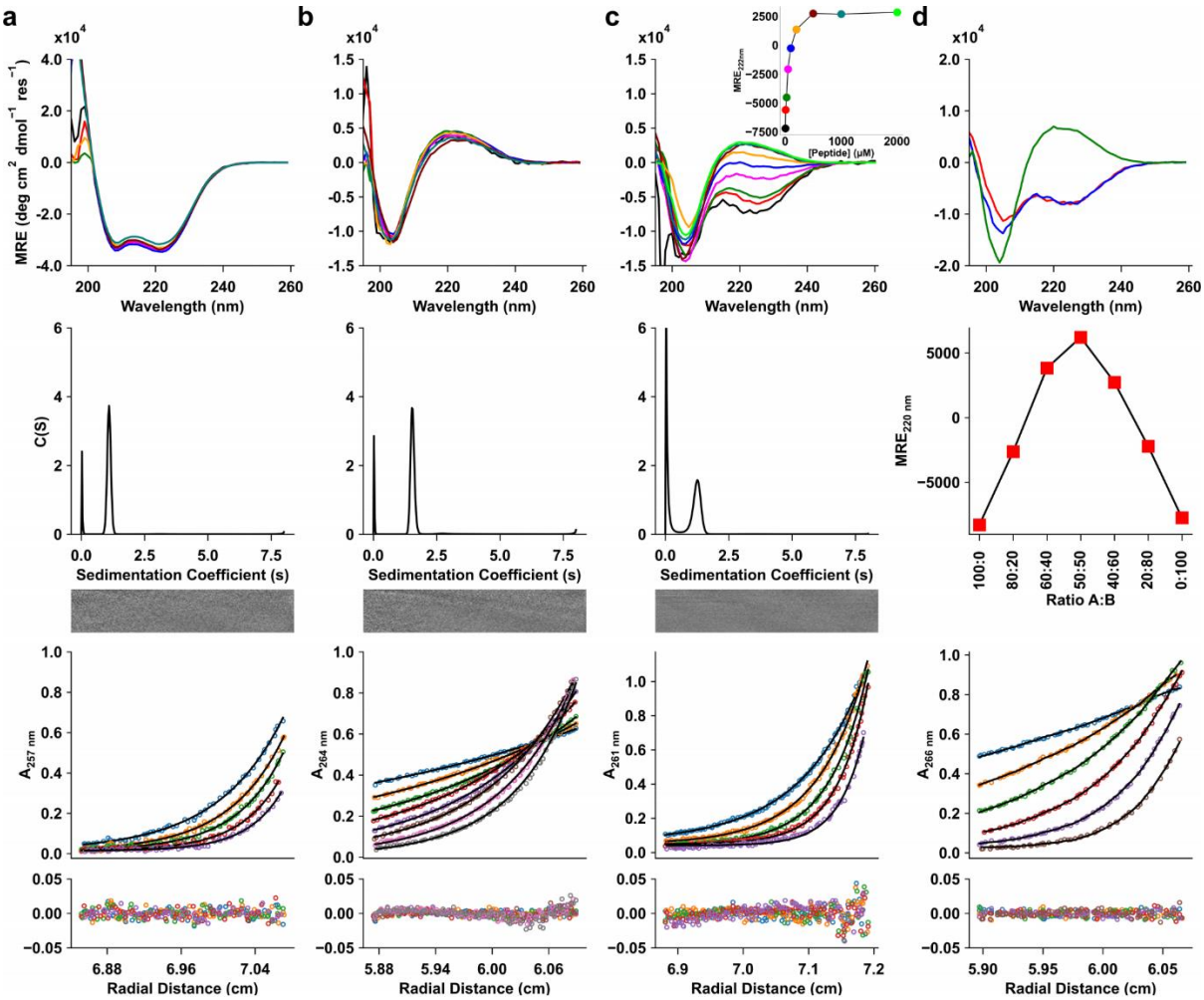

Fig. 2: Biophysical characterization of *de novo* designed peptides. (**a** – **c**) Data for peptides PK-1 (**a**), PK-4 (3_{10}HD) (**b**), and PK-12 (**c**), from top to bottom: CD spectra recorded at 5 °C; SV c(s) distributions; and SE plots. Conditions for CD experiments: PBS, pH 7.4, at peptide concentrations of 5 μM (black), 10 μM (red), 25 μM (green), 50 μM (magenta), 100 μM (blue), 200 μM (orange), 500 μM (maroon), 1000 μM (light green), and 2000 μM (lime; panel **C** only). AUC conditions: 20 °C, PBS, pH 7.4, 100 μM peptide concentration. Rotor speeds for SV: 60K RPM; Rotor speed for SE in panel **a** and c: 44k (blue), 48k (orange), 52k (green), 56k (red) and 60k (purple) RPM; Rotor speed for SE for panel **b**: 15k (blue), 18k (orange), 21k (green), 24k (red), 27k (purple), 30k (brown), 33k (pink) and 36k (grey) RPM. The inset in the top panel of **c** plots the change in MRE_{222} with peptide concentration. (**d**) Heteromeric 3_{10} design (3_{10}HD-AB), from top to bottom: CD spectra at 5 °C, 100 μM peptide concentrations for the acidic (PK-13; red), basic (PK-14; blue) peptides and mixture (green); Job plot for the mixture of acidic and basic peptides; SE plots for the mixture. Rotor speed for SE for panel **d**: 18k (blue), 24k (orange), 30k (green), 36k (red), 42k (purple) and 48k (brown) RPM.

Iterative design and experiments deliver water-soluble 3_{10}-helical assemblies

Circular dichroism (CD) spectroscopy showed that PK−1 adopted a thermally stable α helix in solution, Fig. 2a and Supplementary Information Fig. S3.1, and analytical ultracentrifugation (AUC) indicated that it formed a monodisperse trimeric species, Fig. 2a and Supplementary Information Fig. S4.1. We crystallised and determined an X-ray structure for PK−1, which revealed a parallel, trimeric, α-helical, coiled-coil bundle, Fig. 3a and Supplementary Information Tables S7.2&3, consistent with the coiled-coil design[30-32].

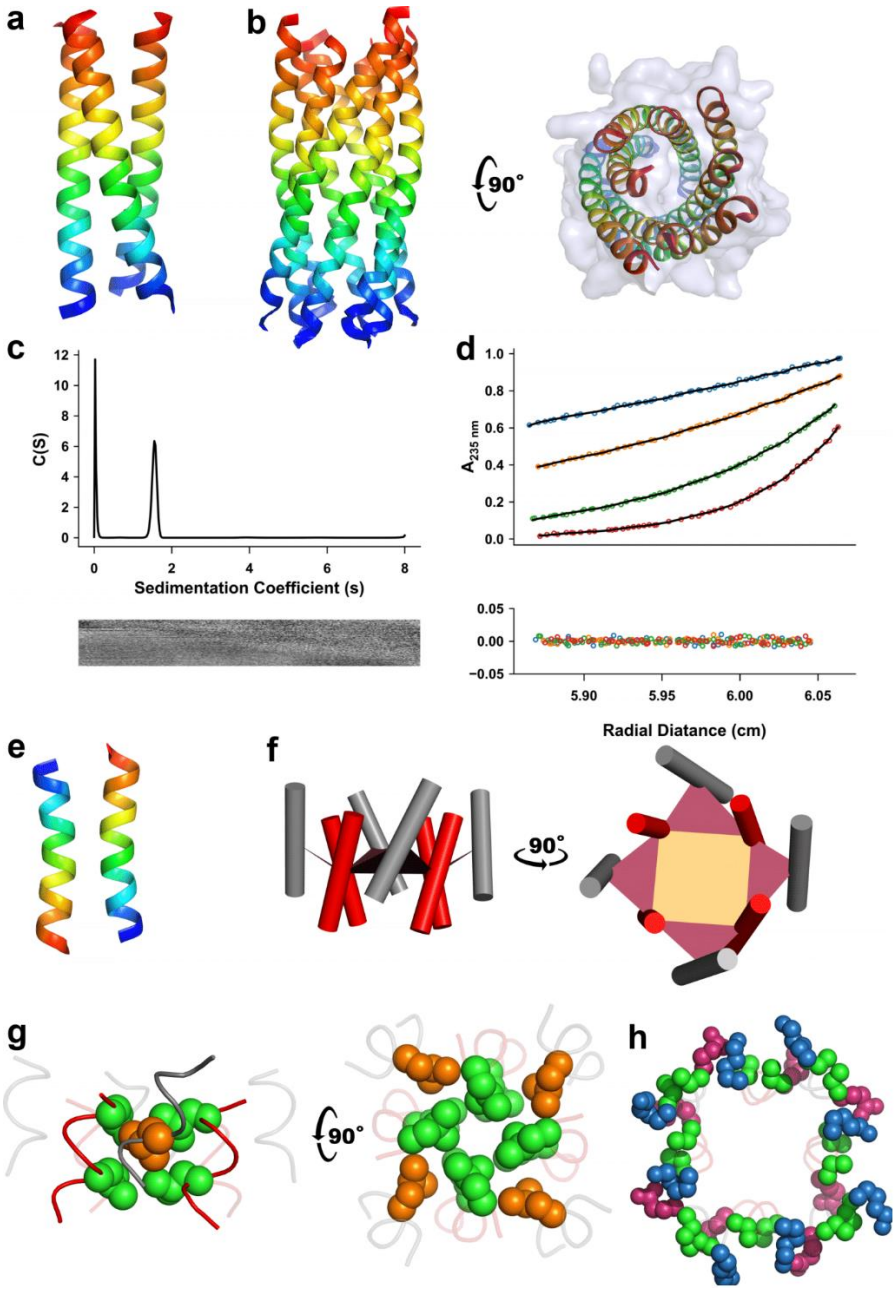

***Fig. 3: Crystal structures of de* novo designed peptides.** Backbone ribbon structures coloured blue to red from the *N* to *C* termini for (**a**) PK-1 (CC-TypeN-L_aL_d) (PDB id 7qdk), and (**b**) orthogonal views of PK-5 (D-3_{10}HD) (7qdi). (c) SV data and fits (**top**) with residuals (**bottom**) for PK-5 ($\bar{v}$ = 0.7957 $cm^3\ g^{-1}$). SV continuous c(s) distribution at 60 krpm (s = 1.552 S, s20,w = 1.624 S, f/f0 = 1.49, Mw = 27474 Da, 8.1x monomer mass at a 95% confidence level). (**d**) SE data and fitted curves (**top**) with residuals (**bottom**) for a single, ideal species model (Mw = 23334.9 Da, 6.9x monomer mass, 95% confidence limits 21818.5 – 24789.9 Da). Conditions: 20 °C, PBS, pH 7.4, and rotor speeds of 15K, 20K, 30K and 40K rpm. SV and SE experiments were conducted at 50 µM concentrations. (**e**) Backbone ribbon structure of the PK-10/PK-11 racemate (7qdj). (**f - h**) More views of the PK-5 (D-3_{10}HD) structure: (**f**) Orthogonal views showing the outer helices (grey cylinders) shifted with respect to the inner helices (red cylinders), with planes drawn from the centres of identical sequence repeats; (**g**) similar to panel **f** with Leu side chains shown in space-filling representation to highlight the intimate packing to form a consolidated hydrophobic core; and (**h**) packing of D-Glu (magenta) and D-Lys (blue) residues, plus Aib (green) residues in space-filling representations. *N.B.* Whilst most of the residues are resolved in the X-ray crystal structure of PK-5, electron density for 10 out of 16 terminal Gly residues could not be located, which we attribute to helix fraying confined to those residues.

Next, as 3_{10}-helix conformations can be stabilised by including Aib residues, we replaced the two Ala residues in each 7-residue repeat of PK-1 to give PK-2, Table 1. However, like its parent, PK-2 remained a highly α-helical, thermally stable, and monodisperse trimer in solution, Supplementary Information Figs. S3.2 and S4.2. Thus, simply placing 3_{10}-helix-favouring Aib residues within 7-residue repeats does not override the propensity of these to form α-helical bundles. Presumably, this is because of the strong tendency of alternating 3,4-pattern of hydrophobic residues to direct amphipathic α-helix formation and tight coiled-coil packing[33]. Therefore, our next step was to break this pattern and find an alternative repeat to favour 3_{10}-helix conformations.

On average, the 3_{10} helices of our dataset had 3.28 residues per turn, Fig. 1c; though the ideal 3_{10}-helix parameter is often stated as 3 residues per turn[5]. Therefore, we reasoned that 3_{10}-helical bundles might be stabilised by reducing the sequence repeat to 3 residues with a single hydrophobic residue per repeat. Our contention was that in an ideal 3_{10} helix, these residues would form a continuous hydrophobic 'seam' of an amphipathic 3_{10} helix to promote intermolecular association in water, Fig. 1g. Also, by analogy with their use in designed coiled-coil bundles[31], we hypothesised that inter-chain salt bridges might stabilise bundles of such helices further. This led us to design two peptides, PK-3 and PK-4 (Table 1), with 4 x 6-residue repeats, E-L-Z-Z-L-K, with Z = all Ala (A) or all Aib (U), respectively.

In aqueous buffer, by CD spectroscopy, PK-3 was largely α helical with a broad thermal unfolding transition, Supplementary Information Fig S3.3. These CD data did not show any dependence on peptide concentration. Consistent with this, AUC experiments confirmed PK-3 as a monodisperse monomer, Supplementary Information Fig. S4.3. The peptide did not crystallise. These data are indicative of single α-helical domains[34]; *i.e.*, α helices that do not associate into higher-order structures. Thus, though the 3-residue hydrophobic repeat removes the drive of the α helix to assemble into bundles, alone it does not stabilise 3_{10} helices or bundles thereof.

By contrast, PK-4 behaved completely differently. First, its CD spectrum was patently different from the three previous peptides, Fig. 2b, with a minimum at ≈205 nm characteristic of right-handed helices, and a maximum at ≈220 nm characteristic of right-handed 3_{10} helices in aqueous buffer[35]. The CD spectrum changed little upon heating, Supplementary Information Fig S3.4. However, in the presence of 3 M guanidinium chloride, we were able to obtain a thermal denaturation curve for PK-4, Supplementary Information Fig. S3.14. This was reversible and sigmoidal indicative of a unique, cooperatively folded species. Consistent with

this, AUC measurements indicated that PK-4 associated into a monodisperse hexamer, Fig. 2b and Supplementary Information Fig. S4.4. *De novo* designed α-helical coiled coils of this size can form barrels[36] with central lumens that bind the environment-sensitive dye, 1,6-diphenylhexatriene (DPH)[37]. However, PK-4 did not bind DPH in solution, Supplementary Information Fig S5.1, suggesting a tightly packed structure with a consolidated hydrophobic core.

X-ray crystallography reveals a new 3_{10}-helix bundle

We made many attempts to crystallise and solve X-ray structures of PK-4 and its derivatives. This included peptides with L- and D-amino acids, and racemic mixtures of these. Eventually, we crystallised and solved a 2.34 Å resolution structure (Fig. 3b and Supplementary Information Table 7.2&3) for a variant, PK-5, with D-Glu, D-Lys and D-Leu and a *C*-terminal D-4-bromophenylalanine (D-Br-Phe), Table 1, Figs. 3c&d, Supplementary Information S3.5 and S4.5. The structure revealed a parallel bundle of eight left-handed 3_{10} helices with otherwise canonical 3_{10}-helix parameters (green asterisks, Fig. 1a-d). Left-handed helices are expected for D-peptide structures. Notably, however, the helices of PK-5 have ≈3.15 residues per turn, and nine contiguous turns (27 residues) of $CO_i \rightarrow NH_{i+3}$ hydrogen bonding. As 3.15 is slightly greater than the 3-residue spacing of the leucine residues, these hydrophobic residues track around the left-handed helices in a right-handed manner. This allows the helices to pack in a right-handed supercoil with an extrapolated pitch of 93 residues or 177 Å. The rare examples of long 3_{10} helices in natural proteins tend to be irregular[38]. By contrast, those in the X-ray crystal structure of PK-5 curve smoothly, Supplementary Information Fig. S6.1[39].

In more detail, the structure of PK-5 has C_4 symmetry with pairs of adjacent helices contributing to an inner four-helix bundle and an outer four-helix ring, Fig. 3f. The packing of the side chains is intimate and reminiscent of 'knobs-into-holes' packing in α-helical coiled coils[33]. The Leu residues of the inner helices point directly into a central core, similar to 'x-layers' in some α-helical coiled coils [40], while those of the outer helices pack into constellations of Leu residues provided by the inner helices, Fig. 3g. The latter requires the outer helices to be slipped relative to the inner by ≈2 Å. The solvent-accessible surface of the assembly comprises entirely D-Glu and D-Lys polar residues. In accord with our design strategy, these form a network of salt bridges, with the closest pairs having C_δ-N_ε distances of 3.8 ± 0.5 Å (*N* = 32). Together with the Aib residues, these shield the hydrophobic core from solvent, Fig. 3h.

Although both are discrete oligomers, the octameric structure of PK-5—which was confirmed in solution by AUC, Fig. 3c&d and Supplementary Information Fig. S4.5—differs in oligomeric state from the hexamer indicated by the solution-phase data of the parent PK-4 (Fig. 2b), which has Trp in place of *p*-Br-Phe. The extended crystal lattice of PK-5 reveals side-by-side and head-to-tail packing of octamers, with the D-*p*-Br-Phe residues packed in a layer of with many edge-to-face aromatic interactions[41], Supplementary Information Fig. S6.2. Modelling D-Trp residues into these sites revealed potential steric clashes between the octamers, which may in part explain why PK-4 did not crystallise. Thus, the powerful tendency to assemble programmed into the amphipathic *de novo* 3_{10}-helical sequences can be modulated by subtleties in sequence leading to varying degrees of oligomerisation.

Design variants uncover further principles for 3_{10}-helix formation

Thus, PK-4 and PK-5 form parallel, bundles of six and eight 3_{10} helices, respectively. Whilst thermostable, these both unfold cooperatively and reversibly with chemical denaturants, Supplementary Information Fig. S3.14. Such assemblies are unprecedented in both the length of the 3_{10}-helices involved and in being the first water-soluble, supramolecular or quaternary assemblies of such helices. With our design goal achieved, we renamed PK-4 and PK-5 as 3_{10}HD and D-3_{10}HD (3_{10}-Helical Designs), respectively. This success raises further questions: why are similar structures not found in natural proteins? And, which features of our *de novo* designed sequence make it so disposed to form stable 3_{10}-helix-based quaternary structures? To address these questions, we made variants of 3_{10}HD.

To start, we tested the effect of hydrophobicity and steric size at the 'leucine sites' of the 6-residue repeats on the folding and assembly of 3_{10}HD. First, we replaced all of the Leu residues of 3_{10}HD with Ala to give PK-6, Table 1. CD and AUC measurements showed that this peptide was a partially α-helical (≈50%) monomer in solution, Supplementary Information Fig. S3.6 and S4.6. Similarly, when the Leu residues were replaced by Aib in PK-7, Table 1, the peptide remained monomeric but with an unusual CD spectrum, Supplementary Information Fig. S3.7 and S4.7. The β-branched hydrophobic residues, Ile and Val, have low propensities for 3_{10} helicity in natural structures, Fig. 1e. To examine this, we changed all of the Leu residues to Ile (PK-8) or Val (PK-9), Table 1. Neither peptide gave a CD spectrum consistent with 3_{10} helicity (Supplementary Information Figs. S3.8 and S3.9), and both were monomers in solution (Supplementary Information Figs. S4.8 and S4.9). Combined, these experiments indicate that, of the proteinogenic aliphatic hydrophobic side chains, Leu best promotes the supramolecular assembly and stabilisation of 3_{10} helices in our system.

Next, we tested the effect of changing peptide length on the stability of the 3_{10}-helix bundle. We kept the overall sequence repeat, ELUULK, but systematically decreased the number of these from 4 to 3 and then to 2 repeats, giving PK-12 and PK-10, respectively, Table 1. The shortest peptide, PK-10, was a partly folded monomer in solution, Supplementary Information Figs. S3.10a and S4.10a; and a structure of a racemic mixture (PK-10+PK-11) at 1.4 Å resolution revealed two antiparallel α helices packed with 'knobs-into-holes' interactions as observed in other heterochiral systems[33,42], Fig. 3e, and Supplementary Information Figs. S3.10 and S4.10. By contrast, and interestingly, CD spectra of PK-12 revealed a cooperative concentration-dependent switch from a partially α-helical to a 3_{10}-helical conformation, Fig. 2c and Supplementary Information Fig. S3.11: as the peptide concentration was increased from 5 μM to 2 mM, the negative maximum at 226 nm changed to a positive maximum at 220 nm. Concomitantly, the oligomeric state of the peptide changed from multiple low-order species at 25 μM to hexamer at higher concentrations, Supplementary Information Figs. S4.11&12. Attempts to crystallise PK-12 failed. Nonetheless, the series PK-4 (3_{10}HD) → PK-12 → PK-10 shows that peptide length, and with it the length of the hydrophobic seam, is critical for the folding, assembly, and stabilisation of 3_{10} helices.

Finally, exploiting the symmetry and salt bridging of the D-3_{10}HD structure, we designed a heteromeric system comprising acidic, PK-13, and basic, PK-14, peptides, Table 1. In isolation in solution, the two peptides were partially folded, α-helical monomers, Fig. 2d (top), and Supplementary Information Figs. S4.13&14. However, when mixed, the CD spectrum switched to that of the parent homomeric assembly, Figs. 2b&d, indicating 3_{10}-helix formation. Moreover, a Job plot revealed a 1:1 stoichiometry for the complex, Fig. 2d (middle) and Supplementary Information S3.13. Finally, AUC measurements returned a monodisperse hexamer similar to the parent peptide, 3_{10}HD, Figs. 2b&d (bottom) and Supplementary Information Figs. S4.4&15. Retrospectively, we named this completely new *de novo* designed heteromeric quaternary complex of 3_{10} helices 3_{10}HD-AB.

Conclusion

In conclusion, we have achieved the rational *de novo* design of unprecedented water-soluble supramolecular assemblies or quaternary structures, constructed from 3_{10}-helical peptides. The designs incorporate: (i) a bioinformatically guided reduced amino-acid alphabet; (ii) the non-proteinogenic α,α-disubstituted amino acid, Aib; (iii) strict 6-residue sequence repeats; and (iv) a minimum number of 3 such repeats. All but one of these features (*i.e.*, number ii) could be achieved through natural ribosomal protein synthesis, which raises the

question: why has nature not found and exploited these or similar structures? The significance of our successful designs is that they require extended and amphipathic 3_{10} helices, which are rare in nature. We suggest that these requirements are critical, otherwise polypeptide chains will adopt alternative α-helical conformations, which are nearby in conformational space and energetically more favourable. Indeed, it is noteworthy that even non-ribosomal peptide natural products rich in Aib, such as alamethicin and the cephaibols, crystallise in α-helical conformations[43,44]. Moreover, although Aib stabilises 3_{10} helices in short peptides[15,16], and all-Aib polymers can adopt 3_{10}-helical conformations [18], analysis of the PDB reveals that Aib is found predominantly in α-helical secondary structures, Supplementary Information Fig. S1.3. Nonetheless, we provide decisive proof that this preference can be overridden to achieve stabilised 3_{10}-helical conformations and assemblies. It will be interesting to see what explorations of new structural and functional peptide chemistry this route into a seemingly unexplored region of protein-structure space opens up. For instance, Aib and similar residues can be incorporated into engineered proteins[45], the chemistry of α,α-disubstituted amino acids is being expanded[46], and other foldamers are being developed as quaternary assemblies[47,48] and exploited as functional frameworks for binding and catalysis[49,50].

References

1 Pauling, L., Corey, R. B. & Branson, H. R. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci U S A* **37**, 205-211 (1951).

2 Chakrabartty, A. & Baldwin, R. L. Stability of alpha-helices. *Adv Protein Chem* **46**, 141-176 (1995).

3 Korendovych, I. V. & DeGrado, W. F. De novo protein design, a retrospective. *Q Rev Biophys* **53**, e3 (2020).

4 Lapenta, F., Aupic, J., Strmsek, Z. & Jerala, R. Coiled coil protein origami: from modular design principles towards biotechnological applications. *Chem Soc Rev* **47**, 3530-3542 (2018).

5 Schulz, G. E. & Schirmer, R. H. *Principles of Protein Structure*. (Springer-Vcrlag, NewYork, 1979).

6 Scholtz, J. M. & Baldwin, R. L. The mechanism of alpha-helix formation by peptides. *Annu Rev Biophys Biomol Struct* **21**, 95-118 (1992).

7 Woolfson, D. N. A Brief History of De Novo Protein Design: Minimal, Rational, and Computational. *J Mol Biol*, 167160 (2021).

8 Dawson, W. M., Rhys, G. G. & Woolfson, D. N. Towards functional de novo designed proteins. *Curr Opin Chem Biol* **52**, 102-111 (2019).

9 Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J Mol Biol* **7**, 95-99 (1963).

10 Kuster, D. J., Liu, C., Fang, Z., Ponder, J. W. & Marshall, G. R. High-resolution crystal structures of protein helices reconciled with three-centered hydrogen bonds and multipole electrostatics. *PLoS One* **10**, e0123146, doi:10.1371/journal.pone.0123146 (2015).

11 Burley, S. K. *et al.* RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res* **49**, D437-D451 (2021).

12 Chothia, C., Levitt, M. & Richardson, D. Helix to helix packing in proteins. *J Mol Biol* **145**, 215-250 (1981).

13 Orengo, C. A. *et al.* CATH--a hierarchic classification of protein domain structures. *Structure* **5**, 1093-1108 (1997).

14 Gessmann, R., Axford, D., Owen, R. L., Bruckner, H. & Petratos, K. Four complete turns of a curved 3(1)(0)-helix at atomic resolution: the crystal structure of the peptaibol trichovirin I-4A in a polar environment suggests a transition to alpha-helix for membrane function. *Acta Crystallogr D Biol Crystallogr* **68**, 109-116 (2012).

15 Toniolo, C. & Brückner, H. *Peptaibiotics*. (John Wiley & Sons, 2009).

16 Toniolo, C. & Benedetti, E. The polypeptide 310-helix. *Trends Biochem Sci* **16**, 350-353 (1991).

17 Gessmann, R., Bruckner, H. & Petratos, K. The crystal structure of Z-(Aib)10-OH at 0.65 A resolution: three complete turns of 310-helix. *J Pept Sci* **22**, 76-81 (2016).

18 Solà, J., Helliwell, M. & Clayden, J. Interruption of a 310-helix by a single Gly residue in a poly-Aib motif: A crystallographic study. *Biopolymers* **95**, 62-69 (2011).

19 Pike, S. J., Boddaert, T., Raftery, J., Webb, S. J. & Clayden, J. Participation of non-aminoisobutyric acid (Aib) residues in the 310 helical conformation of Aib-rich foldamers: a solid state study. *New J Chem* **39**, 3288-3294 (2015).

20 Karle, I. L., Flippen-Anderson, J. L., Gurunath, R. & Balaram, P. Facile transition between 310- and α-helix: Structures of 8-, 9-, and 10-residue peptides containing the -(Leu-Aib-Ala)2-Phe-Aib-fragment. *Protein Sci* **3**, 1547-1555 (1994).

21 Toniolo, C. *et al.* Preferred conformation of the terminally blocked (Aib)10 homo-oligopeptide: A long, regular 310-helix. *Biopolymers* **31**, 129-138 (1991).

22 Nagaraj, R. & Balaram, P. Alamethicin, a transmembrane channel. *Acc Chem Res* **14**, 356-362 (1981).

23 Toniolo, C. *et al.* Conformation of pleionomers of. alpha.-aminoisobutyric acid. *Macromolecules* **18**, 895-902 (1985).

24 Karle, I. L. & Balaram, P. Structural characteristics of alpha-helical peptide molecules containing Aib residues. *Biochemistry* **29**, 6747-6756 (1990).

25 Byrne, L. *et al.* Foldamer-mediated remote stereocontrol: >1,60 asymmetric induction. *Angew Chem Int Ed Engl* **53**, 151-155, doi:10.1002/anie.201308264 (2014).

26 Lister, F. G. A., Le Bailly, B. A. F., Webb, S. J. & Clayden, J. Ligand-modulated conformational switching in a fully synthetic membrane-bound receptor. *Nat Chem* **9**, 420-425 (2017).

27 De Poli, M. *et al.* Conformational photoswitching of a synthetic peptide foldamer bound within a phospholipid bilayer. *Science* **352**, 575-580 (2016).

28 Formaggio, F. *et al.* The First Water-Soluble 310-Helical Peptides. *Chemistry* **6**, 4498-4504 (2000).

29 Zieleniewski, F., Woolfson, D. N. & Clayden, J. Automated solid-phase concatenation of Aib residues to form long, water-soluble, helical peptides. *Chem Commun (Camb)* **56**, 12049-12052 (2020).

30 Woolfson, D. N. Coiled-Coil Design: Updated and Upgraded. *Subcell Biochem* **82**, 35-61 (2017).

31 Fletcher, J. M. *et al.* A basis set of de novo coiled-coil peptide oligomers for rational protein design and synthetic biology. *ACS Synth Biol* **1**, 240-250 (2012).

32 Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* **262**, 1401-1407 (1993).

33 Kumar, P. & Woolfson, D. N. Socket2: A Program for Locating, Visualising, and Analysing Coiled-coil Interfaces in Protein Structures. *Bioinformatics* (2021).

34 Swanson, C. J. & Sivaramakrishnan, S. Harnessing the unique structural properties of isolated alpha-helices. *J Biol Chem* **289**, 25460-25467 (2014).

35 Brown, R. A., Marcelli, T., De Poli, M., Sola, J. & Clayden, J. Induction of unexpected left-handed helicity by an N-terminal L-amino acid in an otherwise achiral peptide chain. *Angew Chem Int Ed Engl* **51**, 1395-1399 (2012).

36 Thomson, A. R. *et al.* Computational design of water-soluble alpha-helical barrels. *Science* **346**, 485-488 (2014).

37 Thomas, F. *et al.* De Novo-Designed alpha-Helical Barrels as Receptors for Small Molecules. *ACS Synth Biol* **7**, 1808-1816 (2018).

38 Enkhbayar, P., Hikichi, K., Osaki, M., Kretsinger, R. H. & Matsushima, N. 3(10)-helices in proteins are parahelices. *Proteins* **64**, 691-699 (2006).

39 Kumar, P. & Bansal, M. HELANAL-Plus: a web server for analysis of helix geometry in protein structures. *J Biomol Struct Dyn* **30**, 773-783 (2012).

40 Lupas, A. N. & Gruber, M. The structure of alpha-helical coiled coils. *Adv Protein Chem* **70**, 37-78 (2005).

41 Hunter, C. A. & Sanders, J. K. M. The nature of π-π interactions. *J Am Chem Soc* **112**, 5525-5534 (1990).

42 Mortenson, D. E. *et al.* High-resolution structures of a heterochiral coiled coil. *Proc Natl Acad Sci U S A* **112**, 13144-13149 (2015).

43 Fox, R. O., Jr. & Richards, F. M. A voltage-gated ion channel model inferred from the crystal structure of alamethicin at 1.5-A resolution. *Nature* **300**, 325-330, doi:10.1038/300325a0 (1982).

44 Bunkoczi, G., Schiell, M., Vertesy, L. & Sheldrick, G. M. Crystal structures of cephaibols. *J Pept Sci* **9**, 745-752, doi:10.1002/psc.496 (2003).

45 Mendel, D., Ellman, J. & Schultz, P. G. Protein biosynthesis with conformationally restricted amino acids. *J Am Chem Soc* **115**, 4359-4360 (1993).

46 Leonard, D. J., Ward, J. W. & Clayden, J. Asymmetric α-arylation of amino acids. *Nature* **562**, 105-109 (2018).

47 Collie, G. W. *et al.* Shaping quaternary assemblies of water-soluble non-peptide helical foldamers by sequence manipulation. *Nat Chem* **7**, 871-878, doi:10.1038/nchem.2353 (2015).

48 Wang, P. S. & Schepartz, A. beta-Peptide bundles: Design. Build. Analyze. Biosynthesize. *Chem Commun (Camb)* **52**, 7420-7432, doi:10.1039/c6cc01546h (2016).

49 Chandramouli, N. *et al.* Iterative design of a helically folded aromatic oligoamide sequence for the selective encapsulation of fructose. *Nat Chem* **7**, 334-341 (2015).

50 Girvin, Z. C., Andrews, M. K., Liu, X. & Gellman, S. H. Foldamer-templated catalysis of macrocycle formation. *Science* **366**, 1528-1531 (2019).

Methods

Composition of dataset

A precompiled list of 10087 PDB IDs with 25% sequence similarity or less, resolution better than 2.0 Å, and R-factor < 0.25 was obtained from Pisces server [51] and downloaded from PDB[11] on 15th August 2020 for further analyses. An additional dataset of 73 PDB files for all proteins with Aib residues as a part of polymer sequence was downloaded on 26th April 2021.

Programs used

In-house code was written either in Python or Perl. DSSP [52] was used to identify secondary structures in proteins and HELANAL-Plus [39] was used to analyse 3_{10} helices. Matplotlib [53] was used for plotting data. Protein-structure figures were generated using PyMol [54]. Figures were assembled in GIMP2. General information of identified α, 3_{10} and π helices in the 10087 protein structures database is provided in Supplementary Information Table S7.1.

Peptide synthesis and purification

Peptides were synthesised by Fmoc methods on a CEM Liberty Blue automated solid-phase peptide synthesis apparatus with inline UV monitoring. Activation was achieved using DIC/Oxyma [29]. All peptides were produced as the *C*-terminal amide on a Rink Amide MBHA solid support. *N*- and *C*-terminal capping were performed using 0.25 mL acetic anhydride and 0.3 mL DIPEA in 10 ml dimethylformamide (DMF) for 30 minutes on a rotator followed by washe with DMF and dichloromethane. To cleave peptides from the solid support, solutions of 25 mL trifluoroacetic acid (TFA), 0.4 mL triisopropylsilane, and 0.4 mL water was added to the resin and left on a rotator for 3 hours. These TFA solutions were reduced to 2 mL under a flow of nitrogen. Crude peptides were precipitated with diethyl ether (45 mL) at 4 °C. The solid was recovered by centrifugation and redissolved in 1:1 acetonitrile:water before freeze-drying to yield crude peptides as white or pale yellow solids.

Fmoc-protected amino acids, DMF, and activators were purchased from AGTC Bioproduct. All other solvents, Rink Amide ChemMatrix and Rink Amide MBHA solid support resin were purchased from Fisher Scientific, UK, PCAS Biomatrix, Canada, and Carbosynth Ltd, UK, respectively.

Analytical HPLC

Analytical HPLC traces were obtained using a Jasco 2000 series HPLC system using a Phenomenex 'Kinetex' 5 μm particle size, 100 Å pore size, C18 column of dimensions 100 ×

4.6 mm. Chromatograms were monitored at 220 and 280 nm and with a gradient of 20 to 80% acetonitrile in water (each containing 0.1% TFA) over 20 minutes.

MALDI-TOF Mass Spectrometry

MALDI-TOF mass spectra were collected on a Bruker UltraFlex MALDI-TOF mass spectrometer operating in negative or positive-ion linear mode. Peptides were spotted on a ground steel target plate using dihydroxybenzoic acid as the matrix. Masses quoted are for the monoisotopic mass as the singly protonated species. Calibration was achieved using the 'nearest neighbour' method, with Bruker Peptide Calibration Standard II as the reference masses.

Circular Dichroism Spectroscopy

A JASCO J-810 spectropolarimeter fitted with a Peltier temperature controller (Jasco UK) was used for collecting circular dichroism (CD) data. Peptide samples were made up as 5, 10, 25, 50, 100, 200, 300, 400, 500 and 1000 μM solutions in phosphate buffered saline (PBS, 8.2 mM sodium phosphate, 1.8 mM potassium phosphate, 137 mM sodium chloride, 2.7 mM potassium chloride at pH 7.4). CD spectra were recorded at 5 °C in quartz cuvettes of path lengths: 5 mm, for 5, 10 and 25 μM solutions; 1 mm for 50, 100 and 200 μM solutions; and 0.1 mm for 500 and 1000 μM solutions. CD spectra were recorded with a scan rate of 100 nm min^{-1}, 1 nm interval, 1 nm bandwidth and 1 s response time; and were an average of 8 scans recorded for the same sample. Single recordings of thermal denaturation curves with same settings and a ramping rate of 60 °C $hour^{-1}$ were acquired at 220 or 222 nm between 5 °C and 95 °C. Baselines recorded using the same buffer, cuvette and parameters were subtracted from each dataset. The spectra were converted from ellipticities (deg) to molar ellipticities (MRE (deg.cm^2.$dmol^{-1}$.res^{-1})) by normalising for the concentration of peptide bonds and the cell path length using formula:

$$\Theta\ (\text{deg. cm}^2\text{. dmol}^{-1}\text{.res}^{-1}) = \frac{\text{Ellipticity (mdeg)} * 10^6}{\text{Path length (mm)} * \text{[peptide] } \mu\text{M} * \text{n}}$$

Where, ellipticity is raw data obtained from the instrument and 'n' is the number of peptide bonds in the peptide. The *N*-terminal acetyl bond was included as a residue contributing to MRE but not the *C*-terminal amide.

The chemical denaturation of three peptides (PK-1, PK-4 and PK-5) at 5 μM concentrations was followed by recording CD spectra in the presence of 0 – 6M guanidinium chloride (GdmCl); and thermal denaturation curves for these samples were obtained at either 3 M (PK-1 and PK-4) or 4 M (PK-5) GdmCl concentration by monitoring the CD signal at 222 nm.

Analytical Ultracentrifugation

Analytical ultracentrifugation (AUC) sedimentation velocity (SV) experiments were conducted at 20 °C in a Beckman Optima XL-A analytical ultracentrifuge using an An-60 Ti rotor. Solutions of 310 µl volume were made up in PBS at 25, 50, 100 and 200 µM peptide concentration, and placed in a SV cell with aluminium centrepiece and sapphire windows. The reference channel was loaded with 320 µl of buffer. The samples were centrifuged at 60000 rpm, with absorbance scans taken across a radial range of 5.8 to 7.3 cm at 5 min intervals to a total of 120 scans. Data were fitted to a continuous c(s) distribution model using Sedfit [55], at 95% confidence level. The baseline, meniscus, frictional coefficient (f/f_0) and systematic time- invariant and radial-invariant noise were fitted. The peptide partial specific volume ($\bar{v}$), and the buffer density and viscosity were calculated using Sednterp [56]. However, Sednterp cannot recognize non-natural amino acids. Hence, for the calculations of $\bar{v}$, pairs of Aib residues or Aib+Gly pairs of a peptide sequence were replaced by Ala+Val or 2 x Ala, respectively. For example, PK-10 sequence was considered as GELAVLKELAVLKAVAWKA. Residuals for SV experiments are shown as a bitmap in which the greyscale shade indicates the difference between the fit and raw data. Scans are ordered vertically, with earlier scans at the top. The horizontal axis is the radial range over which the data were fitted.

AUC sedimentation equilibrium (SE) experiments were conducted at 20 °C in a Beckman Optima XL-I or XL-A analytical ultracentrifuge using an An-50 Ti or An-60 Ti rotor (Beckman-Coulter) respectively. Solutions were made up in PBS at 25, 50, 100 and 200 µM peptide concentration, and the pH were adjusted to 7.4 using 0.1 M NaOH. The samples were centrifuged at speeds in the range 15–36 krpm or 44–60 krpm. Data were fitted to single, ideal species models using Sedphat [57]. 95% confidence limits were achieved via Monte Carlo analyses of the obtained fits.

1,6-diphenylhexatriene (DPH)-binding assay

An epMotion 5070 liquid handler (Eppendorf) was used to generate samples for 1,6-diphenylhexatriene (DPH) binding experiments. The total concentration of the 3_{10} assembly was varied between 0 and 50 µM, while the concentration of DPH was kept constant at 0.1 µM, which was introduced in DMSO to a final concentration of 5% v/v. Data were collected on a Clariostar plate reader (BMG Labtech) using an excitation wavelength of 350 nm and emission monitored at 450 nm. Binding constants were extracted by fitting Equation 1 to the data in SigmaPlot13.0.

$$y = B_{max}\frac{(c + x + K_D) - \sqrt{(c + x + K_D)^2 - 4cx}}{2c}$$

Where c is the total concentration of the constant component (e.g. DPH), x is the concentration of 3_{10} helix assembly, Bmax is the fluorescence signal when all of the constant component is bound, and y is the fraction of bound component being monitored via fluorescence signal.

X-ray crystal structure determination

Diffraction-quality peptide crystals were grown using a sitting-drop vapour-diffusion method. Freeze-dried PK-1 (CC-TypeN-L_aL_d) was resuspended in ultrapure water to ≈10 mg/ml. Since all attempts to crystallize the peptide PK-10 were unsuccessful, racemic crystallography was tried. Equal concentration of PK-10 and PK-11 were mixed and resuspended in ultrapure water to approximate concentration of 10 mg/ml. Numerous attempts to crystallize PK-4 (3_{10}HD) were unsuccessful. Multiple datasets were collected for a racemic mixture, but these could not be solved. Lastly, PK-5 (D-3_{10}HD) was resuspended in 25 μM sodium acetate to ≈4 mg/ml, which produced spherulites that were used as seed to get crystals in various conditions.

Commercially available sparse matrix screens from Molecular Dimensions (Morpheus, JCSG plus™, Structure Screen 1+2, Pact Premier™, ProPlex) were used, and the drops were dispensed using a robot (Oryx8, Douglas Instruments). For each well of an MRC2 drop plate, 0.3 μl of peptide solution and 0.3 μl of reservoir solution were mixed and the plate was incubated at 4 or 20 °C. To improve the crystal quality for PK-10 and D-3_{10}HD, microseeding was performed. Final crystallisation conditions for all peptides are provided in Supplementary Information Table S7.2.

Except for D-3_{10}HD, crystals were observed within two weeks. Crystallization of D-3_{10}HD produced spherulites in B1 condition of JCSG plus (0.2M sodium thiocyanate, 20% w/v PEG3350) that was further used for microseeding to get crystals in multiple conditions. While looping, crystals were soaked in the mixture of their respective reservoir solutions and cryoprotectants such 25% glycerol, ethylene glycol and PEG500MME. X-ray diffraction data were collected at the Diamond Light Source (Didcot, UK) on beamlines I03, I04-1 and I24. For specific wavelengths see Supplementary Information Table S7.3.

Diffraction images were processed using either an automated pipeline (xia2 [58]) or manually (Imosflm [59]). For the latter, data were reduced using POINTLESS [60], AIMLESS [61] and CTRUNCATE [62]. Diffraction images of the racemic mixture of PK-10 and PK-11 were processed using XDS [63]. Shelxt [64] in Olex2 [65] was used to get the initial structure. The PK-1 (CC-TypeN-L_aL_d) structure was determined using ARCIMBOLDO Lite [66] in the coiled-coil mode [67], using a single ideal polyalanine helix as a search model. Final structures were obtained after iterative rounds of model building with COOT [68] and refinement with Phenix.refine [69] or REFMAC5 [70] available in PHENIX [71] and CCP4 [72], respectively. Late-stage models of all structures were submitted to PDB_REDO [73] and further refined with REFMAC5 [70]. Solvent-

exposed atoms lacking map density were either deleted or left at full occupancy. Data collection and refinement statistics are provided in Supplementary Information Table S7.3.

Data availability

The coordinate and structure factor files are available from the RCSB-PDB under the accession code: CCTri-TypeN-LaLd (PDB ID: 7QDK; https://doi.org/10.2210/pdb7QDK/pdb); D-310HD (PDB ID: 7QDI; https://doi.org/10.2210/pdb7QDI/pdb); PK-10+PK-11 (PDB ID: 7QDJ; https://doi.org/10.2210/pdb7QDJ/pdb). The list of PDB files for the bioinformatic analyses was downloaded from Pisces server (http://dunbrack.fccc.edu/pisces/download/). Data for generating figures in the main text are provided with this paper. Additional data to generate figures in supporting information is available at http://coiledcoils.chm.bris.ac.uk/SI-data/PK-310/.

Code availability

The customized scripts used for bioinformatic analyses are available at http://coiledcoils.chm.bris.ac.uk/SI-data/PK-310/.

References

51 Wang, G. & Dunbrack, R. L., Jr. PISCES: a protein sequence culling server. *Bioinformatics* **19**, 1589-1591 (2003).

52 Joosten, R. P. *et al.* A series of PDB related databases for everyday needs. *Nucleic Acids Res* **39**, D411-419 (2011).

53 Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput Sci Eng* **9**, 90-95 (2007).

54 Schrödinger, L. L. C. The PyMOL Molecular Graphics System Open-Source, Version 2.4.0. (2021).

55 Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* **78**, 1606-1619 (2000).

56 Laue, T., Shah, B., Ridgeway, T. & Pelletier, S. in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (ed SE Harding, Rowe, AJ, Horton, JC) (Royal Society of Chemistry, Cambridge, 1992).

57 Zhao, H., Piszczek, G. & Schuck, P. SEDPHAT--a platform for global ITC analysis and global multi-method analysis of molecular interactions. *Methods* **76**, 137-148 (2015).

58 Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *Journal of applied crystallography* **43**, 186-190 (2010).

59 Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* **67**, 271-281 (2011).

60 Evans, P. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**, 72-82 (2006).

61 Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204-1214 (2013).

62 Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr* **67**, 282-292 (2011).
63 Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132 (2010).
64 Sheldrick, G. M. SHELXT - integrated space-group and crystal-structure determination. *Acta Crystallogr A Found Adv* **71**, 3-8 (2015).
65 Dolomanov, O. V., Bourhis, L. J., Gildea, R. J., Howard, J. A. K. & Puschmann, H. OLEX2: a complete structure solution, refinement and analysis program. *Journal of Applied Crystallography* **42**, 339-341 (2009).
66 Sammito, M. *et al.* ARCIMBOLDO_LITE: single-workstation implementation and use. *Acta Crystallogr D Biol Crystallogr* **71**, 1921-1930 (2015).
67 Caballero, I. *et al.* ARCIMBOLDO on coiled coils. *Acta Crystallogr D Biol Crystallogr* **74**, 194-204 (2018).
68 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
69 Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-367 (2012).
70 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-367 (2011).
71 Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Biol Crystallogr* **75**, 861-877 (2019).
72 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242 (2011).
73 Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1**, 213-220 (2014).

Acknowledgments

PK and DNW are supported by Biotechnology and Biological Sciences Research Council (BBSRC; BB/R00661X/1) and European Research Council (340764) grants to DNW. DNW is also supported by BrisSynBio, a BBSRC/EPSRC-funded Synthetic Biology Research Centre (BB/L01386X/1), and a Royal Society Wolfson Research Merit Award (WM140008). JC is supported by ERC Advanced Grant DOGMATRON (Agreement no. 884786), and by an EPSRC Programme Grant (EP/P027067/1). 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), and BrisSynBio for access to peptide synthesizers. We thank Dr Christopher Williams for collecting 1D-1H NMR spectra. We thank Diamond Light Source for access to beamlines I03, I04, I04-1 and I24 (Proposal 23269) and Dr Mark Warren from I19 who helped NGP with the direct methods solution. The authors thank Prof Todd Yeates (UCLA), and Drs Kapil Gupta, Christine Tölzer, and Frank Zieleniewski, and members of the Clayden and Woolfson labs and BrisSynBio for helpful discussions.

Author Contributions

PK, JC and DNW conceived the project. PK and DNW designed the bioinformatics analyses, which were performed by PK. PK and DNW designed the sequences, which were synthesized, characterised, and crystallised by PK. PK and NGP solved the X-ray crystal structures. PK, JC and DNW wrote the manuscript, which was read by all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information: The online version contains supplementary material. Correspondence and requests for additional data/ scripts should be addressed to Prasun Kumar and Derek N Woolfson.