

Cite this: *Chem. Sci.*, 2025, 16, 18632

All publication charges for this article have been paid for by the Royal Society of Chemistry

De novo designed 3-helix bundle peptides and proteins with controlled topology and stability

Xiyue Leng,^{ab} Katherine I. Albanese,^{abc} Lia R. Golub,^a Arthur A. Norman,^d Jonathan Clayden^{ab} and Derek N. Woolfson^{ab*}

Computational protein design is advancing rapidly. However, approaches and methods are needed to increase success rates and to elaborate designs. Here we describe the combination of rational and computational design to deliver three-helix bundle (3HB) peptide assemblies and single-chain proteins with control over topology and thermal stability. First, we garner sequence-to-structure relationships from antiparallel 3HBs in the Protein Data Bank. This gives core-packing rules, including layers of hydrogen-bonded polar residues, which are combined with surface-charge patterning to design complementary sequences for acidic (A), basic (B), and neutral (N) helices. By altering the design of the N helix, two sets of synthetic peptides are generated for clockwise and anticlockwise arrangements of the three-helix assemblies. Solution-phase characterisation shows that both ABN peptide mixtures form stable, heterotrimeric assemblies consistent with the targeted 'up-down-up' topologies. Next, AlphaFold2 models for both designs are used to seed computational designs of single-chain proteins where the helices are connected by loop building. Synthetic genes for these express in *E. coli* to yield soluble, monomeric, and thermally stable proteins. By systematically introducing additional polar layers within the core, the thermal stability of these proteins is varied without compromising the specificity of the helix-helix interactions. Chemical and thermal denaturation reveals comparable thermodynamic parameters to those of highly stable natural proteins. Four X-ray crystal structures confirm that the design models and AlphaFold2 predictions match to sub-Å accuracy.

Received 25th July 2025
Accepted 23rd August 2025

DOI: 10.1039/d5sc05576h

rsc.li/chemical-science

Introduction

Protein design is advancing rapidly, increasingly using computational and artificial intelligence (AI) methods.^{1–3} Despite recent advances, a deeper understanding of sequence-to-structure/function relationships remains essential to deliver designs that are more predictable, yield higher experimental success rates, and are generally more fit for purpose.

To reach this point, *de novo* protein design has undergone several phases, which we can learn from.^{4–7} Historically, minimal protein design employed fundamental chemical principles—such as sequence patterns of hydrophobic and polar residues—to render mimics of simple, natural protein folds. Aided by developments in bioinformatics, rational design emerged, integrating analyses of natural protein sequence and structural databases, such as the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB),^{8,9} into

de novo design. This has led to improved sequence-to-structure relationships to inform better designs.⁴ As the field has matured, computational methods have increasingly been used in design pipelines, including parametric design to sketch out backbones, and physical forcefields to assess sequences that best fit these.^{10,11} Now, AI-based models built from large sequence and structural datasets that capture relationships *en masse* are being used to generate backbones and sequences either separately or simultaneously.^{1,3,6,12,13}

However, while AI-based methods can exploit both known and unknown relationships, they often lack interpretability, making it difficult to uncover how these relationships govern design outcomes. Despite recent success in delivering complex *de novo* protein scaffolds and functions,^{14–17} the explainability and success rates of AI-driven methods remain low. Therefore, to continue advancing the field, these gaps in understanding and applicability need to be filled. This does not necessarily mean replacing 'black-box' methods. Rather, they need to be augmented to add reasoning and insight to enhance their predictive power, efficiency, and robustness. In short, we would like to achieve fully programmable protein design based on understanding sequence-to-structure/functional relationships in proteins. *En route* to this, we advocate combining rational

^aSchool of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK.
E-mail: D.N.Woolfson@bristol.ac.uk

^bMax Plack-Bristol Centre for Minimal Biology, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK

^cDepartment of Chemistry, Wake Forest University, Winston-Salem, NC, USA

^dSchool of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol BS8 1TD, UK



and computation protein design for targets when possible leading to more explainable AI for *de novo* protein design.

α -Helical coiled coils (CCs) have been particularly fruitful targets for *de novo* protein design due to many well-established principles that link primary sequences to a variety tertiary and quaternary structures.^{4,18,19} As such, they are well suited to the challenge of combining rational, computational, and AI-based approaches in protein design. Generally, CC structures have two or more right-handed α helices supercoiled around one another to form usually left-handed helical bundles. These are encoded by 7-residue (heptad) sequence repeats denoted $(abcdefg)_n$. The *a* and *d* sites are predominately occupied by hydrophobic residues, giving the pattern $(hphpppp)_n$. When configured into an α helix, this produces a hydrophobic seam that drives helix association, and sets a framework for specifying helix orientation (parallel or antiparallel), oligomeric states (dimer and above), partner preferences (homo- or heteromeric), and overall topology and handedness of the tertiary or quaternary structure as described below.^{18,20} In addition, the introduction of electrostatic or polar interactions flanking the hydrophobic core—usually at the *e* and *g* sites for dimers to tetramers, or *b* and *c* sites for pentamers and above—can be used to define CC assemblies further.^{21,22} Combined, established combinations of amino acids at the *g-a-d-e* positions can be used to generate toolkits of *de novo* CC assemblies.^{7,23,24} In turn, these peptide assemblies can be repurposed to design functional peptides for catalysis,^{25,26} materials assembly,^{27–30} and in cell and synthetic biology.^{31,32} Most recently, some of the peptide sequences and experimental 3D structures have been used as “seeds” to deliver more-complex, single-chain proteins through computational protein design leveraging the new AI-based methods.^{33,34}

Here, we focus on CC assemblies of three α helices, termed 3-helix bundle (3HB) CCs. These are appealing due to their relative simplicity and small size, making them ideal for studying structure, folding, and stability, and, therefore, for developing further fundamental design rules and principles^{35–37} and new applications.^{38–41} The homotrimeric peptide coil-Ser provides an early example of an ‘up-up-down’ 3HB architecture—*i.e.*, with one helix aligned antiparallel to the other two—discovered serendipitously in an attempt to design a parallel dimeric CC.⁴² Notable progress has also been made towards desymmetrizing 3HBs using attractive ion pairs at the interfacial positions, effectively guiding the specific formation of heterotrimeric assembly over alternative oligomeric states or homo- or mixed-trimers.^{41,43–45} DeGrado and colleagues have advanced this by developing an iterative design transitioning coil-Ser-derived sequences to native-like globular proteins.³⁵ Recent innovations include using buried hydrogen bonds and shape-complementary packing to create highly specific heterotrimers that serve as biological scaffolds.⁴⁶

We build on this foundation here to describe the design of a series of 3HB CC peptide assemblies and single-chain proteins with up-down-up arrangements of three helices but with different and complementary sequences; that is, an ABN system composed of acidic (A), basic (B), and neutral (N) strands. By strategically positioning interhelical charges, we can dictate the

overall conformational handedness of the 3HBs, achieving either clockwise (CW) or anticlockwise (ACW) topologies in solution. Models for each of these peptide assemblies are then used to seed the computational design of single-chain 3HB proteins, which are expressed from synthetic genes in *E. coli*, and confirmed by X-ray structures with the intended topologies and handedness of the tertiary structures. By introducing polar residues into the hydrophobic core, we fine-tune thermal stability without compromising folding specificity. This study demonstrates the feasibility of designing 3HB CCs with precise control of topology and thermal stability by combining rational, computational, and AI-based *de novo* design, and shows how design success rates can be improved through such pipelines.

Results and discussion

Rational design of up-down-up heterotrimeric α -helical peptide assemblies

To garner sequence-to-structure relationships for the target, we searched the CC+ database of structurally validated CC proteins for examples of up-down-up 3HBs (Tables S1 and S2).⁴⁷ This returned 169 examples of antiparallel heterotrimeric peptide assemblies and 3-helix protein bundles with up-down-up connectivity. This was from a total of 285 3-helix CCs in the database. Manual assignment of the subset of 169 structures revealed a strong natural bias: 121 had the anticlockwise topology, and 48 had the clockwise arrangement. Notably, in single-chain proteins, the predominant anticlockwise topology has right-handed helical connectivity consistent with the left-handed supercoiling of CCs. Given the underrepresentation of clockwise topologies in nature and the associated design challenge, we initially prioritised this configuration. The helical sequences of the 169 subset were used to compile a 20×7 amino-acid profile from their component *a-g* heptad repeats (Fig. 1A and B). The SWISS-PROT normalised data revealed preferences for certain amino acids at the different heptad positions, which were used as rules to design peptide sequences.

The core-defining *a* and *d* positions were made Leu, as this was the most preferred residue at both sites. The core-flanking *e* and *g* positions were made combinations of Glu and Lys, as they featured highly at these sites and offered possibilities for directing helix-helix partnering through electrostatic interactions.^{18,48} Building on previous work^{43,44} to design away from homomeric assembly and to favour heteromeric association, we designed one sequence to be acidic (A) with *e* = *g* = Glu, and a second basic (B) peptide with *e* = *g* = Lys. Typically, such AB systems are designed for even oligomers as the alternating charges can be satisfied by C2, C4 or D2 symmetry.^{31,49–51} Therefore, to target an ABX-type heterotrimer, the third helix was made neutral (N) with *e* = Glu plus *g* = Lys to give complementary interactions to both flanking acidic and basic helices (Fig. 1C). As noted previously by DeGrado and co-workers,³⁵ helical wheels suggests an alternative design for the assembly with the opposite (anticlockwise) cyclic order of the three helices, which can be achieved by switching the polarity *e* = Lys and *g* = Glu giving an alternative neutral peptide (N')





Fig. 1 The rational design of new sequences to form antiparallel CC heterotrimers. (A) Amino-acid propensities for each position of the heptad repeats for 3HB CCs pulled from the CC+ database.⁴⁷ Raw counts (Table S1) were normalised using the amino-acid frequencies in SWISS-PROT to provide the propensity scale shown as a heat map (high to low: red to blue). (B) Helical-wheel diagram for a clockwise antiparallel trimeric assembly, showing heptad assignments and with the helical termini closest to the viewer labelled. (C and D) Helical-wheel representations of apCCTri-BAN (clockwise) and apCCTri-BAN' (anticlockwise), respectively. Sequences have canonical heptad repeats, *abcdefg*. The residues at key positions of the designs, *g*, *a*, *d*, and *e*, are highlighted: grey for hydrophobic, Leu; green for polar, Thr and Asn; red for acidic, Glu; and blue for basic, Lys. (E and F) Slices through the third heptad of the AF2 models for apCCTri-BAN and apCCTri-BAN', respectively, with the key side chains shown as sticks. For B–D, arrows indicate the overall handedness of the quaternary structure. This is defined as follows: with the first helix at 12 o'clock and coming out towards the viewer, the assemblies have either a clockwise (B– $\bar{A}$ -N) or an anticlockwise (B– $\bar{A}$ -N') arrangement.

(Fig. 1D). To help specify the antiparallel arrangement further, specifically with the A helix antiparallel (denoted $\bar{A}$) to the B and N helices, we incorporated a layer of polar residues at *d*-*a* sites of the $\bar{A}$ -B-N/N' combinations in the otherwise hydrophobic cores (Fig. 1E and F).^{22,52,53} This was chosen by inspecting the CC+ dataset derived above for buried, three-residue constellations of polar side chains. This revealed STAT5a (1y1u), which has an ordered, hydrogen-bonded network between Thr-155 (*a*), Thr-236 (*d*) and Asn-289 (*a*).⁵⁴ This polar layer was compatible with both the clockwise (CW, BAN) and anticlockwise (ACW, BAN') target assemblies. The designs were completed as 4-heptad sequences with *b* = *c* = Ala to provide

specificity for the assembly, and combinations of polar and aromatic residues at the *f* sites to enhance solubility and to introduce chromophores (Table 1 and S3).

Ahead of experimental work, all combinations of the A, B, and N and A, B, and N' sequences were modelled using AlphaFold2-multimer (AF2, Table S4). The predicted models were consistent with the target heterotrimeric assemblies with the designed helical topologies, *i.e.*, mixed parallel and antiparallel helices and clockwise or anticlockwise peptide assembly. The target and alternate state models (AAA, AAB, ABB, *etc.*) were assessed by predicted template modelling (pTM) and local distance difference test (pLDDT) scores. However, the average pLDDT scores were all above 95%, so, effectively, AF2 did not discriminate between the models (Table S4).

Table 1 Designed sequences of apCCTri-A, B, N and N' peptides

systematic names	Amino acid sequence			
	<i>gabcdef</i>	<i>gabcdef</i>	<i>gabcdef</i>	<i>gabcdef</i>
apCCTri-A	G -LAALEK	ELAATEK	ELAALEW	ELAAL-- G
apCCTri-B	G -LAALKE	KLAALKE	KNAALKY	KLAALK-- G
apCCTri-N	G KLAALEK	KLAALEK	KTAALEY	KLAAL-- G
apCCTri-N'	G ELAALKE	ELAALKE	ETAALKY	ELAAL-- G

Biophysical characterisation of synthetic peptides confirms the design targets

All four peptides A, B, N, and N' were synthesised by Fmoc solid-phase peptide synthesis, purified by reverse-phase HPLC, and verified by MALDI-TOF mass spectrometry (Table S3 and Fig. S1). The peptides were characterised in solution individually and as 1 : 1 : 1 mixtures as described below.

First, circular dichroism (CD) spectroscopy was used to probe the folding and thermal stability of the peptides at 100 μM concentration in phosphate buffered saline at pH 7.4 (PBS; Fig. 2A). In contrast to AF2 predictions, peptide A was completely unfolded under these conditions. Peptide B was

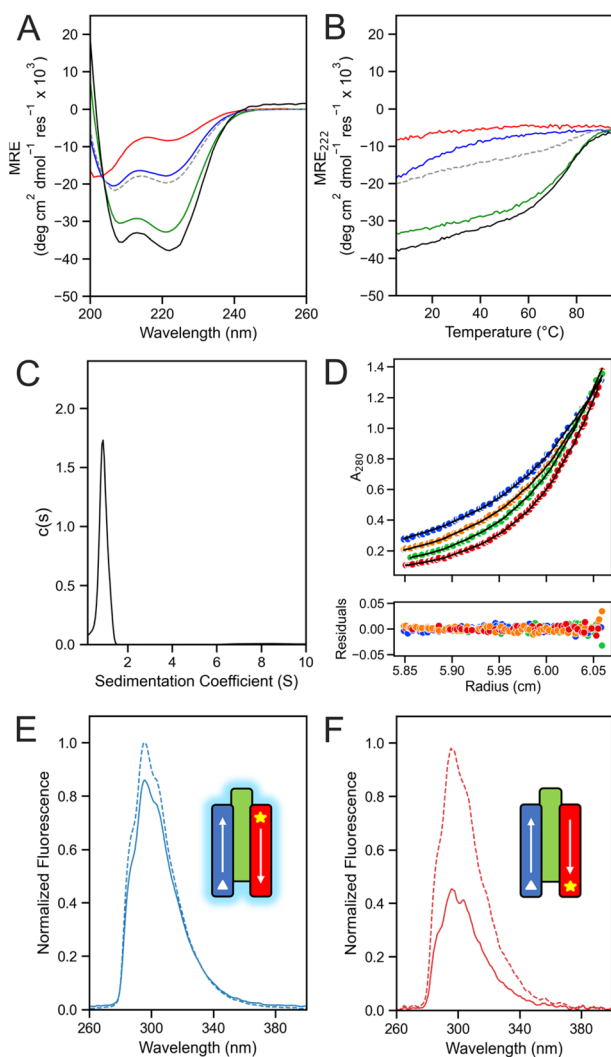


Fig. 2 Biophysical characterisation of the apCCTri A, B and N peptides. (A) CD spectra at 5 °C, and (B) thermal responses of the CD signals at 222 nm (ramping from 5 to 95 °C at 1 °C min⁻¹) of the individual peptides (A, red; B, blue; and N, green), the equimolar 1 : 1 : 1 mixture (black), and the calculated theoretical averages (dashed). Conditions: 100 μM in 1 $\times$ PBS, pH 7.4. (C) AUC-SV and (D) AUC-SE profiles of apCCTri-BAN 1 : 1 : 1 mixture. Fits returned molecular weights consistent with a trimeric assembly (1.1 and 0.9 $\times$ the summed masses, respectively). Conditions: 250 μM in 1 $\times$ PBS, pH 7.4; SV run at 60 krpm, SE run at 40–52 krpm. Key: 40 krpm, blue; 44 krpm, orange; 48 krpm, green; and 52 krpm, red. (E and F) Fluorescence-quenching assays for mixtures of apCCTri-B-nMSE and apCCTri-N plus apCCTri-A-n4CF (E) or apCCTri-A-c4CF (F). Dotted lines are for the fluorescently-labelled A peptides alone, and solid lines are for the 1 : 1 : 1 mixture. In the cartoons, arrows indicate the peptide direction from N to C termini; 4CF is represented by the yellow star and MSE by the white triangle. Conditions: 33 μM of each peptide in phosphate buffer (8.2 mM potassium phosphate dibasic, 1.8 mM potassium phosphate monobasic), pH 7.4.

partially folded at 5 °C but unfolded readily with increased temperature (midpoint of thermal denaturation, $T_M < 15$ °C). These behaviours are consistent with the design hypothesis, as highly charged peptides are not expected to self-associate appreciably. By contrast, the neutral, peptides N and N' were highly helical and thermally stable ($T_M \approx 78$ and 54 °C, respectively). The difference in the T_M values can be explained by the order of Glu and Lys residues in the sequence.²¹ However, both mixtures BAN and BAN' were also highly helical and highly thermally stable ($T_M \approx 73$ and 67 °C, respectively, Fig. S2). For completeness, data for the pairwise combinations were collected and compared with the respective theoretical averages, showing reduced helicity and cooperative folding (Fig. S2 and Table S6). The solution-phase oligomeric states for the stable complexes, N₃, N'₃, BAN and BAN', were determined by sedimentation-velocity (SV) and sedimentation-equilibrium (SE) experiments in analytical ultracentrifugation (AUC) at 20 °C (Fig. 2C and D, Table S6, Fig. S3 and S4). These revealed monodisperse trimeric species as designed.

Despite multiple attempts, we could not crystallise any of the species formed. Therefore, we used a fluorescence-quenching assay to probe helix orientation in the BAN and BAN' systems in solution.⁵⁵ First, guided by the AF2 models, we placed spatially proximal fluorophore (4-cyanophenylalanine, 4CF) and quencher (L-selenomethionine, MSE) pairs on two different peptides. For instance, the B peptide was remade with an N-terminal MSE, apCCTri-B-nMSE; and two variants of the A peptide were made with 4CF at the N-terminal g site or the C-terminal e site, apCCTri-A-n4CF and apCCTri-A-c4CF, respectively. As expected for an antiparallel arrangement of A and B helices, no quenching of 4CF fluorescence was seen when apCCTri-B-nMSE and apCCTri-A-n4CF were mixed with the unlabelled N peptide (apCCTri-N), Fig. 2E. However, quenching did occur with the other mixture with apCCTri-A-c4CF, Fig. 2F. This demonstrated that the A helix was oriented antiparallel to B in solution as designed. To probe the orientation of the N helix, we carried out the analogous experiments using the two 4CF-labelled A peptides, unlabelled B, and with the N peptide N-terminally labelled with MSE, apCCTri-N-nMSE. This gave similar results to those shown in Fig. 2E and F indicating that the N helix also aligns antiparallel to A and, therefore, parallel to B (Fig. S5).

Following this solution-phase characterisation, and consistent with our systematic naming of *de novo* CC peptide,¹⁸ we named the two heterotrimeric peptide assemblies apCCTri-BAN and apCCTri-BAN'.

Peptide assemblies can be used as seeds for single-chain protein designs

Recently, we introduced the concept of rationally seeded design of single-chain proteins from characterised peptide assemblies.^{33,34} This takes computational or experimental structural models of the peptide assemblies and uses computational design to link the C and N termini of neighbouring helices. To do this here, we used MASTER to search the RCSB PDB for loops to connect helices in the AF2 models for the clockwise, apCCTri-



BÄN, and anticlockwise, apCCTri-BÄN', topologies.^{56,57} Models for these were built using AF2 in single-sequence mode, and assessed by pLDDT and root mean square deviation (RMSD) from the peptide models (Tables S4 and S5). Next, ProteinMPNN was used to optimise the loop regions of the single-chain models while keeping the helical seeds completely intact. Based on pLDDT and RMSD, two BÄN and one BÄN' single-chain designs were chosen to test experimentally.

Synthetic genes for the sequences were expressed in *E. coli* (Table S3). As the parent peptide assemblies were thermally stable, the cell lysate was heat-shocked at 65 °C for 10 min. Immobilised metal affinity column chromatography (IMAC) and size exclusion chromatography (SEC) were used to yield purified proteins (Fig. S6). CD spectroscopy of all proteins showed that they were highly α helical and hyper-thermally

stable, and AUC experiments confirmed they were monomeric in solution (Fig. 3A–E).

We solved X-ray crystal structures at 2.10 Å, 2.15 Å and 2.20 Å resolution for the two BÄN- and one BÄN'-based designs, respectively (Fig. 3F–I, Tables S7 and S8). This was done by molecular replacement using the AF2 models of the peptide assemblies as the search models. All confirmed the three-helix bundles with up-down-up topologies for the B-Ä-N/N' helices. All had consolidated hydrophobic cores with knobs-into-holes packing between the core Leu residues confirmed by Socket2 (Fig. 3H–I).⁵⁸ In addition, water-mediated hydrogen-bonded layer of Thr (A) – Asn (B) – Thr (N/N') was present in a clockwise design (Fig. 3I). Moreover, the conformational handedness of the two structures was different and as designed. Therefore, we named these proteins sc-apCC3-CW (for the clockwise BÄN combination, 9rgv and 9rgw) and sc-apCC3-ACW (for the



Fig. 3 Rationally seeded computational designs of the single-chain 3HB proteins based on models for the apCCTri-BÄN peptide assemblies. Two different designs were made and characterised, sc-apCC3-CW (data and images in magenta) and sc-apCC3-ACW (blue). (A) CD spectra at 5 °C and (B) thermal responses of the CD signals at 222 nm (ramping from 5 to 95 °C at 1 °C min^{−1}). Conditions: 10 μ M protein, PBS, pH 7.4. Data for the apCCTri-BÄN peptide assembly are included for comparison (black). (C) AUC-SV and (D&E) AUC-SE data. Fits returned molecular weights of 0.91 and 0.84 $\times$ monomer mass for sc-apCC3-CW, and 0.87 and 0.77 $\times$ monomer mass for sc-apCC3-ACW, respectively. Conditions: 50 μ M protein, PBS, pH 7.4; SV run at 60 krpm, and SE run at 44–60 krpm. (F and G top) Top views of the X-ray crystal structures of sc-apCC3-CW1 (G; 2.10 Å, 9rgv), and sc-apCC3-ACW (F; 2.20 Å, 9rgx) aligned with AF2 model (rank 1, grey) to C α -RMSD of 1.079 and 1.742 Å, respectively. Arrows indicate the overall handedness of the tertiary structure. This is defined as follows: with the N-terminal helix at 12 o'clock and coming out towards the viewer, the chain trace follows either a clockwise or anticlockwise path. (F and G bottom) Side views of the same structures with each helix identified by electron densities of the loop region and the Trp and Tyr side chains, which are contour level (σ) = 1, 2mFo-DFc maps displayed to within 1.3 Å to the region of interest. (H) X-ray crystal structure of sc-apCC3-CW2 (2.15 Å, 9rgw) with CC regions identified by Socket2 coloured as chainbows, KIH packing cutoff is 7 Å. (I) Slices through the structure at the third heptad repeat with the water-mediated polar layer (top), and the second heptad showing KIH packing of the consolidated Leu core, with a knobs in red and d in green (bottom).



apCC3-CW-3NTT, and sc-apCC3-CW-4NTT, respectively (Fig. 4A). This progression ensures that the central regions of the protein are destabilised first and remodelled with polar interactions, while maintaining fully hydrophobic and stabilising heptad repeats at the termini. As a control, we included a variant without a polar layer, namely sc-apCC3-CW-0NTT. All sequences were evaluated using AF2, which consistently predicted clockwise 3HB topologies with high confidence (Table S4). The variants with 0 and 2 N-T-T layers crystallised, and the X-ray crystal structures aligned with the parent design, with C α RMSDs of 1.015 Å and 0.605 Å, respectively (Tables S7 and S8; pdb ids 9rgy, 9rgz). The variants with 3 and 4 N-T-T layers could not be crystallised.

These observations were supported by the CD spectroscopy. CD spectra showed a progressive decrease in helicity with the addition of N-T-T layers (Fig. 4C and S7, Table S6). The most extreme case was sc-apCC3-CW-4NTT, which retained only ~50% helicity relative to the 1NTT and 0NTT variant. Variable-temperature CD measurements revealed that sc-apCC3-CW with 0, 1, and 2 N-T-T layers were hyperthermally stable and did not unfold upon heating (Fig. 4D). In contrast, sc-apCC3-CW-3NTT and -4NTT had measurable T_M values of 86.5 ± 0.8 and 25.3 ± 0.7 °C, respectively (Fig. 4D and S7). Whilst sc-apCC3-CW-3NTT showed good reversibility from the pre- and post-melt CD spectra, sc-apCC3-CW-4NTT remained unfolded after heating and was disregarded from further analysis (Fig. S7).

We extended the AF2 predictions of the 3HBs to these 0–4 NTT layer variants (Table S4). Interestingly, there was little difference between the predicted models and confidence metrics for any of these constructs. Specifically, secondary structure calculations from the models all gave $\approx 74\%$ helical residues. Thus, apart from the 0NTT construct, which coincidentally had 75% helicity by CD spectroscopy, there was no correlation between the predicted and observed helicities (Table S9).

Returning to sc-apCC3-CW-3NTT, this is an example of a *de novo* designed single-chain protein with an accessible, reversible, and cooperative thermal-unfolding transition. This is unusual in contemporary designed single-chain proteins, which are often hyper-stable. Therefore, we characterised the thermodynamics of its unfolding in more detail.

The CD spectra and T_M values for sc-apCC3-CW-3NTT did not change with protein concentration (Fig. S12), confirming that the protein folds as a non-associating monomer as designed. Given its appreciable thermal stability, we used guanidinium hydrochloride (GdmHCl) as a chemical denaturant to access melting transitions at lower T_M values for a full thermodynamic analysis (Fig. S13–S15). This gave well-defined, sigmoidal, thermal-unfolding curves, which were fitted to a two-state model to estimate T_M over a range of GdmHCl concentrations (Fig. S13–S15 and Table S10).⁶³

The resulting T_M values were linearly related to the GdmHCl concentration (Fig. 4D). Encouraged by this, we sought to estimate the thermodynamics of unfolding at 0 M GdmHCl using the linear extrapolation method (LEM). Traditional van't Hoff analysis assumes $\Delta C_p = 0$. However, this neglects the substantial structural reorganisation and exposure to solvent associated

with protein unfolding, both of which contribute to a positive $\Delta C_{p,unf}$.^{5,64} Therefore, we used a global nonlinear Gibbs–Helmholtz fitting procedure to model the free energy of unfolding (ΔG_{unf}) as a function of temperature and GdmHCl concentration under the LEM (Table S10).^{65,66} This gave ΔG_{unf} of 6.16 ± 0.07 kcal mol^{−1} at 25 °C in the absence of denaturant (Fig. S15). This is consistent with the observed stability of the protein under those conditions. Moreover, when expressed per residue, it is comparable to the most stable natural globular proteins under similar conditions (Fig. 4E and S16),^{67–69} most of which are not all- α -helical proteins. The modest change in enthalpy ($\Delta H_{unf} = 24.08 \pm 0.53$ kcal mol^{−1}) and entropy ($\Delta S_{unf} = 0.06 \pm < 0.01$ kcal mol^{−1} K^{−1}) reflect the small size of the 3HB. The small heat capacity change ($\Delta C_p = 0.09 \pm 0.01$ kcal mol^{−1} K^{−1}) indicates opposing hydration effects upon burying polar (negative ΔC_p) and nonpolar side chains (positive ΔC_p), which partly cancel.^{70–72} Nonetheless, the net value, along with its dependence on the denaturant ($\Delta \Delta C_{p,[GdmHCl]} = 0.46 \pm 0.01$ kcal mol^{−1} K^{−1} M^{−1}), are consistent with a compact native state and the progressive solvation of the largely hydrophobic core upon unfolding.

Conclusions

In summary, we have developed sequence-to-structure relationships for the rational design of simple up-down-up, heterotrimeric, coiled-coil peptide assemblies. These relationships are informed by modelling and confirmed in solution using a variety of biophysical measurements. Two distinct topologies of the design target are generated by exploiting electrostatic interactions to specify the helix–helix interfaces. Next, AlphaFold2 models for the two assemblies are used to seed the computational design of single-chain three-helix-bundle proteins with clockwise and anticlockwise arrangements of the component helices. Synthetic genes for these designs express well in *E. coli* to give hyper-stable, water-soluble, monomeric proteins. High-resolution X-ray crystal structures for the two designs match the computational models with atomic accuracy, confirming the design rules and pipeline used. The structures reveal that rational design can precisely specify conformational handedness, reliably accessing the more-common anticlockwise configuration seen in natural helical bundles, but also the less-prevalent clockwise form. This ability to control helix–helix interfaces firmly establishes another level of programmability in CC design.³⁵

Miniprotein designs of the type delivered here provide relatively straightforward model systems to test relationships between designed sequence and structure, stability and function.^{73–76} This is essential for the field to move towards quantitative and fully programmable protein design. As part of this quest, here we show that in addition to sequence-to-structure relationships for the well-defined secondary and tertiary elements of the targeted fold, loop design is important. Specifically, for connecting adjacent elements of secondary structure, design-specific loops appear to be better than flexible linkers, which are better than mismatched loops. In addition, we show that the hyperstability often observed with modern *de*



de novo designed scaffolds can be attenuated through robust rationale. In our case, the core packing of the parent design consists of eight layers, with three residues in each layer. All but one of these layers comprise solely leucine residues. The remaining layer is a hydrogen-bonded constellation of one asparagine and two threonine side chains. By introducing two more of these layers, such that 3/8 of the layers and 9/24 of the core residues are polar, the protein unfolds reversibly below 100 °C in aqueous buffer. Detailed analysis of experimental unfolding data reveals that the modified design has thermodynamic parameters comparable to natural proteins of similar size. This targeted approach adds to and complements the emerging quantitative understanding of stability determinants from high-throughput studies of other *de novo* miniproteins.⁷⁶

We posit that further systematic analyses of this and similar *de novo* proteins will help to uncover sequence-to-structure/stability relationships and advance fully predictive and quantitative *de novo* protein design.^{75–77} In addition, as we^{25,78} and others^{40,79} have demonstrated, small *de novo* peptide assemblies and single-chain proteins of the type delivered here provide robust scaffolds to generate functional *de novo* proteins by grafting or otherwise introducing functional residues such as binding and catalytic sites. Regardless of how functionalisation is introduced—i.e., rationally, computationally, or generatively with AI—an advantage of using pre-designed and well-characterised scaffolds is that the roles and positions of most residues are known and understood at the atomistic level, providing a robust foundation for designing, introducing, and controlling functional modifications.

Author contributions

X. L. and D. N. W. conceived the study and designed the peptide sequences and experiments. X. L., K. I. A. and L. R. G. synthesised and characterised the peptides. X. L. designed and characterised the proteins. A. A. N. contributed to the design concept for distinguishing bundle handedness. X. L. collected the X-ray data and determined the protein X-ray crystal structures. J. C. and D. N. W. provided supervision and sourced funding. X. L. and D. N. W. wrote the paper. All authors have read and contributed to the preparation of the manuscript.

Conflicts of interest

There are no conflicts to declare.

Data availability

Supplementary information: Details of the experimental methods and additional experimental data are provided in the SI. Crystallographic data for sc-apCC3-CW1, sc-apCC3-CW2, sc-apCC3-ACW, sc-apCC3-CW-0NTT, and sc-apCC3-CW-2NTT have been deposited to the PDB and given the accession codes 9rgv, 9rgw, 9rgx, 9rgy, 9rgz respectively. Data for this article are openly available at Zenodo at DOI: <https://doi.org/10.5281/zenodo.15620429>. CC+ search results, peptide and protein sequences, purification data, biophysical and structural

characterisations, and thermodynamic data and fits (PDF). See DOI: <https://doi.org/10.1039/d5sc05576h>.

Acknowledgements

X. L. is supported by an Engineering and Physical Sciences Research Council (EPSRC)-funded PhD studentship through the Centre for Doctoral Training in Technology Enhanced Chemical Synthesis (EP/S024107/1). K. I. A. was supported by a Biotechnology and Biological Sciences Research Council (BBSRC)–National Science Foundation grant (BB/V004220/1 and 2019598). We are also grateful to the Max Planck-Bristol Centre for Minimal Biology, which supports K. I. A. and D. N. W. J. C. is supported by ERC through Advanced grant DOGMATRON (883786). We thank the Mass Spectrometry Facility for access to the EPSRC-funded Bruker UltraFlex MALDI-TOF instrument (EP/K03927X/1) at School of Chemistry, University of Bristol. We thank Diamond Light Source for access to beamlines I03, I04, I24 (proposals mx 37593 and mx31440). We further thank N. J. Savery for scientific discussions and co-supervision of A. A. N. Finally, we thank R. Petrenas for assistance with determining the X-ray crystal structures and members of the Woolfson and Clayden laboratories for many helpful discussions.

References

- 1 T. Kortemme, *Cell*, 2024, **187**, 526–544.
- 2 N. Ferruz, M. Heinzinger, M. Akdel, A. Goncarenco, L. Naef and C. Dallago, *Comput. Struct. Biotechnol. J.*, 2023, **21**, 238–250.
- 3 K. I. Albanese, S. Barbe, S. Tagami, D. N. Woolfson and T. Schiex, *Nat. Rev. Methods Primers*, 2025, **5**, 1–28.
- 4 D. N. Woolfson, *J. Mol. Biol.*, 2021, **433**, 167160.
- 5 P.-S. Huang, S. E. Boyken and D. Baker, *Nature*, 2016, **537**, 320–327.
- 6 X. Pan and T. Kortemme, *J. Biol. Chem.*, 2021, **296**, 100558.
- 7 I. V. Korendovych and W. F. Degrado, *Q. Rev. Biophys.*, 2020, **53**, 1–33.
- 8 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235–242.
- 9 K. B. Stephen, R. Bhatt, C. Bhikadiya, C. Bi, A. Biester, P. Biswas, S. Bittrich, S. Blaumann, R. Brown, H. Chao, V. R. Chithari, A. C. Paul, V. C. Gregg, M. D. Jose, S. Dutta, Z. Feng, W. F. Justin, S. Ghosh, S. G. David, R. K. Green, V. Guranovic, J. Henry, P. H. Brian, M. Joy, T. K. Jason, I. Khokhriakov, J.-S. Lai, L. L. Catherine, Y. Liang, D. Myers-Turnbull, E. Peisach, I. Persikova, W. P. Dennis, A. Pingale, Y. Rose, J. Sagendorf, A. Sali, J. Segura, M. Sekharan, C. Shao, J. Smith, M. Trumbull, B. Vallat, M. Voigt, B. Webb, S. Whetstone, A. Wu-Wu, T. Xing, Y. Y. Jasmine, A. Zalevsky and C. Zardecki, *Nucleic Acids Res.*, 2025, **53**, D564–D574.
- 10 C. W. Wood, J. W. Heal, A. R. Thomson, G. J. Bartlett, A. Á. Ibarra, R. L. Brady, R. B. Sessions and D. N. Woolfson, *Bioinformatics*, 2017, **33**, 3043–3050.



- 11 A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. W. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y.-E. A. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker and P. Bradley, in *Methods Enzymol.*, Elsevier, 2011, vol. 487, pp. 545–574.
- 12 B. Kuhlman and P. Bradley, *Nat. Rev. Mol. Cell Biol.*, 2019, **20**, 681–697.
- 13 J. L. Watson, D. Juergens, N. R. Bennett, B. L. Trippe, J. Yim, H. E. Eisenach, W. Ahern, A. J. Borst, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, N. Hanikel, S. J. Pellock, A. Courbet, W. Sheffler, J. Wang, P. Venkatesh, I. Sappington, S. V. Torres, A. Lauko, V. De Bortoli, E. Mathieu, S. Ovchinnikov, R. Barzilay, T. S. Jaakkola, F. Dimaio, M. Baek and D. Baker, *Nature*, 2023, **620**, 1089–1100.
- 14 R. Krishna, J. Wang, W. Ahern, P. Sturmfels, P. Venkatesh, I. Kalvet, G. R. Lee, F. S. Morey-Burrows, I. Anishchenko, I. R. Humphreys, R. Mchugh, D. Vafeados, X. Li, G. A. Sutherland, A. Hitchcock, C. N. Hunter, A. Kang, E. Brackenbrough, A. K. Bera, M. Baek, F. Dimaio and D. Baker, *Science*, 2024, 384.
- 15 A. H.-W. Yeh, C. Norn, Y. Kipnis, D. Tischer, S. J. Pellock, D. Evans, P. Ma, G. R. Lee, J. Z. Zhang, I. Anishchenko, B. Coventry, L. Cao, J. Dauparas, S. Halabiya, M. Dewitt, L. Carter, K. N. Houk and D. Baker, *Nature*, 2023, **614**, 774–780.
- 16 J. Dauparas, G. R. Lee, R. Pecoraro, L. An, I. Anishchenko, C. Glasscock and D. Baker, *Nat. Methods*, 2025, **22**, 717–723.
- 17 A. M. Hermosilla, C. Berner, S. Ovchinnikov and A. A. Vorobieva, *Protein Sci.*, 2024, **33**, e5033.
- 18 D. N. Woolfson, *J. Biol. Chem.*, 2023, **299**, 104579.
- 19 A. N. Lupas and J. Basser, *Trends Biochem. Sci.*, 2017, **42**, 130–140.
- 20 S. R. Presnell and F. E. Cohen, *Proc. Natl. Acad. Sci. U. S. A.*, 1989, **86**, 6592–6596.
- 21 E. G. Baker, G. J. Bartlett, M. P. Crump, R. B. Sessions, N. Linden, C. F. J. Faul and D. N. Woolfson, *Nat. Chem. Biol.*, 2015, **11**, 221–228.
- 22 S. E. Boyken, Z. Chen, B. Groves, R. A. Langan, G. Oberdorfer, A. Ford, J. M. Gilmore, C. Xu, F. Dimaio, J. H. Pereira, B. Sankaran, G. Seelig, P. H. Zwart and D. Baker, *Science*, 2016, **352**, 680–687.
- 23 J. M. Fletcher, A. L. Boyle, M. Bruning, G. J. Bartlett, T. L. Vincent, N. R. Zaccai, C. T. Armstrong, E. H. C. Bromley, P. J. Booth, R. L. Brady, A. R. Thomson and D. N. Woolfson, *ACS Synth. Biol.*, 2012, **1**, 197–250.
- 24 W. M. Dawson, F. J. O. Martin, G. G. Rhys, K. L. Shelley, R. L. Brady and D. N. Woolfson, *Chem. Sci.*, 2021, **12**, 6923–6928.
- 25 A. J. Burton, A. R. Thomson, W. M. Dawson, R. L. Brady and D. N. Woolfson, *Nat. Chem.*, 2016, **8**, 837–844.
- 26 R. Petrenas, O. A. Hawkins, J. F. Jones, D. A. Scott, J. M. Fletcher, U. Obst, L. Lombardi, F. Pirro, G. J. Leggett, T. A. A. Oliver and D. N. Woolfson, *J. Am. Chem. Soc.*, 2025, **147**, 3796–3803.
- 27 J. Snoj, F. Lapenta and R. Jerala, *Chem. Sci.*, 2024, **15**, 3673–3686.
- 28 H. V. Zhang, F. Polzer, M. J. Haider, Y. Tian, J. A. Villegas, K. L. Kiick, D. J. Pochan and J. G. Saven, *Sci. Adv.*, 2016, **2**, e1600307.
- 29 A. Grosvirt-Dramen, Z. J. Urbach, P. J. Hurst, C. E. Kwok, J. P. Patterson, F. Wang and A. I. Hochbaum, *ACS Nano*, 2025, **19**, 10162–10172.
- 30 J. M. Fletcher, R. L. Harniman, F. R. H. Barnes, A. L. Boyle, A. Collins, J. Mantell, T. H. Sharp, M. Antognozzi, P. J. Booth, N. Linden, M. J. Miles, R. B. Sessions, P. Verkade and D. N. Woolfson, *Science*, 2013, **340**, 595–599.
- 31 G. G. Rhys, J. A. Cross, W. M. Dawson, H. F. Thompson, S. Shanmugaratnam, N. J. Savery, M. P. Dodding, B. Höcker and D. N. Woolfson, *Nat. Chem. Biol.*, 2022, **18**, 999–1004.
- 32 T. Plaper, E. Rihtar, T. Ž. Ramuta, V. Forstnerič, V. Jazbec, F. Ivanovski, M. Benčina and R. Jerala, *Cell Chem. Biol.*, 2024, **31**, 1460–1472.
- 33 E. A. Naudin, K. I. Albanese, A. J. Smith, B. Mylemans, E. G. Baker, O. D. Weiner, D. M. Andrews, N. Tigue, N. J. Savery and D. N. Woolfson, *Chem. Sci.*, 2022, **13**, 11330–11340.
- 34 K. I. Albanese, R. Petrenas, F. Pirro, E. A. Naudin, U. Borucu, W. M. Dawson, D. A. Scott, G. J. Leggett, O. D. Weiner, T. A. A. Oliver and D. N. Woolfson, *Nat. Chem. Biol.*, 2024, **20**, 991–999.
- 35 J. W. Bryson, J. R. Desjarlais, T. M. Handel and W. F. DeGrado, *Protein Sci.*, 1998, **7**, 1404–1414.
- 36 J. A. Boice, G. R. Dieckmann, W. F. DeGrado and R. Fairman, *Biochemistry*, 1996, **35**, 14480–14485.
- 37 J. P. Schneider, A. Lombardi and W. F. DeGrado, *Folding Des.*, 1998, **3**, 29–40.
- 38 G. R. Dieckmann, D. K. McRorie, D. L. Tierney, L. M. Utschig, C. P. Singer, T. V. O'Halloran, J. E. Penner-Hahn, W. F. DeGrado and V. L. Pecoraro, *J. Am. Chem. Soc.*, 1997, **119**, 6195–6196.
- 39 A. Lombardi, J. W. Bryson and W. F. DeGrado, *Biopolymers*, 1996, **40**, 495–504.
- 40 F. Natri, D. D'Alonzo, L. Leone, G. Zambrano, V. Pavone and A. Lombardi, *Trends Biochem. Sci.*, 2019, **44**, 1022–1040.
- 41 A. E. Tolbert, C. S. Ervin, L. Ruckthong, T. J. Paul, V. M. Jayasinghe-Arachchige, K. P. Neupane, J. A. Stuckey, R. Prabhakar and V. L. Pecoraro, *Nat. Chem.*, 2020, **12**, 405–411.
- 42 B. Lovejoy, S. Choe, D. Cascio, D. K. McRorie, W. F. DeGrado and D. Eisenberg, *Science*, 1993, **259**, 1288–1293.
- 43 S. Nautiyal, D. N. Woolfson, D. S. King and T. Alber, *Biochemistry*, 1995, **34**, 11645–11651.
- 44 S. Nautiyal and T. Alber, *Protein Sci.*, 1999, **8**, 84–90.
- 45 N. A. Schnarr and A. J. Kennan, *J. Am. Chem. Soc.*, 2002, **124**.
- 46 S. Bermeo, A. Favor, Y.-T. Chang, A. Norris, S. E. Boyken, Y. Hsia, H. K. Haddox, C. Xu, T. J. Brunette, V. H. Wysocki, G. Bhabha, D. C. Ekiert and D. Baker, *Nat. Struct. Mol. Biol.*, 2022, **29**, 1266–1276.



- 47 P. Kumar, R. Petrenas, W. M. Dawson, H. Schweke, E. D. Levy and D. N. Woolfson, *Protein Sci.*, 2023, **32**, e4789.
- 48 E. K. O'Shea, K. J. Lumb and P. S. Kim, *Curr. Biol.*, 1993, **3**, 658–667.
- 49 F. Thomas, A. L. Boyle, A. J. Burton and D. N. Woolfson, *J. Am. Chem. Soc.*, 2013, **135**, 4901–5228.
- 50 C. L. Edgell, A. J. Smith, J. L. Beesley, N. J. Savery and D. N. Woolfson, *ACS Synth. Biol.*, 2020, **9**, 427–436.
- 51 J. J. Chubb, K. I. Albanese, A. Rodger and D. N. Woolfson, *Biochemistry*, 2025, **64**, 1973–1982.
- 52 D. L. Akey, V. N. Malashkevich and P. S. Kim, *Biochemistry*, 2001, **40**, 6352–6360.
- 53 L. Gonzalez, D. N. Woolfson and T. Alber, *Nat. Struct. Biol.*, 1996, **3**, 1011–1018.
- 54 D. Neculai, A. M. Neculai, S. Verrier, K. Straub, K. Klumpp, E. Pfizner and S. Becker, *J. Biol. Chem.*, 2005, **280**, 40782–40787.
- 55 M. D. Watson, I. Peran and D. P. Raleigh, *Biochemistry*, 2016, **55**, 3685–3691.
- 56 J. Zhou and G. Grigoryan, unpublished work.
- 57 J. Zhou and G. Grigoryan, *Protein Sci.*, 2015, **24**, 508–524.
- 58 P. Kumar and D. N. Woolfson, *Bioinformatics*, 2021, **37**, 4575–4577.
- 59 W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577–2637.
- 60 W. G. Touw, C. Baakman, J. Black, T. A. H. te Beek, E. Krieger, R. P. Joosten and G. Vriend, *Nucleic Acids Res.*, 2015, **43**, D364–D368.
- 61 L.-J. Tai, S. M. Mcfall, K. Huang, B. Demeler, S. G. Fox, K. Brubaker, I. Radhakrishnan and R. I. Morimoto, *J. Biol. Chem.*, 2002, **277**, 735–745.
- 62 K. A. Dill, *Biochemistry*, 2002, **29**, 7133–7356.
- 63 D. F. Kreitler, D. E. Mortenson, K. T. Forest and S. H. Gellman, *J. Am. Chem. Soc.*, 2016, **138**, 6498–6505.
- 64 D. N. Woolfson, A. Cooper, M. M. Harding, D. H. Williams and P. A. Evans, *J. Mol. Biol.*, 1993, **229**, 502–511.
- 65 C. Tanford, *Adv. Protein Chem.*, 1968, **23**, 121–282.
- 66 C. N. Pace, *Methods Enzymol.*, 1986, **131**, 266–280.
- 67 S. E. Jackson, *Folding Des.*, 1998, **3**, R81–R91.
- 68 B. S. McCrary, S. P. Edmondson and J. W. Shriver, *J. Mol. Biol.*, 1996, **264**, 784–805.
- 69 R. Nikam, A. Kulandaisamy, K. Harini, D. Sharma and M. M. Gromiha, *Nucleic Acids Res.*, 2021, **49**, 420–424.
- 70 G. I. Makhatadze and P. L. Privalov, *J. Mol. Biol.*, 1990, **213**, 375–384.
- 71 P. L. Privalov and G. I. Makhatadze, *J. Mol. Biol.*, 1990, **213**, 385–391.
- 72 P. L. Privalov and G. I. Makhatadze, *J. Mol. Biol.*, 1992, **224**, 715–723.
- 73 E. G. Baker, G. J. Bartlett, K. L. P. Goff and D. N. Woolfson, *Acc. Chem. Res.*, 2017, **50**, 2085–2092.
- 74 E. G. Baker, C. Williams, K. L. Hudson, G. J. Bartlett, J. W. Heal, K. L. P. Goff, R. B. Sessions, M. P. Crump and D. N. Woolfson, *Nat. Chem. Biol.*, 2017, **13**, 764–770.
- 75 K. Tsuboyama, J. Dauparas, J. Chen, E. Laine, Y. Mohseni Behbahani, J. J. Weinstein, N. M. Mangan, S. Ovchinnikov and G. J. Rocklin, *Nature*, 2023, **620**, 434–444.
- 76 G. J. Rocklin, T. M. Chidyausiku, I. Goresnik, A. Ford, S. Houliston, A. Lemak, L. Carter, R. Ravichandran, V. K. Mulligan, A. Chevalier, C. H. Arrowsmith and D. Baker, *Science*, 2017, **357**, 168–175.
- 77 K. T. O'Neil and W. F. DeGrado, *Science*, 1990, **250**, 646–651.
- 78 J. M. Fletcher, K. A. Horner, G. J. Bartlett, G. G. Rhys, A. J. Wilson and D. N. Woolfson, *Chem. Sci.*, 2018, **9**, 7656–7665.
- 79 I. V. Korendovych and W. F. DeGrado, *Curr. Opin. Struct. Biol.*, 2014, **27**, 113–121.

