

## **SUPPORTING INFORMATION**

### **From peptides to proteins: coiled-coil tetramers to single-chain 4-helix bundles**

Elise A. Naudin,<sup>1</sup> Katherine I. Albanese,<sup>1,2</sup> Abigail J. Smith,<sup>3</sup> Bram Mylemans,<sup>1,2</sup>  
Emily G. Baker,<sup>1,3</sup> Orion D. Weiner,<sup>4</sup> David M. Andrews,<sup>5</sup> Natalie Tighe,<sup>6</sup>  
Nigel J. Savery,<sup>3,7\*</sup> and Derek N. Woolfson<sup>1,2,3,7\*</sup>

<sup>1</sup>School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK

<sup>2</sup>Max Planck-Bristol Centre for Minimal Biology, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK

<sup>3</sup>School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol BS8 1TD, UK

<sup>4</sup>Cardiovascular Research Institute and Department of Biochemistry and Biophysics, University of California, San Francisco, 555 Mission Bay Blvd. South, San Francisco, CA 94158, USA

<sup>5</sup>Oncology R&D, AstraZeneca, Cambridge Science Park, Darwin Building, Cambridge CB4 0WG, UK

<sup>6</sup>BioPharmaceuticals R&D, AstraZeneca, Granta Park, Cambridge CB21 6GH, UK

<sup>7</sup>BrisEngBio, School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK

\*To whom correspondence should be addressed:  
[N.J.Savery@bristol.ac.uk](mailto:N.J.Savery@bristol.ac.uk) and [D.N.Woolfson@bristol.ac.uk](mailto:D.N.Woolfson@bristol.ac.uk)

## Table of contents

<b>1. Methods</b> .....	<b>3</b>
<b>1.1. General</b> .....	<b>3</b>
<b>1.2. Computational tools</b> .....	<b>3</b>
<b>1.3. Peptide synthesis and purification</b> .....	<b>4</b>
<b>1.4. Protein expression and purification</b> .....	<b>5</b>
<b>1.5. Solution-phase biophysical characterizations</b> .....	<b>6</b>
<b>1.6. <i>In vivo</i> characterization of peptide assembly</b> .....	<b>9</b>
<b>1.7. Structural characterization</b> .....	<b>9</b>
<b>2. Supplementary data</b> .....	<b>11</b>
<b>2.1. Supplementary tables</b> .....	<b>11</b>
<b>2.2. Supplementary figures</b> .....	<b>17</b>
2.2.1. Rational and computational design.....	17
2.2.2. Characterization of homopeptides .....	21
2.2.3. Characterization of heterotetramers .....	30
2.2.4. Characterization of the single-chain protein.....	40
<b>3. References</b> .....	<b>44</b>

## 1. Methods

### 1.1. General

All solvents, chemicals, and reagents were purchased from commercial sources and used without further purification. Fluorenylmethoxycarbonyl(Fmoc)- $\alpha$ -L-amino acids, Rink amide MBHA resin for solid-phase peptide synthesis (SPPS) and N,N-dimethylformamide (DMF) were purchased from Sigma-Aldrich, Fisher Scientific and Cambridge Reagents. Coupling reagents Oxyma Pure and diisopropylcarbodiimide (DIC) were purchased from Fluorochem and Sigma-Aldrich, respectively. Pyridine was purchased from Fisher Scientific; triisopropylsilane (TIPS) was from Acros Organics. Morpholine and trifluoroacetic acid (TFA) as well as all other chemicals (reagent grade) were purchased from Sigma-Aldrich.

### 1.2. Computational tools

#### **CC+ search**

Antiparallel coiled-coil tetramers and bundles were analysed through the online CC+ database.<sup>1</sup> Four-helix coiled coils (CCs) with an antiparallel orientation, homo- or hetero-oligomeric partnering, canonical or non-canonical sequence repeats and comprising many chains of more than 11 residues, were searched with a sequence redundancy cut-off of  $\leq 50\%$ . From the database updated the 9<sup>th</sup> of January 2020, 65 coiled coils were assigned from 179  $\alpha$ -helical sequences. Amino-acid profiles were compiled for each position of the heptad repeat, **abcdefg** (raw counts in Table S1). The profile was normalized using expected amino-acid frequencies from SWISS-PROT to give propensities for each residue (Figure 2B).

Amino-acid frequencies from SWISS-PROT are the following: Ala (A) 8.25, Arg (R) 5.53, Asn (N) 4.06, Asp (D) 5.46, Cys (C) 1.38, Gln (Q) 3.93, Glu (E) 6.72, Gly (G) 7.07, His (H) 2.27, Ile (I) 5.91, Leu (L) 9.65, Lys (K) 5.80, Met (M) 2.41, Phe (F) 3.86, Pro (P) 4.74, Ser (S) 6.64, Thr (T) 5.35, Trp (W) 1.10, Tyr (Y) 2.92, Val (V) 6.86.

#### **AlphaFold2 prediction**

AlphaFold2-multimer predictions were performed using the version implemented in ColabFold.<sup>2-4</sup> Different oligomeric states (4, 6, 8 and 12) were tested for each sequence in order to assess if the tetrameric assembly was still predicted. Predicted structures were relaxed using the

implemented Amber force fields. Predicted IDDT per position and the predicted alignment error (PAE) were collected to evaluate the predictions.

### 1.3. Peptide synthesis and purification

#### ***Automated microwave Fmoc/tBu solid-phase peptide synthesis (SPPS)***

Automated microwave SPPS was performed on a Liberty Blue (CEM) synthesizer with in-line UV monitoring. The synthesis was performed on a 0.1 mmol scale on Rink amide MBHA (0.65 mmol/g loading, 100-200 mesh) resins. Side-chain protections of the amino acids were as follows: Gln(Trt), Glu(OtBu), Lys(Boc), Tyr(tBu), Trp(Boc). The coupling of the amino acids was performed in DMF by adding 5 equivalents of amino acids (2.5 mL, 0.2 M), 10 equivalents of DIC in DMF (1.0 mL, 1 M) and 5 equivalents of Oxyma Pure in DMF (1 mL, 0.5 M). Standard couplings were performed at 90 °C for 4 min (100 W for 20 s, 60 W for 10 s, 35 W for 240 s). Standard deprotections were performed using 20% (v/v) morpholine in DMF at 90 °C for 1 min (125 W 30 s, 32 W 60 s). All peptides were manually capped with *N*-terminal acetyl groups through addition of pyridine (0.5 mL) and acetic anhydride (0.25 mL) in DMF (9.25 mL), with shaking at room temperature (rt) for 20 minutes. The resin was washed 3 times with DMF followed by 6 times with DCM before cleavage. Peptides were cleaved from the resin and deprotected with addition of 10 mL of a mixture 95:2.5:2.5 v/v/v TFA/H<sub>2</sub>O/TIPS, with shaking at rt for 2 hours. The SeMet-containing peptide was cleaved from the resin with addition of 10 mL of a mixture 88.7:5:3.3:3 v/v/v/v TFA/thioanisole/anisole/1,2-ethanedithiol (EDT), with shaking at rt for 2 hours. The TFA solution was filtered to remove the resin beads and was reduced in volume to ≤5 mL using a flow of N<sub>2</sub>. Cleaved peptide was precipitated with cold diethyl ether (≈45 mL), isolated via centrifugation and dissolved in a 1:1 mixture MeCN/H<sub>2</sub>O. Crude peptides were lyophilized to yield white powders.

#### ***Semi-preparative High Performance Liquid Chromatography (HPLC)***

All peptides were purified by reverse phase HPLC (JASCO) using a Luna C18 (Phenomenex) column (150 x 10 mm, 5 μM particle size, 100 Å pore size). Crude peptide was dissolved at 7 mg/mL in 20% v/v MeCN in H<sub>2</sub>O with 0.1% TFA, injected to the column and eluted with a 3 mL/min linear gradient (20-60%) of MeCN in H<sub>2</sub>O with 0.1% TFA each over 30 minutes. Alternatively, acidic peptides were eluted with a linear gradient (20-60%) of MeCN in 25 mM NH<sub>4</sub>HCO<sub>3</sub> in H<sub>2</sub>O over 30 min. Elution of the peptide was detected with in-line UV monitoring at

220 nm and 280 nm simultaneously. When required, a column oven (50 °C) was used to improve separation. Pure fractions were identified by analytical HPLC and matrix-assisted laser desorption/ionization–time of flight (MALDI-TOF) mass spectrometry, then pooled, and freeze-dried.

### ***Analytical HPLC***

Analytical HPLC traces were obtained using a Jasco 2000 series HPLC system and a Phenomenex Kinetex C18 (100 x 4.6 mm, 5 µm particle size, 100 Å pore size) column. Chromatograms were monitored at 220 nm and 280 nm. The linear gradient was 20–80% MeCN in water (each containing 0.1% TFA) over 25 min at a flow rate of 1 mL/min. Alternatively, acidic peptides were analyzed with the same linear gradient of MeCN but with 25 mM NH<sub>4</sub>HCO<sub>3</sub> in H<sub>2</sub>O. When required, a column oven (50 °C) was used to assist peptide elution.

### ***Mass spectrometry***

Matrix-assisted laser desorption/ionization–time of flight (MALDI-TOF) mass spectra were collected on a Bruker UltraFlex MALDI-TOF mass spectrometer operating in positive-ion reflector mode. Peptides were spotted on a ground steel target plate using α-cyano-4-hydroxycinnamic acid dissolved in 1:1 MeCN/H<sub>2</sub>O as the matrix. Masses quoted are for the monoisotopic mass as the singly protonated species. Alternatively, full electrospray ionization (ESI) MS spectra were acquired on a Synapt G2S (Waters) mass spectrometer equipped with an IMS-Q-TOF analyser and using an Advion Nanomate for robot chip-based nanospray ionization in positive mode. Some acidic peptides were analysed in negative mode. 5 µL of a 50 µM peptide solution in 1:1 MeCN/H<sub>2</sub>O were generally injected for the analysis. Masses quoted are for the deconvoluted monoisotopic mass.

## **1.4. Protein expression and purification**

### ***Cloning and protein sequence***

A synthetic gene for sc-apCC-4 cloned directly into pET28a vector using NdeI and XhoI restriction sites was purchased from Twist Biosciences.

Sc-apCC-4 protein sequence:

MGSSHHHHHSSGLVPRGSHMQLEEIAQQLEEIAKQLKKIAWQLKKIAQGEP SAQQGLEEIAQ  
QLEEIAKQLKKIAWQLKKIAQGPDSVQLEEIAQQLEEIAKQLKKIAWQLKKIAQGGTSGGQLEEIA  
QLEEIAKQLKKIAWQLKKIAQ

### ***Protein expression and purification***

The sc-apCC-4 gene was transformed and then recombinantly expressed in *E. coli* BL21-DE3 Gold (Agilent) cells. Flasks containing 1 L of LB-kanamycin (Thermo Scientific, 50 µg/mL) were inoculated with 5 mL overnight cultures and incubated to an OD<sub>600</sub> of ~0.6 at 37 °C with 220 rpm shaking. Expression was induced with 0.5 mM IPTG (Neo Biotech), and cultures were incubated at 18 °C overnight with 220 rpm shaking (Thermo Scientific MaxQ 4000).

Following expression, cultures were pelleted at 5000 rpm for 10 minutes (Thermo Scientific Sorvall Lynx 4000). Cell pellets were resuspended in 20 mL lysis buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 30 mM imidazole, 1 mg/mL lysozyme) for 30 min at 37 °C. Resuspended pellets were sonicated using a Biologics Model 3000 Ultrasonic homogenizer with settings to 50% power, 90% pulser (1 pulse/second) for five minutes, then clarified at 12,000 rpm for 1 hour. The expressed sc-apCC-4 was first purified with Ni-affinity chromatography. Filtered lysate was loaded onto an ÄKTAprime plus (GE Healthcare) equipped with a HisTrap- 5 mL HP column (GE Healthcare). His-tagged proteins were eluted at 55% Buffer B (Buffer A: 50 mM Tris, 150 mM NaCl, 30 mM imidazole, pH 7.4; Buffer B: 50 mM Tris, 150 mM NaCl, 300 mM imidazole, pH 7.4) using a step gradient. Fractions were combined and purified further by size exclusion chromatography using a HiLoad 16/600 Superdex 75 pg size exclusion column (GE Healthcare) equilibrated in buffer containing 50 mM sodium phosphate, pH 7.4, 150 mM NaCl. Eluted fractions were pooled, concentrated, diluted 1:1 in loading buffer (50mM Tris-Cl (pH 6.8), 10 mM DTT, 2% SDS, 0.1% bromophenol blue, 10% glycerol), and heated to 95 °C. 15 µL samples and PageRuler Low Range Unstained Protein Ladder (Thermo Scientific) were separated on a 10-20% SDS-PAGE gel using Tris-Glycine running buffer at 120V (BioRad PowerPac Basic) to confirm identity.

## **1.5. Solution-phase biophysical characterizations**

### ***Peptide and protein concentration determination***

Peptide and protein concentrations were determined from absorbance at 280 nm measured in a Nanodrop 2000 (Thermo Scientific) spectrometer ( $\epsilon_{280}(\text{Trp}) = 5690 \text{ cm}^{-1}$ ;  $\epsilon_{280}(\text{Tyr}) = 1280 \text{ cm}^{-1}$ ;  $\epsilon_{280}(\text{sc-apCC-4}) = 22000 \text{ cm}^{-1}$ ), or at 214 nm using a Cary-100 (Agilent) UV-Visible spectrometer by measuring the peptide bond.<sup>5</sup>

### **Sample annealing**

When necessary, mixtures of acidic and basic peptides were annealed before biophysical measurements. Peptide samples in buffer were heated up to 90 °C in a heat block (Grant), incubated at 90 °C for 10 min and slowly cooled down to room temperature over 90 min.

### **Circular dichroism (CD) spectroscopy**

Circular dichroism (CD) data were collected in the far-UV region on a JASCO J-810 or J-815 spectropolarimeter fitted with a Peltier temperature controller. Peptide samples were made up as 50 µM peptide solution in phosphate buffered saline (PBS; 8.2 mM sodium phosphate dibasic, 1.8 mM potassium phosphate monobasic, 137 mM sodium chloride, 2.4 mM potassium chloride), pH 7.4 at 5 °C. For the protein sc-apCC-4, CD spectra were acquired at 25 µM protein concentration in 50 mM sodium phosphate, 150 mM NaCl, pH 7.4 at 5 °C. Data were collected in a 1 mm quartz cuvette between 190 and 260 nm and the instrument was set as follows: band width 1 nm, data pitch 1 nm, scanning speed 100 nm/min, 1 s response time. Every CD curve was obtained by averaging of 8 scans and subtracting the background signal of buffer and cuvette. For thermal profile experiments, CD signals were monitored at 222 nm wavelength. A temperature range of 5–95 °C and a temperature ramp rate of 60 °C per hour were used with the same settings and peptide or protein concentration as above.

CD data were converted from ellipticities (mdeg) to mean residue ellipticities (MRE, (deg·cm<sup>2</sup>·dmol<sup>-1</sup>·res<sup>-1</sup>)) by normalizing for concentration of peptide bonds and the cell path length using the equation:

$$MRE = \frac{\theta \times 10^6}{c \times l \times n}$$

where the variable  $\theta$  is the measured difference in absorbed in left- and right-handed circularly polarized light in millidegrees,  $c$  is the micromolar concentration of the compound,  $l$  is the path length of the cuvette in mm, and  $n$  is the number of amide bonds in the polypeptide, for which the N-terminal acetyl bond was included but not the C-terminal amide.

### Chemical denaturation:

The stability against chemical denaturation was examined in presence of the chaotropic agent guanidinium hydrochloride (Gn·HCl). 5 µM or 50 µM of apCC-Tet\* or 1 µM of sc-apCC-4 were mixed with increasing concentration of Gn·HCl (0-6 M), and samples were incubated overnight. Data were collected in a 1 or 5 mm quartz cuvette at 5 °C between 190 and 260 nm (100 nm/min,

1 nm interval and bandwidth, 1 s response time). Variable-temperature spectra were collected at 222 nm for the highest concentration of Gn·HCl (i.e., 6 M) using the settings and peptide or protein concentration as above between 5 and 95 °C at a 60 °C/hour ramp rate.

### ***Analytical ultracentrifugation***

Analytical ultracentrifugation (AUC) was performed on a Beckman Optima X-LA or X-LI analytical ultracentrifuge with an An-50-Ti or An-60-Ti rotor (Beckman-Coulter). Buffer densities, viscosities, and peptide partial specific volumes ( $\bar{v}$ ) were calculated using SEDNTERP (<http://rasmb.org/sednterp/>).

For sedimentation velocity (SV), sample solutions of 310 or 410  $\mu\text{L}$  were prepared in PBS at 150  $\mu\text{M}$  peptide concentration and placed in a sedimentation velocity cell with an epon or aluminium, respectively, 2-channel centrepiece and quartz windows. The reference channel was loaded with 320  $\mu\text{L}$  (or 420  $\mu\text{L}$ , respectively) of PBS buffer. For SV experiments with the sc-apCC-4 protein, samples were prepared at 25  $\mu\text{M}$  peptide concentration in 50 mM sodium phosphate, pH 7.4, 150 mM NaCl. The samples were centrifuged at 50 krpm at 20 °C, with absorbance scans taken across a radial range of 5.8–7.3 cm at 5 min intervals to a total of 120 scans. Data from a single run were fitted to a continuous  $c(s)$  distribution model using SEDFIT,<sup>6</sup> at 95% confidence level. Residuals for sedimentation velocity experiments are shown as a bitmap in which the grayscale shade indicates the difference between the fit and raw data (residuals  $< -0.05$  are black and residuals  $> 0.05$  are white). Good fits are uniformly grey without major dark or light streaks.

Sedimentation equilibrium (SE) experiments were performed at 70  $\mu\text{M}$  peptide concentration or 25  $\mu\text{M}$  protein concentration in 110  $\mu\text{L}$  at 20 °C. Experiments were run in triplicate in a six-channel epon centrepiece. The samples were centrifuged at speeds in the range of 18–48 krpm or 20–45 krpm in increments of 6 and 5 krpm respectively. Scans at each speed were duplicated 6 or 8 hours after the increase in speed to confirm equilibration. Data were fitted using SEDPHAT to a single species model.<sup>7</sup> Monte Carlo analysis was performed to give 95% confidence limits.

### ***Fluorescence quenching experiments***

Fluorescent quenching experiments were performed following previously published procedure.<sup>8</sup> Mixture were prepared as an equimolar ratio of acidic and basic peptide analogues at 50  $\mu\text{M}$  of the CN-Phe-containing peptide in phosphate buffer (8.2 mM sodium phosphate dibasic, 1.8 mM

potassium phosphate monobasic), pH 7.4. Fluorescence emission spectra from 260 nm to 400 nm were recorded on a Jasco Fluorimeter, in a 10 x 10 mm quartz cuvette, and with an excitation wavelength of 240 nm. Every fluorescence emission spectrum was obtained by averaging of 3 scans and subtracting the background signal of buffer and cuvette.

## **1.6. *In vivo* characterization of peptide assembly**

### ***Plasmids for GFP repression assays***

The plasmids for the GFP repression assays, pBADLacI, pBADLacI\*, pBADLacI\*-ccDi and pBADLacI\*-apCC-Tet, were used as previously described.<sup>9,10</sup> In order to make pBADLacI\*-apCC-Tet\*, a DNA fragment encoding apCC-Tet\*, with codons optimized for *E. coli*, was designed and was purchased from Eurofins Genomics. The DNA fragment was cloned into pBADLacI\* at Acc65I/XbaI restriction sites. These plasmids expressed WT LacI or a truncated dimerization mutant LacI\* (LacI L251A aa 1-332) with terminal His<sub>6</sub>, T7, and Xpress tags from the P<sub>araBAD</sub> arabinose-inducible promoter. CC peptides were fused in frame and C-terminal to LacI\* by a flexible linker. The GFP reporter plasmids pVRbLacUV5 and pVRblacO1-lacO1 have also been described previously.<sup>9</sup>

### ***GFP repression assays***

In order to assay repression of GFP transcription, TB28 cells (MG1655ΔLacIZYA)<sup>11</sup> were transformed with either pVRbLacUV5 (1 lac operator site) or pVRblacO1-lacO1 (2 lac operator sites) and pBADLacI\* or its derivatives. Repression assays were carried out essentially as described previously,<sup>9</sup> except that fluorescence intensity was measured using a Clariostar plus microplate reader (BMG Labtech) with GFP settings (i.e., monochromator/filter settings: 470-15/515-20).

## **1.7. Structural characterization**

### ***Crystal growth***

Diffraction-quality peptide crystals were grown using a sitting-drop vapor-diffusion method. Freeze-dried peptides were dissolved in ultrapure water and diluted to 10 mg/mL. Purified sc-apCC-4 in 50 mM sodium phosphate, 150 mM NaCl was concentrated, then diluted to 10 mg/mL in deionized water, in this case the final buffer concentration was 40 mM phosphate, 120 mM NaCl. Commercially available sparse matrix screens were used (Morpheus®, JCSG-plus™,

Structure Screen 1 and 2, Pact Premier™, ProPlex™; Molecular Dimensions), and the drops were dispensed using a robot (Oryx8; Douglas Instruments). The peptides were screened in MRC 2 drop plates. Drops consisting of 0.3 µL of peptide solution and 0.3 µL of reservoir solution were screened in parallel with drops consisting of 0.4 µL of the peptide solution and 0.2 µL of reservoir solution. For the sc-apCC-4, 0.4 µL of the protein solution and 0.2 µL of reservoir solution were mixed. All plates were incubated at 20 °C. Crystals generally formed within a month, and after looping were soaked in reservoir solution containing 25% glycerol as a cryoprotectant. Crystals of apCC-Tet\*<sup>3</sup>-A<sub>2</sub>B<sub>2</sub> were obtained by optimization around the condition from Structure Screen 1 and 2 C5 comprising 1.4 M sodium citrate tribasic dihydrate, 0.1 M sodium HEPES, pH 7.5. A range of concentrations of sodium citrate tribasic dihydrate from 0.8 M to 1.8 M was screened as well as peptides concentrations (11 mg/mL, 8.5 mg/mL and 6.4 mg/mL). 0.5 µL of the peptide solution and 0.5 µL of reservoir solution were mixed in MRC Maxi 48 Well plates and incubated at 20 °C.

Final crystallization conditions for all peptides are provided in Table S2.

### ***X-ray crystal structure determination***

Diffraction data for the crystals were obtained at the Diamond Light Source on beamlines I04 or I24 (for apCC-Tet\*<sup>3</sup>-A<sub>2</sub>B<sub>2</sub> and apCC-Tet\*, respectively) or at the European Synchrotron Radiation Facility (ESRF) on beamline ID30B (for apCC-Tet\*<sup>3</sup> and sc-apCC-4).

Data were processed using the automated pipelines: Xia2,<sup>12</sup> which ports data through DIALS<sup>13</sup> or MOSFLM<sup>14</sup> to POINTLESS and AIMLESS<sup>15</sup> as implemented in the CCP4 suite,<sup>16</sup> or XDS to XSCALE.<sup>17</sup> Structure of apCC-Tet\*<sup>3</sup>-A<sub>2</sub>B<sub>2</sub> was phased using *ab initio* phasing using ARCIMBOLDO\_LITE.<sup>18</sup> Structure of apCC-Tet\* was phased using FRAGON,<sup>19</sup> the initial phases were into and refined using BUCCANEER<sup>20</sup> and REFMAC 5.<sup>21</sup> Structure of apCC-Tet\*<sup>3</sup> and sc-apCC-4 were solved by molecular replacement using AlphaFold<sup>2,4</sup> or apCC-Tet\*, respectively, as models using PHASER.<sup>22</sup> Final structures were obtained after iterative rounds of model building with COOT and refinement with PHENIX Refine.<sup>23</sup> Solvent-exposed atoms lacking map density were either deleted or left at full occupancy. Data collection and refinement statistics are provided in Table S3.

## 2. Supplementary data

### 2.1. Supplementary tables

**Table S1. Raw amino-acid counts for antiparallel four-helix coiled coils found by the CC+ database (over 179  $\alpha$ -helical sequences).<sup>a</sup>**

-	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	Sum
A	103	41	42	77	<b>92</b>	69	74	498
C	7	2	5	0	4	0	1	19
D	5	67	44	8	31	58	25	238
E	17	128	183	24	71	148	72	643
F	31	5	7	14	4	10	12	83
G	4	27	18	5	20	25	21	120
H	13	15	12	10	5	11	12	78
I	81	14	13	<b>194</b>	33	18	65	418
K	11	74	76	5	60	75	38	339
L	<b>226</b>	37	18	<b>224</b>	98	30	<b>158</b>	791
M	33	11	12	32	17	4	16	125
N	9	51	33	14	23	52	23	205
P	0	0	2	5	0	0	6	13
Q	21	66	68	33	58	55	<b>67</b>	368
R	8	82	70	6	77	75	39	357
S	40	37	40	36	56	43	46	298
T	36	25	31	22	37	39	34	224
V	126	17	17	82	43	13	39	337
W	0	0	4	3	2	5	8	22
Y	5	7	9	8	20	11	4	64
Sum	780	706	706	804	758	741	766	5240

<sup>a</sup>Residues identified for the design of the new antiparallel tetramer sequences are highlighted in red and bold.

**Table S2. Sequences of *de novo* peptides and proteins in this study.**

Peptide/protein names	Sequences <i>gabcdef gabcdef gabcdef gabcdef loop</i>
pLLL	Ac-G LLEELAQ LLEELAK LLKKLAW LLKKLAQ G-NH <sub>2</sub>
pLLI	Ac-G LLEEIAQ LLEEIAK LLKKIAW LLKKIAQ G-NH <sub>2</sub>
pQLL	Ac-G QLEELAQ QLEELAK QLKKLAW QLKKLAQ G-NH <sub>2</sub>
pQLI apCC-Tet*	Ac-G QLEEIAQ QLEEIAK QLKKIAW QLKKIAQ G-NH <sub>2</sub>
pQLL <sup>3</sup>	Ac-G QLEELAK QLQQLAW QLKKLAQ G-NH <sub>2</sub>
pQLI <sup>3</sup> apCC-Tet* <sup>3</sup>	Ac-G QLEEIAK QLQQIAW QLKKIAQ G-NH <sub>2</sub>
pQLI-A apCC-Tet*-A	Ac-G QLEEIAQ QLEEIAK QLEEIAW QLEEIAQ G-NH <sub>2</sub>
pQLI-B apCC-Tet*-B	Ac-G QLKKIAQ QLKKIAK QLKKIAY QLKKIAQ G-NH <sub>2</sub>
pQLI <sup>3</sup> -A apCC-Tet* <sup>3</sup> -A	Ac-G QLEEIAK QLEEIAW QLEEIAQ G-NH <sub>2</sub>
pQLI <sup>3</sup> -B apCC-Tet* <sup>3</sup> -B	Ac-G QLKKIAK QLKKIAY QLKKIAQ G-NH <sub>2</sub>
pQLI <sup>3</sup> -A-nMSE apCC-Tet* <sup>3</sup> -A-nMSE	Ac-G QLXEIAK QLEEIAK QLEEIAQ G-NH <sub>2</sub>
pQLI <sup>3</sup> -B-c4CF apCC-Tet* <sup>3</sup> -B-c4CF	Ac-G QLKKIAK QLKKIAK QLKKIZQ G-NH <sub>2</sub>
pQLI <sup>3</sup> -B-n4CF apCC-Tet* <sup>3</sup> -B-n4CF	Ac-G QLKZIAK QLKKIAK QLKKIAQ G-NH <sub>2</sub>
<b>sc-apCC-4</b>	QLEEIAQ QLEEIAK QLKKIAW QLKKIAQ <u>GEPSAQG</u> QLEEIAQ QLEEIAK QLKKIAW QLKKIAQ <u>GPDSV</u> QLEEIAQ QLEEIAK QLKKIAW QLKKIAQ <u>GGTSGG</u> QLEEIAQ QLEEIAK QLKKIAW QLKKIAQ
pQLL-A	Ac-G QLEELAQ QLEELAK QLEELAW QLEELAQ G-NH <sub>2</sub>
pQLL-B	Ac-G QLKKLAQ QLKKLAK QLKKLAY QLKKLAQ G-NH <sub>2</sub>
pQLL <sup>3</sup> -A	Ac-G QLEELAK QLEELAW QLEELAQ G-NH <sub>2</sub>
pQLL <sup>3</sup> -B	Ac-G QLKKLAK QLKKLAY QLKKLAQ G-NH <sub>2</sub>

x = Selenomethionine (MSE); z = 4-cyanophenylalanine (4CF)

**Table S3. Crystallization conditions used to obtain the structures discussed in this article.**

<b>Peptide or protein</b>	<b>Molecular dimensions screen</b>	<b>Crystallization conditions of the reservoir</b>	<b>Dilution sample/reservoir (in the drop)</b>
<b>apCC-Tet* (8A3G)</b>	Proplex, G2	2.0 M Ammonium sulphate, 0.1 M Sodium acetate, pH 5.0	1:1
<b>apCC-Tet*<sup>3</sup> (8A3I)</b>	Structure Screen 1 and 2, C5	1.4 M sodium citrate tribasic dihydrate, 0.1 M sodium HEPES, pH 7.5	2:1
<b>apCC-Tet*<sup>3</sup>-A<sub>2</sub>B<sub>2</sub> (8A3J)</b>	Optimization from Structure Screen 1 and 2, C5	1.8 M sodium citrate tribasic dihydrate, 0.1 M sodium HEPES, pH 7.5	1:1
<b>sc-apCC-4 (8A3K)</b>	Morpheus, D9	10% w/v PEG 20 000, 20% v/v PEG MME 550, 0.02 M 1,6-hexanediol, 0.02 M 1-butanol, 0.02 M (RS)-1,2-propanediol, 0.02 M 2-propanol, 0.02 M 1,4-butanediol, 0.2 M 1,3-propanediol, 0.1 M bicine/Trizma base pH 8.5	2:1

**Table S4. Merging and refinement statistics for all X-ray crystal structures.**

	<b>apCC-Tet*</b>	<b>apCC-Tet*<sup>3</sup></b>	<b>apCC-Tet*<sup>3</sup>-A<sub>2</sub>B<sub>2</sub></b>	<b>sc-apCC-4</b>
<b>PDB ID</b>	8A3G	8A3I	8A3J	8A3K
<b>Data collection</b>				
Source	Diamond I24	ESRF ID30B	Diamond I04	ESRF ID30B
Detector	PILATUS 6M	PILATUS3 6M	EIGER2 XE 16M	PILATUS3 6M
Wavelength (Å)	0.85	0.9253	0.9795	0.9253
Resolution range	26.86 - 0.96 (0.9943 - 0.96)	42.46 - 1.42 (1.471 - 1.42)	31.58 - 2.1 (2.175 - 2.1)	27.48 - 2.0 (2.071 - 2.0)
Space group	P 4 <sub>2</sub> 2 <sub>1</sub> 2	P 3 <sub>2</sub> 2 1	P 1 2 <sub>1</sub> 1	P 1 2 <sub>1</sub> 1
Unit cell: <i>a</i> , <i>b</i> , <i>c</i> (Å)	34.01 34.01 87.53	49.03 49.03 32.62	27.66 63.16 47.30	28.08 38.86 56.83
<i>α</i> , <i>β</i> , <i>γ</i> (°)	90 90 90	90 90 120	90 91.6752 90	90 101.884 90
Total reflections	645077 (19488)	28332 (2768)	64987 (6508)	18274 (1751)
Unique reflections	31036 (2293)	8690 (831)	9430 (923)	7660 (751)
Multiplicity	20.8 (8.5)	3.3 (3.2)	6.9 (7.1)	2.4 (2.3)
Completeness (%)	95.51 (72.54)	97.45 (93.53)	98.43 (97.88)	92.85 (92.36)
Mean I/sigma(I)	30.55 (1.32)	38.83 (2.30)	4.37 (0.62)	13.21 (1.29)
Wilson B-factor	8.99	16.65	22.57	47.29
R-merge	0.0429 (0.7173)	0.1061 (0.6371)	0.2294 (1.152)	0.03111 (0.8826)
R-meas	0.04392 (0.7643)	0.1268 (0.7589)	0.2486 (1.243)	0.03907 (1.101)
R-pim	0.009184 (0.2529)	0.06726 (0.4024)	0.09484 (0.4646)	0.02326 (0.6503)
CC1/2	1 (0.94)	0.973 (0.666)	0.986 (0.783)	0.999 (0.602)
CC*	1 (0.984)	0.993 (0.894)	0.997 (0.937)	1 (0.867)
<b>Refinement</b>				
Reflections used in refinement	30910 (2280)	8592 (809)	9413 (922)	7651 (749)
Reflections used for R-free	1514 (114)	427 (51)	389 (42)	345 (24)
R-work	0.1603 (0.2220)	0.1989 (0.3187)	0.2297 (0.2879)	0.2363 (0.3773)
R-free	0.1688 (0.2054)	0.2348 (0.4990)	0.2672 (0.3055)	0.2756 (0.3643)
CC(work)	0.969 (0.939)	0.498 (0.135)	0.907 (0.814)	0.969 (0.695)
CC(free)	0.976 (0.946)	0.752 (0.236)	0.918 (0.848)	0.987 (0.535)
Number of non-hydrogen atoms	577	379	1299	823
macromolecules	533	370	1275	823
ligands	5	0	0	0
solvent	39	9	24	0
Protein residues	64	50	186	127
RMS(bonds)	0.017	0.010	0.009	0.006
RMS(angles)	1.50	1.08	1.06	0.83
Ramachandran favored (%)	100.00	100.00	99.38	96.69
Ramachandran allowed (%)	0.00	0.00	0.62	3.31
Ramachandran	0.00	0.00	0.00	0.00
Rotamer outliers (%)	0.00	0.00	3.09	0.00
Clashscore	0.90	2.68	4.84	4.52
Average B-factor	12.89	28.54	29.38	55.19
macromolecules	11.97	28.00	29.16	55.19
ligands	26.99	-	-	-
solvent	20.69	38.03	33.90	-
Number of TLS groups	-	-	2	4

**Table S5. Socket2 analysis of crystallographic structures and related AlphaFold2 predictions.<sup>24</sup>**

Structures		Number of observed knobs at different cut-offs											
		7 Å				7.5 Å				8 Å			
		<i>g</i>	<i>a</i>	<i>d</i>	<i>e</i>	<i>g</i>	<i>a</i>	<i>d</i>	<i>e</i>	<i>g</i>	<i>a</i>	<i>d</i>	<i>e</i>
X-ray structures	apCC-Tet*	-	-	-	-	-	12	12	10	2	12	12	12
	apCC-Tet* <sup>3</sup>	-	-	-	-	-	8	8	4	-	8	8	8
	apCC-Tet* <sup>3</sup> -A <sub>2</sub> B <sub>2</sub> <sup>a</sup>	-	-	-	-	-	8	7	10	1	8	9	11
	sc-apCC-4	-	-	-	-	-	12	8	13	-	12	9	13
AlphaFold2 predictions	apCC-Tet*	-	12	12	4	-	12	12	12	-	12	12	12
	apCC-Tet* <sup>3</sup>	-	8	8	8	-	8	8	8	-	8	8	8
	apCC-Tet* <sup>3</sup> -A <sub>2</sub> B <sub>2</sub>	-	8	8	8	-	8	8	8	-	8	8	8
	sc-apCC-4	-	12	12	2	-	12	12	14	3	12	12	14

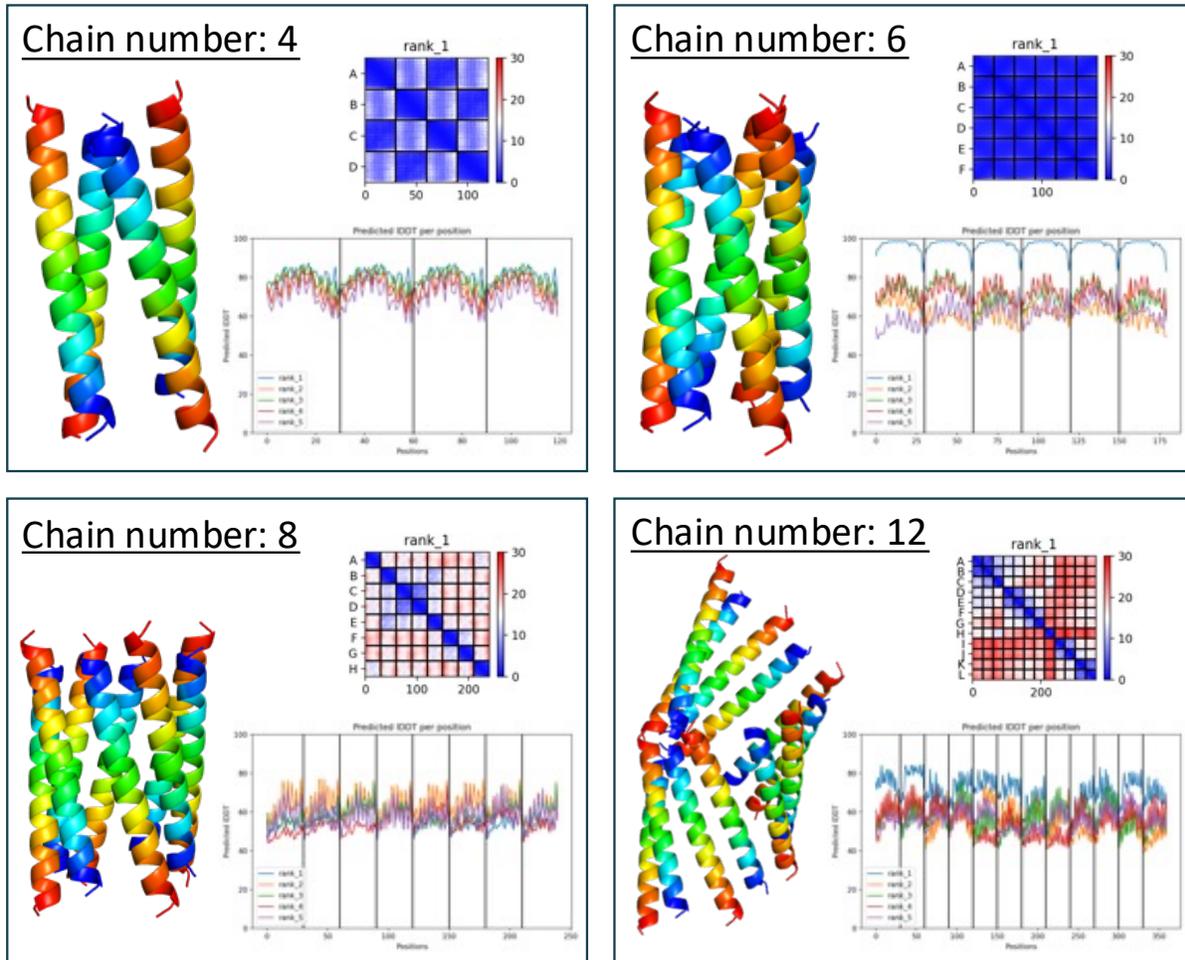
<sup>a</sup>The numbers of knobs correspond to the average per coiled-coil assembly as the crystal structure of apCC-Tet\*<sup>3</sup>-A<sub>2</sub>B<sub>2</sub> was composed of 2 coiled coils in the asymmetric unit.

**Table S6. Sequences comparison of historical *de novo* antiparallel 4-helix bundles.** Structurally resolved antiparallel 4-helix bundles with identified knobs-into-holes packing by SOCKET2<sup>24</sup> were visually inspected. These were mainly detected from the CC+ database.<sup>1</sup> Amino acids found at **g**, **a**, **d** and **e** positions are given below. Antiparallel 4-helix bundles without full structural characterization were not considered.

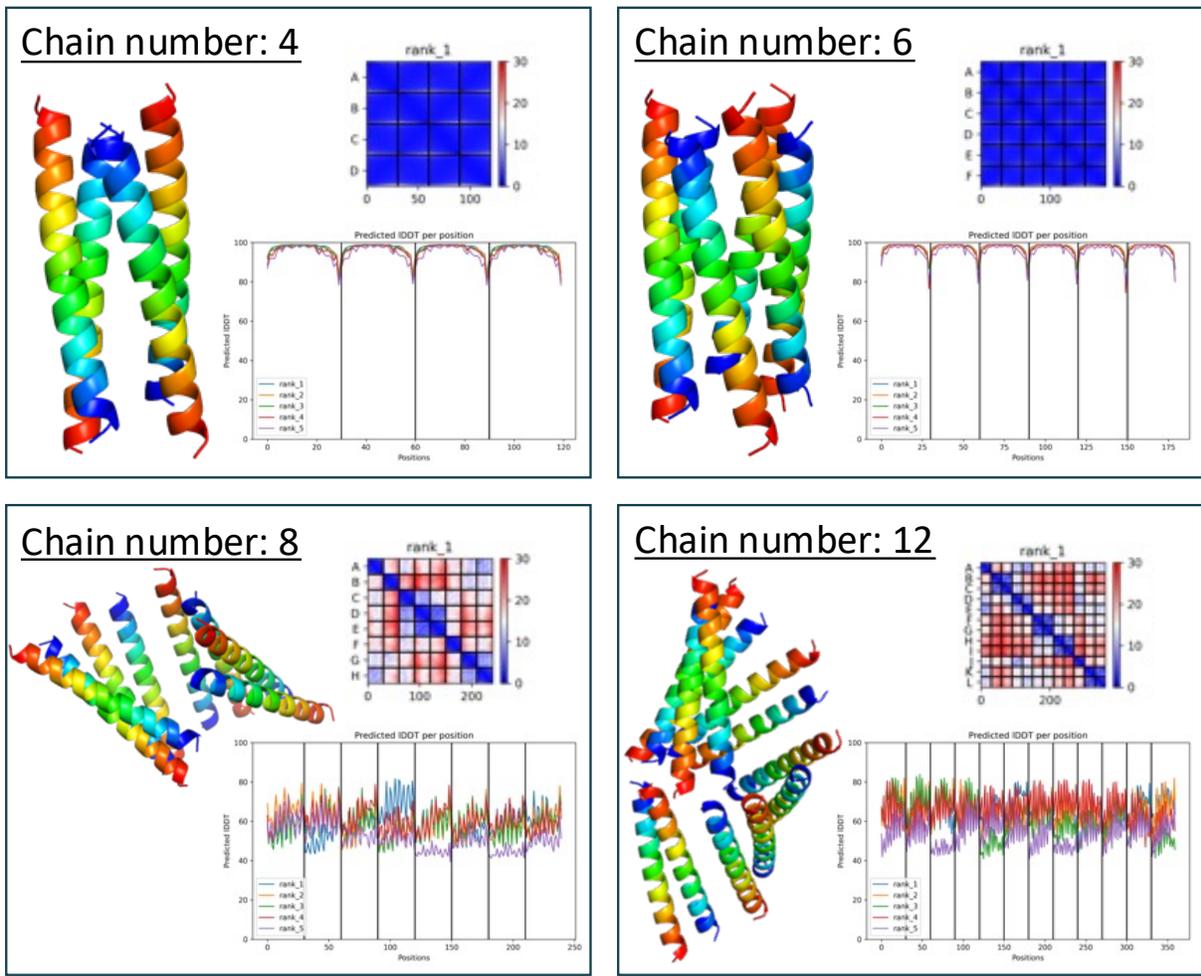
PDB ID	Composition				Reference
	<b>g</b>	<b>a</b>	<b>d</b>	<b>e</b>	
8A3G, 8A3I, 8A3J, 8A3K	Q	L	I	A	This study
1EC5	Q, L, D, W	L, E, Y	L, A, I, H	L, I	Lombardi, et al. (2000) <sup>25</sup>
1MFT, 1U7J, 1U7M, 1Y47, 1JMB, 1JM0	Y, L, W, D	L, E, Y	L, A, I, H	Y, M, S, L, I	Lahr, et al. (2005) <sup>26</sup> Di Costanzo, et al. (2001) <sup>27</sup>
1W5H, 2CCN, 2CCF	K, E, L	L	I	E, L, C (or S), K	Yadav, et al. (2006) <sup>28</sup>
1W5J, 1W5K	R, K, E, L	L	I	E, L, C, R	Yadav, et al. (2006) <sup>28</sup>
2B1F	R, H, E, Q	L	A, L	V, N	Deng, et al. (2006) <sup>29</sup>
2B22	H, V	V, L	L, V	L, N	Deng, et al. (2006) <sup>29</sup>
2NRN	Q, E, H, R	L	V, K, E, N, L	V, N, K, E, Y, A	Deng, et al. (2007) <sup>30</sup>
3S0R	A	L	A	E, Q, R	Grigoryan, et al. (2011) <sup>31</sup>
4HB1	L, A	L, A	A, L	K, Q	Schafmeister, et al. (1997) <sup>32</sup>
5J0K, 6EGC	L, S	L, A, I, V	L, Q, V, I	E, K, R	Boyken, et al. (2016) <sup>33</sup> Chen, et al. (2019) <sup>34</sup>
5J10	L, S, I	V, L, A, I, M	I, Q, L, M	A, K, E, L	Boyken, et al. (2016) <sup>33</sup>
5TGY	A, L, I, V, Q	F, I, L, G	L, I, A, W, F	E, R, A, V	Polizzi, et al. (2017) <sup>35</sup>
5VJS, 5VJU	L, Q, G, T, D	R, A, L, G, E, Y, V	S, H, F, L, A, I, G	Q, L, I, S	Unpublished
6C52	K	L, A	L, A	R, E, Q	Zhang, et al. (2018) <sup>36</sup>
6DLC	L, I, N, R, Q, V, A	L, S, V, F, I, A	F, H, I, A, S, V, L, Y	L, I, V, E, T, N, R	Chen, et al. (2019) <sup>37</sup>
6DMP	L, I, A	L, I, H, V, S	A, Q, L, M, S, I	Q, L, E, K, R	Chen, et al. (2019) <sup>37</sup>
6Q5S, 6Q5R	K, E	L	L	A	Rhys, et al (2019) <sup>38</sup>
6W70	E, V, A, Y, L, K, T, I	F, L, A, G, Y, M, T	V, A, M, L, H, I, F, G	E, A, K, V, L, T	Polizzi, et al. (2020) <sup>39</sup>

## 2.2. Supplementary figures

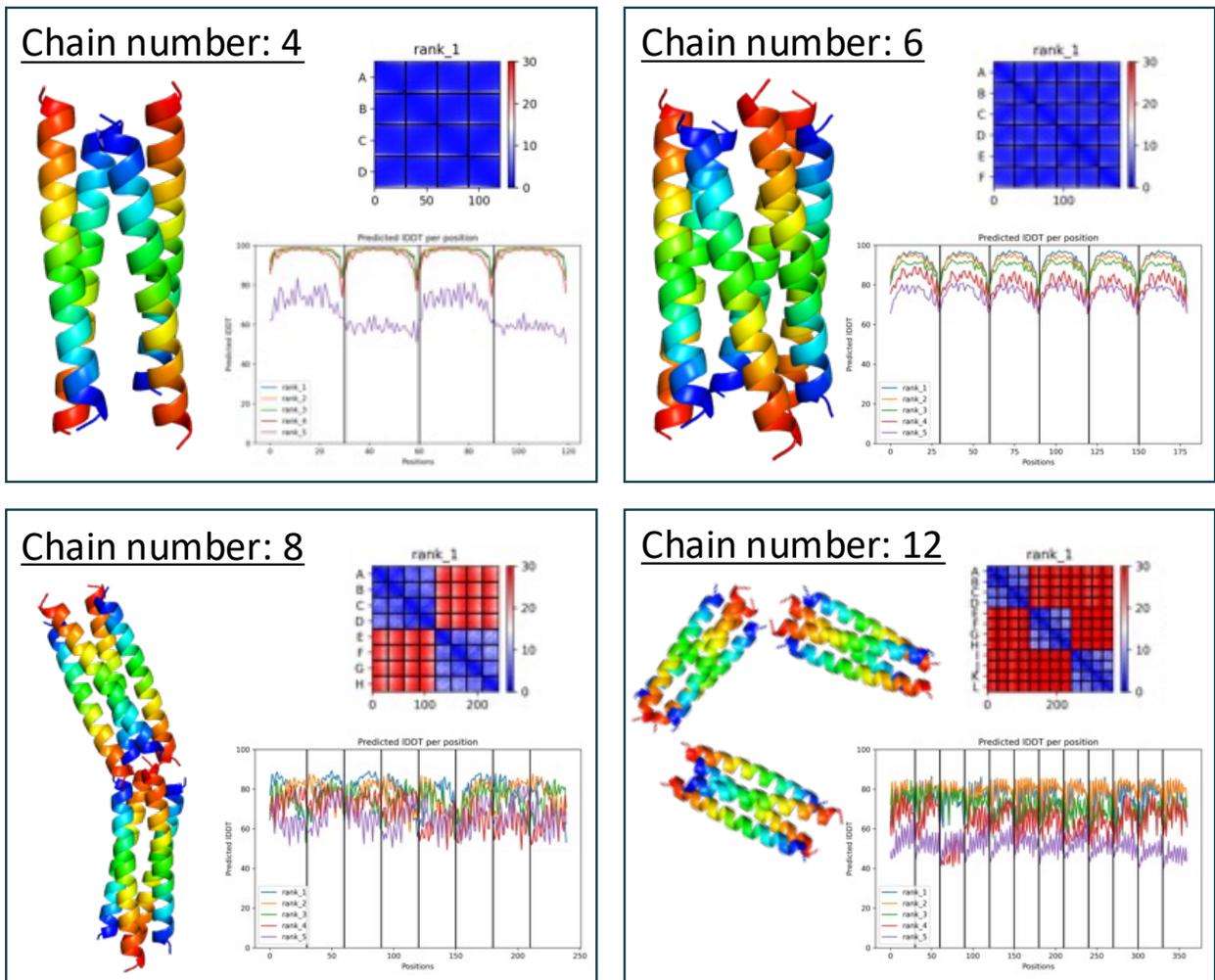
### 2.2.1. Rational and computational design



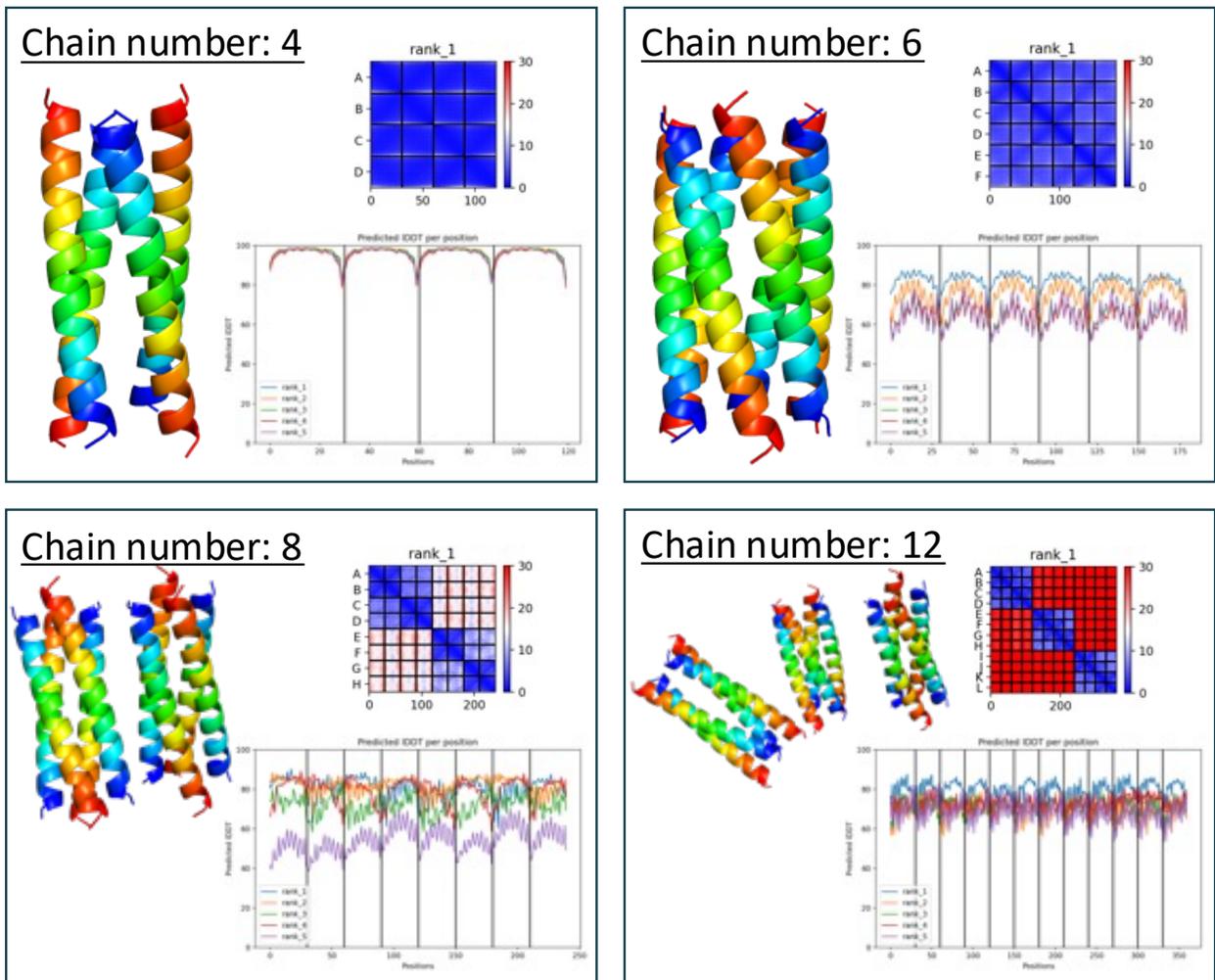
**Figure S1. AlphaFold2 predictions for pLLL.** Assembly predictions for the pLLL sequence with different numbers of chains as the input for AlphaFold2-multimer (ColabFold). Each panel shows the top-ranked prediction (left) with the associated Predicted Alignment Error (PAE) graph (top right) to assess the confidence about the interface, and the predicted IDDT graph (bottom right) to assess model confidence at each position. In this case, AlphaFold2-multimer predicted with high confidence a 6-helix bundle for the pLLL sequence with 6 peptide chains as input.



**Figure S2. AlphaFold2 predictions for pLLI.** Assembly predictions for the pLLI sequence with different numbers of chains as the input for AlphaFold2-multimer (ColabFold). Each panel shows the top-ranked prediction (left) with the associated Predicted Alignment Error (PAE) graph (top right) to assess the confidence about the interface, and the predicted IDDT graph (bottom right) to assess model confidence at each position. In this case, AlphaFold2 predicted with high confidence 4- or 6-helix bundle for pLLI sequence when the requested oligomeric state was 4 or 6, respectively.

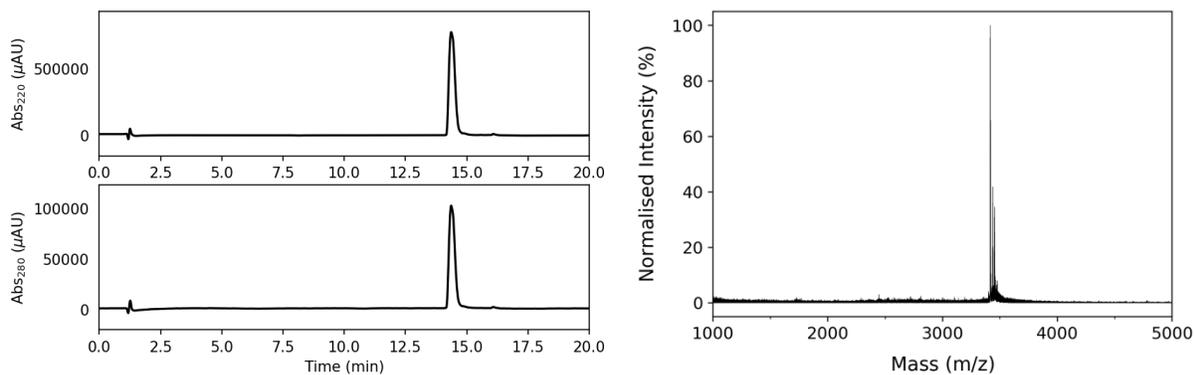


**Figure S3. AlphaFold2 predictions of pQLI (apCC-Tet\*).** Assembly predictions for the pQLI sequence with different numbers of chains as the input for AlphaFold2-multimer (ColabFold). Each panel shows the top-ranked prediction (left) with the associated Predicted Alignment Error (PAE) graph (top right) to assess the confidence about the interface, and the predicted IDDT graph (bottom right) to assess model confidence at each position. In this case, AlphaFold2 slightly predicted better and with high confidence a 4-helix bundle for the pQLI sequence with 4 peptide chains as input. Interestingly, when the requested oligomeric was higher (i.e., 8 or 12), AlphaFold2 predicted the formation of several tetramers with pIDDT around 80%.

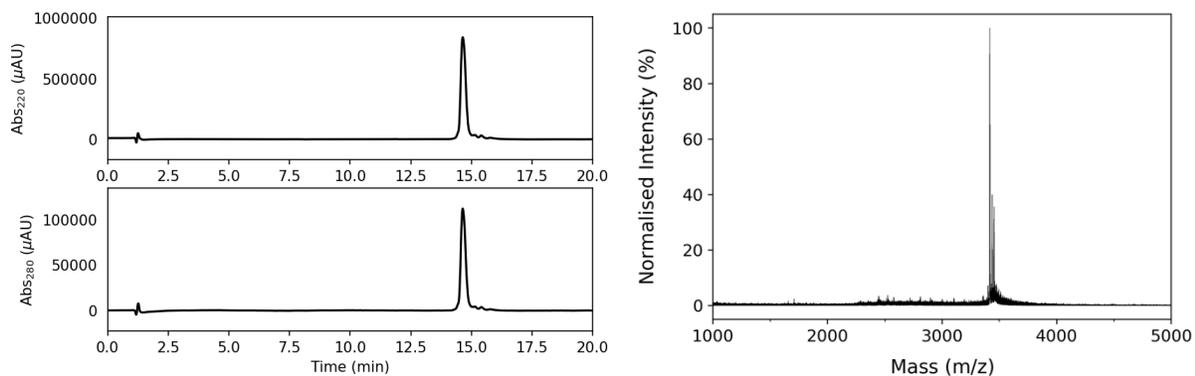


**Figure S4. AlphaFold2 predictions of pQLL** Assembly predictions for the pQLL sequence with different numbers of chains as the input for AlphaFold2-multimer (ColabFold). Each panel shows the top-ranked prediction (left) with the associated Predicted Alignment Error (PAE) graph (top right) to assess the confidence about the interface, and the predicted IDDT graph (bottom right) to assess model confidence at each position. In this case, AlphaFold2 predicted with high confidence a 4-helix bundle for pQLL sequence when the requested oligomeric state was 4. The hexameric assembly was predicted with lower confidence for oligomeric state of 6 as input. Interestingly, when the requested oligomeric was higher (i.e., 8 or 12), AlphaFold2 predicted the formation of several tetramers with pIDDT around 80%.

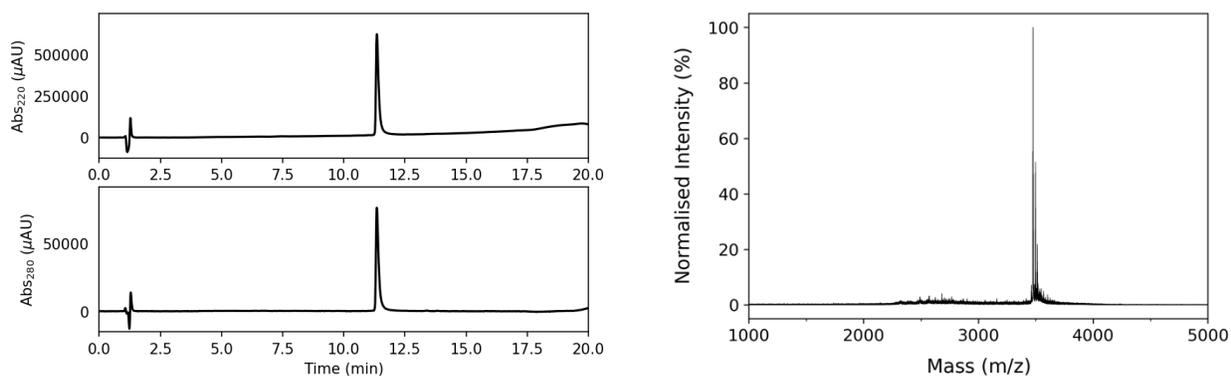
## 2.2.2. Characterization of homopeptides



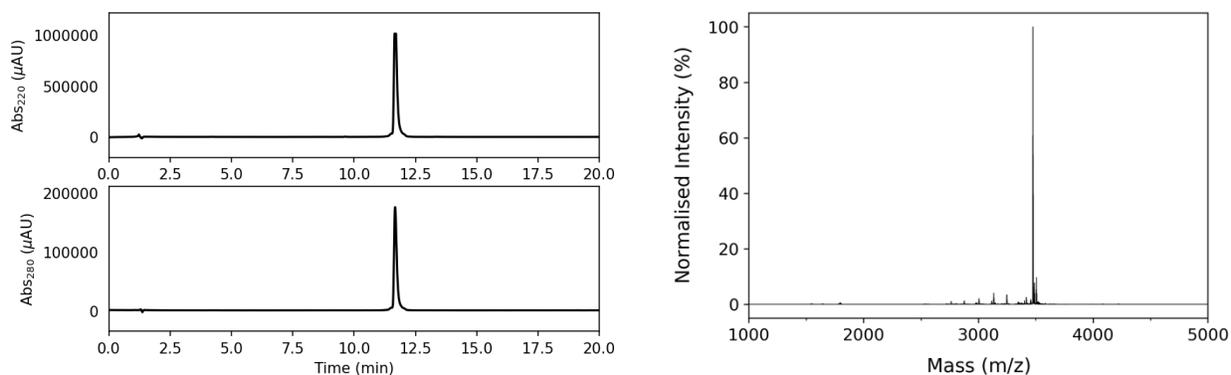
**Figure S5. Characterization of pLLL.** Left: Analytical HPLC chromatogram at 220 nm (top) and 280 nm (bottom) from a gradient of 40 to 100% MeCN (0.1% TFA) in H<sub>2</sub>O (0.1% TFA) with column oven at 50 °C. Right: MALDI-TOF MS. Calculated mass: 3416.2 Da [M+H]<sup>+</sup>. Observed mass: 3417.1 Da [M+H]<sup>+</sup>.



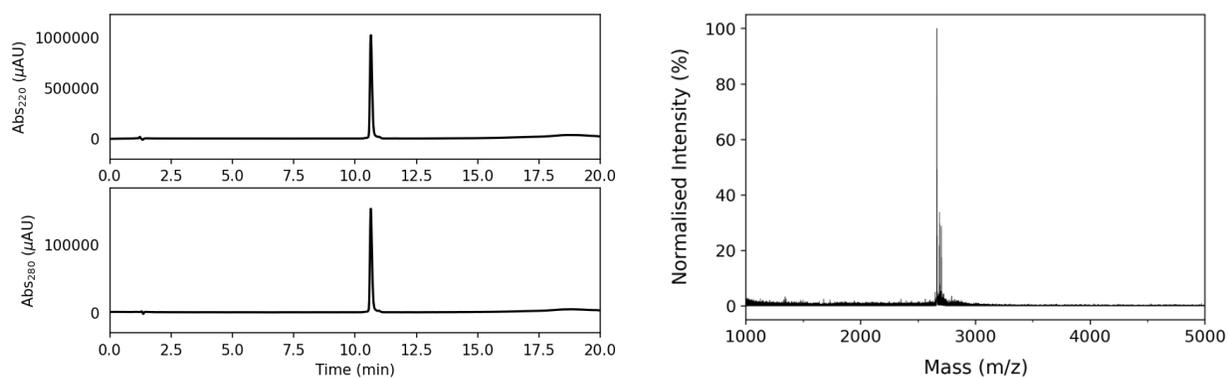
**Figure S6. Characterization of pLLI.** Left: Analytical HPLC chromatogram at 220 nm (top) and 280 nm (bottom) from a gradient of 40 to 100% MeCN (0.1% TFA) in H<sub>2</sub>O (0.1% TFA) with column oven at 50 °C. Right: MALDI-TOF MS. Calculated mass: 3416.2 Da [M+H]<sup>+</sup>. Observed mass: 3417.2 Da [M+H]<sup>+</sup>.



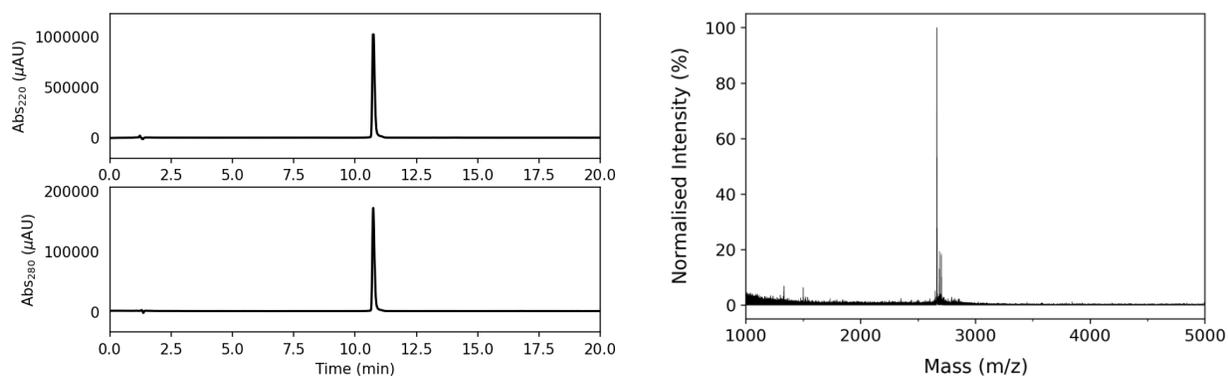
**Figure S7. Characterization of pQLI (apCC-Tet\*).** Left: Analytical HPLC chromatogram at 220 nm (top) and 280 nm (bottom) from a gradient of 20 to 80% MeCN (0.1% TFA) in H<sub>2</sub>O (0.1% TFA). Right: MALDI-TOF MS. Calculated mass: 3476.1 Da [M+H]<sup>+</sup>. Observed mass: 3476.0 Da [M+H]<sup>+</sup>.



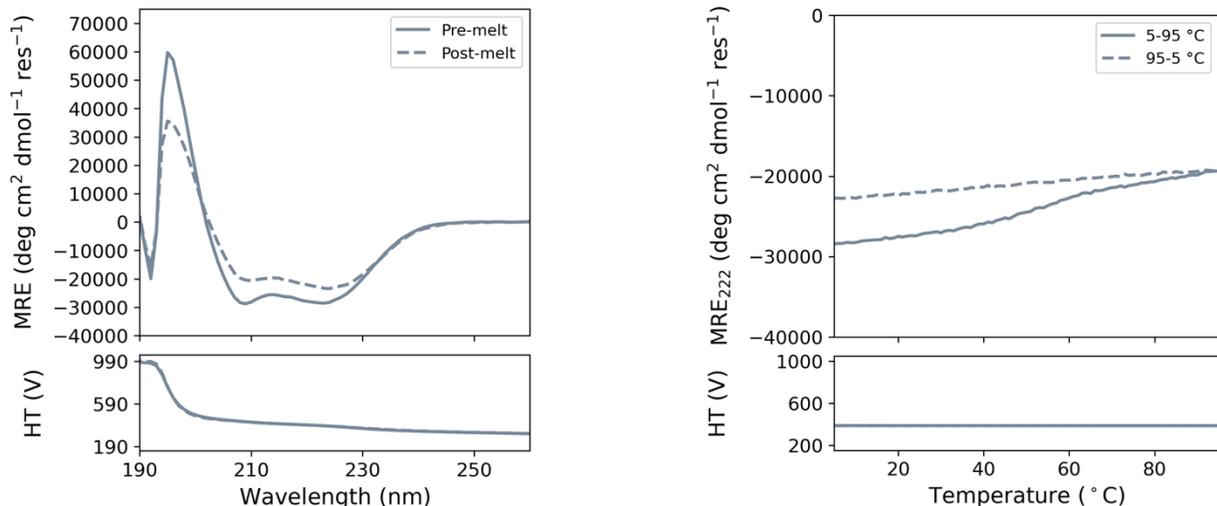
**Figure S8. Characterization of pQLL.** Left: Analytical HPLC chromatogram at 220 nm (top) and 280 nm (bottom) from a gradient of 20 to 80% MeCN (0.1% TFA) in H<sub>2</sub>O (0.1% TFA). Right: Deconvoluted ESI MS. Calculated mass: 3476.1 Da [M+H]<sup>+</sup>. Observed mass: 3476.6 Da [M+H]<sup>+</sup>.



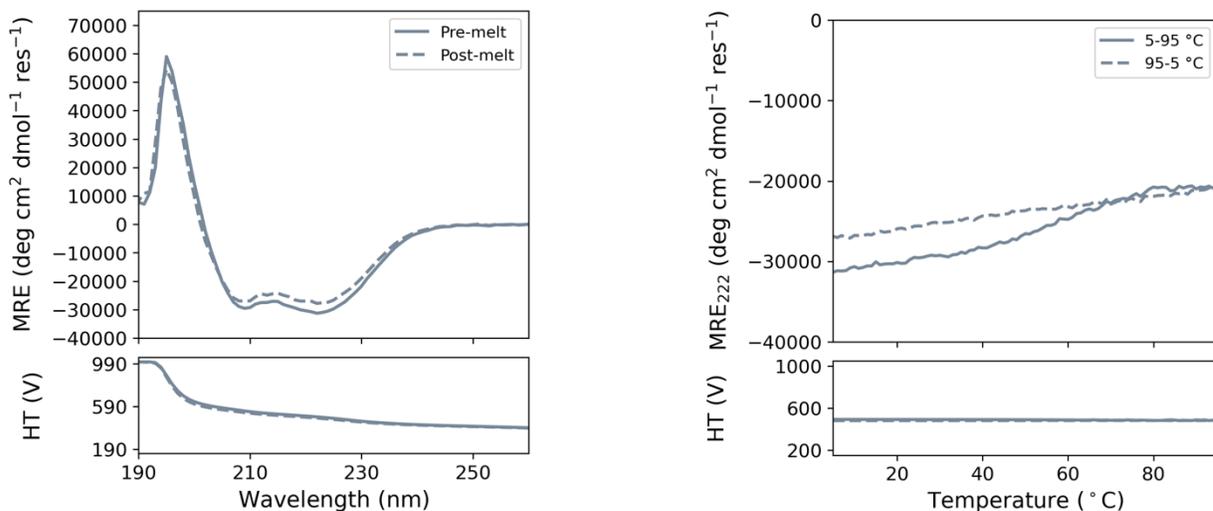
**Figure S9. Characterization of pQLI<sup>3</sup> (apCC-Tet<sup>3</sup>).** Left: Analytical HPLC chromatogram at 220 nm (top) and 280 nm (bottom) from a gradient of 20 to 80% MeCN (0.1% TFA) in H<sub>2</sub>O (0.1% TFA). Right: MALDI-TOF MS. Calculated mass: 2664.1 Da [M+H]<sup>+</sup>. Observed mass: 2664.4 Da [M+H]<sup>+</sup>.



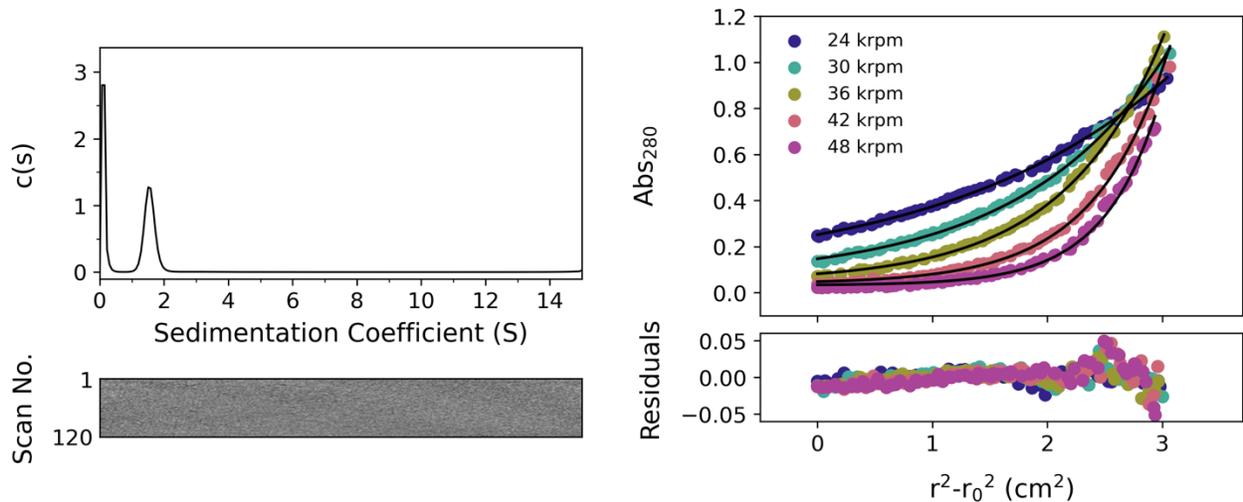
**Figure S10. Characterization of pQLL<sup>3</sup>.** Left: Analytical HPLC chromatogram at 220 nm (top) and 280 nm (bottom) from a gradient of 20 to 80% MeCN (0.1% TFA) in H<sub>2</sub>O (0.1% TFA). Right: MALDI-TOF MS. Calculated mass: 2664.1 Da [M+H]<sup>+</sup>. Observed mass: 2664.4 Da [M+H]<sup>+</sup>.



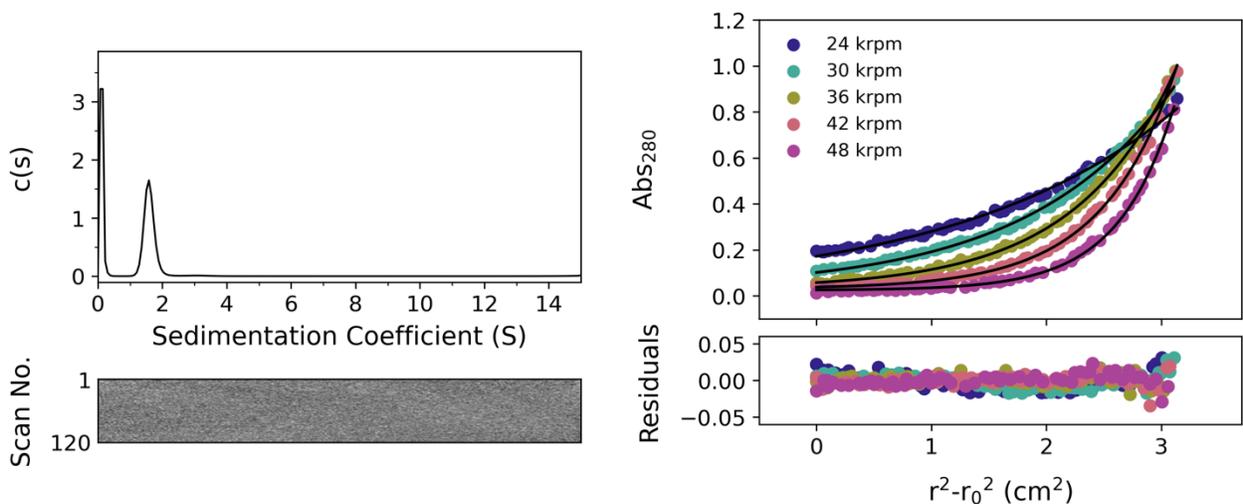
**Figure S11. CD spectroscopy of pLLL.** Left: CD spectra at 5 °C before thermal ramp up (solid line) and after thermal ramp down (dashed line). Right: Thermal profiles (ramping up and ramping down) monitored at 222 nm. Bottom panels show corresponding high-tension (HT) voltages applied to the detector during measurements. Conditions: 50  $\mu$ M peptide concentration, PBS, pH 7.4.



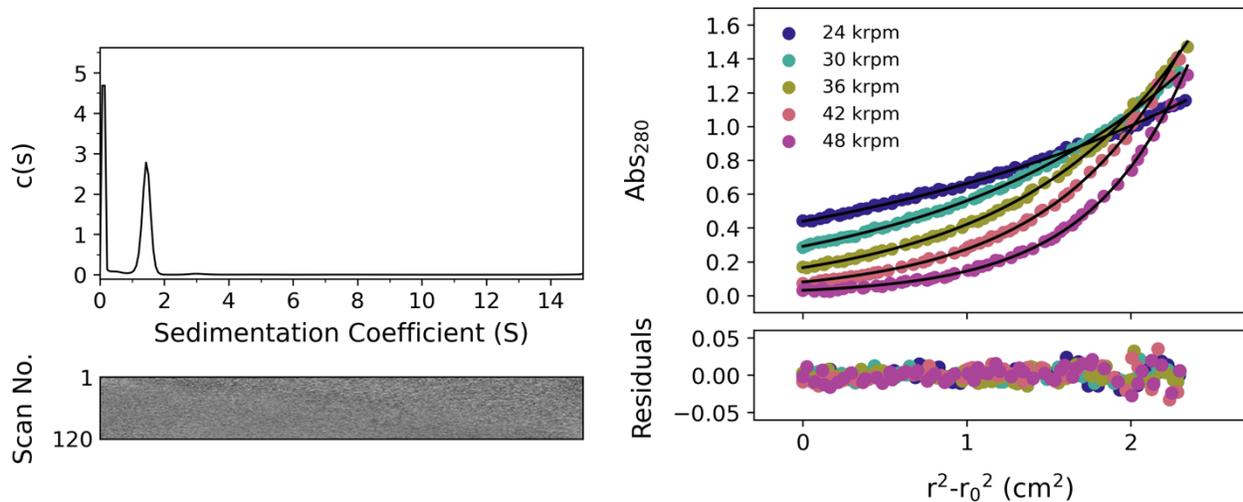
**Figure S12. CD spectroscopy of pLLI.** Left: CD spectra at 5 °C before thermal ramps up (solid line) and after thermal ramps down (dashed line). Right: Thermal profiles (ramping up and ramping down) monitored at 222 nm. Bottom panels show corresponding high-tension (HT) voltages applied to the detector during measurements. Conditions: 50  $\mu$ M peptide concentration, PBS, pH 7.4.



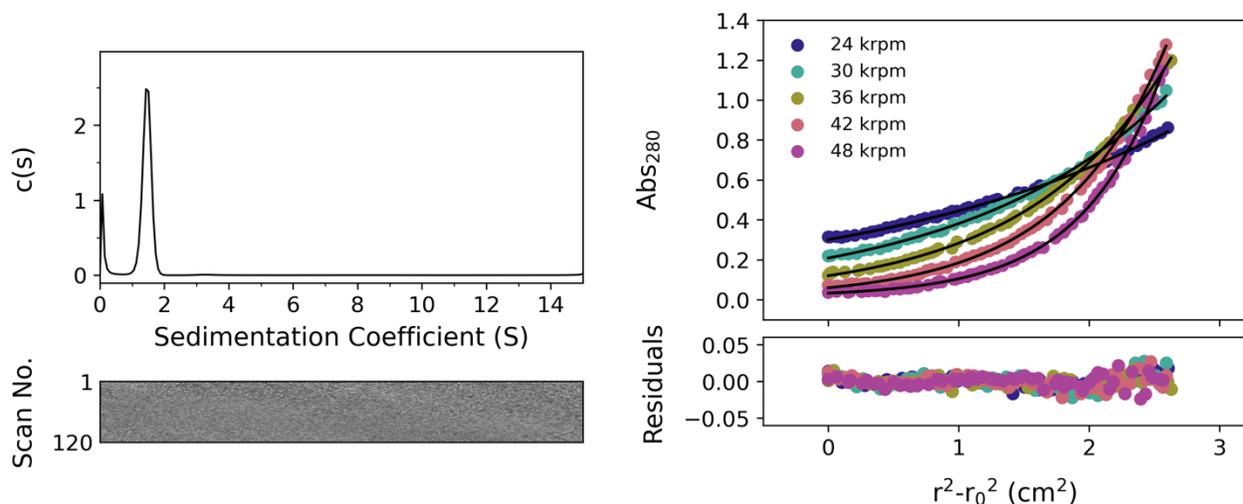
**Figure S13. AUC data for pLLL.** **Left:** Sedimentation velocity (SV) AUC data for pLLL peptide at 50 krpm ( $\bar{v} = 0.799 \text{ cm}^3 \text{ g}^{-1}$ ). Residuals are shown as a bitmap below the fitted data. Continuous  $c(s)$  distribution returned a molecular mass of 20583 Da corresponding to 6.0 x monomer mass at 95% confidence level ( $f/f_0 = 1.205$ ,  $s = 1.556 \text{ S}$ ,  $s_{20,w} = 1.627\text{S}$ ). **Right:** Sedimentation equilibrium (SE) AUC data for pLLL between 24 and 48 krpm at 6 krpm intervals. Fitted single-ideal species model curves are overlaid in black and gave a molecular mass 20049 Da corresponding to 5.9 x monomer mass, 95% confidence limits 19945-20268 Da. Conditions: 150  $\mu\text{M}$  peptide, PBS, pH 7.4, 20  $^\circ\text{C}$ .



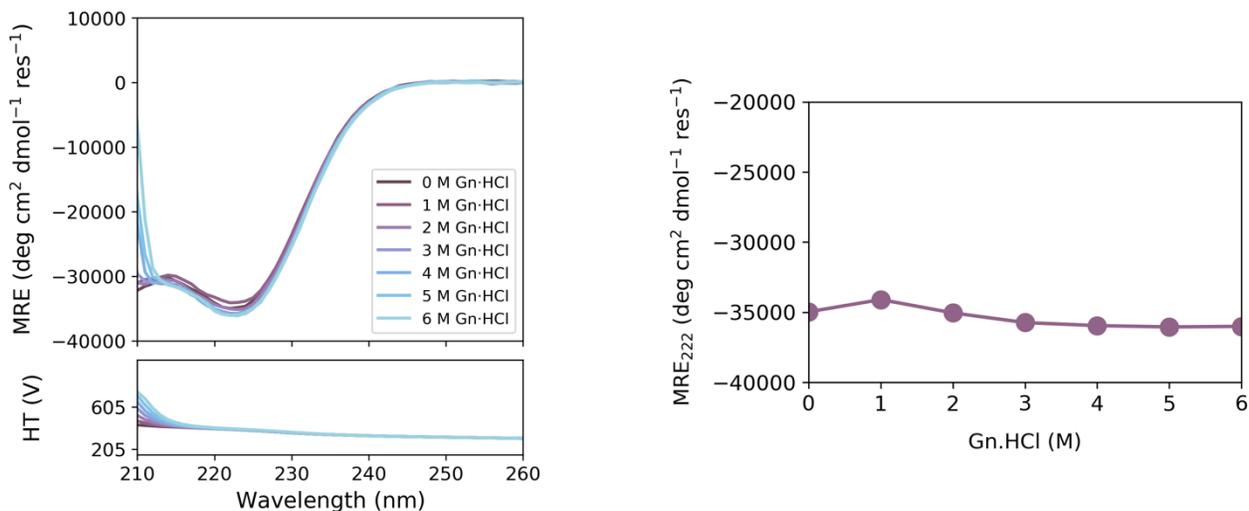
**Figure S14. AUC data for pLLI.** **Left:** Sedimentation velocity (SV) AUC data for pLLI peptide at 50 krpm ( $\bar{v} = 799 \text{ cm}^3 \text{ g}^{-1}$ ). Residuals are shown as a bitmap below the fitted data. Continuous  $c(s)$  distribution returned a molecular mass of 21537 Da corresponding to 6.3 x monomer mass at 95% confidence level ( $f/f_0 = 1.212$ ,  $s = 1.595 \text{ S}$ ,  $s_{20,w} = 1.668 \text{ S}$ ). **Right:** Sedimentation equilibrium (SE) AUC data for pLLI between 24 and 48 krpm at 6 krpm intervals. Fitted single-ideal species model curves are overlaid in black and gave a molecular mass 20079 Da corresponding to 5.9 x monomer mass, 95% confidence limits 19889-20287 Da. Conditions: 150  $\mu\text{M}$  peptide, PBS, pH 7.4, 20  $^\circ\text{C}$ .



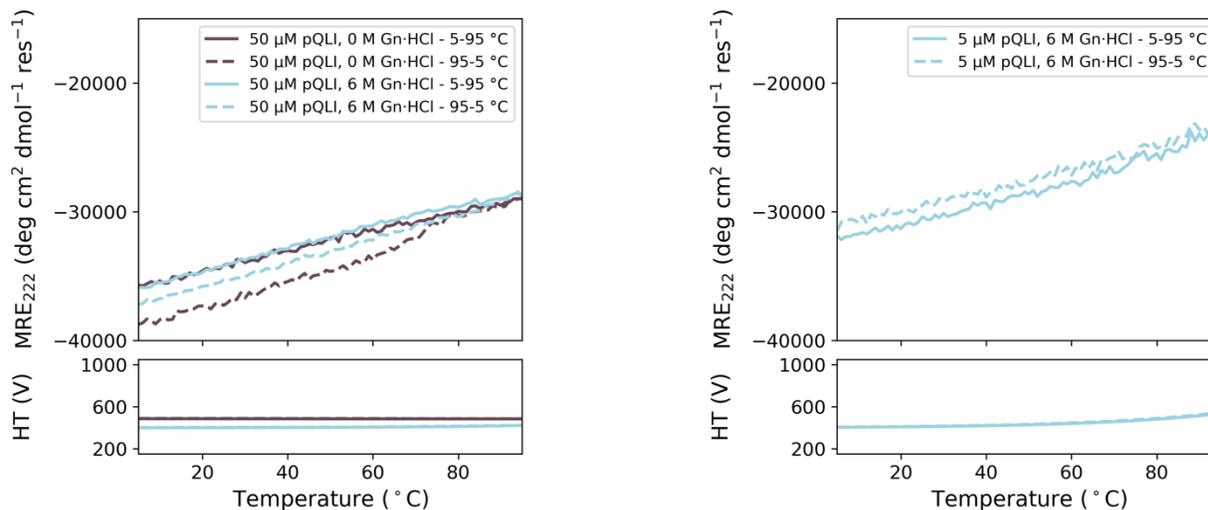
**Figure S15. AUC data for pQLI (apCC-Tet\*).** **Left:** Sedimentation velocity (SV) AUC data for apCC-Tet\* at 50 krpm ( $\bar{v} = 0.766 \text{ cm}^3 \text{ g}^{-1}$ ). Residuals are shown as a bitmap below the fitted data. Continuous  $c(s)$  distribution returned a molecular mass of 15800 Da corresponding to 4.5 x monomer mass at 95% confidence level ( $f/f_0 = 1.298$ ,  $s = 1.434 \text{ S}$ ,  $s_{20,w} = 1.492 \text{ S}$ ). **Right:** Sedimentation equilibrium (SE) AUC data for apCC-Tet\* between 24 and 48 krpm at 6 krpm intervals. Fitted single-ideal species model curves are overlaid in black and gave a molecular mass 14134 Da corresponding to 4.1 x monomer mass, 95% confidence limits 14053-14306 Da. Conditions: 150  $\mu\text{M}$  and 75  $\mu\text{M}$  peptide for SV and SE experiments respectively, PBS, pH 7.4, 20  $^\circ\text{C}$ .



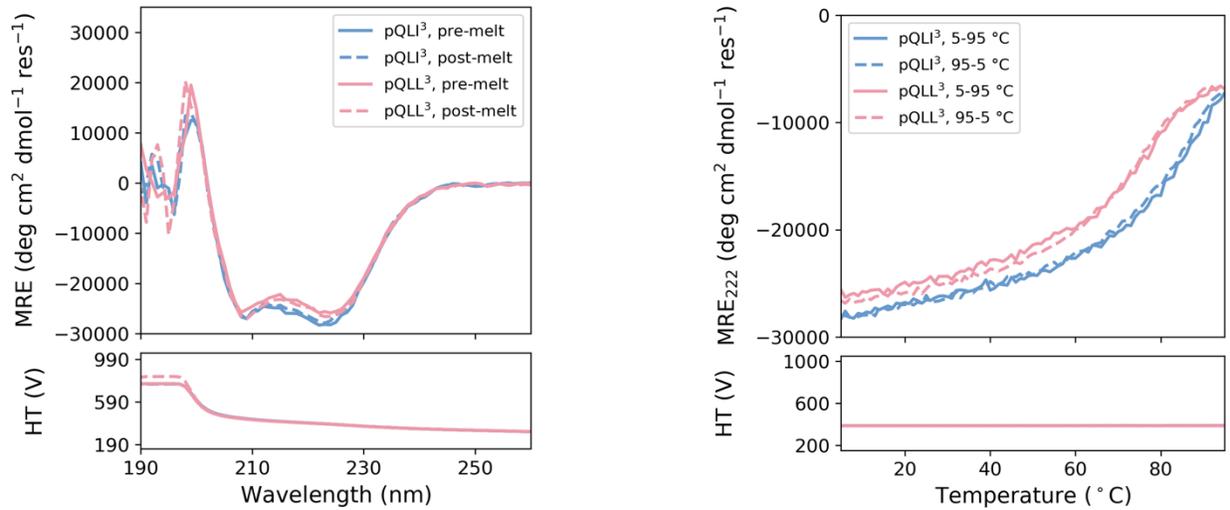
**Figure S16. AUC data for pQLL** **Left:** Sedimentation velocity (SV) AUC data for pQLL peptide at 50 krpm ( $\bar{v} = 0.766 \text{ cm}^3 \text{ g}^{-1}$ ). Residuals are shown as a bitmap below the fitted data. Continuous  $c(s)$  distribution returned a molecular mass of 15064 Da corresponding to 4.3 x monomer mass at 95% confidence level ( $f/f_0 = 1.238$ ,  $s = 1.457 \text{ S}$ ,  $s_{20,w} = 1.516 \text{ S}$ ). **Right:** Sedimentation equilibrium (SE) AUC data for pQLL between 24 and 48 krpm at 6 krpm intervals. Fitted single-ideal species model curves are overlaid in black and gave a molecular mass 13698 Da corresponding to 3.9 x monomer mass, 95% confidence limits 13454-13740 Da. Conditions: 150  $\mu\text{M}$  and 75  $\mu\text{M}$  peptide for SV and SE experiments respectively, PBS, pH 7.4, 20  $^\circ\text{C}$ .



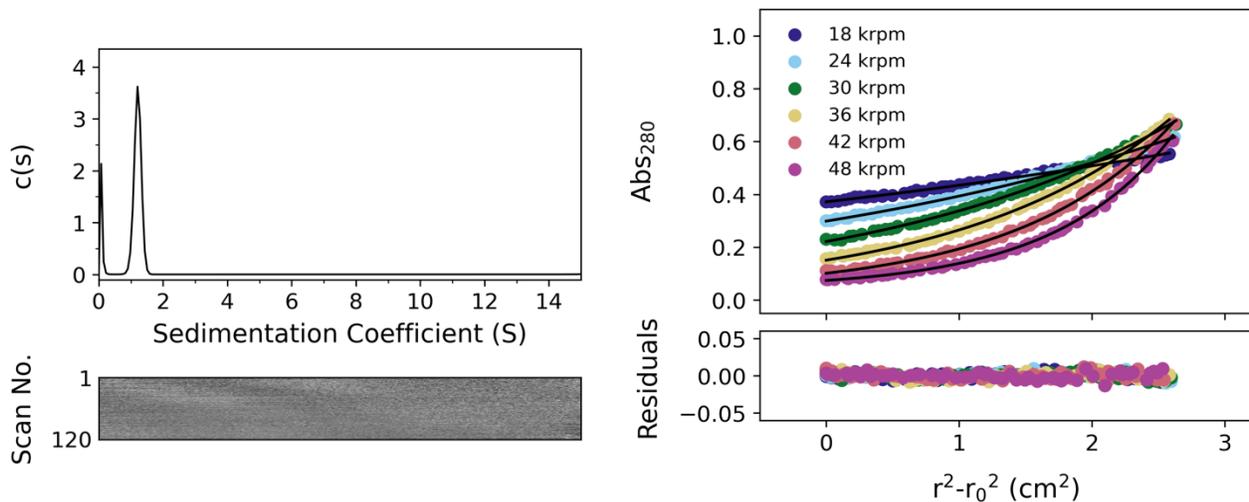
**Figure S17. Chemical denaturation of pQLI (apCC-Tet\*).** Left: CD full scan from 210 to 260 nm at 5 °C of apCC-Tet\* incubated with increasing concentration of guanidinium hydrochloride (Gn·HCl). Right: Chemical denaturation curve measured at 222 nm for apCC-Tet\* (deduced from left spectra). Conditions: 50  $\mu$ M peptide concentration,  $\pm$  Gn·HCl, PBS, pH 7.4 and incubation overnight before measurements.



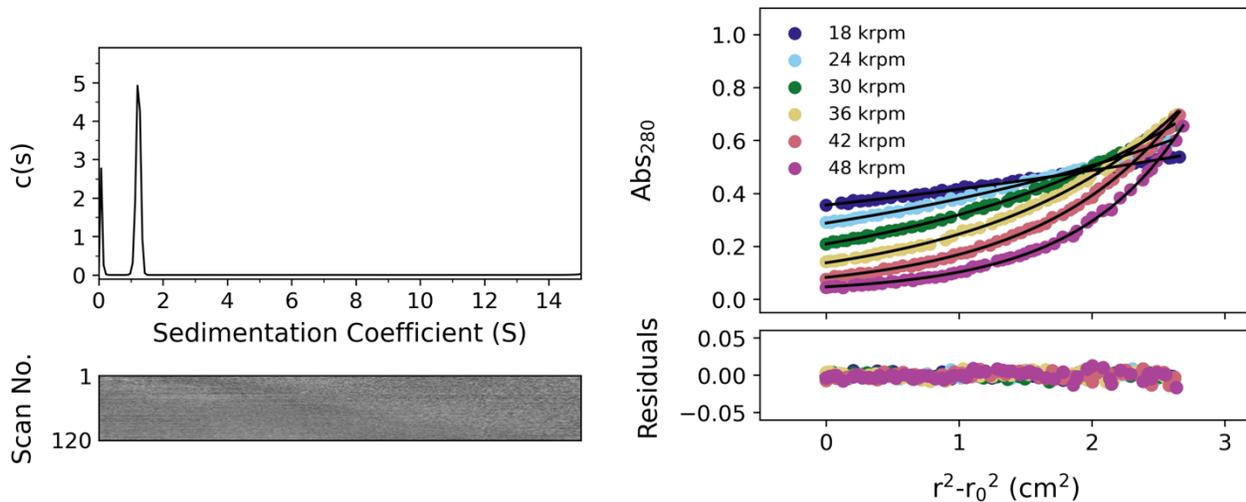
**Figure S18. Temperature profile of pQLI (apCC-Tet\*) in presence of high concentration of denaturant.** Left: Thermal profiles (ramping up and ramping down) monitored at 222 nm for 50  $\mu$ M of apCC-Tet\* with or without 6 M Gn·HCl. Right: Thermal profiles (ramping up and ramping down) monitored at 222 nm in presence of 6 M Gn·HCl for lower concentration of apCC-Tet\* (5  $\mu$ M). Peptide samples were incubated overnight with guanidinium before measurements.



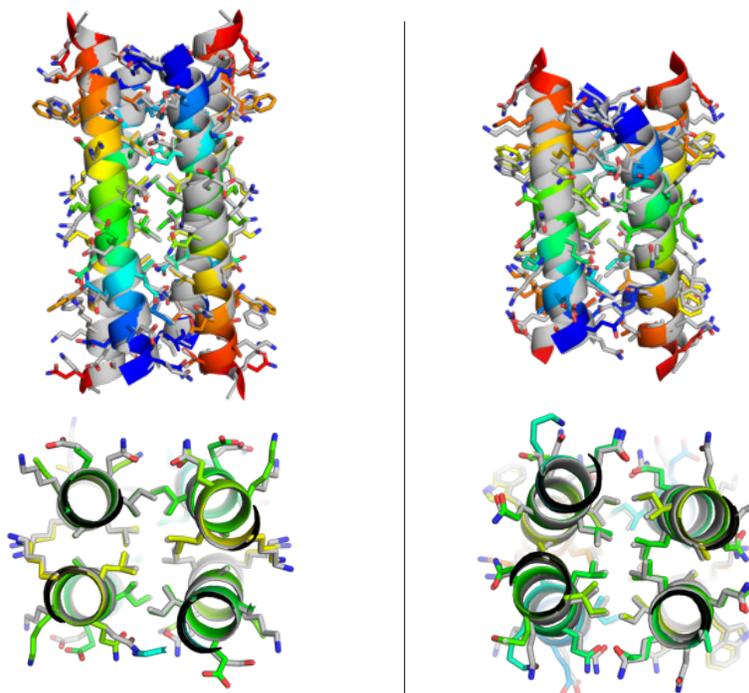
**Figure S19. CD spectroscopy of 3-heptad homotetramers pQLI<sup>3</sup> (apCCTet<sup>\*3</sup>) and pQLL<sup>3</sup> at low concentration.** **Left:** CD spectra at 5 °C before thermal ramp up (solid line) and after thermal ramp down (dashed line) for pQLI<sup>3</sup> (pink) and pQLL<sup>3</sup> (blue). **Right:** Thermal profiles (ramping up and ramping down) monitored at 222 nm. Bottom panels show corresponding high-tension (HT) voltages applied to the detector during measurements. Conditions: 5  $\mu$ M peptide concentration, PBS, pH 7.4.



**Figure S20. AUC data for pQLI<sup>3</sup> (apCC-Tet<sup>\*3</sup>).** **Left:** Sedimentation velocity (SV) AUC data for apCC-Tet<sup>\*3</sup> at 50 krpm ( $\bar{v} = 0.761 \text{ cm}^3 \text{ g}^{-1}$ ). Residuals are shown as a bitmap below the fitted data. Continuous  $c(s)$  distribution returned a molecular mass of 11431 Da corresponding to 4.3 x monomer mass at 95% confidence level ( $f/f_0 = 1.273$ ,  $s = 1.211 \text{ S}$ ,  $s_{20,w} = 1.259 \text{ S}$ ). **Right:** Sedimentation equilibrium (SE) AUC data for apCC-Tet<sup>\*3</sup> between 18 and 48 krpm at 6 krpm intervals. Fitted single-ideal species model curves are overlaid in black and gave a molecular mass 9217 Da corresponding to 3.5 x monomer mass, 95% confidence limits 9052-9298 Da. Conditions: 150  $\mu$ M and 75  $\mu$ M peptide for SV and SE experiments respectively, PBS, pH 7.4, 20 °C.



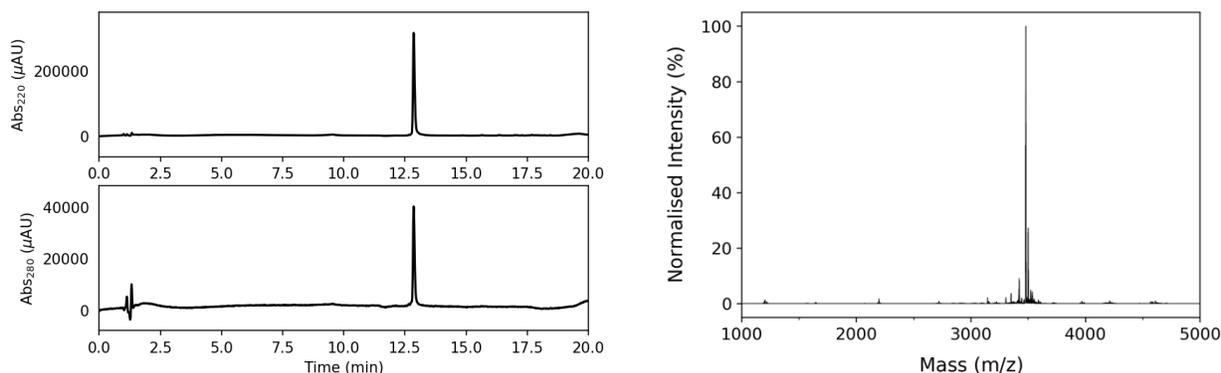
**Figure S21. AUC data for pQLL<sup>3</sup>.** **Left:** Sedimentation velocity (SV) AUC data for pQLL<sup>3</sup> at 50 krpm ( $\bar{v} = 0.761 \text{ cm}^3 \text{ g}^{-1}$ ). Residuals are shown as a bitmap below the fitted data. Continuous  $c(s)$  distribution returned a molecular mass of 10758 Da corresponding to 4.0 x monomer mass at 95% confidence level ( $f/f_0 = 1.205$ ,  $s = 1.229 \text{ S}$ ,  $s_{20,w} = 1.277 \text{ S}$ ). **Right:** Sedimentation equilibrium (SE) AUC data for pQLL<sup>3</sup> between 18 and 48 krpm at 6 krpm intervals. Fitted single-ideal species model curves are overlaid in black and gave a molecular mass 10152 Da corresponding to 3.8 x monomer mass, 95% confidence limits 10039-10227 Da. Conditions: 150  $\mu\text{M}$  and 75  $\mu\text{M}$  peptide for SV and SE experiments respectively, PBS, pH 7.4, 20 °C.



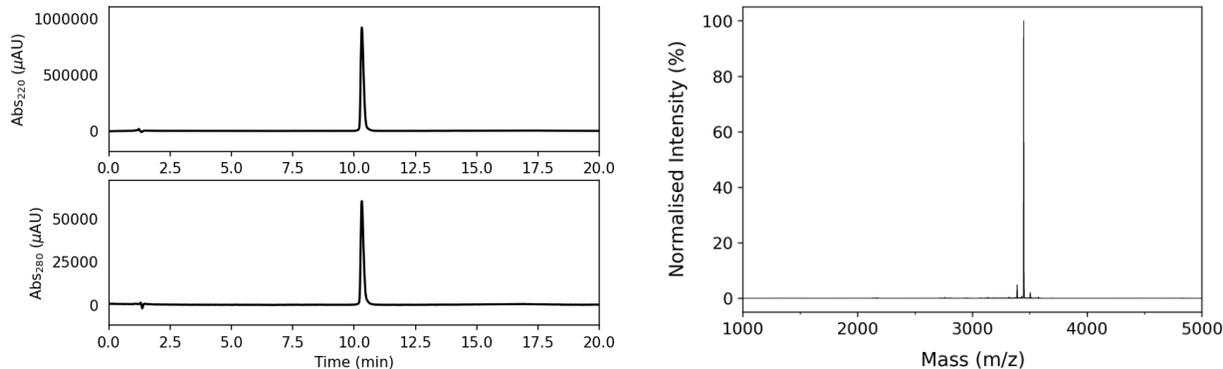
**Figure S22. Structural overlay of crystal structures and related AlphaFold2 predictions for pQLI (apCC-Tet\*) analogues.** **Left:** Overlay of the crystal structure of pQLI (apCC-Tet\*, chainbow colouring) and the corresponding AlphaFold2 prediction (grey) with a

$\text{RMSD}_{\text{all-atom}} = 0.359 \text{ \AA}$  and  $\text{C}\alpha \text{ RMSD} = 0.337 \text{ \AA}$ . **Right:** Overlay of the crystal structure of pQLI<sup>3</sup> (apCC-Tet<sup>\*3</sup>, chainbow colouring) and the corresponding AlphaFold2 prediction (grey) with a  $\text{RMSD}_{\text{all-atom}} = 0.584 \text{ \AA}$  and  $\text{C}\alpha \text{ RMSD} = 0.590 \text{ \AA}$ .

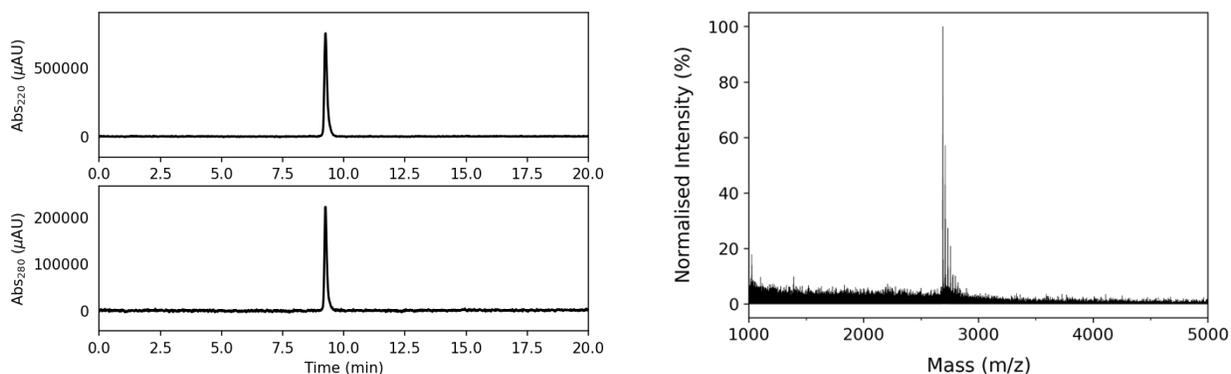
### 2.2.3. Characterization of heterotetramers



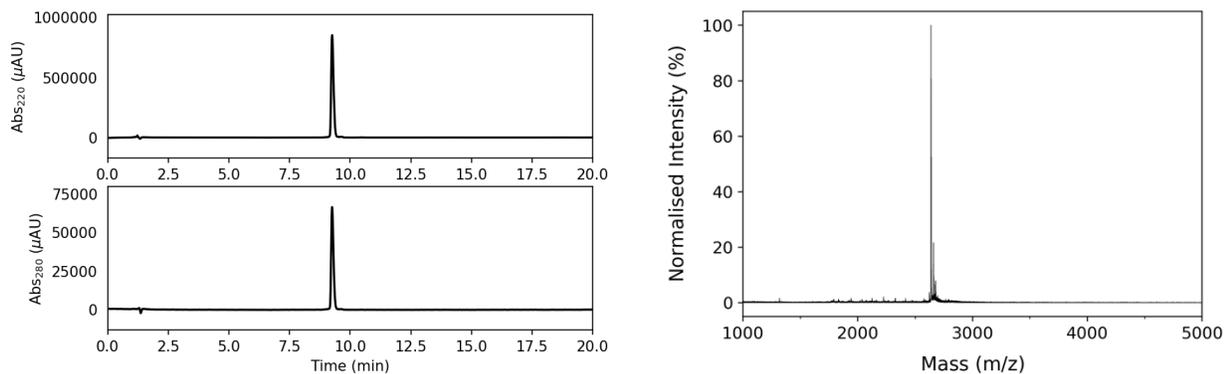
**Figure S23. Characterization of apCC-Tet<sup>\*-A</sup>.** Left: Analytical HPLC chromatogram at 220 nm (top) and 280 nm (bottom) from a gradient of 20 to 80% MeCN in 25 mM NH<sub>4</sub>HCO<sub>3</sub> in H<sub>2</sub>O. Right: Deconvoluted ESI MS, negative mode. Calculated mass: 3478.9 Da [M]. Observed mass: 3479.8 Da [M].



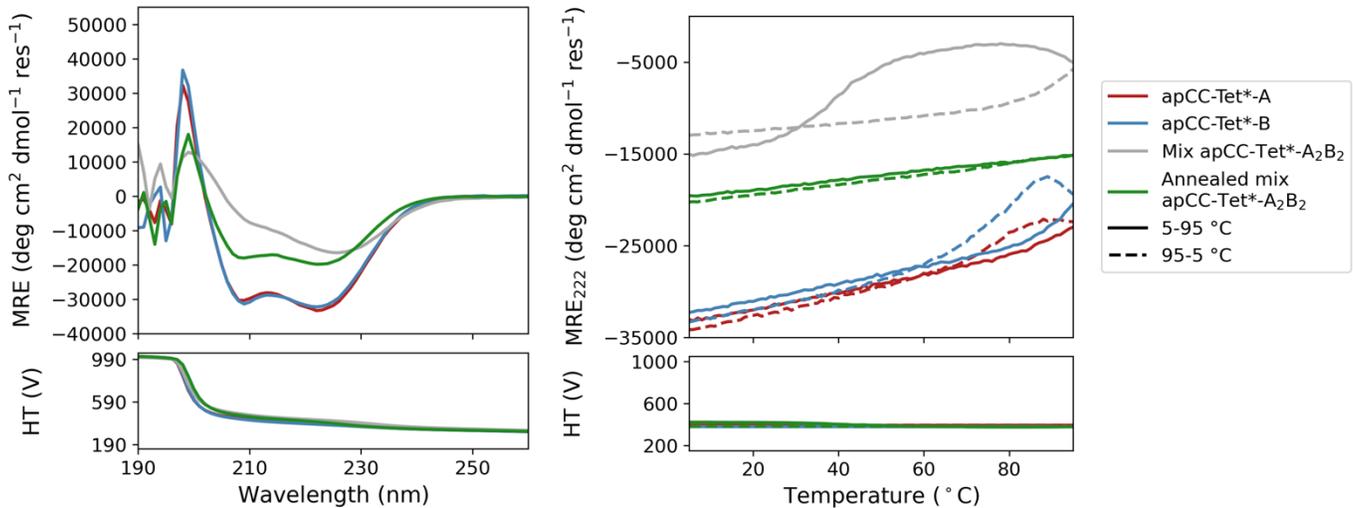
**Figure S24. Characterization of apCC-Tet<sup>\*-B</sup>.** Left: Analytical HPLC chromatogram at 220 nm (top) and 280 nm (bottom) from a gradient of 20 to 80% MeCN (0.1% TFA) in H<sub>2</sub>O (0.1% TFA). Right: Deconvoluted ESI MS. Calculated mass: 3449.3 Da [M+H]<sup>+</sup>. Observed mass: 3448.8 Da [M+H]<sup>+</sup>.



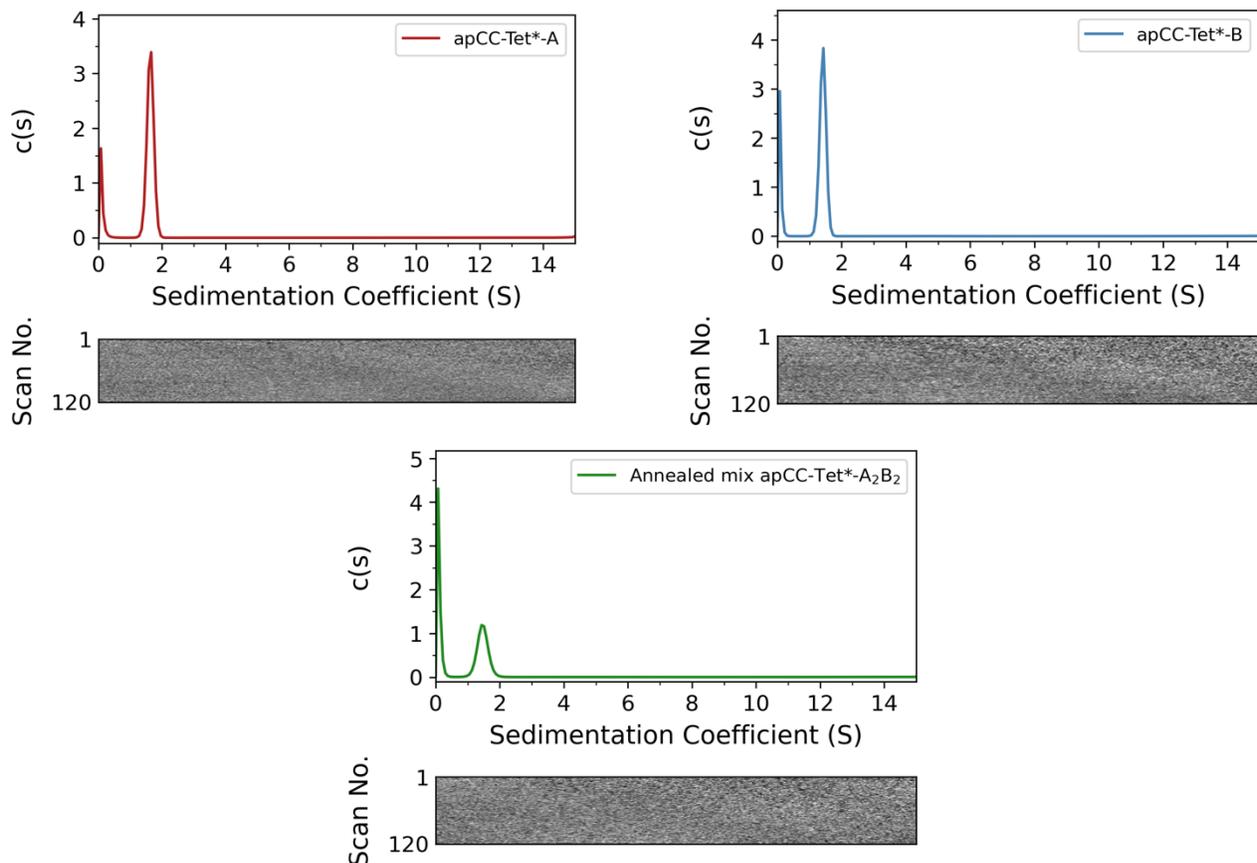
**Figure S25. Characterization of apCC-Tet\*<sup>3</sup>-A.** Left: Analytical HPLC chromatogram at 220 nm (top) and 280 nm (bottom) from a gradient of 20 to 80% MeCN in 25 mM NH<sub>4</sub>HCO<sub>3</sub> in H<sub>2</sub>O. Right: MALDI-TOF MS. Calculated mass: 2668.0 Da [M+H]<sup>+</sup>, 2690.0 Da [M+Na]<sup>+</sup>. Observed mass: 2689.9 Da [M+Na]<sup>+</sup>.



**Figure S26. Characterization of apCC-Tet\*<sup>3</sup>-B.** Left: Analytical HPLC chromatogram at 220 nm (top) and 280 nm (bottom) from a gradient of 20 to 80% MeCN (0.1% TFA) in H<sub>2</sub>O (0.1% TFA). Right: MALDI-TOF MS. Calculated mass: 2639.3 Da [M+H]<sup>+</sup>. Observed mass: 2639.4 Da [M+H]<sup>+</sup>.

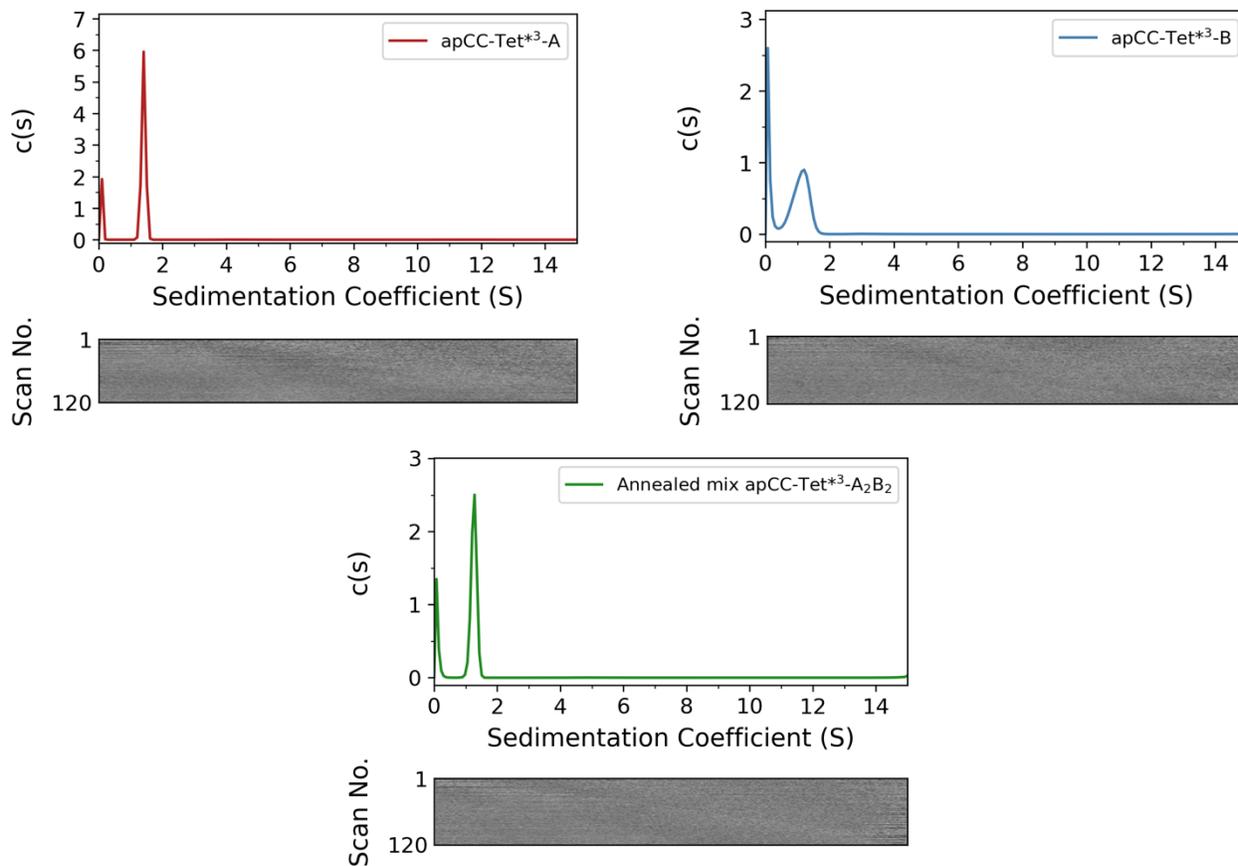


**Figure S27. CD spectroscopy of isolated peptides and mixture for apCC-Tet\*-A<sub>2</sub>B<sub>2</sub>.** Left: CD spectra at 5 °C before thermal ramp up. Right: Thermal profiles (ramping up and ramping down) monitored at 222 nm. Bottom panels show corresponding high-tension (HT) voltages applied to the detector during measurements. Conditions: 10  $\mu$ M peptide concentration, PBS (pH 6.5).

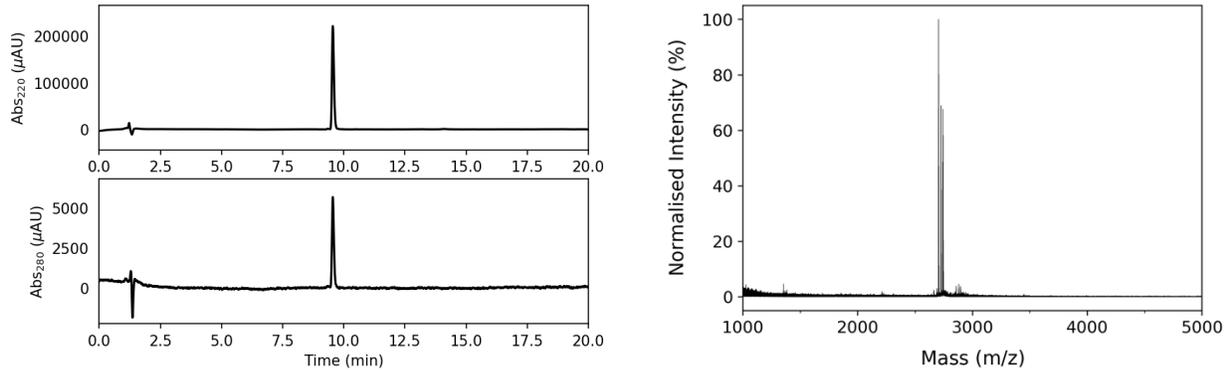


**Figure S28. AUC data isolated peptides and mixture for apCC-Tet\*-A<sub>2</sub>B<sub>2</sub>.** Top left: Sedimentation velocity (SV) AUC data for apCC-Tet\*-A at 50 krpm ( $\bar{v} = 0.742 \text{ cm}^3 \text{ g}^{-1}$ ). Continuous

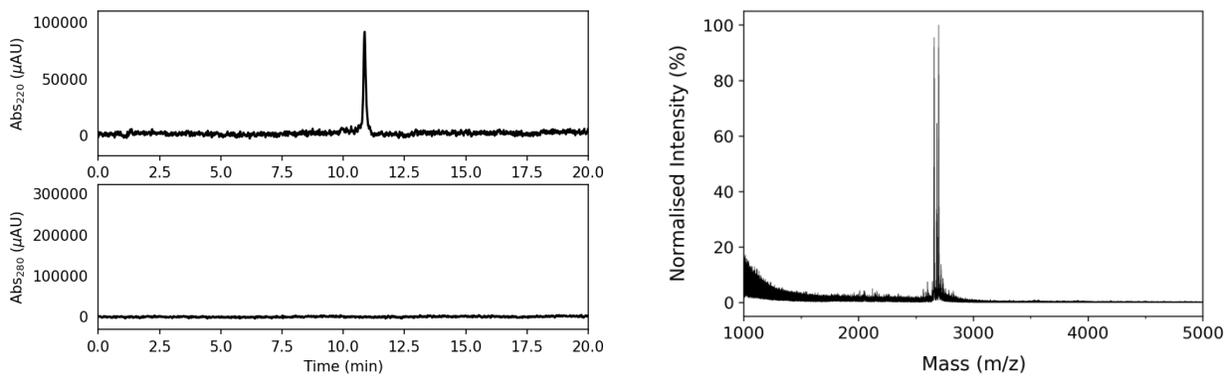
c(s) distribution returned a molecular mass of 15128 Da corresponding to 4.3 x monomer mass at 95% confidence level ( $f/f_0 = 1.238$ ,  $s = 1.632$  S,  $s_{20,w} = 1.693$  S). **Top right:** Sedimentation velocity (SV) AUC data for apCC-Tet<sup>\*3</sup>-B at 50 krpm ( $\bar{v} = 0.789$  cm<sup>3</sup> g<sup>-1</sup>). Continuous c(s) distribution returned a molecular mass of 16248 Da corresponding to 4.7 x monomer mass at 95% confidence level ( $f/f_0 = 1.192$ ,  $s = 1.416$  S,  $s_{20,w} = 1.478$  S). **Bottom middle:** Sedimentation velocity (SV) AUC data for apCC-Tet<sup>\*3</sup>-A<sub>2</sub>B<sub>2</sub> at 50 krpm ( $\bar{v} = 0.766$  cm<sup>3</sup> g<sup>-1</sup>). Continuous c(s) distribution returned a molecular mass of 14030 Da corresponding to 4.1 x average monomer mass at 95% confidence level ( $f/f_0 = 1.176$ ,  $s = 1.467$  S,  $s_{20,w} = 1.525$  S). Conditions: 150 μM peptide for SV experiments, PBS, pH 6.8, 20 °C. Residuals are shown as a bitmap below the fitted data.



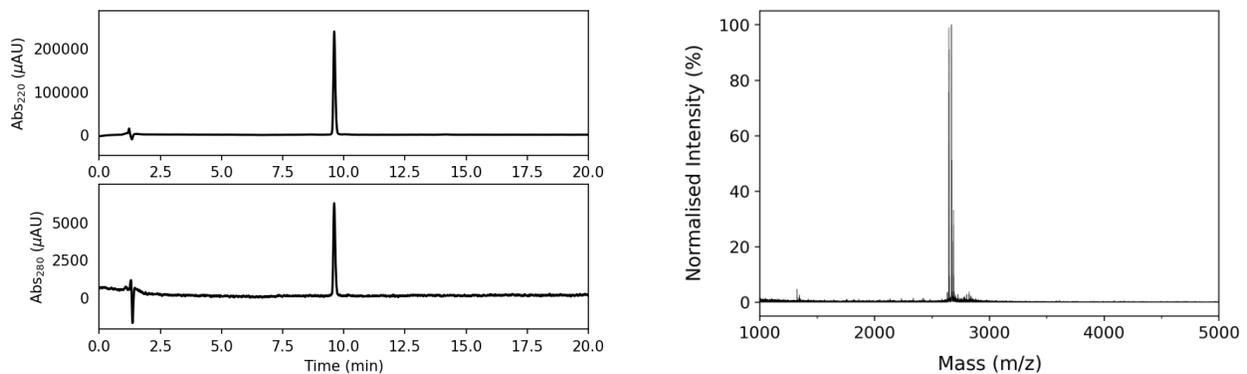
**Figure S29. AUC data isolated peptides and mixture for apCC-Tet<sup>\*3</sup>-A<sub>2</sub>B<sub>2</sub>.** **Top left:** Sedimentation velocity (SV) AUC data for apCC-Tet<sup>\*3</sup>-A at 50 krpm ( $\bar{v} = 0.744$  cm<sup>3</sup> g<sup>-1</sup>). Continuous c(s) distribution returned a molecular mass of 11362 Da corresponding to 4.3 x monomer mass at 95% confidence level ( $f/f_0 = 1.179$ ,  $s = 1.406$  S,  $s_{20,w} = 1.459$  S). **Top right:** Sedimentation velocity (SV) AUC data for apCC-Tet<sup>\*3</sup>-B at 50 krpm ( $\bar{v} = 0.790$  cm<sup>3</sup> g<sup>-1</sup>). Continuous c(s) distribution returned a molecular mass of 10869 Da corresponding to 4.1 x monomer mass at 95% confidence level ( $f/f_0 = 1.180$ ,  $s = 1.092$  S,  $s_{20,w} = 1.140$  S). **Bottom middle:** Sedimentation velocity (SV) AUC data for apCC-Tet<sup>\*3</sup>-A<sub>2</sub>B<sub>2</sub> at 50 krpm ( $\bar{v} = 0.767$  cm<sup>3</sup> g<sup>-1</sup>). Continuous c(s) distribution returned a molecular mass of 10660 Da corresponding to 4.0 x average monomer mass at 95% confidence level ( $f/f_0 = 1.135$ ,  $s = 1.259$  S,  $s_{20,w} = 1.310$  S). Conditions: 150 μM peptide for SV experiments, PBS, pH 7.4, 20 °C. Residuals are shown as a bitmap below the fitted data.



**Figure S30. Characterization of apCC-Tet<sup>\*3</sup>-B-c4CNF.** Left: Analytical HPLC chromatogram at 220 nm (top) and 280 nm (bottom) from a gradient of 20 to 80% MeCN (0.1% TFA) in H<sub>2</sub>O (0.1% TFA). Right: MALDI-TOF MS. Calculated mass: 2705.4 Da [M+H]<sup>+</sup>. Observed mass: 2705.5 Da [M+H]<sup>+</sup>.

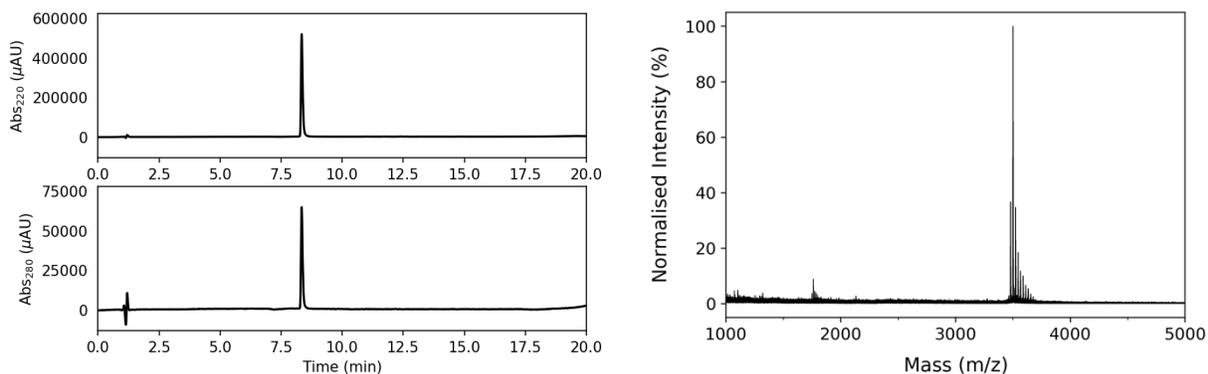


**Figure S31. Characterization of apCC-Tet<sup>\*3</sup>-A-nMSE.** Left: Analytical HPLC chromatogram at 220 nm (top) and 280 nm (bottom) from a gradient of 20 to 80% MeCN in 25 mM NH<sub>4</sub>HCO<sub>3</sub> in H<sub>2</sub>O. No chromophore was present in the sequence of apCC-Tet<sup>\*3</sup>-A-nMSE to absorb at 280 nm. Right: MALDI-TOF MS. Calculated mass: 2658.9 Da [M+H]<sup>+</sup>. Observed mass: 2659.0 Da [M+H]<sup>+</sup>.

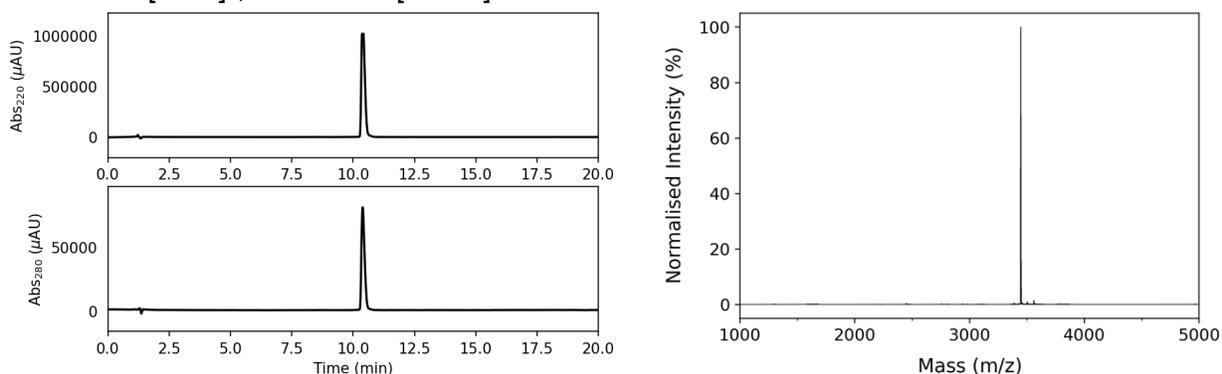


**Figure S32. Characterization of apCC-Tet<sup>\*3</sup>-B-n4CNF.** Left: Analytical HPLC chromatogram at 220 nm (top) and 280 nm (bottom) from a gradient of 20 to 80% MeCN (0.1% TFA) in H<sub>2</sub>O (0.1% TFA).

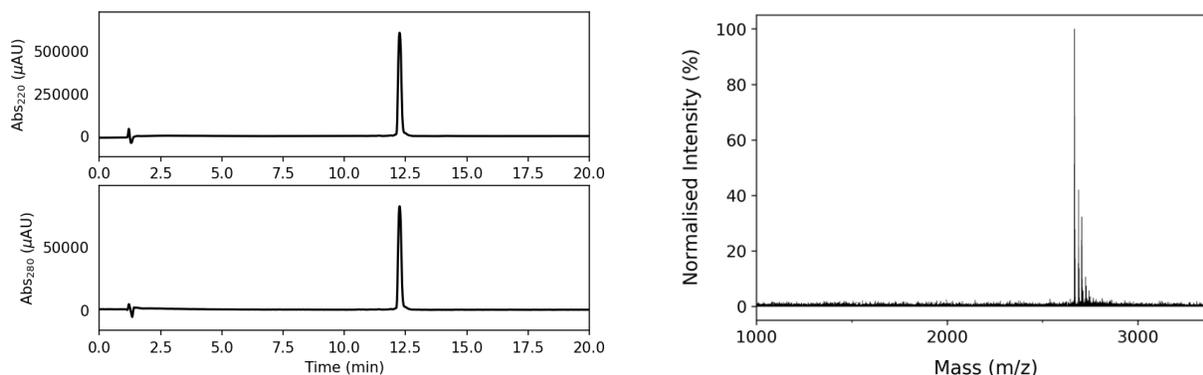
TFA). **Right:** MALDI-TOF MS. Calculated mass: 2648.3 Da  $[M+H]^+$ . Observed mass: 2648.6 Da  $[M+H]^+$ .



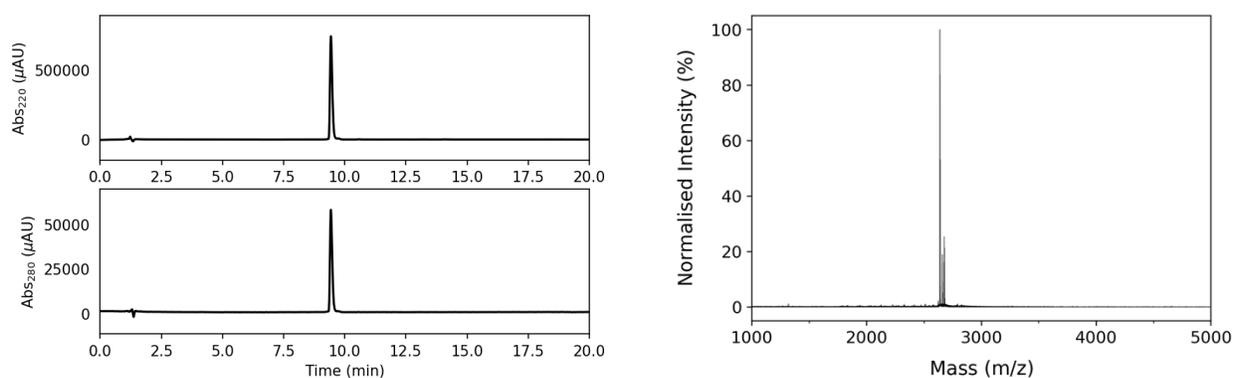
**Figure S33. Characterization of pQLL-A.** Left: Analytical HPLC chromatogram at 220 nm (top) and 280 nm (bottom) from a gradient of 20 to 80% MeCN in 25 mM  $\text{NH}_4\text{HCO}_3$  in  $\text{H}_2\text{O}$ . **Right:** MALDI-TOF MS. Calculated mass: 3479.9 Da  $[M+H]^+$ , 3501.9 Da  $[M+\text{Na}]^+$ . Observed mass: 3479.4 Da  $[M+H]^+$ , 3501.4 Da  $[M+\text{Na}]^+$ .



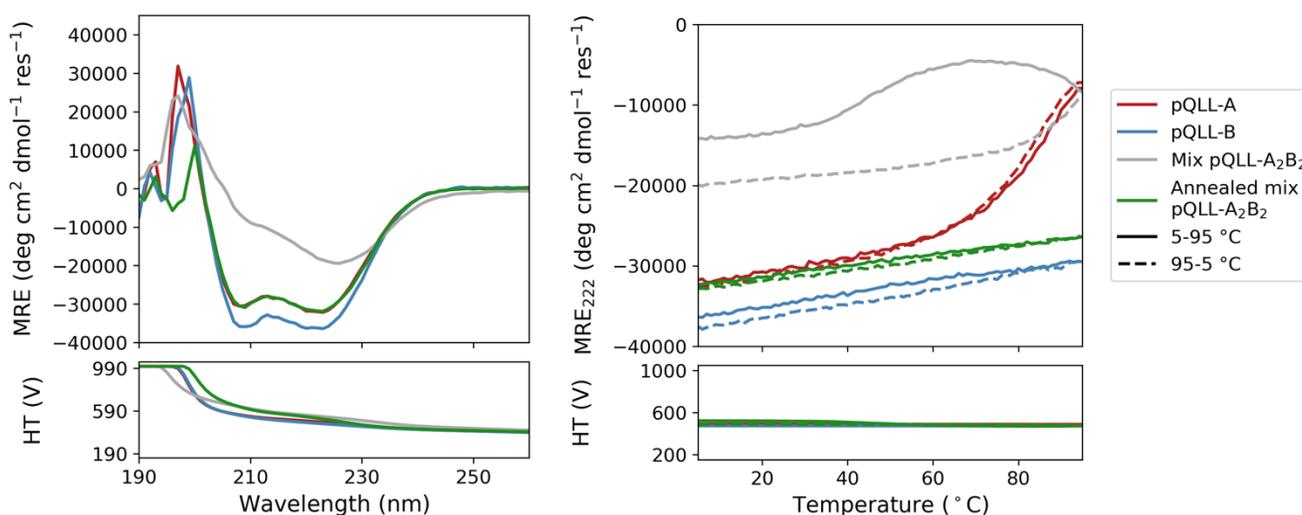
**Figure S34. Characterization of pQLL-B.** Left: Analytical HPLC chromatogram at 220 nm (top) and 280 nm (bottom) from a gradient of 20 to 80% MeCN (0.1% TFA) in  $\text{H}_2\text{O}$  (0.1% TFA). **Right:** Deconvoluted ESI MS. Calculated mass: 3449.3 Da  $[M+H]^+$ . Observed mass: 3449.0 Da  $[M+H]^+$ .



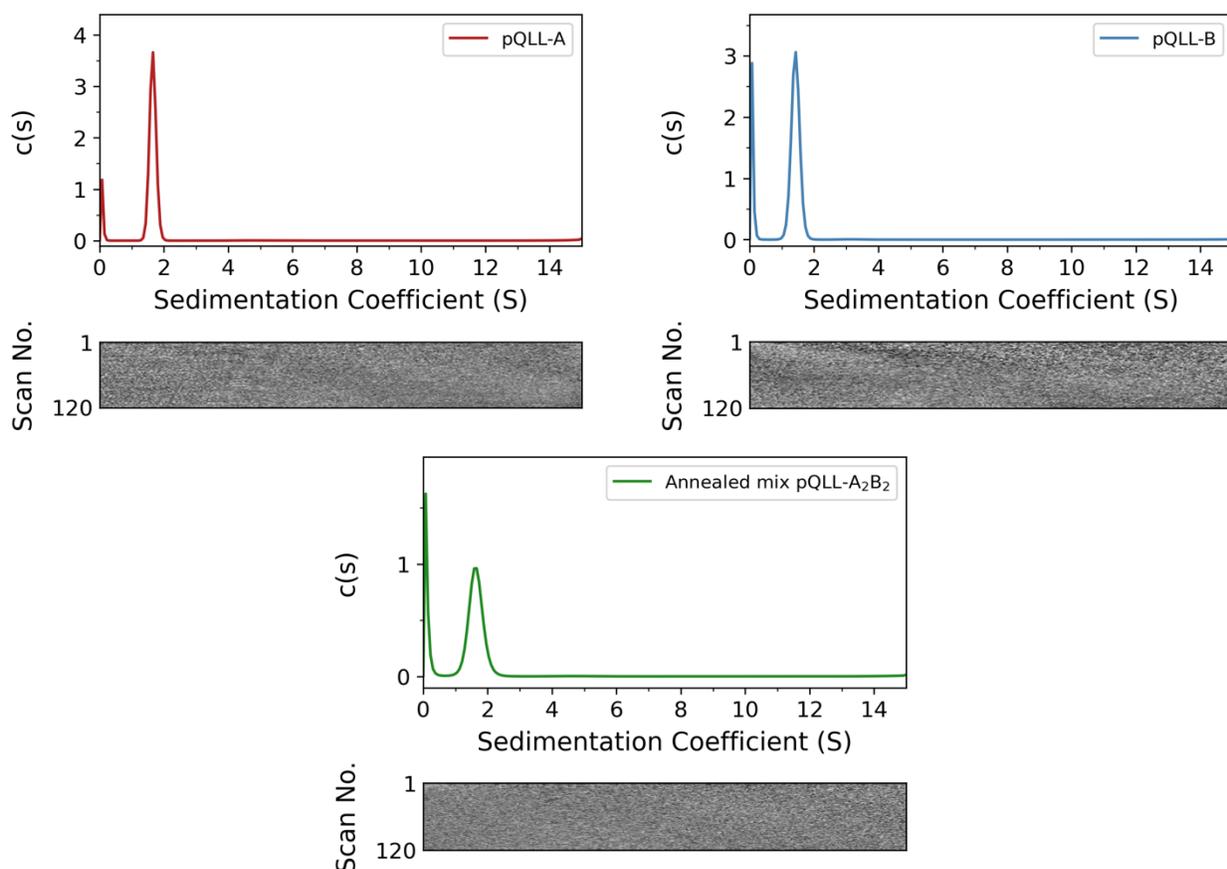
**Figure S35. Characterization of pQLL<sup>3</sup>-A.** Left: Analytical HPLC chromatogram at 220 nm (top) and 280 nm (bottom) from a gradient of 20 to 80% MeCN (0.1% TFA) in  $\text{H}_2\text{O}$  (0.1% TFA). **Right:** MALDI-TOF MS. Calculated mass: 2668.0 Da  $[M+H]^+$ . Observed mass: 2667.9 Da  $[M+H]^+$ .



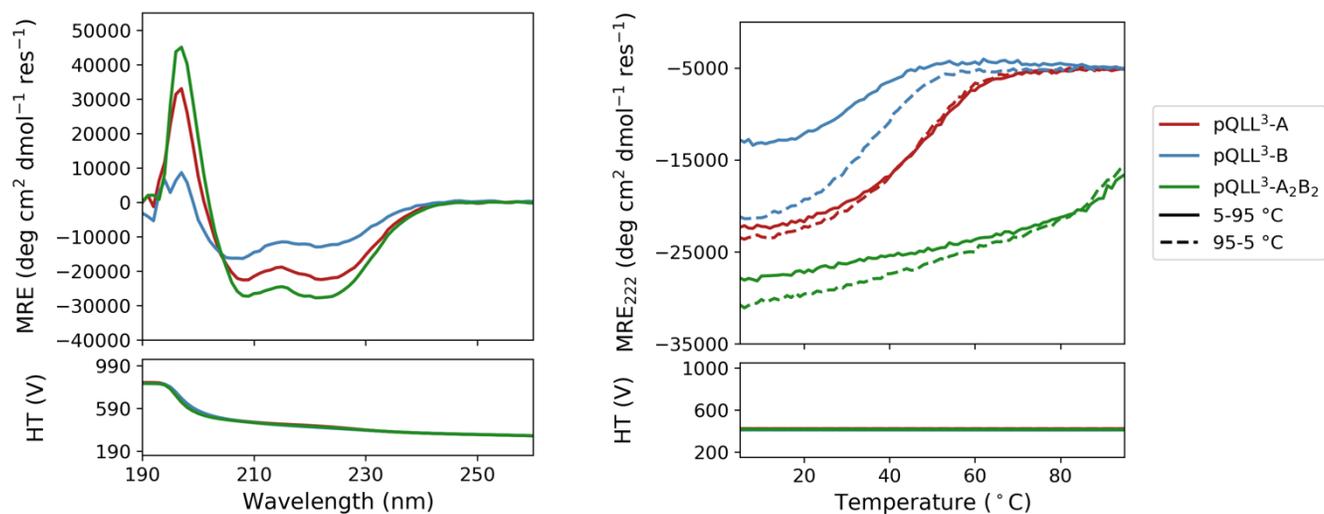
**Figure S36. Characterization of pQLL<sup>3</sup>-B.** Left: Analytical HPLC chromatogram at 220 nm (top) and 280 nm (bottom) from a gradient of 20 to 80% MeCN (0.1% TFA) in H<sub>2</sub>O (0.1% TFA). Right: MALDI-TOF MS. Calculated mass: 2639.3 Da [M+H]<sup>+</sup>. Observed mass: 2639.3 Da [M+H]<sup>+</sup>.



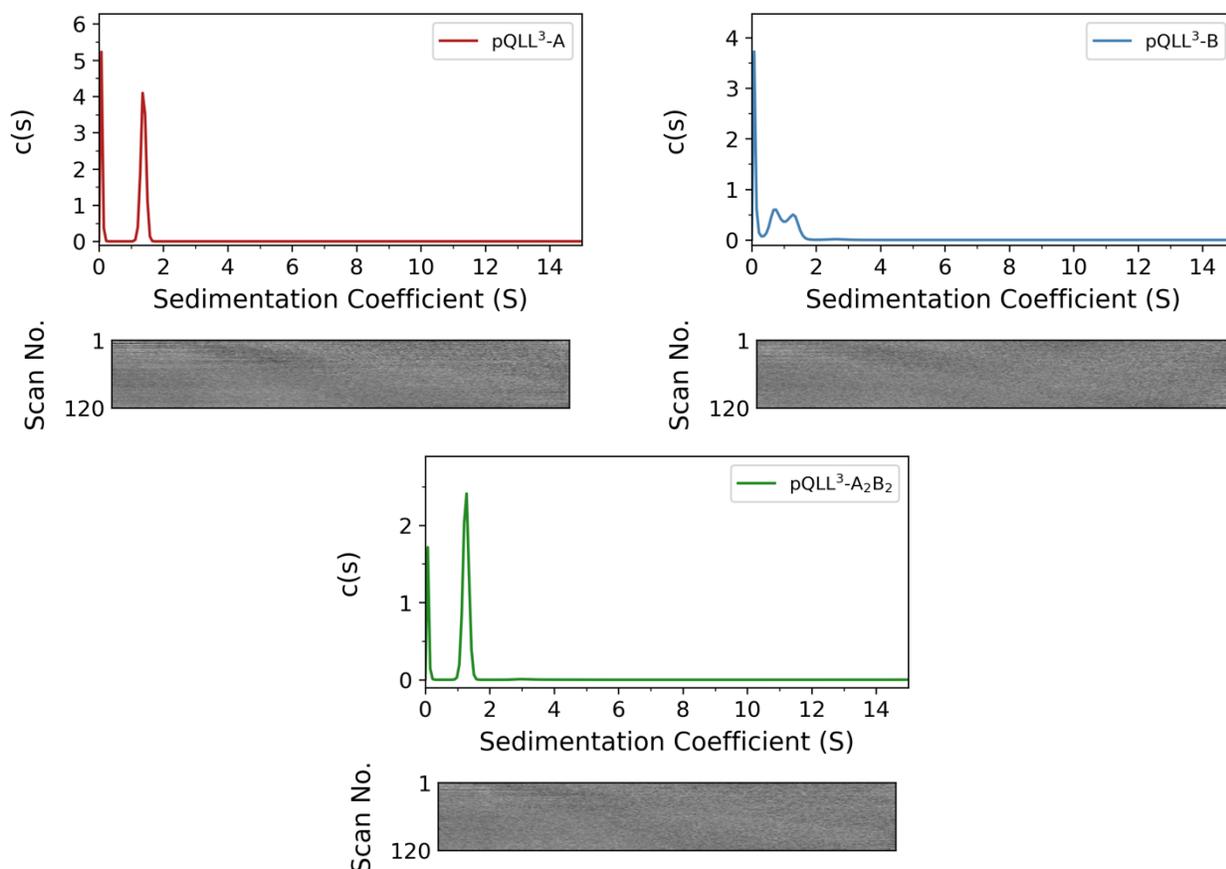
**Figure S37. CD spectroscopy of isolated peptides and mixture for pQLL-A<sub>2</sub>B<sub>2</sub>.** Left: CD spectra at 5 °C before thermal ramp up. Right: Thermal profiles (ramping up and ramping down) monitored at 222 nm. Bottom panels show corresponding high-tension (HT) voltages applied to the detector during measurements. Conditions: 10 μM peptide concentration, PBS (pH 7.4).



**Figure S38. AUC data isolated peptides and mixture for pQLL-A<sub>2</sub>B<sub>2</sub>.** **Top left:** Sedimentation velocity (SV) AUC data for pQLL-A at 50 krpm ( $\bar{v} = 0.742 \text{ cm}^3 \text{ g}^{-1}$ ). Continuous  $c(s)$  distribution returned a molecular mass of 15464 Da corresponding to 4.4 x monomer mass at 95% confidence level ( $f/f_0 = 1.240$ ,  $s = 1.654 \text{ S}$ ,  $s_{20,w} = 1.715 \text{ S}$ ). **Top right:** Sedimentation velocity (SV) AUC data for pQLL-B at 50 krpm ( $\bar{v} = 0.789 \text{ cm}^3 \text{ g}^{-1}$ ). Continuous  $c(s)$  distribution returned a molecular mass of 18432 Da corresponding to 5.3 x monomer mass at 95% confidence level ( $f/f_0 = 1.288$ ,  $s = 1.425 \text{ S}$ ,  $s_{20,w} = 1.488 \text{ S}$ ). **Bottom middle:** Sedimentation velocity (SV) AUC data for pQLL-A<sub>2</sub>B<sub>2</sub> at 50 krpm ( $\bar{v} = 0.766 \text{ cm}^3 \text{ g}^{-1}$ ). Continuous  $c(s)$  distribution returned a molecular mass of 16795 Da corresponding to 4.8 x monomer mass at 95% confidence level ( $f/f_0 = 1.188$ ,  $s = 1.636 \text{ S}$ ,  $s_{20,w} = 1.702 \text{ S}$ ). Conditions: 150  $\mu\text{M}$  peptide for SV experiments, PBS, pH 7.4, 20 °C. Residuals are shown as a bitmap below the fitted data.

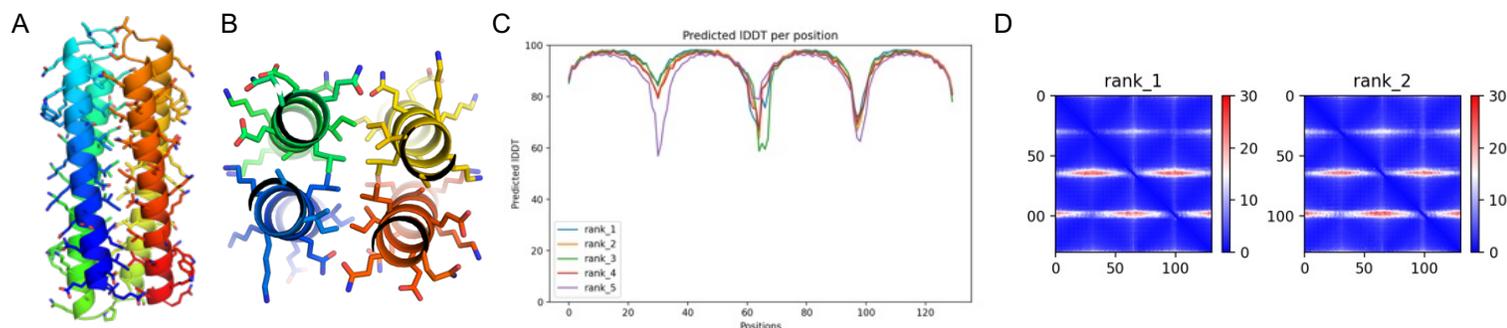


**Figure S39. CD spectroscopy of isolated peptides and mixture for pQLL<sup>3</sup>-A<sub>2</sub>B<sub>2</sub>.** Left: CD spectra at 5 °C before thermal ramp up. Right: Thermal profiles (ramping up and ramping down) monitored at 222 nm. Bottom panels show corresponding high-tension (HT) voltages applied to the detector during measurements. Conditions: 10 μM peptide concentration, PBS (pH 7.4).

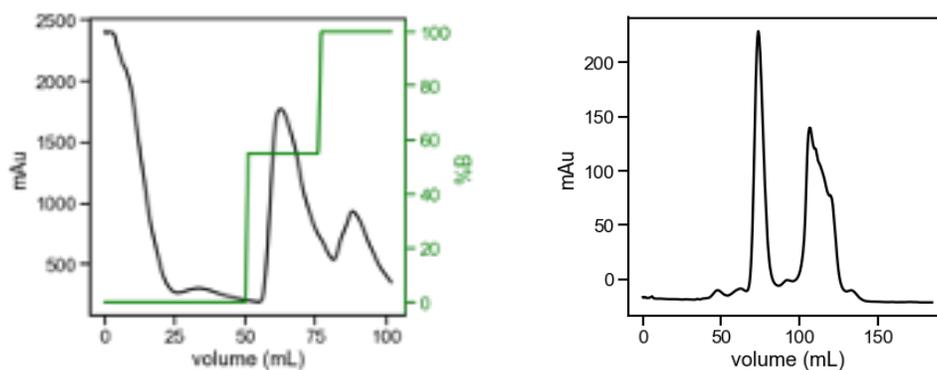


**Figure S40. AUC data isolated peptides and mixture for pQLL<sup>3</sup>-A<sub>2</sub>B<sub>2</sub>.** **Top left:** Sedimentation velocity (SV) AUC data for pQLL<sup>3</sup>-A at 50 krpm ( $\bar{v} = 0.744 \text{ cm}^3 \text{ g}^{-1}$ ). Continuous c(s) distribution returned a molecular mass of 11150 Da corresponding to 4.2 x monomer mass at 95% confidence level ( $f/f_0 = 1.187$ ,  $s = 1.379 \text{ S}$ ,  $s_{20,w} = 1.431 \text{ S}$ ). **Top right:** Sedimentation velocity (SV) AUC data for pQLL<sup>3</sup>-B at 50 krpm ( $\bar{v} = 0.790 \text{ cm}^3 \text{ g}^{-1}$ ). Continuous c(s) distribution returned a molecular mass of 7970 Da corresponding to 3.0 x monomer mass at 95% confidence level ( $f/f_0 = 1.341$ ,  $s = 0.781 \text{ S}$ ,  $s_{20,w} = 0.815 \text{ S}$ ). **Bottom middle:** Sedimentation velocity (SV) AUC data for pQLL<sup>3</sup>-A<sub>2</sub>B<sub>2</sub> at 50 krpm ( $\bar{v} = 0.767 \text{ cm}^3 \text{ g}^{-1}$ ). Continuous c(s) distribution returned a molecular mass of 11556 Da corresponding to 4.4 x monomer mass at 95% confidence level ( $f/f_0 = 1.197$ ,  $s = 1.260 \text{ S}$ ,  $s_{20,w} = 1.311 \text{ S}$ ). Conditions: 150  $\mu\text{M}$  peptide for SV experiments, PBS, pH 7.4, 20 °C. Residuals are shown as a bitmap below the fitted data.

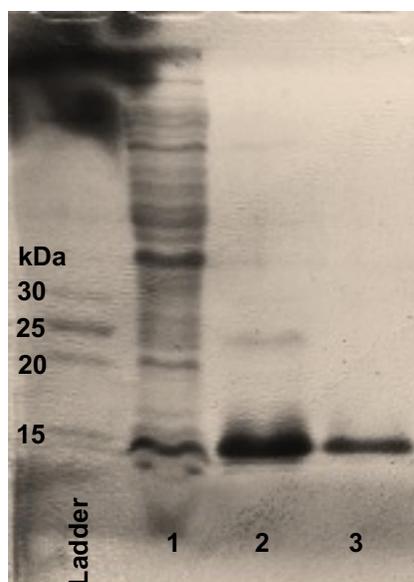
## 2.2.4. Characterization of the single-chain protein



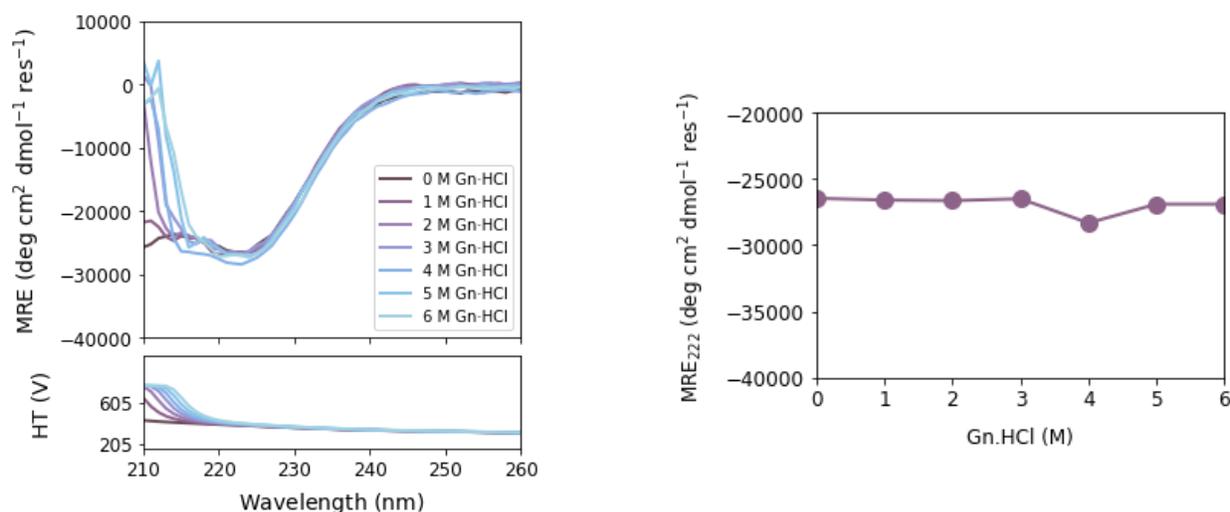
**Figure S41.** AlphaFold2 prediction of sc-apCC-4. (A&B) Orthogonal views of the top-ranked AlphaFold2 prediction for the sequence of sc-apCC-4. It is in good agreement with the narrow and wide interface organization of the parent apCC-Tet\*. (C) Associated predicted IDDT and (D) Predicted Alignment Error (PAE) graphs for the sc-apCC-4 prediction.



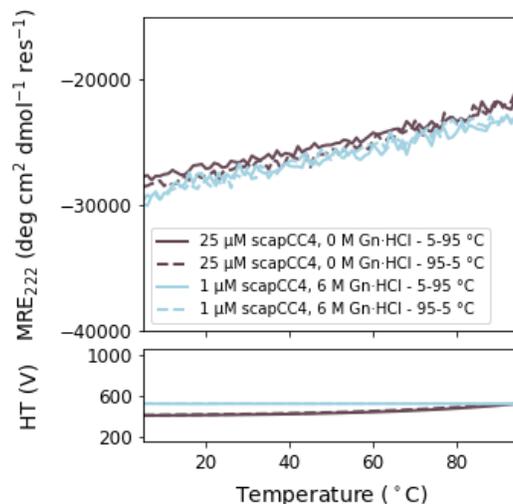
**Figure S42. FPLC of scapCC4 purification.** **Left:** sc-apCC-4 Ni-affinity chromatography. Fractions eluting at 55% B were pulled and concentrated then further purified by size exclusion chromatography. **Right:** sc-apCC-4 size exclusion chromatography. Fractions eluting at 60 mL were pooled and concentrated to be analysed in further experiments.



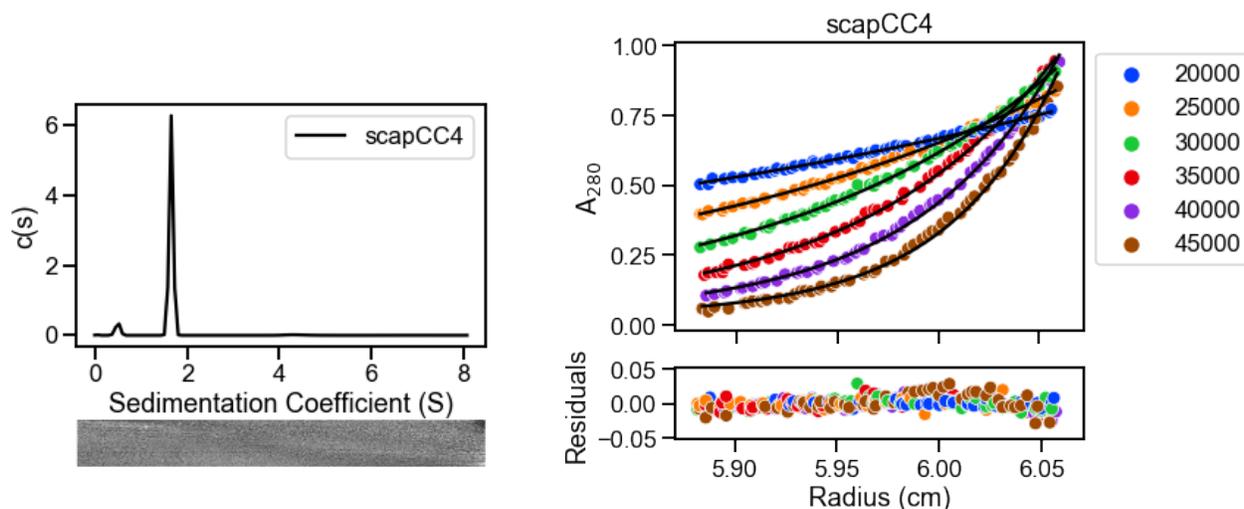
**Figure S43. SDS-PAGE of sc-apCC-4 purification.** Ladder: PageRuler Low Range Unstained Protein Ladder (Thermo Scientific), (1) sc-apCC-4 flow through from Ni-affinity chromatography, (2) combined fractions from Ni-affinity chromatography, (3) combined fractions after SEC. Expected molecular weight: 17 kDa.



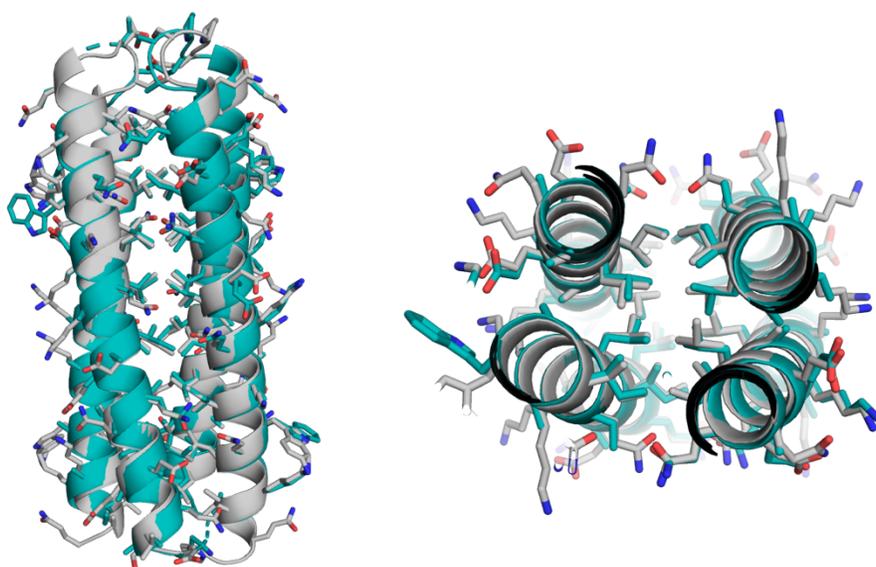
**Figure S44. Chemical denaturation of sc-apCC-4.** **Left:** CD full scan from 210 to 260 nm at 5 °C of sc-apCC-4 incubated with increasing concentration of guanidinium hydrochloride (Gn·HCl). **Right:** Chemical denaturation curve measured at 222 nm for sc-apCC-4 (deduced from left spectra). Conditions: 1  $\mu$ M peptide concentration,  $\pm$  Gn·HCl, phosphate buffer, pH 7.4 and incubation overnight before measurements.



**Figure S45. Temperature profile of sc-apCC-4 in presence of high concentration of denaturant.** Thermal profiles (ramping up and ramping down) monitored at 222 nm for 1  $\mu\text{M}$  of sc-apCC-4 with 6 M Gn·HCl after incubation overnight.



**Figure S46. AUC data for sc-apCC-4.** **Left:** Sedimentation velocity (SV) AUC data for sc-apCC-4 at 50 krpm ( $\bar{v} = 0.752 \text{ cm}^3 \text{ g}^{-1}$ ). Residuals are shown as a bitmap below the fitted data. Continuous  $c(s)$  distribution returned a molecular mass of 16042 Da corresponding to 0.9 x monomer mass at 95% confidence level ( $f/f_0 = 1.21$ ,  $s = 1.658 \text{ S}$ ,  $s_{20,w} = 1.722 \text{ S}$ ). **Right:** Sedimentation equilibrium (SE) AUC data for sc-apCC-4 between 20 and 45 krpm at 5 krpm intervals. Fitted single-ideal species model curves are overlaid in black and gave a molecular mass 17009 Da corresponding to 1.0 x monomer mass, 95% confidence limits 16745-17263 Da. Conditions: 25  $\mu\text{M}$  protein for SV and SE experiments, 50 mM sodium phosphate, 150 mM NaCl, pH 7.4, 20  $^\circ\text{C}$



**Figure S47. Structural overlay of the crystal structure and related AlphaFold2 prediction for sc-apCC-4 protein.** Overlay of the crystal structure of sc-apCC-4 (teal) and the corresponding AlphaFold2 prediction (grey) with a  $\text{RMSD}_{\text{all-atom}} = 0.475 \text{ \AA}$  and  $\text{C}\alpha \text{ RMSD} = 0.463 \text{ \AA}$ .

## References

- 1 O. D. Testa, E. Moutevelis and D. N. Woolfson, *Nucleic Acids Res.*, 2009, **37**, D315-322.
- 2 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, *Nature*, 2021, **596**, 583-589.
- 3 R. Evans, M. O'Neill, A. Pritzel, N. Antropova, A. Senior, T. Green, A. Židek, R. Bates, S. Blackwell, J. Yim, O. Ronneberger, S. Bodenstein, M. Zielinski, A. Bridgland, A. Potapenko, A. Cowie, K. Tunyasuvunakool, R. Jain, E. Clancy, P. Kohli, J. Jumper and D. Hassabis, *bioRxiv*, 2021, DOI: 10.1101/2021.10.04.463034, 2021.2010.2004.463034.
- 4 M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov and M. Steinegger, *Nat. Methods*, 2022, **19**, 679-682.
- 5 B. J. H. Kuipers and H. Gruppen, *J. Agric. Food Chem.*, 2007, **55**, 5445-5451.
- 6 P. Schuck, *Biophys. J.*, 2000, **78**, 1606-1619.
- 7 P. Schuck, *Anal. Biochem.*, 2003, **320**, 104-124.
- 8 M. D. Watson, I. Peran and D. P. Raleigh, *Biochemistry*, 2016, **55**, 3685-3691.
- 9 A. J. Smith, F. Thomas, D. Shoemark, D. N. Woolfson and N. J. Savery, *ACS Synth. Biol.*, 2019, **8**, 1284-1293.
- 10 C. L. Edgell, A. J. Smith, J. L. Beesley, N. J. Savery and D. N. Woolfson, *ACS Synth. Biol.*, 2020, **9**, 427-436.
- 11 T. G. Bernhardt and P. A. de Boer, *Mol. Microbiol.*, 2003, **48**, 1171-1182.
- 12 G. Winter, *J. Appl. Crystallogr.*, 2010, **43**, 186-190.
- 13 G. Winter, D. G. Waterman, J. M. Parkhurst, A. S. Brewster, R. J. Gildea, M. Gerstel, L. Fuentes-Montero, M. Vollmar, T. Michels-Clark, I. D. Young, N. K. Sauter and G. Evans, *Acta Crystallogr. D*, 2018, **74**, 85-97.
- 14 H. Powell, *Acta Crystallogr. D*, 1999, **55**, 1690-1695.
- 15 P. R. Evans and G. N. Murshudov, *Acta Crystallogr. D*, 2013, **69**, 1204-1214.
- 16 M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson, *Acta Crystallogr. D*, 2011, **67**, 235-242.
- 17 W. Kabsch, *Acta Crystallogr. D*, 2010, **66**, 125-132.
- 18 D. D. Rodríguez, C. Grosse, S. Himmel, C. González, I. M. de Ilarduya, S. Becker, G. M. Sheldrick and I. Usón, *Nat. Methods*, 2009, **6**, 651-653.
- 19 H. Jenkins, *Acta Crystallogr. D*, 2018, **74**, 205-214.
- 20 K. Cowtan, *Acta Crystallogr. D*, 2006, **62**, 1002-1011.
- 21 G. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin, *Acta Crystallogr. D*, 2011, **67**, 355-367.
- 22 A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J. Appl. Crystallogr.*, 2007, **40**, 658-674.
- 23 P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart and P. D. Adams, *Acta Crystallogr. D*, 2012, **68**, 352-367.
- 24 P. Kumar and D. N. Woolfson, *Bioinformatics*, 2021, **37**, 4575-4577.
- 25 A. Lombardi, C. M. Summa, S. Geremia, L. Randaccio, V. Pavone and W. F. DeGrado, *Proc. Natl. Acad. Sci. U.S.A.*, 2000, **97**, 6298-6305.
- 26 S. J. Lahr, D. E. Engel, S. E. Stayrook, O. Maglio, B. North, S. Geremia, A. Lombardi and W. F. DeGrado, *J. Mol. Biol.*, 2005, **346**, 1441-1454.
- 27 L. Di Costanzo, H. Wade, S. Geremia, L. Randaccio, V. Pavone, W. F. DeGrado and A. Lombardi, *J. Am. Chem. Soc.*, 2001, **123**, 12749-12757.
- 28 M. K. Yadav, L. J. Leman, D. J. Price, C. L. Brooks, C. D. Stout and M. R. Ghadiri, *Biochemistry*, 2006, **45**, 4463-4473.
- 29 Y. Deng, J. Liu, Q. Zheng, D. Eliezer, N. R. Kallenbach and M. Lu, *Structure*, 2006, **14**, 247-255.
- 30 Y. Deng, Q. Zheng, J. Liu, C.-S. Cheng, N. R. Kallenbach and M. Lu, *Protein Sci.*, 2007, **16**, 323-328.

- 31 G. Grigoryan, Y. H. Kim, R. Acharya, K. Axelrod, R. M. Jain, L. Willis, M. Drndic, J. M. Kikkawa and W. F. DeGrado, *Science*, 2011, **332**, 1071-1076.
- 32 C. E. Schafmeister, S. L. LaPorte, L. J. W. Miercke and R. M. Stroud, *Nat. Struct. Biol.*, 1997, **4**, 1039-1046.
- 33 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.
- 34 Z. Chen, M. C. Johnson, J. Chen, M. J. Bick, S. E. Boyken, B. Lin, J. J. De Yoreo, J. M. Kollman, D. Baker and F. DiMaio, *J. Am. Chem. Soc.*, 2019, **141**, 8891-8895.
- 35 N. F. Polizzi, Y. Wu, T. Lemmin, A. M. Maxwell, S. Q. Zhang, J. Rawson, D. N. Beratan, M. J. Therien and W. F. DeGrado, *Nat. Chem.*, 2017, **9**, 1157-1164.
- 36 S.-Q. Zhang, H. Huang, J. Yang, H. T. Kratochvil, M. Lolicato, Y. Liu, X. Shu, L. Liu and W. F. DeGrado, *Nat. Chem. Biol.*, 2018, **14**, 870-875.
- 37 Z. Chen, S. E. Boyken, M. Jia, F. Busch, D. Flores-Solis, M. J. Bick, P. Lu, Z. L. VanAernum, A. Sahasrabudde, R. A. Langan, S. Bermeo, T. J. Brunette, V. K. Mulligan, L. P. Carter, F. DiMaio, N. G. Sgourakis, V. H. Wysocki and D. Baker, *Nature*, 2019, **565**, 106-111.
- 38 G. G. Rhys, C. W. Wood, J. L. Beesley, N. R. Zaccai, A. J. Burton, R. L. Brady, A. R. Thomson and D. N. Woolfson, *J. Am. Chem. Soc.*, 2019, **141**, 8787-8797.
- 39 N. F. Polizzi and W. F. DeGrado, *Science*, 2020, **369**, 1227-1233.