Supporting Information

A Basis Set of *de novo* Coiled-coil Peptide Oligomers for Rational Protein Design and Synthetic Biology

Jordan M. Fletcher[1,†], Aimee L. Boyle[1,†], Marc Bruning[1,†], Gail J. Bartlett[1], Thomas L. Vincent[1,2], Nathan R. Zaccai[3,4], Craig T. Armstrong[1,3], Elizabeth H. C. Bromley[1,5], Paula J. Booth[3], R. Leo Brady[3], Andrew R. Thomson[1,*] and Derek N. Woolfson[1,3,*]

[1]School of Chemistry, University of Bristol, Cantocks Close, Bristol BS8 1TS, U.K.

[2]Current address: Department of Genetic Medicine, Weill Cornell Medical College, 1305, York Avenue, New York, NY, 10021, U.S.A.

[3]School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol BS8 1TD, U.K.

[4]Current address: Thomas C. Jenkins Department of Biophysics, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, U.S.A.

[5]Current address: Department of Physics, Durham University, South Road, Durham DH1 3LE, U.K.

*To whom correspondence should be addressed: Drew.Thomson@bristol.ac.uk; D.N.Woolfson@bristol.ac.uk

[†]These authors contributed equally to this work.

Table of Contents

Bioinformatic Analysis of CC+ Data

Amino acid	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*Sum*
A	54	90	78	73	55	78	65	493
C	6	4	0	9	0	4	4	27
D	4	75	64	9	23	67	54	296
E	7	115	113	21	156	111	146	669
F	29	9	13	25	7	4	3	90
G	4	16	11	1	10	28	16	86
H	8	11	21	8	6	20	6	80
I	162	10	15	67	20	13	16	303
K	37	80	63	23	83	88	93	467
L	235	30	34	366	57	30	48	800
M	13	4	10	28	10	4	11	80
N	57	47	40	27	49	48	25	293
P	1	3	1	0	0	1	0	6
Q	9	72	89	28	98	67	83	446
R	22	55	61	12	63	66	80	359
S	15	56	53	27	46	49	36	282
T	18	34	46	36	27	42	24	227
V	118	20	23	38	29	21	24	273
W	2	0	0	4	4	5	2	17
Y	15	16	9	17	6	7	13	83
Sum	816	747	744	819	749	753	749	5377

Table S1. Observed numbers of amino acids at each register position in all parallel, homo-oligomeric coiled coils (<= 50% sequence identity, >21 residues, culled from November 2011 release of the CC+ database (http://coiledcoils.chm.bris.ac.uk))

Amino acid	*a*	*b*	*c*	*d*	*e*	*f*	*g*	Sum
A	42	57	34	59	36	56	29	313
C	6	4	1	9	1	4	4	29
D	3	49	44	9	16	44	28	193
E	8	89	88	16	107	84	122	514
F	13	8	10	8	7	5	3	54
G	2	7	8	2	4	14	4	41
H	7	10	14	4	6	15	4	60
I	75	8	13	17	16	10	13	152
K	38	59	49	24	62	53	64	349
L	161	24	20	275	34	25	34	573
M	10	5	11	18	9	4	9	66
N	54	21	25	7	25	33	12	177
P	1	3	2	1	1	1	1	10
Q	3	42	58	13	66	42	64	288
R	21	41	43	12	47	47	48	259
S	11	37	34	22	26	22	19	171
T	12	21	28	21	17	28	17	144
V	66	14	16	15	15	13	18	157
W	3	1	1	5	4	2	3	19
Y	16	10	9	17	5	7	8	72
Total	552	510	508	554	504	509	504	3641

Table S2A. Observed parallel homo-dimers; <= 50% sequence identity; >21 residues long. Full data for calculation of Oligomer-state Discrimination Factors. Data were collected from the November 2011 release of CC+ (http://coiledcoils.chm.bris.ac.uk). A pseudocount of 1 was added to each cell to prevent any errors from division by zero.

Amino acid	*a*	*b*	*c*	*d*	*e*	*f*	*g*	Sum
A	12	30	37	16	18	20	32	165
C	2	2	1	2	1	2	2	12
D	2	20	21	2	9	22	18	94
E	1	28	22	3	47	25	20	146
F	7	3	5	6	2	1	2	26
G	3	11	5	1	8	14	13	55
H	3	3	9	6	2	6	4	33
I	83	2	4	41	5	5	5	145
K	1	17	13	1	19	27	27	105
L	63	8	15	92	17	7	13	215
M	4	1	1	9	2	2	4	23
N	5	24	15	22	20	15	15	116
P	2	1	1	1	1	2	1	9
Q	7	26	25	15	32	24	19	148
R	3	13	16	2	15	20	31	100
S	6	20	19	6	19	25	19	114
T	7	15	19	14	12	15	8	90
V	51	7	9	23	15	10	8	123
W	1	1	1	1	2	4	1	11
Y	1	8	2	2	3	2	7	25
Total	264	240	240	265	249	248	249	1755

Table S2B. Observed parallel homo-trimers; <= 50% sequence identity; >21 residues long. Full data for calculation of Oligomer-state Discrimination Factors. Data were collected from the November 2011 release of CC+ (http://coiledcoils.chm.bris.ac.uk). A pseudocount of 1 was added to each cell to prevent any errors from division by zero.

Amino acid	*a*	*b*	*c*	*d*	*e*	*f*	*g*
A	0.22	-0.05	-0.36	0.25	-0.01	0.13	-0.35
C	0.16	-0.03	-0.33	0.33	-0.31	-0.01	-0.01
D	-0.14	0.06	0.00	0.33	-0.06	-0.01	-0.11
E	0.58	0.17	0.28	0.41	0.05	0.21	0.48
F	-0.05	0.10	-0.02	-0.20	0.24	0.39	-0.13
G	-0.50	-0.52	-0.12	-0.02	-0.61	-0.31	-0.82
H	0.05	0.20	-0.13	-0.50	0.17	0.09	-0.31
I	-0.36	0.27	0.19	-0.70	0.20	-0.01	0.11
K	1.26	0.21	0.25	1.06	0.21	-0.02	0.07
L	0.09	0.15	-0.20	0.16	-0.01	0.24	0.11
M	0.08	0.37	0.72	-0.02	0.35	-0.01	0.05
N	0.71	-0.39	-0.10	-0.82	-0.21	0.03	-0.40
P	-0.62	0.15	-0.02	-0.32	-0.31	-0.61	-0.31
Q	-0.69	-0.12	0.04	-0.38	0.01	-0.07	0.22
R	0.52	0.17	0.10	0.46	0.19	0.06	-0.12
S	-0.06	-0.06	-0.07	0.24	-0.17	-0.37	-0.31
T	-0.09	-0.18	-0.16	-0.14	-0.15	-0.04	0.02
V	-0.21	-0.03	-0.08	-0.51	-0.31	-0.20	0.05
W	0.16	-0.33	-0.33	0.38	-0.01	-0.61	0.17
Y	0.88	-0.23	0.33	0.61	-0.08	0.23	-0.25

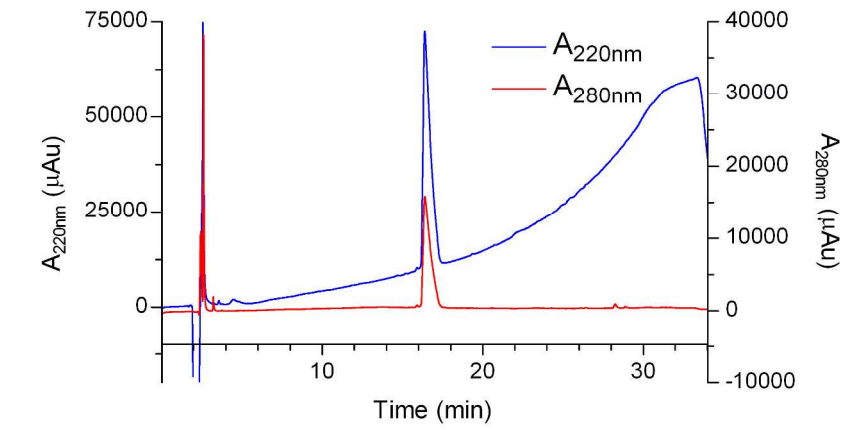
Table S2C. Log_{10} ($observed_{dimer}$ / $observed_{trimer}$) where the observed number is normalized by the total number of amino acid residues at each heptad position; *i.e.* to account for the different sizes of the two datasets.

Peptide Sequences

Peptide Name	Sequence	Mass (Exact/ Molar)
CC-pIL	Ac–GE**I**AA**L**KQ EIAA**L**KK EIAA**L**KW EIAA**L**KQ GYY–NH_2	3571.0/ 3573.2
CC-pII	Ac–GE**I**AA**I**KQ EIAA**I**KK EIAA**I**KW EIAA**I**KQ GYG–NH_2	3465.0/ 3467.1
CC-pLI	Ac–GE**L**AA**I**KQ E**L**AA**I**KK E**L**AA**I**KW E**L**AA**I**KQ GAG–NH_2	3372.9/ 3375.0
CC-pIL-I17N	Ac–GE**I**AA**L**KQ EIAA**L**KK E**N**AA**L**KW EIAA**L**KQ GYY–NH_2	3572.0/ 3574.2
CC-pII-I13N	Ac–GE**I**AA**I**KQ EIAA**N**KK EIAA**I**KW EIAA**I**KQ GYG–NH_2	3465.9/ 3468.0
CC-pIL-W22Φ	Ac–GE**I**AA**L**KQ EIAA**L**KK EIAA**L**K**Φ** EIAA**L**KQ GYY–NH_2	3657.0/ 3660.1
CC-pII-W22Φ	Ac–GE**I**AA**I**KQ EIAA**I**KK EIAA**I**K**Φ** EIAA**I**KQ GYG–NH_2	3551.9/ 3553.9
CC-pIL-I17N-W22Φ	Ac–GE**I**AA**L**KQ EIAA**L**KK E**N**AA**L**K**Φ** EIAA**L**KQ GYY–NH_2	3658.8/ 3661.0
GCN4-pIL	Ac–R MKQ**L**EDK **I**EE**L**LSK **I**YH**L**ENE **I**AR**L**KKL **I**GER–H	4034.3/ 4036.8

Table S3. Peptides used in this study.

Peptide Characterization



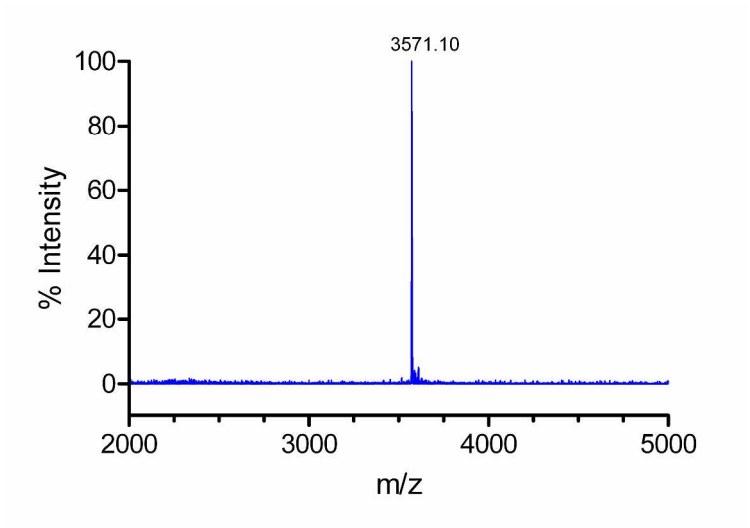


Figure S1A. Analytical HPLC (top) and MALDI-TOF mass spectrum for CC-pIL.

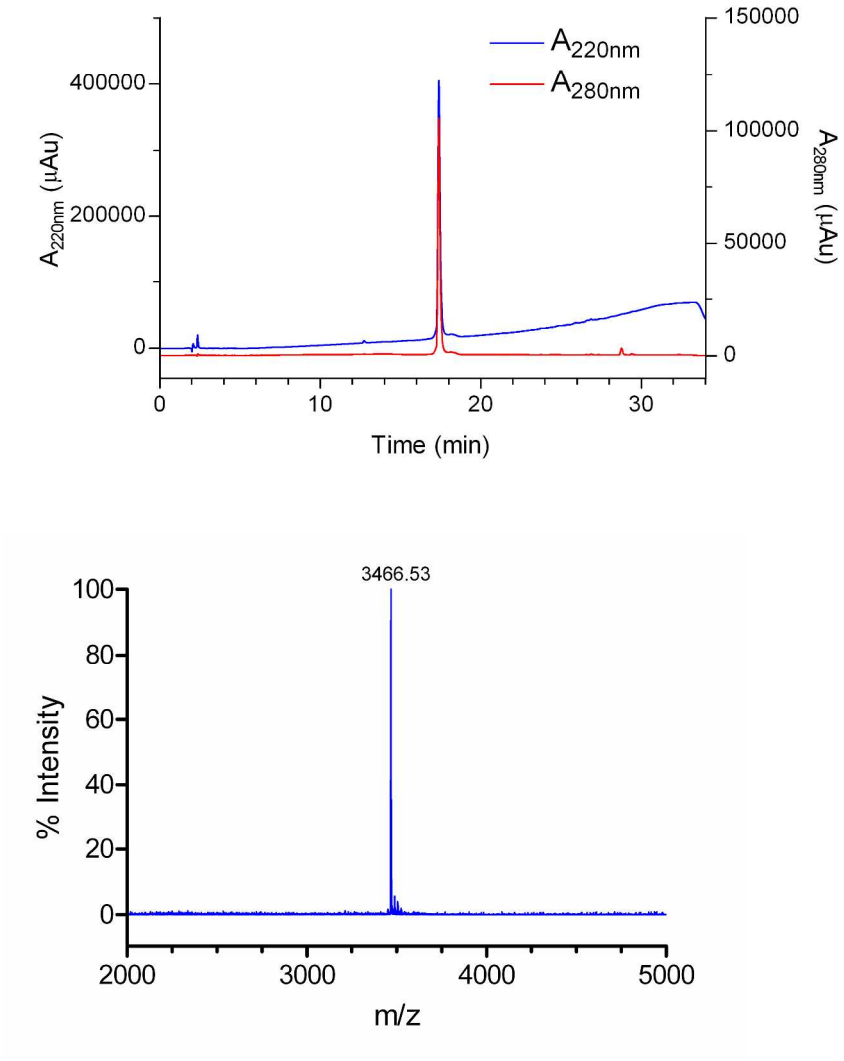


Figure S1B. Analytical HPLC (top) and MALDI-TOF mass spectrum for CC-pII.

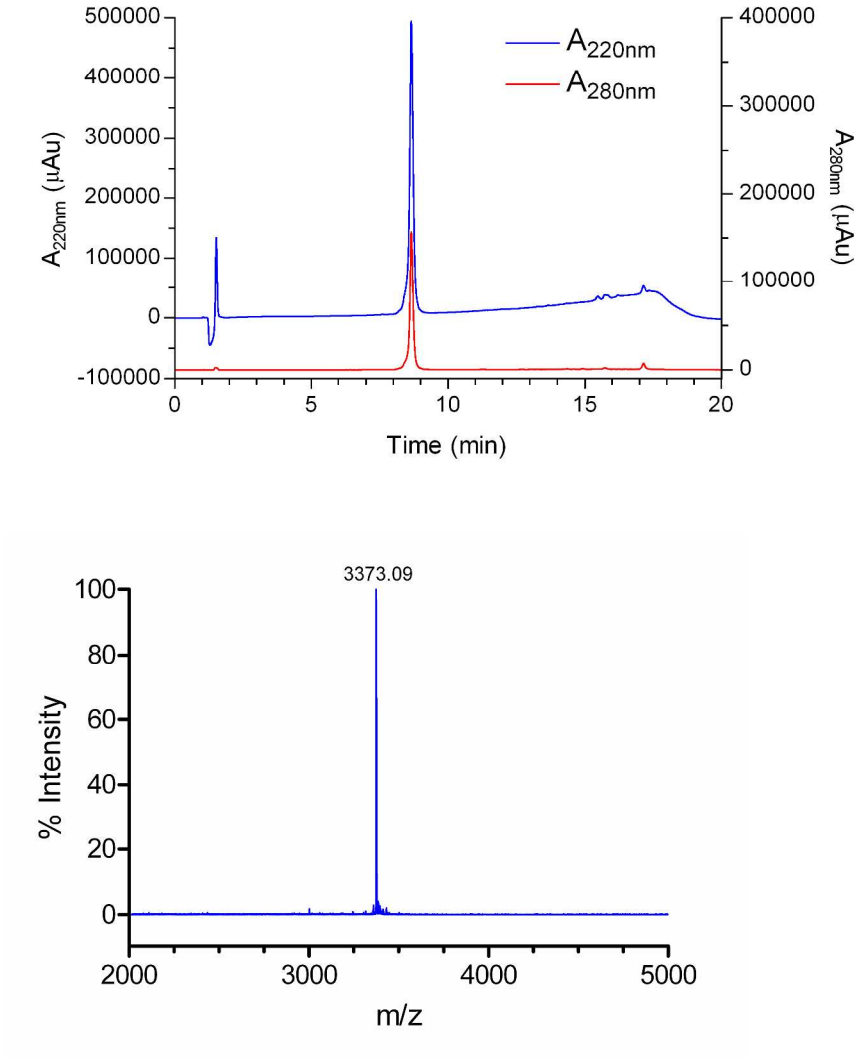


Figure S1C. Analytical HPLC (top) and MALDI-TOF mass spectrum for CC-pLI.

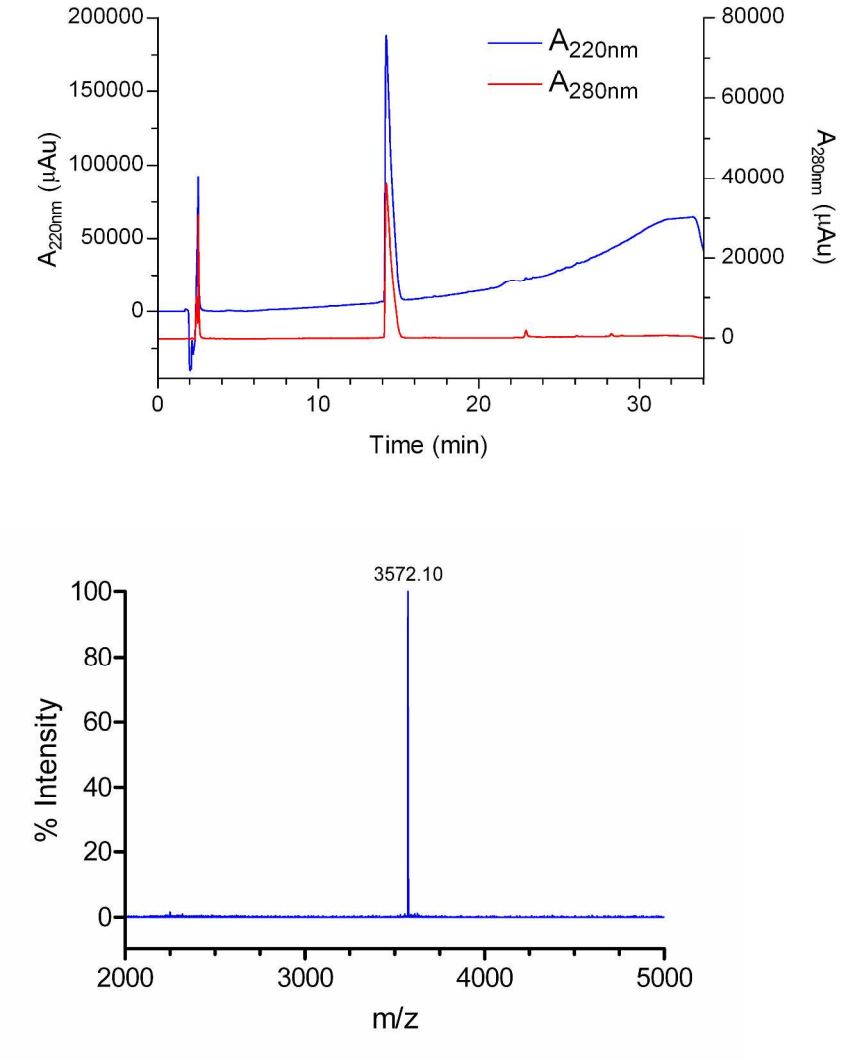


Figure S1D. Analytical HPLC (top) and MALDI-TOF mass spectrum for CC-pIL-I17N.

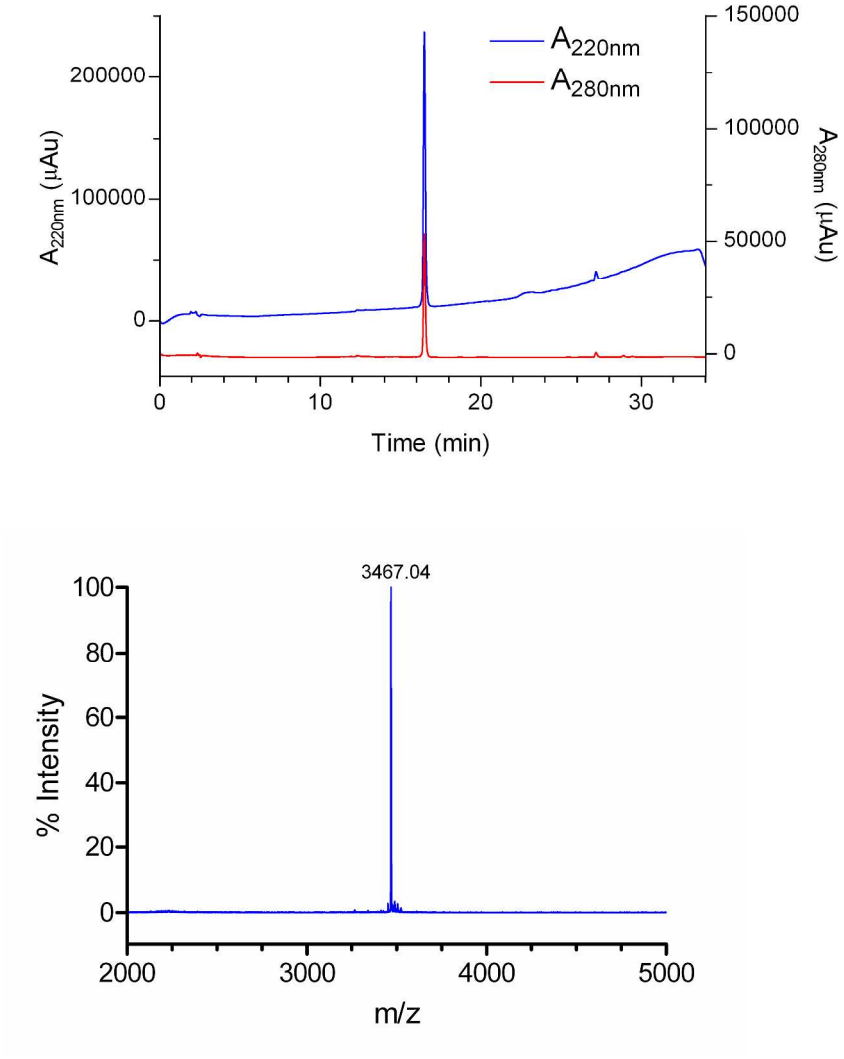


Figure S1E. Analytical HPLC (top) and MALDI-TOF mass spectrum for CC-pII-I13N.

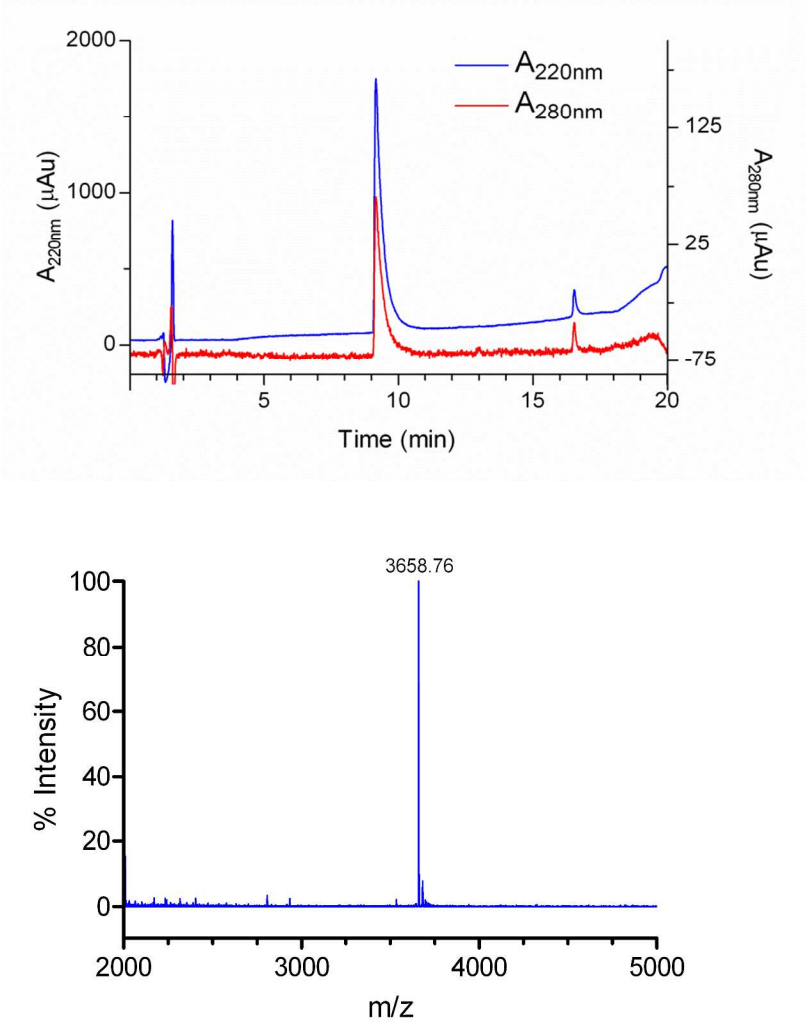


Figure S1F. Analytical HPLC (top) and MALDI-TOF mass spectrum for CC-pIL-W22Φ.

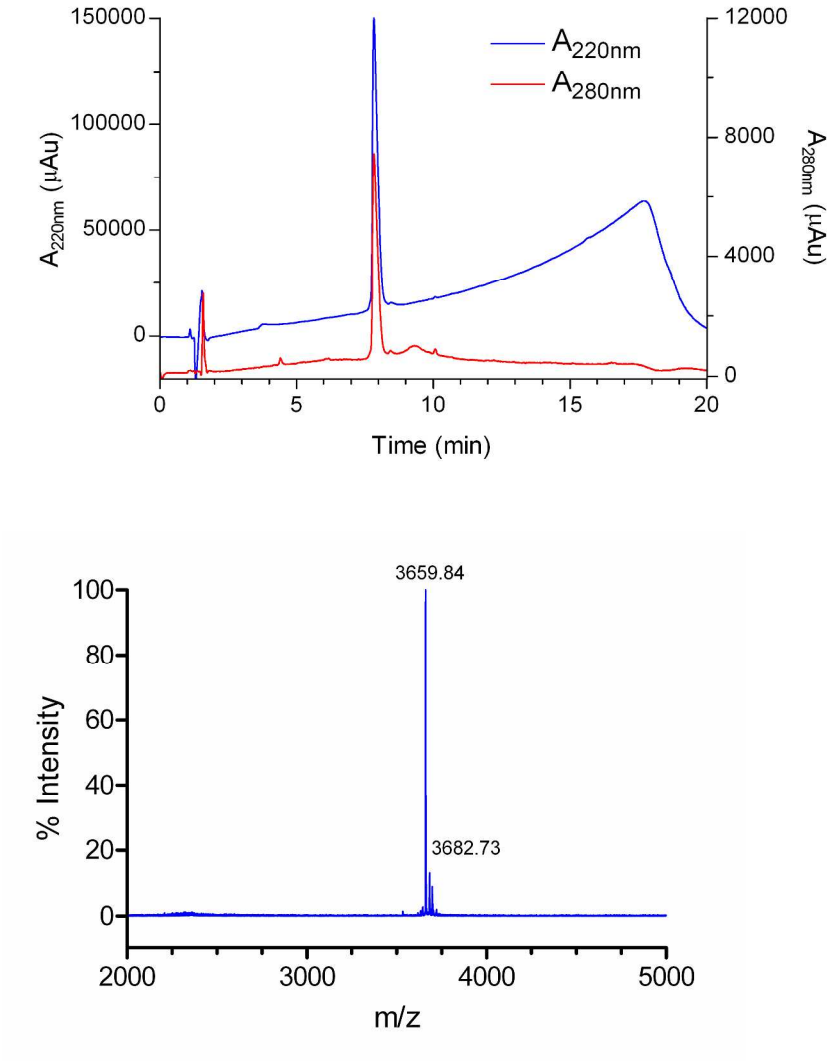


Figure S1G. Analytical HPLC (top) and MALDI-TOF mass spectrum for CC-pIL-I17N-W22Φ.

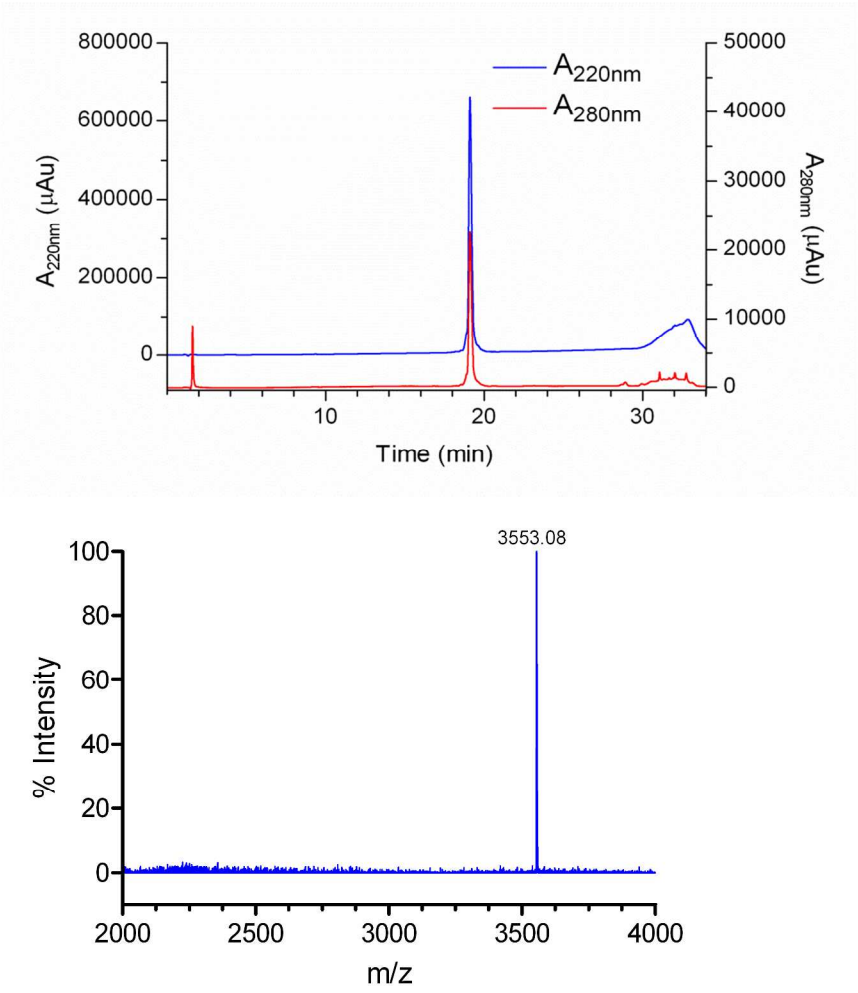


Figure S1H. Analytical HPLC (top) and MALDI-TOF mass spectrum for CC-pII-W22Φ.

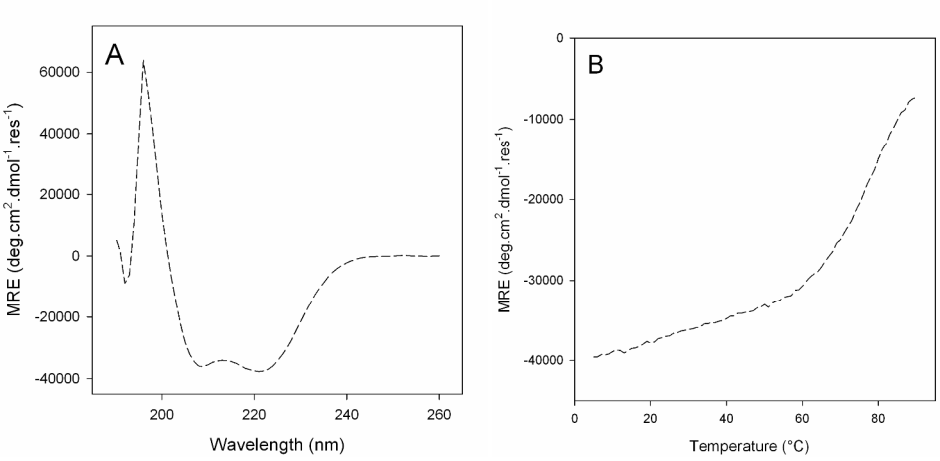


Figure S2A & B. Circular dichroism spectrum (A) and temperature dependence of signal at 222 nm (B) for CC-pIL-I17N at 50 µM concentration in PBS buffer.

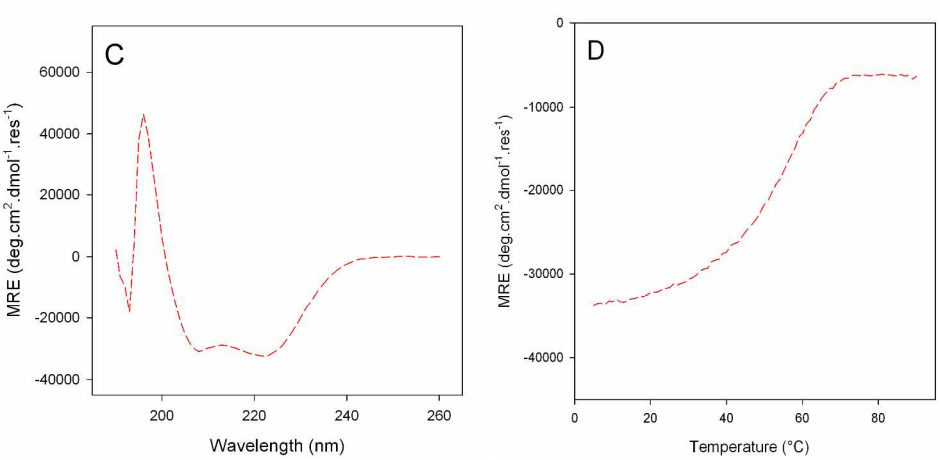


Figure S2C & D. Circular dichroism spectrum (C) and temperature dependence of signal at 222 nm (D) for CC-pII-I13N at 50 µM concentration in PBS buffer.

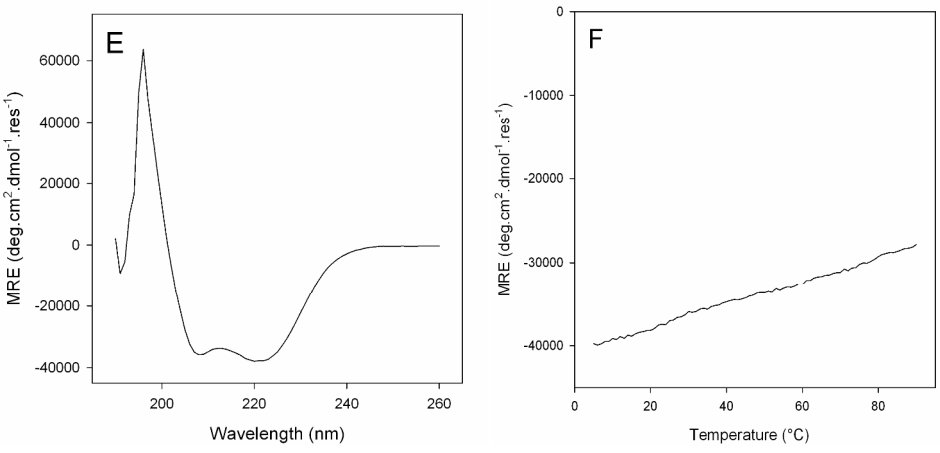


Figure S2E & F. Circular dichroism spectrum (E) and temperature dependence of signal at 222 nm (F) for CC-pIL at 50 µM concentration in PBS buffer.

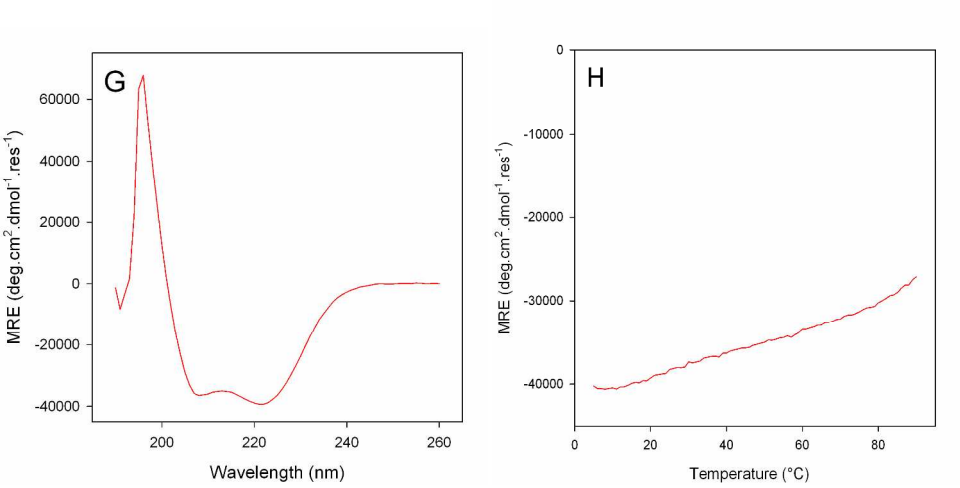


Figure S2G & H. Circular dichroism spectrum (G) and temperature dependence of signal at 222 nm (H) for CC-pII at 50 µM concentration in PBS buffer.

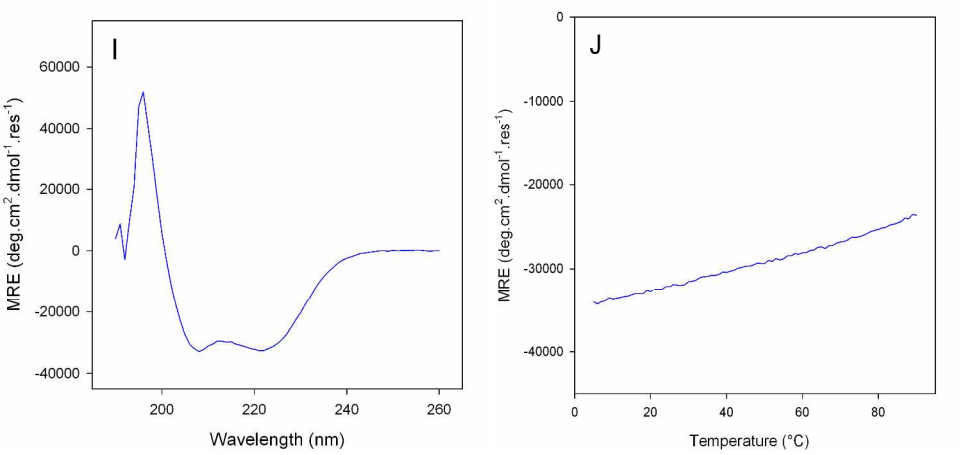


Figure S2I & J. Circular dichroism spectrum (I) and temperature dependence of signal at 222 nm (J) for CC-pLI at 50 µM concentration in PBS buffer.

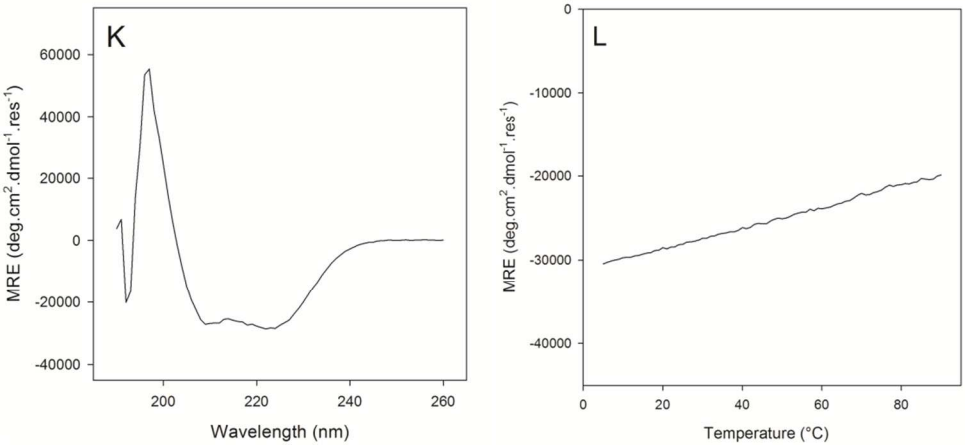


Figure S2K & L. Circular dichroism spectrum (K) and temperature dependence of signal at 222 nm (L) for GCN4-pIL at 50 µM concentration in PBS buffer.

Analysis of B-Factors

B-Factors were examined as a function of sequence position for the peptides in this study.

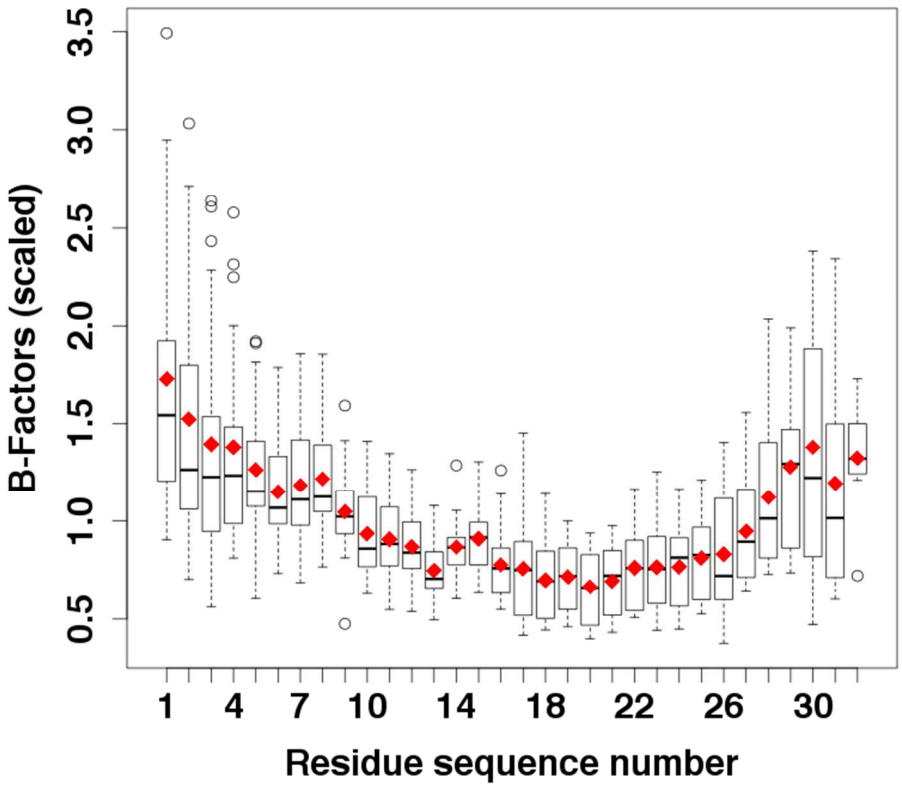


Figure S3. Variation of B-factor relative to sequence position for main chain C-alpha atoms in the Basis-set structures. B-factors were extracted from the main chain C-alpha atoms of the following structures: CC-Tet, PDB I.D.3R4A; CC-Tri-I13N, 4DZK; CC-Tri, 4DZL; CC-Di, 4DZM; CC-pIL, 4DZN. B-factors of individual structures were scaled to the mean B-factor in the respective set and B-factor distributions per residue over all structures are plotted against their sequence number as Box-and-Whisker plots. The box contains the inter-quartile-range (IQR) between 25% (1st quartile) and 75% (3rd quartile) of the data, while the whiskers at the top and bottom show the maximum and minimum values, respectively, excluding outliers (shown as hollow circles). Outliers are defined as less than 1.5 times the 1st or more than 1.5 times the 3rd quartile. The mean and median B-factors per residue are displayed as red diamonds and black bars, respectively. The B-factors per residue show an even distribution over the length of the sequence with a slightly narrower IQR in the two central heptads (residues 8-21) as compared to the C- and N-terminal heptads. In addition, both B-factor mean and median clearly decrease from the two termini towards the center of the coiled-coil chains, where mean B-factors are up to 40% lower than at the termini. Elevated B-factors for the termini as compared to the central regions of the helices indicate a higher mobility in accordance with possible fraying of the chain ends. Data analysis and plotting were carried out using R (R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.)

Circular Dichroism Spectroscopy in Presence of Guanidine Hydrochloride

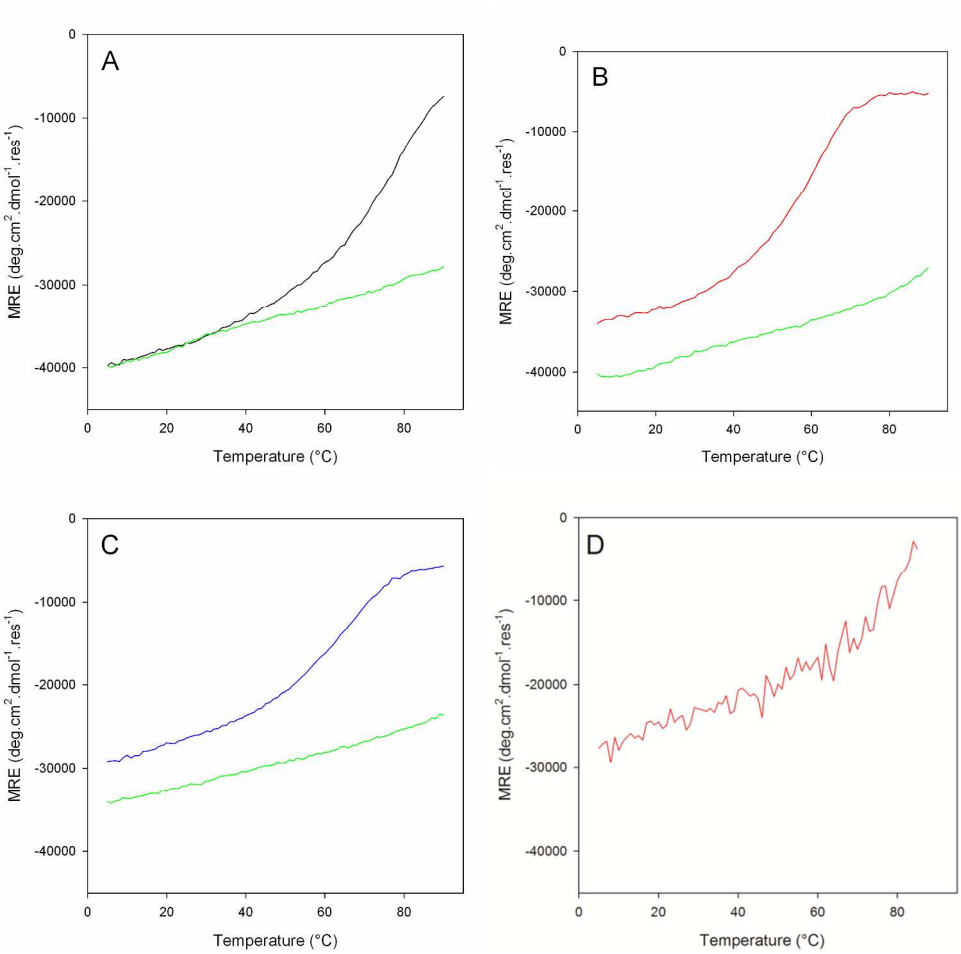


Figure S4A-D. Thermal denaturation of CC-pIL (A), CC-pII (B) and CC-pLI (C) at 50 µM peptide concentration in PBS buffer in the presence of 3M guanidine hydrochloride as monitored by circular dichroism spectroscopy at 222 nm. In all cases the green line shows the equivalent temperature dependent trace in the absence of any denaturant. Panel D shows the thermal denaturation profile for CC-pII at 1 µM peptide concentration. The beginning of a sigmoidal curve may be seen, but this is at the limit of the sensitivity for the apparatus available. Similar behaviour is observed for CC-pLI and CC-pIL

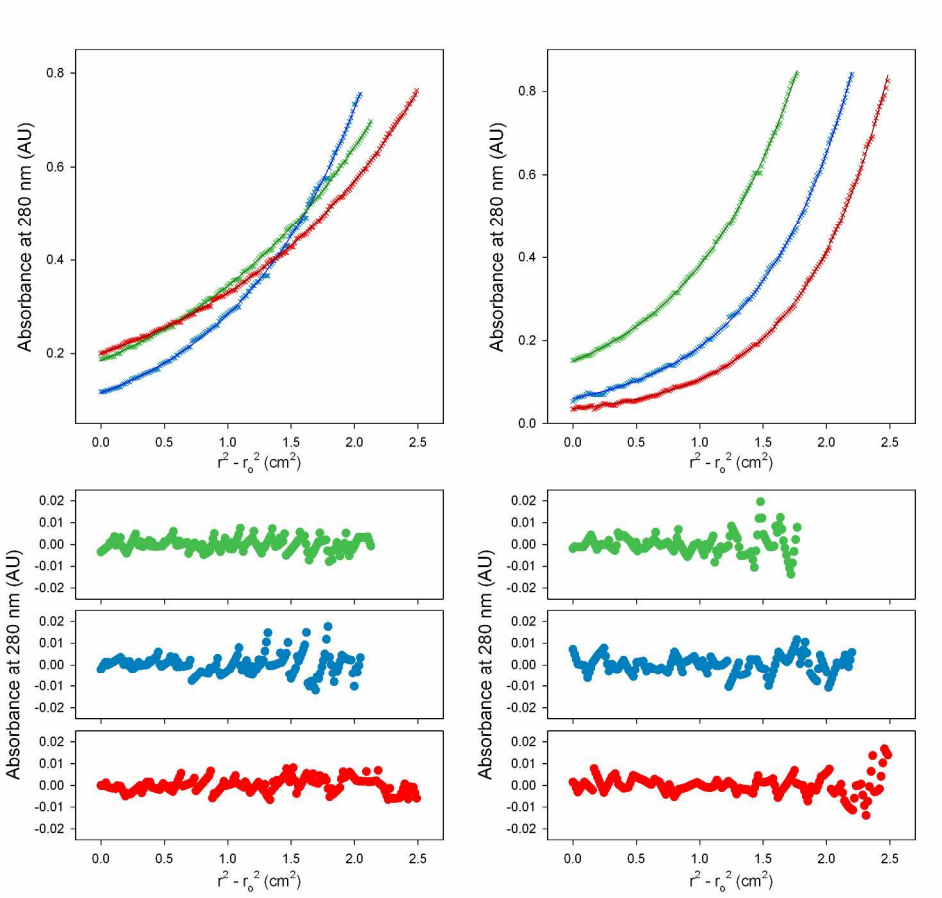


Figure S5A. AUC data (crosses) and fits (lines), and residuals for CC-pIL-I17N (left) and CC-pIL (right). Rotor speeds for CC-pIL-I17N were 40,000 rpm (green), 43,000 rpm (blue) and 50,000 rpm (red), and for CC-pIL were 43,000 rpm (green), 46,000 rpm (blue) and 50,000 rpm (red). The fits shown for CC-pIL-I17N are for a single ideal species with a mass of 7318 Da (2.05 × monomer mass), those for CC-pIL are for a single ideal species of mass 10560 (2.96 × monomer mass).

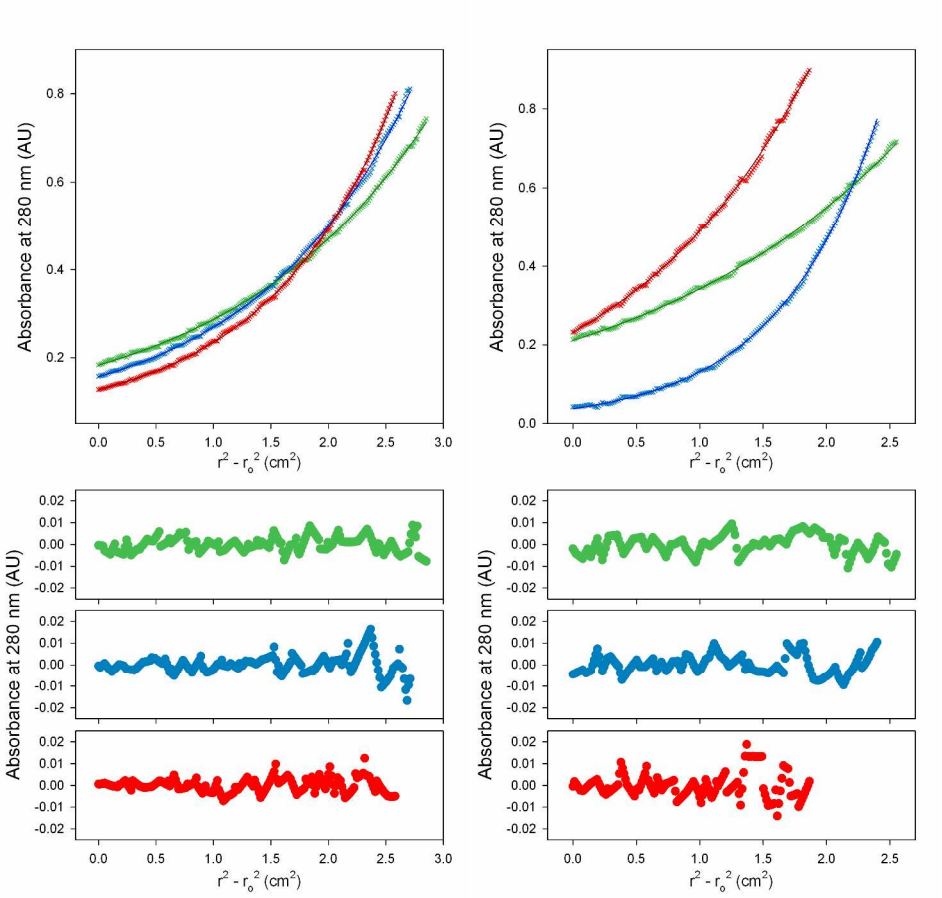


Figure S5B. AUC data (crosses) and fits (lines), and residuals for CC-pII-I13N (left) and CC-pII (right). Rotor speeds for CC-pIL-I17N were 36,000 rpm (green), 40,000 rpm (blue) and 43,000 rpm (red), and for CC-pII were 36,000 rpm (green), 40,000 rpm (blue) and 50,000 rpm (red). The fits shown for CC-pII-I13N are for a single ideal species with a mass of 10850 Da (3.12 × monomer mass), those for CC-pII are for a single ideal species of mass 10020 (2.90 × monomer mass).

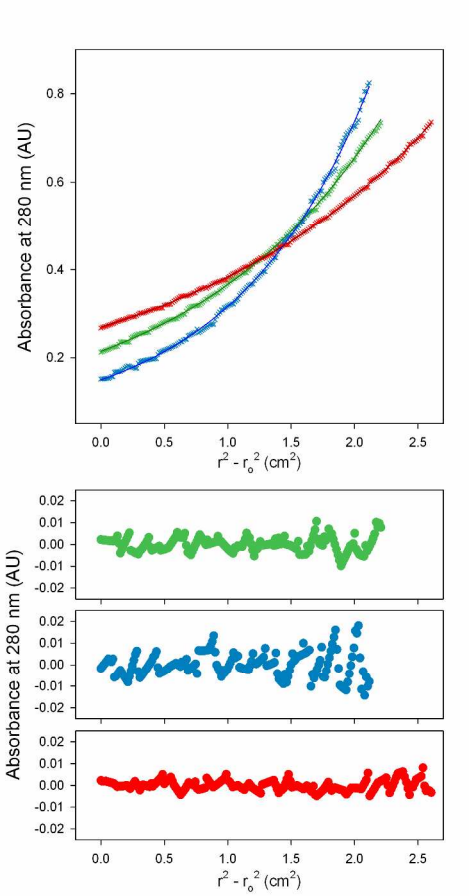


Figure S5C. AUC data (crosses) and fits (lines), and residuals for CC-pLI . Rotor speeds were 26,000 rpm (green), 30,000 rpm (blue) and 36,000 rpm (red). The fits are for a single ideal species with a mass of 13240 Da (3.92 × monomer mass).

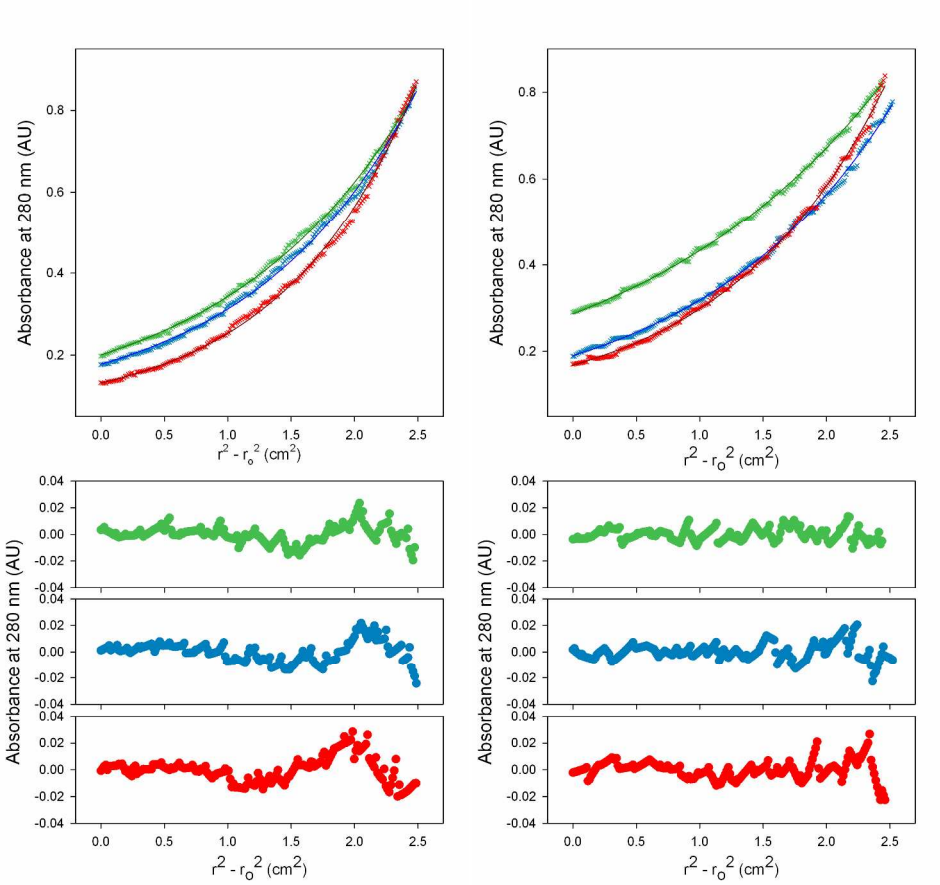


Figure S5D. AUC data (crosses) and fits (lines), and residuals for GCN4-pIL at 250 μM concentration (left) and at 500 μM concentration (right). Rotor speeds for 250 μM were 34,000 rpm (green), 36,000 rpm (blue) and 40,000 rpm (red), and for 500 μM were 34,000 rpm (green), 36,000 rpm (blue) and 40,000 rpm (red). The fits shown for 250 μM are for a single ideal species with a mass of 11070 Da (2.74 × monomer mass), those for 500 μM are for a single ideal species of mass 13600 (3.37 × monomer mass). The systematic deviation exhibited by the residuals indicates that the single ideal species model is likely to be unsuitable, however use of more-complex models did not achieve a better fit.

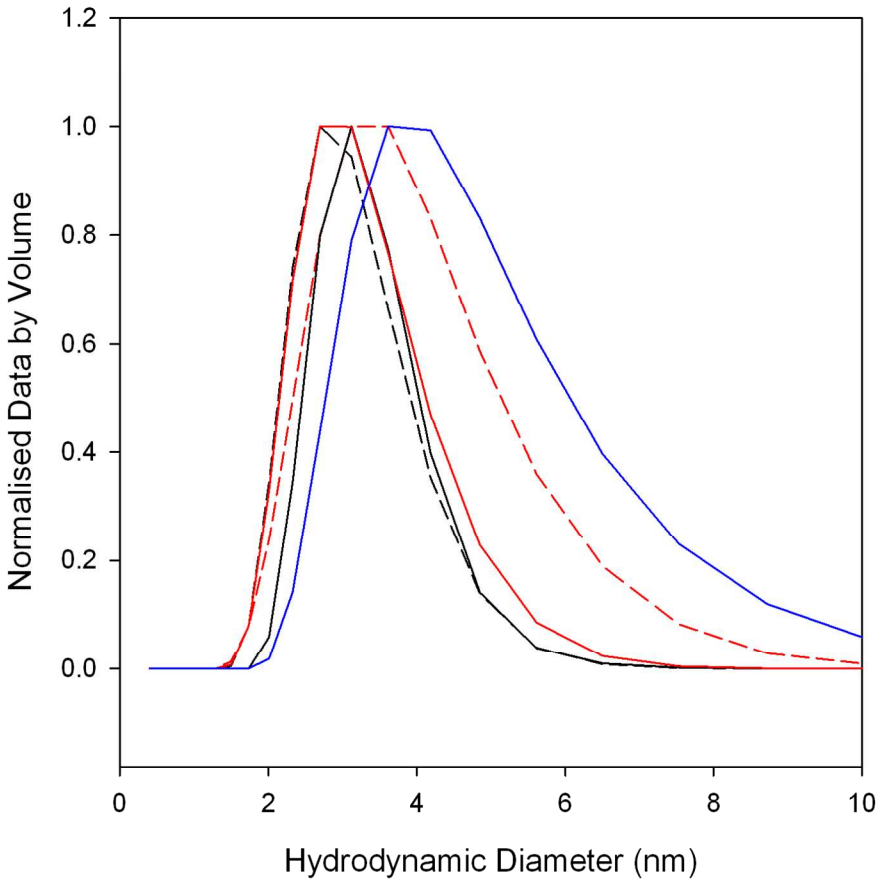


Figure S6. Dynamic light scattering profiles. Key: CC-pIL, solid black lines; CC-pII, solid red lines; CC-pLI, solid blue lines; CC-pIL-I17N, broken black lines; and CC-pII-I13N, broken red lines. All samples were at a peptide concentration of 50 μM and were in PBS (pH 7.4).

Peptide	Salt	Buffer	pH	Precipitant
CC-pIL-I17N-W22Φ	0.2 M NaCl	0.1 M Bis Tris	5.5	25 % w/v PEG 3350
CC-pIL-W22Φ	0.2 M $NH_4H_2PO_4$	0.1 M Tris	8.5	50% v/v MPD
CC-pII-W22Φ	0.2 M $C_4H_4KNaO_6$	None	N/D	20 % w/v PEG 3350
CC-pII-I13N	2.0 M $(NH_4)_2SO_4$	None	N/D	5 % v/v 2-propanol
CC-pLI	None	0.1 M Tris	7.5	3 M HCOONa

Table S4. Crystallisation conditions

	CC-pIL-W22Φ	CC-pIL-I17N-W22Φ
PDB accesion code	4DZN	4DZM
crystal parameters		
space group	$P2_12_12_1$	$P6_2$
unit cell a, b, c (Å)	24.7, 40.9, 87.4	23.5, 23.5, 189.6
data collection statistics		
wavelength (Å)	1.7	1.7
resolution(Å)	41.0 - 1.6 (1.7 - 1.6)	19.6 - 1.9 (2.0 - 1.9)
total reflections	72412 (9837)	64111 (7470)
unique reflections	11991 (1691)	8395 (1172)
R_{merge}	9.5 (29.6)	11.5 (43.9)
mean I/σ(I)	11.9 (4.6)	13.1 (3.9)
completeness (%)	96.1 (95.4)	97.1 (87.5)
redundancy	6.0 (5.8)	20.8 (20.9)
Wilson B-factor	16	19.2
refinement statistics		
peptide molecules per A.U.	3	2
residues	96	64
water molecules	78	32
ligand atoms	---	---
total number of atoms	833	544
R_{cryst}/R_{free} (%)	15.5/19.8	19.2/24.4
rmsd of bond lengths (Å)	0.0116	0.0118
rmsd of bond angles (deg)	1.115	1.665

Table S5A. Data collection and refinement statistics

	CC-pII-W22Φ	CC-pII-I13N
PDB accession code	4DZL	4DZK
crystal parameters		
space group	$P2_12_12$	P321
unit cell a, b, c (Å)	102.8, 108.4, 41.6	38.2, 38.2, 44.3
data collection statistics		
wavelength (Å)	1.7	0.98
resolution(Å)	20.0 - 2.3 (2.4 - 2.3)	44.3 - 1.8 (1.9 - 1.8)
total reflections	269604 (12038)	36152 (5371)
unique reflections	21306 (1997)	3777 (541)
R_{merge}	18.0 (56.0)	10.4 (27.7)
mean I/σ(I)	27.8 (2.0)	15.5 (7.1)
completeness (%)	99.6 (99.5)	99.7 (100.0)
redundancy	12.7 (6.4)	9.6 (9.9)
Wilson B-factor	35.6	24.4
refinement statistics		
peptide molecules per A.U.	12	1
residues	361	29
water molecules	188	14
ligand atoms	---	1
total number of atoms	2937	250
R_{cryst}/R_{free} (%)	24.8/28.8	23.2/26.9
rmsd of bond lengths (Å)	0.0083	0.0068
rmsd of bond angles (deg)	1.021	1.033

Table S5B. Data collection and refinement statistics

Analysis of Core-packing Angles

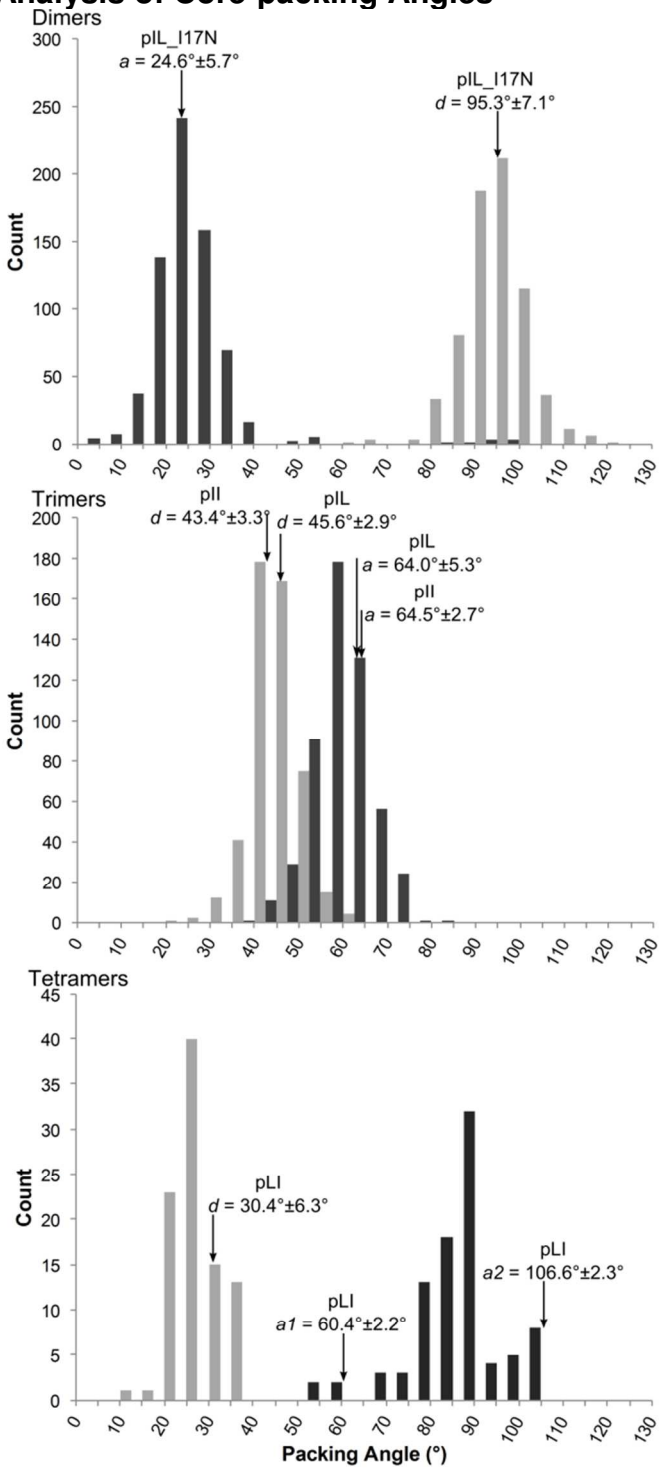


Figure S7. Analysis of packing angles of core residues at ***a*** (dark grey) and ***d*** (light grey) heptad positions in dimers, trimers and tetramers.These angles were generated by SOCKET and were measured as the C_α-C_β bond vector of the knob residues to the C_α-C_α vector of the hole residues on the partnering helix. Mean packing angles for basis-set peptides are indicated by arrows, with values and standard deviations given. Note that the packing angles for the ***a*** positions of the tetramer, CC-pLI, fall into two different distributions; the mean and standard deviation of each is given. This is because of some deviation from C_4 symmetry in the structure.

SOCKET Analyses of Crystal Structures

SOCKET outputs from analysis of the crystal structures for CC-pIL-I17N, CC-pII-I13N, CC-pIL, CC-pII and CC-pLI' The X-ray crystal structures for peptides CC-pIL-I17N, CC-pII-I13N, CC-pIL, CC-pII and CC-pLI were subjected to a SOCKET (Walshaw, J.; Woolfson, D. N. *J Mol Biol* **2001**, *307*, 1427) analysis using a cut-off of 7.0 Å for identifying knobs-into-holes interactions.

Sequence	EIAALKQEIAALKKENAALKXEIAALKQG
Register	abcdefgabcdefgabcdefgabcd
Helix 1	----Y---Y--Y---Y--Y---Y------
Helix 2	----X---X--X---X--X---X------

Table S6A. SOCKET output for CC-pIL-I17N. Heptad assignment for Helix 1-2 from Residue 3-27.

Sequence	EIAAIKQEIAANKKEIAAIKWEIAAIK
Register	abcdefgabcdefgabcdefgabcd
Helix 1	----Z---Y--Z---Y--Z---Y----
Helix 2	----X---Z--X---Z--X---Z----
Helix 3	----Y---X--Y---X--Y---X----

Table S6B. SOCKET output for CC-pII-I13N. Heptad assignment for Helix 1-3 from Residue 3-27

Sequence	EIAALKQEIAALKKEIAALKXEIAALKQGY
Register	abcdefgabcdefgabcdefgabcd
Helix 1	----Z---Y--Z---Y--Z---Y-------
Helix 2	----X---Z--X---Z--X---Z--X----
Helix 3	----Y---X--Y---X--Y---X--Y----

Table S6C. SOCKET output for CC-pIL. Heptad assignment for Helix 1-3 from Residue 3-27.

Sequence	EIAAIKQEIAAIKKEIAAIKXEIAAIK
Register	defgabcdefgabcdefgabcd
Helix 1	----Z---Y--Z---Y--Z---Y----
Helix 2	--------Z--X---Z--X---Z----
Helix 3	----Y---X--Y---X--Y---X----

Table S6D. SOCKET output for CC-pII. Heptad assignment for Helix 1-3 from Residue 6-27

Sequence	ELAAIKQELAAIKKELAAIKWELAAIK
Register	bcdefgabcdefgabcdefgabcde
Helix 1	--------Z--XZ--Z---Z--Z----
Helix 2	----Y--YW--Y--YW--Y--Y-----
Helix 3	-----X--X--ZX--X---X--X----
Helix 4	----W--WY--W--WY--W--------

Table S6E. SOCKET output for CC-pLI. Heptad assignment for Helix 1-2 from Residue 3-28

TWISTER Analysis of Crystal Structures

	CC-pIL-Φ	CC-pIL-I17N-Φ	CC-pII-Φ	CC-pII-I13N	CC-Tet-Φ
CC pitch (Å)	182.7±11.5	174.3±17.6	162.7±7.2	172.7±14.6	201.4±15.8
CC radius (Å)	6.41±0.06	4.85±0.03	6.5±0.04	6.25±0.17	7.09±0.05

Table S7. Coiled-coil parameters as determined by TWISTER analysis of crystal structures. Calculations were carried out for the two central heptads (residues 10-23). Standard deviations as given by TWISTER describe variations of coiled-coil parameters between heptad register positions along the helix.

ODF Values for *e* and *g* Positions

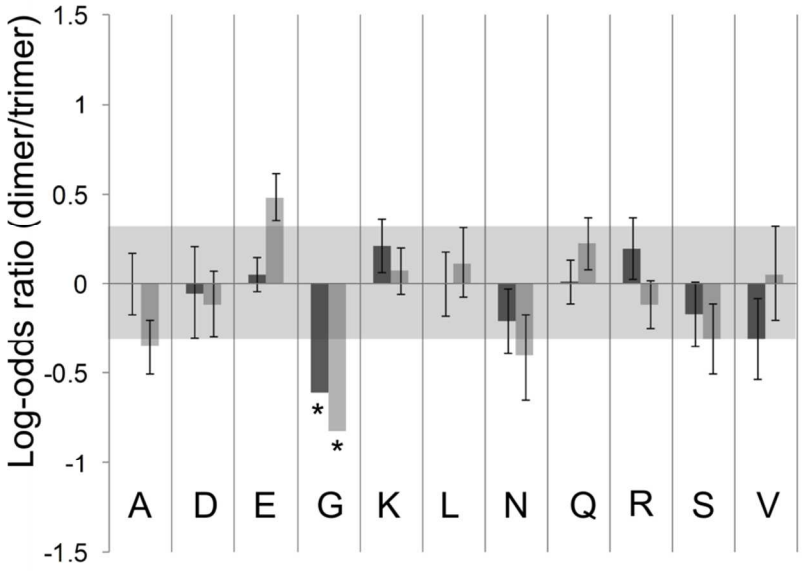


Figure S8. Oligomer-state discrimination factors (ODFs) for individual amino acids in parallel, homomeric coiled-coil dimers and trimers. ODFs were calculated as the $\log_{10}$ of the ratio of the normalized percentages of occurrence of each amino acid at the specified positions: ***e*** (dark gray bars) and ***g*** (light gray bars), in the dimer and trimer data sets. The shaded region highlights ODFs in the range +0.3 and -0.3; *i.e.*, preferences for dimer and trimer, respectively, of no more than twice the alternative oligomer state. Errors on the count data for the dimers and trimers were estimated using the normal distribution and propagated through the ODF calculation, with the error bars shown indicating 1 standard deviation. Starred data shown with no error bars failed to meet the criteria for the normal approximation to the binomial distribution due to one or more of the contributing count data elements being less than 5.

ODF Versus Coiled-coil Propensity

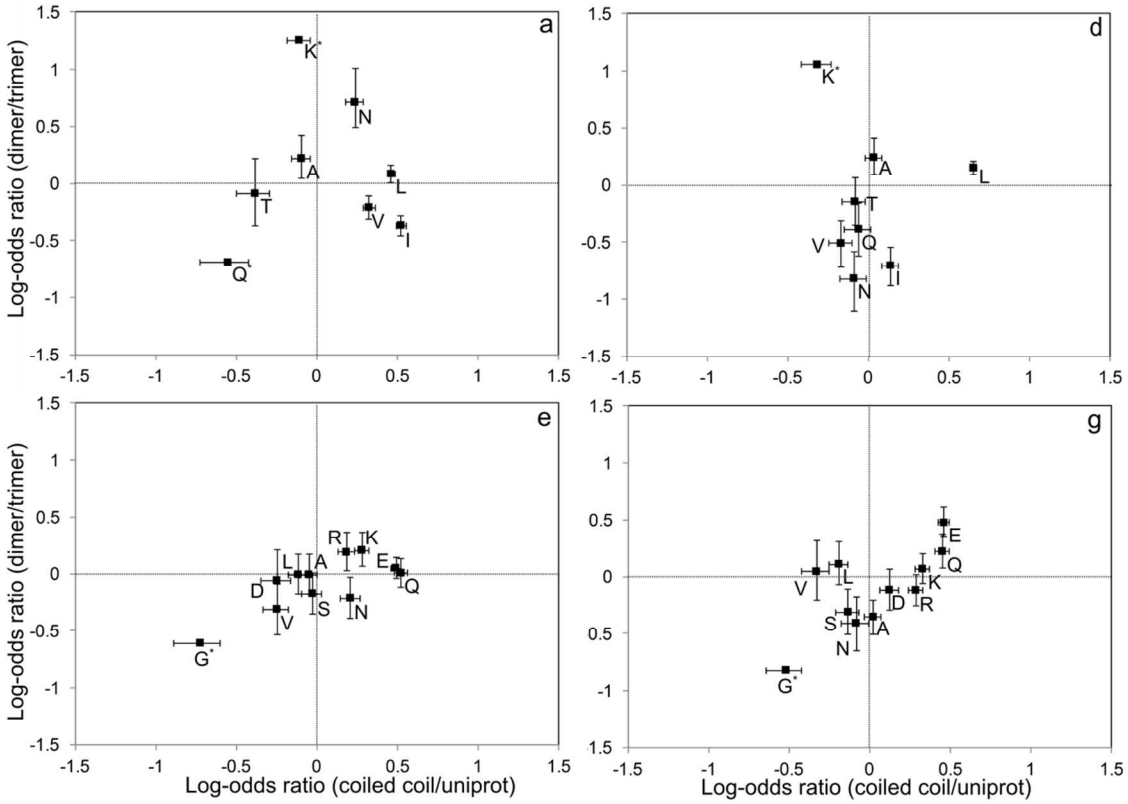


Figure S9. Comparison of oligomer-state discrimination factor (ODF) and amino acid coiled-coil propensity. ODFs were calculated as the log10 of the ratio of normalized percentages of occurrence of each amino acid at the ***a*** (top left), ***d*** (top right), ***e*** (bottom left) and ***g*** (bottom left) positions in the dimer and trimer data sets. Only amino acids that formed ≥ 5% of residues at any of the ***a***/***d*** or ***e***/***g*** positions were included. Coiled-coil propensity was calculated as the log10 of the ratio of normalized percentages of occurrence of each amino acid in a non-redundant (≤ 50% sequence identity), parallel subset of the coiled-coil database (CC+, http://coiledcoils.chm.bris.ac.uk/ccplus/search) and the proportion of that amino acid in the Uniprot sequence database. Errors in both datasets were estimated using the normal distribution and propagated through both calculations, with the error bars shown indicating 1 standard deviation. Data points marked * have no error bars on the dimer/trimer log-odds ratio value and failed to meet the criteria for the normal approximation to the binomial distribution, due to one or more of the contributing count data elements being less than 5.

Analysis of Salt-bridge Distances

Distances (Å)					
	CC-pIL**			CC-pIL-I17N*	
Potential Salt-bridge	Chain A	B	C	Mol 1	2
Glu 2 - Lys 7	**3.7**	**3.48**	**3.6**	5.68	8.57
Glu 9 - Lys 14	**3.42**	**3.66**	**3.87**	4.19	6.48
Glu 16 - Lys 21	**3.38**	**3.61**	**3.59**	4.5	5.68
Glu 23 - Lys 28	**3.61**	**3.42**	**3.75**	5.62	7.63

* Chains are symmetry related within dimers with two dimers in the ASU

** Chains are not symmetry related with one trimer in the ASU

Table S8. Measurements of putative salt-bridging interactions for CC-pIL and CC-pIL-I17N. Distances are measured between N_{ε} of Lys and C_{δ} of Glu. Values in bold are below the distance cutoff of 4 Å and are classified as salt bridges.

Structural Analysis of Asparagine Residues at *a* Positions

A set of parallel coiled coils with sequence identity <= 50% was identified from the CC+ database (http://coiledcoils.chm.bris.ac.uk). From these sequences, a set of 44 homodimeric coiled coils containing just one asparagine at an ***a*** position was identified. The local structure around the knob-into-hole examined. The distance of closest approach between the two asparagine residues was identified, and the structures were overlaid on one of the asparagine residues so that the spatial distribution of the opposing partner asparagine could be examined.

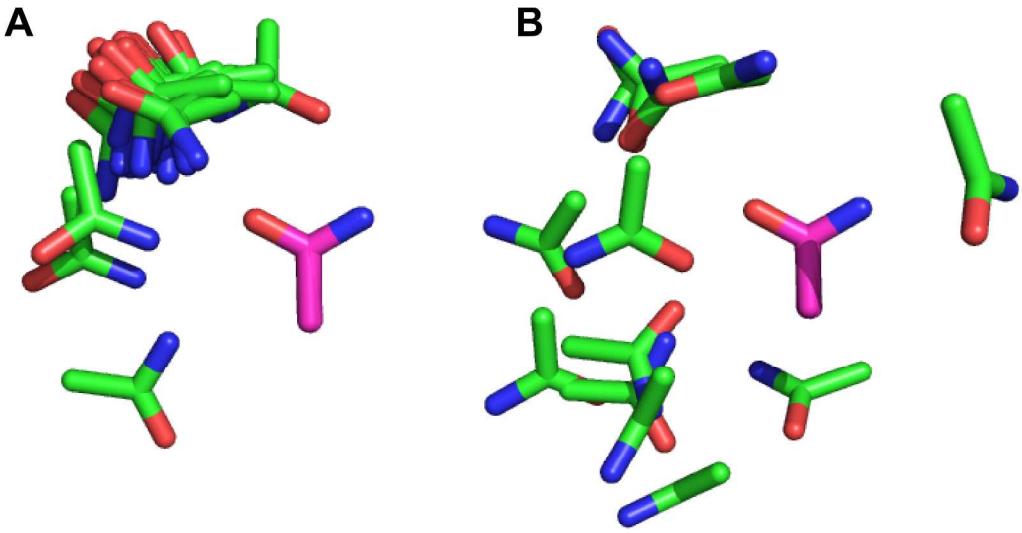


Figure S10. Superposition of asparagine at ***a*** in homodimeric coiled-coil structures. Structures were overlaid on the C_{γ}, $O_{\delta 1}$ and $N_{\delta 2}$ atoms of one of the asparagine residues (magenta). The asparagine residue on the opposing helix is coloured green. (A) Asparagine pairs where the closest O...N distance is ≤ 3.5 Å (B) Asparagine pairs where the closest O...N distance is > 3.5 Å. Structures were identified from CC+ and asparagine atom assignment was corrected using MolProbity. Structures were overlaid using the McLachlan algorithm (Mclachlan, A. D. *Acta Crystallogr A* **1982**, *38*, 871) as implemented in the program ProFit (http://www.bioinf.org.uk/software/profit/) over asparagine side-chain atoms only. Images were made with PyMOL.

Thermodynamic Analysis of Circular Dichroism Data

Through an in-depth analysis of CD data, a variety of thermodynamic parameters for the folding of coiled coil peptides can be determined (Marky, L. A.; Breslauer, K. J. *Biopolymers* **1987**, *26*, 1601). In the case of peptides containing a polar core residue (i.e. CC-pIL-I17N and CC-pII-I13N) a series of thermal denaturation experiments was performed across a range of peptide concentrations (100 µM, 50 µM, 10 µM, 5 µM, 1 µM). From each individual thermal denaturation, a two-state transition between monomer and the oligomer state determined by AUC was assumed. Dissociation constants were calculated from the following analysis. From each thermal denaturation, T_M values were determined (in Kelvin) and plotted as their reciprocal versus the natural logarithm of peptide concentration to give a linear relationship. Using these data, one can extrapolate to determine the peptide concentration that would give a T_M of the temperature of interest (defined here as $K^{T}_{½}$). This parameter may be determined as follows for the dissociation of a coiled coil (A_n) of oligomer state *n*.

$$A_n \leftrightarrow nA$$

$$K_d = \frac{[A]^n}{[A_n]} = \frac{[(1-\alpha)C_T]^n}{\alpha(C_T/n)} = \frac{nC_T^{n-1}(1-\alpha)^n}{\alpha}$$

Where C_T = total peptide concentration, *n* is the oligomeric state of the folded coiled coil, A_n, and α is the fraction of coiled coil in the folded state, such that α = 0 when fully unfolded, and α = 1 in the fully folded state. Since at the melting temperature, T_m, α = ½, it follows:

$$K^T_{1/2} = \frac{nC_T^{n-1}(1/2)^n}{1/2} = n(C_T/2)^{n-1}$$

Peptide	T_M (50 µM)	MRE† (deg $cm^2.dmol^{-1}res^{-1}$)		
CC-pIL-I17N	78.15 ℃	-37,844	6.67×10^{-5} µM	5.54×10^{-3} µM
CC-pII-I13N	56.47 ℃	-32,232	1.74×10^{-2} µM	9.42×10^{-2} µM

Table S9. Summary of CD data and dissociation constants. † Mean Residue Elipticity (MRE) at 222 nm and peptide concentration of 50 µM.

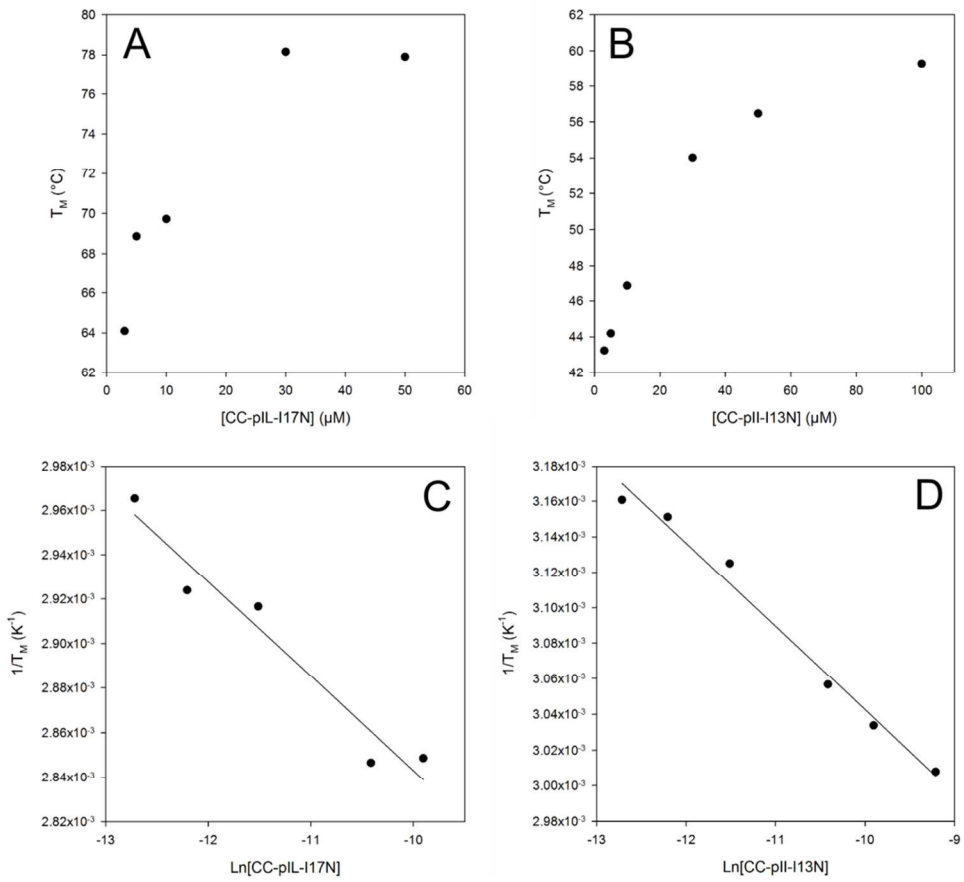


Figure S11A-D. Panels A and B: T_M from thermal denaturation plots versus peptide concentration for peptides CC-pIL-I17N (A) and CC-pII-I13N (B). Panels C and D: plots depicting $1/T_M$ versus the natural log of peptide concentration for peptide CC-pIL-I17N (C) and CC-pII-I13N (D). Straight lines of best fit are also depicted, and can be used to extrapolate a peptide concentration from a desired T_M or vice versa.

Example PComp Datasheet

Pcomp Element ID: BS2N

Pcomp Element Datasheet

General

Date Deposited	Peptide Name	Project Name	Authors	Contact
29/2/2012	CC-pIL-I17N	Basis Set	ART,GB	d.n.woolfson@bristol.ac.uk
Reference	Fletcher et. al., JACS (submitted); PDB: 4DZM			

Summary

1. Source / design rationale
Designed coiled-coil dimer for basis set project: follows Harbury rules I@a, L@d, plus N@a to further specify dimer

2. Preparation
Chemical synthesis. Standard HBTU coupling on CEM liberty system. ChemMatrix rink amide resin, acetic anhydride/pyridine capping

3. Biophysical characterization
CD, AUC, crystallographic studies, DLS

4. Biological characterization

Sequence details

Sequence	Absorbance at 280 nm (μ^{-1} cm^{-1})	Charge at pH 7.0	Molecular Weight
G EIAALKQ EIAALKK ENAALKW EIAALKQ GYY gabcdef gabcdef gabcdef gabcdef	8250	+1	3574

Method of preparation

Recombinant Expression		Chemical synthesis	☒	Other	

Level of characterization

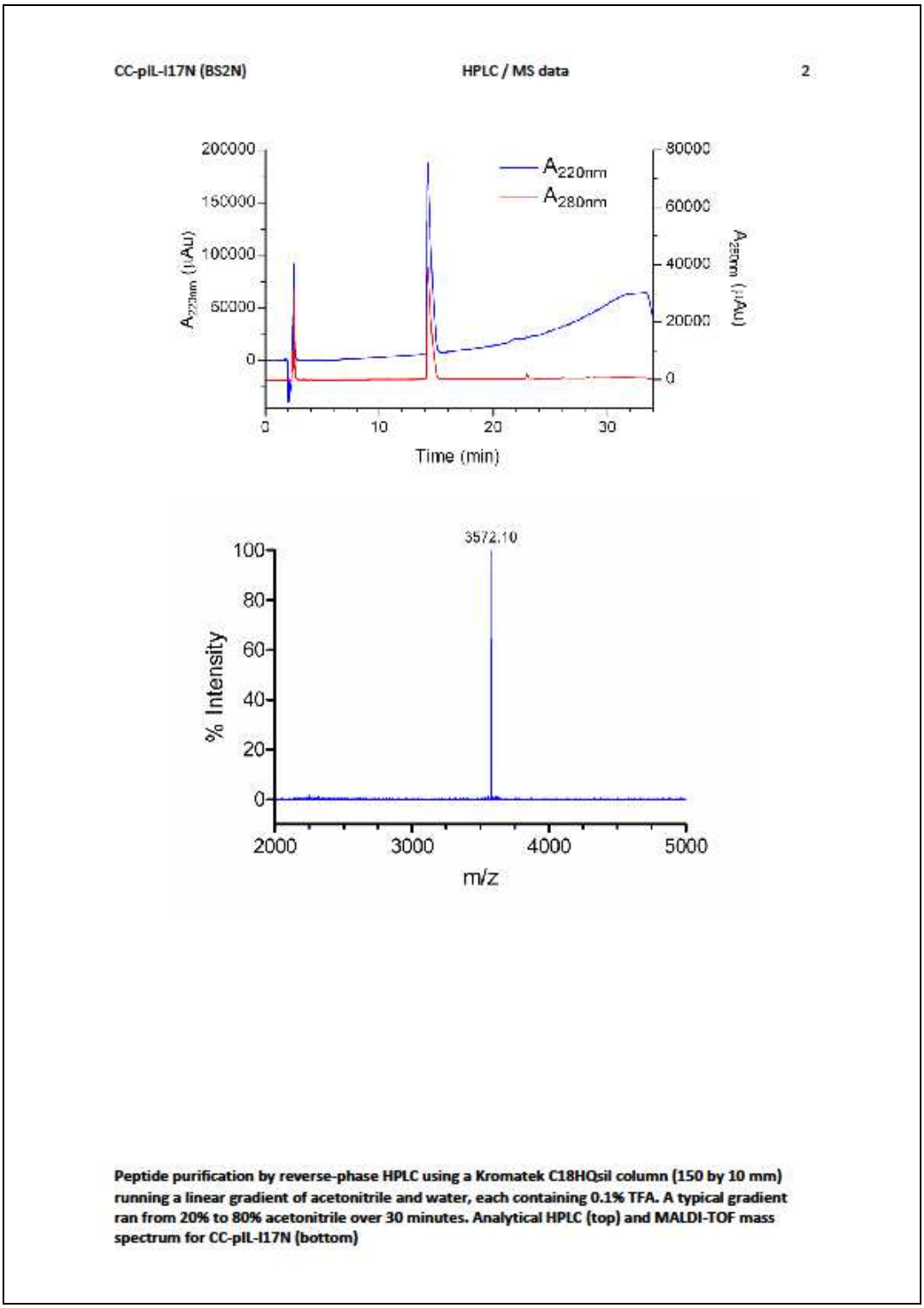
	Proof of purity (e.g. HPLC)	Proof of identity (e.g. MS)	3D structure (NMR or XRD)	Biological Characterization
Complete?	☒	☒	☒	☐
Refer to page	2	2	6	

Analysis experiments:

CD ☒	DLS ☒	AUC ☒	FT-IR ☐	NMR ☐	XRD ☒
Page:3	Page:4	Page:5	Page:	Page:	Page:6

Activity:

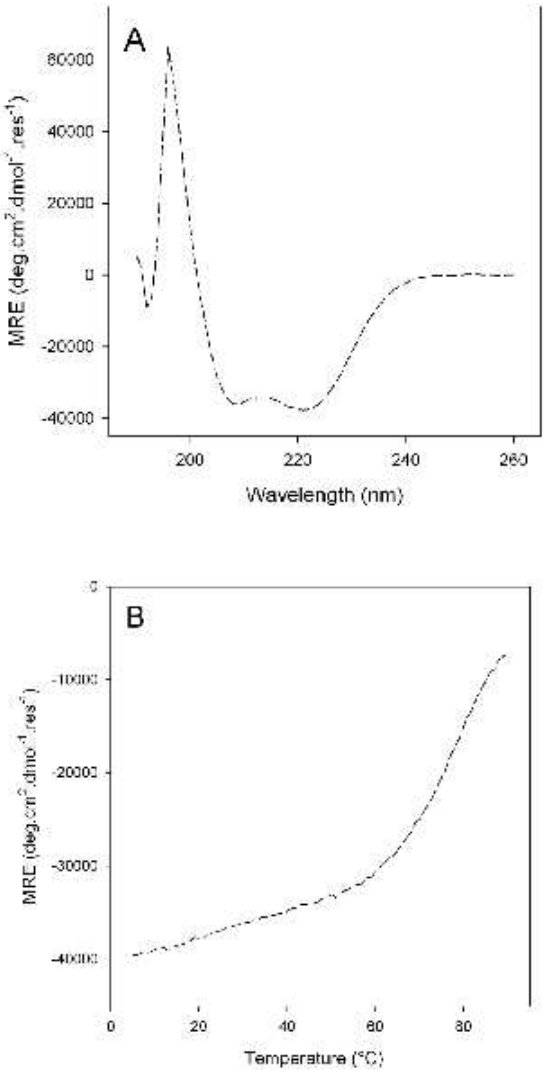
Figure S12A. Example Pcomp datasheet page 1.



Peptide purification by reverse-phase HPLC using a Kromatek C18HQsil column (150 by 10 mm) running a linear gradient of acetonitrile and water, each containing 0.1% TFA. A typical gradient ran from 20% to 80% acetonitrile over 30 minutes. Analytical HPLC (top) and MALDI-TOF mass spectrum for CC-pIL-I17N (bottom)

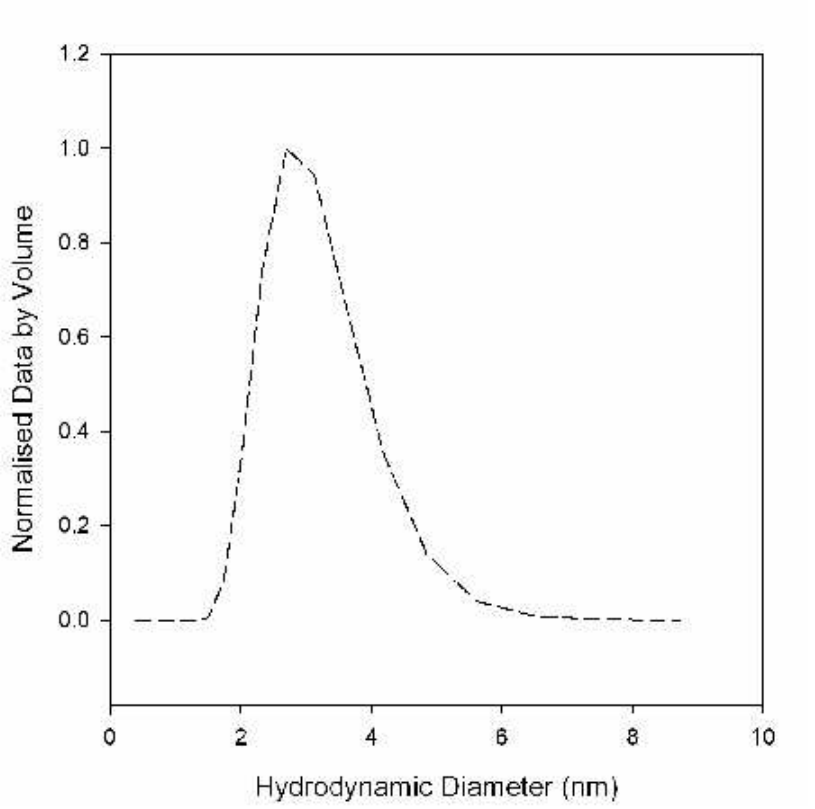
Figure S12B. Example Pcomp datasheet page 2.

CC-pIL-I17N (BS2N) CD data 3



CD spectrum (top) and temperature dependence of signal at 222nm (bottom) for CC-pIL-I17N at 50 µM concentration in PBS buffer

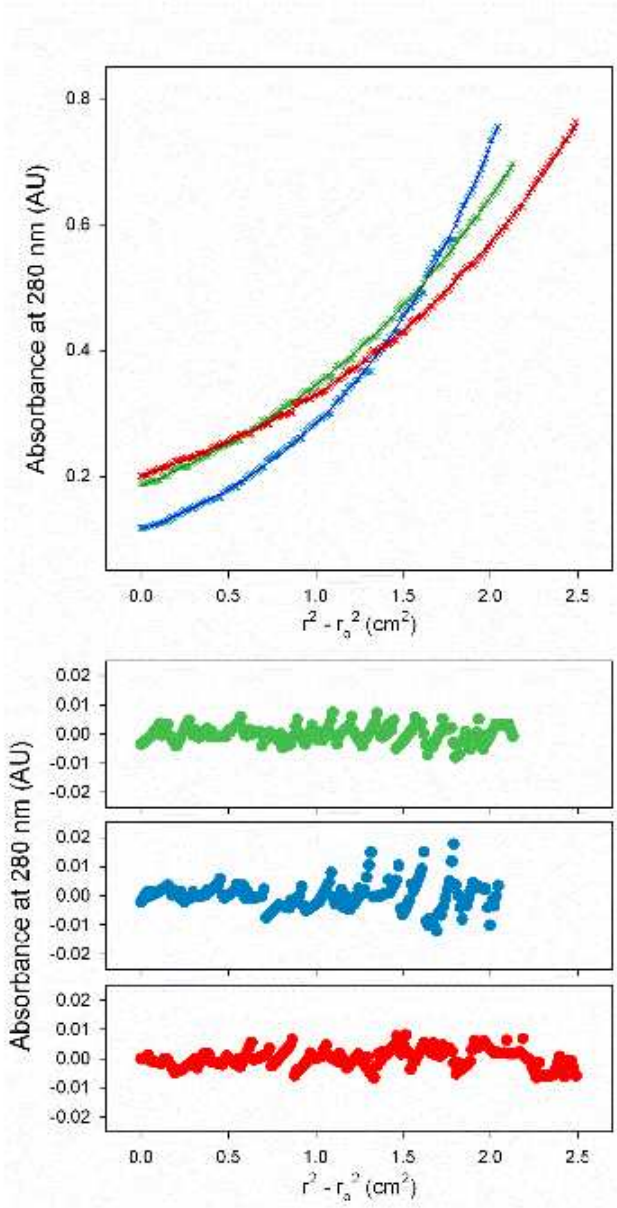
Figure S12C. Example Pcomp datasheet page 3.



Samples were prepared at 100 µM peptide concentration in PBS. Samples were centrifuged prior to analysis in order to remove any large particulate material. Measurements were made at 20°C using automated settings. The data was analyzed using the associated DTS Nano particle sizing software.

Figure S12D. Example Pcomp datasheet page 4.

CC-pIL-I17N (BS2N) AUC data 2



AUC fits for CC-pIL-I17N (top) and residuals (bottom). Rotor speeds were 46,000 rpm (green), 50,000 rpm (blue) and 53,000 rpm (red)

Figure S12E. Example Pcomp datasheet page 5.

CC-pIL-I17N (BS2N) Crystallography data 6

1. Crystallization conditions:

Peptide	Salt	Buffer	pH	Precipitant
CC-pIL-I17N	0.2 M NaCl	0.1 M Bis Tris	5.5	25 % w/v PEG 3350

2. Peptide used for crystallization:

Ac-GEIAALKQ EIAALKK EIAALKΦ EIAALKQ GYY-NH_2

where Φ is iodo-phenylalanine

3. Data collection and refinement statistics

	CC-pIL-I17N-Φ
PDB accesion code	4DZM
crystal parameters	
space group	$P6_2$
unit cell a, b, c (Å)	23.5, 23.5, 189.6
data collection statistics	
wavelength (Å)	1.7
resolution(Å)	19.6 - 1.9 (2.0 - 1.9)
total reflections	64111 (7470)
unique reflections	8395 (1172)
R_{merge}	11.5 (43.9)
mean I/σ(I)	13.1 (3.9)
completeness (%)	97.1 (87.5)
redundancy	20.8 (20.9)
Wilson B-factor	19.2
refinement statistics	
peptide molecules per A.U.	2
residues	64
water molecules	32
ligand atoms	—
total number of atoms	544
R_{cryst}/R_{free} (%)	19.2/24.4
rmsd of bond lengths (Å)	0.0118
rmsd of bond angles (deg)	1.665

Figure S12F. Example Pcomp datasheet page 6.