Supplementary Information

Local and macroscopic electrostatic interactions in single α-helices

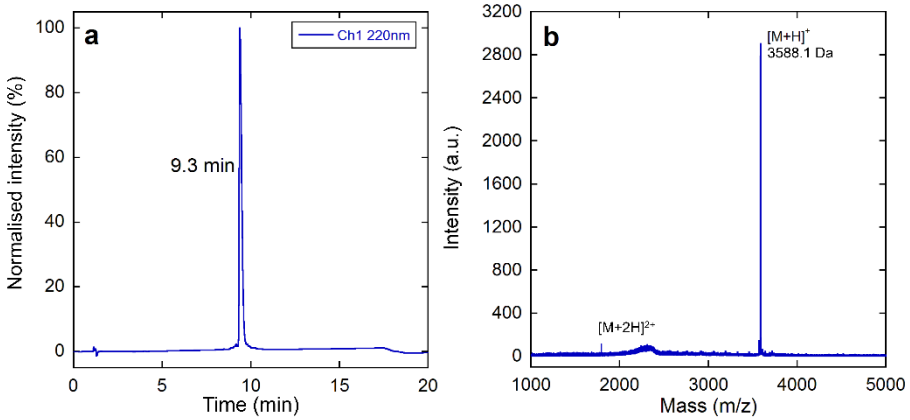
Emily G. Baker,[1] Gail J. Bartlett,[1] Matthew P. Crump,[1] Richard B. Sessions,[2] Noah Linden,[3] Charl F. J. Faul[1] and Derek N. Woolfson[1,2]

[1]School of Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS, UK
[2]School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol, BS8 1TD, UK
[3]School of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, UK

Supplementary Results

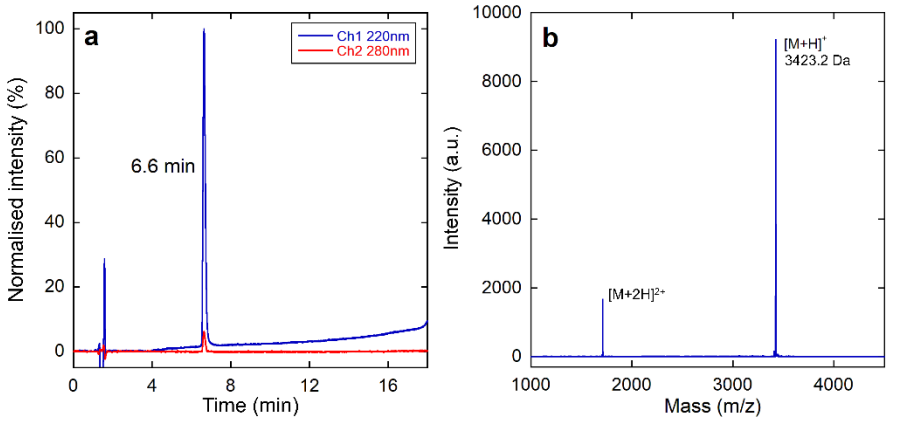
Supplementary Figures 1 – 27 (**a**) HPLC chromatograms and (**b**) MALDI-TOF spectra confirming the purity and identity of peptides in this study.

$(E_2K_2)_6$: Ac–GEEKKEEKKEEKKEEKKEEKKEEKKGYY–NH_2



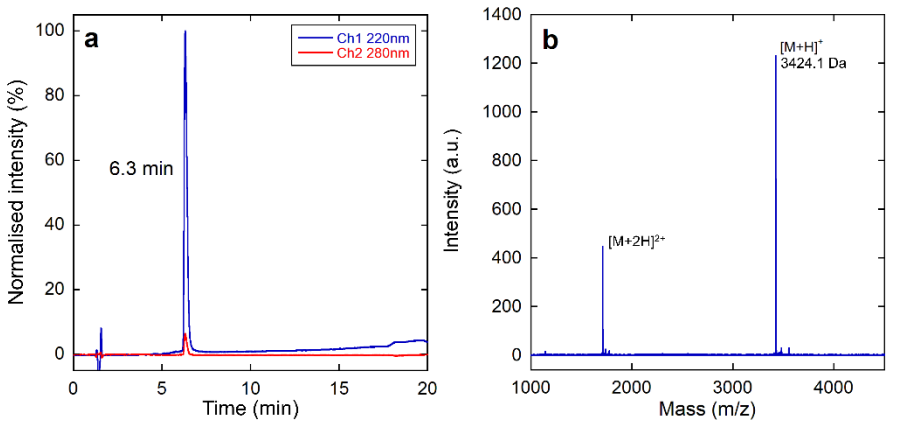
Supplementary Figure 1: (**a**) HPLC traces, gradient: 5-35%B, 16 min, 220 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 3588.0 Da)

$(EK)_{12}$: Ac–GEKEKEKEKEKEKEKEKEKEKEKEKEKGY–NH_2



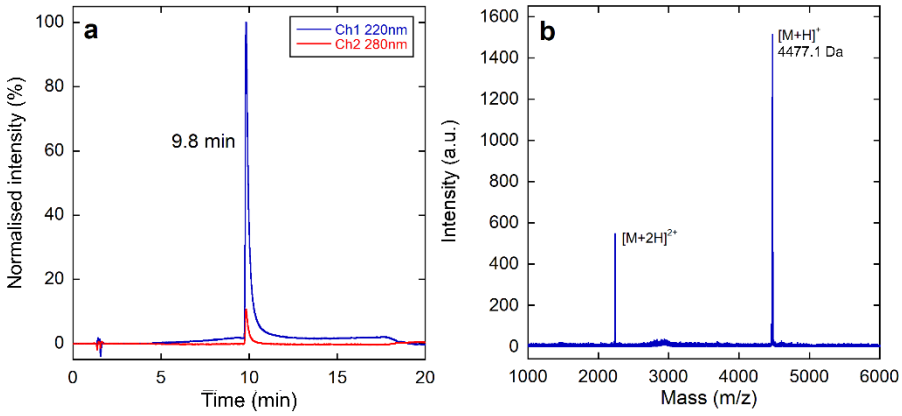
Supplementary Figure 2: (**a**) HPLC traces, gradient: 10-40%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 3424.8 Da)

$(KE)_{12}$: Ac–GKEKEKEKEKEKEKEKEKEKEKEKEKEGY–NH_2



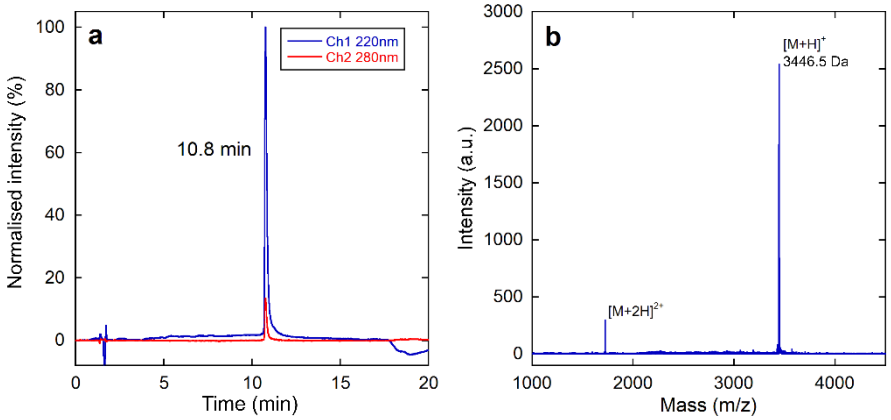
Supplementary Figure 3: (**a**) HPLC traces, gradient: 10-40%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 3424.8 Da)

$(E_4K_4)_4$: Ac-GEEEEKKKKEEEEKKKKEEEEKKKKEEEEKKKKGW-NH_2



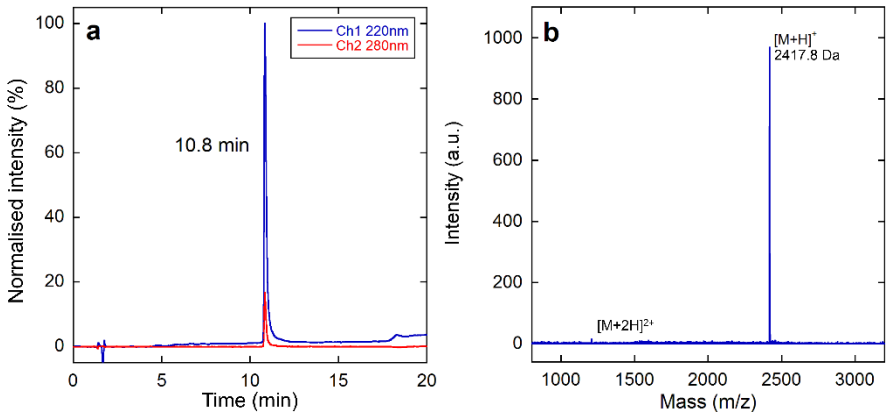
Supplementary Figure 4: (**a**) HPLC traces, gradient: 5-35%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 4477.0 Da)

$(E_4K_4)_3$: Ac-GEEEEKKKKEEEEKKKKEEEEKKKKGW-NH_2



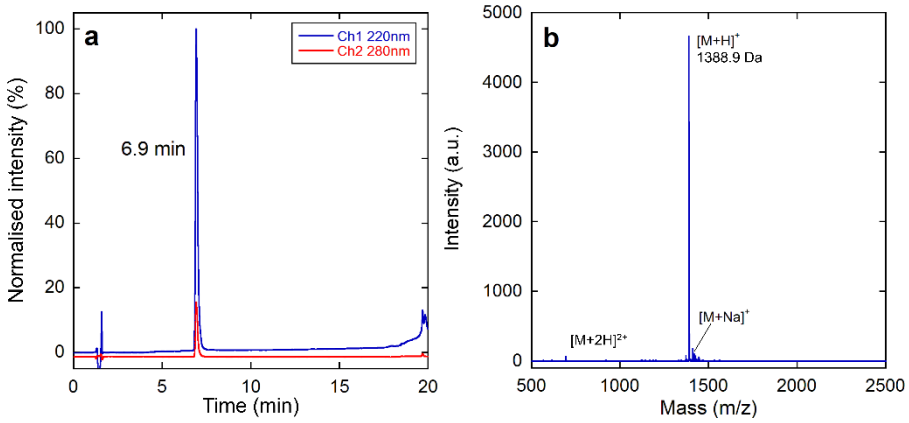
Supplementary Figure 5: (**a**) HPLC traces, gradient: 5-35%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 3447.8 Da)

$(E_4K_4)_2$: Ac-GEEEEKKKKEEEEKKKKGW-NH_2



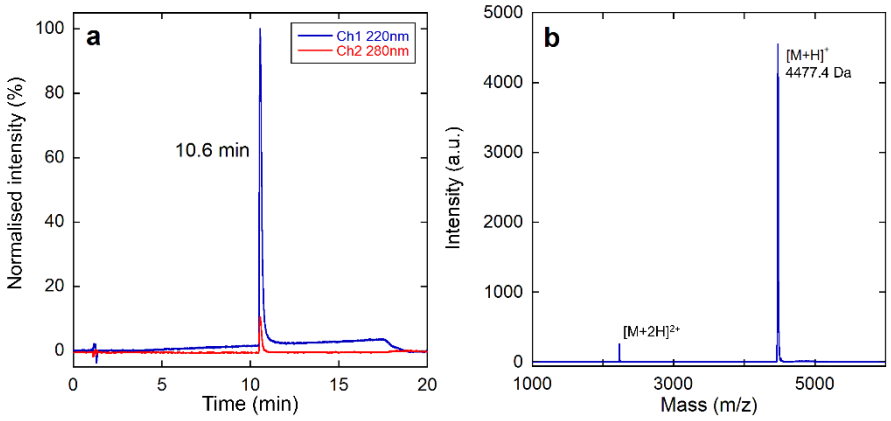
Supplementary Figure 6: (**a**) HPLC traces, gradient: 5-35%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 2418.7 Da)

(E_4K_4): Ac–G**EEEE**KKKKGW–NH_2



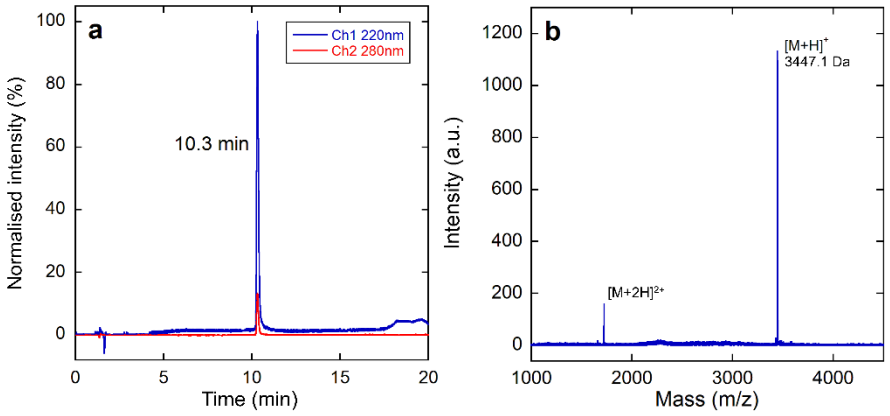
Supplementary Figure 7: (**a**) HPLC traces, gradient: 5-45%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 1389.5 Da)

($K_4E_4)_4$: Ac–G**KKKK**EEEEKKKKEEEEKKKKEEEEKKKKEEEEGW–NH_2



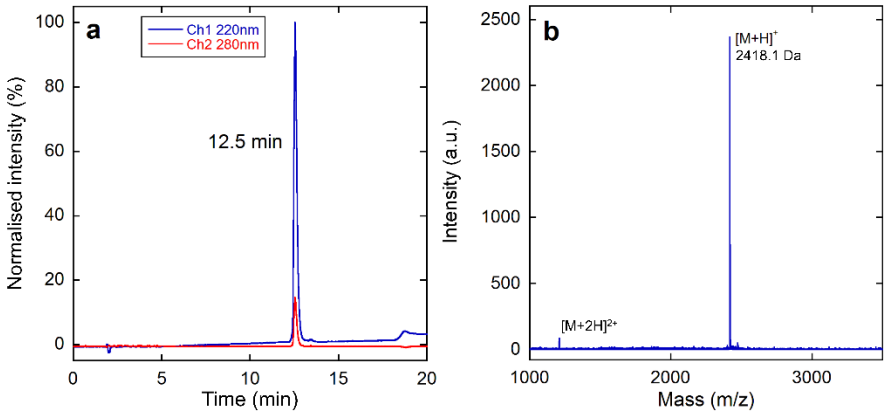
Supplementary Figure 8: (**a**) HPLC traces, gradient: 5-35%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, linear mode ($[M+H]^+_{calc}$: 4477.0 Da)

($K_4E_4)_3$: Ac–G**KKKK**EEEEKKKKEEEEKKKKEEEEGW–NH_2



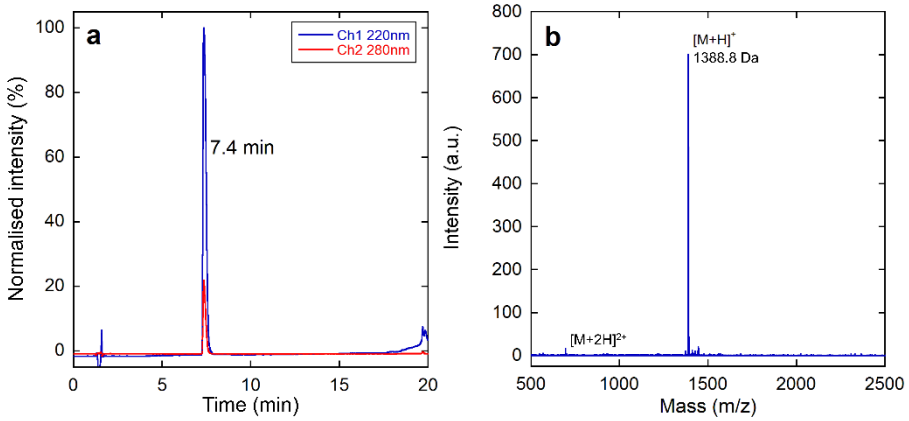
Supplementary Figure 9: (**a**) HPLC traces, gradient: 5-35%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 3447.8 Da)

$(K_4E_4)_2$: Ac–GKKKKEEEEKKKKEEEEGW–NH_2



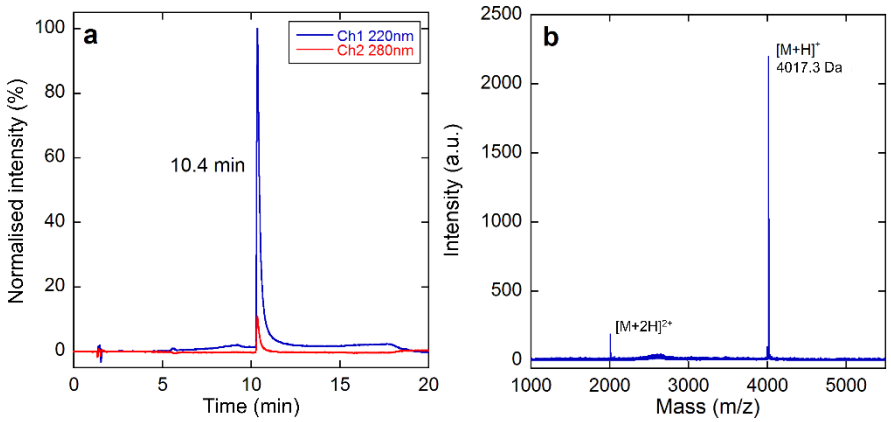
Supplementary Figure 10: (**a**) HPLC traces, gradient: 5-35%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 2418.7 Da)

(K_4E_4): Ac–GKKKKEEEEGW–NH_2



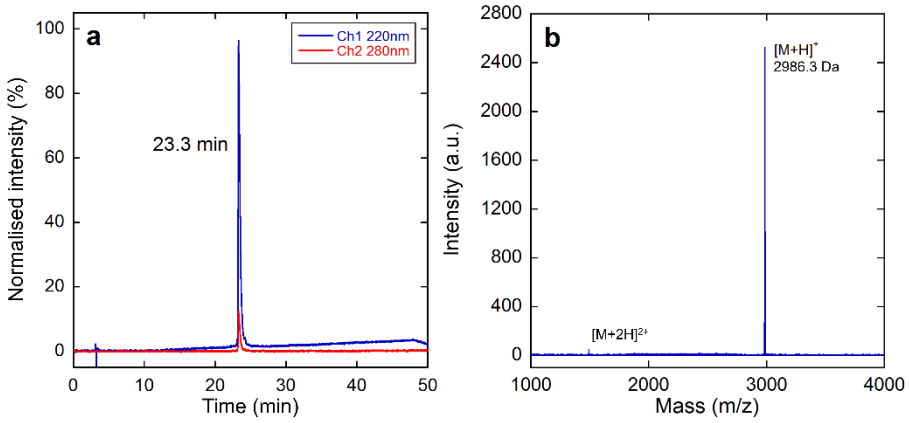
Supplementary Figure 11: (**a**) HPLC traces, gradient: 5-45%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 1389.5 Da)

$A_4(E_4K_4)_3A_4$: Ac–GAAAAEEEEKKKKEEEEKKKKEEEEKKKKAAAAGW–NH_2



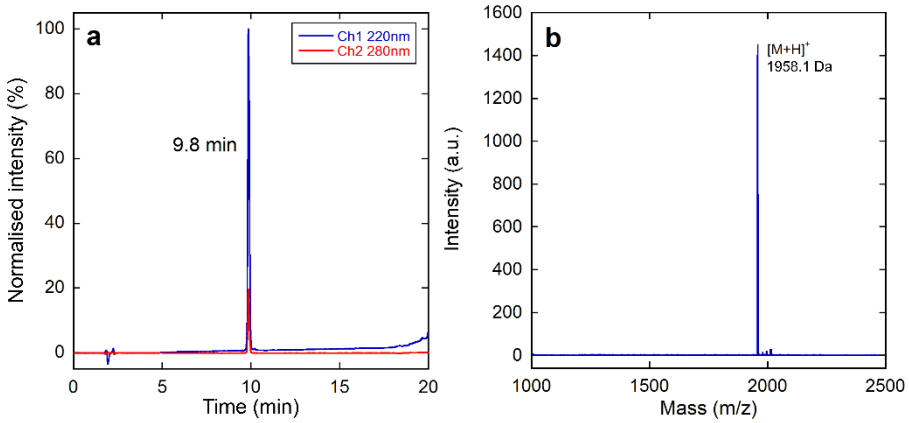
Supplementary Figure 12: (**a**) HPLC traces, gradient: 10-35%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 4016.5 Da)

$A_4(E_4K_4)_2A_4$: Ac-GAAAAEEEEKKKKEEEEKKKKAAAAGW-NH_2



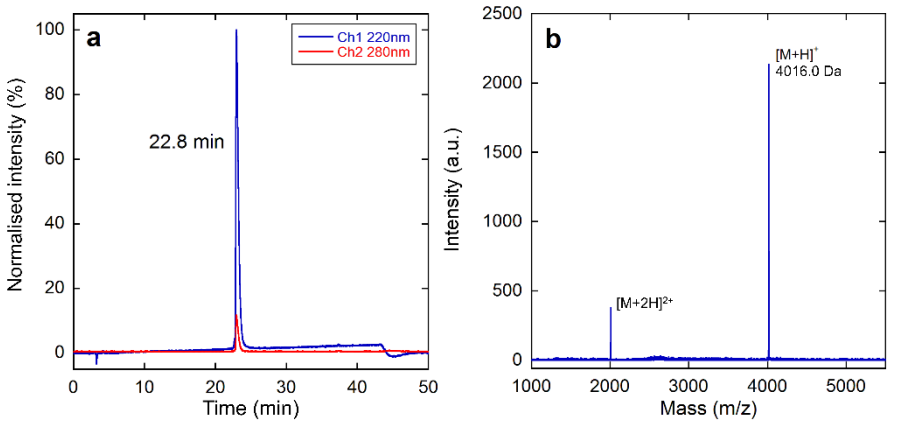
Supplementary Figure 13: (**a**) HPLC traces, gradient: 10-35%B, 37 min, 220 nm and 280 nm, reverse-phase C8 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 2987.3 Da)

$A_4(E_4K_4)A_4$: Ac-GAAAAEEEEKKKKAAAAGW-NH_2



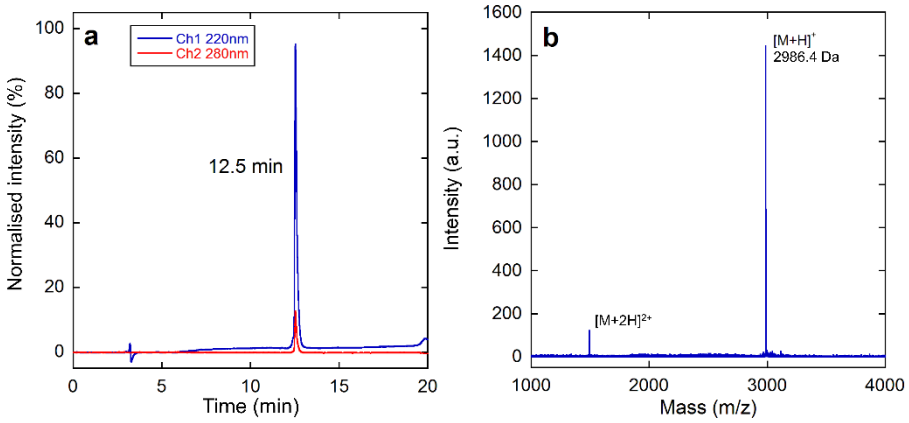
Supplementary Figure 14: (**a**) HPLC traces, gradient: 10-50%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 1958.1 Da)

$A_4(K_4E_4)_3A_4$: Ac-GAAAAKKKKEEEEKKKKEEEEKKKKEEEEAAAAGW-NH_2



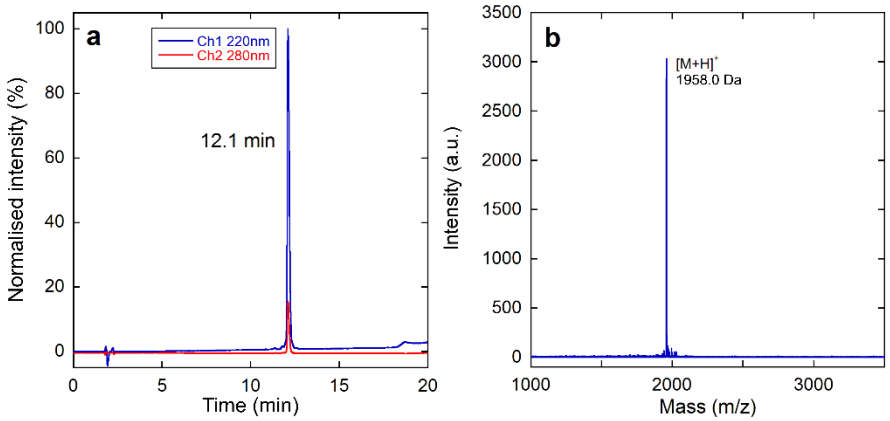
Supplementary Figure 15: (**a**) HPLC traces, gradient: 5-35%B, 37 min, 220 nm and 280 nm, C8 reverse-phase column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 4016.5 Da)

$A_4(K_4E_4)_2A_4$: Ac-GAAAAKKKKEEEEKKKKEEEEAAAAGW-NH_2



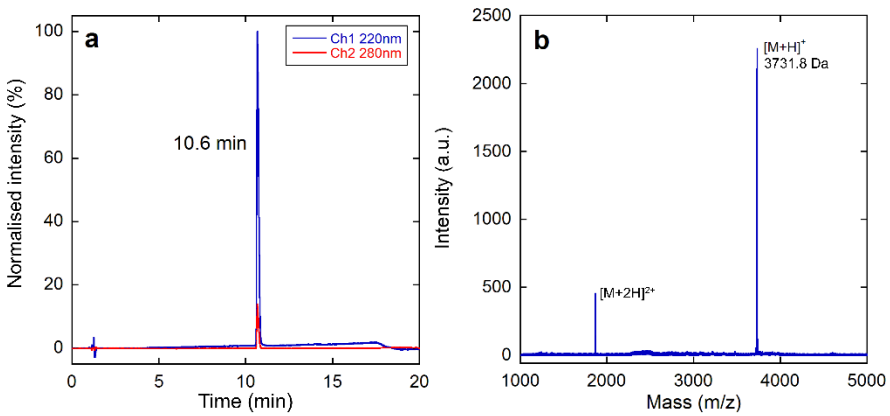
Supplementary Figure 16: (**a**) HPLC traces (gradient: 5-45%B, 16 min, 220 nm and 280 nm, C18 reverse-phase column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 2987.3 Da)

$A_4(K_4E_4)A_4$: Ac-GAAAAKKKKEEEEAAAAGW-NH_2



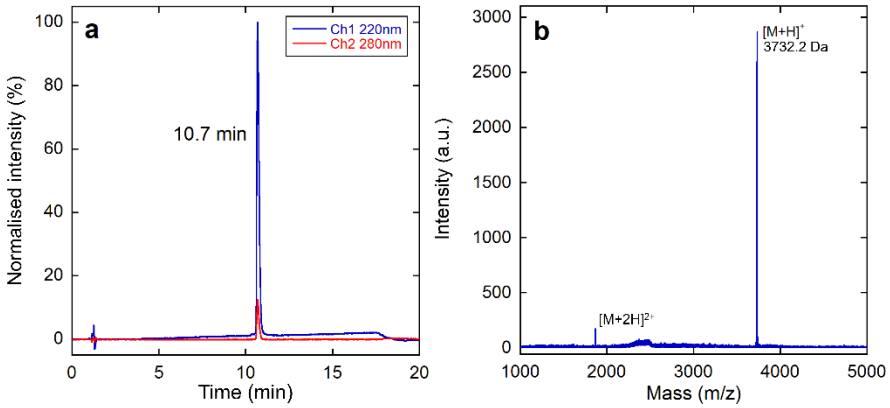
Supplementary Figure 17: (**a**) HPLC traces, gradient: 10-35%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 1958.1 Da)

$A_4(E_4K_4)_3$: Ac-GAAAAEEEEKKKKEEEEKKKKEEEEKKKKGW-NH_2



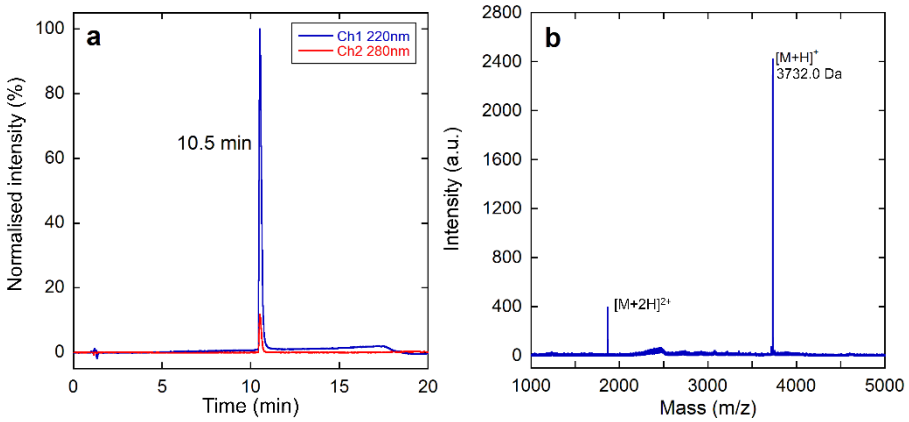
Supplementary Figure 18: (**a**) HPLC traces, gradient: 5-35%B, 16 min, 220 nm and 280 nm, C18 reverse-phase column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 3732.2 Da)

$(E_4K_4)_3A_4$: Ac–GEEEEKKKKEEEEKKKKEEEEKKKKAAAAGW–NH_2



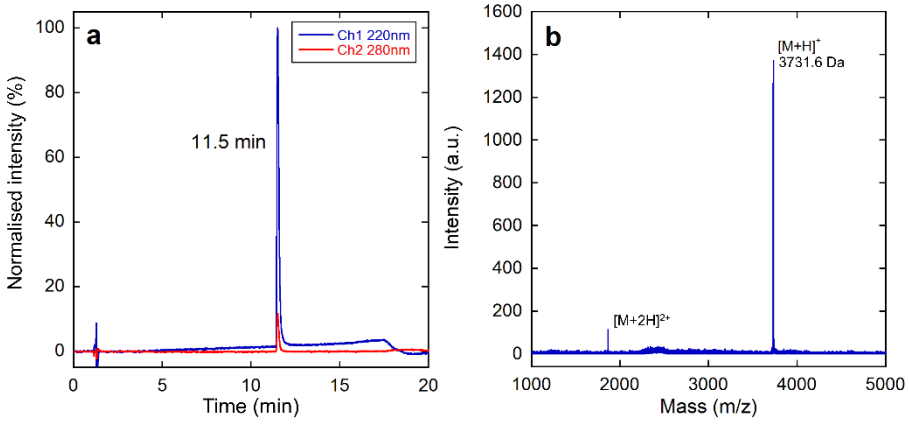
Supplementary Figure 19: (**a**) HPLC traces, gradient: 5-35%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 3732.2 Da)

$A_4(K_4E_4)_3$: Ac–GAAAAKKKKEEEEKKKKEEEEKKKKEEEEGW–NH_2



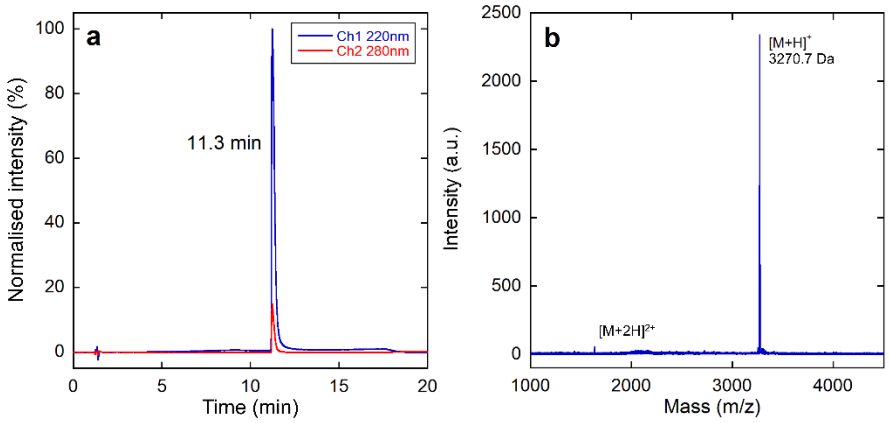
Supplementary Figure 20: (**a**) HPLC traces, gradient: 5-35%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 3732.2 Da)

$(K_4E_4)_3A_4$: Ac–GKKKKEEEEKKKKEEEEKKKKEEEEAAAAGW–NH_2



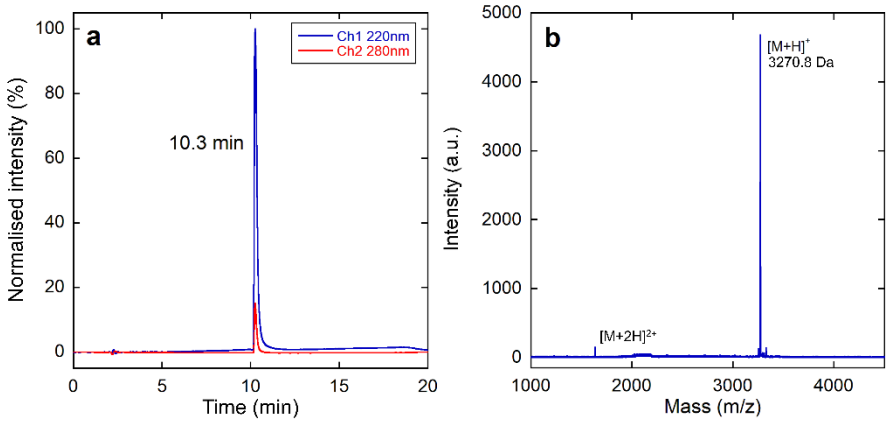
Supplementary Figure 21: (**a**) HPLC traces, gradient: 5-35%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 3732.2 Da)

$A_4(E_4K_4)A_4(E_4K_4)A_4$: Ac-GAAAAEEEEKKKKAAAAEEEEKKKKAAAAGW-NH_2



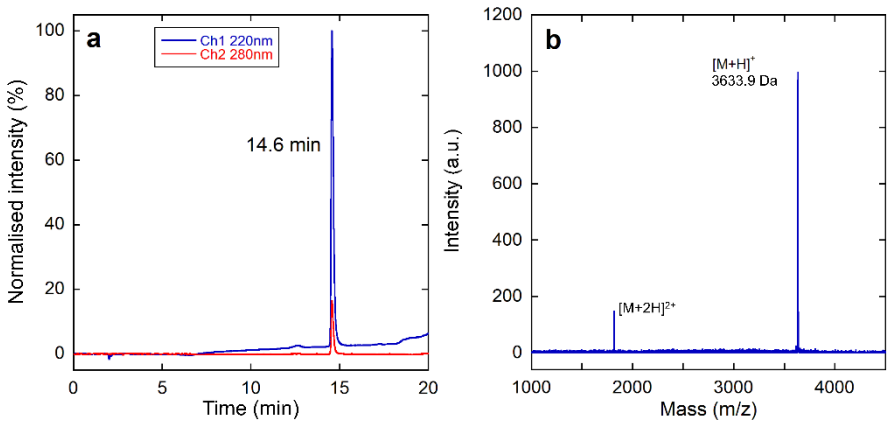
Supplementary Figure 22: (**a**) HPLC traces, gradient: 5-35%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 3271.6 Da)

$A_4(K_4E_4)A_4(K_4E_4)A_4$: Ac-GAAAAKKKKEEEEAAAAKKKKEEEEAAAAGW-NH_2



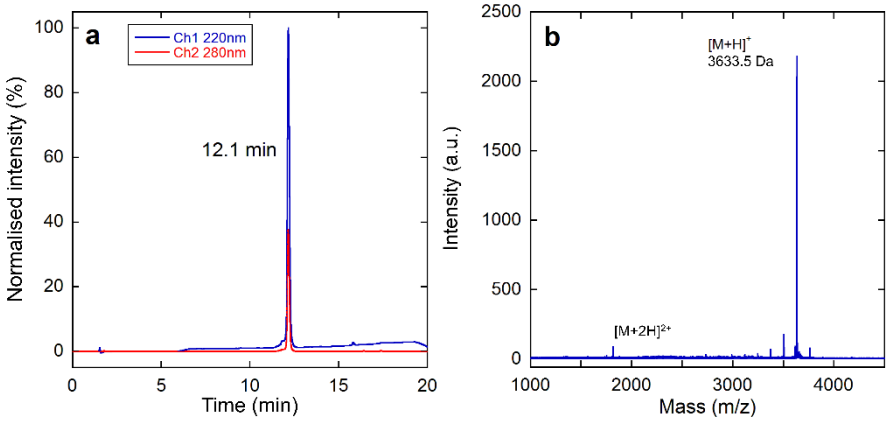
Supplementary Figure 23: (**a**) HPLC traces, gradient: 5-35%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 3271.6 Da)

$(E_3K_3)_4$: Ac-GEEEKKKEEEKKKEEEKKKEEEKKKGWW-NH_2



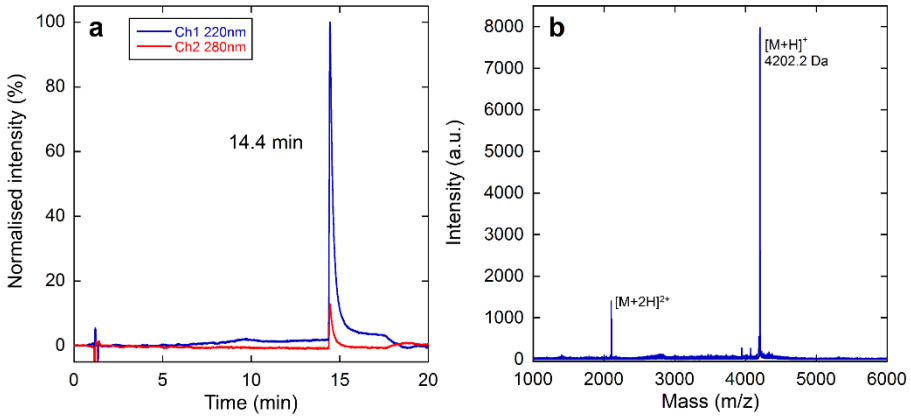
Supplementary Figure 24: (**a**) HPLC traces, gradient: 10-40%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 3634.1 Da)

$(K_3E_3)_4$: Ac-GKKKEEEKKKEEEKKKEEEKKKEEEGWW-NH_2



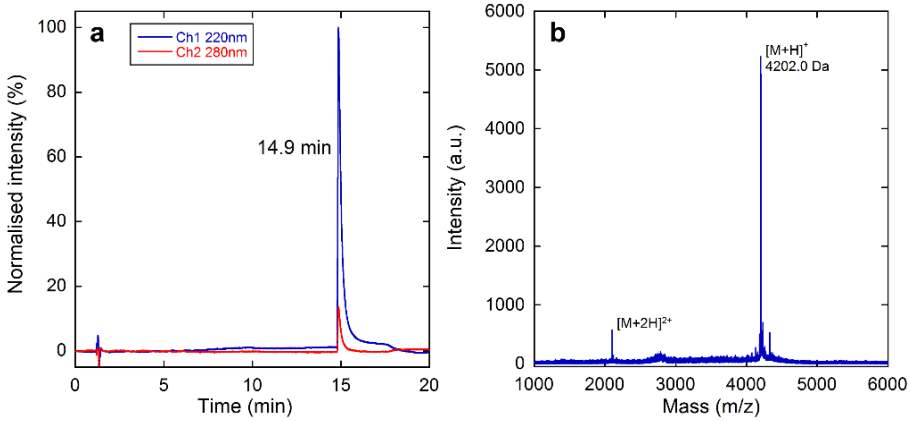
Supplementary Figure 25: (**a**) HPLC traces, gradient: 0-50%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 3634.1 Da)

$A_4(E_3K_3)_4A_4$: Ac-GAAAAEEEKKKEEEKKKEEEKKKEEEKKKAAAAGWW-NH_2

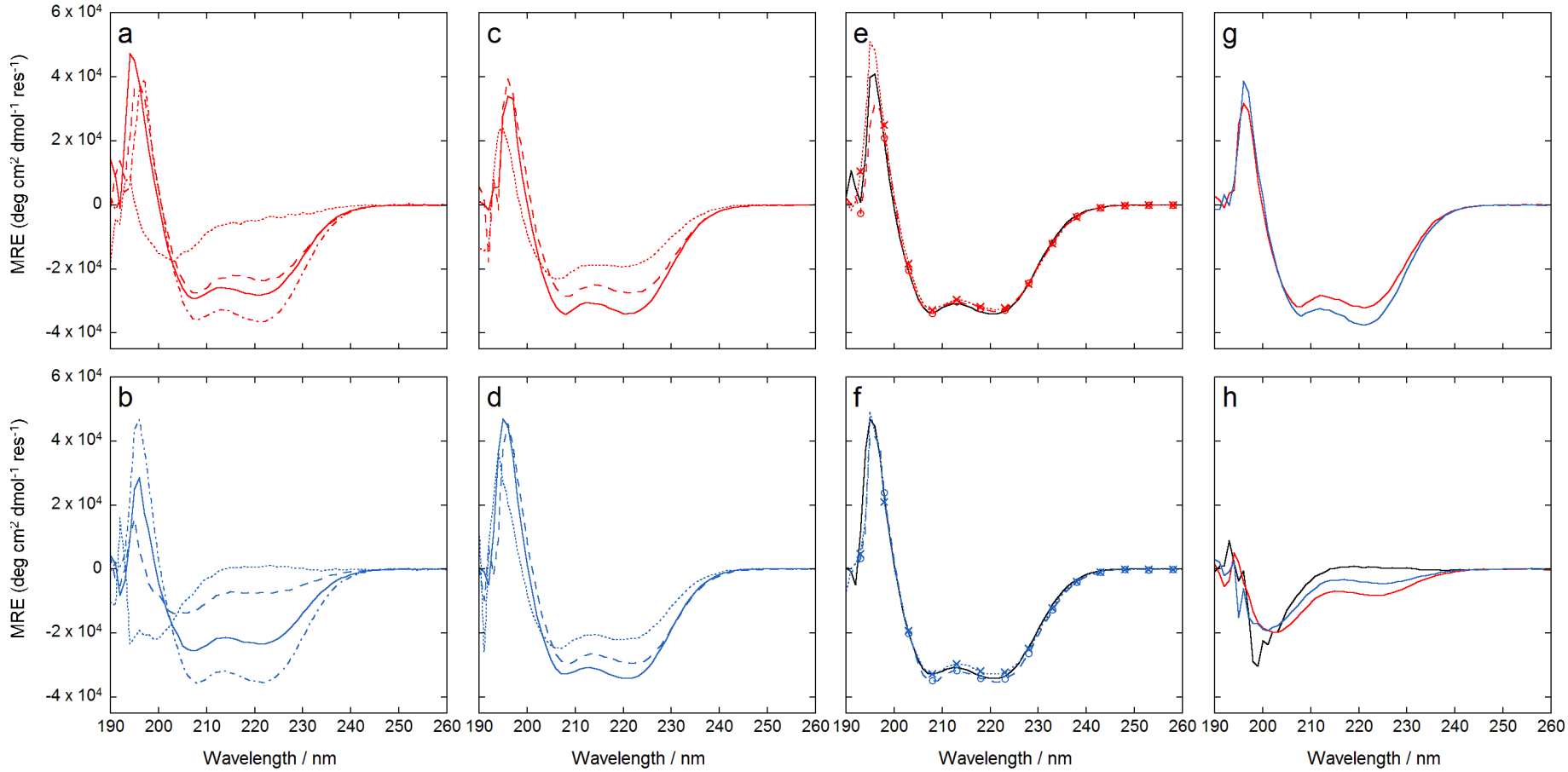


Supplementary Figure 26: (**a**) HPLC traces, gradient: 5-35%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS, reflector mode ($[M+H]^+_{calc}$: 4202.7 Da).

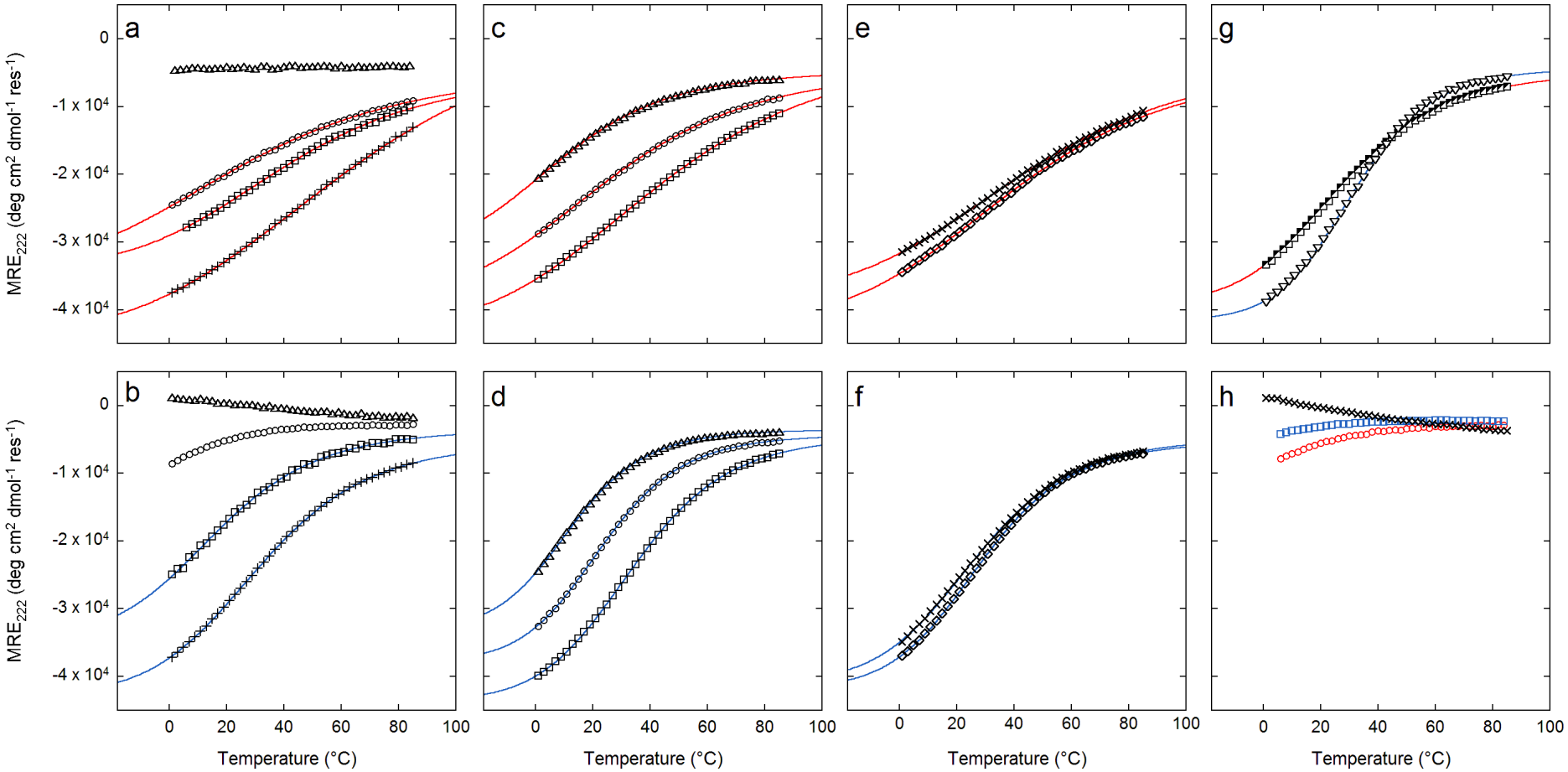
$A_4(K_3E_3)_4A_4$: Ac-GAAAAKKKEEEKKKEEEKKKEEEKKKEEEAAAAGWW-NH_2



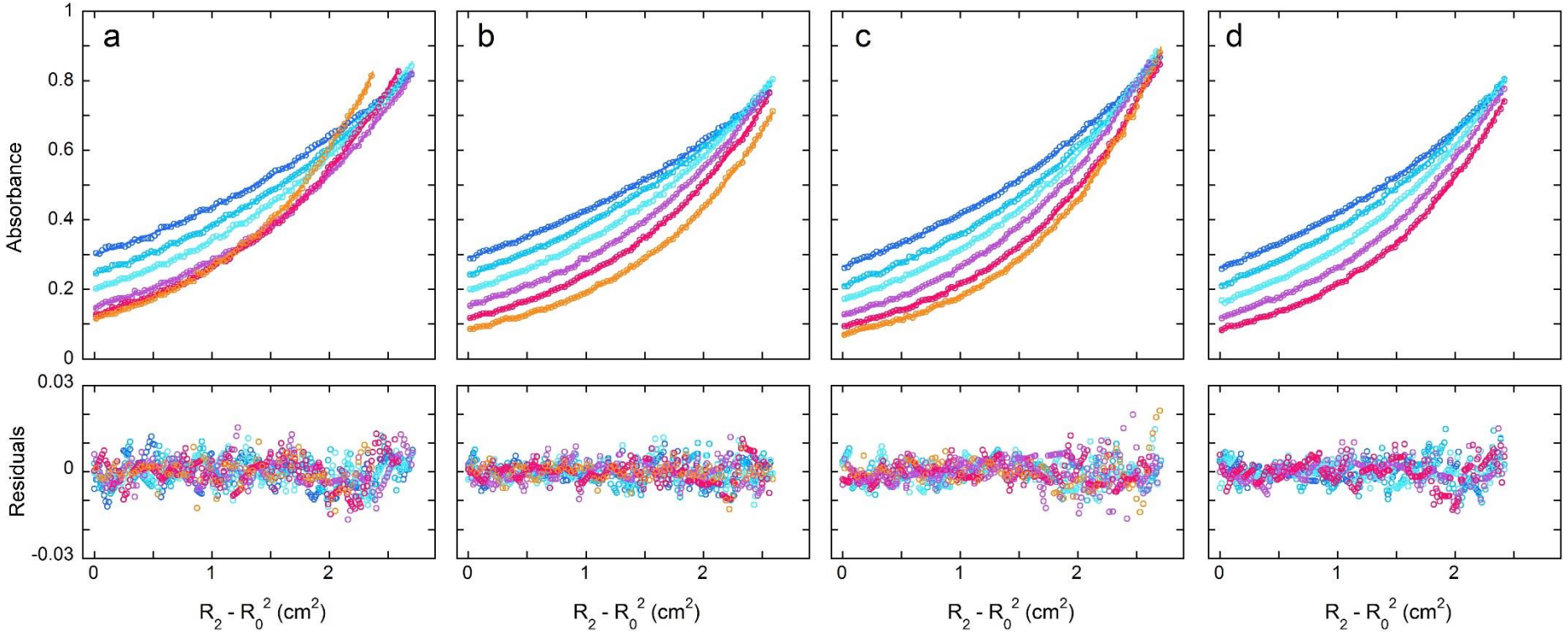
Supplementary Figure 27: (**a**) HPLC traces, gradient: 5-35%B, 16 min, 220 nm and 280 nm, reverse-phase C18 column (**b**) MALDI-TOF MS reflector mode ($[M+H]^+_{calc}$: 4202.7 Da).



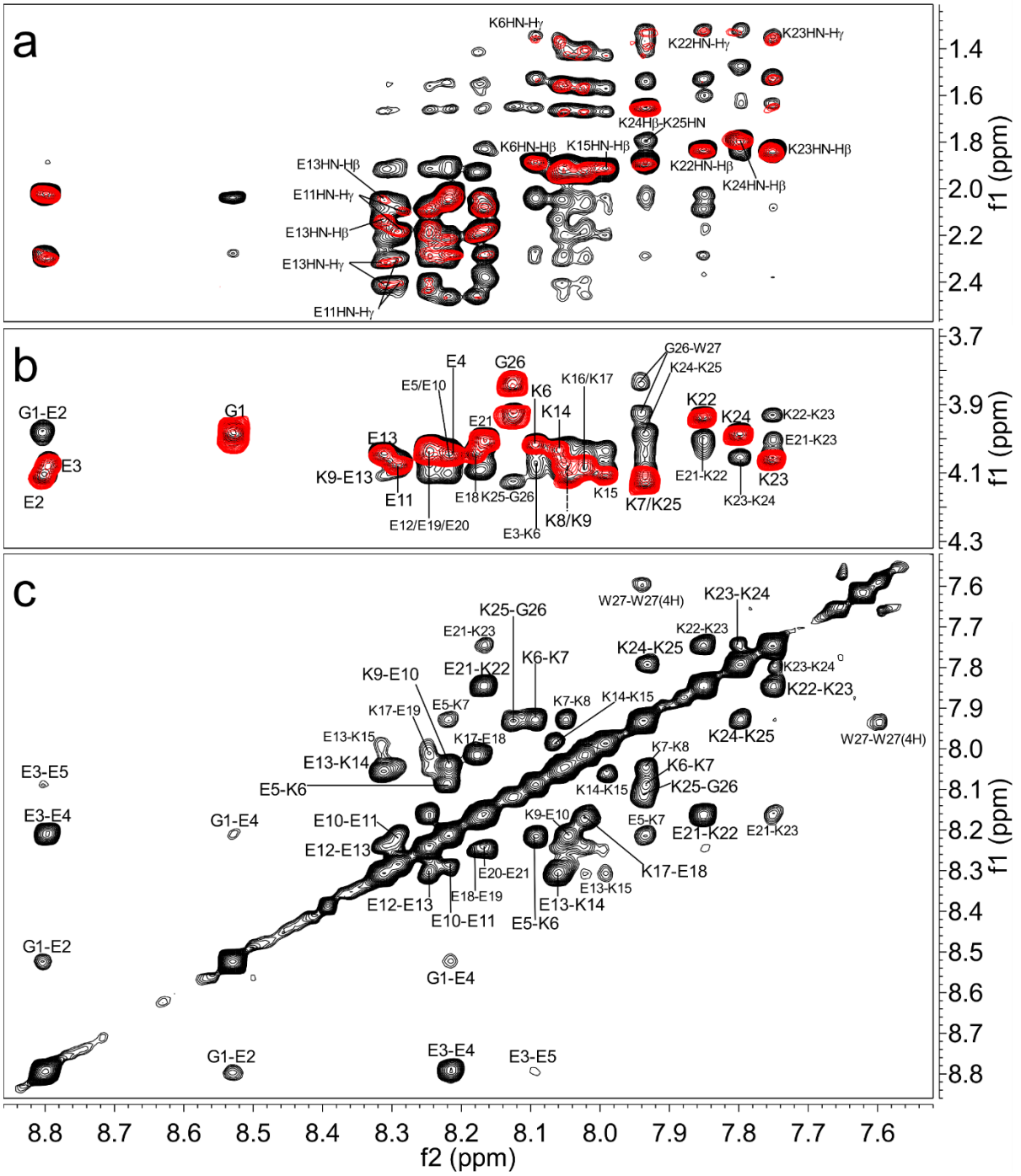
Supplementary Figure 28: CD spectra (5 °C) for peptides in this study. (**a**) **$(E_4K_4)_n$**, (**b**) **$(K_4E_4)_n$**, (**c**) **$A_4(E_4K_4)_nA_4$**, (**d**) **$A_4(K_4E_4)_nA_4$**. Key (**a**-**d**): n=4, alternating dash and dotted line; n=3, solid line; n=2, dashed line; n=1, dotted line. (**e**-**f**): A_4-flanked peptides: **$A_4(E_4K_4)_3A_4$** black solid line in (**e**); **$A_4(K_4E_4)_3A_4$** black solid line in (**f**); *N*-terminally flanked peptides **$A_4(E_4K_4)_3$**, red dashed line (circles, **e**), **$A_4(K_4E_4)_3$**, blue dashed line (circles, **f**); and *C*-terminally flanked peptides, **$(E_4K_4)_3A_4$** red dotted line (crosses, **e**), and **$(K_4E_4)_3A_4$**, blue dotted line (crosses, **f**). (**g**) **$A_4(E_4K_4)A_4(E_4K_4)A_4$**, red and **$A_4(K_4E_4)A_4(K_4E_4)A_4$**, blue. (**h**) **$(E_2K_2)_6$**, black; **$(EK)_{12}$**, red; **$(KE)_{12}$**, blue. In all parts (**a**-**h**) spectra for peptides with E→K directionality are coloured red, and those with K→E directionality, blue. Conditions: 5°C, 100 μM peptide concentration, at pH 7.4 in PBS.



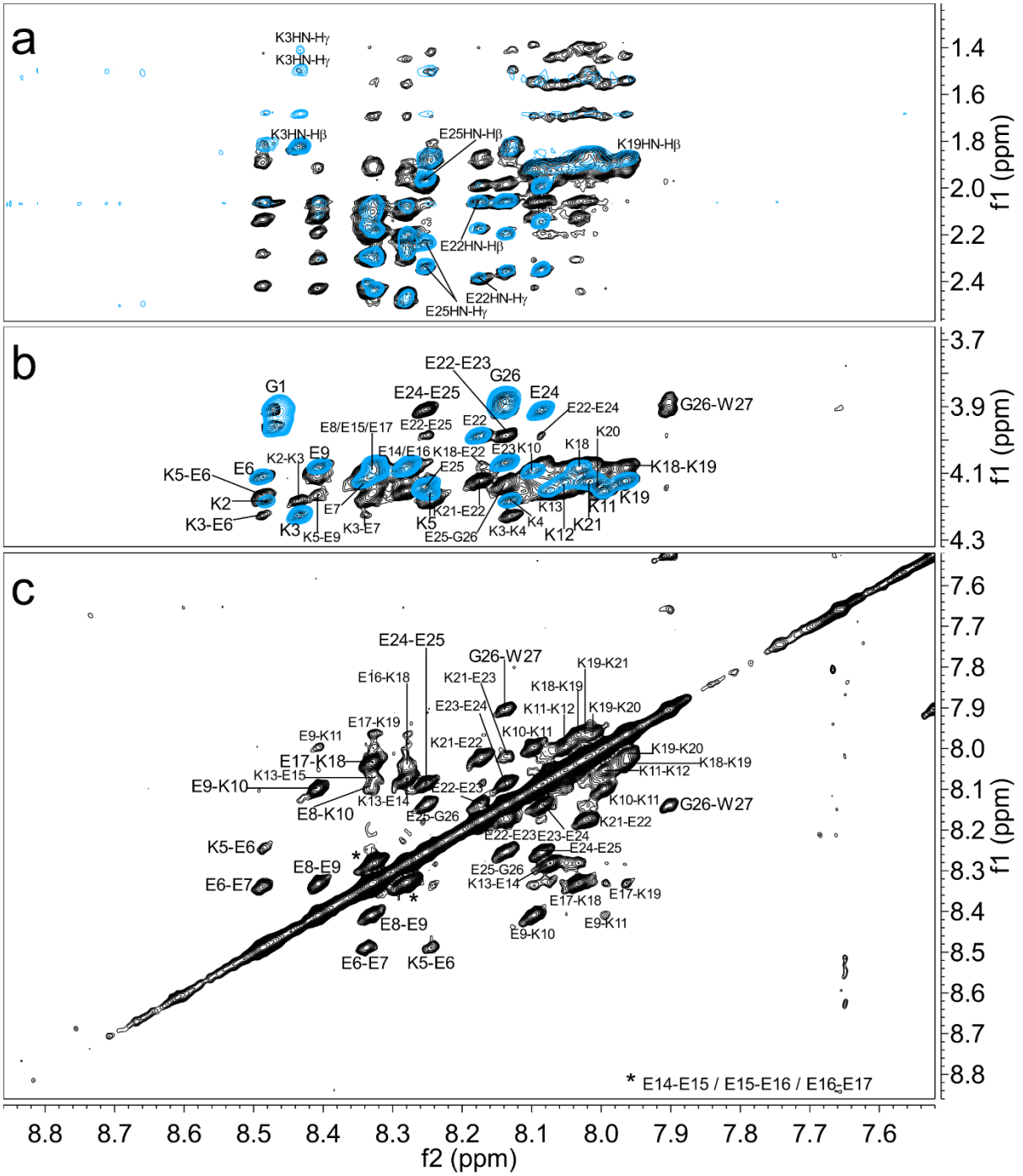
Supplementary Figure 29: Thermal denaturation curves for peptides in this study. (**a**) **$(E_4K_4)_n$**, (**b**) **$(K_4E_4)_n$**, (**c**) **$A_4(E_4K_4)_nA_4$**, (**d**) **$A_4(K_4E_4)_nA_4$**. Key (**a**-**d**): n=4, crosses; n=3 squares; n=2, circles; n=1, upward-pointing triangles. (**e**) **$A_4(E_4K_4)_3$**, diamonds and **$(E_4K_4)_3A_4$**, saltires. (**f**) **$A_4(K_4E_4)_3$**, diamonds and **$(K_4E_4)_3A_4$**, saltires. (**g**) **$A_4(E_4K_4)A_4(E_4K_4)A_4$**, half-filled squares and **$A_4(K_4E_4)A_4(K_4E_4)A_4$**, downward-pointing triangles. (**h**) **$(E_2K_2)_6$**, black crosses; **$(EK)_{12}$**, red circles; **$(KE)_{12}$**, blue squares. In all parts (**a**-**h**) every other data point has been plotted for clarity. The fits for the data are shown by solid lines; peptides with E→K directionality are coloured red, and those with K→E directionality, blue; note the sharper more sigmoidal transitions for the latter. Conditions: 100 μM peptide concentration, at pH 7.4 in PBS.



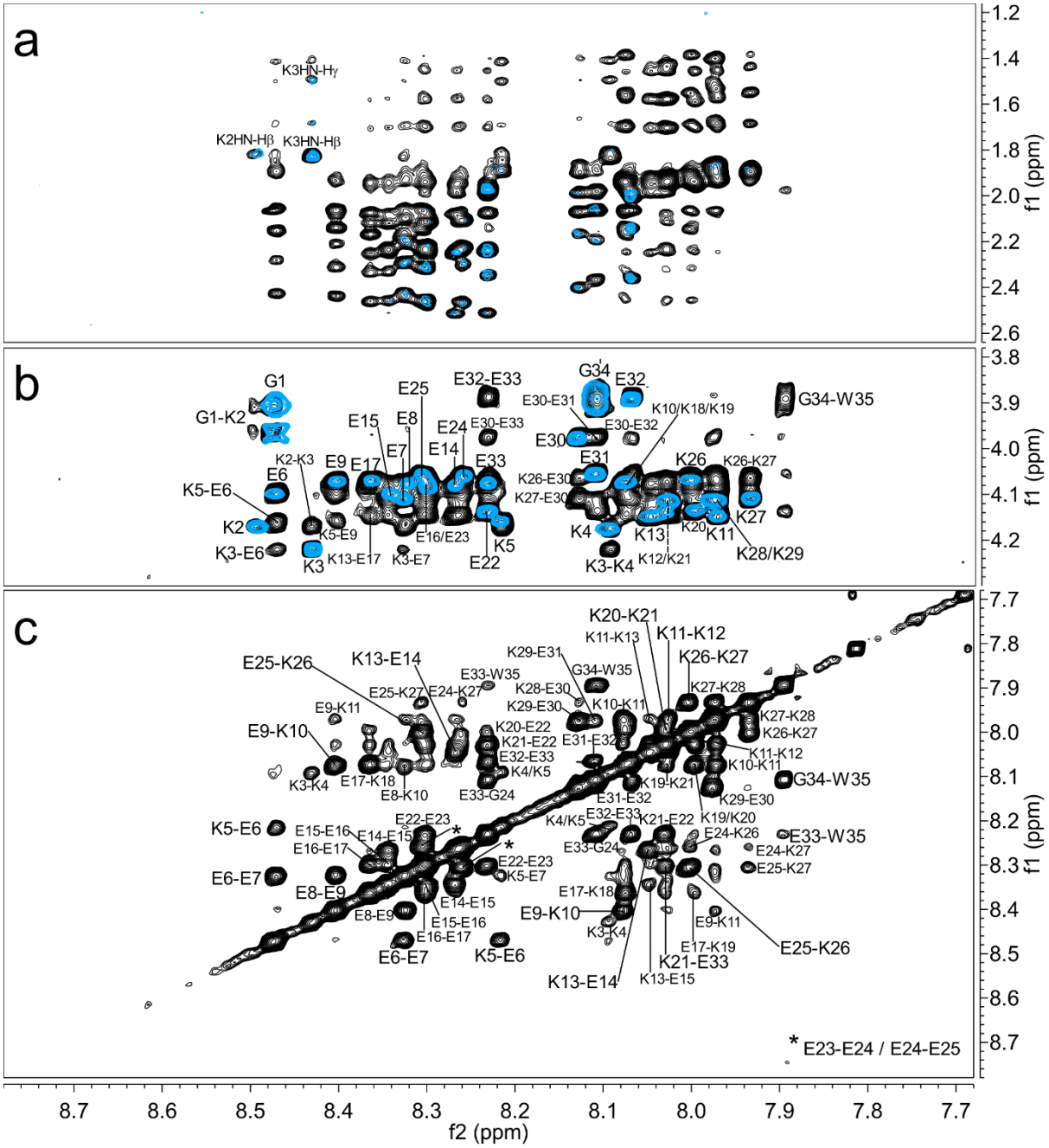
Supplementary Figure 30 AUC sedimentation equilibrium data for: (**a**) **$(E_4K_4)_3$**; (**b**) **$(K_4E_4)_3$**; (**c**) **$A_4(E_4K_4)_3A_4$**; and (**d**) **$A_4(K_4E_4)_3A_4$**. The mass by AUC of each peptide is 1.1x that of the molecular mass (see Supplementary Table 2). Top panels: data (circles) and single ideal species fits (solid line). Lower panels: residuals of the fits. Key: blue, 40,000 rpm; light blue, 44,000 rpm; cyan, 48,000 rpm; purple, 52,000 rpm; magenta, 56,000 rpm; and orange, 60,000 rpm. Conditions: 88 μM peptide concentration, 20 °C, pH 7.4 in PBS.



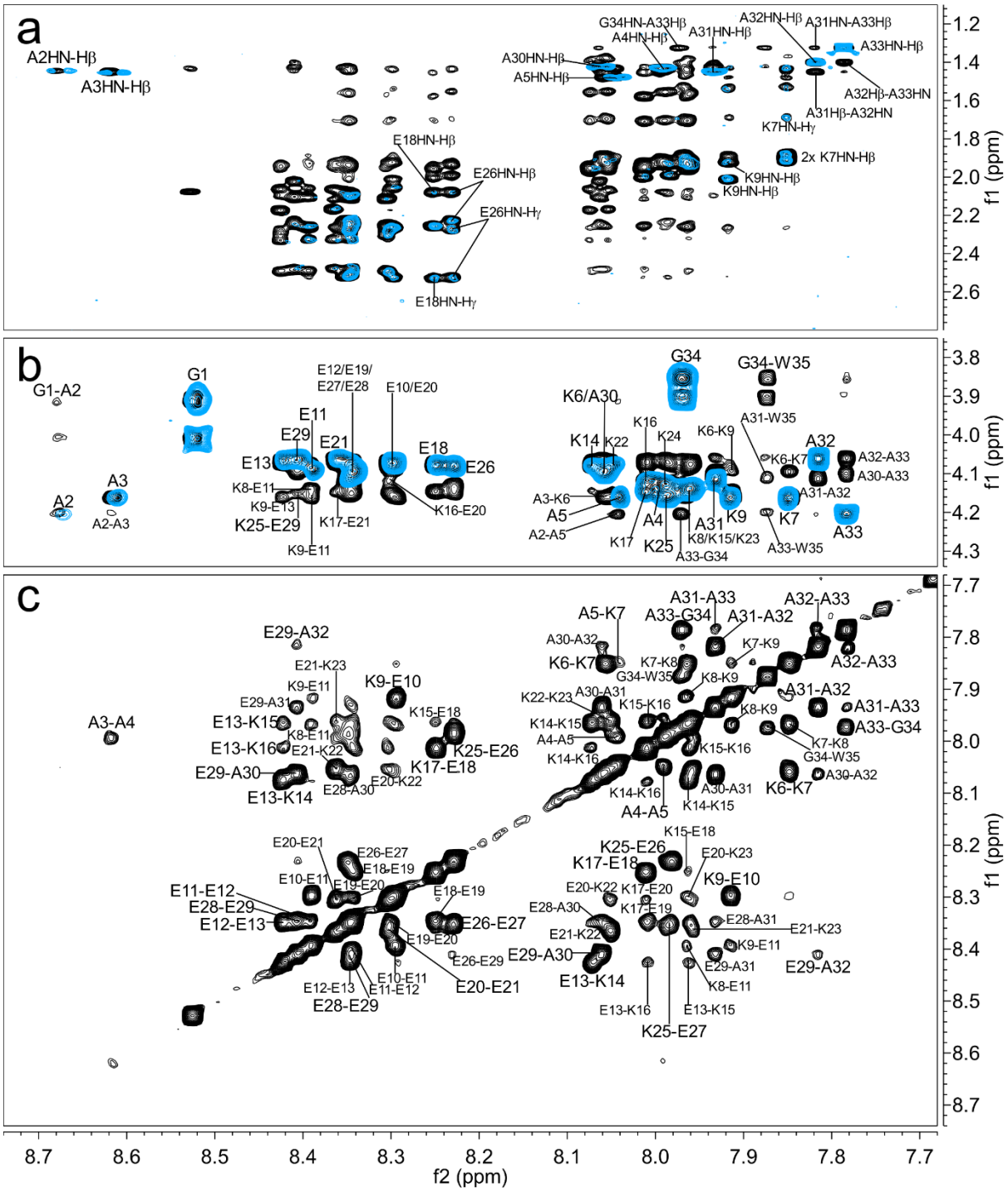
Supplementary Figure 31: ^{1}H-NMR spectra for **$(E_4K_4)_3$**. Overlaid (**a**) side chain (**b**) fingerprint and (**c**) amide regions of the TOCSY (red contours) and NOESY (black contours) spectra. Much of the side chain region could not be assigned due to signal overlap. Conditions: 5 °C, 600 MHz, 1 mM peptide concentration, pH 7.4, in PBS and 10% D_2O.



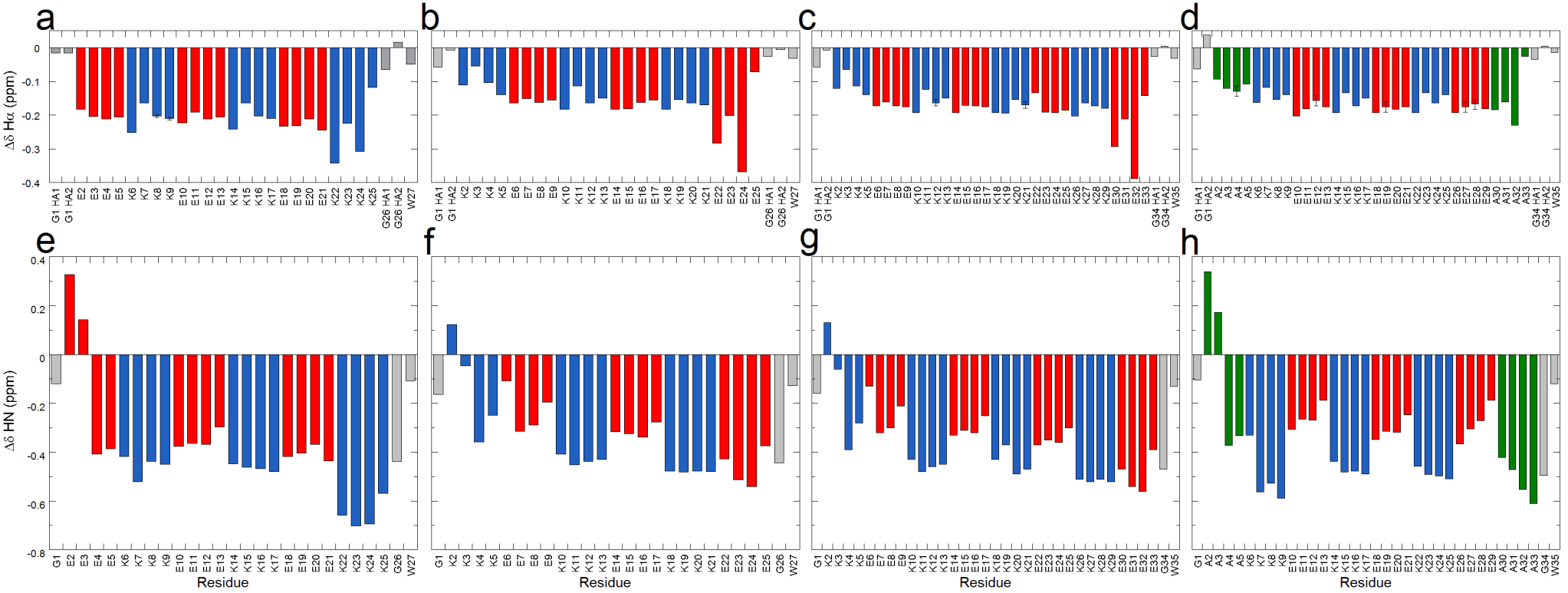
Supplementary Figure 32: ^{1}H-NMR spectra for **$(K_4E_4)_3$**. Overlaid (**a**) side chain (**b**) fingerprint and (**c**) amide regions of the TOCSY (blue contours) and NOESY (black contours) spectra. Much of the side chain region could not be assigned due to signal overlap. Conditions: 5 °C, 900 MHz, 1 mM peptide concentration, pH 7.4, in PBS and 10% D_2O.



Supplementary Figure 33: ^{1}H-NMR spectra for **$(K_4E_4)_4$**. Overlaid (**a**) side chain (**b**) fingerprint and (**c**) amide regions of the TOCSY (blue contours) and NOESY (black contours) spectra. Much of the side chain region could not be assigned due to signal overlap. Conditions: 5 °C, 900 MHz, 1 mM peptide concentration, pH 7.4, in PBS and 5% D_2O.

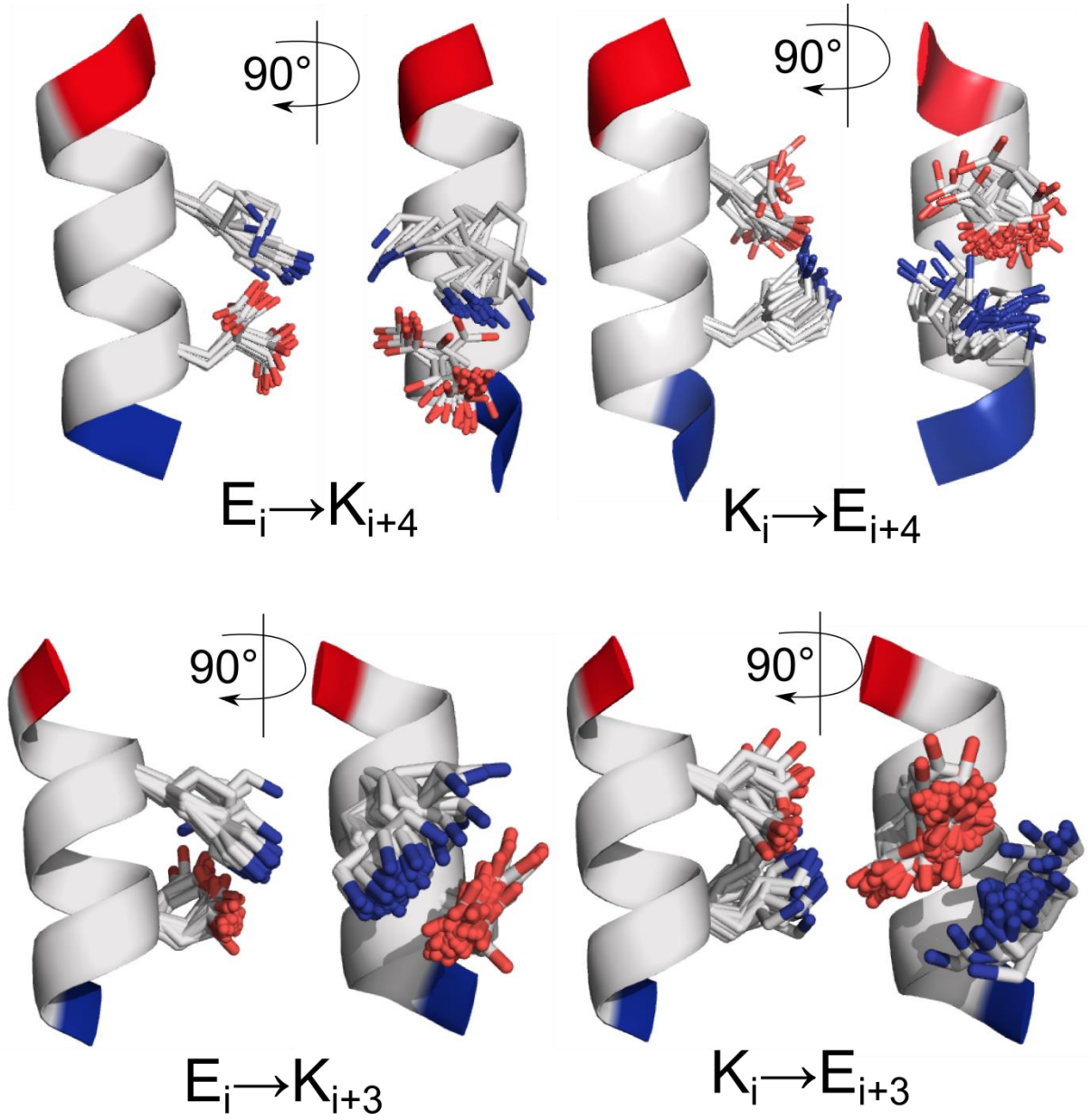


Supplementary Figure 34: ^{1}H-NMR spectra for **$A_4(K_4E_4)_3A_4$**. Overlaid (**a**) side chain (**b**) fingerprint and (**c**) amide regions of the TOCSY (blue contours) and NOESY (black contours) spectra. Much of the side chain region could not be assigned due to signal overlap. Conditions: 5 °C, 900 MHz, 1 mM peptide concentration, pH 7.4, in PBS and 10% D_2O.

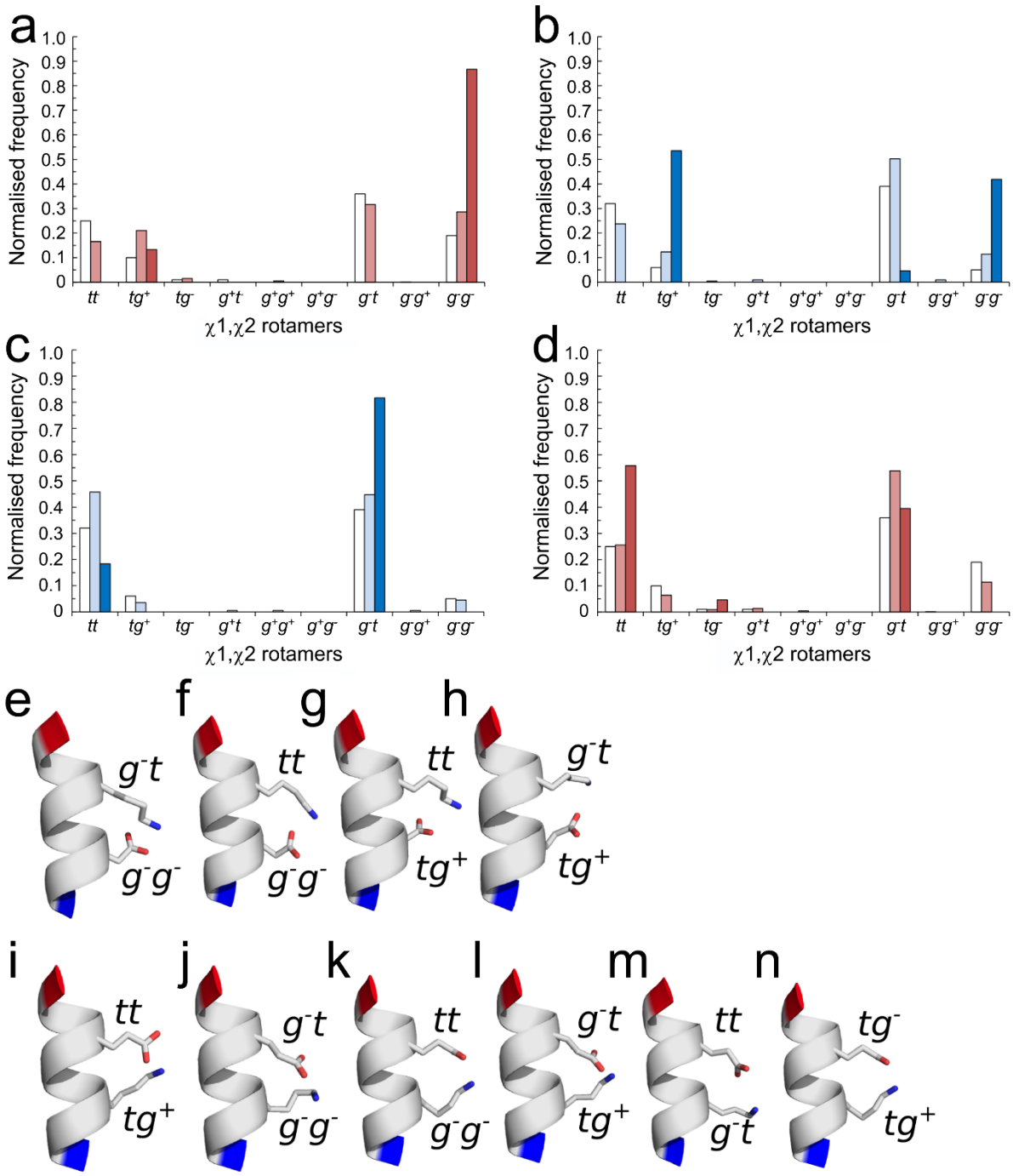


Supplementary Figure 35: ΔδHα (**a – d**) and ΔδHN (**e – h**) plots for (**a&e**) **$(E_4K_4)_3$**, (**b&f**) **$(K_4E_4)_3$**, (**c&g**) **$(K_4E_4)_4$** and (**d&h**) **$A_4(K_4E_4)_3A_4$**. ΔδHα values for residues at the *N*- and *C*-termini are consistent with these regions being more disordered than helix-central positions. ΔδHN values exhibit a square wave pattern which we do not understand but is linked to the change in residue type from Glu-to-Lys and vice versa. Key: Glu, red; Lys; blue; Ala, green; Gly and Trp, grey. Error bars show ambiguity in the assignment due to signal overlap.

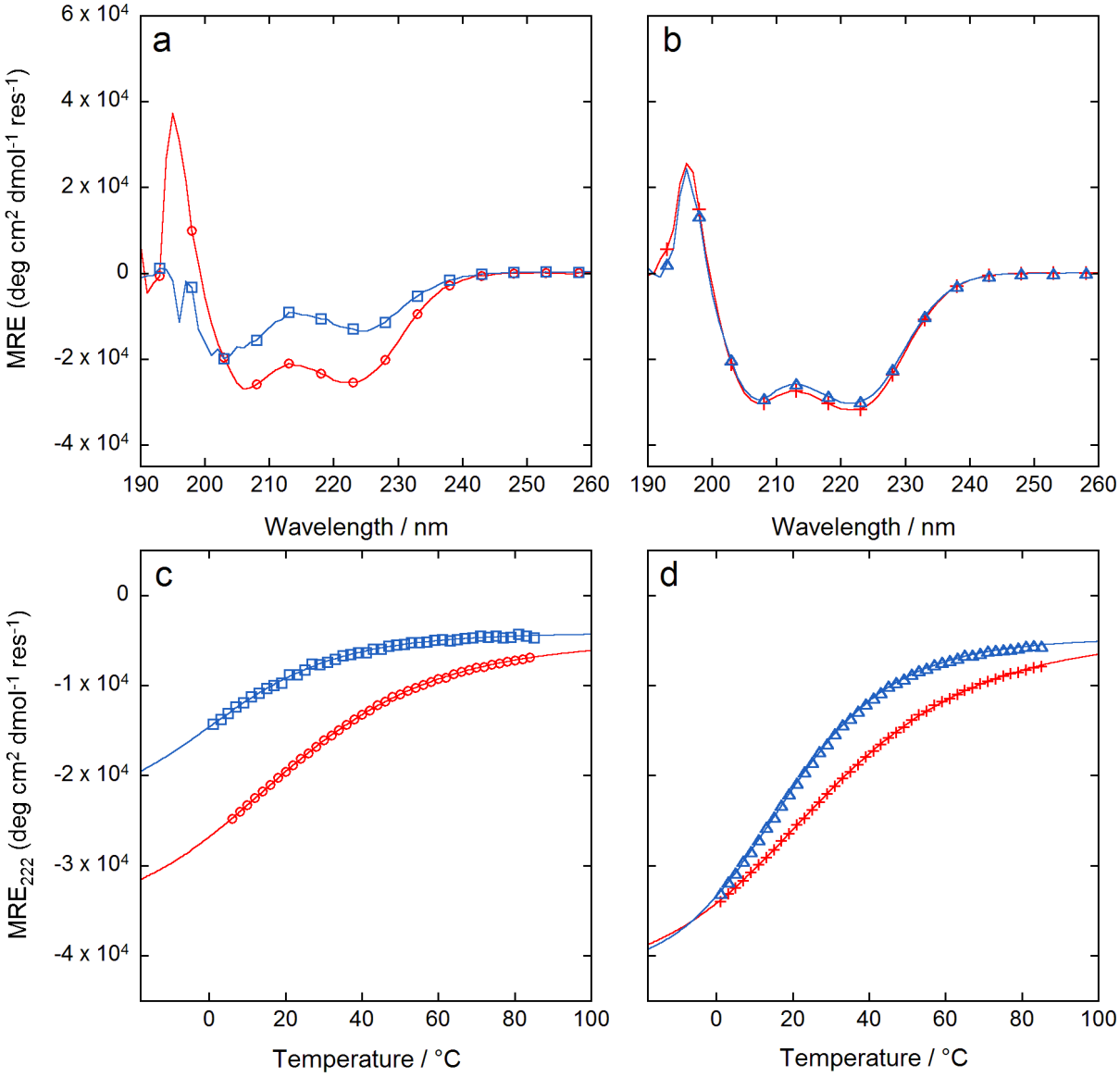
Supplementary Figure 36: Distributions of hydrogen bonds made by (**a**) Glu (**b**) Lys and (**c**) Asn residues at specific positions in α-helices in the PDB.[1] All residue positions at least 4 positions in sequence away from the termini were categorized as 'central'. Hydrogen bonds made by the side chain of each residue were categorized according to the hydrogen bond donor type (blue, no hydrogen bonds; red, ≥ 1 hydrogen bond to water; green, ≥ 1 hydrogen bond to a side-chain atom; purple, ≥ 1 hydrogen bond to a main-chain atom). The total of these were expressed as a percentage for each category at each position.



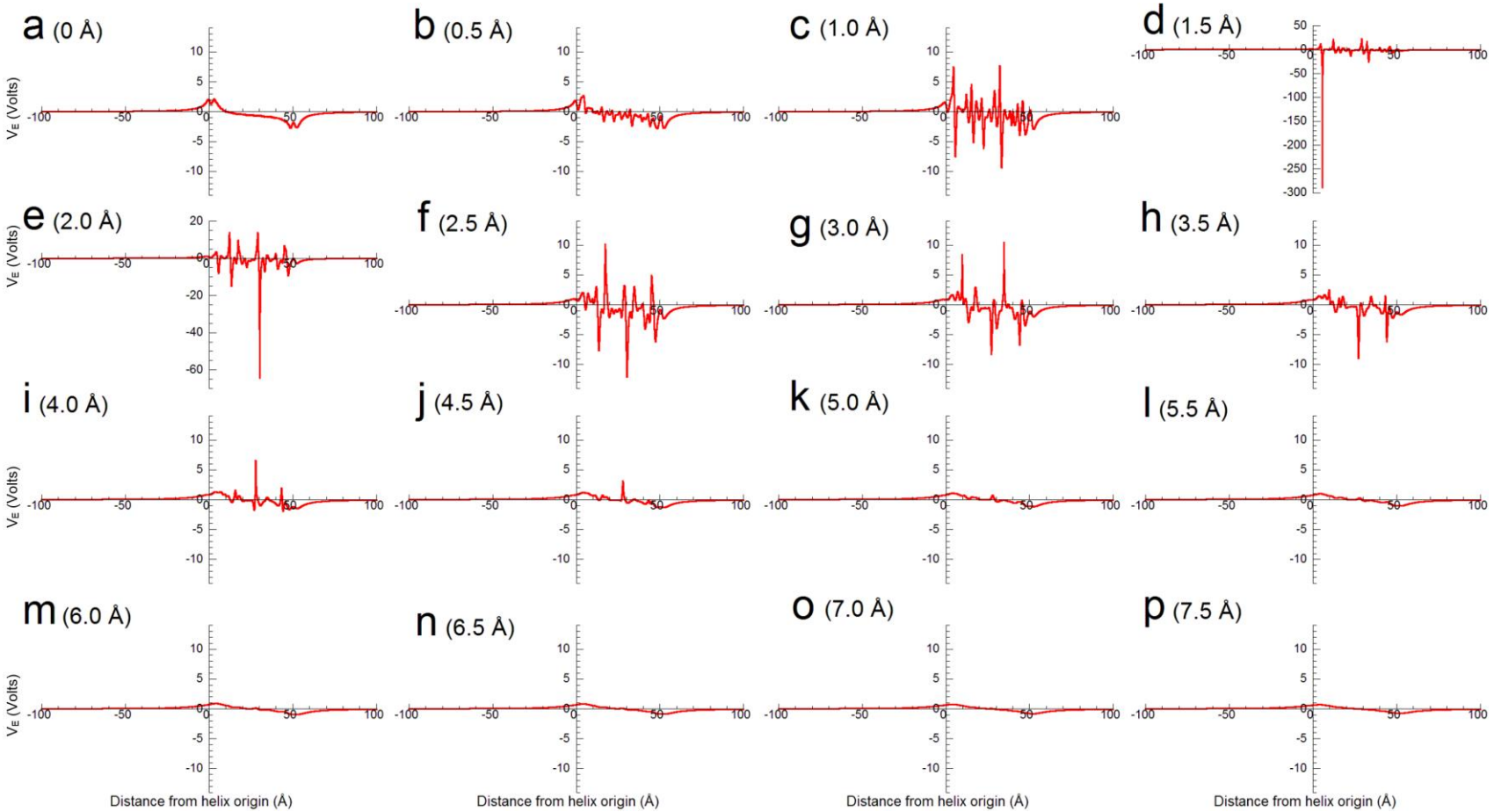
Supplementary Figure 37: Overlays of different types of salt bridge examined in this analysis, taken from central helical pairs only. $E_i \rightarrow K_{i+4}$ average RMSD, 1.79 Å ± 1.03, n=25; $K_i \rightarrow E_{i+4}$ average RMSD, 1.08 Å ± 0.72, n=52; $E_i \rightarrow K_{i+3}$ average RMSD 1.09 Å ± 0.66, n=60; $K_i \rightarrow E_{i+3}$ average RMSD 2.27Å ± 0.84, n=43.



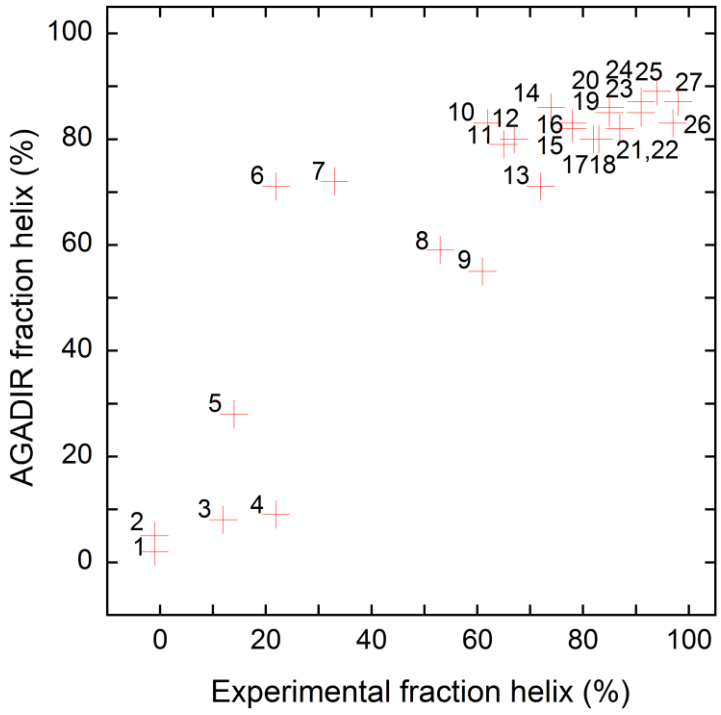
Supplementary Figure 38: $\chi1,\chi2$ distributions and conformers for $E_i \rightarrow K_{i+3}$ and $K_i \rightarrow E_{i+3}$ pairs in high-resolution X-ray crystal structures. (**a-d**) Normalized frequency plots of preferred χ_{1},χ_{2} angles for glutamate (**a&d**) and lysine (**b&c**) residues in $E_i \rightarrow K_{i+3}$ (**a&c**) and $K_i \rightarrow E_{i+3}$ (**b&d**) pairs. Key: white bars indicate the frequency of each rotamer found in all α-helices; pale bars indicate pairs where no salt bridge is made; dark bars indicate pairs where a salt bridge is formed; and red colouring is for Glu and blue for Lys. For the $E_i \rightarrow K_{i+3}$ pairs, there is one dominant rotamer combination, with Glu (*g*$^-$*g*$^-$) plus Lys (*g*$^-$*t*) (**e**, 46 examples (77%)), and three minor combinations (**f,g,h**, 6, 5 and 3 examples respectively) with Glu;Lys (*g*$^-$*g*$^-$;*tt*), Glu;Lys (*tg*$^+$,*tt*) and Gly;Lys (*tg*$^+$;*g*$^-$*t*) respectively; whereas for the $K_i \rightarrow E_{i+3}$ pairs, there are two preferred combinations (**i&j**, Lys;Glu (*tg*$^+$;*tt*), 15 examples and Lys;Glu (*g*$^-$*g*$^-$;*g*$^-$*t*), 11 examples; 60% of all pairs), two less preferred combinations (**k&l**, Lys;Glu (*g*$^-$*g*$^-$;*tt*), 7 examples and Lys;Glu (*tg*$^+$;*g*$^-$*t*), 6 examples) and two minor combinations (**m&n**, Lys;Glu (*g*$^-$*t*;*tt*) and Lys;Glu (*tg*$^+$;*tg*$^-$), 2 examples each. Examples of these are shown in panels **e** – **n**, which are taken from (**e**), PDB 1c1d, A159-A162; (**f**), PDB 3vmk, A20-A23; (**g**), PDB 3eki, A74-A77; (**h**), PDB 2w6a, A462-A465; (**i**), PDB 1f1e, A122-A125; (**j**) PDB 1fcy, A352-A355; (**k**) PDB 1iom, A292-A295; (**l**) PDB 1xoc, A468-A471; (**m**) PDB 1t1u, A74-A77; (**n**) PDB 3mxz, A26-A29. Images generated with PyMol (www.pymol.org).



Supplementary Figure 39: Secondary structure of **-(E$_3$K$_3$)$_4$-** and **-(K$_3$E$_3$)$_4$-** based peptides in solution. (**a&b**) CD spectra at 5°C: (**a**) **(E$_3$K$_3$)$_4$** (red circles) and **(K$_3$E$_3$)$_4$** (blue squares); (**b**) **A$_4$(E$_3$K$_3$)$_4$A$_4$** (red crosses) and **A$_4$(K$_3$E$_3$)$_4$A$_4$** (blue triangles). (**c&d**) Thermal denaturation curves of the peptides followed at 222 nm. Fits to the data using the Gibbs-Helmholtz equation are shown by solid lines: (**c**) **(E$_3$K$_3$)$_4$** (red circles) and **(K$_3$E$_3$)$_4$** (blue squares); (**d**) **A$_4$(E$_3$K$_3$)$_4$A$_4$** (red crosses) and **A$_4$(K$_3$E$_3$)$_4$A$_4$** (blue triangles). Note the sharper transition for **A$_4$(K$_3$E$_3$)$_4$A$_4$** in (**d**). Each measurement in **b&d** was repeated three times using freshly prepared samples each time, and the average plotted.

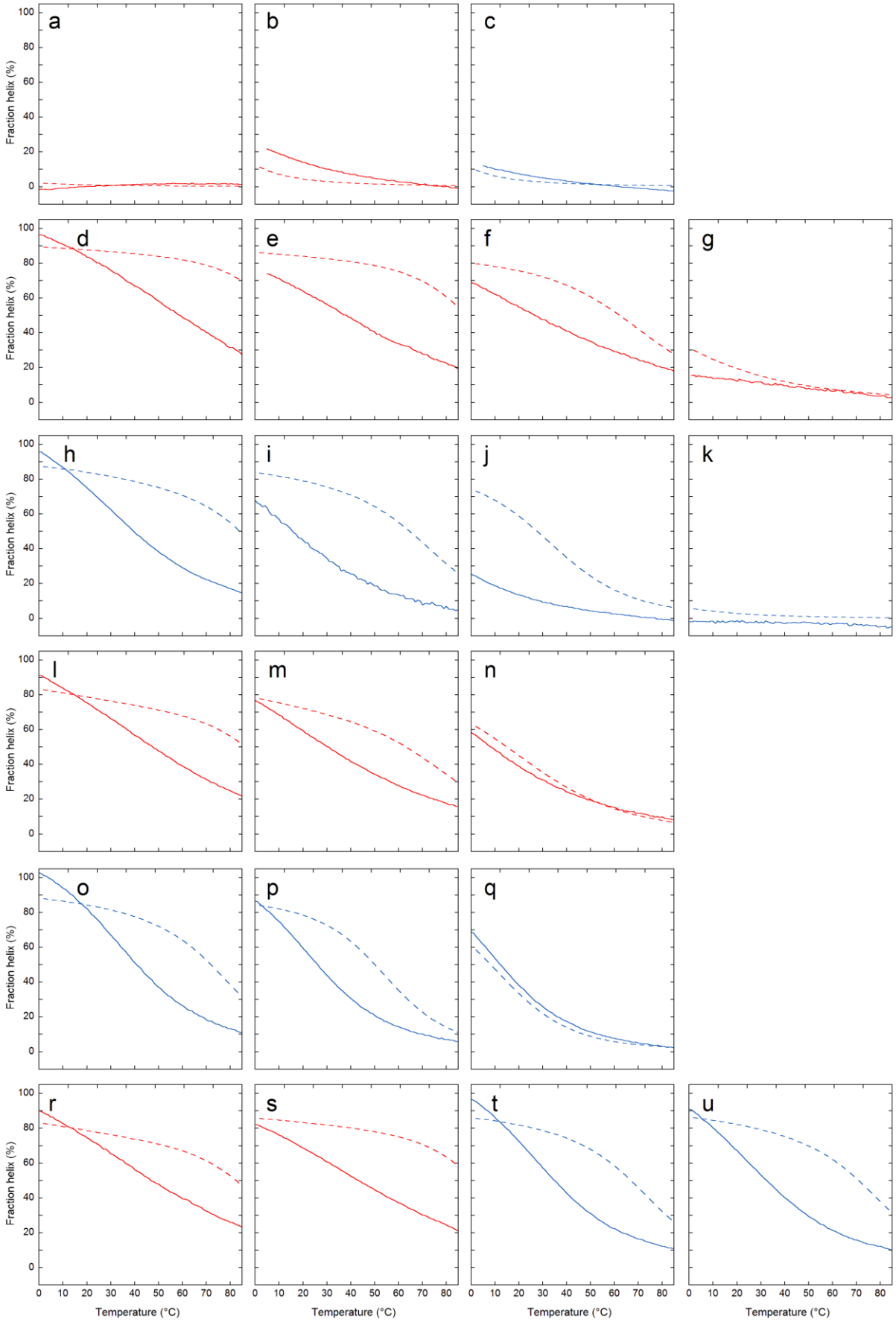


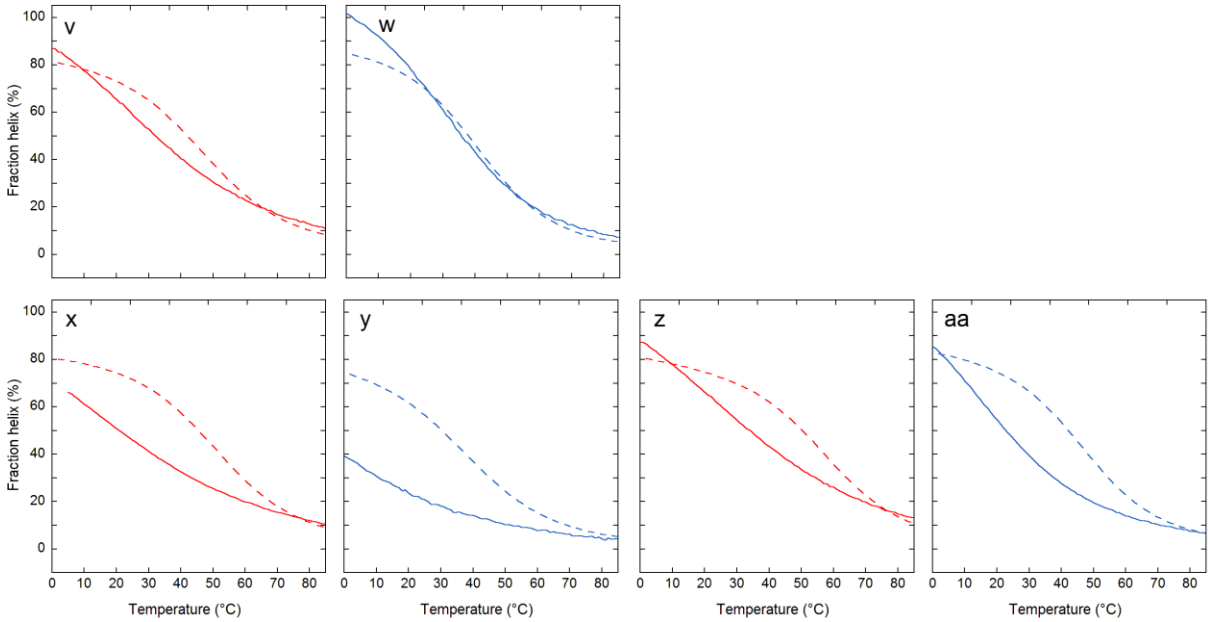
Supplementary Figure 40: The electrostatic potential, $V_E(x,0,z)$, of a 32 residue polyalanine α-helix. V_E was calculated at various distances, x (shown in brackets), parallel to the long helix axis (z), y = 0. The N-terminus of the helix is at the origin. *N.B.* the increased scale for V_E in parts (**d**) and (**f**) is the result of x being very close to a single point charge.



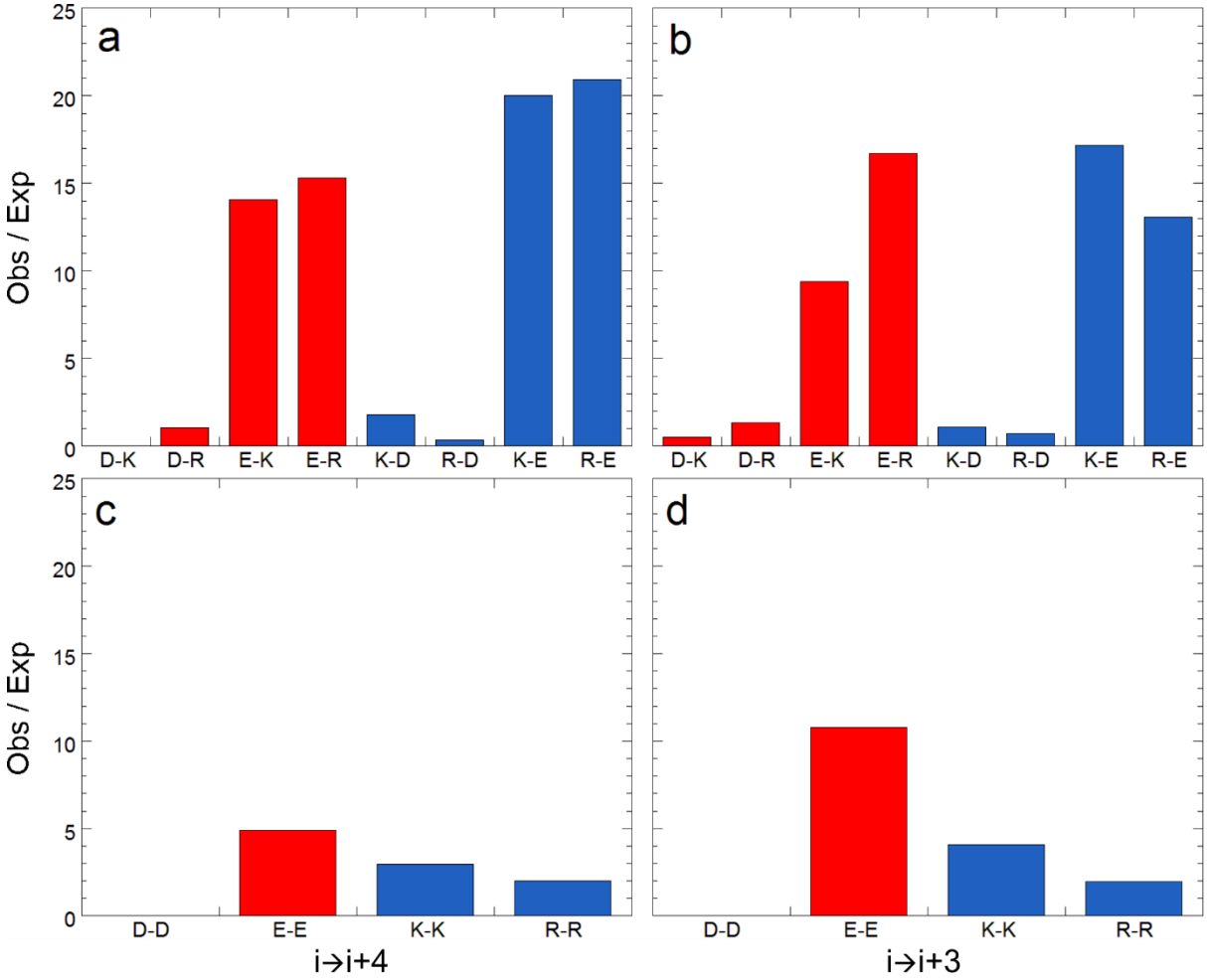
Supplementary Figure 41: Percent helicities of the following peptides predicted by AGADIR[2] (at 5 °C and ionic strength 0.1661 M) vs. those observed experimentally by CD spectroscopy:

1	$(E_2K_2)_6$	10	$(K_4E_4)_3$	19	$(E_4K_4)_3A_4$
2	(K_4E_4)	11	$(E_4K_4)_2$	20	$(K_4E_4)_3A_4$
3	$(KE)_{12}$	12	$(E_3K_3)_4$	21	$A_4(E_4K_4)_3$
4	$(EK)_{12}$	13	$A_4(E_4K_4)_2A_4$	22	$A_4(E_4K_4)_3A_4$
5	(E_4K_4)	14	$(E_4K_4)_3$	23	$A_4(K_4E_4)_3$
6	$(K_4E_4)_2$	15	$A_4(K_3E_3)_4A_4$	24	$(K_4E_4)_4$
7	$(K_3E_3)_4$	16	$A_4(K_4E_4)_2A_4$	25	$(E_4K_4)_4$
8	$A_4(E_4K_4)A_4$	17	$A_4(E_3K_3)_4A_4$	26	$A_4(K_4E_4)A_4(K_4E_4)A_4$
9	$A_4(K_4E_4)A_4$	18	$A_4(E_4K_4)A_4(E_4K_4)A_4$	27	$A_4(K_4E_4)_3A_4$


a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
Fraction helix (%)
0
20
40
60
80
100
0 10 20 30 40 50 60 70 80
Temperature (°C)



Supplementary Figure 42: AGADIR-predicted[2] thermal denaturation curves (dashed lines) vs. experiment (solid lines). (**a**) **$(E_2K_2)_6$**, (**b**) **$(EK)_{12}$**, (**c**) **$(KE)_{12}$**, (**d**) **$(E_4K_4)_4$**, (**e**) **$(E_4K_4)_3$**, (**f**) **$(E_4K_4)_2$**, (**g**) **(E_4K_4)**, (**h**) **$(K_4E_4)_4$**, (**i**) **$(K_4E_4)_3$**, (**j**) **$(K_4E_4)_2$**, (**k**) **(K_4E_4)**, (**l**) **$A_4(E_4K_4)_3A_4$**, (**m**) **$A_4(E_4K_4)_2A_4$**, (**n**) **$A_4(E_4K_4)A_4$**, (**o**) **$A_4(K_4E_4)_3A_4$**, (**p**) **$A_4(K_4E_4)_2A_4$**, (**q**) **$A_4(K_4E_4)A_4$**, (**r**) **$A_4(E_4K_4)_3$**, (**s**) **$(E_4K_4)_3A_4$**, (**t**) **$A_4(K_4E_4)_3$**, (**u**) **$(K_4E_4)_3A_4$**, (**v**) **$A_4(E_4K_4)A_4(E_4K_4)A_4$**, (**w**) **$A_4(K_4E_4)A_4(K_4E_4)A_4$**, (**x**) **$(E_3K_3)_4$**, (**y**) **$(K_3E_3)_4$**, (**z**) **$A_4(E_3K_3)_4A_4$**, (**aa**) **$A_4(K_3E_3)_4A_4$**. Key: E→K directional peptides, red; K→E directional peptides, blue. The predictions made by AGADIR are poor, apart for peptides with low α-helicity. The authors of AGADIR do acknowledge that in cases where there are multiple possible salt bridges the algorithm can over predict α-helical content. We suggest this should be addressed.



Supplementary Figure 43: Observed/expected numbers of D/E/K/R→D/E/K/R pairs in a set of nine SAHs* (determined as such by virtue of CD spectroscopy and/or electron microscopy techniques by others) from non-homologous proteins (<30% overall pairwise sequence identity by CDHit).[3] Expected numbers are calculated from our PDB analysis of α-helices. (**a&b**) Favourably matched i→i+4 and i→i+3 pairs respectively and (**c&d**) unfavourably matched i→i+4 and i→i+3 pairs respectively. Key: D, aspartic acid; E, glutamic acid; K, lysine; R, arginine. Where the side chain of residue 'i' bears a negative charge (D,E) the data is coloured red; blue is used for K or R (positively charged side chain) when at this position. Where possible the sequence from the human protein was used. If there was more than one assignment in the literature for which region of the protein constitutes the SAH, we took the longest.

***Uniprot codes and SAH sequences used (D and E are coloured red, K and R blue, to highlight the repeating blocks of four E followed by four of either R or K typical to most SAHs):

>Q9HD67|MYO10_HUMAN Unconventional myosin-X SAH domain: 808-931[4,5]
YRQLLAEKREQEEKKKQEEEEKKKREEEERERERERRRAELRAQQEEETRKQQELEALQKSQKEAELT
RELEKQKENKQVEEILRLEKEIEDLQRMKEQQELSLTEASLQKLQERRDQELRRLE

>Q9H3P7|GCP60_HUMAN Golgi resident protein SAH domain: 174-247[4,6]
TYVASHKIEKEEQEKKRKEEEERRRREEEERERLQKEEEKRRREEEERLRREEEERRRIEEERLRLEQ
QKQQIMAALNSQTAVQ

>Q9UM54|MYO6_HUMAN Unconventional myosin-VI SAH domain: 808-931[4,7]
KSSEELLSALQKKKQQEEEAERLRRIQEEMEKERKRREEDEKRRRKEEEERRMKLEMEAKRKQEEEER
KKREDDEKRIQAEVEAQLARQKEEESQQQAVLEQERRDRELALRIAQSEAEL

>O95819|M4K4_HUMAN Mitogen-activated protein kinase kinase kinase kinase 4 SAH domain: 361-480[4,6]
QENKERSEALRRQQLLQEQQLREQEEYKRQLLAERQKRIEQQKEQRRRLEEQQRREREARRQQEREQR
RREQEEKRRLEELERRRKEEEERRRAEEEKRRVEREQEYIRRQLEEEQRHLE

>Q05682|CALD1_HUMAN Caldesmon SAH domain: 181-420[4,8]
EENKKEDKEKEEEEEEKPKRGSIGENQVEVMVEEKTTESQEETVVMSLKNGQISSEEPKQEEEREQGS
DEISHHEKMEEEDKERAEAERARLEAEERERIKAEQDKKIADERARIEAEEKAAAQERERREAEERER
MREEEKRAAEERQRIKEEEKRAAEERQRIKEEEKRAAEERQRIKEEEKRAAEERQRARAEEEEKAKVE
EQKRNKQLEEKKHAMQETKIKGEKVEQKIEGKWVNE

>**Q13402|MYO7A_HUMAN** Unconventional myosin-VIIa SAH domain: 870-935[4,9]
EKMRLAEEEKLRKEMSAKKAKEEAERKHQERLAQLAREDAERELKEKEAARRKKELLEQMERARHE

>**A2EUZ9|A2EUZ9_TRIVA** Kelch motif family protein SAH domain: 869-1068[10]
KKKEEEEKKQKEEQERLAKEEAERKQKEEQERLAKEEAERKQKEEEERKQKEEEERKQKEEEERKLKE
EQERKAAEEKKAKEEAERKAKEEQERKAEEERKKKEEEERLERERKEREEQEKKAKEEAERIAKLEAE
KKAEEERKAKEEEERKAKEEEERKKKEEQERLAKEKEEAERKAAEEKKAKEEQERKEKEEAERK

>**Q9TW28|MYOM_DICDI** Myosin-M heavy chain SAH domain: 931-1042[11]
DFEQLVILENKRKEEERKKELERQRKEEEERQKELERQRREEEKELERKRKEEERELERQRKEEEKEQ
ERKRKEEEKEQERKKKEEKEIEKKRKEEEKKKKKNEQNLSLPSL

>**P02417|RL9_GEOSE** 50S ribosomal protein SAH domain: 41-74[12]
PANLKALEAQKQKEQRQAAEELANAKKLKEQLEK

	i→i+4				i→i+3						
Peptide name	**EE**	**EK**	**KE**	**KK**	**EE**	**EK**	**KE**	**KK**	**Total no. i→i+4 pairs** (opposite/same charge)	**Total no. i→i+3 pairs** (opposite/same charge)	**Fraction helix / %**
$(E_2K_2)_6$	**10** (50)	**0** (0)	**0** (0)	**10** (50)	**5** (23.8)	**6** (28.6)	**5** (23.8)	**5** (23.8)	**20** (0 / 20)	**21** (11 / 10)	**-1**
$(EK)_{12}$	**10** (50)	**0** (0)	**0** (0)	**10** (50)	**0** (0)	**11** (52.4)	**10** (47.6)	**0** (0)	**20** (0 / 20)	**21** (21 / 0)	**22**
$(KE)_{12}$	**10** (50)	**0** (0)	**0** (0)	**10** (50)	**0** (0)	**10** (47.6)	**11** (52.4)	**0** (0)	**20** (0 / 20)	**21** (21 / 0)	**12**
$(E_4K_4)_4$	**0** (0)	**16** (57.1)	**12** (42.9)	**0** (0)	**4** (13.8)	**12** (41.4)	**9** (31)	**4** (13.8)	**28** (28 / 0)	**29** (21 / 8)	**94**
$(E_4K_4)_3$	**0** (0)	**12** (60)	**8** (40)	**0** (0)	**3** (14.3)	**9** (42.9)	**6** (28.6)	**3** (14.3)	**20** (20 / 0)	**21** (15 / 6)	**74**
$(E_4K_4)_2$	**0** (0)	**8** (66.7)	**4** (33.3)	**0** (0)	**2** (15.4)	**6** (46.2)	**3** (23.1)	**2** (15.4)	**12** (12 / 0)	**13** (9 / 4)	**65**
$(E_4K_4)_1$	**0** (0)	**4** (100)	**0** (0)	**0** (0)	**1** (20)	**3** (60)	**0** (0)	**1** (20)	**4** (4 / 0)	**5** (3 / 2)	**14**
$(K_4E_4)_4$	**0** (0)	**12** (42.9)	**16** (57.1)	**0** (0)	**4** (13.8)	**9** (31)	**12** (41.4)	**4** (13.8)	**28** (28 / 0)	**29** (21 / 8)	**91**
$(K_4E_4)_3$	**0** (0)	**8** (40)	**12** (60)	**0** (0)	**3** (14.3)	**6** (28.6)	**9** (42.9)	**3** (14.3)	**20** (20 / 0)	**21** (15 / 6)	**62**
$(K_4E_4)_2$	**0** (0)	**4** (33.3)	**8** (66.7)	**0** (0)	**2** (15.4)	**3** (23.1)	**6** (46.2)	**2** (15.4)	**12** (12 / 0)	**13** (9 / 4)	**22**
$(K_4E_4)_1$	**0** (0)	**0** (0)	**4** (100)	**0** (0)	**1** (20)	**0** (0)	**3** (60)	**1** (20)	**4** (4 / 0)	**5** (3 / 2)	**-1**
$A_4(E_4K_4)_3A_4$	**0** (0)	**12** (42.9)	**8** (28.6)	**0** (0)	**3** (10.3)	**9** (31)	**6** (20.7)	**3** (10.3)	**28** (20 / 0)	**29** (15 / 6)	**87**
$A_4(E_4K_4)_2A_4$	**0** (0)	**8** (40)	**4** (20)	**0** (0)	**2** (9.5)	**6** (28.6)	**3** (14.3)	**2** (9.5)	**20** (12 / 0)	**21** (9 / 4)	**72**
$A_4(E_4K_4)_1A_4$	**0** (0)	**4** (33.3)	**0** (0)	**0** (0)	**1** (7.7)	**3** (23.1)	**0** (0)	**1** (7.7)	**12** (4 / 0)	**13** (3 / 2)	**53**
$A_4(K_4E_4)_3A_4$	**0** (0)	**8** (28.6)	**12** (42.9)	**0** (0)	**3** (10.3)	**6** (20.7)	**9** (31)	**3** (10.3)	**28** (20 / 0)	**29** (15 / 6)	**98**
$A_4(K_4E_4)_2A_4$	**0** (0)	**4** (20)	**8** (40)	**0** (0)	**2** (9.5)	**3** (14.3)	**6** (28.6)	**2** (9.5)	**20** (12 / 0)	**21** (9 / 4)	**78**
$A_4(K_4E_4)_1A_4$	**0** (0)	**0** (0)	**4** (33.3)	**0** (0)	**1** (7.7)	**0** (0)	**3** (23.1)	**1** (7.7)	**12** (4 / 0)	**13** (3 / 2)	**61**
$A_4(E_4K_4)_3$	**0** (0)	**12** (50)	**8** (33.3)	**0** (0)	**3** (12)	**9** (36)	**6** (24)	**3** (12)	**24** (20 / 0)	**25** (15 / 6)	**87**
$(E_4K_4)_3A_4$	**0** (0)	**12** (50)	**8** (33.3)	**0** (0)	**3** (12)	**9** (36)	**6** (24)	**3** (12)	**24** (20 / 0)	**25** (15 / 6)	**85**
$A_4(K_4E_4)_3$	**0** (0)	**8** (33.3)	**12** (50)	**0** (0)	**3** (12)	**6** (24)	**9** (36)	**3** (12)	**24** (20 / 0)	**25** (15 / 6)	**91**
$(K_4E_4)_3A_4$	**0** (0)	**8** (33.3)	**12** (50)	**0** (0)	**3** (12)	**6** (24)	**9** (36)	**3** (12)	**24** (20 / 0)	**25** (15 / 6)	**85**
$A_4(E_4K_4)A_4(E_4K_4)A_4$	**0** (0)	**8** (33.3)	**0** (0)	**0** (0)	**2** (8)	**6** (24)	**0** (0)	**2** (8)	**24** (8 / 0)	**25** (6 / 4)	**83**
$A_4(K_4E_4)A_4(K_4E_4)A_4$	**0** (0)	**0** (0)	**8** (33.3)	**0** (0)	**2** (8)	**0** (0)	**6** (24)	**2** (8)	**24** (8 / 0)	**25** (6 / 4)	**97**
$(E_3K_3)_4$	**3** (12.5)	**8** (40)	**6** (30)	**3** (12.5)	**0** (0)	**12** (57.1)	**9** (42.9)	**0** (0)	**20** (14 / 0)	**21** (21 / 0)	**67**
$(K_3E_3)_4$	**3** (12.5)	**6** (30)	**8** (40)	**3** (12.5)	**0** (0)	**9** (42.9)	**12** (57.1)	**0** (0)	**20** (14 / 0)	**21** (21 / 0)	**33**
$A_4(E_3K_3)_4A_4$	**3** (9.38)	**8** (28.6)	**6** (21.4)	**3** (9.38)	**0** (0)	**12** (41.4)	**9** (31)	**0** (0)	**28** (14 / 0)	**29** (21 / 0)	**82**
$A_4(K_3E_3)_4A_4$	**3** (9.38)	**6** (21.4)	**8** (28.6)	**3** (9.38)	**0** (0)	**9** (31)	**12** (41.4)	**0** (0)	**28** (14 / 0)	**29** (21 / 0)	**78**

Supplementary Table 1: Inventory of all possible i→i+3/4 E/K→E/K interactions in the fully helical states of the designed peptides. Numbers of potential ion pairs are in bold type, and the percentage of possible pairs is given in brackets (terminal glycine and tryptophan/tyrosine residues were disregarded when counting the total number of possible i→i+3/4 pairs).

Peptide	Peptide mass / Da	AUC mass / Da	x monomer mass	99% confidence limits
$(E_4K_4)_3$	3447	3958	1.1	3928 – 3988
$(K_4E_4)_3$	3447	3786	1.1	3749 – 3797
$A_4(E_4K_4)_3A_4$	4015	4431	1.1	4404 – 4457
$A_4(K_4E_4)_3A_4$	4015	4599	1.1	4560 – 4638

Supplementary Table 2: AUC sedimentation equilibrium data.

Supplementary Tables 3-6: ^{1}H chemical shifts of **$(E_4K_4)_3$**, **$(K_4E_4)_3$**, **$(K_4E_4)_4$** and **$A_4(K_4E_4)_3A_4$**.

Residue	HN / ppm	Hα / ppm	Hβ / ppm	Others
G1	8.53	3.98, 3.98		
E2	8.80	4.11	-	-
E3	8.80	4.08	-	-
E4	8.21	4.04	-	-
E5	8.22	4.04	-	-
K6	8.09	4.02	1.88	Hγ 1.36
K7	7.93	4.11	-	-
K8	8.05	4.11 – 4.07*	-	-
K9	8.05	4.11 – 4.07*	-	-
E10	8.22	4.05	-	-
E11	8.29	4.08	2.18, 2.10	Hγ 2.41, 2.30
E12	8.25	4.04	-	-
E13	8.31	4.04	2.13, 2.05	Hγ 2.41, 2.32
K14	8.06	4.03	-	-
K15	7.99	4.11	1.91	-
K16	8.02	4.09	-	-
K17	8.02	4.09	-	-
E18	8.18	4.04	-	-
E19	8.25	4.04	-	-
E20	8.25	4.04	-	-
E21	8.17	4.00	-	-
K22	7.85	3.93	1.84	-
K23	7.75	4.05	1.84	Hγ 1.35
K24	7.80	3.99	1.80	-
K25	7.93	4.12	-	-
G26	8.12	3.92, 3.84		
W27	7.95	4.63	3.31, 3.24	2H 7.14 4H 7.62 5H 7.12 6H 7.21 7H 7.46 NH 10.19

Supplementary Table 3: ^{1}H chemical shifts of **$(E_4K_4)_3$**. Much of the side chain region could not be assigned due to signal overlap.
*A range is given where the assignment is ambiguous due to spectral crowding.

Residue	HN / ppm	Hα / ppm	Hβ / ppm	Others
G1	8.47	3.96, 3.91		
K2	8.48	4.18	-	-
K3	8.44	4.23	1.83	Hγ 1.50, 1.41
K4	8.13	4.19	-	-
K5	8.25	4.16	-	-
E6	8.49	4.11	-	-
E7	8.34	4.12	-	-
E8	8.33	4.09	-	-
E9	8.41	4.09	-	-
K10	8.10	4.09	-	-
K11	8.00	4.16	-	-
K12	8.05	4.13	-	-
K13	8.07	4.15	-	-
E14	8.28	4.09	-	-
E15	8.33	4.09	-	-
E16	8.28	4.09	-	-
E17	8.33	4.09	-	-
K18	8.03	4.09	-	-
K19	7.97	4.12	1.88	-
K20	8.01	4.13	-	-
K21	8.02	4.13	-	-
E22	8.17	3.99	2.06	Hγ 2.39
E23	8.14	4.07	-	-
E24	8.09	3.91	-	-
E25	8.25	4.14	1.97	Hγ 2.34, 2.24
G26	8.14	3.90, 3.88		
W27	7.90	4.65	3.32, 3.29	2H 7.22 4H 7.66 5H 7.14 6H 7.22 7H 7.48 NH 10.23

Supplementary Table 4: ^{1}H chemical shifts of **$(K_4E_4)_3$**. Much of the side chain region could not be assigned due to signal overlap.

Residue	HN / ppm	Hα / ppm	Hβ / ppm	Others
G1	8.47	3.96, 3.91		
K2	8.49	4.17	1.82	-
K3	8.43	4.22	1.83	Hγ 1.49
K4	8.10	4.18	-	-
K5	8.22	4.16	-	-
E6	8.47	4.1	-	-
E7	8.33	4.11	-	-
E8	8.32	4.08	-	-
E9	8.40	4.07	-	-
K10	8.08	4.08	-	-
K11	7.97	4.15	-	-
K12	8.03	4.14 – 4.12*	-	-
K13	8.05	4.15	-	-
E14	8.27	4.08	-	-
E15	8.34	4.1	-	-
E16	8.30	4.08	-	-
E17	8.36	4.07	-	-
K18	8.08	4.08	-	-
K19	8.08	4.08	-	-
K20	8.00	4.14	-	-
K21	8.03	4.14 – 4.12*	-	-
E22	8.23	4.14	-	-
E23	8.30	4.08	-	-
E24	8.26	4.06	-	-
E25	8.31	4.06	-	-
K26	8.00	4.07	-	-
K27	7.93	4.11	-	-
K28	7.98	4.12	-	-
K29	7.98	4.12	-	-
E30	8.13	3.98	-	-
E31	8.11	4.06	-	-
E32	8.07	3.89	-	-
E33	8.23	4.07	-	-
G34	8.11	3.91, 3.88		
W35	7.90	4.65	3.33, 3.29	2H 7.22 4H 7.66 5H 7.14 6H 7.22 7H 7.47 NH 10.25

Supplementary Table 5: ^{1}H chemical shifts of **$(K_4E_4)_4$**. Much of the side chain region could not be assigned due to signal overlap.
*A range is given where the assignment is ambiguous due to spectral crowding.

Residue	HN / ppm	Hα / ppm	Hβ / ppm	Others
G1	8.52	4.01, 3.91		
A2	8.67	4.21	1.45	-
A3	8.61	4.16	1.46	-
A4	7.99	4.15 – 4.12*	1.43	-
A5	8.04	4.16	1.48	-
K6	8.06	4.10	-	
K7	7.85	4.16	1.91, 1.89	Hγ 1.69
K8	7.96	4.14	-	-
K9	7.91	4.16	2.01, 1.93	-
E10	8.29	4.07	-	-
E11	8.39	4.09	-	-
E12	8.35	4.11 – 4.08*	-	-
E13	8.42	4.07	-	-
K14	8.07	4.08	-	-
K15	7.97	4.14	-	-
K16	8.01	4.12	-	-
K17	8.01	4.15	-	-
E18	8.25	4.08	2.08	Hγ 2.53
E19	8.34	4.11 – 4.08*	-	-
E20	8.3	4.07	-	-
E21	8.36	4.07	-	-
K22	8.05	4.08	-	-
K23	7.96	4.14	-	-
K24	7.99	4.13	-	-
K25	7.99	4.16	-	-
E26	8.23	4.08	2.23, 2.08	2.53, 2.28
E27	8.35	4.11 – 4.08*	-	-
E28	8.35	4.11 – 4.08*	-	-
E29	8.41	4.07	-	-
A30	8.06	4.10	1.42	-
A31	7.93	4.11	1.45	-
A32	7.82	4.06	1.40	-
A33	7.78	4.21	1.33	-
G34	7.97	3.90, 3.86		
W35	7.87	4.67	3.33, 3.27	2H 7.22 4H 7.66 5H 7.14 6H 7.22 7H 7.48 NH 10.25

Supplementary Table 6: ^{1}H chemical shifts of **$A_4(K_4E_4)_3A_4$**. Much of the side chain region could not be assigned due to signal overlap.
*A range is given where the assignment is ambiguous due to spectral crowding.

Peptide	ΔG_f / kJ mol^{-1}	ΔH_f / kJ mol^{-1}	ΔS_f / J K^{-1} mol^{-1}	Fraction helix / % (calc.)	Fraction folded / % (data fitting)
$(E_4K_4)_4$	-4.0	-23.7	-71	94	85
$(E_4K_4)_3$	-3.2	-29.8	-96	74	80
$(E_4K_4)_2$	-1.4	-22.7	-77	65	65
(E_4K_4)	-	-	-	14	-
$(K_4E_4)_4$	-3.7	-39.9	-131	91	81
$(K_4E_4)_3$	-1.5	-38.1	-132	62	65
$(K_4E_4)_2$	-	-	-	22	-
(K_4E_4)	-	-	-	-1	-
$A_4(E_4K_4)_3A_4$	-3.0	-26.2	-83	87	79
$A_4(E_4K_4)_2A_4$	-1.7	-27.2	-92	72	68
$A_4(E_4K_4)A_4$	-0.3	-33.1	-118	53	53
$A_4(K_4E_4)_3A_4$	-4.4	-47.0	-153	98	87
$A_4(K_4E_4)_2A_4$	-3.1	-51.2	-173	79	79
$A_4(K_4E_4)A_4$	-1.3	-51.7	-181	61	64
$A_4(E_4K_4)_3$	-2.9	-26.6	-85	87	78
$(E_4K_4)_3A_4$	-3.0	-25.8	-82	85	79
$A_4(K_4E_4)_3$	-3.6	-47.4	-158	91	83
$(K_4E_4)_3A_4$	-3.0	-43.0	-144	85	79
$A_4(E_4K_4)A_4(E_4K_4)A_4$	-3.0	-41.0	-137	83	79
$A_4(K_4E_4)A_4(K_4E_4)A_4$	-4.8	-56.5	-186	97	89
$(E_3K_3)_4$	-1.7	-34.6	-119	67	68
$(K_3E_3)_4$	0.06	-40.7	-147	33	48
$A_4(E_3K_3)_4A_4$	-2.6	-35.5	-118	82	74
$A_4(K_3E_3)_4A_4$	-2.1	-48.4	-167	78	71

Supplementary Table 7: Gibbs' free energies (at 5 °C) and enthalpies of folding with estimated entropies for each peptide calculated from a direct fit of the thermal denaturation data. The enthalpies of folding for all **-(K_xE_x)-** based peptides are better than equivalent **-(E_xK_x)-** peptides. For reference, experimental fractions helix (5 °C) are given as calculated directly from the MRE at 222 nm (calc.),[13,14] (Table 1) and by the data fitting.

a	**$E_i \rightarrow K_{i+4}$**					**$K_i \rightarrow E_{i+4}$**				
Helix region	Observed in sequence	Expected in sequence[a]	Observed / Expected	Observed in structure[b]	% salt bridges made	Observed in sequence	Expected in sequence[a]	Observed / Expected	Observed in structure[b]	% salt bridges made
N-terminal region	309	195	1.58	33	10.7	162	107	1.51	29	17.9
Central helical region	259	146	1.77	25	9.7	251	155	1.62	52	20.7
C-terminal region	292	175	1.67	34	11.6	305	152	2.00	71	23.2
Total	860	523	1.64	92	10.7	718	413	1.74	152	21.2

b	**$E_i \rightarrow K_{i+3}$**					**$K_i \rightarrow E_{i+3}$**				
Helix region	Observed in sequence	Expected in sequence[a]	Observed / Expected	Observed in structure[b]	% salt bridges made	Observed in sequence	Expected in sequence[a]	Observed / Expected	Observed in structure[b]	% salt bridges made
N-terminal region	277	193	1.44	56	20.2	160	101	1.58	28	17.5
Central helical region	271	175	1.55	60	22.1	272	177	1.54	43	15.8
C-terminal region	303	179	1.69	83	27.4	246	153	1.61	37	15.0
Total	851	547	1.56	199	23.4	678	431	1.57	108	15.6

c	**$E_i \rightarrow E_{i+4}$**					**$K_i \rightarrow K_{i+4}$**				
Helix region	Observed in sequence	Expected in sequence[a]	Observed / Expected	Mean side-chain atom distance[c] (Å)	% residues ≤ 4Å apart[d]	Observed in sequence	Expected in sequence[a]	Observed / Expected	Mean side-chain atom distance[c] (Å)	% residues ≤ 4Å apart[d]
N-terminal region	289	223	1.30	6.7 (2.2)	9.7	100	94	1.07	8.6 (2.7)	1.0
Central helical region	231	187	1.23	6.5 (2.1)	10.0	147	121	1.21	9.7 (2.8)	1.0
C-terminal region	222	186	1.19	6.0 (1.9)	12.6	155	144	1.08	8.5 (2.5)	1.3
Total	742	596	1.25	6.4 (2.1)	10.6	402	359	1.12	9.0 (2.7)	1.0

d	**$E_i \rightarrow E_{i+3}$**					**$K_i \rightarrow K_{i+3}$**				
Helix region	Observed in sequence	Expected in sequence[a]	Observed / Expected	Mean side-chain atom distance[c] (Å)	% residues ≤ 4Å apart[d]	Observed in sequence	Expected in sequence[a]	Observed / Expected	Mean side-chain atom distance[c] (Å)	% residues ≤ 4Å apart[d]
N-terminal region	266	218	1.22	7.7 (2.0)	4.5	89	91	0.97	9.0 (2.2)	1.1
Central helical region	222	214	1.23	7.6 (2.2)	4.5	168	145	1.16	9.7 (2.8)	1.2
C-terminal region	197	189	1.04	7.2 (2.1)	7.6	131	144	0.91	9.8 (2.6)	1.5
Total	685	621	1.10	7.5 (2.1)	5.1	388	380	1.02	9.5 (2.4)	1.3

Supplementary Table 8: Results from sequence and structural analysis of the Protein Data Bank. Numbers of attractive (**a&b**) $E_i \rightarrow K_{i+4}$, $K_i \rightarrow E_{i+4}$, $E_i \rightarrow K_{i+3}$, and $K_i \rightarrow E_{i+3}$ and repulsive (**c&d**) $E_i \rightarrow E_{i+4}$, $K_i \rightarrow K_{i+4}$, $E_i \rightarrow E_{i+3}$, and $K_i \rightarrow K_{i+3}$ pairs in α-helices. [a]Expected numbers of pairs were estimated using the natural abundance of each residue in the dataset. [b]Salt bridges were considered to be formed if the distance between Lys Nζ and the centroid of (Glu Oε1, Oε2) was ≤ 4Å. [c]For the repulsive pairs, the mean shortest distance between side-chain atoms was recorded (standard deviation in brackets), along with [d]the percentage of repulsive residue pairs with sub-4 Å distances between them. For comparison, the propensities of Glu vs Lys at various positions within α-helical structure are as follows: *N*-terminal region, 1.24 vs 1.26; central region, 0.88 vs 0.98; and *C*-terminal region (1 vs 1.26).

E_i→ K_{i+4}↓	*tt*		*tg*$^+$		*tg*$^-$		*g*$^+$*t*		*g*$^+$*g*$^+$		*g*$^+$*g*$^-$		*g*$^-$*t*		*g*$^-$*g*$^+$		*g*$^-$*g*$^-$	
tt	8.58	6.88	4.06	1.62	0.54	0.40	0.18	0.00	0.07	0.00	0.00	0.00	15.40	19.84	0.00	0.00	7.57	6.88
	1	0.00	×	0.00	3	0.40	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00
tg$^+$	2.02	2.43	0.96	0.40	0.13	0.00	0.04	0.00	0.02	0.00	0.00	0.00	3.63	1.62	0.00	0.00	1.79	0.40
	×	0.00	2	0.00	4	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00
tg$^-$	0.28	0.81	0.13	0.00	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.25	0.81
	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00
g$^+$***t***	0.16	0.00	0.08	0.40	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.40	0.00	0.00	0.14	0.00
	1	0.00	×	0.00	8	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00
g$^+$***g***$^+$	0.05	0.40	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.04	0.00
	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00
g$^+$***g***$^-$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00
g$^-$***t***	10.23	20.24	4.84	3.24	0.64	0.00	0.21	0.40	0.09	0.00	0.00	0.00	18.36	15.38	0.00	0.00	9.03	7.29
	12	6.88	8	2.45	22	0.00	5	0.00	×	0.00	×	0.00	1	0.00	×	0.00	2	0.00
g$^-$***g***$^+$	0.09	0.00	0.04	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.40	0.00	0.00	0.08	0.00
	13	0.00	21	0.00	26	0.00	5	0.00	×	0.00	×	0.00	1	0.00	×	0.00	7	0.00
g$^-$***g***$^-$	2.16	2.43	1.02	4.05	0.14	0.00	0.05	0.00	0.02	0.00	0.00	0.00	3.88	2.43	0.00	0.00	1.91	0.81
	6	0.00	10	0.40	15	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	2	0.00

Totals	
100	247
n/a	25

E_i→ K_{i+4}↓	***X***	
Y	(a)	(b)
	(c)	(d)

X: χ_1, χ_2 rotamers for Glu
Y: χ_1, χ_2 rotamers for Lys
(a) Expected percentage of dataset with this rotamer combination, based on joint probabilities of individual amino acid rotamers
(b) Observed percentage of pairs with this rotamer combination in sequence
(c) Number of full rotamer combinations (max 27) that have the potential to make a salt bridge, × = rotamer combination doesn't exist
(d) Observed percentage of pairs with this rotamer combination that make a salt bridge ($O\varepsilon1/O\varepsilon2\cdots N\zeta < 4$Å)

E_{i+4}→ K_i↓	***tt***		***tg⁺***		***tg⁻***		***g⁺t***		***g⁺g⁺***		***g⁺g⁻***		***g⁻t***		***g⁻g⁺***		***g⁻g⁻***	
tt	8.58	8.13	4.06	3.66	0.54	0.00	0.18	0.00	0.07	0.00	0.00	0.00	15.40	34.96	0.00	0.00	7.57	4.88
	3	0.00[d]	×	0.00	×	0.00	5	0.00	×	0.00	×	0.00	12	15.60	×	0.00	5	0.82
tg⁺	2.02	0.00	0.96	1.22	0.13	0.00	0.04	0.41	0.02	0.00	0.00	0.00	3.63	2.03	0.00	0.00	1.79	3.66
	3	0.00	5	0.41	×	0.00	2	0.41	×	0.00	×	0.00	12	0.41	×	0.00	14	0.82
tg⁻	0.28	1.63	0.13	0.81	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.50	0.41	0.00	0.00	0.25	0.00
	13	1.20	11	0.00	2	0.00	18	0.00	×	0.00	×	0.00	26	0.41	×	0.00	14	0.00
g⁺t	0.16	0.41	0.08	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.41	0.00	0.00	0.14	0.00
	×	0.00	×	0.00	×	0.00	2	0.00	×	0.00	×	0.00	5	0.41	×	0.00	×	0.00
g⁺g⁺	0.05	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.04	0.00
	×	0.00	×	0.00	×	0.00	17	0.00	×	0.00	×	0.00	12	0.00	×	0.00	×	0.00
g⁺g⁻	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00
g⁻t	10.23	6.10	4.84	3.66	0.64	2.44	0.21	0.00	0.09	0.00	0.00	0.00	18.36	13.82	0.00	0.00	9.03	4.47
	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	2	0.00	×	0.00	×	0.00
g⁻g⁺	0.09	0.00	0.04	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.08	0.00
	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00
g⁻g⁻	2.16	1.63	1.02	0.81	0.14	0.41	0.05	0.00	0.02	0.00	0.00	0.00	3.88	2.03	0.00	0.00	1.91	2.03
	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	3	0.41	×	0.00	3	0.00

Totals	
100	246
n/a	52

$E_i \rightarrow$ / $K_{i+3} \downarrow$	***tt***		***tg⁺***		***tg⁻***		***g⁺t***		***g⁺g⁺***		***g⁺g⁻***		***g⁻t***		***g⁻g⁺***		***g⁻g⁻***	
tt	8.58	6.95	4.06	10.42	0.54	0.39	0.18	0.00	0.07	0.39	0.00	0.00	15.40	9.65	0.00	0.00	7.57	11.58
	×	0.00	9	1.93	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	9	2.32
tg⁺	2.02	0.39	0.96	0.00	0.13	0.00	0.04	0.00	0.02	0.00	0.00	0.00	3.63	1.93	0.00	0.00	1.79	0.39
	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00
tg⁻	0.28	0.00	0.13	0.00	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.25	0.00
	×	0.00	8	0.00	1	0.00	2	0.00	×	0.00	×	0.00	×	0.00	×	0.00	6	0.00
g⁺t	0.16	0.39	0.08	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.00	0.00	0.00	0.14	0.00
	9	0.00	24	0.00	8	0.00	×	0.00	×	0.00	×	0.00	6	0.00	×	0.00	25	0.00
g⁺g⁺	0.05	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.39	0.00	0.00	0.04	0.00
	5	0.00	21	0.00	14	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	8	0.00
g⁺g⁻	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00
g⁻t	10.23	3.86	4.84	8.11	0.64	0.39	0.21	0.00	0.09	0.00	0.00	0.00	18.36	11.58	0.00	0.00	9.03	29.34
	×	0.00	5	1.20	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	9	17.76
g⁻g⁺	0.09	0.00	0.04	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.08	0.39
	×	0.00	2	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	5	0.00
g⁻g⁻	2.16	1.16	1.02	0.77	0.14	0.39	0.05	0.00	0.02	0.00	0.00	0.00	3.88	0.77	0.00	0.00	1.91	0.39
	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00

Totals	
100	259
n/a	60

E_{i+3}→ K_i↓	*tt*		*tg⁺*		*tg⁻*		*g⁺t*		*g⁺g⁺*		*g⁺g⁻*		*g⁻t*		*g⁻g⁺*		*g⁻g⁻*	
tt	8.58	4.20	4.06	1.53	0.54	0.00	0.18	0.00	0.07	0.00	0.00	0.00	15.40	10.31	0.00	0.00	7.57	3.44
	×	0.00	×	0.00	×	0.00	3	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00
tg⁺	2.02	8.40	0.96	0.00	0.13	0.00	0.04	0.00	0.02	0.38	0.00	0.00	3.63	8.78	0.00	0.00	1.79	0.76
	7	5.73	×	0.00	6	0.00	22	0.00	×	0.00	×	0.00	5	2.30	×	0.00	×	0.00
tg⁻	0.28	0.38	0.13	0.00	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.25	0.00
	×	0.00	×	0.00	2	0.00	6	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00
g⁺t	0.16	0.00	0.08	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.76	0.00	0.00	0.14	0.00
	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00
g⁺g⁺	0.05	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.04	0.00
	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00
g⁺g⁻	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00
g⁻t	10.23	12.98	4.84	3.05	0.64	0.38	0.21	1.15	0.09	0.00	0.00	0.00	18.36	20.61	0.00	0.00	9.03	4.58
	×	0.76	×	0.00	×	0.00	1	0.00	×	0.00	×	0.00	×	0.00	×	0.00	×	0.00
g⁻g⁺	0.09	0.00	0.04	0.38	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.38	0.00	0.00	0.08	0.00
	×	0.00	×	0.00	×	0.00	7	0.00	×	0.00	×	0.00	5	0.00	×	0.00	×	0.00
g⁻g⁻	2.16	4.58	1.02	0.38	0.14	0.00	0.05	0.00	0.02	0.00	0.00	0.00	3.88	10.69	0.00	0.00	1.91	0.76
	7	2.67	×	0.00	×	0.00	18	0.00	×	0.00	×	0.00	9	4.20	×	0.00	×	0.00

Totals	
100	262
n/a	43

Supplementary Table 9: χ_1,χ_2 rotamer combinations for $E_i \rightarrow K_{i+4}$, $K_i \rightarrow E_{i+4}$, $E_i \rightarrow K_{i+3}$, $K_i \rightarrow E_{i+3}$ pairs. Grey shaded boxes indicate rotamer combinations that are disallowed in α-helices and therefore they were not counted in the analysis. Green shaded boxes identify the rotamer combinations that do have the potential to form salt bridges. The number in the green box expresses the number of rotamers within each χ_1,χ_2 category which have the potential to form salt bridges. Within each χ_1,χ_2 category, there are 3 possible rotamers for Glu and 9 possible rotamers for Lys, giving a maximum 27 rotamers per box. Yellow shaded boxes identify the rotamer combinations making salt bridges that are observed in α-helices in the PDB. Numbers (a), (b) and (d) in the table are expressed as percentages for the purposes of comparison: the total for each category is in the lower right-hand box for raw number extrapolation. Total possible rotamer combinations for each pair: $E_i \rightarrow K_{i+4}$, 175 (13.5%);

$K_i \rightarrow E_{i+4}$, 189 (14.6%); $E_i \rightarrow K_{i+3}$, 176 (13.6%); $K_i \rightarrow E_{i+3}$, 98 (7.6%). Notably, $K_i \rightarrow E_{i+4}$ has the better orientations to make salt bridge interactions with 21% made/14.6% expected compared with 11% made/13.5% expected for $E \rightarrow K_{i+4}$. Moreover, for 38% of the rotamer combinations for $E \rightarrow K_{i+4}$ pairs, salt bridges were not possible, whereas for $K_i \rightarrow E_{i+4}$ this number reduced to 24%.

1 Berman, H.M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235-242, (2000).
2 Lacroix, E., Viguera, A.R. & Serrano, L. Elucidating the folding problem of α-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J. Mol. Biol.* **284**, 173-191, (1998).
3 Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659, (2006).
4 Peckham, M. & Knight, P.J. When a predicted coiled coil is really a single α-helix, in myosins and other proteins. *Soft Matter* **5**, 2493-2503, (2009).
5 Knight, P.J. *et al.* The predicted coiled-coil domain of myosin 10 forms a novel elongated domain that lengthens the head. *J. Biol. Chem.* **280**, 34702-34708, (2005).
6 Süveges, D., Gáspári, Z., Tóth, G. & Nyitray, L. Charged single α-helix: A versatile protein structural motif. *Proteins: Structure Function and Bioinformatics* **74**, 905-916, (2009).
7 Spink, B.J., Sivaramakrishnan, S., Lipfert, J., Doniach, S. & Spudich, J.A. Long single α-helical tail domains bridge the gap between structure and function of myosin VI. *Nat. Struct. Mol. Biol.* **15**, 591-597, (2008).
8 Wang, C.L. *et al.* A long helix from the central region of smooth muscle caldesmon. *J. Biol. Chem.* **266**, 13958-13963, (1991).
9 Yang, Y. *et al.* A FERM domain autoregulates Drosophila myosin 7a activity. *Proc. Natl. Acad. Sci. USA* **106**, 4189-4194, (2009).
10 Sivaramakrishnan, S. *et al.* Combining single-molecule optical trapping and small-angle X-ray scattering measurements to compute the persistence length of a protein ER/K α-helix. *Biophys. J.* **97**, 2993-2999, (2009).
11 Baboolal, T.G. *et al.* The SAH domain extends the functional length of the myosin lever. *Proc. Natl. Acad. Sci. USA* **106**, 22193-22198, (2009).
12 Kuhlman, B., Yang, H.Y., Boice, J.A., Fairman, R. & Raleigh, D.P. An exceptionally stable helix from the ribosomal protein L9: implications for protein folding and stability. *J. Mol. Biol.* **270**, 640-647, (1997).
13 Scholtz, J.M., Qian, H., York, E.J., Stewart, J.M. & Baldwin, R.L. Parameters of helix-coil transition theory for alanine-based peptides of varying chain lengths in water. *Biopolymers* **31**, 1463-1470, (1991).
14 Myers, J.K., Pace, C.N. & Scholtz, J.M. A direct comparison of helix propensity in proteins and peptides. *Proc. Natl. Acad. Sci. USA* **94**, 2833-2837, (1997).