What Are the Minimal Folding Seeds in Proteins? Experimental and Theoretical Assessment of Secondary Structure Propensities of Small Peptide Fragments

Supporting Information

Zuzana Osifová [a,b,‡] Tadeáš Kalvoda,[a,‡,] Jakub Galgonek,[a] Martin Culka,[a] Jiří Vondrášek,[a] Petr Bouř,[a] Lucie Bednárová,[a] Valery Andrushchenko,[a,*] Martin Dračínský,[a,*] and Lubomír Rulíšek[a,*]*

[a] Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Flemingovo náměstí 2, 160 00, Praha 6, Czech Republic

[b] Department of Organic Chemistry, Faculty of Science, Charles University, Hlavova 2030, Prague 128 00, Czech Republic

* rulisek@uochb.cas.cz; dracinsky@uochb.cas.cz; andrushchenko@uochb.cas.cz, tadeas.kalvoda@uochb.cas.cz

‡ These authors contributed equally to this work

Supporting Information Contents

1. CATWEAMEKCK undecapeptide within the 20-α-hydroxysteroid dehydrogenase.

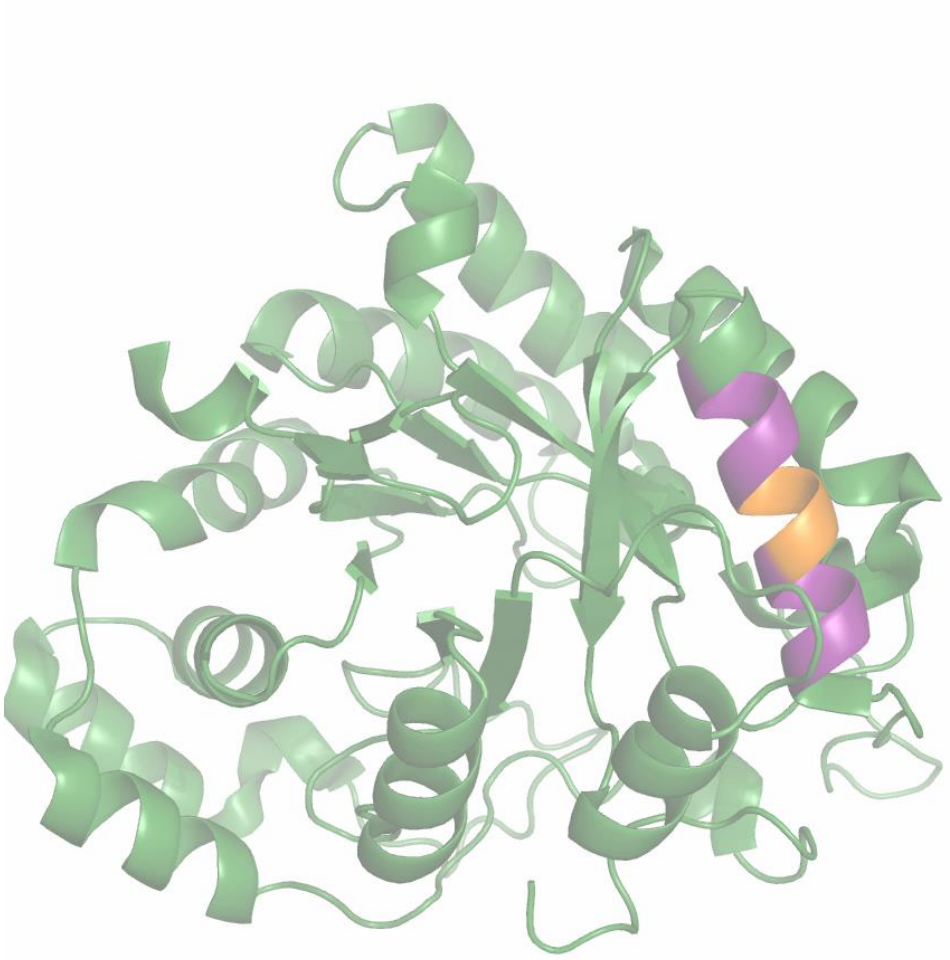


Fig. S1 The crystallographic structure of the 20-α-hydroxysteroid dehydrogenase (PDBID 1Q5M). The CATWEAMEKCK undecapeptide in the chain B (purple) and its EAM core (orange) are highlighted. Side chains of the undecapeptide were omitted for clarity.

2. Validation of the computational protocol for calculating VCD spectra.

To determine the accuracy of the computational protocol, we first evaluated the quality of the protocol for calculating VCD spectra. As model systems, we choose alanine tripeptide in all-α-helical, all-extended and all-polyproline II conformations, as it has been shown that the side chains of the peptide do not affect the resulting spectrum[1]. The resulting spectra are shown in **Fig. S2**. The **Fig. S2** depicts a pronounced positive couplet in amide I region at ~1700(−)/~1650(+) cm^{-1} in case of α-helical conformer. PPII conformer has a negative couplet at ~1680(+)/~1660(−) cm^{-1}. In the case of extended conformer, there are two negative maxima at ~1680 and ~1660 cm^{-1} and one positive maximum at ~1675 cm^{-1}. Other maxima in the spectrum of extended conformer are much less pronounced. While there is an unexpected positive maximum in the extended spectrum at ~1675 cm^{-1}, all other significant bands agree with experimental spectra of protein structures. It is therefore safe to conclude that our computational protocol is robust and usable for the predictions of VCD spectra.

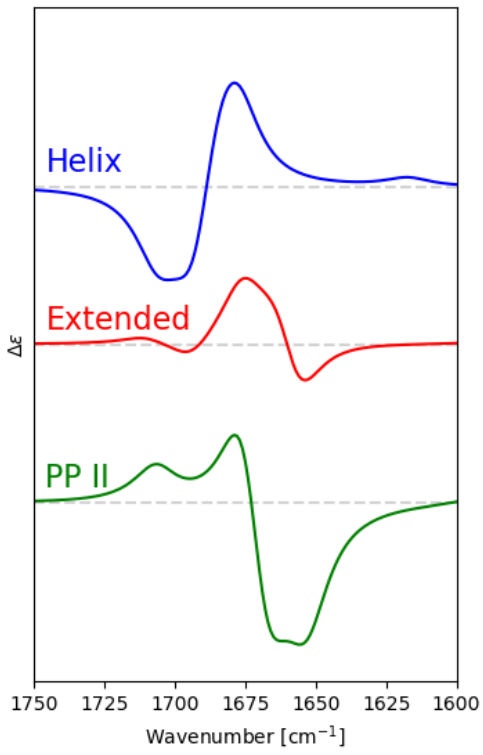


Fig. S2 The amide I region of VCD spectra of all-α-helical, all-extended and all-polyproline II conformations of alanine tripeptide, as reference spectra of pure α-helix, β-sheet and polyproline II.

3. Experimental IR and VCD spectra of EAM and VIV tripeptides

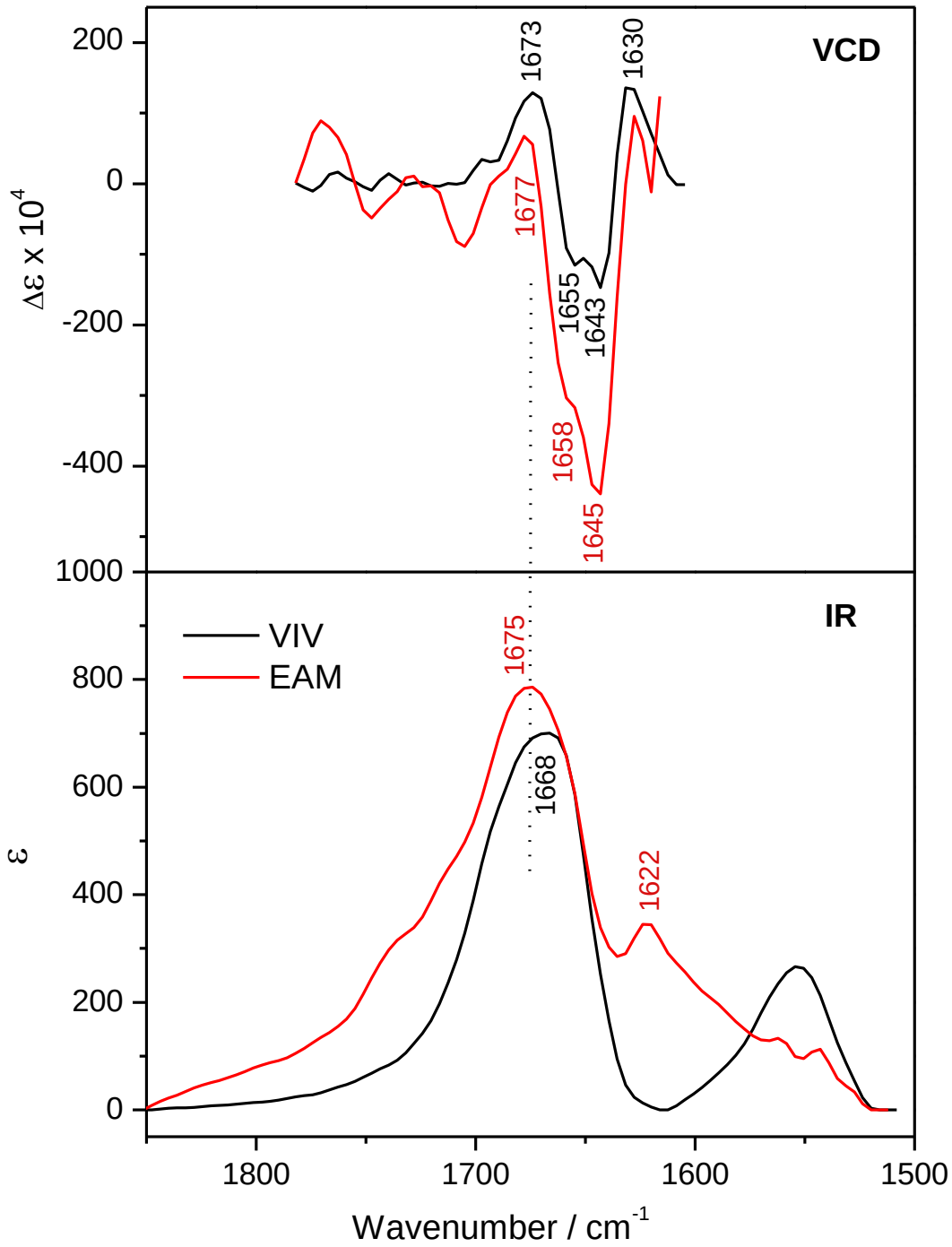


Fig. S3 Experimental IR absorption (bottom) and VCD (top) spectra of VIV and EAM peptides. The vertical dotted line shows the position of the IR band maxima with respect to the VCD spectra to demonstrate the large shift of the corresponding negative VCD bands to lower wavenumbers, discussed in the text.

4. Examples of VCD spectral shapes obtained by convolution of spectra characteristic for different secondary structures

Fig. S4 illustrates VCD spectral patterns obtained by convolution of characteristic α-helix (positive couplet at ~1670(-)/1645(+) cm^{-1}) and PPII (negative couplet at ~1685(+)/1645(-) cm^{-1}) spectra at different ratios. The composite trace obtained by 50:50 convolution shows a strong negative band at ~1665 cm^{-1} coming from the α-helical couplet accompanied by a weaker positive band at ~1700 cm^{-1} contributed by the PPII couplet. Depending on the ratio of α-helix/PPII contributions, the positions of both convoluted bands are slightly shifted. Furthermore, a stronger positive band at ~1700 cm^{-1} reflects lager PPII contribution.

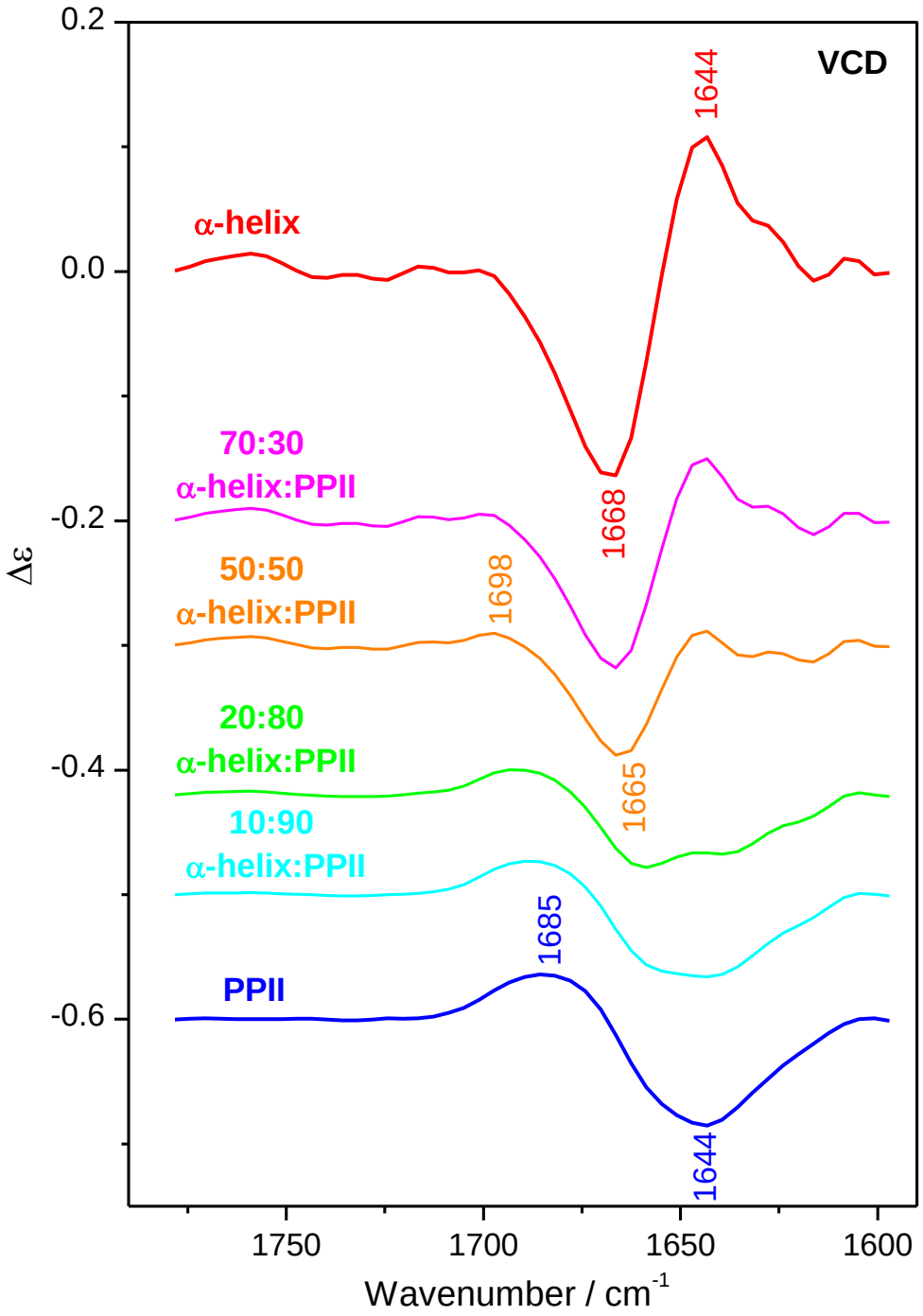


Fig. S4 Experimental VCD spectra characteristic for α-helical (red) and PPII (blue) peptide secondary structures and spectral traces obtained by their convolution at different ratios.

5. ECD experiments

ECD spectrum for the undecapeptide (CATWEAMEKCK) in methanol (**Fig. 5** in the main text) with characteristic minima at 207 nm and 223 nm shows typical α-helical conformation, confirmed also by secondary structure assignment (55% α-helix, 16% 3_{10} helix, 1% β-sheet, 9.6% β-turn, 2% PPII and 16% unordered structure). The ECD spectrum of the undecapeptide in water was characterized by negative maximum at 201 nm accompanied by a shoulder at 222 nm showing that the peptide is less structured, which is confirmed also by structural analysis (**Table S1**). As expected, the unordered fraction increases with temperature but the peptide core preserves its original structure to some extent (**Fig. S5**).

For tripeptides alone, the pattern of ECD spectra in water or methanol is in general negative spectral band at ~ 198 nm accompanied by negative shoulder at ~225 nm, which differs in intensity for various tripeptides (**Fig. 5** in the main text). According to the spectral analysis, these tripeptides have low content of right-handed helical structure but relatively high content of β-sheet and PPII structure (**Table S1**). The fraction of PPII structure decreases with temperature in favour of unordered structure: with increasing temperature the spectral band at ~198 nm shifts to 199 nm while its intensity decreases. Simultaneously, the positive spectral band at ~214 nm appears with isosbestic point at ~204 nm accompanied by negative shoulder at ~ 230 nm (**Fig. S5**).[2]

ECD spectra of tripeptide VIV in water shows characteristic minima at ~200 nm and ~220 nm with intensity, which in the analysis represents ~30% right-handed helical structure (**Table S1**). On the other hand, when VIV is dissolved in methanol, ECD spectrum clearly shows smaller portion of right-handed helical structure. As for other tripeptides, the increase of the temperature causes decrease of the amount of PPII structure. It should be pointed out that the numbers in **Table S1** are *very* approximate.

Table S1. Secondary structure assignment of six tripeptides and undecapeptide measured with ECD in water and methanol.

Water						
Tripeptide	**α-helix**	**3_{10} helix**	**β-sheet**	**β-turn**	**PPII**	**unordered**
ALA	1%	5%	11%	17%	19%	48%
DIC	7%	3%	21%	15%	10%	44%
EAM	1%	4%	17%	16%	16%	46%
EKF	<1%	4%	24%	13%	15%	43%
KAM	<1%	2%	17%	14%	19%	48%
VIV	18%	5%	7%	16%	11%	44%
CATWEAMEKCK	22%	6%	8%	16%	9%	39%
Methanol						
Tripeptide	**α-helix**	**3_{10} helix**	**β-sheet**	**β-turn**	**PPII**	**unordered**
ALA	<1%	3%	17%	18%	13%	49%
DIC	11%	4%	15%	16%	11%	44%
EAM	1%	3%	20%	15%	15%	46%
EKF	<1%	5%	23%	15%	14%	43%
KAM	1%	2%	19%	15%	16%	47%
VIV	5%	3%	21%	16%	11%	45%
CATWEAMEKCK	55%	32%	1%	10%	2%	16%

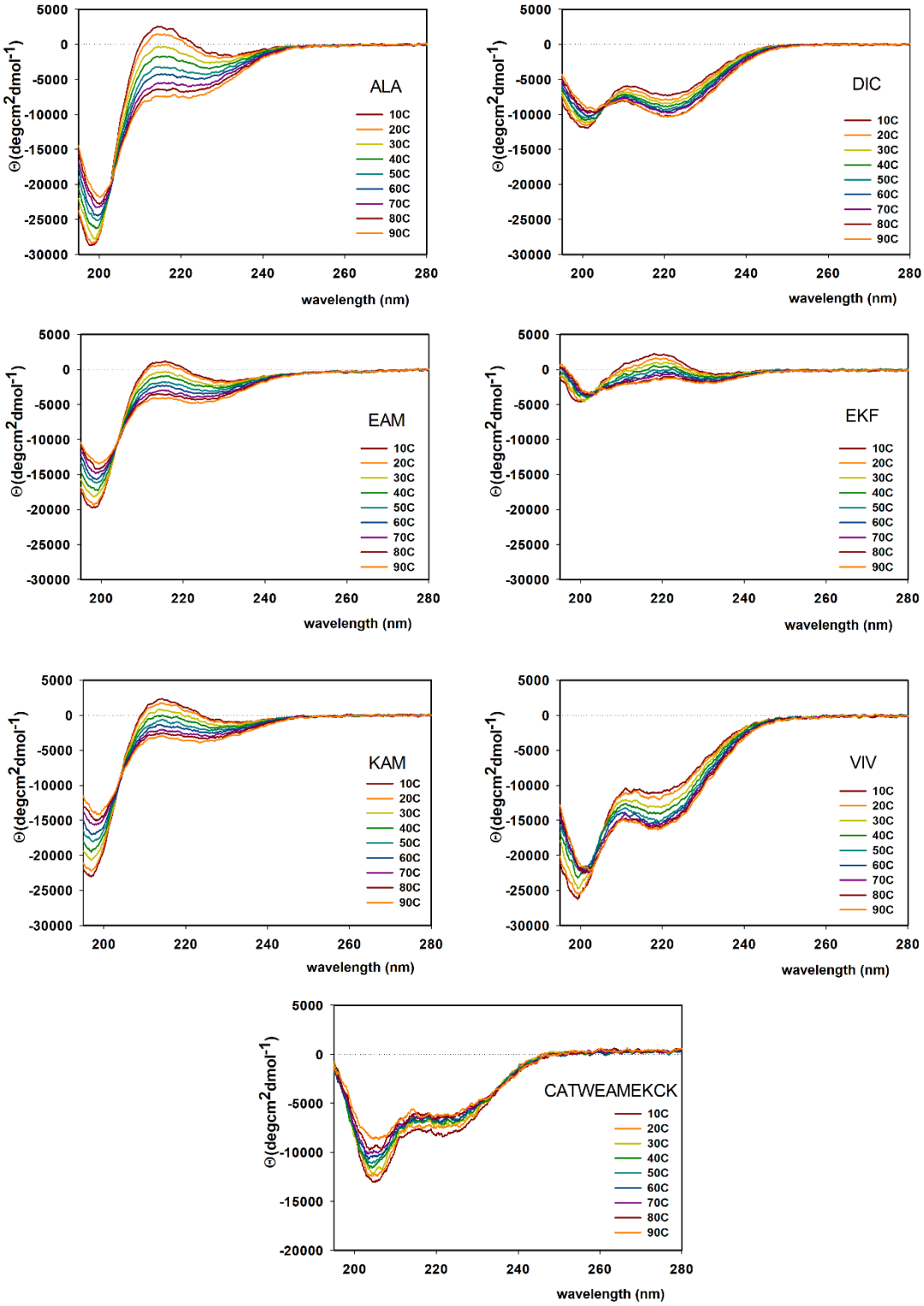


Fig. S5 Temperature dependence of ECD spectra of six tripeptides and undecapeptide with water as a solvent.

6. NMR Spectroscopy

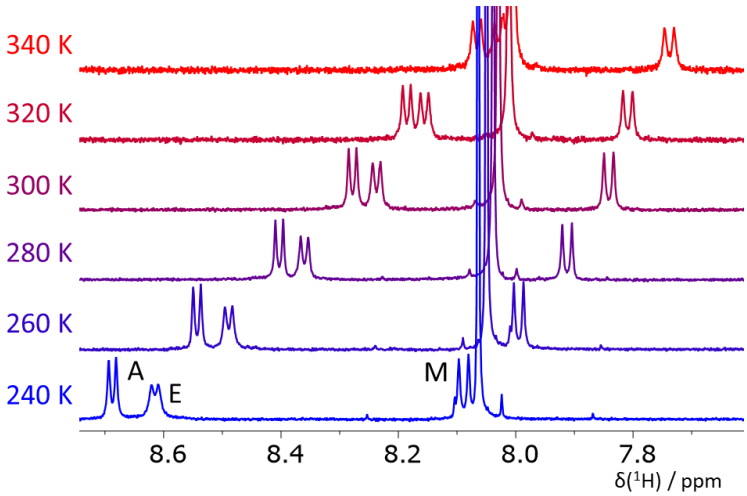


Fig. S6 The NH region of ^{1}H NMR spectra of the tripeptide EAM in DMF at T = 240–340 K.

Table S2. Experimentally determined $^{3}J_{NH,H\alpha}$ coupling values (Hz) in the tripeptide ALA in DMF and methanol.

				DMF					
T / K	360	340	320	300	280	260	240	220	
A1	[a]	6.5	6.4	6.1	6.1	6	5.8	5.8	
L	7.6	7.8	7.8	7.7	7.7	7.7	7.6	7.6	
A3	7.6	7.6	7.6	7.4	7.5	7.5	[a]	7.3	
							methanol		
T / K				300	280	260	240	220	200
A1				[a]	5.8	5.5	5.3	5.1	5.0
L				[a]	7.4	7.4	7.2	7.2	7.1
A3				[a]	7.0	6.8	6.7	6.3	6.1

[a]Not determined because of a signal overlap, signal broadening or fast chemical exchange process

Table S3. Experimentally determined $^{3}J_{NH,H\alpha}$ coupling values (Hz) in the tripeptide EAM in DMF and methanol.

				DMF					
T / K	360	340	320	300	280	260	240	220	
E	6.7	7.1	6.7	6.5	6.3	6.2	6.0	[a]	
A	6.9	6.9	6.7	6.6	6.6	6.4	6.3	6.1	
M	8.0	8.2	8.1	8.1	8.1	8.2	8.2	8.0	
						Methanol			
T / K			320	300	280	260	240	220	200
E			[a]	6.6	6.3	6.3	6.1	6.0	5.7
A			[a]	6.2	6.1	6.0	5.7	5.5	5.2
M			7.9	7.9	7.9	7.8	7.7	7.6	7.5

[a]Not determined because of a signal overlap, signal broadening or fast chemical exchange process

Table S4. Experimentally determined $^{3}J_{NH,H\alpha}$ coupling values (Hz) in the tripeptide DIC in water and methanol.

				methanol			
T / K	320	300	280	260	240	220	200
D	[a]	8.0	8.1	8.1	8.1	8.0	8.0
I	[a]	7.1	7.0	6.7	6.6	6.7	6.3
C	7.5	7.4	7.3	7.1	7.0	6.9	6.7
				H_2O-D_2O			
T / K	320	300	280	260	240	220	200
D		[a]	7.6				
I		7.6	7.5				
C		[a]	6.9				

[a]Not determined because of a signal overlap, signal broadening or fast chemical exchange process

Table S5. Experimentally determined $^{3}J_{NH,H\alpha}$ coupling values (Hz) in the tripeptide EKF in water and methanol.

				methanol			
T / K	320	300	280	260	240	220	200
E	6.6	6.4	6.3	6.1	5.8	5.6	[a]
K	7.6	7.5	7.4	7.4	7.2	7.1	[a]
F	7.9	8.0	7.9	7.9	7.8	7.8	7.5
				H_2O-D_2O			
T / K	320	300	280	260	240	220	200
E		6.3	6.2				
K		7.2	7.1				
F		7.7	7.6				

[a]Not determined because of a signal overlap, signal broadening or fast chemical exchange process

Table S6. Experimentally determined $^{3}J_{NH,H\alpha}$ coupling values (Hz) in the tripeptide KAM in DMF and methanol.

				DMF				
T / K	360	340	320	300	280	260	240	220
K	7.3	7.0	6.8	6.9	6.7	6.6	6.5	6.3
A	70.	6.9	6.9	6.8	6.7	6.6	6.5	6.3
M	8.1	8.1	8.2	8.2	8.1	8.2	8.0	8.0
				methanol				
T / K				300	280	260	240	220	200
K				7.0	6.9	6.7	6.5	[a]	6.2

A	6.2	6.1	5.9	5.7	5.5	5.2
M	7.8	7.7	7.7	7.6	[a]	7.2

[a]Not determined because of a signal overlap, signal broadening or fast chemical exchange process

Table S7. Experimentally determined $^{3}J_{NH,H\alpha}$ coupling values (Hz) in the tripeptide VIV in methanol and DMF. This tripeptide is very poorly soluble in DMF.

			methanol			
T / K	300	280	260	240	220	200
V1[b]	7.9	[a]	[a]	7.7	7.4	6.9
I	8.6	[a]	[a]	8.6	8.5	[a]
V3[b]	8.7	8.5	8.3	8.3	8.1	8.0
			DMF			
T / K	300	280	260	240	220	200
V1	8.6	[a]	[a]	[a]	[a]	[a]
I	8.7	[a]	[a]	[a]	[a]	[a]
V3	8.8	[a]	[a]	[a]	[a]	[a]

[a]Not determined because of a signal overlap, signal broadening or fast chemical exchange process

[b]The assignment of V1 and V3 in methanol may be interchanged

Table S8. Experimentally determined $^{3}J_{NH,H\alpha}$ coupling values (Hz) in the tripeptide IYI in DMF and methanol.

				DMF				
T / K	360	340	320	300	280	260	240	220
I1	8.0	7.9	7.8	7.9	[a]	7.7	7.7	7.6
Y	8.0	8.1	8.1	[a]	8.1	8.2	8.1	8.3
I3	8.7	8.8	8.8	9.0	9.1	9.2	9.0	[a]
					methanol			
T / K			320	300	280	260	240	220
I1[b]			8.1	8.0	8.1	[a]	7.7	[a]
Y			7.6	7.7	8.1	[a]	7.7	7.8
I3[b]			8.7	8.7	[a]	[a]	8.7	[a]

[a]Not determined because of a signal overlap, signal broadening or fast chemical exchange process

[b]The assignment of I1 and I3 in methanol may be interchanged

Table S9. Experimental (at T = 300 and 200 K) and calculated[a] $^3J_{NH,H\alpha}$ coupling values (Hz) in the studied tripeptides. Only values for the central amino acid are shown.

	Exp 300 K	Exp 200 K	Calc α-helix	Calc extended	Calc PPII
ALA		7.1	4.3	9.6	5.7
DIC	7.1	6.3	3.2	6.8[b]	9.9[b]
EAM	6.2	5.2	3.4	9.6	6.4
EKF	7.5		3.8	9.5	6.3
KAM	6.2	5.2	3.9	9.7	4.9
VIV	8.6		5.4	9.3	5.7
IYI	7.7		5.0	9.8	6.8

[a]The structures of proposed ideal tripeptide conformers were optimized at DFT level (B3LYP/cc-pvdz/gd3/pcm-methanol) using Gaussian 16 software. The structures were partially optimized with the dihedral angles of the tripeptide backbone freezed in values $\varphi = -139°$, $\psi = 135°$, $\omega = 180°$ for extended structures, $\varphi = -58°$, $\psi = -47°$, $\omega = 180°$ for α-helices and $\varphi = -75°$, $\psi = 145°$, $\omega = 180°$ for polyproline II conformations. The $^3J_{NH,H\alpha}$ coupling values were calculated for the optimized structures at the B3LYP/6-31+G*/GD3/PCM level.

[b]The unusual coupling values for this tripeptide are caused by the formation of an intramolecular hydrogen bond of the NH group with the aspartic acid carboxylate leading to substantial changes of the H–N–Cα–H torsion angle.

7. NMR experiments: CATWEAMEKCK undecapeptide

An insight into the conformation of the undecapeptide was obtained from the values of the $^3J_{NH,H\alpha}$ coupling values, temperature dependence of the chemical shifts of the amido protons, and chemical shifts of the alpha protons. The amino acid sequence of the undecapeptide was independently confirmed by a combination of ^{1}H, ^{13}C, ^{1}H,^{13}C-HSQC, ^{1}H,^{13}C-HMBC, ^{1}H,^{1}H-TOCSY and ^{1}H,^{1}H-ROESY experiments (**Fig. S7**).

All obtained $^3J_{NH,H\alpha}$ coupling values are smaller than or close to 5 Hz at 300 K except for residue C10 (note that the coupling was not determined in residues T3 and K11 because of a signal overlap). The smallest coupling values were observed for the amino acid residues in the middle of the sequence (W4–A6). The coupling values further decrease at low temperature (**Table 2** in the main text). These observations thus indicate that the undecapeptide tends to form a helical structure.

Another NMR observable sensitive to the formation of a secondary structure is the chemical shift of hydrogen atoms in position alpha. A simple and straightforward methodology of chemical shift index (CSI) is widely used in structural biology studies. The chemical shift of

each amino acid residue is compared to that obtained for random-coil peptides and sites with chemical shifts lower by more than 0.1 ppm indicate helical structures while sites with chemical shifts higher by more than 0.1 ppm indicate a β sheet. Although the random-coil chemical shifts were obtained for peptides dissolved in water and are not transferrable to the results obtained for the methanol solution, we can compare the chemical shifts in the undecapeptide with those obtained for the EAM tripeptide in the same solution. The chemical shifts of protons H alpha of the EAM fragment in the undecapeptide are 0.25–0.35 ppm lower than in the EAM tripeptide. This significant decrease of the chemical shifts thus confirms that the central part of the undecapeptide is in a helical conformation.

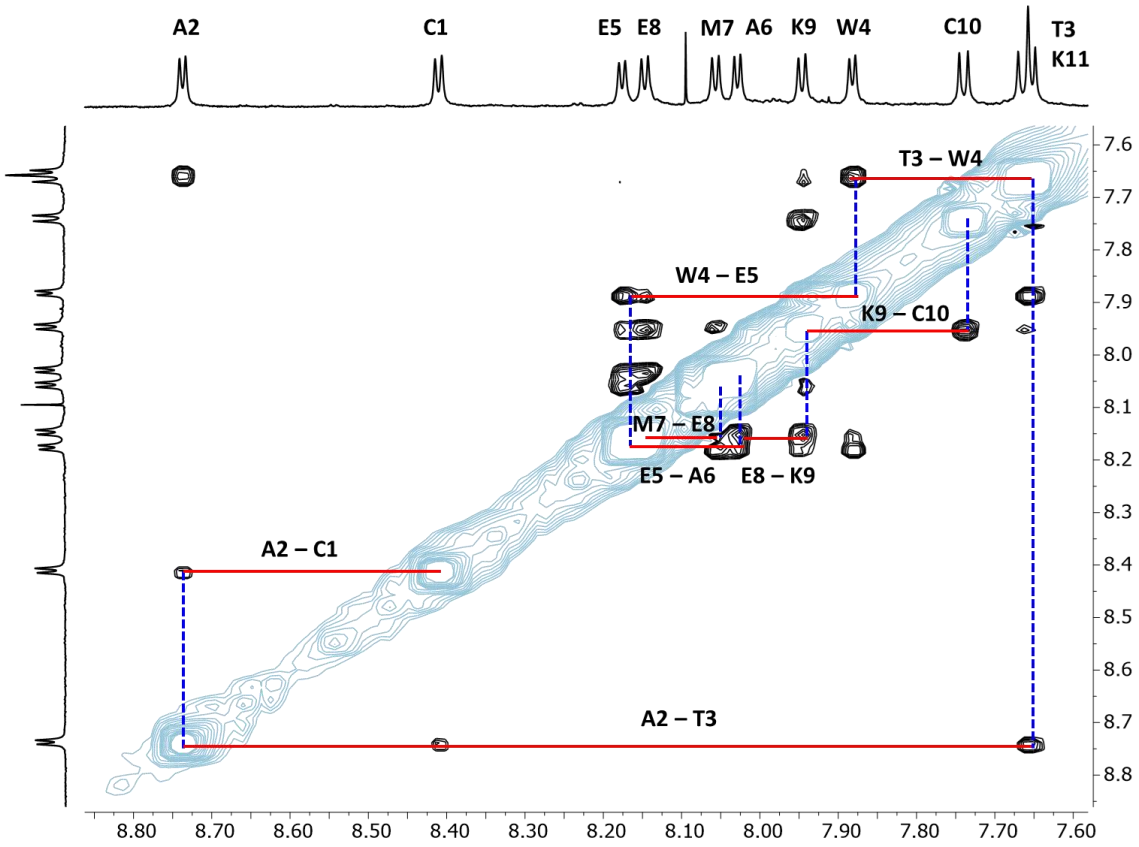


Fig. S7 The amido-hydrogen part of ^{1}H,^{1}H-ROESY spectra of Ac-CATWEAMEKCK in CD_3OH at 300 K with the backbone sequence assignment. The signals belonging to hydrogen atoms within individual residues were assigned using ^{1}H,^{1}H-TOCSY spectra and the sequence assignment was confirmed by the ^{1}H,^{1}H-ROESY data, where cross-peaks between neighboring backbone amido protons (dNN(i,i+1)) can be observed.

Additional insight into the secondary structure formation can be obtained from the temperature dependence of chemical shifts of amido protons. Generally, the chemical shifts of amido protons decrease with increasing temperatures due to chemical exchange with water protons. However, the temperature dependence of the chemical shifts is less pronounced when a secondary structure with intramolecular hydrogen bonds is formed. Typical values of the temperature dependence of random-coil structures are known for peptides in water solution and these values can significantly differ for those in methanol solution. Therefore, we compare the temperature coefficients of the amido protons in the undecapeptide with those obtained for the tripeptide EAM in the same solvent (**Table 2** in the main text, **Figures S8** and **S9**).

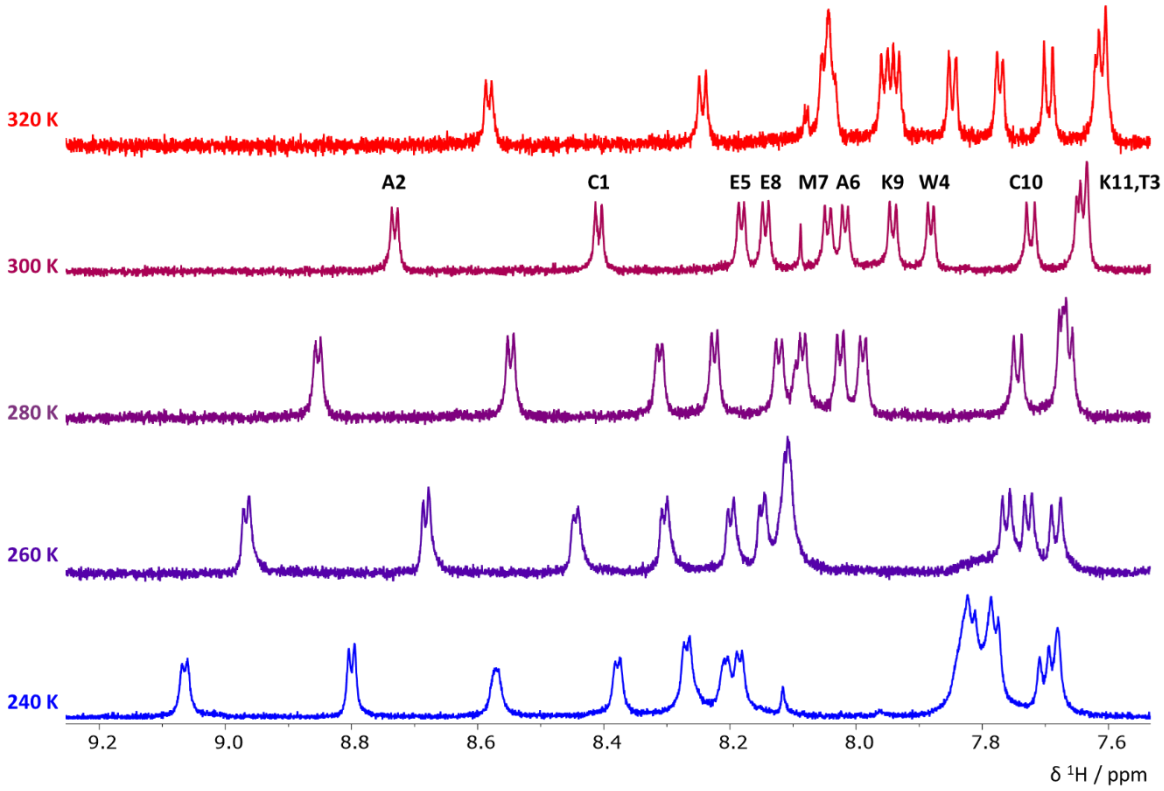


Fig. S8 The NH region of variable-temperature ^{1}H NMR spectra of Ac-CATWEAMEKCK in CD_3OH. The spectra were recorded on a 500 MHz spectrometer and the chemical shifts were referenced to $CD_2\mathit{H}OH$.

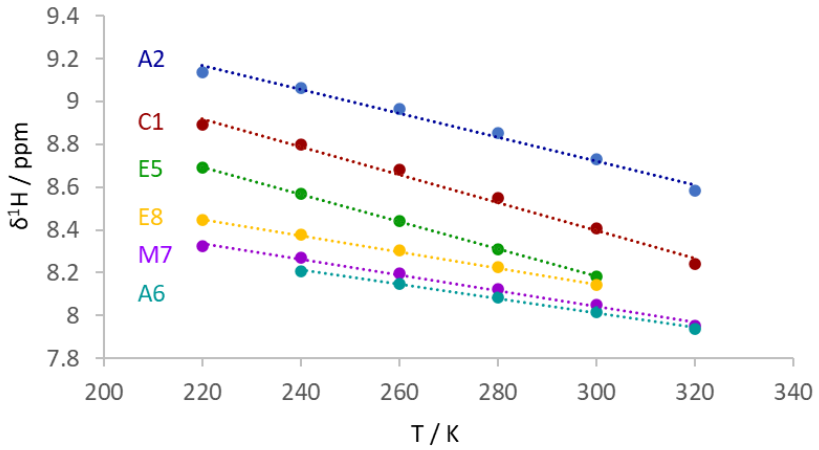


Fig. S9 The temperature dependence of chemical shifts of selected amido protons of Ac-CATWEAMEKCK in CD_3OH.

8. Effect of solvent on conformational energies

To quantify the effect of solvent on peptide secondary structure, we decided to carry on a conformational sampling in three common solvents: water (as more polar environment on the surface of the protein), methanol (less polar, more resembling the environment inside the protein and also the solvent used in VCD and NMR experiments), and *N*,*N*-dimethylformamide (DMF, the solvent also used in NMR experiments). We choose two tripeptides: EAM (present in α-helices in 80 % in proteins), and IYI (present in 69 % of extended structures in proteins).

Fig. S10 compares the energetic distribution of helical and extended conformers for pro-helical EAM and pro-extended IYI in water and in two solvents used in herein-presented VCD and NMR experiments, methanol and DMF. For the more pro-helical EAM, the helical and extended conformers occupy roughly the same energetic window. On the other hand, for the more pro-extended IYI tripeptide, the extended conformers are on average significantly lower in energy than the helical conformers. The results are very robust and differ only marginally among different solvents. Histograms in **Fig. S10** imply that helical conformers should be energetically accessible for EAM at room temperature and even lower temperatures, while they are expected to be less populated in case of the IYI tripeptide, and this effect is independent of solvent (at least out of three solvents we worked with).

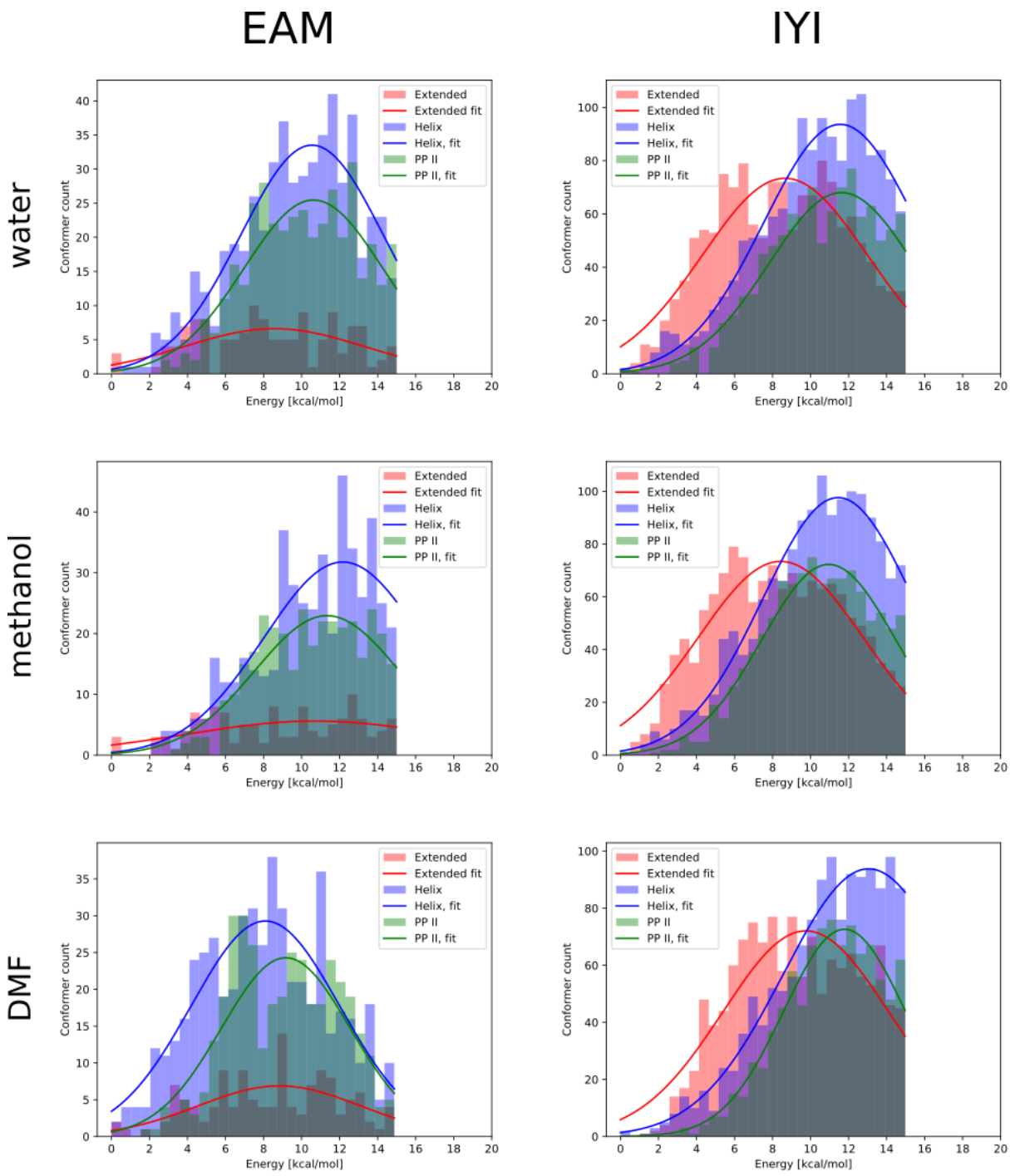
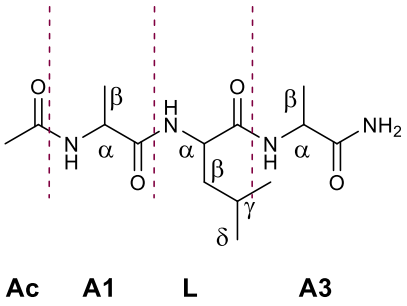


Fig. S10 Histograms of conformer energies for conformers of EAM and IYI tripeptides in three different solvents. The secondary structure type of the central residue is depicted.

9. NMR – characterization spectra

ALA



^{1}H NMR (600 MHz, d_6-DMSO): δ = 0.83 (d, $^{3}J_{H\delta,H\gamma}$ = 6.5 Hz, 3H, L-δ_b), 0.88 (d, $^{3}J_{H\delta,H\gamma}$ = 6.6 Hz, 3H, L-δ_a), 1.18 (d, $^{3}J_{H\beta,H\alpha}$ = 7.1 Hz, 3H, A1-β), 1.19 (d, $^{3}J_{H\beta,H\alpha}$ = 7.1 Hz, 3H, A3-β), 1.42–1.50 (m, 2H, L-β), 1.59 (m, 1H, L-γ), 1.83 (s, 3H, Ac-C**H**$_3$), 4.15 (p, $^{3}J_{H\alpha,NH}$ = 7.1 Hz, $^{3}J_{H\alpha,H\beta}$ = 7.1 Hz, 1H, A3-α), 4.20–4.25 (m, 2H, A1-α, L-α), 7.00 (bs, 1H, N**H**$_2$), 7.22 (bs, 1H, N**H**$_2$), 7.71 (d, $^{3}J_{NH,H\alpha}$ = 7.4 Hz, 1H, A3-N**H**), 7.94 (d, $^{3}J_{NH,H\alpha}$ = 8.1 Hz, 1H, L-N**H**), 8.06 (d, $^{3}J_{NH,H\alpha}$ = 7.1 Hz, 1H, A1-N**H**).

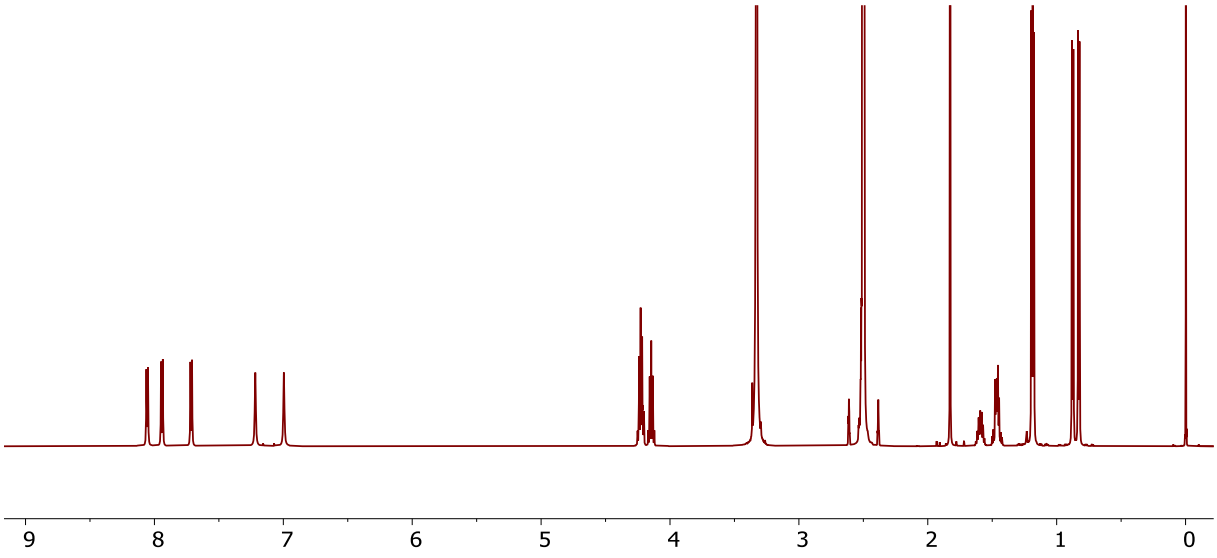


Fig. S11 ^{1}H NMR spectrum of tripeptide ALA (600 MHz, d_6-DMSO).

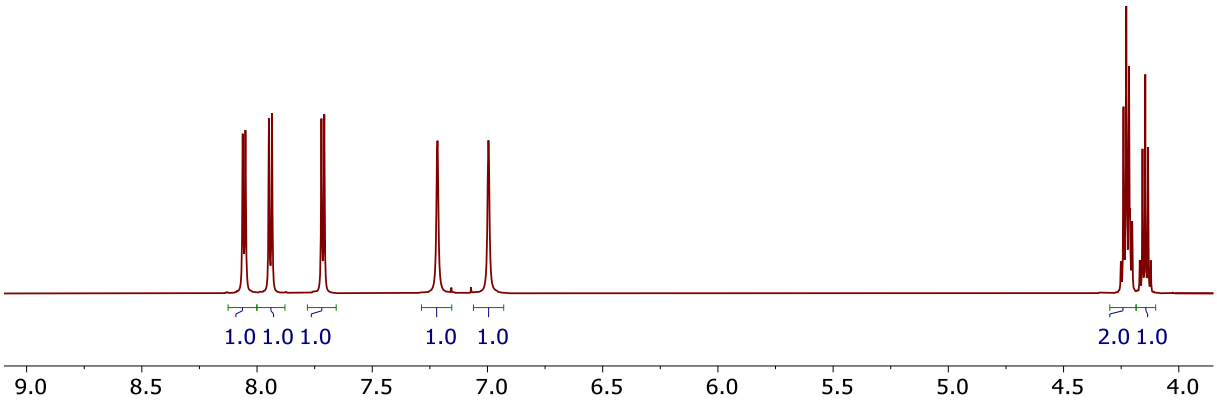
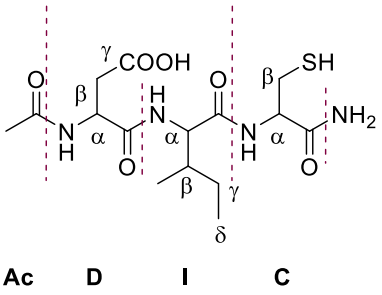


Fig. S12 Hα and NH region of ^{1}H NMR (600 MHz, d_6-DMSO) of tripeptide ALA.

DIC



^{1}H NMR (500 MHz, d_4-methanol): δ = 0.93 (t, $^{3}J_{H\delta,H\gamma}$ = 7.4 Hz, 3H, I-δ), 0.98 (d, $^{3}J_{H\beta}$ = 6.9 Hz, 3H, I-β-CH_3), 1.26 (m, 1H, I-γ_b), 1.57 (m, 1H, I-γ_a), 1.95 (m, 1H, I-β), 1.97 (s, 3H, Ac-CH_3), 2.68 (dd, ^{2}J = 16.8 Hz, ^{3}J = 6.4 Hz, 1H, C-β_b), 2.83 (dd, 1H, D-β_b), 2.88 (dd, 1H, C-β_a), 2.97 (dd, ^{2}J = 13.9 Hz, ^{3}J = 4.9 Hz, 1H, D-β_a), 4.20 (d, $^{3}J_{H\alpha,H\beta}$ = 6.1 Hz, 1H, I-α), 4.42 (dd, ^{3}J = 8.6 and 4.9 Hz, D-α), 4.76 (dd, ^{3}J = 7.8 and 6.4 Hz, C-α),
^{1}H NMR (500 MHz, d_3-methanol): δ = 7.14 (bs, 1H, N**H**$_2$), 7.35 (bs, 1H, N**H**$_2$), 8.07 (d, $^{3}J_{NH,H\alpha}$ = 8.0 Hz, 1H, D-N**H**), 8.17 (d, $^{3}J_{NH,H\alpha}$ = 7.1 Hz, 1H, I-N**H**), 8.42 (d, $^{3}J_{NH,H\alpha}$ = 7.4 Hz, 1H, C-N**H**).

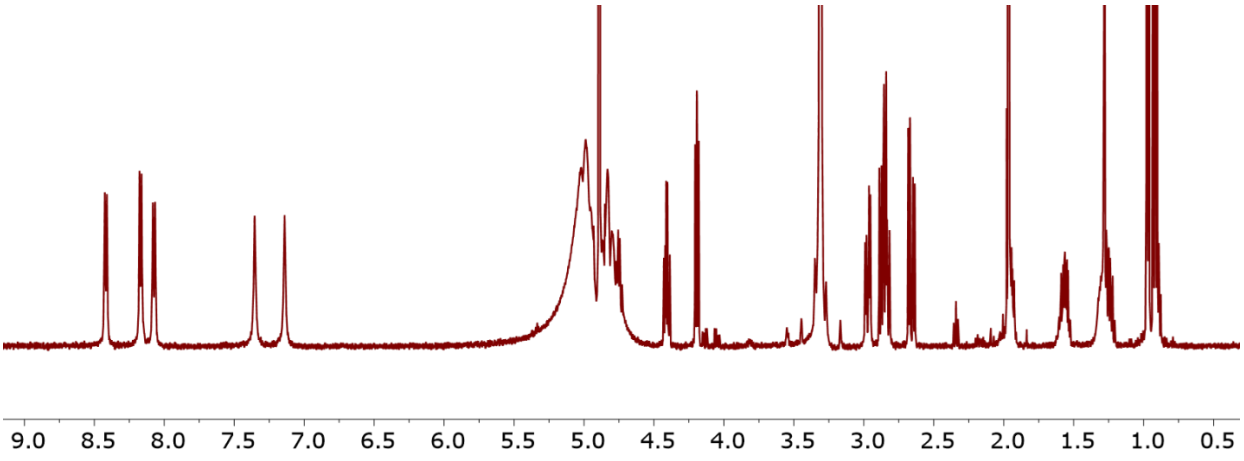


Fig. S13 ^{1}H NMR spectrum of tripeptide DIC (500 MHz, d_3-methanol).

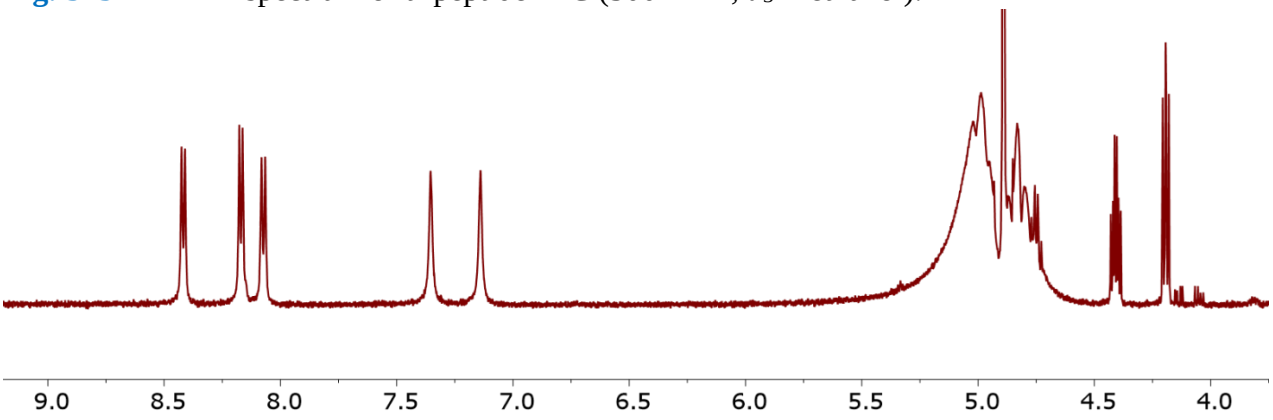
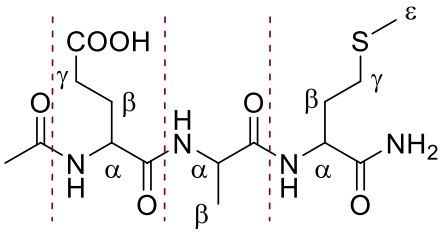


Fig. S14 Hα and NH region of ^{1}H NMR (500 MHz, d_3-methanol) of tripeptide DIC.

EAM



^{1}H NMR (600 MHz, d_6-DMSO): δ = 1.21 (d, $^3J_{H\beta,H\alpha}$ = 7.1 Hz, 3H, A-β), 1.71 (m, 1H, E-β_b), 1.78 (m, 1H, M-β_b), 1.84 (s, 3H, Ac-C**H**$_3$), 1.85 (m, 1H, E-β_a), 1.92 (m, 1H, M-β_a), 2.03 (s, 3H, S-C**H**$_3$), 2.24 (m, 2H, E-γ), 2.38 (ddd, ^2J = 13.1 Hz, ^3J = 3.7 and 6.4 Hz, 1H, M-γ_b), 2.45 (ddd, ^2J = 13.1 Hz, ^3J = 10.0 and 4.9 Hz, 1H, M-γ_a), 4.18–4.24 (m, 3H, E-α, M-α, A-α), 7.06 (bs, 1H, N**H**$_2$), 7.73 (bs, 1H, N**H**$_2$), 7.83 (d, $^3J_{NH,H\alpha}$ = 8.1 Hz, 1H, M-N**H**), 8.07 (d, $^3J_{NH,H\alpha}$ = 7.7 Hz, 1H, E-N**H**), 8.09 (d, $^3J_{NH,H\alpha}$ = 6.9 Hz, 1H, A-N**H**), 12.08 (bs, 1H, E-COO**H**).

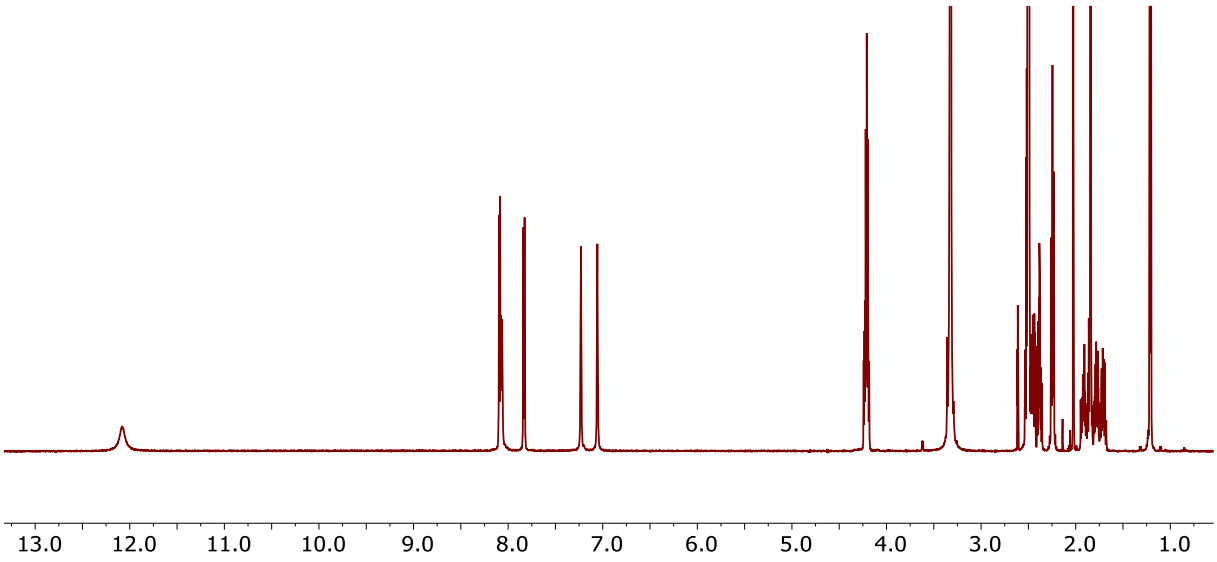


Fig. S15 ^{1}H NMR spectrum of tripeptide EAM (600 MHz, d_6-DMSO).

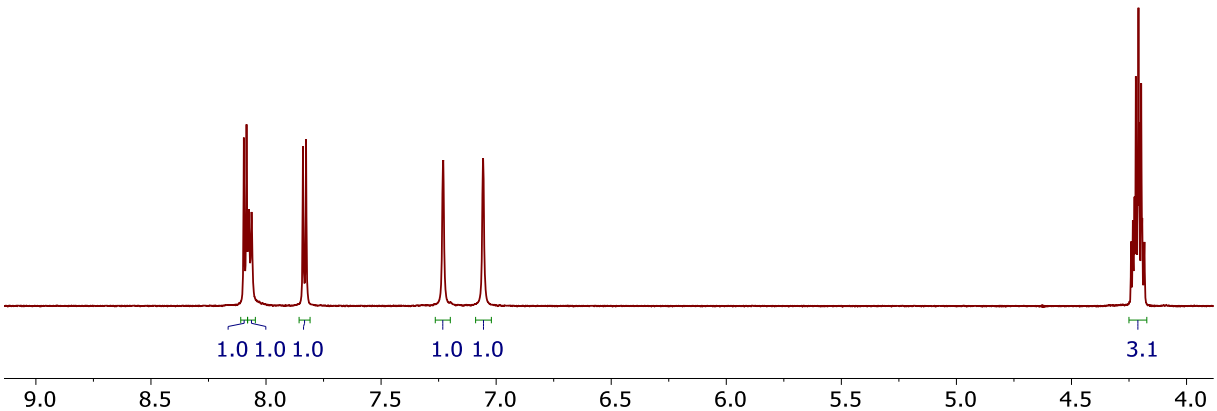
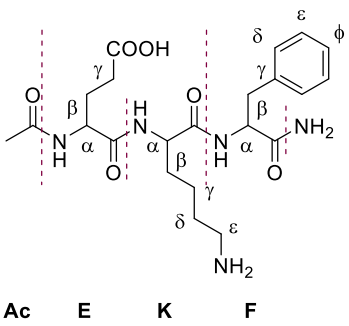


Fig. S16 Hα and NH region of ^{1}H NMR (600 MHz, d_6-DMSO) of tripeptide EAM.

EKF



^{1}H NMR (500 MHz, d_4-methanol): δ = 1.37 (m, 2H, K-γ), 1.61 (m, 3H, K-β$_b$, K-δ), 1.74 (m, 1H, K-β$_a$), 1.90 (m, 1H, E-β$_b$), 1.99 (m, 1H, E-β$_a$), 1.99 (s, 3H, Ac-CH$_3$), 2.37 (m, 2H, E-γ), 2.88 (t, $^{3}J_{Hε,Hδ}$ = 7.4 Hz, 2H, K-ε), 2.96 (dd, ^{2}J = 14.0 Hz, ^{3}J = 9.1 Hz, 1H, F-β$_b$), 3.16 (m, 1H, F-β$_a$), 4.22 (d, $^{3}J_{Hα,Hβ}$ = 8.4 and 5.9 Hz, 1H, E-α), 4.29 (dd, ^{3}J = 8.9 and 5.3 Hz, K-α), 4.60 (dd, ^{3}J = 9.1 and 5.6 Hz, F-α), 7.20 (m, 1H, F-φ), 7.24–7.29 (m, 4H, F-δ, F-ε),
^{1}H NMR (500 MHz, d_3-methanol): δ = 7.07 (bs, 1H, N**H**$_2$), 7.58 (bs, 1H, N**H**$_2$), 8.03 (d, $^{3}J_{NH,Hα}$ = 8.0 Hz, 1H, F-N**H**), 8.17 (d, $^{3}J_{NH,Hα}$ = 7.5 Hz, 1H, K-N**H**), 8.31 (d, $^{3}J_{NH,Hα}$ = 6.4 Hz, 1H, E-N**H**).

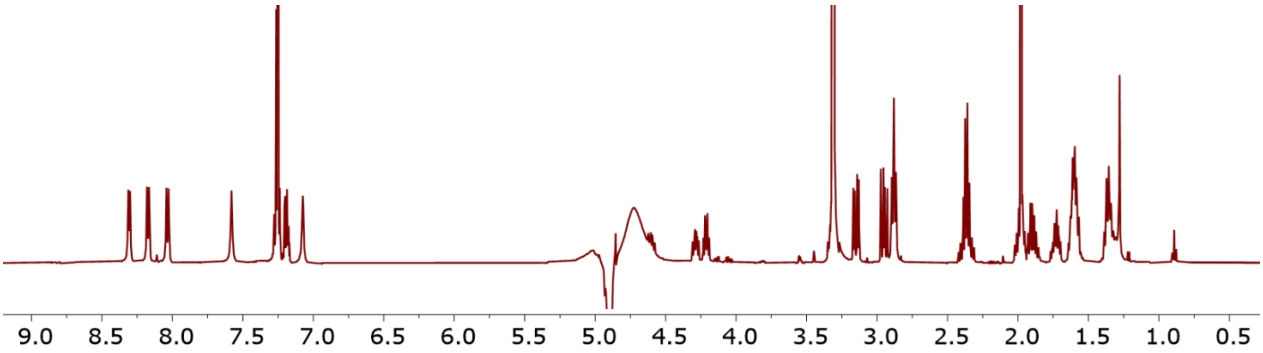


Fig. S17 ^{1}H NMR spectrum of tripeptide EKF (500 MHz, d_3-methanol).

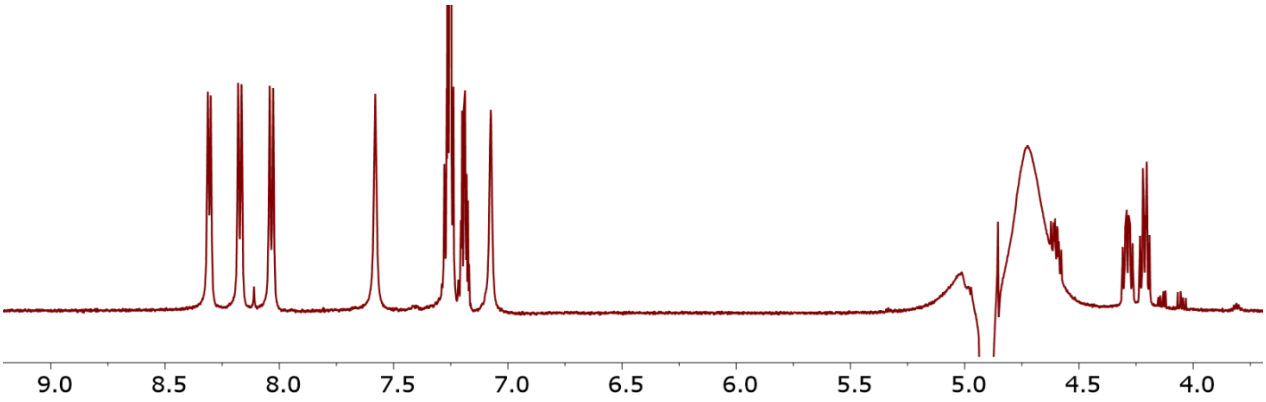
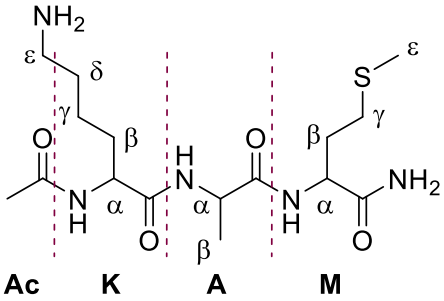


Fig. S18 Hα and NH region of ^{1}H NMR (600 MHz, d_3-methanol) of tripeptide EKF.

KAM



^{1}H NMR (600 MHz, d_6-DMSO): δ = 1.22 (d, $^{3}J_{H\beta,H\alpha}$ = 7.1 Hz, 3H, A-β), 1.25–1.35 (m, 2H, K-γ), 1.45–1.54 (m, 3H, K-β_b, K-δ), 1.62 (m, 1H, K-β_a), 1.77 (m, 1H, M-β_b), 1.85 (s, 3H, Ac-C**H**$_3$), 1.99 (m, 1H, M-β_a), 2.03 (s, 3H, S-C**H**$_3$), 2.36–2.46 (m, 2H, M-γ), 2.73–2.78 (m, 2H, K-ε), 4.19–4.25 (m, 3H, A-α, K-α, M-α), 7.08 (bs, 1H, N**H**$_2$), 7.27 (bs, 1H, N**H**$_2$), 7.62 (bs, 3H, K-ε-N**H**$_3^+$), 7.80 (d, $^{3}J_{NH,H\alpha}$ = 8.1 Hz, 1H, M-N**H**), 8.04 (d, $^{3}J_{NH,H\alpha}$ = 7.8 Hz, 1H, K-N**H**), 8.11 (d, $^{3}J_{NH,H\alpha}$ = 7.0 Hz, 1H, A-N**H**).

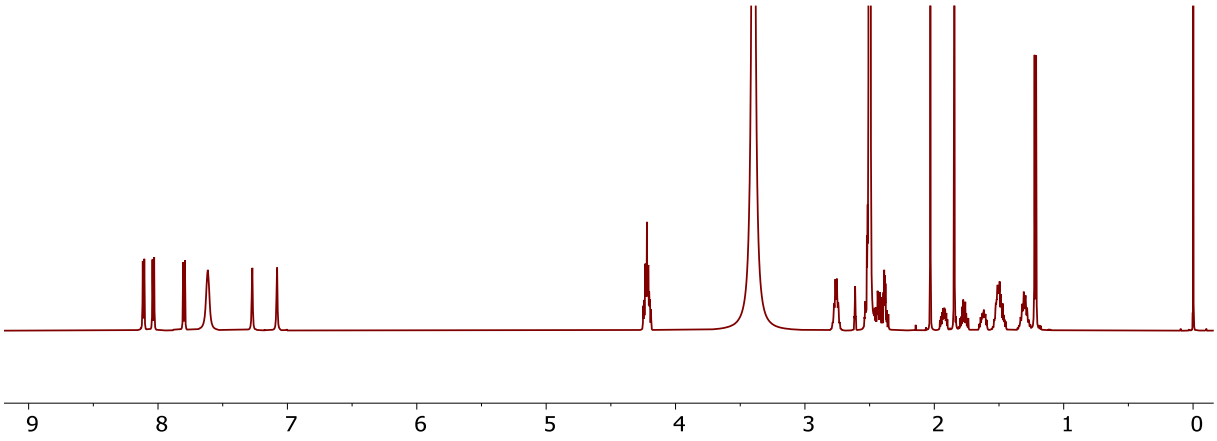


Fig. S19 ^{1}H NMR spectrum of tripeptide KAM (600 MHz, d_6-DMSO).

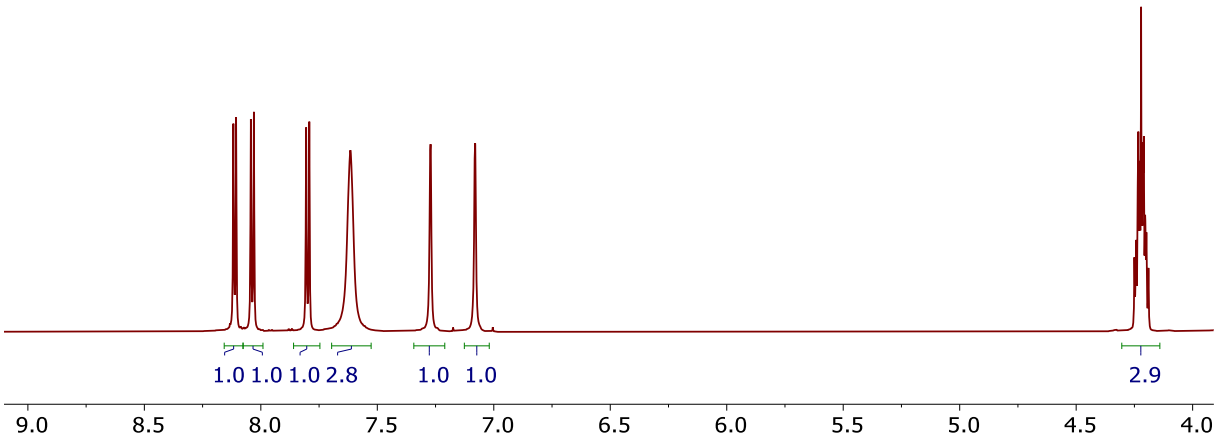
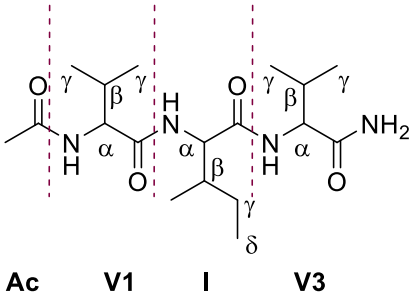


Fig. S20 Hα and NH region of ^{1}H NMR (600 MHz, d_6-DMSO) of tripeptide KAM.

VIV



^{1}H NMR (600 MHz, d_7-DMF): δ = 0.84 (t, $^{3}J_{H\delta,H\gamma}$ = 7.4 Hz, 3H, I-δ), 0.89–0.93 (m, 15H, V1-γ, V3-γ, I-C**H**$_3$), 1.17 (m, 1H, I-γ$_b$), 1.53 (m, 1H, I-γ$_a$), 1.88 (m, 1H, I-β), 1.96 (s, 3H, Ac-C**H**$_3$), 2.07–2.13 (m, 2H, V1-β, V3-β), 4.30 (dd, $^{3}J_{H\alpha,NH}$ = 8.8 Hz, $^{3}J_{H\alpha,H\beta}$ = 6.0 Hz, 1H, V3-α), 4.33 (dd, $^{3}J_{H\alpha,NH}$ = 8.6 Hz, $^{3}J_{H\alpha,H\beta}$ = 6.5 Hz, 1H, V1-α), 4.36 (dd, $^{3}J_{H\alpha,NH}$ = 8.6 Hz, $^{3}J_{H\alpha,H\beta}$ = 7.2 Hz, 1H, I-α), 7.07 (bs, 1H, N**H**$_2$), 7.40 (bs, 1H, N**H**$_2$), 7.63 (d, $^{3}J_{NH,H\alpha}$ = 8.8 Hz, 1H, V3-N**H**), 7.88 (d, $^{3}J_{NH,H\alpha}$ = 8.7 Hz, 1H, I-N**H**), 7.90 (d, $^{3}J_{NH,H\alpha}$ = 8.6 Hz, 1H, V1-N**H**).

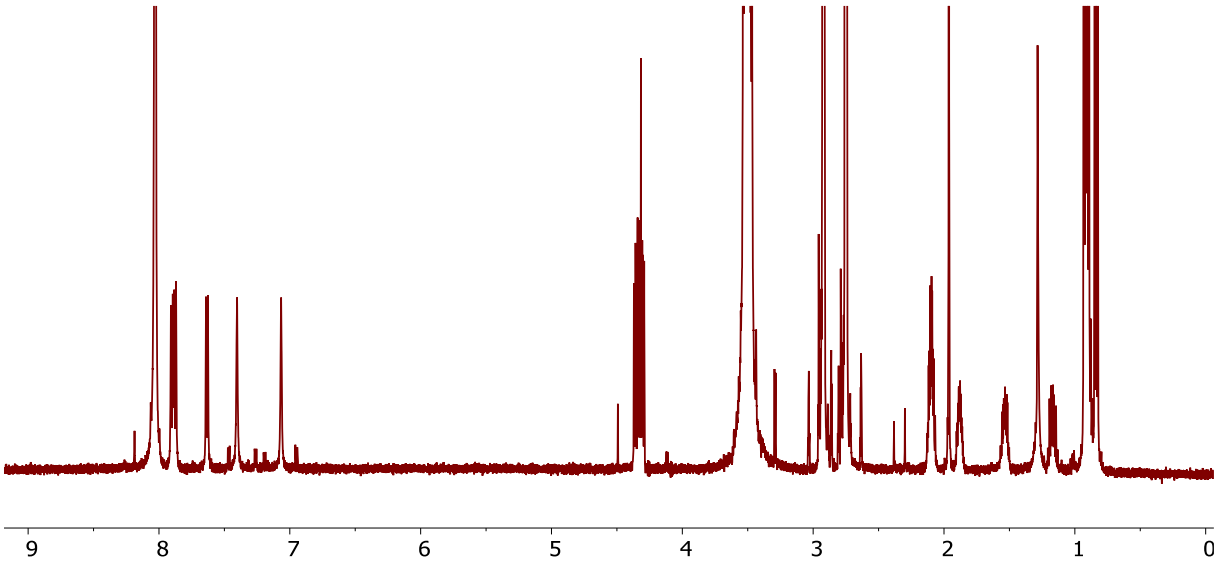


Fig. S21 ^{1}H NMR spectrum of tripeptide VIV (600 MHz, d_7-DMF).

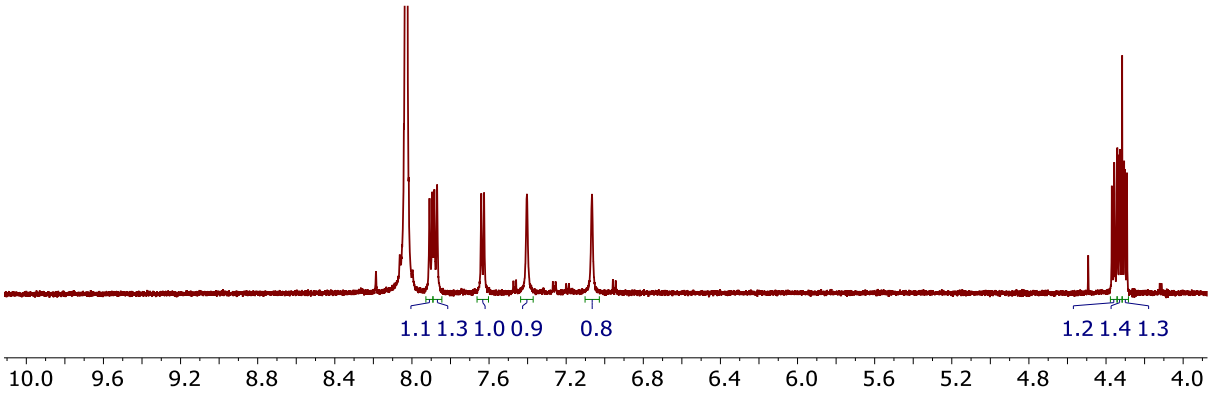
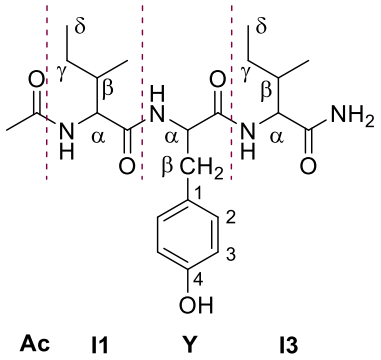


Fig. S22 Hα and NH region of ^{1}H NMR (600 MHz, d_7-DMF) of tripeptide VIV.

IYI



^{1}H NMR (600 MHz, d_7-DMF): δ = 0.79–0.82 (m, 6H, I1-β-C**H**$_3$, I1-δ), 0.85 (t, $^{3}J_{H\delta,H\gamma}$ = 7.4 Hz, 3H, I3-δ), δ = 0.91 (d, ^{3}J = 6.9 Hz, 3H, I3-β-C**H**$_3$), 1.07–1.19 (m, 2H, I1-γ_b, I3-γ_b), 1.39 (m, 1H, I1-γ_a), 1.51 (m, 1H, I3-γ_a), 1.76 (m, 1H, I1-β), 1.85 (m, 1H, I3-β), 1.94 (s, 3H, Ac-C**H**$_3$), 2.85 (m, 1H, Y-β_b), 3.07 (dd, ^{2}J = 14.0 Hz, $^{3}J_{H\beta,H\alpha}$ = 4.7 Hz, 1H, Y-β_a), 4.21 (dd, $^{3}J_{H\alpha,NH}$ = 7.8 Hz, $^{3}J_{H\alpha,H\beta}$ = 6.6 Hz, 1H, I1-α), 4.30 (dd, $^{3}J_{H\alpha,NH}$ = 8.9 Hz, $^{3}J_{H\alpha,H\beta}$ = 6.6 Hz, 1H, I3-α), 4.64 (td, $^{3}J_{H\alpha,NH}$ = 8.7 Hz, $^{3}J_{H\alpha,H\beta}$ = 4.8 Hz, 1H, Y-α), 6.72 (m, 2H, Y-3), 7.07 (bs, 1H, N**H**$_2$), 7.08 (m, 2H, Y-2), 7.29 (bs, 1H, N**H**$_2$), 7.63 (d, $^{3}J_{NH,H\alpha}$ = 8.9 Hz, 1H, I3-N**H**), 7.92 (d, $^{3}J_{NH,H\alpha}$ = 7.9 Hz, 1H, I1-N**H**), 8.03 (d, 1H, Y-N**H**), 9.35 (bs, 1H, Y-O**H**).

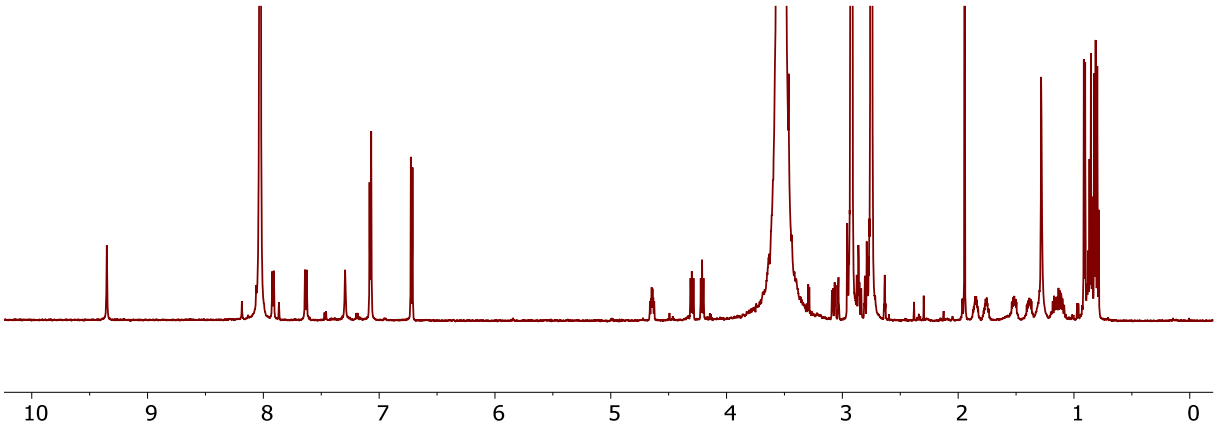


Fig. S23 ^{1}H NMR spectrum of tripeptide IYI (600 MHz, d_7-DMF).

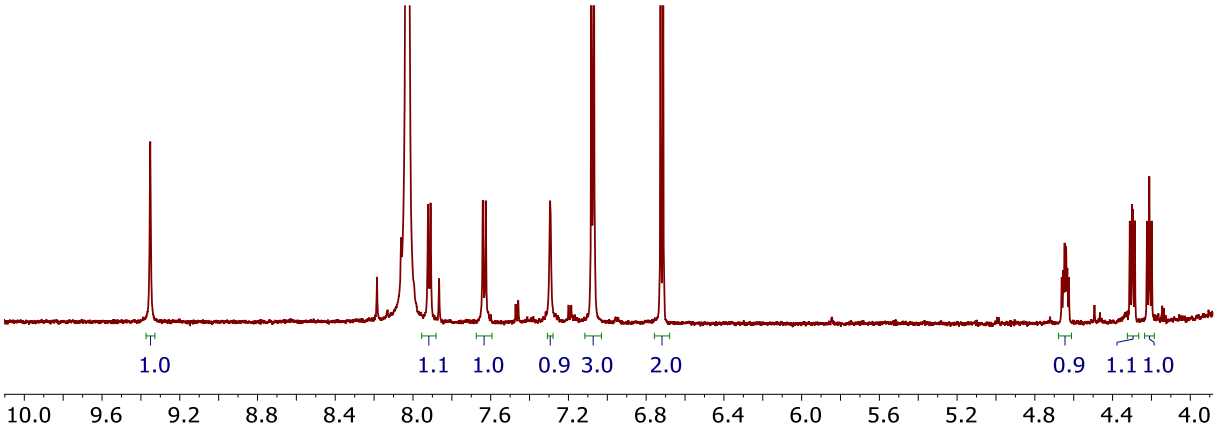
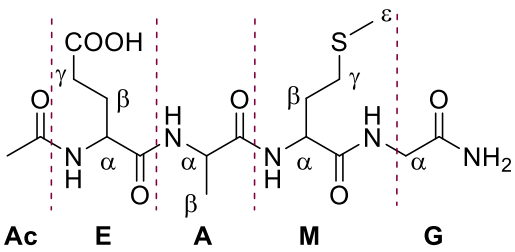
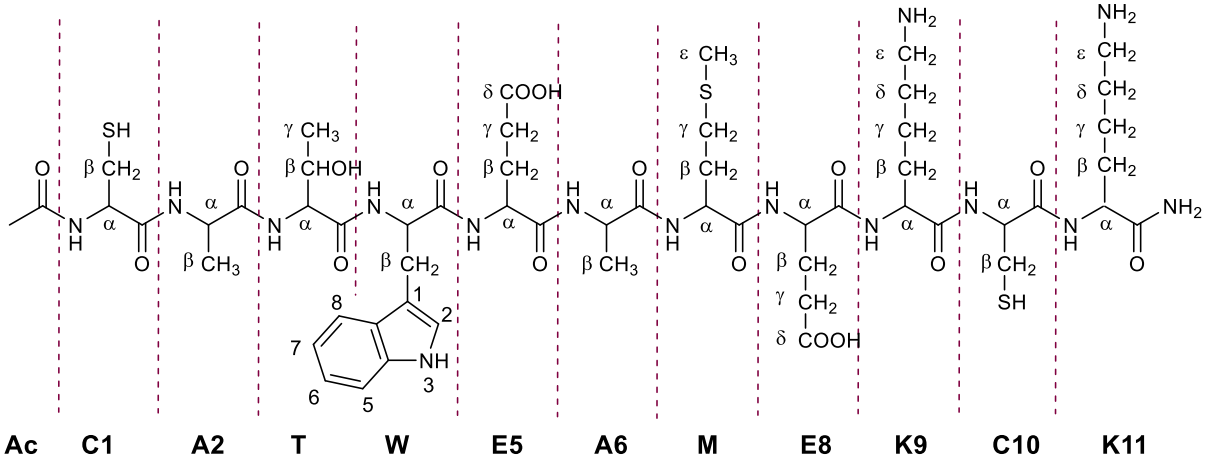


Fig. S24 Hα and aromatic region of ^{1}H NMR (600 MHz, d_7-DMF) of tripeptide IYI.

CATWEAMEKCK





H_α, NH and aromatic region of ^{1}H NMR (600 MHz, d_3-methanol): δ = 3.93 (m, 1H, E5-α), 3.97 (m, 1H, E8-α), 4.00 (m, 1H, T-α), 4.03 (m, 1H, A6-α), 4.08 (m, 1H, K9-α), 4.13 (m, 1H, M-α), 4.17 (m, 1H, T-β), 4.24 (m, 1H, K11-α), 4.27 (m, 1H, A2-α), 4.30 (m, 2H, C1-α, C10-α), 4.39 (m, 1H, W-α), 7.01 (m, 1H, W-7), 7.08 (m, 1H, W-6), 7.16 (m, 3H, W-2, N**H**$_2$), 7.32 (m, 1H, W-5), 7.50 (m, 1H, W-8), 7.63–7.64 (d, $^3J_{NH,H\alpha}$ = 7.8 Hz, 1H; d, $^3J_{NH,H\alpha}$ = 5.1 Hz, 1H, K9-N**H**, T-N**H**), 7.72 (d, $^3J_{NH,H\alpha}$ = 6.7 Hz, 1H, C10-N**H**), 7.87 (d, $^3J_{NH,H\alpha}$ = 4.6 Hz, 1H, W-N**H**), 7.93 (d, $^3J_{NH,H\alpha}$ = 5.4 Hz, 1H, K9-N**H**), 8.01 (d, $^3J_{NH,H\alpha}$ = 4.8 Hz, 1H, A6-N**H**), 8.03 (d, $^3J_{NH,H\alpha}$ = 5.1 Hz, 1H, M-N**H**), 8.13 (d, $^3J_{NH,H\alpha}$ = 5.1 Hz, 1H, E8-N**H**), 8.17 (d, $^3J_{NH,H\alpha}$ = 4.7 Hz, 1H, E5-N**H**), 8.39 (d, $^3J_{NH,H\alpha}$ = 5.2 Hz, 1H, C1-N**H**), 8.72 (d, $^3J_{NH,H\alpha}$ = 4.7 Hz, 1H, A2-N**H**), 10.31 (s, 1H, W-3).

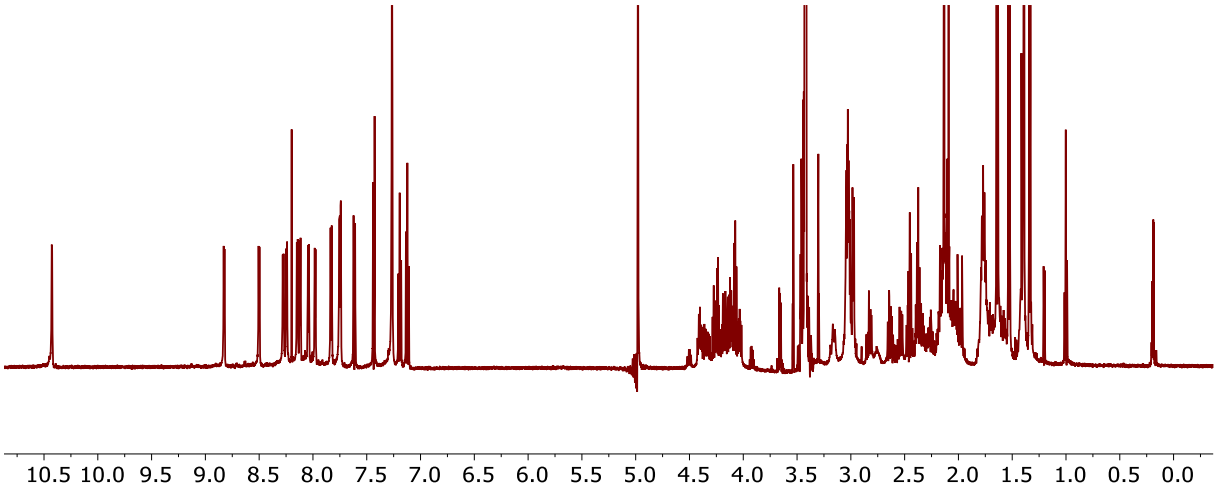


Fig. S25 ^{1}H NMR spectrum of undecapeptide CATWEAMEKCK (600 MHz, d_3-methanol).

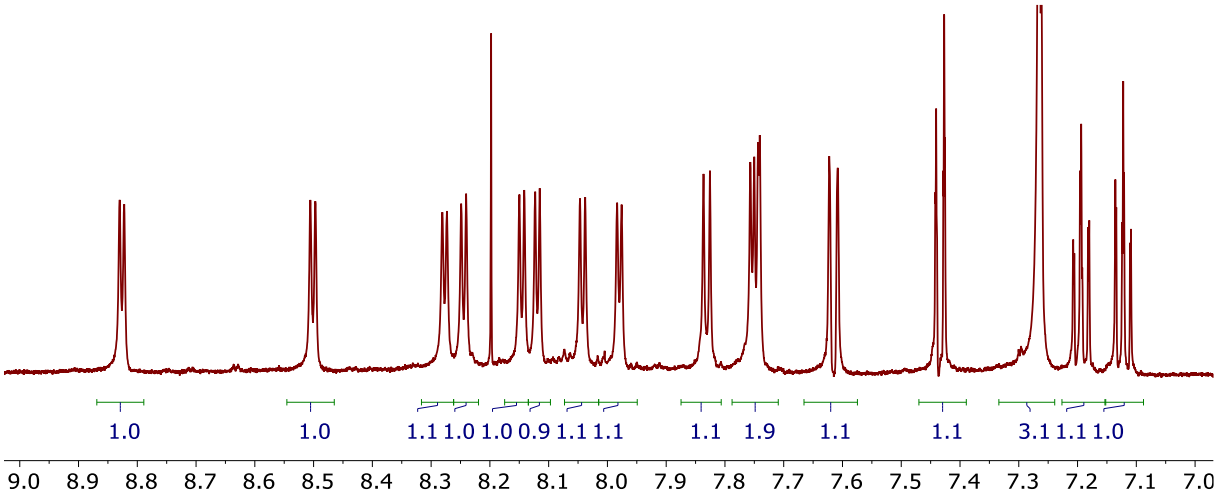


Fig. S26 NH and aromatic region of ^{1}H NMR (600 MHz, d_3-methanol) of undecapeptide CATWEAMEKCK.

References

1 J. Kessler, V. Andrushchenko, J. Kapitán and P. Bouř, *Phys. Chem. Chem. Phys.*, 2018, **20**, 4926–4935.
2 A. F. Drake, G. Siligardi and W. A. Gibbons, *Biophys. Chem.*, 1988, **31**, 143–146.