PNAS
www.pnas.org

Supplementary Information for

How electrostatic networks modulate stability and specificity of collagen

Hongning Zheng, Cheng Lu, Jun Lan, Shilong Fan*, Vikas Nanda*, Fei Xu*

Shilong Fan, Vikas Nanda, Fei Xu
Email: fanshilong@mail.tsinghua.edu.cn, nanda@cabm.rutgers.edu, feixu@jiangnan.edu.cn,

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1802171115

Supplementary Information Text

Structural Analysis Methods and Calculation of Step Parameters. Collagen is treated as a series of triangles or Cα-triangles whose vertices are the Cα atoms from a Gly in one strand and the Cα atoms from the nearest non-Gly residues in the two other strands. The six step parameters relating Cα-triangles were computed* and listed in Table S2.

* Xu F, et al. (2017) Parallels between DNA and collagen - comparing elastic models of the double and triple helix. Sci Rep 7(1):12802.

Circular Dichroism (CD) Spectroscopy. Peptides were dialyzed in deionized water, freeze dried, weighed and re-dissolved in 10 mM phosphate buffer at pH 7.0 to make the stock solutions. The stock solutions of *a*, *b*, *c* and the corresponding and corresponding substitute sequences were diluted to a concentration of 0.2mM and combined at 1:1:1 ratio and heated at 80 °C for 10 min to ensure complete unfolding. Mixtures were subsequently incubated at 4 °C for ~24 hours.

Circular Dichroism (CD) experiments were performed on a Chirascan Instrument (Applied Photophysics Ltd, England) using optically matched quartz cuvettes with path lengths of 0.1 cm (Model 110-OS, Hellma USA). Wavelength scans were conducted from 190 to 260 nm at 4°C with a 0.5 nm increment per step and a 0.5 s averaging time, and repeated three times. For temperature melting experiments, the data were recorded every 1 °C/step with 6-min equilibration time from 4 to 80 °C. The temperature melting curves were acquired by monitoring the ellipticity at 225 nm. Apparent melting temperatures, T_m, were estimated as below:

$$\boldsymbol{F(T)} = \frac{\boldsymbol{\theta(T) - \theta_U(T)}}{\boldsymbol{\theta_F(T) - \theta_U(T)}}$$

where $\boldsymbol{\theta(T)}$ was the observed ellipticity and $\boldsymbol{\theta_F(T)}$ and $\boldsymbol{\theta_U(T)}$ were estimated ellipticities derived from linear fits to the folded and unfolded baselines. The melting temperature was estimated as *T*, where *F*(*T*) = 0.5.

Melting Temperature Predictions. We used a weighted sum of stabilities for various charge pair networks based on the experimental and computational analyses of *abc* (Eq. 1 from the main document):

$$T_m = \sum_{i=1}^{4} n_i p_i \Delta T_{m,i} + \sum_{j=1}^{4} n_j \Delta T_{m,j}$$

(attractive) (repulsive)

For lateral DK and axial KD, $\Delta T_{m,i}$ was calculated in the isolated and complex states, respectively (**Table S4**). As the attractive ion pairs should stabilize the triple helix, the observed marginal negative $\Delta T_{m,i}$ values of the lateral KD and axial DK were assumed to be zero. The same nomenclature is used in a type of repulsive interactions, *j*, except that the formation probability was assumed to be one. The cutoff for effective electrostatic interactions was set as 6 Å. The weight, $\boldsymbol{p_i}$, is the total probability of forming an effective electrostatic interaction. Using backbone propensity terms developed by Persikov, Ramshaw and Brodsky* resulted in negative stabilities for all sequences, due to their low Pro/Hyp content. Therefore, these terms were not used in the current scoring function.

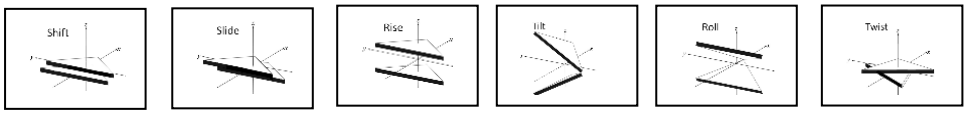
*Persikov, A.V., Ramshaw, J.A.M., Brodsky, B. (2005) "Prediction of collagen stability from amino acid sequence." *J. Biol. Chem.* **280**(19): 19343-19349.

SUPPLEMENTAL TABLES

Table S1. Statistics of crystal data collection and structure refinement. One crystal was used for each structure. Values in parentheses are for the highest resolution shell. R_{merge}=ΣhΣi|*Ih,i-Ih*|/ΣhΣi*Ih,i*, where *Ih* is the mean intensity of the *i* observations of symmetry related reflections of *h*. R=Σ|F_{obs}-F_{calc}|/ΣF_{obs}, where F_{calc} is the calculated protein structure factor from the atomic model (R_{free} was calculated with 5% of the reflections selected randomly).

Data collection	
Space Group	$P2_12_12_1$
Cell dimensions	
a, b, c (Å)	27.59, 51.39,128.52
α, β, γ, (°)	90, 90, 90
Resolution (Å)	47.72~1.77
R_{merge} (%)	5.5 (78.8)
I / σI	20.1 (1.38)
Completeness (%)	98.1 (77.3)
Redundancy	5.7 (3.2)
Refinement	
Resolution (Å)	47.72~1.77
No. reflections	18,280
R_{work} / R_{free} (%)	24.1/ 28.3
No. atoms	
Protein	988
B-factors	
Protein	30.2
R.m.s. deviations	
Bond lengths (Å)	0.005
Bond angles (°)	1.407
Ramachandran plot statistics (%)	
Most favoured	94.1
Additional allowed	5.9
Generously allowed	0.0
Disallowed	0.0

Table S2. The average step parameters and the standard deviations of the X-crystal triple helix structures with different helical periodicity are listed. The schematic presentations of the step parameters are shown at the top.



PDB ID	Shift (Å)	Slide (Å)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)
5YAN	−4.54	0.82	3.16	−7.94	6.25	−107.57
(this study)	(0.38)	(0.20)	(0.12)	(2.57)	(2.48)	(4.59)
1K6F	−4.16	0.92	3.25	−10.98	9.75	−101.92
$[(PPG)_{10}]_3$	(0.08)	(0.06)	(0.04)	(0.67)	(0.67)	(1.71)
3DMW	−4.45	0.67	3.16	−7.59	7.64	−106.81
(step 1-16)	(0.17)	(0.25)	(0.08)	(1.96)	(2.29)	(4.55)
3DMW	−4.35	0.73	3.17	−9.00	7.87	−103.27
(step 17-34)	(0.20)	(0.16)	(0.07)	(2.31)	(2.85)	(3.70)

Table S3. Predicted salt bridges and melting temperatures of all possible registries of the three chains *a*, *b*, and *c*.

	Lateral		Axial		Lateral		Axial		Predicted
	KD	DK	KD	DK	KK	DD	KK	DD	T_m
abc	4	6	6	5	0	0	0	0	28.9
acb	1	1	0	0	4	4	1	3	−28.0
bac	0	1	0	0	3	4	2	3	−28.1
bca	5	6	3	4	1	0	0	0	17.1
cab	4	3	5	5	1	0	0	0	17.5
cba	1	0	0	1	2	4	1	4	−32.0

Table S4. The parameters used in Eq. (1) to predict melting temperatures.

$$T_m = \sum_{i=1}^{4} n_i p_i \Delta T_{m,i} + \sum_{j=1}^{4} n_j \Delta T_{m,j}$$

	Lateral			Axial			Lateral		Axial	
	KD	DK		KD		DK	KK	DD	KK	DD
		isolated	complex	isolated	complex					
ΔT_m (˚C)	0	3.3	2.8	5.5	2.9	0	−1.15	−2.6	−1.25	−4.5
p	0.242	0.527	0.540	0.958	0.975	0.098	1	1	1	1

Table S5. Comparison between the measured and predicted melting temperatures using the original* and updated (Eq.(1)) scoring functions. Isolated contacts occur between residue pairs where no other contacts are possible for either residue. Complex contacts occur where either or both residues are also involved in other charge-pair interactions. For each association stoichiometry, the melting temperatures of all possible chain registries are predicted and the highest one is listed here. Peptide heterotrimer system names are consistent with the associated references.

*Xu F, Zhang L, Koder RL, & Nanda V (2010) "De novo self-assembling collagen heterotrimers using explicit positive and negative design." *Biochemistry* 49(11):2307-2316.

peptide system	# lateral contacts			# axial contacts			# lateral		# axial		T_m (°C)			Ref.
	KD	DK		KD		DK	KK	DD	KK	DD	Predicted updated	Predicted original	Measured	
		isolated	complex	isolated	complex									
abc	4	2	4	1	5	5	0	0	0	0	28.9	33.8	29	[1]
pa1:b:c	5	3	3	0	4	4	1	0	0	0	19.9	32.3	24	[2]
pa4:e:f	4	2	4	1	5	4	0	0	0	0	28.9	33.4	24	
d:e:f	4	2	4	0	5	4	0	0	0	0	23.7	33.1	24	
pa2:b:c	3	2	3	1	3	3	0	0	2	0	19.3	27.9	19.8	
pa3:e:f	4	3	3	0	4	3	0	1	0	1	14.0	32.3	17.6	
e:f:c	0	2	5	2	0	5	1	0	0	0	20.4	30.5	15	
αβγ	5	0	0	5	9	0	0	0	0	0	51.8	35.9	58	[3]
α:γ	6	0	0	3	5	0	0	0	0	0	29.9	34.0	44	
β:γ	6	0	0	4	4	0	0	0	0	0	32.4	34.0	44	

[1] Xu F, Zahid S, Silva T, & Nanda V (2011) Computational design of a collagen A: B: C-type heterotrimer. *JACS* 133(39):15260-15263.

[2] Xu F, Silva T, Joshi M, Zahid S, & Nanda V (2013) Circular Permutation Directs Orthogonal Assembly in Complex Collagen Peptide Mixtures. *JBC* 288(44):31616-31623.

[3] Fallas JA & Hartgerink JD (2012) Computational design of self-assembling register-specific collagen heterotrimers. *Nature Comm*. 3:1087.

Table S6. Predicted melting temperatures (see Eq. 1) of the best competing association among the six associations of the heterotrimers bearing point mutations (abc, acb, bca, bac, cab, cba). The alanine mutations are listed in **(A)**, and the mutations of the reversed charges are listed in **(B)**.

(A)

Peptides			best competing state	Predicted T_m (°C) [§]
a	*b*	*c*	bca	17.1
a	*b*	*c*:K15A	bca	15.3
a	*b*	*c*:D10A	bca	17.1
a	*b*:D21A	*c*	bca	14.3
a	*b*:K18A	*c*	bca	15.3
a	*b*	*c*:D7A	bca	17.1
a:D24A	*b*	*c*	bca	17.1
a:D18A	*b*	*c*	bca	19.5

[§] The predicted melting temperatures were calculated with Eq. (1).

(B)

Peptides			best competing state	Predicted T_m (°C)
a	b	c:K15D	cba	17.5
a	b	c:D10K	cba	14.9
a	b:D21K	c	bca	14.7
a	b:K18D	c	bca	14.1
a	b	c:D7K	bca	17.5
a:D24K	b	c	bca	17.5

SUPPLEMENTAL FIGURES

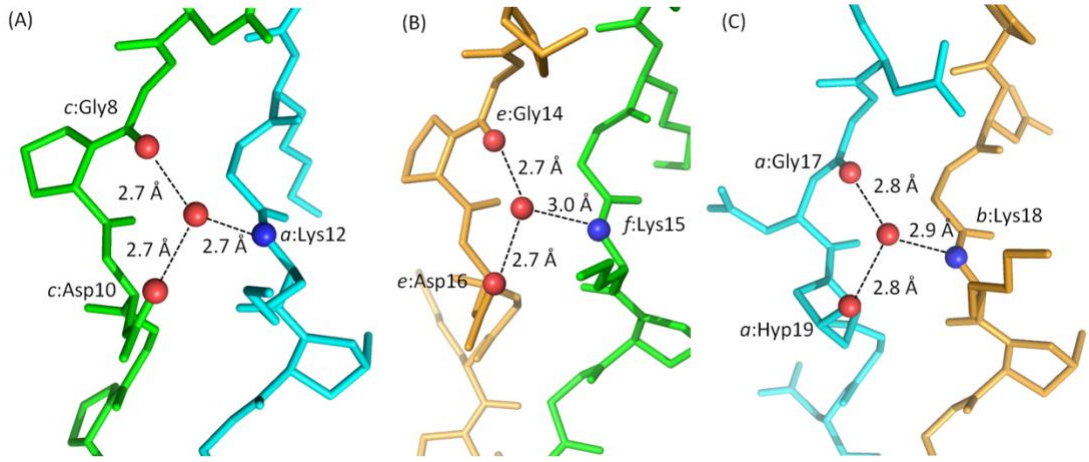


Figure S1. Examples from the *abc* structure of the extensive, repeated water-mediated hydrogen bond networks between the backbone and side chain atoms along the triple-helix surface.

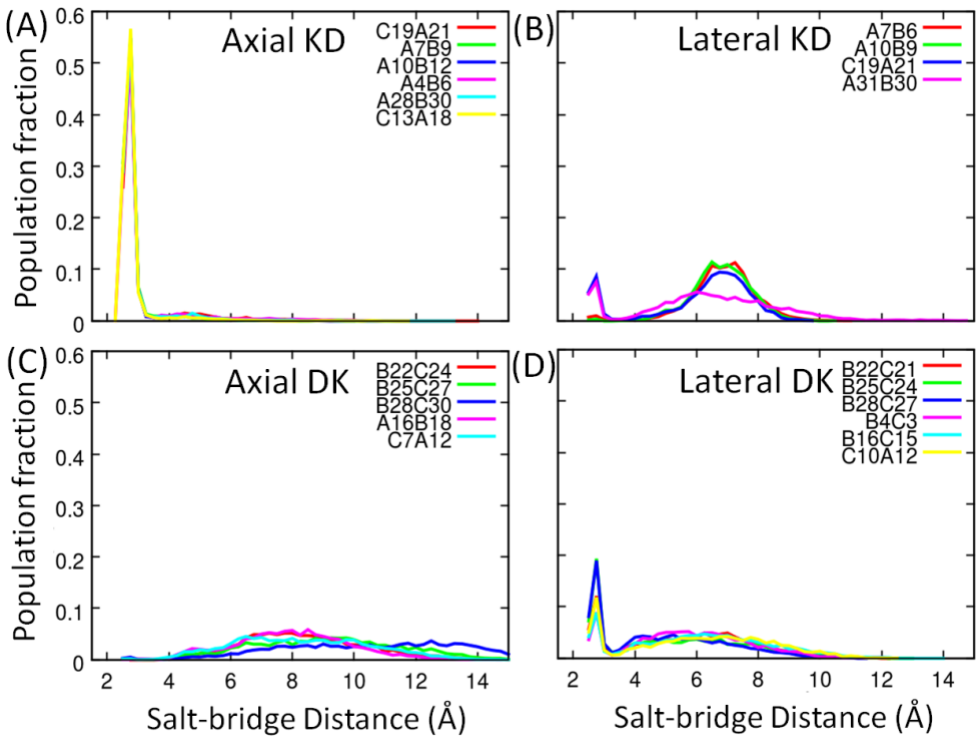


Figure S2. Salt-bridge distance distributions sampled by molecular dynamics (MD) simulations of *abc*. The chain identifications and residue numbers of the salt bridges are as labeled. Distances are between the closes carboxylate oxygen of Asp and the sidechain amine nitrogen of Lys.

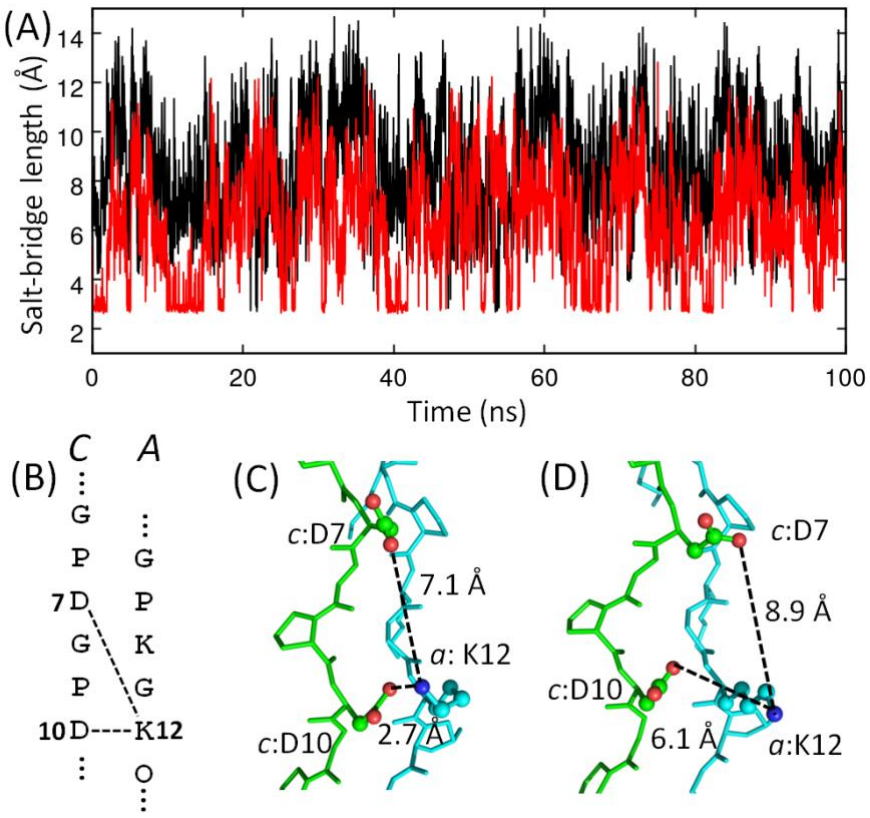


Figure S3. (A) Salt-bridge distance as a function of the simulation time for a three-residue network containing a lateral DK (*c*:D10-*a*:K12) and axial DK (*c*:D7-*a*:K12). The residue locations are shown in the diagram **(B)**. Two representative conformations are shown in **(C)** and **(D)**.

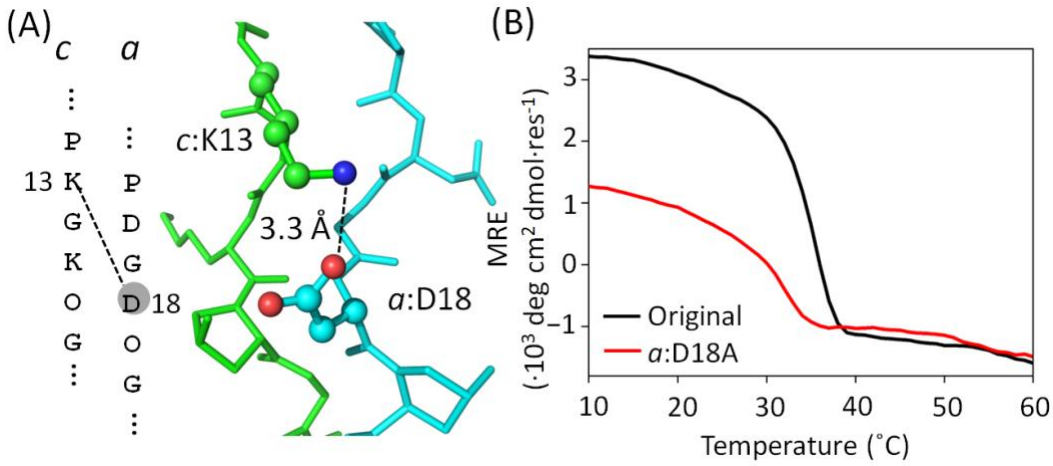


Figure S4. Characterization of an isolated axial KD salt bridge. **(A)** Sequence and structural representations of the *c*:K13-*a*:D18 interaction. **(B)** The *a*:D18A substitution resulted in an observed T_m = 28.6ºC, significantly less stable than the original abc T_m = 34.1ºC. MRE was monitored by CD at 225 nm.

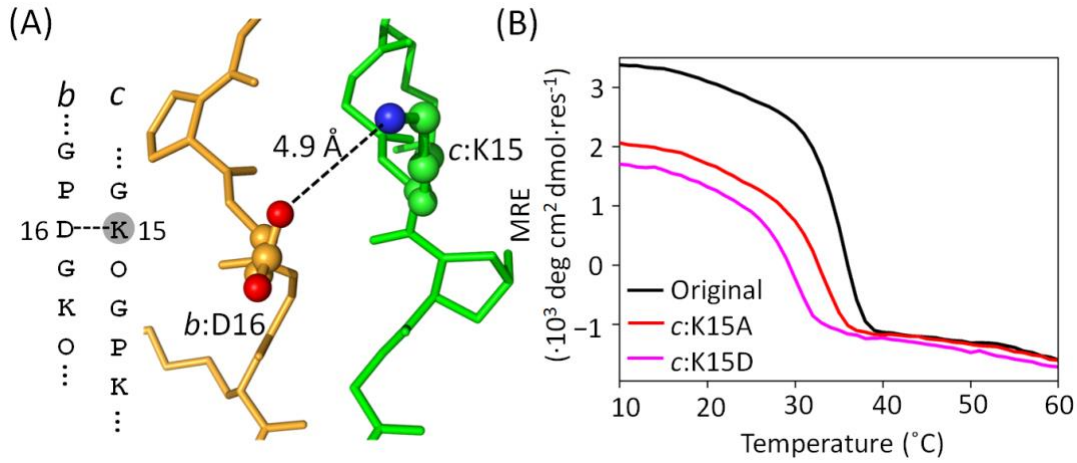


Figure S5. Characterization of an isolated lateral DK salt-bridge. **(A)** Sequence and structural representations of the *b*:D16-*c*:K15 interaction. **(B)** *c*:K15A T_m=30.8ºC, *c*:K15D T_m=28.2ºC. MRE was monitored by CD at 225 nm.

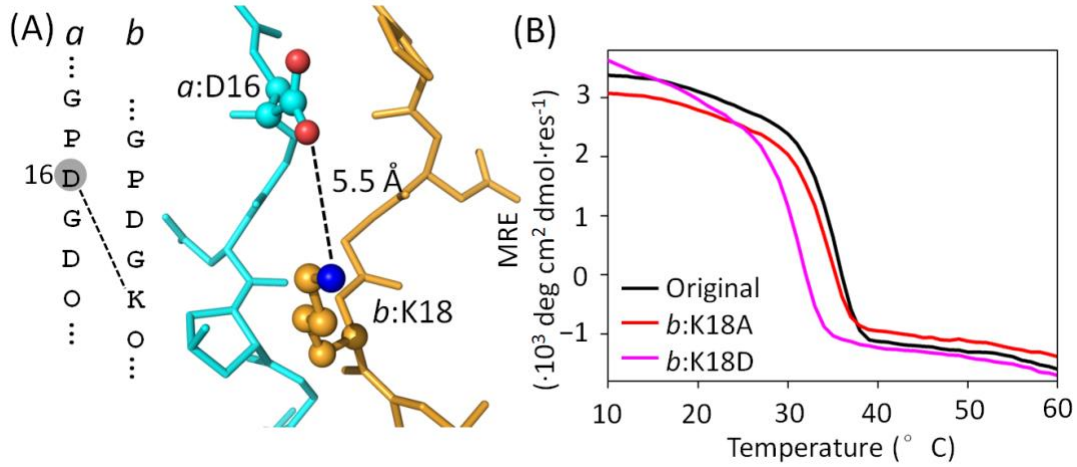


Figure S6. Characterization of an isolated axial DK salt bridge. **(A)** Sequence and structural representations of the *a*:D16-*b*:K18 interaction. **(B)** *b*:K18A T_m=33.5ºC, *b*:K18D T_m=29.0ºC. MRE was monitored by CD at 225 nm.

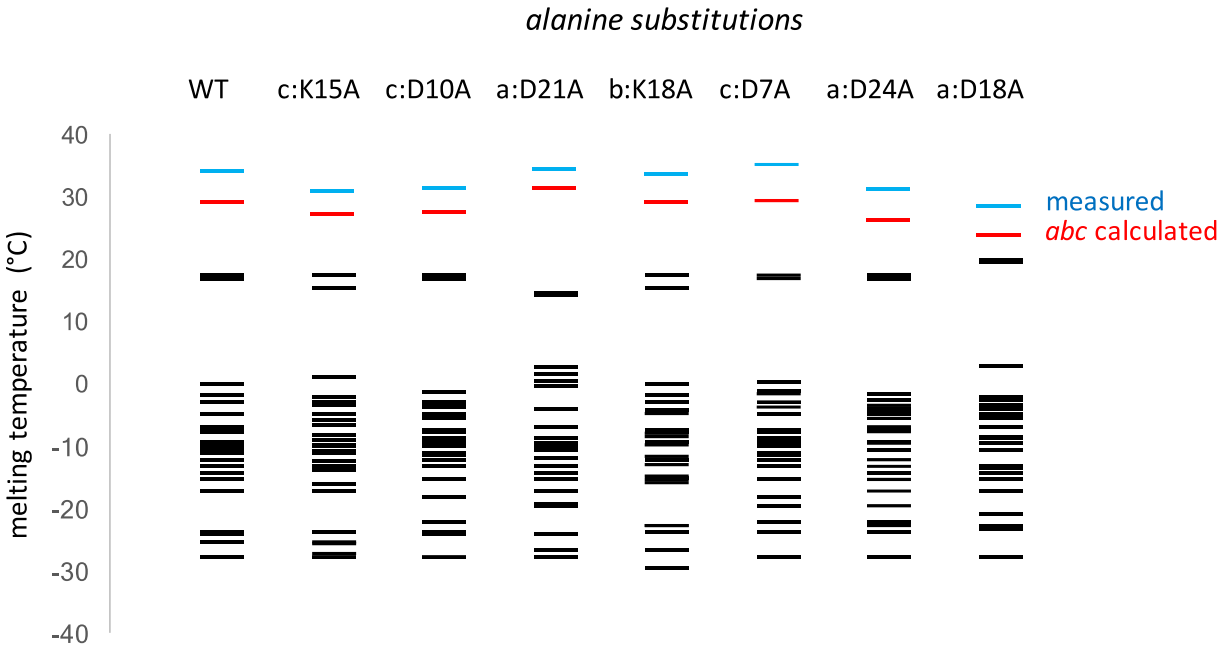


Figure S7. Predicted melting temperatures of Ala-substituted peptides using the updated model (see Eq. 1 in the main document). Stabilities were calculated for all 27 association states (*aaa, aab, aac, aba, abb … ccc*) for wild-type (WT) *abc* and peptide substitutions. The Tm of the *abc* target association (red) state is higher than the remaining 26 competing states for all substitutions. Measured stabilities (blue) correlate with the stability of *abc*. Stability values of most stable competing states are found in **Table S6**. An addition of tyrosine-glycine at the N-terminus of each chain increased the melting temperature (T_m) of the a:b:c mixture from the previously observed 29 °C to 34°C.

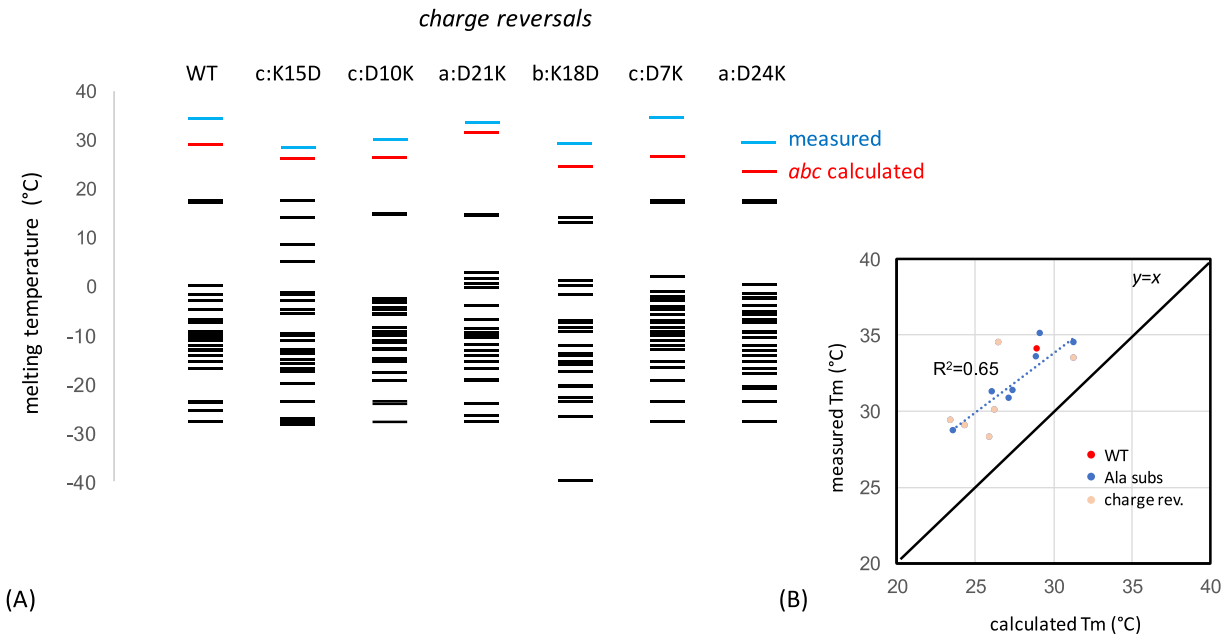


Figure S8. Predicted melting temperatures of charge-reversal substitutions to *a*, *b* and *c* using the updated model (see Eq. 1 in the main document). **(A)** Stabilities were calculated for all 27 association states (*aaa, aab, aac, aba, abb … ccc*) for wild-type (WT) *abc* and peptide substitutions. The Tm of the *abc* target association (red) state is higher than the remaining 26 competing states for all substitutions. **(B)** Measured and computed melting temperatures show a positive correlation with Ala (blue circles) and charge-reversal (brown circles) substitutions. The calculation consistently over-predicts the stability by about 5 degrees (y=x shown for comparison). Stability values of most stable competing states are found in **Table S6**. An addition of tyrosine-glycine at the N-terminus of each chain increased the melting temperature (T_m) of the a:b:c mixture from the previously observed 29 ºC to 34ºC.

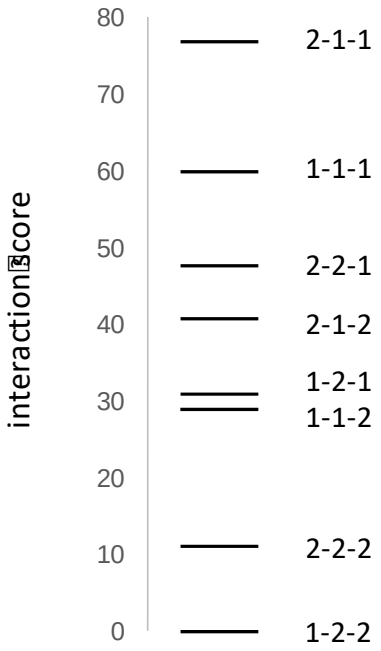


Figure S9. Computed stabilities of Type I collagen association states. Using Eq. 1 on the ~1000 residue-long triple-helical regions of human COL1A1 and COL1A2, the 2-1-1 ($\alpha2\alpha1\alpha1$) association state with $\alpha2$ chain in the leading position has the most favorable score. Refer to **Fig. S10** for positions specific interaction energies.

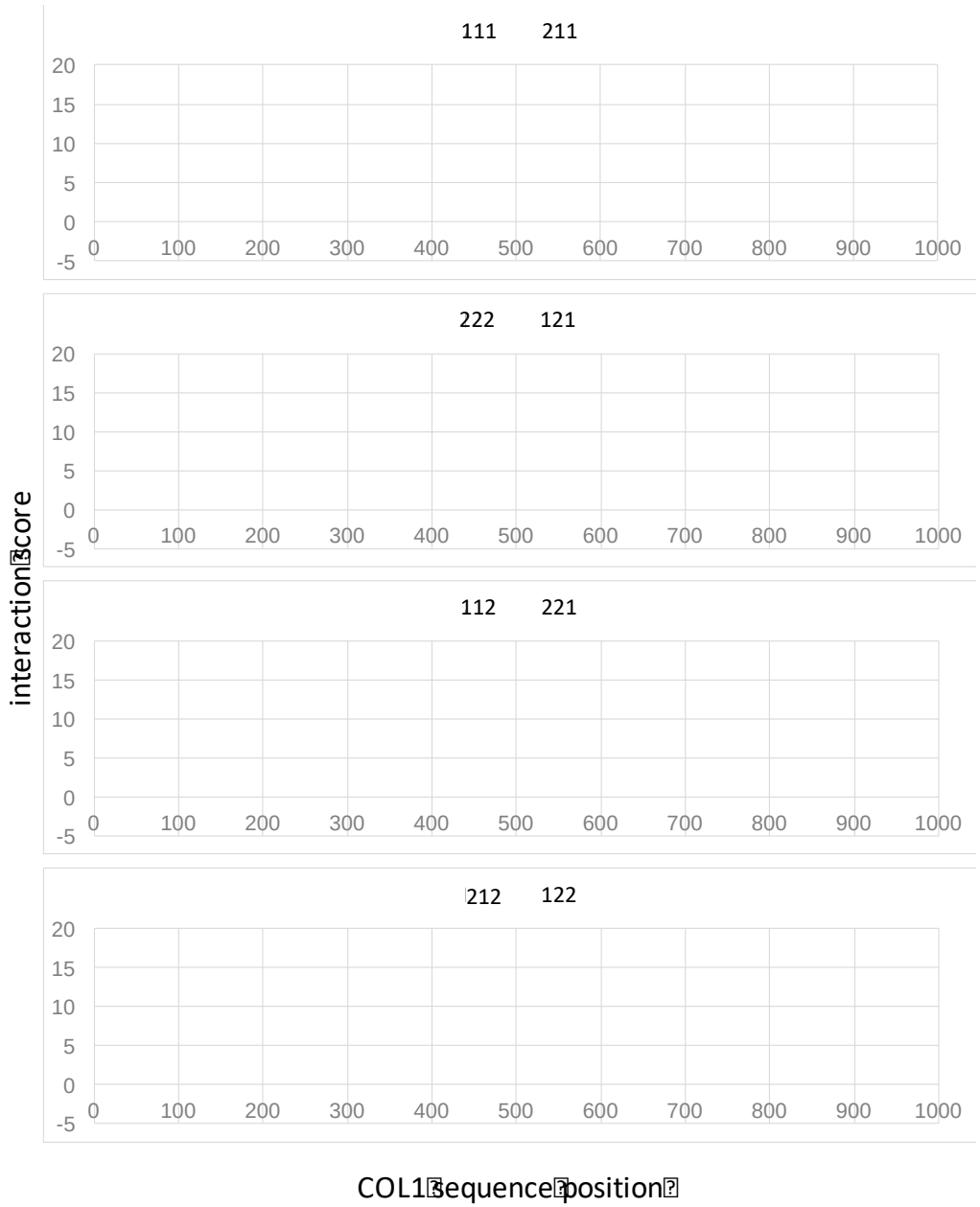


Figure S10. Position specific computed stabilities of Type I collagen association states using Eq. 1 on the ~1000 residue-long triple-helical regions of human COL1A1 and COL1A2. Although the 2-1-1 association state is favored when all interactions are summed along the entire sequence (see Fig. S9), for many charge-bearing domains, both 1-1-1 and 2-1-1 have equal computed stabilities – and in some regions, 1-1-1 is more stable.

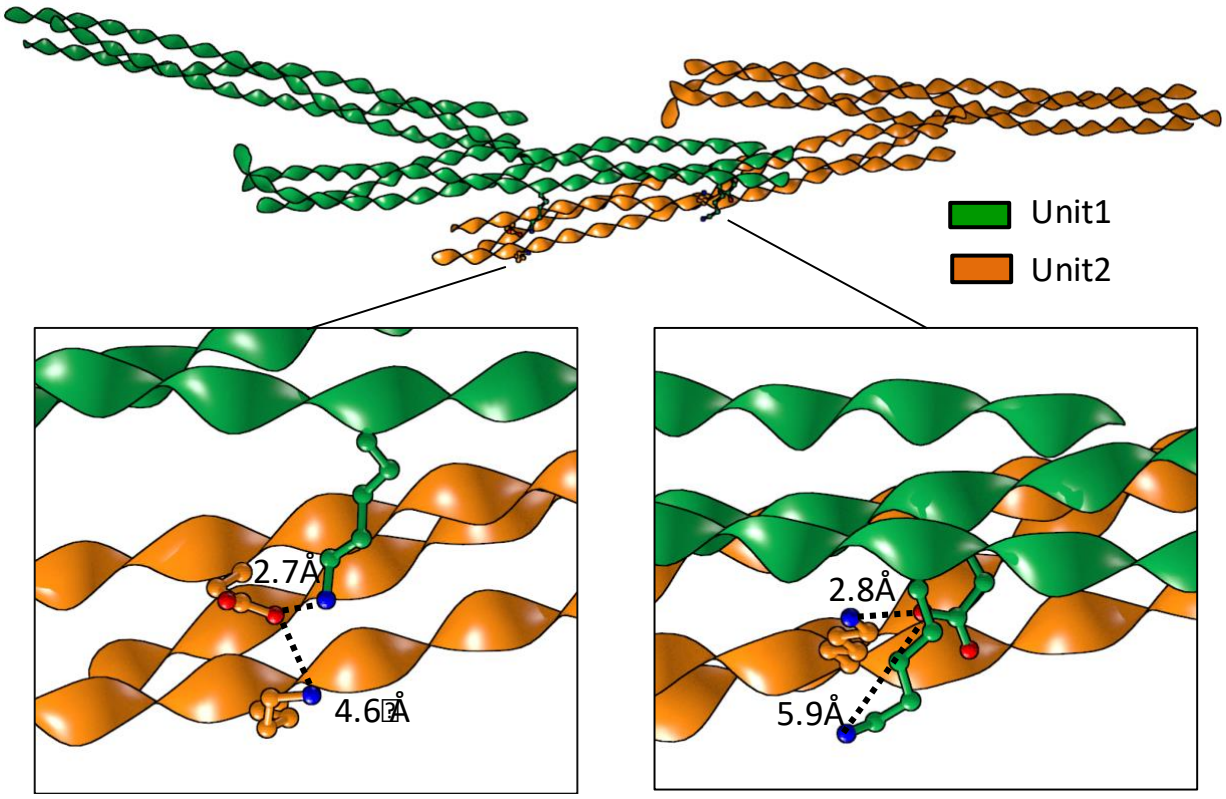
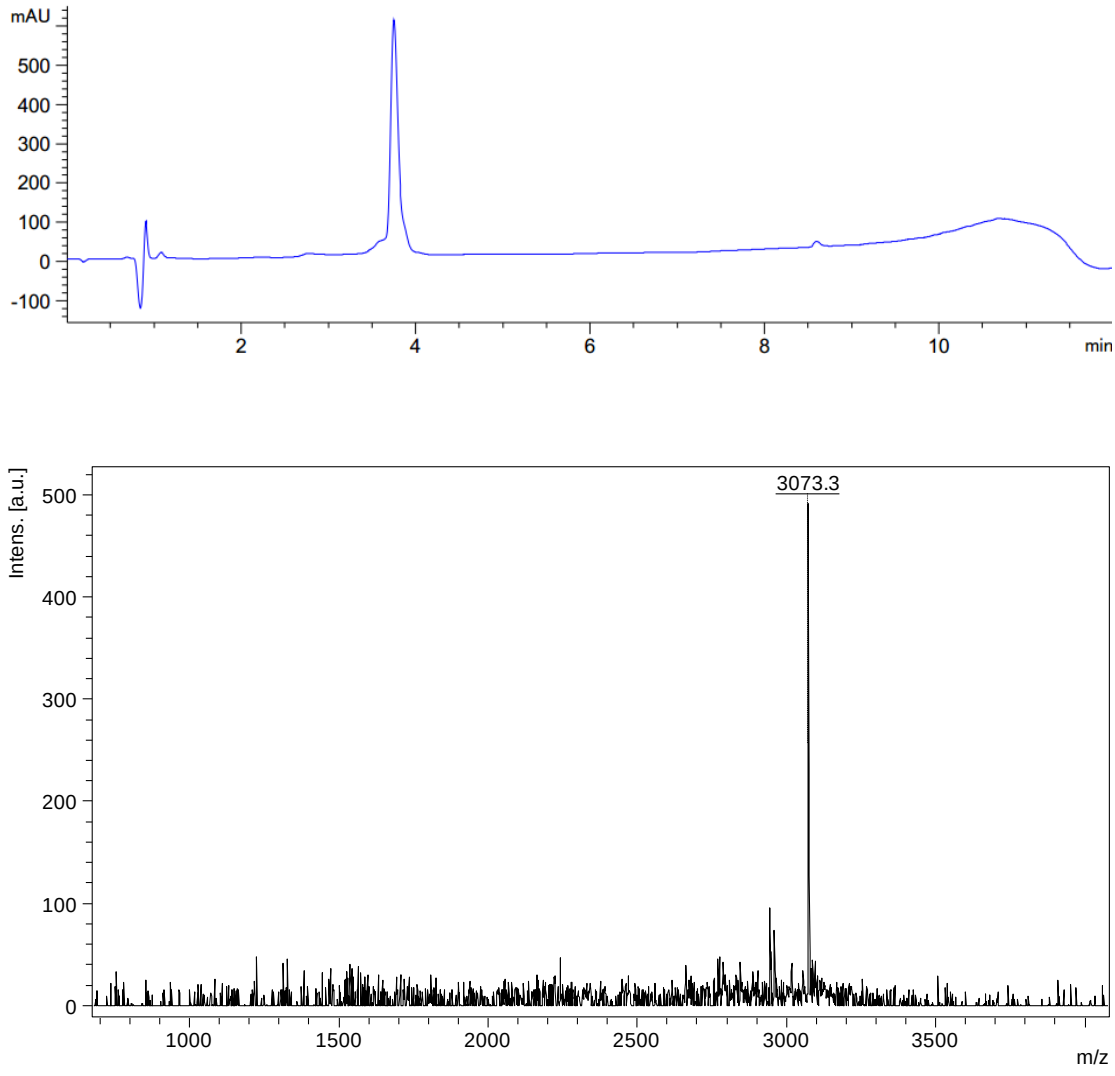


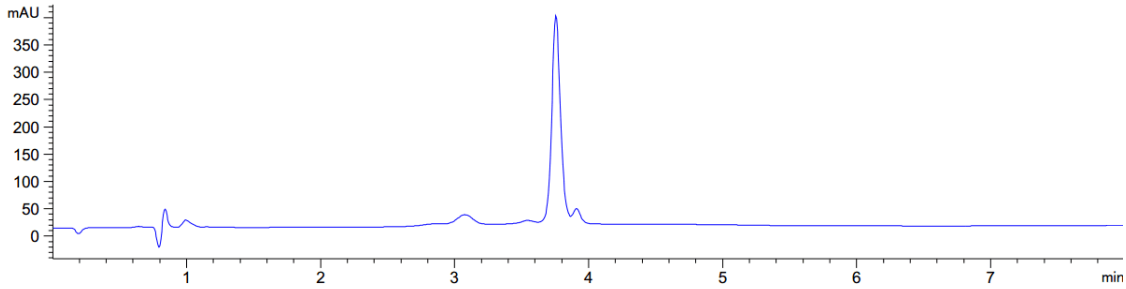
Figure S11. Lattice contacts between asymmetric units in the abc structure are primarily mediated by solvent. Interhelical charge-pair networks are observed. *b*:D28 makes two favorable charge pairs, one with *c*:K27 on the same helix, and a second with *b*:K8 on an adjacent unit. This interaction is seen on both ends of an antiparallel interhelical interface.

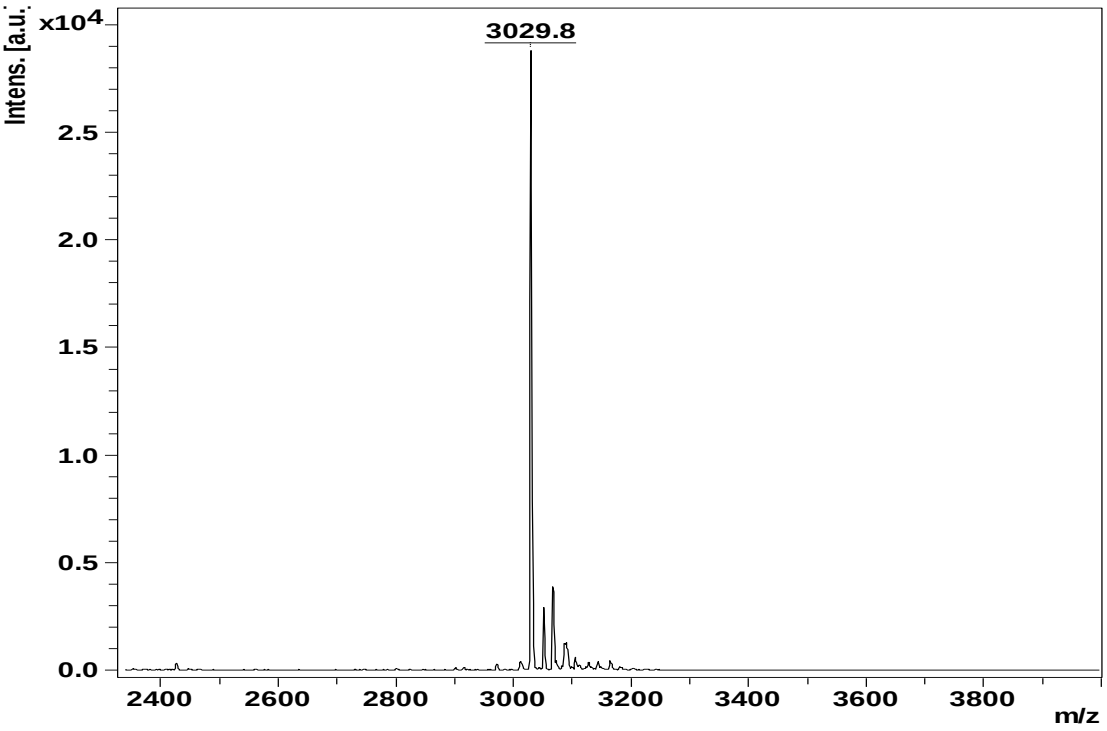
Figure S12. HPLC Chromatograms and Mass Spectrometry analysis of peptides. Purification and analysis of all peptides was performed by Reversed-Phase HPLC on an Agilent 1260 system using an AdvanceBio Peptide Map column (C18), 2.1×150 mm, pore size 2.7-micron. Mass spectrometric (MS) analysis was performed on an Ultraflex II TOF instrument (Bruker). A 10 mg/ml solution of 2,5-dihydroxy-benzoic acid in water, 50% acetonitrile was used for the MALDI matrix. Spectra were acquired using a dual-stage reflectron mirror.

a YGPKGPKGPKGKOGPDGDOGDOGDOGPKGPKG

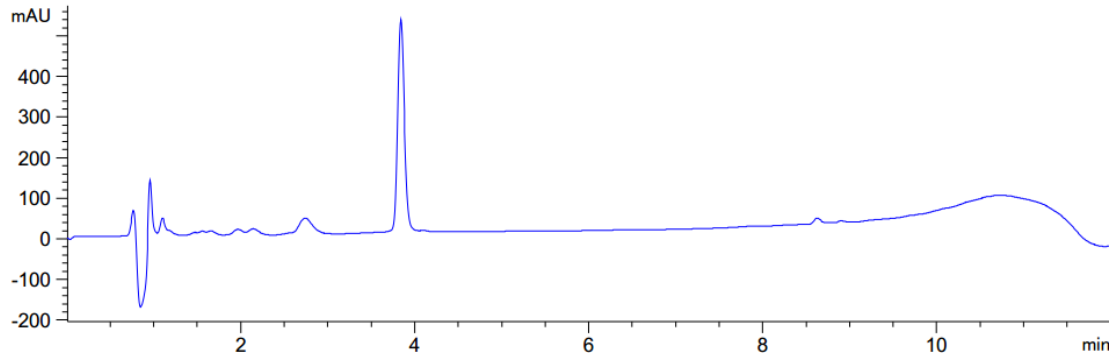


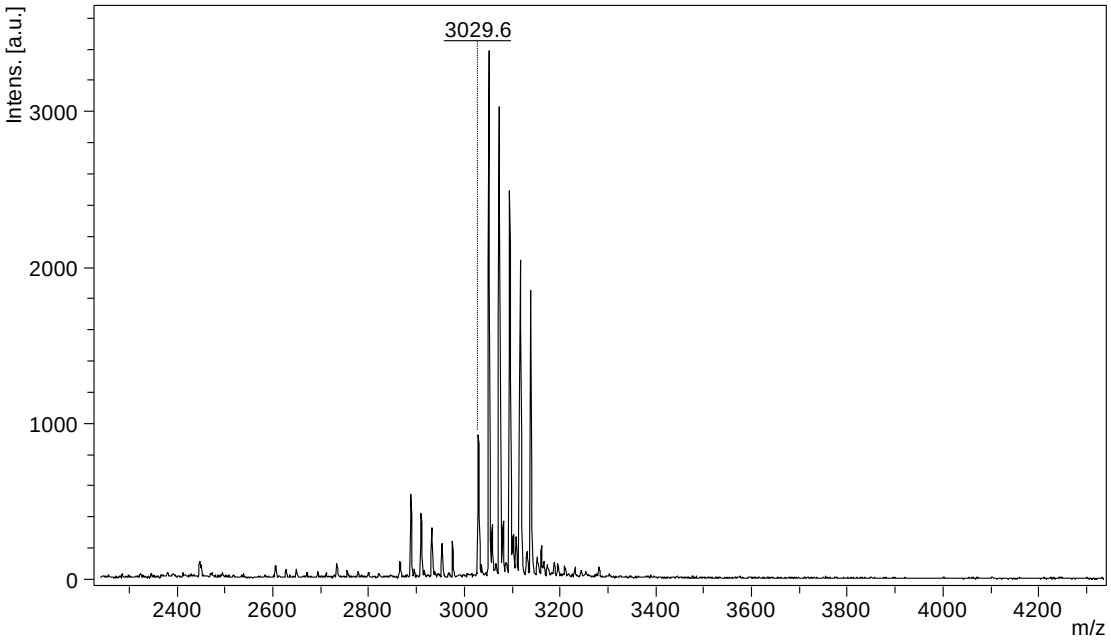
a:D18A YGPKGPKGPKGKOGPDG**A**OGDOGDOGPKGPKG



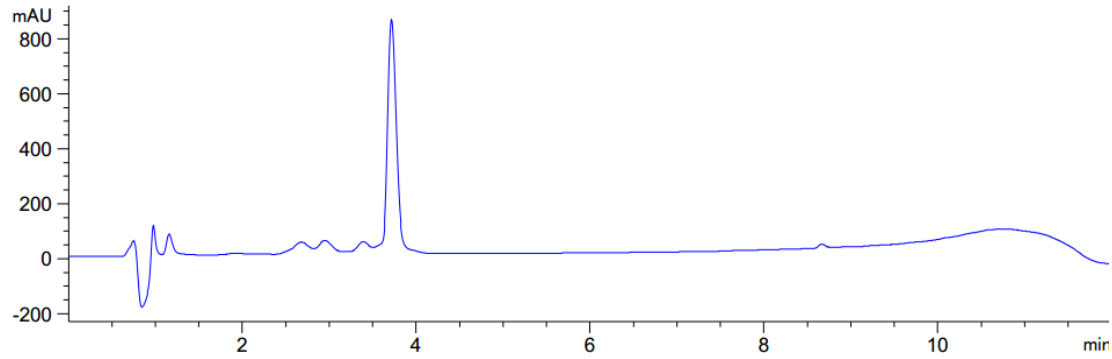


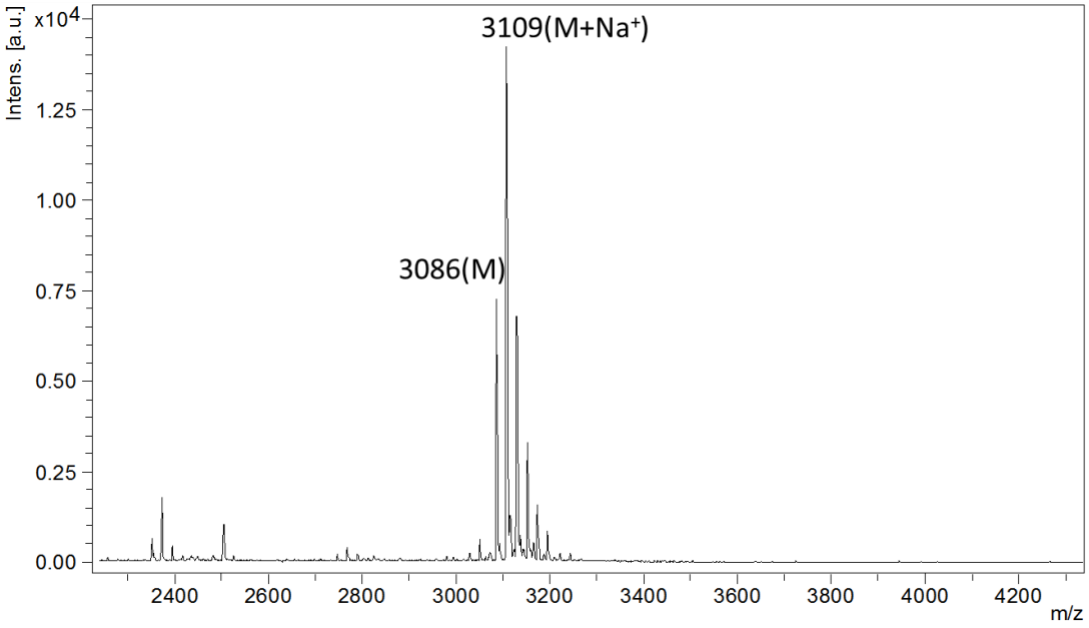
a:D21A YGPKGPKGPKGKOGPDGDOG**A**OGDOGPKGPKG



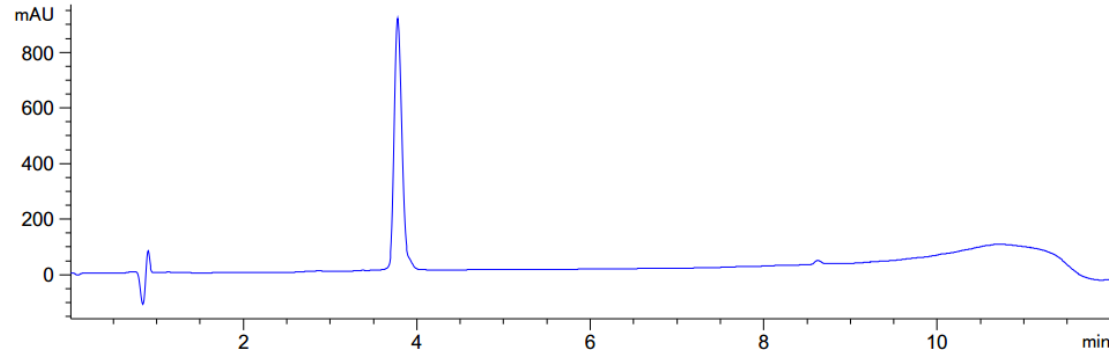


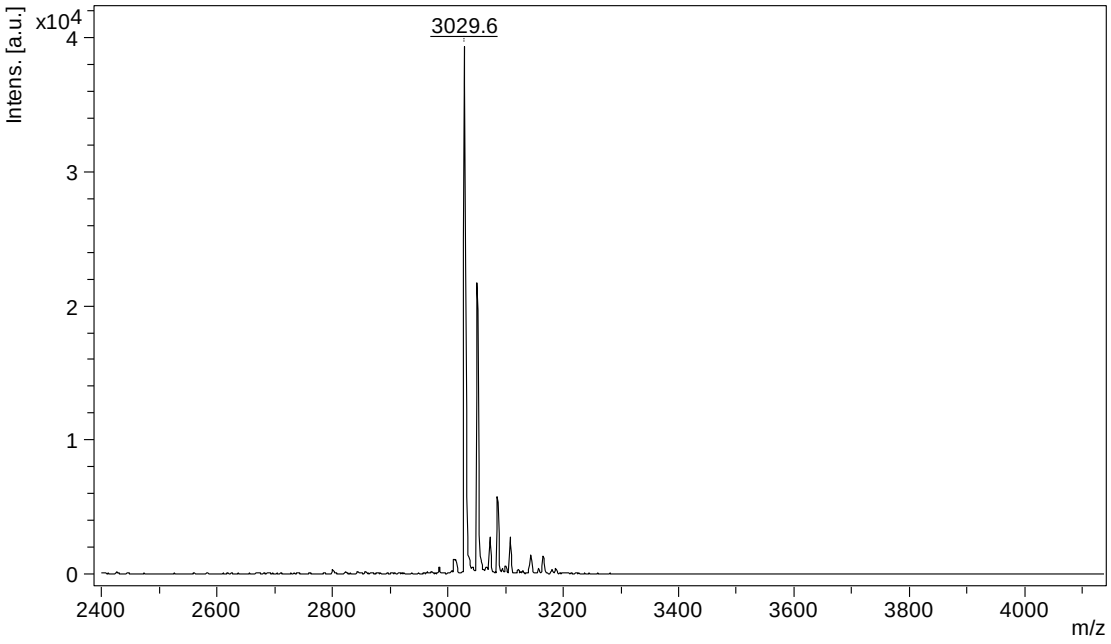
a:D21K YGPKGPKGPKGKOGPDGDOG**K**OGDOGPKGPKG



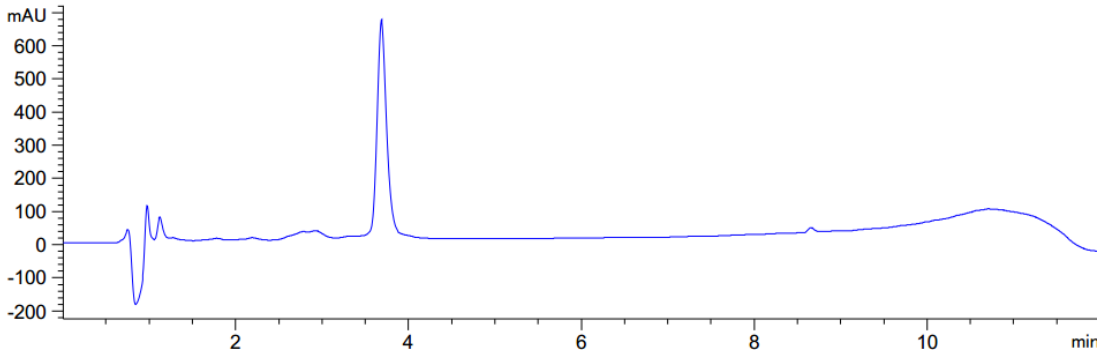


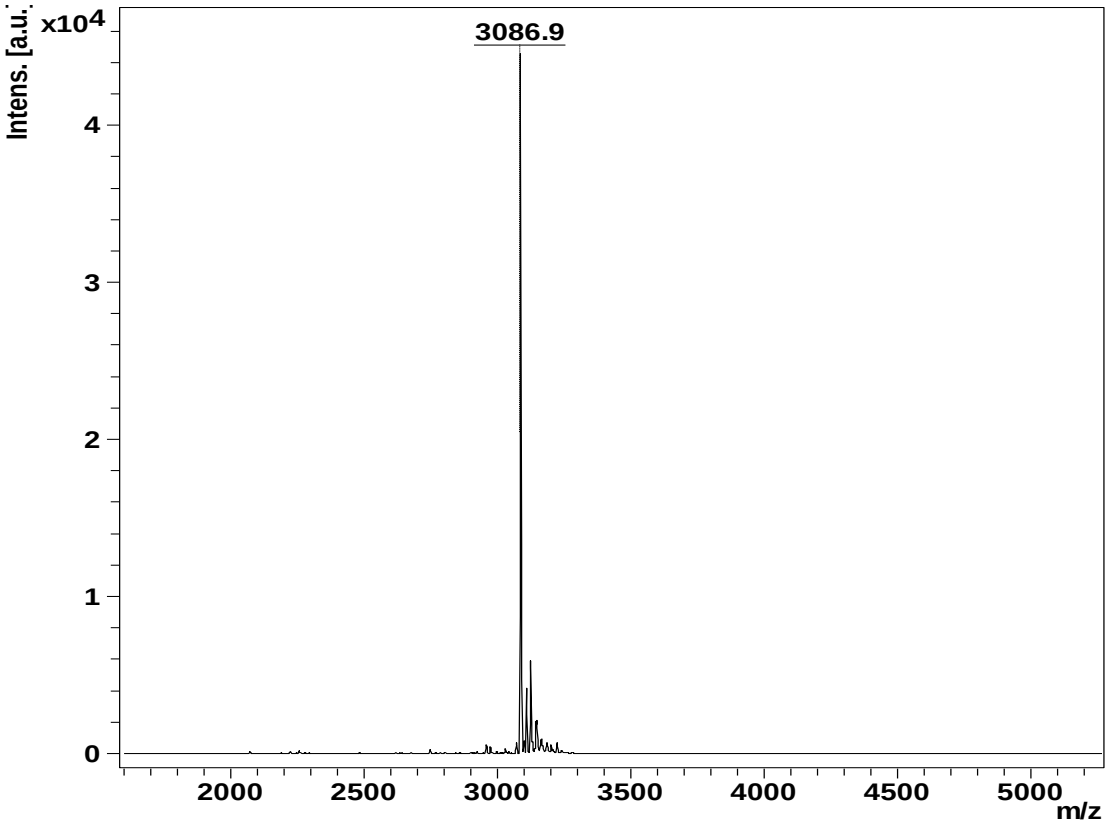
a:D24A YGPKGPKGPKGKOGPDGDOGDOG**A**OGPKGPKG



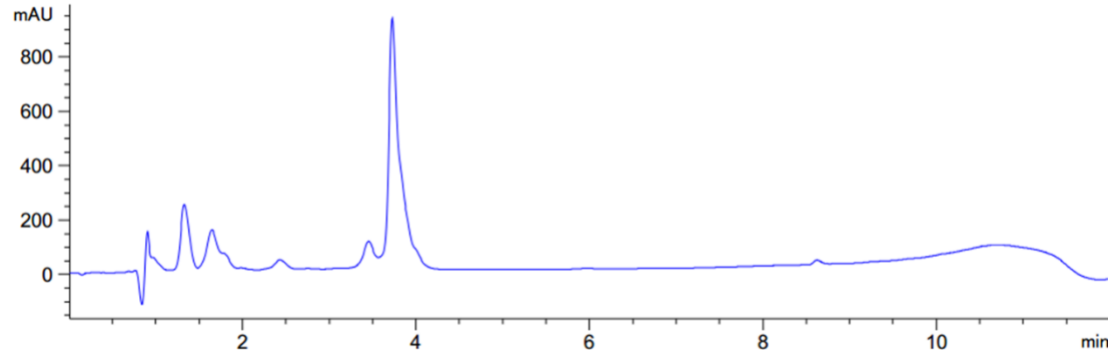


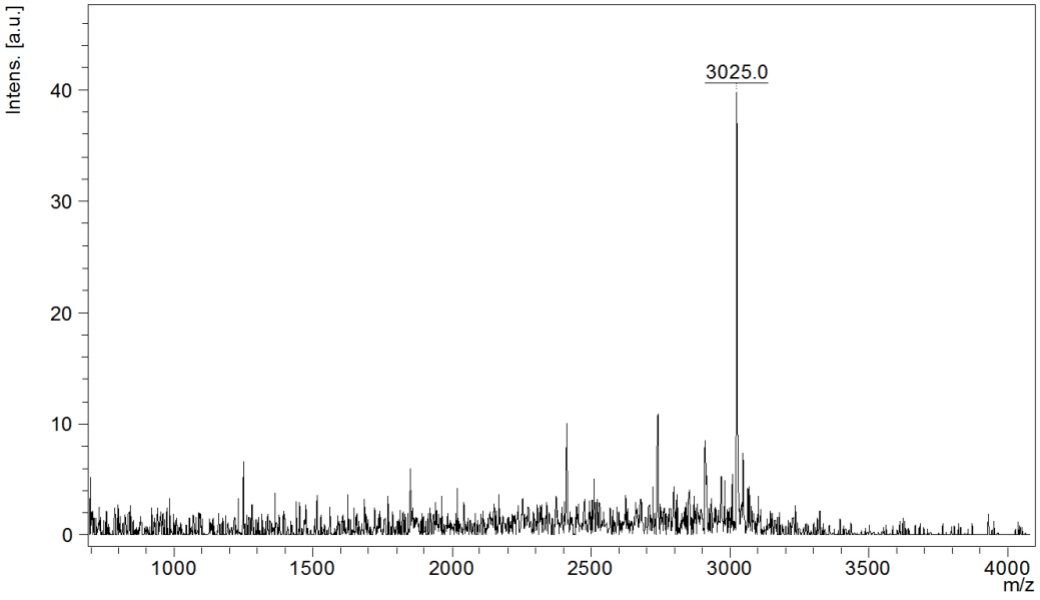
a:D24K YGPKGPKGPKGKOGPDGDOGDOG**K**OGPKGPKG



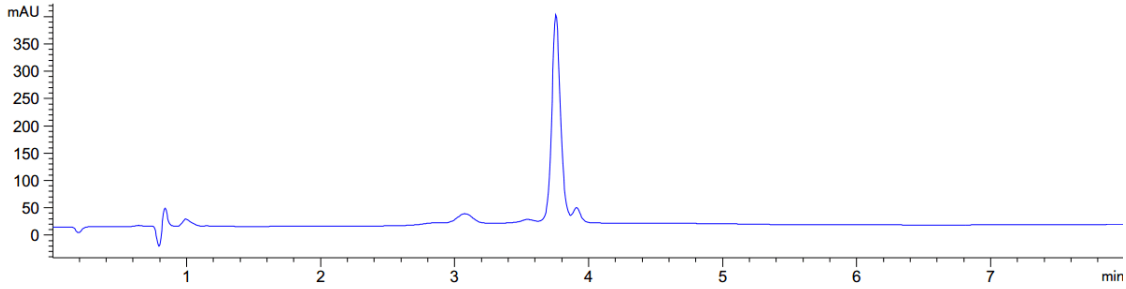


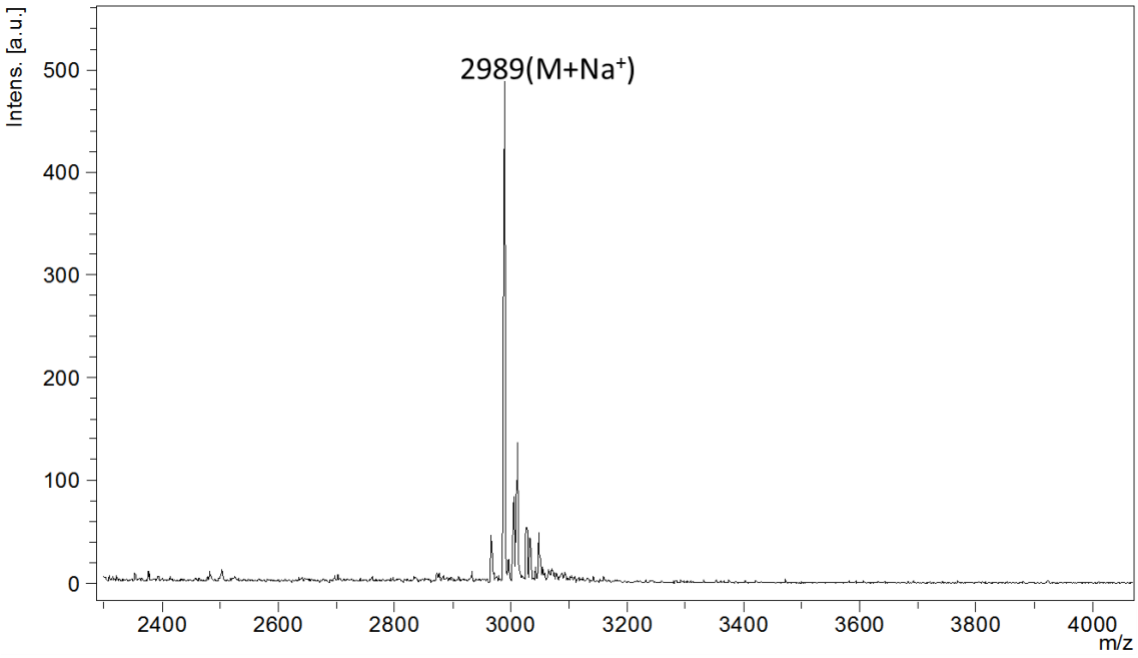
b YGPDGDOGDOGDOGPDGKOGPDGPDGPDGDOG



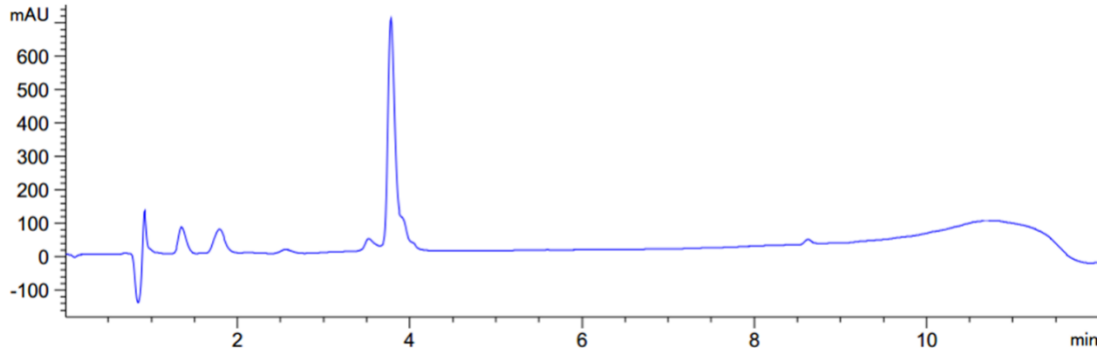


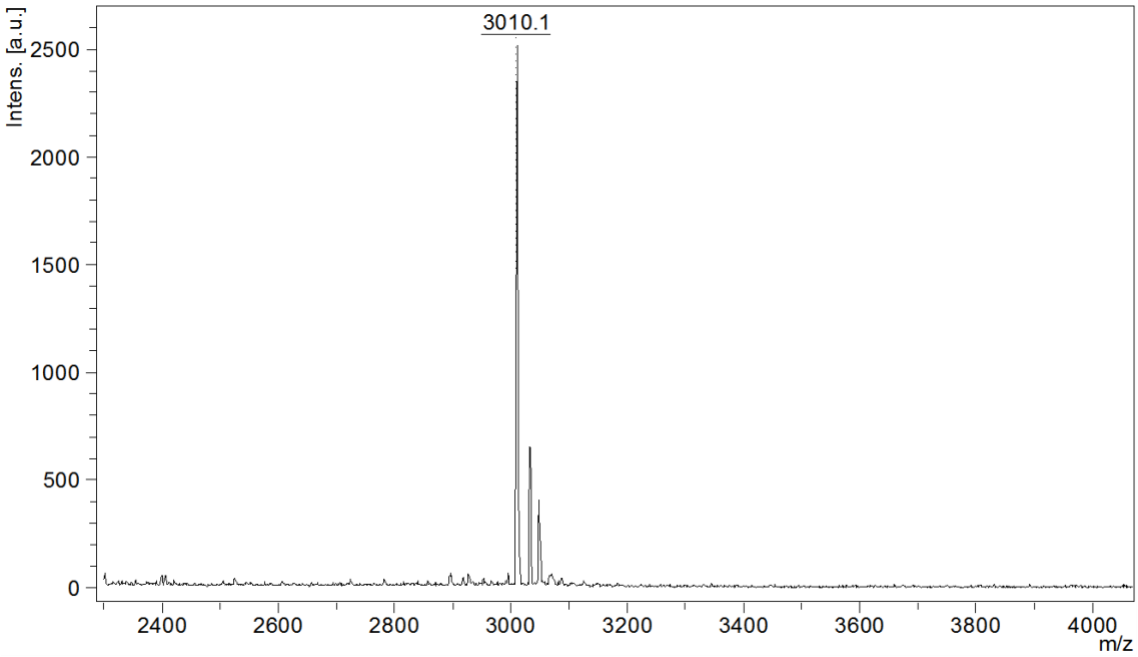
b:K18A YGPDGDOGDOGDOGPDG**A**OGPDGPDGPDGDOG



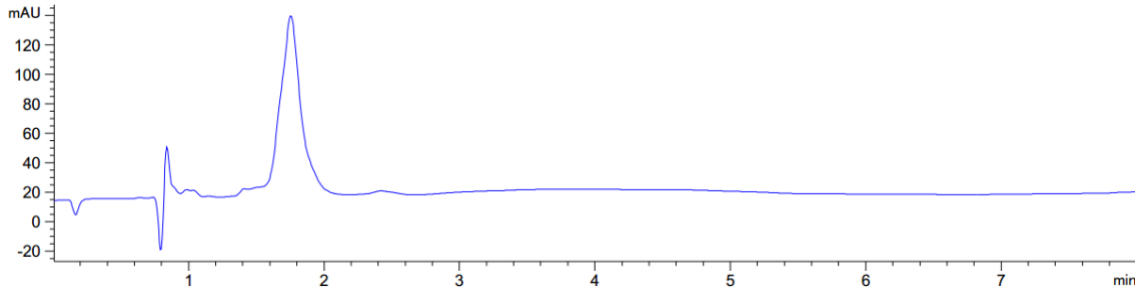


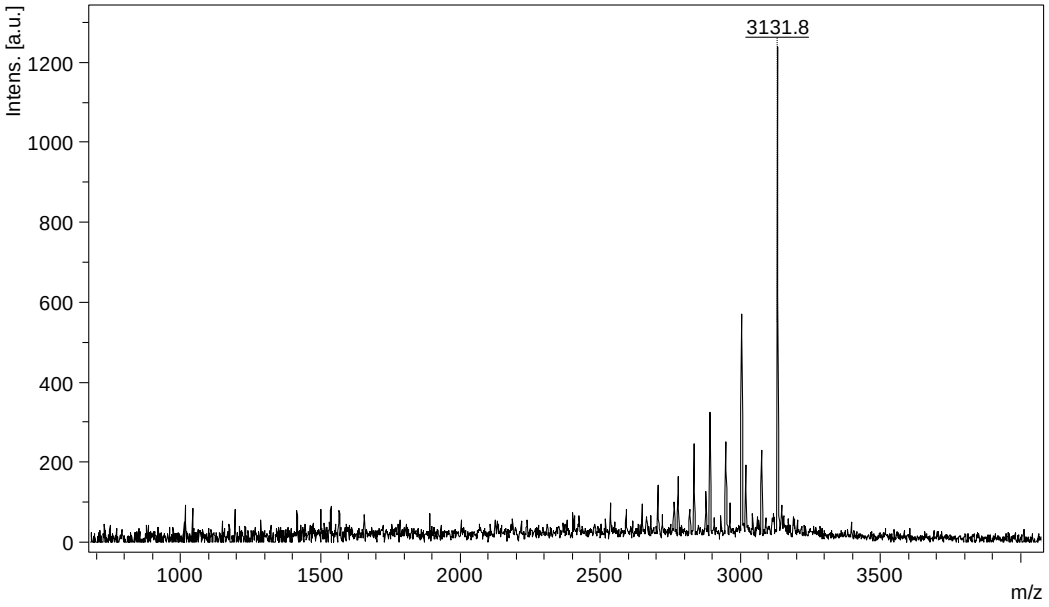
b:K18D YGPDGDOGDOGDOGPDG**D**OGPDGPDGPDGDOG



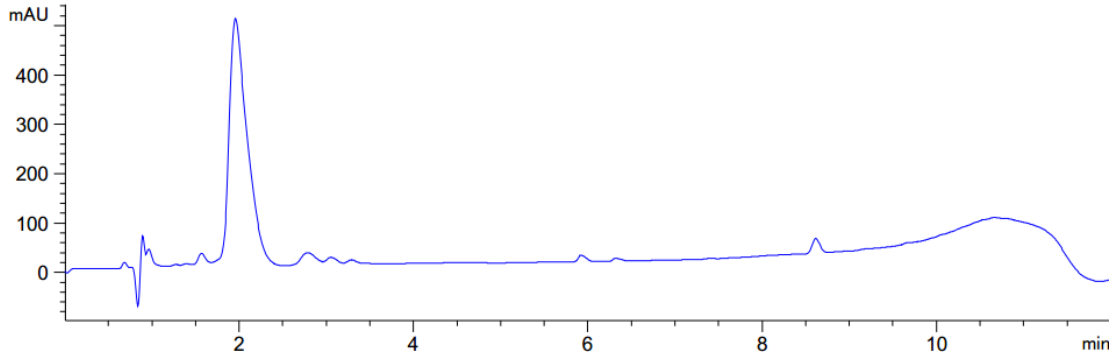


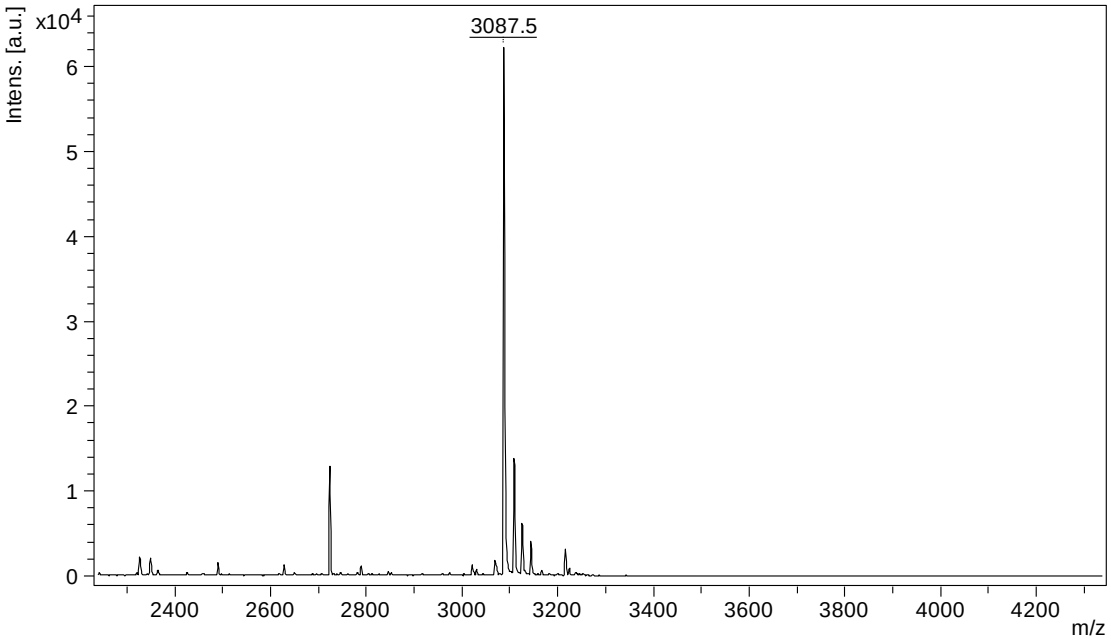
c YGKOGPDGPDGPKGKOGPKGKOGKOGKOGKOG



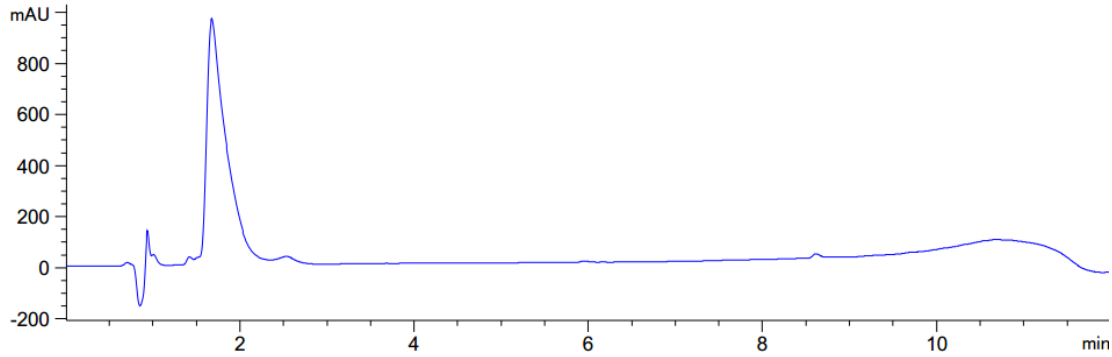


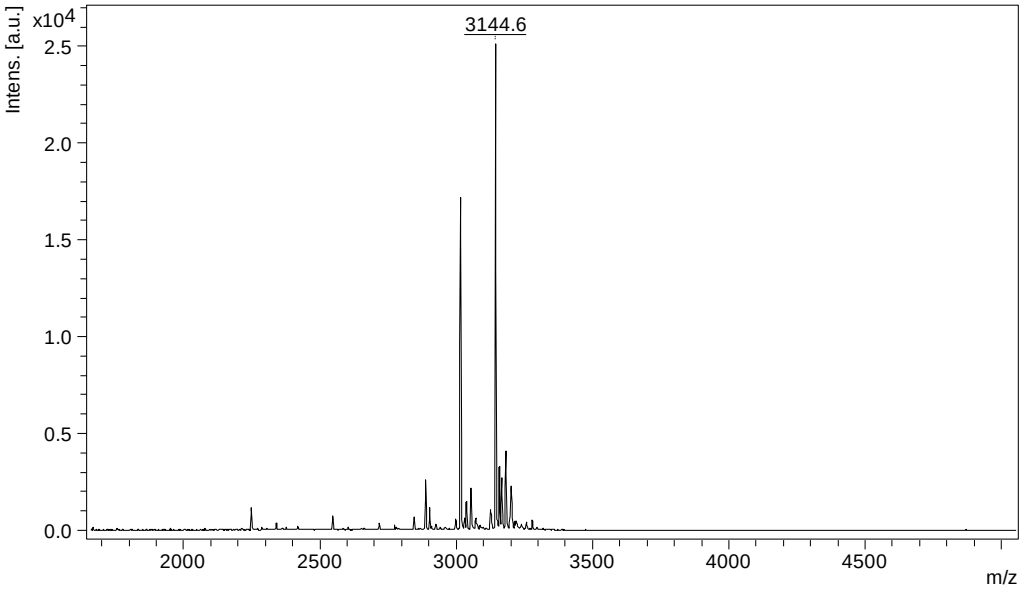
c:D7A YGKOGP**A**GPDGPKGKOGPKGKOGKOGKOGKOG



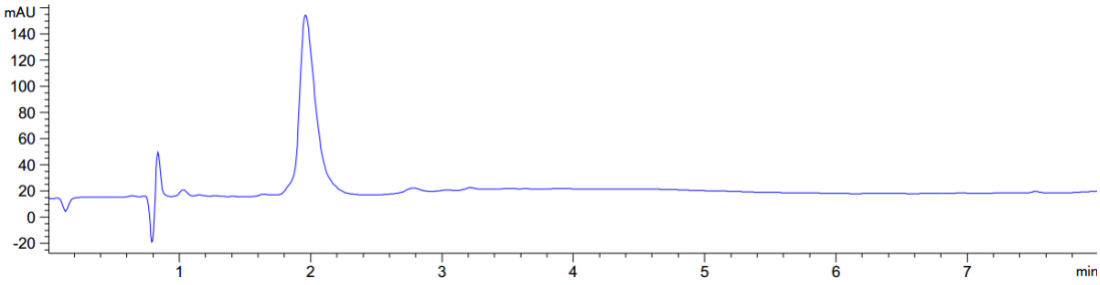


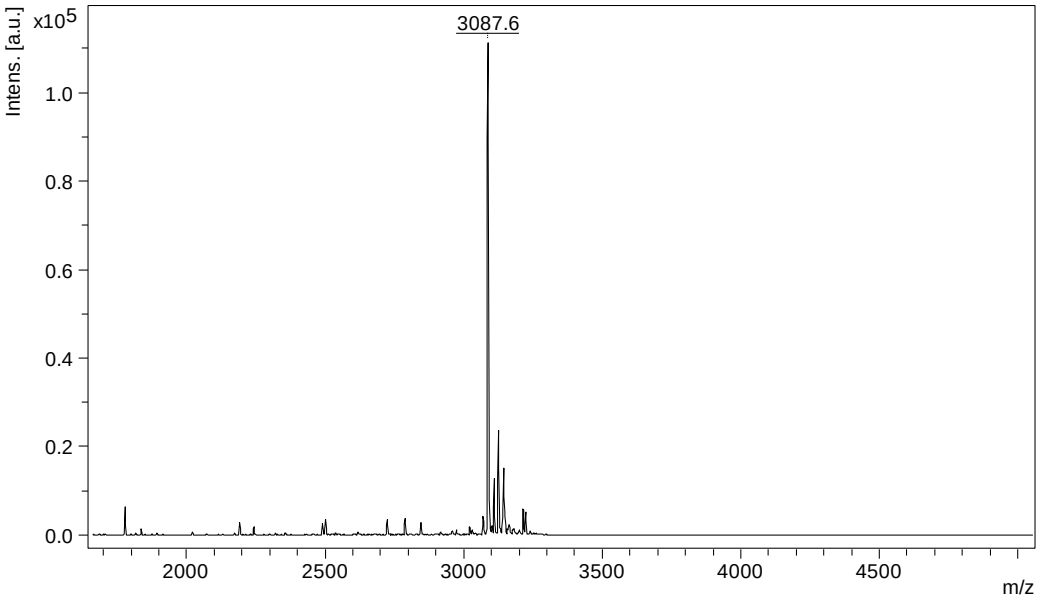
c:D7K YGKOGP**K**GPDGPKGKOGPKGKOGKOGKOGKOG



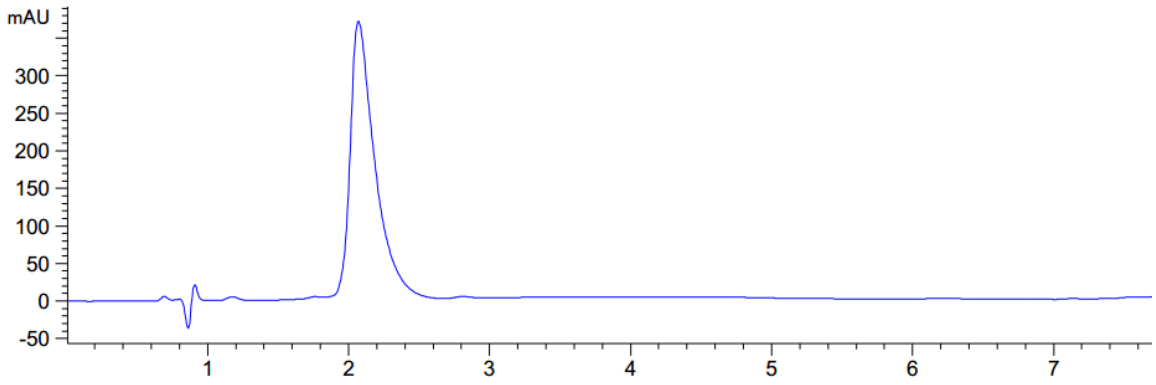


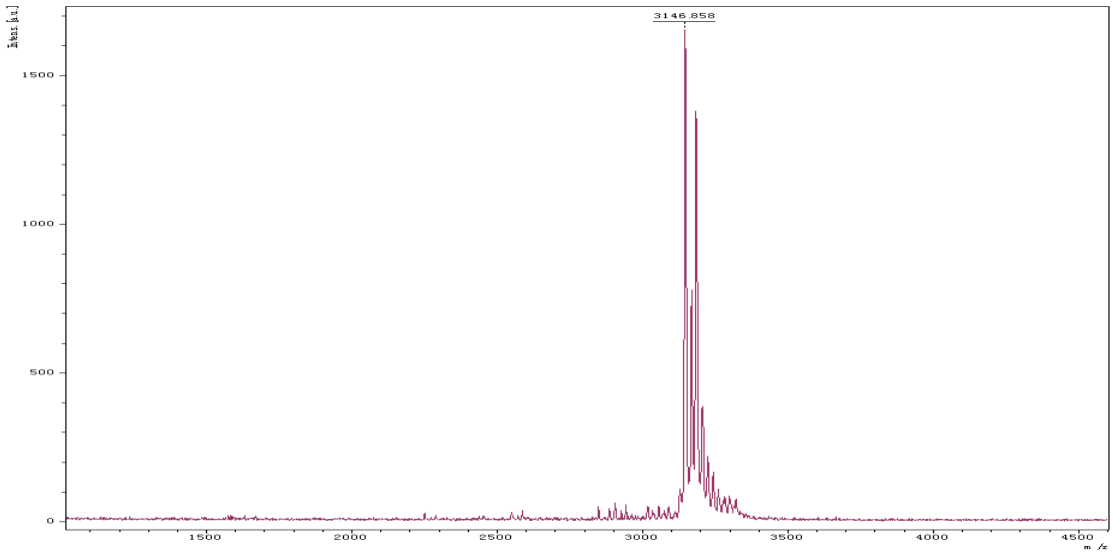
c:D10A YGKOGPDGP**A**GPKGKOGPKGKOGKOGKOGKOG



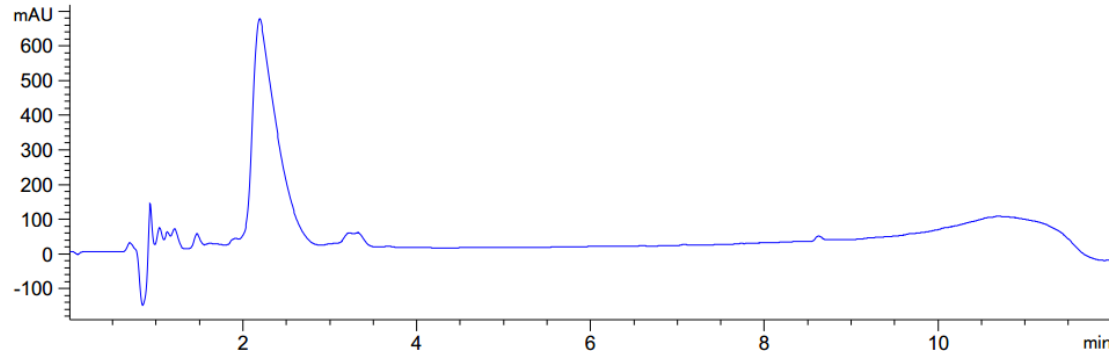


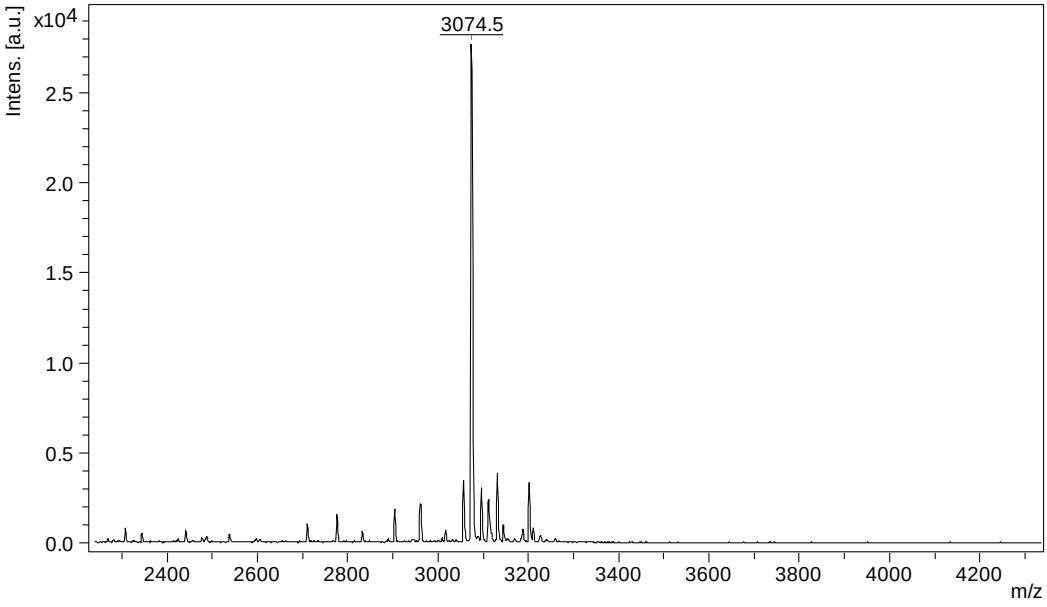
c:D10K YGKOGPDGP**K**GPKGKOGPKGKOGKOGKOGKOG





c:K15A YGKOGPDGPDGPKG**A**OGPKGKOGKOGKOGKOG





c:K15D YGKOGPDGPDGPKG**D**OGPKGKOGKOGKOGKOG