Electronic Supplementary Information

A computationally designed β-amino acid-containing miniprotein

Magdalena Bejger,[b]+ Paulina Fortuna,[a,c]+ Magda Drewniak-Świtalska,[a] Jacek Plewka,[d] Wojciech Rypniewski,[b] Łukasz Berlicki*[a]

[a] Dr. P. Fortuna, Dr. M. Drewniak-Świtalska, Prof. Ł. Berlicki
Department of Bioorganic Chemistry
Wrocław University of Science and Technology
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
E-mail: lukasz.berlicki@pwr.edu.pl

[b] Dr. M. Bejger, Prof. W. Rypniewski
Institute of Bioorganic Chemistry,
Polish Academy of Sciences,
Noskowskiego 12/14, 61-704 Poznań, Poland

[c] Dr. P. Fortuna,
Department of Medical Biochemistry,
Wrocław Medical University,
Pasteura 1, 50-368 Wroclaw, Poland

[d] Dr. J. Plewka
Faculty of Chemistry
Jagiellonian Univeristy
Gronostajowa 2, 30-387 Kraków, Poland

Table of contents

Peptide design

Peptides were designed using Rosetta software (v 3.8) using Rosetta Scripts interface.[S1] The initial conformation of the backbone was created by joining two α-helical fragments and one helical fragment build using ααβαααβ sequence pattern with application of (1*S*,2*S*)-2-aminocyclopentanecarboxylic acid. The dihedral angles of the backbone of β-amino acid-containing ware copied from published crystal structure containing fragment of the same pattern (PDB id 3F4Z).[S2] The *resfile* was created assuming that side chains forming the hydrophobic core are chosen from hydrophobic (one of I, L, F, Y, W, V, M), while that designed to be on the surface are polar. The FastDesign protocol[S3] was applied using beta_nov15 scoring function and 5000 iterations.

Peptide synthesis and purification

All commercially available reagents and solvents were purchased from Sigma-Aldrich, Merck, Iris Biotech or Bachem and used without further purification. Peptides were obtained with an automated solid-phase peptide synthesizer (Biotage® Initiator+Alastra™) on H-Rink amide ChemMatrix® resin (loading: 0.59 mmol/g). Fmoc deprotection was done using 20% piperidine in DMF for 3 + 10 min at room temperature. A double-coupling procedure for α-amino acids (15 min at 75 °C) was performed with 5eq of Fmoc amino acid derivative, 0.5 M solution of DIC and 0.5 M solution of OXYMA (1:1) in DMF. While β-amino acids were coupled for 30 min at 75 °C (followed by 30 min at room temperature) using 3 eq of Fmoc-trans-ACPC, 0.5 M solution of DIC and 0.5 M solution of OXYMA (1:1) in DMF. Cleavage of the peptides from the resin was accomplished with the mixture of TFA/TIS/H_2O (95:2.5:2.5) during 3 h of shaking at room temperature. The crude peptide was precipitated with ice-cold diethyl ether and centrifuged (14 500 rpm, 2 x 5 min, 4 °C). Obtained crude peptides were purified using the HPLC (Knauer Prep) with

a preparative column 250 mm x 30 mm, Thermo Scientific™ Hypersil GOLD™ (C18, 12µm) and analyzed using an analytical column 150 mm x 4.6 mm, Kinetex 100A (C18, 5µm). Solvents and gradients are given below in results section (Table S1). Peptides were analyzed by WATERS LCT Premier XE mass spectrometry system consisting of high resolution mass spectrometer with a time of flight (TOF) analyzer. All peptide showed purity >98% as estimated by analytical HPLC (Table S1).

Circular dichroism

CD spectra were recorded on JASCO J-815 at 20 °C between 300 and 180 nm in 50 mM phosphate buffer (pH 7.0) using following parameters: 0.2 nm resolution, 1.0 nm band width, 20 mdeg sensitivity, 0.25 s response, 50 nm/min scanning speed, 10 scans, 0.02 cm cuvette path length. The CD spectra of the solvent alone were recorded and subtracted from the raw data. The CD intensity is given as mean residue molar ellipticity (θ [deg x cm^2 x $dmol^{-1}$]).

To examine the thermal unfolding of the peptide, stock solutions were prepared containing 0.2 mg/mL peptide in 20 mM potassium phosphate buffer, pH 7.0, with concentrations of guanidine hydrochloride ranging from 0 M to 6.0 M with 0.5M interval. The temperature was increased from 4 to 96 ºC in increments of 2 ºC. Ellipticity measurements were recorded at 222 nm. The determination of T_m value was based on nonlinear fitting of the following function[S4] to data obtained at guanidine hydrochloride concentration equals to 0.0 M:

$$\theta_{obs} = \frac{1}{1 + e^{\frac{-\Delta H_m \cdot \left(1 - \frac{T}{T_m}\right)}{RT}}} \cdot (b_f - b_u - m_u \cdot T + m_f \cdot T) + b_u + m_u \cdot T$$

Where: θ_{obs} – observed values of ellipticity; T – temperature, ΔH_m – entalphy; R – gas constant; T_m – melting point temperature, b_u and m_u – constants describing unfolded state, b_u and m_u – constants describing folded state.

Unfolding of peptides were monitored using CD spectra at 222 nm. Raw data were fitted to previously reported equations[S5] using nonlinear regression in Matlab R2016a (The MathWorks, Inc.). The observed elipticity (Q_{obs}) is dependent on equilibrium constant K and ellipticities of folded (Q_f) and unfolded states (Q_u):

$$Q_{obs} = \frac{1}{1+K}(Q_u * K + Q_f)$$

Where Q_f and Q_u are dependent on both temperature and concentration of denaturant:

$$Q_u = a + bT + c[GuHCl]$$

$$Q_f = d + eT + f[GuHCl]$$

Equilibrium constant (K) is related to peptide folding free energy (ΔG):

$$K = e^{\frac{-\Delta G^0}{RT}}$$

While ΔG is given by formula:

$$\Delta G = \Delta H^0 - T\Delta S^0 + \Delta C_p\left(T - T_0 + Tln\left(\frac{T_0}{T}\right)\right) - m[GuHCl]$$

Crystallization, data collection, processing, structure determination and refinement

Prior to crystallization, purified peptide powder was dissolved in distilled water to the concentration of 50mg/ml. Crystallization conditions were screened using the Gryphon crystallization robot (Art Robbins

Instruments). This was followed by optimization of the conditions by the hanging-drop vapor-diffusion method. Crystallization drops were mixed with the precipitant solution in different volume ratios: from 1:1 up to 1:4. All crystals were grown at 19°C. The successful crystallization conditions were: 100 mM HEPES pH 7.5, 200 mM $MgCl_2$, 25%(w/v) PEG 3350 (HAMPTON crystallization screen INDEX (MD1-37)) for peptide **1**, and 100 mM Tris pH 8.5, 200 mM $MgCl_2$, 20%(w/v) PEG 8000 (MDL crystallization screen JCSG (HR2-134)) for peptide **1-Se**.

X-ray diffraction data were collected using synchrotron radiation at beamline P13, operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany).[S6] The crystals were flash-cooled at 100 K in a cold nitrogen-gas stream without additional cryoprotection.

The diffraction data were indexed, integrated and scaled using XDS.[S7] The initial structure of the peptide was solved by the SAD method using selenomethionine derivative and the HKL2MAP interface[S8] for phasing with the ShelxCDE programs.[S9] Initial model of the peptide was further built in ARP/wARP.[S10] The structures of the peptide **1** and peptide **1-Se** were solved by molecular replacement using Phaser.[S11] Refinement of both structures was done in Refmac[S12] from the CCP4 suite.[S13] Manual model rebuilding according to electron-density maps was performed in Coot.[S14] Structural figures were prepared in PyMOL.[S15] The data collection, processing and refinement statistics are summarized in Table S3.

sr-SAXS measurements

SAXS measurements were performed in batch mode at 1, 2.5 and 5 mg/mL concentrations at ESRF, Grenoble, France beamline BM29. Scattering images were recorded for 1 s using a 2D Pilatus 1M at a photon energy of 12.5 keV (λ=0.99 Å) at 2.8 sample-to-detector distance. Four frames were merged, integrated and background subtracted on the matching buffer. For a better signal-to-noise ratio signals from 1 and 5 mg/mL were merged with each other using Primus software.[S16] Final results were analyzed using ATSAS package software.[S17] For the molecular envelope determination a 10 round refinement in DAMMIF was applied using P2 symmetry.[S18]

Table S1. Rosetta-designed sequences and their scores beta-Nov15

No	Sequence	Score
1	LSEEEIQRIFGLSSEQIKSLPEEXYKKXVEXTGYL	-131.718
2	LSEEEIQKVFGMSKEELSSLPEEXLKKXIEXKGYL	-131.598
3	YSEEEIQRLLGISKEQFKSLPEEXVKKXVEXSGVL	-130.219
4	VSEEEIQRIFGMSSEQLKSLPEEXFKKXVKXKGYV	-130.264
5	LSEEQIQKLLGMSKEQFKSLPEEXVRKXVEXSGYL	-130.203

Table S2. MS and HPLC data for studied peptides

No	MS		Preparative HPLC		Analytical HPLC	
	Calc.	**Exp.**	**t_R [min]**	**gradient**	**t_R [min]**	**gradient**
1	816.2403 1020.0484	816.2432 1020.0397	17.0	0-5min 90%A, 10-18min 50%A, 28min 35%A, 30-33min 10%A, 35min 90%A, flow: 10ml/min	8.93	0-2min 90%A, 11-12min 10%A, 15min 90%A, Flow: 1ml/min
2	806.2495 1007.5599 1343.0773	806.2462 1007.5557 1343.0764	25.6	0-5min 90%A, 10min 60%A, 18min 55%A, 28min 40%A, 30-33min 10%A, 35min 90%A, flow: 10ml/min	9.01	0-2min 90%A, 11-12min 10%A, 15min 90%A, Flow: 1ml/min
3	671.8743 806.0476 1007.3076	671.8719 806.0536 1007.3056	21.1	0-5min 90%A, 10min 60%A, 18min 55%A, 28min 40%A, 30-33min 10%A, 35min 90%A, flow: 10ml/min	8.36	0-2min 90%A, 11-12min 10%A, 15min 90%A, Flow: 1ml/min
4	812.6550 1015.5668 1353.7531	812.7979 1015.5552 1353.7863	18.2	0-5min 90%A, 10min 60%A, 18min 55%A, 28min 40%A, 30-33min 10%A, 35min 90%A, Flow: 10ml/min	8.62	0-2min 90%A, 11-12min 10%A, 15min 90%A, Flow: 1ml/min
5	808.2528 1010.5655 1347.0847	808.2524 1010.5865 1347.0864	23.9	0-5min 90%A, 10min 60%A, 18min 55%A, 28min 40%A, 30-33min 10%A, 35min 90%A, Flow: 10ml/min	8.92	0-2min 90%A, 11-12min 10%A, 15min 90%A, Flow: 1ml/min
1-Se	829.2205 1036.2737 1381.3623	829.2184 1036.2714 1381.3448	14.8	0-5min 90%A, 10min 50%A, 18min 50%A, 28min 38%A, 30-33min 10%A, 35min 90%A, Flow: 10ml/min	8.60	0-2min 90%A, 11-12min 10%A, 15min 90%A, Flow: 1ml/min

Table S3. The crystallographic data collection, processing and refinement statistics

Peptide no	**1-Se**	**1**
PDB code	7ARS	7ARR
Space group	C2	P1
Cell constants *a, b, c, α, β, γ* [Å, °]	51.1, 35.5, 41.9 90.0, 111.2, 90.0	30.9, 32.1, 43.1 114.2, 91.9, 109.9
Resolution (Å)	39.0 - 1.15	29.0 - 1.10
Crystal mosaicity (°)	0.1	0.4
Unique reflections	24712	49390
Multiplicity	6.4	3.5
% Data completeness	98	88
R_{merge}	0.055	0.031
$\langle I/\sigma(I) \rangle$ in highest resolution bin	1.06	7.25
Number of molecules in ASU	2	4
Total number of atoms	835	1585
R_{free} test set	1001 reflections (4%)	998 reflections (2%)
R/R_{free}	0.174/0.220	0.129/0.161
R.m.s.d. - bond lengths (Å)	0.020	0.024
R.m.s.d. - bond angles (º)	2.6	2.8
Ramachandran outlies	0	0
Average B-factor, all atoms ($Å^2$)	21	16
Wilson B-factor ($Å^2$)	14	11
Anisotropy	0.498	0.568
Bulk solvent K_{solv} ($e/Å^3$), B_{solv} ($Å^2$)	0.34 , 47	0.30 , 47
F_o ,F_c correlation	0.96	0.98

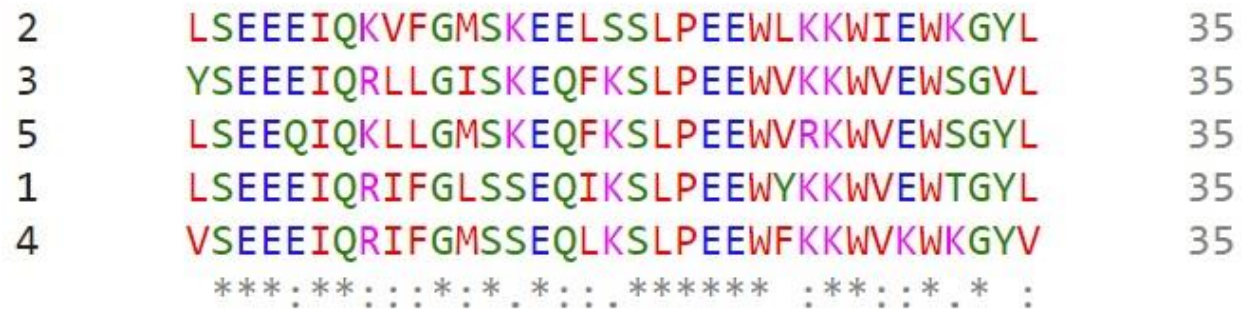


Figure S1. Colored multiple sequence alignment of peptides **1-5** done using CLUSTAL O(1.2.4).[S20]

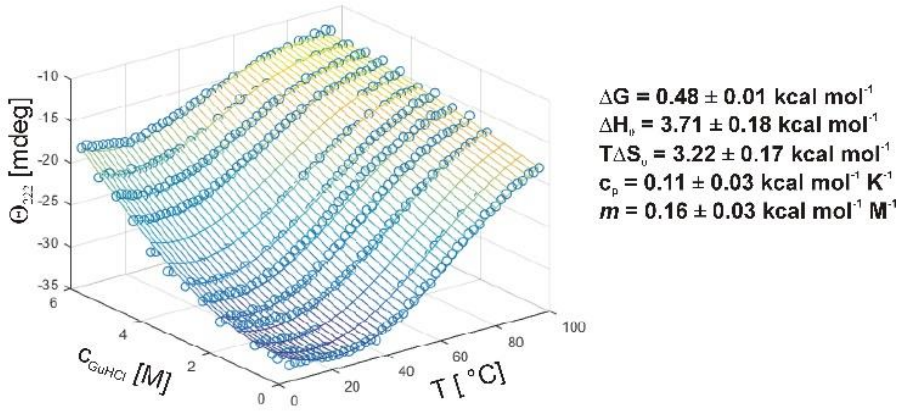


Figure S2. Circular dichroism (ellipticity at 222 nm) as a function of temperature and guanidine hydrochloride (denaturant) concentration for peptide **1**. Experimental data are shown as blue open circles, while the surface represents the fitted model with thermodynamic parameters (for 293 K) of the unfolding equilibrium is shown on the right.

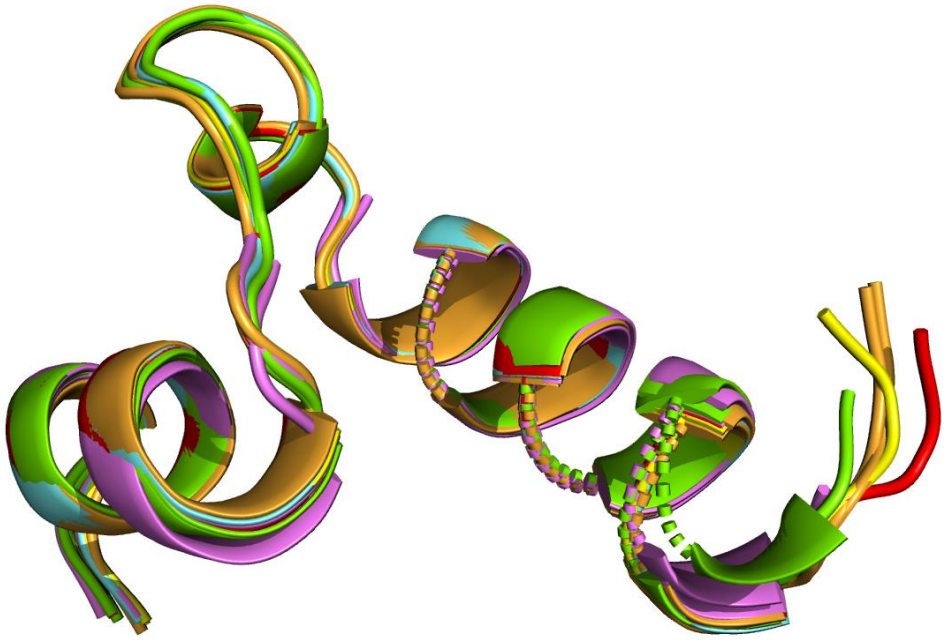
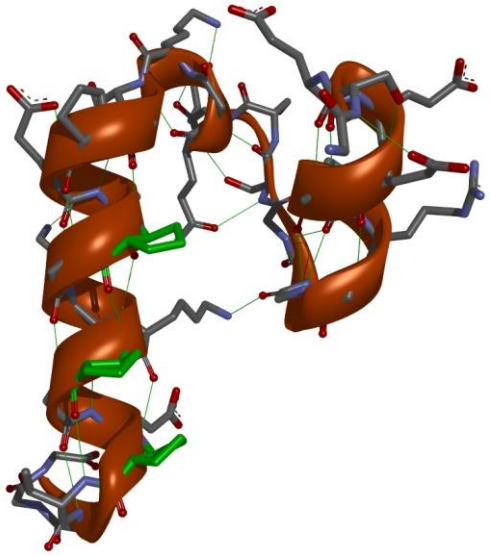
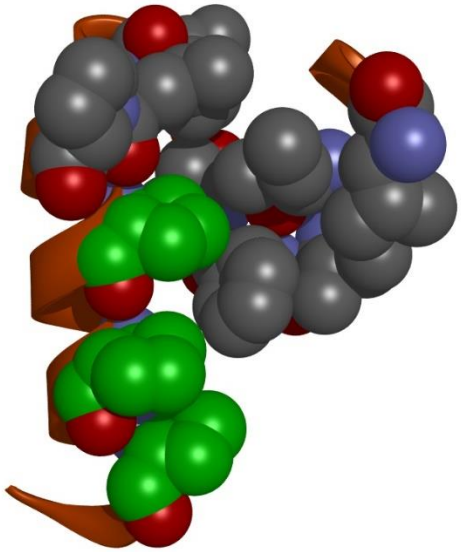


Figure S3. Superposition of all the crystallographically independent peptides, including double conformations: peptide **1** chains A (*yellow*), B (*red*), C (*blue*), D (*green*); peptide **1-Se** chains A (*orange*) and B (*violet*).

A



B



C

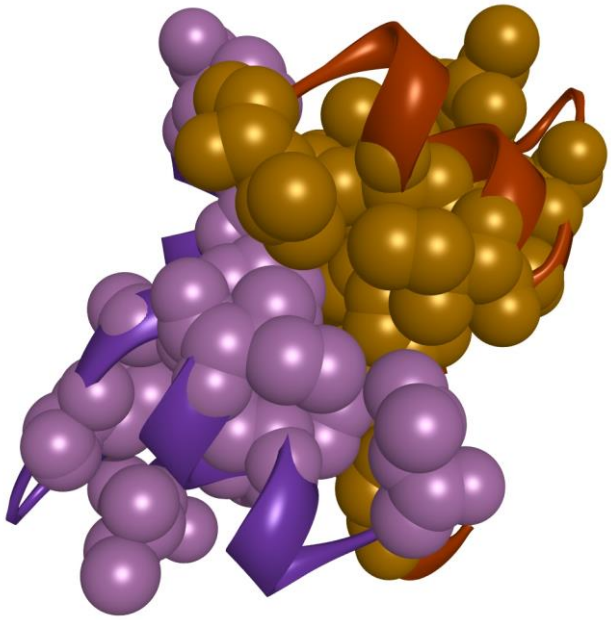


Figure S4. The crystal structure of peptide **1**. The hydrogen bond network (thin green lines, A) and the packing of the hydrophobic core (space-filling spheres, B) in the monomer. Carbon atoms of the β-amino acid residues are shown in green. The hydrophobic residues' interface (space-filling spheres, C) in the dimer. The main chain of peptide **1** is shown as a solid ribbon.

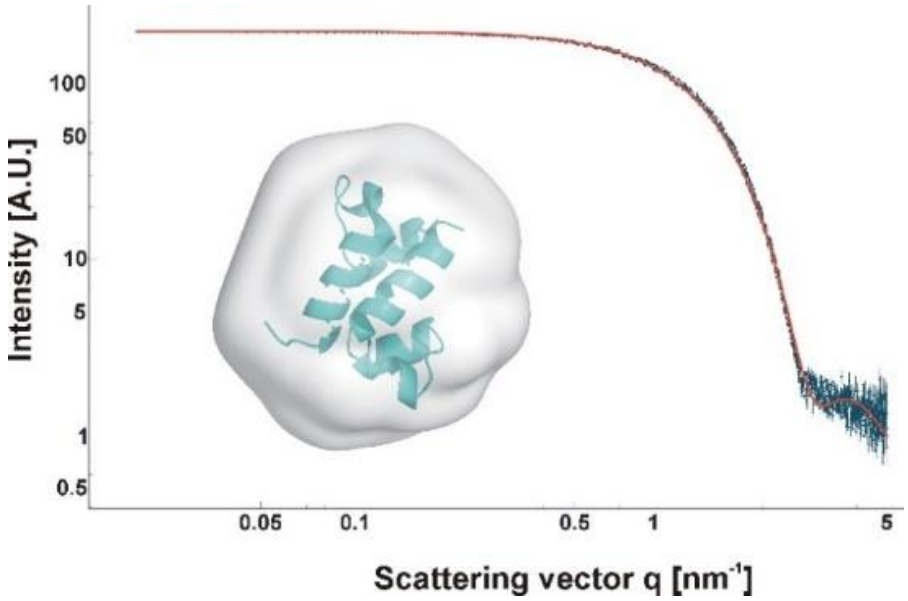


Figure S5. Overlay of the experimental scattering curve from peptide **1** (blue points) and the putative signal from the miniprotein dimer (red line). The inset is a superimposed structure of the **1** dimer in a molecular envelope calculated from the experimental sr-SAXS curve.

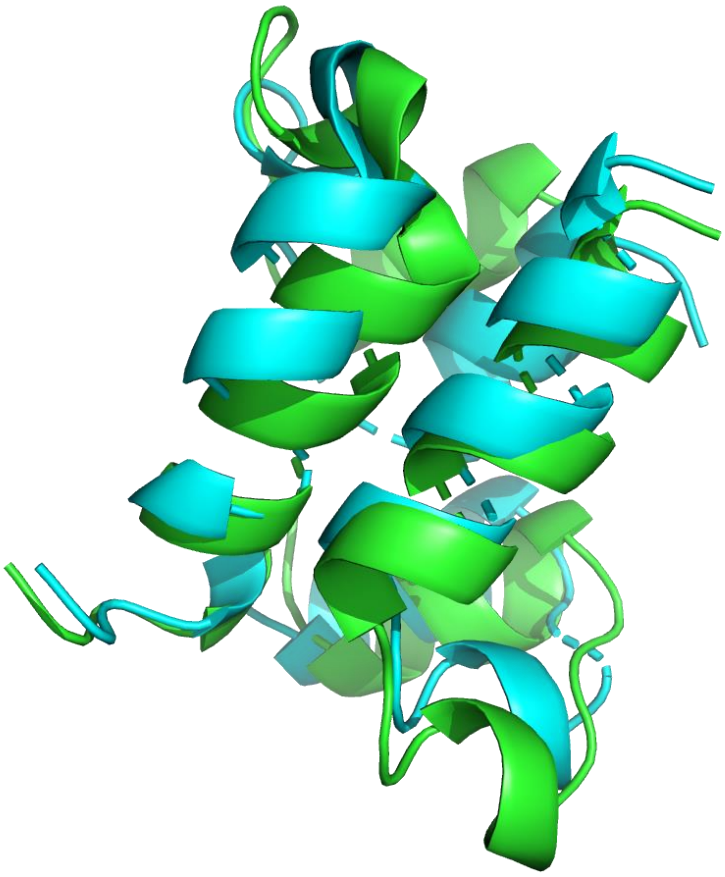
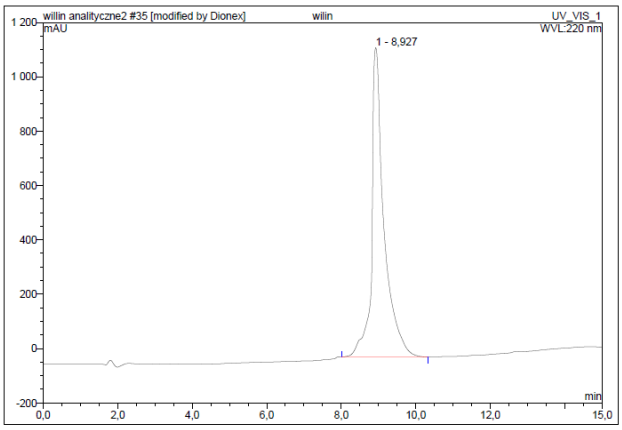


Figure S6. The comparison of the miniprotein **1** dimer (cyan) and its structure modified to fit the experimental SAXS curve better (green). The modelling of subunits suggests that the dimer is slightly less compact in the solution as compared to the crystal.

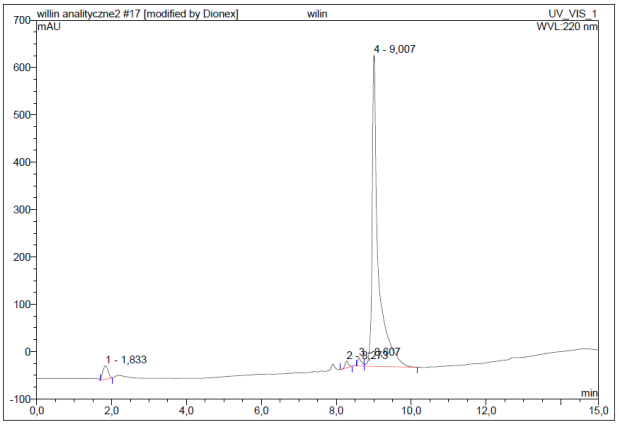
Table S4. Root-mean-square deviation (r.m.s.d.) in Å for pairwise superpositions of C_α atoms of the peptide **1** chains (peptide **1** chains: A, B, C, D and peptide **1-Se** chains: A, B), where double conformation residues are separated (conformations: a and b for each chain). Superpositions were analyzed using the program Align.[S19] The program did not take into account C_α atoms from β-amino acid residues (24XCP, 28XCP, 31XCP). Numbers of C_α atom pairs included in the comparison are given in brackets.

peptide - chain -- conformation	N-A-b	N-B-a	N-B-b	N-C-a	N-C-b	N-D-a	N-D-b	Se-A-a	Se-A-b	Se-B-a	Se-B-b
N-A-a	0.07 (32)	0.58 (32) 0.36 (31)	0.58 (32) 0.36 (31)	0.20 (30)	0.16 (30)	0.49 (30)	0.38 (30)	0.27 (31)	0.30 (31)	0.81 (24)	0.23 (24)
N-A-b	- - -	0.57 (32) 0.35 (31)	0.57 (32) 0.35 (31)	0.18 (30)	0.14 (30)	0.48 (30)	0.36 (30)	0.27 (31)	0.30 (31)	1.16 (24)	0.21 (24)
N-B-a	- - -	- - -	0.03 (32) 0.00 (31)	0.30 (30)	0.26 (30)	0.54 (30) 0.27 (29)	0.23 (30)	0.49 (31) 0.42 (30)	0.32 (31)	0.62 (24)	0.22 (24)
N-B-b	- - -	- - -	- - -	0.29 (30)	0.26 (30)	0.55 (30) 0.28 (29)	0.23 (30)	0.48 (31)	0.32 (31)	0.62 (24)	0.22 (24)
N-C-a	- - -	- - -	- - -	- - -	0.25 (30)	1.08 (30) 0.47 (29)	0.35 (30)	0.25 (30)	0.30 (30)	0.79 (24)	0.20 (24)
N-C-b	- - -	- - -	- - -	- - -	- - -	0.99 (30) 0.33 (29)	0.37 (30)	0.32 (30)	0.31 (30)	0.78 (24)	0.17 (24)
N-D-a	- - -	- - -	- - -	- - -	- - -	- - -	0.99 (30) 0.30 (29)	0.61 (30)	0.45 (30) 0.36 (29)	0.91 (24) 0.53 (23)	1.17 (24) 0.32 (23)
N-D-b	- - -	- - -	- - -	- - -	- - -	- - -	- - -	0.46 (30)	0.30 (30)	0.55 (24)	0.32 (24)
Se-A-a	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	0.43 (31)	0.93 (24)	0.32 (24)
Se-A-b	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	0.67 (24)	0.29 (24)
Se-B-a	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	0.68 (24)

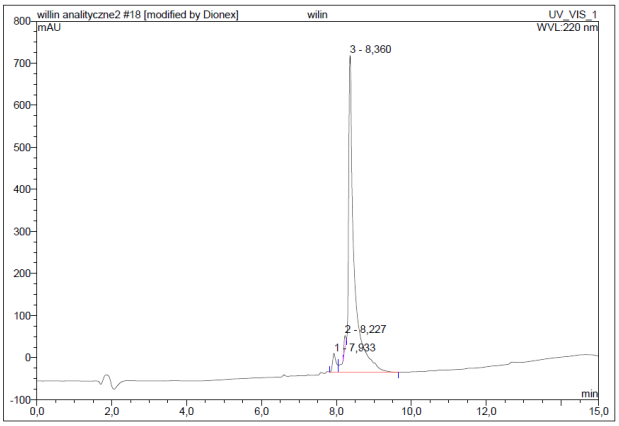
1



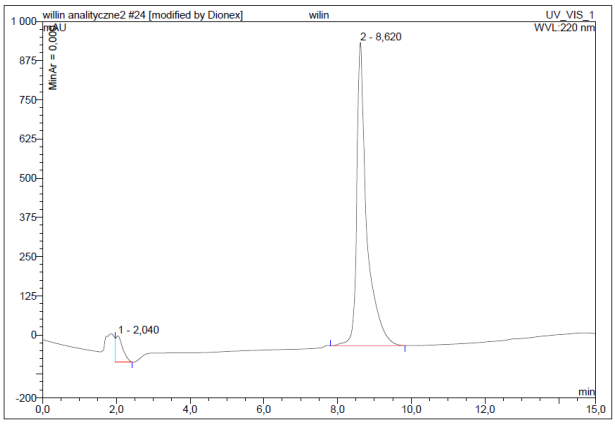
2



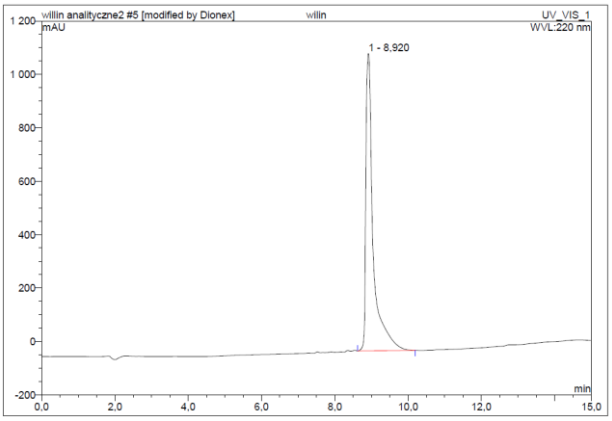
3



4



5



1Se

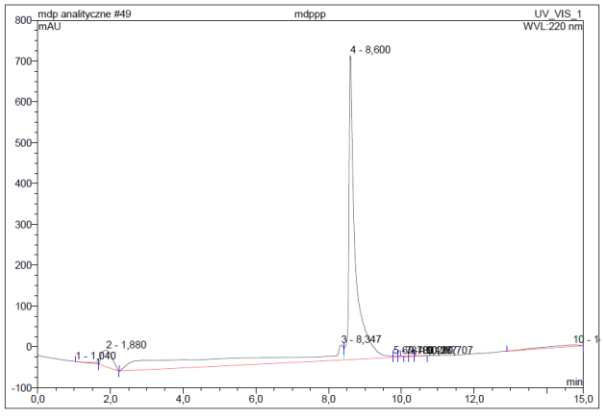
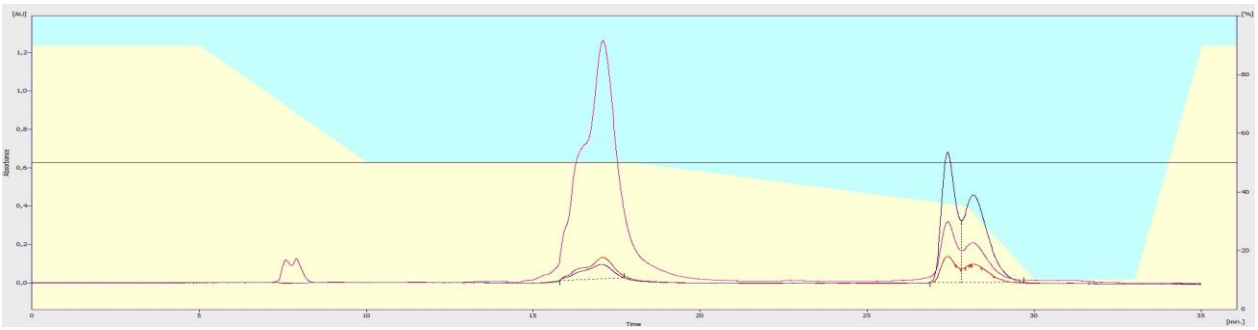
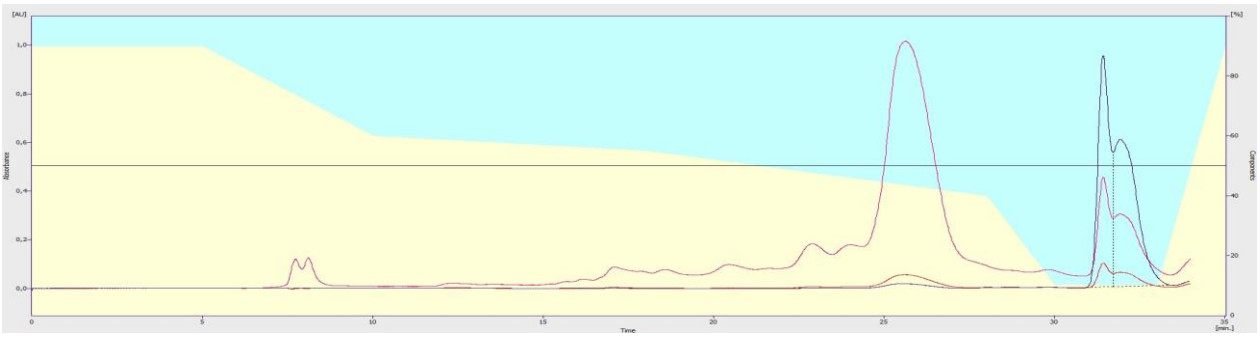


Figure S7. Analytical chromatograms of purified peptides **1-5** and **1Se** (absorbance measured at 220 nm).

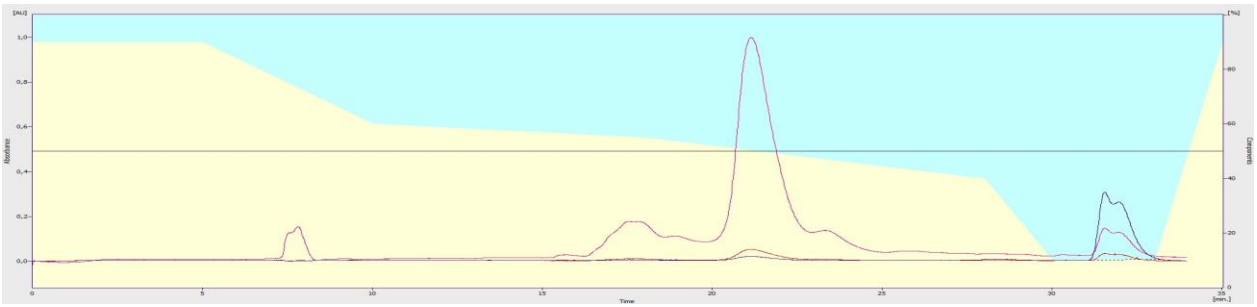
1



2



3



4

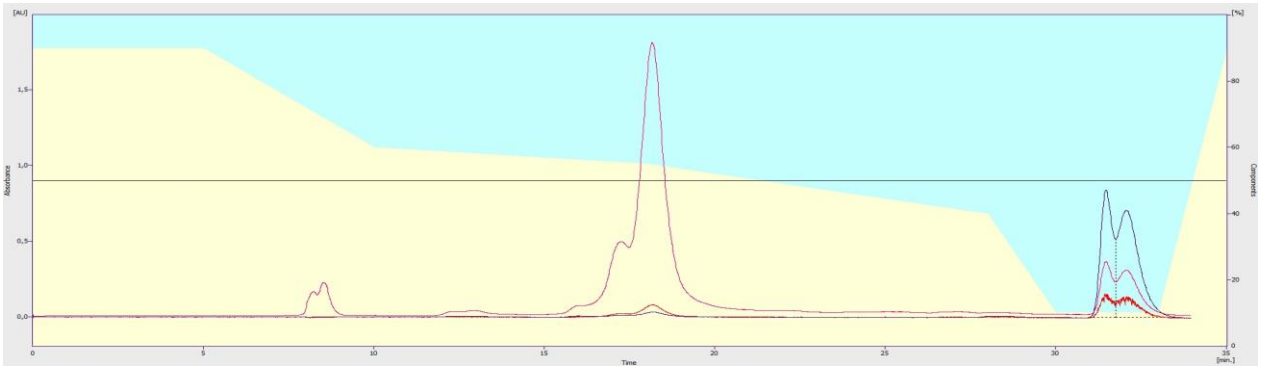
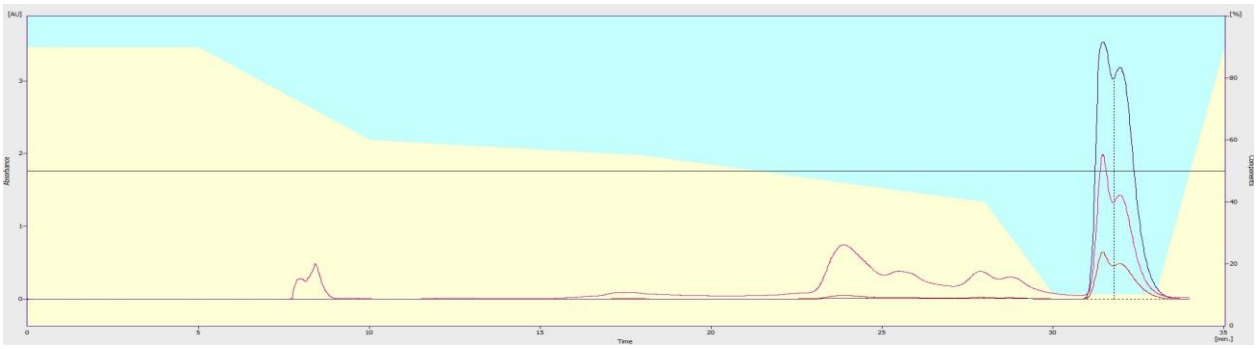


Figure S8. Preparative chromatograms of crude peptides **1-5** and **1Se** (blue, pink and red lines indicate absorbance measured at 254nm, 220nm and 280 nm)**.**

5



1Se

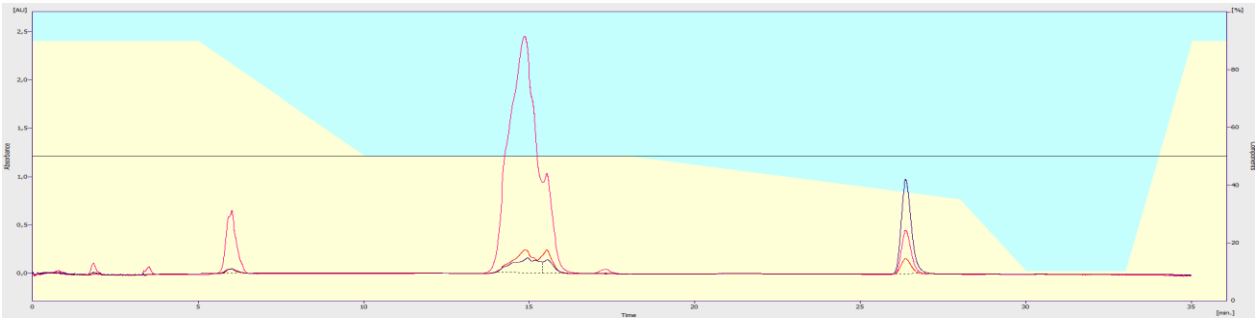


Figure S8 (CONTINUED). Preparative chromatograms of crude peptides **1-5** and **1Se** (blue, pink and red lines indicate absorbance measured at 254nm, 220nm and 280 nm)**.**

References

S1. S. J. Fleishman, A. Leaver-Fay, J. E. Corn, E. M. Strauch, S. D. Khare, N. Koga, J. Ashworth, P. Murphy, F. Richter, G. Lemmon, J. Meiler, D. Baker *Plos One* **2011**, *6*, e20161.

S2. W. S. Horne, L. M. Johnson, T. J. Ketas, P. J. Klasse, M. Lu, J. P. Moore, S. H. Gellman, *Proc Natl Acad Sci U S A* **2009**, *106*, 14751-14756.

S3. G. J. Rocklin, T. M. Chidyausiku, I. Goreshnik, A. Ford, S. Houliston, A. Lemak, L. Carter, R. Ravichandran, V. K. Mulligan, A. Chevalier, C. H. Arrowsmith, D. Baker, *Science* **2017**, *357*, 168-175.

S4. D. F. Kreitler, D. E. Mortenson, K. T. Forest, S. H. Gellman, *J. Am. Chem. Soc.* **2016**, *138*, 6498-6505.

S5. Z. E. Reinert, G. A. Lengyel, W. S. Horne, *J. Am. Chem. Soc.* **2013**, *135*, 12528-12531.

S6. M. Gerlach, U. Mueller, M. S. Weiss, *J. Large-scale Res. Facilities* **2016**, *2*, A47.

S7. W. Kabsch, *Acta Crystallogr D* **2010**, *66*, 125-132.

S8. T. Pape, T. R. Schneider, *J. Appl. Cryst.* **2004**, *37*, 843-844.

S9. G. M. Sheldrick, *Acta Crystallogr. D* **2008**, *64*, 112-122.

S10. G. Langer, S. X. Cohen, V. S. Lamzin, A. Perrakis, *Nat. Protoc.* **2008**, *3*, 1171-1179.

S11. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Cryst.* **2007**, *40*, 658-674.

S12. G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 355–367.

S13. M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson *Acta. Cryst. D* **2011**, 67 , 235-242.

S14. P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, *Acta Crystallographica.* D **2010**, *66*, 486–501.

S15. W. L. DeLano, PyMOL. **2002** DeLano Scientific, San Carlos, CA, 700.

S16. P.V. Konarev, V.V. Volkov, A.V. Sokolova, M.H.J. Koch D. I. Svergun *J. Appl. Cryst.* **2003**, *36*, 1277-1282.

S17. D. Franke, M. V. Petoukhov, P. V. Konarev, A. Panjkovich, A. Tuukkanen, H. D. T. Mertens, A. G. Kikhney, N. R. Hajizadeh, J. M. Franklin, C. M. Jeffries, D. I. Svergun, *J. Appl. Cryst.* **2017**, *50*, 1212-1225.

S18. D. Franke, D. I. Svergun, *J. Appl. Cryst.* **2009**, *42*, 342-346.

S19. G. H. Cohen, *J. Mol. Biol.* **1986**, *190*, 593-604

S20. F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson, D.G. Higgins, *Mol. Syst. Biol.* **2011**, *7*, 539.