Supporting Information for

Shape-specific Nanostructured Protein Mimics from *de novo* Designed Chimeric Peptides

Linhai Jiang, Su Yang, Reidar Lund, He Dong*

*Corresponding author: E-mail: hdong@clarkson.edu

Materials

Fmoc protected amino acids and 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) were purchased from Novabiochem. N,N-Diisopropylethylamine (DIPEA), acetic anhydride, triisopropylsilane (TIS), piperidine, 1,2-ethanedithiol (EDT), 2,2'-dithiodipyridine, α-lactalbumin, albumin from chicken egg white (ovalbumin), bovine serum albumin (BSA) and catalase were purchased from Sigma-Aldrich. N,N-dimethylformamide (DMF), trifluoroacetic acid (TFA), diethyl ether, dithiothreitol(DTT), methanol, Tris buffer (1M, pH=7.5), sodium chloride, monosodium phosphate, disodium phosphate and glycine were purchased from Fisher Scientific. Proto gel (30%, w/v), N,N,N',N'-tetramethylethylene diamine (TEMED) and ammonium persulfate was purchased from National Diagnostics. All the materials were used as received.

Peptide synthesis

The peptides used in this study were synthesized through Fmoc-solid phase peptide synthesis method on a PS3 automatic synthesizer at a scale of 0.05 mmol. 26r and P26 were synthesized on MBHA rink-amide resins using DMF as solvents. Piperidine (25% by volume) was used to deprotect the Fmoc group followed by animo acid coupling using HBTU as the coupling reagents in the presence of DIPEA. HBTU and Fmoc protected amino acids were added in 5 equivalents and DIPEA in 10 equivalents to the amino groups on the resin. Acetylation was done in the presence of 50 equivalents of acetic anhydride and 6 equivalent of DIPEA. 5 ml of cocktail solution containing trifluoroacetic acid(TFA)/triisopropylsilane/water/ethanedithiol (91/3/3/3 by volume) was prepared to cleave all peptides from the MBHA resin. After 3 hrs, solid resin was removed by gravity filtration and TFA was evaporated. Cold diethyl ether was added to the concentrated TFA solution to precipitate the peptides which were centrifuged and thoroughly washed by diethyl ether for four times. Residual diethyl ether was removed under vacuum. Crude peptides were purified by reverse phase HPLC using a binary mobile phase composed of water and acetonitrile in the presence of 0.05 % TFA by volume and was lyophilized for three days. Ligation between 26r and P26 was achieved as follows.¹ 10 ml of 26r solution (1mg/ml in 20 mM Tris buffer, pH=7.4) was mixed with 1 ml of 2,2'-dithiodipyridine methanol solution (1 mg/ml) to activate the thiol group. After stirring overnight, the excess of 2,2'-dithiodipyridine was removed by extraction using diethyl ether for three times. The aqueous layer containing activated 26r was

purified by HPLC and lyophilized for three days. Activated 26r solution (1 mg/ml in 20 mM Tris buffer, pH=7.4) was mixed with equal volume of P26 solution (1 mg/ml in 20 mM Tris buffer, pH=7.4) to react overnight. The chimeric peptide conjugate P26-26r was purified by HPLC and lyophilized for three days. The molecular weight of each peptide was determined by MALDI-TOF mass spectrometry (Bruker) using α -cyano-4-hydroxycinnamic acid as the matrix.

Native Polyacrylamide Gel Electrophoresis (PAGE) and Ferguson plot analysis

As shown on **Table S1**, polyacrylamide gels with five different resolving gel concentrations were prepared 24 hrs before the experiments. Each gel contains 4% of stacking gel on top of the resolving gel. For every 10 ml of gel solution, 100 μ l of ammonium persulfate (10 wt% in DI water) and 10 μ l TEMED were added to induce the polymerization. 1.5 M TrisHCl (pH=8.8), 0.6 M TrisHCl (pH=6.8) and 0.192 M glycine-0.025 M TrisHCl (pH=8.3) were used as the resolving gel buffer, stacking gel buffer and running buffer, respectively.

Peptide samples were prepared by dissolving appropriate amounts of lyophilized peptide powder in sodium phosphate buffer (10 mM, pH=7.4). For 26r containing peptides, the concentration was accurately determined by UV spectroscopy ($\varepsilon_{280,Trp}$ =5560M⁻¹cm⁻¹). P26 does not have aromatic amino acids and the concentration was calculated based on the weight from balance reading. The concentration of each sample was adjusted to 100 µM and annealed for 10 minutes at 70°C. 5 µL of peptide or protein standard solution was pre-mixed with 5 µl of loading buffer. Each electrophoresis was run at 136 Volt for 2 hrs followed by Coomassie G-250 (SimplyBlueTM SafeStain) staining for 2 hrs and extensive washing by DI water. The gels were imaged by a Bio-Rad ChemiDocTM MP Imaging System. Migration distance of P26-26 assembly and each protein standard along the resolving gel was measured with Preview application. For any given sample, the relative mobility (\mathbf{R}_r) was defined as the ratio of its migration distance to that of the protein sample travelled to the bottom of the gel. Linear fitting of logarithmic \mathbf{R}_r versus gel concentration yielded retardation coefficient (K_r) which is the opposite of the slope of the linear fit. A calibration curve was obtained by linear fitting of the logarithmic of Kr vs logarithmic of M_w , from which the molecular weight of P26-26r assembly was determined.

Hydrodynamic Analysis

The sedimentation coefficient of P26-26r assembly was determined by sucrose density gradient centrifugation. 10 mM sodium phosphate buffer was used to prepare the sucrose gradient solution at 5 wt% on the top and 20 wt% at the bottom in a 5 ml centrifuge tube (Beckman #Part: 326819). 20 µl sample (1 mg/ml in 10 mM sodium phosphate buffer, pH=7.4) was gently placed on the top of sucrose gradient solution. P26r-26r assembly and protein standards were subject to centrifugation (10 krpm for 80 hrs at 10°C) in parallel. After centrifugation, a small hole was made at the bottom of the centrifuge tube by using a fine needle. 20 fractions were collected for each sample. All fractions were run on Native PAGE to identify the fraction number at which peptide or protein standards were most populated. Plot of that fraction number versus the sedimentation coefficient of protein standards allows for estimation of the sedimentation coefficient of the P26-26r assembly.

Gel permeation chromatography (elution buffer: 10 mM sodium phosphate buffer, pH=7.4) was used to determine the Stoke's radius of P26-26r assembly. Calibration curve was obtained by linear fitting of the Stoke's radius versus the elution volume of protein standards (BSA, ovalbumin, α -lactalbumin).

Small Angle X-ray Scattering (SAXS)

Samples for SAXS measurements were prepared by direct dissolution of peptide powders in 10 mM sodium phosphate buffer (pH=7.4) to reach a concentration in the range of 500 μ M to 600 μ M. For 26r and P26, 10X DTT was added to prevent disulfide bond formation. SAXS experiments were performed using the automated BM29 bioSAXS beamline at the ESRF, Grenoble, France. More technical details were referred to the literature.² Scattering data were obtained using an energy of 12.5 keV and a detector distance of 2.87 m covering a Q-range (Q= $4\pi \sin(\theta/2)/\lambda$, λ is the wavelength, θ is the scattering angle) of about 0.0047 Å⁻¹< Q < 0.5 Å⁻¹. The data were calibrated to absolute intensity scale using water as a primary standard. The pair-distribution function, p(r), was calculated from the experimental scattering data using the indirect Fourier transform method with the program GNOM in the ATSAS-package (EMBL, Hamburg).

Transmission Electron Microscopy (TEM)

TEM samples were prepared the same as those for CD measurement. 10 μ l of sample solution was dropped onto a TEM grid and leave it on for one minute. The excess solution was gently wicked off by filter paper. After air-drying for 30 seconds, 10 μ l negatively staining reagent (2 wt% uranyl acetate in deionized water) was added onto the grid and left on the grid for 1 minute. The excess of staining solution was wicked off by filter paper. The specimen was air-dried overnight before imaging on JEOL 2010 high-resolution TEM.

Circular Dichroism (CD) Spectroscopy

All samples used for CD experiments were prepared by dissolving certain amount of lyophilized peptide powder in sodium phosphate buffer (10 mM, pH=7.4). For 26r containing peptides, the concentrations were accurately determined by the UV absorption of tryptophan ($\epsilon_{280,Trp}$ =5560 M⁻¹cm⁻¹). The concentration of P26 was determined based on its balance weight. The concentration of each sample was adjusted to 100 µM and annealed for 10 minutes at 70°C. All samples were incubated at 4°C overnight before CD experiments.

CD spectra were acquired on Jasco-J715 spectropolarimeter between 250 nm and 190 nm at room temperature with a scanning rate 100 nm/min (response time: 2 seconds, bandwidth: 1 nm). The thermal unfolding profiles were measured by monitoring the ellipticity at 225 nm from 5°C to 95°C with a heating rate at 20°C per hour. Thermal unfolding profile of P26 was fitted to Boltzmann function. Melting temperature was defined as the peak on the first order derivative of the fitted curve.

Molecular Dynamics (MD) Simulation

The coiled-coil dimer from 26r was generated on PyMOL by placing two α -helices in parallel with the hydrophobic residues embedded at the helix-helix interface. The initial distance between the longitudinal axes of two helices was set to be 13 Å to ensure favorable electrostatic interactions between the glutamic acid at the g position on one strand and the lysine residue at the e position on the other strand.³ Crystal structure of the collagen triple helix structure based on [(PPG)₁₀]₃ (pdb ID:1K6F) was used as the template to build the initial structure of P26 trimer. Mutation was made on selected residues in accordance with the primary sequence of P26 ((POG)₂PCG(POG)₆) with an acetylated N-terminus and amidated C-terminus. Thermal average structures of 26r dimer and P26 trimer were used as the building blocks to construct the initial structure of the hexameric assembly for MD simulation. (1) Three 26r dimers were evenly distributed on the x-y plane of a trigonal bipyrimidal geometry with their N-terminus pointing to the center of the equatorial triangle and the residue at the f positions oriented along the z-axis. (2) One P26 trimer was positioned above the plane of the 26r dimers with its longitudinal axis at the z-axis and its N-terminus pointing down. (3) The dihedral angle of (POG)₂PCG segment at the Nterminus was iteratively randomized until the distance between the sulfur atoms on P26 and 26r is within favorable disulfide bond length at 2.05 Å. (4) Placement of the second P26 trimer followed the same steps, but as a mirror image to the first trimer. (5) The offset of the N-termini of coiled coil dimers from the equatorial triangle center were iteratively refined to generate a model where the generated SAXS profile match the experimental data to a maximum extent.

Desmond package⁴ (non-commercial version 2016.3) with OPLS_2005 force field was applied to carry out all molecular dynamic simulations in high performance computing environment (Orion in Clarkson University). Hydrogen atoms were added before solvation with TIP4P water model⁵⁻⁷ in an orthorhombic simulation box. The size of the simulation box was set to ensure that the initial minimum thickness of water layer between any atom on peptides and the nearest boundary of the simulation box was 10 Å. All glutamic acid residues were deprotonated and all lysine residues were protonated under neutral pH in 10 mM sodium phosphate buffer (pH=7.4). Proper amounts of sodium counter ions (6 for 26r dimer and 18 for hexameric assembly) were added to neutralize the solute.

For all simulations, a constant temperature was maintained by Nose-Hoover chain thermostat method with a relaxation time of 1 ps. The pressure was kept constant for NPT ensemble at 1.01 Bar by Martyna-Tobias-Klein method (a relaxation time of 2 ps and isotropic coupling style). A time step of 0.5 fs was used for energy integration. Particle Mesh Ewald method was used to calculate the electrostatic interaction under periodic boundary condition with a grid spacing of 1 Å. The cutoff for electrostatic interaction was 10 Å with a switching distance of 9 Å. The van der Waals interaction was represented by the classic Lennard-Jones potential with a gradual cutoff from 10 to 12 Å. Pair-list distance for both non-bonded interactions was 14 Å with updating time of 12 fs. Before each production simulation, the system was relaxed by performing (a) energy minimization through Brownian Dynamics (T=10 K) for 100 ps, (b) 12 ps simulation in NVT ensemble with restraints on solute heavy atoms, (c) 12 ps simulation in NPT (T=10 K, P=1.01 Bar) ensemble with restraints on solute heavy atoms, (d) 12 ps simulation in NPT (T=293 K, P=1.01 Bar) ensemble with restraints on solute heavy atoms, (f) 24 ps simulation in NPT (T=293 K, P=1.01 Bar) ensemble without restraints. Production simulation was performed in NPT (T=293 K, P=1.01 Bar) ensemble for several nanoseconds with 1.2 ps interval time for recording both trajectory and energy.

Gel %	Resolving gel buffer	Deionized water	ProtoGel (30%) solution
	(ml)	(ml)	(ml)
8	2.5	4.84	2.67
8.5	2.5	4.67	2.83
9	2.5	4.5	3
9.5	2.5	4.33	3.17
10	2.5	4.17	3.33
Stacking gel (4%)	2.5 (Stacking gel buffer)	6.17	1.33

Table S1. Formulation of native resolving polyacrylamide gel and stacking polyacrylamide gel.



Figure S1. (Left panel) HPLC of 26r, P26 and P26-26r. (Right panel) MALDI-TOF mass spectrometry of 26r, expected mass [M+H]⁺: 2984.48, observed mass: 2984.86, P26, expected mass [M+Na]⁺: 2477.74, observed mass: 2477.66, P26-26r, expected mass [M+H]⁺: 5437.22, observed mass: 5437.64.



Figure S2. Images of gel with different concentration after electrophoresis and staining. From left to right in each gel image: commercial protein marker (molecular weight in Dalton of each protein from bottom to top: 20k, 66k, 146k, 242k, 480k, 720k, 1048k), ovalbumin, P26-26r assembly and α -lactalbumin.



Figure S3. Native-PAGE analysis result of fractions from sucrose density gradient centrifugation.



Figure S4. Gel permeation chromatography result of P26-26r assembly and three protein standards.



Figure S5. Subtraction of the melting curve of 26r from P26-26r. Circle: experimental data; Solid line, sigmoidal fitting of experimental data; Dashed line, first order derivative of fitted sigmoidal curve.



Figure S6. Plot of root mean square deviation (RMSD) of P26-26r hexamer upon MD simulation.



Figure S7. Comparison of the Kratky plot of experimental SAXS data and that from the thermal average structure.



Figure S8. Plot of root mean square fluctuation (RMSF) of the peptide backbone of P26-26r hexamer during MD simulation.



Figure S9. Comparison of the thermal unfolding profiles of (a) 26r coiled-coil dimer, (b) P26-26r assemblies and (c) P26 collagen triple helices in 10 mM sodium phosphate buffer (pH=7.4) with/without 150 mM NaCl. Peptide concentration: 100 μ M. Wavelength monitored: 225nm.



Figure S10. Plot of root mean square deviation (RMSD) of 26r dimer upon MD simulation.



Figure S11. Plot of root mean square deviation (RMSD) of P26 trimer upon MD simulation.

References

- 1 M. Ruiz-Gayo, F. Alberxlo, M. Pons, M. Royo, E. Pedroso and E. Giralt, *Tetrahedron Lett.*, 1988, **29**, 3845-3848.
- 2 P. Pernot, A. Round, R. Barrett, A. De Maria Antolinos, A. Gobbo, E. Gordon, J. Huet, J. Kieffer, M. Lentini, M. Mattenet, C. Morawe, C. Mueller-Dieckmann, S. Ohlsson, W. Schmid, J. Surr, P. Theveneau, L. Zerrad and S. McSweeney, *J. Synchrotron Radiat.*, 2013, 20, 660-664.
- 3 G. Offer and R. Sessions, J. Mol. Biol., 1995, 249, 967-987.
- 4 K. J. Bowers, E. Chow, H. Xu, R. O. Dror, M. P. Eastwood, B. A. Gregersen, J. L. Klepeis, I. Kolossvary, M. A. Moraes, F. D. Sacerdoti, J. K. Salmon, Y. Shan and D. E. Shaw, *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida*, 2006, November 11–17.
- 5 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
- 6 P. S. Nerenberg and T. Head-Gordon, J. Chem. Theory Comput., 2011, 7, 1220-1230.
- 7 R. Shevchuk, D. Prada-Gracia and F. Rao, J. Phys. Chem. B, 2012, 116, 7538-7543.