

Supplementary Information

Title: Covalently-assembled single-chain protein nanostructures with ultra-high stability

Authors: Wenqin Bai¹, Cameron J. Sargent², Jeong-Mo Choi³, Rohit V. Pappu³, Fuzhong Zhang^{1, 2, 4*}

Affiliations:

¹Department of Energy, Environmental and Chemical Engineering,

²Division of Biological & Biomedical Sciences,

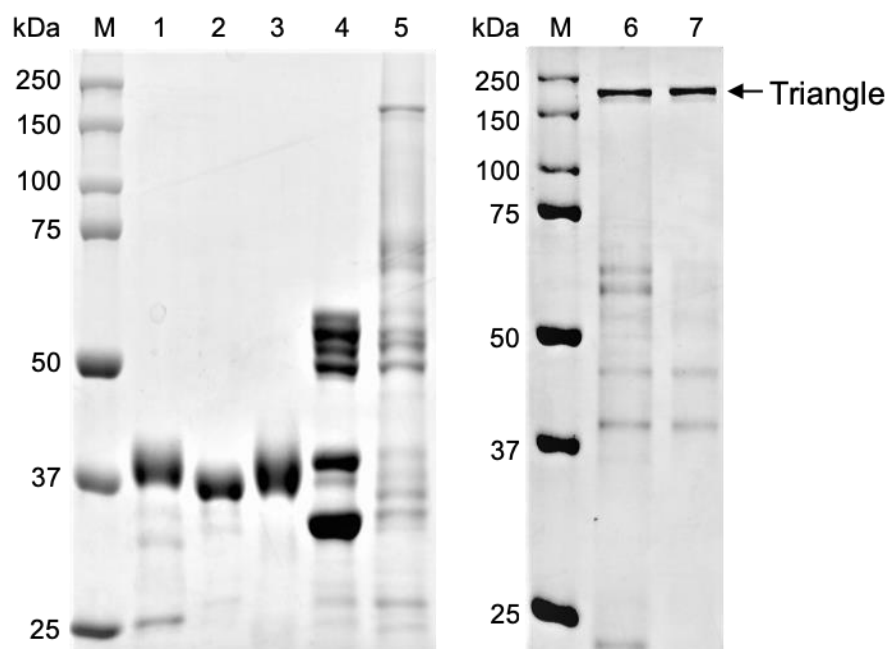
³Department of Biomedical Engineering and Center for the Science & Engineering of Living Systems,

⁴Institute of Materials Science & Engineering,

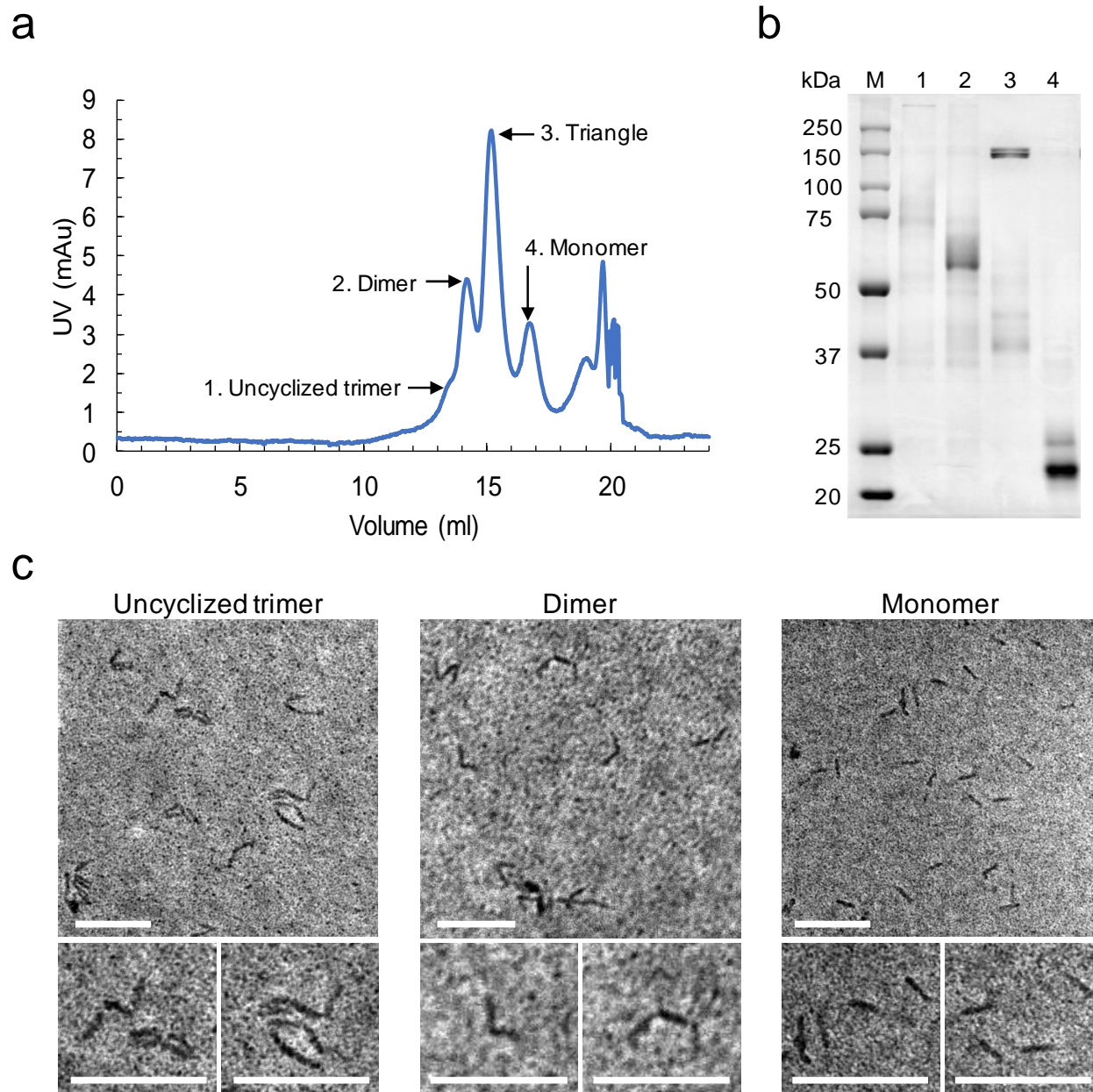
Washington University in St. Louis, Saint Louis, MO 63130, USA

*Correspondence to: fzhang@seas.wustl.edu

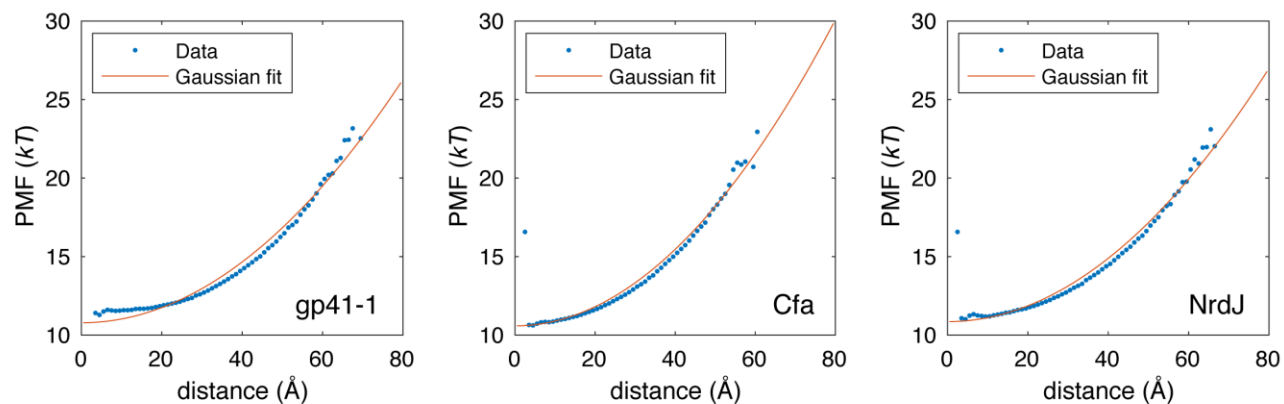
Supplementary Figures



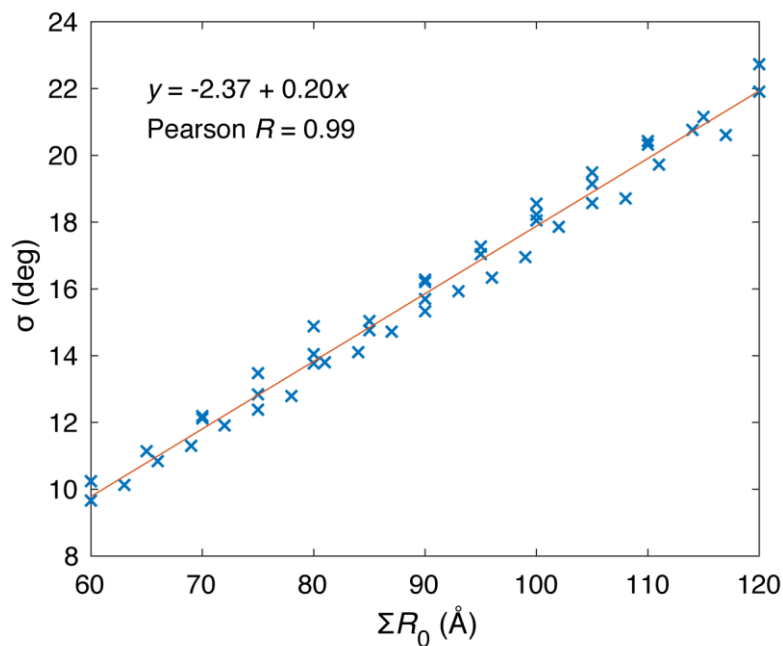
Supplementary Figure 1 SDS-PAGE gels of the construction and purification of triangular nanostructure Tri18. Lane M, protein marker; lane 1, purified $\text{Gp}^{\text{C}_{18}\text{-3HB-Cfa}^{\text{N}_{14}}}$; lane 2, purified $\text{Cfa}^{\text{C}_{14}\text{-3HB-NrdJ}^{\text{N}_{18}}}$; lane 3, purified $\text{NrdJ}^{\text{C}_{18}\text{-3HB-Gp}^{\text{N}_{18}}}$; lane 4, ligation product of $\text{Cfa}^{\text{C}_{14}\text{-3HB-NrdJ}^{\text{N}_{18}}}$ and $\text{NrdJ}^{\text{C}_{18}\text{-3HB-Gp}^{\text{N}_{18}}}$; lane 5, ligation product of $\text{Cfa}^{\text{C}_{14}\text{-3HB-NrdJ}^{\text{N}_{18}}}$, $\text{NrdJ}^{\text{C}_{18}\text{-3HB-Gp}^{\text{N}_{18}}}$, and $\text{Gp}^{\text{C}_{18}\text{-3HB-Cfa}^{\text{N}_{14}}}$; lane 6, triangle purified by Ni-NTA column; lane 7, triangle further purified by SEC. Source data are provided as a Source Data file.



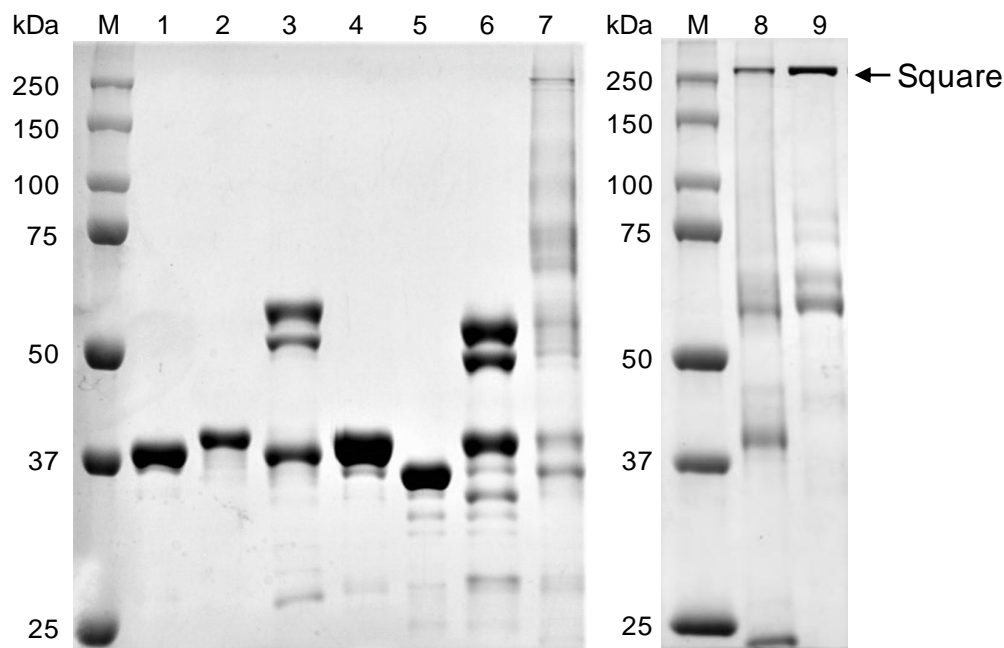
Supplementary Figure 2 Identification of side products from triangle ligation. **a** SEC profile of the triangle purification, with the peaks corresponding to protein species indicated. **b** SDS-PAGE gel showing the corresponding fractions from the SEC purification. Lane M, protein marker; lane 1, uncyclized trimer; lane 2, dimer; lane 3, triangular nanostructure; lane 4, monomer. **c** STEM images of the three SEC fractions, showing primarily uncyclized trimer, dimer, and monomer (STEM image of triangle shown in Figure 2). Scale bars, 50 nm. Source data are provided as a Source Data file.



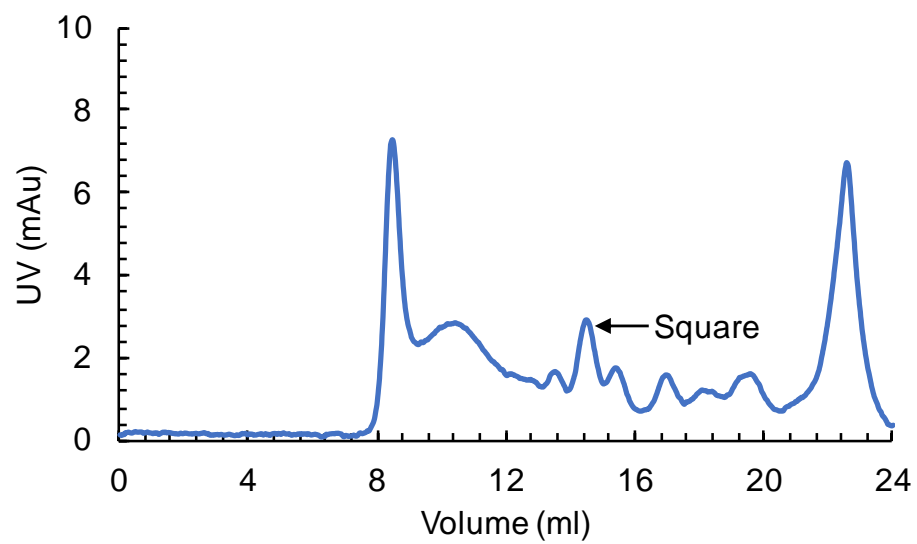
Supplementary Figure 3 Potential of mean force (PMF) data from atomistic simulations. PMF data shown as blue dots and the fitted curves as red lines. The fitting function is $f(x) = A - 3(x/R_0)^2$, where A and R_0 are fitting parameters. The Pearson correlation coefficients (R) are 0.99 (Gp), 0.97 (Cfa), and 0.97 (NrdJ). Source data are provided as a Source Data file.



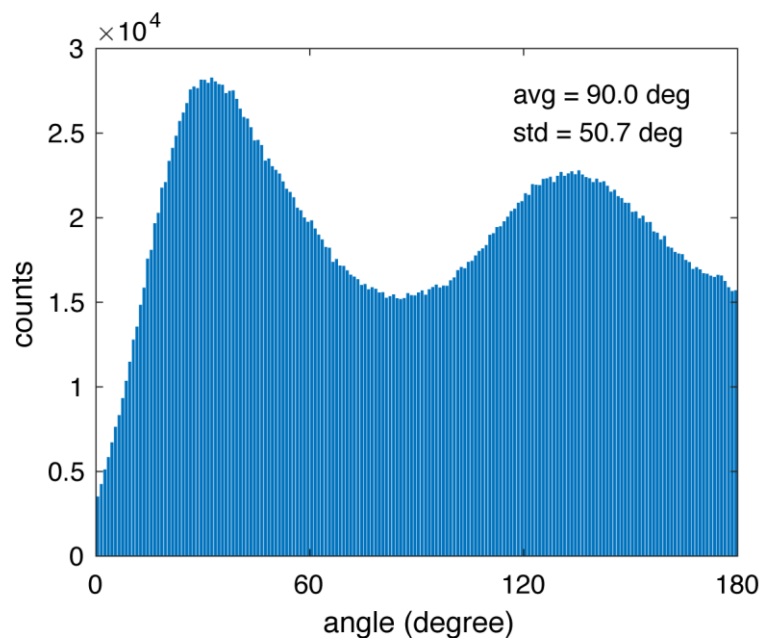
Supplementary Figure 4 Linear relationship of the summation of characteristic lengths of three Gaussian-chain linkers (ΣR_0) to the standard deviation (σ) of angles. Blue crosses indicate data points from different combinations of characteristic lengths, and the red line shows the regression line. The Pearson correlation coefficient (R) is 0.99. Source data are provided as a Source Data file.



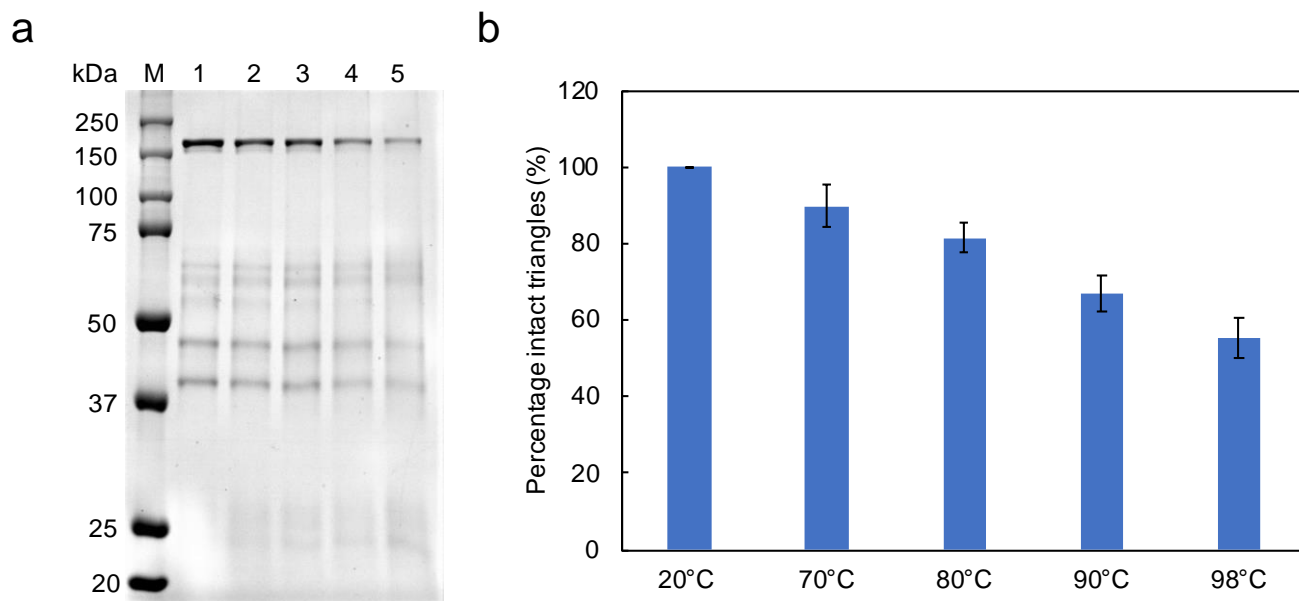
Supplementary Figure 5 SDS-PAGE gel of the construction and purification of the square nanostructure. Lane M, protein marker; lane 1, purified $\text{Gp}^{\text{C}}_{10}\text{-3HBt-Cfa}^{\text{N}}_{10}$; lane 2, purified $\text{Cfa}^{\text{C}}_{10}\text{-3HBt-NrdJ}^{\text{N}}_{10}$; lane 3, ligation product of $\text{Gp}^{\text{C}}_{10}\text{-3HBt-Cfa}^{\text{N}}_{10}$ and $\text{Cfa}^{\text{C}}_{10}\text{-3HBt-NrdJ}^{\text{N}}_{10}$; lane 4, purified $\text{NrdJ}^{\text{C}}_{10}\text{-3HBt-Cfa}^{\text{N}}_{10}$; lane 5, purified $\text{Cfa}^{\text{C}}_{10}\text{-3HBt-Gp}^{\text{N}}_{10}$; lane 6, ligation product of $\text{NrdJ}^{\text{C}}_{10}\text{-3HBt-Cfa}^{\text{N}}_{10}$ and $\text{Cfa}^{\text{C}}_{10}\text{-3HBt-Gp}^{\text{N}}_{10}$; lane 7, ligation product of $\text{Gp}^{\text{C}}_{10}\text{-3HBt-Cfa}^{\text{N}}_{10}$, $\text{Cfa}^{\text{C}}_{10}\text{-3HBt-NrdJ}^{\text{N}}_{10}$, $\text{NrdJ}^{\text{C}}_{10}\text{-3HBt-Cfa}^{\text{N}}_{10}$, and $\text{Cfa}^{\text{C}}_{10}\text{-3HBt-Gp}^{\text{N}}_{10}$; lane 8, square purified by Ni-NTA column; lane 9, square further purified by SEC. Source data are provided as a Source Data file.



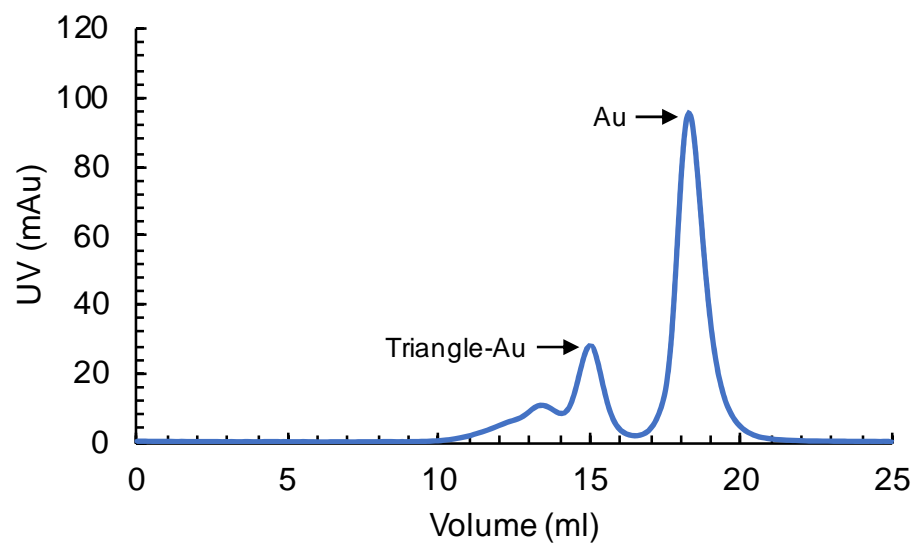
Supplementary Figure 6 SEC profile of the square purification. The peak corresponding to the desired square nanostructure is indicated. Source data are provided as a Source Data file.



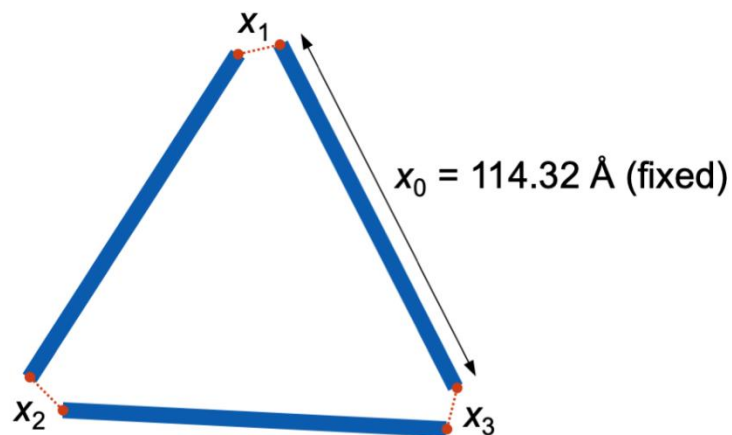
Supplementary Figure 7 Simulated distribution of the square structure angles with no constraints on angle flexibility. Note that the simulated standard deviation (50.7°) is significantly higher than the measured value (27.6°) and that the simulation produces a bimodal distribution instead of the trimodal distribution observed experimentally. Source data are provided as a Source Data file.



Supplementary Figure 8 SDS-PAGE analysis of protein triangle thermostability. **a** Representative SDS-PAGE gel showing Tri10t sample before and after incubation at elevated temperatures. Lane M, protein marker; lane 1, Tri10t incubated at 20°C; lane 2, Tri10t incubated at 70°C; lane 3, Tri10t incubated at 80°C; lane 4, Tri10t incubated at 90°C; lane 5, Tri10t incubated at 98°C. **b** Ratio of intact triangles after incubation at different temperatures based on band intensities observed in SDS-PAGE. Temperature melts were performed in triplicate and normalized by the sample incubated at 20°C. Error bars represent the s.d. ($n = 3$ temperature melt replicates). Source data are provided as a Source Data file.



Supplementary Figure 9 SEC profile of the purification of the triangular nanostructures labeled with AuNPs. Peaks corresponding to the labeled nanostructures (Triangle-Au) and unbound AuNPs (Au) are identified. Source data are provided as a Source Data file.



$$\begin{aligned}
 E_{\text{tot}} &= E_1 + E_2 + E_3 \\
 &= 2(x_1/R_{01})^2 + 2(x_2/R_{02})^2 + 2(x_3/R_{03})^2
 \end{aligned}$$

Supplementary Figure 10 Schematic representation of the triangular structures used in coarse-grained simulations. The main part of 3HB is modeled as a rigid rod with fixed length of 114.32 Å, and the flexible linker is modeled as a 2-dimensional harmonic spring. The total energy of the system is a simple arithmetic sum of all spring energies.

Supplementary Tables

Supplementary Table 1 Linker information of triangular and square nanostructures.

Protein nanostructure	Protein building block	Linker length	Linker amino acid sequence
Tri18	Gp ^C ₁₈ -3HB-Cfa ^N ₁₄	14	Cfa ₁₄ : SGPG <u>AEYCFN</u> SGPG
Tri18	Cfa ^C ₁₄ -3HB-NrdJ ^N ₁₈	18	NrdJ ₁₈ : SGPGGTNP <u>CSEIVL</u> SGPG
Tri18	NrdJ ^C ₁₈ -3HB-Gp ^N ₁₈	18	Gp ₁₈ : SGPGTRSGYSSSDVSGPG
Tri10	Gp ^C ₁₀ -3HB-Cfa ^N ₁₀	10	Cfa ₁₀ : PG <u>AEYCFN</u> GP
Tri10	Cfa ^C ₁₀ -3HB-NrdJ ^N ₁₀	10	NrdJ ₁₀ : PGN <u>PCSEI</u> GP
Tri10	NrdJ ^C ₁₀ -3HB-Gp ^N ₁₀	10	Gp ₁₀ : PG <u>SGYSSS</u> GP
Tri7	Gp ^C ₇ -3HB-Cfa ^N ₇	7	Cfa ₇ : PG <u>CFN</u> GP
Tri7	Cfa ^C ₇ -3HB-NrdJ ^N ₇	7	NrdJ ₇ : PG <u>SEI</u> GP
Tri7	NrdJ ^C ₇ -3HB-Gp ^N ₇	7	Gp ₇ : PG <u>SSS</u> GP
Tri18t	Gp ^C ₁₈ -3HB _t -Cfa ^N ₁₄	14	Cfa ₁₄ : SGPG <u>AEYCFN</u> SGPG
Tri18t	Cfa ^C ₁₄ -3HB _t -NrdJ ^N ₁₈	18	NrdJ ₁₈ : SGPGGTNP <u>CSEIVL</u> SGPG
Tri18t	NrdJ ^C ₁₈ -3HB _t -Gp ^N ₁₈	18	Gp ₁₈ : SGPGTRSGYSSSDVSGPG
Tri14t	Gp ^C ₁₄ -3HB _t -Cfa ^N ₁₀	10	Cfa ₁₀ : PG <u>AEYCFN</u> GP
Tri14t	Cfa ^C ₁₀ -3HB _t -NrdJ ^N ₁₄	14	NrdJ ₁₄ : PGGTNP <u>CSEIVL</u> GP
Tri14t	NrdJ ^C ₁₄ -3HB _t -Gp ^N ₁₄	14	Gp ₁₄ : PGTRSGYSSSDVGP
Tri10t	Gp ^C ₁₀ -3HB _t -Cfa ^N ₁₀	10	Cfa ₁₀ : PG <u>AEYCFN</u> GP
Tri10t	Cfa ^C ₁₀ -3HB _t -NrdJ ^N ₁₀	10	NrdJ ₁₀ : PGN <u>PCSEI</u> GP
Tri10t	NrdJ ^C ₁₀ -3HB _t -Gp ^N ₁₀	10	Gp ₁₀ : PG <u>SGYSSS</u> GP
Square	Gp ^C ₁₀ -3HB _t -Cfa ^N ₁₀	10	Cfa ₁₀ : PG <u>AEYCFN</u> GP
Square	Cfa ^C ₁₀ -3HB _t -NrdJ ^N ₁₀	10	NrdJ ₁₀ : PGN <u>PCSEI</u> GP
Square	NrdJ ^C ₁₀ -3HB _t -Cfa ^N ₁₀	10	Cfa ₁₀ : PG <u>AEYCFN</u> GP
Square	Cfa ^C ₁₀ -3HB _t -Gp ^N ₁₀	10	Gp ₁₀ : PG <u>SGYSSS</u> GP

Note: Residual extein sequences, specific to the SI group used, are underlined.

Supplementary Table 2 Primer sequences used in plasmid construction.

Primer name	Primer sequence
prWB01	GGTCTCGGTCCAGGTAATGAAGACGACATGAAAAAACTG
prWB02	GGTCTCTCCGGATTTCGAGGCCTTTCAGCACTTTTTTCG
prWB03	GGTCTCCATATGGCTAAGACTAAAATGCTGAAAAAAATTC
prWB04	GGTCTCTGGACCGGAAACATCGCTAGAAGAGTTGTGG
prWB05	GGTCTCTCCGGACCGGGTGCAGAATATTGCCTGTCTTACG
prWB06	GGTCTCGCTACCCGGCAAACCATCAACTTGTTTCAGG
prWB07	GGTCTCCATATGTATATCTCCTTCTTAAAAGATCTTTTGAATTC
prWB08	GGTCTCGGTAGCCATCACCATCACCATCATCACCACC
prWB09	GGTCTCCATATGGCTAAGACTAAAGTCAAGATC
prWB10	GGTCTCTGGACCGGAGTTGAAGCAGTTAGAGGCCAC
prWB11	GGTCTCTCCGGACCGGGTGGCACTAACCCGTGCTG
prWB12	GGTCTCGCTACCAATAGCCACCACCAGTTC
prWB13	GGTCTCCATATGGCTAAGACTAAAGAGGCGAAG
prWB14	GGTCTCTGGACCGGACAGCACGATTTCCGGAGTTG
prWB15	GGTCTCTCCGGACCGGGTACCCGTTCTGGTTACTGTCTG
prWB16	GGTCTCGCTACCTTCTTTTCACGTACAGGCACATAC
prWB17	GGTCTCGGTCCAAATGAAGACGACATGAAAAAACTG
prWB18	GGTCTCACCCGGTTCGAGGCCTTTCAGCACTTTTTTC
prWB19	GGTCTCTGGACCGCTAGAAGAGTTGTGGGTCAG
prWB20	GGTCTCCCGGGTGCAGAATATTGCCTGTCTTACG
prWB21	GGTCTCTGGACCGTTGAAGCAGTTAGAGGCCAC
prWB22	GGTCTCCCGGGTAACCCGTGCTGCCTGGTTG
prWB23	GGTCTCTGGACCGATTTCCGGAGTTGTGTACCAGAATG
prWB24	GGTCTCCCGGGTTCTGGTTACTGTCTGGACCTG
prWB25	GGTCTCCCGGGTTGCCTGTCTTACGACACAGAG
prWB26	GGTCTCCCGGGTTGCTGCCTGGTTGGCAGC
prWB27	GGTCTCCCGGGTTGCTGTCTGGACCTGAAGACCC
prWB28	GGTCTCGGTCCAGGTAAAAACTGTATAAACAAATGGTGCAGG
prWB29	GGTCTCTCCGGATTTTCAGCACTTTTTTCGGCGATTTTC
prWB30	GGTCTCGGTCCAAAAAACTGTATAAACAAATGGTGCAG
prWB31	GGTCTCACCCGGGCCGCACAGCACTTTTTTCGGCGA
prWB32	GGTCTCTGGACCAACATCGCTAGAAGAGTTGTG
prWB33	GGTCTCCCGGGTGGCACTAACCCGTGCTGC
prWB34	GGTCTCTGGACCCAGCACGATTTCCGGAGTTG
prWB35	GGTCTCCCGGGTACCCGTTCTGGTTACTGTCTGGAC

Supplementary Table 3 Percentages of component structures present following purification by affinity chromatography and SEC, as determined by SDS-PAGE gel densitometry.

Component	After affinity chromatography (%)	After SEC (%)
Triangle or square	60.4	81.0
Uncyclized trimer	0.6	ND
Dimer	21.0	8.3
Monomer with SI	10.0	10.7
Monomer without SI	8.1	ND

Note: ND, not detected. The relative amounts of the components of triangular nanostructure sample were quantified based on Supplementary Figure 1. Source data are provided as a Source Data file.

Supplementary Table 4 Secondary structure contents of simulated linker proteins, as quantified by the DSSP scores.

Name	Sequence	Avg. α content (H-score)	Avg. β content (E-score)
3HB ^C -Cfa ₁₄ -3HB ^N	<i>GLESGPGA</i> EYCFNSGPG NEDDM	8.04 %	1.42 %
3HB ^C -NrdJ ₁₈ -3HB ^N	<i>GLESGPGGT</i> NPCSEIVLSGPG NEDDM	3.47 %	1.95 %
3HB ^C -Gp ₁₈ -3HB ^N	<i>GLESGPGTR</i> SGYSSSDVSGPG NEDDM	2.05 %	3.06 %

Note: The flexible N- and C-terminal tail regions of 3HB are shown in bold and italic text, respectively.

Supplementary Table 5 Gaussian characteristic lengths of linkers extracted from atomistic simulations.

The fitting function is $f(x) = A - 3(x/R_0)^2$, where A and R_0 are fitting parameters.

Name	Sequence	R_0 (Å)	Pearson's R
Cfa ₁₀	PGA EYCFNGP	20.3	0.86
NrdJ ₁₀	PGNPCSEIGP	20.1	0.88
Gp ₁₀	PGSGYSSSGP	21.7	0.83
NrdJ ₁₄	PGGTNPCSEIVLGP	24.6	0.97
Gp ₁₄	PGTRSGYSSSDVGP	24.3	0.93
Cfa ₁₄	SGPGA EYCFNSGPG	23.1	0.96
NrdJ ₁₈	SGPGGTNPCSEIVLSGPG	28.8	0.96
Gp ₁₈	SGPGTRSGYSSSDVSGPG	27.7	0.97
3HB ^C -Cfa ₁₀ -3HB ^N	<i>GLE</i> PGA EYCFNGP NEDDM	28.0	1.00
3HB ^C -NrdJ ₁₀ -3HB ^N	<i>GLE</i> PGNPCSEIGP NEDDM	29.0	0.97
3HB ^C -Gp ₁₀ -3HB ^N	<i>GLE</i> PGSGYSSSGP NEDDM	30.4	0.95
3HB ^C -Cfa ₁₄ -3HB ^N	<i>GLE</i> SGPGA EYCFNSGPG NEDDM	31.4	0.97
3HB ^C -NrdJ ₁₈ -3HB ^N	<i>GLE</i> SGPGGTNPCSEIVLSGPG NEDDM	35.2	0.99
3HB ^C -Gp ₁₈ -3HB ^N	<i>GLE</i> SGPGTRSGYSSSDVSGPG NEDDM	34.5	0.97

Note: The flexible N- and C-terminal tail regions of 3HB are shown in bold and italic text, respectively.