Supplementary Information

Exploration of the Hierarchical Assembly Space of Collagen-like Peptides Beyond the Triple Helix

Le Tracy Yu[1], Mark A. B. Kreutzberger[2], Thi H. Bui[1], Maria C. Hancu[1], Adam C. Farsheed[3], Edward H. Egelman[2], Jeffrey D. Hartgerink[1,3*]

[1] Department of Chemistry, Rice University, 6100 Main Street, Houston, TX, 77005, USA
[2] Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, VA, 22903, USA
[3] Department of Bioengineering, Rice University, 6100 Main Street, Houston, TX, 77005, USA

* To whom correspondence should be addressed: jdh@rice.edu

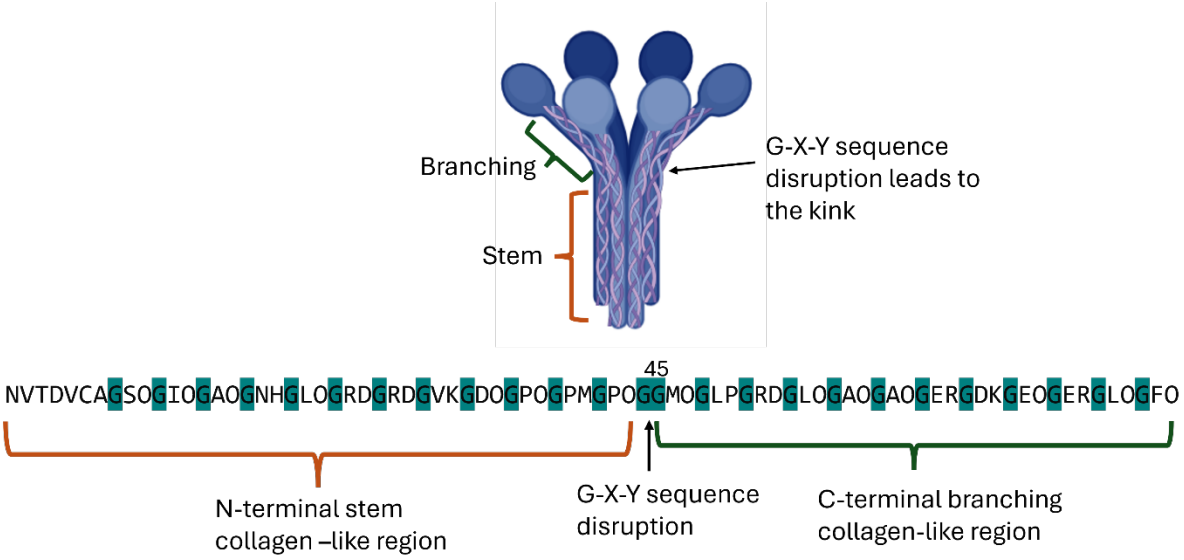

Figure S1. *Illustration of amino acid sequence dissection of protein rat SP-A. The full-length protein sequence was obtained from the database.[1] "G" residue was highlighted. The amino acid sequence that conforms to the "G-X-Y" sequence pattern was considered as the collagen-like region. Disruption of the sequence pattern was observed at the 45th Gly residue (numbering starts after the signal peptide). This discontinuity was considered to cause the collagen triple helices to bend, forming the kink. Therefore, the N-terminal 1-44 residues were considered as the stem collagen-like region and the C-terminal 45-81 residues are considered as the branching collagen-like region. Created in BioRender. Yu, T. (2023) BioRender.com/p84z508*

Table S1 Peptide amino acid sequence and molecular weight information.

Peptides	Sequence	Expected	Observed
SPA-Rat	NVTDVAAGSOGIOGAOGNHGLOGRDGRDGVKGDOGPOGPMGPOG	4206.5	4207.0
SPA-Human	EVKDVAVGSOGIOGAOGNHGLOGRDGRDGVKGDOGPOGPMGPOG	4282.0	4281.5
SPA-bcb	AAGSOGIOGAOGNHGLOGRDGRDGVKGDOGPOGPMGPOG	3479.7	3480.7
SPA-R-P	NVTDVAAGSOGIOGAOGNHGLOGPDGRDGVKGDOGPOGPMGPOG	4147.4	4147.9
SPA-D-O	NVTDVAAGSOGIOGAOGNHGLOGROGRDGVKGDOGPOGPMGPOG	4204.5	4204.4
SPA-I-P	NVTDVAAGSOGPOGAOGNHGLOGRDGRDGVKGDOGPOGPMGPOG	4191.4	4192.0
SPA-I-S	NVTDVAAGSOGSOGAOGNHGLOGRDGRDGVKGDOGPOGPMGPOG	4180.4	4180.9
DES-1	DAAGIOGPOGPOGLOGRDGRDGVOGPOGPOGPM	3162.5	3163.0
DES-1-no DAA	GIOGPOGPOGLOGRDGRDGVOGPOGPOGPM	2906.4	2907.2
DES-H	DAAGIOGPOGPOGLOGPOGPOGVOGPOGPOGPM	3063.3	3064.0
DES-H-no DAA	GIOGPOGPOGLOGPOGPOGVOGPOGPOGPM	2783.3	2806.5 $[M+Na]^+$
DES-P	DAAGPOGPOGPOGRDGRDGPOGPOGPOGPOGPO	3110.4	3111.4
DES-P-no DAA	GPOGPOGPOGRDGRDGPOGPOGPOGPOGPO	2853.3	2853.2
RD1	DAAGPOGPOGPOGRDGRDGRDGRDGPOGPOGPO	1617.8	1617.9
RD1-no DAA	GPOGPOGPOGRDGRDGRDGRDGPOGPOGPO	2975.4	2976.0
RD2	DAAGPOGRDGRDGPOGPOGPOGPOGRDGRDGPO	3233.5	3234.1
RD2-no DAA	GPOGRDGRDGPOGPOGPOGPOGRDGRDGPO	2976.0	2976.5

Expected / Observed masses are for $[M+H]^+$ in all cases with the exception of RD1 where $[M+2H]^{2+}$ is reported.

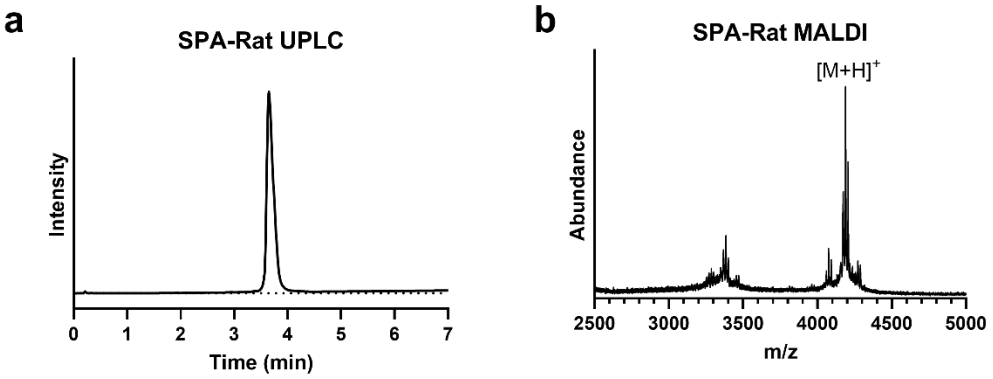

Figure S2. *Peptide characterizations. a) Analytical HPLC (UPLC) of pure peptide SPA-Rat. The chromatographic method was run using a 5-35% B gradient with a 10% B/min ramping rate. b) MALDI mass spectrometry of peptide SPA-Rat.*

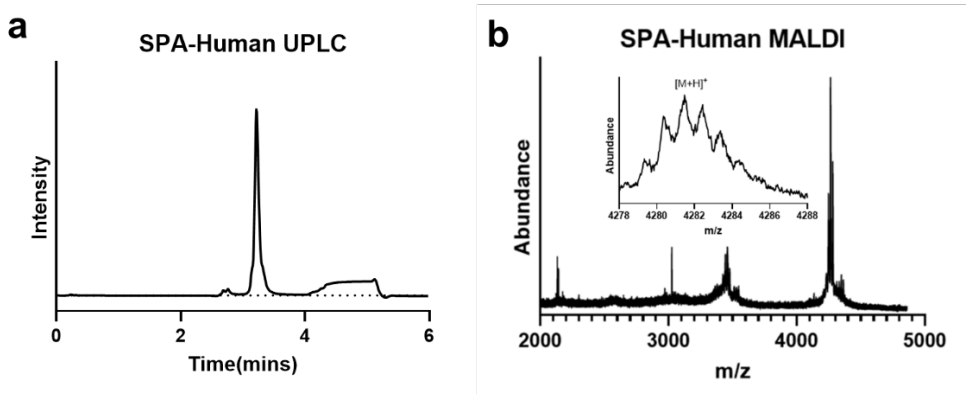

Figure S3. *Peptide characterizations. a) Analytical HPLC (UPLC) of pure peptide SPA-Human. The chromatographic method was run using a 5-35% B gradient with a 10% B/min ramping rate. b) MALDI mass spectrometry of peptide SPA-human.*

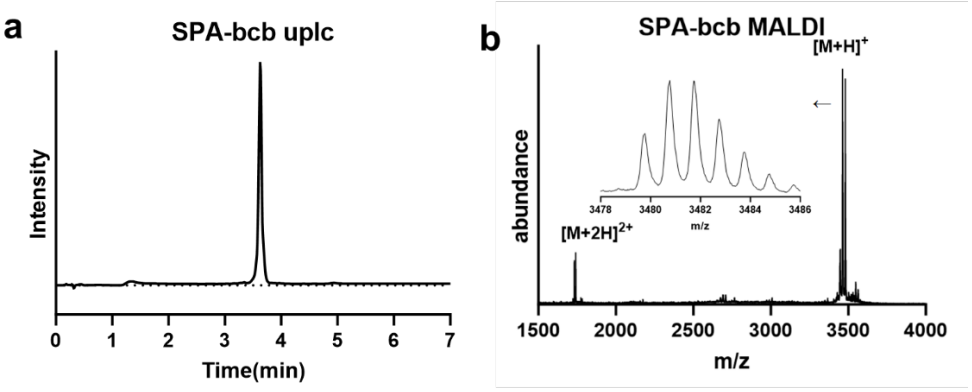

Figure S4. *Peptide characterizations. a) Analytical HPLC (UPLC) of pure peptide SPA-bcb. The chromatographic method was run using a 5-35% B gradient with a 10% B/min ramping rate. b) MALDI mass spectrometry of peptide SPA-bcb.*

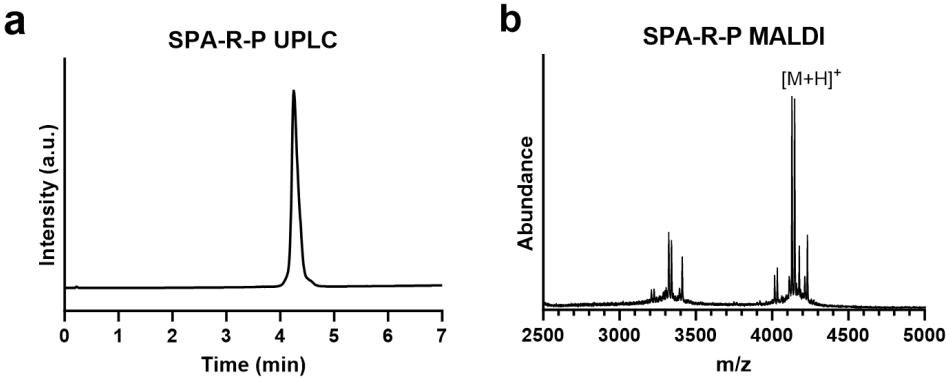

Figure S5. *Peptide characterizations. a) Analytical HPLC (UPLC) of pure peptide SPA-R-P. The chromatographic method was run using a 5-35% B gradient with a 10% B/min ramping rate. b) MALDI mass spectrometry of peptide SPA-R-P.*

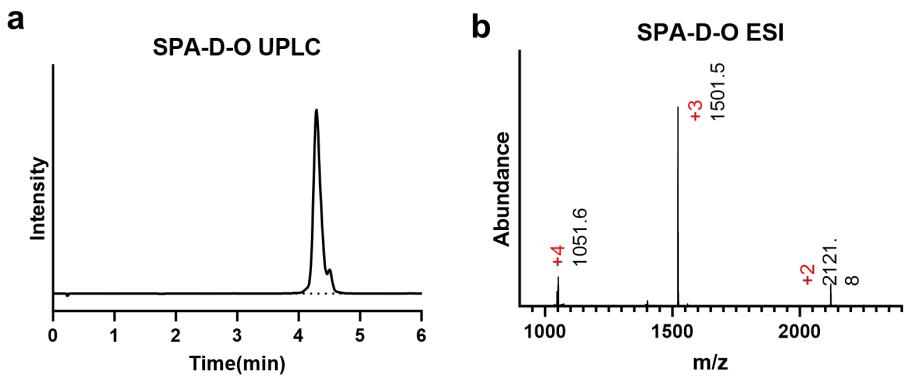

Figure S6. *Peptide characterizations. a) Analytical HPLC (UPLC) of pure peptide SPA-D-O. The chromatographic method was run using a 5-35% B gradient with a 10% B/min ramping rate. b) LC-MS (ESI) mass spectrometry of peptide SPA-D-O.*

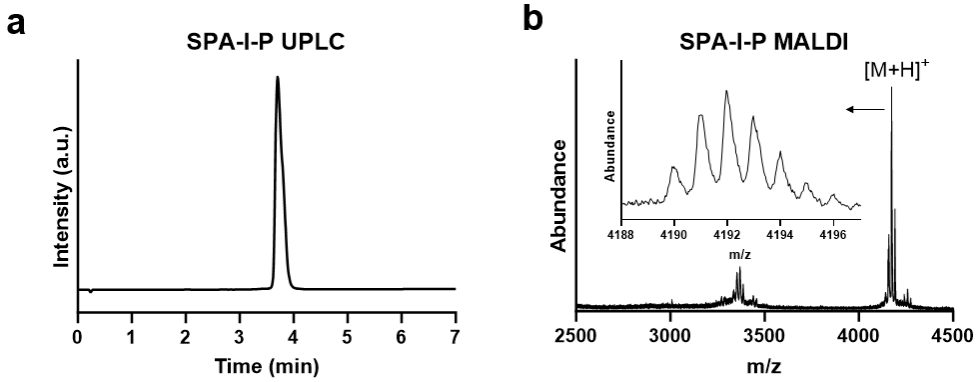

Figure S7. *Peptide Characterizations. a) Analytical HPLC (UPLC) of pure peptide SPA-I-P. The chromatographic method was run using a 5-35% B gradient with a 10% B/min ramping rate. b) Mass spectrometry of peptide SPA-I-P.*

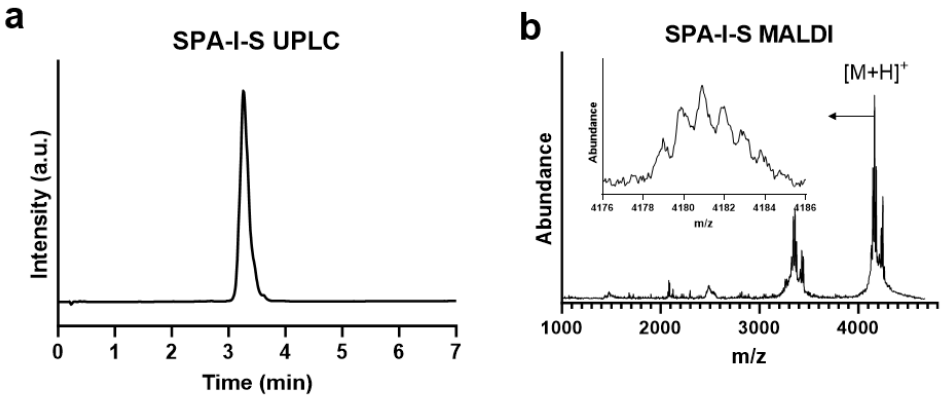

Figure S8. *Peptide characterizations. a) Analytical HPLC (UPLC) of pure peptide SPA-I-S. The chromatographic method was run using a 5-35% B gradient with a 10% B/min ramping rate. b) Mass spectrometry of peptide SPA-I-S.*

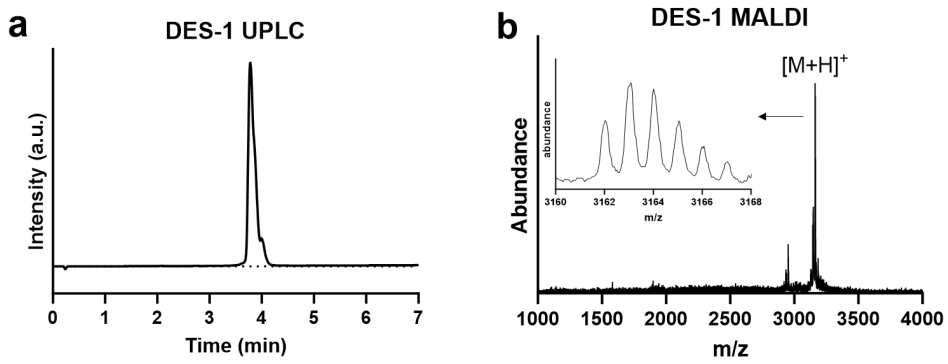

Figure S9. *Peptide characterizations. a) Analytical HPLC (UPLC) of pure peptide DES-1. The chromatographic method was run using a 5-35% B gradient with a 10% B/min ramping rate. b) Mass spectrometry of peptide DES-1.*

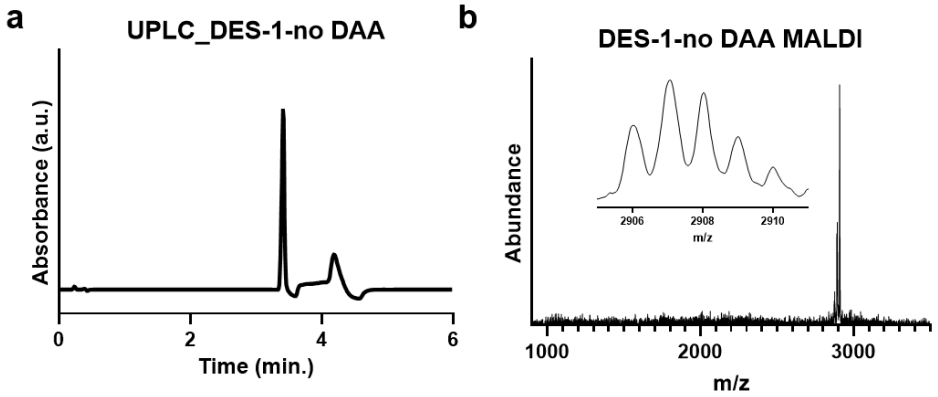

Figure S10 Peptide characterizations. a) Analytical HPLC (UPLC) of pure peptide DES-1-no DAA. The chromatographic method was run using a 5-35% B gradient with a 10% B/min ramping rate. b) Mass spectrometry of peptide DES-1-no DAA.

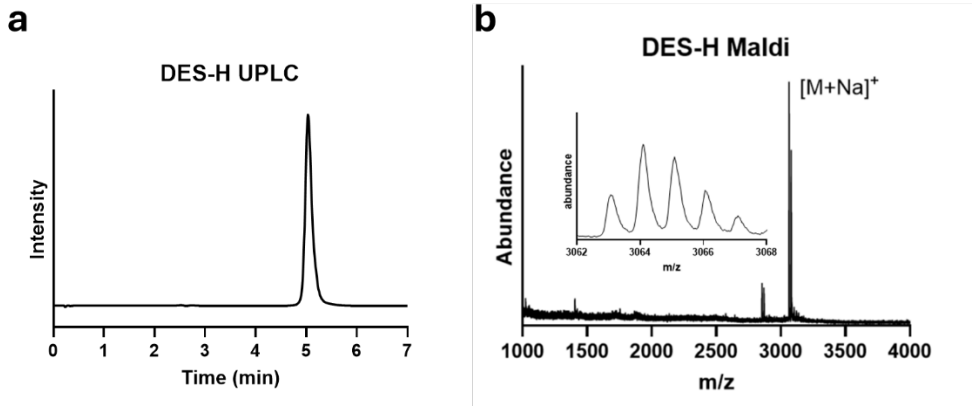

Figure S11. *Peptide characterizations. a) Analytical HPLC (UPLC) of pure peptide DES-H. The chromatographic method was run using a 5-35% B gradient with a 10% B/min ramping rate. b) Mass spectrometry of peptide DES-H.*

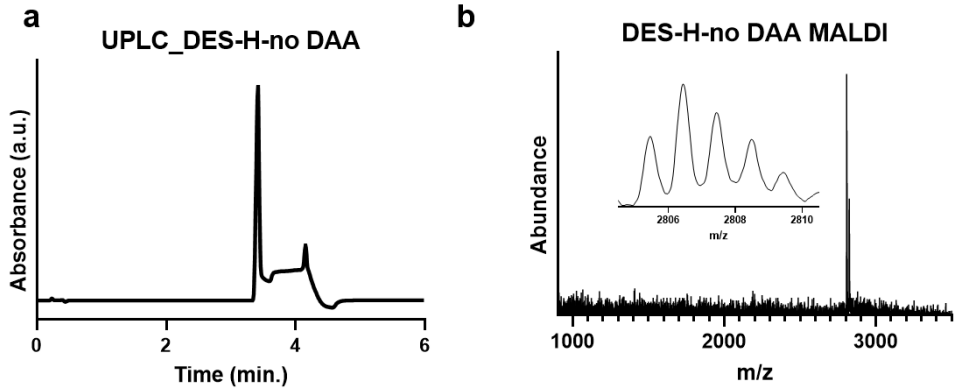

Figure S12 *Peptide characterizations. a) Analytical HPLC (UPLC) of pure peptide DES-H-no DAA. The chromatographic method was run using a 5-35% B gradient with a 10% B/min ramping rate. b) Mass spectrometry of peptide DES-H-no DAA.*

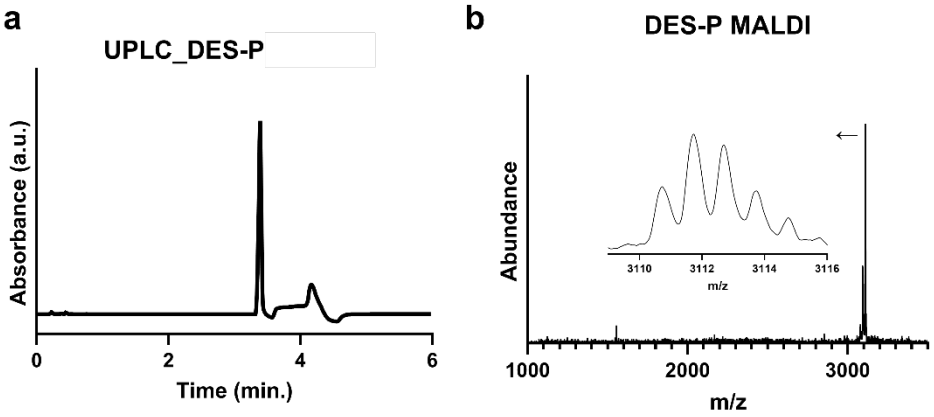

Figure S13 *Peptide characterizations. a) Analytical HPLC (UPLC) of pure peptide DES-P. The chromatographic method was run using a 5-35% B gradient with a 10% B/min ramping rate. b) Mass spectrometry of peptide DES-P.*

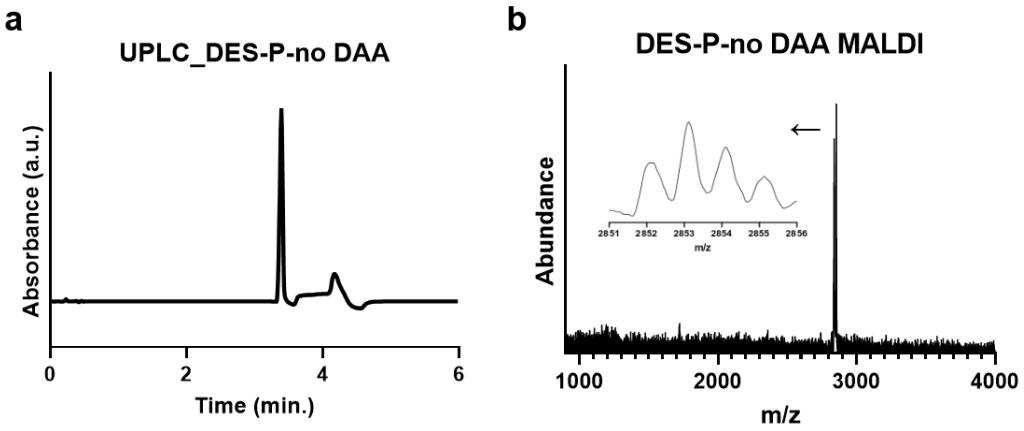

Figure S14 *Peptide characterizations. a) Analytical HPLC (UPLC) of pure peptide DES-P-no DAA. The chromatographic method was run using a 5-35% B gradient with a 10% B/min ramping rate. b) Mass spectrometry of peptide DES-P-no DAA.*

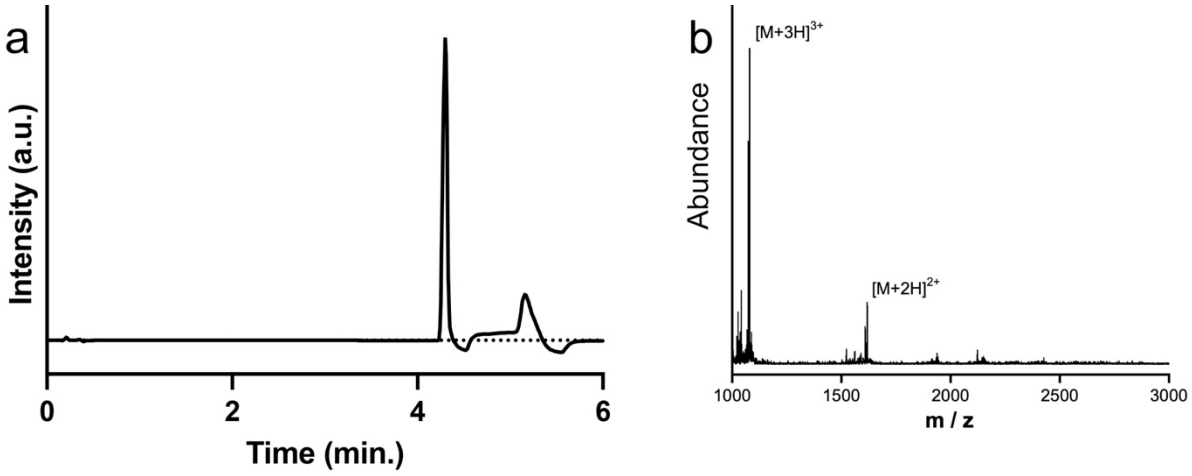

Figure S15. Peptide characterizations. a) Analytical HPLC (UPLC) of pure peptide RD1.b) ESI mass spectrometry of peptide RD1 showing the $[M+3H]^{3+}$ (Expected 1078.8, observed 1078.7) and $[M+2H]^{2+}$ (Expected $[M+2H]^{2+}$ 1617.8, observed 1617.9) ions.

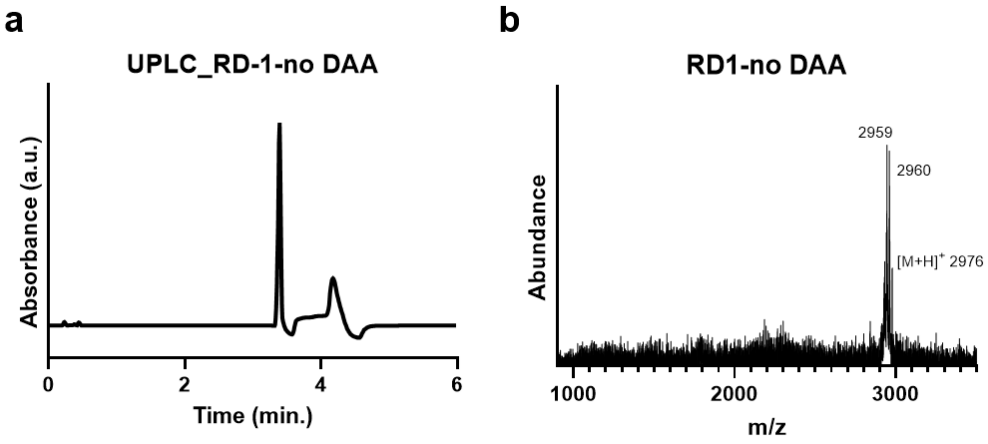

Figure S16 Peptide characterizations. a) Analytical HPLC (UPLC) of pure peptide RD1-no DAA. The chromatographic method was run using a 5-35% B gradient with a 10% B/min ramping rate. b) Mass spectrometry of peptide RD1-no DAA.

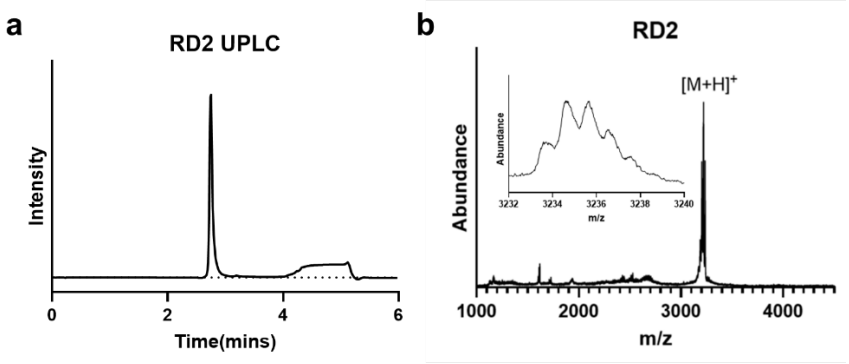

Figure S17. *Peptide characterizations. a) Analytical HPLC (UPLC) of pure peptide RD2. The chromatographic method was run using a 5-35% B gradient with a 10% B/min ramping rate. b) MALDI mass spectrometry of peptide RD2.*

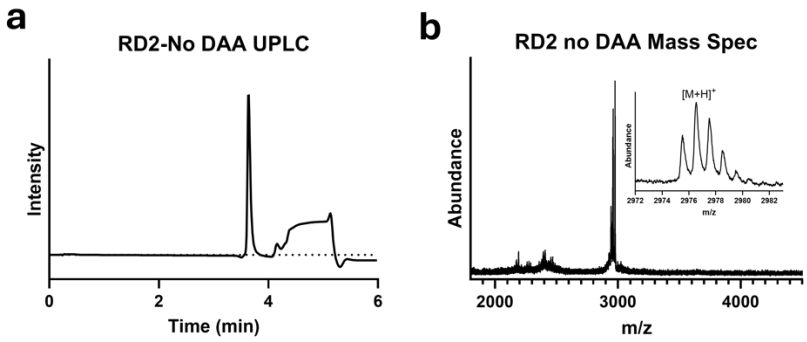

Figure S18. *Peptide characterizations. a) Analytical HPLC (UPLC) of pure peptide RD2-No DAA. The chromatographic method was run using a 5-35% B gradient with a 10% B/min ramping rate. The peak elutes after 4 min is in the wash peak and does not represent biomacromolecules. b) MALDI mass spectrometry of peptide RD2-No DAA.*

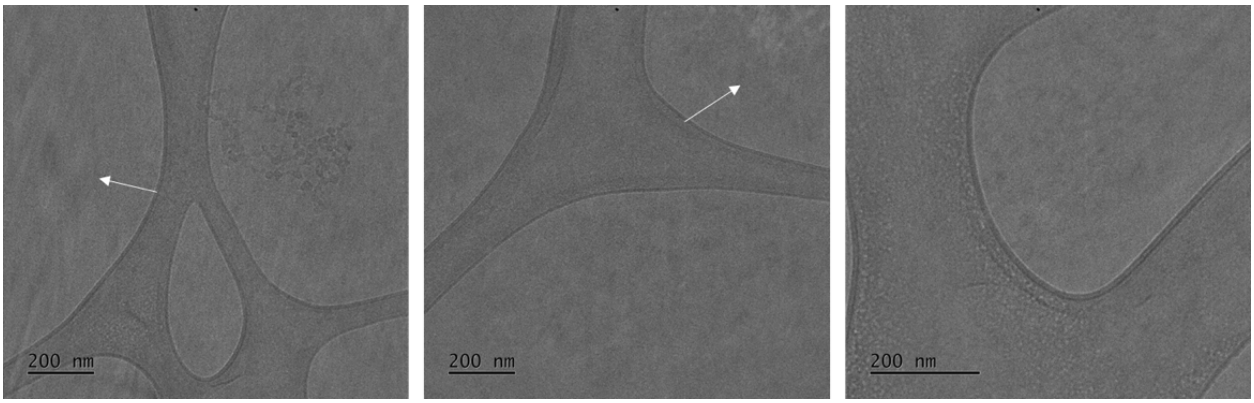

Figure S19. *Cryo-TEM characterization of SPA-bcb aggregates. White arrows highlight the amorphous features of the aggregates. No uniform assembly was observed.*

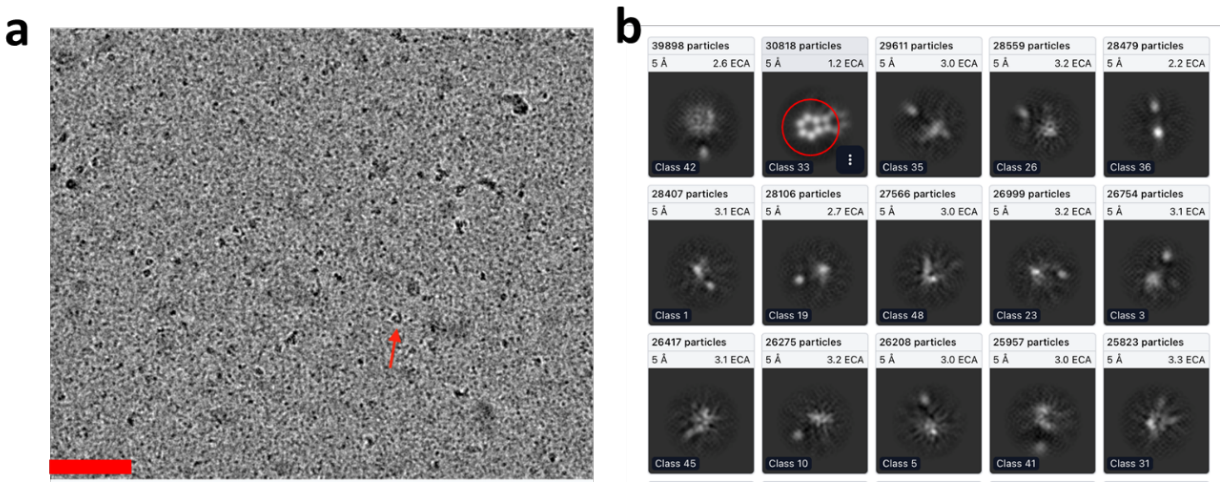

Figure S20. Cryo-EM characterization of SPA-Rat peptide assemblies. *a) representative cryo-EM image of SPA-Rat peptide. The scale bar is 50nm. b) 2D class averaging of the cryo-EM images collected on peptide SPA-Rat.*

Table S2 *Parameters of the models constructed from the cryo-EM data.*

Reconstruction	EMDB number	Resolution (Å)
DES-1 Doublet	EMD-44405	10.2
DES-1 Octadecamer (C6)	EMD-44406	5.3
DES-1 Pentadecamer (C5)	EMD-44409	5.0
SPA-human octadecamer	EMD-44418	5.4

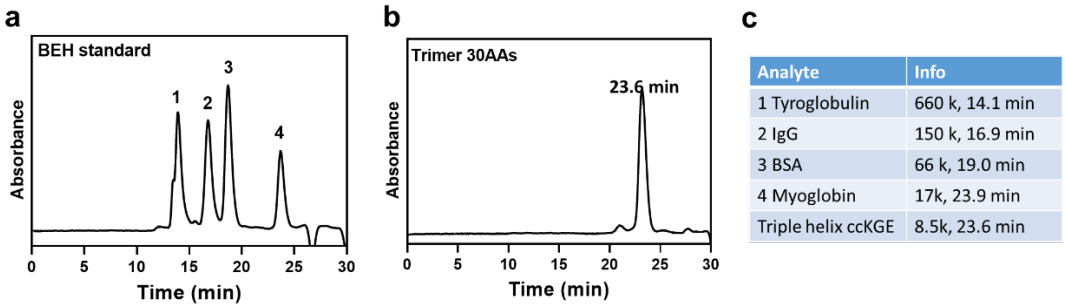

Figure S21. *SEC characterization of protein standards in the Superdex 200 Increase column with a flow rate of 0.75 mL/min. a) SEC of the commercial BEH standard from Waters. B) SEC of a covalently captured collagen trimer with a molecular weight of approximately 8.5 kDa. The trimer was made in our previous work.[2] c) molecular weight information of the standard protein species.*

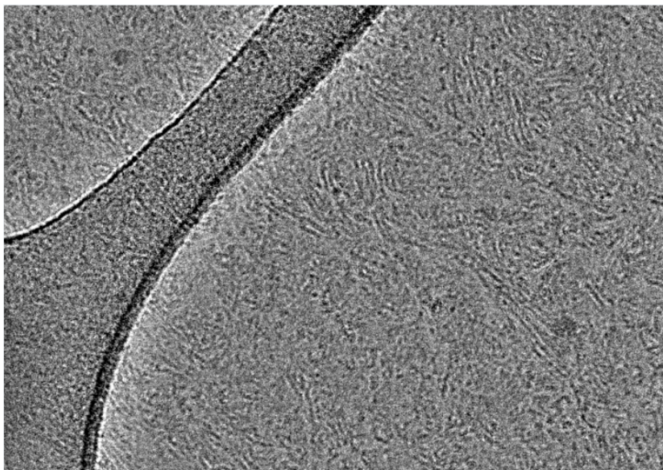

Figure S22. *Additional representative cryo-EM image of SPA-Human assembly.*

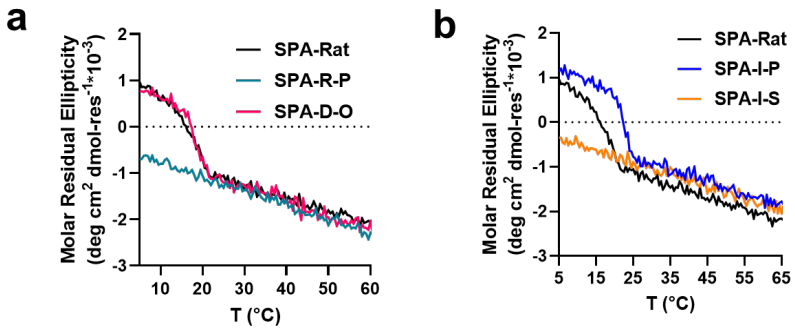

Figure S23. *CD melting results of SPA peptides. a) CD melting curves of SPA-Rat, SPA-R-P and SPA-D-O peptide mutants. b) CD melting curves of SPA-Rat, SPA-I-P, and SPA-I-S samples.*

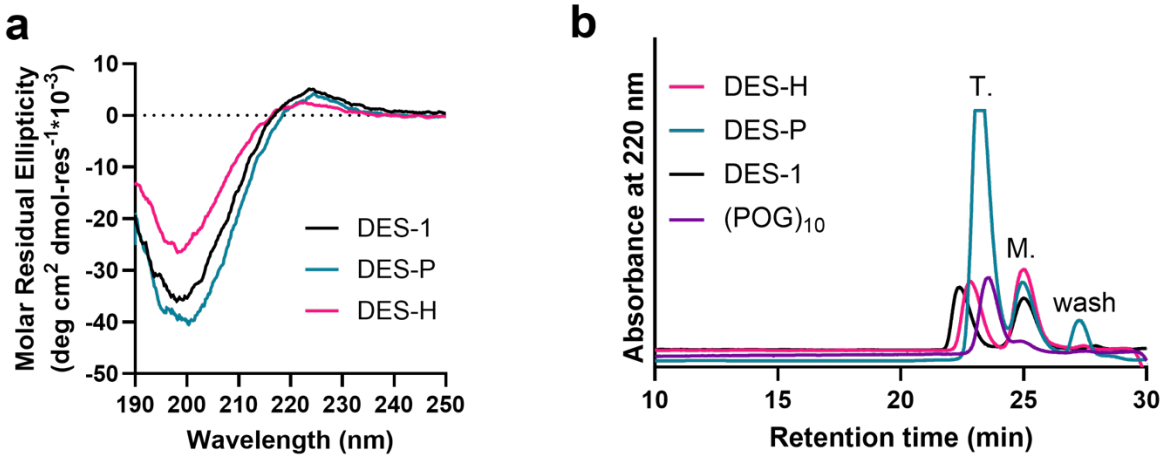

Figure S24 *CD and SEC characterizations of DES-1, H, P peptides. a) CD spectra curve. b) SEC curves collected in a superdex 200 10/300G column with a flowrate of 0.75 ml/min.*

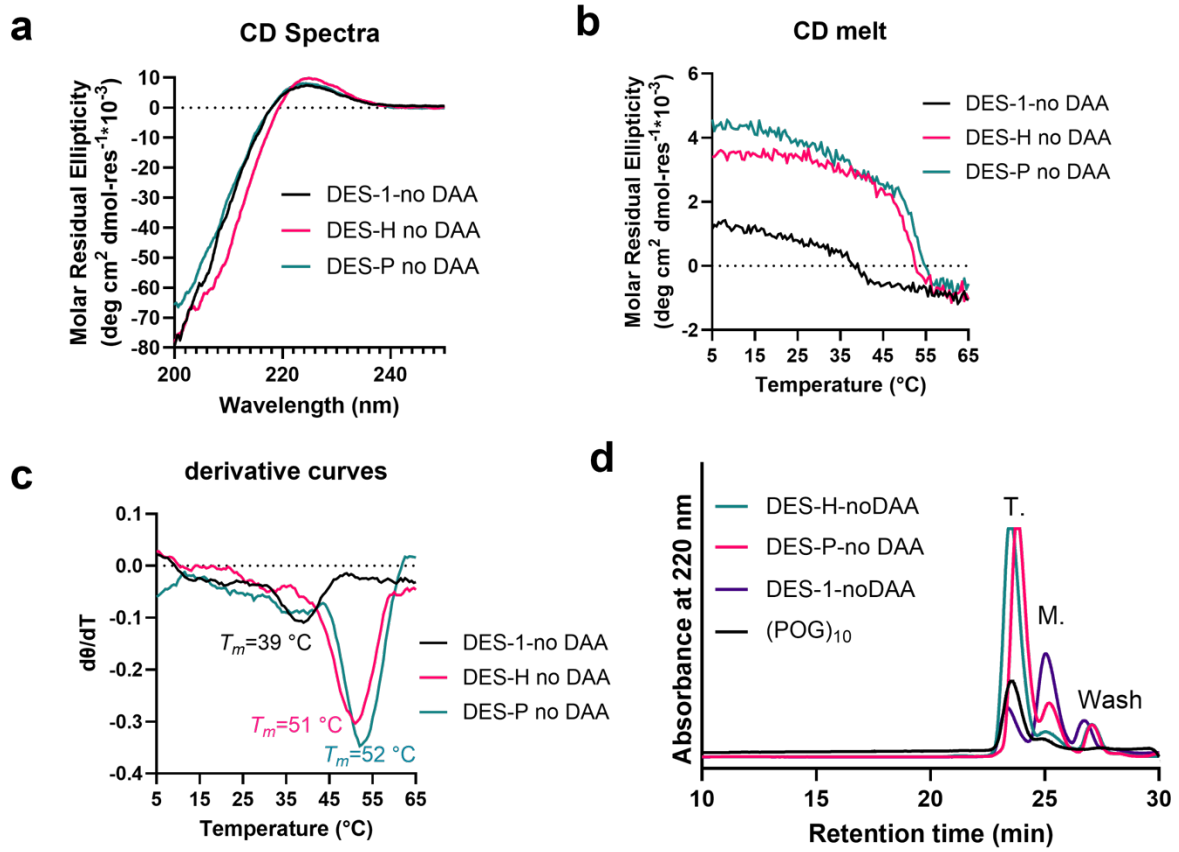

Figure S25 *CD characterization of design peptides with no "DAA" noncollagenous domain. a) CD spectra, b) CD melting curves, c) CD derivative curves. d) SEC chromatography.*

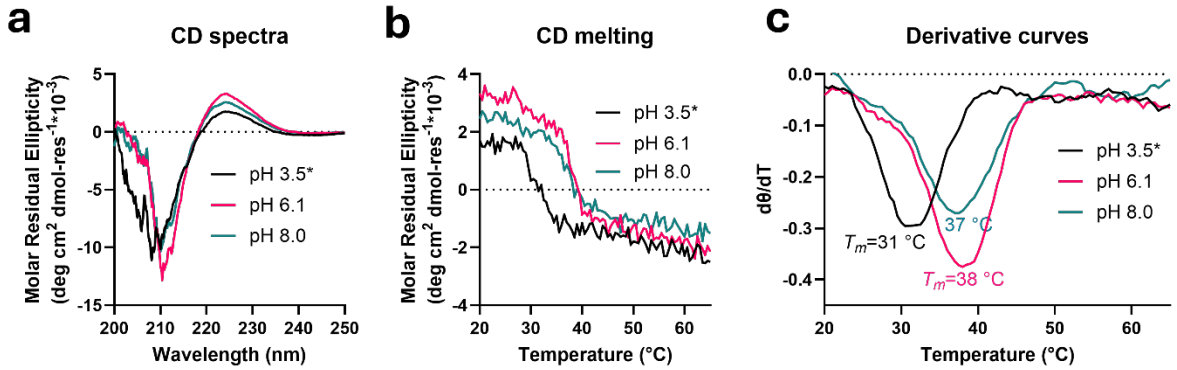

Figure S26. *CD characterization of DES-1 peptide folded at pH 3.5, 6.1 and 8.0. After 3 weeks of equilibration, the peptide folded in pH 3.5 precipitated and the solution became turbid which may impact the quality of CD measurement. a) CD spectra. b) CD melting curves. Signal was monitored at 224 nm. c) The first-order derivative curves of the melting curves in b).*

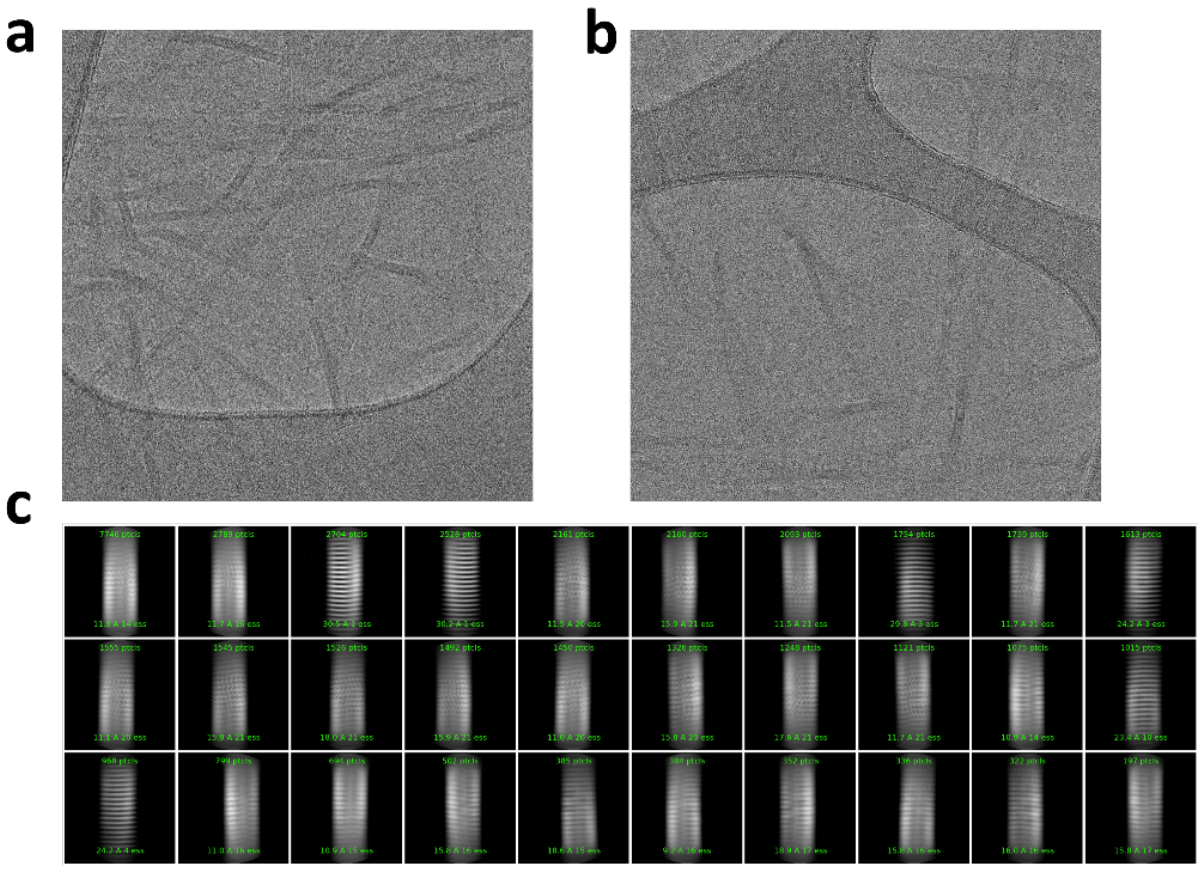

Figure S27. *Additional representative cryo-EM results of DES-1 peptide at pH 3.5.*

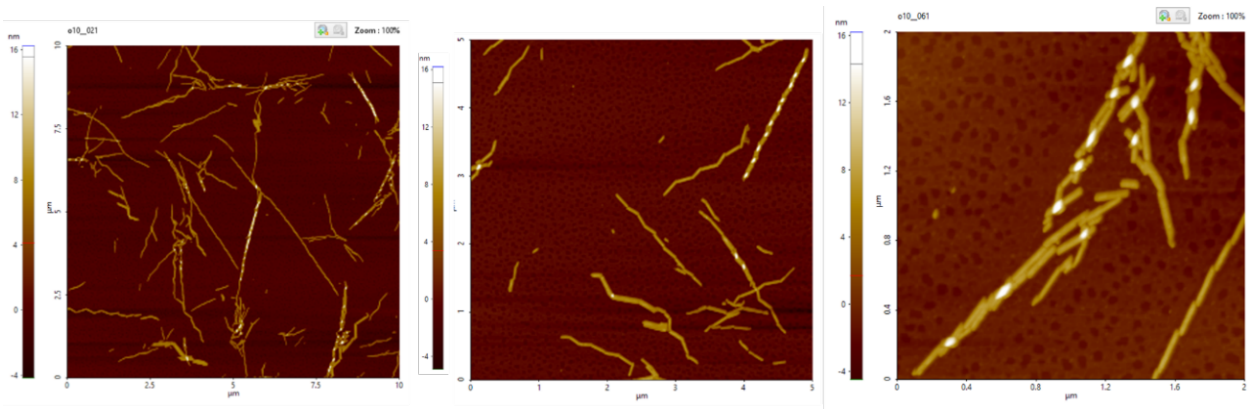

Figure S28. *Additional AFM results of DES-1 peptide assembled at pH 3.5.*

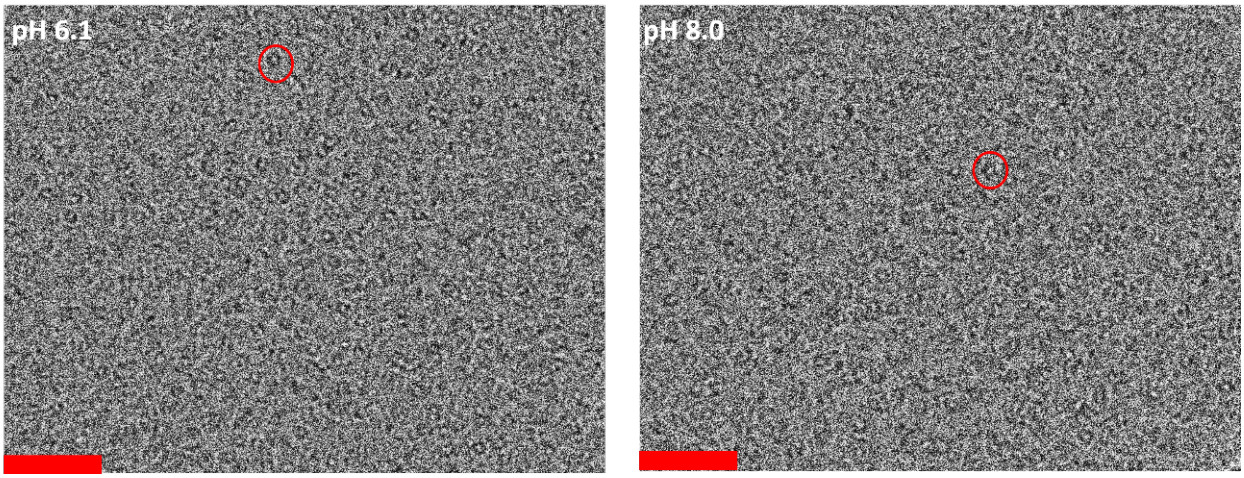

Figure S29. *Representative raw cryo-EM image of DES-1 peptide at pH 6.1 and 8.0. The scale bars are 50 nm.*

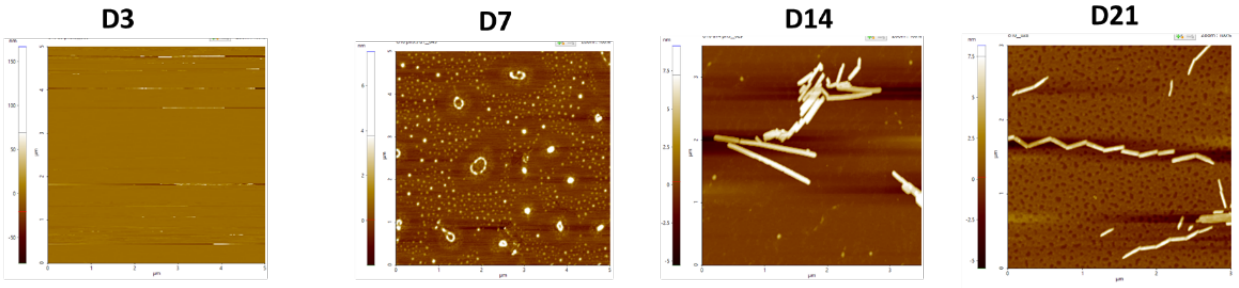

Figure S30. *AFM results of DES-1 peptide assembled at pH 3.5 after different time of equilibration.*

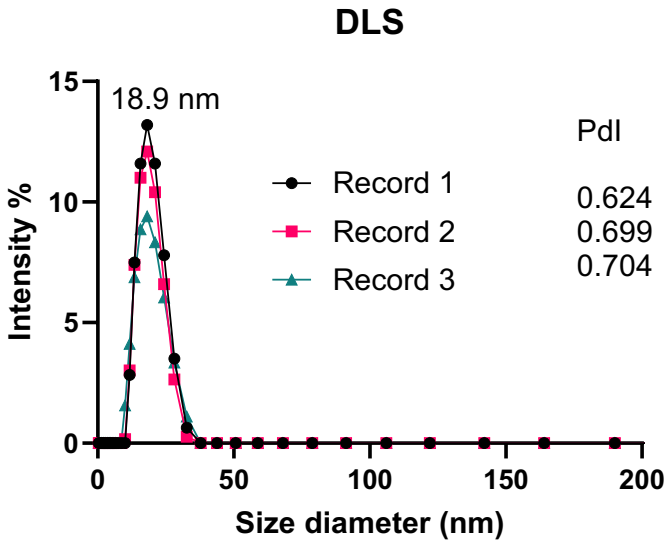

Figure S31. *Dynamic light scattering of RD1 nanoparticles.*

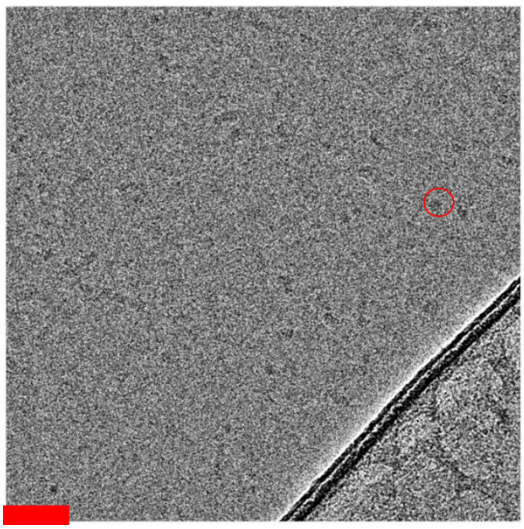

Figure S32. *Representative cryo-EM image of RD1 peptide assembly. The red circle highlights one of the RD1 protein nanoparticles. Scale bar is 50 nm.*

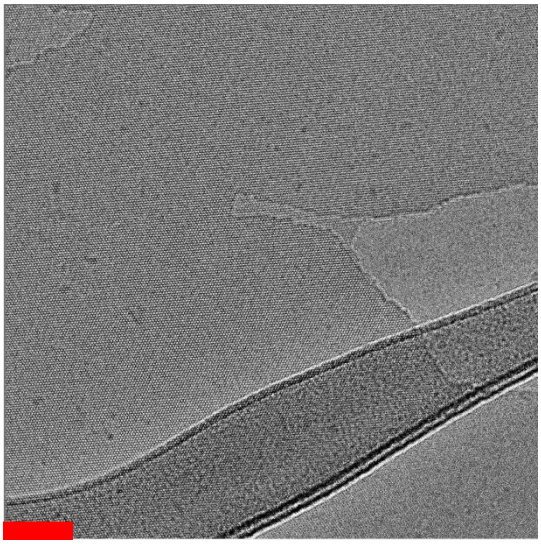

Figure S33. *Additional representative cryo-EM image of the RD2 nanosheets. The scale bar is 50 nm.*

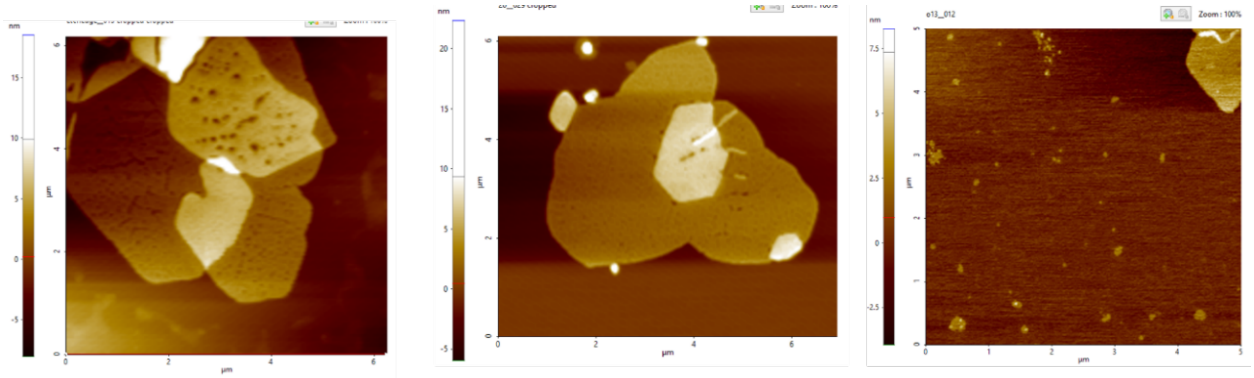

Figure S34. *Additional AFM characterization results of RD2 two-dimensional nanosheets.*

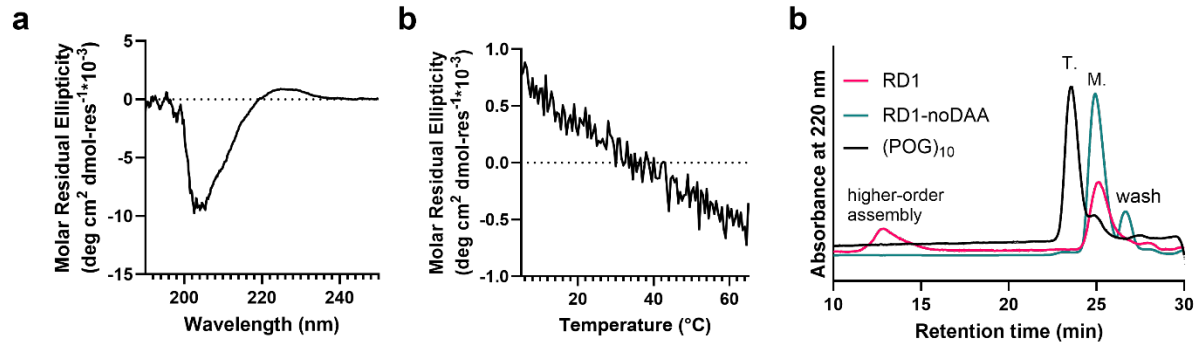

Figure S35 *CD and SEC characterization of RD1-no DAA assemblies. a) CD spectra, b) CD melting curve, c) SEC chromatography and RD1 and RD1-no DAA and the trimer standard of (POG)$_{10}$.*

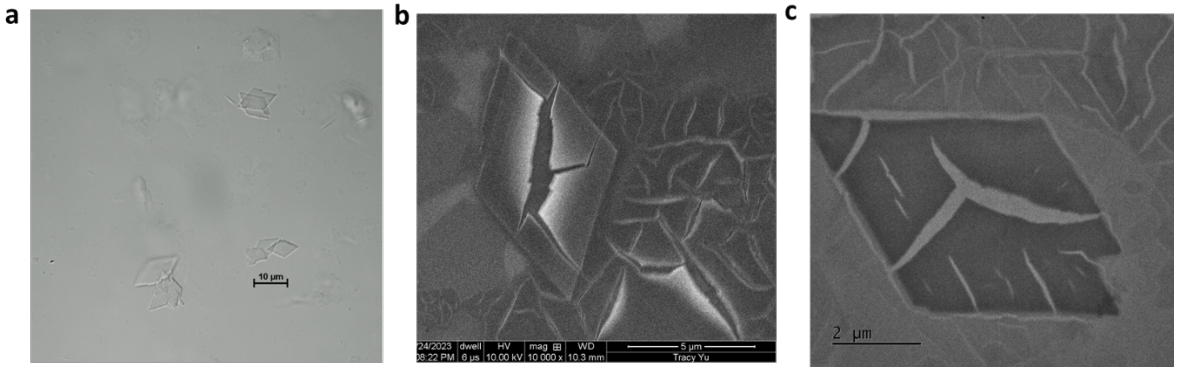

Figure S36. *a) optical microscope of RD2-No DAA peptide. Microcrystals were observed. b) SEM of RD2-No DAA. c) TEM of RD2-No DAA microcrystals.*

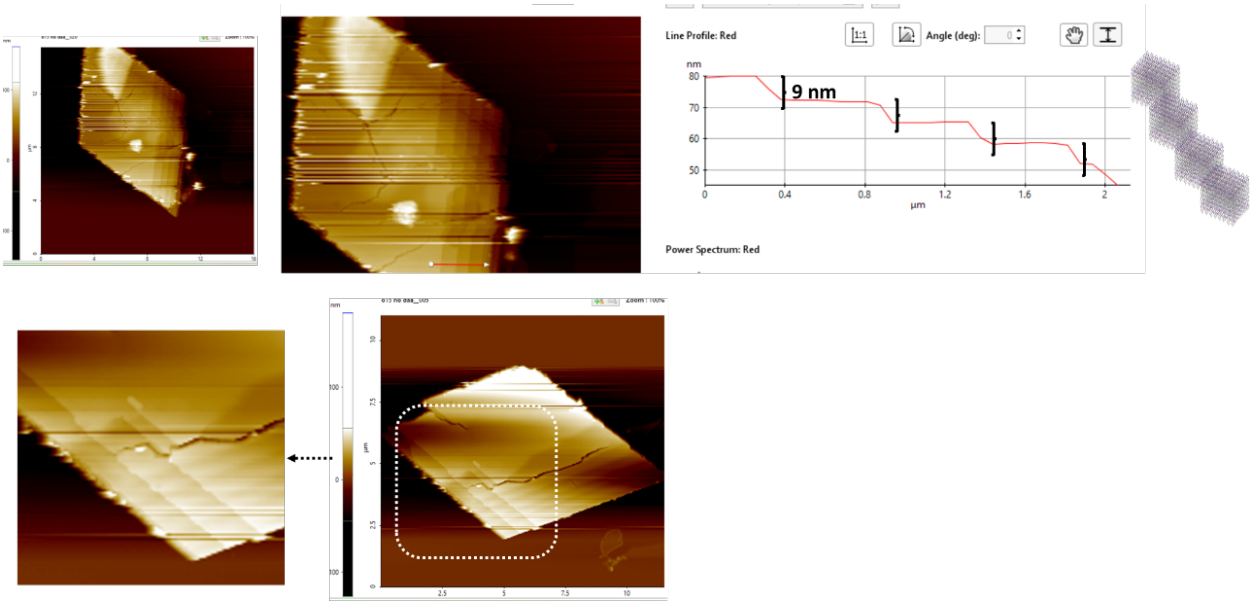

Figure S37. *AFM characterization of RD2-No DAA microcrystals.*

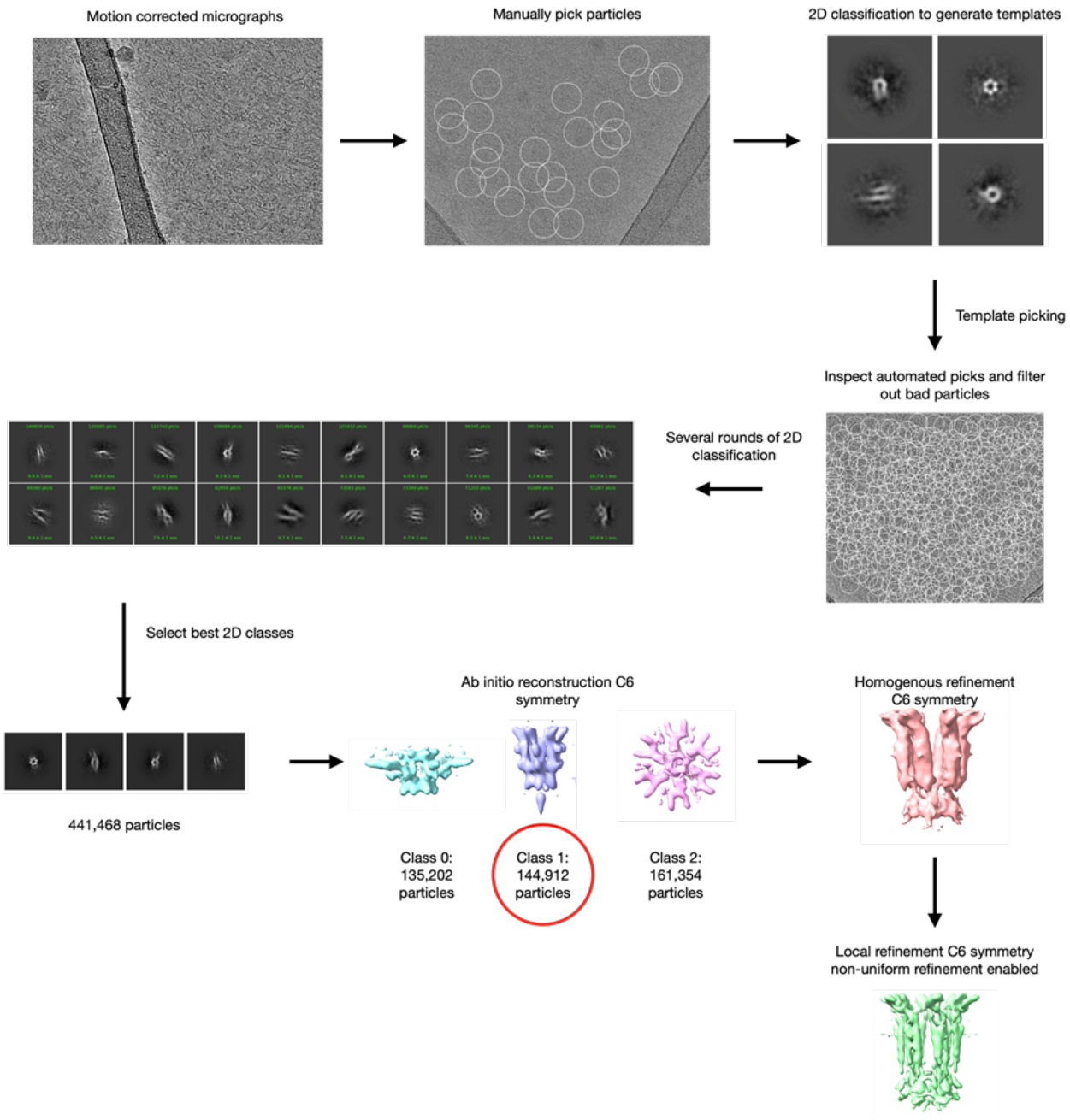

Figure S38*. General workflow for single particle cryo-EM processing and 3D reconstruction of the SPA octadecamer.*

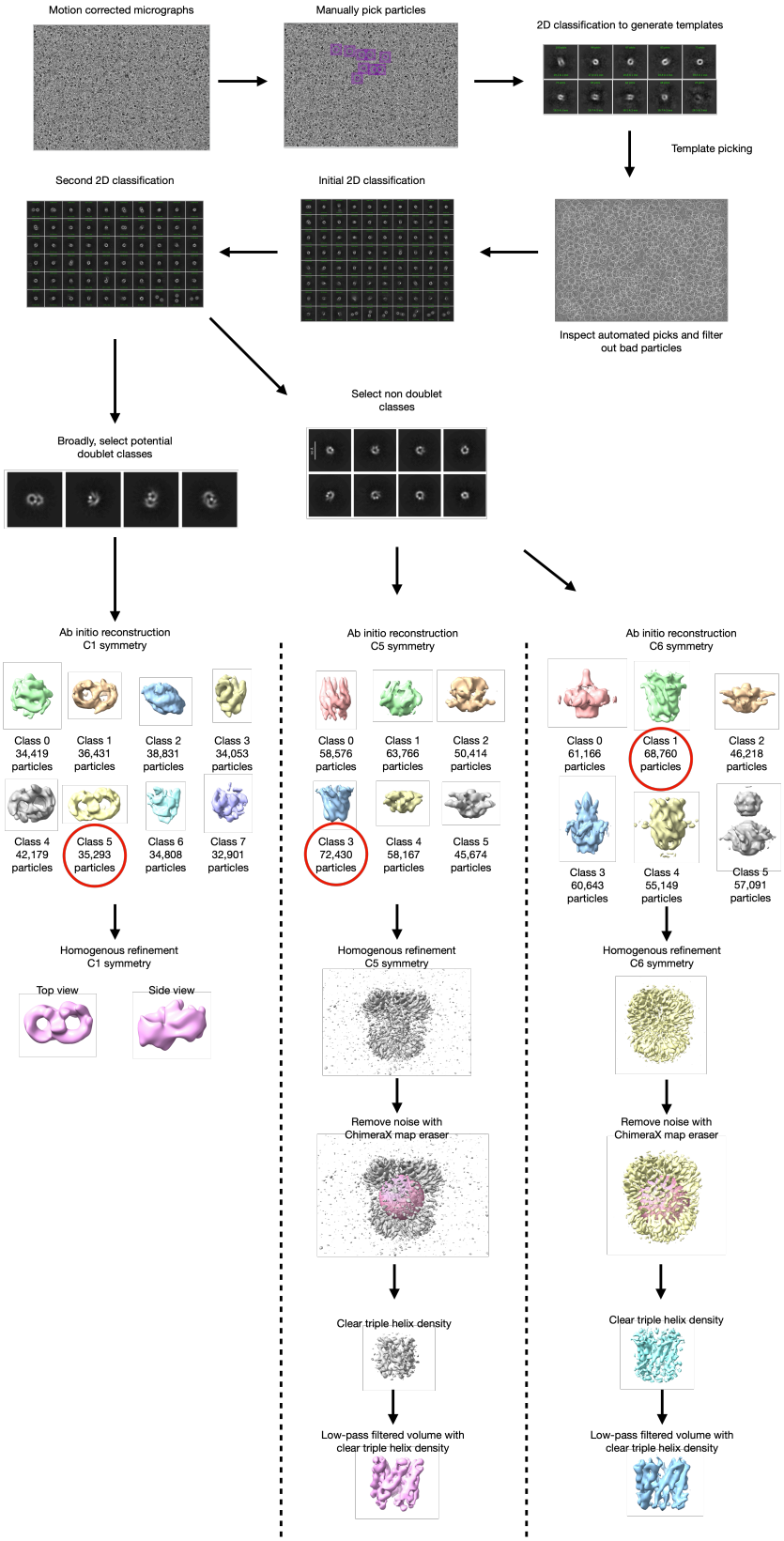

***Figure S39**. Cryo-EM single particle reconstruction workflow for the DES-1 peptide assemblies at pH 6.1.*

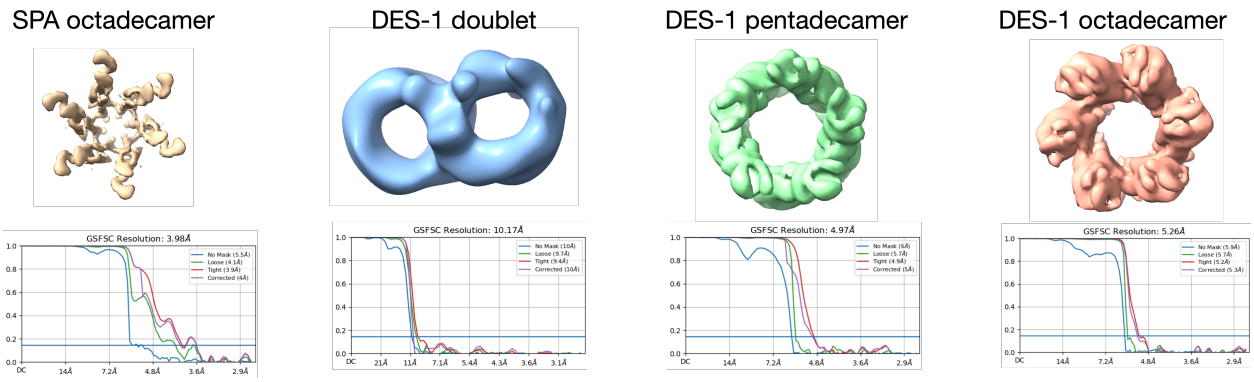

Figure S40 *Map:map Fourier shell correlation(FSC) curves for the structures solved in this study. The resolutions were estimated using the 0.143 FSC criteria. Note that for the SPA assembly the resolution estimate of 4 Å appears to be an overestimate when compared to the information present in the density map (4.5-5 Å resolution).*

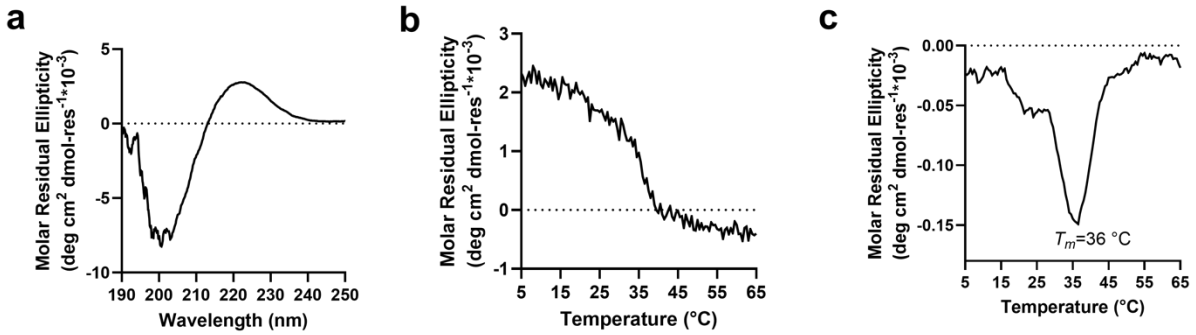

Figure S41. *CD characterization of RD2-noDAA assembly. A) CD spectrum b) thermal denaturation c) First derivative of the thermal denaturation presented in (b).*

Supplementary References

(1) The UniProt Consortium. UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* **2023**, *51* (D1), D523–D531. https://doi.org/10.1093/nar/gkac1052.

(2) Yu, L. T.; Hartgerink, J. D. Selective Covalent Capture of Collagen Triple Helices with a Minimal Protecting Group Strategy. *Chem. Sci.* **2022**, *13* (9), 2789–2796. https://doi.org/10.1039/D1SC06361H.