Electronic Supplementary Material (ESI) for Soft Matter.

Electronic Supporting Information

Impact of collagen-like peptide (CLP) heterotrimeric triple helix design on melting transition and assembly: A coarse-grained molecular dynamics simulation study

Phillip A. Taylor[1], April M. Kloxin[1,2]*, Arthi Jayaraman[1,2]*

Affiliations:
[1]Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE 19716 USA
[2]Department of Materials Science and Engineering, University of Delaware, Newark, DE 19716 USA

*Corresponding Authors
Email: akloxin@udel.edu
Email: arthij@udel.edu

Table S1. Peptide designs and solutions conditions explored in this study. A total number of 84 unique combinations of peptides and solution conditions are explored in this study.

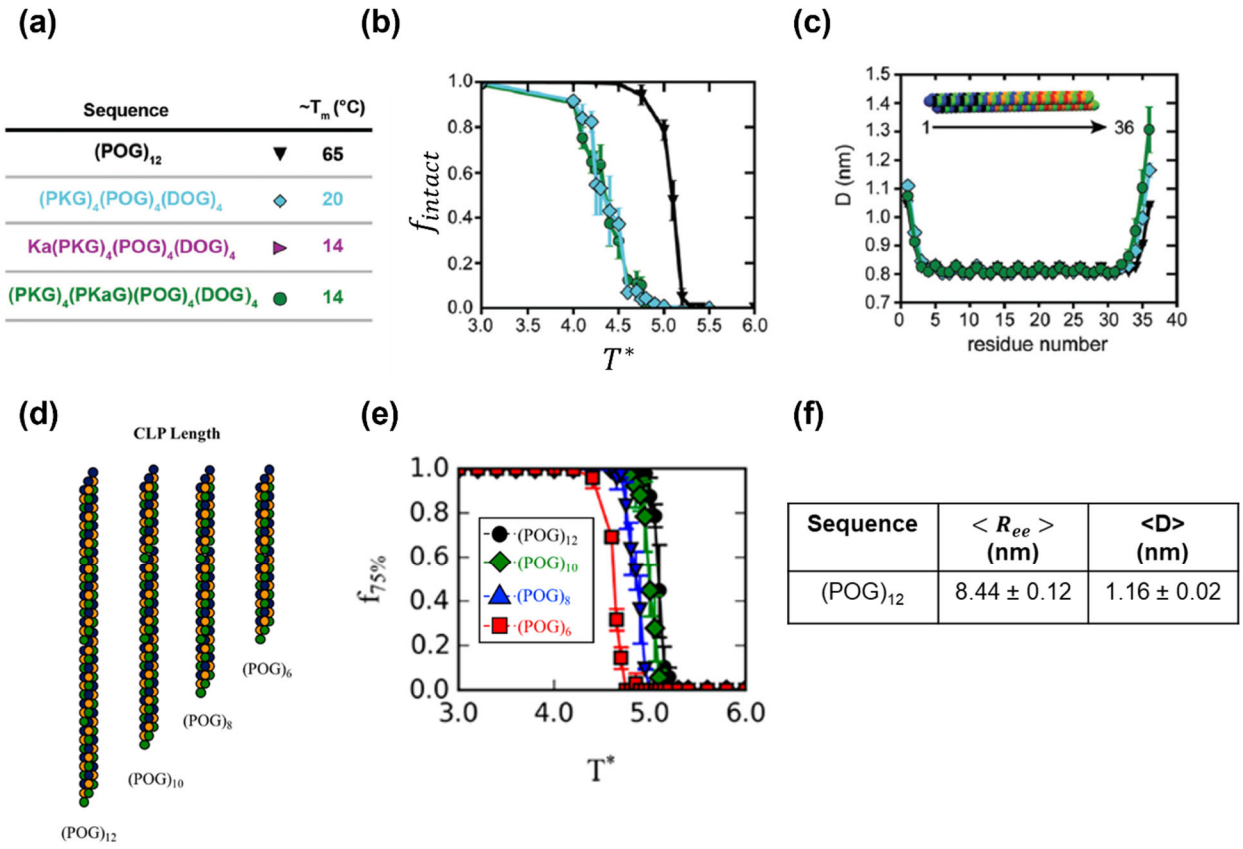
Δ	Number of sticky ends	CLP concentrations	(POG) families
2	1	0	10
4	2	0.3	12
6	-	1	-
-	-	2.5	-
-	-	4	-
-	-	5	-
-	-	20	-

Table S2. Number of CLP heterotrimeric triple helices as a function of CLP concentration in a cubic simulation box of size 110σ. All concentrations are reported as total CLP molar concentrations and therefore the concentration of each individual chain within a heterotrimer triple helix contributes a third of the total molar concentration.

CLP concentration (mM)	Number of triple helices
0	1
0.3	10
1	34
2.5	84
4	134
5	167
20	668

Experimental validation of coarse-grained (CG) CLP model

Previous efforts by Condon and Jayaraman[1] and Kloxin and co-workers[2] have shown that coarse-grained melting curves and triple helical diameters show semi-quantitative agreement with experimental measurements. Computational melting curves obtained using the CG model reproduce the destabilizing effect of charged amino acids on the melting transition of the triple helix and triple helical diameters show the destabilizing impact of nonnatural amino acids on triple helical configurations. The CG model of Condon and Jayaraman[1] also reproduced the experimentally observed effect of CLP chain length on triple helix melting and yields end-to-end distances triple helical diameters that are in quantitative agreement with experimental measurements.



Sequence	$< R_{ee} >$ (nm)	<D> (nm)
$(POG)_{12}$	8.44 ± 0.12	1.16 ± 0.02

Figure S1. Comparison of coarse-grained CLP model results with experiments. (a) Experimental melting temperatures for CLPs containing uncharged, charged, and nonnatural amino acids. (b) Computational melting curves and (c) triple helical diameters obtained using the CG CLP model of Condon and Jayaraman. (d) Schematic showing four triple helices of varying CLP chain lengths, (e) corresponding computational melting curves, and (f) triple helical end-to-end distances (R_{ee}) and diameters obtained using the CG CLP model of Condon and Jayaraman. Reproduced from refs. 1 and 2 with permission from the American Chemical Society and Royal Society of Chemistry, respectively.

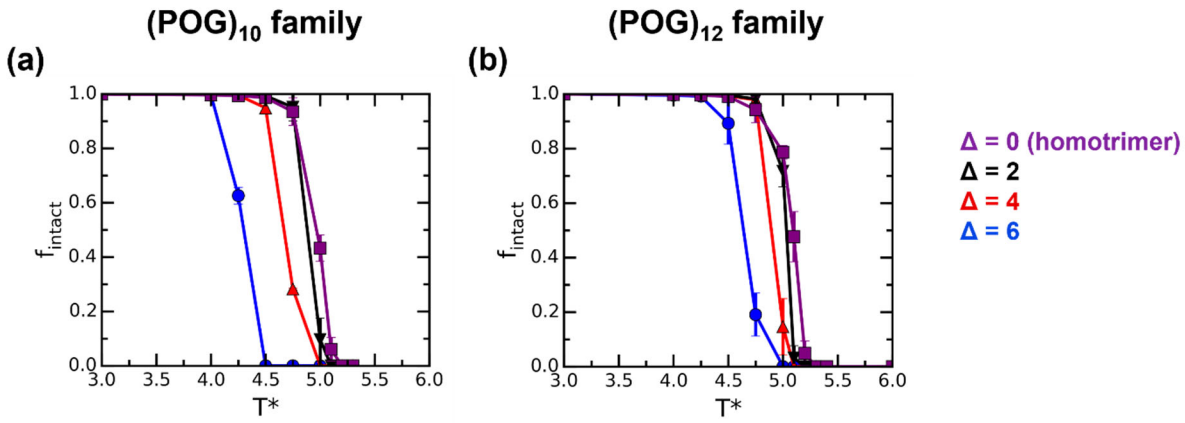


Figure S2. Melting curves of CLP heterotrimers obtained from CG MD simulations for *two sticky ended* heterotrimeric triple helices. Ensemble average fraction of intact glycine triplets (f_{intact}) as a function of reduced temperature (T*) for sticky end lengths, Δ, of 2, 4, and 6 for (a) the $(POG)_{10}$ family and (b) the $(POG)_{12}$ family. All melting curves were obtained at a CLP concentration of 0.3 mM.

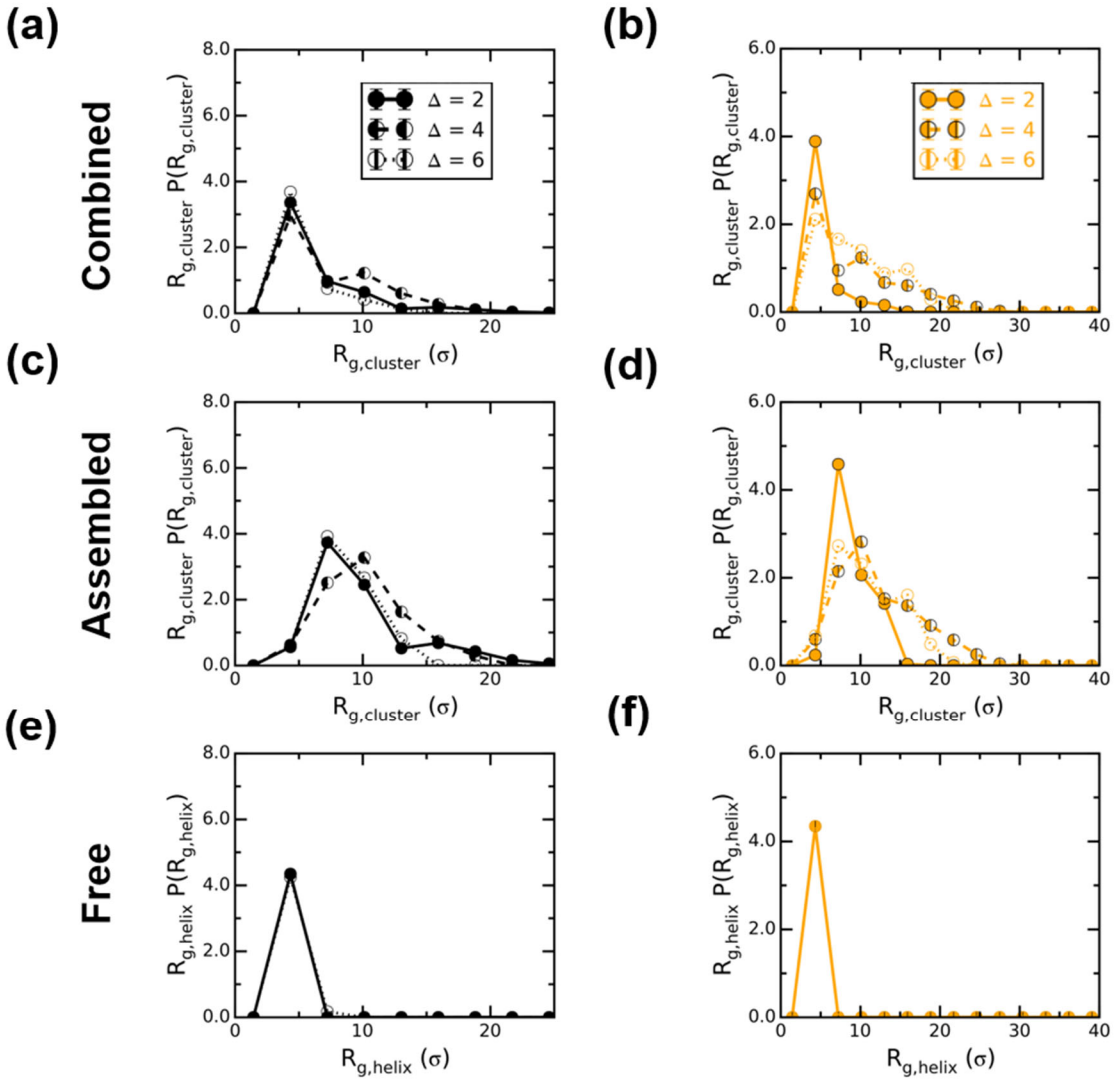


Figure S3. Weighted probability distributions of radii of gyration, $R_{g,cluster}$, and number of helices per cluster, $N_{helix,cluster}$, of CLP heterotrimers obtained from CG MD simulations for the (a,c,e) one-sticky ended and (b,d,f) two-sticky ended $(POG)_{10}$ family at a CLP concentration of 1 mM at T* = 3.0. Distributions of $R_{g,cluster}$ (a,b) for clusters of CLP heterotrimers with at least 1 helix per cluster. Distributions of $R_{g,cluster}$ (c,d) for clusters of CLP heterotrimers with at least 2 helices per cluster. (e,f) $R_{g,cluster}$ distribution for clusters made up of a single "free" triple helix.

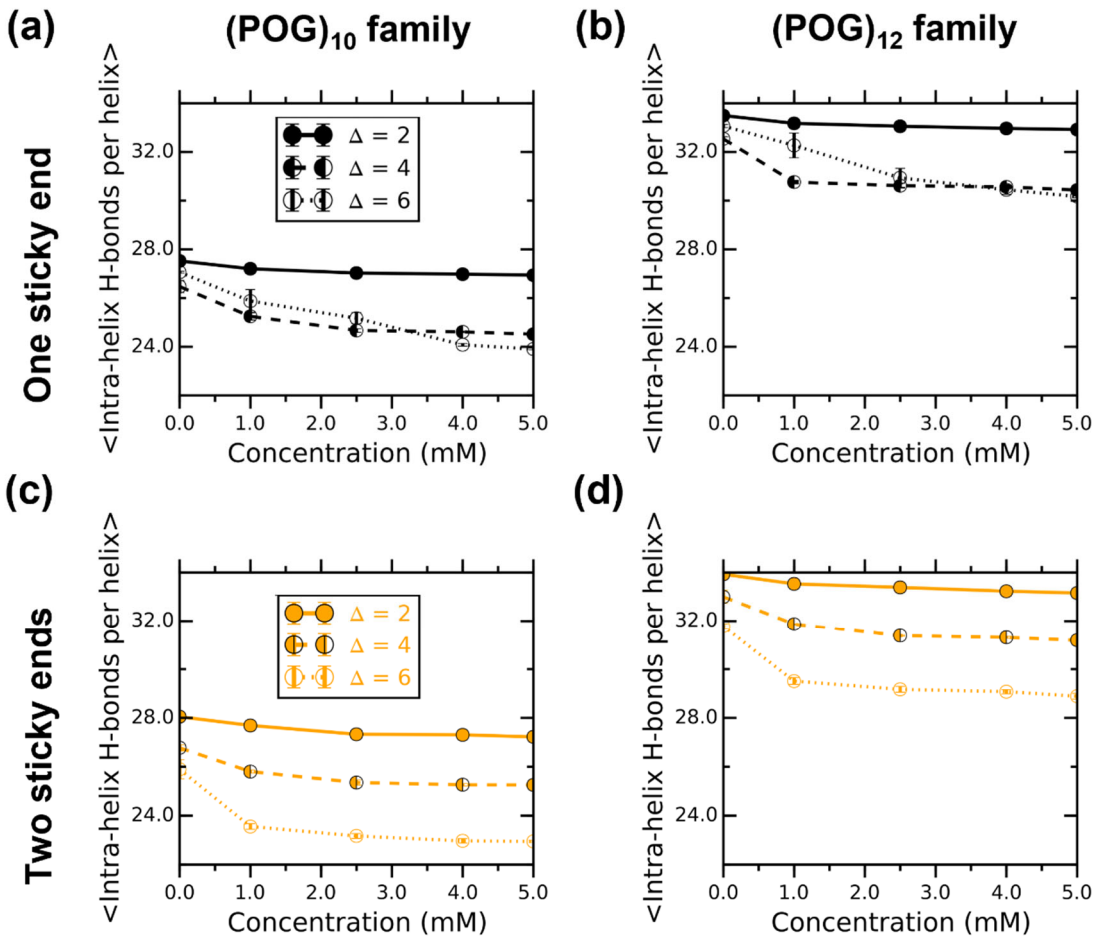


Figure S4. Intra-helix hydrogen bonds per helix as a function of CLP concentration, sticky end length, Δ, and number of sticky ends obtained from CG MD simulations for the $(POG)_{10}$ (a,b) and $(POG)_{12}$ families at T* = 3.0. Inter-helix H-bonds per helix are also shown for a one sticky ended design (a,c) and a two sticky ended design (b,c).

Quantification of loop structures in CG MD simulations

To quantify the extent of loop formation as a function of CLP heterotrimer design and solution conditions, we calculate the ensemble averaged fraction of CLP strands forming loop structures obtained from CG MD simulations. Since the CG model of Condon and Jayaraman[1] was not designed to accurately reproduce secondary structures in terms of backbone and side chain conformations (e.g., dihedrals), we define a loop structure as a single CLP chain that forms at least 4 intra-chain hydrogen bonds. We choose this metric as it defines a loop structure based on the maximum number intra-chain hydrogen bonds observed for all systems (maximum at $\Delta = 6$ at 0 mM) and ensures that multiple intra-chain hydrogen bonds are required to define a loop. Moreover, this definition also yields results consistent with visualization (Figure 6) and inter-helix hydrogen bonding analyses (Figure 5) which show the presence of loop structures for one sticky ended designs and low CLP concentrations (e.g., 0 mM).

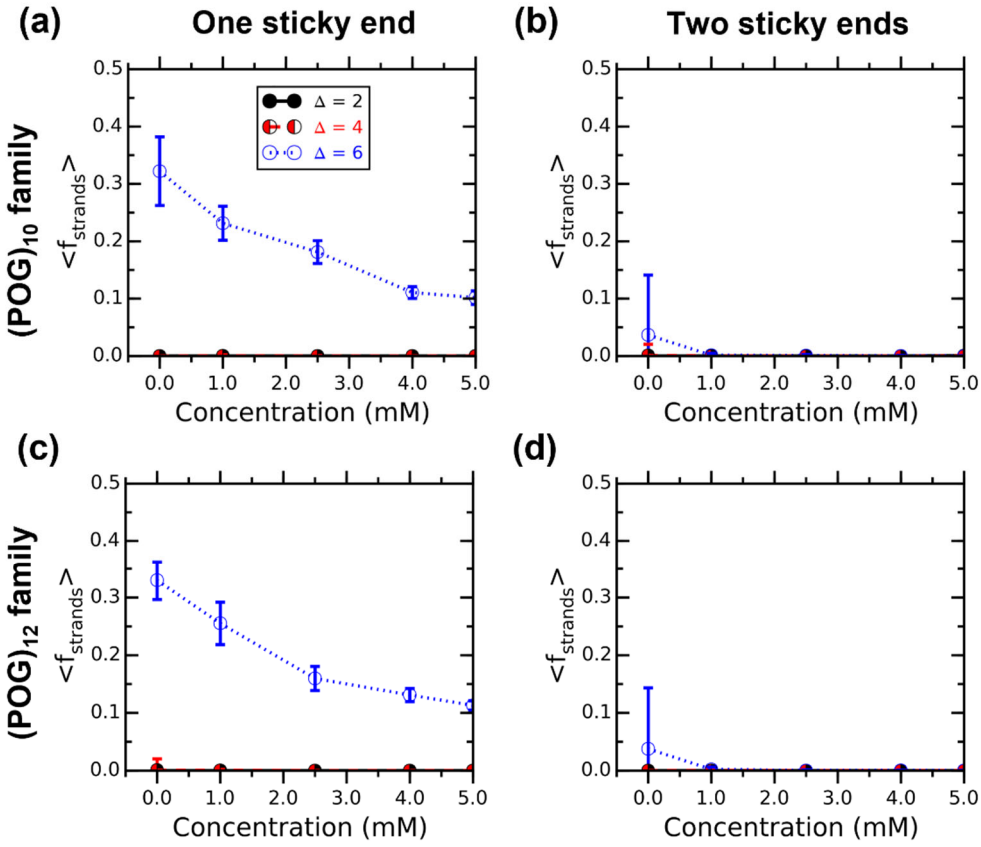


Figure S5. Fraction of CLP strands forming loop structures ($<f_{strands}>$) obtained from CG MD simulations for the $(POG)_{10}$ (a,b) and $(POG)_{12}$ families. Fraction of CLP strands forming loop structures are shown for a one sticky ended design (a,c) and a two sticky ended design (b,d).

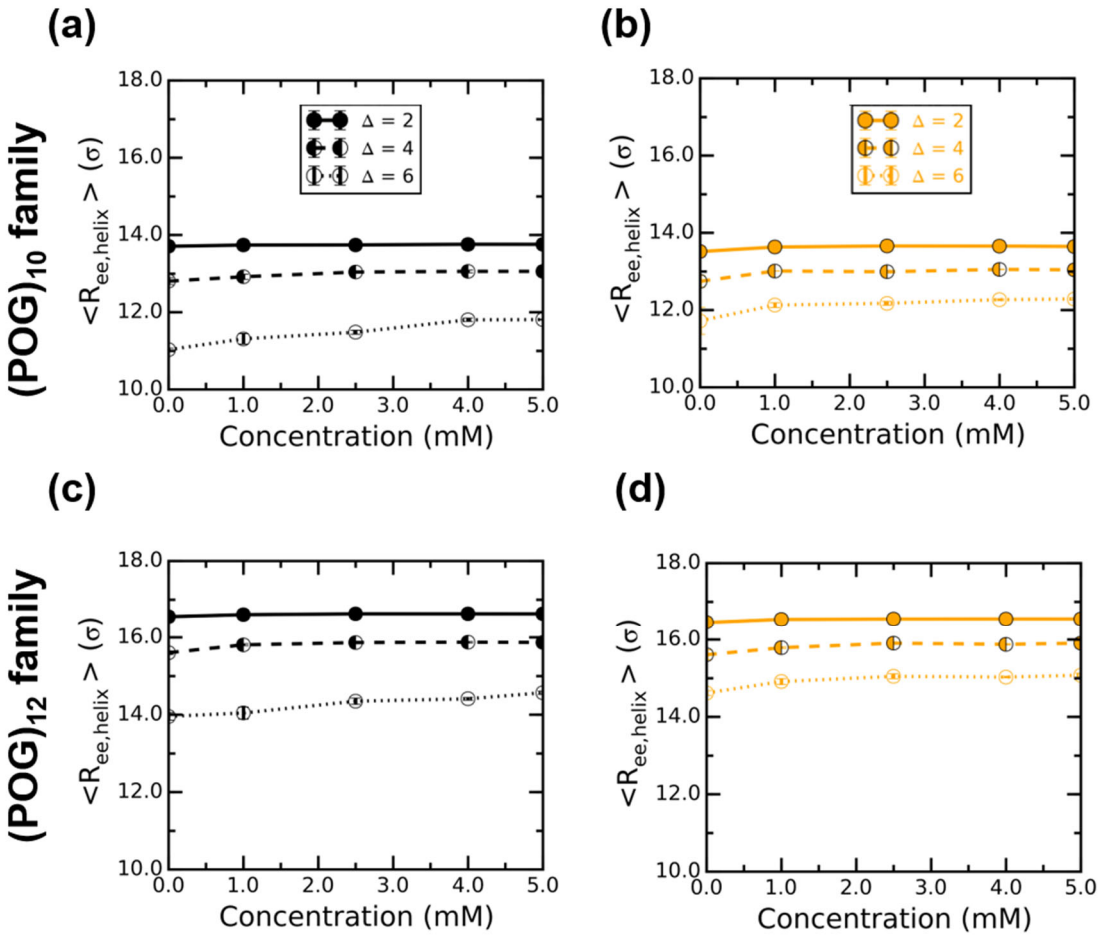


Figure S6. Ensemble averaged end-to-end distances (<R_{ee}>) of (a,c) one sticky ended and (b,d) two sticky ended heterotrimeric triple helices as a function of CLP concentration and sticky end length, Δ, obtained at a reduced temperature of T* = 3.0. <R_{ee}> are shown for the $(POG)_{10}$ family of peptides (a,b) and the $(POG)_{12}$ family of peptides (c,d). <R_{ee}> values are obtained from three trials per CLP concentration, (POG) family, and Δ.

Calculation of minimum repeat units required for loop formation

Below we show a simple calculation to obtain a lower bound estimate of the minimum sticky end length (i.e., (POG) repeat units) that is required to form a loop structure by ignoring entropic penalties associated with forming a loop from a free CLP chain. We note, however, that the energetic-entropic considerations of the formation of loop in a chain are not independent of energetic and entropic considerations arising from the solution concentration (i.e., presence of other chains in the system which affects crowding as well as provides opportunities for hydrogen-bond formations). However, writing out expressions for such entropic terms is not trivial. Therefore, we conduct simulations that include all entropic and energetic considerations leading to the 'optimal design' information.

If one completely ignored all the above entropic considerations, at the infinitely dilute limit of one CLP strand, the number of intra-strand hydrogen bonds required to balance the bending cost of the chain upon forming the loop can be written as

$$(\Delta U)_{h-bond} = (\Delta U)_{bending} \tag{S1}$$

where $(\Delta U)_{h-bond}$ is the change in potential energy upon the formation of a hydrogen bond and $(\Delta U)_{bending}$ is the change in potential energy upon bending of three adjacent backbone beads.

The terms in equation (1) can be rewritten as follows:

1) Using the number of hydrogen bonds formed, *x*, times the hydrogen bond strength, obtained from the well-depth of the Lennard-Jones interaction between hydrogen bonding donors and acceptors, $\varepsilon_{D-A}^{HB} = 50.4\varepsilon$, we get

$$(\Delta U)_{h-bond} = (50.4\varepsilon)x \tag{S2}$$

2) The force constant for harmonic angle potentials involving adjacent backbone beads, BB-BB-BB, is defined as $k_{BB-BB-BB} = 20\frac{\varepsilon}{rad^2}$, as described in the methods section.

3) If we assume an angle of $\theta = \frac{\pi}{2}$ rad (90°) for loop formation with an equilibrium BB-BB-BB angle, θ_0, of π rad (180°) as described in the methods section.

4) As two equivalent bending terms with θ are required for the bending energy to form a loop, we have a factor of 2 for the bending energy.

$$(\Delta U)_{bending} = 2k(\theta - \theta_0)^2 \quad \text{(S3)}$$

Combining equations (1-3) we get:

$$(50.4\varepsilon)x = 2(20\frac{\varepsilon}{rad^2})(\theta - \pi\, rad)^2$$

For $\theta = \frac{\pi}{2}$ rad (90°), x yields a value of 1.96, i.e., ~ 2 intra-chain hydrogen bonds. As θ decreases from $\frac{\pi}{2}$ rad, x increases. Regardless of the value of θ needed for formation of the loop, this calculation ignores the entropic penalty associated with forming a loop. As a result, the value of x is an underestimate, and a larger energetic gain would be required to balance the entropic penalty. We can state that a minimum number of repeat units (i.e., sticky end length) of two repeat units would be needed for loop formation, assuming each intra-chain hydrogen bond occurs between only two unique repeat units. There are, however, two hydrogen bonding sites per repeat unit (i.e., one on P and another on G), so the reported value is the minimum number of two repeat units also accounts for the inability of a single repeat unit (POG) to form a loop due to the 90° angle potential (constraint) involving HB-BB-BB angles which prevents hydrogen bonding within a single repeat unit in a loop.

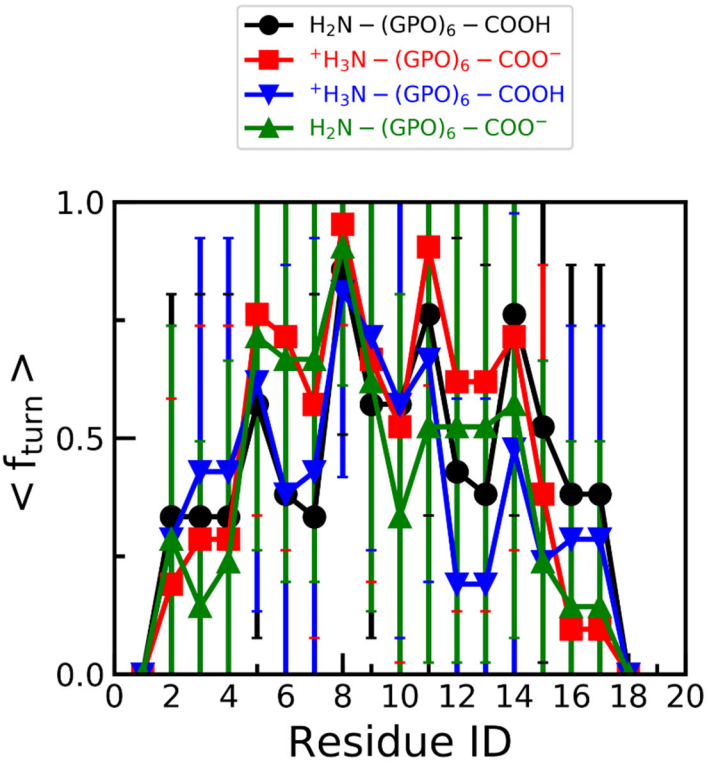


Figure S7. Secondary structure analyses obtained using the STRIDE algorithm[3] for four $(GPO)_6$ systems obtained using atomistic simulations. Data is shown for the ensemble averaged fraction of configurations in which each, unique residue forms turn structures, $\langle f_{turn} \rangle$, as a function of end group functionalization.

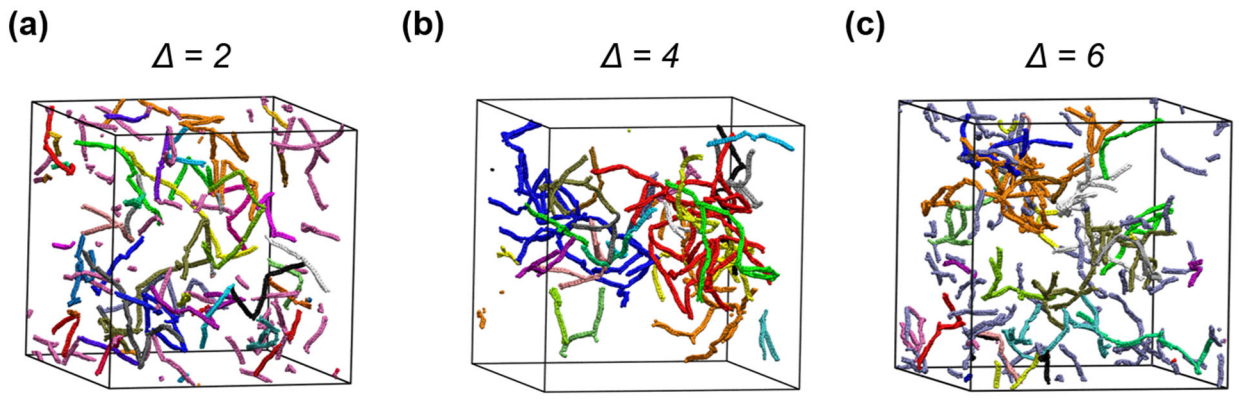


Figure S8. Representative simulation snapshots of *one sticky ended* heterotrimeric triple helices for the $(POG)_{12}$ family as a function of sticky end length, Δ at T* = 3.0. Snapshots are shown at the end of production simulations of 10^7 time steps for Δ = 2, 4, and 6 (a,b,c). Clusters of CLP heterotrimeric triple helices are color coded based on cluster ID.

Table S3. Ensemble average fraction of percolated configurations for systems of one sticky ended CLP heterotrimers at a CLP concentration of 20 mM and reduced temperature of T* = 3.0 Ensemble averages were obtained by averaging over the entire trajectory and all three trials and error bars show the standard deviation across three trials. Percolation analyses are shown for systems in which the combined and assembled $R_{g,cluster}$ distribution showed a cluster which spanned the simulation box (i.e., with a peak near half of the box length, 55 σ).

Δ	**(POG) family**	$< \tau_{perc} >$
4	10	0.95 ± 0.05
4	12	1.00 ± 0.00
6	12	1.00 ± 0.00
6	10	1.00 ± 0.00

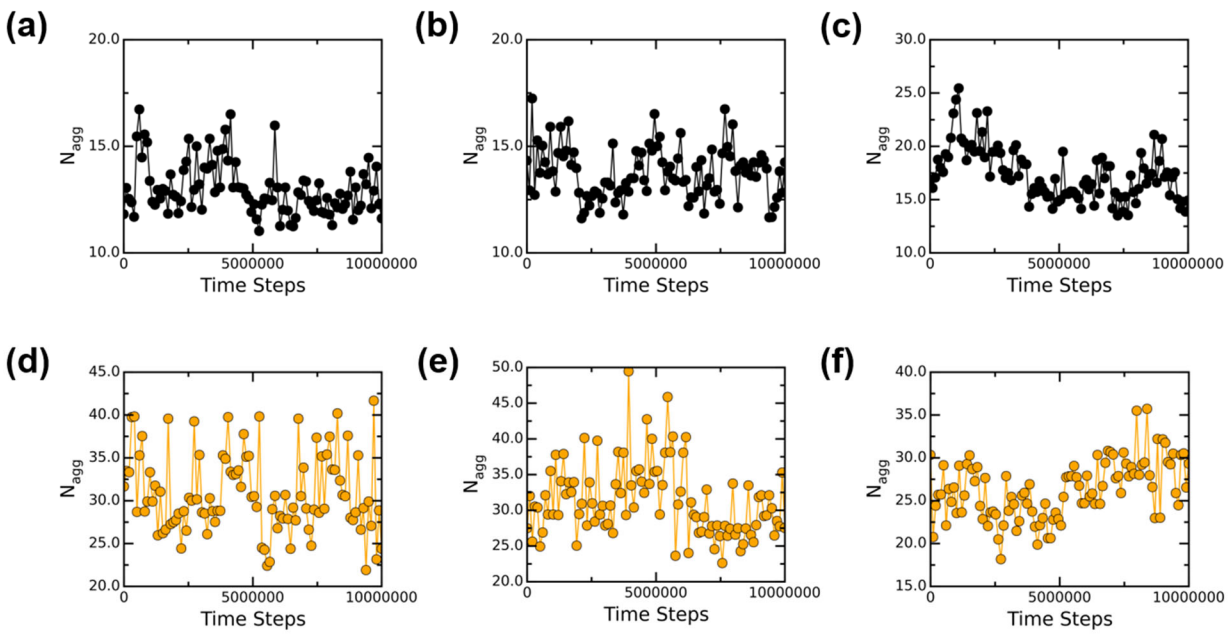


Figure S9. Average number of helices per cluster or aggregation number, N_{agg}, as a function of time steps for a one sticky ended triple helices (a,b,c) and two sticky ended triple helices (d,e,f) for the $(POG)_{12}$ family obtained from CG simulations at T* = 3.0. Data is shown for all three trials for one sticky ended triple helices (a,b,c) and all three trials for two sticky ended triple helices (d,e,f).

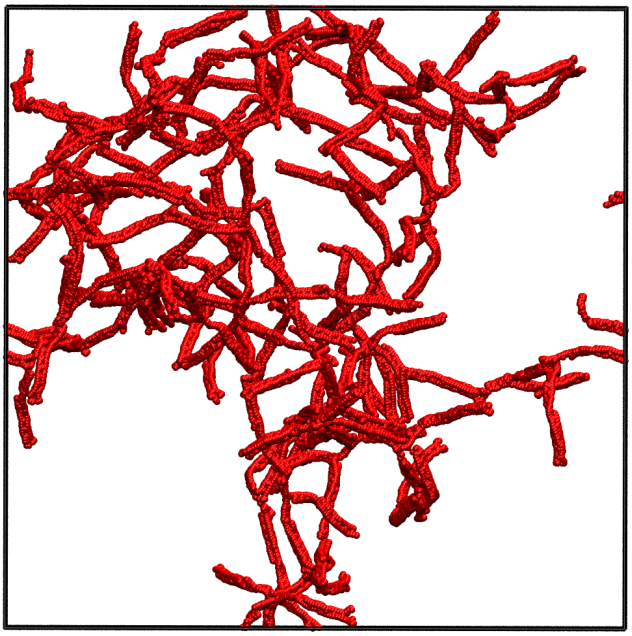


Figure S10. Representative simulation snapshot of one sticky ended heterotrimeric triple helices for the $(POG)_{10}$ family for $\Delta = 4$ at $T^* = 3.0$. For clarity, only the largest (red) cluster is shown at the end of production simulations of 10^7 time steps. Clusters of CLP heterotrimeric triple helices are color coded based on cluster ID.

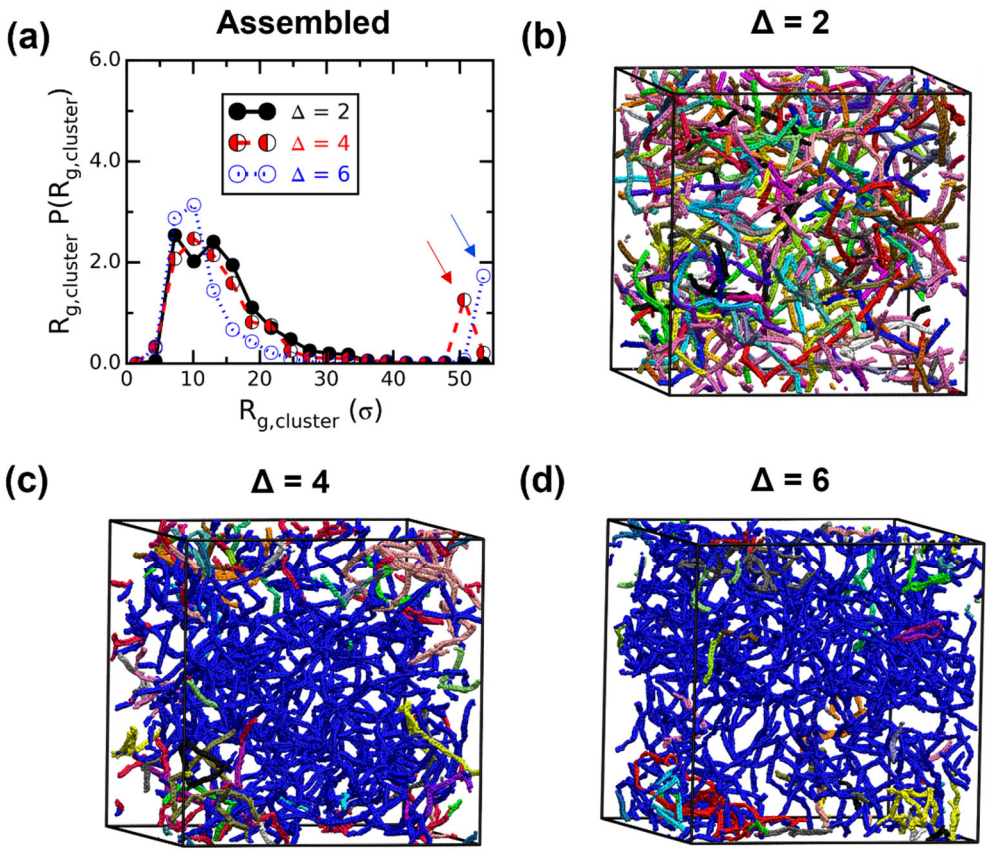


Figure S11. Weighted probability distributions of radii of gyration, $R_{g,cluster}$ of CLP heterotrimers for the $(POG)_{10}$ family obtained from CG MD simulations for the two-sticky end design at a CLP concentration of 20 mM (a,c,e) and representative simulation snapshots at T* = 3.0 (b,d,f). Distributions of $R_{g,cluster}$ for clusters of CLP heterotrimers with at least 1 helix per cluster (a), clusters of CLP heterotrimers with at least 2 helices per cluster (c), and clusters made up of a single "free" triple helix (e). Simulation snapshots are shown for Δ = 2 (b), 4 (d), and 6 (f).

Table S4. Ensemble average fraction of percolated configurations for systems of two sticky ended CLP heterotrimers at a CLP concentration of 20 mM and reduced temperature of T* = 3.0 Ensemble averages were obtained by averaging over the entire trajectory and all three trials and error bars show the standard deviation across three trials. Percolation analyses were conducted for systems in which the combined and assembled $R_{g,cluster}$ distribution showed a cluster which spanned the simulation box (i.e., with a peak near half of the box length, 55 σ).

Δ	**(POG) family**	$< \tau_{perc} >$
4	12	1.00 ± 0.00
6	12	1.00 ± 0.00
4	10	1.00 ± 0.00
6	10	1.00 ± 0.00

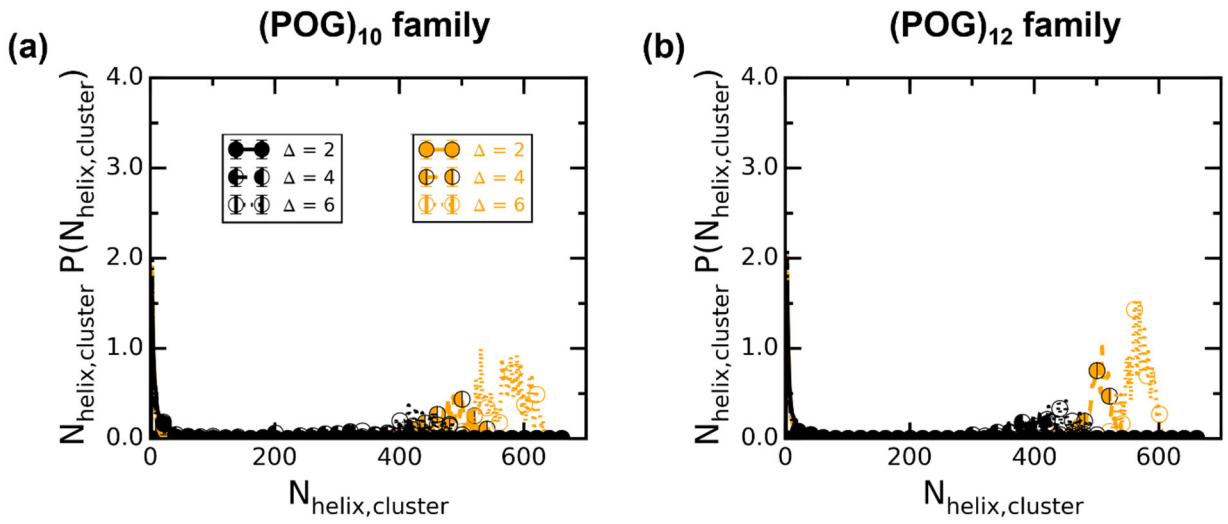


Figure S12. Weighted probability distributions of $N_{helix,cluster}$ for one sticky ended triple helices (black) and two sticky ended triple helices (orange) for the (a) $(POG)_{10}$ and (b) $(POG)_{12}$ families as a function of sticky end length, Δ at T* = 3.0. Distributions of $N_{helix,cluster}$ are shown for all clusters of CLP heterotrimers with at least two helices per cluster (assembled).

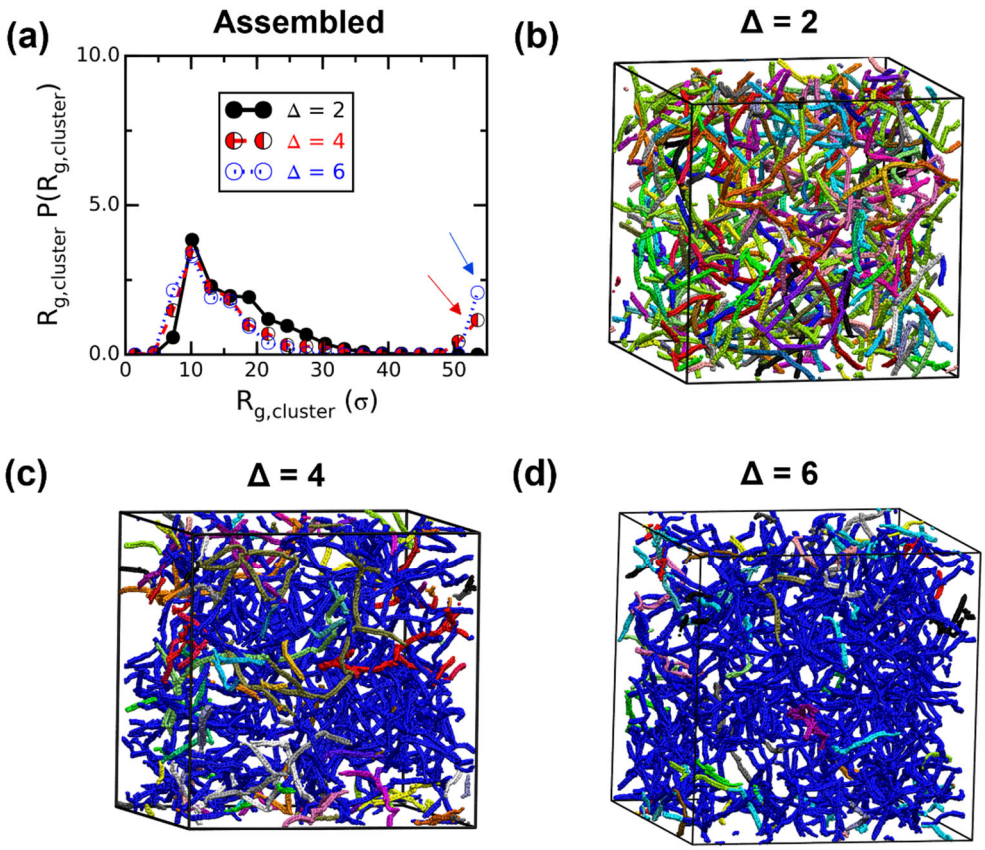


Figure S13. Weighted probability distributions of radii of gyration, $R_{g,cluster}$ of CLP heterotrimers for the $(POG)_{12}$ family obtained from CG MD simulations for the two-sticky end design at a CLP concentration of 20 mM (a,c,e) and representative simulation snapshots at T* = 3.0 (b,d,f). Distributions of $R_{g,cluster}$ for clusters of CLP heterotrimers with at least 1 helix per cluster (a), clusters of CLP heterotrimers with at least 2 helices per cluster (c), and clusters made up of a single "free" triple helix (e). Simulation snapshots are shown for Δ = 2 (b), 4 (d), and 6 (f).

References

1. Condon, J. E.; Jayaraman, A., Development of a Coarse-Grained Model of Collagen-Like Peptide (CLP) for Studies of CLP Triple Helix Melting. *Journal of Physical Chemistry B* **2018,** *122* (6), 1929-1939.
2. Hilderbrand, A. M.; Taylor, P. A.; Stanzione, F.; LaRue, M.; Guo, C.; Jayaraman, A.; Kloxin, A. M., Combining simulations and experiments for the molecular engineering of multifunctional collagen mimetic peptide-based materials. *Soft Matter* **2021,** *17* (7), 1985-1998.
3. Heinig, M.; Frishman, D., STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic acids research* **2004,** *32* (suppl_2), W500-W502.