Effects of Solvent Conditions on the Self-Assembly of Heterotrimeric Collagen-Like Peptide (CLP) Triple Helices: A Coarse-Grained Simulation Study

Journal:	*Soft Matter*
Manuscript ID	SM-ART-03-2023-000374.R1
Article Type:	Paper
Date Submitted by the Author:	25-May-2023
Complete List of Authors:	Taylor, Phillip; University of Delaware, Dept. of Chemical and Biomolecular Engineering Kronenberger, Stephen; University of Delaware, Chemical and Biomolecular Engineering Kloxin, April; University of Delaware, Chemical & Biomolecular Engineering and Materials Science & Engineering Jayaraman, Arthi; University of Delaware, Chemical and Biomolecular Engineering

SCHOLARONE™
Manuscripts

Effects of Solvent Conditions on the Self-Assembly of Heterotrimeric Collagen-Like Peptide (CLP) Triple Helices: A Coarse-Grained Simulation Study

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

1. Department of Chemical and Biomolecular Engineering, University of Delaware, Colburn Lab, 150 Academy St, Newark, DE 19716

2. Department of Materials Science and Engineering, University of Delaware, Pierre S. Du Pont Hall, 127 The Green, Newark, DE 19716

\# Equal contributions

*Corresponding author arthij@udel.edu

Abstract

We perform coarse-grained (CG) molecular dynamics (MD) simulations to investigate the self-assembly of collagen-like peptide (CLP) triple helices into fibrillar structures and percolated networks as a function of solvent quality. The focus of this study is on CLP triple helices whose strands are different lengths (i.e., heterotrimers), leading to dangling 'sticky ends'. These 'sticky ends' are segments of the CLP strands that have unbonded hydrogen-bonding donor/acceptor sites that drive heterotrimeric CLP triple helices to physically associate with one another, leading to assembly into higher-order structures. We use a validated CG model for CLP in implicit solvent and capture varying solvent quality through changing strength of attraction between CG beads representing the amino acids in the CLP strands. Our CG MD simulations show that, at lower CLP concentrations, CLP heterotrimers assemble into fibrils and, at higher CLP concentrations, into percolated networks. At higher concentrations, decreasing solvent quality causes i) the formation of heterogeneous network structures with a lower degree of branching at network junctions and ii) increases in the diameter of network strands and pore sizes. We also observe a nonmonotonic effect of solvent quality on distances between network junctions due to the balance between heterotrimer end-end associations driven by hydrogen bonding and side-side associations driven by worsening solvent quality. Below the percolation threshold, we observe that decreasing solvent quality leads to the formation of fibrils composed of multiple aligned CLP triple helices, while the number of 'sticky ends' governs the spatial extent (radius of gyration) of the assembled fibrils.

1. Introduction

Peptides are one of the major building blocks of biological systems (e.g., extracellular matrix of tissues and organs) and act as a template for synthetic biomimetic materials with controllable structure, physical properties, and responsiveness to stimuli. The programmable structure due to peptide self- or directed assembly and their resulting properties can be tuned by varying chemical composition (i.e., choice of amino acids), sequence, and length (i.e., number of amino acids). Further, their ability to change structure in response to external stimuli (e.g., heat,[1] and light[2]) and or/solvent conditions (e.g., solvents' chemistry, pH,[3] and salt[4]) makes these peptide-based soft materials ideal for drug delivery, cell culture, and tissue engineering. One such class of self-assembling peptides of interest to the biomaterials community is collagen like peptides (CLPs), also denoted as collagen mimetic peptides (CMPs). CLPs are biomimetic polymers that mimic the naturally occurring extracellular matrix protein, collagen. CLPs or CMPs are made up of repeat units of amino acid triplets, (X-Y-G), where X and Y are usually proline (P), and hydroxyproline (O), respectively, and G is glycine. Like native (natural) collagen, CLP exhibits hierarchical structure with CLP single strands forming triple helices primarily due to inter-strand hydrogen bonding and the CLP triple helices assembling to form fibrils and networks of fibrils in aqueous solutions due to a combination of hydrophobic interactions and hydrogen bonding.

The self-assembly of CLPs into triple helices and the higher order fibrillar assembly in aqueous solutions can be further manipulated with a multitude of covalent or noncovalent mechanisms which stabilize interactions between individual strands and their subsequent interactions between triple helices. For example, Hartgerink and coworkers have utilized lateral (i.e., radial) and axial salt bridges within triple helices to stabilize assembled triple helical, fibril structure, and network assembly.[5, 6] Such studies have shown that charged residues at the ends of

the CLP strands serve as 'sticky ends' of the CLP trimer helix which in turn promote fibrillar assembly into fibrils via electrostatic interactions and assembly of fibrils into networks (hydrogels). Similarly, Pal et al. synthesized two oppositely charged CLP-based pentapeptides that self-assembled to form self-supporting hydrogels at either acidic or basic pH, respectively, depending on overall surface charge.[7] Mixtures of the two oppositely charged peptides resulted in the formation of self-supporting hydrogels at physiological pH and thus, enhanced their potential for biomedical applications. Other studies have shown that other driving forces such as metal-ligand binding and π-π stacking interactions can also be controlled to facilitate fibrillar and network assembly of CLPs in aqueous solutions. For example, Pires et al. designed a collagen-based peptide containing three distinct metal-binding moieties (nitrilotriacetic acid, histidine, and bipyridyl) that self-assembled into a three-dimensional network through the bidirectional coordination of metal ions.[8] The use of noncovalent interactions as physical crosslinks also offers the ability to create responsive, dynamic biomaterials with reversible thermo-responsive crosslinks, such as CLPs as crosslinkers in PEG-based hydrogels.[9]

Over several decades, the self-assembly behavior of charged and metal-binding CLPs in dilute aqueous solutions has been studied extensively, as described above. In contrast, self-assembly of CLPs into higher order structures such as fibrils and networks purely using amino acids' inherent hydrogen bonding ability and solvent-mediated physical interactions (e.g., hydrophobic, hydrogen bonding), both of which are easier for researchers to tune/tailor, has received less attention. Some experimental studies have examined CLP triple helices assembling to form micron-long fibrils via hydrogen bonding due to the presence of uncharged, hydrogen bonding 'sticky ends' in the triple helix.[10, 11] Experimental reports have described the use of cysteine residues to cross-link triple helices into staggered helical registers (i.e., relative alignment

of CLP strands within the triple helix) which expose hydrogen bonding groups at the ends of the triple helix.[10] Such covalently associated, uncharged sticky ends result in assembly of triple helices via hydrogen bonding to produce fibrils. Similar reports by Hartgerink and coworkers[11] on CLP polymers synthesized via native chemical ligation showed that triple helices with sticky ends resulting from dispersity in polymerization also demonstrated fibrillar assembly with fibril lengths on the order of microns. Computational studies by Taylor et al.[12] showed how sticky ended CLP triple helix architecture with varying sticky end length, number of sticky ends, and CLP chain lengths can be used to tailor the propensity of fibrillar and percolated network assembly. While there are experimental and computational reports of hydrogen bonding-mediated assembly of CLPs into higher order structures (fibril and networks) in aqueous environments,[6] fewer studies have examined the additional impact of solvent quality on this hydrogen bonding driven CLP higher order structure and morphology.

CLPs are often synthesized in organic solvents and there have been reports of effects of solvent identity and quality for a limited range of systems, specifically effects of solvent quality at the triple helical and fibrillar length scales of CLP self-assembly.[13-15] Studies of collagen films by Kuznetsova et al. showed that the swelling behavior of such collagen films differed greatly depending on the hydrogen bonding propensity of the solvent.[13] Moreover, their experimental swelling curves showed hard shell repulsion between triple helices for solvents with low hydrogen bonding propensity (e.g., ethanol, 2-propanol, and NN-dimethylformamide) and softer repulsion for solvents with high hydrogen bonding propensity attributing these to the ability of the solvent to form a hydration network mediated via hydrogen bonds surrounding CLP triple helices. Other reports by Fiorani et al. examined the impact of solvent conditions [e.g., trifluoroethanol (TFE) versus dilute acetic acid (AcOH) solvents] on electrospun collagen fibrils and concluded that

nonaqueous solvents result in fibers with a low content of triple helices as opposed to natural (pristine) collagen.[14] Similarly, experimental reports of staggered triple helical registers by Raines and coworkers[10] were achieved in nonaqueous solutions (mixtures of acetic acid and methanol) and showed fibrillar assembly with fibril diameters on the order of a single triple helical diameter, as opposed to other studies of CLP fibrillar assembly with diameters on the order of hundreds of nanometers[16] in aqueous systems. Clearly, previous studies have demonstrated that solvent identity can be used to further control CLP assembly at multiple length scales (triple helix, fibril).

In this study, using coarse-grained (CG) molecular dynamics (MD) simulations we explore the impact of solvent quality on the self-assembly of hydrogen bonding driven assembly of CLP triple helices with varying strand lengths (i.e., CLP heterotrimers). The solvent and its quality on the CLP strands are modeled implicitly using attractive-repulsive (Lennard Jones potential) or repulsive-only (Weeks Chandler Andersen potential) effective interactions between the CG beads representing the CLP strands' amino acids. We also examine the impact of CLP heterotrimer design – specifically, if the CLP strand length variation is manifested at one end ('one sticky-ended'), or on both ends ('two sticky-ended') - on the resulting CLP assembled structure. At higher concentrations of CLP in solution, our CG MD simulations show that as the solvent quality on CLP heterotrimers worsens, they form more heterogenous percolated CLP networks with larger fibril diameters and pore sizes than the CLP heterotrimers in better solvent quality. One sticky-ended CLP heterotrimers form network structures with larger pore sizes with deceasing solvent quality versus two sticky-ended CLP heterotrimers, indicating the utility of using solvent quality and CLP design as handles for controlling CLP's fibrillar network structure. We also note that similar network structures to those seen in this study can be obtained from linear chains with attractive/patchy sites.[17-20] However, the architecture of the sticky ended CLP triple helix has

distinct differences from that of a simple linear polymer chains in terms of flexibility (e.g., intact sections of the CLP triple helix are stiffer than melted sections of the triple helix) and functionality (hydrogen bonding sticky ends). Therefore, it is not straightforward to predict *a priori* how the competition between hydrogen bonding and solvent quality would impact self-assembled structure. Moreover, we explore in this study solvent quality as a driving force for self-assembly which is underutilized and not well understood within the context of CLP-based biomaterials.

This paper is organized as follows. First, we describe our CG model, MD simulation details, and relevant structural analyses in this study. Then, we describe the results of our CG MD simulations focused on how solvent quality, CLP strands' design, and CLP concentrations impact the CLP heterotrimer higher order assembly into network morphologies. Finally, we summarize our results and implications in future/ongoing experiments, in the conclusions section.

2. Approach

2.1 Model

In this study we use the coarse-grained (CG) CLP model initially developed by Condon and Jayaraman.[21] In that CG model, each (POG) repeat unit is represented using five coarse-grained beads: a proline backbone bead (PB), proline hydrogen bonding acceptor bead (PH), hydroxyproline backbone bead (OB), glycine backbone bead (GB), and glycine hydrogen bonding donor bead (GH). For this study we modify that CLP CG model to incorporate effective attractive-repulsive or repulsive-only interactions between backbone beads to mimic the impact of (implicitly represented) solvent quality on the assembly of sticky-ended, CLP heterotrimer triple helices (**Figures 1a** and **1b**).
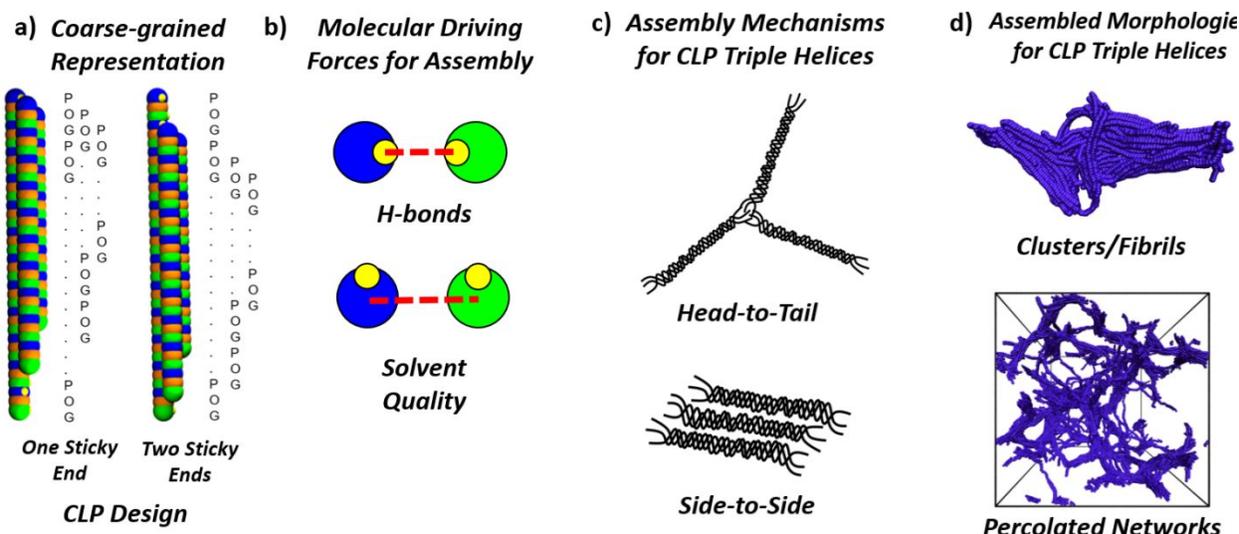
Figure 1. a) Coarse-grained model representation of CLP and the amino acid sequences for a CLP triple helix with one and two sticky ends resulting from exposed hydrogen bonding groups due to different chain lengths of the three CLP strands in the triple helix. b) Dashed lines show the interactions between CLP segments - hydrogen bonding interactions between H-bonding donors and acceptors (top) and effective solvent-induced CLP backbone-backbone interactions (bottom); these interactions are the driving forces for CLP triple helix assembly into networks. c) Schematics showing head-to-tail assembly of CLP triple helices due to hydrogen bonding of sticky ends and side-to-side assembly of CLP triple helices due to backbone-backbone attraction resulting from selected solvent quality. d) Simulation snapshots of CLP triple helices assembled into fibrillar clusters at low concentration (1mM) and percolated networks at high concentration (25mM). These snapshots are taken from the simulation trajectory at the end of the 10 million-timestep production run for CLP triple helices with the worst solvent quality considered in this study.

In our CG CLP model, all backbone (BB) beads have a diameter of 1.0 σ and a mass of 3.0 m while H-bond beads have a diameter of 0.3 σ and a mass of 1.0 m. All distances in our CG model are specified in terms of σ = 0.5 nm and all masses in terms of an arbitrary m as the model was not designed to capture the correct dynamics of these CLP systems. All energy terms are listed in terms of ε = 0.1 kcal/mol.

For each CLP chain, adjacent BB beads are connected using a harmonic bond potential with a bond length of 0.5 σ and force constant of 1000 ε/σ². Each H-bond bead is connected to its "parent" BB bead using a harmonic bond potential with a bond length of 0.37 σ and force constant

of 1000 ε/σ^2. The SHAKE algorithm is used to constrain all BB-HB bond lengths.[22] To mimic the rigidity of the polyproline type II conformation of the CLP strands,[23] there is a harmonic angle potential between three adjacent BB beads with the angle constant set at 20 $\varepsilon/radian^2$ and equilibrium bond angle set at 180° (i.e., π radians). To ensure that H-bond formation occurs perpendicular to the backbone, a H-bond bead-parent BB bead-adjacent BB bead angle potential is included with a force constant of 300 $\varepsilon/radian^2$ and equilibrium angle of 90°. To capture the directionality of H-bond beads with respect to neighboring H-bond beads along the same strand, there are two H-bond bead—BB bead—BB bead—H-bond bead dihedral angle potentials,

$$U^{dih}(\varphi) = k_d(1 + \cos(\varphi - \varphi_0)) \tag{1}$$

in which the dihedral constant k_d is set at 15 ε and the reference dihedral, φ_0 is set at −120° for (GH-GB-PB-PH) dihedrals and 120° for (PH-PB-GB-GH) dihedrals.

Non-bonded H-bond donor (D) - acceptor (A) interactions are represented using the Lennard-Jones (LJ) potential,[24]

$$U^{HB}_{D-A}(r) = \varepsilon^{HB}_{D-A}\left[\left(\frac{\sigma^{HB}}{r}\right)^{12} - \left(\frac{\sigma^{HB}}{r}\right)^6\right] \tag{2}$$

where ε^{HB}_{D-A} is the interaction strength between D and A beads and is set at 50.4 ε and σ^{HB} is the diameter of a H-bond bead. A cutoff of $1.9\sigma^{HB}$ is used with a switching function taken from GROMACS[25] and implemented in the LAMMPS package that smoothly ramps both the force and potential to zero at $2.0\sigma^{HB}$. We note that with this specific model, we are aiming to describe the assembly resulting from hydrogen bonding and place the D and A beads accordingly. However, it would likely be possible to use similar attractive beads in a different orientation to represent other assembly mechanisms (metal-ligand binding or π-π stacking interactions, for example).

We also include Lennard Jones interactions between backbone beads (PB, OB, and GB) to mimic interactions between CLP triple helices due to worsening solvent conditions.

$$U_{BB-BB}(r) = \varepsilon_{BB}\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{3}$$

where increasing interaction strength ε_{BB} is used to mimic decreasing solvent quality and a cutoff of 2.5σ is used for all BB-BB interactions.

All other pairwise interactions besides the H-bond donor-acceptor interactions and CLP backbone bead-backbone bead interactions, are modeled using the purely repulsive, Weeks-Chandler-Andersen (WCA) potential.[26]

$$U_{ij}(r) = \begin{cases} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right] + \varepsilon_{ij} \; ; r < 2^{\frac{1}{6}}\sigma_{ij} \\ 0 \qquad\qquad\qquad\qquad\qquad\quad ;otherwise \end{cases} \tag{4}$$

where σ_{ij} is the arithmetic mean of the diameters of beads i and j, and ε_{ij}, the interaction strength, is set at $1.0\ \varepsilon$. For donor-donor and acceptor-acceptor interactions, σ_{ii} is set at $0.7\ \sigma$ instead of $0.3\ \sigma$ to ensure that a donor HB bead only forms an H-bond with a single HB acceptor bead, and vice versa. **Section S.1** in the Supplementary Information includes tables with a summary of the bonded and nonbonded potential parameters for the CG CLP model (**Tables S1 – S5**).

Since the model we use in this study is a phenomenological CG model, the form of pair-wise interaction potentials is not chosen based on chemistry but instead a choice of a mathematical expression of the U(r) that reproduces the desired phenomenon. Such a model for homotrimeric CLP triple helices has successfully captured experimental trends in melting transition of CLP triple helices as a function of CLP design[21] (length and amino acid sequences) and has also successfully been used to explore the self-assembly of sticky-ended heterotrimeric triple helices at multiple lengths scales (helix, fiber, network) in aqueous solutions[12]. In this study, we also use an implicit solvent model due to the significant increase in computational resources that would be required to simulate explicit solvent for the large system sizes ($110\sigma = 55$nm) explored in this study. Explicit solvent simulations would require approximately 6 million additional coarse-grained beads per

system and would make exploring a large design space (i.e., solvent quality and CLP design) computationally intractable.

2.2 Parameters Varied

In this study, we focus on CLP triple helices which have three CLP strands of different lengths (the leading strand, middle or intermediate strand, and lagging strand, from longest to shortest), resulting in 'sticky ends' that can interact with other 'sticky ends' via hydrogen bonding. We examine CLP triple helices with a total number of 36 (POG) repeat units per triple helix and a sticky end length of 6 (POG) units. With this number of (POG) units and sticky end length, the length of the leading strand is 18 (POG) units, the length of the middle strand is 12 (POG) units, and the length of the lagging strand is 6 (POG) units, giving the sequence $(POG)_{18}$-$(POG)_{12}$-$(POG)_6$. As done in Taylor et al.'s work, we investigate both one sticky-ended and two sticky-ended heterotrimeric triple helices.[12] For a one sticky-ended design, the entire offset of 6 repeat units is placed on a single end of the helix while for a two sticky-ended design half of the offset (3 repeat units) is placed on each side of the triple helix. Thus, both a one sticky-ended helix and two sticky-ended helix each have the same total sticky end length and same number of available hydrogen bonding groups. The choice of the triple helices with 36 total (POG) units and a sticky end length of 6 (POG) units over other CLP sizes and sticky end lengths is dictated by our previous work, which showed longer sticky ends had the highest proficiency for assembling into percolated networks, independent of the number of (POG) units in the triple helix.[12] The design parameters explored in this study such as the number of sticky ends and CLP concentrations are also inspired by previous experimental studies by Hartgerink and co-workers[5, 6] and Kloxin and co-workers[27, 28]. Going beyond our previous work, in this study, we vary the effective Lennard-Jones interaction

strength between backbone (BB) beads, ε_{BB}, between 0 (corresponding to a WCA potential between backbone beads) and 0.3ε to mimic relatively good (low ε_{BB}) and worse solvent conditions (high ε_{BB}). We focus most of the results and discussion on the network-forming solution conditions (i.e., high concentration of CLP in solution) to understand how the varying BB interactions impact the features of the network structure. We only briefly discuss the low concentration results at the end of this paper.

2.3 Simulation Details

We perform Langevin dynamics simulations in the NVT ensemble using the LAMMPS simulation package.[29] Based on the work of Taylor et al.,[12] all simulations are performed at T* = 3.0; this temperature is below the melting temperatures of the CLP triple helices calculated using simulations with this CLP CG model and ensures that 100% of triple helices are intact. For the WCA backbone-backbone attraction and weak backbone-backbone LJ attraction, $\varepsilon_{BB} = 0.1\varepsilon$, an initial configuration is obtained by randomly placing intact CLP triple helices in a cubic simulation box of size 110 σ with periodic boundary conditions. For systems with stronger backbone-backbone LJ attractions, $\varepsilon_{BB} = 0.2\varepsilon, 0.3\varepsilon$, initial configurations are obtained from final equilibrated configurations of CLP systems with weaker LJ attraction. In other words, final equilibrated snapshots from $\varepsilon_{BB} = 0.1\varepsilon$ are used as initial configurations for $\varepsilon_{BB} = 0.2\varepsilon$, and final equilibrated snapshots from $\varepsilon_{BB} = 0.2\varepsilon$ are used as initial configurations for $\varepsilon_{BB} = 0.3\varepsilon$. This protocol mimics a simulated annealing approach that is useful to avoid kinetic trapping that one may observe if the simulations are run directly at the higher attraction strength from a random initial configuration. For all systems, we ensure that results from different independent simulation trials of the same system are reproducible and report the error between independent simulation

trials. Small variability in structural data between multiple trials for the same system suggests we have likely avoided kinetic trapping.

In all simulations the friction coefficient is set to 10 τ, where 0.001 τ is approximately 6 fs, guided by previous work with similar systems.[30] A two-level RESPA[31] integrator is also implemented so that non-bonded and bonded interactions are integrated with a time step of 0.001 τ and 0.0005 τ, respectively. All systems are first equilibrated for 10^8 timesteps followed by a 10^7 timesteps-long production run during which configurations are collected every 100,000 timesteps.

In these simulations the CLP triple helices self-assemble head-to-tail via the H-bonding interactions between CLP beads in the 'sticky ends' and assemble side-to-side via backbone bead-backbone bead attraction (**Figure 1c**). At low concentrations, the self-assembled structures are finite clusters **(Figure 1d)** of CLP helices with head-to-tail and side-to-side assembly the extents of which depend on the dominant interaction. As the concentration increases, the clusters grow in size and eventually form percolated networks (**Figure 1d**).

2.4 Analyses

2.4.1. Network Structure Characterization
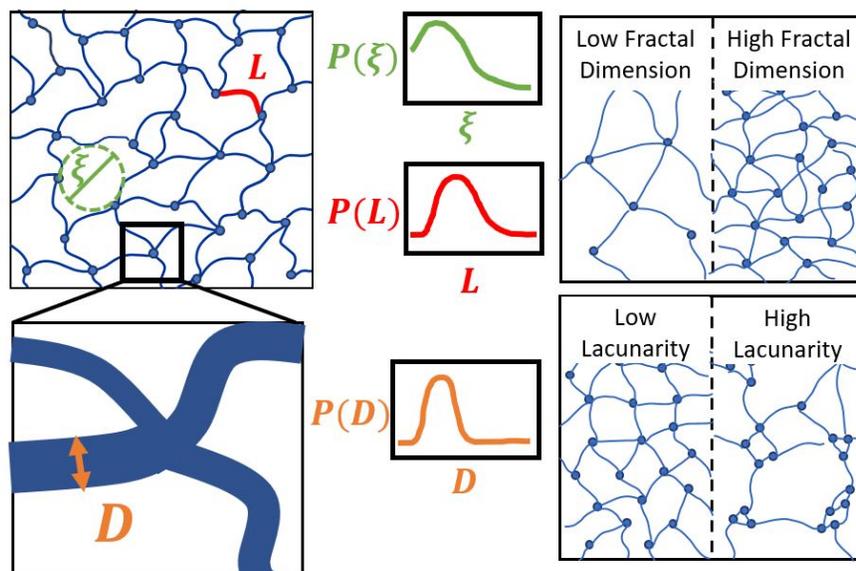
Figure 2. Schematics illustrating the various morphological features that we calculate to characterize the assembled CLP network structure observed at CLP concentrations 15, 20, and 25mM. These networks have junctions where multiple CLP triple helices physically interact via the hydrogen bonding and/or solvent-induced attractive interactions, with network strands connecting these junctions being either a single CLP triple helix or multiple aggregated CLP helices. The network structure is quantitatively described by the distributions of pore sizes (ξ), network strand lengths (L), and network strand diameters (D), as well as the fractal dimension and lacunarity.

At CLP concentrations of greater than 5 mM, Taylor et al.,[12] found the formation of a percolating network. In this study, at concentration above this percolation threshold of 5mM, we analyze various aspects of the percolated network structure (**Figure 2**). For each analysis technique described in **Section 2.4.1**, we perform the analysis every one million timesteps over the course of the ten-million-timestep production run, corresponding to 10 total snapshots for analysis. This is repeated for three independent trials. Generally, we find that the variation between trials is greater than the variation between the timesteps of an individual trial, so we compute the mean over the ten analyzed timesteps for each trial and report the mean and standard deviation of the trial means.

Network pore size distribution: We first compute the network pore size probability distributions, P(ξ) vs. ξ. We accomplish this by randomly sampling 1000 coordinates in the simulation box for each frame and defining the pore size for each sampled coordinate as the diameter of the largest sphere that avoids overlap with the surrounding coarse-grained CLP beads and encompasses the randomly selected coordinate; this calculation is inspired by the methods of Wang et al. on DNA crosslinked hydrogels[32] and methods of Bhattacharya and Gubbins on porous silica.[33] This method has also been used to quantify pore size distributions in a wide range of other material types, including polymeric solutions and particulate systems.[34, 35]

We also analyze the fractal dimension, lacunarity, network strand length distribution, and network strand diameter distribution (as shown in schematics in **Figure 2**). We calculate these structural features because these have been shown to impact the properties of fibrillar networks, such as the transport of solvents, nanoparticles, or other chemical species through the network and the mechanical properties of the network, both of which would be important for CLP networks used in biomedical applications.[36-38]

Fractal dimension: Fractal dimension (see **Figure 2**) quantifies the exponent with which the mass or volume of a system scales with length scale, as described by

$$m \sim R^{D_f} \tag{5}$$

where m represents the mass of a system, R represents the length scale, and D_f is the fractal dimension. To calculate fractal dimension, we use the box counting tool implemented in the PoresPy package.[39] We include the specific details regarding the box-counting method of estimating fractal dimension and examples of the calculation in the Supporting Information **Section S.2**. The values of fractal dimension for 3-dimensional systems, in general, range from 1 to 3, with a fractal dimension of 1 corresponding to a thin rod shape, and a fractal dimension of 3

corresponding to a solid cube or sphere. For our specific case of CLP networks, the fractal dimension can be used as an indicator of both the size of the network strands and the branching of the network. Given equal levels of branching, a network with higher diameter network strands will have a higher fractal dimension, and given two networks with similar network strand diameters, the one with greater branching will have a higher fractal dimension.

Lacunarity: Lacunarity (**Figure 2**) describes the heterogeneity of an assembled structure at different length scales. Mathematically, the lacunarity is computed on a binary grid representation of the structure and involves breaking the structure into cubes containing different numbers of voxels. For each cube size, the distribution of the number of voxels valued 1 (representative of the CLP network in our case) is computed, and the lacunarity at that cube size is reported as the squared ratio of the standard deviation of the distribution to the mean of the distribution. The lacunarity values, in general, range from 1 (at the length scale of the entire voxel representation) to $n_{voxels}/n_{voxels=1}$ (at the length scale of a single voxel), where n_{voxels} is the total number of voxels and $n_{voxels=1}$ is the number of voxels valued 1 in the grid representation of the structure. Higher values of lacunarity indicate a greater level of heterogeneity. For the lacunarity calculation, we use our in house code based on the method described by Sebők, et al.[40] Additional details regarding this method and our specific implementation are included in the Supporting Information **Section S.2**.
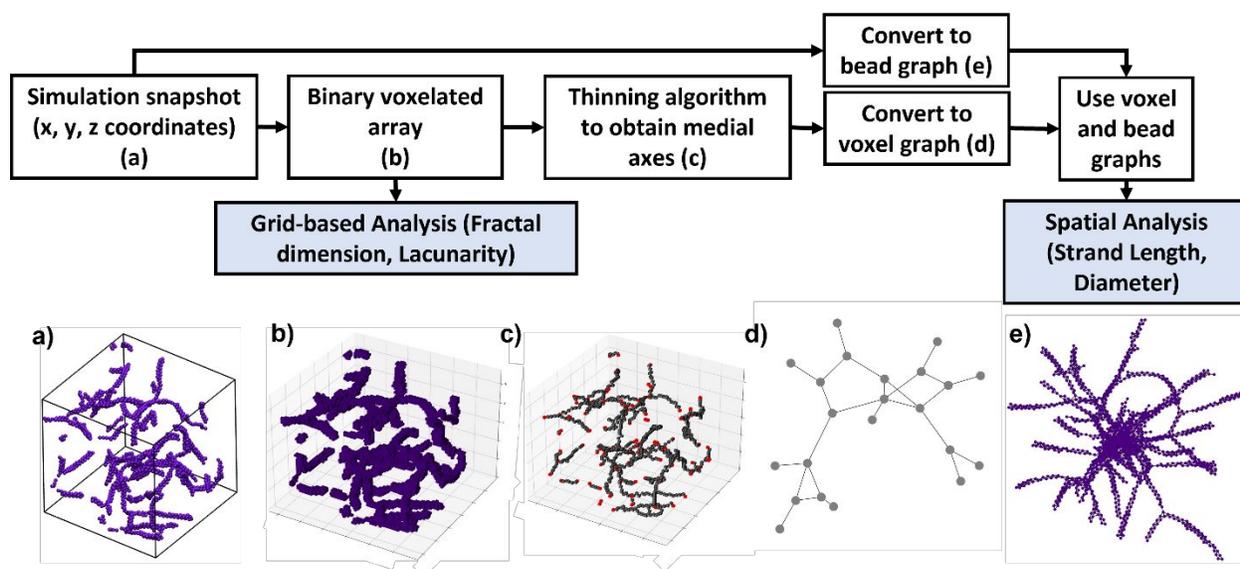
Figure 3. Flowchart for our graph-based analysis of assembled CLP network structures. From left to right, we start with the x, y, and z coordinates of the CLP coarse-grained beads (part a). Then we convert the x, y, and z coordinates into a voxel representation (part b). This voxel representation is used to compute the fractal dimension, D_f, and lacunarity. We then extract the skeletonized representation (part c) where junctions are shown in red. The skeletonized representation is then converted to a graph representation (part d). Lastly, we construct a graph based on the positions of the simulation beads (part e), where each purple node represents a simulation bead, and edges connect nearby beads. The voxel representation (part b) and the bead graphs (part e) are used in conjunction to compute network strand lengths, l and diameter, D.

Network strand length and diameter distributions: For the network strand length and diameter calculations, we extract a graph representation of the network assembly, and use the graph and simulation bead coordinates to compute the network strand lengths and fibril diameters. We summarize the graph-based analysis of CLP network assemblies in the flow diagram shown in **Figure 3** and in the text below; each individual step is described in greater detail in Supporting Information **Section S.2c**.

First, we construct a binary voxel representation of the simulation, where a voxel is assigned the value 1 if any part of a simulation bead is within the voxel and 0 if no beads intersect the voxel. For all simulations, we use a voxel size of 1 σ. The fractal dimension and lacunarity are computed using grid-based methods on this voxel representation. We then use a skeletonization

procedure to thin the voxel representation,[41] extracting the medial axes of the network structure. We follow the skeletonization with a creation of a graph based on the skeletonized representation with junctions in the network represented as nodes, and strands in the network represented as edges. Each voxel in the skeleton is added to the graph as a node, with edges connecting to any other voxels in the representation that are in the neighboring 26 sites. All nodes with degree 2 are removed, iteratively joining the adjacent nodes until the graph contains no nodes with 2 connected edges. This process reduces the graph, such that each node is a junction in the network, and each edge is a network strand, providing a wholistic graphical representation of the simulation snapshot.

The skeletonization procedure can result in 'blobs' of many voxels surrounding true junctions in the network (examples of these artifacts are shown in the Supporting Information **Section S.2**). We remove artifacts from the voxel graph by clustering together all adjacent voxels, replacing them with a node in the graph with the average position of the cluster, and connecting that node to all other nodes in the graph that shared an edge with any node in the original cluster. Second, the voxel representation could lead to two nearby simulation beads being classified as connected, when they are nearby and not in contact (we show some instances of this occurring in the Supporting Information **Section S.2**). To remedy this issue, we construct another graph based on the simulation beads themselves, where each simulation bead is represented as a node, and beads within a cutoff distance of 2σ are connected by edges. We use only OB beads for the construction of both graphs to reduce the computational intensity of the analysis. For each edge in the voxel graph, we compute the shortest path between the two corresponding nodes in the bead graph using Dijkstra's algorithm implemented in the NetworkX Python package.[42, 43] If no path exists between the two nodes, we discard the edge from the graph and do not use it for further analysis.

Using the voxel graph, which now has an accurate graphical representation of the assembly, we link the two graphs by mapping each node in the voxel graph to the nearest bead coordinate in the bead graph. The network strand length is computed for each edge again using the shortest path length between the two node beads along the bead graph. To compute the diameter of a network strand, we similarly map each voxel along an edge of the voxel graph to its nearest beads. We compute the direction vector of the strand using the beads at either end of the strand and compute the diameter as twice the furthest distance of any bead in the network strand to this vector. To avoid inaccuracies resulting from a tortuous network strand, we break long network strands into shorter segments and repeat this calculation on each segment, taking the final diameter of a network strand as the average of the diameters computed for the shorter segments of the network strand.

2.4.2.   Assembled CLP Clusters' Structural Characterization

At CLP concentrations below the percolation threshold, we observe clusters of CLP heterotrimers. To quantify the finite sizes of these clusters and the number of heterotrimers in each cluster, we compute the radius of gyration of each cluster of CLP heterotrimeric triple helices, $R_{g,cluster}$, as well as the number of helices per cluster, $N_{helix,cluster}$. Two CLP triple helices are defined as being part of the same cluster if a) there exists at least a single hydrogen bond between the two helices with a hydrogen bond defined as a pair of PH and GH beads that are separated by a distance of 0.45 σ or less, or b) there exists an inter-helix BB-BB pair separated by a distance of 2.5 σ or less. By looping through all pairs of triple helices, we identify clusters as groups of triple helices that satisfy the above criterion. Then, the radius of gyration of each cluster of CLP heterotrimeric triple helices, $R_{g,cluster}$, is calculated as

$$R_{g,cluster} = \sqrt{\frac{1}{N}\Sigma_i|r_i - r_{COM}|^2} \tag{6}$$

where N is the number of CLP beads in the cluster, r_i, is the position of bead i, and r_{com} is the cluster's center of mass. We compute weighted probability distributions of $R_{g,cluster}$ and $N_{helix,cluster}$ such that the probability of each bin is weighted by its respective value ($R_{g,cluster}$ or $N_{helix,cluster}$). For all calculations, we focus on clusters with at least two helices per cluster and calculate weighted distributions of $R_{g,cluster}$ and $N_{helix,cluster}$. For each analysis technique described in **Section 2.4.2**, three independent simulation trials each consisting of 100 independent configurations are pooled to generate probability distributions of radii of gyration of CLP clusters ($R_{g,cluster}$) and number of helices per clusters ($N_{helix,cluster}$). The 100 independent configurations for each trial correspond to a 10^7 time step production run with configurations stored every 100,000 time steps. In this work, we emphasize the qualitative trends in how aggregate sizes decrease or increase as a function of CLP design and solvent quality rather than numerical values, as the numerical value of cluster sizes are a function of the system size (i.e., number of CLP helices simulated)

3.   Results and Discussion

In this study, we explore the effects of solvent quality on the self-assembly of sticky-ended, heterotrimeric CLP triple helices. Previous work by Taylor et al.[12] explored the impact of peptide designs such as sticky end lengths, number of sticky ends, CLP concentration, and chain length (i.e., (POG) family) on the thermal stability of the triple helix and the assembly of helices into fibrillar structures and supramolecular networks. In their work, Taylor et al.[12] assumed that hydrogen bonding was the dominant driving force via which triple helices assembled to form fibril and network structures, as compared to other effects such as solvent-induced attractions of triple helices. The dominance of hydrogen bonding is in alignment with experimental reports of solvent conditions, such as solutions of acetic acid, methanol, and mixtures of acetic acid and methanol,

which promote head-to-tail assembly of triple helices via hydrogen bonding at sticky ends rather than side-to-side assembly via solvent induced attractions (e,g., hydrophobic interactions). We explore in this study the competition between decreasing solvent quality (leading to increasing effective CLP helix-helix attraction) and hydrogen bonding strength and the result of such competition on CLP structure and morphology.

We first examine CLP concentrations above the percolation threshold of sticky-ended CLP systems, determined by Taylor et al.,[12] and understand the impact of solvent quality on supramolecular assembly in CLP networks, and then explore CLP concentrations below the percolation threshold to determine the impact of solvent quality on fibrillar assembly in CLP solutions.
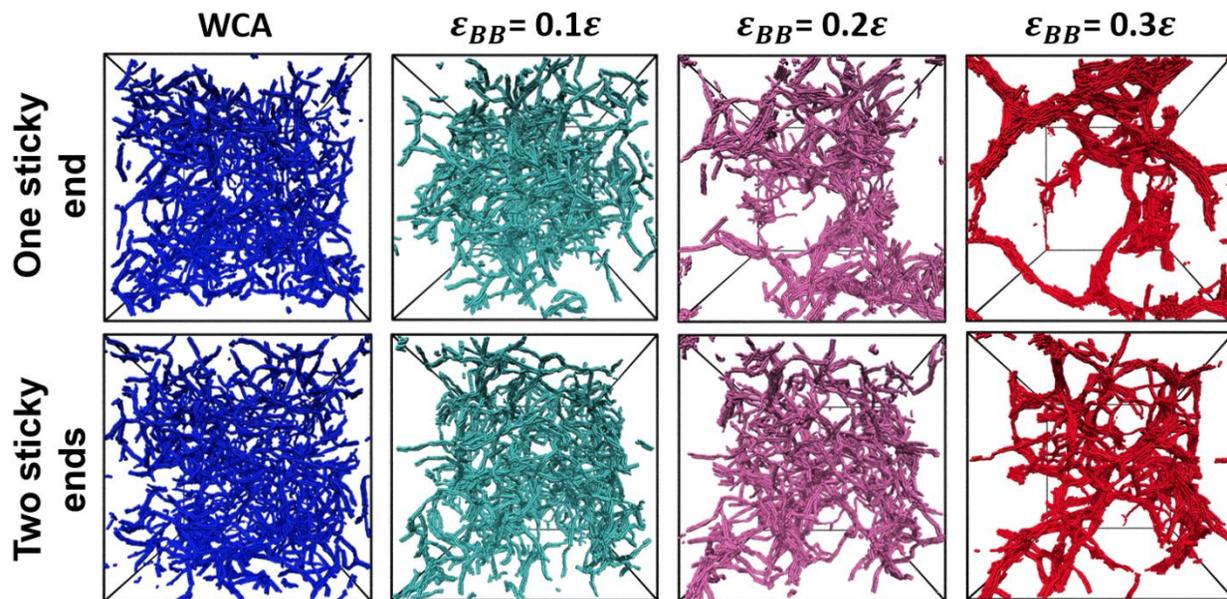
Figure 4. Representative snapshots from simulations of one and two sticky-ended CLP heterotrimers with sticky end length of 6 POG units and CLP triple helix made of 36 POG units, at different solvent qualities modeled as BB-BB interactions: WCA (repulsive only) and Lennard Jones strength ε_{BB} = 0.1, 0.2, and 0.3 ε at a CLP concentration of 20mM. Snapshots are taken from the simulation trajectory at the end of the 10 million time steps production run.
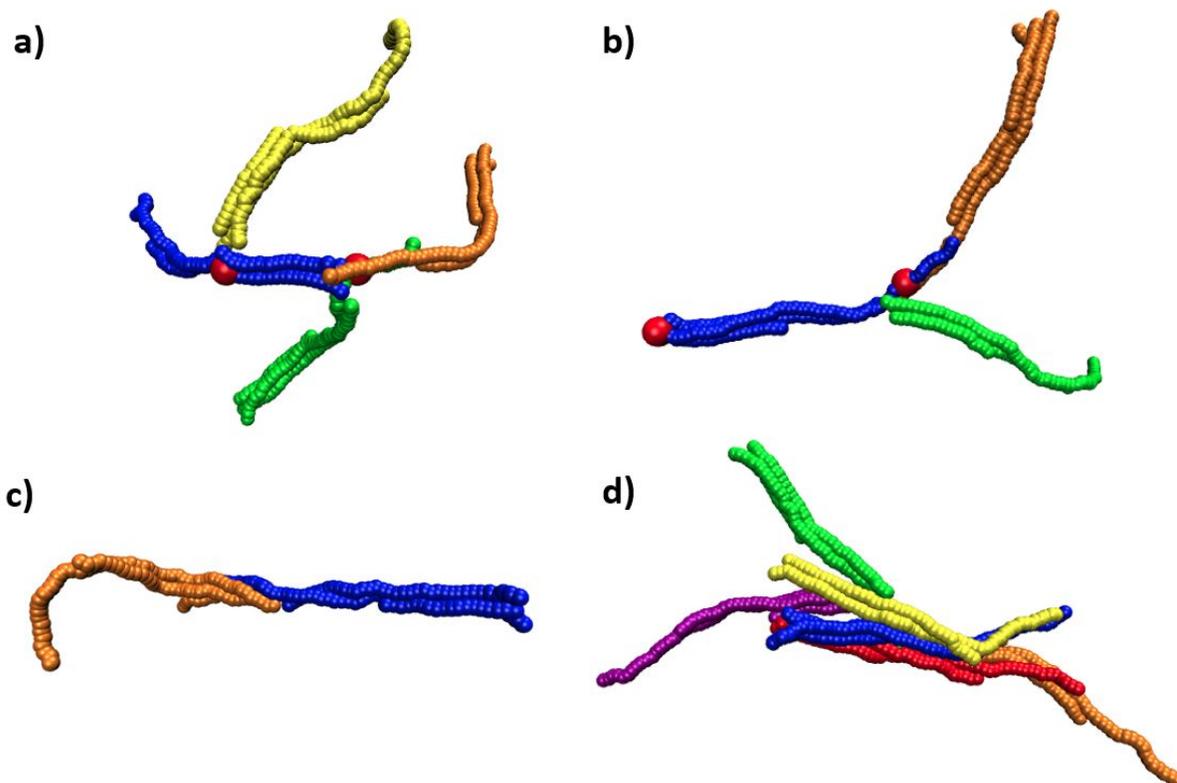
Figure 5. Snapshots of certain sections of the percolating network to show how CLP heterotrimeric triple helices assemble to form network strands and network junctions. In all images, individual CLP triple-helices are shown in their own color, and in (a) and (b), the large red beads denote the location that the graph-based algorithm identifies as the junctions for the given network strand. Short network strands with lengths around 6-10σ result from CLP heterotrimers forming junctions with lengths approximately equal to the length of the lagging CLP strand (shortest CLP strand) of the CLP heterotrimer (a). Longer network strands with lengths around 16-20σ result from CLP heterotrimers forming junctions with network strand lengths approximately equal to intermediate CLP strand length in the CLP triple helix (b). CLP helices can assemble due to hydrogen bonding at sites on the end of individual CLP strands or side-to side as a result of backbone-backbone attraction. An example of the former, with network strand diameter ~3σ (c), and an example of the latter, with network strand diameter ~7σ (d). All images are from snapshots of 20mM networks of one sticky-ended CLP helices at the end of the 10 million timestep production run. (a) and (b) are with WCA interactions (good solvent), and (c) and (d) are with $\varepsilon_{BB} = 0.2\ \varepsilon$.

3.1 Effect of Solvent Quality on Structural Features of CLP helices Assembled into Percolated Networks at High Concentrations

Before we describe our results it is important to state that a network strand is different from the CLP strand in a heterotrimer. Based on our observation, a network strand can be composed of a part of a single CLP triple helix, an entire CLP triple helix, parts of multiple CLP triple helices, or multiple complete CLP triple helices. We show these ideas via some representative snapshots of network strands of different length and diameter and the associated CLP triple helices in **Figure 5.** For this discussion, it is also useful to remind the reader that the CLP heterotrimeric triple helices considered in this study have three CLP strands of different lengths, a leading (long) strand with 18 (POG) units, a middle (intermediate) strand with 12 (POG) units, and a lagging (short) strand with 6 (POG) units (as described in **Section 2.2**). Considering the equilibrium bond lengths for the bonded backbone P, O, and G beads, these CLP strands have lengths of approximately 9, 18, and 27 σ, respectively.

In our simulations we observe that with decreasing solvent quality (WCA to increasing ε_{BB}), simulation snapshots show fibrillar network structures with increasing network strand diameters and pore sizes for one sticky-ended triple helices (top row, **Figure 4)**. For two sticky-ended CLP heterotrimers, we also observe an increase in network strand diameters and pore sizes with decreasing solvent quality (bottom row, **Figure 4**); however, as solvent quality worsens the two sticky-ended triple helices exhibit smaller network strand diameters and pore sizes than one sticky-ended triple helices do. There are no differences between one and two sticky-ended triple helices in terms of network strand diameter and pore structure at good solvent conditions (WCA).

Next, we compute network strand lengths for CLP systems with varying number of sticky ends, CLP concentration, and solvent quality. We first compare the distributions of network strand lengths for varying CLP concentration, solvent quality, and number of sticky ends in **Figure 6**. Under good solvent conditions (WCA backbone interactions), we observe two primary peaks in

the distribution of network strand lengths, independent of CLP concentration and number of sticky ends. Networks composed of one sticky-ended and two sticky-ended CLP triple helices have a primary peak at approximately 9 σ and a secondary peak at about 18 σ. We show examples of individual CLP triple helices assembling into network strands of these lengths in **Figure 5a** and **5b** to aid in visualization.

As solvent quality worsens to intermediate values of ε_{BB} ($\varepsilon_{BB} = 0.1, 0.2\ \varepsilon$), for both one and two sticky-ended CLP triple helices and at all concentrations, we observe shorter network strands becoming more prominent, while longer network strands (associated with the secondary peak) become less prominent. This transition could be the result of these longer network strands beginning to assemble in a side-to-side fashion, with shorter strand lengths becoming more frequent as branches in the network occur more frequently from staggered side-to-side assembly of CLP triple helices. As solvent quality worsens further to $\varepsilon_{BB} = 0.3\ \varepsilon$, the propensity for short network strand lengths decreases and we see an increase in the population of larger network strand lengths, seen as a decrease in the primary peak height at shorter strand lengths and longer tails in the strand length distribution. Despite this shift to larger strand lengths, the mode of the strand length distribution remains below 10 σ, indicating that the networks still assemble with many smaller strands.

Figure 6. Distribution of network strand lengths, L, observed for the networks formed from CLP triple helices at 15, 20, and 25mM. The top row plots correspond to results from one sticky-ended CLP helices while the bottom row plots correspond to two sticky-ended CLP helices; each column corresponds to the value of CLP concentration as indicated. In each plot, the results for each solvent quality are depicted with a dark red curve (ε_{BB} = 0.3 ε; worst solvent), light red curve (ε_{BB} = 0.2 ε), cyan curve (ε_{BB} = 0.1 ε), dark blue (WCA interaction; good solvent). For each simulation, we compute the mean length distribution over ten snapshots collected every one million timesteps for the 10 million timesteps of the production simulation. Error bars represent the standard deviation of these means across three independent trials as described in the methods section.

To understand the impact of CLP design, concentration, and solvent quality on the side-to-side assembly of triple helices and resulting network strand dimensions, next we examined strand diameters (**Figure 7**). Independent of CLP concentration, as solvent quality worsens, one sticky-ended triple helices form fibrils with larger network strand diameters (**Figures 7a – 7c**). A similar trend is observed for two sticky-ended triple helices (**Figures 7d – 7f**). As solvent quality worsens, one sticky-ended triple helices display larger network strand diameters than two sticky-ended triple helices, in agreement with their respective simulation snapshots (**Figure 4**). In contrast to networks formed from one sticky-ended triple helices, networks with two sticky-ended CLPs maintain a higher number of network strands with diameters of approximately 3-4 σ as the solvent quality

worsens ($\varepsilon_{BB} = 0.2$ and $0.3\ \varepsilon$). This indicates that the strands of two sticky-ended networks have a higher tendency to be composed of a single CLP triple helix than networks of one sticky-ended triple helices. This is likely because having sticky ends at both ends of the CLP triple helix promotes head-to-tail assembly over side-to-side assembly as the triple helices assemble into networks.
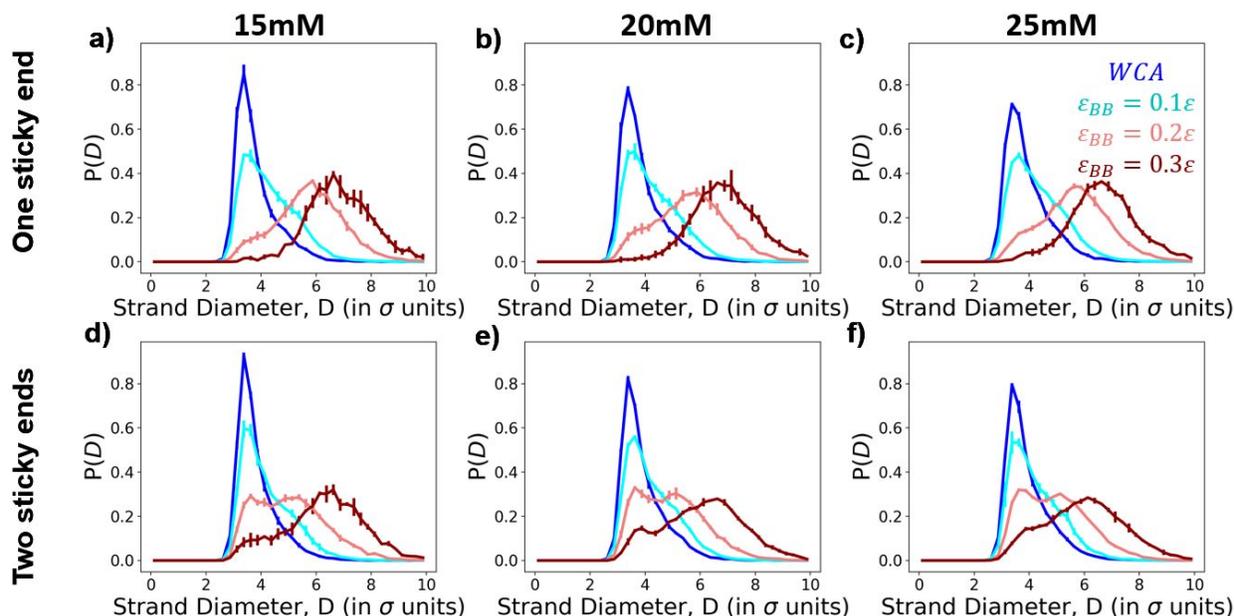
Figure 7. Distribution of network strand diameters, **D**, observed for the networks formed from CLP triple helices at 15, 20, and 25mM. The top row plots correspond to results from one sticky-ended CLP helices while the bottom row plots correspond to two sticky-ended CLP helices; each column corresponds to the value of CLP concentration as indicated. In each plot, the results for each solvent quality are depicted with dark red curve ($\varepsilon_{BB} = 0.3\ \varepsilon$; worst solvent), light red curve ($\varepsilon_{BB} = 0.2\ \varepsilon$), cyan curve ($\varepsilon_{BB} = 0.1\ \varepsilon$), dark blue (WCA interaction; good solvent). For each simulation, we compute the mean diameter distribution over ten snapshots collected every one million timesteps for the 10 million timesteps of the production simulation. Error bars represent the standard deviation of these means across three independent trials as described in the methods section.

We have also computed the pore size distribution of the network structures resulting from CLP heterotrimeric triple helix assembly. The pore size distribution is known to affect transport of particles and other chemical species through the network as well as the mechanical strength of

fibrillar networks.[38] When helices assemble in a good solvent, the pore sizes follow a relatively narrow distribution centered at approximately 10 σ (**Figure 8**). At all concentrations and regardless of one or two sticky-ended CLP, as the solvent quality worsens, the network pore sizes increase. At $\varepsilon_{BB}= 0.1\,\varepsilon$, the distributions exhibit larger tails than at WCA (good solvent), while maintaining a similar mode (peak location), indicating that most of the pores remain the same size, but we begin to see instances of larger pores occurring as solvent quality worsens.

Under good solvent conditions, the pore size distributions are similar between networks of one and two sticky-ended CLP helices. However, when in a worse solvent, the two CLP helix designs show significant differences in pore sizes. At higher ε_{BB} values, we see substantial growth of the pore sizes and greater differences between the networks assembled from one and two sticky-ended CLP helices. While the networks of two sticky-ended CLP helices maintain a mode pore size less than 50σ, the one sticky-ended counterparts show broader, flatter distributions with pore sizes approaching the simulation box size. We include plots directly comparing the pore sizes one and two sticky-ended CLP helices in **Figure S11.**
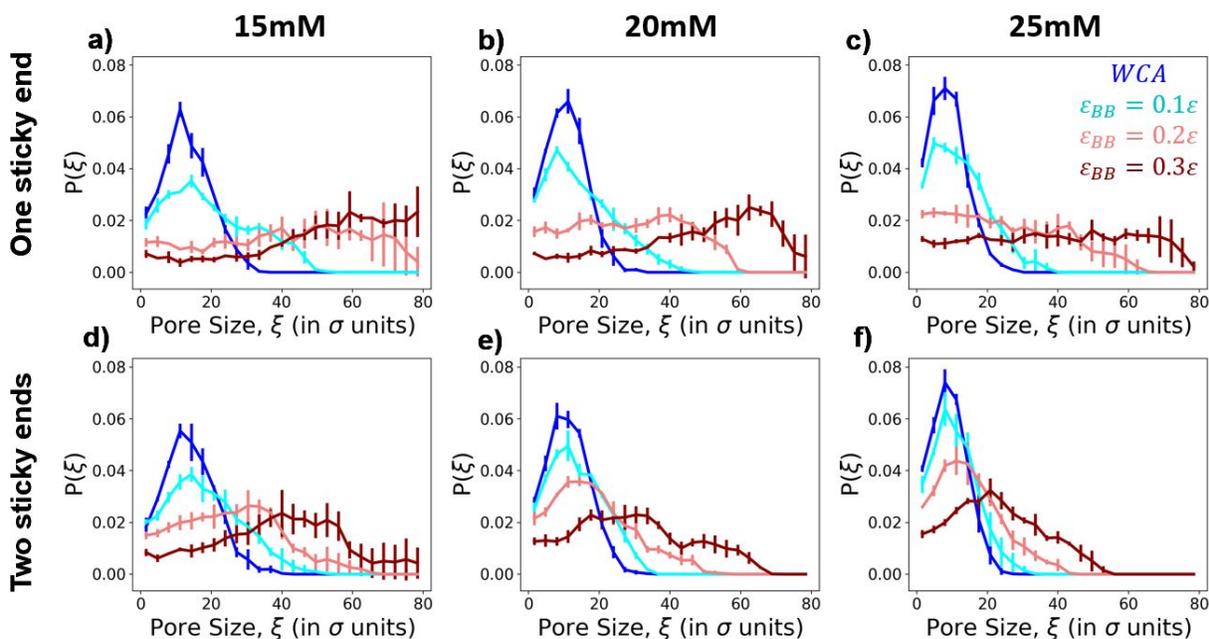
Figure 8. Distribution of pore sizes, ξ, observed for the networks formed from CLP triple helices at 15, 20, and 25*mM*. The top row plots correspond to results from one sticky-ended CLP helices

while the bottom row plots correspond to two sticky-ended CLP helices; each column corresponds to the value of CLP concentration as indicated. In each plot, the results for each solvent quality are depicted with a dark red curve (ε_{BB} = 0.3 ε; worst solvent), light red curve (ε_{BB} = 0.2 ε), cyan curve (ε_{BB} = 0.1 ε), and dark blue (WCA interaction; good solvent). For each simulation, we compute the mean pore size distribution over ten snapshots collected every one million timesteps for the 10 million timesteps of the production simulation. Error bars represent the standard deviation of these means across three independent trials as described in the methods section.

In the analysis of networks, the fractal dimension can be used as an indicator of both the size of the network strands and the branching of the network. Given equal levels of branching, a network with larger diameter strands will have a higher fractal dimension, and for two networks with similar fibril diameters, the one with greater branching will have a higher fractal dimension. We report the computed fractal dimensions in **Figure 9.** For all systems studied in this paper, we see a decrease in fractal dimension as solvent quality decreases. Since a worsening solvent quality is associated with an increase in strand diameters (**Figure 5)**, we can attribute this decrease in fractal dimension to a lower degree of network branching. Visually we can see the decreasing fractal dimension (less branching) with worsening solvent, in the snapshots in **Figure 4**; there are more side-to-side aggregation at worse solvent conditions, resulting in network strands that have larger diameters and branch less frequently. We also see that this fractal dimension decrease is more substantial for networks formed from one sticky-ended CLP triple helices than for those formed from two sticky-ended CLP triple helices, as indicated by the slopes seen in adjacent panels of **Figure 9**.

The contrast in the behavior of network pore sizes and fractal dimension between one and two sticky-ended networks as solvent quality worsens can be explained by the propensity of each type of CLP triple helix to assemble head-to-tail or side-to-side. Two sticky-ended triple helices have a higher tendency to assemble in a head-to-tail fashion as a result of the hydrogen bonding

sites on both ends of the triple helix, resulting in a network that is more branched with smaller pore sizes. In contrast, one sticky-ended triple helices have a higher tendency to assemble in a side-to-side fashion since they only have hydrogen bonding sites on one end of the helix which ultimately results in a network that exhibits less branching and larger pores.

For assembled networks of both one and two sticky-ended CLP helices and all investigated solvent qualities, the fractal dimension is positively correlated with the CLP concentration. This result makes sense, as higher CLP concentration leads to CLP helices being in closer proximity, which will increase the tendency for a highly branching network.
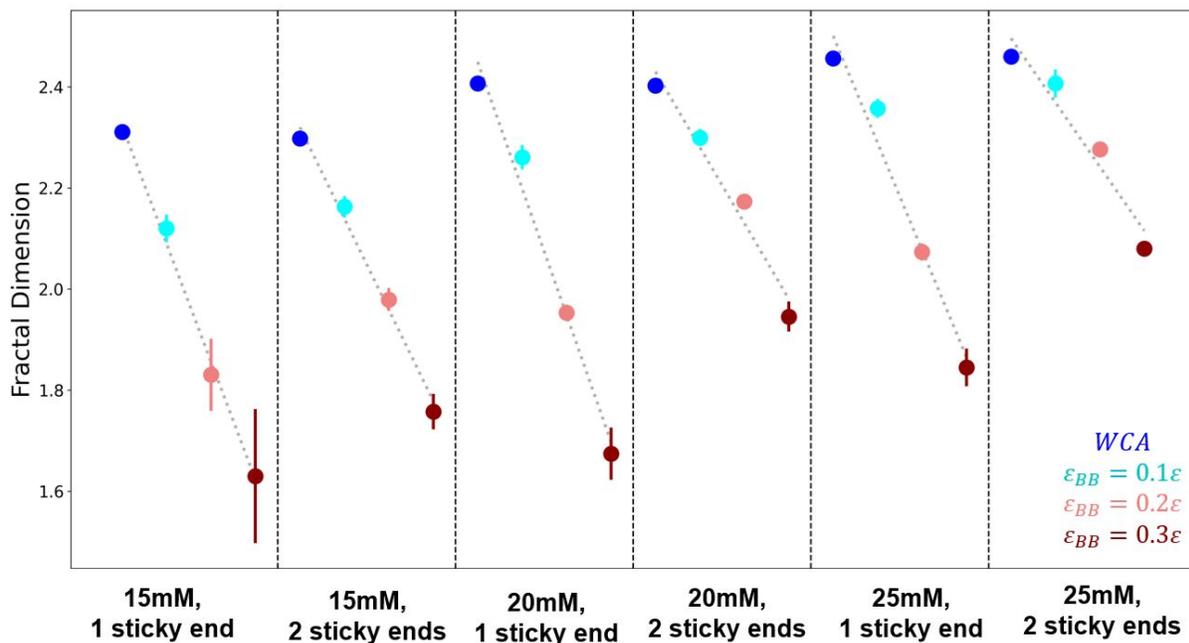
Figure 9. The fractal dimension for each CLP concentration and for one and two CLP sticky ends. For each simulation, we compute the mean fractal dimension over ten snapshots collected every one million timesteps for the 10 million timesteps of the production simulation. Error bars represent the standard deviation of these means across three independent trials as described in the methods section. For each data point, the solvent quality improves as the color changes from dark red to light red to cyan to dark blue (corresponding to ε_{BB} values of 0.3 ε, 0.2 ε, 0.1 ε, and WCA, respectively). Light gray dotted lines show linear fits to the data to enable easier comparison of the rate at which fractal dimension decreases as solvent quality worsens.

Lacunarity describes the heterogeneity of an assembled structure at different length scales. Lacunarity values range from 1 (which occurs at the length scale of the entire voxel representation) to $n_{voxels}/n_{voxels=1}$ (which occurs at the length scale of a single voxel), with higher values of lacunarity indicating a greater level of heterogeneity. In **Figure 10** we see that for all CLP concentrations and sequences studied, the lacunarity increases across all length scales as the solvent quality worsens. The one sticky-ended CLP triple helices form assembled networks that are more heterogenous at larger length scales than their two sticky-ended counterparts, as indicated by their higher lacunarity values at length scales approaching the simulation box size. The direct comparison between the lacunarities of one and two sticky-ended CLP helix networks is shown in **Figure S13.**
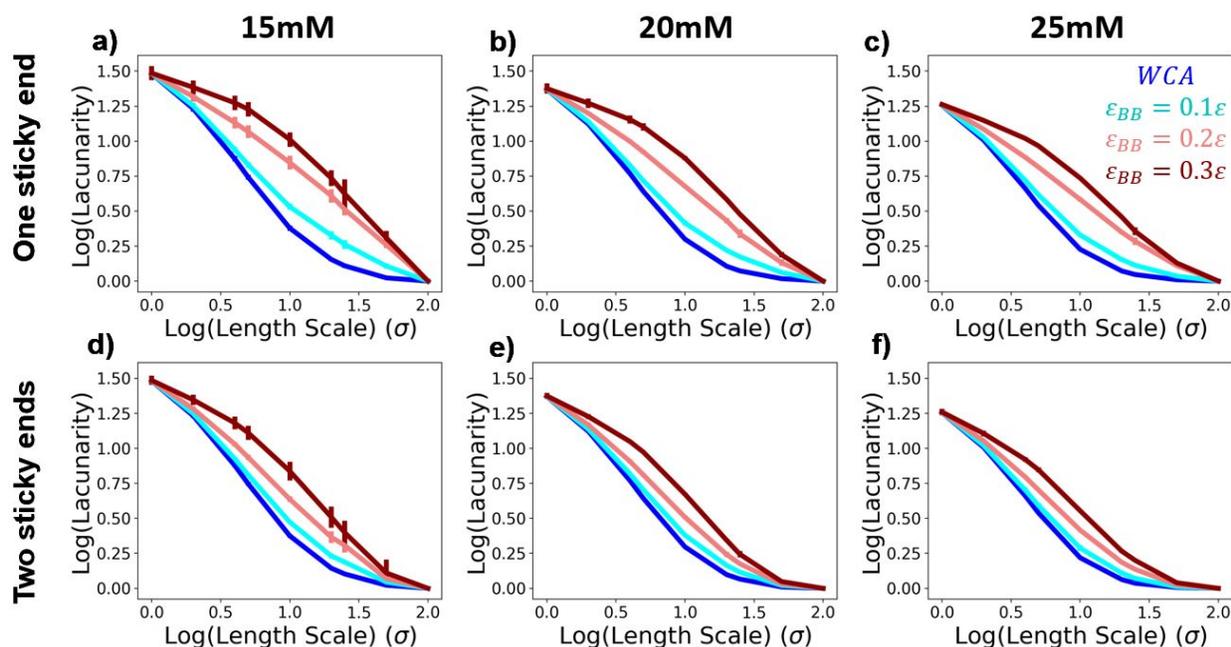
Figure 10. Lacunarity curves observed for the networks formed from CLP triple helices at 15, 20, and $25mM$. The quantification of the length scale used for the x-axis is described in more detail in the Supporting Information. The top row plots correspond to results from one sticky-ended CLP helices while the bottom row plots correspond to two sticky-ended CLP helices; each column corresponds to the value of CLP concentration as indicated. In each plot, the results for each solvent quality are depicted with a dark red curve (ε_{BB} = 0.3 ε; worst solvent), light red curve (ε_{BB} = 0.2 ε), cyan curve (ε_{BB} = 0.1 ε), dark blue (WCA interaction; good solvent). For each simulation, we compute the mean lacunarity distribution over ten snapshots collected every one million

timesteps for the 10 million timesteps of the production simulation. Error bars represent the standard deviation of these means across three independent trials as described in the methods section.

3.2. Effect of Solvent Quality on Assembly of CLP helices at Low Concentration.

We also investigate the impact of solvent quality on CLP triple helix assembly at CLP concentration of 1 mM, which is well below the percolation threshold quantified in our previous work.[12] At this concentration of 1 mM we observe formation of finite fibrillar clusters of CLP helices.

We observe that a moderate solvent quality results in partially aligned one sticky-ended CLP heterotrimeric triple helices within fibrillar structure (**Figure 11a**), while further decreasing solvent quality results in enhanced alignment of triple helices within fibrils (**Figure 11b**). As expected, simulations snapshots also indicate the formation of fibrils with a larger number of helices per cluster with decreasing solvent quality, yielding a system with a less disperse arrangement of CLP helices. Similar results are seen in simulation snapshots for two sticky-ended triple helices, in which dispersed, fibrillar structures are observed at good solvent qualities (**Figure S8a**) and compact clusters of CLP triple helices with highly aligned triple helices within a single cluster arise with worsening solvent quality (**Figure S8b**).

Next, we quantify sizes of clusters of heterotrimeric triple helices assembled via a combination of hydrogen bonding and solvent-induced interactions, in terms of the radius of gyration of the assembled cluster, $R_{g,cluster}$, and number of helices per cluster, $N_{helix,cluster}$. We observe minimal effects of decreasing solvent quality on $R_{g,cluster}$ (**Figure 11c**) for one sticky-ended helices, indicating that at lower concentrations, decreasing solvent quality does not increase the radius of gyration (spatial extent) of clusters, despite the enhanced side-to-side contacts. This is

the result of helices packing more tightly in the cluster so that a larger number of helices can

associate via side-to-side contacts to form fibrils, while the overall size or radius of gyration of the

cluster is unaffected. In agreement with this hypothesis, we observe an increase in the number of

helices per cluster (**Figure S8c**). For two sticky-ended triple helices, we observe that worsening

solvent quality results in clusters of triple helices with smaller radii of gyration (**Figure 11d**). The

smaller $R_{g,cluster}$ values arise due to the compact and tighter packing of CLP triple helices within

fibrils with increasing interactions between backbone beads. Like the one sticky-ended triple

helices, two sticky-ended triple helices form clusters with a greater number of helices per cluster

with decreasing solvent quality (**Figure S8d**). Therefore, although the spatial extent of fibrillar

structures decreases in size with decreasing solvent quality for a two sticky-ended design, more

helices are assembling to form a tightly packed cluster.
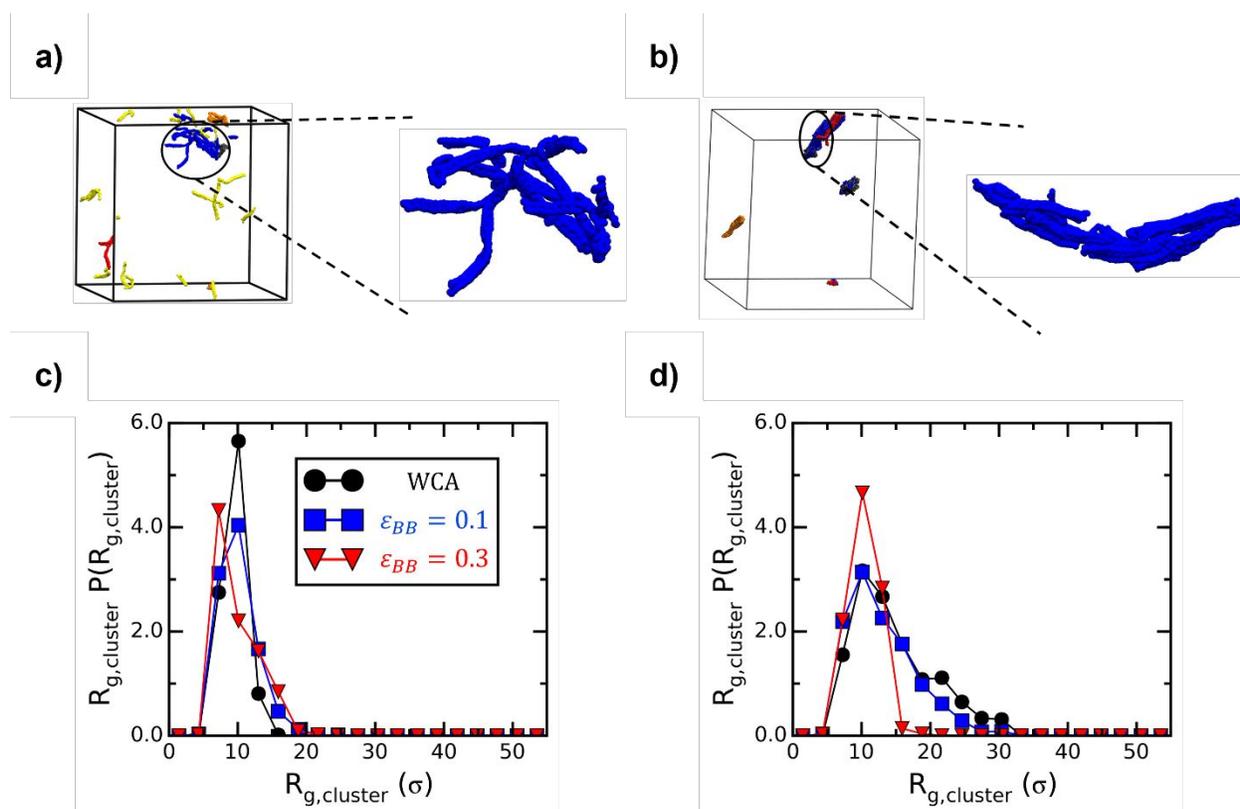
Figure 11. Simulation snapshots of one sticky ended coarse-grained CLP triple helices and weighted probability distributions of radii of gyration of the cluster, $R_{g,cluster}$ for one and two sticky ended heterotrimeric CLP triple helices as a function of solvent quality at a CLP concentration of 1 mM. Snapshots for each solvent quality are depicted with a red curve ($\varepsilon_{BB} = 0.3\ \varepsilon$; worst solvent), blue curve ($\varepsilon_{BB} = 0.1\ \varepsilon$), and black curve (WCA interaction; good solvent). Weighted distributions of radii of gyration are shown for c) one sticky ended and d) two sticky ended CLP triple helices.

4. Conclusion

4.1. Summary of this computational study

In this study, we examined the impact of solvent conditions on the self-assembly of CLP triple helices into fibrillar structures and percolated networks. Specifically, we studied CLP triple helices made with different CLP strand lengths with POG repeating units in all strands; the differences in CLP strand lengths within the same triple helix lead to 'sticky ends' that have exposed hydrogen bonding donor or acceptor groups. We elucidated the extent to which hydrogen bonding between CLP heterotrimer triple helices and the solvent quality impact the structure of CLP triple helix assembly into clusters and networks for one sticky-ended and two sticky-ended CLP heterotrimers.

Using coarse-grained molecular dynamics simulations, we simulated the assembly of CLP triple helices in an implicit solvent and varied the solvent quality by changing the effective interactions between the CG beads representing the CLP strands' amino acids. At higher concentration, we compared the structures of the CLP networks composed of different CLP triple helix designs, at different solvent qualities, through the distributions of network strand lengths and diameters, pore size distributions, fractal dimensions, and lacunarity (heterogeneity) of the simulated self-assembly of CLP heterotrimers into networks. At lower concentrations, we computed the distribution of numbers of CLP triple helices in each assembled cluster and the effective size of each cluster (via the radius of gyration).

At CLP concentrations above the percolation threshold, we observed that a decrease in solvent quality results in more heterogeneous networks with a lower degree of branching, quantified by the lacunarity and fractal dimension, as interactions between individual CLP triple helices become more dominant. We also observed an increase in the diameter of network strands and pore sizes within the network as solvent quality worsens. There was also a nonmonotonic effect of solvent quality on network strand length (distance between physical crosslinks) in which intermediate solvent quality favored smaller network strand lengths, followed by larger network strand lengths observed for the worst solvent quality. When comparing the networks assembled from one and two sticky-ended CLP helices, we found that at good solvent qualities, the two assemble into networks that are similar. As the solvent quality worsened, differences between the one and two sticky-ended networks emerged, with one sticky-ended networks showing lower branching, greater heterogeneity, and larger and broader network strand diameter and pore size distributions.

At CLP concentrations below the percolation threshold, we observed that solvent quality played a larger role in increasing the number of helices per cluster as compared to the radius of gyration (i.e., spatial extent) of clusters of one sticky-ended CLP heterotrimers. For two sticky-ended CLP heterotrimers, simulations showed larger numbers of helices per cluster with decreasing solvent quality but also showed more tightly packed helices with smaller radii of gyration per cluster.

Next, we offer the implications of the above design rules and future work involving the use of solvent quality to study nanoparticle diffusion in heterogeneous networks and leveraging solvent quality as a means for tailoring network structure and mechanical properties of polymer networks.

4.2.   Implications of CLP fibrillar network's structural features on macroscopic properties

The rational design of percolated networks (e.g., hydrogels) for various applications requires a fundamental understanding of the impact of molecular design and solution conditions on the resulting features of assembled network structure such as distribution of pore sizes, physical crosslinks (junctions) density, branching at the crosslinks, and distributions of network strand lengths and network strand diameters. For example, hydrogels are frequently used as carriers for drug delivery with the pore structure of the network governing drug storage, diffusion, and release kinetics.[38, 44] Studies have shown that defects in polyethylene glycol-based networks result in anomalous nanoparticle (quantum dot) diffusion; specifically, nanoparticles smaller than the average pore size were found to be immobile in highly defective networks with an open network structure.[38]   The defective network resulted in bimodal pore size distributions, leading to entrapment of small particles in regions within the hydrogel that have below average pore sizes. In our study in this paper, we have demonstrated that pore structure and network heterogeneity (lacunarity) for hydrogen bonding driven network formation of CLP heterotrimers can be tuned via varying solvent quality, potentially allowing for creation of networks with desired heterogeneity and enabling future studies on the effects of network heterogeneity on drug/nanoparticle diffusion.

The importance of solvent quality on polymer networks has also been demonstrated in several studies in which non-aqueous solvents have been used to tailor the mechanical properties of polymer networks.[45-47] Ahn, et al. demonstrated that the addition of glycerol/water mixtures to alginate hydrogels resulted in loss of intra-chain alginate hydrogen bonding and enhanced chain

flexibility.[47] Moreover, glycerol improved the solvent quality for the alginate molecule while also promoting entanglement of alginate molecules with neighboring chains and superior mechanical properties. While our study in this paper has focused on short, unentangled peptide chains, our simulations suggest that good solvent quality may promote smaller pore sizes and potentially enhanced mechanical properties, in alignment with studies demonstrating that networks with larger pores result in mechanically weak networks due to the large amounts of void volume.[48, 49] Other studies have shown that trace amounts of organic solvent (methanol) in water-organic solvent mixtures promote robust fibrillar assembly of short amyloid peptide fragments containing di-phenylalanine motifs, while such peptide fragments are insoluble in water.[50] These results mirror the simulation results in our study in this paper which show a higher propensity for large diameter fibrils with decreasing solvent quality. Mechanical models, such as the one developed by MacKintosh, Käs, and Jamney,[37] have also been used to explain the scaling of the shear modulus with respect to the mesh size, fibril stiffness, entanglement length, and temperature in a broad range of semiflexible fibrillar networks.[36, 51, 52] Specifically, their model states the shear modulus scales with the inverse square of the network mesh size. The geometric mesh size has been shown to be roughly equivalent to the mean pore size in polymeric solutions,[34] so we expect qualitatively similar scaling of the modulus based on our computation of the pore size distribution.

Overall, this work highlights the complex interplay between solvent quality and CLP design in governing CLP self-assembly into networks and offers design rules regarding the impact of solvent conditions on CLP assembled structure and morphology. The computational work in this paper is meant to guide experimentalists on ways to tailor peptide network structure using polypeptide architecture, polypeptide concentration, and solvent quality with widespread

applications including the design of mechanically robust, fibrillar hydrogels with tailorable drug delivery capabilities.

Supporting Information. Includes additional details for analysis implementation of fractal dimensions, lacunarity, graph-based analysis of network strand lengths and diameters. Additional simulation snapshots, cluster size analyses (radii of gyration of clusters, $R_{g,cluster}$, and number of helices per cluster, $N_{helix,cluster}$), and figures directly comparing results for different CLP designs (one vs. two sticky ends) are reported.

Competing interests: The authors declare no competing interests.

References

1.   M. Zhou, Y. Shmidov, J. B. Matson and R. Bitton, *Colloids and Surfaces B: Biointerfaces*, 2017, **153**, 141-151.
2.   M. E. Roth-Konforti, M. Comune, M. Halperin-Sternfeld, I. Grigoriants, D. Shabat and L. Adler-Abramovich, *Macromolecular rapid communications*, 2018, **39**, 1800588.
3.   Y. Liu, Y. Yang, C. Wang and X. Zhao, *Nanoscale*, 2013, **5**, 6413-6421.
4.   C. Zhang, C. Liu, X. Xue, X. Zhang, S. Huo, Y. Jiang, W.-Q. Chen, G. Zou and X.-J. Liang, *ACS Applied Materials & Interfaces*, 2014, **6**, 757-762.
5.   A. A. Jalan, K. A. Jochim and J. D. Hartgerink, *Journal of the American Chemical Society*, 2014, **136**, 7535-7538.

6.   L. E. O'leary, J. A. Fallas, E. L. Bakota, M. K. Kang and J. D. Hartgerink, *Nature chemistry*, 2011, **3**, 821-828.
7.   V. K. Pal, R. Jain and S. Roy, *Langmuir*, 2019, **36**, 1003-1013.
8.   M. M. Pires, D. E. Przybyla and J. Chmielewski, *Angewandte Chemie International Edition*, 2009, **48**, 7813-7817.
9.   P. J. Stahl, N. H. Romano, D. Wirtz and S. M. Yu, *Biomacromolecules*, 2010, **11**, 2336-2344.
10.   F. W. Kotch and R. T. Raines, *Proceedings of the National Academy of Sciences*, 2006, **103**, 3028-3033.
11.   S. E. Paramonov, V. Gauba and J. D. Hartgerink, *Macromolecules*, 2005, **38**, 7555-7561.
12.   P. A. Taylor, A. M. Kloxin and A. Jayaraman, *Soft Matter*, 2022.
13.   N. Kuznetsova, D. Rau, V. Parsegian and S. Leikin, *Biophysical journal*, 1997, **72**, 353-362.
14.   A. Fiorani, C. Gualandi, S. Panseri, M. Montesi, M. Marcacci, M. L. Focarete and A. Bigi, *Journal of Materials Science: Materials in Medicine*, 2014, **25**, 2313-2321.
15.   J. Engel, H. T. Chen, D. J. Prockop and H. Klump, *Biopolymers: Original Research on Biomolecules*, 1977, **16**, 601-622.
16.   M. A. Cejas, W. A. Kinney, C. Chen, G. C. Leo, B. A. Tounge, J. G. Vinter, P. P. Joshi and B. E. Maryanoff, *Journal of the American Chemical Society*, 2007, **129**, 2202-2203.
17.   I. Malhotra and S. B. Babu, *Pure and Applied Chemistry*, 2018, **90**, 1085-1098.
18.   I. Malhotra and S. B. Babu, *The Journal of Chemical Physics*, 2019, **151**, 084901.
19.   N. Kazem, C. Majidi and C. E. Maloney, *Soft Matter*, 2015, **11**, 7877-7887.
20.   B. Huisman, P. Bolhuis and A. Fasolino, *Physical Review Letters*, 2008, **100**, 188301.
21.   J. E. Condon and A. Jayaraman, *The Journal of Physical Chemistry B*, 2018, **122**, 1929-1939.
22.   H. C. Andersen, *Journal of Computational Physics*, 1983, **52**, 24-34.
23.   M. D. Shoulders and R. T. Raines, *Annual review of biochemistry*, 2009, **78**, 929-958.
24.   J. E. Lennard-Jones, *Proc. R. Soc. Lond. A*, 1924, **106**, 463-477.
25.   M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1**, 19-25.
26.   J. D. Weeks, D. Chandler and H. C. Andersen, *The Journal of Chemical Physics*, 1971, **54**, 5237-5247.
27.   A. M. Hilderbrand, E. M. Ford, C. Guo, J. D. Sloppy and A. M. Kloxin, *Biomaterials science*, 2020, **8**, 1256-1269.
28.   A. M. Hilderbrand, P. A. Taylor, F. Stanzione, M. LaRue, C. Guo, A. Jayaraman and A. M. Kloxin, *Soft Matter*, 2021, **17**, 1985-1998.
29.   S. Plimpton, *Journal of computational physics*, 1995, **117**, 1-19.
30.   A. F. Ghobadi and A. Jayaraman, *Soft matter*, 2016, **12**, 2276-2287.
31.   M. Tuckerman, B. J. Berne and G. J. Martyna, *The Journal of Chemical Physics*, 1992, **97**, 1990-2001.
32.   K. W. Wang, T. Betancourt and C. K. Hall, *Macromolecules*, 2018, **51**, 9758-9768.
33.   S. Bhattacharya and K. E. Gubbins, *Langmuir*, 2006, **22**, 7726-7731.
34.   V. Sorichetti, V. Hugouvieux and W. Kob, *Macromolecules*, 2020, **53**, 2568-2581.
35.   S. Agrawal, S. Galmarini and M. Kröger, *Phys. Rev. E*, 2023, **107**, 015307.
36.   M. Gardel, J. H. Shin, F. MacKintosh, L. Mahadevan, P. Matsudaira and D. A. Weitz, *Science*, 2004, **304**, 1301-1305.
37.   F. MacKintosh, J. Käs and P. Janmey, *Physical review letters*, 1995, **75**, 4425.
38.   K. A. Rose, E. Marino, C. S. O'Bryan, C. B. Murray, D. Lee and R. J. Composto, *Soft Matter*, 2022, **18**, 9045-9056.
39.   J. T. Gostick, Z. A. Khan, T. G. Tranter, M. D. Kok, M. Agnaou, M. Sadeghi and R. Jervis, *Journal of Open Source Software*, 2019, **4**, 1296.

40. D. Sebők, L. Vásárhelyi, I. Szenti, R. Vajtai, Z. Kónya and Á. Kukovecz, *Acta Materialia*, 2021, **214**, 116970.

41. T.-C. Lee, R. L. Kashyap and C.-N. Chu, *CVGIP: Graphical Models and Image Processing*, 1994, **56**, 462-478.

42. A. Hagberg, P. Swart and D. S Chult, *Exploring network structure, dynamics, and function using NetworkX*, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

43. E. W. Dijkstra, 1959.

44. R. Huang, W. Qi, L. Feng, R. Su and Z. He, *Soft Matter*, 2011, **7**, 6222-6230.

45. A. Gravelle, M. Davidovich-Pinhas, A. Zetzl, S. Barbut and A. Marangoni, *Carbohydrate polymers*, 2016, **135**, 169-179.

46. Y. R. Sliozberg, R. A. Mrozek, J. D. Schieber, M. Kröger, J. L. Lenhart and J. W. Andzelm, *Polymer*, 2013, **54**, 2555-2564.

47. Y. Ahn, H. Kim and S.-Y. Kwak, *European Polymer Journal*, 2019, **116**, 480-487.

48. V. Karageorgiou and D. Kaplan, *Biomaterials*, 2005, **26**, 5474-5491.

49. Q. L. Loh and C. Choong, 2013.

50. M. Krysmann, V. Castelletto and I. W. Hamley, *Soft Matter*, 2007, **3**, 1401-1406.

51. N. Y. Yao, C. P. Broedersz, Y.-C. Lin, K. E. Kasza, F. C. MacKintosh and D. A. Weitz, *Biophysical journal*, 2010, **98**, 2147-2153.

52. P. W. Schmidt, S. Morozova, P. M. Owens, R. Adden, Y. Li, F. S. Bates and T. P. Lodge, *Macromolecules*, 2018, **51**, 7767-7775.