Megamolecule Self-Assembly Networks: A Combined Computational and Experimental Design Strategy

Jiangbo Wu,[#,†] *Zhaoyi Gu,*[#,§] *Justin A. Modica,*[§] *Sijia Chen*,[†] *Milan Mrksich*,[*,§] *and Gregory A. Voth*[*,†]

[†] Department of Chemistry, Chicago Center for Theoretical Chemistry, The James Franck Institute, and Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL 60637

[§] Departments of Chemistry and Biomedical Engineering, Northwestern University, Evanston, Illinois 60208, United States

ABSTRACT: This work describes the use of computational strategies to design megamolecule building blocks for the self-assembly of lattice networks. The megamolecules are prepared by attaching four Cutinase-SnapTag fusion proteins (CS fusions) to a four-armed linker, followed by functionalizing each fusion with a terpyridine linker. This functionality is designed to participate in a metal-mediated self-assembly process to give networks. This manuscript describes a simulation-guided strategy for the design of megamolecules to optimize the peptide linker in the fusion protein to give conformations that are best suited for self-assembly, and therefore streamlines the typically time-consuming and labor-intensive experimental process. We designed eleven candidate megamolecules and identified the most promising linker, $(EAAAK)_2$, along with the optimal experimental conditions through a combination of all-atom molecular dynamics, enhanced sampling, and larger-scale coarse-grained molecular dynamics simulations. Our simulation findings were validated and found to be consistent with experimental results. Significantly, this study offers valuable insight into the self-assembly of megamolecule networks and provides a novel and general strategy for large biomolecular material designs using systematic bottom-up coarse-grained simulations.

[#]Contributed equally to this work.

INTRODUCTION

Proteins are versatile building blocks for self-assembled materials, particularly the network-like assemblies that offer unique opportunities for biomedical and material science applications.[1-7] Some protein assembly networks provide exciting applications in cellular functions, drug delivery, diagnostics, and protein therapeutics.[8-11] Recent work has used natural metal-binding amino acids, synthetic ligands, as well as other biomolecular modifications to drive the large-scale assembly of protein networks.[11-16] These approaches often require significant protein engineering efforts, involving the introduction of multiple mutations or chemical modifications to the protein building blocks to establish necessary protein-protein interfaces for geometric specificity.[7,17] These examples demonstrate methods to install synthetic ligands on the protein binding surface, but these strategies are not general or straightforward.[14,18] Furthermore, the choice of protein building blocks is often limited to naturally symmetric proteins or non-symmetric units with pre-existing protein-protein interactions.[12] We previously reported a series of synthetic strategies for synthesizing precisely-defined multiprotein scaffolds, or megamolecules, by way of reactions between enzyme domains and their covalent inhibitors.[19-23]. These approaches have been used to prepare linear, branched, cyclic, and dendritic megamolecules having defined size, orientation, and connectivity,[19-21,23,24] and has been used to prepare megamolecule mimics of therapeutic antibodies which has promising biomedical applications.[10,22]

In this work, we use a four-armed megamolecule that is functionalized with a terpyridine ligand (tpy) on each arm (Figure 1). In this way, the megamolecule serves as a building block that can assemble into a large-scale protein network with addition of divalent metal ions, which form a complex with two terpyridine ligands on neighboring proteins. To synthesize the tetravalent building block, we first expressed a fusion protein of Cutinase and SnapTag (CS fusion) in *E. coli*

hosts. This fusion protein can selectively react with a chloro-pyrimidine (CP) group at SnapTag Cys145[25,26] and with an ethyl p-nitrophenyl phosphonate (pNPP) group at Cutinase Ser120,[27] respectively, to give covalent adducts (Figure S1 of Supporting Information). We then functionalized the SnapTag domain (colored *orange* in Figure 1) at the catalytic residue Cys145 with a heterodimeric linker consisting of a CP group (*orange*) and a terpyridine ligand (colored *purple* in Figure 1), which has a strong coordination affinity with transition metals.[28] Four tpy-functionalized CS fusions (CSt) were then reacted with a symmetrical tetrameric linker to connect the Cutinase domain (colored *blue* in Figure 1) on each arm, forming the Tetra-(Cutinase-SnapTag-Terpyridine) megamolecule, or $(CSt)_4$. Additional details and characterization regarding the synthesis and characterization can be found in the Supporting Information.

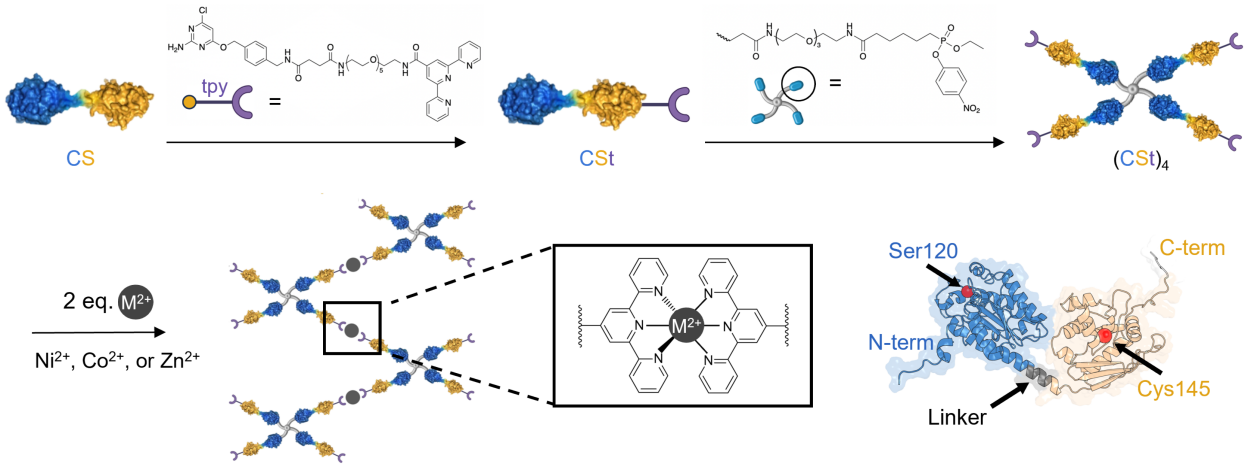

Figure 1. Synthesis and metal-dependent self-assembly of the tetravalent megamolecule building block $(CSt)_4$. A Cutinase-SnapTag fusion protein (CS fusion) was reacted with a terpyridine (tpy) linker to give the functionalized fusion protein (CSt). Four equivalents of CSt were reacted with the tetrameric linker to yield the building block $(CSt)_4$. In the presence of two equivalents of divalent metal ions, the building block is designed to self-assemble into the lattice, mediated by the chelation between two tpy ligands and a divalent metal ion (complex shown in the box). The bottom-right structure depicts the CS fusion with Cutinase Ser120 colored *blue*, SnapTag Cys145 colored *orange*, and the engineered peptide linker colored *grey*.

The CS fusion protein should have certain properties for optimizing the self-assembly process, including the proper positioning of active amino acids (Ser120 and Cys145) and a rigid $(CSt)_4$ skeleton to support and maintain the network. These properties will primarily depend on the peptide sequence (linker) that joins the two proteins in the CS fusion.[29-32] The characteristics of the peptide linker, such as its length, flexibility, and composition, will influence the relative positions of the active sites, with a direct effect on the orientation of the tpy ligands and how the metal-dependent complexation determines the overall symmetry of the self-assembled network. Specifically, the position of the tpy ligand on the fusion protein determines whether interactions are more likely to occur intramolecularly or intermolecularly, which either terminates or facilitates self-assembly, respectively. However, the synthesis and characterization for large-scale screening through experiments are time-consuming, labor-intensive, and expensive. We therefore leverage computer simulations in this work to assist in the design of CS fusions through molecular dynamics (MD) simulations and validate the computational results with experiments.[33,34] On the other hand, regular all-atom (AA) MD alone encounters limitations in handling large-scale biomolecular processes, such as the self-assembly of cytoskeleton and viral capsids,[35-38] long timescale biomolecular processes,[39-41] and the megamolecules explored in this study. To overcome this limitation, we employed a systematic bottom-up coarse-grained (CG) modeling approach,[42-44] which represents a megamolecule with only a small subset of CG beads while preserving their essential dynamics and significantly simplifying the complexity of the system.[44-50]

In this work, we demonstrate how the use of both AA and CG simulations can facilitate the design and screening of candidate CS fusions for use in self-assembling networks. The workflow is illustrated in Figure 2, including AA MD simulations augmented by enhanced free energy sampling, while both are employed to calculate the favorable conformation and rigidity of CS

fusions. Subsequently, intensive CG simulations are conducted to explicitly mimic the actual assembly process under diverse experimental conditions, including variations in megamolecule concentrations, types of metal ions, metal-to-megamolecule ratios, and linker types. We also include experimental validation to confirm the computational results. Finally, guided by insights from the computational investigation, we synthesized and characterized the candidate $(CSt)_4$ megamolecule building blocks under optimized conditions. As predicted by simulations, we observed significant conversion of the optimal building blocks into a macroscopic gelatinous product, while the negative controls showed minimal assembly.

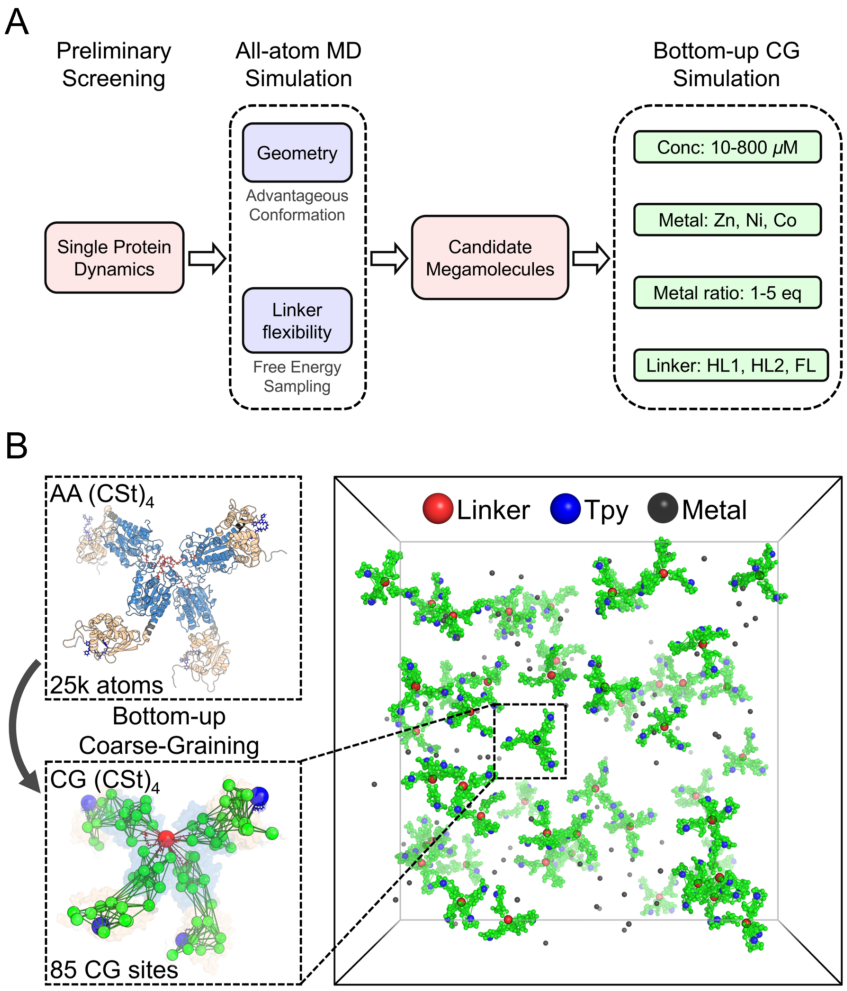

Figure 2. (A) The computational workflow used in this study. The strategy starts with the exploration of single protein dynamics of Cutinase-SnapTag proteins (CS fusion) through all-atom MD simulations, which are used to screen fusion proteins and identify those with an optimal geometry and flexibility for assembly into networks. Next, the candidate CS fusions are assembled into $(CSt)_4$ megamolecules, and their self-assembly is explicitly simulated using bottom-up coarse-grained MD simulations under different experimental conditions. (B) A representative snapshot of a cubic simulation box with periodic boundary conditions containing fifty $(CSt)_4$ megamolecules, with the central tetrameric linker, tpy, and metal colored *red*, *blue*, and *black*, respectively. An 85-bead CG model is constructed based on the corresponding all-atom model comprising ~25,000 atoms.

EXPERIMENTAL SECTION

Computational Methods. The input structures of CS fusions for AA MD simulations were predicted using AlphaFold2,[51] as displayed in Figure S2. Each CS fusion was solvated by packing TIP3P water in a cubic box within at least 10 Å from the box edges, neutralized with a minimal amount of Na^+ ions, and parameterized by CHARMM36m force field.[52] For each CS fusion, we conducted ~1000 ns × 3 replica simulations to relax the AlphaFold-predicted structures using GROMACS 2021.5 MD software.[53,54] The initial 100 ns of each simulation was excluded as equilibration. Umbrella sampling[55] was then employed to calculate the free energy profile associated with pulling the Cutinase and SnapTag domains apart. The potential of mean force (PMF) was then reconstructed using the weighted histogram analysis method (WHAM).[56]

For comparison with the AA MD results and validation with the experimental data, we constructed CG models of $(CSt)_4$ to simulate the megamolecule self-assembly. Gromologist was used to set up the AA simulations system of $(CSt)_4$.[57] The development of each CG model was based on the corresponding AA trajectories. Essential dynamics coarse-graining (ED-CG)[45] was used to generate the CG mapping with a resolution of ~20 residues/CG bead. Additionally, five CG beads were assigned to one tetrameric linker and four tpy ligands (Figure 2B). Thereafter, a heterogeneous elastic network model[58] (HeteroENM) was used to capture the internal motion and

parameterize the intramolecular interactions. The intermolecular interactions between CG beads were represented using a mixed potential of soft and inverted Gaussian functions to mimic their repulsive and attractive nature. All of the previously mentioned methods are implemented in the OpenMSCG software package.[59] The overall workflow of the CG modeling is shown in Figure S12. The interactions between terpyridine and divalent metal ions were determined using Gaussian16 software[60] and fitted by a Morse potential. Packmol[61] was implemented to sample the initial configurations in a cubic box containing fifty $(CSt)_4$ megamolecules. For each condition, ~10^9 CG MD steps × 6 replica CG MD simulations were carried out using LAMMPS 20Sep2021 MD software[62] by integrating Langevin dynsmics[63] with a timestep of 10 fs at 300 K. The initial 8 μs of the simulation was excluded from the subsequent analysis. A complete description of the simulation details can be found in the Supporting Information.

Synthesis of $(CSt)_4$ Megamolecules. The synthesis of $(CSt)_4$ is outlined in Figure 1 and detailed in the Supporting Information (Figure S1). Briefly, the CS fusion protein was reacted with the tpy linker to give CSt, which was subsequently treated with a tetrameric linker to produce the final product $(CSt)_4$. The functionalization of the SnapTag domain with the tpy linker was confirmed by Electrospray Ionization Mass Spectrometry (ESI) (Table S1). The purity and molecular weight of $(CSt)_4$ were verified by gel-electrophoresis (Figure S10) and mass photometry (Figure S11 and Table S2), respectively. The purification histag was removed using Tev protease to avoid unwanted metal coordination. A control experiment was conducted to validate that metal-mediated assembly occurred exclusively with tpy-functionalized protein (Figure S5).

Mass Photometry (MP) Measurement of $(CSt)_4$ Megamolecules Assembly. The $(tSC\ HL2)_4$ Megamolecules (20 mM HEPES, pH 7.4, 150 mM NaCl, 10 μM EDTA) (39 μM) were mixed with 2 eq of $NiCl_2$, $CoCl_2$, or $ZnCl_2$, and incubated at room temperature overnight. For the metal

equivalence experiment, the (tSC HL2)$_4$ megamolecule (39 μM) was mixed with $NiCl_2$ ranging from 0-5 equivalents. EDTA (10uM) was added to the buffer to prevent protein monomer assembly with trace metals in the buffer. An additional 10 μM of metal was added at each condition to neutralize the EDTA. The resulting reaction mixtures were centrifuged at 6000 rpm for 5 minutes.

All MP data were acquired using a Refeyn TwoMP mass photometer. For each acquisition, an initial 10 μL of filtered buffer was applied on a silicone gasket on a microscopy slide to focus the microscope. The diluted protein solution (10 μL, 80 nM) was prepared immediately before measurement, added to the gasket containing 10 μL of buffer, and mixed well with a micropipette to achieve a final concentration of 40 nM. Movies were recorded for 60 seconds and processed using DiscoverMP. Contrast values were converted into molecular weights using a calibration curve constructed with Thyroglobulin (670 kDa) and β-amylase (56, 112, and 224 kDa), with calibration performed every 2 hours to maintain accuracy. Species counts were obtained from Gaussian fitting to the MP distribution. The percentage of monomer and oligomer was calculated by dividing the number of each species by the total counts, and the average multimer size was determined by accounting for the number of valencies. Each sample was measured independently three times.

UV-Vis Analysis of (CSt)$_4$ Megamolecules Assembly. The (CSt HL1)$_4$, (CSt HL2)$_4$, (CSt FL)$_4$, and (tSC HL2)$_4$ megamolecules ranging from 19.5 μM to 156 μM were mixed with 1.5 eq of $NiCl_2$ in solution and incubated at room temperature overnight, followed by centrifugation at 6000 rpm for 5 minutes. The absorbance of the supernatant was measured at 280 nm to determine protein concentration in the supernatant and at 330 nm to assess Ni^{2+} coordination with the tpy ligand. Conversion of the (CSt)$_4$ network was calculated using the measured absorbance and the extinction coefficient of protein and terpyridine.

Preparation and Macroscopic Characterization of $(CSt)_4$ Megamolecules Assembly. The assembly solution (8 μL) was prepared on silicone gaskets (Grace Bio-Labs reusable CultureWellTM gasket, diameter x depth 3 mm x1 mm) on a glass slide. The protein solution was mixed with 1.5 eq of $NiCl_2$ in HEPES buffer and incubated at room temperature overnight. The $(CSt\ HL2)_4$ Megamolecule (156, 117, and 78 μM, respectively) rapidly formed a dense, gelatinous material, which was imaged using an optical microscope. Assemblies of $(CSt\ HL1)_4$, $(CSt\ HL2)_4$, $(CSt\ FL)_4$, $(tSC\ HL2)_4$ ranging from 19.5 μM to 156 μM were prepared similarly. Their images were captured by a camera.

RESULTS AND DISCUSSION

Design of the CS Fusions. We identified a series of tetrameric proteins that were designed to self-assemble into regular network lattices. They were each based on a CS fusion protein (**CS**) and differed in either the peptide linker that tethered the Cutinase and SnapTag proteins or the orientation of the fusion protein. Specifically, we designed eight CS fusions that lacked a linker (CS w/o linker), that had helical linkers having one or two copies of EAAAK (CS HL1, CS HL2), XTEN flexible linkers (CS FL), (Pro)3 or (Pro)9 linkers (CS P3, CS P9), three repeats of the Asp-Pro dipeptide linker (CS DP3), HL2 linkers without the first 16 amino acids (CS HL2 w/o pro), and three N-terminus SnapTag fusions (SC HL1, SC HL2, SC HL5). The complete structures of those CS fusions are shown in Figure S2. The sequences of the fusion proteins can be found in the Supporting Information.

The Screening of Candidate CS Fusions. We hypothesized that a CS fusion connected by a rigid linker would be important for holding the two enzyme domains apart and for minimizing the potential conformational heterogeneity. To measure the rigidity of various CS fusions, we

employed umbrella sampling to calculate the PMF associated with pulling the Cutinase and SnapTag domains apart. The reaction coordinate was chosen as the center of mass (COM) distance between the Cutinase and SnapTag domains, with the equilibrium $d_{COM} = 0$ Å. As shown in Figure 3B, the HL2 linker provides the most rigid fusions due to its alpha-helix structure.[30] Notably, most linkers enhance the rigidity of CS fusions compared to the unmodified fusion. On the other hand, the solvent-accessible surface area (SASA) of the Cutinase and SnapTag domains can measure the structure compactness between the two domains. A larger SASA indicates an extended framework of a CS fusion. As shown in Figure 3C, CS HL2 and CS HL1 have the largest SASA values, indicating that the domains are held apart from one another. The Ser120-Cys145 distances for all CS fusions are shown in Figure 3D. SC HL1 and SC HL2 have the smallest distances, suggesting a more compact structure. These smaller distances likely reduce conformational flexibility, effectively locking the ligands in positions that favor intramolecular assembly. This is further supported by the expanded angle observed, which will be discussed in detail in the next paragraph.

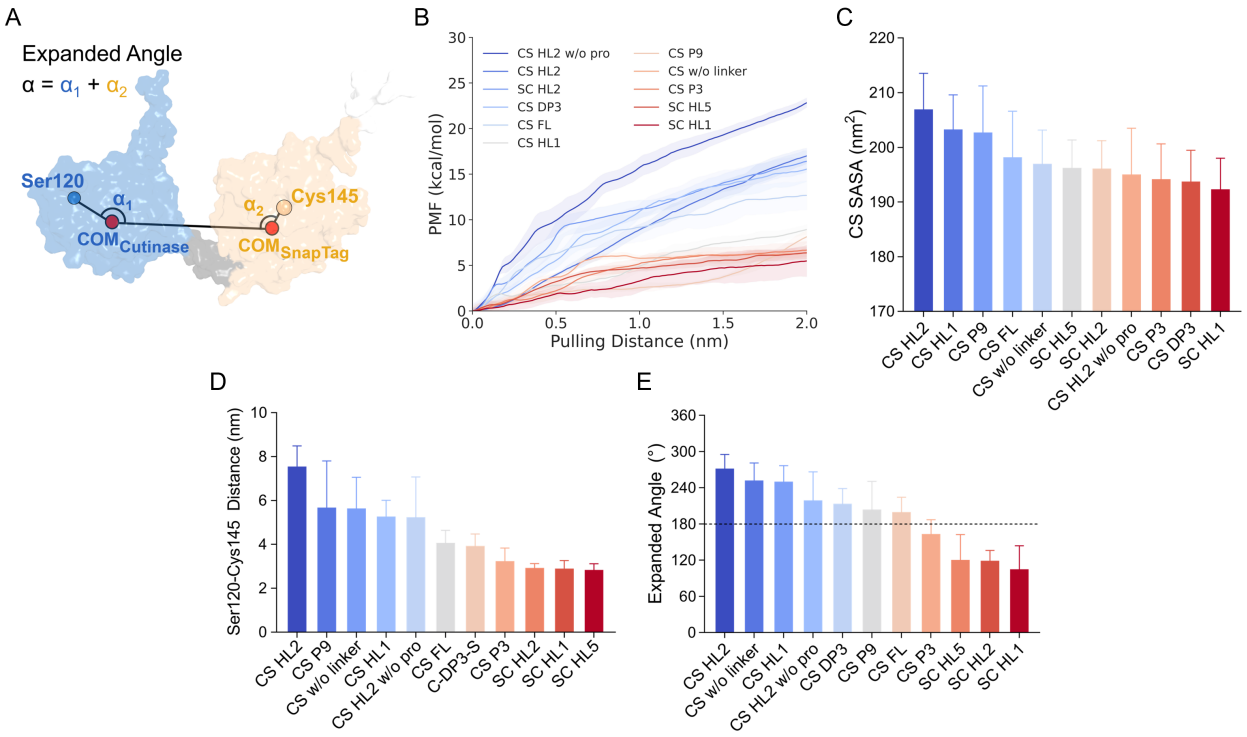

Figure 3. Insights from all-atom simulations of the CS fusion proteins to compare their conformations and rigidities. (A) Definition of the expanded angle. (B) Potential of mean force (PMF) as a function of the pulling distance between the Cutinase and SnapTag domain, starting from the equilibrium state. (C) Solvent-accessible surface area (SASA), (D) Ser120-Cys145 distance and (E) expanded angles for all the CS fusions, respectively. Error bars indicate the standard deviations throughout MD trajectories after 100 ns.

Further, we believe that the best fusion proteins would have the Cutinase and SnapTag active sites positioned at opposite ends of the fusion proteins and not, for example, at the sides or even oriented towards the center. This orientation would promote an end-to-end assembly of the $(CSt)_4$ building block in the network. We define this conformation using an expanded angle, defined as the sum of the two angles formed by each active site and the COM of the two domains (Figure 3A). The results for the expanded angles are presented in Figure 3E, where CS HL2 gives the most open structure, while SC HL1 and SC HL2 maintain a more closed configuration. Notably, the directionality of CS fusions significantly impacts their behavior due to the asymmetry of their active sites. When Cutinase is positioned at the N-terminus and connected with the helical HL2

linker, the helix at its C-terminus can extend further, resulting in a longer helix and a more rigid, structured scaffold. In contrast, when SnapTag is at the N-terminus, it terminates with a flexible loop, which does not support the same helical extension. The Ser120-Cys145 distance and the extended angles also reveal that Cutinase adopts the desired active site orientation when it is included as the N-terminus protein. Suggested by the insight from AA MD simulations, we identified CS HL2 as the optimal CS fusion for assembling $(CSt)_4$. Based on this analysis we selected four megamolecules—CSt HL1, CSt HL2, CSt FL, and SC HL2—for modeling using the CG method, as described next.

Varying the Length of Tpy Linker. To ensure an accessible presentation of the tpy ligand at Cys145, we introduced variations in the length of the linker used to attach the tpy ligands to the Cutinase domain. We did this by modifying the number of $[CH_2CH_2O]_n$ (EG groups), to include one, two, three and five glycol units (Figure S3). We measured the distance between terpyridine and Cys145 within the same CSt domain of a $(CSt)_4$ megamolecule. As Figure 4B shows, there is no significant difference for the number of EG groups. We believe this finding can be explained by a hydrophobic effect, where the $(EG)_n$ chains tend to coil rather than adopt an extended conformation.[64] This result is also consistent with data from dynamic light scattering experiments (Figure S3C), where the Z-average diameters of different tpy linker lengths do not exhibit significant differences or a clear trend. We therefore used the $(EG)_5$ linker in the subsequent theoretical computations and experiments because it offers sufficient EG groups to improve tpy linker solubility in aqueous solutions, making it a practical choice for experimental procedures.[65]

Concentration Dependence. The AA MD simulations described above are unable to handle the complex self-assembly process for the megamolecule system. For example, a cubic AA box containing fifty $(CSt)_4$ at a concentration of 50 mM would contain ~168 million atoms. Hence, CG

MD simulations were conducted to probe the assembly behavior. We defined the connection number (CN) as a metric, similar to the coordination number commonly used in crystallography, to quantify the average number of intermolecular connections of all $(CSt)_4$ in a simulation box, taking into account the periodic boundary conditions. A value of CN = 1, for example, describes a system where each $(CSt)_4$ building block is metal-coordinated to one other molecule in a dimer. A CN value of 2 describes a system where the building blocks are present in a polymer-chain configuration, or a one-dimensional assembled network. A value of CN = 4 represents the maximum possible CN that $(CSt)_4$ megamolecule can achieve, which is a fully interconnected network structure. We calculated the CN of $(CSt\ HL2)_4$ at different concentrations of the molecule and find the expected result that CN increases with higher concentrations, demonstrating that elevated concentrations favor the assembly (Figure 4A). This is in good agreement with the DLS data and SEC traces depicted in Figure S3C and Figure S4. A concentration of ~200 μM was found to be necessary to sustain extended self-assembly.

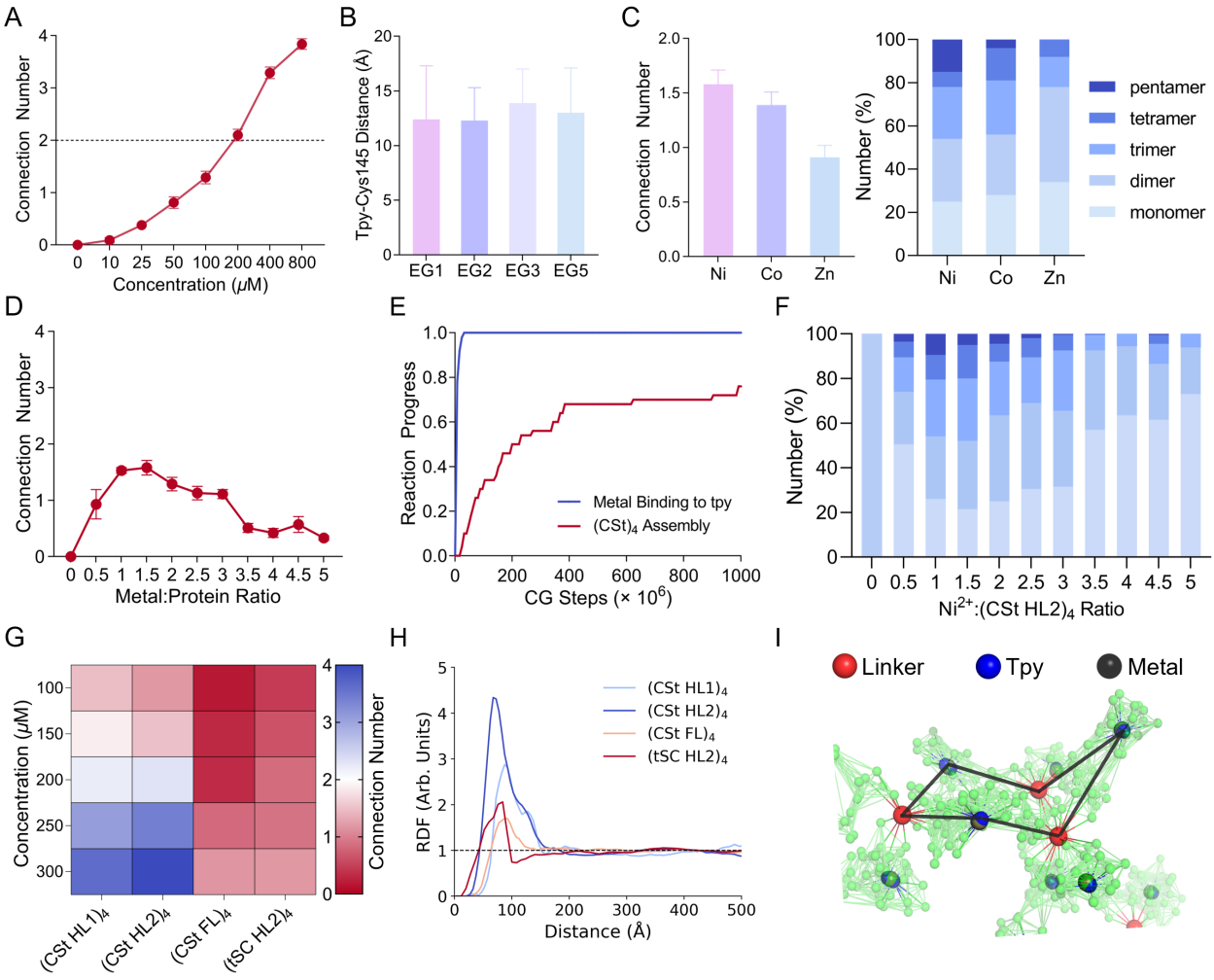

Figure 4. Coarse-grained molecular dynamics simulations. (A) Connection number (CN) for various concentrations of $(CSt\ HL2)_4$, with a dotted *black* line indicating the threshold CN = 2. (B) Distances between Tpy-Cys145 within the same CSt domain of a $(CSt)_4$, varying with the number of $(EG)_n$ groups. (C) CN and system compositions of 100 mM $(CSt\ HL2)_4$ megamolecules reacting with 2 eq of different metal ions. (D) CN values of 100 mM $(CSt\ HL2)_4$ varying with different metal:protein ratios. (E) Reaction progress of $(CSt\ HL2)_4$ assembly (*red*) and metal binding to tpy (*blue*) as a function of simulation timesteps. (F) Megamolecules compositions of panel E. (G) CN grids for various megamolecule types at different concentrations. (H) Radial distribution functions (RDF) with respect to the distances between all the central tetrameric linkers in different 200 μM megamolecules reacting with 2 eq of Ni^{2+} ions. (I) A snapshot of the chair conformation of the $(CSt\ HL2)_4$ assembly system, highlighting the central tetrameric linker in *red*, the metal ions in *grey*, and the tpy ligand in *blue*.

Varying Divalent Metal Types. Terpyridine forms dimeric complexes with a wide range of divalent transition metals.[66] Ni^{2+}, Co^{2+}, and Zn^{2+} are three commonly used transition metals that have been used in protein assembly and exhibit different orders of magnitude of kinetic and

thermodynamic constants for complexation.[67,68] We used these three metal ions to mediate the self-assembly process. The affinity of the metal ion for the ligand can affect the partitioning of intramolecular and intermolecular complexation (which also depends on the conformational properties of the $(CSt)_4$ megamolecule), though in non-obvious ways. We therefore used CG MD simulations to explicitly compare the self-assembly process in the presence of these metal ions. We added the metal ions (2 eq) to $(CSt\ HL2)_4$ (100 μM). As shown in Figure 4C, Ni^{2+} gives the most favorable results with a CN value of 1.58 ± 0.13, while Zn^{2+} demonstrates the lowest CN value of 0.91 ± 0.11. To characterize the megamolecule compositions for the different metal ions, we calculated the molecular percentage of monomer, dimer, trimer, tetramer, and pentamer of each system (Figure 4C). The Ni^{2+} ion gives the lowest proportion of intramolecular monomers and the highest proportion of pentamer assemblies.

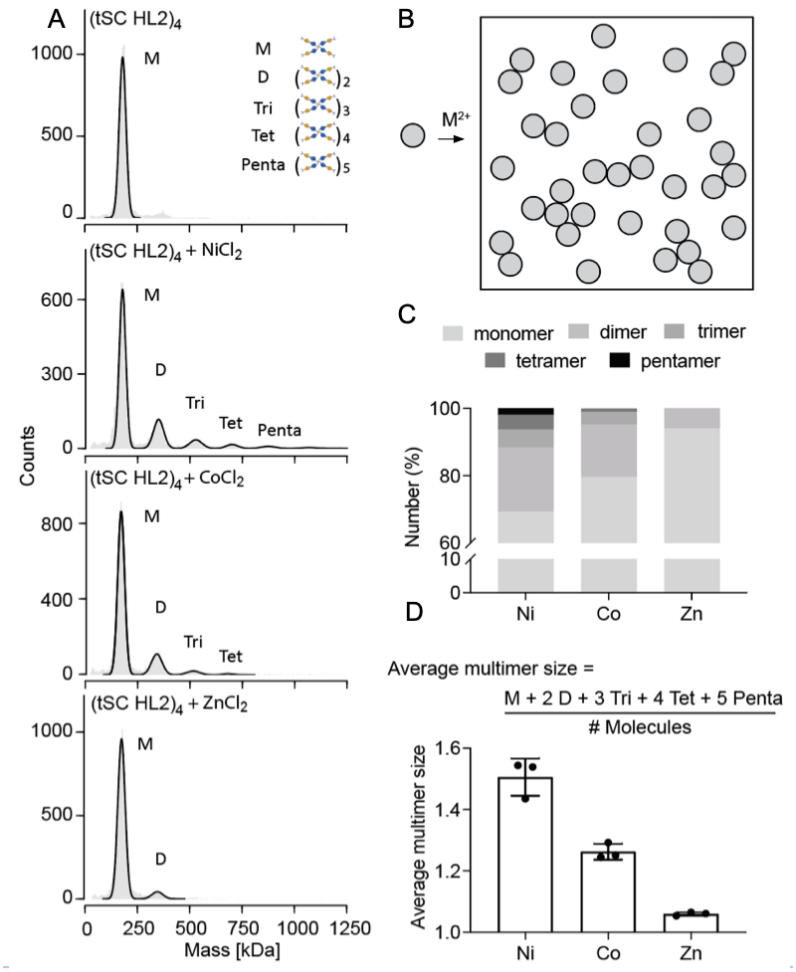

Figure 5. Mass photometry (MP) analysis of $(CSt)_4$ assembly using various metal ions. (A) MP distribution of $(tSC\ HL2)_4$ control and multimers assembled with $NiCl_2$, $CoCl_2$, and $ZnCl_2$. (B) Cartoon representation of $(CSt)_4$ oligomers mediated by metal coordination. (C) Monomer, dimer, trimer, tetramer, and pentamer populations calculated from the areas of the Gaussian peaks based on the MP distribution. (D) Average size of multimers in the $(tSC\ HL2)_4$ assembly. $NiCl_2$ resulted in the most extensive assembly.

To compare these results to experiment, we separately treated the $(CSt\ HL2)_4$ building block with the three metal ions. However, in the presence of $NiCl_2$ (2 eq), we observed significant precipitation with $(CSt\ HL2)_4$, even at a low protein concentration, as evidenced by a decrease in absorbance at 280 nm in the supernatant (Figure S6). This precipitation made it challenging to quantitatively analyze the assembly distribution in solution. We therefore switched to the (tSC

HL2)$_4$ building block, which did not precipitate under the same conditions. This allowed more straightforward quantification of assembly composition in the solution phase using mass photometry, a label-free single-molecule technique.[69] The outcomes of various conditions were quantified by the percentage and average multimer size of soluble oligomers in solution.

We expected the choice of metal to significantly influence assembly outcomes due to varying binding affinity for terpyridine. We tested the three metals simulated previously ($NiCl_2$, $CoCl_2$, and $ZnCl_2$). We again mixed 2 eq of metal with (tSC HL2)$_4$ (39 μM) and incubated the solutions at room temperature overnight to allow the assembly to equilibrate. The (tSC HL2)$_4$ monomers were expected to form a distribution of monomers and oligomers, as depicted in Figure 5B. We used mass photometry (MP) to identify the amounts of each oligomeric form and we observed formation of dimers, trimers, tetramers, and pentamers when $NiCl_2$ was used (Figure 5A and Figure S7). Mass photometry provides the number of counts for each molecule at a given mass, with the peak area corresponding to the total number of molecules. Using these counts, we calculated the percentage of monomers and oligomers, as well as the average multimer size. Approximately 30% of total molecules comprised species larger than monomers, indicating that 55% of starting monomers assembled into larger species (Figure 5C). In contrast, the final assembly mixture contained 85% and 94% of monomers with $CoCl_2$ and $ZnCl_2$, respectively. Figure 5D summarizes the average multimer size for these three metals, with Ni^{2+} producing the largest average multimer size of 1.51 ± 0.06. These results align with CG simulation trends (Figure 4C) and previous reports on their thermodynamic constants.[67] Even though the rate constants for formation of the single liganded metal ion complex follows the order $Ni^{2+} < Co^{2+} < Zn^{2+}$, the Ni^{2+} complexes exhibit significantly higher stability (β_2 = 21.8), with a stability constant three orders of magnitude higher compared to Co^{2+} (β_2 = 18.3), which is dictated by the dissociative rate

constant (Ni^{2+} < Co^{2+} < Zn^{2+}).[66] This strong Ni-tpy coordination serves as a potent driving force for protein assembling into large oligomeric complexes.

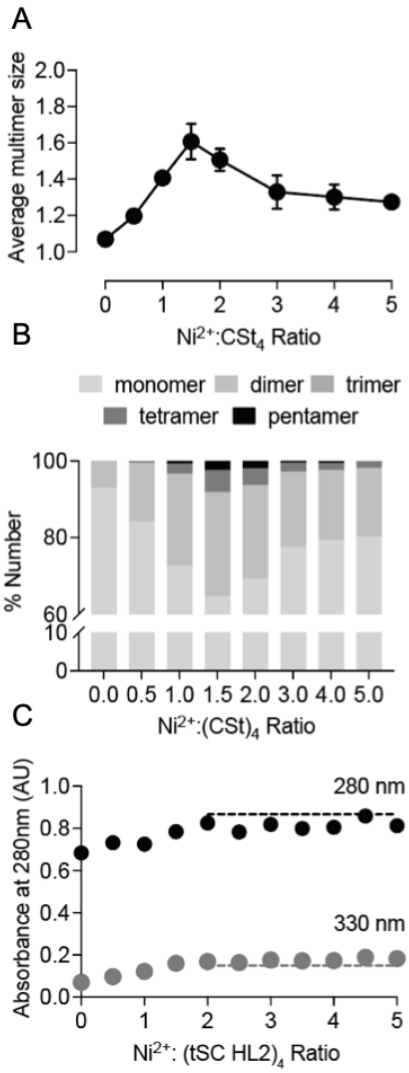

Figure 6. Analysis of $(CSt)_4$ assembly mediated by different equivalents of $NiCl_2$. (A) Average size of multimers in the $(tSC\ HL2)_4$ assembly when treated with 0-5 eq of $NiCl_2$ relative to $(tSC\ HL2)_4$. (B) Monomer, dimer, trimer, and tetramer populations calculated from the MP distribution. (C) Absorbance at 280 nm and 330 nm of the $(tSC\ HL2)_4$ assembly when treated with 0-5 equivalents of NiCl2. The dashed line represents the expected absorbance value when tpy ligands are saturated with $NiCl_2$.

Varying the Ratio between Metal Ions and $(CSt)_4$. Despite our initial assumption that 2 eq of metal is ideal for the formation of fully connected networks, here we use simulations to examine

how the CN varies with different Ni^{2+}:$(CSt\ HL2)_4$ ratios, for a concentration of megamolecule of 100 μM. We find that the CN number reaches its maximum value when metal:protein = 1.5 and then gradually diminishes (Figure 4D). To further explore this result, we analyzed the reaction progress of $(CSt)_4$ assembly and metal binding to tpy ligands throughout the course of simulation timesteps. Figure 4E reveals that the binding of metal to tpy ligands occurs rapidly within the initial 30×10^6 CG MD steps and much faster than the subsequent assembly of $(CSt)_4$. This suggests that metal ions primarily occupy almost all the possible tpy ligands first before megamolecule self-assembly occurs. The compositions of megamolecules at different metal:protein ratios are shown in Figure 4F. The simulation results indicate that directly incorporating 2 eq or more metal ions into the megamolecules leaves fewer free tpy sites for coordinating with metal-occupied tpy sites. By introducing fewer equivalents of metal ions, more tpy sites are left available to the binding of a second metal-occupied tpy ligand. Notably, the monomer population exhibits a basin-like pattern with its bottom at metal:protein = 1.5. Once a tpy ligand coordinates with a metal ion, it can quickly engage with its neighboring tpy ligand through fast intramolecular interaction to form a bis-terpyridine complex.[67] Although CS HL2 was designed with optimized active sites, expanded angles, and domain positions to favor intermolecular interactions, competition between intra- and intermolecular interactions still occurs. Due to the inherent conformational flexibility of fusion proteins, and because bis-complex formation is fast and does not require a second protein megamolecule to diffuse into proximity, intramolecular interaction can happen rapidly. This process reduces the number of free tpy sites for intermolecular interaction. Adding slightly fewer equivalents of metal ions preserves more free tpy sites, even in the presence of rapid intramolecular interactions. These additional free tpy sites are crucial for enabling the intermolecular interactions and forming large assembly networks.

We next experimentally determined the influence of metal-to-protein ratio on the assembly of $(tSC\ HL2)_4$. Mass photometry analysis revealed that mixing $NiCl_2$ and $(tSC\ HL2)_4$ at a molar ratio of 1.5 eq yielded the largest sized oligomers (Figure 6A). At 1.5 eq $NiCl_2$, the percentage of larger oligomers (trimer, tetramer, and pentamer) and its average multimer size were the highest (Figure 6A and 6B). The average multimer size decreases with lower metal equivalents due to insufficient metal ions for coordination, and with higher metal equivalents due to the depletion of free tpy ligands for coordination. This finding is consistent with the results of the CG MD simulation in Figure 4F. Additionally, achieving full saturation of the tpy ligand on the protein required a stoichiometric ratio of 2 eq $NiCl_2$, as shown by UV-vis analysis in Figure 6C. We monitored the absorbance of the protein solution at 280 nm and 330 nm after $(tSC\ HL2)_4$ was treated with 0-5 eq $NiCl_2$. Both absorbance curves plateaued over 2 eq of $NiCl_2$, indicating that at least 2 eq $NiCl_2$ is required to fully saturate the tpy ligands on $(CSt)_4$. The absorbance at 280 nm did not decrease, suggesting that all protein assembly complexes remained soluble in the solution.

The Impact of Different Linkers. The flexibility of the peptide linker connecting the two protein domains and the domain directionality are two other important factors for assembly. We used simulation to analyze the assembly results of different megamolecules reacting with 2 eq Ni^{2+} ions. Here, $(CSt\ HL2)_4$ exhibits the highest CN, confirming it as the most promising candidate for formation of self-assembled networks (Figure 4G). In contrast, both $(CSt\ FL)_4$ and $(tSC\ HL2)_4$ exhibit small CN values (<1.5) across all the grid points, indicating their inherent unsuitability for assembly. To further understand the specific connectivity of the assembly system, we calculated the radial distribution function (RDF) between the central tetrameric linker for the assembly of different megamolecules, at 200 μM, with 2 eq of Ni^{2+} ion. The RDF, as shown in Figure 4H,

illustrates different first-peak intensities unique to each megamolecule assembly process, where $(CSt\ HL2)_4$ features the highest and sharpest first-peak, followed by $(CSt\ HL1)_4$, $(tSC\ HL2)_4$ and $(CSt\ FL)_4$, which indicate that $(CSt\ HL2)_4$ forms more ordered and compact assemblies. For comparison, $(tSC\ HL2)_4$ and $(CSt\ FL)_4$ exhibit lower peaks, indicating fewer neighboring molecules. To avoid further confusion, it is important to note that due to the inefficiency in simulating diffusion, the box was limited to having fifty $(CSt)_4$ megamolecule building blocks. This restriction is necessary to manage the simulation box size effectively. Also, the simulations were conducted using the constant *NVT* ensemble. Consequently, only a few peaks can be observed in Figure 4H. Figure 4I depicts a snapshot from the $(CSt\ HL2)_4$ system, showing an ideal chair conformation of a megamolecule trimer.

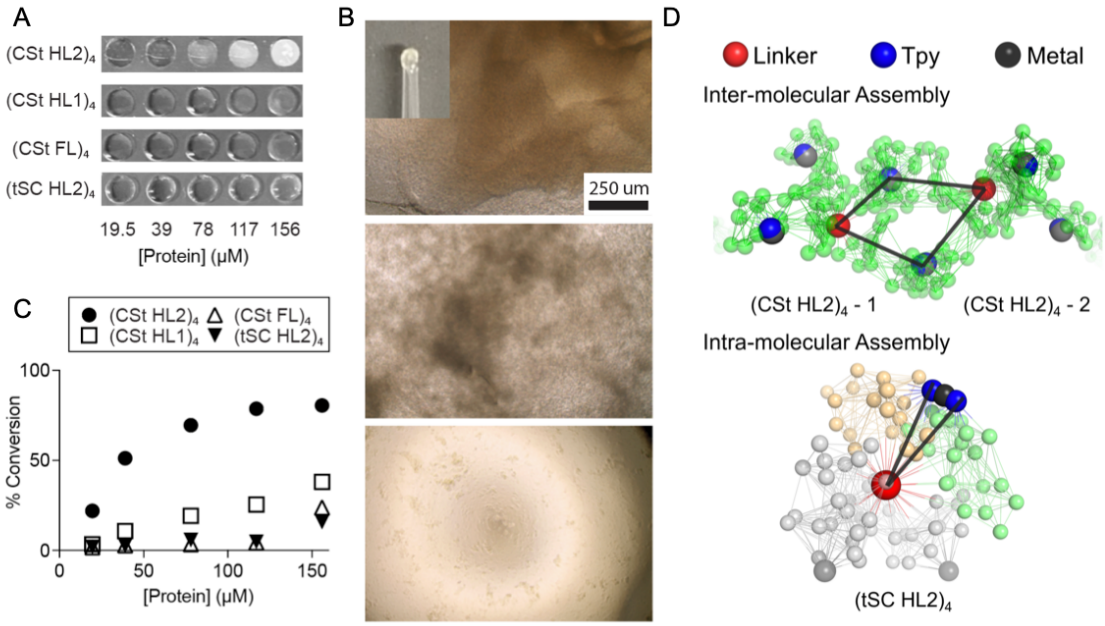

Figure 7. Comparison of $(CSt)_4$ assembly with various CS building blocks. (A) Images of gelatinous products for the assembly of four megamolecule building blocks: CS HL2, CS HL1, CS FL, and SC HL2. Proteins at concentrations ranging from 19.5 to 156 μM were treated with $NiCl_2$ and incubated at room temperature for 24 hr. (B) Macroscopic characterization of $(CSt\ HL2)_4$ assembly at 156 μM (top), 117 μM (middle), and 78 μM (bottom) reveals the formation of gelatinous assemblies across all three concentrations. (C) Precipitated $(CSt)_4$ versus the protein concentration. (D) The snapshots of the intermolecular assembly of two $(CSt\ HL2)_4$ megamolecules

and the intramolecular assembly of two tpy ligands within the same (tSC HL2)$_4$, highlighting the central tetrameric linker in *red*, the metal ions in *grey* and the tpy ligands in *blue*.

Finally, we experimentally explored assembly of the CS fusion candidates into extended networks. We mixed either (CSt HL2)$_4$, (CSt HL1)$_4$, (CSt FL)$_4$, or (tSC HL2)$_4$ with $NiCl_2$ at protein concentrations ranging from 156 μM down to 19.5 μM. To clearly visualize the gel formation, we prepared the assembly mixtures in a gasket (Figure 7A). The (CSt HL2)$_4$ solution immediately gave a viscous, gelatinous material upon mixing with $NiCl_2$ solution. The material was sticky and would cling to a pipette tip when brought into contact (Figure 7B). To determine the fraction of megamolecule building blocks that were incorporated into this insoluble network, we prepared the reactions under the same conditions, allowed them to incubate overnight, and then centrifuged any insoluble material and measured the absorbance of the supernatant at 280 nm. (CSt HL2)$_4$ displayed the highest conversion, starting at 50% at 39 μM, 70% at 78 μM, and 80% at 156 μM (Figure 7C). Scaffolds containing rigid linkers (CS HL1 and CS HL2) yielded higher conversion of gelatinous products, while those with flexible linkers or with N-terminal SnapTag did not produce any macroscopic material until the concentration reached 156 μM. This result is consistent with the simulated (CSt HL2)$_4$ assembling in the concentration range shown in Figure 4G, where a connection number of approximately 2 was achieved when concentration was increased to 150 μM and 200 μM. To elucidate the molecular-level differences of these building blocks, we reexamined the conformation and connectivity of individual molecules in the resulting assembly mixture generated by CG simulation. We notice that the conformation of (CSt HL2)$_4$ is more extended. This extended conformation likely renders intramolecular interaction less favorable and allows more time for the tpy ligands to coordinate with another protein through intermolecular interactions. In contrast, the assembly of many (tSC HL2)$_4$ molecules is prematurely terminated

due to the intramolecular interaction of tpy ligands on the arms of the same molecule (Figure 7D). The position of tpy ligands is less exposed for intramolecular assembly. This observation is consistent with the expanded angle and active site separation in Figure 3C-E.

CONCLUSIONS

In this work, we elucidated an efficient and robust computational framework that combines AlphaFold,[51] all-atom, and systematic bottom-up coarse-grained molecular dynamics simulations for explicitly simulating the structures and conformational properties of $(CSt)_4$ megamolecule building blocks and their self-assembly into networks. The computational studies served to identify building blocks that had desired properties—orientation of the ligands and stiffness of the fusion proteins—and understand parameters that would optimize metal ion-dependent assembly. This approach allows for precise and efficient CG MD simulation, using only 0.0025% of the number of particles to represent the corresponding all-atom simulation with the capability to complete each replica within one day using half of a Cascade Lake computer node, and the computational results were largely in good agreement with experimental studies.

Specifically, we first used AA MD simulations to investigate the conformational and mechanical properties of the eleven candidate megamolecules. Based on the computational results, we selected CS HL1, CS HL2, CS FL, and SC HL2 fusion proteins for subsequent construction of CG models. Next, the CG models were constructed based on the reference AA trajectories to ensure they could capture the essential dynamics of the AA models.[45,49] Extensive CG simulations were performed under varying experimental conditions, including concentration, choice of metal ion, metal:protein ratio, and choice of megamolecule building block. This comprehensive analysis highlighted $(CSt\ HL2)_4$ as the most promising candidate, demonstrating the highest CN value, a lower threshold for

achieving CN = 2, and a favorable RDF curve. These computational findings then aligned well with the experimental results. Notably, most of these computational results are inaccessible using AA MD or high-resolution CG modeling, but required the implementation of low-resolution systematic bottom-up CG modeling, which significantly reduces the computational complexity but preserves the essential dynamics of megamolecules.[45,49]

Guided by these computational insights, we synthesized the $(CSt)_4$ megamolecules and report preliminary studies of their assembly into networks, and network precursors. This work lays the foundation for further work that will demonstrate and characterize long-range ordered networks. What is important here, though, is that the computational approach has identified building blocks and experimental conditions for attaining the network structures, and specifically has revealed optimized peptide linkers and conditions for metal-chelator interactions crucial for network formation. The key properties of the building blocks identified here for the macroscopic assembly can be leveraged in future studies. For example, by varying the connection valency through altering the central symmetry of megamolecule linkers, we can further refine the physical properties, such as viscosity, elasticity, and porosity, of these macroscopic networks. To achieve these goals, the same computational approach can be used to guide the experimental tuning of these network. These materials can be further customized to incorporate functional protein domains, through genetic fusion or encapsulation within the network to expand its functionalities.[2,3,6,70,71]

This work significantly reduces experimental effort and advances our understanding of the factors that influence megamolecule assembly, thereby facilitating the design of more effective strategies for the construction of structured and functionalized networks. It is especially noteworthy that the CG modeling successfully predicted how the macroscopic assembly depends

on a number of parameters, including: (1) Concentration dependence: Extensive CG simulations were conducted for four selected $(CSt)_4$ candidates, which can estimate their threshold concentrations required to form networks. (2) Metal ion affinity: We evaluated the tpy binding strength by varying the types of divalent metal ions. A high-affinity metal ion, such as Ni^{2+}, is favorable for forming stable assemblies. (3) Metal-to-protein Ratio: This parameter affects the availability of tpy ligands for bis-terpyridine complex formation. 1.5 eq of $NiCl_2$ yielded the largest average assembly population, although two equivalents were necessary to fully saturate the terpyridine ligand. (4) Peptide linker rigidity and directionality of the Cutinase and SnapTag domains are two other non-trivial factors that influence the optimal orientation and accessibility of the tpy ligands for intra- or inter-assembly. These key insights gained from MD simulations prior to synthesis also help to mitigate the cost associated with time-consuming trial and error experimental effort, thereby avoiding additional potential pitfalls. This design framework thus provides a versatile platform for the development of a variety of megamolecules with potential therapeutic and materials science applications.[10,22]

At the same time, we note that the bottom-up CG modeling described in this paper allows researchers to tailor the resolution of their biomolecular systems, i.e., by defining the number of CG sites per a biomolecule, based on the desired tradeoff between system size and simulation accuracy that one would like to achieve. The resolution can extend to over 30 amino acids per CG site, which has been demonstrated as being particularly useful for the modeling of complex biomolecular processes, such as the actin filament growth and virus capsid assembly.[36,72] By adopting a lower CG resolution, the simulation can become significantly more efficient since it only maintains the essential dynamics of the biomolecules and ignores non-crucial details, such as the conformational fluctuations in this study. The computationally guided framework employed

here can be transferable to other biomolecular assembly systems to aid in the design of biomaterials and in the study of complex biomolecular processes.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge.

The software is a part of the OpenMSCG package[59] (https://software.rcc.uchicago.edu/mscg/)

Additional experimental and computational methods and results (PDF)

AUTHOR INFORMATION

Corresponding Author

*Gregory A. Voth - *Department of Chemistry, Chicago Center for Theoretical Chemistry, James Franck Institute, and Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois 60637 United States*; https://orcid.org/0000-0002-3267-6748; Email: gavoth@uchicago.edu

*Milan Mrksich - *Departments of Chemistry and Biomedical Engineering, Northwestern University, Evanston, Illinois 60208, United States*; http://orcid.org/0000-0002-4964-796X; Email: milan.mrksich@northwestern.edu

Authors

Jiangbo Wu - *Department of Chemistry, Chicago Center for Theoretical Chemistry, James Franck Institute, and Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois 60637 United States*

Zhaoyi Gu - *Departments of Chemistry and Biomedical Engineering, Northwestern University, Evanston, Illinois 60208, United States*

Justin A. Modica - *Departments of Chemistry and Biomedical Engineering, Northwestern University, Evanston, Illinois 60208, United States*

Sijia Chen - *Department of Chemistry, Chicago Center for Theoretical Chemistry, James Franck Institute, and Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois 60637 United States*

Author Contributions

J.W. and Z.G. contributed equally to this work. J.W., Z.G., J.A.M., M.M. and G.A.V. designed the research. J.W. performed the simulations. Z.G. and J.A.M. performed the experiments. M.M. and G.A.V. directed the study. All authors analyzed the results and wrote the study.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported by the Army Research Office (W911NF1810200). This work also used the IMSERC facility and Keck facility at Northwestern University and the Biophysics Core at University of Illinois Chicago. The computational resources for this research were provided by the University of Chicago Research Computing Center (RCC) Midway supercomputer and the NIH-funded Beagle3 HPC cluster (Award Number S10OD028655).

REFERENCES

(1) Kim, J.; Grate, J. W. Single-Enzyme Nanoparticles Armored by a Nanometer-Scale Organic/Inorganic Network. *Nano Lett.* **2003**, *3*, 1219-1222.

(2) Peschke, T.; Bitterwolf, P.; Gallus, S.; Hu, Y.; Oelschlaeger, C.; Willenbacher, N.; Rabe, K. S.; Niemeyer, C. M. Self-Assembling All-Enzyme Hydrogels for Flow Biocatalysis. *Angew. Chem. Int. Ed.* **2018**, *57*, 17028-17032.

(3) Bitterwolf, P.; Gallus, S.; Peschke, T.; Mittmann, E.; Oelschlaeger, C.; Willenbacher, N.; Rabe, K. S.; Niemeyer, C. M. Valency Engineering of Monomeric Enzymes for Self-Assembling Biocatalytic Hydrogels. *Chem. Sci.* **2019**, *10*, 9752-9757.

(4) Fang, Y.; Han, E.; Zhang, X.-X.; Jiang, Y.; Lin, Y.; Shi, J.; Wu, J.; Meng, L.; Gao, X.; Griffin, P. J.; Xiao, X.; Tsai, H.-M.; Zhou, H.; Zuo, X.; Zhang, Q.; Chu, M.; Zhang, Q.; Gao, Y.; Roth, L. K.; Bleher, R.; Ma, Z.; Jiang, Z.; Yue, J.; Kao, C.-M.; Chen, C.-T.; Tokmakoff, A.; Wang, J.; Jaeger, H. M.; Tian, B. Dynamic and Programmable Cellular-Scale Granules Enable Tissue-like Materials. *Matter* **2020**, *2*, 948-964.

(5) Arpino, J. A. J.; Polizzi, K. M. A Modular Method for Directing Protein Self-Assembly. *ACS Synth. Biol.* **2020**, *9*, 993-1002.

(6) Heater, B. S.; Yang, Z.; Lee, M. M.; Chan, M. K. In Vivo Enzyme Entrapment in a Protein Crystal. *J. Am. Chem. Soc.* **2020**, *142*, 9879-9883.

(7) Galloway, J. M.; Bray, H. E. V.; Shoemark, D. K.; Hodgson, L. R.; Coombs, J.; Mantell, J. M.; Rose, R. S.; Ross, J. F.; Morris, C.; Harniman, R. L.; Wood, C. W.; Arthur, C.; Verkade, P.; Woolfson, D. N. De Novo Designed Peptide and Protein Hairpins Self-Assemble into Sheets and Nanoparticles. *Small* **2021**, *17*, e2100472.

(8) Lutolf, M. P.; Gilbert, P. M.; Blau, H. M. Designing Materials to Direct Stem-Cell Fate. *Nature* **2009**, *462*, 433-441.

(9) Zhang, Y. S.; Khademhosseini, A. Advances in Engineering Hydrogels. *Science* **2017**, *356*, eaaf3627.

(10) Metcalf, K. J.; Kimmel, B. R.; Sykora, D. J.; Modica, J. A.; Parker, K. A.; Berens, E.; Dai, R.; Dravid, V. P.; Werb, Z.; Mrksich, M. Synthetic Tuning of Domain Stoichiometry in Nanobody-Enzyme Megamolecules. *Bioconjug. Chem.* **2021**, *32*, 143-152.

(11) Huang, C.; Huang, J.; Zhu, S.; Tang, T.; Chen, Y.; Qian, F. Multivalent Nanobodies with Rationally Optimized Linker and Valency for Intravitreal VEGF Neutralization. *Chem. Eng. Sci.* **2023**, *270*, 118521.

(12) Pieters, B. J.; van Eldijk, M. B.; Nolte, R. J.; Mecinovic, J. Natural Supramolecular Protein Assemblies. *Chem. Soc. Rev.* **2016**, *45*, 24-39.

(13) Suzuki, Y.; Cardone, G.; Restrepo, D.; Zavattieri, P. D.; Baker, T. S.; Tezcan, F. A. Self-Assembly of Coherently Dynamic, Auxetic, Two-Dimensional Protein Crystals. *Nature* **2016**, *533*, 369-373.

(14) Yang, M.; Song, W. J. Diverse Protein Assembly Driven by Metal and Chelating Amino Acids with Selectivity and Tunability. *Nat. Commun.* **2019**, *10*, 5545.

(15) Li, Z.; Chen, S.; Gao, C.; Yang, Z.; Shih, K.-C.; Kochovski, Z.; Yang, G.; Gou, L.; Nieh, M.-P.; Jiang, M.; Zhang, L.; Chen, G. Chemically Controlled Helical Polymorphism in Protein Tubes by Selective Modulation of Supramolecular Interactions. *J. Am. Chem. Soc.* **2019**, *141*, 19448-19457.

(16) Mout, R.; Bretherton, R. C.; Decarreau, J.; Lee, S.; Gregorio, N.; Edman, N. I.; Ahlrichs, M.; Hsia, Y.; Sahtoe, D. D.; Ueda, G.; Sharma, A.; Schulman, R.; Deforest, C. A.; Baker, D. De Novo Design of Modular Protein Hydrogels With Programmable Intra- And Extracellular Viscoelasticity. *Proc. Natl. Acad. Sci. U. S. A.* **2024**, *121*.

(17) Manea, F.; Garda, V. G.; Rad, B.; Ajo-Franklin, C. M. Programmable Assembly of 2D Crystalline Protein Arrays into Covalently Stacked 3D Bionanomaterials. *Biotechnol. Bioeng.* **2020**, *117*, 912-923.

(18) Liu, R.; Kochovski, Z.; Li, L.; Yin, Y. W.; Yang, J.; Yang, G.; Tao, G.; Xu, A.; Zhang, E.; Ding, H. M.; Lu, Y.; Chen, G.; Jiang, M. Fabrication of Pascal-triangle Lattice of Proteins by Inducing Ligand Strategy. *Angew. Chem. Int. Ed.* **2020**, *59*, 9617-9623.

(19) Modica, J. A.; Skarpathiotis, S.; Mrksich, M. Modular Assembly of Protein Building Blocks to Create Precisely Defined Megamolecules. *Chembiochem* **2012**, *13*, 2331-2334.

(20) Modica, J. A.; Lin, Y.; Mrksich, M. Synthesis of Cyclic Megamolecules. *J. Am. Chem. Soc.* **2018**, *140*, 6391-6399.

(21) Kimmel, B. R.; Modica, J. A.; Parker, K.; Dravid, V.; Mrksich, M. Solid-Phase Synthesis of Megamolecules. *J. Am. Chem. Soc.* **2020**, *142*, 4534-4538.

(22) Modica, J. A.; Iderzorig, T.; Mrksich, M. Design and Synthesis of Megamolecule Mimics of a Therapeutic Antibody. *J. Am. Chem. Soc.* **2020**, *142*, 13657-13661.

(23) Zhou, S.; He, P.; Dhindwal, S.; Grum-Tokars, V. L.; Li, Y.; Parker, K.; Modica, J. A.; Bleher, R.; Dos Reis, R.; Zuchniarz, J.; Dravid, V. P.; Voth, G. A.; Roux, B.; Mrksich, M.

Synthesis, Characterization, and Simulation of Four-Armed Megamolecules. *Biomacromolecules* **2021**, *22*, 2363-2372.

(24) Kimmel, B. R.; Mrksich, M. Development of an Enzyme-Inhibitor Reaction Using Cellular Retinoic Acid Binding Protein II for One-Pot Megamolecule Assembly. *Chemistry* **2021**, *27*, 17843-17848.

(25) Juillerat, A.; Gronemeyer, T.; Keppler, A.; Gendreizig, S.; Pick, H.; Vogel, H.; Johnsson, K. Directed Evolution of O6-Alkylguanine-DNA Alkyltransferase for Efficient Labeling of Fusion Proteins with Small Molecules In Vivo. *Chem. Biol.* **2003**, *10*, 313-317.

(26) Hoehnel, S.; Lutolf, M. P. Capturing Cell–Cell Interactions via SNAP-tag and CLIP-tag Technology. *Bioconjugate Chem.* **2015**, *26*, 1678-1686.

(27) Kwon, Y.; Han, Z.; Karatan, E.; Mrksich, M.; Kay, B. K. Antibody Arrays Prepared by Cutinase-Mediated Immobilization on Self-Assembled Monolayers. *Anal. Chem.* **2004**, *76*, 5713-5720.

(28) Förster, C.; Dorn, M.; Reuter, T.; Otto, S.; Davarci, G.; Reich, T.; Carrella, L.; Rentschler, E.; Heinze, K. Ddpd as Expanded Terpyridine: Dramatic Effects of Symmetry and Electronic Properties in First Row Transition Metal Complexes. *Inorganics* **2018**, *6*.

(29) Kakinoki, S.; Hirano, Y.; Oka, M. On the Stability of Polyproline-I and II Structures of Proline Oligopeptides. *Polym. Bull.* **2005**, *53*, 109-115.

(30) Chen, X.; Zaro, J. L.; Shen, W.-C. Fusion Protein Linkers: Property, Design and Functionality. *Adv. Drug Deliv. Rev.* **2013**, *65*, 1357-1369.

(31) Haga, T.; Hirakawa, H.; Nagamune, T. Fine Tuning of Spatial Arrangement of Enzymes in a PCNA-Mediated Multienzyme Complex Using a Rigid Poly-L-Proline Linker. *PLOS ONE* **2013**, *8*, e75114.

(32) Komor, A. C.; Kim, Y. B.; Packer, M. S.; Zuris, J. A.; Liu, D. R. Programmable Editing of a Target Base in Genomic DNA without Double-Stranded DNA Cleavage. *Nature* **2016**, *533*, 420-424.

(33) Narayanan, R. P.; Procyk, J.; Nandi, P.; Prasad, A.; Xu, Y.; Poppleton, E.; Williams, D.; Zhang, F.; Yan, H.; Chiu, P. L.; Stephanopoulos, N.; Sulc, P. Coarse-Grained Simulations for the Characterization and Optimization of Hybrid Protein-DNA Nanostructures. *ACS Nano* **2022**, *16*, 14086-14096.

(34) Su, Z.; Wu, Y. Dissecting the General Mechanisms of Protein Cage Self‐Assembly by Coarse‐Grained Simulations. *Protein Sci.* **2023**, *32*.

(35) Pollard, T. D.; Cooper, J. A. Actin, a Central Player in Cell Shape and Movement. *Science* **2009**, *326*, 1208-1212.

(36) Mani, S.; Katkar, H. H.; Voth, G. A. Compressive and Tensile Deformations Alter ATP Hydrolysis and Phosphate Release Rates in Actin Filaments. *J. Chem. Theory Comput.* **2021**, *17*, 1900-1913.

(37) Pak, A. J.; Gupta, M.; Yeager, M.; Voth, G. A. Inositol Hexakisphosphate (IP6) Accelerates Immature HIV-1 Gag Protein Assembly toward Kinetically Trapped Morphologies. *J. Am. Chem. Soc.* **2022**, *144*, 10417-10428.

(38) Gupta, M.; Pak, A. J.; Voth, G. A. Critical Mechanistic Features of HIV-1 Viral Capsid Assembly. *Sci. Adv.* **2023**, *9*, eadd7434.

(39) McCullagh, M.; Saunders, M. G.; Voth, G. A. Unraveling the Mystery of ATP Hydrolysis in Actin Filaments. *J. Am. Chem. Soc.* **2014**, *136*, 13053-13058.

(40) Wang, Y.; Wu, J.; Zsolnay, V.; Pollard, T. D.; Voth, G. A. Mechanism of Phosphate Release from Actin Filaments. *Proc. Natl. Acad. Sci. U.S.A.* **2024**, *121*, e2408156121.

(41) Yue, Z.; Wu, J.; Teng, D.; Wang, Z.; Voth, G. A. Activation of the Influenza B M2 Proton Channel (BM2). *bioRxiv* **2024**, 2024.2007.2026.605324.

(42) Izvekov, S.; Voth, G. A. A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B* **2005**, *109*, 2469-2473.

(43) Noid, W. G.; Chu, J. W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. The Multiscale Coarse-Graining Method. I. A Rigorous Bridge Between Atomistic and Coarse-Grained Models. *J. Chem. Phys.* **2008**, *128*, 244114.

(44) Jin, J.; Pak, A. J.; Durumeric, A. E. P.; Loose, T. D.; Voth, G. A. Bottom-up Coarse-Graining: Principles and Perspectives. *J. Chem. Theory Comput.* **2022**, *18*, 5759-5791.

(45) Zhang, Z.; Lu, L.; Noid, W. G.; Krishna, V.; Pfaendtner, J.; Voth, G. A. A Systematic Methodology for Defining Coarse-Grained Sites in Large Biomolecules. *Biophys. J.* **2008**, *95*, 5073-5083.

(46) Saunders, M. G.; Voth, G. A. Coarse-Graining Methods for Computational Biology. *Annu. Rev. Biophys.* **2013**, *42*, 73-93.

(47) Noid, W. G. Perspective: Coarse-Grained Models for Biomolecular Systems. *J. Chem. Phys.* **2013**, *139*, 090901.

(48) Noid, W. G. Perspective: Advances, Challenges, and Insight for Predictive Coarse-Grained Models. *J. Phys. Chem. B* **2023**, *127*, 4174-4207.

(49) Wu, J.; Xue, W.; Voth, G. A. K-Means Clustering Coarse-Graining (KMC-CG): A Next Generation Methodology for Determining Optimal Coarse-Grained Mappings of Large Biomolecules. *J. Chem. Theory Comput.* **2023**, *19*, 8987-8997.

(50) Teng, D.; Mironenko, A. V.; Voth, G. A. QM/CG-MM: Systematic Embedding of Quantum Mechanical Systems in a Coarse-Grained Environment with Accurate Electrostatics. *J. Phys. Chem. A* **2024**, *128*, 6061-6071.

(51) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with Alphafold. *Nature* **2021**, *596*, 583-589.

(52) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., Jr. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017**, *14*, 71-73.

(53) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

(54) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701-1718.

(55) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187-199.

(56) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011-1021.

(57) Wieczor, M.; Czub, J. Gromologist: a Gromacs-Oriented Utility Library for Structure and Topology Manipulation. *ChemRxiv* **2022**.

(58) Lyman, E.; Pfaendtner, J.; Voth, G. A. Systematic Multiscale Parameterization of Heterogeneous Elastic Network Models of Proteins. *Biophys. J.* **2008**, *95*, 4183-4192.

(59) Peng, Y.; Pak, A. J.; Durumeric, A. E. P.; Sahrmann, P. G.; Mani, S.; Jin, J.; Loose, T. D.; Beiter, J.; Voth, G. A. OpenMSCG: A Software Tool for Bottom-Up Coarse-Graining. *J. Phys. Chem. B* **2023**, *127*, 8537-8550.

(60) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.;

Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian 16 Rev. C.01. **2016**.

(61) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comput. Chem.* **2009**, *30*, 2157-2164.

(62) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1-19.

(63) Vanden-Eijnden, E.; Ciccotti, G. Second-Order Integrators for Langevin Equations With Holonomic Constraints. *Chem. Phys. Lett.* **2006**, *429*, 310-316.

(64) Alessi, M. L.; Norman, A. I.; Knowlton, S. E.; Ho, D. L.; Greer, S. C. Helical and Coil Conformations of Poly(thylene glycol) in Isobutyric Acid and Water. *Macromolecules* **2005**, *38*, 9333-9340.

(65) Banerjee, S. S.; Aher, N.; Patil, R.; Khandare, J. Poly(ethylene glycol)-Prodrug Conjugates: Concept, Design, and Applications. *J. Drug Deliv.* **2012**, *2012*, 103973.

(66) Schubert, U. S.; Eschbaumer, C. Macromolecules Containing Bipyridine and Terpyridine Metal Complexes: Towards Metallosupramolecular Polymers. *Angew. Chem. Int. Ed.* **2002**, *41*, 2892-2926.

(67) Hogg, R.; Wilkins, R. G. 57. Exchange Studies of Certain Chelate Compounds of the Transitional Metals. Part VIII. 2,2′,2″-Terpyridine Complexes. *J. Chem. Soc.* **1962**, 341-350.

(68) Holyer, R. H.; Hubbard, C. D.; Kettle, S. F. A.; Wilkins, R. G. The Kinetics of Replacement Reactions of Complexes of the Transition Metals with 1,10-Phenanthroline and 2,2'-Bipyridine. *Inorg. Chem.* **1965**, *4*, 929-935.

(69) Sonn-Segev, A.; Belacic, K.; Bodrug, T.; Young, G.; VanderLinden, R. T.; Schulman, B. A.; Schimpf, J.; Friedrich, T.; Dip, P. V.; Schwartz, T. U.; Bauer, B.; Peters, J.-M.; Struwe, W. B.; Benesch, J. L. P.; Brown, N. G.; Haselbach, D.; Kukura, P. Quantifying the Heterogeneity of Macromolecular Machines by Mass Photometry. *Nat. Commun.* **2020**, *11*, 1772.

(70) Gao, X.; Fang, J.; Xue, B.; Fu, L.; Li, H. Engineering Protein Hydrogels Using SpyCatcher-SpyTag Chemistry. *Biomacromolecules* **2016**, *17*, 2812-2819.

(71) Mallin, H.; Ward, T. R. Streptavidin–Enzyme Linked Aggregates for the One-Step Assembly and Purification of Enzyme Cascades. *ChemCatChem* **2018**, *10*, 2810-2816.

(72) Grime, J. M. A.; Dama, J. F.; Ganser-Pornillos, B. K.; Woodward, C. L.; Jensen, G. J.; Yeager, M.; Voth, G. A. Coarse-Grained Simulation Reveals Key Features of HIV-1 Capsid Self-Assembly. *Nat. Commun.* **2016**, *7*, 11568.

TOC Graphic