pubs.acs.org/Biomac

Article

Synthesis, Characterization, and Simulation of Four-Armed Megamolecules

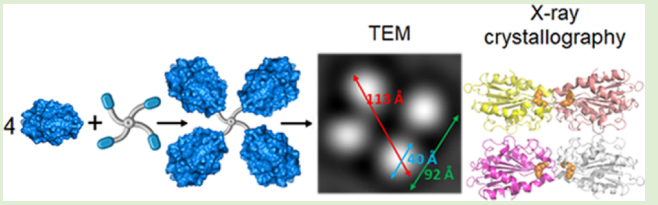
Shengwang Zhou, Peng He, Sonali Dhindwal, Valerie L. Grum-Tokars, Ying Li, Kelly Parker, Justin A. Modica, Reiner Bleher, Roberto dos Reis, Joshua Zuchniarz, Vinayak P. Dravid, Gregory A. Voth, Benoît Roux, and Milan Mrksich*

Cite This: https://doi.org/10.1021/acs.biomac.1c00118

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: This paper describes the synthesis, characterization, and modeling of a series of molecules having four protein domains attached to a central core. The molecules were assembled with the "megamolecule" strategy, wherein enzymes react with their covalent inhibitors that are substituted on a linker. Three linkers were synthesized, where each had four oligo(ethylene glycol)-based arms terminated in a *para*-nitrophenyl phosphonate group that is a covalent inhibitor for cutinase. This enzyme is a serine hydrolase and reacts efficiently with the phosphonate to give a new ester linkage at the Ser-120 residue in the active site of the enzyme. Negative-stain transmission electron microscopy (TEM) images confirmed the architecture of the four-armed megamolecules. These cutinase tetramers were also characterized by X-ray crystallography, which confirmed the active-site serine-phosphonate linkage by electron-density maps. Molecular dynamics simulations of the tetracutinase megamolecules using three different force field setups were performed and compared with the TEM observations. Using the Amberff99SB-disp + pH7 force field, the two-dimensional projection distances of the megamolecules were found to agree with the measured dimensions from TEM. The study described here, which combines high-resolution characterization with molecular dynamics simulations, will lead to a comprehensive understanding of the molecular structures and dynamics for this new class of molecules.

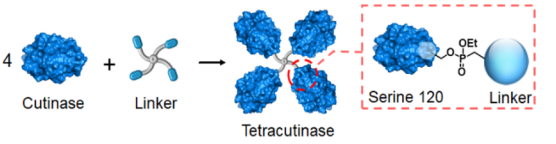


INTRODUCTION

We recently described a strategy to synthesize very large molecules by joining fusion proteins with multifunctional linkers.[1−5] The reactions occur between a covalent inhibitor on the linker and its target enzyme domain and proceed in high yield and with essentially no side products. We have demonstrated this "megamolecule" approach for the synthesis of linear[1] and cyclic[2] molecules, to organize fluorescent proteins for studies of energy transfer,[3] and for the preparation of antibody mimics.[4,5] However, the characterization of these structures and understanding their conformations and dynamics are very challenging. Here, we report the synthesis and comprehensive characterization of a four-armed megamolecule with transmission electron microscopy (TEM),[6,7] X-ray crystallography,[8,9] and simulation[10,11] and demonstrate how this comprehensive approach contributes to an understanding of the structures of this new class of molecules.[12−16]

In this work, we use the reaction of a four-armed linker with a cutinase enzyme to generate a tetracutinase product (Scheme 1). The linker is based on a tetra-aza scaffold having four acetic acid arms, each of which is linked to a *para*-nitrophenylphosphonate (referred to as the DOTA-(amide)$_4$-(EG)$_n$-tetraphosphonate or DOTP).[17−19] The phosphonate group is an irreversible inhibitor of the fungal esterase cutinase and reacts with an active-site serine residue (Ser 120) to give a phosphonate linkage. We further use three versions of this linker, where the arms are increased in length by incorporating additional ethylene glycol units. These three linkers were chosen to evaluate steric interactions that might prevent the reaction with all four enzyme domains—since shorter linkers might not accommodate the attachment of four cutinase domains—and also to determine if they would affect the

Scheme 1. Reaction of Four Equivalents of the Enzyme Cutinase with a Four-Armed Linker Presenting a Covalent Inhibitor for Cutinase Results in a Tetracutinase Product



Received: January 28, 2021
Revised: April 16, 2021

overall dimensions and dynamics of the tetracutinase products.[20−22]

MATERIALS AND METHODS

Materials. All chemical reagents were purchased from Sigma-Aldrich and Alfa Aesar, unless noted. Commercial materials were directly used without further purification. BenchMark Protein Ladder was purchased from Thermo Fisher Scientific. 4−15% Mini-Protean TGX precast protein gels and 10× Tris/glycine/SDS running buffer were purchased from Bio-Rad.

Plasmid Construction. Expression plasmids were constructed using the pET21b(+) backbone in *Escherichia coli* DH5α (NEB). Cloning reactions were performed using the Golden Gate method with BsaI and T4 ligase. Polymerase chain reactions were carried out using the Q5 master mix (NEB).

Protein Expression and Purification. Proteins were expressed in the Shuffle T7 Express *E. coli* cell line (NEB). A stab of stocked cells was incubated in a 2xYT medium (5 mL) containing carbenicillin (100 μg/mL) at 30 °C with orbital shaking. After overnight incubation, the culture was diluted to 500 mL with a 2xYT medium containing the same concentration of antibiotics in a baffled flask. The culture was then shaken at 30 °C until the OD_{600} value of the cell suspension reached 0.8−1.0. To induce protein expression, isopropyl β-D-1-thiogalactopyranoside was added to a final concentration of 1 mM. The culture was then shaken in an Innova 44R (New Brunswick Scientific) incubator at 20 °C for 16−18 h. The cell pellet was collected by centrifugation and stored at −20 °C. For lysis, the pellet was resuspended in 1× phosphate-buffered saline (PBS) buffer and sonicated on ice for 3 min (5 s on, 10 s off). Cell debris was removed by centrifugation, and the resulting cell lysate was incubated with 5 mL of HisPur cobalt resin (Thermo Fisher Scientific) in a 50 mL Kontes Flex column (Kimbal Kontes Glassware). The column was rotated at 4 °C for 1 h and then washed with 1× PBS buffer. The bound protein was then eluted with PBS buffer containing 150 mM imidazole and concentrated. Size exclusion chromatography (SEC) was then carried out on a Hi-Load 16/600 Superdex 200pg column using an Akta FPLC (GE Healthcare). The purified fractions were identified by SDS-PAGE pooled and concentrated. The concentration of protein was calculated based on the absorbance at 280 nm using the molar absorption coefficient predicted with the ExPASy ProtParam tool.

Linker Synthesis. The synthesis process of three DOTP derivatives is described in detail in the Supporting Information. ^1H NMR spectra were recorded on a Bruker AVANCE III HD 500 MHZ system (TXO 5 mm Prodigy probe w/Z-Gradient). Mass spectrometry (MS) of the linkers was performed on a Bruker amaZon SL LC/MS mass spectrometer using electrospray ionization (ESI) with direct injection and a 5800 MALDI-TOF/TOF mass spectrometer (AB Sciex, Framingham, MA) using a positive-ion reflector mode. Stock solutions of the linkers were prepared in dimethyl sulfoxide at a final concentration of 100 mM.

Synthesis and Purification of Tetracutinases. All protein assembly reactions were performed in 1× PBS buffer after incubation with different stoichiometric ratios of protein—linker for 3 h at room temperature. After the reaction, solutions were concentrated by an Amicon Ultra centrifugal filter unit with 10 kDa cutoff. Purification was carried out as above on a Hi-Load 16/60 Superdex 200 column using a mobile phase of 1× PBS containing 0.02 wt % NaN_3. Pure fractions were identified by SDS-PAGE on a 4−15% Mini-Protean TGX precast protein gel. Electrophoresis was run at 200 V for 30 min, and the gel was stained by Coomassie-R-250. A NanoDrop 2000 UV−Vis spectrophotometer (Thermo Fisher) was used to quantify the concentration of purified protein.

Protein Mass Spectrometry. Liquid chromatography−MS (LC−MS) protein analysis was carried out on an Agilent 1200 series HPLC system with an Agilent 6210A time-of-flight mass spectrometer. Protein samples with a concentration of 1 μM were prepared for LC−MS by dilution in water. Separation was carried out on a C18 trap column (Waters) with 10 μL of sample for each injection. A gradient elution method using 5 to 95% acetonitrile in water with 0.1% formic acid and a flow rate of 0.25 mL/min was used. Data analysis was performed with Agilent Mass Hunter Qualitative Analysis B.04.00.

Negative-Staining Procedure. Both the samples, Cut_4-L and Cut_4-M, were diluted to a concentration of 500 and 200 nM, respectively, using HEPES buffer (20 mM, pH 7.4, with 150 mM NaCl). To image the samples by TEM, grids with a continuous carbon film supported on 300-mesh copper (Ted Pella, CA) were used. For each sample of Cut_4-L and Cut_4-M, grids were first plasma cleaned for 20 s, followed by the application of 3.5 μL of sample. The sample was absorbed on to the grid for 60 s and blotted using a filter paper to leave a thin film of the sample on the grid. The grid was then water washed twice by floating the grid with the sample side on a drop of distilled water for 2 s and blotting with a filter paper. This was followed by staining the sample by floating the grid on a drop of 0.75% (w/v) uranyl formate for 2 and 20 s, followed by blotting the grid each time with a filter paper. Finally, the grid was left to air-dry completely.

Electron Microscopy. Samples were imaged using an aberration-corrected JEOL ARM200CF microscope operated at 200 kV, and images were acquired using a Gatan OneView CMOS camera (Gatan, CA). Around 100 micrographs were collected for each sample, Cut_4-L and Cut_4-M with a pixel size of 1.8 Å. Images were collected with an exposure time of 1 s using the smallest size condenser aperture (10 μm) to obtain the most coherent beam and defocus (C1) was varied from 1.5 to 2.5 μm. The contrast transfer function parameters were estimated using CTFFIND-4.1.13,[23] and the micrographs were screened for drift and astigmatism. Two-dimensional (2D) classification was performed using Relion 3.0.7.[24] For this, particles were picked using the Autopick procedure by the Laplacian-of-Gaussian filter[25] in Relion and later screened manually to give 2989 particles from 96 selected micrographs for Cut_4-L and 6452 particles from 87 selected micrographs for Cut_4-M. These particles were extracted and subjected to reference-free 2D classification.

Model Generation of the Three-Dimensional Structure. GaussianView[26] was used to generate the three-dimensional (3D) structure of all three linkers. The energy minimization of the linker was performed with quantum mechanism simulations by Gaussian09[27] to acquire an extended regular tetrahedron confirmation. Four cutinases (PDB ID: 1CEX) were manually connected to each linker arm at the catalytically active Ser120. The linker was modeled in its tetrahedron conformation to best avoid steric clash among the four cutinases.

Amberff14SB + TIP3P Force Field.[28,29] The DOTP linkers were constructed as an unnatural amino acid. The partial charges were obtained from the Ambertools18/antechamber AM1-bcc charge, and the vdw and the bonded parameters are constructed from Amber14SB and AMBERgaff.[30,31] The topology and parameters of each cutinase were generated from AMBER14SB. Then, the system was solvated in a TIP3P water box with an additional 9.0 Å shell on each side. Cl^- ions were added to neutralize the net charge of the whole system. To generate the model under pH 5.0 conditions, Propka3.0 was used to obtain the protonation state of each residue and the organic linker.[32,33]

Amberff99SB-Disp Force Field.[34] With the new force field parameters obtained from DE Shaw research, the exact same procedure aforementioned was followed to construct the Amber input files.

Simulation Details. All simulations were conducted by a GPU version of AMBER18 PMEMD using the University of Chicago's Research Computing Center (RCC) Midway supercomputer.[35,36] Simulation details of each system are also listed in Table S3. For each simulation, a 500-step minimization, 1 ns restrained *NVT* equilibrium, and 1 ns *NPT* equilibrium were performed before the production runs at 300 K. ∼200 ns × 4 replica simulations were performed for each tetracutinase system, and a 200 ns simulation was performed for each tricutinase system.

Conformation Analysis of the MD Simulations. Cpptraj and mdtraj were used to calculate all the MD properties presented.[37,38]

https://doi.org/10.1021/acs.biomac.1c00118
Biomacromolecules XXXX, XXX, XXX−XXX

The first 100 ns with Amberff14SB + tip3p at pH7 and 50 ns for the other simulations were excluded as equilibration, and the trajectory averages of all characteristics for each model are reported in Table S3. To compare the molecule size in simulation with the TEM image straightforwardly, we performed a projection of the heavy atoms of each snapshot in the trajectory from the 3D space to a few randomly defined 2D surfaces. Then, all the heavy atoms were mapped into a 40 × 40 grid with a cell size of 4 Å × 4 Å. Then, the center-to-center distance between the furthest-separated nonempty cells was calculated. This process was repeated 20 times for each frame in the trajectory and the largest distance among the 20 was taken as the molecular size. To quantify whether a snapshot in the trajectory was a regular tetrahedron or a planar square, we measured all 12 improper dihedral angles defined by four Ser120 CA atoms. As a reference, the dihedral angle of the methane molecule with a regular tetrahedron conformation is 1.23 rad and the dihedral angle of a planar square is 0 or π. To explain why the S linker is not able to accommodate a tetracutinase, we measured the distribution of minimum distances between the free linker tail (phosphate atom) and the other three cutinase proteins (CA atoms) along the tricutinase simulation trajectory as an extension capability, and the distribution of minimum distances between the one linker tail (phosphate atom) and the other three cutinases (CA atoms) in the tetracutinase simulation. Supposing the bumpy protein surface, using the real minimum would be unrealistic. Thus, for each snapshot, the average value of the 15 minimum distances was taken. The most extended simulation (Amberff99SB-disp) was used to calculate the extension capability and the most compact simulation (Amberff14SB/TIP3P) was used for the minimum distance required.

Crystallization and Structure Determination of a Tetrameric Megamolecule. For data collection, crystals of tetracutinase were flash frozen with liquid nitrogen using 10% glycerol in the mother liquor as a cryoprotectant. X-ray diffraction data were collected at 100 K at the Life Sciences Collaborative Access Team (LS-CAT) beamlines at the Advanced Photon Source (APS), Argonne National Laboratory, using a Rayonix CCD detector (Table S2). Data processing was conducted by XDS[39] and AIMLESS.[40] The crystals showed anisotropic diffraction, extending to 1.27 Å (Cut_4-L) and 1.36 Å (Cut_4-M) in the best direction, and the data were further processed with the Staraniso server[41] to produce an anisotropic data set. The crystal structures of tetracutinase were solved by molecular replacement using a cutinase molecule[42] (PDB ID 2CUT) as a search model. All iterative rounds of model building and refinement were performed with Coot[43] and Refmac.[44] For Cut_4-L crystals, the final model consists of one cutinase monomer (residues 17−212), one PEG linker with 8 atoms visible around the covalent bond formed with serine 120, and one linker region corresponding to the PEG region and 185 water molecules and has a R/R_{free} = 0.175/0.192−1.267 Å for the anisotropic data set. The model has excellent geometry, with a root-mean-square deviation of 0.014 Å and 1.97° for bonds and angles, respectively, and 98.97% in the favored region of the Ramachandran plot. For Cut_4-M crystals, the final model consists of one cutinase monomer (residues 16−212) and one M linker with only 8 atoms visible around the covalent bond formed with serine 120 and 202 water molecules and has a R/R_{free} = 0.186/0.205−1.359 Å for the anisotropic data set. The model has excellent geometry, with a root-mean-square deviation of 0.012 Å and 1.86° for bonds and angles, respectively, and 98.97% in the favored region of the Ramachandran plot.

RESULTS AND DISCUSSION

Design and Synthesis of Linkers. The three linkers we use are shown in Figure 1, where each is based on a 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA) core, which is symmetrically substituted with four oligo(ethylene glycol)-based arms that terminate in a phosphonate group. The first linker has the longest arm and includes three ethylene glycol groups (linker **1**); the second has two fewer ethylene glycol groups (linker **2**); and the third also has one ethylene

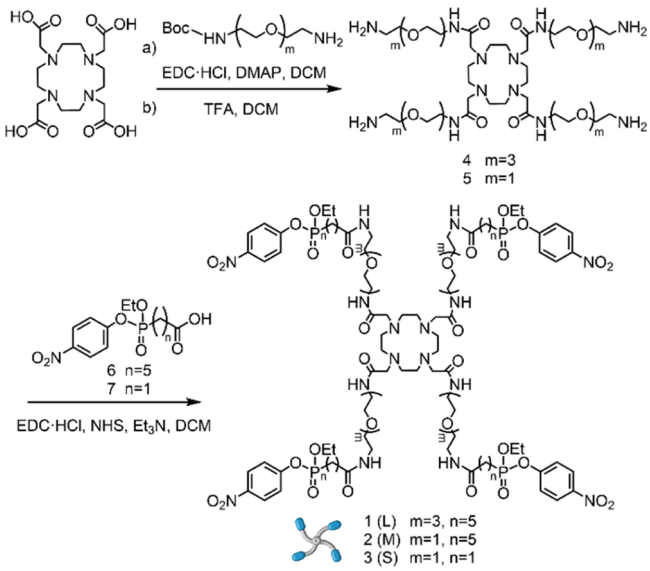


Figure 1. Synthetic route for preparing linker molecules with long, medium, and short arms (L, M, and S).

glycol group but a shorter alkyl chain that tethers the phosphonate group (linker **3**). The synthesis of these three linkers was straightforward and started with the activation of the four acids of DOTA and treatment with an excess of BocNH-PEG_m-$CH_2CH_2NH_2$ to give the tetrasubstituted core and finally deprotection with TFA to reveal the four amines. The tetraamine was then treated with 4-nitrophenyl-ethyl-(carboxyalkyl)phosphonate after activation with EDC/NHS to give the final products **1**−**3** (see Supporting Information for details).

Synthesis and Characterization of Four-Armed Megamolecules. Cutinase was expressed in *E. coli*, as described previously, as a C-terminal 6xHis-tag fusion.[45] The expressed cutinase has a molecular weight of 23,380 Da and an approximate size of 45 × 30 × 30 $Å^3$ in a compact domain.[46] The cutinase monomer was purified using SEC and showed one clear band of 23 kDa on SDS-PAGE (Figure 2A). The yield of cutinase was approximately 5 mg/L after purification by cobalt-immobilized metal affinity chromatography and SEC (Figure 2B).

Cutinase is irreversibly inhibited by *para*-nitrophenyl phosphonate groups through esterification of the Ser-120 residue in the active site.[45] Hence, we treated linker **1** (5 μM) with an excess of cutinase (50 μM) in a PBS solution for 3 h and analyzed the reaction products with SDS-PAGE (Figures 2A and S1). For the reaction of cutinase and linker **1**, we analyzed the crude reaction mixture and observed the tetrameric cutinase product (Cut_4-L, for the long linker) as a clear band at the expected mass of 95 kDa. The gel also revealed bands for free cutinase and a small amount of dimeric and trimeric cutinase intermediates. The Cut_4-L product was easily purified by SEC (Figure 2B) and showed a single band in the gel (Figure 2A). Furthermore, ESI-MS analysis of the product gave a mass of 95,375 Da, in excellent agreement with the calculated mass (Figure 2C, Table S1). We then repeated this reaction using the medium-length linker **2** where SDS-PAGE of the crude mixture again showed the desired tetracutinase product (Cut_4-M, Figures 2A and S2) together with small amounts of the intermediates. Again, the product could be readily purified with SEC and gave a mass that was

https://doi.org/10.1021/acs.biomac.1c00118
Biomacromolecules XXXX, XXX, XXX−XXX

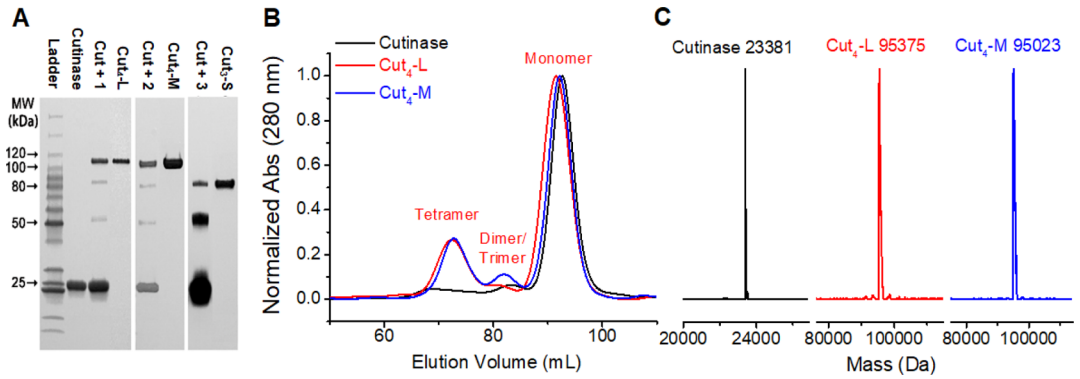


Figure 2. Synthesis, purification, and characterization of four-armed megamolecules. (A) SDS-PAGE analysis of cutinase monomer, reaction products with the linker, and purified tetracutinases. Cutinase was incubated with linkers **1**, **2**, and **3** for 3 h, respectively, at cutinase concentrations of 50, 50, and 200 μM, respectively. The cutinase/linker molar ratio was set to 10/1 for each reaction. (B) Size-exclusion chromatograms of cutinase, Cut_4-L, and Cut_4-M. (C) Deconvoluted ESI-MS spectra of cutinase, Cut_4-L, and Cut_4-M.

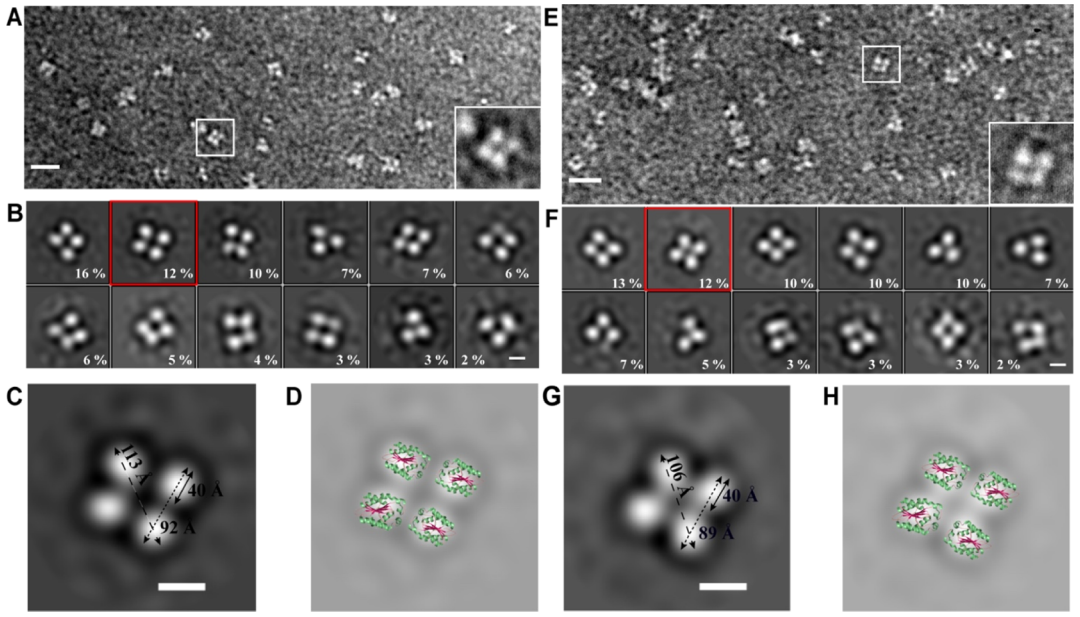


Figure 3. Negative stain TEM exhibits the architecture of the four-armed megamolecules. (A) Representative TEM image of the negatively stained Cut_4-L molecule. Scale bar represents 20 nm. (B) Reference-free 2D class averages of all the particles picked from the micrographs of Cut_4-L. Percentage of particles in each class is mentioned. Scale bar represents 50 Å. (C) Enlarged view of one of the classes [red box in (B), fully extended top view] of the Cut_4-L molecule. Scale bar represents 50 Å. (D) Coordinates of four Cutinase molecules (PDB ID: 1CEX) overlaid on the 2D average from (C) to represent the architecture of the Cut_4-L molecule. (E−H) The same TEM characterization and analysis corresponding to (A−D) for Cut_4-M, respectively. Insets in (A,E) show representative orientations of Cut_4-L and Cut_4-M in the raw TEM image.

consistent with its expected mass. However, when we repeated the reaction using the shortest linker **3**, the gel revealed that no tetracutinase product was detected. We repeated this reaction using four-fold higher concentrations of the reactants and still could not detect any final product (Figures 2A and S3). Instead, we observed only di- and tricutinase intermediates. Hence, we reason that linker **3**, for steric reasons, cannot accommodate the attachment of four enzyme domains, and in the work that follows, we only address Cut_4-L and Cut_4-M.

Electron Microscopy Reveals the Overall Architecture of Cut_4-L and Cut_4-M. We next used TEM to directly image the two tetracutinase products, which revealed the overall architecture of Cut_4-L and Cut_4-M. The proteins were applied to a grid and stained with uranyl formate at a pH of ∼5 to enhance the contrast and to stabilize the tetracutinases by rapid fixation.[47,48] TEM images for Cut_4-L and Cut_4-M reveal the four-lobed structures throughout the imaging area, where each of these lobes corresponds to one cutinase domain (Figures 3, S4, and S5). These images are consistent with the expected four-armed structure of the megamolecules and demonstrate that the resolution level necessary for their structural analysis can be easily achieved through the negative-stain imaging used in this study.[49]

In order to improve the signal-to-noise ratio of the TEM images, we obtained reference-free class averages using the method described in Figures S6 and S7. 2D classification of Cut_4-L and Cut_4-M both gave 12 unique orientations of the tetrameric structure (Figure 3B,F) with sufficient signal-to-noise; data collection was unbiased and all classes obtained are shown in Figure 3. The 2D class with a fully extended top view was selected for measuring dimensions of the Cut_4-L and Cut_4-M molecules, as this most closely resembles the orientation analyzed by molecular dynamics (MD) simulations. The size of Cut_4-L was determined to be ∼113 Å diagonally and ∼92 Å

https://doi.org/10.1021/acs.biomac.1c00118
Biomacromolecules XXXX, XXX, XXX−XXX

laterally (Figure 3C). Similarly, the size of Cut_4-M was ~106 Å diagonally and ~89 Å laterally (Figure 3G). For both analogues, the size of a single cutinase domain was measured to be ~40 Å, which matched the size observed in a prior X-ray crystallographic study (PDB ID: 1CEX) and with overlays of these structures shown in Figure 3D,H. These measurements include an error range of 4 Å due to the grain size of the negative stain. Because both gel electrophoresis and ESI-MS (Figure 2A,C) confirmed that the sample was homogeneous, the different projections in the 2D classes are assigned to different orientations of the megamolecules (and not contamination by intermediates in the synthesis). Although negative staining may not accurately represent the distribution of specific conformations due to preferential adherence to the grid surface, the visualized orientations are representative of the possible protein configurations. The signal-to-noise ratio is clearly improved by class-averaging (see inset in Figure 3A,B), leading to improved visualization and size estimation. We also confirm that TEM of the cutinase monomer (prior to reaction with the linker) showed a monodisperse collection of proteins having a single lobe (Figure S8).

X-ray Crystallography of Tetracutinase. We also used X-ray crystallography to obtain high-resolution structures of Cut_4-L and Cut_4-M. Crystals were grown at 25 °C by vapor diffusion in a sitting-drop format using 150 mM ammonium sulfate, 0.1 M MES (pH 5.5), and 25% PEG 4000 as a precipitant. Single crystals readily grew from a sample with a megamolecule concentration of 5 mg/mL. X-ray diffraction data were collected at the Argonne APS (Table S2) and were processed using XDS[39] and AIMLESS.[40] The structures of the tetracutinase molecules were resolved to ~1.3 Å and clearly show the expected phosphonate linkage between Ser-120 of the enzyme and the linker (Figure 4A). This modification is very similar to the one observed in cutinase inhibited by diethyl-*p*-nitrophenyl phosphate[42] (PDB ID 2CUT), a molecule that was used for molecular replacement, with the exception that the L/M linker molecule has an oxygen atom replaced by a carbon atom in the linker. Interestingly, the tetracutinase molecule crystallizes with the same cell parameters and space group as cutinase covalently modified with an organometallic phosphonate-pincer-metal complex (PDB ID 3EF3),[50] even though the crystallization conditions and the covalently attached group at Ser 120 are different. In the vicinity of the active site, the hydrogen-bond interactions with Ser42 and Gln 121 and the hydrogen bond between the imidazole ring of His 188 and the gamma oxygen of Ser 120 are preserved, as was observed in 2CUT and 3EF3. Clearly, the replacement of an oxygen molecule by a carbon molecule in the covalent linker region does not affect the overall conformation of the active site of the protein at this resolution. There is, however, a slight difference in the planarity of the atoms for the L and M linkers due to the introduction of the carbon, consistent with geometries expected for these atoms. Similar to other deposited structures, we observe a covalent bond between Cys 31 and Cys 109.

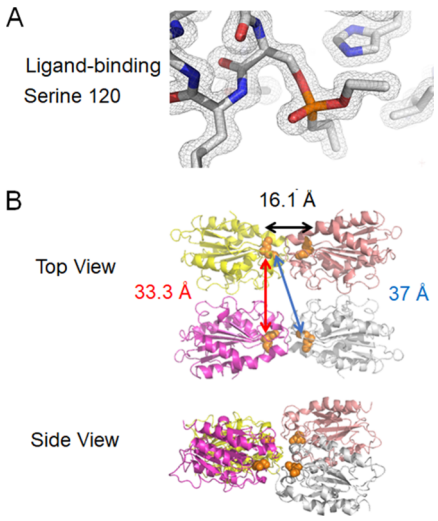


Figure 4. Representative structures of Cut_4-L obtained by X-ray crystallography. (A) Stick diagram of the Ser-120 residue of cutinase and its reaction with L linker **1** via phosphonylation. The mesh corresponds to the 2mF0-DFc electron-density map. (B) Overall structure of Cut_4-L is displayed on top view and side view. The phosphonate binding sites are colored orange. Intermolecular phosphate distances are measured for a possible packed tetramer of Cut_4-L.

Interestingly, the Cut_4-L structure shows clear density for what we interpret as a portion of the ethylene glycol linker. Even though the L- and M-modified tetracutinases were crystallized in the presence of PEG 4000, we only observe the PEG molecule in complex with Cut_4-L. Furthermore, a structure of the pincer/cutinase complex[50] (PDB ID 3EF3) was also crystallized in PEG and shows the same unit cell but does not contain a PEG molecule. These observations suggest that the PEG molecule observed in the Cut_4-L complex is a part of the linker and does not represent the ethylene glycol unit that derives from the crystallization mix. Finally, because this PEG linker is associated with the surface of cutinase between two symmetry-related molecules (and quite distal from Ser-120), we believe it represents a different arm of the linker.

We note that our use of gadolinium was critical for achieving ordered crystals. The original conditions that lacked gadolinium gave crystals characterized by stacked plates. Molecular replacement yielded clear solutions for these original crystals, but the packing was unusual in that the 2D layers were connected by head-to-head cutinase dimers, but there was a wide gap between adjacent layers consistent with a disordered layer. The addition of gadolinium promoted order in a different packing, leading to 3D crystals amenable for X-ray diffraction analysis (Figure S9).

The Cut_4-L/M megamolecules are more compact in the crystal than observed by TEM (Figures 4B and S10). In the crystals, we reason that packing forces led to a collapse of the linker regions. Nevertheless, extracting dimers of tetracutinase from the symmetry-related molecules in the crystal showed that these dimers are consistent in geometry and orientation with those observed in images obtained from TEM. Intermolecular phosphate distances for possible packed tetramers of Cut_4-L and Cut_4-M were measured in the crystals, which matched the average distance from the simulation (see below and Figures 4B, S10, and S11). The tetrameric structures in the crystals are distorted and related to the conformational flexibility of the linker and phase transition during crystallization. The distance between the active site of one molecule and the active site from a symmetry-related molecule is ~15 Å and would allow for the presence of the linker. Cut_4-L has a longer linker region than Cut_4-M and it appears that the protein accommodates these additional atoms

https://doi.org/10.1021/acs.biomac.1c00118
Biomacromolecules XXXX, XXX, XXX−XXX

by binding a portion of the linker in between molecules, as mentioned above.

Simulation of Protein Assembly. To gain insights into the equilibrium and dynamical behavior of the tetracutinase megamolecules, we prepared atomistic models and conducted MD simulations for the three structures derived from the L, M, and S linkers. For each megamolecule, we generated MD simulations over 200 ns, with four replicas each under three different force field setups, Amberff14SB + TIP3P at pH7, Amberff14SB + TIP3P at pH5, and Amberff99SB-disp at pH7.[28,29,34]

Compared to the gradient trend in molecular sizes from Cut_4-L to Cut_4-M observed in TEM, neither Amberff14SB + TIP3P + pH7 nor Amberff14SB + TIP3P + pH5 simulations were capable of reproducing the TEM molecular assembly. The Amberff14SB + TIP3P + pH7 simulations display a conformation where all cutinase molecules are aggregated and in direct contact with the others for all linker sizes. At the same time, the Amberff14SB + TIP3P + pH5 simulation, which follows the sample preparation pH used in TEM, presents a noticeable size gradient, albeit not as significant as the TEM observation.

Thus, we evaluated a new force field, Amberff99SB-disp, recently developed from Amberff99SB-ildn + TIP4PD to increase the solvation stability and flexibility of intrinsically disordered proteins.[34] By modifying the charge and Lennard-Jones parameters of the water molecules together with side-chain torsions of selected residues, the force field increases the protein solvation stability and thus reduces the improper aggregation among solute molecules. In contrast to Amberff14SB + TIP3P + pH7 and Amberff14SB + TIP3P + pH5, the Amberff99SB-disp simulations reproduce the TEM-observed sizes and shapes for both Cut_4-L and Cut_4-M. Details of the comparison of the MD simulation results with all three force field setups are shown in Table S3 and Figures S12 and S13.

To directly characterize the size of the megamolecules from the MD simulations, we calculated the molecular size and the radius of gyration for all simulations (Figure 5A).[51−54] The molecular size calculation, which corresponds to the maximum atom−atom distance on a projected 2D surface, is a close analogue to the molecular size measured on the 2D TEM images. The molecular sizes of Cut_4-L and Cut_4-M in simulations are 117.7 ± 6.4 and 101.6 ± 7.8 Å, respectively, which agree with the distances of 113 and 106 Å found by TEM. More importantly, the Amberff99SB-disp simulation correctly captures the increase in molecular size from Cut_4-M to Cut_4-L observed in TEM, whereas the other two force field setups do not. In all MD simulations of Cut_4-L and Cut_4-M, cutinase monomers are neither in direct contact nor separated by fully extended linkers but instead display an intermediate extended conformation of the linkers. The stable molecular size is approximately 30−40 Å smaller than the maximally extended conformation that is used as the starting configuration for the simulations. We surmise that this partially extended conformation is responsible for the dark region at the center of the Cut_4-L molecule in the TEM images (Figure 3C).

The overall shape of all tetracutinases, Cut_4-L, Cut_4-M, and Cut_4-S, under Amberff99SB-disp + pH7 approximates a regular tetrahedron (Figure 5B), which minimizes the intercutinase interactions. To quantify their overall shapes and fluctuations, we measured the linker distances in the radial orientation—which represent how far each cutinase extends from the core of the organic linker—and the tetrahedron regularity in the angular orientation, which measures all possible improper dihedral angles from four cutinase molecules. Going from Cut_4-L, Cut_4-M, to Cut_4-S, the linker fluctuations decrease with decreasing length, indicating that the longest linker L can sample more conformational space with a longer cutinase-core distance than can a shorter linker S. On the other hand, the decrease in the fluctuation of the tetrahedron regularity from Cut_4-L to Cut_4-M and Cut_4-S suggests an increasing stability promoted by direct intercutinase interactions.

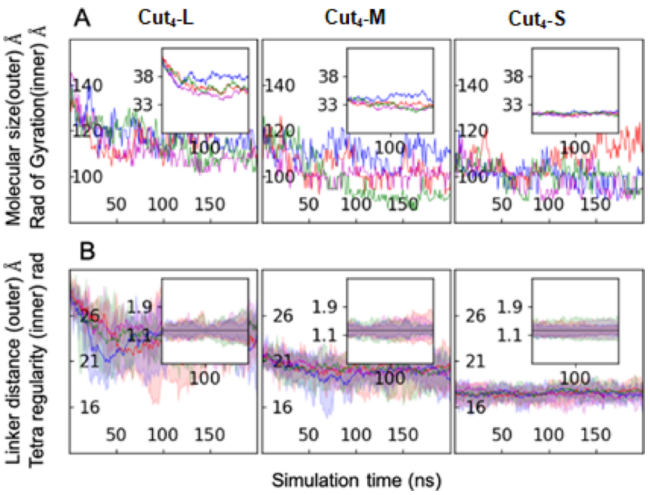


Figure 5. (A) Molecular size (outer) and radius of gyration (inner) measurements in the MD simulations with the force field setup of Amberff99SB-disp + pH7 for Cut_4-L, Cut_4-M, and Cut_4-S. A total of four replicas of the simulation are colored differently. (B) Linker distance (outer) and tetrahedron regularity (inner) measurements of MD simulations with the model of Amberff99SB-disp + pH7 for Cut_4-L, Cut_4-M, and Cut_4-S. The four replicas of the simulation are colored differently. The black line in the inner plot corresponds to the value of a regular tetrahedron.

In addition to the characterization just described, the MD simulations also explain why the shortest linker S can react with only three cutinases and is not able to form a tetramer. The trimer-to-tetramer reaction can only occur if the extension capability, which describes the maximum extension of the fourth arm in the tricutinase intermediate from the other three connected cutinases, exceeds the minimum distance required to avoid the steric clash. If it does not, the reaction steric hindrance does not allow this last reaction. The extension capability and the minimum distances required for the reaction can be calculated from tricutinase and the tetracutinase MD simulations. With the additional simulations of tricutinases, Cut_3-S and Cut_3-M, those comparisons were performed and are shown in Figure 6. For Cut_3-S, the extension capability has no overlap with the minimum distance required for another connection and hence does not permit the final reaction. For Cut_3-M, the linker arm of M is just long enough to have some overlap, which indicates that the reaction is possible. Hence, this analysis is consistent with the observed reactivity.

The tricutinase megamolecule also presents a different configuration from the tetracutinase megamolecule. Both Cut_3-M and Cut_3-S prefer a trigonal planar configuration, which is also the best configuration to maximize the separation of the cutinase domains. The TEM observations, along with tricutinase and the tetracutinase MD simulations, all imply

https://doi.org/10.1021/acs.biomac.1c00118
Biomacromolecules XXXX, XXX, XXX−XXX

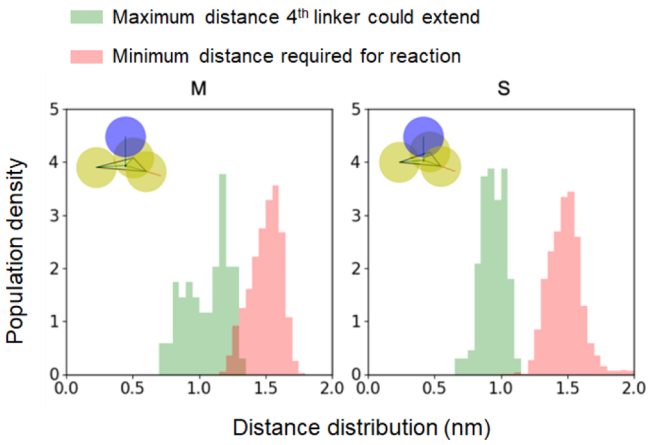


Figure 6. Distributions of linker extension capability (green) and minimum distance required for the tetramerization (red) of Cut_3-M and Cut_3-S. Distribution overlap for M but not for S indicates a possible tetramerization reaction for M but not for S. The blue (4th) and yellow (connected) spheres shown at the upper left indicate protein space for the reaction (left) and clash (right).

separation over aggregation for cutinase monomers in the megamolecule.

DISCUSSION

Our development of megamolecules offers a new approach to preparing extremely large yet structurally homogeneous molecules.[55−57] The use of reactions between enzymes and their covalent inhibitors allows for the efficient and modular assembly of protein building blocks into a broad range of architectures—including linear, cyclic, and branched scaffolds—by choosing linker molecules with distinct symmetries.[58−60] The approach is appealing because these scaffolds can be further functionalized with diverse groups including proteins, molecules, nanoparticles, and others.[61−65] These functions can be introduced by fusing proteins to the enzymes that are used to assemble the scaffolds. In this way, the enzyme domain becomes a structural element of the scaffold and the functional domain is attached to the scaffold, as we have demonstrated with the preparation of antibody mimics and antibody−enzyme conjugates.[4,5] Another benefit of the megamolecule approach is that functional molecules can also be attached to the linkers that are used in the assembly, allowing a broader range of chemistries to introduce those groups that are compatible with protein substrates.

However, our understanding of the structures and dynamics of this new class of molecules is still at an early stage, in part because the methods that have been important for understanding small molecules and proteins are challenging to apply to these large structures. To address this challenge, we demonstrate in this paper how high-resolution characterization—using both TEM and X-ray crystallography—can be combined with MD simulations to gain insights into the structures of these molecules and to understand how linker length affects the overall structure as well as aiding the design of linkers that will accommodate a target structure. Indeed, the correspondence of the imaging results and the MD simulations for the sizes of Cut_4-L and Cut_4-M is significant, as is the ability of the MD simulations to predict that steric factors would prevent the formation of Cut_4-S.

CONCLUSIONS

This work demonstrates the synthesis, characterization, and simulation of structurally defined four-armed megamolecules. The assembly of megamolecules through the reactions of enzymes and irreversible inhibitors has the benefits of high specificity and yield in each step of the synthesis. Moreover, the development of linkers that are functionalized with different irreversible inhibitors allows the preparation of complex megamolecule architectures. This work shows how the combination of high-resolution characterization and MD simulations can guide the development of megamolecules having defined conformations and dynamics. We believe that the megamolecules will find applications in many areas,[66−70] including in the development of antibody mimics for therapeutics.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.biomac.1c00118.

Protein sequences, SDS-PAGE analyses, single-particle analyses, TEM images, simulation results, simulation movies, X-ray crystallographic statistics, and linker syntheses (PDF)

MD simulation of linker **1** (MOV)

MD simulation of Cut_4-L (MOV)

MD simulation of Cut_4-M (MOV)

MD simulation of Cut_3-S (MOV)

AUTHOR INFORMATION

Corresponding Author

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

Authors

Shengwang Zhou − *School of Pharmacy, Jiangsu University, Zhenjiang 212013, P. R. China*

Peng He − *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*

Sonali Dhindwal − *Department of Materials Science, Northwestern University, Evanston, Illinois 60208, United States*

Valerie L. Grum-Tokars − *Department of Molecular Pharmacology and Biological Chemistry, Northwestern University, Chicago, Illinois 60611, United States*

Ying Li − *Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637, United States;* orcid.org/0000-0002-0124-781X

Kelly Parker − *Department of Materials Science, Northwestern University, Evanston, Illinois 60208, United States*

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

Reiner Bleher − *Department of Materials Science, Northwestern University, Evanston, Illinois 60208, United States;* orcid.org/0000-0003-0962-797X

https://doi.org/10.1021/acs.biomac.1c00118
Biomacromolecules XXXX, XXX, XXX−XXX

Roberto dos Reis − *Department of Materials Science, Northwestern University, Evanston, Illinois 60208, United States*
Joshua Zuchniarz − *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*
Vinayak P. Dravid − *Department of Materials Science, Northwestern University, Evanston, Illinois 60208, United States;* orcid.org/0000-0002-6007-3063
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;* orcid.org/0000-0002-3267-6748
Benoît Roux − *Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637, United States;* orcid.org/0000-0002-5254-2712

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.biomac.1c00118

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Army Research Office (ARO W911NF1810200 to M.M.) and the Air Force Office of Scientific Research (AFOSR FA9550-16-1-0150 to M.M.). K.P. gratefully acknowledges support from the Ryan Fellowship, the International Institute for Nanotechnology at Northwestern University, and the National Science Foundation Graduate Research Fellowship under grant DGE-1842165. This work made use of the IMSERC at Northwestern University and the BioCryo Facility and EPIC of Northwestern University's NUANCE Center, which has received support from the Soft and Hybrid Nanotechnology Experimental (SHyNE) Resource (NSF ECCS-1542205); the MRSEC program (NSF DMR-1720139) at the Materials Research Center; the International Institute for Nanotechnology (IIN); and the State of Illinois, through the IIN. LS-CAT/Sector 21 at the Advanced Photon Source, Argonne National Laboratory.

■ REFERENCES

(1) Modica, J. A.; Skarpathiotis, S.; Mrksich, M. Modular Assembly of Protein Building Blocks To Create Precisely Defined Megamolecules. *Chembiochem* **2012**, *13*, 2331−2334.
(2) Modica, J. A.; Lin, Y.; Mrksich, M. Synthesis of cyclic megamolecules. *J. Am. Chem. Soc.* **2018**, *140*, 6391−6399.
(3) Taylor, E. L.; Metcalf, K. J.; Carlotti, B.; Lai, C.-T.; Modica, J. A.; Schatz, G. C.; Mrksich, M.; Goodson, T. Long-Range Energy Transfer in Protein Megamolecules. *J. Am. Chem. Soc.* **2018**, *140*, 15731−15743.
(4) 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.
(5) 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. *Bioconjugate Chem.* **2021**, *32*, 143−152.
(6) Sagredo, S.; Pirzer, T.; Aghebat Rafat, A.; Goetzfried, M. A.; Moncalian, G.; Simmel, F. C.; de la Cruz, F. Orthogonal Protein Assembly on DNA Nanostructures Using Relaxases. *Angew. Chem., Int. Ed.* **2016**, *55*, 4348−4352.
(7) Yang, G.; Ding, H.-M.; Kochovski, Z.; Hu, R.; Lu, Y.; Ma, Y.-Q.; Chen, G.; Jiang, M. Highly Ordered Self-Assembly of Native Proteins into 1D, 2D, and 3D Structures Modulated by the Tether Length of Assembly-Inducing Ligands. *Angew. Chem., Int. Ed.* **2017**, *56*, 10691−10695.
(8) Pyles, H.; Zhang, S.; De Yoreo, J. J.; Baker, D. Controlling protein assembly on inorganic crystals through designed protein interfaces. *Nature* **2019**, *571*, 251−256.
(9) Rennie, M. L.; Fox, G. C.; Pérez, J.; Crowley, P. B. Auto-regulated Protein Assembly on a Supramolecular Scaffold. *Angew. Chem., Int. Ed.* **2018**, *57*, 13764−13769.
(10) Zacharias, M. Atomic Resolution Insight into Sac7d Protein Binding to DNA and Associated Global Changes by Molecular Dynamics Simulations. *Angew. Chem., Int. Ed.* **2019**, *58*, 5967−5972.
(11) Periole, X.; Huber, T.; Marrink, S.-J.; Sakmar, T. P. G Protein-Coupled Receptors Self-Assemble in Dynamics Simulations of Model Bilayers. *J. Am. Chem. Soc.* **2007**, *129*, 10126−10132.
(12) Kim, Y. E.; Kim, Y.-N.; Kim, J. A.; Kim, H. M.; Jung, Y. Green fluorescent protein nanopolygons as monodisperse supramolecular assemblies of functional proteins with defined valency. *Nat. Commun.* **2015**, *6*, 7134.
(13) Alberstein, R.; Suzuki, Y.; Paesani, F.; Tezcan, F. A. Engineering the entropy-driven free-energy landscape of a dynamic nanoporous protein assembly. *Nat. Chem.* **2018**, *10*, 732−739.
(14) Azuma, Y.; Zschoche, R.; Tinzl, M.; Hilvert, D. Quantitative Packaging of Active Enzymes into a Protein Cage. *Angew. Chem., Int. Ed.* **2016**, *55*, 1531−1534.
(15) Lapenta, F.; Aupič, J.; Strmšek, Ž.; Jerala, R. Coiled coil protein origami: from modular design principles towards biotechnological applications. *Chem. Soc. Rev.* **2018**, *47*, 3530−3542.
(16) Abendroth, J. M.; Cheung, K. M.; Stemer, D. M.; El Hadri, M. S.; Zhao, C.; Fullerton, E. E.; Weiss, P. S. Spin-Dependent Ionization of Chiral Molecular Films. *J. Am. Chem. Soc.* **2019**, *141*, 3863−3874.
(17) Stasiuk, G. J.; Long, N. J. The ubiquitous DOTA and its derivatives: the impact of 1, 4, 7, 10-tetraazacyclododecane-1, 4, 7, 10-tetraacetic acid on biomedical imaging. *Chem. Commun.* **2013**, *49*, 2732−2746.
(18) Vinjamuri, S.; Gilbert, T. M.; Banks, M.; McKane, G.; Maltby, P.; Poston, G.; Weissman, H.; Palmer, D. H.; Vora, J.; Pritchard, D. M.; Cuthbertson, D. J. Peptide receptor radionuclide therapy with ^{90}Y-DOTATATE/^{90}Y-DOTATOC in patients with progressive metastatic neuroendocrine tumours: assessment of response, survival and toxicity. *Br. J. Cancer* **2013**, *108*, 1440.
(19) Mueller, D.; Breeman, W. A. P.; Klette, I.; Gottschaldt, M.; Odparlik, A.; Baehre, M.; Tworowska, I.; Schultz, M. K. Radiolabeling of DOTA-like conjugated peptides with generator-produced 68Ga and using NaCl-based cationic elution method. *Nat. Protoc.* **2016**, *11*, 1057.
(20) Buehler, M. J.; Yung, Y. C. Deformation and failure of protein materials in physiologically extreme conditions and disease. *Nat. Mater.* **2009**, *8*, 175.
(21) Gatterdam, V.; Ramadass, R.; Stoess, T.; Fichte, M. A. H.; Wachtveitl, J.; Heckel, A.; Tampé, R. Three-Dimensional Protein Networks Assembled by Two-Photon Activation. *Angew. Chem., Int. Ed.* **2014**, *53*, 5680−5684.
(22) Maugeri, P. T.; Griese, J. J.; Branca, R. M.; Miller, E. K.; Smith, Z. R.; Eirich, J.; Högbom, M.; Shafaat, H. S. Driving Protein Conformational Changes with Light: Photoinduced Structural Rearrangement in a Heterobimetallic Oxidase. *J. Am. Chem. Soc.* **2018**, *140*, 1471−1480.
(23) Rohou, A.; Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **2015**, *192*, 216−221.
(24) Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **2012**, *180*, 519−530.
(25) Zivanov, J.; Nakane, T.; Forsberg, B. O.; Kimanius, D.; Hagen, W. J.; Lindahl, E.; Scheres, S. H. New tools for automated high-

resolution cryo-EM structure determination in RELION-3. *eLife* **2018**, *7*, No. e42166.
(26) Dennington, R.; Keith, T. A.; Millam, J. M. *GaussView*, version 6; Semichem Inc.: Shawnee Mission, KS, 2016.
(27) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A. *Gaussian 09*, Revision E.01; Gaussian, Inc.: Wallingford CT, 2016.
(28) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(29) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(30) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.
(31) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(32) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525−537.
(33) Søndergaard, C. R.; Olsson, M. H.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284−2295.
(34) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, No. E4758.
(35) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878−3888.
(36) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **2013**, *184*, 374−380.
(37) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084−3095.
(38) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528−1532.
(39) Kabsch, W. XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125−132.
(40) Evans, P. R.; Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013**, *69*, 1204−1214.
(41) Tickle, I. J.; Flensburg, C.; Keller, P.; Paciorek, W.; Sharff, A.; Vonrhein, C.; Bricogne, G. *Staraniso*; Global Phasing Ltd: Cambridge, United Kingdom, 2018. http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi.
(42) Martinez, C.; Nicolas, A.; van Tilbeurgh, H.; Egloff, M. P.; Cudrey, C.; Verger, R.; Cambillau, C. Cutinase, a lipolytic enzyme with a preformed oxyanion hole. *Biochemistry* **1994**, *33*, 83−89.
(43) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126−2132.
(44) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1997**, *53*, 240−255.
(45) Hodneland, C. D.; Lee, Y.-S.; Min, D.-H.; Mrksich, M. Selective immobilization of proteins to self-assembled monolayers presenting active site-directed capture ligands. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5048−5052.
(46) Martinez, C.; De Geus, P.; Lauwereys, M.; Matthyssens, G.; Cambillau, C. Fusarium solani cutinase is a lipolytic enzyme with a catalytic serine accessible to solvent. *Nature* **1992**, *356*, 615−618.
(47) Zhao, F.-Q.; Craig, R. Capturing time-resolved changes in molecular structure by negative staining. *J. Struct. Biol.* **2003**, *141*, 43−52.
(48) Cabra, V.; Murayama, T.; Samsó, M. Ultrastructural Analysis of Self-Associated RyR2s. *Biophys. J.* **2016**, *110*, 2651−2662.
(49) Ohi, M.; Li, Y.; Cheng, Y.; Walz, T. Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. *Biol. Proced. Online* **2004**, *6*, 23−34.
(50) Rutten, L.; Wieczorek, B.; Mannie, J.-P. B. A.; Kruithof, C. A.; Dijkstra, H. P.; Egmond, M. R.; Lutz, M.; Klein Gebbink, R. J. M.; Gros, P.; van Koten, G. Solid-State Structural Characterization of Cutinase−ECE-Pincer−Metal Hybrids. *Chem.—Eur. J.* **2009**, *15*, 4270−4280.
(51) Zhang, B. W.; Cui, D.; Matubayasi, N.; Levy, R. M. The Excess Chemical Potential of Water at the Interface with a Protein from End Point Simulations. *J. Phys. Chem. B* **2018**, *122*, 4700−4707.
(52) Sakae, Y.; Zhang, B. W.; Levy, R. M.; Deng, N. Absolute Protein Binding Free Energy Simulations for Ligands with Multiple Poses, a Thermodynamic Path That Avoids Exhaustive Enumeration of the Poses. *J. Comput. Chem.* **2020**, *41*, 56−68.
(53) Dominguez, L.; Foster, L.; Straub, J. E.; Thirumalai, D. Impact of membrane lipid composition on the structure and stability of the transmembrane domain of amyloid precursor protein. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, No. E5281.
(54) Dominguez, L.; Foster, L.; Meredith, S. C.; Straub, J. E.; Thirumalai, D. Structural Heterogeneity in Transmembrane Amyloid Precursor Protein Homodimer Is a Consequence of Environmental Selection. *J. Am. Chem. Soc.* **2014**, *136*, 9619−9626.
(55) Luo, Q.; Hou, C.; Bai, Y.; Wang, R.; Liu, J. Protein assembly: Versatile approaches to construct highly ordered nanostructures. *Chem. Rev.* **2016**, *116*, 13571−13632.
(56) Sciore, A.; Su, M.; Koldewey, P.; Eschweiler, J. D.; Diffley, K. A.; Linhares, B. M.; Ruotolo, B. T.; Bardwell, J. C. A.; Skiniotis, G.; Marsh, E. N. G. Flexible, symmetry-directed approach to assembling protein cages. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 8681.
(57) Sasaki, E.; Böhringer, D.; van de Waterbeemd, M.; Leibundgut, M.; Zschoche, R.; Heck, A. J. R.; Ban, N.; Hilvert, D. Structure and assembly of scalable porous protein cages. *Nat. Commun.* **2017**, *8*, 14663.
(58) Park, W. M.; Bedewy, M.; Berggren, K. K.; Keating, A. E. Modular assembly of a protein nanotriangle using orthogonally interacting coiled coils. *Sci. Rep.* **2017**, *7*, 10577.
(59) Ljubetič, A.; Lapenta, F.; Gradišar, H.; Drobnak, I.; Aupič, J.; Strmšek, Ž.; Lainšček, D.; Hafner-Bratkovič, I.; Majerle, A.; Krivec, N.; Benčina, M.; Pisanski, T.; Veličković, T. Ć.; Round, A.; Carazo, J. M.; Melero, R.; Jerala, R. Design of coiled-coil protein-origami cages that self-assemble in vitro and in vivo. *Nat. Biotechnol.* **2017**, *35*, 1094.
(60) Lai, Y.-T.; Cascio, D.; Yeates, T. O. Structure of a 16-nm cage designed by using protein oligomers. *Science* **2012**, *336*, 1129.
(61) Rudd, A. K.; Valls Cuevas, J. M.; Devaraj, N. K. SNAP-Tag-Reactive Lipid Anchors Enable Targeted and Spatiotemporally Controlled Localization of Proteins to Phospholipid Membranes. *J. Am. Chem. Soc.* **2015**, *137*, 4884−4887.
(62) Petershans, A.; Wedlich, D.; Fruk, L. Bioconjugation of CdSe/ZnS nanoparticles with SNAP tagged proteins. *Chem. Commun.* **2011**, *47*, 10671−10673.
(63) Wilkins, L. E.; Hasan, M.; Fayter, A. E. R.; Biggs, C.; Walker, M.; Gibson, M. I. Site-specific conjugation of antifreeze proteins onto polymer-stabilized nanoparticles. *Polym. Chem.* **2019**, *10*, 2986−2990.
(64) Liu, D. S.; Phipps, W. S.; Loh, K. H.; Howarth, M.; Ting, A. Y. Quantum Dot Targeting with Lipoic Acid Ligase and HaloTag for Single-Molecule Imaging on Living Cells. *ACS Nano* **2012**, *6*, 11080−11087.
(65) Popa, I.; Berkovich, R.; Alegre-Cebollada, J.; Badilla, C. L.; Rivas-Pardo, J. A.; Taniguchi, Y.; Kawakami, M.; Fernandez, J. M.

https://doi.org/10.1021/acs.biomac.1c00118
Biomacromolecules XXXX, XXX, XXX−XXX

Nanomechanics of HaloTag Tethers. *J. Am. Chem. Soc.* **2013**, *135*, 12762−12771.
(66) MaHam, A.; Tang, Z.; Wu, H.; Wang, J.; Lin, Y. Protein-Based Nanomedicine Platforms for Drug Delivery. *Small* **2009**, *5*, 1706−1721.
(67) Kang, H. J.; Kang, Y. J.; Lee, Y.-M.; Shin, H.-H.; Chung, S. J.; Kang, S. Developing an antibody-binding protein cage as a molecular recognition drug modular nanoplatform. *Biomaterials* **2012**, *33*, 5423−5430.
(68) Malay, A. D.; Miyazaki, N.; Biela, A.; Chakraborti, S.; Majsterkiewicz, K.; Stupka, I.; Kaplan, C. S.; Kowalczyk, A.; Piette, B. M. A. G.; Hochberg, G. K. A.; Wu, D.; Wrobel, T. P.; Fineberg, A.; Kushwah, M. S.; Kelemen, M.; Vavpetič, P.; Pelicon, P.; Kukura, P.; Benesch, J. L. P.; Iwasaki, K.; Heddle, J. G. An ultra-stable gold-coordinated protein cage displaying reversible assembly. *Nature* **2019**, *569*, 438.
(69) Strauch, R. C.; Mastarone, D. J.; Sukerkar, P. A.; Song, Y.; Ipsaro, J. J.; Meade, T. J. Reporter protein-targeted probes for magnetic resonance imaging. *J. Am. Chem. Soc.* **2011**, *133*, 16346−16349.
(70) Li, C.; Winnard, P. T.; Takagi, T.; Artemov, D.; Bhujwalla, Z. M. Multimodal image-guided enzyme/prodrug cancer therapy. *J. Am. Chem. Soc.* **2006**, *128*, 15072−15073.

https://doi.org/10.1021/acs.biomac.1c00118
Biomacromolecules XXXX, XXX, XXX−XXX