

2019

Covalently-assembled single-chain protein nanostructures with ultra-high stability

Wenqin Bai

Washington University in St. Louis

Cameron J. Sargent

Washington University School of Medicine in St. Louis

Jeong-Mo Choi

Washington University in St. Louis

Rohit V. Pappu

Washington University in St. Louis

Fuzhong Zhang

Washington University School of Medicine in St. Louis

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs

Recommended Citation





Bai, Wenqin; Sargent, Cameron J.; Choi, Jeong-Mo; Pappu, Rohit V.; and Zhang, Fuzhong, "Covalently-assembled single-chain protein nanostructures with ultra-high stability." *Nature Communications*, . . (2019).
https://digitalcommons.wustl.edu/open_access_pubs/8039

ARTICLE

<https://doi.org/10.1038/s41467-019-11285-8>

OPEN

Covalently-assembled single-chain protein nanostructures with ultra-high stability

Wenqin Bai¹, Cameron J. Sargent ², Jeong-Mo Choi ³, Rohit V. Pappu ³ & Fuzhong Zhang ^{1,2,4}

Protein nanostructures with precisely defined geometries have many potential applications in catalysis, sensing, signal processing, and drug delivery. While many *de novo* protein nanostructures have been assembled via non-covalent intramolecular and intermolecular interactions, a largely unexplored strategy is to construct nanostructures by covalently linking multiple individually folded proteins through site-specific ligations. Here, we report the synthesis of single-chain protein nanostructures with triangular and square shapes made using multiple copies of a three-helix bundle protein and split intein chemistry. Coarse-grained simulations confirm the experimentally observed flexibility of these nanostructures, which is optimized to produce triangular structures with high regularity. These single-chain nanostructures also display ultra-high thermostability, resist denaturation by chaotropes and organic solvents, and have applicability as scaffolds for assembling materials with nanometer resolution. Our results show that site-specific covalent ligation can be used to assemble individually folded proteins into single-chain nanostructures with bespoke architectures and high stabilities.

¹Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, Saint Louis, MO 63130, USA. ²Division of Biological & Biomedical Sciences, Washington University in St. Louis, Saint Louis, MO 63130, USA. ³Department of Biomedical Engineering and Center for the Science & Engineering of Living Systems, Washington University in St. Louis, Saint Louis, MO 63130, USA. ⁴Institute of Materials Science & Engineering, Washington University in St. Louis, Saint Louis, MO 63130, USA. Correspondence and requests for materials should be addressed to F.Z. (email: fzhang@seas.wustl.edu)

Over the past decade, the design rules and computational tools for DNA origami have been well established, and a wide variety of intricate and impressive DNA-based nanostructures have been demonstrated^{1–4}. In recent years, much attention has also been turned towards the creation of complex and tailor-made nanostructures using proteins, the class of biomolecules that nature employs to make both functional and structural materials. Proteins contain a wide variety of chemical functional groups, are capable of complex folding possibilities⁵, and can be sustainably produced *in vivo*. Accordingly, protein-based nanostructures have great potential to be used in valuable applications ranging from drug delivery vehicles^{5–7} to scaffolds for assembling inorganic nanoparticles or enzymes for sensing, catalysis, or signal transduction^{6,7}.

Although the field is still quite nascent, several design strategies currently exist for constructing protein nanostructures. Top-down approaches rely on naturally occurring protein structures, such as bacterial microcompartments, engineering these structures to allow regiospecific labeling of functional molecules and particles or to alter their dimensions and shapes⁸. However, because the extent of changes that can be made to these natural structures is fairly limited, rationally designing tailor-made structures through top-down approaches remains difficult. Meanwhile, bottom-up approaches to create *de novo* nanostructures using computationally designed protein structures and/or assembly interfaces have greatly expanded the diversity and potential applications of protein structures^{5,9}. Despite the difficulty of *de novo* protein design, recent developments have resulted in the creation of impressive protein-based nanostructures with precisely defined symmetries, geometries ranging from 2D arrays¹⁰ to 3D tetrahedra^{5,7,11} and icosahedra^{9,12}, and length scales ranging from nanometers⁵ to hundreds of nanometers⁶. A variety of protein interfaces using non-covalent interactions, such as hydrogen bonding, electrostatic interactions, the hydrophobic effect, and metal-coordinated interactions^{5,7,9,10,12–14}, have been employed to assemble these complexes. These efforts demonstrate the viability of creating protein-based nanostructures using bottom-up approaches.

A largely unexplored strategy is to link individually folded proteins via site-specific covalent ligation to generate homogeneous and controllable multi-subunit nanostructures. Although covalent ligation has been used in some instances to reinforce protein structures assembled via noncovalent interactions^{6,15,16} or to attach other functional proteins to such structures^{7,17}, the post-translational assembly of proteinaceous nanostructures exclusively through covalent interactions has not been widely explored. Site-specific covalent assembly could result in fewer defects from incorrectly assembled or kinetically trapped intermediate structures, which are often seen in complexes constructed via non-covalent interactions¹⁸. Additionally, site-specific covalent assembly could be used in conjunction with both top-down and bottom-up approaches, positioning it as a versatile tool to use in creating protein nanostructures. Furthermore, it stands to reason that structures assembled using covalent bonds will generally have higher thermodynamic stability at the assembly interfaces than structures assembled through non-covalent interactions, thus enabling broader applications in extreme temperatures and chemical conditions^{19,20}.

In this report, we explore this design strategy by covalently connecting individually folded protein blocks end-on-end to form cyclized, single-chain nanostructures with defined geometry. Specifically, we use split inteins (SI) to post-translationally ligate multiple copies of a hyper-stable three-helix bundle (3HB), creating 2D nanostructures with triangle and square shapes. By optimizing the lengths of the peptide linkers between each 3HB copy, we then produce nanostructures of increased uniformity

and tunable flexibility. These structures exhibit high stability in harsh denaturing conditions as well as utility as a scaffold for precisely assembling inorganic nanoparticles, demonstrating the merit of covalent assembly in developing protein-based nanostructures with exceptional and customizable properties.

Results

Design and synthesis of protein nanostructures. To construct single-polypeptide chain nanostructures with defined geometry, we used SI chemistry²¹ to covalently ligate multiple individually folded protein structures (referred to here as building blocks) into one cyclized structure. Of the various protein ligation approaches, SIs are well-suited for the assembly of protein nanostructures because they are small, autocatalytic domains and have several orthogonal pairings that allow for tightly controlled assembly through sequence-specific interactions. Compatible SIs fused to the termini of two protein building blocks can be used to form a peptide bond between them, leaving only a few scar residues at the ligation site. If building blocks are flanked by orthogonal SIs in the form of Int^C-building block-Int^N (with Int^C and Int^N representing the C- and N-terminal halves, respectively, of SI pairs), multiple building blocks can be covalently assembled. The large number of orthogonal SI pairs from both natural sources^{21,22} and engineering efforts²³ allows the covalent assembly of multiple building blocks in a defined order. Together with a selection of high-yield, fast-ligating SIs^{22,23}, we chose to use a recently engineered 3HB protein²⁴ as a building block to create simple yet highly stable 2D nanostructures. When folded, this 3HB forms a rigid protein rod that is 12 nm long and exceptionally stable (folding energy > 60 kcal/mol). Furthermore, the two termini of the 3HB are located on opposite ends of the rod-shaped protein, meaning that SI groups can be fused to both termini to build the structure bidirectionally. By cyclically ligating together a set number of orthogonally reactive 3HB blocks, 2D shapes with the same number of sides can be constructed (Fig. 1).

To create these structures, we designed and constructed multiple plasmids each encoding a 3HB block fused to non-complementary SI domains on both termini (Supplementary Tables 1 and 2, Supplementary Datasets 1 and 2). Additionally, a proline-glycine dipeptide was placed between 3HB and each of the SI groups of the fusion constructs to prevent the propagation of α -helical structure from the 3HB to the SIs³. Three blocks with complementary SIs were prepared: Gp^C₁₈-3HB-Cfa^N₁₄, Cfa^C₁₄-3HB-NrdJ^N₁₈, and NrdJ^C₁₈-3HB-Gp^N₁₈, where Gp, Cfa, and NrdJ are the orthogonal SI pairs^{22,23}, the superscripts denote the C- or N- half of each SI pair, and the subscripts represent the total number of residues in the corresponding peptide linker between copies of 3HB in the final structure after ligation. Each fusion protein was individually expressed and purified, and the three were then mixed together to form a covalently linked triangular nanostructure called Tri18 (Supplementary Fig. 1). When imaged by scanning transmission electron microscope (STEM; Fig. 2a), each edge of Tri18 was ~12 nm long, agreeing with the reported length of the original 3HB rod and indicating that the three 3HB copies in the assembled nanostructure folded properly. The vertex angles in each structure varied considerably, with a standard deviation of 17.4° from the mean value of 60°, the value expected for each angle in an equilateral triangle (Fig. 2b). In some irregular shapes, the angles deviated from 60° by as much as 40°. However, the edge lengths remained ~12 nm in all measured shapes, suggesting that these angle variations are caused by the flexible linkers rather than partially unfolded 3HB blocks.

During the synthesis of Tri18, we observed a few faint, low molecular weight bands on SDS-PAGE gels (Supplementary Fig. 1, Supplementary Table 3) after purification with affinity

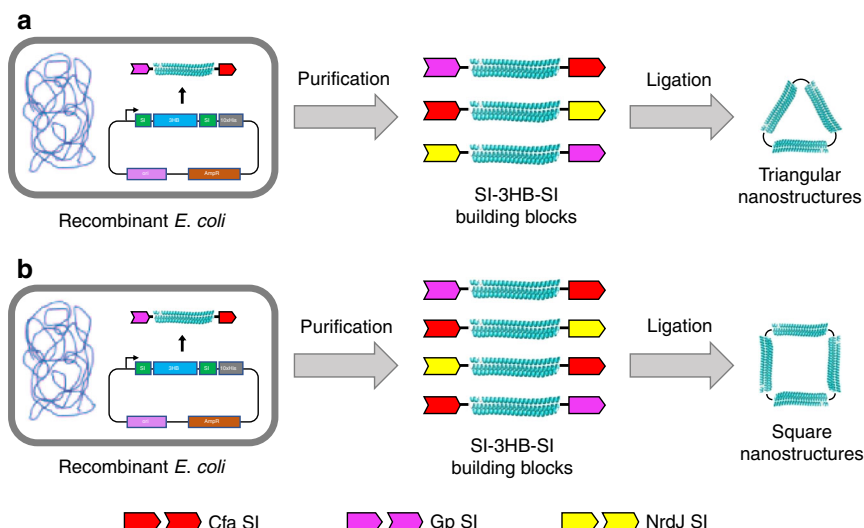


Fig. 1 Schematic illustration of the covalent assembly of triangle and square protein nanostructures. **a** By genetically fusing cognate SI pairs to the N- and C-termini, three copies of 3HB (PDB:4TQL) can be linked together to form triangular nanostructures. **b** Using constructs with a different combination of the same orthogonal SI pairs, square nanostructures containing four copies of 3HB can also be made

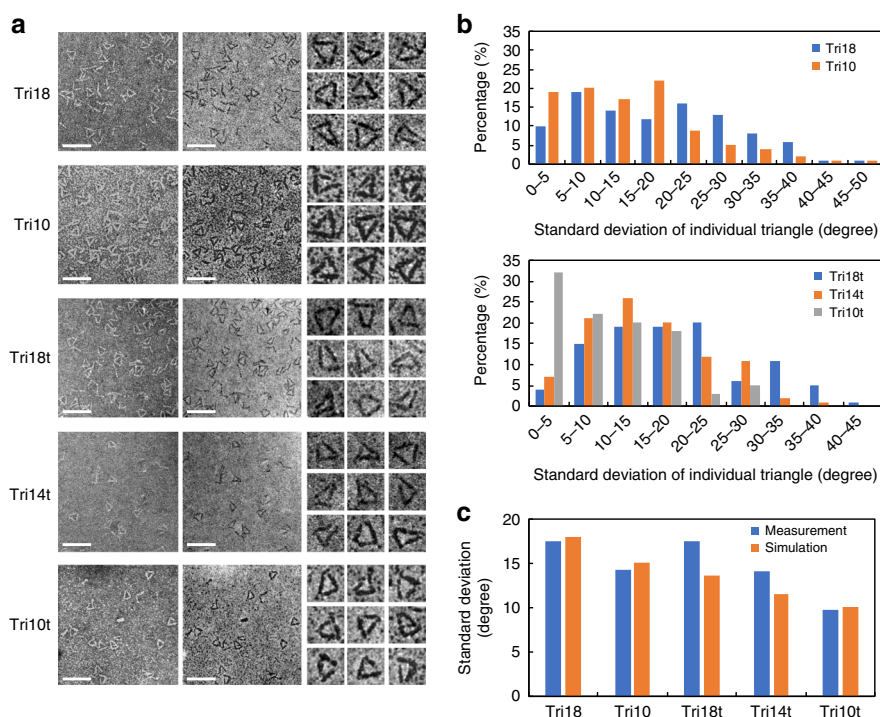


Fig. 2 Construction of protein triangles and optimization of vertex linker lengths. **a** STEM images of protein triangles. The images in the left column are bright-field overview images of the triangles, the images in the middle column are dark-field overview images of the same triangles, and the images in the right column are magnified dark-field images of individual triangles. Scale bars, 50 nm. Magnified images, 25 nm × 25 nm. **b** Histograms displaying the standard deviations of the three vertex angles within each individual triangle for the five triangular nanostructures made in this study using full length (above) and truncated (below) 3HB constructs ($n = 100$ triangles for each structure). **c** Standard deviations of all individual vertex angles in each of the five constructed triangles from STEM image measurements and computational simulations ($n = 100$ triangles for each structure). Source data are provided as a Source Data file

chromatography and, to a lesser extent, after further purification by size-exclusion chromatography (SEC) (Supplementary Table 3). To identify whether these were unreacted intermediates or side-products, these low molecular weight species were isolated by SEC and further analyzed by SDS-PAGE and STEM

(Supplementary Fig. 2). These species were identified as monomeric, dimeric, and uncyclized trimeric 3HB structures based on their relative SDS-PAGE gel migration rates (i.e. molecular weights; Supplementary Fig. 2b) and STEM images (Supplementary Fig. 2c). The SEC-isolated monomeric 3HB had

an apparent molecular weight ~ 15 kDa smaller than the unreacted, SI-flanked 3HB building blocks (Supplementary Figs. 1 and 2b), suggesting the loss of both SI domains. This result is in agreement with previous reports of SI side reactions, in which Int^C and/or Int^N groups are cleaved off the proteins prior to normal SI ligation²¹. Cleavage of SI domains without ligation could occur at several stages during the multi-step ligation reaction, generating monomers, dimers, or uncyclized trimers that can no longer react. Meanwhile, we also observed unreacted monomer, which can be explained by previously reported ligation yields of the SIs used (85–95%)²⁵. Further engineering of SIs to improve SI ligation yield could improve the overall yield of triangular nanostructures.

Next, to check whether the linkers between 3HB building blocks form any regular secondary structures or instead form heterogeneous, random coil-like ensembles, we performed atomistic simulations for those linkers using the ABSINTH implicit solvation model and forcefield paradigm. This simulation approach has been deployed successfully for uncovering sequence-ensemble relationships of a wide range of intrinsically disordered proteins^{26–33}. The results indicate a low level of secondary structure formation in the linkers (<10% of the ensemble in each case is part of an ordered secondary structure; Supplementary Table 4), and that the linkers indeed behave like Gaussian random chains, as their end-to-end distributions can be fitted well to a Gaussian model (Supplementary Fig. 3). End-to-end distance distributions deviate from being Gaussian if the effective solvent quality deviates from the theta point. Concordance with the Gaussian distribution implies that the effective solvent quality is theta-like for the linker sequences. Further, the linker lengths are larger than the Kuhn length but smaller than the thermal blob size, meaning that the chain behavior can therefore be simply explained by ideal-like interactions of Kuhn monomers.

Fine-tuning the flexibility of triangular nanostructures. To reduce the flexibility of triangular nanostructures, we truncated the peptides between the 3HB and SI groups to produce linkers in the final structure with 10 residues, yielding Gp^C₁₀-3HB-Cfa^N₁₀, Cfa^C₁₀-3HB-NrdJ^N₁₀, and NrdJ^C₁₀-3HB-Gp^N₁₀ building blocks. Ligation of these three blocks yielded the Tri10 nanostructure (Fig. 2a), which displayed slightly improved regularity (Fig. 2b). The number of residues between the 3HB and SIs were then further reduced to make blocks that would yield a final structure with linkers seven residues long. However, when mixed together, these blocks yielded linear products but no cyclized triangular products, suggesting that the reduced flexibility of these shortened linkers prevented cyclization. Meanwhile, we noticed that, based on the crystal structure, a few residues at the N- and C-termini of 3HB, namely NEDDM (residues 1–5) and GLE (residues 239–241), are not included in the α -helices and therefore do not contribute to the rigid portions of the structure but instead might contribute to the flexibility observed in the linkers. Accordingly, we prepared fusion proteins using a truncated 3HB lacking these residues together with different linker lengths. We ligated these to create three triangular nanostructures, Tri18t, Tri14t, and Tri10t (Supplementary Table 1). STEM imaging showed a dramatic reduction of vertex flexibility in Tri10t, with a large proportion of these structures being close to equilateral triangles (Fig. 2a, b). The uniformity of these triangles demonstrates that our approach can successfully create nanostructures with consistent and reliable dimensions, a necessary characteristic of synthetic nanochemistry.

To understand the relationship between the structural regularity and the linker length, we performed simple coarse-

grained simulations in which 3HB was modeled as a rigid rod and the linkers were represented by a two-dimensional Gaussian chain (see “Methods” section). We tested various combinations of spring constants for the linker chains and obtained a linear relationship of the summation of characteristic lengths of three Gaussian-chain linkers (ΣR_0) to the standard deviation (σ) of angles (Supplementary Fig. 4):

$$\sigma = -2.37^\circ + (0.20^\circ/\text{\AA}) \times \Sigma R_0. \quad (1)$$

This implies that chain extensibility (as quantified by the characteristic length) almost additively contributes to the triangle flexibility (as quantified by the standard deviation). The characteristic lengths of the experimental linkers can be determined as done above by Gaussian fitting of atomistic simulation data (Supplementary Table 5), which successfully reproduces the experimentally determined standard deviations (Fig. 2c). This shows that the flexibility of a triangular structure can be explained mostly by the near-ideal behavior of linkers, which have the requisite sequence features to ensure this behavior.

Synthesis of square-shaped nanostructures. Using the shortest linkers that allowed the formation of triangular structures, we designed four fusion proteins (Gp^C₁₀-3HB_t-Cfa^N₁₀, Cfa^C₁₀-3HB_t-NrdJ^N₁₀, NrdJ^C₁₀-3HB_t-Cfa^N₁₀, and Cfa^C₁₀-3HB_t-Gp^N₁₀) to create square-shaped protein nanostructures (Fig. 3a, Supplementary Figs. 5 and 6). While the distribution of the observed angles in these ligated squares peaked around the mean value of 90°, peaks of nearly the same magnitude around 70° and 110° and a standard deviation of 27.6° were also observed, indicating considerably greater angular freedom in the squares than in the triangles (Fig. 3b). Coarse-grained simulations suggest that simulated squares would show a significantly higher standard deviation (50.7°) than the measured value (27.6°) without any constraint on individual angles. These simulations also predict two asymmetric peaks around 30° and 135° (Supplementary Fig. 7), which can be explained by fluctuations around the rhombus-like nature of many of the structures imaged using STEM. Since the length of 3HB is much longer than the length scale of the linkers, the structure behaves like a rhombus with four equal sides and four angles: θ , $180^\circ - \theta$, θ , and $180^\circ - \theta$, where θ is a free parameter. Therefore, the linker flexibility not only broadens peaks but also produces a bimodal angular distribution unless it peaks around 90°. The experimentally observed trimodality thus implies that there may be two populations, one close to a perfect square and another with two pairs of angles $\sim 70^\circ$ and $\sim 110^\circ$. The simulations are based on a coarse-grained model and the assumption of rigid rods connected by flexible linkers. While this model captures the overall features of the structural ensemble, the distinct populations inferred from the experimental data, specifically those that yield the apparent trimodality in the θ parameter distribution, were not observed in the simulated ensemble. It would appear that the discrete states observed experimentally are reflective of the bendability of individual angles that are feasible for the current architecture. This requires further investigation and systematic titrations of sequence parameters of flexible linkers including the amino acid composition, which tunes the persistence length, and the overall lengths of linkers. At this juncture, such titrations are challenging because they directly impact the efficiency of cyclic ligation. Other methods for titrating the sequence parameters of linkers are necessary to probe for ways to increase the uniformity of square structures.

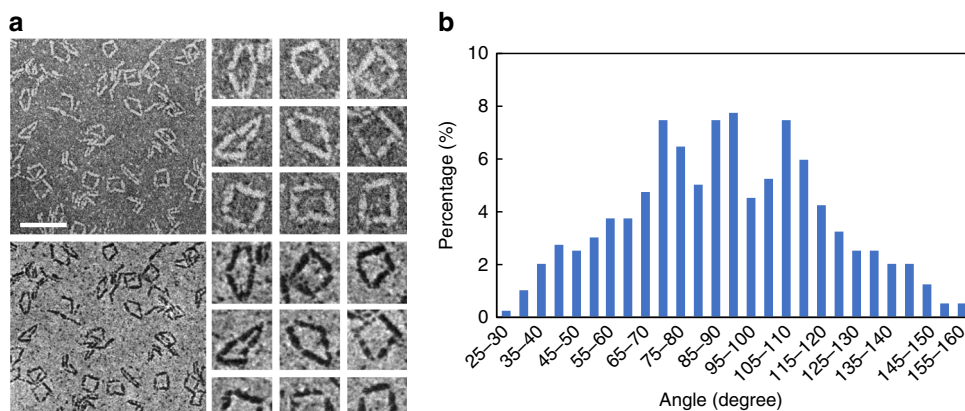


Fig. 3 Construction of protein squares and their angle distributions. **a** Bright-field (top) and dark-field (bottom) STEM images of the protein squares, with individual squares magnified on the right. Scale bars, 50 nm. Magnified images, 25 nm × 25 nm. **b** Experimental distribution of angles measured in the square nanostructures ($n = 400$, with four angles measured from each of 100 squares). Source data are provided as a Source Data file

Protein nanostructures display ultra-high thermostability. We next examined the stabilities of our single-chain nanostructures using the rigid Tri10t. Secondary structure stability was first assessed using circular dichroism (CD). Tri10t in phosphate-buffered saline (PBS; pH 7.4) with an additional 500 mM NaCl was incubated at 90 °C for 4 h, during which little to no unfolding was observed, consistent with the CD spectral changes reported previously for 3HB²⁴ (Fig. 4a). When the temperature was returned to 20 °C, the CD spectrum fully recovered its original shape, suggesting that there is either no unfolding or no irreversible unfolding of the α -helices in Tri10t at 90 °C over an extended period of time. We then examined the stability of Tri10t at high temperatures using STEM. Tri10t in PBS buffer (pH 7.4) with 500 mM NaCl was incubated at elevated temperatures for 1 h and then immediately prepared for STEM imaging. A large proportion of intact triangles at all tested temperatures and minimal degradation below 80 °C were observed, indicating that the cyclized nanostructures are highly stable (Fig. 4b, c). The fraction of opened structures increased as the incubation temperature increased, which could be attributed to more peptide hydrolysis in the linker regions at high temperatures. We also examined the thermostability of Tri10t in bulk using SDS-PAGE and observed similar thermostability (Supplementary Fig. 8a, b). We then evaluated the stability of Tri10t in harsh chemical conditions. Tri10t was incubated in chaotropic (6 M guanidine hydrochloride at 70 °C for 1 h) and organic solvents (ethanol, acetone, and DMSO for 24 h) and then added directly to grids and fixed. STEM imaging showed that the nanostructures remained almost entirely intact in all of these solvents, demonstrating their stability in harsh chemical environments (Fig. 4c, d).

Gold nanoparticle assembly using triangular nanostructures.

Finally, to demonstrate how these covalently assembled structures might be used as scaffolds for assembling other molecules of interest, we harnessed protein side-chain chemistry for site-specific labeling. Our designed Tri10t structure contains one cysteine residue located at each vertex of the triangular nanostructure, with no other cysteine residues on the edges. Tri10t was reacted with 1.4 nm, maleimide-functionalized gold nanoparticles (AuNPs) to form maleimide-thiol conjugates (Fig. 5a). After purification by gel filtration chromatography (Supplementary Fig. 9), the Tri10t-AuNP assembly was visualized by STEM using methylamine vanadate staining (Fig. 5b). As expected, the AuNPs

were site-specifically assembled at the three vertices of each triangle, proofing the concept of using these protein nanostructures as scaffolds. This method can be potentially extended to assemble AuNPs at other locations of a triangle by incorporating cysteine mutations into specific sites of the 3HB edges or into other geometries (e.g. square shape) by using nanostructures with alternative shapes.

Discussion

Herein we demonstrate the covalent assembly of individually folded 3HB proteins into hyperstable 2D protein nanostructures with triangle and square shapes using SI-mediated post-translational ligation. We tested different linker lengths for the triangle structures and found that shorter linkers lead to reduced structural flexibility, as expected, and that the linker length alone can explain the observed angle statistics with quantitative precision. The square structures provide an interesting symmetric trimodal angle distribution centered at 90° that is a reflection of the square and rhombic natures of these structures. In addition to the 3HB selected for this study, other individually folded proteins can be used as building blocks for covalent assemblies, greatly expanding the diversity and complexity of possible nanostructures. Natural discoveries and engineering efforts have produced a large library of orthogonal SIs²¹ that can be used to assemble more complex structures. Furthermore, site-specific, post-translational, covalent assembly is not limited to only SIs; other biochemical tools like sortase³⁴ and SpyTag/SpyCatcher³⁵ can be used in place of or even together with SIs. Incorporating multiple orthogonal ligation domains for covalent assembly can lead to more complex geometries, including 3D shapes by ligation using multi-component junctions.

The exceptional stability and precise scaffolding capabilities of our covalently assembled protein nanostructures allow them to be used for valuable applications not possible for many other nanostructures, protein or not. Their stability in extreme thermal conditions makes them ideal candidates for use as scaffolds for assembling and spatially coordinating thermophilic enzymes that operate most efficiently in elevated temperatures, such as many cellulolytic enzymes³⁶. These nanostructures could also be used to coordinate multi-step enzymatic reactions in harsh chemical conditions, acting as scaffolds for industrial enzymes in processes where organic solvents are unavoidable³⁷. In addition to their thermal and chemical stability, these cyclized structures are also expected to be protected from proteolysis, especially by

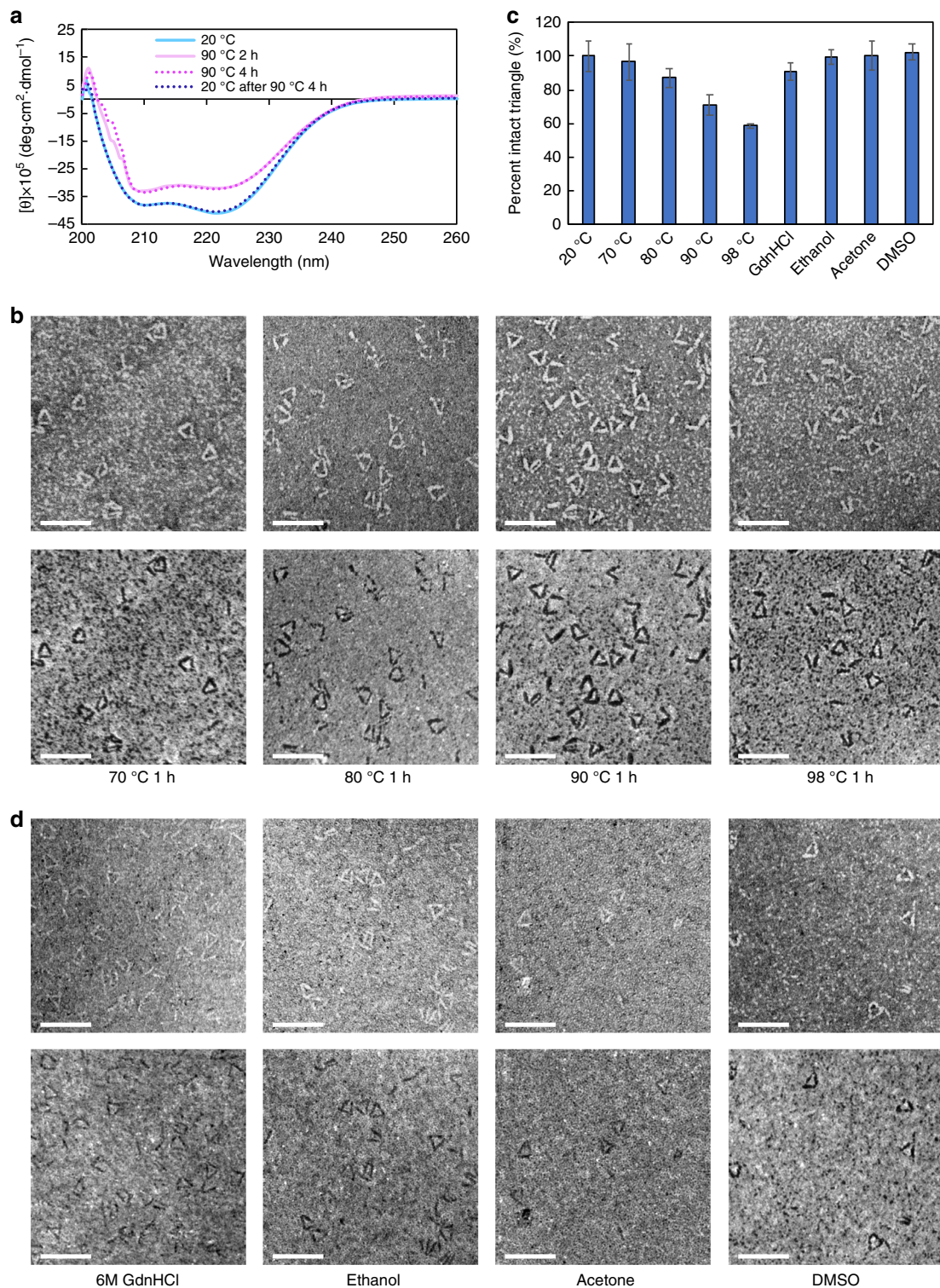


Fig. 4 Thermal and chemical stability of the protein triangles. **a** CD spectra of the Tri10t sample incubated at 20 and 90 °C. **b** Bright-field (top) and dark-field (bottom) STEM images of protein triangles after 1 h incubation at different temperatures. Scale bars, 50 nm. **c** Percentages of 3HB copies contained in intact triangular nanostructures after incubation in different conditions, normalized by pre-incubation values. Error bars represent the s.d. ($n = 3$ denaturation experiment replicates). **d** Bright-field (top) and dark-field (bottom) STEM images of the protein triangles after incubation in chaotropic and organic solvents. Scale bars, 50 nm. Source data are provided as a Source Data file

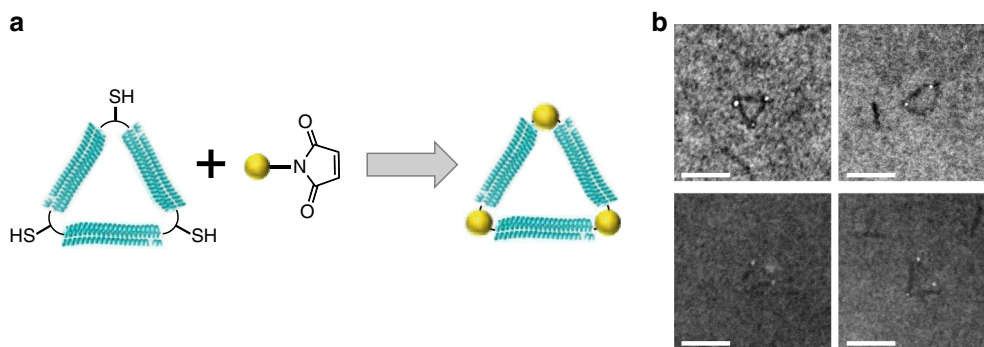


Fig. 5 Assembly of gold nanoparticles (AuNPs) using Tri10t as a molecular scaffold. **a** Schematic depicting the conjugation of AuNPs to the vertices of the Tri10t protein nanostructures. **b** Dark-field STEM images showing AuNPs (lighter than background) attached to the three vertices of the protein nanostructures (darker than background). Scale bars, 20 nm

exopeptidases¹⁹, making them prime candidates for biomedical applications such as the generation of advanced multivalent vaccine nanoparticles, which have been demonstrated to elicit improved immune responses^{5,38}.

Methods

Strains and growth conditions. For molecular cloning, *Escherichia coli* NEB 10 β was cultured at 37 °C in Luria-Bertani (LB) medium with 100 μ g/mL ampicillin. For protein production, *E. coli* BL21(DE3) was cultured at 37 °C in Terrific Broth (TB) containing 24 g/L yeast extract, 20 g/L tryptone, 0.4% v/v glycerol, 17 mM KH_2PO_4 , and 72 mM K_2HPO_4 with 100 μ g/mL ampicillin.

Chemicals and reagents. Unless otherwise noted, all chemicals and reagents were obtained from Millipore Sigma. Plasmid purification and gel extraction kits were purchased from iNTRON Biotechnology. FastDigest restriction enzymes and T4 DNA ligase were purchased from Thermo Fisher Scientific.

Plasmid construction. Amino acid sequences of the orthogonal split inteins (SIs; Gp41-1, Cfa, and NrdJ) were obtained from published papers^{22,23,39} and codon-optimized for expression in *E. coli* using the SI-Bricks method⁴⁰. The optimized DNA sequences were chemically synthesized by Integrated DNA Technologies, Inc. (IDT) and PCR-amplified using different sets of primers (Supplementary Table 2 and Supplementary Dataset 1; also acquired from IDT) for Golden Gate DNA Assembly. The DNA sequence of the three-helix bundle (3HB) was amplified from the 3H5L_2 gene²⁴ using primers as listed (Supplementary Table 2 and Supplementary Dataset 1). PCR-amplified DNA fragments were inserted into expression plasmids using Golden Gate DNA Assembly⁴¹. The amino acid sequences of the resulting fusion proteins are listed in Supplementary Dataset 2. Sequences of all assembled plasmids were confirmed by DNA sequencing (Eurofins Genomics, USA).

Expression and purification of 3HB-SI fusion proteins. Sequenced plasmids were transformed into host cells for protein overexpression. Transformed cells were pre-cultured at 37 °C overnight in 5 mL TB containing 100 μ g/mL ampicillin. Overnight cultures were used to inoculate 1 L TB with 100 μ g/mL ampicillin. The cultures were grown at 37 °C until OD₆₀₀ reached 2 and then induced by the addition of 20% arabinose to a final concentration of 0.1%. Cells were then cultured at 18 °C for 20 h and harvested by centrifugation. To purify the 3HB-SI fusion proteins, cell pellets were resuspended in 1X PBS (pH 7.4) with 10 mM imidazole and lysed by sonication. Cell lysate was then centrifuged at 4 °C and 18,000 \times g for 30 min, and the precipitate was resuspended in Buffer A (1X PBS, 8 M urea, 10 mM Imidazole, 1 mM TCEP, pH 7.4) and loaded onto a pre-equilibrated Ni-NTA column (Qiagen). After loading, the column was washed with 10 column volumes (CV) of Buffer B (1X PBS, 8 M urea, 60 mM imidazole, 1 mM TCEP, pH 7.4), followed by washing with 10 CV of Buffer C (1X PBS, 8 M urea, 1 M NaCl, 1 mM TCEP, pH 7.4). The desired proteins were eluted with 10 CV of Buffer D (1X PBS, 8 M urea, 300 mM imidazole, 1 mM TCEP, pH 7.4), and the eluted fractions were dialyzed against 1X PBS.

Synthesis and purification of protein triangles and squares. To construct triangular nanostructures, two 3HB-SI building blocks (e.g. in the case of Tri18, Cfa^C₁₄-3HB-NrdJ^N₁₈ and NrdJ^C₁₈-3HB-Gp^N₁₈) were first mixed at a 1:1 ratio in Buffer E (1X PBS, 1 M NaCl, 1 mM TCEP, pH 7.4) at 37 °C for 8 h. The third building block (in this case Gp^C₁₈-3HB-Cfa^N₁₄) was then added in the same molar ratio and incubated at 37 °C for another 12 h to allow complete ligation. To

construct square nanostructures, two 3HB-SI building blocks (Gp^C₁₀-3HB-Cfa^N₁₀ and Cfa^C₁₀-3HB-NrdJ^N₁₀) were mixed at 1:1 ratio in Buffer E at 37 °C for 8 h. Meanwhile, the other two building blocks (NrdJ^C₁₀-3HB-Cfa^N₁₀ and Cfa^C₁₀-3HB-Gp^N₁₀) were also ligated using the same conditions. The two reaction mixtures were then mixed at a 1:1 molar ratio and incubated at 37 °C for another 12 h.

Because the histidine tag on each building block is removed during SI ligation, the cyclized protein nanostructures lacked a histidine tag and were thereby purified from both unreacted building blocks and partially reacted intermediates (e.g. dimers, trimers) by affinity chromatography. Each ligation product was loaded onto a pre-equilibrated Ni-NTA column, and the flow through containing the cyclized nanostructures was collected. The flow through was subsequently purified by SEC using a Superose 6 Increase 10/300 GL column (GE Healthcare). The samples were loaded with 1X PBS buffer with 500 mM NaCl and eluted with an isocratic elution.

SDS-PAGE analysis. All SDS-PAGE gels were 1 mm thick and discontinuous with a 6% stacking gel and 10% separating gel. Protein building block samples, ligation intermediate/side product, and circularized triangle and square were prepared in Laemmli sample buffer (2% SDS, 10% glycerol, 60 mM Tris, pH 6.8, 0.01% bromophenol blue, 50 μ M DTT) and boiled at 100 °C for 15 min. The Dual Color Precision Plus Protein™ Standard (BioRad) was used as reference. Gels were run on Mini-PROTEAN Tetra Cells (Bio-Rad) in 1x Tris-glycine SDS buffer (25 mM Tris base, 250 mM glycine, 0.1% w/v SDS) until just before the dye front exited the gel. Gels were stained in Coomassie Blue solution (50% v/v methanol, 10% v/v acetic acid, 1 g/L Coomassie Brilliant Blue) for a minimum of 1 h at room temperature with gentle agitation and destained in destaining buffer (40% v/v methanol, 10% v/v acetic acid) for a minimum of 1 h. Gels were imaged on an Azure c600 Imager (Azure Biosystems). Uncropped and unprocessed scans of SDS-PAGE gels are included in the Source Data file.

Negative stain electron microscopy. A 10 μ L drop of purified triangular or square nanostructures or of ligation intermediates/side products was applied to UV-treated (Novascan), pure carbon 400-mesh copper grids (Ted Pella, Inc.) and allowed to incubate for 5 min. Grids were then washed with three drops of Milli-Q water and a 10 μ L drop of 0.75% uranyl formate, after which they were stained using a 10 μ L drop of 0.75% uranyl formate, incubated for 3 min⁴². Liquids were wicked off of the grids between each step using filter paper, and the grids were allowed to air-dry. Samples were imaged on a STEM (JEM-2100F, JEOL, Japan). STEM images were collected using both a bright-field (BF) detector and a high-angle annular dark-field (HAADF) detector at 200 kV. Images were simultaneously recorded from both STEM detectors. The camera length was 12 cm and magnification was \times 1,000,000.

Data analysis by ImageJ. For each sample, 100 triangles or squares were randomly selected by manually identifying triangle-shaped or square-shaped arrangements of 3HB proteins that had, based on the results from atomistic simulations, no more than 3.5 nm separation between each pair of adjacent 3HB edges. The interior angles for these cyclized structures were measured using ImageJ software, and the standard deviation of the three vertex angles within each individual triangle for the five triangular nanostructures was calculated as

$$SD = \sqrt{\frac{\sum(X_i - \langle 60 \rangle)^2}{n - 1}}, \quad (2)$$

where n is 3 and x_i is the vertex angle value of each individual triangle in that nanostructure. The standard deviation for all angles in triangular nanostructures

was calculated as

$$SD = \sqrt{\frac{\sum(x_i - 60)^2}{n - 1}} \quad (3)$$

where n is 300 and x_i is the measured angle values.

Atomistic simulations of protein linkers. We performed atomistic simulations for a series of peptide linkers, modeled as autonomous units, and for the tethered systems conforming to the architecture of 3HB-linker-3HB. In all, we performed simulations for 14 different systems. In these simulations, the N- and C-termini of each of the peptides were capped using acetyl and N-methylamide groups, respectively. Details of the sequences used in each of the simulations are shown in Supplementary Table 5. The goal is to obtain converged, statistically robust descriptions of conformational ensembles for each of the simulated systems.

Accordingly, we performed simulations using the ABSINTH implicit solvation model and forcefield paradigm⁴³ as implemented in the CAMPARI modeling suite v.2⁴⁴. The ABSINTH model is interoperable with standard molecular mechanics forcefields and used parameters from the all atom OPLS-AA/L forcefield as implemented in the `abs_ompls3.2.prm` parameter set within CAMPARI. For each peptide, we performed 200 independent Markov Chain Metropolis Monte Carlo (MC) simulations. The starting structures for these simulations were extracted randomly from a distribution of self-avoiding conformations. Each simulation consists of 5.1×10^6 MC steps, and the first 0.1×10^6 steps were discarded for equilibration. The atomic coordinates were registered every 5000 steps. The simulation temperature was set to 300 K, and the background NaCl concentration was set to ~150 mM while preserving charge neutrality. Ions are modeled explicitly while the solvent is treated in a mean field manner using the ABSINTH model. Our choice of solvation model, forcefield, and sampling protocols were motivated by the following considerations: We sought a combination of solvation model and forcefield with a proven track record of accuracy for simulating folded domains and disordered regions. Many of the established combinations of explicit solvent models and molecular mechanics models lack the accuracy to be applied to systems that include hyper-stable ordered domains and disordered regions with significant conformational heterogeneity. The ABSINTH model serves this purpose, as was shown recently in a high-throughput study of disordered regions tethered as linkers and/or tails to ordered domains²⁹. Further, the computational demands increase with the size of the system, and, for molecules corresponding to the architecture 3HB-linker-3HB, simulations with explicit representations of solvent molecules become orders of magnitude more expensive when compared to simulations of the linkers alone. This increase in computational complexity is offset by the choice of the highly efficient ABSINTH model. Because the ABSINTH model relies on a fast estimation of conformational specific free energies of solvation, referenced as they are to experimentally derived free energies of solvation for model compounds, we are in a position to assess the interplay between solvation and desolvation effects through the interplay of the direct mean field interaction with the solvent and the polar term. Details of how the ABSINTH model works and why it works well have been previously published^{28,43,45}. The attached key file that is interoperable with version 2.0 of CAMPARI enables reproduction of the simulation results.

To obtain the end-to-end distance distribution for each peptide, we computed the distances between the two sp^3 carbon atoms at the terminal acetyl and N-methylamide groups and constructed a histogram with bin size of 1 Å. The histogram was then converted into potential of mean force (PMF) by taking a logarithm and dividing it by $4\pi r^2$. We fitted the PMF with the fitting function,

$$f(x) = A + 3(x/R_0)^2, \quad (4)$$

to extract the Gaussian characteristic length R_0 . If the fitting does not work well, it indicates that the chain cannot be assumed to be Gaussian, which was not the case (Supplementary Fig. 3). The secondary structure analysis was performed by the DSSP analysis module⁴⁶ implemented in the CAMPARI engine. The fractions of residues assigned as 'H' and 'E' were used as proxies to α and β contents, respectively.

Coarse-grained simulations of protein nanostructures. We developed a phenomenological, rods-and-springs coarse-grained model to perform simulations of triangular and square nanostructures. A schematic of the model for triangular nanostructures is shown in Supplementary Fig. 10. In the rod-and-springs representation, residues 6–238 of 3HB are modeled as rigid rods. These rods are of fixed length 114.32 Å, which corresponds to the distance between the two alpha carbons of residues 6 and 238 in the 3HB crystal structure. The rigid rod model is based on the assumption that the rod is not significantly distorted when compared to the flexible tethers, at least on experimentally relevant time scales. The linkers are modeled as Gaussian chains with an elastic energy written in terms of a harmonic potential:

$$E(x) = 2(x/R_0)^2; \quad (5)$$

note the factor of 2, which is appropriate given that the dimensionality is 2 for planar structures. Here, R_0 is determined by regression analysis where a second-order polynomial $f(x) = A - 3(x/R_0)^2$ is used to fit the PMF written in terms of the end-to-end distances extracted from atomistic simulations. The regression analysis

yields linker specific values for R_0 and these are used in the coarse-grained simulations.

Starting from an initial default structure, which is an equilateral triangle or square with side length of 114.32 Å, a MC procedure was used to equilibrate the system. At each MC step, one side is randomly picked among three and tilted by a random angle variable that follows a Gaussian distribution of mean 0 and standard deviation 0.1° . Then the tip-to-tip distance x is converted into the Gaussian chain energy $E(x)$. The Gaussian chain energy is calculated for every linker, and the system energy is obtained by a simple sum of all 3 or 4 linker contributions. If the trial move gives lower energy, the program accepts the move. Otherwise, it accepts the move according to the Metropolis MC criterion. We performed 100 independent simulations for triangular structures, each of which consists of 10^7 MC steps, and the coordinates were registered every 10^4 steps, discarding the first 5×10^5 steps. For square structures, we also performed 100 simulations, but the step number was increased to 10^9 and we discarded the first 10^8 MC steps. The distributions of bending angles were constructed from the coordinate information, from which the standard deviation was also calculated.

CD measurement. CD measurements were performed on a JASCO J-810 CD spectrometer equipped with a Lauda RM 6 refrigerated circulator and a JASCO PTC-423S peltier. Each protein nanostructure was dissolved in 1X PBS at a concentration of 10 μ M and loaded into a capped 1 mm quartz cuvette (Hellma, Germany). CD spectra were scanned from 200 to 260 nm using a 1 nm step, 1 nm bandwidth, 50 nm/min scanning speed, and 2 s response time. The stability of nanostructures was measured by scanning CD spectra from 200 to 260 nm at 20 or 90 °C after incubating for varying amounts of time as indicated. All CD spectra were recorded in triplicate and averaged.

Thermal and chemical stability of protein nanostructures. The triangular nanostructure Tri10t (5 μ M in 1X PBS) was incubated for 1 h at four different temperatures (70, 80, 90, and 98 °C). Each sample was then mixed with Laemmli sample buffer and run in 1 mm SDS-PAGE gels in 1x Tris-glycine SDS buffer until just before the dye front exited the gel. Gels were then stained, destained, and imaged as before. Each sample was prepared and run on a gel three separate times. Protein band intensities were integrated using GelAnalyzer 2010 software (www.gelanalyzer.com). After incubation, samples were also immediately loaded to a UV-treated, pure carbon copper grid and prepared by 0.75% uranyl formate negative staining for STEM imaging. For each sample, ~150–400 copies of 3HB were randomly selected. The percentage of these 3HB copies that were contained in intact triangular nanostructures, as defined previously, was calculated. Percentages were reported after being normalized by the corresponding percentage calculated for an unincubated sample (prepared at 20 °C). The same experiment was repeated in triplicate for each sample to obtain mean percent intact triangle and standard deviation.

To evaluate the stability of these nanostructures in organic solvents, Tri10t was incubated at 50 nM in ethanol, acetone, or dimethylsulfoxide (DMSO) at 4 °C for 24 h. Samples were directly loaded to a carbon-coated copper grid and stained in triplicate as before, followed by STEM imaging. To assess the stability of these nanostructures in chaotropic conditions, Tri10t was dissolved at 500 nM in 1X PBS with 6 M guanidine hydrochloride (GdnHCl) and then incubated at 70 °C for 1 h. After incubation, the sample was diluted into 50 mM HEPES buffer (pH 7.5) and directly loaded to a carbon-coated copper grid and stained as before. All images were recorded by STEM and analyzed as described above. On each of the three grids per sample, ~50–120 copies of 3HB were counted for samples incubated in ethanol and GdnHCl, while ~30–50 copies were counted for samples incubated in acetone and DMSO due to protein aggregation. Percentages of 3HB copies contained in intact triangular nanostructures were calculated as before and normalized by the corresponding percentage from a sample incubated in 1X PBS with 500 mM NaCl. Samples were prepared and analyzed in triplicate.

AuNP labeling of triangular nanostructure vertices. AuNP labeling of triangular nanostructure vertices was performed based on the protocol provided for the Monomaleimido Nanogold[®] Labeling Reagent (Nanoprobe), with minor modifications. The Tri10t nanostructure purified using affinity chromatography was incubated in 1X PBS (pH 7.4) with 500 mM NaCl and 2 mM TCEP to keep the sulfhydryl group reduced at room temperature for 2 h. TCEP was then removed by SEC using a Superose 6 Increase 10/300 GL column and an isocratic elution buffer (0.02 M sodium phosphate, pH 6.5, 500 mM NaCl, and 1 mM EDTA). The reduced Tri10t sample (1 nmol) was then mixed with 0.2 ml Nanogold[®] reagent (6 nmol) in 0.02 M sodium phosphate (pH 6.5) with 500 mM NaCl. The reaction mixture was incubated at room temperature for 2 h. Tri10t nanostructure conjugates were separated from unreacted AuNPs by gel filtration chromatography using a Superose 6 Increase 10/300 GL column. The Tri10t-AuNP conjugates were eluted with 0.02 M sodium phosphate (pH 7.4) with 500 mM NaCl.

The purified Tri10t nanostructure conjugates were imaged by negative stain electron microscopy as described above, only using 2% methylamine vanadium (Nanoprobe) instead of 0.75% uranyl formate.

Data availability

The source data underlying Figs. 2b, c, 3b, 4a, c, Supplementary Figs. 1–9, and Supplementary Table 3 are provided as a Source Data file. Other relevant data are available from the corresponding author upon reasonable request.

Received: 4 February 2019 Accepted: 2 July 2019

Published online: 25 July 2019

References

- Rothmund, P. W. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
- Tikhomirov, G., Petersen, P. & Qian, L. Fractal assembly of micrometre-scale DNA origami arrays with arbitrary patterns. *Nature* **552**, 67–71 (2017).
- Ong, L. L. et al. Programmable self-assembly of three-dimensional nanostructures from 10,000 unique components. *Nature* **552**, 72–77 (2017).
- Benson, E. et al. DNA rendering of polyhedral meshes at the nanoscale. *Nature* **523**, 441–444 (2015).
- Ljubetic, A. et al. Design of coiled-coil protein-origami cages that self-assemble in vitro and in vivo. *Nat. Biotechnol.* **35**, 1094–1101 (2017).
- Fletcher, J. M. et al. Self-assembling cages from coiled-coil peptide modules. *Science* **340**, 595–599 (2013).
- Gradisar, H. et al. Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nat. Chem. Biol.* **9**, 362–366 (2013).
- Sasaki, E. et al. Structure and assembly of scalable porous protein cages. *Nat. Commun.* **8**, 14663 (2017).
- Hsia, Y. et al. Design of a hyperstable 60-subunit protein icosahedron. *Nature* **535**, 136–139 (2016).
- Gonen, S., DiMaio, F., Gonen, T. & Baker, D. Design of ordered two-dimensional arrays mediated by noncovalent protein–protein interfaces. *Science* **348**, 1365–1368 (2015).
- King, N. P. et al. Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103–108 (2014).
- Bale, J. B. et al. Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389–394 (2016).
- Brodin, J. D. et al. Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem.* **4**, 375–382 (2012).
- Radford, R. J. & Tezcan, F. A. A superprotein triangle driven by nickel(II) coordination: exploiting non-natural metal ligands in protein self-assembly. *J. Am. Chem. Soc.* **131**, 9136–9137 (2009).
- Padilla, J. E., Colovos, C. & Yeates, T. O. Nanohedra: using symmetry to design self-assembling protein cages, layers, crystals, and filaments. *Proc. Natl Acad. Sci. USA* **98**, 2217–2221 (2001).
- Kobayashi, N. et al. Self-assembling supramolecular nanostructures constructed from de novo extender protein nanobuilding blocks. *ACS Synth. Biol.* **7**, 1381–1394 (2018).
- Shekhawat, S. S., Porter, J. R., Sriprasad, A. & Ghosh, I. An autoinhibited coiled-coil design strategy for split-protein protease sensors. *J. Am. Chem. Soc.* **131**, 15284–15290 (2009).
- Bai, Y., Luo, Q. & Liu, J. Protein self-assembly via supramolecular strategies. *Chem. Soc. Rev.* **45**, 2756–2767 (2016).
- Iwai, H. & Pluckthun, A. Circular beta-lactamase: stability enhancement by cyclizing the backbone. *FEBS Lett.* **459**, 166–172 (1999).
- Iwai, H., Lingel, A. & Pluckthun, A. Cyclic green fluorescent protein produced in vivo using an artificially split PI-Pful intein from *Pyrococcus furiosus*. *J. Biol. Chem.* **276**, 16548–16554 (2001).
- Shah, N. H. & Muir, T. W. Inteins: nature's gift to protein chemists. *Chem. Sci.* **5**, 446–461 (2014).
- Dassa, B., London, N., Stoddard, B. L., Schueler-Furman, O. & Pietrokovski, S. Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. *Nucleic Acids Res.* **37**, 2560–2573 (2009).
- Stevens, A. J. et al. Design of a split intein with exceptional protein splicing activity. *J. Am. Chem. Soc.* **138**, 2162–2165 (2016).
- Huang, P. S. et al. High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481–485 (2014).
- Shi, J. & Muir, T. W. Development of a tandem protein trans-splicing system based on native and engineered split inteins. *J. Am. Chem. Soc.* **127**, 6198–6206 (2005).
- Newcombe, E. A. et al. Tadpole-like conformations of huntingtin exon 1 are characterized by conformational heterogeneity that persists regardless of polyglutamine length. *J. Mol. Biol.* **430**, 1442–1458 (2018).
- Harmon, T. S., Holehouse, A. S., Rosen, M. K. & Pappu, R. V. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* **6**, e30294 (2017).
- Mittal, A., Das, R. K., Vitalis, A. & Pappu, R. V. ABSINTH implicit solvation model and force field paradigm for use in simulations of intrinsically disordered proteins. in *Computational Approaches to Protein Dynamics: From Quantum to Coarse-Grained Methods* (ed. Fuxreiter, M.) (CRC Press, Boca Raton, 2015).
- Mittal, A., Holehouse, A. S., Cohan, M. C. & Pappu, R. V. Sequence-to-conformation relationships of disordered regions tethered to folded domains of proteins. *J. Mol. Biol.* **430**, 2403–2421 (2018).
- Sherry, K. P., Das, R. K., Pappu, R. V. & Barrick, D. Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor. *Proc. Natl Acad. Sci. USA* **114**, E9243–E9252 (2017).
- Fuertes, G. et al. Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements. *Proc. Natl Acad. Sci. USA* **114**, E6342–E6351 (2017).
- Martin, E. W. et al. Sequence determinants of the conformational properties of an intrinsically disordered protein prior to and upon multisite phosphorylation. *J. Am. Chem. Soc.* **138**, 15323–15335 (2016).
- Das, R. K., Ruff, K. M. & Pappu, R. V. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **32**, 102–112 (2015).
- Mao, H. Y., Hart, S. A., Schink, A. & Pollok, B. A. Sortase-mediated protein ligation: a new method for protein engineering. *J. Am. Chem. Soc.* **126**, 2670–2671 (2004).
- Zakeri, B. et al. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl Acad. Sci. USA* **109**, E690–E697 (2012).
- Kim, D. M. et al. Enhancement of cellulolytic enzyme activity by clustering cellulose binding domains on nanoscaffolds. *Small* **7**, 656–664 (2011).
- Stepankova, V. et al. Strategies for stabilization of enzymes in organic solvents. *ACS Catal.* **3**, 2823–2836 (2013).
- Lopez-Sagaseta, J., Malito, E., Rappuoli, R. & Bottomley, M. J. Self-assembling protein nanoparticles in the design of vaccines. *Comput. Struct. Biotechnol. J.* **14**, 58–68 (2016).
- Carvajal-Vallejos, P., Pallisse, R., Mootz, H. D. & Schmidt, S. R. Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources. *J. Biol. Chem.* **287**, 28686–28696 (2012).
- Bowen, C. H. et al. Recombinant spidroins fully replicate primary mechanical properties of natural spider silk. *Biomacromolecules* **19**, 3853–3860 (2018).
- Engler, C. & Marillonnet, S. Golden Gate cloning. *Methods Mol. Biol.* **1116**, 119–131 (2014).
- Nannenga, B. L., Iadanza, M. G., Vollmar, B. S. & Gonen, T. Overview of electron crystallography of membrane proteins: crystallization and screening strategies using negative stain electron microscopy. *Curr. Protoc. Protein Sci.* **17**, 17.15 (2013).
- Vitalis, A. & Pappu, R. V. ABSINTH: a new continuum solvation model for simulations of polypeptides in aqueous solutions. *J. Comput. Chem.* **30**, 673–699 (2009).
- Vitalis, A. & Pappu, R. V. Methods for Monte Carlo simulations of biomacromolecules. *Annu. Rep. Comput. Chem.* **5**, 49–76 (2009).
- Choi, J. M. & Pappu, R. V. Improvements to the ABSINTH force field for proteins based on experimentally derived amino acid specific backbone conformational statistics. *J. Chem. Theory Comput.* **15**, 1367–1382 (2019).
- Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

Acknowledgements

We thank Prof. David Baker, Lauren Carter, Sydney Gordon, and Gustav Oberdorfer at the University of Washington for providing the 3HB plasmid and production advice. We thank Huaifang Li at the Washington University in St. Louis (WUSTL) Institute of Materials Science and Engineering (IMSE); Remya Nair at the WUSTL Nano Research Facility (NRF); and James Fitzpatrick, Matthew Joens, Gregory Strout, and Daniel Geanon at the Washington University Center for Cellular Imaging (WUCCI) for help with sample preparation and TEM and STEM imaging. We thank Prof. Srikanth Siganamani, Bin Dai, and Hongcheng Sun at WUSTL and Jonathan Galazka at NASA for useful suggestions on this work. We thank Prof. Matthew Lew for providing access to lab equipment for UV-treating the EM grids. STEM and TEM were performed within the facilities of IMSE, NRF, and WUCCI. This work was supported by the Office of Naval Research under the award number N000141512515 (to F.Z.), by a NASA Space Technology Research Fellowship under the award number 80NSSC18K1145 (to C.J.S. and F.Z.), the Human Frontier Science Program under the award number RGP0034/2017 (to R.V.P.), and the US National Science Foundation under the award number MCB-1614766 (to R.V.P.).

Author contributions

F.Z. conceived the project. F.Z., W.B., and C.J.S. designed the experiments. W.B. prepared the nanostructure and performed the characterization. C.J.S. performed some STEM characterization. J.-M.C. and R.V.P. performed the computational simulations. All authors analyzed the data and wrote the paper.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-019-11285-8>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Peer review information: *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019