Construction of a Quadrangular Tetramer and a Cage-Like Hexamer from Three-Helix Bundle-Linked Fusion Proteins

Takaaki Miyamoto,[1†] Yugo Hayashi,[1] Keito Yoshida,[1] Hiroki Watanabe,[2] Takayuki Uchihashi,[2,3] Kento Yonezawa,[4] Nobutaka Shimizu,[4] Hironari Kamikubo,[1,4] and Shun Hirota[1*]

[1]Division of Materials Science, Graduate School of Science and Technology, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan

[2]Exploratory Research Center on Life and Living Systems, Higashiyama, Myodaiji, 444-8787 Okazaki, Aichi, Japan

[3]Department of Physics, Nagoya University, Chikusa-ku, Nagoya, Aichi 464-8602, Japan

[4]Structural Biology Research Center, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki, 305-0801, Japan

ABSTRACT

Self-assembled protein nanostructures have gained interest, owing to their potential applications in biomaterials; however, successful design and construction of protein nanostructures are limited. Herein, we constructed fusion protein **1** by linking the C-terminus of a dimerization domain and the N-terminus of another dimerization domain with a three-helix bundle protein, where it self-assembled mainly into tetramers. By replacing the C-terminal dimerization domain of **1** with a trimerization domain (fusion protein **2**), hexamers were mainly obtained. According to *ab initio* structural models reconstructed from the small angle X-ray scattering data, the tetramer of **1** and hexamer of **2** adopted quadrangle and cage-like structures, respectively, although they were combinations of different conformations. High-speed atomic force microscopy observations indicated that the tetramer and hexamer exhibit conformational dynamics. These results show that the present method utilizing three-helix bundle-linked fusion proteins is useful in the construction of protein nanostructures.

KEYWORDS

fusion protein, high-speed atomic force microscopy, oligomer formation, protein nanostructure, self-assembly, small angle X-ray scattering analysis.

INTRODUCTION

In nature, protein assemblies exhibit diverse functions on the basis of their well-ordered nanoscale structures, including filament, ring, tube, and cage structures.[1] Protein nanostructures are gaining interest in the fields of synthetic biology and nanomedicine. For example, virus capsids have been considered as potential carriers for gene and drug delivery.[2] As well as natural protein nanostructures, artificial protein nanostructures have received attention as tailor-made biomaterials.[3] Recently, artificial protein nanostructures have been constructed by various methods, such as computational interface design,[4-8] metal coordination,[9-17] fusion of oligomeric protein domains,[18-28] etc.[29-41] However, successful constructions of artificial protein nanostructures remain limited.

Computational designs of new protein–protein interfaces have provided tetrahedral,[4] octahedral,[5] and icosahedral protein cages with atomic level accuracy.[6-8] Metal-mediated protein complexes have been achieved in crystals, although they often exhibit dynamic equilibriums among metal coordinated and non-coordinated structures in solution.[9,10,12] Fusion proteins have also been utilized to construct protein oligomer cages.[18,20-22,24,25,27] For an example, Yeates and co-workers constructed a building block protein by fusing two oligomerization domains; the C-terminal α-helix of one domain was connected to the N-terminal α-helix of the other domain with a short α-helical linker.[18,20-22] The relative orientations of the two oligomerization domains were restricted by the elongated α-helices in

the building block proteins, which resulted in the formation of self-assemblies with specific nanostructures, including polymorphic cage-like structures.[18] Later, a homologous tetrahedral protein cage was constructed by reducing the steric hindrance between the trimeric component and α-helical linker,[20,21] whereas a porous cube-shaped protein cage was constructed by a similar method.[22]

Fusion proteins of two oligomerization domains connected with flexible linkers have also been used to construct nanostructures.[23-25] The flexible linker allows a wide variety of oligomerization domains to be used in the building block proteins, regardless of the secondary structures at the N- and C-terminal domains, although the flexibility of the connection between oligomerization domains often causes polydispersion of the assemblies.[42,43] To circumvent the heterogeneous property, Marsh and co-workers constructed a building block protein by fusing C_3- and C_4-symmetric domains with a short flexible linker without controlling the relative orientation between the domains, since only the octahedral structure is available by the combination of C_3- and C_4-symmetric domains.[24] However, limited combinations of symmetric domains, such as the C_3/C_4 combination, are available to construct nanostructures when fusing two oligomerization domains with a flexible linker.[24,25]

In this study, we designed fusion proteins **1** and **2** by linking the C-terminus of an oligomerization domain and the N-terminus of another oligomerization domain with a three-helix bundle protein. Fusion proteins **1** and **2** formed a quadrangular tetramer and a cage-like

hexamer, respectively, demonstrating that the present method is useful in the construction of protein nanostructures.

RESULTS

Design of fusion proteins. Our design strategy utilizes a dimeric protein in which the C-terminal helices of the subunits protrude orthogonally. We envisage that when the C-terminal helix of the subunit of the dimeric protein is directly fused to the N-terminal helix of a three-helix bundle protein, the resulting protein will adopt a rigid V-shaped structure (Figure 1). Connecting the C-terminus of the V-shaped protein to N-termini of other oligomeric proteins will likely result in building block proteins, forming unique nanostructures.

To realize this design strategy, we surveyed the Protein Data Bank (PDB) for a dimeric protein with C-terminal helices that protrude orthogonally. A dimeric protein (the L44C mutant of HP0242 from *Helicobacter pylori*,[44] D_{HP} is used for abbreviation, PDB ID: 3MLI) was selected as the N-terminal domain of a building block protein due to its rigid structure and well-characterized thermodynamic parameters.[44] We also examined the PDB for other components of the building block protein, and selected a three-helix bundle protein (the N-terminal domain of syntaxin-1A from *Rattus norvegicus*,[45] $3HB_{s\text{-}1A}$, PDB ID: 1BR0) and another dimeric protein (Trp repressor from *Escherichia coli*,[46] D_{TR}, PDB ID: 1WRT), owing to their structural stabilities.[45,47] We constructed fusion protein **1** as a building block for a quadrangular tetramer

by genetically linking these three protein domains (Figure 1A). The fusion protein possessed a rigid α-helical junction and a flexible non-structural junction; the C-terminal helix of D_{HP} was connected directly to the N-terminal helix of $3HB_{s\text{-}1A}$, whereas the C-terminus of $3HB_{s\text{-}1A}$ was connected to the N-terminus of D_{TR} with a flexible linker (GSGSG) (Figure 1A, left). We anticipated that **1** would form a quadrangular tetramer, owing to the following properties. (1) The C-terminal helices of two D_{HP} domains protrude orthogonally, owing to the interaction of the remaining parts of the subunits between each other. (2) $3HB_{s\text{-}1A}$ is structurally rigid and assists the separation of D_{HP} and D_{TR} domains. (3) The D_{TR} dimer is stable (it forms a dimer at least up to 65 °C).[47] Thus, two **1** may form a V-shaped dimer by the intermolecular interaction of the D_{HP} domains (Figure 1A, middle), and the two V-shaped dimers assemble into a quadrangular tetramer by the intermolecular interaction through the two D_{TR} domains of each V-shaped dimer (Figure 1A, right).

We extended our approach to construct a cage-like hexamer. Fusion protein **2** was constructed by replacing D_{TR} of **1** with a trimerization domain (a *de novo* designed coiled-coil trimer domain,[48] T_{CC}, PDB ID 4DZL) and genetically linking three protein domains; D_{HP}, $3HB_{s\text{-}1A}$, and T_{CC} (Figure 1B). The N-terminus of T_{CC} was connected to the C-terminus of $3HB_{s\text{-}1A}$ with a flexible four-residue linker (GSGG) (Figure 1B, left). We envisaged that **2** forms a cage-like hexamer by the intermolecular interaction of three V-shaped dimers through the two T_{CC} domains of each V-shaped dimer (Figure 1B, middle and right).

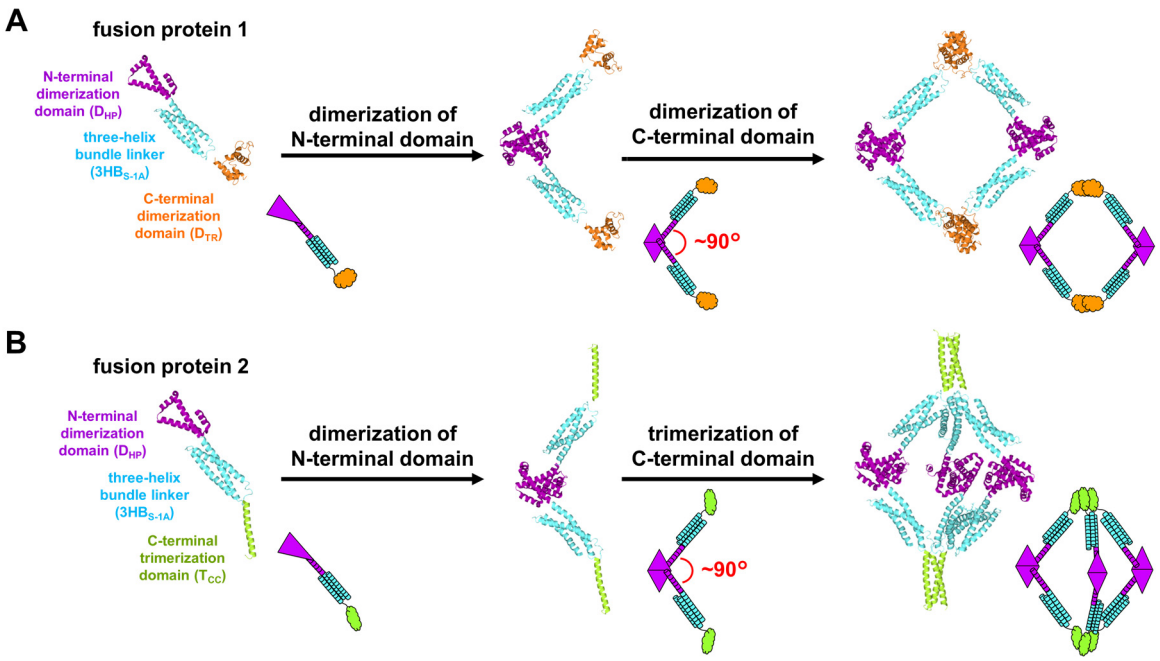


Figure 1. Schematic representation of the components of fusion proteins **1** and **2** and their assemblies. (A) Fusion protein **1** is composed of N-terminal dimerization (D_{HP}, purple), three-helix bundle linker ($3HB_{s-1A}$, cyan), and C-terminal dimerization (D_{TR}, orange) domains, and forms a tetramer. (B) Fusion protein **2** is composed of N-terminal dimerization (D_{HP}, purple), three-helix bundle linker ($3HB_{s-1A}$, cyan), and C-terminal trimerization (T_{CC}, green) domains, and forms a hexamer.

Oligomer formation of fusion proteins. Fusion proteins **1** and **2** with N-terminal His_6 tags (Figure S1) were overexpressed in *E. coli* and obtained as inclusion bodies. Soluble **1** and **2** were obtained by refolding the proteins from inclusion bodies, and purified by immobilized metal affinity chromatography (IMAC). Single bands were detected in the sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of IMAC-purified **1** and **2** (Figure S2), whereas several peaks were observed in their elution curves of size exclusion-fast

protein liquid chromatography (SEC-FPLC) (Figures 2A and 2B), indicating that **1** and **2** form various sizes of homo-oligomers. In the SEC-FPLC elution curve, peaks were observed at 155 and 180 mL in the elution curve of **1**, whereas a broad peak was observed at 110–140 mL, which corresponded to high-order oligomers (Figure 2A, IMAC-purified **1**). Since the peak intensity of the fraction eluting at 155 mL was larger than that eluting at 180 mL, the oligomer eluting at 155 mL was purified further by repeating size exclusion chromatography (SEC) (Figure 2A, main oligomer of **1**). The purified main oligomer eluting at 155 mL did not dissociate at a protein concentration of 25 μM (Figure S3). For the SEC-FPLC elution curve of **2**, three peaks were detected at 110–138, 151, and 178 mL (Figure 2B, IMAC-purified **2**). The 151 mL-peak was the main peak in the elution curve of **2**; thus, we purified the oligomer eluting at 151 mL by repeating SEC (Figure 2B, main oligomer of **2**). Native-PAGE analysis of IMAC-purified **1** and **2** revealed that **1** and **2** form several oligomeric species larger than the main oligomer whereas **1** also exhibits a species smaller than its main oligomer (Figures 2C and 2D).

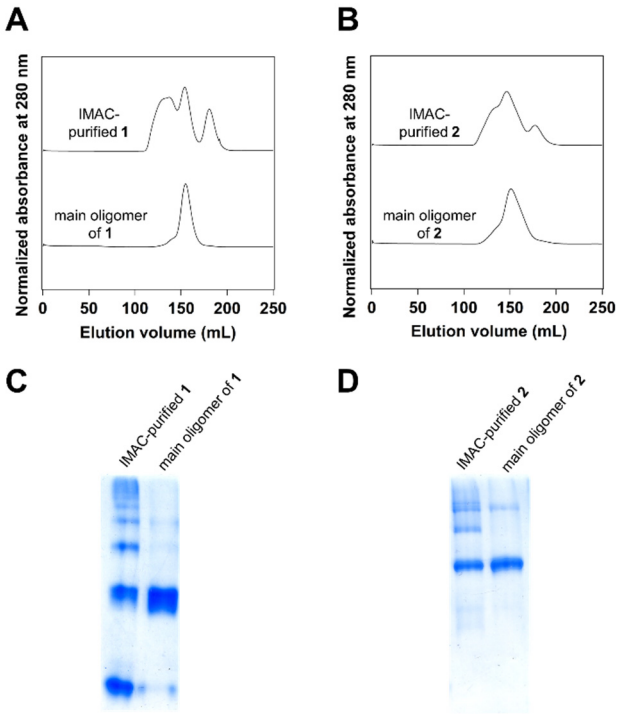


Figure 2. SEC-FPLC and native-PAGE analyses of fusion proteins **1** and **2**. (A,B) Chromatograms of IMAC-purified fusion protein and its SEC-FPLC-purified main oligomer: (A) **1** and (B) **2**. The concentration of the main oligomer was 250 μM. (C,D) Native-PAGE analysis of IMAC-purified fusion protein and its SEC-FPLC-purified main oligomer: (C) **1** and (D) **2**.

The molecular weights of the main oligomers of **1** and **2** were estimated by small angle X-ray scattering (SAXS) analysis. The forward scattering intensities $I(0)$ of the main oligomers were obtained from their Guinier plots (Figures 3A and 3B). $I(0)$/conc., where conc. represents the protein concentration, is proportional to the protein molecular weight. The $I(0)$/conc. values were corrected by the linear extrapolation of the $I(0)$/conc. vs. protein concentration plots to zero protein concentrations (Figures 3C and 3D). We compared the corrected $I(0)$/conc. values

of the tetramer and hexamer with that of a standard protein, ovalbumin, to estimate their apparent molecular weights. Accordingly, the molecular weights of the main oligomers of **1** and **2** were estimated as 177 and 187 kDa, respectively. We also performed size exclusion chromatography-multi-angle light scattering (SEC-MALS) analysis for the main oligomers of **1** and **2** to verify their molecular weights (Figures 3E and 3F). The molecular weights of the main oligomers of **1** and **2** were estimated as 144 and 173 kDa, respectively, from the SEC-MALS analysis. These molecular weights obtained from the SAXS and SEC-MALS analyses corresponded well to those of the tetramer of **1** (150 kDa) and hexamer of **2** (180 kDa), which were calculated from the molecular weights of the fusion proteins (37.4 kDa for **1** and 30 kDa for **2**), supporting the hypothesis that the main oligomers of **1** and **2** are a tetramer and a hexamer, respectively.

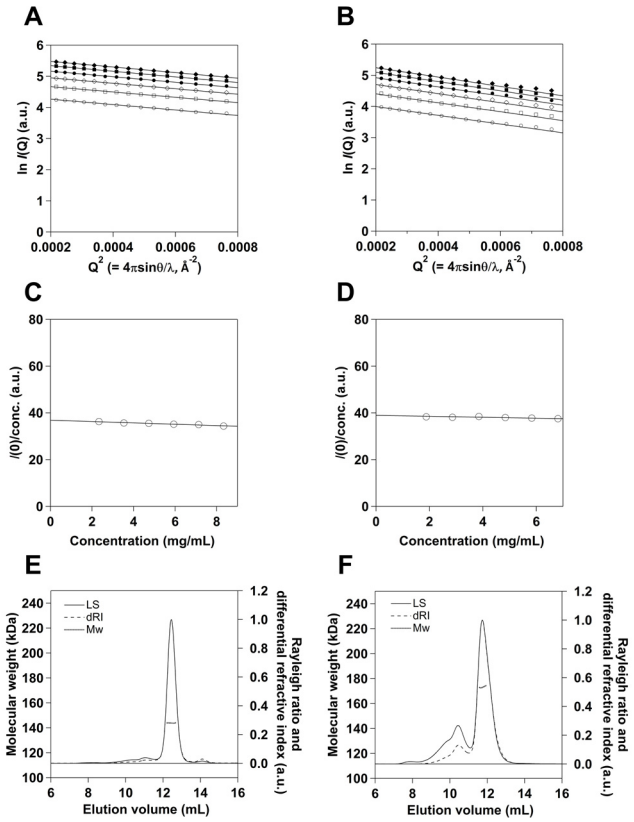


Figure 3. SAXS and SEC-MALS analyses of main oligomers of fusion proteins **1** and **2**. (A) Guinier plots of the main oligomer of **1** at protein concentrations of 2.33 (open circle), 3.53 (open square), 4.73 (open diamond), 5.93 (solid circle), 7.13 (solid square), and 8.33 mg/mL (solid diamond). The least-square fitted line of the data is shown for each plot. (B) Guinier plots of the main oligomer of **2** at protein concentrations of 1.88 (open circle), 2.86 (open square), 3.85 (open diamond), 4.84 (solid circle), 5.83 (solid square), and 6.81 mg/mL (solid diamond). The least-square fitted line of the data points is shown for each plot. (C,D) Protein concentration dependence of $I(0)$/conc. for the main oligomer: (C) **1** and (D) **2**. (E,F) SEC-MALS analysis of the main oligomers: (E) **1** and (F) **2**. Light scattering (LS, solid line), differential reflective index (dRI, dashed line), and the molecular weight of the protein (Mw, dotted line) are plotted against the elution volume.

Structures of the tetramer of fusion protein 1 and hexamer of fusion protein 2. The

structures of the tetramer of **1** and hexamer of **2** were estimated by SAXS analysis. The radii of gyration (R_g) were calculated to be 50.1 ± 1.3 and 64.5 ± 5.9 Å for the tetramer and hexamer, respectively, based on the slope ($-R_g^2/3$) of the Guinier plots (Figures 3A, 3B, S4, and Table S1). We also obtained the maximum dimension (D_{max}), which is defined as the point that the pair-distance distribution function $P(r)$ becomes zero (Figures 4A and 4B). The estimated D_{max} values of the tetramer (200 Å) and hexamer (240 Å) were larger than those estimated from the predicted structures (tetramer, 180 Å; hexamer, 215 Å) (Figure 1), suggesting that the oligomers adopt elongated structures. The *ab initio* bead and rigid body models of the tetramer and hexamer were obtained from the SAXS data using the dummy atom minimization (DAMMIF) program and the complexes with random loops (CORAL) program, respectively (Figures 4C–H). The possible conformations of the tetramer and hexamer exhibited quadrangular and cage-like structures with quasi D_2 and quasi D_3 symmetries, respectively (Figures 4E–H). Although the structures of the *ab initio* models were generally similar to the predicted structures, no hole was observed in the *ab initio* tetramer model, whereas no cavity was observed in the hexamer model (Figures 4E–H).

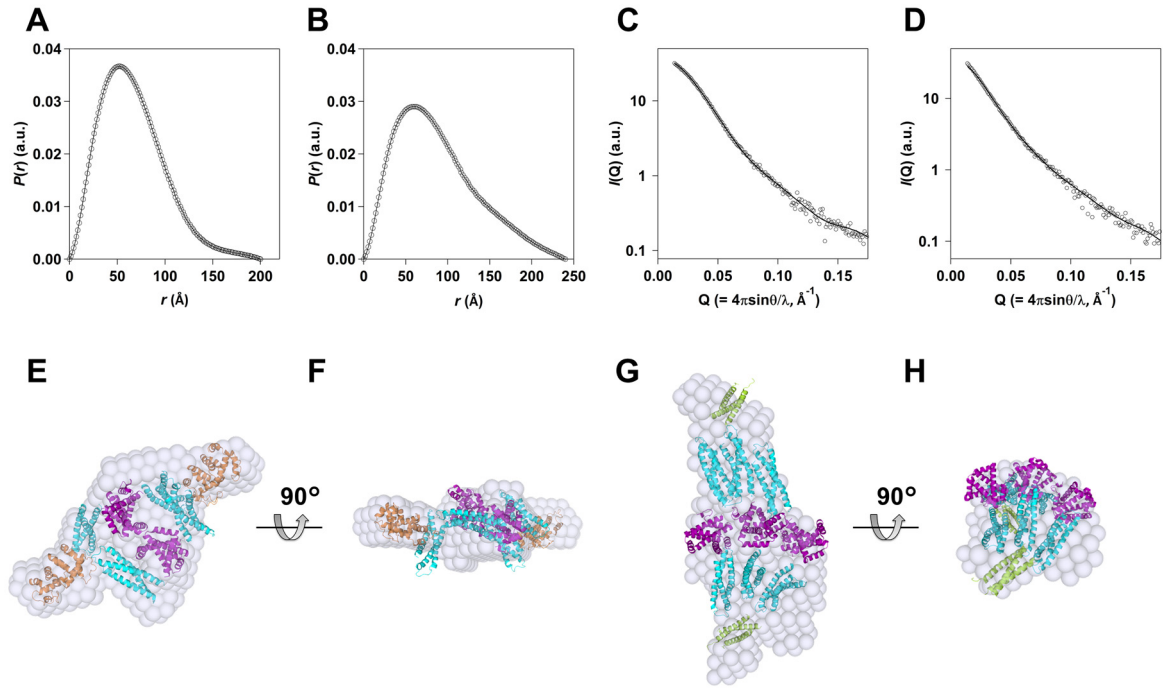


Figure 4. Structural characterizations of the tetramer of **1** and hexamer of **2** by SAXS analysis. (A,B) The pair distance distribution functions obtained from the SAXS scattering curves extrapolated to zero protein concentrations: (A) tetramer and (B) hexamer. (C,D) Fitting of the experimental SAXS scattering curves (open circles) and theoretical curves obtained from the CORAL models (black lines): (C) tetramer and (D) hexamer. (E–H) The *ab initio* bead models (gray) depicted with the CORAL models composed of the N-terminal dimerization (purple), three-helix bundle linker (cyan), and C-terminal dimerization (orange) or trimerization (green) domains (SASBDB ID: tetramer, SASDFR4; hexamer, SASDFS4): (E,F) tetramer and (G,H) hexamer. (F) and (H) are 90°-rotated views of (E) and (G), respectively.

Conformational dynamics of the tetramer of fusion protein 1 and hexamer of fusion protein 2. We performed high-speed atomic force microscopy (HS-AFM) imaging of the tetramer of **1** and hexamer of **2** to investigate their structural arrangements and dynamics (Figure 5, Movies S1 and S2). In the HS-AFM images of the tetramer, two globular units,

which were diagonal to each other, were fixed on the mica substrate modified by (3-aminopropyl)triethoxysilane (APTES mica), whereas the other two globular units were mobile (Figures 5A and 5B). The surface charge of D_{HP} is more negative than that of D_{TR} (Figure S4), and thus D_{HP} was more readily absorbed on the APTES-mica surface with abundant positive charges. We attribute the two globular units absorbed on the substrate to D_{HP}, and the other two mobile globular units to D_{TR}. Although negative charges are partially distributed on the surface of $3HB_{s\text{-}1A}$ (Figure S4), the negatively-charged region may be too small to let $3HB_{s\text{-}1A}$ of the tetramer absorb on the substrate. These observations indicate that the tetramer adopts various conformations, apparently owing to the flexible linker between $3HB_{s\text{-}1A}$ and D_{TR}.

According to the HS-AFM images for the hexamer of **2**, three globular units were fixed on the APTES-mica surface and arranged in a triangle orientation (Figure 5C). The three units may correspond to D_{HP}, since the surface charge of D_{HP} is more negative than those of $3HB_{s\text{-}1A}$ and T_{CC} (Figures 5C, 5D, and S4). In addition to the three units, two obscure units were also detected (Figure 5C, white arrows). We attribute the two obscure units to the moving T_{CC} domains (Figure 5D), since $3HB_{s\text{-}1A}$ was not observed in the HS-AFM images of the tetramer of **1** (Figure 5A), although $3HB_{s\text{-}1A}$ and T_{CC} both exhibit partial negative charges on their surfaces (Figure S4). The flexible GSGG linker between T_{CC} and $3HB_{s\text{-}1A}$ may allow the hexamer to adopt dynamic conformations.

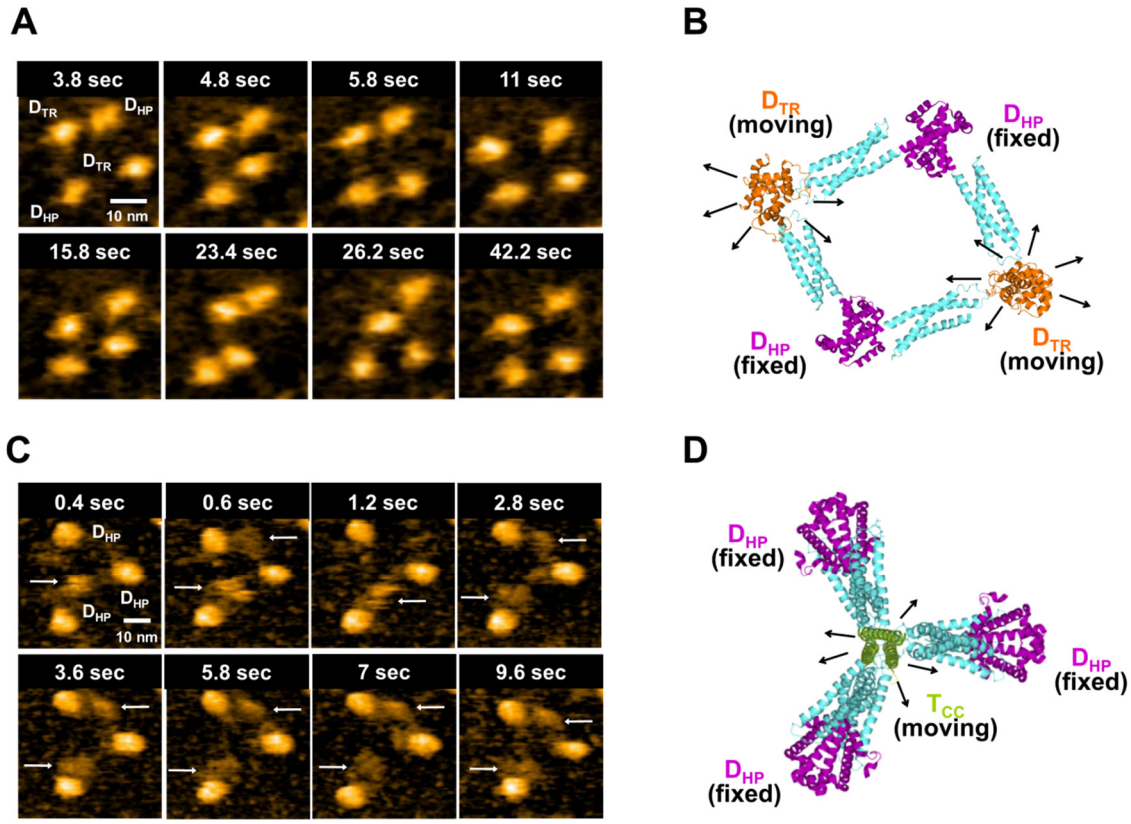


Figure 5. HS-AFM observations of the tetramer of **1** and hexamer of **2** on the APTES-mica substrates. (A,C) Clipped HS-AFM images: (A) tetramer and (C) hexamer. (B,D) Schematic views of the fixed N-terminal dimerization domains (purple) on the substrate and the mobile C-terminal dimerization or trimerization domains (orange or green): (B) tetramer and (D) hexamer. The three-helix bundle linkers are shown in cyan. The white arrows in (C) indicate the C-terminal trimerization domains of the hexamer.

Discussion

In the fusion protein strategy, building block proteins have been constructed by linking two oligomerization domains through a rigid α-helical or a flexible linker.[49] The oligomerization domain is required to possess an α-helical terminus when a rigid α-helical linker is attached to

the domain, whereas specific combinations of symmetry elements of oligomerization domains are necessary when a flexible linker is used. These limitations hamper the design of diverse nanostructures. We utilized an α-helical linker and a flexible linker to fuse three domains; an N-terminal dimerization domain, a three helix bundle domain, and a C-terminal oligomerization domain (Figure 1). Fusion protein **1** with a C-terminal dimerization domain formed a quadrangle (Figures 4E and 4F), whereas fusion protein **2** with a C-terminal trimerization domain formed a hexameric cage-like structure (Figures 4G and 4H), demonstrating that a simple alteration of the C-terminal domain of the fusion protein controls the oligomer structure. A wide variety of oligomerization domains is available at the C-terminal region of the fusion protein, resulting in unique protein nanostructures.

Various sizes of oligomers were obtained by refolding fusion proteins **1** and **2** from inclusion bodies (Figure 2), indicating that the present design strategy needs improvement. Various sizes of oligomers formed, apparently due to the flexibility of the linker connecting the three-helix bundle domain and the C-terminal oligomerization domain in the fusion proteins. Marsh and co-workers have shown that the length of the flexible linker between the two oligomerization domains affects the oligomer formation;[24,25] thus, a more selective formation of oligomers for **1** and **2** may be achieved by tuning the length of the flexible linker. Additionally, the stability of the presumed helical region connecting the N-terminal and three-helix bundle domains may decrease. For example, the bending of a helical linker of a fusion

protein induces polymorphism in the assemblies.[22,50] Recently, a chemical cross linker has been used to stabilize a helical linker that connects two protein components,[51] indicating the improvement in the oligomer homogeneity by combining the three-helix bundle-linking approach with the helix-stabilizing cross linker method.

The molecular weights of the main oligomers of **1** (177 kDa) and **2** (187 kDa), which were estimated from the *I*(0)/conc. values (Figure 3), were consistent with those of the tetramer of **1** (150 kDa) and hexamer of **2** (180 kDa), respectively. However, the molecular weights estimated as 215 and 245 kDa for the main oligomers of **1** and **2** from the elution volumes in the SEC chromatograms (Figure 2) were larger than those of the tetramer and hexamer, respectively. These discrepancies imply elongated structures for the tetramer and hexamer (Figures 4E–H). The hydrodynamic radius (*i.e.*, Stokes radius) of a protein with an elongated structure is larger than that of a globular protein, and the elongated-structural protein elutes earlier in SEC compared to a globular protein when the molecular weights are the same.[52] Thus, the tetramer and hexamer may have eluted earlier than standard globular proteins with similar molecular weights, resulting in the overestimation of their molecular weights by SEC.

The *ab initio* models reconstructed from the SAXS data indicated that the tetramer of **1** and hexamer of **2** formed closed structures. However, no hole was observed in the *ab initio* model of the tetramer of **1** (Figures 4E and 4F), and no cavity was observed in the model of the hexamer of **2** (Figures 4G and 4H). The HS-AFM observation indicated that the tetramer and

hexamer adopt various conformations (Figure 5), apparently due to the flexibility of the linker between the three-helix bundle and C-terminal domains. The *ab initio* models obtained by the SAXS data for the tetramer and hexamer are average structures with various conformations. Since conformational dynamics are essential for biological functions in many natural protein nanostructures, such as ATP synthase,[53] thermosome,[54] RNA polymerase II,[55] and protease-activating nucleotidase,[56] artificial protein nanostructures with conformational dynamics may be useful to develop functional biomaterials. However, structural flexibility leads to structural polymorphism as observed in the tetramer of **1** and hexamer of **2**, and it would be necessary to control the structural flexibility in the protein nanostructures maintaining the conformational dynamics for future oligomer design studies.

Conclusions

In the present study, we designed two fusion proteins **1** and **2**, comprising an N-terminal dimerization domain, a three-helix bundle linker domain, and a C-terminal oligomerization domain. According to the *ab initio* models reconstructed from the SAXS data, the tetramer of **1**, possessing a C-terminal dimerization domain, formed a quadrangle. The hexamer of **2**, possessing a C-terminal trimerization domain, formed a cage-like structure. However, the tetramer and hexamer of the fusion proteins revealed conformational dynamics, according to the HS-AFM observations. The present method of fusion proteins linking oligomerization

protein domains with a three-helix bundle protein may expand the variety of protein nanostructures.

METHODS

Construction of expression systems for fusion proteins 1 and 2. DNA fragments encoding fusion proteins **1** and **2** (Figure S1) were synthesized (Life Technologies Japan, Tokyo) and inserted into pET15b vectors using an In-Fusion HD cloning kit (Takara). The amino acid sequences of **1** and **2** were checked by DNA sequencing with a BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems, Inc., Foster City, CA) and an ABI PRISM 3100 genetic analyzer sequencing system (Applied Biosystems, Inc.). The obtained plasmids were transformed into *E. coli* Rosetta2 (DE3) (Novagen).

Preparation of the tetramer of fusion protein 1 and hexamer of fusion protein 2. *E. coli* Rosetta2 (DE3) cells overproducing fusion proteins **1** and **2** were grown at 37 °C in LB broth containing ampicillin (100 μg/mL) and chloramphenicol (30 μg/mL). Isopropyl β-D-1-thiogalactopyranoside was added to the culture (final concentration, 0.5 mM) when the OD_{600} value reached 0.6. The culture was incubated at 37 °C for 4 h, and harvested by centrifugation. The cell pellet was suspended in a minimal volume of 25 mM tris(hydroxymethyl)aminomethane-hydrochloric acid (Tris-HCl) buffer (pH 8.0) containing 150 mM sodium chloride (NaCl), and lysed by freeze-thaw and sonication. After the

centrifugation of the cell lysate, the fusion proteins were obtained as inclusion bodies. The inclusion body was solubilized with 25 mM Tris buffer (pH 8.0) containing 6 M guanidine hydrochloride, 0.1 M L-arginine, 0.1 M dithiothreitol, 0.01 M ethylene diamine tetraacetic acid, and 0.5 M NaCl. The refolding of the fusion protein was performed by dialysis in 25 mM Tris buffer (pH 8.0) containing 0.2 M NaCl. After the dialysis, the protein solution was purified by IMAC with a HisTrap HP column (GE Healthcare) using a FPLC system (AKTA prime plus, GE Healthcare) (flow rate, 3.0 mL/min; monitoring wavelength, 280 nm; solvent, 25 mM Tris-HCl buffer (pH 8.0) containing 0.5 M NaCl and 0–0.5 M imidazole; temperature, 4 °C). The tetramer of **1** and hexamer of **2** were purified by repeating SEC (HiLoad 26/60 Superdex200, GE healthcare, Buckinghamshire) using the FPLC system (AKTA prime plus, GE Healthcare) (flow rate, 2.5 mL/min; monitoring wavelength, 280 nm; solvent, 25 mM Tris-HCl buffer (pH 8.0) containing 150 mM NaCl; temperature, 4 °C). The purities of the fusion proteins were checked by SDS-PAGE analysis. The protein concentration was determined by the Bradford method.[57] Native-PAGE analysis of IMAC-purified fusion proteins and their SEC-FPLC-purified main oligomers were performed with a 7.5% acrylamide gel and by staining the proteins with Coomassie Brilliant Blue.

Small angle X-ray scattering measurement. The purified tetramer of **1** and hexamer of **2** were prepared in 25 mM Tris-HCl (pH 8.0) containing 150 mM NaCl for SAXS measurements. SAXS measurements were carried out using a Nano-Viewer diffractometer

system equipped with a MicroMax-007 X-ray generator (RIGAKU), possessing a Cu target (λ = 1.5418 Å) and a PILATUS 200 K detector (DECTRIS) (Table S1). The exposure time was 25 min for each measurement. Scattering intensities, I(Q), of the tetramer and hexamer were collected at various protein concentrations in the momentum transfer (Q) range of 0.012–0.25 $Å^{-1}$, where Q is expressed as $4\pi\sin\theta / \lambda$, 2θ is the scattering angle, and λ is the X-ray wavelength. The background scattering intensity of the buffer solution was subtracted from those of the tetramer and hexamer. The scattering intensities were used to generate Guinier plots, in which the natural logarithm of scattering intensity, ln I(Q), was plotted against squared momentum transfer (Q^2). The I(0) values of the tetramer and hexamer were obtained by the linear extrapolation of Guinier plots to zero angle. The I(0) value normalized by protein concentration, I(0)/conc., is proportional to the protein molecular weight, but varies with the protein concentration. To avoid the effect of protein concentration, the I(0)/conc. values of the tetramer and hexamer were plotted against the protein concentrations and corrected by the linear extrapolation of the plots to zero protein concentrations (Figures 3C and 3D, Table S1). The extrapolated I(0)/conc. values of the tetramer and hexamer were compared with that of a standard protein, ovalbumin (45 kDa; Sigma-Aldrich), to estimate their apparent molecular weights. The R_g value was obtained using the relation in which the slope of the Guinier plot corresponds to $-R_g^2/3$ at low Q regions. Since the R_g^2 value depends on the protein concentration, the R_g^2 values of the tetramer and hexamer were corrected by the linear

extrapolation of their R_g^2 vs. protein concentration plots to zero protein concentrations (Figure S5). The pair-distance distribution function, $P(r)$, of the tetramer and hexamer were calculated from the scattering intensities using the GNOM software.[58,59] The D_{max} values of the tetramer and hexamer were obtained from their $P(r)$ values. *Ab initio* bead modelling was performed using the DAMMIF software without structural restrictions,[60] such as point symmetry and particle anisometry. The 19 independently calculated models were averaged using the DAMAVER software.[61] Using the average model as a starting model, we refined the bead model using the DAMMIN software.[62] The rigid body models of the tetramer and hexamer were obtained from the CORAL program.[58] The initial models were constructed from the crystal structures of the N-terminal dimerization (PDB ID: 3MLI), three-helix bundle linker (PDB ID: 1BR0), and C-terminal dimerization (1WRT) domains for the quadrangular tetramer, and from the N-terminal dimerization (PDB ID: 3MLI), three-helix bundle linker (PDB ID: 1BR0), and C-terminal trimerization (PDB ID 4DZL) domains for the cage-like hexamer. The positions of the three-helix bundle linker and N-terminal domains were firstly optimized; subsequently, setting the obtained positions of the three-helix bundle linker and N-terminal domains, the positions of the C-terminal domains were optimized until converged.

Size-exclusion chromatography-multi-angle static light scattering. SEC-MALS analysis of the main oligomers of **1** and **2** were performed with a Superdex 200 increase 10/300 GL column (GE Healthcare) using an Alliance 2695 high performance liquid chromatography

system (Waters). Light scattering and refraction index were measured using a DAWN HELEOS II detector (Wyatt Technology) and a 2414 Refractive Index (RI) detector (Waters), respectively. The column was equilibrated at 20 ºC with 25 mM Tris-HCl buffer, pH 8.0, containing 150 mM NaCl. Samples (30 µl) were injected under the buffer flow rate of 0.6 ml/min. The obtained data were recorded and processed using the ASTRA 6.1 software (Wyatt Technologies).

High-speed atomic force microscopy observation. AFM images were capture by a laboratory-built HS-AFM operated with tapping mode. For the HS-AFM imaging, a small cantilever with dimensions of 7-µm long, 2-µm wide, and 90-nm thick was used (BL-AC7, Olympus). Its nominal spring constant and resonant frequency were ~0.2 N/m and ~800 kHz, respectively, in a liquid. An amorphous carbon tip was deposited on the original bird-beak tip using electron beam deposition (EBD), and then sharpened by a plasma etching in an argon environment. The typical radius of the EBD tip was reduced to approximately 2 nm after the sharpening. For the tapping-mode imaging, the cantilever was oscillated with an amplitude of about 1 nm for the free oscillation, and the amplitude was reduced to be ~ 90 % of the free oscillation amplitude for a feedback control. As a sample substrate, we used mica surface treated by (3-aminopropyl)triethoxysilane (APTES, Shin-Etsu Chemical) with volume concentrations of 0.1 % and 0.03% for the observations of tetramer and hexamer, respectively. After cleavage of a mica sheet 1.5 mm in a diameter, APTES with a concentration of 0.1% was deposited on the mica surface. After 3-min incubation, the mica surface was thoroughly washed

with pure water. The solution containing the tetramer of **1** (3.8 nM) or hexamer of **2** (46.8 nM) was deposited on the APTES-mica and incubated for 3 min. After the incubation, the residual molecules were thoroughly washed with an observation buffer (20 mM Tris-HCl (pH 8.0) containing 150 mM NaCl). After the washing, the tip was approached, and the HS-AFM imaging was performed under the buffer solution. The HS-AFM images were taken at a frame rate of 10 fps.

ASSOCIATED CONTENT

Supporting Information

Amino acid sequences and 3D model structures of fusion proteins **1** and **2**; SDS-PAGE analysis results of **1** and **2**; Size exclusion chromatograms of the main oligomers of **1** and **2**; Surface charge distribution comparison of the domains of **1** and **2**; HS-AFM movies of the tetramer of **1** and hexamer of **2**; SAXS data collection and analyzed parameters of the tetramer of **1** and hexamer of **2**; Figures S1–5; Movies S1 and S2; Table S1.

AUTHOR INFORMATION

Corresponding Author

*Phone: +81-743-72-6110. Fax: +81-743-72-6119. E-mail: hirota@ms.naist.jp.

Present Address

†T. Miyamoto: Biomacromolecules Research Team, RIKEN Center for Sustainable Resource Science, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan.

Author Contributions

T.M. and S.H. conceived and designed the research. T.M. designed and prepared the fusion proteins. T.M. performed the biochemical experiments. Y.H., K.Yoshida, and H.K. performed the SAXS experiments. H.W. and T.U. performed the HS-AFM experiments. K.Yonezawa and N.S. performed the SEC-MALS experiments. All the authors analyzed the data. T.M. and S.H. wrote the manuscript with help of Y.H., H.W., T.U., K. Yonezawa, H.K., and N.S.

Notes

The authors declare no competing financial interest.

ACKOWLEDGMENT

We are grateful to Prof. Leigh McDowell, Nara Institute of Science and Technology, for his advice on manuscript preparation. The preliminary SAXS experiments were performed under the approval of the Photon Factory Program Advisory Committee (Proposal No. 2016G077 and No. 2018G119). This work was partially supported by Grants-in-Aid from JSPS for Scientific Research (Category B, No. JP18H02088 (S.H.); Category S, No. JP17H06165 (H.K.); Innovative Areas, No. JP16H00839 (S.H.), No. JP16H00830 (T.U.), and No. JP18H04512 (T.U.)) and from the Japan Science Society for Sasakawa Scientific Research Grant (No. 28-

327 (T.M.)).

REFERENCES

(1) Pieters, B. J. G. E., van Eldijk, M. B., Nolte, R. J. M., and Mecinović, J. (2016) Natural supramolecular protein assemblies. *Chem. Soc. Rev. 45*, 24–39.

(2) Sahandi-Zangabad, P., Karimi, M., Mehdizadeh, F., Malekzad, H., Ghasemi, A., Bahrami, S., Zare, H., Moghoofei, M., Hekmatmanesh, A., and Hamblin, M. R. (2017) Nanocaged platforms: modification, drug delivery and nanotoxicity. Opening synthetic cages to release the tiger. *Nanoscale 9*, 1356–1392.

(3) Sun, H., Luo, Q., Hou, C., and Liu, J. (2017) Nanostructures based on protein self-assembly: From hierarchical construction to bioinspired materials. *Nano Today 14*, 16–41.

(4) King, N. P., Sheffler, W., Sawaya, M. R., Vollmar, B. S., Sumida, J. P., Andre, I., Gonen, T., Yeates, T. O., and Baker, D. (2012) Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science 336*, 1171–1174.

(5) King, N. P., Bale, J. B., Sheffler, W., McNamara, D. E., Gonen, S., Gonen, T., Yeates, T. O., and Baker, D. (2014) Accurate design of co-assembling multi-component protein nanomaterials. *Nature 510*, 103–108.

(6) Bale, J. B., Gonen, S., Liu, Y., Sheffler, W., Ellis, D., Thomas, C., Cascio, D., Yeates, T. O., Gonen, T., King, N. P., and Baker, D. (2016) Accurate design of megadalton-scale two-

component icosahedral protein complexes. *Science 353*, 389–394.

(7) Hsia, Y., Bale, J. B., Gonen, S., Shi, D., Sheffler, W., Fong, K. K., Nattermann, U., Xu, C., Huang, P. S., Ravichandran, R., Yi, S., Davis, T. N., Gonen, T., King, N. P., and Baker, D. (2016) Design of a hyperstable 60-subunit protein icosahedron. *Nature 535*, 136–139.

(8) Butterfield, G. L., Lajoie, M. J., Gustafson, H. H., Sellers, D. L., Nattermann, U., Ellis, D., Bale, J. B., Ke, S., Lenz, G. H., Yehdego, A., Ravichandran, R., Pun, S. H., King, N. P., and Baker, D. (2017) Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature 552*, 415–420.

(9) Salgado, E. N., Faraone-Mennella, J., and Tezcan, F. A. (2007) Controlling protein−protein interactions through metal coordination: Assembly of a 16-helix bundle protein. *J. Am. Chem. Soc. 129*, 13374–13375.

(10) Salgado, E. N., Lewis, R. A., Faraone-Mennella, J., and Tezcan, F. A. (2008) Metal-mediated self-assembly of protein superstructures: Influence of secondary interactions on protein oligomerization and aggregation. *J. Am. Chem. Soc. 130*, 6082–6084.

(11) Biswas, S., Kinbara, K., Oya, N., Ishii, N., Taguchi, H., and Aida, T. (2009) A tubular biocontainer: Metal ion-induced 1D assembly of a molecularly engineered chaperonin. *J. Am. Chem. Soc. 131*, 7556–7557.

(12) Salgado, E. N., Lewis, R. A., Mossin, S., Rheingold, A. L., and Tezcan, F. A. (2009) Control of protein oligomerization symmetry by metal coordination: C_2 and C_3 symmetrical

assemblies through Cu^{II} and Ni^{II} coordination. *Inorg. Chem. 48*, 2726–2728.

(13) Ni, T. W., and Tezcan, F. A. (2010) Structural characterization of a microperoxidase inside a metal-directed protein cage. *Angew. Chem. Int. Ed. 49*, 7014–7018.

(14) Brodin, J. D., Ambroggio, X. I., Tang, C., Parent, K. N., Baker, T. S., and Tezcan, F. A. (2012) Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem. 4*, 375–382.

(15) Bai, Y., Luo, Q., Zhang, W., Miao, L., Xu, J., Li, H., and Liu, J. (2013) Highly ordered protein nanorings designed by accurate control of glutathione S-transferase self-assembly. *J. Am. Chem. Soc. 135*, 10966–10969.

(16) Biswas, S., Kinbara, K., Niwa, T., Taguchi, H., Ishii, N., Watanabe, S., Miyata, K., Kataoka, K., and Aida, T. (2013) Biomolecular robotics for chemomechanically driven guest delivery fuelled by intracellular ATP. *Nat. Chem. 5*, 613–620.

(17) Suzuki, Y., Cardone, G., Restrepo, D., Zavattieri, P. D., Baker, T. S., and Tezcan, F. A. (2016) Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals. *Nature 533*, 369–373.

(18) Padilla, J. E., Colovos, C., and Yeates, T. O. (2001) Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl. Acad. Sci. U. S. A. 98*, 2217–2221.

(19) Sinclair, J. C., Davies, K. M., Vénien-Bryan, C., and Noble, M. E. M. (2011) Generation

of protein lattices by fusing proteins with matching rotational symmetry. *Nat. Nanotechnol. 6*, 558–562.

(20) Lai, Y. T., Cascio, D., and Yeates, T. O. (2012) Structure of a 16-nm cage designed by using protein oligomers. *Science 336*, 1129–1129.

(21) Lai, Y. T., Tsai, K. L., Sawaya, M. R., Asturias, F. J., and Yeates, T. O. (2013) Structure and flexibility of nanoscale protein cages designed by symmetric self-assembly. *J. Am. Chem. Soc. 135*, 7738–7743.

(22) Lai, Y. T., Reading, E., Hura, G. L., Tsai, K. L., Laganowsky, A., Asturias, F. J., Tainer, J. A., Robinson, C. V., and Yeates, T. O. (2014) Structure of a designed protein cage that self-assembles into a highly porous cube. *Nat. Chem. 6*, 1065–1071.

(23) Kobayashi, N., Yanase, K., Sato, T., Unzai, S., Hecht, M. H., and Arai, R. (2015) Self-assembling nano-architectures created from a protein nano-building block using an intermolecularly folded dimeric *de novo* protein. *J. Am. Chem. Soc. 137*, 11285–11293.

(24) Sciore, A., Su, M., Koldewey, P., Eschweiler, J. D., Diffley, K. A., Linhares, B. M., Ruotolo, B. T., Bardwell, J. C. A., Skiniotis, G., and Marsh, E. N. G. (2016) Flexible, symmetry-directed approach to assembling protein cages. *Proc. Natl. Acad. Sci. U. S. A. 113*, 8681–8686.

(25) Badieyan, S., Sciore, A., Eschweiler, J. D., Koldewey, P., Cristie-David, A. S., Ruotolo, B. T., Bardwell, J. C. A., Su, M., and Marsh, E. N. G. (2017) Symmetry-directed self-assembly

of a tetrahedral protein cage mediated by *de novo*-designed coiled coils. *ChemBioChem 18*, 1888–1892.

(26) Cristie-David, A. S., Koldewey, P., Meinen, B. A., Bardwell, J. C. A., and Marsh, E. N. G. (2018) Elaborating a coiled-coil-assembled octahedral protein cage with additional protein domains. *Protein Sci. 27*, 1893–1900.

(27) Kawakami, N., Kondo, H., Matsuzawa, Y., Hayasaka, K., Nasu, E., Sasahara, K., Arai, R., and Miyamoto, K. (2018) Design of hollow protein nanoparticles with modifiable interior and exterior surfaces. *Angew. Chem. Int. Ed. 57*, 12400–12404.

(28) Kobayashi, N., Inano, K., Sasahara, K., Sato, T., Miyazawa, K., Fukuma, T., Hecht, M. H., Song, C., Murata, K., and Arai, R. (2018) Self-assembling supramolecular nanostructures constructed from *de novo* extender protein nanobuilding blocks. *ACS Synth. Biol. 7*, 1381–1394.

(29) Ballister, E. R., Lai, A. H., Zuckermann, R. N., Cheng, Y., and Mougous, J. D. (2008) *In vitro* self-assembly of tailorable nanotubes from a simple protein building block. *Proc. Natl. Acad. Sci. U. S. A. 105*, 3733–3738.

(30) Chou, T. F., So, C., White, B. R., Carlson, J. C. T., Sarikaya, M., and Wagner, C. R. (2008) Enzyme nanorings. *ACS Nano 2*, 2519–2525.

(31) Kitagishi, H., Kakikura, Y., Yamaguchi, H., Oohora, K., Harada, A., and Hayashi, T. (2009) Self-assembly of one- and two-dimensional hemoprotein systems by polymerization

through heme–heme pocket interactions. *Angew. Chem. Int. Ed. 121*, 1297–1300.

(32) Matsuura, K., Watanabe, K., Matsuzaki, T., Sakurai, K., and Kimizuka, N. (2010) Self-assembled synthetic viral capsids from a 24-mer viral peptide fragment. *Angew. Chem. Int. Ed. 49*, 9662–9665.

(33) Lanci, C. J., MacDermaid, C. M., Kang, S.-g., Acharya, R., North, B., Yang, X., Qiu, X. J., DeGrado, W. F., and Saven, J. G. (2012) Computational design of a protein crystal. *Proc. Natl. Acad. Sci. U. S. A. 109*, 7304–7309.

(34) Gradišar, H., Božič, S., Doles, T., Vengust, D., Hafner-Bratkovič, I., Mertelj, A., Webb, B., Šali, A., Klavžar, S., and Jerala, R. (2013) Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nat. Chem. Biol. 9*, 362–366.

(35) Miyamoto, T., Kuribayashi, M., Nagao, S., Shomura, Y., Higuchi, Y., and Hirota, S. (2015) Domain-swapped cytochrome *cb*$_{562}$ dimer and its nanocage encapsulating a Zn–SO_4 cluster in the internal cavity. *Chem. Sci. 6*, 7336–7342.

(36) Jorda, J., Leibly, D. J., Thompson, M. C., and Yeates, T. O. (2016) Structure of a novel 13 nm dodecahedral nanocage assembled from a redesigned bacterial microcompartment shell protein. *Chem. Commun. 52*, 5041–5044.

(37) Ross, J. F., Bridges, A., Fletcher, J. M., Shoemark, D., Alibhai, D., Bray, H. E. V., Beesley, J. L., Dawson, W. M., Hodgson, L. R., Mantell, J., Verkade, P., Edge, C. M., Sessions, R. B., Tew, D., and Woolfson, D. N. (2017) Decorating self-assembled peptide cages with

proteins. *ACS Nano 11*, 7901–7914.

(38) Yang, G., Ding, H.-m., Kochovski, Z., Hu, R., Lu, Y., Ma, Y.-q., Chen, G., and Jiang, M. (2017) 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. 56*, 10691–10695.

(39) Oohora, K., Onuma, Y., Tanaka, Y., Onoda, A., and Hayashi, T. (2017) A supramolecular assembly based on an engineered hemoprotein exhibiting a thermal stimulus-driven conversion to a new distinct supramolecular structure. *Chem. Commun. 53*, 6879–6882.

(40) Oda, A., Nagao, S., Yamanaka, M., Ueda, I., Watanabe, H., Uchihashi, T., Shibata, N., Higuchi, Y., and Hirota, S. (2018) Construction of a triangle-shaped trimer and a tetrahedron using an α-helix-inserted circular permutant of cytochrome c_{555}. *Chem. Asian J. 13*, 964–967.

(41) Wilson, C. J., Bommarius, A. S., Champion, J. A., Chernoff, Y. O., Lynn, D. G., Paravastu, A. K., Liang, C., Hsieh, M.-C., and Heemstra, J. M. (2018) Biomolecular assemblies: moving from observation to predictive design. *Chem. Rev. 118*, 11519–11574.

(42) Patterson, D. P., Su, M., Franzmann, T. M., Sciore, A., Skiniotis, G., and Marsh, E. N. G. (2014) Characterization of a highly flexible self-assembling protein system designed to form nanocages. *Protein Sci. 23*, 190–199.

(43) Patterson, D. P., Desai, A. M., Holl, M. M. B., and Marsh, E. N. G. (2011) Evaluation of

a symmetry-based strategy for assembling protein complexes. *RSC Advances 1*, 1004–1012.

(44) King, N. P., Jacobitz, A. W., Sawaya, M. R., Goldschmidt, L., and Yeates, T. O. (2010) Structure and folding of a designed knotted protein. *Proc. Natl. Acad. Sci. U. S. A. 107*, 20732–20737.

(45) Fernandez, I., Ubach, J., Dulubova, I., Zhang, X. Y., Sudhof, T. C., and Rizo, J. (1998) Three-dimensional structure of an evolutionarily conserved N-terminal domain of syntaxin 1A. *Cell 94*, 841–849.

(46) Zhao, D., Arrowsmith, C. H., Jia, X., and Jardetzky, O. (1993) Refined solution structures of the *Escherichia coli* Trp holo- and aporepressor. *J. Mol. Biol. 229*, 735–746.

(47) Gloss, L. M., and Matthews, C. R. (1997) Urea and thermal equilibrium denaturation studies on the dimerization domain of *Escherichia coli* Trp repressor. *Biochemistry 36*, 5612–5623.

(48) Fletcher, J. M., Boyle, A. L., Bruning, M., Bartlett, G. J., Vincent, T. L., Zaccai, N. R., Armstrong, C. T., Bromley, E. H. C., Booth, P. J., Brady, R. L., Thomson, A. R., and Woolfson, D. N. (2012) A basis set of *de novo* coiled-coil peptide oligomers for rational protein design and synthetic biology. *ACS Synth. Biol. 1*, 240–250.

(49) Yeates, T. O. (2017) Geometric principles for designing highly symmetric self-assembling protein nanomaterials. *Annu. Rev. Biophys. 46*, 23–42.

(50) Lai, Y. T., Jiang, L., Chen, W., and Yeates, T. O. (2015) On the predictability of the

orientation of protein domains joined by a spanning alpha-helical linker. *Protein Eng. Des. Sel. 28*, 491–499.

(51) Jeong, W. H., Lee, H., Song, D. H., Eom, J. H., Kim, S. C., Lee, H. S., Lee, H., and Lee, J. O. (2016) Connecting two proteins using a fusion alpha helix stabilized by a chemical cross linker. *Nat. Commun. 7*, 1–9.

(52) Irvine, G. B. (2000) Determination of molecular size by size-exclusion chromatography (gel filtration). *Curr. Protoc. Cell Biol. 6*, 5.5.1–5.5.16.

(53) Stock, D. (1999) Molecular architecture of the rotary motor in ATP synthase. *Science 286*, 1700–1705.

(54) Ditzel, L., Löwe, J., Stock, D., Stetter, K. O., Huber, H., Huber, R., and Steinbacher, S. (1998) Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. *Cell 93*, 125–138.

(55) Cramer, P. (2001) Structural basis of transcription: RNA polymerase II at 2.8 Ångstrom resolution. *Science 292*, 1863–1876.

(56) Latham, M. P., Sekhar, A., and Kay, L. E. (2014) Understanding the mechanism of proteasome 20S core particle gating. *Proc. Natl. Acad. Sci. U. S. A. 111*, 5532–5537.

(57) Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem. 72*, 248–254.

(58) Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V., and Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr. 45*, 342–350.

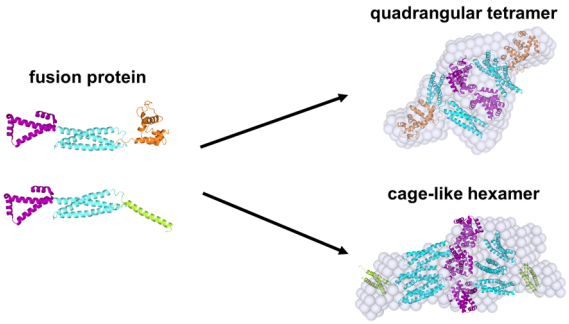
(59) Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr. 25*, 495–503.

(60) Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering. *J. Appl. Crystallogr. 42*, 342–346.

(61) Volkov, V. V., and Svergun, D. I. (2003) Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr. 36*, 860–864.

(62) Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J. 76*, 2879–2886.

TOC



Supporting Information

Construction of a Quadrangular Tetramer and a Cage-Like Hexamer from Three-Helix Bundle-Linked Fusion Proteins

Takaaki Miyamoto,[1] Yugo Hayashi,[1] Keito Yoshida,[1] Hiroki Watanabe,[2] Takayuki Uchihashi,[2,3] Kento Yonezawa,[4] Nobutaka Shimizu,[4] Hironari Kamikubo,[1,4] and Shun Hirota[1*]

[1]Division of Materials Science, Graduate School of Science and Technology, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan
[2]Exploratory Research Center on Life and Living Systems, Higashiyama, Myodaiji, Okazaki, Aichi 444-8787, Japan
[3]Department of Physics, Nagoya University, Chikusa-ku, Nagoya, Aichi 464-8602, Japan
[4]Structural Biology Research Center, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki, 305-0801, Japan

Contents

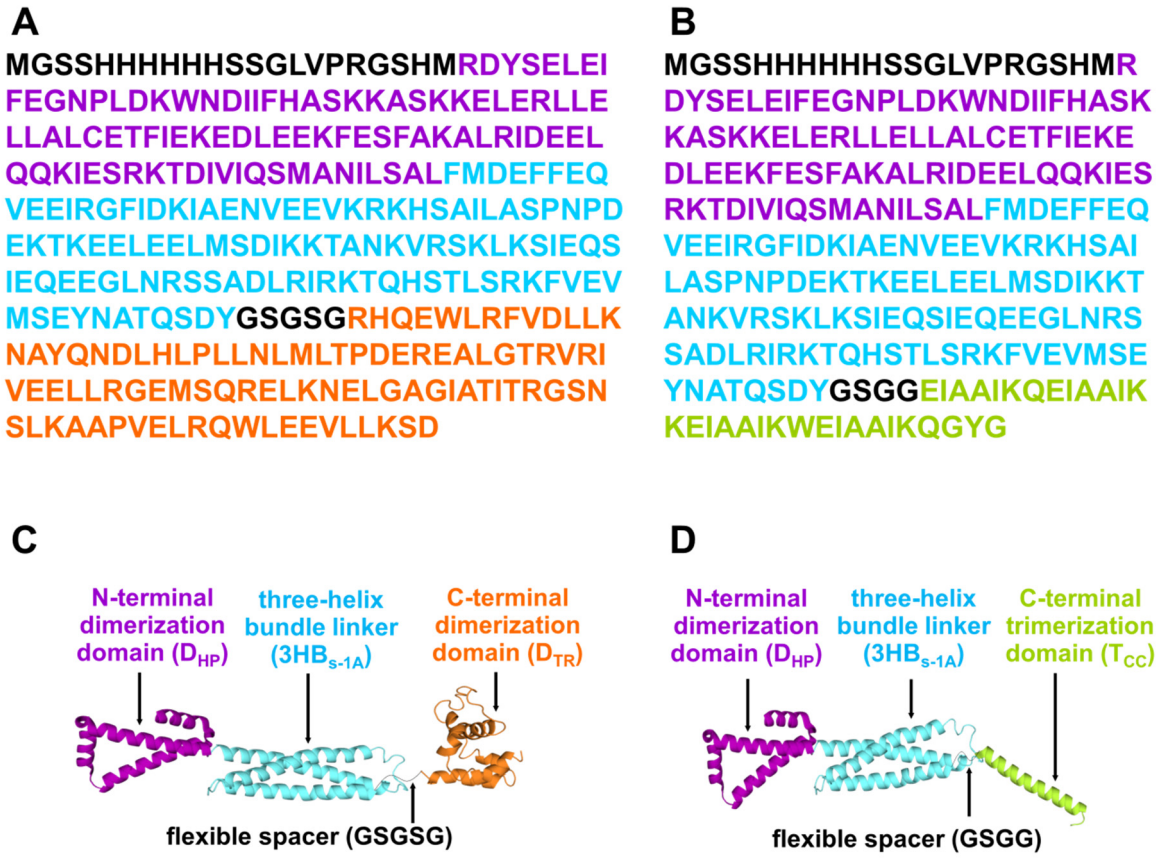


Figure S1. Amino acid sequences and 3D model structures of fusion proteins **1** and **2**. Amino acid sequences of (A) **1** and (B) **2** and cartoon representations of (C) **1** and (D) **2** are depicted. The residues corresponding to the N-terminal dimerization, three-helix bundle, C-terminal dimerization, and C-terminal trimerization domains are depicted in purple, light blue, orange, and light green, respectively.

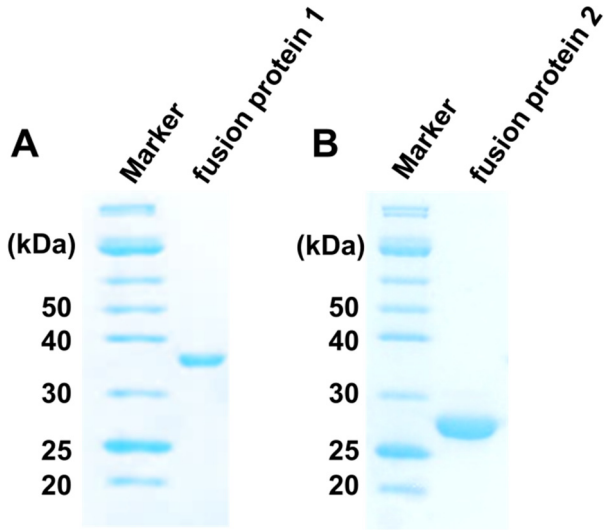


Figure S2. SDS-PAGE analysis of fusion proteins (A) **1** and (B) **2** purified by immobilized metal affinity chromatography.

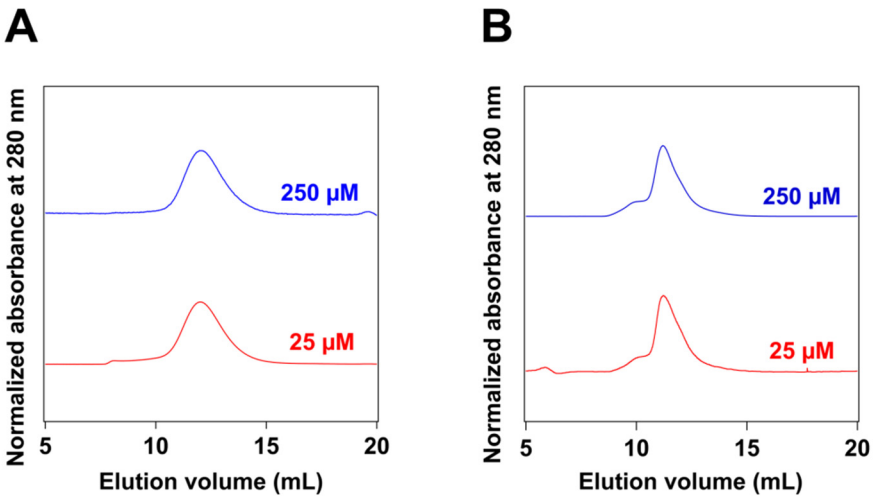


Figure S3. Size exclusion chromatograms of the main oligomers of fusion proteins **1** and **2**. (A) Chromatograms of the main oligomer of **1** at protein concentrations of 250 µM (blue) and 25 µM (red). (B) Chromatograms of the main oligomer of **2** at protein concentrations of 250 µM (blue) and 25 µM (red).

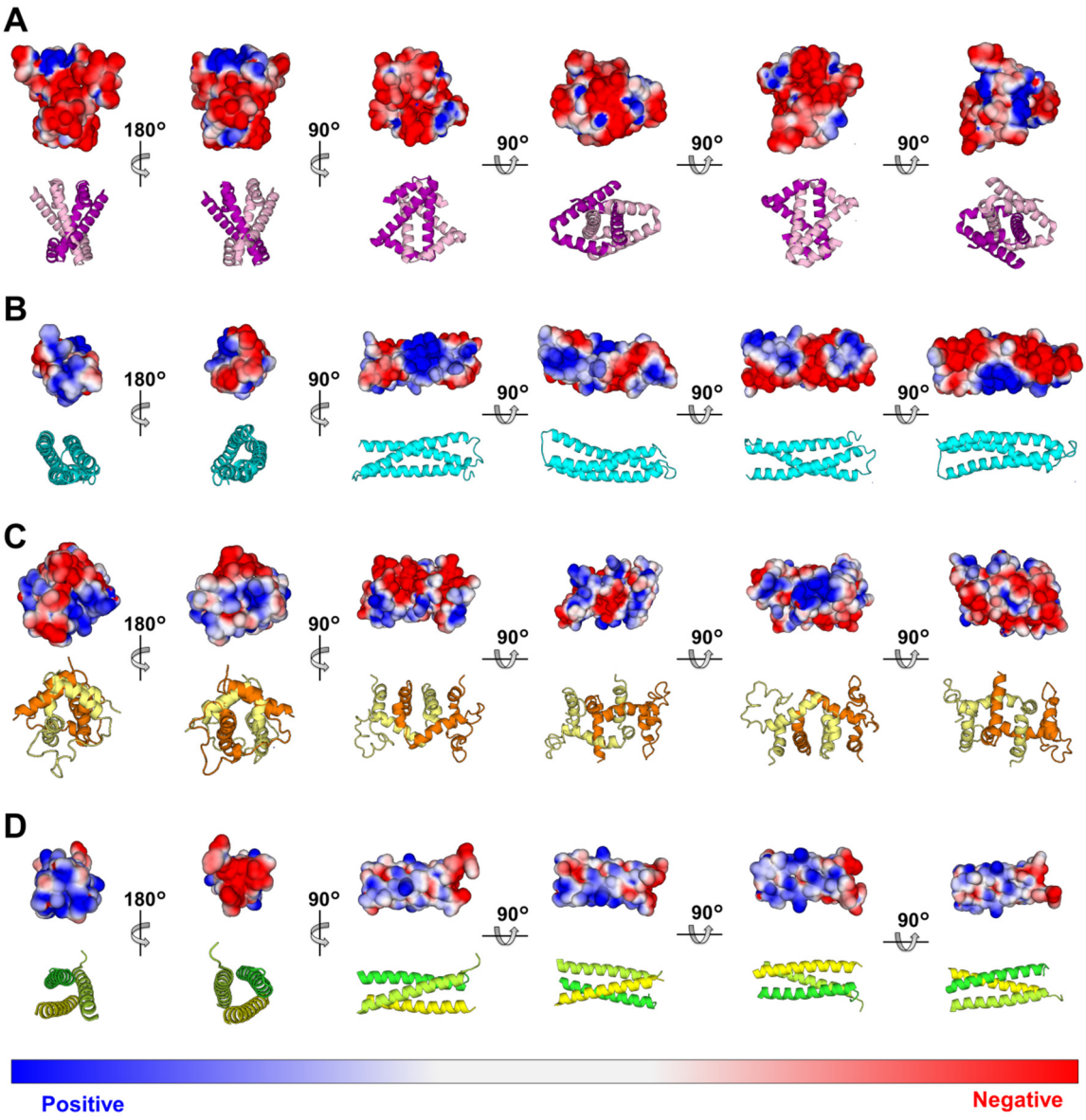


Figure S4. Surface charge distribution comparison of the domains of fusion proteins **1** and **2**: (A) N-terminal domain (D_{HP}, PDB ID: 3MLI), (B) three-helix bundle domain ($3HB_{s\text{-}1A}$, PDB ID: 1BR0), (C) C-terminal domains of **1** (D_{TR}, PDB ID: 1WRT) and (D) C-terminal domains of **2** (T_{cc}, PDB ID: 4DZL). Electrostatic calculations were performed using the APBS program tool and visualized by PyMOL. Positively-charged, neutral, and negatively-charged regions are shown in blue, gray, and red, respectively. D_{HP}, $3HB_{s\text{-}1A}$, D_{TR}, and T_{cc} are also shown in ribbon models in (A) purple/pink, (B) cyan, (C) orange/yellow, and (D) yellow/yellow-green/green, respectively.

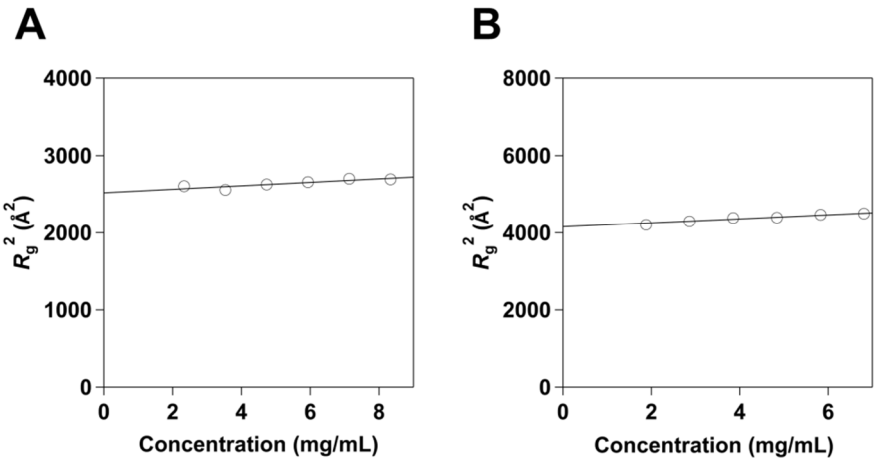


Figure S5. Protein concentration dependence of R_g^2 for the main oligomers of fusion proteins **1** and **2**. (A) R_g^2 vs. protein concentration plot for the main oligomer of **1**. (B) R_g^2 vs. protein concentration plots for the main oligomer of **2**.

Movie S1. HS-AFM movie of the tetramer of fusion protein **1** on an APTES-mica substrate.

Movie S2. HS-AFM movie of the hexamer of fusion protein **2** on an APTES-mica substrate.

Table S1. SAXS data collection and analyzed parameters of the tetramer of **1** and hexamer of **2**.

Data-collection parameters	the tetramer of **1**	the hexamer of **2**
Diffractometer	Nano-Viewer (RIGAKU)	Nano-Viewer (RIGAKU)
Detector	PILATUS 200K	PILATUS 200K
Beam geometry	Pin hole slit (Φ0.8 mm)	Pin hole slit (Φ0.8 mm)
Wavelength (Å)	1.5418	1.5418
Q range	0.012–0.25	0.012–0.25
Exposure time (min)	25	25
Concentration range (mg/mL)	2.3–8.3	1.9–6.8
Analyzed parameters		
$I(0)$/conc. (from Guinier)[a]	36.8 ± 0.7	38.9 ± 3.0
R_g (Å) (from Guinier)	50.1 ± 1.3	64.5 ± 5.9
$I(0)$/conc. (from $P[r]$)	37.4 ± 0.9	40.9 ± 2.2
R_g (Å) (from $P[r]$)	52.4 ± 1.6	69.9 ± 3.1
D_{max} (Å)	200	240
Porod volume estimate (Å^3)	379000	642000
Mw (kDa) (from $I(0)$/conc.)	177	187
Calculated Mw (kDa)	150	180

[a] The $I(0)$/conc. values of the tetramer of **1** and hexamer of **2** were obtained by the linear extrapolation of the $I(0)$/conc. vs. protein concentration plots to zero protein concentrations.