ADVANCED MATERIALS
www.advmat.de

Polymorphic Self-Assembly with Procedural Flexibility for Monodisperse Quaternary Protein Structures of DegQ Enzymes

Hanul Jeon, Ah-reum Han, Sangmin Oh, Jin-Gyeong Park, Myeong Namkoong, Kyeong-Mi Bang, Ho Min Kim, Nak-Kyoon Kim, Kwang Yeon Hwang, Kahyun Hur, Bong-Jin Lee, Jeongyun Heo, Sehoon Kim, Hyun Kyu Song, Hyesung Cho, and In-Gyun Lee**

As large molecular tertiary structures, some proteins can act as small robots that find, bind, and chaperone target protein clients, showing the potential to serve as smart building blocks in self-assembly fields. Instead of using such intrinsic functions, most self-assembly methodologies for proteins aim for de novo-designed structures with accurate geometric assemblies, which can limit procedural flexibility. Here, a strategy enabling polymorphic clustering of quaternary proteins, exhibiting simplicity and flexibility of self-assembling paths for proteins in forming monodisperse quaternary cage particles is presented. It is proposed that the enzyme protomer DegQ, previously solved at low resolution, may potentially be usable as a threefold symmetric building block, which can form polyhedral cages incorporated by the chaperone action of DegQ in the presence of protein clients. To obtain highly monodisperse cage particles, soft, and hence, less resistive client proteins, which can program the inherent chaperone activity of DegQ to efficient formations of polymorphic cages, depending on the size of clients are utilized. By reconstructing the atomic resolution cryogenic electron microscopy DegQ structures using obtained 12- and 24-meric clusters, the polymorphic clustering of DegQ enzymes is validated in terms of soft and rigid domains, which will provide effective routes for protein self-assemblies with procedural flexibility.

1. Introduction

The programming of self-assembled geometric clusters[1] is a broad topic that has attracted increasing interest over the past few decades.[2] Often, polyhedral assemblies, for example, tetrahedrons, cubes, and octahedrons, and others,[3] are targets for fabrication from materials at every scale; these assemblies include DNAs,[4] peptides,[5] proteins,[6] colloids,[7] and others,[8] and the construction of feasible paths to program these assembly structures is important for diverse applications. In materials science, the programming of such paths for polyhedral geometries generally falls into the following two categories: i) self-assembly and ii) origami folding. Self-assembly starts from disassembled units, and the units must bind together accurately.[9] In contrast, origami strategies use deployable layouts, that is, units already connected by soft hinges,[10] and are highly advantageous in flexible procedures during their morphing.[11]

H. Jeon, J.-G. Park, J. Heo, S. Kim, I.-G. Lee
Biomedical Research Division
Korea Institute of Science and Technology
5, Hwarang-ro 14-gil, Seongbuk-gu, Seoul 02792, Republic of Korea
E-mail: ik86@kist.re.kr

H. Jeon, J.-G. Park, K. Y. Hwang
Department of Biotechnology
Korea University
145, Anam-ro, Seongbuk-gu, Seoul 02841, Republic of Korea

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adma.202308837

DOI: 10.1002/adma.202308837

A.-reum Han, H. M. Kim
Center for Biomolecular and Cellular Structure
Life Science Cluster
Institute for Basic Science (IBS)
55, Expo-ro, Daejeon 34126, Republic of Korea

S. Oh, M. Namkoong, K. Hur, H. Cho
Extreme Materials Research Center
Korea Institute of Science and Technology
5, Hwarang-ro 14-gil, Seongbuk-gu, Seoul 02792, Republic of Korea
E-mail: hcho@kist.re.kr

K.-M. Bang, N.-K. Kim
Advanced Analysis Center
Korea Institute of Science and Technology
5, Hwarang-ro 14-gil, Seongbuk-gu, Seoul 02792, Republic of Korea

The experimental routines in self-assembly strategies for polyhedral cages consist of designing geometric features (e.g., shapes and coordinators at vertices or edges) and functions with specific binding and affinity, as well as evaluating clusters in terms of defects, which are involved in forward engineering. When unfavorable defects are found in sets or subsets of target assemblies in the assembling paths, the "player" returns to one of the past stages in the routine in forward engineering. This process may successfully find a feasible path but may also be prone to many failures during the iteration. Thanks to recent computational approaches, the number of iterations in the path construction has been effectively reduced, as evidenced in examples of assembled clusters along fascinating and feasible paths.[12] For quaternary protein clusters with tertiary protein units, however, it is highly challenging to make the units follow such programmed paths, mainly owing to their structural complexity (i.e., large multicomponent proteins).[6c,13] Proteins feature local rigidity differences in folding structures (e.g., alpha-helices and beta-strands vs loop regions) and hence are difficult to obtain ordered forms (e.g., cages and crystals),[6a,14] unlike other low molecular building blocks.

For procedural flexibility, origami strategies have been reported in recent materials sciences for programmed assemblies; in the models, approximate (i.e., not precise) yet feasibly working paths via low-energy deformation of soft hinges locally confined by rigid bodies (i.e., nonstretchable/compressive tiles) are employed.[10,11b,15] Unlike de novo designs with highly accurate path construction,[16] herein we present a strategy that can form monodisperse protein cage particles by exploiting the rigid body and soft hinge characteristics in tertiary proteins. Using the low-resolution structures in previous studies, we assume DegQ, a protomer threefold symmetric trimer, can retain soft parts in subdomains during clustering, which can be guaranteed when soft client proteins are introduced to take advantage of a chaperone action of DegQ. In our experiments, self-assembled clusters of DegQ in polyhedral forms of 12-mers and 24-mers can be formed uniformly, wherein the trapped client should be less resistive than the domains of DegQ to accommodate facile deformation during clustering via symmetry preference.[17] To support our claim, the reconstruction of atomic cryogenic electron microscopy (cryo-EM) structures of DegQ quaternary structures are presented at high resolution (2.56–3.65 Å), providing a basis for classifying rigidity differences based on the tendencies of order and disorder of clustered protein domains.

K.-M. Bang, H. K. Song
Department of Life Science
Korea University
145, Anam-ro, Seongbuk-gu, Seoul 02841, Republic of Korea
H. M. Kim
Graduate School of Medical Science and Engineering
Korea Advanced Institute of Science and Technology (KAIST)
291, Daehak-ro, Daejeon 34126, Republic of Korea
B.-J. Lee
The Research Institute of Pharmaceutical Science
Seoul National University
599, Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea
B.-J. Lee
College of Pharmacy
Ajou University
206, Worldcup-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 16499, Republic of Korea
S. Kim
KU-KIST Graduate School of Converging Science and Technology
Korea University
145, Anam-ro, Seongbuk-gu, Seoul 02841, Republic of Korea

2. Results and Discussion

2.1. Reverse Engineering of Polyhedral Cages to Conjecture Feasible Paths for Protein Cage Particles with Procedural Flexibility in Symmetric Clustering

Figure 1a compares the forward and reverse engineering of polyhedral clusters, wherein a tetrahedron consisting of four triangular surfaces is exemplified as the target. Forward engineering involves continuous assembly procedures described by a known, programmed, concrete blueprint, whereas reverse engineering focuses on a logical, rational, and gradual dismantling of the target into working subsets from which to conjecture underlying hidden paths,[18] especially for the consideration of reassembly. Before starting to build the target tetrahedron, forward engineering may be needed to define design factors in unit triangles (for some materials, forward engineering starts from much smaller primitives, for example, lines or points);[4,12b] at this stage, the functions that allow specific binding between blocks should be programmed. This is highly essential because, at a small scale, it is usually impossible to manipulate the disassembly of blocks connected in unfavorable configurations in local energy-minimum states; thus, particular coordinators[6a] or linkers[9,19] must be accurately designed at each hinge or vertex to avoid such misconfigurations. Most protein self-assemblies strictly follow this rule (i.e., one-way assembly procedures from the starting strand).[20]

In contrast, inspired by the procedural flexibility in morphing of origami structures[11a] and reverse engineering (see Figures S1 and S2, Supporing Information), we propose reverse paths from the given tetrahedron cage. Here, our insight is to harness the intrinsic chaperone action of a protein, DegQ, in the presence of client proteins (Figure 1b), eventually creating polyhedral cages via symmetry preference.[6c] Notably, some enzymes act as small robots, targeting, binding, and chaperoning client proteins; these actions are desirable function for researchers who must program such functions one-by-one into their building blocks in programming self-assembled structures. Importantly, the chaperone action of DegQ can easily create approximate clusters (i.e., rough ones in asymmetric configurations) with client proteins; thus, we may avoid burdens in searching and programming additional details of subpaths, which is advantageous than other strategies in forward engineering (e.g., de novo structure design). Instead, we focus on how to guide the rough, approximate clusters to symmetric polyhedral cages without defects. The key is to use very soft client proteins in implementing clustering in a symmetric configuration with cage-forming proteins that feature chaperone action, thereby undergoing less resistive deformation of clients. Based on this strategy, monodisperse polyhedral cages of proteins can be self-assembled, enabling the elucidation of the high-resolution quaternary structures of the protein *Escherichia coli* (*E. coli*) DegQ for the first time in this study. A detailed explanation is given below.

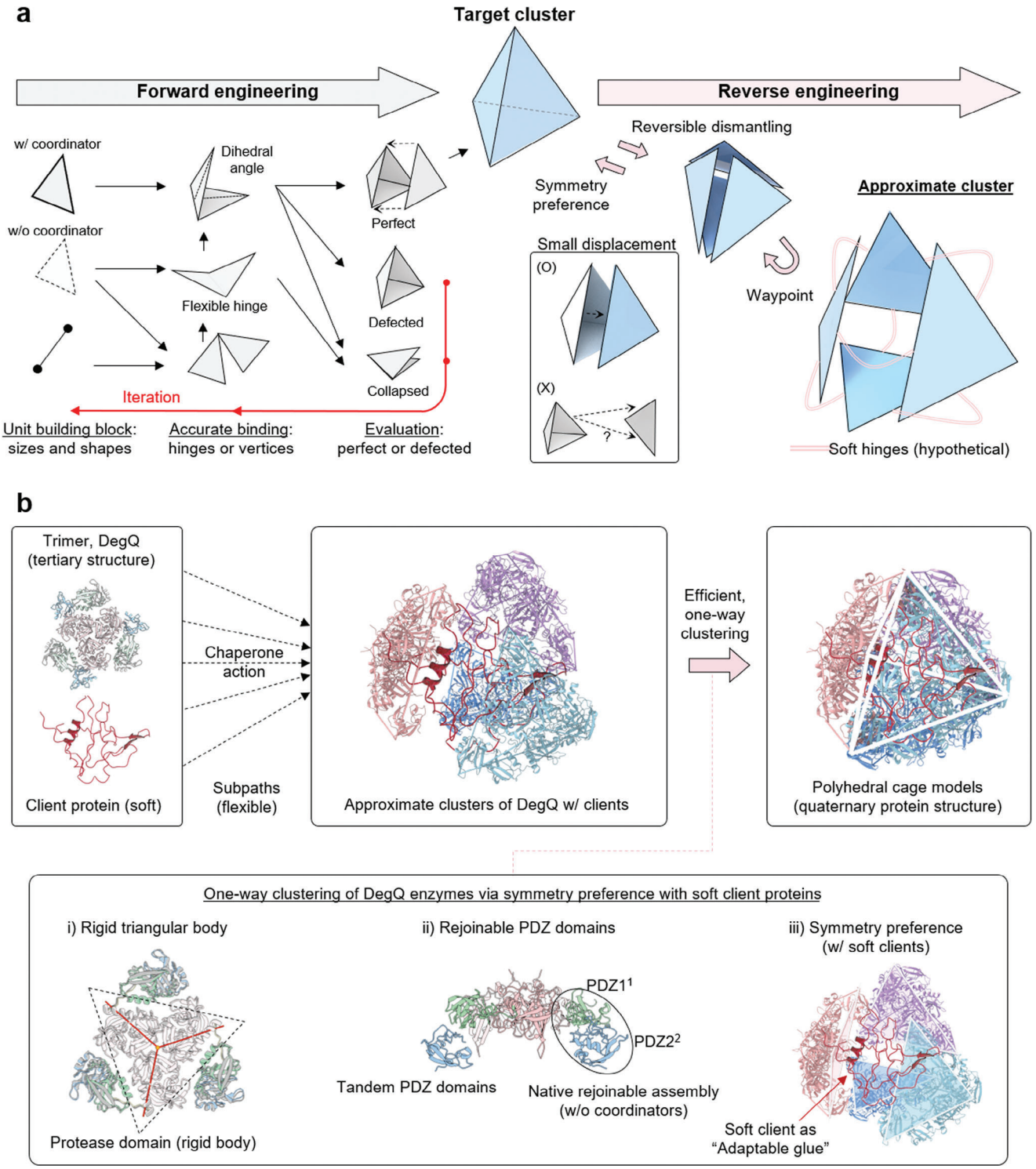

Figure 1. a) Comparison between forward and reverse engineering to achieve a target structure, in which a tetrahedron is used as an example. Notably, forward engineering involves a series of iterations to find feasible paths with highly accurate and programmed building blocks, while reverse engineering aims for the deterministic reassembly of subunits. In the inset, the small displacement (marked as O) represents the dismantling of a triangular unit from a tetrahedron, where the original edge pairs can be rejoined, and the large displacement (marked as X) causes losing the paths toward the original edge pairs. b) A model path including the flexible subpaths obtained by the chaperone action of the trimer DegQ working with soft client proteins and one-way clustering. Procedural flexibility via symmetry preference is proposed for the clustering from the approximate clusters.

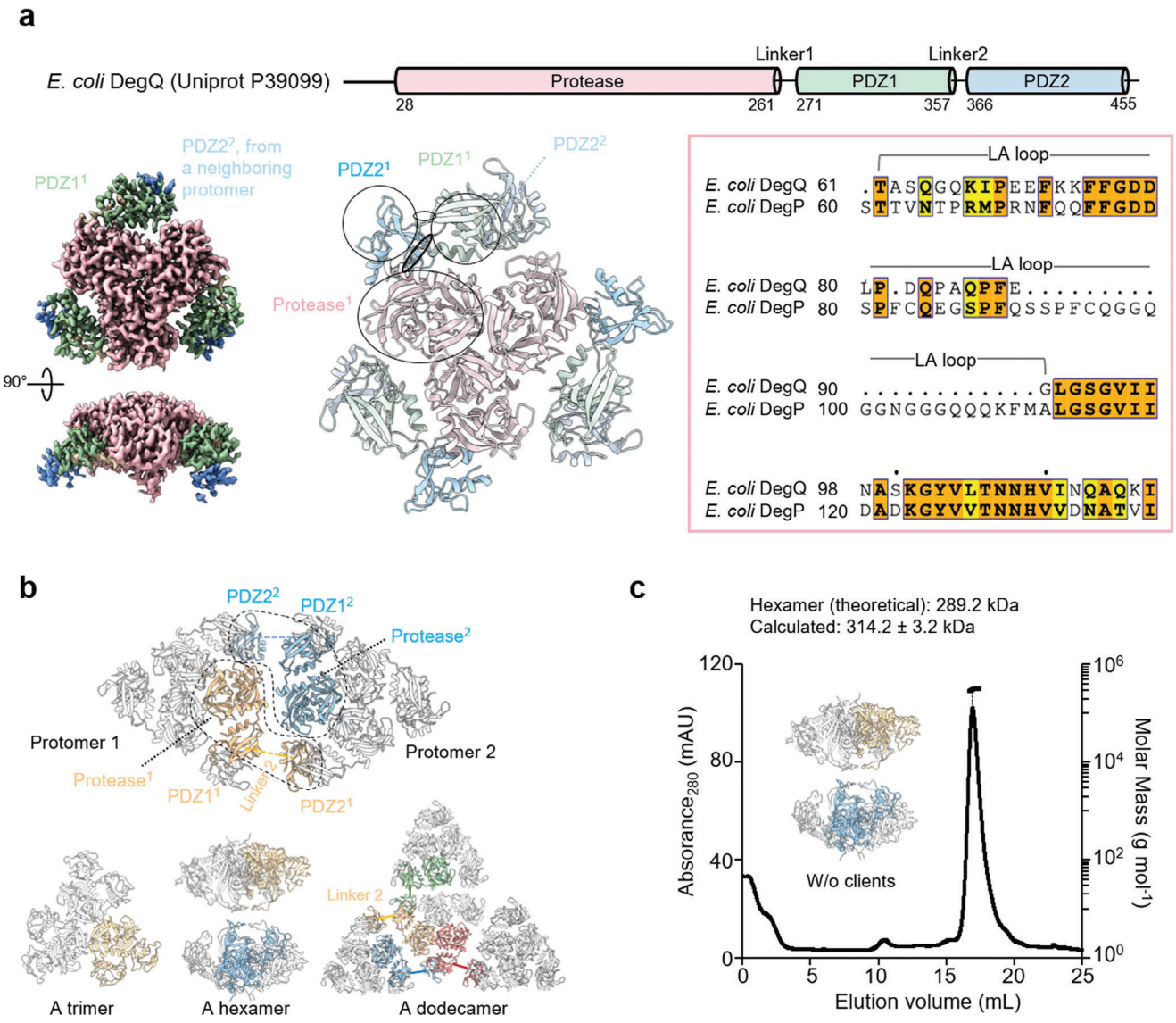

Figure 2. a) The domain architectures of DegQ enzyme, as a model unit for symmetric clustering; the protease domains with a short LA loop act as the triangular building block, and the DegQ units are connected at PDZ domains. Cryo-EM density map (left) and atomic model (middle) for a trimeric unit of DegQ. A comparison between the protease domains of DegQ and DegP is presented in a box colored pink (right). A single DegQ polypeptide chain starting from the protease domain, PDZ1 to PDZ2, is highlighted in circles in black. b) A few possible connections of DegQ, where the superscript denotes each protomer. c) SEC-MALS analysis of apo DegQ. The molar mass determined from light scattering (right *y*-axis) and UV absorption at 280 nm (left *y*-axis) is plotted as a function of the elution volume.

2.2. DegQ Trimer, as a Building Block for Polyhedral Polymorphic Cages

DegQ and DegP, members of the high-temperature requirement protein A (HtrA) serine protease family, were observed in biology with their structural evidence in the form of self-assembled cages, including shapes of tetrahedrons, octahedrons, and icosahedrons in combination with client proteins (or substrates).[21] However, structural information on DegQ quaternary structures has remained limited to low resolution (see Table S1, Supporting Information). Our initial goal was to set paths for the creation of highly uniform protein clusters and then find new information on the structural biology of the obtained quaternary protein structures at high resolution via cryo-EM analysis to evaluate our logic. The domain architectures of DegQ consist of the *N*-terminal protease domains followed by tandem PDZ domains (PDZ1 and PDZ2), as shown in **Figure** 2a. Early in this study, the structural characteristics of DegQ were revisited, as this could support consideration of the rigid body and soft hinge models; note that we initially assumed such models from the low-resolution structures of DegP and DegQ in previous studies[21b,c,22] but subsequently validated them by using the high-resolution cryo-EM structures from our study after the successful construction of one-way routes for monodisperse cluster formation. The structures presented in Figure 2 are from our high-resolution cryo-EM structures, but the assumption below may be conjectured

(but not verified) from the low-resolution structure. First, the protease domains interact closely with each other to form a symmetric trimer, that is, a threefold DegQ oligomer block. Second, DegQ contains a markedly shorter LA loop than DegP, offering locally rigid mechanical characteristics in the protease domain (see the amino acid sequences in the pink box in Figure 2a; for comparisons between full-length sequences, see Figure S3, Supporting Information). Third, some domains of the trimer may be softer, among PDZ1, PDZ2, and linkers connecting the domains (Linker 1, residues 262–270 and Linker 2, residues 358–365) than the protease domains. Unless other specific coordinators are bounded, the native connection of the two trimers, occurred by the connection between $PDZ1^{i}$:$PDZ2^{j}$ ($i \neq j$, for different trimer building blocks), remains flexible, which may allow them to function as soft hinges (Figure 2b). Based on these conjectures, DegQ could be adopted as a trimer unit for the construction of 12-mers (i.e., tetramers of trimers) or 24-mers (i.e., octamers of trimers) with procedural flexibility from the approximate assemblies.

However, using DegQ alone (i.e., without the chaperone action) could not guarantee efficient formations of target clusters; under such a condition, various configuration could be simultaneously generated. Indeed, there is disagreement on the stable form of DegQ in solution without clients; the clusters in solution can be hexamers[21b] or other forms (including trimers and 12-mers; see.[21c]) We observed the footprint of the hexamers under our experimental conditions (Methods), as measured by size exclusion chromatography coupled with multiangle light scattering (SEC-MALS) (Figure 2c). Note that the calculated mass (i.e., 314.2 kDa) corresponds very well with the theoretical molar mass of a hexameric state (i.e., 289.2 kDa), suggesting that most clusters in the solution are hexamers. Presumably, DegQ might assemble into 12-mers as one of the background distributions in our solution; as it is known that $PDZ1^{i}$: $PDZ2^{j}$ can be rejoinable, the hexamers may have a hypothetical route toward the formation of 12-mers. Among those metastable configurations, utilization of soft clients and the chaperone action of DegQ can be the key to program the stable clustering of symmetric cage particles.

2.3. Soft Client Protein, as the Key in Constructing Arbitrary Paths toward Less Resistive Clustering and Deterministically Forming Symmetric Cages

Unsurprisingly, it has been observed in structural biology that DegQ and DegP can form tetrahedral clusters when introduced to some clients; surprisingly, however, the role of softness in the client remains poorly understood. This lack of knowledge is due to the differences in goals between biology and engineering. The goal of the former is to observe and report the biological implications of such quaternary structures, while the goal of the latter is to design paths for self-assembled clusters without defects (i.e., high-throughput yielding) for applications. Interestingly, we show that these two goals can converge; once a path is feasible for programming highly monodisperse cluster formation, high-resolution observation, that is, cryo-EM analysis, can reveal further structural information in structural biology. To achieve both goals, we first report the importance of employing soft client proteins in forming monodisperse clusters of DegQ by locally adjusting the units bounded to soft clients, thereafter ensuring the formation of symmetric conformations as polyhedral cages, as presented in **Figure 3**.

As a client protein, the native lysozyme structure is shown in Figure 3a, which shows the alpha-helices and beta-strands. Our assumption is that more rigid clients result in clusters with lower orders of symmetry or defective clusters because the soft parts in DegQ have a limited ability to deform rigid clients (Figure 3b). To verify this assumption, native and denatured clients were introduced into the solution with DegQ and compared (see Methods). As shown in Figure 3c, the denatured lysozyme yielded a uniform conformation of DegQ 12-mers whereas the native lysozyme failed to form such a high-order symmetric conformation under our experimental conditions. This finding can be explained by the fraction of total alpha-helices and beta-strands because of their effect on rigidity in both cases (Figure 3d and Methods). The circular dichroism (CD) spectrum of denatured lysozyme showed a weak far-UV CD signature, suggesting the less existence of alpha-beta structure such that the client can be less resistive against mechanical distortion during symmetry preference of DegQ cage particles. When denatured, lysozyme mostly consists of loop regions, lacks secondary structural elements, and can be easily captured by hexamers.

DegQ naturally has dual functions as a chaperone and a protease. For researchers using the origami strategy, designing a mechanism to fold hinges is a task, for which many driving mechanisms have been found, including swelling,[23] magnetic forces,[24] and others.[25] However, because of its chaperone function, DegQ can find, capture, and even deform the client proteins, acting as self-operating robotic building blocks. At the same time, as a protease, DegQ can break the captured client into pieces, followed by the disassembly into primitives (e.g., trimers or hexamers) until it finds another client. In our reverse engineering strategy, we deactivated the protease function (here, we mutated the catalytically important residue Ser214 to Ala) to maintain the stability of the formed cluster for high-resolution cryo-EM imaging.

It should be noted that the role of the soft client can be explained by various analogous systems in other research fields, which aim at establishing a symmetric conformation of units via softness. Recently, Lee et al.[26] reported the self-arrangement of microdevices wherein the molten solder between a device and a sapphire substrate can spontaneously find the center of each device via symmetry preference. Similarly, we reported that even a random number of expanding soft shells can form symmetric clusters when they are confined in circular holes.[27] Thus, even though we leave subpaths undefined, soft clients allow for high degrees of freedom in capturing procedures by surrounding hexamers; for example, some hinges in a hexamer can be disconnected by introducing soft clients, and then the dihedral angles at the remaining hinges can be varied, allowing for flexible attachment of another hexamer in a similar fashion to form an approximate cluster. Regardless of which subpaths would be implemented, more significantly, once such a cluster is formed with denatured lysozyme, the symmetry preference of four DegQ units can easily lead to the conformation of 12-mers symmetrically. To validate this assumption with procedural flexibility, we present high-resolution cryo-EM structures of DegQ cages.

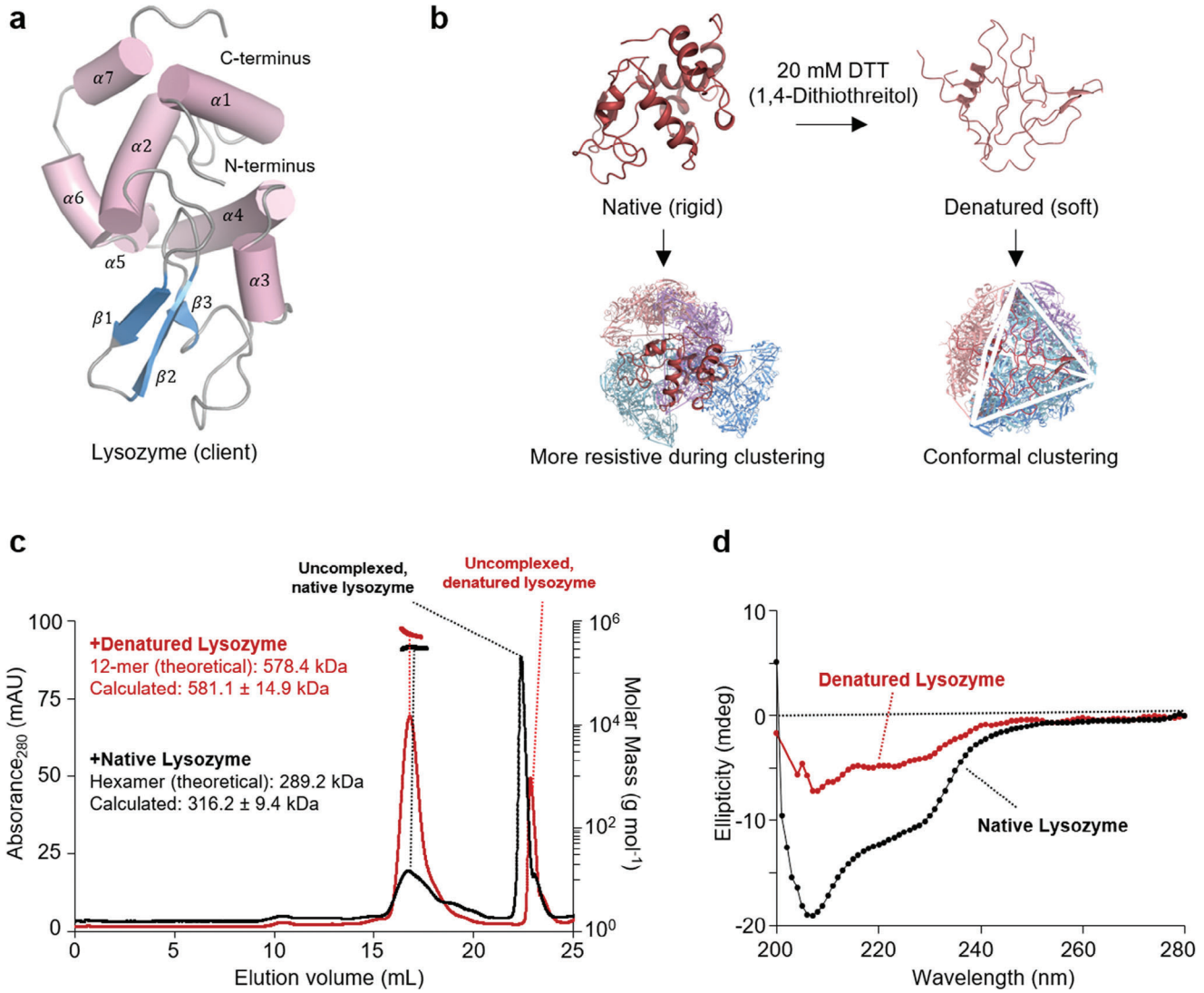

Figure 3. a) Ribbon representation of the crystal structure of lysozyme (UniProt ID: P00698, PDB ID: 193L). b) Comparison of the role of the softness of clients in our strategy. Soft clients are prone to forming conformal clusters. c) SEC-MALS analysis of DegQ enzyme complexed with denatured (red) and native (black) lysozyme. d) Circular dichroism (CD) spectra of native (black) and denatured (red) lysozyme.

2.4. High-Resolution Cryo-EM Structures and Validation of the Rigid Body and Soft Hinge Model with Soft Clients

As mentioned above, DegQ quaternary structures were previously solved only at low resolution. The cryo-EM structures in **Figure** 4a,b present the symmetric cages of the 12-mer and 24-mer, respectively, which are reconstructed at overall resolutions reaching 4.06 and 4.17 Å, respectively. Local refinement using a mask on each protomer yielded maps at 2.56 and 3.65 Å, respectively. It is worth noting that this work presents the first atomic model for both the 12-mer and 24-mer of *E. coli* DegQ. The assembled clusters of 12-mer and 24-mer cages are highly uniform, where denatured lysozyme and β-casein were used as the clients for their assembling, respectively (see Figures S4 and S5, Supporting Information and Experimental Section). We elucidate the path with simplicity for accessing the quaternary structures of DegQ, primarily based on soft hinges in subdomains of DegQ, where the role of the soft client is analogous to that of an "adaptable soft glue" as found in other studies.[28] Despite their disorder (see the Poly-A model without side chains in Figures S6 and S7, Supporting Information) owing to the softness, the substrates were consistently found to bind to DegQ in the vicinity of protease domain catalytic cores consisting of activation loops L1-L3, LD, and LA and the catalytic triad (His-109, Asp-139, and Ser-214 (mutated to Ala)), suggesting that soft clients can still be specifically recognized by the protease domains and also less resistive to the clustering of highly monodisperse cage particles. In our findings, the trimer domains (protease, PDZ1, and PDZ2 domains) were very similar between 12-mers and 24-mers, with root mean square deviations of ≈1.0 Å for equivalent Cα atoms. Regarding the configurations (Figures 4c and S9, Supporting Information), the reconstructed structures for both cases have dihedral angles of ≈70.4° and ≈109.7°, which are comparable to the angles of a tetrahedron and an octahedron, respectively. Once the hinges in

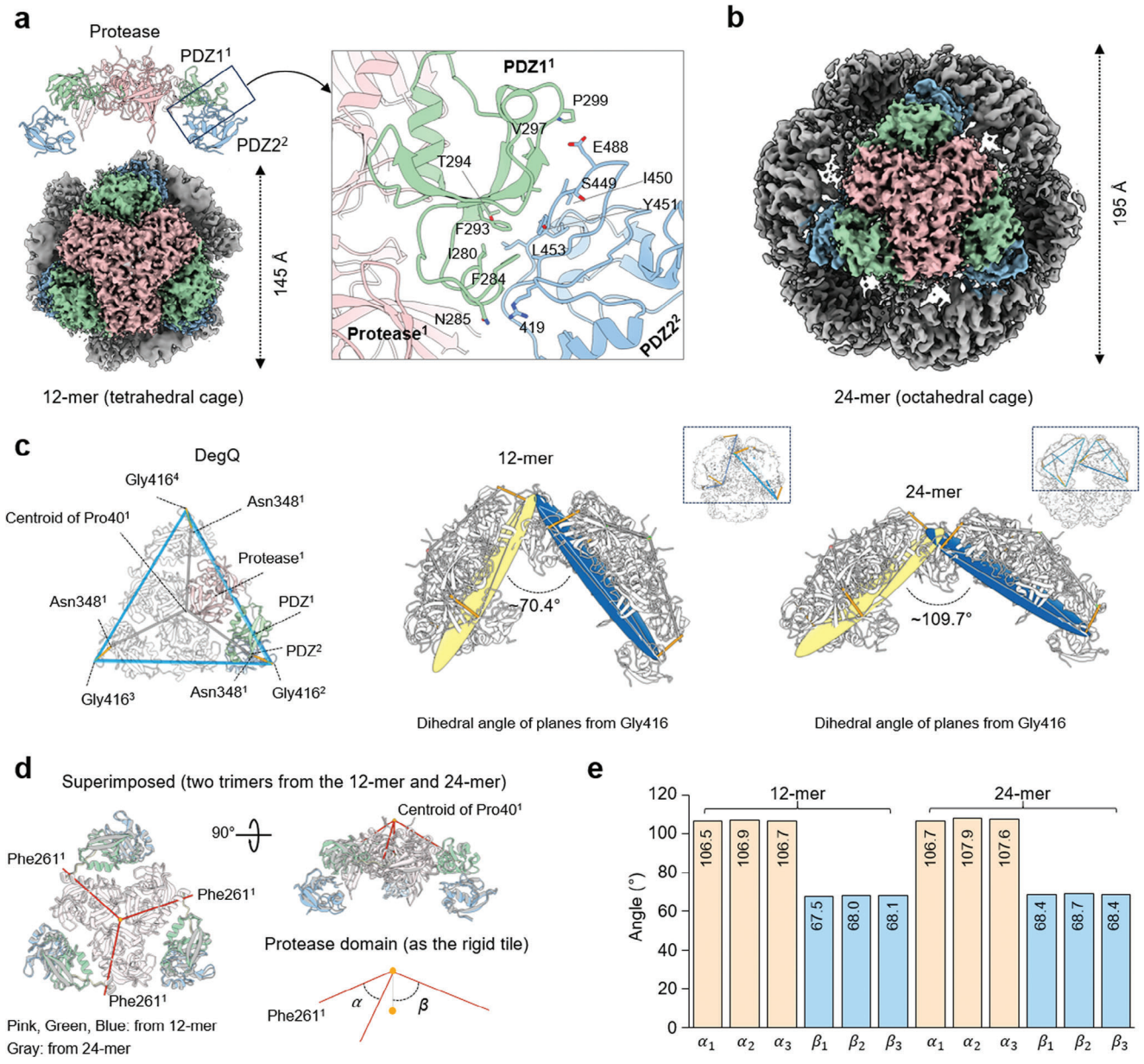

Figure 4. a) Consensus cryo-EM map of the lysozyme complexed 12-meric DegQ enzyme cage. The atomic model of the trimeric protomer is shown on top, and a zoomed-in view of the PDZ1[1]:PDZ2[2] interaction is shown on the right. b) Consensus cryo-EM map of the β-casein complexed 24-meric DegQ cage. c) Geometric points used to define dihedral angles for both cages in the verification of the soft hinge model. d) Superimposed structures of trimers from the 12-mer and 24-mer, where the C-terminal residue of the protease domain, Phe261, is chosen to evaluate the rigid body model. e) Comparison of geometric characteristics where the two cases showed similar values, confirming the rigid body model.

the cages have specific coordinators to lock the angle, no such polymorphic assembly can be formed. In our reconstructed cryo-EM structures, Linker 2 and the clients are disordered and hence can be soft (Figure S8, Supporting Information), which play key roles in low-energy deformation during symmetric one-way clustering. To verify the assumption for the rigid body model, we compared the cryo-EM structures of both 12-mers and 24-mers, the results of which are presented in Figure 4d. We measured the angles at the protease domains for both cages, revealing that they remained mostly unchanged when measured at certain geometric points, such as Phe261[1]. The measured angles between the points and the centroid are summarized in Figure 4e, and overall, they confirm the validity of all the assumptions in our study.

3. Conclusion

We have presented a route with procedural flexibility to program monodisperse self-assembled clusters of a protomer, DegQ, and verified this strategy by solving high-resolution cryo-EM structures via structural biology approaches. Paths toward self-assembled quaternary structures of DegQ, including 12-mers and 24-mers, were revealed at high resolution for the first time,

enabling detailed measurements of the geometric characteristics of both protease domains (i.e., rigid bodies) and other domains (i.e., PDZ1, PDZ2, and linkers) in polyhedral clusters. The clusters obtained were highly monodisperse, supporting highly reproducible reconstruction of the high-resolution structures by cryo-EM analysis. In the route we programmed, the role of soft clients was critical to the formation of symmetric cages, whereas a rigid client confined to natural hinges could not be deformed enough to allow symmetric cages. Instead, soft clients could lead to highly symmetric clusters along flexible paths in symmetric reconfiguration. Moreover, no specific function was added to DegQ; instead, the native chaperone action of the protein on soft clients was applied to create the cluster. For path construction for self-assembled cages, reverse engineering was performed to drive flexible paths via unique intermediate states, that is, approximate clusters with threefold units bound to the soft clients, which eventually and deterministically formed symmetric cages of DegQ. Along with reverse engineering, the protease function was deactivated to prevent the disassembly of the formed clusters. The strategy presented here is feasible, simple, and practical for programming polyhedral protein cages via unique procedural flexibility, which would enable the construction of additional protein assemblies by reducing the burden of current forward engineering.

4. Experimental Section

Cloning, Expression, and Purification of Recombinant Proteins: The cDNA coding for *E. coli* DegQ (UniProt ID: P39099) was synthesized by IDT (Integrated DNA Technologies, Coralville, LA) as gBlocks. *E. coli* full-length DegQ was cloned and inserted into pET28a (NEB, Ipswich, MA), which added a His_6-tag at the C-terminus of DegQ, using a Gibson assembly method. A point mutation (ser214 to Ala) was introduced using a method described previously.[29] DegQ protein was expressed in *E. coli* BL21 (DE3) cells (Invitrogen, Carlsbad, CA) grown in Terrific Broth medium at 37 °C until the OD_{600} reached a value of 1.5-2, followed by 16 h at 19 °C in the presence of 0.25 mM isopropyl-β-D-thiogalactoside. The cells were collected by centrifugation, resuspended in 50 mM Tris (pH 8.0) and 500 mM NaCl (Buffer A), and lysed using ultrasonication. The protein was loaded on a Ni-NTA-affinity column equilibrated with Buffer A. The proteins were eluted with Buffer A supplemented with 100 mM imidazole. The proteins were additionally purified by gel filtration on a Superose 6 Increase 10/300 GL column (GE Healthcare, Chicago, IL) in 50 mM Tris (pH 7.0) and 100 mM NaCl (Buffer B). Lysozyme and β-casein were purchased from Sigma Aldrich. To obtain the DegQ-lysozyme complexes, purified protease-deficient $DegQ_{S214A}$ was added at a molar ratio of 1:2 (DegQ: lysozyme) to denatured lysozyme that was pretreated with 20 mM DTT.

Size Exclusion Chromatography Coupled with Multiangle Light Scattering (SEC-MALS): Protein samples were separated by size exclusion chromatography on a Superose 6 Increase 10/300 GL column equilibrated with 50 mM Tris (pH 7.0) and 100 mM NaCl. Multiangle light scattering was measured in line using a DAWN-HELEOS multiangle light scattering detector and an Optilab rEX refractive index detector. The scattering data were analyzed with ASTRA software (Wyatt Technology, Santa Barbara, CA). In Figure 3c, the purified 100 µM DegQ was pre-mixed with an equimolar concentration of native lysozyme or lysozyme denatured with 20 mM DTT.

Circular Dichroism (CD) Spectroscopy: The CD spectra of proteins and of buffer alone were recorded on a Chirascan-Plus CD Spectropolarimeter (Applied Photophysics, Leatherhead, UK) in 1 nm steps from 280 to 200 nm using a 0.1 cm path-length quartz cuvette and a scan rate of 20 nm min^{-1}. The proteins were concentrated to 30 µM in 20 mM Tris (pH 7.5), 100 mM NaCl. The reported spectra corresponded to merged spectra after buffer subtraction.

Cryo-EM Sample Preparation and Data Collection: Holey carbon grids (Quantifoil, Cu, 300-mesh, R1.2/1.3) were glow-discharged for 60 s at 15 mA using a PELCO easiGlow glow discharge cleaning system (Ted Pella). After blotting for 5 s using Whatman filter paper, the grids were quickly submerged into liquid ethane that was cooled by liquid nitrogen using Vitrobot Mark IV (Thermo Fisher) at 4 °C and 100% humidity. DegQ-lysozyme complex datasets were collected on a 300 kV Titan Krios (FEI) equipped with a Gatan K3 BioQuantum direct electron detector in tiff format. Movies were recorded at a nominal magnification of ×105 000 corresponding to a calibrated pixel size of 0.848 Å $pixel^{-1}$. Micrographs were automatically collected for each sample with a defocus ranging from −0.7 to −1.7 µm with a total exposure time of 6.1 s fragmented into 57 frames. Total dose was 67.7 e^{-} $Å^{-2}$. 11135 movies were collected using the automated EPU software package. DegQ- β-casein complex datasets were collected on a Glacios electron microscope (Thermo Fisher) operated at 200 kV and images were recorded with a Falcon 4i Direct Electron Detector in EER format. Movies were recorded at a nominal magnification of ×73 000 corresponding to a calibrated pixel size of 0.874 Å $pixel^{-1}$. Micrographs were automatically collected for each sample with a defocus ranging from −0.9 to −2.0 µm. Total dose was 60 e^{-} $Å^{-2}$. Movies (6067) were collected using the automated EPU software package.

Cryo-EM Data Processing: All data were processed using cryoSPARC v4.2.1[30] The flowcharts for data processing for DegQ-lysozyme and DegQ-β-casein are presented in (Figures S4 and S5, Supporing Information), respectively. For the DegQ-lysozyme complex, a total of 657936 particles from the best classes were subjected to Ab initio reconstruction and heterogeneous refinement. After multiple rounds of refinement, 382427 particles were further subjected to nonuniform refinement, yielding a reconstruction map at an average resolution of 4.06 Å. To improve the local map quality, different generated local masks were used to perform particle subtraction or local refinement using cryoSPARC. This practice yielded improved EM reconstructions with a local resolution of 2.56 Å. The resolution was determined according to the Fourier shell correlation 0.143 criterion with a high-resolution noise substitution method. For the dataset of the DegQ-β-casein complex, a similar data processing procedure was carried out with slight modification. After particle extraction, several rounds of 2D classification were performed, which gave rise to a final dataset containing 73850 particles. The same 2D classification procedure was performed as described above, and the data were subjected to ab initio reconstruction, heterogeneous refinement, and non-uniform refinement. The process yielded a final reconstruction at 4.17 Å. To improve the local map quality, different generated local masks were used to perform particle subtraction or local refinement using cryoSPARC. This practice yielded improved EM reconstructions with a local resolution of 3.65 Å.

Model Building and Refinement: The atomic coordinates of DegQ (22) were docked into the EM map using Chimera[31] and manually adjusted and rebuilt in COOT[32] to yield the final atomic model. All structural refinements were performed in PHENIX in real space with secondary structure and geometry restraints.[33] Overfitting of the models was monitored by refining the model against one of the two independent half maps from the gold-standard refinement approach and testing the refined model against the other map.[34] The structures were validated through an examination of the MolProbity[35] scores and the statistics of the Ramachandran plots (Table S2, Supporting Information).

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work was supported by the research support program of the Lim Sung Ki Foundation (LF-RSP2022-02) and by the intramural grants of the Korea

Institute of Science and Technology (KIST). This research was also supported by the National Research Council of Science & Technology (NST) grant by the Korean government (MSIT) (No. CAP23011-100). This research was supported by a KIST internal project (2E32527). The authors are grateful to the staff of the Research Solution Center at IBS for help with cryo-EM data collection. Computational work for this research was performed on the data analysis hub (Olaf) in the IBS Research Solution Center.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

procedural flexibility, quaternary protein structures, self-assemblies

Received: August 30, 2023
Revised: February 8, 2024
Published online: February 28, 2024

[1] P. F. Damasceno, M. Engel, S. C. Glotzer, *Science* **2012**, *337*, 453.
[2] a) A. S. Keys, C. R. Iacovella, S. C. Glotzer, *Annu. Rev. Condens. Matter Phys.* **2011**, *2*, 263; b) V. N. Manoharan, *Science* **2015**, *349*, 1253751.
[3] C. Avci, I. Imaz, A. Carné-Sánchez, J. A. Pariente, N. Tasios, J. Pérez-Carvajal, M. I. Alonso, A. Blanco, M. Dijkstra, C. López, D. Maspoch, *Nat. Chem.* **2018**, *10*, 78.
[4] a) Y. He, T. Ye, M. Su, C. Zhang, A. E. Ribbe, W. Jiang, C. D. Mao, *Nature* **2008**, *452*, 198; b) R. Iinuma, Y. G. Ke, R. Jungmann, T. Schlichthaerle, J. B. Woehrstein, P. Yin, *Science* **2014**, *344*, 65.
[5] A. Levin, T. A. Hakala, L. Schnaider, G. J. L. Bernardes, E. Gazit, T. P. J. Knowles, *Nat. Rev. Chem.* **2020**, *4*, 615.
[6] a) A. D. Malay, N. Miyazaki, A. Biela, S. Chakraborti, K. Majsterkiewicz, I. Stupka, C. S. Kaplan, A. Kowalczyk, B. M. A. G. Piette, G. K. A. Hochberg, D. Wu, T. P. Wrobel, A. Fineberg, M. S. Kushwah, M. Kelemen, P. Vavpetic, P. Pelicon, P. Kukura, J. L. P. Benesch, K. Iwasaki, J. G. Heddle, *Nature* **2019**, *569*, 438; b) E. Golub, R. H. Subramanian, J. Esselborn, R. G. Alberstein, J. B. Bailey, J. A. Chiong, X. D. Yan, T. Booth, T. S. Baker, F. A. Tezcan, *Nature* **2020**, *578*, 172; c) Y. Nakamichi, N. Miyazaki, K. Tsutsumi, A. Higashiura, H. Narita, K. Murata, A. Nakagawa, *Structure* **2019**, *27*, 439.
[7] a) Y. J. Kim, J. B. Moon, H. Hwang, Y. S. Kim, G. R. Yi, *Adv. Mater.* **2022**, *35*, 2203045. b) Z. W. Li, Q. S. Fan, Y. D. Yin, *Chem. Rev.* **2022**, *122*, 4976.
[8] a) X. Y. Ji, Q. Li, H. Y. Song, C. H. Fan, *Adv. Mater.* **2022**, *34*; b) D. Fujita, Y. Ueda, S. Sato, N. Mizuno, T. Kumasaka, M. Fujita, *Nature* **2016**, *540*, 563; c) Z. Ji, R. Freund, C. S. Diercks, P. Hirschle, O. M. Yaghi, S. Wuttke, *Adv. Mater.* **2021**, *33*, 2103808.
[9] L. X. Cao, B. Coventry, I. Goreshnik, B. W. Huang, W. Sheffler, J. S. Park, K. M. Jude, I. Markovic, R. U. Kadam, K. H. G. Verschueren, K. Verstraete, S. T. R. Walsh, N. Bennett, A. Phal, A. Yang, L. Kozodoy, M. DeWitt, L. Picton, L. Miller, E. M. Strauch, N. D. DeBouver, A. Pires, A. K. Bera, S. Halabiya, B. Hammerson, W. Yang, S. Bernard, L. Stewart, I. A. Wilson, H. Ruohola-Baker, et al., *Nature* **2022**, *605*, 551.
[10] S. Y. Li, H. B. Fang, S. Sadeghi, P. Bhovad, K. W. Wang, *Adv. Mater.* **2019**, *31*, 1805282.
[11] a) S. Pandey, M. Ewing, A. Kunas, N. Nguyen, D. H. Gracias, G. Menon, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 19885; b) S. S. Chen, J. F. Chen, X. D. Zhang, Z. Y. Li, J. F. Li, *Light Sci. Appl.* **2020**, *9*, 75. c) Y. H. Zhang, F. Zhang, Z. Yan, Q. Ma, X. L. Li, Y. G. Huang, J. A. Rogers, *Nat. Rev. Mater.* **2017**, *2*, 17019.
[12] a) P. S. Huang, S. E. Boyken, D. Baker, *Nature* **2016**, *537*, 320; b) N. P. King, W. Sheffler, M. R. Sawaya, B. S. Vollmar, J. P. Sumida, I. Andre, T. Gonen, T. O. Yeates, D. Baker, *Science* **2012**, *336*, 1171.
[13] a) N. Nelson, A. Ben-Shem, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 971; b) J. A. Marsh, S. A. Teichmann, *Annu. Rev. Biochem.* **2015**, *84*, 551; c) J. T. Canty, R. Tan, E. Kusakci, J. Fernandes, A. Yildiz, *Annu. Rev. Biophys.* **2021**, *50*, 549.
[14] N. E. Chayen, E. Saridakis, *Nat. Methods* **2008**, *5*, 147.
[15] S. Leanza, S. Wu, X. Sun, H. J. Qi, R. R. Zhao, *Adv. Mater.* **2023**, 2302066.
[16] a) J. Zhu, N. Avakyan, A. Kakkis, A. M. Hoffnagle, K. Han, Y. Y. Li, Z. Y. Zhang, T. S. Choi, Y. Na, C. J. Yu, F. A. Tezcan, *Chem. Rev.* **2021**, *121*, 13701; b) T. G. W. Edwardson, M. D. Levasseur, S. Tetter, A. Steinauer, M. Hori, D. Hilvert, *Chem. Rev.* **2022**, *122*, 9145.
[17] a) J. E. Padilla, C. Colovos, T. O. Yeates, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 2217; b) I. G. Johnston, K. Dingle, S. F. Greenbury, C. Q. Camargo, J. P. K. Doye, S. E. Ahnert, A. A. Louis, *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119*, 2113883119.
[18] Y. N. Geng, G. van Anders, P. M. Dodd, J. Dshemuchadse, S. C. Glotzer, *Sci. Adv.* **2019**, *5*, aaw0514.
[19] I. Stupka, Y. Azuma, A. P. Biela, M. Imamura, S. Scheuring, E. Pyza, O. Woznicka, D. P. Maskell, J. G. Heddle, *Sci. Adv.* **2022**, *8*, eabj9424.
[20] a) R. Divine, H. Dang, G. Ueda, J. A. Fallas, I. Vulovic, W. Sheffler, S. Saini, Y. T. Zhao, I. X. Raj, P. A. Morawski, M. F. Jennewein, L. J. Homad, Y. H. Wan, M. R. Tooley, F. Seeger, A. Etemadi, M. L. Fahning, J. Lazarovits, A. Roederer, A. C. Walls, L. Stewart, M. Mazloomi, N. P. King, D. J. Campbell, A. T. McGuire, L. Stamatatos, H. Ruohola-Baker, J. Mathieu, D. Veesler, D. Baker, *Science* **2021**, *372*, 47; b) S. L. Lovelock, R. Crawshaw, S. Basler, C. Levy, D. Baker, D. Hilvert, A. P. Green, *Nature* **2022**, *606*, 49.
[21] a) R. W. Harkness, Z. A. Ripstein, J. M. Di Trani, L. E. Kay, *J. Am. Chem. Soc.* **2023**, *145*, 13015; b) H. Malet, F. Canellas, J. Sawa, J. Yan, K. Thalassinos, M. Ehrmann, T. Clausen, H. R. Saibil, *Nat. Struct. Mol. Biol.* **2012**, *19*, 152; c) X. C. Bai, X. J. Pan, X. J. Wang, Y. Y. Ye, L. F. Chang, D. Leng, J. L. Lei, S. F. Sui, *Structure* **2011**, *19*, 1328; d) T. Krojer, J. Sawa, E. Schafer, H. R. Saibil, M. Ehrmann, T. Clausen, *Nature* **2008**, *453*, 885; e) T. Krojer, M. Garrido-Franco, R. Huber, M. Ehrmann, T. Clausen, *Nature* **2002**, *416*, 455; f) J. S. Jiang, X. F. Zhang, Y. Chen, Y. Wu, Z. H. Zhou, Z. Chang, S. F. Sui, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 11939; g) S. Kim, R. A. Grant, R. T. Sauer, *Cell* **2011**, *145*, 67.
[22] J. Sawa, H. Malet, T. Krojer, F. Canellas, M. Ehrmann, T. Clausen, *J. Biol. Chem.* **2011**, *286*, 30680.
[23] J. H. Na, A. A. Evans, J. Bae, M. C. Chiappelli, C. D. Santangelo, R. J. Lang, T. C. Hull, R. C. Hayward, *Adv. Mater.* **2015**, *27*, 79.
[24] Y. Kim, H. Yuk, R. K. Zhao, S. A. Chester, X. H. Zhao, *Nature* **2018**, *558*, 274.
[25] K. Kuribayashi-Shigetomi, H. Onoe, S. Takeuchi, *PLoS One* **2012**, *7*, 51085.
[26] D. Lee, S. Cho, C. Park, K. R. Park, J. Lee, J. Nam, K. Ahn, C. Park, K. Jeon, H. Yuh, W. Choi, C. H. Lim, T. Kwon, Y. H. Min, M. Joo, Y. H. Choi, J. S. Lee, C. Kim, S. Kwon, *Nature* **2023**, *619*, 755.
[27] S. Yu, H. Cho, J. P. Hong, H. Park, J. C. Jolly, H. S. Kang, J. H. Lee, J. Kim, S. H. Lee, A. S. Lee, S. M. Hong, C. Park, S. Yang, C. M. Koo, *Nat. Commun.* **2017**, *8*, 721.
[28] T. Arhar, A. Shkedi, C. M. Nadel, J. E. Gestwicki, *J. Biol. Chem.* **2021**, *297*, 101282.

[29] D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, H. O. Smith, *Nat. Methods* **2009**, *6*, 343.
[30] A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, *Nat. Methods* **2017**, *14*, 290.
[31] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605.
[32] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Cryst.* **2010**, *66*, 486.
[33] P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *Acta Cryst.* **2010**, *66*, 213.
[34] S. H. W. Scheres, S. X. Chen, *Nat. Methods* **2012**, *9*, 853.
[35] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *Acta Cryst.* **2010**, *66*, 12.