Title: Dynamic assembly of pentamer-based protein nanotubes

Authors: Lukasz Koziej[1], Farzad Fatehi[2], Marta Aleksejczuk[1], Matthew J. Byrne[3†], Jonathan G. Heddle[1‡], Reidun Twarock[2,4], and Yusuke Azuma[1]*

Affiliations:

[1]Malopolska Centre of Biotechnology, Jagiellonian University; Krakow 30-387, Poland

[2]Departments of Mathematics, University of York; York YO10 5DD, United Kingdom

[3]Astbury Centre for Structural Molecular Biology, School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds; Leeds LS2 9JT, United Kingdom

[4]Department of Biology, University of York; York YO10 5DD, United Kingdom

*Corresponding author. Email: yusuke.azuma@uj.edu.pl

[†]Present address: Exscientia; Heatley Road, Oxford Science Park, Oxford OX4 4GE, United Kingdom

[‡]Present address: Department of Biosciences, Durham University; South Road, DH1 3LE, United Kingdom

Abstract: The molecular mechanisms and the geometrical theory underlying the polymorphic behavior of protein cages provide a basis for designing ones with the desired morphology and assembly properties. We show here that a circularly permuted cage-forming enzyme can controllably assemble into a variety of hollow spherical and cylindrical structures composed entirely of pentamers. A dramatic cage-to-tube transformation is facilitated by an untethered α-helix domain that prevents the 3-fold symmetry interaction and imparts a torsion between the building blocks. The unique double- and triple-stranded helical arrangements of subunits are mathematically optimal tiling patterns for this type of pentamer. These structural insights afford guidelines for the design of customized protein nanotubes for smart delivery and nanoreactor systems.

Engineering of biomolecular assemblies with precisely defined structure and functionality is the ultimate goal of bionanotechnology. While nucleic acids are the favored materials in the field (*1*), substantial efforts have been directed to the design of nanoarchitectures based on protein building blocks (*2*). Hollow spherical or cylindrical structures, called protein cages (*3-11*) or nanotubes (*12-15*), respectively, are particularly interesting in this context because of their prospective applications in delivery (*16-18*), catalysis (*19-21*), and nanomaterial construction (*22, 23*). Naturally occurring protein assemblies, such as viral capsids, provide design concepts as well as reengineering platforms for customized nanodevice development (*24, 25*).

Many viral coat proteins can assemble into particles with a range of sizes and shapes when reconstituted *in vitro* (*26, 27*). Geometrical patterns defined by quasi-equivalence (Casper-Klug) theory (*28*) explain such polymorphic behavior. The variable number of hexamers filling in the gap between pentameric subunits at the vertices results in cage expansions and irregular forms (*29, 30*). Such a dynamic, polymorphic feature is useful for customizing assemblies (*27, 31*), as demonstrated with the cowpea chlorotic mottle virus (CCMV) coat protein, which assembles around DNA origami structures and protects them from degradation (*32*).

Unique polymorphic behavior has been observed for engineered variants of *Aquifex aeolicus* lumazine synthase (*33, 34*). While the wild-type protein, AaLS-wt, self-assembles into a ~16-nm dodecahedral structure composed of 60 identical monomers (Fig. 1A) (*35*), the negatively supercharged variants AaLS-neg and AaLS-13 (*36, 37*) adopt expanded 360- and 720-mer assemblies, respectively, constructed entirely from pentameric capsomers (*38*). Moreover, a circularly permuted variant of AaLS, cpAaLS(119) (Fig. 1B), forms not only ~24-nm and ~28-nm expanded spherical cages, but also straight ~24-nm-wide tubes of variable length (*39*). Although such characteristics potentially present novel design principles (*31*), the geometric blueprints and the molecular mechanisms underlying this polymorphic behavior have remained unknown. Using cryo-EM and mathematical modeling, we here elucidate how pentameric building blocks can controllably and dynamically assemble into specific spherical and tubular structures in preference to the other possible particle morphologies.

Salt- and pH-dependent assembly of cpAaLS(119)

We previously designed two circularly permuted variants of AaLS and confirmed that their morphologies depend on the positions of the newly generated N- and C- termini (*39-41*). The native terminal amino acids were connected via an octapeptide linker using genetic fusion and new sequence termini were introduced either between residues 84 and 85 or 119 and 120, yielding cpAaLS(84) or cpAaLS(119), respectively (Fig. 1B). While cpAaLS(119) exhibits polymorphic behavior as discussed above, the cpAaLS(84) protein forms a ~16-nm diameter homogeneous cage structure, similar to AaLS-wt, and serves as a control for the experiments described below (Fig. 1C).

While characterizing the cpAaLS(119) variant, we unexpectedly found that the cage-like structures disassemble into fragments at low ionic strength and alkaline conditions. The protein was heterologously produced in *Escherichia coli* and isolated using ion exchange chromatography. Subsequent buffer exchange to 5 mM Tris-HCl buffer at pH 8.5 resulted in almost completely disassembled fragments, confirmed by size-exclusion chromatography coupled with right/low-angle light scattering detectors (SEC-RALS/LALS) (fig. S1).
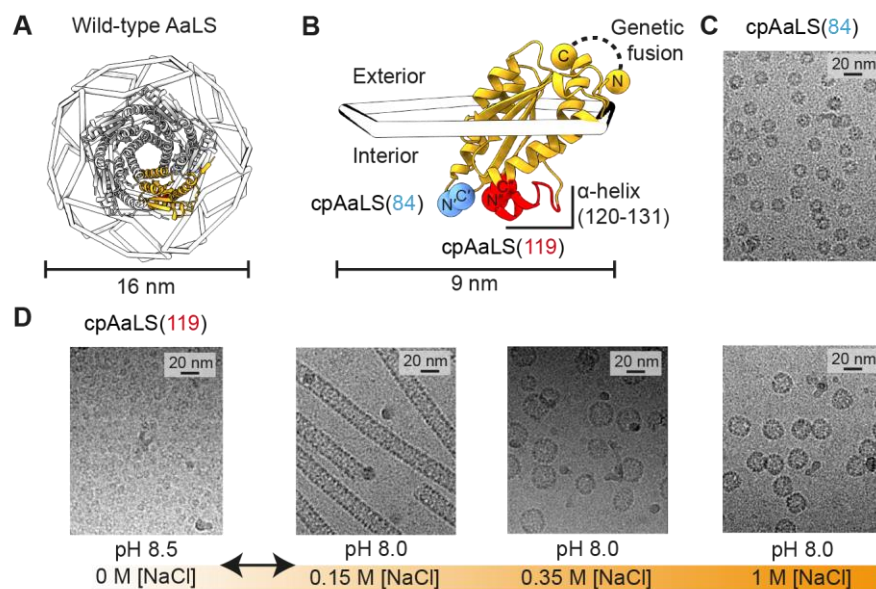
Fig. 1 Assembly control of circularly permuted AaLS. (**A**) Structure of the dodecahedral wild-type AaLS cage (PDB ID 1HQK), shown as 12 wire pentagons with a ribbon diagram of a representative pentamer (grey) and protomer (orange). (**B**) Design of the circularly permuted variants, cpAaLS(84) and cpAaLS(119). The peptide linker connecting the native N- and C-termini (GTGGSGSS) is shown with a black dashed line. The new termini, C′(84) and N′(85) (blue) or C″(119) and N″(120) (red), are indicated by spheres. The α-helix(120-131), untethered by circular permutation for cpAaLS(119), is highlighted in red. (**C,D**) Cryo-EM micrographs of the cpAaLS(84) cage (C) and the NaCl- and pH-dependent cpAaLS(119) assemblies (D).

Isolation of the cpAaLS(119) capsomers enabled systematic investigation of the reassembly process. Cage fragments were subjected to a rapid buffer exchange to 50 mM Tris-HCl buffer at pH 8.0 containing varying concentrations of NaCl (fig. S2A), and the resulting assemblies were analyzed by SEC and cryo-electron microscopy (cryo-EM) (Fig. 1D, and fig. S2, B to E). While remaining as unassembled fragments in the absence of salt (fig. S2B), adding 0.15 M NaCl facilitated cpAaLS(119) tubular formation with ~92% yield (fig. S2C). Further increasing NaCl concentration to 0.35 M yielded a mixture of the tubes and ~28-nm spherical cages ~44% and ~46%, respectively (fig. S2D). In 1 M NaCl, the tubes remained a 2% minority, while the ~28-nm cage constituted ~30% and the smaller ~24-nm cages dominated at ~59% (fig. S2E). These results demonstrated that cpAaLS(119) assembly can be controlled by adjusting the ionic strength of the solution. As NaCl concentration increases, unassembled fragments are preferentially transformed into 24-nm wide tubes, as well as 28-nm-, and 24-nm spheres.

Small changes in buffer pH can modulate the salt-dependent assembly of cpAaLS(119). Lowering pH tends to favor the assemblies that appear at high ionic strength and *vice versa*. For example, with 1 M NaCl at pH 8.5, 8.0, and 7.5, the proportion of 24-nm spherical cages was 9%, 59%, and 74%, respectively (fig. S2E). Furthermore, salt/pH-dependent cpAaLS(119) cage assembly is fully reversible. When the tube and spherical cages were subjected to buffer exchange into 50 mM Tris-HCl buffer at pH 8.5, the protein was completely converted into disassembled cage fragments (fig. S3).

To test if the cpAaLS variants retained the extreme thermal stability from the parent AaLS-wt, which has a melting temperature (T_m) greater than 120°C (*35, 42*), we performed a thermal shift assay based on tryptophan fluorescence, coupled with dynamic light scattering, at different pH and NaCl concentrations (fig. S4A). Irrespective of the tested buffer conditions, cpAaLS(84) did not exhibit substantial fluorescent changes up to 110°C, indicating little protein unfolding (fig. S4B). Temperature-dependent protein denaturation was observed for cpAaLS(119) with a T_m of 88-103°C, where higher ionic strength and lower pH tend to increase the thermal stability. This variant also showed a decrease in size, probably due to partial fragmentation, at 70-80°C under conditions that promote tube formation (fig. S4C, pH 7.5 and 0.15 M NaCl). Topological and/or morphological alteration likely leads to a relatively lower heat tolerance of cpAaLS(119) than that of AaLS-wt.

Geometric blueprints of the cpAaLS assemblies

Control over cpAaLS(119) assembly by salt and pH allowed us to determine the structures of individual morphologies using cryo-EM single particle and helical reconstruction (Fig. 2, and figs. S5 and S6). As previously hypothesized (*39*), all the structures were found to consist exclusively of the pentameric building blocks. The cpAaLS(84) assembly resembles the AaLS-wt cage, where each pentamer interacts with five neighboring subunits in a dodecahedral arrangement (fig. S5A). In contrast, cpAaLS(119) assemblies are non-quasi equivalent, where at least 1 interface of the pentameric subunits remains uncontacted. The 24-nm and 28-nm spherical cages are composed of 24- and 36-pentamers, respectively, and both have tetrahedral symmetry (Fig. 2, C and D), resembling those formed by the previously engineered AaLS variants, NC-1 and AaLS-neg (*38, 43*). In the tubular structure, the building blocks are arranged as a triple-stranded helix (Fig. 2B), which is unique and has never been seen for any natural or engineered proteins.

The transformation of cpAaLS assemblies is accompanied by changes in the number of connections between the constituent pentamers. While the number of pentamer-pentamer contacts remains 0 or 1 without NaCl at pH 8.5, judged by the 2D-averaged cryo-EM images (Fig. 2, A and E), the mean contact number per pentamer increases to 3, 3.7, and 4 in the tubular, 28-nm-, and 24-nm spherical cages, respectively (Fig. 2, F to H). This trend suggests that increasing salt and/or lowering pH stabilizes inter-pentamer interactions.

AaLS is known to have a higher number of ionic interactions and hydrogen bonds at the pentamer-pentamer interface compared to an analogous lumazine synthase derived from *Bacillus subtilis* (*35*). These charge-driven interactions have been hypothesized to contribute to its extreme thermal stability. However, the salt-dependent increase in the number of contacts observed for cpAaLS(119) structures suggests that hydrophobic interactions are the major driving force for their assembly. Meanwhile, the pH-dependency can be explained by the net negative charge of AaLS protein, which has a theoretical isoelectric point of 5.8. An increase in pH potentially endows the pentamers with increased negative surface charge, weakening inter-pentameric interactions due to charge repulsion.
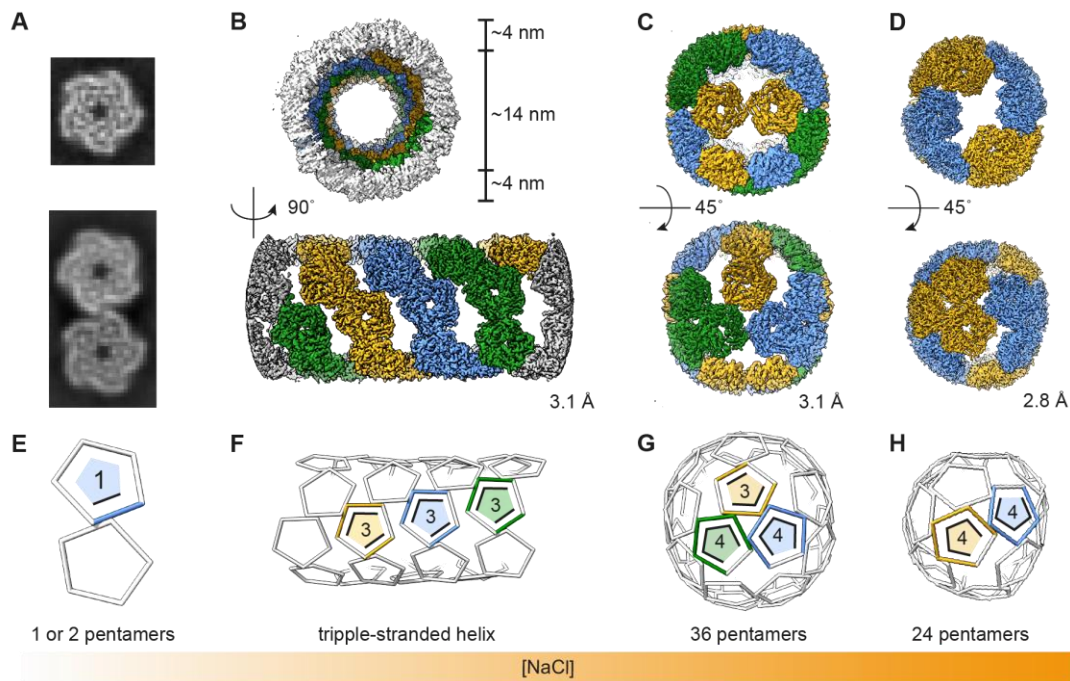
Fig. 2 Cryo-EM structures of the cpAaLS(119) assemblies. (**A-D**) 2D classes (A) and 3D maps (B-D) of the cpAaLS assemblies, where colors (blue, orange, or green) indicate individual threads in the helical structure (B) or symmetry-related pentameric subunits in the spherical cages (C,D). The resolution of the final 3D reconstructions (GS-FSC at 0.143 cut-off) is provided at the right corner of each map. (**E-H**) The corresponding wire representation of the cpAaLS assemblies with the number of contacts per each asymmetric pentamer. Images are not to scale.

Molecular mechanisms underlying the polymorphic behavior

The atomic model of the cpAaLS(84) cage revealed that all the constituent pentamers interact with each other in the same manner as in the AaLS-wt assembly (Fig 3A). The 2-fold symmetrical interface consists of a hydrophobic patch (L8, L141, and W137) surrounded by a hydrogen bond (H41) and ionic interactions (e.g. R40 and E5) (fig. S7, A and B) while an α-helix region (120-131) forms 3-fold symmetric hydrophobic clusters (I121 and I125) in the cage interior (Fig 3, B and C).

The bonding network observed at the 2-fold symmetry interface of AaLS-wt is approximately preserved for all the interacting pentamers in the cpAaLS(119) assemblies (fig. S7, C to F). There is no swapping in the amino acid interaction partners in this region upon circular permutation and morphology change. In marked contrast, the (pseudo) 3-fold symmetrical interaction interface of cpAaLS(119) assemblies lacks cryoEM density corresponding to the N-terminal α-helix(120-131) (Fig. 3, D to F). This finding suggests that the α-helix is unable to form the native-like hydrophobic cluster, likely due to the disconnection of residues 119 and 120 for circular permutation (Fig. 1B).
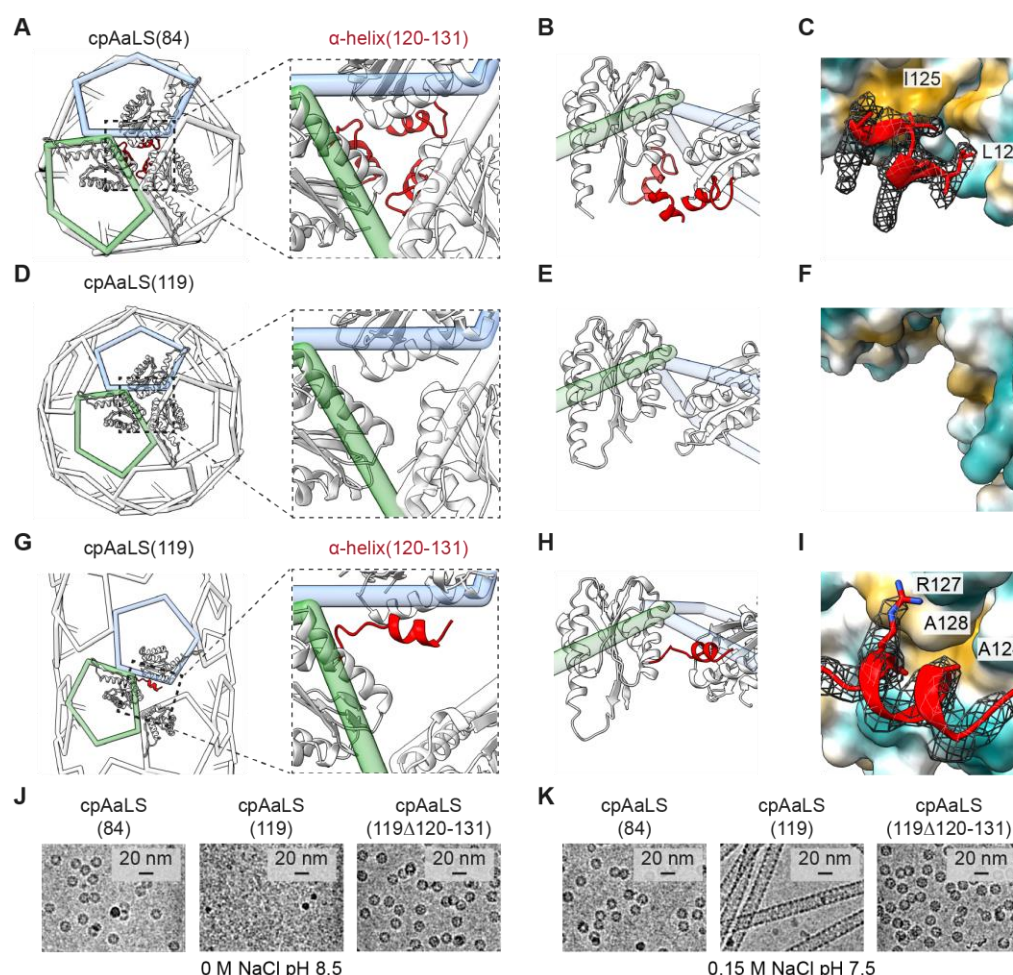
Fig. 3 'Untethered' α-helix(120-131) facilitating the dynamic assembly of cpAaLS(119). (A) Wire diagram of the cpAaLS (84) assembly with an enlarged view of the 3-fold symmetry region. Three interacting monomers are shown as a ribbon with α-helix(120-131) highlighted in red. (**B**) Rotated side view of a pentamer pair (green and blue wire). The α-helix(120-131) domain from another monomer at the front is also shown to present the interaction at the 3-fold symmetry region in the cpAaLS(84) cage lumen. (**C**) Atomic interaction mode of the α-helix(120-131) domain in the cpAaLS(84) assembly. A unit is shown as a ribbon with amino acid side chains at the interface with the corresponding cryo-EM density map (mesh), and the interacting partners as hydrophilic (cyan) and hydrophobic (light brown) surfaces. (**D-I**) The corresponding representations for the cpAaLS(119) 24-pentameric spherical cage (D-F) and the tubular assembly (G-I), where the α-helix(120-131) is structurally disordered and was not modeled (D-F), or flipped to interact with an alternative surface (G-I), respectively. Panel (F) shows the same region as (C) to present the lack of cryoEM density corresponding to the α-helix(120-131) region. (**J,K**) Cryo-EM images of the cpAaLS(119) variant lacking α-helix(120-131), cpAaLS(119Δ120-131), compared to those of cpAaLS(84) and cpAaLS(119).

In cpAaLS(119) tubes, the 'untethered' α-helix(120-131) binds to the adjacent pentamer surface, which is a solvent-exposed region in wild-type-like assemblies (Fig. 3, G to I). Because of this non-native interaction, the α-helix(120-131) appears to block 3-fold symmetrical pentamer-pentamer interactions, which explains the uncontacted interfaces observed in the

cpAaLS(119) assemblies (Fig. 2, E to H). Indeed, the deletion variant lacking the α-helix(120-131) domain, cpAaLS(119Δ120-131), does not exhibit polymorphic behavior, but assembles into only wild-type-like ~16-nm spherical cages (Fig. 3, J and K). These results prove that the untethered α-helix(120-131) domain is essential for the non-native, expanded cage formation of cpAaLS(119).

In thermal shift assays, the cpAaLS(119Δ120-131) cage showed a similar behavior to cpAaLS(84): having no denaturation up to 110 °C (fig. S4). These results suggest that the reduced thermal stability of cpAaLS(119) is mainly due to the morphology, in which constituent pentamers lack one or two interactions with their neighbors, rather than the loss of the hydrophobic core in the region of the 3-fold symmetry axis.

The binding of the untethered α-helix(120-131) to neighboring pentamers appears to be weak and occasional. This is suggested by an atomic-level interaction mode in which an arginine (R127) and two alanine (A128 and A124) residues from the helical domain contact a serine in the linker connecting the native termini and a hydrophobic cleft formed between intra-pentameric monomers, respectively (Fig. 3I). Furthermore, substantial cryo-EM density corresponding to the α-helix(120-131) was found only for 2 protomers in each pentamer constituting the tubular assembly and invisible in the spherical assemblies. The dynamic nature of the binding, which partially blocks pentamer-pentamer interactions, likely reflects multiple assembly states of cpAaLS(119).

Capsomer interaction angles facilitating the tubular formation

The unique triple helix is not the only tubular structure formed by cpAaLS(119). In the cryo-EM micrographs, short and bent caterpillar-like objects, referred to as "twisted tubes" were also observed albeit at a very low frequency (Fig. 4A, left). Due to the limited number of particles, helical reconstruction was not initially possible. However, introduction of C37S and A85C mutations into cpAaLS(119) to give cpAaLS(119, C37S, A85C) and a modified assembly protocol (fig. S8) surreptitiously resulted in enrichment of the twisted tubular structure and allowed subsequent cryo-EM analysis (Fig. 4A, right). Since the twisted tubular assemblies were heterogeneous, we divided particle images into 3 classes and determined the structures individually (fig. S9).

All three distinct twisted tube structures are composed of two compacted helical threads with 0-, 18-, and 28-Å gaps between them (Fig. 4B). This spring-like arrangement rationalizes the bending tendency. We also realized that one of the four pentamer interactions in the twisted tube has an acute bending angle (Fig. 4C, P_0-P_4 interaction). A possible disulfide bridge between two of the introduced cysteine residues at position 85 seems to support the unusual interaction mode, enhancing the twisted tubular formation (fig. S10).

As highlighted by the twisted tubes, the capsomer's interaction angles and the entire morphology are interlinked. While each pentamer-pentamer interface in the regular AaLS-wt assembly is tightly fixed in a single pattern by the 2-fold and two 3-fold symmetry interactions, removal or modulation of these contacts by reengineering likely leads to a flexible connection and expanded cage-like structures. To analyze the relative positioning of the pentamers in the cpAaLS assemblies systematically, we built a script that calculates the bending and torsion angles of two interacting polygon-shaped multimers (Fig. 4D). We named the computational angle generator "AngelaR".
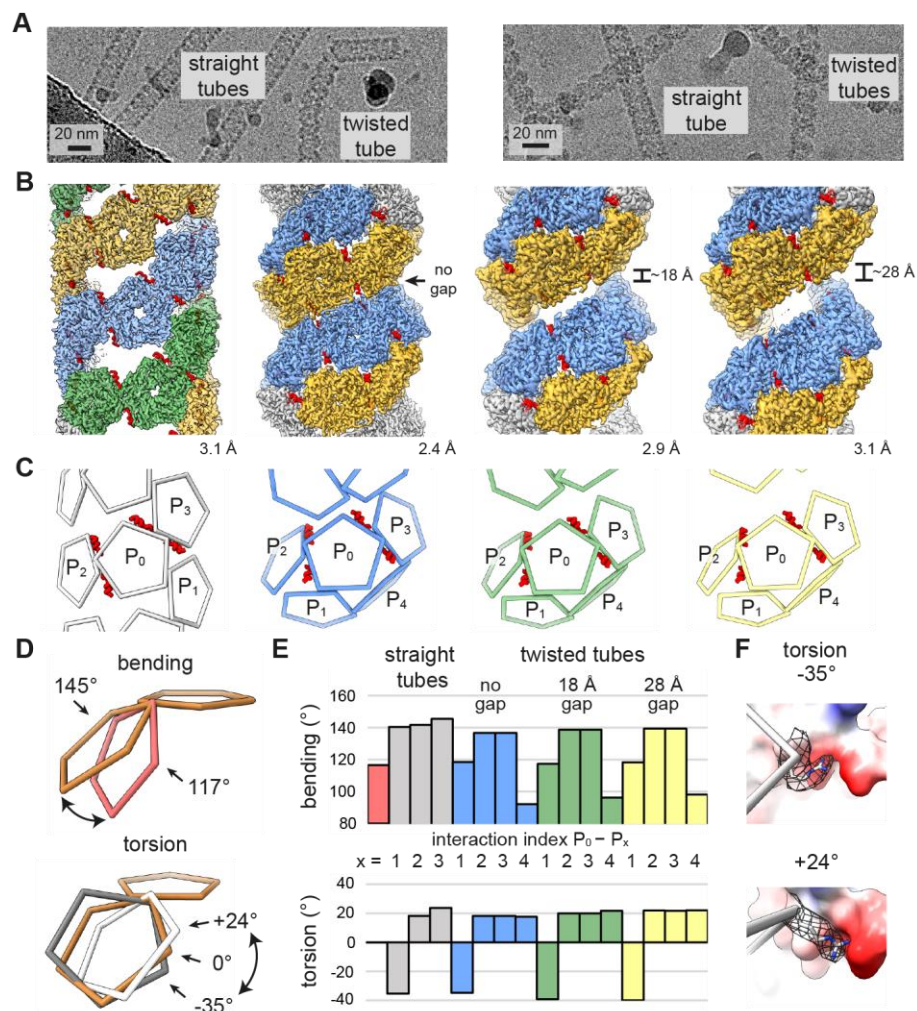
Fig. 4 Capsomer interaction angles in the cpAaLS(119) tubes. (**A**) Cryo-EM micrographs of cpAaLS(119) (left) and cpAaLS(119, C37S, A85C) (right). (**B**) Cryo-EM maps of the straight tube composed of 3 evenly spaced helical strips (green, orange, and blue) and the twisted tubes featuring a variable gap (0, ~18, or ~28 Å) between dual strips (blue and orange). The resolution of the final 3D reconstructions (GS-FSC at 0.143 cut-off) is shown at the right corner of each map. The maps are not to scale. (**C**) The corresponding wire representations showing interactions of a pentamer (P_0) with neighbors (P_{1-4}). The α-helix(120-131) domains that accompany the intrathread interactions (P_0-$P_{2/3}$), otherwise invisible, are highlighted as red ribbons. (**D, E**) Conceptual representation (D) and the measured values (E) of the bending (top) and the torsion angles (bottom) between two pentamers. The wire diagrams show a pentamer-pentamer interaction in the cpAaLS(84) cage (red, top) and the cpAaLS(119) straight tubes (white and grey, bottom) with a reference having 0° torsion and 145° bending angles (brown). The interactions in the bar graph are colored and indexed as in (D), with those of the cpAaLS(84) cage (red bar) shown for comparison. (**F**) Two edges between interacting pentamers with external torsion angles (-35° top, 24° bottom) in the cpAaLS(119) straight tube. The arginine side chain at the right vertex (R40, shown as grey sticks with mesh for the corresponding cryo-EM map) is flipped to maintain its interaction with a negatively charged patch (red surface) in the opposite pentamer.

Analysis of the cpAaLS spherical assemblies using AngelaR showed that an increase in the bending angles between interacting pentamers results in larger-sized structures, as reported previously (*38*) (fig. S11). In the twisted tubes, the increased gap between the compacted helical threads is accompanied by decreased bending and an increased torsion angle, a tendency agreeing with computational simulation using a pentagon-based helical model (fig. S12). In both the straight and twisted tubes, the pentamer-pentamer interaction adopts a wide range of torsion angles: -35°, 18°, and 24° for the straight tube, for example. The prominent negative torsion angles enable interthread dockings in these helical structures (Fig. 4, C and E, P_0-P_1 interaction).

AaLS proteins accept variable interaction angles using the flexible nature of the amino acid side chain. At the 2-fold symmetry interfaces of the cpAaLS(119) straight tube, for instance, an arginine residue (R40) is flipped to retain the interaction with the same negatively charged surface of the neighboring pentamer (Fig. 4F), being accompanied by the interaction torsion-angle change.

When focusing on the individual thread of the tubular assemblies, the constituent pentamers are connected through consistently positive torsion angles (Fig. 4, C and E, P_0-$P_{2/3}$ interactions). This is required for helical arrangements of this type of pentameric building blocks, and the torsion angle defines the helical rise. If the torsion angle is zero degrees, as in the case of AaLS-wt assembly, the pentamers can form only a ring-shaped arrangement with no helical rise (Fig 4E, red bar). Notably, these intrathread capsomer interactions involve a pair of the untethered α-helices(120-131) that bind with a neighbor on the flipped position (Fig. 4, B and C, P_0-$P_{2/3}$ interactions), probably supporting their torsion angles in a certain range.

The α-helix (120-131) domain is structurally disordered and unseen in cryo-EM analysis of the cpAaLS(119) spherical cages (Fig. 3, D to F). Tubular assemblies have never been observed for other AaLS variants in which the α-helix(120-131) is tethered in the native position (*38, 43-45*) or removed (Fig 3, J and K). Considering these observations together, we conclude that the untethered α-helix(120-131), which blocks the 3-fold symmetry interaction (Fig. 3G) while imparting a consistent torsion angle to the subunits (Fig. 4, D and F, P_0-$P_{2/3}$ interactions), is the essential element for the transformation of the AaLS assembly from spherical to tubular structures.

Mathematical rationale for the pentagon-based tubular structures

The spectrum of protein assemblies formed by a specific building block can be classified via tiling theory (*29, 30, 46-49*). To gain geometrical insight into the cpAaLS(119) tubular assemblies, we simulated all the possible helical arrangements formed by pentagons on a planar lattice (Fig. 5A and fig. S13). The model consists of a pair of pentamers (red) arranged with vectors pointing along ($\vec{P}$) and across ($\vec{H}$) helical threads, where the geometrically possible tubular structures are characterized by the number of distinct strips (helicity number n_h) and the translation steps along $\vec{P}$ (periodicity number n_p). Some structures defined by these parameters, e.g. $(n_h, n_p) = (4, 4)$, require flipped pentagons, which can not be realized by asymmetric interaction surfaces of proteins (fig. S14). The cpAaLS(119) straight tubes adopt the tiling pattern with $n_h = 3$ and $n_p = 4$, one of the smallest structures among all the geometrically and biologically possible options (Fig. 5, A and B).
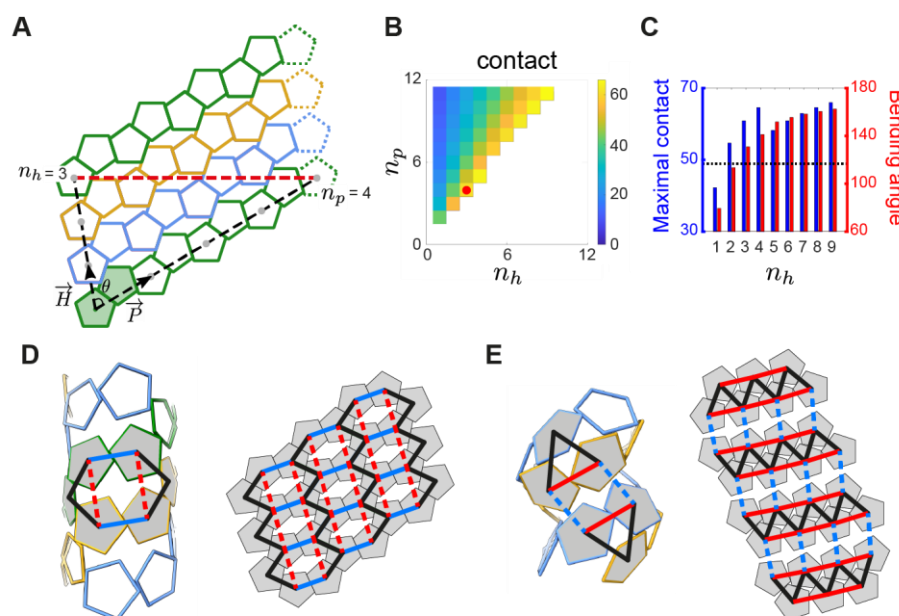
Fig. 5 Geometric rationale for the cpAaLS(119) tubular assembly. (**A**) A tiling representation of the cpAaLS(119) straight tube composed of three pentamer strips (green, blue, and orange). The lattice model is formed from periodic repeats of pentamer pairs (shaded in green) with vectors pointing along ($\vec{P}$) and across ($\vec{H}$) helical threads at an angle (θ). The tubular architecture is defined by the translation steps, helicity (n_h) and periodicity (n_p) numbers, for two pentamers distanced by a single helical turn (grey dots connected by a black dashed line). The pentamers connected with the red dashed line are identical in the 3D tubular structure. (**B**) Heatmap presenting the percentage contact area between pentagonal edges in the tubular model for different (n_h, n_p) combinations. The experimentally observed (n_h, n_p) = (3,4) is indicated by a red dot. The structures in the blank area are geometrically or biologically forbidden. (**C**) The maximally possible contact area (blue) and the bending angle (red) for given n_h over all the possible choices of n_p. The bending angle for the wild-type-like cpAaLS(84) cage is shown as a dotted line. (**D-E**) The interaction network rewiring between pentamers (shaded in grey) in the straight (D) and twisted tube (E), superimposed onto the 3D models (left) and the 2D tiling (right). Solid and dashed lines indicate contact and non-contact between pentamers, respectively. In the network transformation from the straight to the twisted tube, the blue contacts were lost, while the red ones were gained.

The parameters (n_h, n_p) = (3, 4) appear to provide the optimal tiling pattern for the capsomer interaction of the AaLS protein. We next analyzed the percentage contact length between adjacent pentagons for different (n_h, n_p) combinations, finding that the maximal contact for each n_h is consistently obtained with the smallest possible n_p (Fig. 5B). Comparison of the maximal contact for different n_h identifies (n_h, n_p) = (4, 5) as the local maximum and (n_h, n_p) = (3, 4) as the second-best option among the smaller n_h values (Fig. 5C, blue bars). We further benchmarked the bending angles between adjacent pentagons in the different geometric options (Fig. 5C, red bars), indicating that the (n_h, n_p) = (3, 4) gives the smallest possible values larger than that of the wildtype particle (Fig. 5C, dashed line). This angle is probably the most

10

favored for the contact surface of cpAaLS(119) pentamers. Meanwhile, these mathematical analyses suggest the possibility of other pentagon-based tubular assemblies by modulating the contact surface or bending angles between capsomers.

The seemingly distinct structure of the cpAaLS twisted tubes is geometrically related to that of the corresponding straight tubes. This is illustrated by the interaction network analysis that can predict alternatives by rewiring the connections of a given structure (*46*). Applying this approach, the cpAaLS straight tube features squashed hexagons drawn by connecting the centers of interacting pentamers (Fig. 5D, black and blue solid lines). The only alternative to this can be obtained by deleting two existing contacts (blue solid lines) while generating two new ones (red dashed lines), yielding a triangle-based network that corresponds to the tiling blueprint of the twisted tubes (Fig. 5E). This geometrical similarity could explain why a subtle change in the amino acid sequence and assembly condition resulted in the transformation between straight and twisted tubes. Furthermore, the network rewiring analysis implies that no other tiling patterns exist with this type of pentamer.

Conclusion

Successful control over the assemblies and their near-atomic resolution structures revealed the molecular origin of the dynamic, polymorphic nature of a circularly permuted cage-forming protein. A short peptide domain, which is untethered from the native position by topological rearrangement, inhibits the 3-fold symmetry interaction and holds the building blocks with a certain torsion angle, leading to the dramatic conversion of the wild-type dodecahedron to the previously unknown helical arrangements of pentameric subunits. Considering another cpAaLS variant, NC-4, adopts a quasi-equivalent assembly consisting of pentamers and hexamers (*43*), these results highlight the morphological plasticity of AaLS and circular permutation as a powerful approach for modulating or potentially customizing cage-like structures. The specific tiling patterns observed for the tubular structures are the most optimal blueprints for this particular pentamer. Meanwhile, geometrically feasible structures are more diverse, suggesting the potential existence of other types of assemblies formed by naturally occurring or engineered proteins. Notably, the mathematical approaches used in this study, AngelaR, the helix builder, the planar lattice model, and the interaction network analysis (*46*), are general and should be, therefore, useful for characterizing or predicting yet unknown polygon-based structures. While modular and readily modifiable cpAaLS(119) tubes are themselves an attractive platform for the prospective nanodevice development (*16*), our molecular and mathematical insights into the unique pentamer-based nanoarchitectures afford guidelines for the design of protein nanotubes with customized morphology and assembly dynamics.

References and Notes

1. N. C. Seeman, H. F. Sleiman, DNA nanotechnology. *Nat. Rev. Mater.* **3**, 17068 (2017).
2. R. V. Ulijn, R. Jerala, Peptide and protein nanotechnology into the 2020s: beyond biology. *Chem. Soc. Rev.* **47**, 3391-3394 (2018).
3. J. B. Bale *et al.*, Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389-394 (2016).

4. N. P. King *et al.*, Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103-108 (2014).

5. E. Golub *et al.*, Constructing protein polyhedra via orthogonal chemical interactions. *Nature* **578**, 172-176 (2020).

6. Y.-T. Lai, D. Cascio, T. O. Yeates, Structure of a 16-nm Cage Designed by Using Protein Oligomers. *Science* **336**, 1129-1129 (2012).

7. A. D. Malay *et al.*, An ultra-stable gold-coordinated protein cage displaying reversible assembly. *Nature* **569**, 438-442 (2019).

8. J. M. Fletcher *et al.*, Self-assembling cages from coiled-coil peptide modules. *Science* **340**, 595-599 (2013).

9. A. Ljubetič *et al.*, Design of coiled-coil protein-origami cages that self-assemble in vitro and in vivo. *Nat. Biotechnol.* **35**, 1094-1101 (2017).

10. A. Sciore *et al.*, Flexible, symmetry-directed approach to assembling protein cages. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 8681-8686 (2016).

11. I. Stupka *et al.*, Chemically induced protein cage assembly with programmable opening and cargo release. *Sci. Adv.* **8**, eabj9424 (2022).

12. H. Shen *et al.*, De novo design of self-assembling helical protein filaments. *Science* **362**, 705-709 (2018).

13. J. D. Brodin, S. J. Smith, J. R. Carr, F. A. Tezcan, Designed, Helical Protein Nanotubes with Variable Diameters from a Single Building Block. *J. Am. Chem. Soc.* **137**, 10468-10471 (2015).

14. M. Noji, Y. Sugita, Y. Yamazaki, M. Miyazaki, Y. Suzuki, Protein design of two-component tubular assemblies like cytoskeletons. *bioRxiv*, 2024.2004.2017.589732 (2024).

15. N. P. Bethel *et al.*, Precisely patterned nanofibres made from extendable protein multiplexes. *Nat. Chem.* **15**, 1664-1671 (2023).

16. B. Liu *et al.*, Protein Nanotubes as Advanced Material Platforms and Delivery Systems. *Adv. Mater.* **36**, 2307627 (2024).

17. Y. Wang, T. Douglas, Protein nanocage architectures for the delivery of therapeutic proteins. *Curr. Opin. Colloid Interface Sci.* **51**, 101395 (2021).

18. S. Bhaskar, S. Lim, Engineering protein nanocages as carriers for biomedical applications. *NPG Asia Materials* **9**, e371-e371 (2017).

19. D. J. Glover, D. S. Clark, Protein Calligraphy: A New Concept Begins To Take Shape. *ACS Cent. Sci.* **2**, 438-444 (2016).

20. R. A. Miller, A. D. Presley, M. B. Francis, Self-assembling light-harvesting systems from synthetically modified tobacco mosaic virus coat proteins. *J. Am. Chem. Soc.* **129**, 3104-3109 (2007).

21. Y. Wang, T. Douglas, Tuning properties of biocatalysis using protein cage architectures. *J. Mater. Chem. B* **11**, 3567-3578 (2023).

22. Y. J. Lee *et al.*, Fabricating Genetically Engineered High-Power Lithium-Ion Batteries Using Multiple Virus Genes. *Science* **324**, 1051-1055 (2009).

23. M. A. Kostiainen *et al.*, Electrostatic assembly of binary nanoparticle superlattices using protein cages. *Nat. Nanotechnol.* **8**, 52-56 (2013).

24. T. G. W. Edwardson *et al.*, Protein Cages: From Fundamentals to Advanced Applications. *Chem. Rev.* **122**, 9145-9197 (2022).

25. W. M. Aumiller, M. Uchida, T. Douglas, Protein cage assembly across multiple length scales. *Chem. Soc. Rev.* **47**, 3433-3469 (2018).

26. K. Majsterkiewicz, Y. Azuma, J. G. Heddle, Connectability of protein cages. *Nanoscale Adv.* **2**, 2255-2264 (2020).

27. F. Lie, T. N. Szyszka, Y. H. Lau, Structural polymorphism in protein cages and virus-like particles. *J. Mater. Chem. B* **11**, 6516-6526 (2023).

28. D. L. Caspar, A. Klug, Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1-24 (1962).

29. R. Twarock, A. Luque, Structural puzzles in virology solved with an overarching icosahedral design principle. *Nat. Commun.* **10**, 4414 (2019).

30. R. Twarock, A tiling approach to virus capsid assembly explaining a structural puzzle in virology. *J. Theor. Biol.* **226**, 477-482 (2004).

31. A. Khmelinskaia *et al.*, Local structural flexibility drives oligomorphism in computationally designed protein assemblies. *bioRxiv*, 2023.2010.2018.562842 (2023).

32. I. Seitz *et al.*, DNA-origami-directed virus capsid polymorphism. *Nat. Nanotechnol.* **18**, 1205-1212 (2023).

33. R. Ladenstein, M. Fischer, A. Bacher, The lumazine synthase/riboflavin synthase complex: shapes and functions of a highly variable enzyme system. *FEBS J.* **280**, 2537-2563 (2013).

34. Y. Azuma, T. G. W. Edwardson, D. Hilvert, Tailoring lumazine synthase assemblies for bionanotechnology. *Chem. Soc. Rev.* **47**, 3543-3557 (2018).

35. X. Zhang, W. Meining, M. Fischer, A. Bacher, R. Ladenstein, X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile *Aquifex aeolicus* at 1.6 A resolution: determinants of thermostability revealed from structural comparisons. *J. Mol. Biol.* **306**, 1099-1114 (2001).

36. F. P. Seebeck, K. J. Woycechowsky, W. Zhuang, J. P. Rabe, D. Hilvert, A simple tagging system for protein encapsulation. *J. Am. Chem. Soc.* **128**, 4516-4517 (2006).

37. B. Wörsdörfer, K. J. Woycechowsky, D. Hilvert, Directed evolution of a protein container. *Science* **331**, 589-592 (2011).

38. E. Sasaki *et al.*, Structure and assembly of scalable porous protein cages. *Nat. Commun.* **8**, 14663 (2017).

39. Y. Azuma, M. Herger, D. Hilvert, Diversification of protein cage structure using circularly permuted subunits. *J. Am. Chem. Soc.* **140**, 558-561 (2018).

40. L. Koziej, A. Gawin, Y. Azuma, "Lumazine Synthase Nanocompartments" in *Microbial Production of High-Value Products,* B. H. A. Rehm, D. Wibowo, Eds. (Springer International Publishing, Cham, 2022), pp. 335-355.

41. N. Terasaka, Y. Azuma, D. Hilvert, Laboratory evolution of virus-like nucleocapsids from nonviral protein cages. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 5432 (2018).

42. Y. Azuma, R. Zschoche, D. Hilvert, The C-terminal peptide of *Aquifex aeolicus* riboflavin synthase directs encapsulation of native and foreign guests by a cage-forming lumazine synthase. *J. Biol. Chem.* **292**, 10321-10327 (2017).

43. S. Tetter *et al.*, Evolution of a virus-like architecture and packaging mechanism in a repurposed bacterial protein. *Science* **372**, 1220-1224 (2021).

44. X. Zhang *et al.*, A structure-based model of the reaction catalyzed by lumazine synthase from *Aquifex aeolicus*. *J. Mol. Biol.* **328**, 167-182 (2003).

45. M. Hori *et al.*, Stimulus-responsive assembly of nonviral nucleocapsids. *Nat. Commun.* **15**, 3576 (2024).

46. F. Fatehi, R. Twarock, An interaction network approach predicts protein cage architectures in bionanotechnology. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2303580120 (2023).

47. T. Keef, J. P. Wardman, N. A. Ranson, P. G. Stockley, R. Twarock, Structural constraints on the three-dimensional geometry of simple viruses: case studies of a new predictive tool. *Acta Cryst. A* **69**, 140-150 (2013).

48. T. Keef, A. Taormina, R. Twarock, Classification of capped tubular viral particles in the family of Papovaviridae. *J. Phys.: Condens. Matter* **18**, S375 (2006).

49. T. Keef, A. Taormina, R. Twarock, Assembly models for Papovaviridae based on tiling theory. *Phys. Biol.* **2**, 175 (2005).

50. J. Schindelin *et al.*, Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676-682 (2012).

51. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290-296 (2017).

52. J. Zivanov, T. Nakane, S. H. W. Scheres, Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCrJ* **7**, 253-267 (2020).

53. S. He, S. H. W. Scheres, Helical reconstruction in RELION. *J. Struct. Biol.* **198**, 163-176 (2017).

54. R. Sanchez-Garcia *et al.*, DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* **4**, 874 (2021).

55. E. C. Meng *et al.*, UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).

56. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D* **66**, 486-501 (2010).

57. I. W. Davis *et al.*, MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375-W383 (2007).

58. E. Gasteiger, *et al.* "Protein Identification and Analysis Tools on the ExPASy Server" in *The Proteomics Protocols Handbook*, J. M. Walker, Ed. (Humana Press, Totowa, 2005), pp 571-607.

Author contributions:

Conceptualization: YA

Methodology: LK, FF, MB, RT, YA

Investigation: LK, FF, MA, JGH, RT

Visualization: LK, FF

Funding acquisition: RT, YA

Project administration: YA

Supervision: RT, YA

Writing – original draft: LK, FF, RT, YA

Writing – review & editing: LK, FF, JGH, RT, YA

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: The principal data supporting the findings of this study are provided in the figures and supplementary materials. The code implementing the mathematical analysis is available from GitHub: https://github.com/MathematicalComputationalVirology/TubeModeler. Cryo-EM maps and atomic models have been deposited in the Electron Microscopy Data Bank (EMDB) and the Worldwide Protein Data Bank (wwPDB), respectively, with the following accession codes: EMDB-51006 and PDB 9G3P (12-pentamer cage), EMDB-51005 and PDB 9G3O (24-pentamer cage), EMDB-51004 and PDB 9G3N (36-pentamer cage), EMDB-51003 and PDB 9G3M (straight tube), EMDB-51001 and PDB 9G3J (28-Å gap twisted tube), EMDB-51000 and PDB 9G3I (18-Å gap twisted tube), and EMDB-50999 and PDB 9G3H (0-Å gap twisted tube).

Supplementary Materials

Materials and Methods

Figs. S1 to S16

Tables S1 to S3