

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Jessalyn G Miller

Date

Design and characterization of natural and naturally-inspired helical protein assemblies

By

Jessalyn G. Miller
Doctor of Philosophy

Chemistry

Dr. Vincent P. Conticello
Advisor

Dr. Katherine Davis
Committee Member

Dr. Yonggang Ke
Committee Member

Accepted:

Kimberly Jacob Arriola
Dean of the James T. Laney School of Graduate Studies

Date

Design and characterization of natural and naturally-inspired helical protein assemblies

By

Jessalyn G. Miller

B.S., Western Washington University, 2016

M.S., Emory University, 2020

Advisor: Vincent P. Conticello, Ph.D.

An abstract of a dissertation submitted to the
Faculty of the James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Chemistry

2023

Design and characterization of natural and naturally-inspired helical protein assemblies

By Jessalyn G. Miller

Abstract

Protein helical assemblies are abundant in nature because they are useful to the cell and because they can form via a great diversity of mechanisms. In this work, multiple such assemblies are examined, from native-context *A. tumefaciens* T-pili to naturally derived CanA and Hyper2 proteins to α Reps, which are naturally-inspired but rationally engineered. Nature has designed proteins that are extremely durable against thermal energy and enzymatic activity. Human scientists have developed predictive algorithms to make educated guesses about the tertiary and even quaternary structures of proteins. However, these powerful tools are primarily informed by proteins from organisms that either maintain their own body temperature (e.g., humans) or grow within animals that do (e.g., *E. coli*). The proteins described in this work do not always fully auto-assemble; instead, their assembly is mediated by lipids or divalent cations, or they are co-assemblies of multiple proteins. In other words, these modeling tools are very good at predicting how thermally unstable proteins fold. But hyperthermostable proteins such as those described in this paper are ostensibly better for designing tools. They are more likely to be correctly folded in a given condition and they will not degrade as quickly. For these reasons, study in thermostable proteins is as valuable as ever for protein design and structural biology in general.

Design and characterization of natural and naturally-inspired helical protein assemblies

By

Jessalyn G. Miller

B.S., Western Washington University, 2016

M.S., Emory University, 2020

Advisor: Vincent P. Conticello, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Chemistry

2023

Acknowledgements

It goes without saying that I never could have done this research without my adviser, Vince Conticello. Thank you for taking a chance on me and bringing me onto your team when my other offers fell through. With hindsight, this was the right choice for me all along. You supported me to grow into a new area of research and begin a career in a rapidly developing field. You nominated me for the Chateaubriand Fellowship, which altered the course of my career in a very good way. Thank you for giving me the support and the projects I needed to become a better scientist.

Thank you to members of I2BC, especially Philippe Minard and Agathe Urvoas, who mentored me through a large part of my PhD work, and Agnès Mesneau and Marie Valerio-Lepiniec, who helped me settle into life in France.

Thank you to Ricardo Guerrero-Ferreira, the director of the Robert Apkarian Electron Microscopy Core during most of my time at Emory, for your patient mentorship during my degree and my job search. You were an important resource for planning the next part of my life.

Thank you to my friends in my lab in and the department. It was so good to see you every day at work. I am so excited for every one of your futures.

Thank you to my beloved friends, especially my friends at my weekly stitch'n'bitch. It was so refreshing to spend time with brilliant, talented women who didn't have the same job as me. And to my 20s 30s community at All Saints Episcopal Church, which is really a graduate student support group. At least we were all doing PhDs in different things!

My husband Joe dropped everything in San Francisco to move in a little Ford Focus with me to Atlanta. You have supported me in every way through this process, from holding the fort while I was in France for six months, to feeding the dog when I was in the lab late, and being a reliable friend during the hard parts of grad school and life. I hope I could have done it without you, but I'm glad I didn't have to. Thank you.

Throughout my academic career, I have always been grateful for the support of my loving family. My twin sister Michaela has gotten me through the roughest times and been there to celebrate the best times. I am so glad I get to go through life with you. My parents Linda and Charles inspired me to be curious about science and encouraged me to never stop learning. And thank you to my grandmother Margaret, who had three children when she was my age, at a time when research PhDs weren't an option for married women.

Lastly, even though he can't read, I am so grateful for my sweet hound dog Radish. You were my grad school puppy and now you're my post-grad school adventure buddy. I will always appreciate that you made me go walk outside, no matter what, and kept me company while I wrote this dissertation.

Abstract

Acknowledgements

Chapter 1 | Introduction: Structures of synthetic helical filaments and tubes based on peptide and peptidomimetic polymers

Introduction

Structural characterization of synthetic helical filaments

Structures of helical peptide assemblies

Short-peptide assemblies

Cross- β filaments

Cross- β nanotubes

Coiled-coil filaments

Tandem repeat assemblies

Foldamer-based helical assemblies

Conclusions

References

Chapter 2 | Deciphering the rules governing synthetic protein assemblies based on repeat proteins

Background

Results

Stabilizing the assembly by optimizing Brick:Brick binding interface

Kinetics of Brick C crystallization

Changing morphology of the Brick to alter assembly morphology

Thermostability

Stabilizing the interior of the filament

Stabilizing the exterior of the filament

Solubilizing filaments

Capping

Mixing Bricks

Discussion

Materials and Methods

Expression plasmids

Plasmid preparation.

Protein expression and preparation

Filament assembly.

Crystallization rates.

Circular dichroism.

Negative staining sample protocol.

Cryo imaging protocol.

References

Supplemental Information

Chapter 3 | Archaeal DNA-import apparatus is homologous to bacterial conjugation machinery

Introduction

Results

Archaeal conjugation pili are stoichiometric complexes of pilins and lipids

Conjugation pili of *Agrobacterium tumefaciens*

Different helical symmetries still allow quasi-equivalent interactions

Discussion

Methods

Cultivation of archaeal cells and preparation of pili samples

pED208 pili purification

A. tumefaciens pili purification

Cryo-EM sample preparation and data collection

Data processing and helical reconstruction

Sequence analyses

Mass spectrometry and lipidomics of *A. pernix*

Electron spray ionization and tandem MS-MS of pED208 F-pilus and pC58 T-pilus lipidomics

Nomenclature.

Lipid extraction.

PLA2 treatment.

Lipid standards.

Lipid spectrum acquisition.

Lipid identification and quantification.

Chemical and physical treatments of pED208 F-pilus and pC58 T-pilus.

Negative stain transmission electron microscopy of treated pED208 F-pilus and pTiC58 T-pilus

Reporting summary

Data availability

References

Acknowledgements

Author contributions

Competing interests

Additional information

Supplementary Information

Chapter 4 | Stacked cyclic peptide assemblies mediated by pyrene stacking

Introduction

Results and Discussion

Materials and Methods

Preparing filaments

Electron Microscopy:

Structural determination

Circular dichroism spectropolarimetry.

Conclusions

References

Chapter 5 | Structure and characterization of self-assembling archaeal cannula from metagenomic data

Introduction

Results

CanA

Hyper2

Materials and Methods

Protein expression

Nanotube assembly

Negative Stain transmission electron microscopy.

Circular dichroism spectropolarimetry.

Cryo-electron microscopy and structural determination.

Atomic force microscopy and quantitative nanomechanical mapping.

Analytical ultracentrifugation.

Discussion

Structural motifs

Role of ion or heat

Outlook

References

Supplemental Information

Figure 1.1 (a) Top and side projections of a helical polymer in which each dot represents an asymmetric unit of the corresponding assembly. (b) Helical net diagram corresponding to the symmetry of the helical polymer. A solid line connects protomers within the 1-start helix of the assembly. (c) Power spectrum derived from the amplitudes of the Fourier transform of this helical polymer in which layer lines are indicated corresponding to pitch (P) and rise (z) in reciprocal space. From S. A. Fromm and C. Sachse (2016) *Methods in Enzymology*, 579, 307–328, Academic Press Inc. Reprinted with permission from Elsevier..... 4

Figure 1.2. Side (a) and top (b) views of the atomic model of the phage IKe capsid. Side (c) and top (d) views of the atomic model of the phage Pf4 capsid. Single protofilaments corresponding to the 10-start and 11-start helices of the IKe and Pf4 capsids, respectively, are highlighted in red. 6

Figure 1.3 (a) Helical net diagram of IKe phage. (b) Helical net diagram of Pf4 prophage. (c) Calculated power spectrum of IKe phage. (d) Calculated power spectrum of Pf4 phage..... 9

Figure 1.4 Dimeric structure of A β ₁₋₄₂ as solved by Colvin et al. by ssNMR measurements. Only residues Q15-A42 are shown. (a) Overlay of the 10 lowest-energy NMR structures. Left monomer is shown in dark colors; right monomer is shown in pale colors. (b) CPK model showing backbone in gray, hydrophobic side chains are shown in yellow (solvent-exposed), gold, and orange (buried clusters). (c) Surface model of A β ₁₋₄₂. (d) Ribbon model of A β ₁₋₄₂. Reprinted with permission from Colvin et al. (2016). *Journal of the American Chemical Society*, 138 (30), 9663–9674. DOI: 10.1021/jacs.6b05129. Copyright 2016 American Chemical Society. 20

Figure 1.5 Dimeric structure of A β ₁₋₄₂ as solved by Gremer et al. by cryo-EM with helical reconstruction. (a) Opposing assemblies are staggered, producing a pseudo-2₁ screw axis in the filament. (b) Surface hydrophobicity of a representative segment the A β ₁₋₄₂ peptide filament. Brown represents hydrophobicity value of 4.5, white represents 0.0 according to the Kyte–Doolittle scale. (c, d) Model of the ‘ridge’ and ‘groove’ ends of the A β ₁₋₄₂ filament. Colors correspond to layers in (a). From Gremer et al. (2017). *Science*, 358 (6359), 116–119. Reprinted with permission from AAAS. 22

Figure 1.6. (a) View down the helical channel within the crystal structure of the Phe–Phe dipeptide. Density due to water molecules can be observed in the lumen of the tube. (b) View down the channel in the atomic model of the C7 helical filament derived from self-assembly of tetrapeptide 1-KMe ₃ . The ^D Phe– ^D Phe segment of each protofilament is highlighted in red.....	24
Figure 1.7. (a) Facial and side views within the atomic model corresponding to the tilted bilayer filament of peptide RADA-16. Arrows indicate the complementary electrostatic interactions between cross-strand arginines and aspartic acid residues. (b) Facial view of the atomic model of the cross-β fibrils of peptide HHQ. Zinc ions (gray spheres) bridge adjacent strands through coordination to histidine residues. (c) Facial view of the atomic model of the MAX1 bilayer filament. Adjacent hairpins in each layer are syn to each other, but anti to the corresponding hairpin in the other layer.....	28
Figure 1.8. Morphological variants of Aβ ₁₆₋₂₂ . (a) Atomic model of the laminated filament of Aβ ₁₆₋₂₂ observed at neutral pH. (b) Proposed unit cell describing the packing of peptides in the laminated filament of Aβ ₁₆₋₂₂ . (c) Atomic model of the bilayer nanotube of Aβ ₁₆₋₂₂ observed at acidic pH. (d) Expanded representation of the tilted packing of β-strands within a bilayer nanotube of Aβ ₁₆₋₂₂ . Reprinted with permission from Mehta et al. (2008), Journal of the American Chemical Society, 130 (30), 9829–9835. Copyright 2021 American Chemical Society.	33
Figure 1.9. Proposed mechanism of assembly and structural model of nanotubes derived from self-assembly of α-Sβ1 peptides. (a) Model of the initially formed helical ribbon with associated TEM image. (b) Model of the mature tubes with associated TEM image. (c) Cross-sectional depiction of the mature tube indicating the packing arrangement of peptides in the amphiphilic monolayer. (d) Sequence and stick model of the α-Sβ1 peptide. Residues are colored to indicate amphiphilic character based on water-octanol transfer free energies. (e) The orientation of the α-Sβ1 strands are shown in the context of the tape then leading to the nanotubes. The single peptides are represented as lines with hydrophobicity and hydrophilicity shown as orange and cyan, respectively. Reprinted with permission from Morris et al. (2013), Angewandte Chemie International Edition, 52 (8), 2279–2283.	35

Figure 1.10. (a) Sequence of macrocyclic β -sheet 1. (b) Macrocyclic β -sheet 1 nanotubes pack into a honeycomb-like crystal lattice. (c) Top and side views of nanotube formed by macrocyclic β -sheet 1. (d) Structure of dimeric subunit at the inner wall of the nanotube. (e, f) Different views of the tetrameric subunit at the outer wall of the nanotube. Reprinted with permission from Chen et al. (2017), *Journal of the American Chemical Society*, 139 (24), 8102–8105. Copyright 2021 American Chemical Society. ... 38

Figure 1.11 (a) Structural model of a nanotube derived from self-assembly of peptide A₆K. A single protofilament within the nanotube inclines at an angle of 52° with respect to the long axis of the assembly. Inset indicates the proposed antiparallel packing arrangement of peptides within a cross- β protofilament. (b) Structural model of a laminated ribbon derived from self-assembly of peptide A₈K or A₁₀K. The trajectory of single cross- β protofilament within the ribbon is highlighted. The quantity λ_r corresponds to the helical pitch of the ribbon. In both representations, the imposed helical hand was based on an arbitrary decision. Reproduced from Rüter et al. (2020), *Physical Chemistry Chemical Physics*, 22 (33), 18320–18327, with permission from the Royal Society of Chemistry. 40

Figure 1.12. Schematic illustration of the self-assembly of peptide nanotubes derived from site-directed mutagenesis of peptide Ac-KI₄K-NH₂. (a) Proposed model for self-assembly of the parent peptide Ac-KI₄K-NH₂. (b) Thin tubes are formed by changing side-chain branching and hydrophobicity, which influences the degree of lateral stacking of β -sheets. (c) A single mutation of isoleucine to valine in Ac-KI₃VK-NH₂ leads to an intermediate degree of sheet lamination and modest decrease in nanotube diameter. The assignment of helical hand was based on AFM measurements. Reprinted with permission from Zhao et al. (2018), *Small*, 14 (12), 1703216. 43

Figure 1.13. Initial model for the self-assembled nanotubes of the lanreotide peptide. Freeze-fracture micrographs of longitudinally (a, b) and transversely (c) fractured nanotubes indicated the formation of uniform-diameter assemblies. (d) Proposed arrangement of lanreotide monomers in the internal and external leaflets of the bilayer nanotube. (e) Initial structural model of the bilayer nanotube with an expansion of the internal and external leaflets corresponding to 2 × 3 asymmetric units. (f) Cross-sectional model of the bilayer nanotube indicating the proposed side-chain packing arrangement between

monomers at the leaflet interface. Reproduced from Chervy et al. (2019), *Langmuir*, 35 (32), 10648–10657, with permission from the American Chemical Society..... 45

Figure 1.14. Cryo-EM structure of lanreotide nanotubes. (a) Representative cryo-EM image of peptide nanotubes derived from self-assembly of lanreotide acetate. (b) Representative 2D class average derived from cryo-EM analysis of lanreotide nanotubes. (c) Density maps for the two main conformations, capped versus away, of the lanreotide peptide within the nanotubes. (d) Density map for the lanreotide nanotube in which the eight peptides in an asymmetric unit are highlighted within the assembly. An expansion of the asymmetric unit is depicted on the right in which the capped (orange) and away (cyan) conformations are indicated. 46

Figure 1.15. Peptide KFE8 self-assembles into helical ribbons and nanotubes. Representative cryo-EM images of KFE8 assembled at ambient temperature (a) or annealed at 90 °C (b). Scale bar = 20 nm. Red, black, and white arrows indicate the presence of bilayer ribbons, thinner tubes, and thicker tubes, respectively. Cross-sectional and transverse views of the atomic models for the ribbons (c, f), thinner tubes (d, g), and thicker tubes (e, h). The helical hand was assigned based on imaging data from AFM measurements. Reference: Wang, F.; Gnewou, O.; Wang, S.; Osinski, T.; Zuo, X.; Egelman, E. H.; Conticello, V. P. Deterministic chaos in the self-assembly of β -sheet nanotubes from an amphipathic oligopeptide. *Matter*, DOI:10.1016/j.matt.2021.06.037..... 47

Figure 1.16. (a) Helical-wheel diagrams of the SAF-p1, SAF-p2, and SAF-p2a peptides indicating the heptad periodicity. (b) Proposed model for the heteromeric association of SAF-p1 and SAF-p2 (a) peptides into a dimeric coiled-coil through sticky-ended self-assembly. (c, d) Transmission electron micrographs of the paracrystalline assembly of SAF-p1 and SAF-p2a co-assembly. (e, f) Simulated model of the packing of coiled-coil peptides within the SAF assembly. Reprinted from Sharp et al. (2012), *Proceedings of the National Academy of Sciences of the United States of America*, 109 (33), 13266–13271..... 50

Figure 1.17 Mechanism of self-assembly of coiled-coil lock washers into helical filaments. (a) Helical wheel diagram of designed coil-coil peptide 7HSAP1. (b) Sequence of the 7HSAP1 peptide indicating the

heptad registry. Basic and acidic residues involved in the intra-bundle and inter-bundle (axial stacking) interactions are highlighted in blue and red, respectively. (c) Proposed mechanism of self-assembly of heptameric lock washers of 7HSAP1 into filaments through electrostatic interactions between oppositely charged termini. The seven-helix bundle structure was derived from the crystal structure of GCN4-pAA (PDB: 2HY6). Reproduced with permission from Xu et al. (2013), *Journal of the American Chemical Society*, 135 (41), 15565–15578. Copyright 2021 American Chemical Society. 55

Figure 1.18. Blunt-ended coiled-coil oligomers self-assemble into peptide nanotubes. (a) Ribbon diagrams and orthogonal views of space-filling models for designed coiled-coil oligomers; CC-Di (red; PDB: 4DZM), CC-Tri (orange; 4DZL), CC-Tet (green; 3R4A), CC-Pent (turquoise; 4PN8), CC-Hex (blue; 3R3K), and CC-Hept (purple; 4PNA). (b) Proposed model for self-assembly of CC-Hex-T involving axial stacking of hexameric barrels into a filament. (c) Helical wheel diagram depicting the heptad register of the CC-Hex-T sequence. (d) Representative cryo-EM image of thermally annealed CC-Hex-T assemblies (scale bar = 100 nm). (e) Tomographic slice of a CC-Hex-T assembly from cryo-ET analysis depicting an approximately cylindrical cross section. (f) Tetragonal lattice model for packing of CC-Hex-T filaments in a para-crystalline array. Reprinted with permission from Burgess et al. (2015), *Journal of the American Chemical Society*, 137 (33), 10554–10562. Copyright 2021 American Chemical Society. 58

Figure 1.19. (a, b) Helical wheel diagrams indicating the heptad register of the Form I and Form II peptide sequences. (c, d) Representative cryo-EM images of the Form I (c) and Form II (d) nanotubes. (e, f) Atomic models fit into the 3D reconstructions of the Form I (e) and the Form II (f) nanotubes derived from cryo-EM analysis. Reprinted with permission from Egelman et al. (2015), *Structure*, 23 (2), 280–289. 62

Figure 1.20. (a) Peptide sequences of Form I variants in which the position of the arginine clasp motif is highlighted in red. Structural analyses of the 15-10-3 (b), the 29-24-3 (c), and the 36-31-3 (d) filaments. Representative cryo-EM images are shown on the left (scale bar = 50 nm). Atomic models fit into the respective 3D reconstructions are in the middle. Top views of the respective assemblies are shown on the

right. Reproduced under the Creative Commons License from Wang et al. (2021a). Nature Communications, 12 (1), 407.....	64
Figure 1.21. Single tandem repeat motifs self-assemble into stable nanotubes. (a, d) Representative cryo-EM images of LRV_M3Δ1 (a) and HEAT_R1 (d) filaments. (Scale bar = 100 nm.) (b, e) Atomic models fit into the 3D reconstructions of the LRV_M3Δ1 (b) and HEAT_R1 (e) nanotubes. The asymmetric units are highlighted for the respective filaments in blue, which, in the case of HEAT_R1, corresponds to a dimer of peptides. (c, f) Helical net diagrams for the LRV_M3Δ1 (c) and HEAT_R1 (f) nanotubes, in which the difference in helical hand between the respective 1-start helices is apparent. The helical nets show the unrolled surface lattice viewed from the outside of the filament. From Hughes et al. (2019), Proceedings of the National Academy of Sciences of the United States of America, 116 (29), 14456–14464.	69
Figure 1.22. Computational design of helical filaments from designed TRPs. Left to right: Computational models, representative cryo-EM images, cryo-EM-derived atomic models, and structural overlays between the computational model and experimental structure for designed helical filaments DHF58 (a), DHF119 (b), DHF91 (c), DHF46 (d), DHF79 (e), and DHF38 (f). From Shen et al. (2018), Science, 362 (6415), 705–709. Reprinted with permission from AAAS.....	71
Figure 1.23. Diameter of designed helical filament DHF58 can be controlled through the number of tandem repeats in the protomer. (a) Cross sections and side views of computational models based on the four-repeat cryo-EM structure. The number of repeats (n) is shown at the top. (b) Representative negative stain electron micrographs of the corresponding helical filaments. (c) 2D-class averages derived from electron micrographs of the respective filaments. From Shen et al. (2018), Science, 362 (6415), 705–709. Reprinted with permission from AAAS.....	73
Figure 1.24 De novo design of self-assembling βαβ peptides. (a) The computational design is based on the crystal structure of the amyloidogenic heptapeptide sequence GNNQQNY derived from the sup35 N-terminal prion domain. A loop–helix–loop segment constructed using flexible backbone design connects two β-strands with backbone conformations taken from the sup35 peptide. (b) Cross section	

of the computational model of the fibril in a view perpendicular to the fiber axis. The interactions between the two layers of sheets are stabilized by a steric zipper interface formed by interdigitating side chains at the strand interfaces. (c) Structural model the peptide filament with estimated lateral dimensions for a cross section. (d) Sequences of computationally designed peptides $\beta\alpha\beta\text{Zip}$ and $\beta\alpha\beta\text{Zip2}$. Residues highlighted in green correspond to the steric zipper interface, while residues highlighted in red correspond to the α -helix. From Kaltofen et al. (2015), *Journal of Molecular Biology*, 427 (2), 550–562. Reprinted with permission from Elsevier. 75

Figure 1.25. Amphiphilic peptoids block copolymer tiles self-assemble into hollow nanotubes. (a) Sequence and chemical structure of pNdc₁₈-b-pNte₁₈. (b) Schematic depiction of the structure of pNdc₁₈-b-pNte₁₈ tile in which the hydrophobic and hydrophilic domains are highlighted in green and blue, respectively. The gray arrow indicates the N to C chain trajectory. (c) Schematic representation of the packing of four layers of peptoid tiles within the proposed nanotube structures. (d) Cross section of a four-layer nanotube segment with depiction of side-chain arrangement and the stacking repeat distance. (e) Schematic representation of the proposed nanotube structure with amphiphilic tile packing indicated. (f) Cross-sectional view of the proposed packing of amphiphilic tiles for three peptoid polymers with approximate tube diameter for each cross section. Reprinted with permission from Sun et al. (2016), *Proceedings of the National Academy of Sciences of the United States of America*, 113 (15), 3954–3959. 76

Figure 1.26. Amphiphilic peptoid oligomers (APOs) self-assemble into highly ordered, crystalline, single-walled peptoid nanotubes. (a) Sequence and chemical structure of the APOs. Hydrophilic and hydrophobic blocks are indicated in blue and magenta, respectively. (b) Schematic representation of the proposed structure of the monolayer nanotube with color-coding of the hydrophilic and hydrophobic blocks. (c) Cross-sectional representation of the proposed model for the APO2 nanotubes. The molecular packing of the peptoid protomers is depicted along with the critical distances that define the packing arrangement. Reprinted under Creative Commons License from Jin et al. (2018), *Nature Communications*, 9 (1), 270..... 78

Figure 1.27. Oligoureases can adopt amphiphilic helical secondary structures that promote self-assembly into nanotubes. (a, b) Helical wheel diagrams of the decameric (H2) and dodecameric (H5) oligourea sequences indicating the pentad registry. The facial amphiphilicity can be detected from the pattern of polar residues (red) and non-polar residues (blue). (^uXaa indicates the urea analog of a canonical amino acid.) (c, e) Side and top views of the H2 nanotube derived from the corresponding crystal structure. (d, f) Side and top views of the H5 nanotube derived from the corresponding crystal structure. In each case, the individual protofilaments within the respective structures are color coded..... 80

Figure 1.28 (a) Helical wheel diagram of tripeptide Ac-β³[hLhIhA] indicating the triad registry in a lefthanded 14-helix conformation. (b) Side view of the triple-helical ribbon model of the Ac-β³[hLhIhA] filament in water. (c) Expansion of a four-layer helical stack from (b) highlighting packing at the trimeric interface in a stick representation. (d) Trimeric packing interface of a single-layer of the triple helix in which inter-residue contacts are indicated using a mesh depiction. 83

Figure 2.1. Three Brick B/Backbinder filaments in a crystalline arrangement. Surface charges of Brick B and Backbinder coassembly are thought to further stabilize ridges-into-grooves packing, leading to crystal formation. Negative charges (red) line the “groove” created by Backbinders and Bricks on successive turns of the helix. The positively-charged “ridge” along the convex surface of Backbinder fits into this groove and makes contact with the convex surface of the Brick. Figure is reused from Moreaud, et al., 2023. 114

Figure 2.2. A. Sequence of Brick C, arranged to show repeats. I1 and I2 from the Bait protein are highlighted in gold, and I3 is highlighted in orange. B. Model of Brick C showing the highlighted repeats. C,D. Schematic model of Brick C showing how Backbinder (purple) recognizes the reconstituted binding site. E. Model of Backbinder bound to two Brick C proteins. 115

Figure 2.3 A. Brick C/BBST assembly gel showing both Brick C and back binder in the pellet fraction after assembly at concentrations of 8, 4, 2, or 1 μM each protein. B. Negative-stain TEM micrograph showing crystalline superassemblies of Brick C and BBST. C. Cryo-EM micrograph confirming crystalline arrangement of proteins in B. D. Model of Brick C (lavender) and BBST (pink) coassembly. E.

Rate of crystallization monitored by absorbance at 350 nm for Brick B and Brick C, at two temperatures. Brick C is an improved version of Brick B with more nonpolar contacts at the Brick:Brick interface, and assembles faster than Brick B with less temperature dependence.	116
Figure 2.4 A. Model of Brick D (blue) assembled with Backbinder (pink) with cartoon showing screw symmetry. B. Negative-stain TEM image of Brick D/Backbinder crystals. Curving and cracking under the vacuum of the electron microscope is normal for nanoparticles that are flat in solution. C. Cryo-EM image of Brick D/back binder co-crystal with distinct moiré patterning. The crystal is flat in its hydrated state. D and E. Assembly curves of Brick D and Backbinder at 10 or 20 μ M concentration showing temperature and concentration dependence of assembly rate.	121
Figure 2.5. Thermostability monitored by circular dichroism spectropolarimetry. A. All α Reps have a distinctive α helical signature. B. At 95°C, there is some helicity but signatures are generally degraded. C. Thermal melt curves from 4 to 95°C. Bricks C and D do not have a thermal transition, but Backbinder begins to melt around 70°C. Backbinder melts at the same temperature whether soluble or in an assembly.....	124
Figure 2.6. A. Sequence of Brick F, arranged to show repeats. I1 and I2 from the Bait protein are highlighted in gold, and I3 is highlighted in orange. Repeats from A3 protein are highlighted in blue. B. Model of Brick F showing the highlighted repeats on the concave and convex faces of the protein. C,D. Schematic models of Brick F showing how Backbinder (purple) and α 2 recognize their reconstituted binding sites. E. Model of Backbinder and α 2 bound to two Brick F proteins.	125
Figure 2.7 A. Very long, twisted filaments of Brick F and α 2 only. B. Close-up from same sample as A showing two or three filaments in each bundle. C. Tape-like crystals of Brick F and α 2. D. Model of Brick F (blue, violet, green) assembled with only α 2 (yellow, red).	126
Figure 2.8. SDS-PAGE gel showing all constituent proteins in pellet fractions. α 2 is difficult to see in the Brick F/ α 2/Backbinder pellet sample.....	127
Figure 2.9. A. Negative stain TEM image of Brick F/Backbinder crystals, identical to Brick C/Backbinder. B. Cryo-EM image of Brick F/Backbinder/ α 2 condensate with a few free filaments in the	

thin ice patch toward the top of the hole. C. Model of Brick F (blue), Backbinder (pink), and a2 (green) complex..... 128

Figure 2.10. A. Brick C assembled with A3E3. Filaments form and join disordered, somewhat linear aggregates. B. Brick F and A3E3. Identical to Brick C. C. Short, irregular assemblies of Brick D and A3E3..... 129

Figure 2.11. A. 1:4 Caps to Brick C ratio abolishes filament formation and produces disordered aggregates of free protein. 1:16 (B) and 1:80 (C) ratios of Caps to Brick C both produce crystals, indicating that crystals can form regardless of filament length. No stoichiometric relationship is observed. However, Brick D filaments were solubilized by a 1:16 Caps:Brick D ratio, with no crystals observed (D) (scale bar = 50 nm). 1:32 Caps:Brick D produced crystals (E) (scale bar =50 nm). 130

Figure 2.12. Mixing 7-repeat Brick D with 8-repeat Brick C to introduce randomness to the Brick/back binder filament and prevent crystallization. Mixing equal amounts Brick C and Brick D (left) failed to prevent crystallization because the Brick C/back binder crystal forms very rapidly. However, adding 5% Brick C to 95% Brick D (right) prevents crystals from forming and produces a high concentration of short, free filaments. 132

Figure 3.1|Archaeal conjugative pili. a, d Representative cryo-electron micrographs of *A. pernix* and *P. calidifontis* pili, respectively, from >8000 images collected for each. Scale bars, 50 nm. b, e Side and top views of the *A. pernix* (b, yellow) and *P. calidifontis* (e, blue) cryo-EM density maps at resolutions of 3.3 Å and 4.0 Å, respectively. The front half of the filament has been removed in the side views, so that lumens are visible. c, f Atomic models in ribbon representation of the *A. pernix* (c, yellow) and *P. calidifontis* (f, blue) pili docked within their respective transparent cryo-EM density maps. The asymmetric unit of the *A. pernix* pilin CedA1 (c, bright yellow) shows two bound lipids (magenta and blue), while a single asymmetric unit of the *P. calidifontis* pilin TedC (f, dark blue) shows one bound tetraether lipid (orange). g Genomic loci encompassing the Ted system in different members of the order Thermoproteales. Each of the five genera within the Thermoproteales (genera *Caldivirga*, *Pyrobaculum*, *Thermocladium*, *Thermoproteus*, and *Vulcanisaeta*) is represented. Genes encoding the

conserved VirB2/ CedA1-like pilin protein TedC, VirB6/CedA-like transmembrane channel TedA, and VirB4/CedB-like AAA + ATPase TedB are shown as red, green, and cyan arrows, respectively. Additional conserved genes are also indicated, including VirC1/ParAlike, Rad50, and HerA-like helicase which are shown as orange, magenta, and dark blue arrows, respectively. Other genes are shown in gray. Genomic loci are aligned using the TedC gene and indicated with the corresponding UniProt accession numbers, followed by the organism name. In some species, the components of the Ted systems are encoded within distal genomic loci which are separated from the TedC-encoding loci by dashed lines, and the corresponding genes are identified with their UniProt accession numbers..... 158

Figure 3.2| Intricate pilin-lipid interaction networks within archaeal conjugative pili. a Front and back views of a single asymmetric unit of the *P. calidifontis* pilus (blue) containing one pilin (TedC) and one phospholipid GDGT-0 (orange). One polar headgroup of the tetraether lipid faces the lumen and the other faces the outside of the pilus. A top view of a subunit, looking down the helical axis, is shown on the left for orientation. The isoprenoid chains of the GDGT-0 lipids are buried between hydrophobic helices and interact closely with helix $\alpha 1$ of the pilin shown but will also interact with $\alpha 2$ of neighboring pilins. b An atomic model of the cyclic GDGT-0 lipid docked within the lipid density. c Lipid density from *P. calidifontis* (blue). The density is very well resolved and shows that the GDGT-0 lipids have one head group facing the outside of the filament and the other head group facing the lumen. d A single pilin (blue) contacts four surrounding GDGT-0 lipids (orange). e The front and back view of the CedA1 pilin (yellow) of *A. pernix*. The asymmetric unit contains two lipids and one pilin, with the lipid in two different conformations: one having a partially folded shape (blue) and the other a crescent-like shape (magenta). The crescent head group is facing the outside of the pilus while the isoprenoid chains are buried between the pilin subunits. The partially folded lipid's phosphate head group is facing the lumen of the pilus and the isoprenoid chains are bent and buried between the subunits. A top view of a subunit, looking down the helical axis, is shown on the left for orientation. For both lipids the contacts with the pilin are mediated by hydrophobic residues such as leucine, isoleucine and valine from helices $\alpha 1$ and $\alpha 2$. The crescent-like lipid has one hydrophobic interaction with the partially folded lipid. f Atomic models for the crescent-

shaped and partially folded lipid docked into the lipid densities. g Lipid density from *A. pernix* (yellow).

The lipid density is less resolved than in *P. calidifontis* but shows two C25-C25-diether lipids, one of which forms a crescent-like shape (arrow 1) and the other forms a partially folded shape (arrow 2). Both lipids are capped with extra density (arrow 3) which is likely a dihexose sugar attached to the phosphate head group. h A single protein subunit (yellow) makes contacts with ten lipids (five crescent-shaped (pink), and five partially folded (blue))...... 163

Figure 3.3 | Structure of the *A. tumefaciens* T-pilus. a Representative cryo-electron micrograph of *A. tumefaciens* T-pilus, from 8,363 images collected. Scale bar, 50 nm. b Side and top view of the T-pilus cryo-EM density map at a resolution of 3.5 Å. The front half of the filament has been removed in the side view, and we are looking at the lumen. c While there is a 1:1 stoichiometry of lipids to pilins in the filament, a pilin (red) makes contact with four lipids (green). d Atomic model of the T-pilus in ribbon representation docked within the transparent cryo-EM density map. A single subunit model is shown in red. An inset on the right shows front and back views of the VirB2 subunit. 166

Figure 3.4 | Comparison of archaeal and bacterial conjugative pili. a Sequence alignments of archaeal (top) and bacterial (bottom) mating pili. Sequences of mature pilins are shown with the secondary structure elements determined from the structure for the pilins shown above and below the corresponding sequences. Kinks in the α -helices are indicated with triangles. b, e Comparison of the lumen and outer diameter of archaeal *P. calidifontis* (blue) and *A. pernix* (yellow) with bacterial pili: *A. tumefaciens* T-pilus (red), pED208 (mauve), F-pilus (purple) and pKpQIL (green). The outer diameters range between 74 Å to 87 Å. The lumen diameters range between 16 Å to 26 Å, but in some cases cannot be easily reduced to a single number. c, d Atomic models with a single strand, shown in gray, show connectivity between the subunits. This connectivity is also represented with helical nets for the archaeal and bacterial pili. The helical nets show the unrolled surface lattice viewed from the outside of the filament. Each point represents a subunit, and the dotted lines are drawn to highlight the fact that all have right-handed 5-start helices. All of the pili have substantial connectivity between subunits along these 5-start helices..... 169

Figure 3.5| Proposed conjugation mechanisms between donor and recipient cells in archaea (left) and bacteria (right). The schematic shows how ssDNA substrates are generated by the HerA-NurA machinery in the donor archaeal cells and by the plasmid-encoded relaxosome in bacteria. Note that Ceda and, potentially, Ted systems function as DNA importers rather than DNA exporters, contrary to the bacterial plasmid conjugative machinery..... 174

Figure 4.2. Possible assembly strategy for 2TP and gel. Increasing the pH of the solution above 8 is thought to induce assembly into single or paired filaments. CPs, cyclic peptides..... 214

Figure 4.1. Model of 2TP assembly. 2TP is thought to assemble under basic conditions into stacked filaments, with pyrene appendages stacked in register. Filaments are expected to laterally associate by interleaving pyrenes into two stacks. 2TP forms a gel at high concentration (11 mM) and pH 8-8.5, inset. 214

Figure 4.3. Designs of 1TP and 2TP cyclic peptides. 1TP features one pyrene moiety opposite lysine at position 5, whereas 2TP features two pyrenes at positions 4 and 6..... 216

Figure 4.4. Circular dichroism spectra for 1TP and 2TP. The two peptides appear to polarize light at the same wavelengths, but in opposite directions. 2TP also shows a strong signal around 200 nm. Solving the structures of these filaments will also help identify these CD peaks..... 217

Figure 4.5. Negative-stain TEM images from 1TP and 2TP assembly. A. 1TP, 11 μM in water presents as very long, single filaments with excellent dispersion on the film. Some filaments are paired but most are single. B. 2TP, 11 μM in water. Filaments are much shorter than 1TP, about 50-100 nm long. Filaments are almost always in bundles which appear flat. C. 1TP, 11 μM in HEPES pH 8. Filaments are still very long, but are now bundled into groups of 3-5 filaments. The length and bundling behavior combine to create a network of laterally associated filaments. D. 2TP 11 μM in HEPES pH 8. Filaments are still bundled and are too short to develop a network of filaments. 218

Figure 4.6. Cryo-EM micrograph of 225 μM 1TP in 300 mM NaCl. Background of the image is not clear ice, but a monolayer of filaments..... 219

Figure 5.1. a) "Sea urchin" appearance of clusters of CanA tubes. b) Closer view of tubes showing ridged surface.....	230
Figure 5.2. a) TEM micrograph showing poorly-formed CanA assemblies mediated by Zn ²⁺ . b) TEM micrograph showing high density of long, well-dispersed CanA assemblies mediated by Tb ³⁺ . Circles are not short rings of CanA, but other debris, possibly aggregates of a contaminating protein.	231
Figure 5.3. Heat and EDTA effects on CanA tubes. A) CanA tubes made with Ca ²⁺ incubated at 60°C in excess EDTA. Tubes are completely dissolved; solution was optically clear. B) CanA tubes made with Tb ³⁺ incubated at 60°C in excess EDTA. Tubes are badly damaged, especially along the azimuthal dimension. Solution is optically clear. C) CanA tubes incubated at 100°C for 30 minutes without EDTA. Tubes are badly damaged, especially along the azimuthal dimension. Solution is optically clear. D) CanA tubes made with Tb ³⁺ incubated at 100°C for 30 minutes without EDTA. Tubes appear unaffected except for debris attached to some tubes, but not all.....	232
Figure 5.4. A) N20 peptide alone. β fibrils formed as soon as peptide was resuspended in low-salt buffer. B) N20 and terbium-assembled CanA tubes. The peptide failed to incorporate into the nanotube assembly.	233
Figure 5.5. A) CanA tubes under 300 kV cryo-electron microscope. Ridged surface is clearly visible. Single proteins are visible in D class inset. B) Helical map diagram of CanA showing right-handed 2-start helical symmetry. C) Density map of CanA filament. D) Single CanA protein in map (L) and cartoon (R) showing beta-roll fold and calcium ions.....	234
Figure 5.6. CanA intra-filament interactions. A) Front slice of entire filament. B) Close-up view of A with space-filling models of two proteins. The N-terminus of the orange CanA protein associates with the yellow protein via donor strand complementation. C and D) Two angles of view of the calcium ions (green) mediating interaction between blue, teal, and yellow proteins.....	236
Figure 5.7. A. CD spectra for CanA at 20°C and at 95°C. The spectrum at 20°C appears to indicate β-sheet content, with a minimum at 217 nm. There is also some evidence of α-helical character around 200 nm. At 95°C, the signal degrades but shows more β-sheet features, especially in the 200-210 nm region.	

B. CanA melt between 25 and 95°C at 216 nm. The signal trends toward 0 at 80°C but is not indicative of complete unfolding. There is likely some conformation change at high temperatures.....	238
Figure 5.8. CanA under AFM in air. Images and data are consistent with filaments that collapse on the dry carbon-coated mica. A. Height of filament in boxed area average 2 nm. B. Peak force error, average 0.146 nN. C. Adhesion average 0.589 nN. D. Dissipation average 64.9 eV. E. Deformation average 1.21 nm.	239
Figure 5.9. AFM in fluid showed that filament is extremely brittle. QNM cannot discriminate stiffness of assembly vs mica.	240
Figure 5.10. Sedimentation coefficients for CanA±Ca ²⁺ . There is no difference between the samples because no assemblies are observed under AUC conditions.	242
Figure 5.11. Negative-stain TEM micrographs of Hyper2 assembled in a) 40 mM CaCl ₂ , b) 40 mM MgCl ₂ , c) 80°C heat only with no added cation.....	244
Figure 5.12. a) CD spectra for Hyper2 at 25°C and at 95°C after melting. b) Melt curve showing no change in ellipticity from 4°C to 95°C at 216 nm wavelength. Both graphs are consistent with no to negligible thermal conformational change in this range.	245
Figure 5.13. Cryo-EM micrograph showing Hyper2 assemblies.....	245
Figure 5.14 A. Model of 8-start Hyper2 assembly. Calcium ions shown in pink. Green and blue space-filling models demonstrate donor strand complementation mediating axial interactions. B. Close view of calcium ion coordinating residues on azimuthally adjacent green and yellow Hyper2 proteins. C. Lipophilicity diagram for Hyper2, showing overall low lipophilicity/high hydrophilicity and high tyrosine content of the N-terminus. D. Simplified model of Hyper2 showing high B-sheet content.	246
Figure 5.15. Cryo-EM maps of single-walled Hyper2 tubes. A-E) The C7 symmetry assembly has a diameter of about 240 Å and was solved to a resolution of >5Å. F-J) The C8 symmetry assembly is about 270 Å in diameter and was solved to 3.5Å resolution, the best out of the observed single-walled tubes. K-O) The C9 symmetry assembly is about 290 Å in diameter and was solved to >6 Å resolution. All assemblies are right-handed and have similar rise and differ slightly in twist. Layer lines are also very similar between all classes.	247

Figure 5.16. a) EDTA added to Hyper2 assemblies in Ca^{2+} , incubated at room temperature. b) EDTA added to Hyper2 assemblies in Ca^{2+} , incubated at 80°C for 20 minutes. In A and B tubes appear to be unaffected by EDTA. c) Hyper2 mixed with EDTA, then Ca^{2+} , incubated at 80°C for 20 minutes. d) Hyper2 assembled in Ca^{2+} , then boiled 20 minutes. Assemblies are present but appear to be degraded. 248

Figure 5.17. QNM for Hyper2 filament in air. Images and data are consistent with filaments that collapse on the dry carbon-coated mica and appear to be more brittle than CanA filaments. They show characteristic shape but variable thickness and height, consistent with their appearance in electron micrographs. A. Height of filament, 2.9-3.9 nm in boxed area. B. Peak force error, average 0.193 nN. C. Adhesion average 3.13 nN. D. Dissipation average 480 eV. E. Deformation average 8.76 nm..... 249

Figure 5.18. A,B) Surface charges of inner and outer surfaces of A) CanA and B) Hyper2 filaments. C,D) N-X-S/T glycosylation sequon shown in pink on surfaces of c) CanA and d) Hyper2 filaments. 257

SI Figure 2-1. Sequences of proteins described in this chapter. 150

SI Figure 2-2. SDS-PAGE gel showing Brick C expression. 150

SI Figure 2-3. SDS-PAGE gel showing Brick D expression. 151

SI Figure 2-4. SDS-PAGE gel showing Backbinder-ST expression. 151

SI Figure 2-5. SDS-PAGE gel showing Brick F expression, purification, and TEV cleavage..... 152

SI Figure 2-6. SDS-PAGE gel showing $\alpha 2$ overnight expression and purification. 153

SI Figure 2-7. A. SDS-PAGE gel showing that sfGFP-SC and Backbinder-ST do form a covalent bond. B. Brick C assembled with Backbinder-ST, then decorated with sfGFP-SC. No ordered assemblies were observed. 153

SI Figure 3-1. **(A)** An alignment of the CedA1 (yellow) with the TedC (blue) subunits shows an RMSD of 6.5 Å between the two subunits. A distinct kink in helix $\alpha 2$ of CedA1, while the kink in helix $\alpha 2$ of TedC is present but less prominent. **(B)** Alignment of the VirB2 subunit (red) with the TraA subunits of pED208 (orange). There is a 3.9 Å RMSD between VirB2 and pED208 across 58 atom pairs. There are two distinct

kinks in VirB2 not present in pED208. The kink in VirB2 helix α 1 occurs at phenylalanine 61 and the kink in α 3 occurs at proline 23..... 198

SI Figure 3-2. Results of the profile-profile comparisons using HHsearch for different components of the conjugative machineries of bacteria and archaea. **(A)** Comparison of the archaeal membrane pore proteins TedA and CedA. **(B)** Comparison of the archaeal protein TedA and bacterial protein VirB6. **(C)** Comparison of CedA and CedA1. **(D)** Comparison of bacterial conjugation proteins VirB2 and VirB6. H(h), α -helix; E(e), β -strand; C(c), coil. 200

SI Figure 3-3. Mass spectrometry of the lipids extracted from *A. pernix* cell culture. **(A)** A chromatogram of the core lipids found in the sample showing C25-C25 diether lipids as the most abundant species. **(B)** A chromatogram of the most abundant lipid headgroups reveals phospho- dihexose, i.e., a dihexose sugar connected to the glycerol by a phosphate group, as the most prevalent lipid headgroup and phospho-hexose, one sugar connected via a phosphate-group, as the second most prevalent..... 201

SI Figure 3-4 Projections of the 3D reconstructions at different thresholds reveal if pili are extensively glycosylated. T4P pili of *Neisseria gonorrhoeae* **(A)**, EMDB deposition EMD-8739, [<https://www.emdataresource.org/EMD-8739>] and *Pseudomonas aeruginosa* **(B)**, EMDB deposition EMD-8740, [<https://www.emdataresource.org/EMD-8740>] are shown as negative controls, whereas the highly glycosylated type IV pilus of *Saccharolobus islandicus* LAL14/1 **(C)**, EMDB deposition EMD-0397, [<https://www.emdataresource.org/EMD-0397>] shows peripheral density at low threshold, and serves as a positive control. Both the Ted pilus of *P. calidifontis* **(D)** and the Ced pilus of *A. pernix* **(E)** show some surrounding density, which could represent glycosylation. In contrast, the T-pilus of *A. tumefaciens* **(F)** shows no obvious surrounding density at low threshold and behaves like controls A and B..... 202

SI Figure 3-5. PLA2 was applied to purified pili from VirB2 and pED208 to digest any unprotected phospholipids. The remaining lipids were extracted, then identified by mass spectrometry and quantified using internal standards. The major phospholipid classes in the pili samples are shown by headgroup class,

and the acyl chain compositions are shown for the major headgroup classes (PE for VirB2, PG for pED208). Unlabeled bars represent minor species. 203

SI Figure 3-6. Electrostatic potential surface for the lumen of the pED208 (left) and T-pilus (right). The side and top view of the pED208 pilus shows a more electronegative lumen compared to the T-pilus. The top view of the T-pilus shows an alternating positive to negative charge arising from the arginine and the lipid, respectively. All electrostatic potentials and surfaces were calculated using UCSF Chimera. The units for the potential are kT/e 204

SI Figure 3-7. Negative stain electron micrographs showing the effects on pilus structure of different harsh environmental conditions for *A. tumefaciens* T-pilus (left) and *E. coli* pED208 (right), scale bars 50 nm. The morphology of the T-pilus under no treatment, glycerol (50%) and 4M urea appears fully intact. The high-temperature (70°C) and SDS (0.1%) partially degrades the pilus and results in flexuous filaments. Triton-X (1%) fully depolymerizes the filament and creates aggregated clusters. The pED208 pilus morphology is mostly in a flexuous state with partial bundling at the ends of the pilus under SDS (0.1%). Each condition/treatment was replicated three times for both the T-pilus and pED208. Over 15 grid squares were chosen at random and examined for each sample to ensure a broad representation of pilus morphology across the entire electron microscopy grid. See Table S2 for more details. 205

SI Figure 3-8. Heat-map of global structural similarity between pilins from all known structures of prokaryotic conjugation pili: *P. calidifontis* (blue), *A. pernix* (yellow), *A. tumefaciens* (red), F-pilus (purple), pED208 (light brown) and pKpQIL (green). The color scale corresponds to the Dali Z-score values. The dark red indicates increased similarity, the lighter red indicates some similarity and grey is no to low structure similarity. Each pilin consists of two or three hydrophobic α -helices with kinks appearing in both of the archaeal pili and the *A. tumefaciens* T-pilin. All share a common helix-turn-helix architecture and have obvious structural homology. 206

SI Figure 3-9. Conjugative pili cannot transport dsDNA. Side views (top) and top views (bottom) of the atomic models of *P. calidifontis* (blue), *A. pernix* (yellow), and *A. tumefaciens* (red) with a model for B-form dsDNA placed within the lumen. The narrow lumen of all the prokaryotic conjugative pili would

result in extensive clashes with dsDNA suggesting that only ssDNA could be accommodated within the lumen. The atomic model of <i>P. calidifontis</i> (blue) does not include a lipid model with phospho-dihexose headgroups. The addition of these sugar headgroups would decrease the pore size and result in even further clashes with dsDNA.....	207
SI Figure 3-10. Averaged power spectra of the <i>P. calidifontis</i> , <i>A. pernix</i> and <i>A. tumefaciens</i> pili. (A.) Averaged power spectrum generated for <i>P. calidifontis</i> from 54,000 segments. Two layer lines are indexed with their Bessel orders: the $n=1$ and $n=5$. (B.) Averaged power spectrum generated for <i>A. pernix</i> from 44,000 segments. Two layer lines are indexed with their Bessel orders: the $n=1$ and $n=5$. (C.) Averaged power spectrum generated for <i>A. tumefaciens</i> from 49,000 segments. Two layer lines are indexed with their Bessel orders: the $n=0$ and $n=5$. To increase the dynamic range, the log of the intensities are shown for <i>P. calidifontis</i> and <i>A. tumefaciens</i>	208
SI Figure 5-1. Sequence of CanA used in this work. The first 10 N-terminal residues, highlighted in red, were omitted from the sequence to improve solubility.....	265
SI Figure 5-2. Sequence of Hyper2 used in this work, with signal sequence omitted.....	265
SI Figure 5-3. SDS-PAGE gels from CanA and Hyper2 expression and purification. A. 19 kDa CanA is not expressed without IPTG, but is visible after 6 hours IPTG-induced expression. Neither the pellet from the enzyme lysis with lysozyme and Benzonase nor the pellet from incubating the enzyme supernatant at 80°C contain CanA. Supernatant from the heating step contains CanA at high concentration and purity. B. Hyper2 expresses well under IPTG expression. Some Hyper2 partitions to the enzyme lysis pellet fraction, but most remains soluble. Some Hyper2 also crashes out of the heat purification, which was later determined to be heat-mediated nanotube assembly. Hyper2 is present in the heat-purified supernatant at high concentration and good purity.	266
SI Figure 5-4. The secondary structure propensities were obtained from the chemical shifts of C', C ^α , C ^β , N, H ^N and H ^α atoms by the program TALOS-N (Shen and Bax 2013). Li coil regions, Ei β-strand regions, Hi helix. Reused under Creative Commons CC BY license from Kreitner, R., Munte, C.E., Singer, K. et al. Complete sequential assignment and secondary structure prediction of the cannulae forming protein	

CanA from the hyperthermophilic archaeon <i>Pyrodicticum abyssi</i> . <i>Biomol NMR Assign</i> 14 , 141–146 (2020). https://doi.org/10.1007/s12104-020-09934-x	267
SI Figure 5-5. ClustalW alignment of CanA and Hyper2. Residues involved in metal ion binding are highlighted.....	268
SI Figure 5-6. FATCAT rigid alignment of single Hyper2 (blue) and CanA (pink) proteins.....	269
SI Figure 5-7. Hyper2 ESI-Mass spec readout confirming expected mass at 16.1 kDa.....	270
SI Figure 5-8. CanA ESI-Mass spec readout confirming expected mass at 19.9 kDa.....	271
SI Figure 5-9. AlphaFold Multimer predictions for Hyper2 and CanA trimers.....	272
SI Figure 5-10. 40 mM CaCl ₂ and 5 μM CanA have been demonstrated to produce nanotubes, though they are not as easy to confirm via TEM simply because they are few and far between.....	273
SI Figure 5-11. CanA FSC.....	274
SI Figure 5-12. Hyper2 C7 FSC.....	274
SI Figure 5-13. Hyper2 C8 FSC.....	275
SI Figure 5-14. Hyper2 C9 FSC.....	275
SI Figure 5-15. AlphaFold2.0 prediction for CanA, colored by residue.....	276
SI Figure 5-16. pLDDT score for single CanA protein prediction in AlphaFold.....	276
SI Figure 5-17. AlphaFold2.0 prediction for Hyper2, colored by residue.....	277
SI Figure 5-18. pLDDT score for single Hyper2 protein prediction in AlphaFold.....	277
SI Figure 5-19. Helical net maps for the three best classes of Hyper2.....	278

Chapter 1 |

Introduction: Structures of synthetic helical filaments and tubes based on peptide and peptido-mimetic polymers

Note: The following chapter is reprinted from Miller JG, Hughes SA, Modlin C, Conticello VP (2022).

Structures of synthetic helical filaments and tubes based on peptide and peptido-mimetic polymers.

Quarterly Reviews of Biophysics <https://doi.org/10.1017/S0033583522000014> under Creative

Commons CC BY license. It has been reformatted to meet Emory's dissertation format guidelines and to be consistent with the rest of this dissertation.

Quarterly Reviews of Biophysics

Citation: Miller JG, Hughes SA, Modlin C, Conticello VP (2022). Structures of synthetic helical filaments and tubes based on peptide and peptido-mimetic polymers. *Quarterly Reviews of Biophysics* <https://doi.org/10.1017/S0033583522000014>

© The Author (s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Structures of synthetic helical filaments and tubes based on peptide and peptido-mimetic polymers

Jessalyn G. Miller, Spencer A. Hughes, Charles Modlin and Vincent P. Conticello 

Department of Chemistry, Emory University, 1515 Dickey Drive, Atlanta, GA 30322,

USA

Synthetic peptide and peptido-mimetic filaments and tubes represent a diverse class of nanomaterials with a broad range of potential applications, such as drug delivery, vaccine development, synthetic catalyst design, encapsulation, and energy transduction. The structures of these filaments comprise supramolecular polymers based on helical arrangements of subunits that can be derived from self-assembly of monomers based on diverse structural motifs. In recent years, structural analyses of these materials at near-atomic resolution (NAR) have yielded critical insights into the relationship between sequence, local conformation, and higher-order structure and morphology. This structural information offers the opportunity for development of new tools to facilitate the predictable and reproducible *de novo* design of synthetic helical filaments. However, these studies have also revealed several significant impediments to the latter process – most notably, the common occurrence of structural polymorphism due to the lability of helical symmetry in structural space. This article summarizes the current state of knowledge on the structures of designed peptide and peptido-mimetic filamentous assemblies, with a focus on structures that have been solved to NAR for which reliable atomic models are available.

Introduction

Cellular biology provides many examples of filamentous nanomaterials in which control of higher-order structure enables emergent function. Extracellular protein filaments (e.g. pili, flagella, secretory needles and tubes) (Egelman, 2017) and filamentous phage and viruses (Stubbs and Kendall, 2012) represent protein and nucleoprotein assemblies, respectively, in which regulated fabrication from the macromolecular components has led to the evolution of complex function. These extracellular filaments perform a diverse range of functions that would be desirable to emulate in synthetic systems, including chemo-mechanical energy transduction (Poweleit *et al.*, 2016; Wang *et al.*, 2017), controlled delivery (Loquet *et al.*, 2012; Costa *et al.*, 2016), selective and tunable catalysis (Lynch *et al.*, 2017, 2020), and, as recently discovered, electron transfer over multi-micron length scales (Wang *et*

al., 2019). While many synthetic peptide and protein filaments have been proposed as substrates for directed applications in medicine and nanotechnology, the limited availability of structural information at high resolution has hindered the development of these assemblies as functional materials (Haines-Butterick *et al.*, 2007; Yan

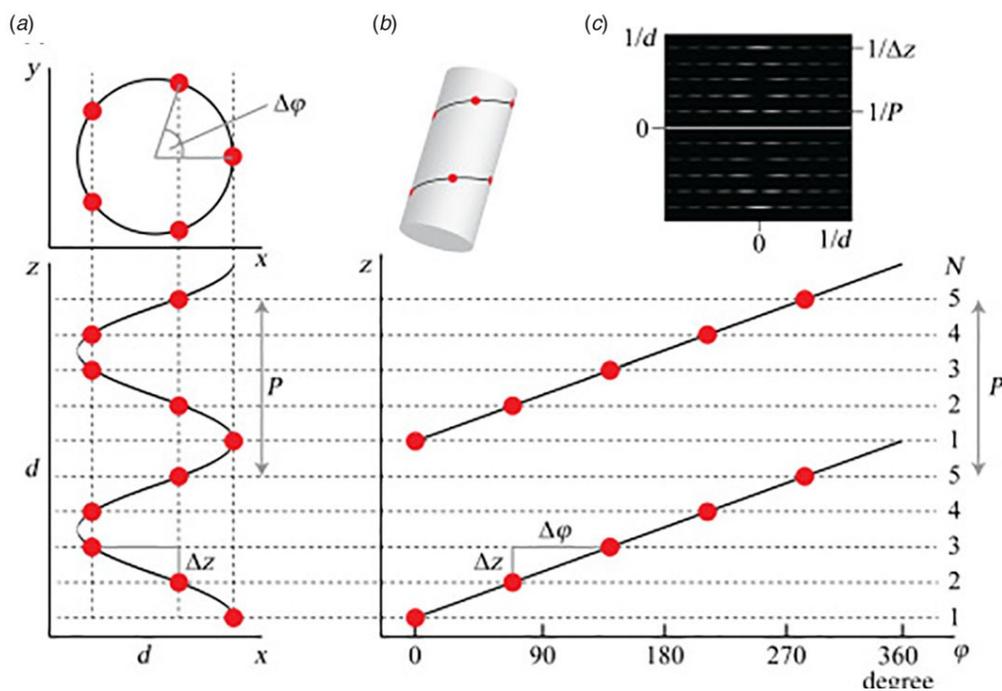


Figure 1.1 (a) Top and side projections of a helical polymer in which each dot represents an asymmetric unit of the corresponding assembly. (b) Helical net diagram corresponding to the symmetry of the helical polymer. A solid line connects protomers within the 1-start helix of the assembly. (c) Power spectrum derived from the amplitudes of the Fourier transform of this helical polymer in which layer lines are indicated corresponding to pitch (P) and rise (z) in reciprocal space. From S. A. Fromm and C. Sachse (2016) *Methods in Enzymology*, 579, 307–328, Academic Press Inc. Reprinted with *et al.*, 2008; Moore and Hartgerink, 2017; Wu *et al.*, 2017; Gelain *et al.*, 2020).

The key to understanding and emulating the diverse functions of protein filaments resides in the ability to deconvolute the structural principles that enable their highly specific assembly. Native protein and nucleoprotein filaments achieve this specificity through a combination of structural control of interactions at protein–protein interfaces within the assembly and spatiotemporal regulation of the post-synthetic processing of the protein subunits into structurally defined filaments. The latter process

enables controlled fabrication of the corresponding protein or nucleoprotein filaments but lies beyond our current capabilities to replicate *in vitro* for synthetic assemblies. In contrast to the majority of synthetic peptide and protein filaments that have been studied *in vitro*, biologically derived protein filaments often *do not self-assemble*, but instead the protomers are fabricated into ‘assembled’ structures under conditions that are far from equilibrium (Costa *et al.*, 2015).

Synthetic peptide and protein filaments, while differing in the mechanism of formation from native biological assemblies, display similarities in supramolecular structure. Both classes of filaments are based on non-covalent self-association of subunits (protomers) into high aspect-ratio supramolecular polymers ($\geq 1 \mu\text{m}$ in length) that display helical symmetry (Figure 1.1). This simple symmetry operation involves rotation through a characteristic helical twist or azimuthal angle (ϕ) with a commensurate axial translation (z). Continuous application of this symmetry operation upon successive addition of protomers generates a filamentous assembly, that is, a helical filament or tube. Such filaments can be generated from protomers having any possible molecular structure as long as sufficient interfacial interactions between subunits are present that stabilize the filament vis-à-vis monomers or discrete oligomers. Within these simple geometrical constraints, an infinite number of structural variations are possible that involve different helical symmetries, as well as superimposed rotational and dihedral symmetry. Helical assemblies are usually characterized in terms of the number of subunits per helical turn, N ($=2\pi/\phi$), and the helical pitch, P ($=Nz$). In nonrotationally symmetric cases, i.e. C_1 symmetry, these parameters can be determined from assignment of the 1-start helix, i.e. the helix that passes through every subunit in the assembly (Figure 1.1). For helical assemblies containing a rotational symmetry axis C_n , the helical symmetry can be understood in terms of the rise and rotation of the n -start helices subject to this rotational symmetry element.

Helical symmetry can be best understood from consideration of the corresponding helical net diagram, in which the helix is unrolled in two dimensions with the axial rise, z , as the ordinate and the azimuthal angle, ϕ , as the abscissa (Figure 1.1b). The perspective is usually presented from the outside of the helical assembly and is critical to identification of the helical hand, i.e. righthanded *versus* left-handed

screw sense, of the respective n -start helices. Each n -start helix passes through every n th protomer in the helical assembly.

The 1-start helix (solid line in the helical net of Figure 1.1**b**) is right-handed since the axial rise increases from left to right in the helical net diagram. In contrast, the same helical net diagram indicates the presence of five 5-start helices within the same assembly, which pass through different sets of subunits at a frequency of every fifth protomer. The 5-start helices are left-handed due to an opposite

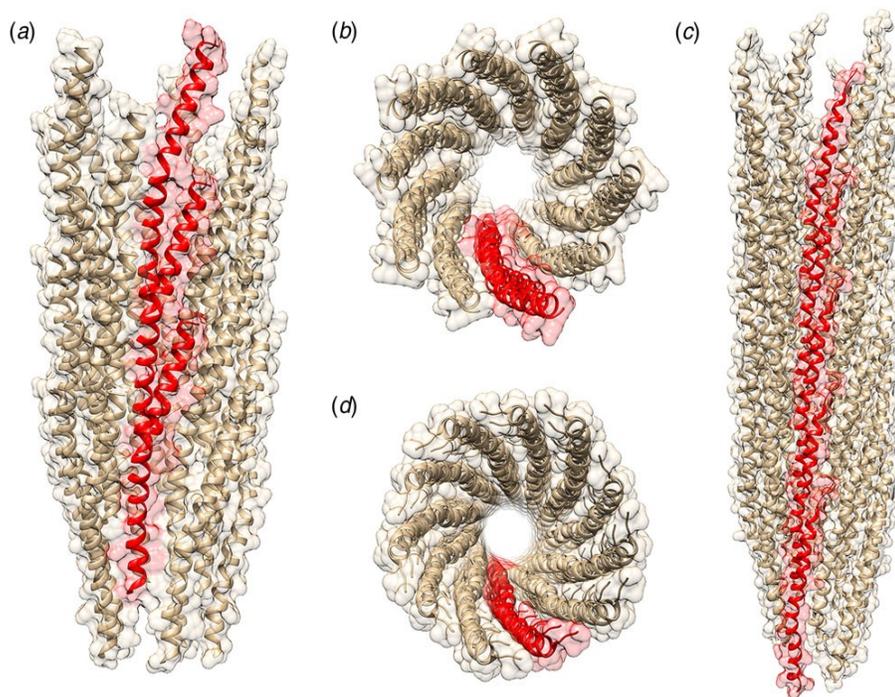


Figure 1.2. Side (a) and top (b) views of the atomic model of the phage IKe capsid. Side (c) and top (d) views of the atomic model of the phage Pf4 capsid. Single protofilaments corresponding to the 10-start and 11-start helices of the IKe and Pf4 capsids, respectively, are highlighted in red.

right-to-left inclination in the helical net diagram. Helical symmetry can be determined from analysis of the averaged power spectrum of the filaments (Figure 1.1**c**) (Wang *et al.*, 2006; Egelman, 2010, 2014), which is derived from the Fourier transform of filament segments that are present in the projection images from electron microscopy (EM). The Fourier transform is complex (having both amplitudes and

phases), while the power spectrum is simply the intensities (squared amplitudes) of the Fourier transform. As a result, the individual power spectra from each segment can be added together without any need for alignment. The layer lines within the averaged power spectrum correspond to repeat spacings in reciprocal space associated with the spatial frequencies of the different n -start helices that are present within the assembly. To assign helical symmetry, two independent orders, n , of the corresponding Bessel functions must be assigned to sets of observed layer lines in the averaged power spectrum. This process often relies on a trial-and-error approach in which multiple different helical symmetries are assessed as workable solutions. The details of structural determination using cryo-EM image analysis are beyond the scope of this article. However, a recent review highlights the critical considerations involved in assignment of helical symmetry in the structural analysis of synthetic peptide filaments (Wang *et al.*, 2022). Prior structural analyses of peptide and protein filaments suggest that helical symmetry can be labile in structural space (Egelman *et al.*, 2015; Lu *et al.*, 2015; Wang *et al.*, 2021a). Small changes in packing at the protomer interfaces within the assembly can result in different helical symmetries, even under conditions in which the individual protomer structures are highly conserved. For example, filamentous bacteriophages comprise two distinct structural families that display C_5 and C_1 symmetry, respectively, for the helical arrangement of capsid proteins around the single-stranded DNA genome (Marvin *et al.*, 2014). Helical reconstruction from high-resolution cryo-EM image analysis enabled structural determination of assemblies from representative members of these two classes of filamentous bacteriophage, IKe (C_5) and Pf4 (C_1), at near-atomic resolution (NAR) (Xu *et al.*, 2019; Tarafder *et al.*, 2020). Comparison of the two structures is illustrative of the challenges encountered in the structural analysis of filamentous assemblies. The capsid assemblies consist of simple α -helical protomers. Superimposition of individual protomers derived from the two structures indicated a backbone root-mean-square deviation of ca. 1.5 Å despite low sequence similarity. Since the side-chain interactions define the protein–protein interfaces within the respective assemblies, the differences in higher-order structure presumably result from differences in sequence since the protomer fold is conserved.

The structures of the corresponding helical assemblies of the IKe and Pf4 phage capsids are depicted in Figure 1.2 at 3.4 Å (PDB: 6A7F) and 3.2 Å (PDB: 6TUQ) resolution, respectively. In each case, the capsid proteins are arranged in a similar orientation around the central axis of the assembly. Single protofilaments are highlighted in red within the assembly. For phage IKe, the highlighted protofilament coincides with one of the right-handed 10-start helices under C_5 symmetry. In contrast, under the different symmetry of phage Pf4 capsid, the protofilaments coincide with the right-handed 11-start helices.

These differences can be most easily visualized through a comparison of the respective helical net diagrams. The structural subunits associated with the respective protofilaments are highlighted in red (Figure 1.3*a* and *b*). For phage IKe, a representative set of protomers related through the C_5 rotational axis is depicted in cyan. Similarly, the right-handed 1-start helix of phage Pf4 is highlighted in the corresponding helical net. Figure 1.3-*c* and *d* depicts a section of calculated power spectra of the respective assemblies, initially sampled at a frequency of 2 Å pixel⁻¹. Arrows indicate the spatial frequencies associated with the axial rise corresponding to the meridional ($n = 0$) for IKe and the 1-start helix ($n = +1$) for Pf4. In either case, the corresponding axial repeats are quite similar between the two assemblies despite the differences in helical symmetry.

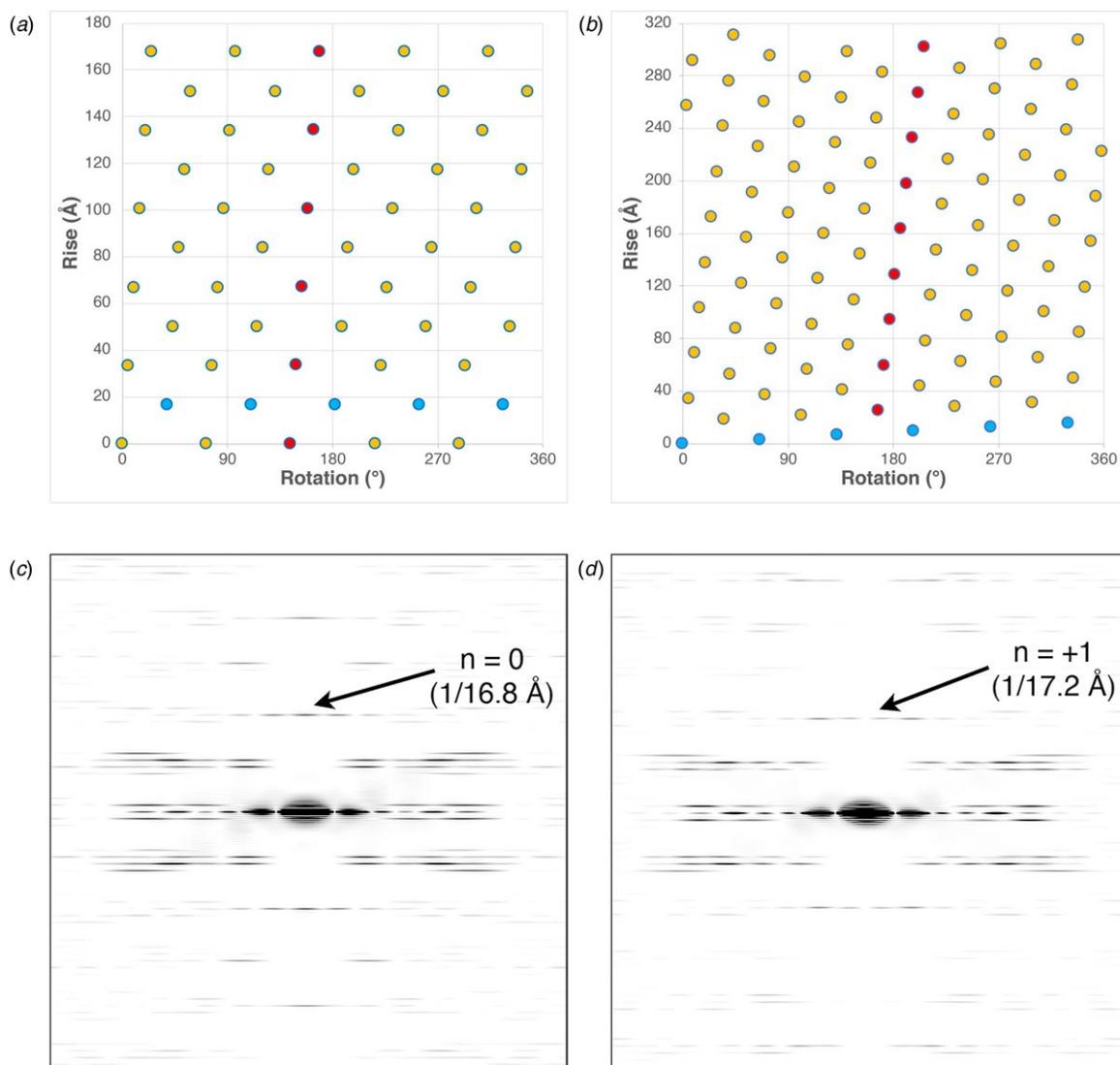


Figure 1.3 (a) Helical net diagram of IKE phage. (b) Helical net diagram of Pf4 prophage. (c) Calculated power spectrum of IKE phage. (d) Calculated power spectrum of Pf4 phage.

In this review, the term ‘helical assemblies’ describes supramolecular polymers of proteins, peptides, and structurally related synthetic foldamers that display helical symmetry. This symmetry operation can be applied to protomers based on any structural motif as long as the interfaces between subunits are stable under the set of experimental conditions under which self-assembly occurs. These assemblies need not form closed cylindrical structures, i.e. assemblies in which successive turns of the

helix make physical contact through an axial interface. The latter structures are often denoted as nanotubes since self-association generates a water-filled central cavity (lumen) that runs through the structure and is oriented parallel to the helical axis (Bong *et al.*, 2001; Hamley, 2014). Peptide nanotubes can also result from other modes of self-assembly, e.g. axial stacking of cyclic peptides (Ghadiri *et al.*, 1993; Insua and Montenegro, 2020). The latter subject has been reviewed recently and will not be covered here (Song *et al.*, 2021).

This review will discuss the structure of biomimetic helical assemblies derived from synthetic peptides and peptido-mimetics. Where possible, we focus on helical assemblies in which the structures have been solved at high resolution using biophysical methods such as single-crystal diffraction, solid-state nuclear magnetic resonance (ssNMR) spectroscopic measurements, or helical reconstruction from cryo-EM. Such NAR methods for structural determination have revolutionized the study of helical assemblies. These techniques afford atomic models that provide structural insights into the interfacial packing interactions between protomers that guide the formation of these helical assemblies. The information gleaned from structural studies of biological helical assemblies has led to the emergence of a research field focusing on the design of biomimetic helical assemblies from synthetic peptides and structurally related foldamers (i.e. peptido-mimetic oligomers). The goal of the latter studies is to recapture the complex function and responsive behavior of biologically derived helical assemblies using simple structural motifs that are amenable to synthetic control.

The initial discussion will focus on an introduction to high-resolution methods for structural analysis of helical assemblies in conjunction with their relative merits and limitations. Subsequently, different classes of helical assemblies will be discussed in terms of the conformational features of the protomers. In each case, we will refer to specific examples of helical filaments that have been characterized to NAR.

A recurrent theme in the field of designed filamentous assemblies is the plasticity of quaternary structure in sequence space (Egelman *et al.*, 2015; Lu *et al.*, 2015). Currently, few high resolution structural models have been generated for filamentous assemblies of designed peptides or proteins

(Kajander *et al.*, 2007; Cormier *et al.*, 2013; Egelman *et al.*, 2015; Nagy-Smith *et al.*, 2015, 2017; Chen *et al.*, 2017, 2020; Lee *et al.*, 2017; Shen *et al.*, 2018; Zhang *et al.*, 2018; Hughes *et al.*, 2019; Feng *et al.*, 2020; Wang *et al.*, 2021a; Pieri *et al.*, 2022). It has been observed that the experimental structures often display significant differences from the conceptual model upon which the initial design was based. Structural polymorphism is commonly observed for synthetic peptide assemblies or native proteins assembled *in vitro*. The observed structures may depend strongly on the initial conditions, i.e. temperature, pH, ionic strength, etc., that were employed for peptide self-assembly (Close *et al.*, 2018; Guenther *et al.*, 2018; Zhang *et al.*, 2019; Wang *et al.*, 2021b). In addition, the experimental methods employed to prepare the sample for high-resolution structural analysis, e.g. crystallization, lyophilization, cryo-vitrification, may bias the system toward a particular structural variant (Egelman *et al.*, 2015).

Structural characterization of synthetic helical filaments

Structural analysis of helical filaments requires application of experimental methods that interrogate structure across multiple length scales. A number of methods can analyze structure at low resolution and are applicable to relatively rapid screening and analysis of peptide specimens for the presence of ordered conformations and the formation of well-defined supramolecular structure. Circular dichroism (CD) spectropolarimetry and Fourier-transform infrared (FTIR) and Raman spectroscopy can provide evidence for the formation of secondary and tertiary structure, which can be subsequently correlated with self-assembly behavior. Atomic force microscopy (AFM) and conventional transmission electron microscopy (TEM) can reveal the presence of supramolecular structure and provide insight into the morphological features of the resultant assemblies at medium resolution. Small-angle X-ray and neutron scattering measurements, i.e. SAXS and SANS, afford information on the structural features of the assemblies in solution at low-to-medium resolution in the range from 1 to 100 nm from analysis of the form factor scattering (Guilbaud and Saiani, 2011). AFM and scanning electron microscopy (SEM) can also be employed for assignment of the helical hand of peptide assemblies for cases in which the chirality of the surface features can be resolved (Wang *et al.*, 2022). However, only a small set of experimental methods can give information at

NAR on filamentous helical assemblies, namely X-ray and electron diffraction methods, ssNMR spectroscopy, and cryo-EM.

NAR structural information is essential for understanding the internal structure of the protomers and the interfacial contacts that stabilize the helical assemblies. The meaning of the term ‘near-atomic resolution’ is a subject of active debate (Chiu *et al.*, 2017; Wlodawer and Dauter, 2017), especially as regards the level of resolution that can be achieved using different methods of structural analysis (Wlodawer *et al.*, 2017). True atomic resolution, ca. 1.2 Å, refers to the ability to distinguish between separated atoms in an electron density map (Sheldrick, 1990; Morris and Bricogne, 2003). NAR is less well-defined, but it is usually employed to describe the resolution of density maps that can serve as the basis for the construction of reliable atomic models. In this latter case, the resolution limit has been claimed to be as low as ca. 4 Å for the corresponding density maps. The reliability of the corresponding atomic model depends on the map quality, which in turn depends on intrinsic properties of the sample, e.g. the thermodynamic stability of the interfacial interactions between protomers within the assembly. Most peptide and protein filaments are intrinsically flexible. Flexibility can introduce disorder into the sample, which can lower the quality of the density maps resulting from helical reconstruction (Egelman *et al.*, 1982; Orlova and Egelman, 1993).

Recent improvements in hardware and software have made the currently available experimental methods for structural determination much more powerful in terms of the limits of resolution and the scope of substrates that can be examined. Until recently, atomic-level structural data could only be obtained on fibrillar assemblies using X-ray diffraction methods (crystallography and fiber diffraction). However, an increasing number of filamentous structures have been solved to NAR using either ssNMR spectroscopic analysis or, more commonly, cryo-EM. In addition, recent developments in microcrystal electron diffraction (MicroED) have enabled a wider application of diffraction methods to protein crystals, especially for situations in which only sub-micron-sized crystals can be obtained as is often the case for peptide filaments (Rodriguez *et al.*, 2015, 2017; Sawaya *et al.*, 2016; Warmack *et al.*, 2019).

Historically, the application of X-ray fiber diffraction enabled the first structural determinations of

filamentous protein assemblies at NAR. However, the necessity of generating highly oriented specimens restricted these analyses to filaments that displayed persistent rod-like structures that promoted mesophase formation, which facilitated alignment of the filaments for diffraction experiments. The most common substrates were rigid filamentous viruses such as tobacco mosaic virus (Namba *et al.*, 1989). Specimens that displayed greater flexibility were less amenable to the formation of highly oriented sols. Given that most helical assemblies fall into the category of flexible or semi-flexible filaments, the utility of fiber diffraction for NAR structural determination of helical assemblies remains limited. Nevertheless, fiber diffraction can provide valuable structural information on peptide assemblies. For example, fiber diffraction of oriented amyloid filaments was critical for the identification of the cross- β conformation (Eanes and Glenner, 1968; Sunde *et al.*, 1997; Diaz-Avalos *et al.*, 2003).

Single-crystal X-ray diffraction was the first experimental method to be widely applied to the analysis of peptide assemblies at NAR. This technique has been particularly valuable for structural determination of assemblies derived from short amyloidogenic peptide sequences (Eisenberg and Sawaya, 2017; Ke *et al.*, 2020). However, two distinct challenges are encountered in the application of this method. The first is the experimental difficulty associated with growing suitable single crystals, which is a necessity for high-resolution diffraction experiments. Helical assemblies present several unique challenges to crystallization including limited solubility, length variability, and structural polymorphism. High-throughput screening of crystallization conditions has facilitated identification of effective crystallization conditions. However, crystallization of fibrillogenic peptides and proteins seems to work best under conditions in which the protomers are weakly associated such that self-assembly occurs concurrently with crystal growth. Achieving this condition may require introduction of mutations that weaken the interfacial interaction between protomers or enhance the solubility of the peptide or protein (Spencer *et al.*, 2013).

The second challenge for single-crystal diffraction analysis is imposition of crystallographic symmetry on the assembly. Helical arrays of chiral rod-like molecules cannot easily accommodate the translational symmetry required for crystallization without imposition of an energy penalty associated

with elastic distortion of the protomers (Rodriguez *et al.*, 2015). The resultant packing frustration can limit the dimensions of the crystals. MicroED methods using cryo-EM have alleviated this problem to a degree by allowing single-crystal diffraction experiments to be performed on sub-micron crystals. In addition, crystallographic space symmetry restricts the screw axes to two-, three-, four-, and six-fold rotations. While other helical symmetries can be accommodated in crystal structures (Dauter and Jaskolski, 2018), the protomers cannot occupy symmetry-equivalent positions under these non-crystallographic screw axes. In contrast, the structures of peptide and protein filaments can vary over a wide range of helical symmetries that include non-crystallographic screw symmetry and non-integral helical repeats.

The implication of these limitations can be best understood through comparison of peptide and protein filament structures solved using crystallography *versus* those determined using complementary methods such as ssNMR spectroscopy or cryo-EM, which are not subject to these symmetry restrictions (*vide infra*). Currently, a number of structures are available at comparable resolution for helical assemblies derived from the same peptide sequence using two independent methods of high-resolution structural analysis. In the limited cases in which these comparisons have been made, the crystallographically determined structures have been observed to differ from those determined from ssNMR or cryo-EM (Guenther *et al.*, 2018; Guerrero-Ferreira *et al.*, 2018, 2019). Crystallization imposes conditions that may select for a specific structural form that happens to crystallize well, while a sample of filaments suspended in solution may display significant structural polymorphism. Despite this caveat, crystallography has yielded significant structural insights into the packing of protomers within peptide and protein assemblies and has been especially important in understanding the structural factors that control association of amyloidogenic peptides that adopt cross- β structures (Eisenberg and Sawaya, 2017). In addition to the aforementioned concerns, crystallographic studies of helical filaments often employ mutants or truncations of the native sequences that promote crystallization. The effect of these modifications, *vis-à-vis* the structure of filaments derived from full-length proteins, needs to be evaluated on an individual basis. ssNMR measurement has also been applied as a method for structural determination of helical peptide filaments (Huang *et al.*, 2018).

Similar to solution NMR structural determination of soluble proteins, selective labeling of the protons with magnetically active nuclei enables determination of inter-atomic distances between labeled residues within the protein sequence. These distance measurements place constraints on the spatial arrangement of the polypeptide backbone and side chains from which a structural model can be generated. The accuracy of the atomic model depends on the number of constraints, the selectivity of labeling, and the size of the protein.

NMR distance measurements have an upper limit that depends on the identity of labeled nuclei and the experimental method. For ^{13}C – ^{15}N contacts, the most commonly employed isotopic spin pair in protein and peptide structural determinations, the upper limit for accurate distance measurements is approximately 6 Å. However, inter-atomic distances up to 10–12 Å can be measured using magnetically active nuclei that are less common in peptides and protein sequences, e.g. ^{19}F and ^{31}P (Mehta *et al.*, 2004). Multiple long-range distance constraints can be challenging to acquire, which hinders assignment of helical symmetry and accurate determination of supramolecular structure. To address this issue, ssNMR structural analysis has been combined with scanning transmission electron microscopy (STEM) mass-per-length measurement or cryo-EM imaging (Colvin *et al.*, 2016).

In addition, the structural polymorphism observed for many helical assemblies can multiply the effective number of resonances in ssNMR experiments, which can complicate structural analysis as it often results in significant spectroscopic overlap and ambiguity in resonance assignment. Therefore, monomorphic filaments represent the best substrates for ssNMR structural analysis, although these may be challenging to isolate and purify to homogeneity due to the intrinsic polymorphism associated with peptide assemblies. Nevertheless, ssNMR methods have been employed for the structural determination of helical assemblies to NAR, including the structures of extended β -sheet filaments such as the HET-s prion domain (Siemer *et al.*, 2005), the A β (1-42) amyloid assembly (Colvin *et al.*, 2016), and the α -synuclein (α -syn) polymeric filament (Heise *et al.*, 2005).

Historically, TEM has been one of the most important methods for morphological analysis of peptide and protein filaments. The development of cryo-vitrification methods enabled preservation of

isolated and dispersed filaments as thin films in the frozen, hydrated state (Dubochet, 2012). In addition, EM imaging at cryogenic temperatures significantly reduces beam damage due to interaction of the substrate with high-energy electrons. Cryo-EM imaging is less susceptible to introduction of structural artifacts than conventional TEM imaging, which often involves application of a heavy-atom stain (Lepault *et al.*, 1983). However, until the emergence of direct electron detectors, the lower limit of resolution for cryo-EM structural determination was ~ 10 Å, which was insufficient to resolve secondary structure elements (Li *et al.*, 2013; Kühlbrandt, 2014; Subramaniam *et al.*, 2016). Higher-resolution reconstructions have been performed using film images (Ge and Zhou, 2011; Sachse, 2015), but this method is labor-intensive and has been superseded by use of direct electron detection cameras with high detective quantum efficiency (Song *et al.*, 2019).

Initially, pseudo-atomic models were built using rigid body modeling of structural subunits into the lower-resolution electron density maps resulting from cryo-EM measurements. The subunit structures were usually determined from X-ray crystallography or solution NMR spectroscopic determinations. These pseudoatomic models were limited in accuracy due to the incompleteness of individual subunit structures. NMR structures display some degree of disorder in the more flexible regions of the molecule. Crystal structures are often derived from protein fragments rather than the entire peptide sequence. In addition, these input structures often provided little information about the packing interfaces in the helical filament, which are frequently more ordered within the assembly than in the isolated, soluble monomeric precursors.

The development of electron cryo-microscopy with direct electron detection has led to a revolution in structural determination of helical assemblies at NAR (Kühlbrandt, 2014). These detectors can routinely achieve resolutions ≤ 4.0 Å. This level of resolution permits *de novo* building of atomic models of helical filaments from cryo-EM data alone. The main method for structure analysis of helical filaments involves iterative real-space reconstruction from cryo-EM images (Egelman, 2000, 2007, 2010). This single-particle method enables reconstruction from randomly oriented helical filaments that result from immobilization in a thin film of vitreous ice on a cryo-EM grid. Programs such as Relion and

cryoSPARC are available to assist in image processing and reconstruction after assignment of helical symmetry (He and Scheres, 2017; Punjani *et al.*, 2017).

Cryo-EM helical reconstruction has many significant advantages for structural analysis of helical assemblies. As a single-particle method, the polymorphism typically observed for helical assemblies can be dealt with through manual or automated classification of the different structural variants. If enough images can be collected, different morphic variants can be structurally analyzed within the same grid. In addition, cryo-EM can tolerate the presence of impurities in the sample as long as the images of the contaminants can be distinguished from those of the analyte (Spaulding *et al.*, 2018). Cryo-EM analysis can be applied to small sample sizes (a few microliters of 0.1–1 mg ml⁻¹ solution) and does not require heavy-atom or isotopic labeling as does crystallographic analysis and NMR measurements, respectively. In addition, the method is amenable to structure determinations on helical filaments directly derived from biological samples available in limited quantity, such as tau filaments from human brain tissue (Scheres *et al.*, 2020) or pili harvested directly from bacterial cells (Egelman, 2017). This procedure precludes the necessity of *in vitro* assembly of protein filaments, which can often result in structural polymorphism or deviations from the native filament structure (Wang *et al.*, 2006).

Comparison between NAR structures determined using different experimental methods highlights the challenges associated with the structural polymorphism of helical assemblies. Guenther *et al.* (2018) characterized the structure that resulted from crystallization of an eleven amino acid, amyloidogenic peptide segment, ²⁴⁷DLIKGISVHI²⁵⁷, from human TAR DNA binding protein 43 (TDP-43). MicroED was employed to determine the structure of ²⁴⁷DLIKGISVHI²⁵⁷ nanocrystals grown at 37 °C in aqueous CHES buffer (pH 8.5) (PDB: 5W52). The resultant structure comprised of a two-fold symmetric filament in which each protofilament displayed a parallel cross-β spine. The two protofilaments interacted through a steric zipper interface in which a face-to-face interaction resulted in side-chain interdigitation.

Surprisingly, at pH values ≤7.5, cryo-EM analysis of assemblies derived from the same ²⁴⁷DLIKGISVHI²⁵⁷ peptide segment indicated the presence of a multitude of filamentous structures. Helical reconstruction of the most abundant population of filaments resulted in an atomic model that

displayed 3_2 screw symmetry (PDB: 5W7V) (Guenther *et al.*, 2018). In addition to the observed difference in helical symmetry between the crystal structure and cryo-EM model, the structure of the protein filament observed in the cryo-EM analysis consisted of three protofilaments that were each based on an asymmetric unit composed of nine peptides. The protofilaments corresponded to the left-handed 3-start helices, in which each of the nine $^{247}\text{DLIIKGISVHI}^{257}$ segments in the asymmetric unit could adopt one of three different possible conformations. This conformational disparity between the crystal structure and cryo-EM structures of $^{247}\text{DLIIKGISVHI}^{25}$ illustrates a critical consideration in the structural analysis of helical assemblies, namely, the potential plasticity of quaternary structure in sequence space and the dependence of the filament structure on the assembly conditions.

Structural differences have also been observed between atomic models generated for assemblies derived from the same peptide using approaches based on either ssNMR spectroscopy or cryo-EM analysis. Colvin *et al.* reported the structure of a monomorphous variant of a peptide assembly of $\text{A}\beta_{1-42}$ using distance constraints derived from ssNMR measurements (Colvin *et al.*, 2016). This peptide has attracted significant scientific interest as it is a primary component of the amyloid fibers associated with the pathology of Alzheimer's disease. The filament structure comprised of an interacting pair of helical assemblies, in which each protofilament displayed the cross- β spine that is typical of amyloid assemblies. Hydrophobic interactions across the mating interface mediated association between protofilaments over an extended interface (Figure 1.4).

Independently, Gremer *et al.* employed helical reconstruction from cryo-EM images to build an atomic model of the filament structure for a different monomorphous variant of $\text{A}\beta_{1-42}$ (PDB: 5OQV) (Gremer *et al.*, 2017). The cryo-EM structure differed significantly in detail from that reported by Colvin *et al.* The protofilament displayed a dimeric structure in which the two assemblies were related by the pseudo- 2_1 screw axis (Figure 1.5) (Scheres, 2020). In addition, a slight left-handed super-helical twist of -1.45° per 4.67 \AA axial rise was observed for the protofilament, which resulted in a large helical pitch of circa 1200 \AA . The helical reconstruction confirmed both the presence of the cross- β spine within the protofilaments and the steric zipper interactions between protofilaments at the inter-protomer interface.

Despite these observations, the details of the packing interaction across the hydrophobic interface and the N-terminal structure of the peptide were distinctly different from the previously reported NMR analysis. The features of the cryo-EM structure were independently confirmed by ssNMR measurements on the corresponding filaments. The two structures represent two distinct morphic variants of A β ₁₋₄₂. The experimental conditions for self-assembly differed significantly between the two A β ₁₋₄₂ preparations, which reinforces the idea that specimen preparation has a critical influence on the structure of the corresponding filaments.

Structures of helical peptide assemblies

The diverse functional properties and exquisite responsive behaviors of biologically derived protein filaments have stimulated interest in the design of synthetic peptide assemblies that mimic the structure and function of the native congeners (Bowerman and Nilsson, 2012; Hamley, 2014; Beesley and Woolfson, 2019). However, rational and predictive design of helical assemblies remains a significant challenge, primarily due to the lability of helical symmetry in structural space and the resultant potential for structural polymorphism. Through necessity, the design of synthetic peptide filaments has primarily drawn from sequence– structure correlations established from structural analysis of super-secondary and tertiary structures of soluble proteins and discrete oligomers. Several examples of these peptide designs are described in the following sections. However, recent structural evidence suggests this approach may be insufficient to uniquely specify the supramolecular structure of the corresponding filaments. Supramolecular polymerization can result in the formation of helical filaments that display periodic

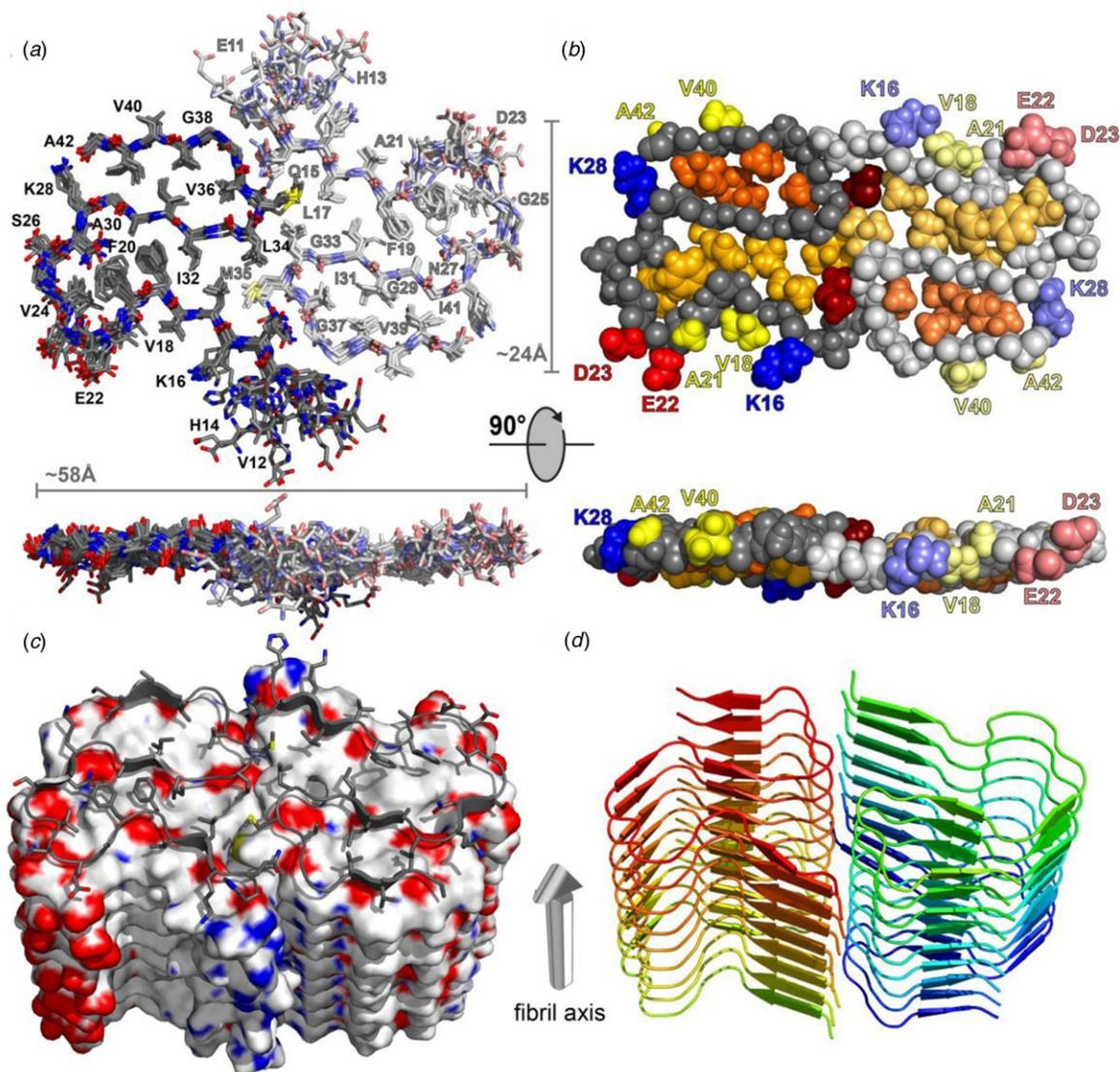


Figure 1.4 Dimeric structure of Aβ₁₋₄₂ as solved by Colvin *et al.* by ssNMR measurements. Only residues Q15-A42 are shown. (a) Overlay of the 10 lowest-energy NMR structures. Left monomer is shown in dark colors; right monomer is shown in pale colors. (b) CPK model showing backbone in gray, hydrophobic side chains are shown in yellow (solvent-exposed), gold, and orange (buried clusters). (c) Surface model of Aβ₁₋₄₂. (d) Ribbon model of Aβ₁₋₄₂. Reprinted with permission from Colvin *et al.* (2016). *Journal of the American Chemical Society*, 138 (30), 9663–9674. DOI: 10.1021/jacs.6b05129. Copyright 2016 American Chemical Society.

chemical functionality along the contour length of the assembly. The high density of functional groups

along the surface of the filaments can often promote self-association into higher-order structures (Wang *et al.*, 2021a). The influence of these interactions can be difficult to accurately predict since, in isolation, the energetic contributions of such local interactions to filament stability may be relatively small. In addition, researchers have observed that a small number of directed mutations at surface positions may convert soluble proteins into filamentous assemblies without perturbing the folded structure of the protomer (Garcia-Seisdedos *et al.*, 2017, 2019). Due to the lability of helical symmetry and the resultant plasticity of quaternary structure, many synthetic peptide filaments display structural heterogeneity, often in conjunction with kinetically controlled and potentially chaotic self-assembly behavior (Wang *et al.*, 2021b). High-resolution structural analysis provides critical structural insight that can inform experimental studies directed toward development of a mechanistic understanding of the elements that determine supramolecular structure and that can potentially enable control of function.

In subsequent sections, representative structural analyses of different classes of synthetic filaments are described in order to convey our current understanding of the relationship between sequence design and structure. Since the majority of assemblies described here derived from chiral substrates, i.e. peptides and enantiopure peptido-mimetics, these structural studies can potentially provide insight into the relationship between molecular chirality and supramolecular chirality. However, while chiral monomers may display a preference for a given helical hand in the corresponding assemblies, this correlation is not necessarily absolute (Harper *et al.*, 1997; Chamberlain *et al.*, 2000). Twist polymorphism, in which a chiral peptide monomer can assemble into either a right-handed or left-handed supramolecular enantiomorph, has been observed for amyloid fibrils assembled *in vivo* or *in vitro* (Usov *et al.*, 2013; Kollmer *et al.*, 2019; Wu *et al.*, 2021). This phenomenon may also apply to helical filaments derived from self-assembly of designed peptides and peptido-mimetic

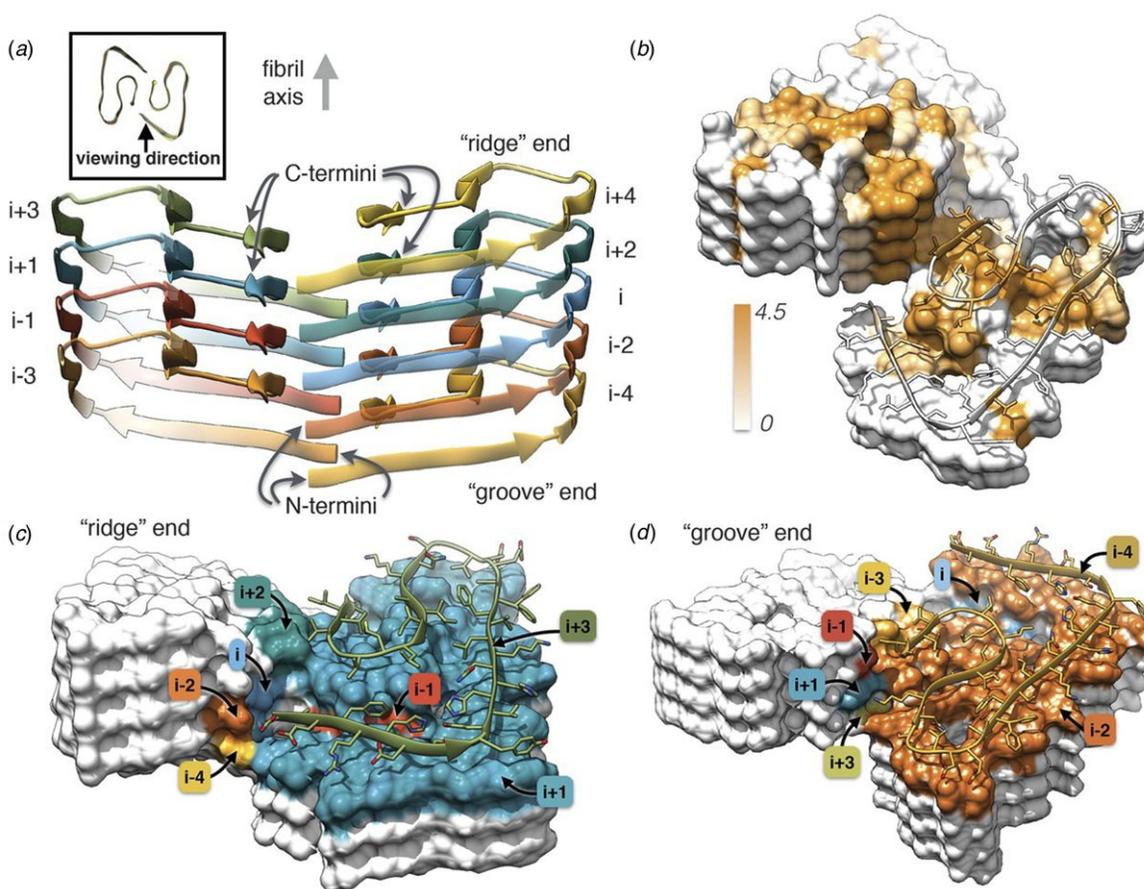


Figure 1.5 Dimeric structure of $A\beta_{1-42}$ as solved by Gremer *et al.* by cryo-EM with helical reconstruction. (a) Opposing assemblies are staggered, producing a pseudo- 2_1 screw axis in the filament. (b) Surface hydrophobicity of a representative segment the $A\beta_{1-42}$ peptide filament. Brown represents hydrophobicity value of 4.5, white represents 0.0 according to the Kyte–Doolittle scale. (c, d) Model of the ‘ridge’ and ‘groove’ ends of the $A\beta_{1-42}$ filament. Colors correspond to layers in (a). From Gremer *et al.* (2017). *Science*, 358 (6359), 116–119. Reprinted with permission from AAAS.

foldamers. Helical hand information is lost in two-dimensional (2D) projection images obtained from TEM and cryo-EM analysis and in many cases cannot be recovered through building of the atomic model into the two enantiomorphic representations of the three-dimensional (3D) volume (Wang *et al.*, 2022). In the absence of structural determinations at true-atomic resolution or easily resolved helical surface features, it may not be possible to unambiguously assign the helical hand of synthetic helical filaments

and tubes.

Short-peptide assemblies

The conventional wisdom in peptide science posits that oligopeptides of length less than five amino acids might represent the least attractive substrates for construction of synthetic assemblies. Due to their limited size, these peptides would not be expected to adopt persistent secondary structures or form stable, structurally ordered assemblies. However, dipeptides and tripeptides can be more easily crystallized than longer-peptide sequences and the corresponding crystal structures often exhibit extensive intermolecular interactions between peptides that mimic the interfaces proposed to exist within self-assembled peptide filaments. Similar to crystallographic analyses reported for amyloidogenic peptides, high-resolution data from these crystal structures can provide insight into the nature of the interfacial interactions between peptides that can potentially inform the design of more complex sequences. Consequently, the structural analysis of assemblies derived from short peptides has been important to the intellectual development of the field of peptide self-assembly. Structural investigations of simple dipeptides date back more than two decades and reveal surprisingly complex supramolecular arrangements between peptides in the crystalline state. In 2001, Görbitz demonstrated that simple dipeptides could crystallize into nanotube arrays (Görbitz, 2001). Subsequently, Reches and Gazit analyzed the self-assembly behavior of the Phe–Phe dipeptide and observed that it could self-associate into high aspect-ratio, nanoporous crystalline filaments (Reches and Gazit, 2003). Since these initial reports, significant research effort has been directed toward exploration of the potential of short peptides for the formation of filamentous assemblies for diverse applications (Raeburn *et al.*, 2013). Oligopeptides have several advantages as materials, most notably ease of preparative-scale synthesis, which provides access to sufficiently large quantities of pure oligopeptides for detailed experimental studies. In addition, Tuttle, Ulijn, and co-workers demonstrated that coarse-grain molecular dynamics simulations could be employed to screen combinatorial sequence space *in silico* to identify dipeptides and tripeptides that might display the potential to form ordered supramolecular assemblies (Frederix *et al.*,

2011, 2015). Subsequent experimental studies have borne out the hypothesis that these computationally screened peptides can form ordered supramolecular assemblies.

Phe–Phe, the best studied synthetic dipeptide system, crystallizes in a unit cell with $P6_1$ space symmetry (Görbitz, 2001). The dipeptide subunits form a head-to-tail supramolecular polymer in the crystal (Figure 1.6a). The chain of Phe–Phe molecules winds into a helical arrangement that defines a water-lined channel. Within a helical turn, the dipeptide segments are held together through a network of hydrogen-bonding interactions. The phenylalanine side chains are directed outward from the nanotube core and mediate lateral interactions between the helical arrays. Reches *et al.* described conditions that enabled the Phe–Phe peptide to self-assemble into crystalline nanotubes with an average diameter of 100–150 nm. Self-assembly of Phe–Phe from a mixed solvent system promoted crystal growth preferentially along the long axis of the tubular filament (Reches and Gazit, 2003). Subsequently, Görbitz and Kim, and co-workers, independently demonstrated that the peptide structure within the self-assembled tubes was identical to that observed in single crystals of the Phe–Phe dipeptide (Görbitz, 2006; Kim *et al.*, 2010).

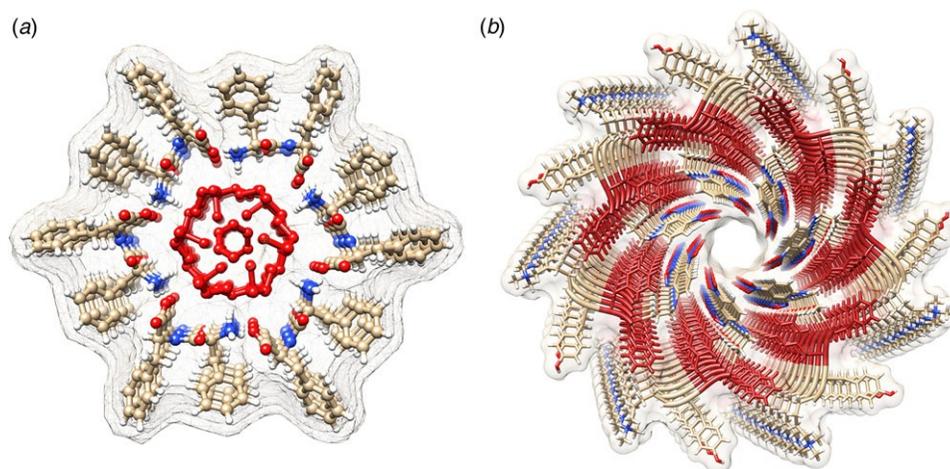


Figure 1.6. (a) View down the helical channel within the crystal structure of the Phe–Phe dipeptide. Density due to water molecules can be observed in the lumen of the tube. (b) View down the channel in the atomic model of the $C7$ helical filament derived from self-assembly of tetrapeptide 1-KMe₃. The ^DPhe–^DPhe segment of each protofilament is highlighted in red.

Many examples of peptides containing Phe–Phe segments have been prepared and examined in terms of their self-assembly behavior and potential use as nanomaterials (Marchesan *et al.*, 2015). These systems display significant promise as hydrogelators, especially under conditions in which a solution pH change drives filament formation (Raeburn *et al.*, 2013). Recently, Feng *et al.* reported the cryo-EM analysis of helical filaments derived from a fluorophore-modified tetrapeptide containing a ^DPhe–^DPhe sequence (Feng *et al.*, 2020). The non-natural stereochemistry of the Phe–Phe sequence was employed to stabilize the resultant assemblies with respect to proteolysis *in vivo*. Two distinct structural polymorphs were observed for the resultant filaments, which were based on either C_7 or C_2 (distorted C_6) helical symmetry. An atomic model (PDB: 6X5I) was generated for the C_7 filament at

4.3 Å resolution (Figure 1.6*b*). A superficial similarity was observed between this filament structure and the crystal structure of the Phe–Phe peptide. In both assemblies, the protofilaments were based on stacking of peptides such that the backbone was oriented in a plane perpendicular to the central helical axis. In either case, the protofilaments associated to form an oligomeric nanotube that defined a solvent-accessible central lumen. However, the structure of the tetrapeptide assembly was distinct from the crystal structure of Phe–Phe nanotube in that the peptide backbone of 1-KMe₃ extended radially outward with respect to the helical axis of the nanotube, in contrast to the circumferential arrangement observed in the crystal structure of the Phe–Phe dipeptide. These results suggested that while Phe–Phe may be employed as a minimalist self-assembling peptide segment, the helical structures of the resultant assemblies can vary significantly depending on the sequence context even for relatively short peptides.

In further support of this hypothesis, Bera *et al.* recently reported the structural analysis of assemblies derived from two related tripeptides, Pro–Phe–Phe and Hyp–Phe–Phe in which imino acids were positioned N-terminal to the well-studied Phe–Phe sequence (Bera *et al.*, 2019). The presence of proline derivatives within a peptide sequence has often been observed to disrupt periodic secondary structures due to conformational restrictions, particularly in the accessible range of ϕ torsions, and the inability of the endocyclic imide group to serve as a hydrogen bond donor (Reiersen and Rees, 2001).

Crystallographic analysis of the respective peptides indicated that they adopted similar structures that displayed a local helical conformation rather than an extended, β -strand conformation that is more typically observed for peptides containing Phe–Phe segments. The tripeptide units were stacked into an extended helix in which the phenylalanine side chains radiating outward. Peptide helices interacted laterally through formation of a phenylalanine zipper at the respective helix–helix interfaces. In contrast, the sequence permuted variants, Phe–Pro–Phe and Phe–Phe–Pro, formed β -sheet structures. As observed for longer peptides (*vide infra*), small sequence modifications can drive the assembly down an alternative folding pathway that results in different quaternary structures.

Cross- β filaments

Cross- β filaments were the first synthetic peptide assemblies that were investigated that resulted from rational design efforts. Initial designs focused on sequences that, when assembled into a β -strand, displayed facial amphiphilicity such that self-assembly resulted in the formation of an amphipathic β -sheet (Bowerman and Nilsson, 2012). These sequences usually comprised alternating patterns of hydrophobic and hydrophilic amino acids such that the polar repeat pattern reinforced the fiber repeat of a β -sheet (Pauling and Corey, 1951). The β -strands readily selfassociate into cross- β filaments through hydrogen bonding interactions. Pairwise association of protofilaments usually accompanies self-assembly, which results in burial of the hydrophobic surfaces of the two sheets at the interface between protofilaments. The rational design of synthetic cross- β filaments was initially reported in the 1990s when several research groups realized that introduction of self-complementary electrostatic and hydrogenbonding interactions between side chains could strongly bias an oligopeptide sequence toward formation of a stable β -sheet filament (Zhang *et al.*, 1993, 1994; Aggeli *et al.*, 1997, 2001; Marini *et al.*, 2002). Zhang and co-workers were among the early entrants in this field and demonstrated that ionic selfcomplementary peptide sequences could form robust β -sheet assemblies. Most of these peptide designs assumed that β -strands would pack in an antiparallel arrangement in the fibrils, which would result in cross-strand pairing of oppositely charged residues (Zhang *et al.*,

1993). However, Cormier *et al.* demonstrated, using ssNMR measurements, that RADA-16, one of the most thoroughly investigated of these peptides, assembled into a bilayer filament derived from self-association of parallel β -sheet protofilaments (Cormier *et al.*, 2013). The atomic model of the filament was based on a face-to-face (homotypic) sheet packing interface between two protofilaments corresponding to the individual β -sheets. Peptides within a protofilament adopted a progressive two-residue shift that enabled electrostatic

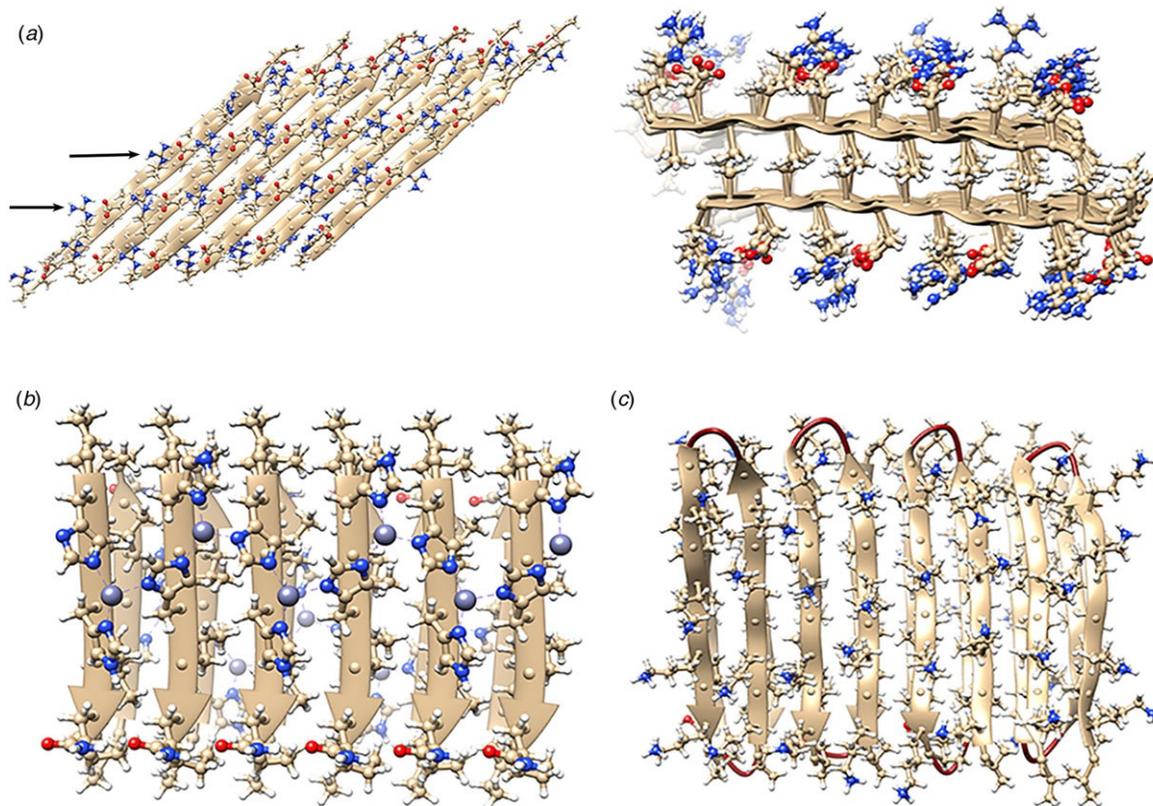


Figure 1.7. (a) Facial and side views within the atomic model corresponding to the tilted bilayer filament of peptide RADA-16. Arrows indicate the complementary electrostatic interactions between cross-strand arginines and aspartic acid residues. (b) Facial view of the atomic model of the cross- β fibrils of peptide HHQ. Zinc ions (gray spheres) bridge adjacent strands through coordination to histidine residues. (c) Facial view of the atomic model of the MAX1 bilayer filament. Adjacent hairpins in each layer are *syn* to each other, but *anti* to the corresponding hairpin in the other layer.

interaction between oppositely charged arginine and aspartic acid residues on adjacent strands in the parallel β -sheet (Figure 1.7a). To accommodate this registry shift, the chain axes of the peptides within the filament were tilted 35° with respect to a plane parallel to the fibril axis.

Since these initial studies, a number of β -sheet-forming peptides have been designed based on similar sequence considerations (Bowerman and Nilsson, 2012). However, only a few of these synthetic

cross- β filaments have been structurally investigated at NAR. Lee *et al.* reported the ssNMR structure (PDB: 5UGK) of a cross- β fibril based on an amphipathic heptapeptide sequence, HHQ (Ac-IHVHLQI-NH₂) (Lee *et al.*, 2017). In the presence of Zn²⁺ ions, the peptide assembled into cross- β fibril in which the parallel, in-register alignment was maintained through coordination of zinc ions by facially proximal histidine residues (Figure 1.7b). These results demonstrate that metal ion coordination can serve as a method to introduce structural specificity within a peptide filament as well as a selective mechanism to drive peptide self-assembly.

Schneider, Pochan, and co-workers, designed a series of synthetic peptides in which a synthetic type-II' β -turn enforced a β -hairpin conformation between two antiparallel β -sheet strands (Schneider *et al.*, 2002). MAX1 (Ac-VKVKVKVKV^{DP}LPTKVKVKVKV-NH₂), a typical peptide in this series, undergoes pH-driven or salt-driven self-assembly into hydrogels that involved the formation of cross- β fibrils. The MAX1 peptide design was based on an alternating sequence of hydrophilic and hydrophobic amino acids. This polar patterning should favor the formation of an amphipathic β -sheet upon charge neutralization or electrostatic screening. The heterochiral turn sequence, V^{DP}LPT, limits the accessible conformational space of the peptide and serves as a constraint that promotes self-assembly (Lamm *et al.*, 2005). Nagy-Smith *et al.* reported the structural analysis of MAX1 using ssNMR measurements (PDB: 2N1E) (Nagy-Smith *et al.*, 2015), which indicated that MAX1 formed a monomorphic bilayer fibril in which adjacent β -hairpins within a fibril adopted a *syn* geometry (Figure 1.7c). The two protofilaments pack face-to-face and are oriented in an *anti*-arrangement such that the turn surfaces are on opposite edges of the protofilaments that comprise the bilayer fibril. Self-assembly is driven through burial of hydrophobic surface area at the interface between the cross- β protofilaments. These results demonstrate the power of NAR structural methods to interrogate the details of peptide packing within synthetic β -sheet fibrils.

While cryo-EM analysis has been employed extensively for structural analysis of amyloid fibrils (Gallardo *et al.*, 2020), it has not been extensively applied to the structural analysis of designed cross- β assemblies (see the following section for examples of the cryo-EM analysis of β -sheet nanotubes). While

these designed assemblies have been routinely employed for the construction of synthetic biomaterials for medical applications, the limited availability of structural information at NAR hinders the development of this field (Wang *et al.*, 2021b). One potential benefit of these analyses would be to gain insight into the surface structure of the filament, which has a critical role in mediating interactions at the cell–biomaterial interface. This structural information is essential for development of functional biomaterials for biomedical applications. Cryo-EM may represent the best experimental approach for the structural analysis of synthetic cross- β fibrils in that it can potentially unravel the speciation of different structural polymorphs, as has been accomplished for amyloid fibrils (Zhang *et al.*, 2019).

Cross- β nanotubes

Structural analyses of filaments derived from self-assembly of a number of β -sheet forming peptides have provided evidence for the formation of thin-walled, wide-diameter nanotubes based on a cross- β fibril supramolecular architecture. Initial studies focused on oligopeptides derived from amyloidogenic peptide sequences. While not resulting from *de novo* design, the corresponding peptide sequences assembled into filamentous structures that appeared quite distinct from the cross- β fibrils typically observed from self-assembly of amyloidogenic peptides and proteins having much longer sequences. The short length of these peptide sequences provided an opportunity to examine the effect of site-directed mutagenesis on self-assembly within a well-defined sequence context. *De novo* design of oligopeptide sequences for cross- β nanotube formation has lagged to some degree due to the absence of structural information at NAR. However, cryo-EM analysis has provided the opportunity to generate reliable atomic models for β -sheet nanotubes, as recently illustrated for two separate peptide assemblies (Wang *et al.*, 2021b; Pieri *et al.*, 2022).

In 2003, Lynn and coworkers reported the formation of wide-diameter, thin-walled nanotubes from self-assembly of a hydrophobic heptapeptide derived from the A β sequence. This experimental study represented a tipping point in the structural analysis of β -sheet nanotubes from biologically derived to bio-inspired assemblies (Lu *et al.*, 2003). The truncated sequence Ac-¹⁶KL¹⁷V¹⁸F¹⁹F²⁰A²¹E²²-NH₂ (A β ₁₆₋₂₂)

self-associated into nanotubes under acidic conditions (pH 2) in which the negative charge of the carboxylate side chain of the terminal glutamic acid was neutralized. The outer diameter of the nanotubes was estimated as ~ 520 Å with a wall thickness of 40 Å. In contrast, at neutral pH, A β_{16-22} formed thin filaments consisting of a pair of twisted protofilaments of ~ 50 Å in diameter. The latter assemblies were consistent with the ultrastructure typically observed in low-resolution TEM images of amyloid fibrils (Schmidt *et al.*, 2016). A combination of SAXS/SANS measurements, powder X-ray and electron diffraction, and ssNMR was employed to provide insight into the structural differences between the two morphological forms (Mehta *et al.*, 2008). The protofilament structure at neutral pH was based on a stack of five cross- β fibrils (Figure 1.8*a* and *b*), in which adjacent A β_{16-22} peptides were arranged in exact registry in an antiparallel β -sheet. In contrast, the A β_{16-22} nanotubes are arranged in an antiparallel β -sheet in which adjacent peptides adopt an alternating single-residue offset (Figure 1.8*c* and *d*). Each nanotube consists of an indeterminate number of cross- β protofilaments based on a bilayer structure. Individual protofilaments are oriented at an angle of $\sim 23^\circ$ with respect to the helical axis of the tube. Conventional TEM and AFM measurements, performed under acidic conditions, indicated that bilayer ribbons formed in solution within 30 h. The ribbons eventually closed to form the nanotube through fusion of the edges. The evolution of morphology from twisted ribbons to helical ribbons to tubes has been observed often for self-assembly processes involving chiral monomers, including amyloidogenic peptides and proteins (Selinger *et al.*, 2004; Ziserman *et al.*, 2011; Adamcik and Mezzenga, 2018).

Like A β , α -synuclein (α -syn) is a natural protein that self-assembles into cross- β fibrils. The presence of these fibrils has been associated with degeneration of dopaminergic neurons that is a symptom of Parkinson's disease (Spillantini *et al.*, 1997). Morris *et al.* demonstrated that an eight-residue truncation product derived from α -syn, α -S β 1, NH₂-³⁷VLYVGSK⁴⁴T-COOH, was able to form a helical ribbon structure (Figure 1.9*a*) (Morris *et al.*, 2013). The initial assemblies evolved over time; subsequently forming a closed nanotube through sealing of the edges of the tape (Figure 1.9*b*). Structural models were constructed for the ribbon and tube using medium-resolution data acquired from a

combination of conventional TEM, FTIR, and X-ray fiber diffraction analyses (Figure 1.9c-e). The peptide sequence displays contour-length amphiphilicity with a hydrophobic N-terminal segment and hydrophilic C-terminal segment. Self-assembly resulted in the formation of a nanotube based on an amphiphilic bilayer in which the hydrophobic segments were sequestered in the interior and the hydrophilic segments decorated the convex and concave surfaces. Each leaflet within the bilayer consisted of peptides packed in parallel β -sheet in which the peptides were oriented in an antiparallel arrangement across the bilayer interface. The α -S β 1 tubes were ~ 2400 Å in diameter, which was significantly wider than most peptide-based nanotubes resulting from self-assembly of either biologically derived or synthetic β -sheet peptides.

The self-association behavior of the α -S β 1 peptide was similar to that of A β_{16-22} in that wide-diameter, thin-walled nanotubes resulted from *in vitro* assembly under ambient conditions. While the sequences were different, both peptides displayed contour-length amphiphilicity, i.e. the sequence could be formally parsed into distinct polar and non-polar segments along the peptide backbone. This sequence pattern more closely resembled conventional amphiphiles

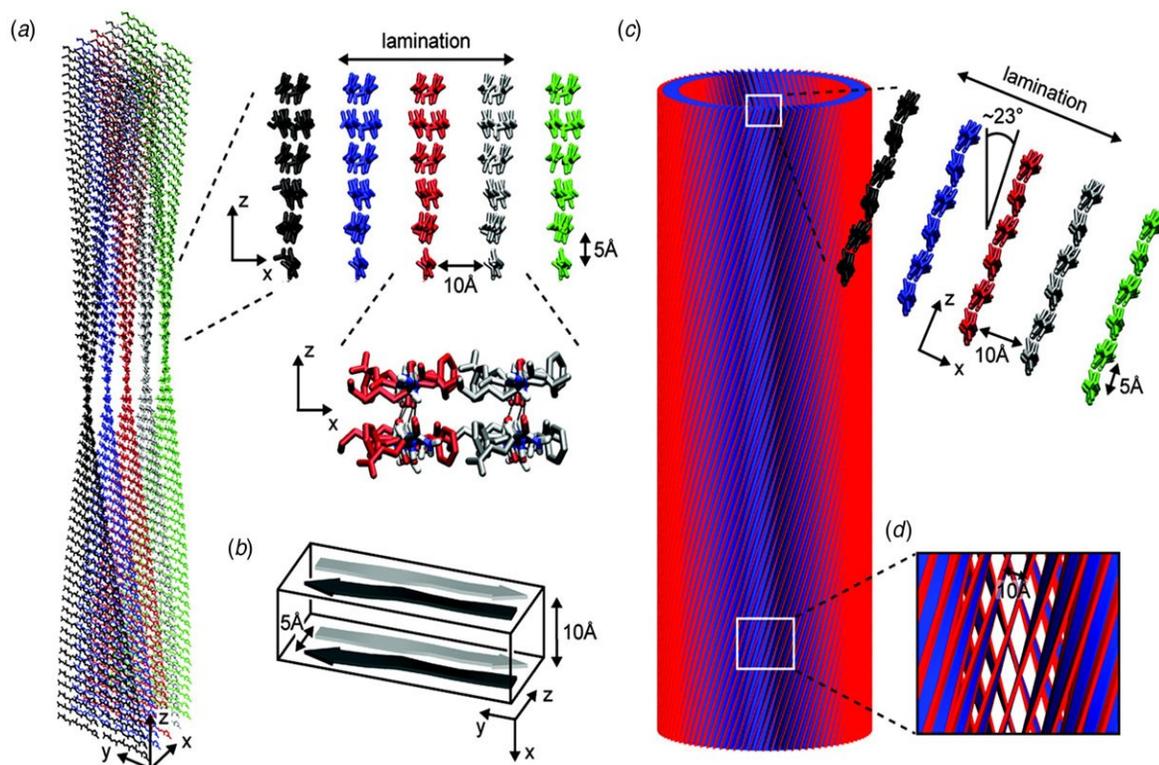


Figure 1.8. Morphological variants of $A\beta_{16-22}$. (a) Atomic model of the laminated filament of $A\beta_{16-22}$ observed at neutral pH. (b) Proposed unit cell describing the packing of peptides in the laminated filament of $A\beta_{16-22}$. (c) Atomic model of the bilayer nanotube of $A\beta_{16-22}$ observed at acidic pH. (d) Expanded representation of the tilted packing of β -strands within a bilayer nanotube of $A\beta_{16-22}$. Reprinted with permission from Mehta *et al.* (2008), *Journal of the American Chemical Society*, 130 (30), 9829–9835. Copyright 2021 American Chemical Society.

such as phospholipids or amphiphilic block copolymers, rather than the facially amphiphilic cross- β fibrils described in the preceding section. Consequently, the layered packing of peptides within the respective nanotubes is reminiscent of the leaflet structure of lipids, in which the hydrophobic portions of the peptide sequence are buried in the core of a monolayer or bilayer with the polar groups decorating the solvent-contacting surfaces (Figure 1.8, Figure 1.9).

Amyloidogenic peptide segments can be incorporated into more complex sequence architectures to promote unique modes of nanotube self-assembly. Nowick and coworkers reported the crystal

structure of a macrocyclic β -sheet peptide, 1, derived from the A β_{16-22} sequence (PDB: 5VF1) (Chen *et al.*, 2017). The KLVFFAE sequence was combined with EAFFVLK, its retrosequence, in a cyclic arrangement in which the two peptide segments were covalently connected with δ -ornithine amide-bond linkages (Figure 1.10*a*). For this peptide sequence, macrocyclization limited the accessible conformational space that the peptide could adopt, which presumably restricted the range of supramolecular structures that could result from self-assembly. Crystallographic analysis revealed the presence of nanotubes that formed a porous, hexagonally close-packed lattice under the $P6_122$ space group symmetry of the crystal structure (Figure 1.10*b*). Individual nanotubes within the crystal were double-walled and the asymmetric unit consisted of six macrocyclic peptides, in which the inner (concave) and outer (convex) walls of the nanotube were composed of helical arrays of structurally distinct dimers and tetramers, respectively. The dimers at the concave surface self-associated through the formation of hydrogen bonds between the backbone of each monomer and were further stabilized through complementary charge interactions between β -strands (Figure 1.10*c* and *d*). The tetramers associated via a combination of hydrogen bonds and hydrophobic interactions in a β -barrel-like structure (Figure 1.10*c* and *e*). Each tetramer was associated with four other tetramers by hydrogen bonds.

The hexagonal close-packed

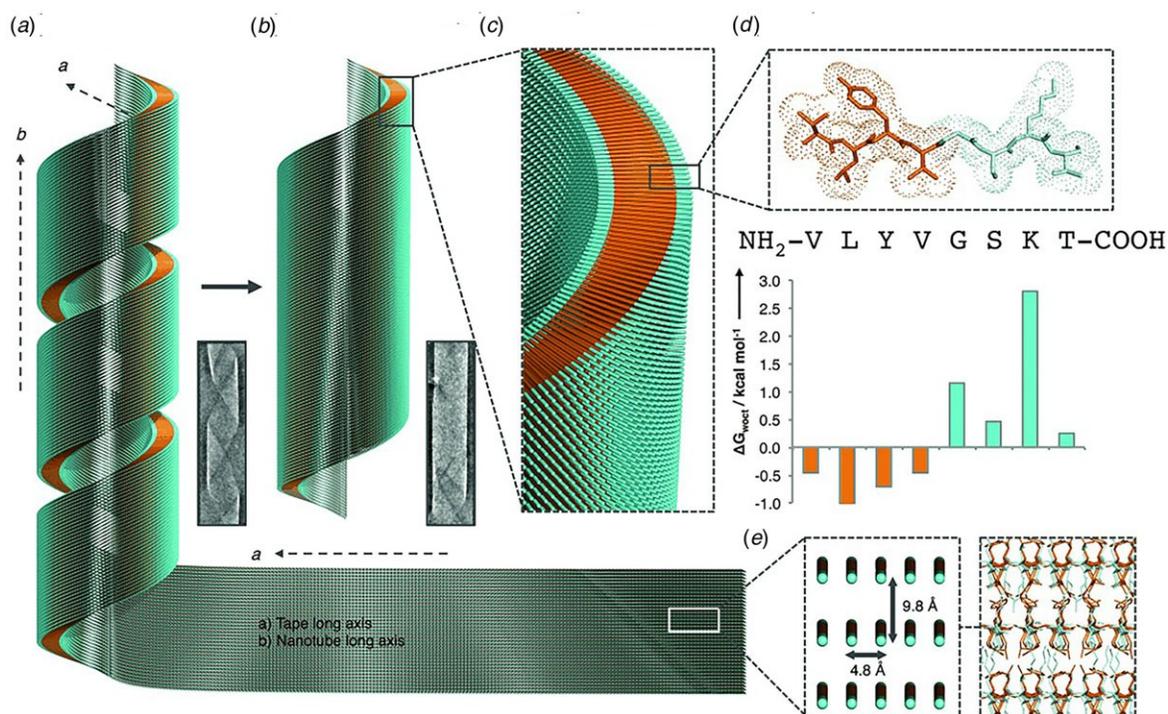


Figure 1.9. Proposed mechanism of assembly and structural model of nanotubes derived from self-assembly of α -S β 1 peptides. (a) Model of the initially formed helical ribbon with associated TEM image. (b) Model of the mature tubes with associated TEM image. (c) Cross-sectional depiction of the mature tube indicating the packing arrangement of peptides in the amphiphilic monolayer. (d) Sequence and stick model of the α -S β 1 peptide. Residues are colored to indicate amphiphilic character based on water-octanol transfer free energies. (e) The orientation of the α -S β 1 strands are shown in the context of the tape then leading to the nanotubes. The single peptides are represented as lines with hydrophobicity and hydrophilicity shown as orange and cyan, respectively. Reprinted with permission from Morris *et al.* (2013), *Angewandte Chemie International Edition*, 52 (8), 2279–2283.

arrangement of tubes, viewed along the crystallographic *c*-axis, was consistent with the formation of assemblies in solution and subsequent lateral association in the crystal structure. A similar mechanism has been proposed for the formation of nematic liquid crystalline phases in the self-assembly of nanotubes from the A6K peptide (*vide infra*) (Bucak *et al.*, 2009). It is interesting to note that despite

the conformational constraints that macrocyclization imposed on the peptide sequence, the resultant structure, especially the outer wall of the nanotube, would have been difficult to predict based on our current knowledge of peptide and protein quaternary structure.

Contour-length amphiphilicity has been employed as an explicit design principle to engineer the sequences of synthetic self-assembling peptides. When these designs incorporate amino acids that display a preference for the formation of β -strand conformation, self-assembly of the corresponding sequences often resulted in supramolecular architectures based on β -sheet formation. Zhang and co-workers were the first researchers to employ this concept to design a class of synthetic surfactant-like peptides (SLPs) based on sequences that displayed contour-length amphiphilicity (Vauthey *et al.*, 2002). In contrast to the adventitious amphiphilicity of the $A\beta_{16-22}$ -derived and α -S β 1 peptides, Zhang's SLP designs were directly analogous to small-molecule amphiphiles. The peptide sequences consisted of a polar head group of one or two charged amino acid residues, and a non-polar tail composed of a short sequence of hydrophobic amino acids. The properties of these peptide surfactants have been studied extensively over the past two decades (Hamley, 2011). Thus far, however, the structures of the resultant assemblies have not been described at NAR.

The surfactant peptide, NH_2 -AAAAAAK-COOH (A6K), has been the most thoroughly investigated member of this class (Bucak *et al.*, 2009; Castelletto *et al.*, 2010; Cenker *et al.*, 2012; Middleton *et al.*, 2013). Above a critical volume fraction, A6K formed a nematic meso-phase composed of an apparently homogeneous population of nanotubes with an approximate mean diameter of 550 Å and an estimated shell thickness of 33 Å. SAXS, cryo-EM, ssNMR, and X-ray diffraction of flow-aligned nanotube solutions contributed structural insights that have led to a proposed structural model for the tubes in which protofilaments derived from antiparallel β -sheet architectures were arranged in an amphiphilic monolayer. In the proposed model, structurally adjacent A6K peptides within the cross- β protofilaments were oriented in register (Figure 1.11*a*). Tube formation was proposed to result from lamination of the protofilaments. The H-bonded network of the protofilaments was determined to be oriented at a pitch angle of $\sim 52^\circ$ with respect to the central axis of the nanotube. In contrast to A6K, the longer-peptide

variants, A8K and A10K, form twisted ribbon architectures despite remarkably similar local packing interactions within the unit cells of the respective peptide systems (Rüter *et al.*, 2020) (Figure 1.11b).

Peptide surfactants based on bolaamphiphile architectures have also been demonstrated to self-assemble into cross- β nanotubes. Bolaamphiphiles display an alternate mode of contour length amphiphilicity, in which the sequence was composed of a hydrophobic core with flanking terminal polar residues, e.g. RFL₄FR, EFL₄FE, or KI₄K (Zhao *et al.*, 2013; Da Silva *et al.*, 2015a, 2015b; Hamley *et al.*, 2017). For example, the bolaamphiphilic peptide Ac-KI₄K-NH₂ assembled into wide-diameter, thin-walled nanotubes, despite the potential for electrostatic repulsion between the terminal lysine residues. In contrast, the amphiphilic peptide Ac-I₄K₂-NH₂, in which charges are localized at the C-terminus, formed thin, twisted filaments, which suggested that the polar sequence pattern had an influence on supramolecular structure despite the fact that both peptides adopted a β -sheet conformation in the assembled state. Zhao *et al.* demonstrated that, in aqueous solution, hydrophobic interaction between the isoleucines in Ac-KI₄K-NH₂ drove the formation of nanotubes (Zhao *et al.*, 2015). In contrast, upon addition of acetonitrile (ACN), the peptides assembled into twisted tapes (20% ACN) or thin fibrils (80% ACN). According to CD data, the β -sheet structure was not disrupted in the presence of ACN. However, the addition of ACN reduced the polarity and dielectric constant of the aqueous solvent, which weakened the hydrophobic interaction between the side chains of the isoleucine residues. The weakening of the hydrophobic interactions was proposed to result in a lower degree of sheet lamination, which was hypothesized to underlie the morphological transition from tubes to ribbons to fibrils.

The short sequences of these bolaamphiphilic peptides represent a flexible platform to examine the effect of mutagenesis on the resultant supramolecular structure of the corresponding nanotubes.

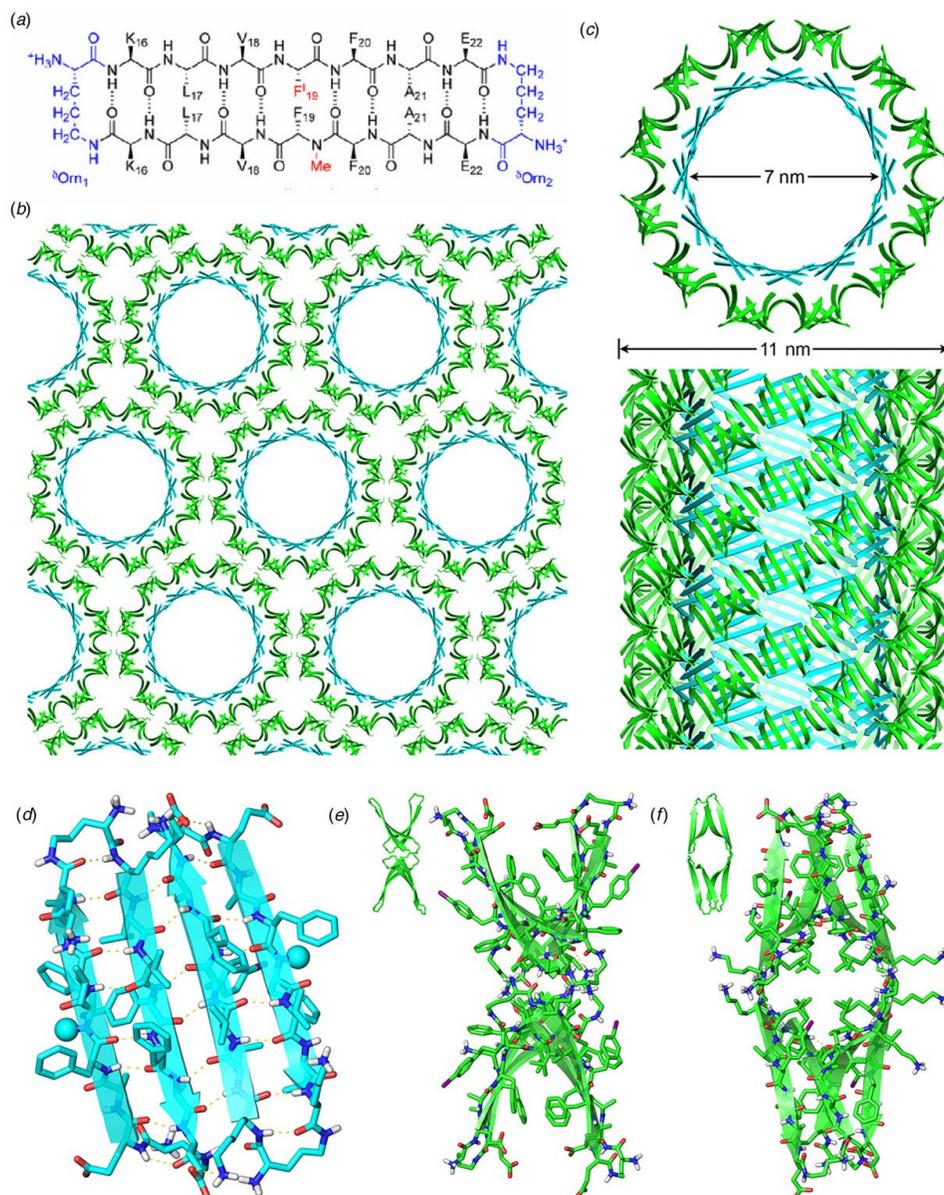


Figure 1.10. (a) Sequence of macrocyclic β -sheet 1. (b) Macrocyclic β -sheet 1 nanotubes pack into a honeycomb-like crystal lattice. (c) Top and side views of nanotube formed by macrocyclic β -sheet 1. (d) Structure of dimeric subunit at the inner wall of the nanotube. (e, f) Different views of the tetrameric subunit at the outer wall of the nanotube. Reprinted with permission from Chen *et al.* (2017), *Journal of the American Chemical Society*, 139 (24), 8102–8105. Copyright 2021 American Chemical

Interestingly, substitution of residues within the hydrophobic core was observed to influence the diameter of the corresponding nanotubes (Zhao *et al.*, 2018). The parent sequence, Ac-KI₄K-NH₂, was modified through introduction of hydrophobic amino acids having different steric properties. In the series Ac-KI_{*n*}V_{4-*n*}K-NH₂ (*n* = 1–4), SANS, TEM, and cryo-EM measurements indicated that the diameter of the corresponding tubes decreased as the number of isoleucine residues (*n*) decreased, i.e. with increasing numbers of valine residues within the sequence (Figure 1.12). A similar effect was observed for substitutions of the C-terminal isoleucine in the parent sequence with either leucine or norleucine. In the latter situation, the peptides differed from isoleucine in the side-chain structure, which resulted in constriction of the diameter of the resultant nanotubes. Although the structures were not analyzed at NAR, these results indicated that the supramolecular structure of the nanotubes could be rationally varied through sequence modifications. More detailed structural analyses should provide insight into the interplay of factors that control higher-order assembly within this peptide family.

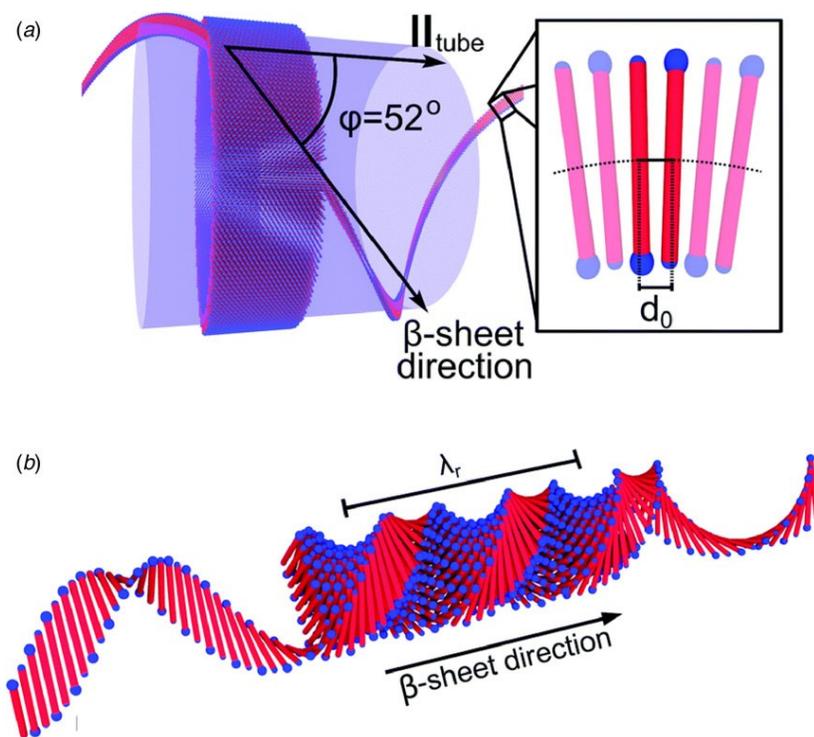


Figure 1.11 (a) Structural model of a nanotube derived from self-assembly of peptide A_6K . A single protofilament within the nanotube inclines at an angle of 52° with respect to the long axis of the assembly. Inset indicates the proposed antiparallel packing arrangement of peptides within a cross- β protofilament. (b) Structural model of a laminated ribbon derived from self-assembly of peptide A_8K or $A_{10}K$. The trajectory of single cross- β protofilament within the ribbon is highlighted. The quantity λ_r corresponds to the helical pitch of the ribbon. In both representations, the imposed helical hand was based on an arbitrary decision. Reproduced from Rüter *et al.* (2020), *Physical Chemistry Chemical Physics*, 22 (33), 18320–18327, with permission from the Royal Society of Chemistry.

Nanotube formation was also observed in a series of amphiphilic block co-polypeptides $Ac-[ALV]_x-b-[KGE]_y-NH_2$ ($x + y = 7$) (Van Rijt *et al.*, 2019). TEM analysis indicated that peptide $Ac-[ALV]_3-b-[KGE]_4-NH_2$ self-assembled to form high aspect-ratio nanotubes of ~ 90 Å in diameter that were stable over a wide pH range (pH = 2–12) and at temperatures up to 80 °C. Cryo-EM and SAXS

analysis confirmed the nanotube morphology, while FTIR spectroscopy supported the formation of β -sheet within the assembly. The proposed structure was based on a monolayer nanotube in which adjacent peptides were aligned in an antiparallel orientation such that the hydrophilic blocks decorated the inner and outer surfaces of the tube. A sequence with a smaller hydrophobic block, Ac-[ALV]₂-b-[KGE]₅-NH₂, remained as an unassociated monomer under the same conditions. Longer hydrophobic blocks resulted in kinetically trapped mixtures of nanotubes and cylindrical micelles or in macro-phase separation. In analogy to synthetic amphiphilic block copolymers, these results suggested that the hydrophilic–hydrophobic balance could be employed as a parameter to control self-assembly and supramolecular structure as has been observed for synthetic block copolymers (Jiao *et al.*, 2020). In the cases of peptide-based and peptido-mimetic materials based on sequence-specific oligomers, the hydrophilic–hydrophobic balance should be easily amenable to synthetic control.

Lanreotide acetate, an 8-residue, disulfide-linked, cyclic peptide hormone, has also been demonstrated to form nanotubes of homogeneous diameter (~ 280 Å) above a critical aggregation concentration in aqueous solution (Valéry *et al.*, 2003, 2004; Pouget *et al.*, 2010). Nanotube assembly coincided with formation of a hexagonal columnar meso-phase. The structure of the lanreotide nanotubes was initially investigated using a combination of EM, FTIR spectroscopy, and SAXS/WAXS diffraction analysis. Based on this evidence, the octapeptide was proposed to fold into a conformationally constrained β -hairpin. Monomers were suggested to initially associate to form face-to-face dimers (Figure 1.13) due to attraction between hydrophobic residues and repulsion between positively charged residues. The dimers were hypothesized to subsequently assemble into helical ribbons through aromatic interactions and hydrogen bonds (Valéry *et al.*, 2003, 2004; Pouget *et al.*, 2010). For the parent lanreotide sequence, the initial structural model for the lanreotide nanotubes was based on the self-association of 26 bilayer protofilaments, in which the peptide backbones are arranged circumferentially around the periphery of the nanotube (Chervy *et al.*, 2019).

The structure of the lanreotide tubes has been recently determined using cryo-EM analysis (Pieri *et al.*, 2022). Helical reconstruction from the projection images afforded a 3D density map at 2.5 Å resolution

(PDB: 7Q5A), which represented the highest resolution achieved thus far for a synthetic peptide nanotube. At this level of resolution, the helical hand of the symmetry could be assigned directly based on a structural comparison of atomic models constructed through direct fitting into the two different enantiomorphs that resulted from mirroring the density map. The helical symmetry of the lanreotide model was best described in terms of a right-handed 1-start helix with a rise of 1.04 Å and a rotation of 26.2° (Figure 1.14). Surprisingly, the asymmetric unit was based on eight peptides arranged in a monolayer shell in which the peptide backbones extend radially outward from the central axis of the assembly. The hydrogen-bonding direction, corresponding to the cross- β structural interaction, occurs between peptides in every 27th asymmetric unit. The left-handed twist (-12.1°) of this 27-start helix is consistent with the left-handed twist observed for β -sheet in globular proteins and amyloid fibers (Chothia, 1973; Chamberlain *et al.*, 2000). The lanreotide nanotube structure differed significantly from the original model and highlighted the necessity of high-resolution structural analysis to fully understand peptide packing within

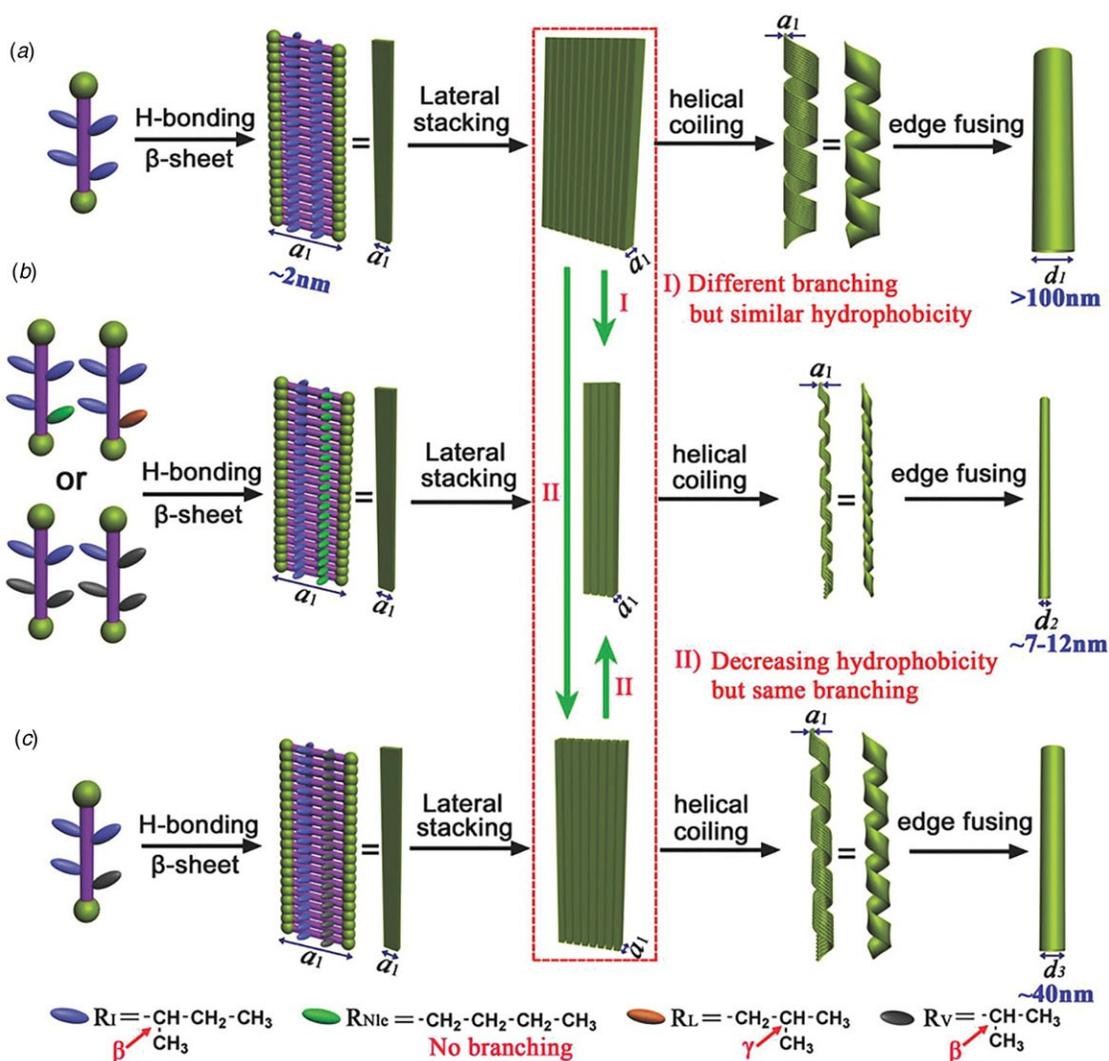


Figure 1.12. Schematic illustration of the self-assembly of peptide nanotubes derived from site-directed mutagenesis of peptide Ac-KI₄K-NH₂. (a) Proposed model for self-assembly of the parent peptide Ac-KI₄K-NH₂. (b) Thin tubes are formed by changing side-chain branching and hydrophobicity, which influences the degree of lateral stacking of β -sheets. (c) A single mutation of isoleucine to valine in Ac-KI₃VK-NH₂ leads to an intermediate degree of sheet lamination and modest decrease in nanotube diameter. The assignment of helical hand was based on AFM measurements. Reprinted with permission from Zhao *et al.* (2018), *Small*, 14 (12), 1703216.

helical filaments. In addition, unlike many synthetic peptide filaments, the population of lanreotide nanotubes was uniform in diameter with negligible polymorphism. The high degree of structural

uniformity within the assembly may have derived from the conformational constraints of the cyclic peptide monomer, which could have restricted interfacial interactions to specific geometries within the assembly through pre-organization.

The small size and constrained conformation of lanreotide acetate made it an attractive substrate to investigate the effect of sequence modifications and counterion effects on nanotube structure. Tarabout *et al.* demonstrated that substitution of the D-Trp4 residue with other aromatic amino acids led, in most cases, to retention of nanotube formation (Tarabout *et al.*, 2011). However, the diameter of the resultant monodisperse nanotubes depended on the size of the amino acid side chain at position 4 and ranged from 100 to 360 Å. Larger side chains increased the nanotube diameter, while smaller side chains had the opposite effect. The observed results were rationalized in terms of the influence of the molecular size parameter of the residue on the curvature at the concave interface between protofilaments. Counterion identity was also observed to influence diameter, although the effect could not be rationalized solely in terms of counterion size (Gobeaux *et al.*, 2012). The high-resolution structure of the parent lanreotide nanotubes should prompt a re-analysis of these mutagenesis results in terms of the structural influence of the substitutions on the packing of peptides in atomic model (Figure 1.14).

Paternostre and coworkers (Valery *et al.*, 2015) subsequently reported the structural analysis of assemblies derived from triptorelin, a decapeptide that acts as a gonadotropin-releasing hormone agonist. Like lanreotide, the sequence of triptorelin contains a D-Trp residue, which was proposed to be involved in the formation of a reverse turn. In aqueous solution, triptorelin formed nanotubes in which the diameter depended on the pH of the buffer. At low pH (<6.5), small diameter (~11 Å) tubes were formed, while larger diameter (~50 Å) tubes were formed at higher pH (>7.5). This transition was proposed to involve a switch in the protonation state of a histidine side chain within the peptide sequence, which induced a structural transition within the protomer. X-ray crystallographic analysis was performed

on

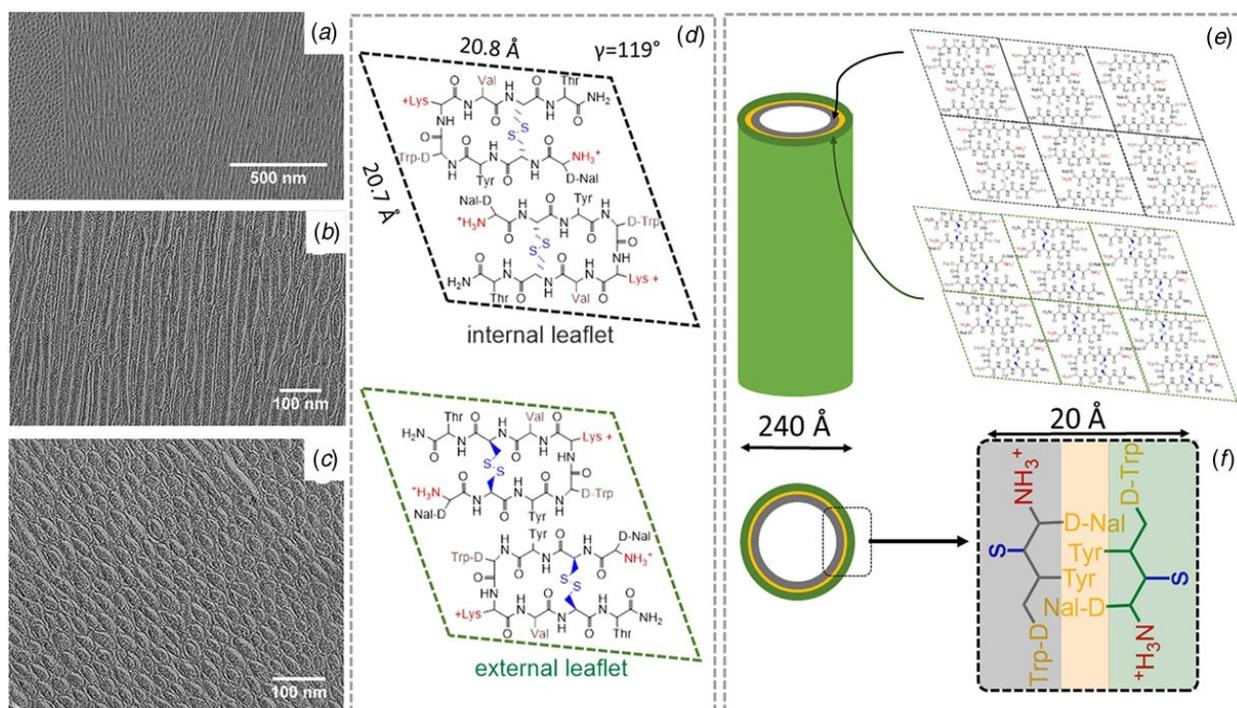


Figure 1.13. Initial model for the self-assembled nanotubes of the lanreotide peptide. Freeze-fracture micrographs of longitudinally (*a*, *b*) and transversely (*c*) fractured nanotubes indicated the formation of uniform-diameter assemblies. (*d*) Proposed arrangement of lanreotide monomers in the internal and external leaflets of the bilayer nanotube. (*e*) Initial structural model of the bilayer nanotube with an expansion of the internal and external leaflets corresponding to 2×3 asymmetric units. (*f*) Cross-sectional model of the bilayer nanotube indicating the proposed side-chain packing arrangement between monomers at the leaflet interface. Reproduced from Chervy *et al.* (2019), *Langmuir*, 35 (32), 10648–10657, with permission from the American Chemical Society.

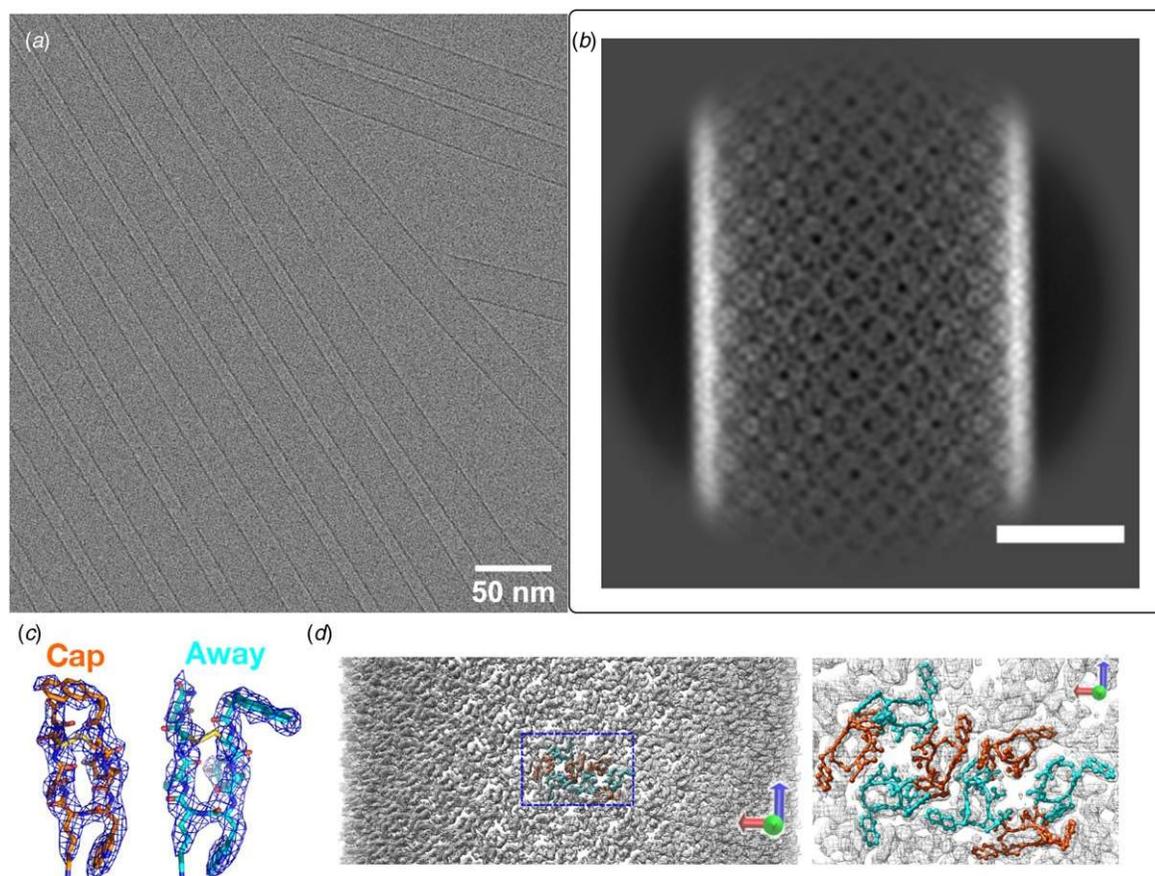
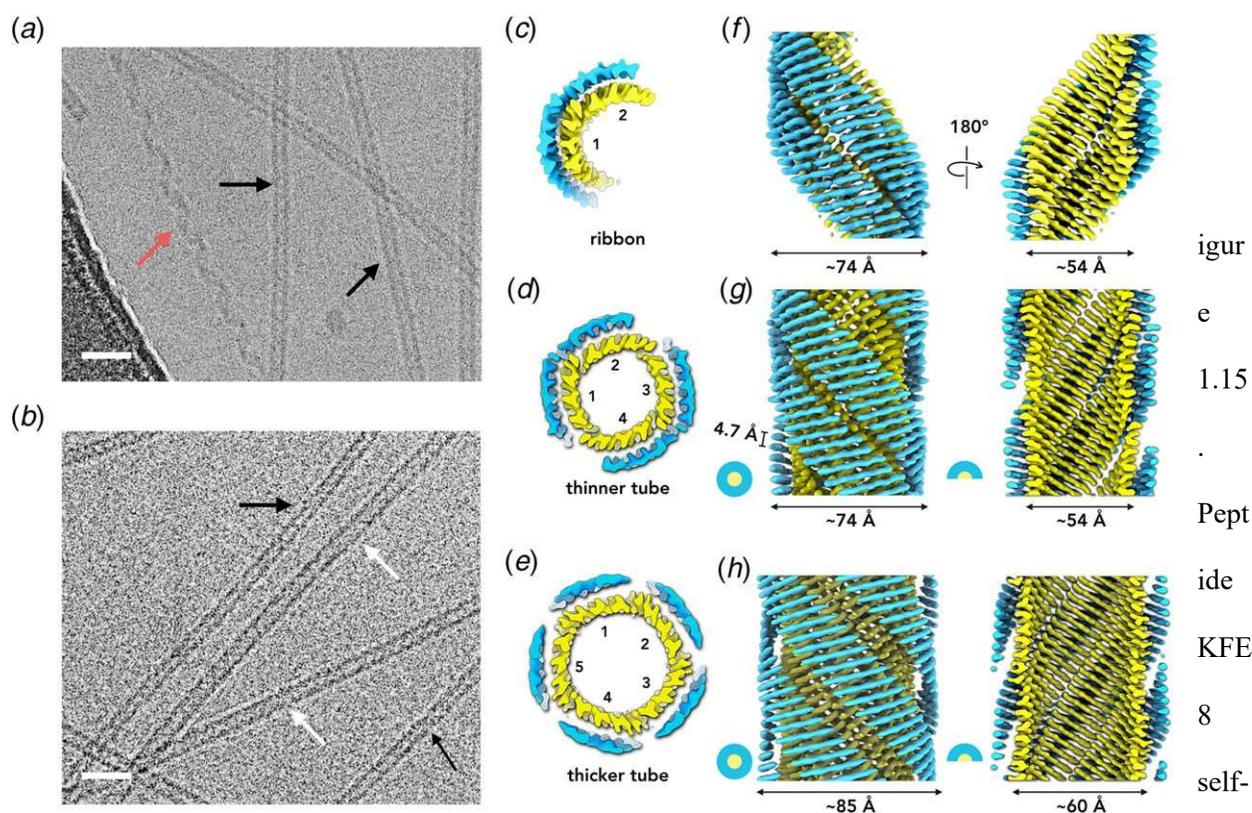


Figure 1.14. Cryo-EM structure of lanreotide nanotubes. (a) Representative cryo-EM image of peptide nanotubes derived from self-assembly of lanreotide acetate. (b) Representative 2D class average derived from cryo-EM analysis of lanreotide nanotubes. (c) Density maps for the two main conformations, capped *versus* away, of the lanreotide peptide within the nanotubes. (d) Density map for the lanreotide nanotube in which the eight peptides in an asymmetric unit are highlighted within the assembly. An expansion of the asymmetric unit is depicted on the right in which the capped (orange) and away (cyan) conformations are indicated.



assembles into helical ribbons and nanotubes. Representative cryo-EM images of KFE8 assembled at ambient temperature (a) or annealed at 90 °C (b). Scale bar = 20 nm. Red, black, and white arrows indicate the presence of bilayer ribbons, thinner tubes, and thicker tubes, respectively. Cross-sectional and transverse views of the atomic models for the ribbons (c, f), thinner tubes (d, g), and thicker tubes (e, h). The helical hand was assigned based on imaging data from AFM measurements. Reference: Wang, F.; Gnewou, O.; Wang, S.; Osinski, T.; Zuo, X.; Egelman, E. H.; Conticello, V. P. Deterministic chaos in the self-assembly of β -sheet nanotubes from an amphipathic oligopeptide. *Matter*, DOI:10.1016/j.matt.2021.06.037.

triptorelin at high pH under conditions that resulted in flat lamellar structures rather than wide-diameter nanotubes. The structure (PDB: 4D5M) indicated the absence of β -sheet and the formation of a small globular fold that was stabilized through a hydrogen bonding interaction between the side chains of His

(2) and Ser (4). The authors suggested that this interaction was lost at low pH, which resulted in the formation of β -sheet conformation.

Conticello and coworkers reported the cryo-EM structural analysis of assemblies derived from an amphipathic peptide KFE8, Ac-FKFEFKFE-NH₂ (Wang et al., 2021b). This peptide had been previously reported to self-assemble into left-handed helical ribbons (Marini et al., 2002; Hwang et al., 2003). However, over the course of assembly, ribbons and nanotubes were observed, although the structure of the nanotubes depended on the assembly conditions. Cryo-EM analysis was employed to generate NAR models for two ribbons and three nanotubes (PDB: LQE, 7LQF, 7LQG, 7LQH, and 7LQI). At ambient temperature, the nanotubes were based on a bilayer structure consisting of four β -sandwich protofilaments, while under annealing the peptide assembled into nanotubes comprising five β -sandwich protofilaments (

Figure 1.15). Although the structures of the nanotubes differed, each protofilament was based on an unprecedented packing arrangement in which an inner layer of parallel β -sheet was packed against an outer layer of antiparallel β -sheet. The dependence of nanotube structure on minor changes under preparative conditions suggested that self-assembly was a kinetically controlled process. The resultant polymorphism could critically impact use of these resultant assemblies in biological applications through alteration of the structural interactions at the cell–biomaterial interface.

Taken together, these data results provide convincing evidence that helical filaments and nanotubes can result from self-assembly of appropriately designed peptides that adopt a cross- β conformation. Thus far, it has been challenging to reliably and accurately predict the supramolecular structure of such helical filaments from peptide sequence information, which contrasts with recent advances in prediction of tertiary structure (Baek et al., 2021; Jumper and Hassabis, 2022). The structural information within the Protein Data Bank may not be sufficient at present to provide insight that would enable reliable prediction of atomic structure for most assemblies described here. In combination with the frequently observed polymorphism of β -sheet assemblies, the opportunities for reliable de novo design of synthetic β -sheet

filaments remain relatively limited at present. However, the explosive growth of high-resolution structural data on amyloids over the past two decades provides evidence that supramolecular structural prediction might at some point in the near future become a tractable problem that would enable de novo design of β -sheet assemblies.

Coiled-coil filaments

Similar to β -sheet assemblies, α -helical filaments can be constructed through molecular design approaches based on polar sequence patterns (Beesley and Woolfson, [2019](#)). However, in

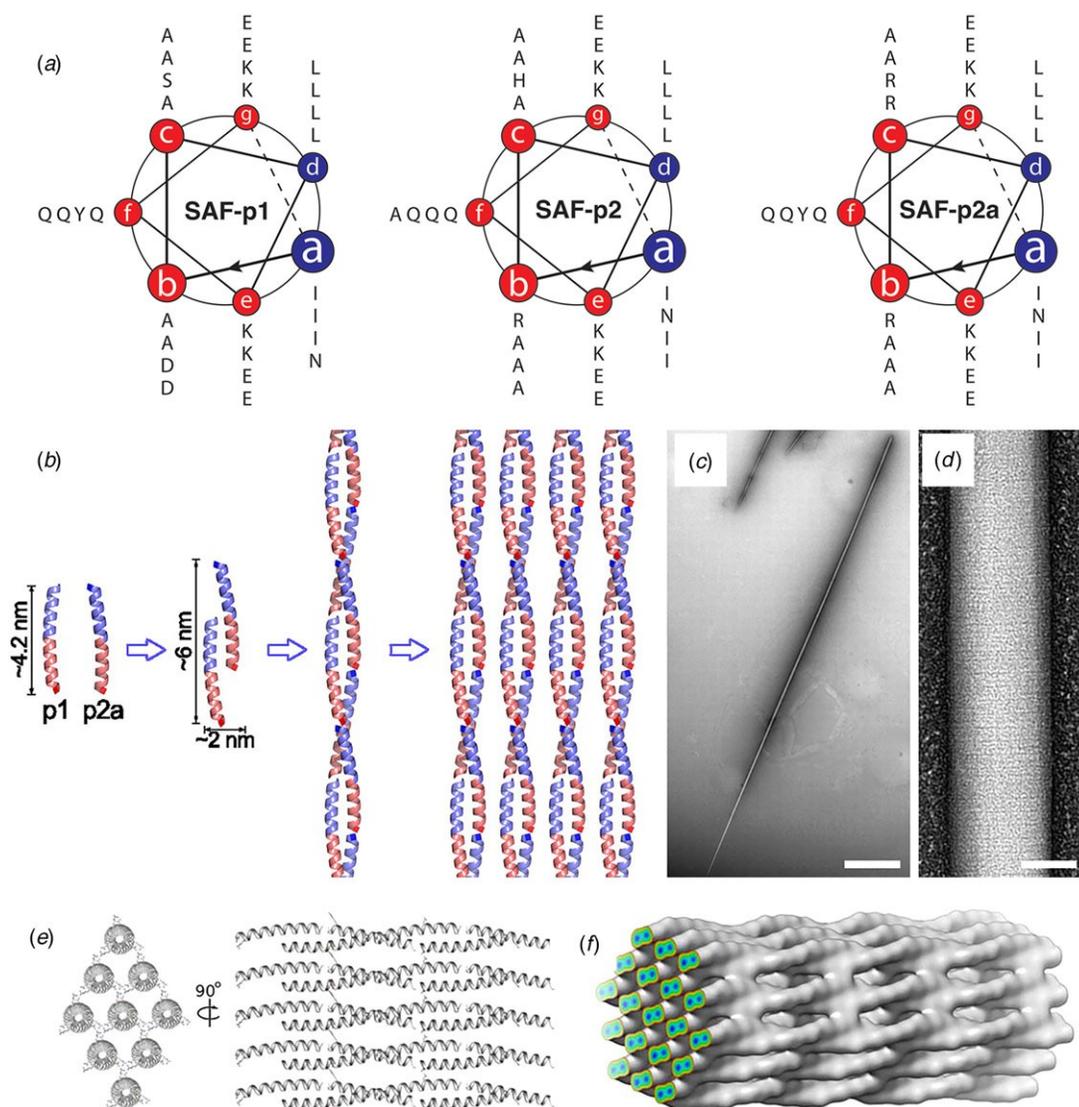


Figure 1.16. (a) Helical-wheel diagrams of the SAF-p1, SAF-p2, and SAF-p2a peptides indicating the heptad periodicity. (b) Proposed model for the heteromeric association of SAF-p1 and SAF-p2 (a) peptides into a dimeric coiled-coil through sticky-ended self-assembly. (c, d) Transmission electron micrographs of the *paracrystalline* assembly of SAF-p1 and SAF-p2a co-assembly. (e, f) Simulated model of the packing of coiled-coil peptides within the SAF assembly. Reprinted from Sharp *et al.* (2012), *Proceedings of the National Academy of Sciences of the United States of America*, 109 (33), 13266–13271.

contrast to β -strands, the hydrogen bonding interactions in α -helices occur within, rather than

between, secondary structure elements. Consequently, α -helical sequences can potentially adopt a stable conformation without self-association, while β -strand formation usually requires self-association between two or more intra or intermolecular peptide segments. Conversely, β -sheet assemblies derived from short peptides are often more stable than α -helical assemblies, as self-association in the former case is primarily driven through directional hydrogen-bonding between β -strands rather than the weaker hydrophobic interactions that hold together α -helical assemblies (Knowles *et al.*, 2007).

The most common mode of self-association between α -helices involves coiling of multiple helices into super-helical architectures known as coiled-coils (Lupas and Bassler, 2017; Woolfson, 2017). Coiled-coil sequences are defined by repeat patterns based on a heptad motif (*a-b-c-d-e-f-g*), in which the residues at the *a* and *d*-positions of the heptad repeat are usually hydrophobic amino acids. Self-association of helices is mediated through ‘knobs-into-holes’ packing interactions, which were first proposed by Crick in 1953 (Crick, 1953). The knobs correspond to the side chains of residues at *a* and *d*-positions, which pack into complementary holes on adjacent helices in the assembly. The pitch of an α -helix corresponds to 3.6 residues, which differs slightly from the 3.5 residues/turn of a coiled-coil. Therefore, the helices in coiled-coil super-coil with a left-handed helical sense to accommodate the ‘knobs-into-holes’ interactions at the hydrophobic interface within the assemblies. In the helical wheel diagram of a coiled-coil, lettered positions are used to describe the architecture of individual α -helices (Figure 1.16a). Sighting down the helix with the amino-terminus closest to the viewer, each residue is represented as a vertex of the helical wheel diagram in super-helix space. This diagram represents the relative orientation of residues that form the interacting faces within a given helix with respect to other helices in the axial projection of the coiled-coil.

The self-association of coiled-coil sequences usually results in discrete oligomeric states displaying rotational, i.e. cyclic (C_n), symmetry. The degree of oligomerization depends on the identity of the core (*a*, *d*) and proximal (*e*, *g*) residues of the heptad repeat sequences. Directed mutagenesis and computational design were employed to identify sequences that form C_n -symmetric homomeric oligomers consisting of two to nine α -helices (Zaccai *et al.*, 2011; Thomson *et al.*, 2014; Rhys *et*

al., 2018; Dawson *et al.*, 2021). In contrast to native coiled-coil filaments, e.g. intermediate filaments, tropomyosin, etc., designed coiled-coils usually form blunt-ended assemblies due to the short length of synthetic peptides, the in-register alignment of helices within the assembly, and the absence of discrete terminal interaction domains.

A convenient strategy to promote the formation of helical filaments from simple helical bundles involves alteration of the helix–helix registry within the assembly to promote staggered self-association between the α -helices (Pandya *et al.*, 2000; Potekhin *et al.*, 2001; Zimenkov *et al.*, 2004). Strand registry can be specified through the pattern of electrostatic interactions between helices, which usually occur between the residues at *e* and *g*-positions on structurally adjacent helices. In addition, buried polar interactions can be introduced at either the *a* or *d*-positions within the sequences of the helical protomers to control helix–helix alignment within the assembly (Oakley and Kim, 1998; Akey *et al.*, 2001).

Woolfson and co-workers employed this approach to fabricate helical assemblies based on a heterodimeric coiled-coil motif (Pandya *et al.*, 2000; Smith *et al.*, 2006). A pair of peptides, SAF-p1 and SAF-p2, was designed such that selective heteromeric association would result in a two-heptad offset between the pair of four-heptad peptides. This staggered orientation was enforced through a combination of complementary electrostatic interactions and a buried polar interaction (Figure 1.16*b*). The formation of high aspect-ratio helical filaments was observed through sticky-ended association. The width of the filaments suggested that the dimeric coiled-coil protofilaments associated into bundles through weak non-specific lateral interactions. To promote a more specific lateral association, a modified peptide sequence, SAF-p2a, was designed to introduce electrostatic interactions between filaments through incorporation of arginine and aspartic acid at surface-exposed positions in the dimeric coiled-coil. Self-assembly resulted in the formation of highly ordered, *paracrystalline* filaments (Figure 1.16*c* and *d*). Cryo-EM analysis of these filaments at intermediate resolution indicated that the dimers packed in a parallel, hexagonal array. The individual protofilaments within the *para*-crystalline assemblies corresponded to pseudo-infinite heterodimeric coiled-coils, which were arranged in parallel along the long axis of the *para*-crystalline assembly (Figure 1.16*e* and *f*) (Papapostolou *et al.*, 2007; Sharp *et al.*,

2012).

A similar approach has been employed for the design of coiled-coil filaments based on homomeric self-association of peptides derived from coiled-coil dimer and trimer sequences (Zimenkov *et al.*, 2004, 2006; Gribbon *et al.*, 2008). However, reliance on a single-peptide sequence significantly constrained design space due to the necessity to incorporate the intermolecular interactions that specify a staggered orientation into single-peptide sequences. In the case of coiled-coil trimers, metal ion binding motifs could be incorporated into the sequence to direct or reinforce helix alignment to trigger assembly or disassembly in response to the presence of specific metal ions (Dublin and Conticello, 2008; Anzini *et al.*, 2013).

An alternative strategy for the fabrication of filamentous coiled-coil assemblies involved the stacking of discrete oligomeric bundles. Crystallographic analyses of coiled-coil structures provided evidence that discrete helical bundles could stack to form continuous superhelices along the highest order rotational axis (Ogihara *et al.*, 1997; Zaccai *et al.*, 2011; Lanci *et al.*, 2012). While these coiled-coil assemblies do not usually persist outside of the crystal, the termini of the peptide sequences could be modified to promote axial interactions that would result in the formation of thermodynamically stable helical filaments. This approach is particularly appealing for larger oligomers since the interfacial surface area buried upon the formation of helical stacks would be sufficient to stabilize the resultant assemblies, especially if coupled with other non-covalent interactions, e.g. electrostatic attraction between oppositely charged termini (Xu *et al.*, 2013; Burgess *et al.*, 2015). An additional advantage of this approach is that larger helical bundles, i.e. oligomerization states greater than a tetramer, encompass a central cavity of sufficient size to encapsulate appropriately shaped small molecules (Thomas *et al.*, 2018).

The latter approach was employed for the design of a helical filament based on a seven-helix bundle structure. Starting with the natural two-stranded coiled-coil GCN4 leucine zipper, Lu and coworkers replaced residues at *e* and *g* with alanine, which generated the modified peptide GCN4-pAA (Liu *et al.*, 2006), which assembled into a heptameric coiled-coil. The seven-helix bundle displayed a successive single-residue offset between structurally adjacent α -helices, which resulted in a discrete

helically symmetric structure that resembled a lock washer (Figure 1.17). The heptameric lock washers remained unassociated in solution and in the corresponding crystal structure of GCN4-pAA (PDB: 2HY6). A positively charged arginine at the C-terminal position served as a gatekeeper residue, which prevented axial stacking through introduction of repulsive interactions between the termini of the seven-helix bundles.

Conticello and coworkers designed a self-assembling peptide nanotube by rationally modifying the sequence of GCN4-pAA (Xu *et al.*, 2013). The C-terminal arginine at the *d*-position was replaced with a leucine, in accord with the canonical sequence preferences for coiled-coils. This substitution restored the continuous hydrophobic interface. In addition, residues at the *b* and *c*-positions of the heptad repeat sequence of GCN4-pAA were replaced with glutamate and lysine, respectively, to introduce electrostatic interactions that reinforced the association between helices in the heptamer. Two arginine residues were added at *f*-positions to prevent lateral association between helical bundles. The peptide termini were uncapped to promote axial stacking of lock washers through electrostatic interactions (Figure 1.17*a* and *b*). Spectroscopic analysis of the corresponding peptide, 7HSAP1, confirmed the presence of α -helical assemblies. TEM measurements indicated that 7HSAP1 spontaneously formed long fibrils under a wide range of monomer concentrations. STEM determined that most fibrils were ~ 31 Å in diameter, which was consistent with the diameter of the heptameric GCN4-pAA bundle observed in the corresponding crystal structure. X-ray fiber diffraction was employed to analyze the structure of the filament, which indicated the presence of stacked helical bundles in the fibril. ssNMR distance measurements on labeled 7HSAP1 peptides provided evidence that the registry shift between helices was retained in the filament. The solvatochromic fluorophore 6-propionyl-2- (*N,N*-dimethylamino)-naphthalene (PRODAN) was used to probe the accessibility of the lumen of the 7HSAP1 nanotube toward guest molecules. A dose-dependent shift in the PRODAN fluorescent emission was observed that was consistent with encapsulation within the hydrophobic lumen of the 7HSAP1 nanotube.

Woolfson and coworkers significantly advanced this approach by extending the scope of the process to a range of different blunt-ended coiled-coil oligomers (Burgess *et al.*, 2015). The self-assembly of helical filaments was observed for coiled-coils having oligomeric states from three to seven, i.e. trimers through heptamers (Figure 1.18*a* and *b*). TEM measurements indicated that the extent of lateral association was variable between the different peptide systems. However, filaments derived from a hexameric coiled-coil, CC-Hex_T (Figure 1.18*c*), were observed to form highly ordered, *para*-crystalline

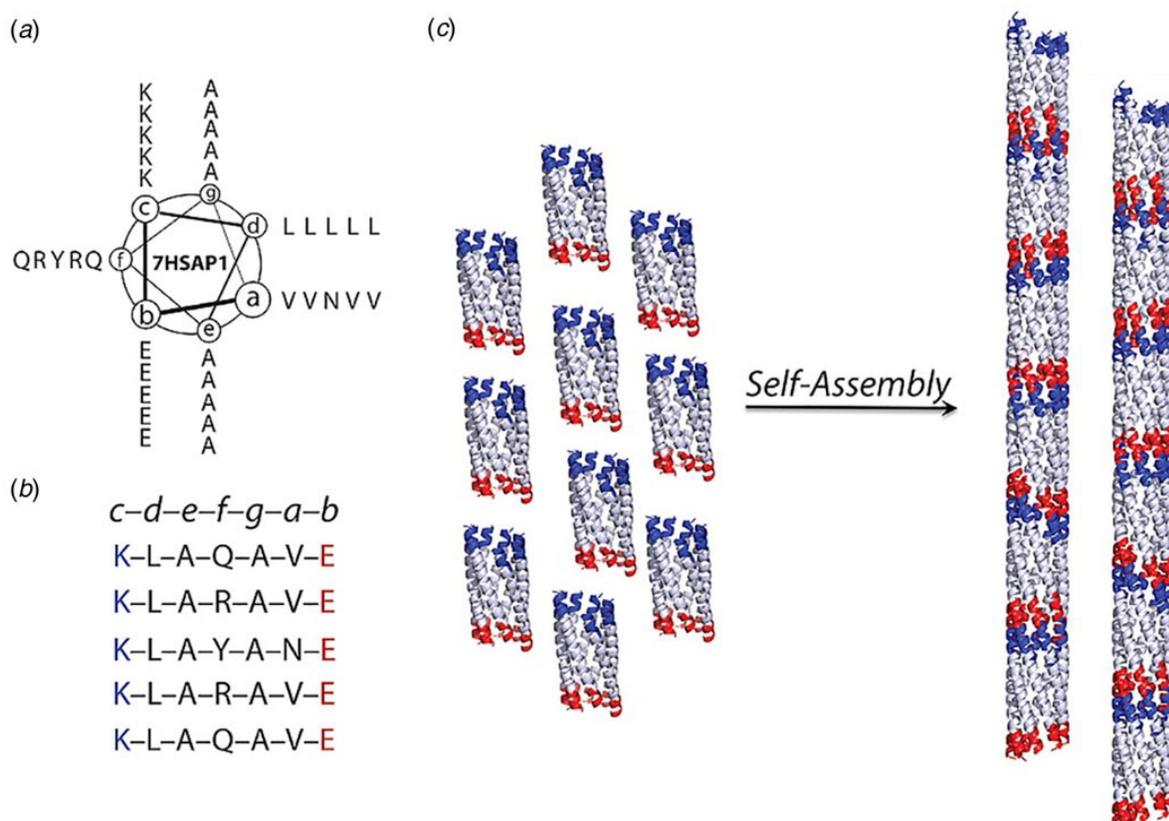


Figure 1.17 Mechanism of self-assembly of coiled-coil lock washers into helical filaments. (a) Helical wheel diagram of designed coil-coil peptide 7HSAP1. (b) Sequence of the 7HSAP1 peptide indicating the heptad registry. Basic and acidic residues involved in the intra-bundle and inter-bundle (axial stacking) interactions are highlighted in blue and red, respectively. (c) Proposed mechanism of self-assembly of heptameric lock washers of 7HSAP1 into filaments through electrostatic interactions between oppositely charged termini. The seven-helix bundle structure was derived from the crystal structure of GCN4-pAA (PDB: 2HY6). Reproduced with permission from Xu *et al.* (2013), *Journal of the American Chemical Society*, 135 (41), 15565–15578. Copyright 2021 American Chemical Society.

assemblies (Figure 1.18*d* and *e*). X-ray fiber diffraction and cryo-EM analysis provided evidence that the hexameric barrels stack axially in the filament and were arranged in a 2D lattice based on a tetragonal unit cell (Figure 1.18*f*). The high degree of internal order within the assembly was rationalized based on the observation that the super-helical pitch of the hexameric coiled-coil corresponded to an integral number of six stacked assemblies, which would result in the formation of a continuous superhelical array along the contour length of the protofilaments. The ordered presentation of function groups at the periphery of the protofilaments could potentially reinforce lateral association into a *para*-crystalline assembly. Fluorescence binding assays demonstrated that the linear hydrophobic dye 1,6-diphenylhexatriene could bind within the lumen of filamentous nanotubes derived from the larger barrels ($n = 5-7$), which suggested that the internal channel of the assemblies was accessible to appropriately shaped substrates.

The coiled-coil assemblies described thus far display the classical architecture in which the individual helices were aligned nearly parallel to the superhelical axis (Lupas and Bassler, 2017; Woolfson, 2017). Tayeb-Fligelman *et al.* recently reported the first example of a cross- α fibril structure from crystallographic analysis of PSM α 3, a short cytolytic peptide secreted from virulent strains of *Staphylococcus aureus* (Tayeb-Fligelman *et al.*, 2017). The helices were observed to pack in a stacked bilayer array, in which individual peptides were oriented perpendicular to the long axis of the crystallographically defined filament. Subsequently, Zhang *et al.* reported a series of peptides based on designed coiled-coil sequences that assembled into filaments (Zhang *et al.*, 2018). Crystallographic analyses demonstrated that the peptides formed cross- α super-helical arrays through stacking of parallel dimers in an alternating antiparallel orientation. However, these filaments displayed limited interfaces over which KIH packing was observed between protomers.

Egelman *et al.* employed helical reconstruction from cryo-EM images to characterize two cross- α filaments based on coiled-coil peptide sequences (Egelman *et al.*, 2015). Two 29-residue peptide sequences were designed based on a type 3 coiled-coil architecture, first proposed by Walshaw and Woolfson (2001), from analysis of the coiled-coil region of the bacterial protein tolC and subsequently observed in bacteriophage transit tubes (Koronakis *et al.*, 2000; Sun *et al.*, 2014). To decrease the degree of

curvature and create a larger nanotube, residues at the interface between protomers (*c* and *d*-positions,

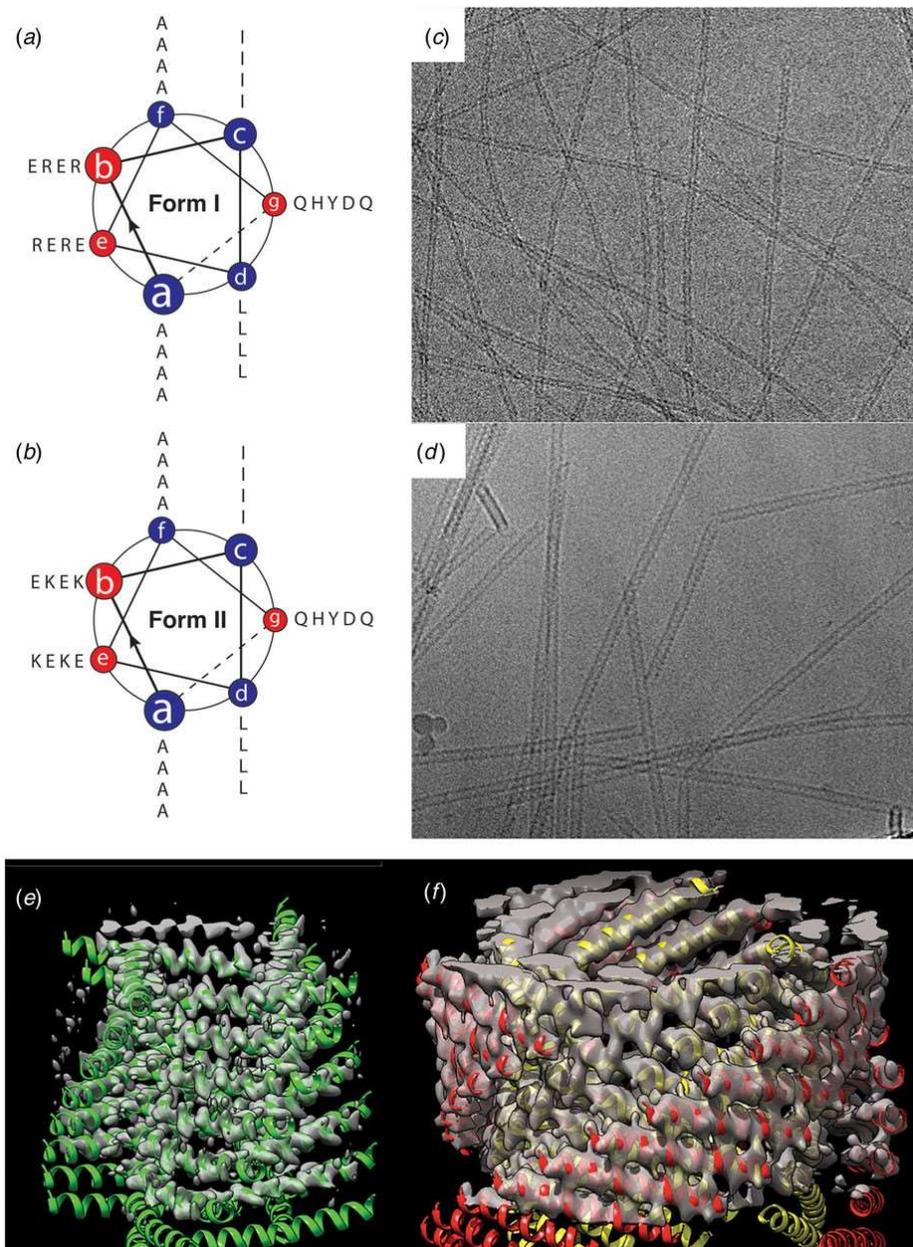


Figure 1.19 *a* and *b*) were replaced with larger isoleucine and leucine, and residues on the outside (*a* and *f*-positions) were exchanged for smaller alanine. Charged residues were substituted at the *b* and *e*-positions to increase solubility and direct heterotypic facial interaction between adjacent helices in a parallel orientation. The resulting peptides, Form I and Form II, differed only in the presence of arginine *versus* lysine at identical positions within the respective sequences.

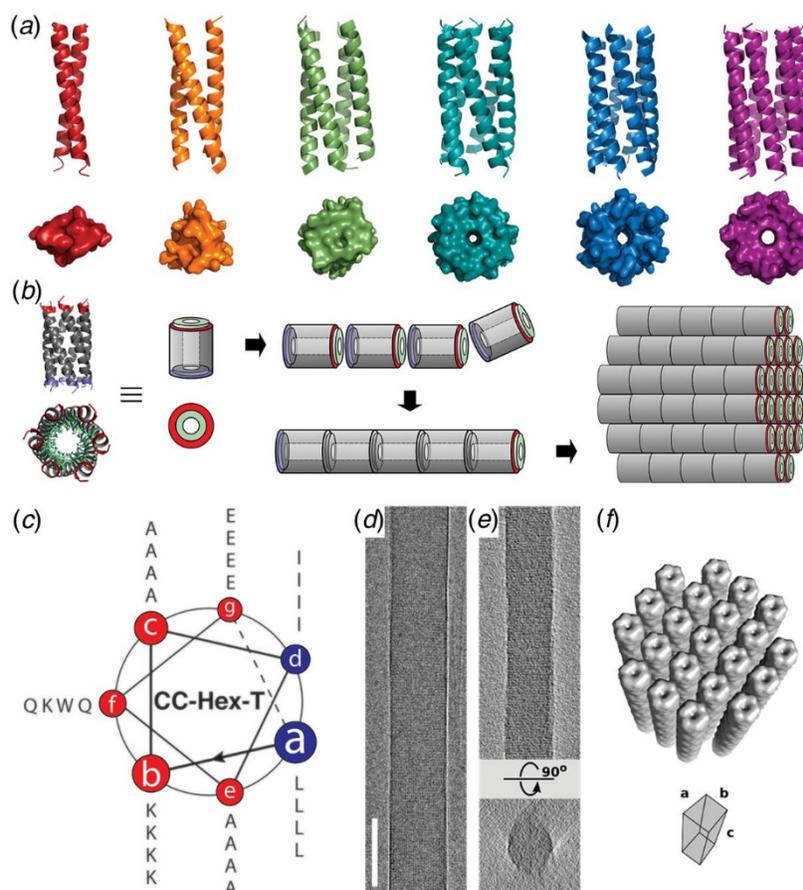


Figure 1.18. Blunt-ended coiled-coil oligomers self-assemble into peptide nanotubes. (a) Ribbon diagrams and orthogonal views of space-filling models for designed coiled-coil oligomers; CC-Di (red; PDB: 4DZM), CC-Tri (orange; 4DZL), CC-Tet (green; 3R4A), CC-Pent (turquoise; 4PN8), CC-Hex (blue; 3R3K), and CC-Hept (purple; 4PNA). (b) Proposed model for self-assembly of CC-Hex-T involving axial stacking of hexameric barrels into a filament. (c) Helical wheel diagram depicting the heptad register of the CC-Hex-T sequence. (d) Representative cryo-EM image of thermally annealed CC-Hex-T assemblies (scale bar = 100 nm). (e) Tomographic slice of a CC-Hex-T assembly from cryo-ET analysis depicting an approximately cylindrical cross section. (f) Tetragonal lattice model for packing of CC-Hex-T filaments in a *para*-crystalline array. Reprinted with permission from Burgess *et al.* (2015), Journal of the American Chemical Society, 137 (33), 10554–10562. Copyright 2021 American Chemical Society.

The Form I and Form II peptides assembled into nanotubes from buffered aqueous solution at ambient temperature. CD spectropolarimetry indicated the presence of an α -helical conformation, which was consistent with the formation of a coiled-coil assembly. However, conventional TEM and SAXS measurements provided evidence that the diameter of the Form II filaments (~ 120 Å) was double the diameter of the Form I filaments (~ 60 Å). These results suggested that the two structures must be different despite the similarity between the two peptide sequences.

Cryo-EM analysis with direct electron detection was employed to generate 2D projection images for helical reconstruction of the thin filaments (

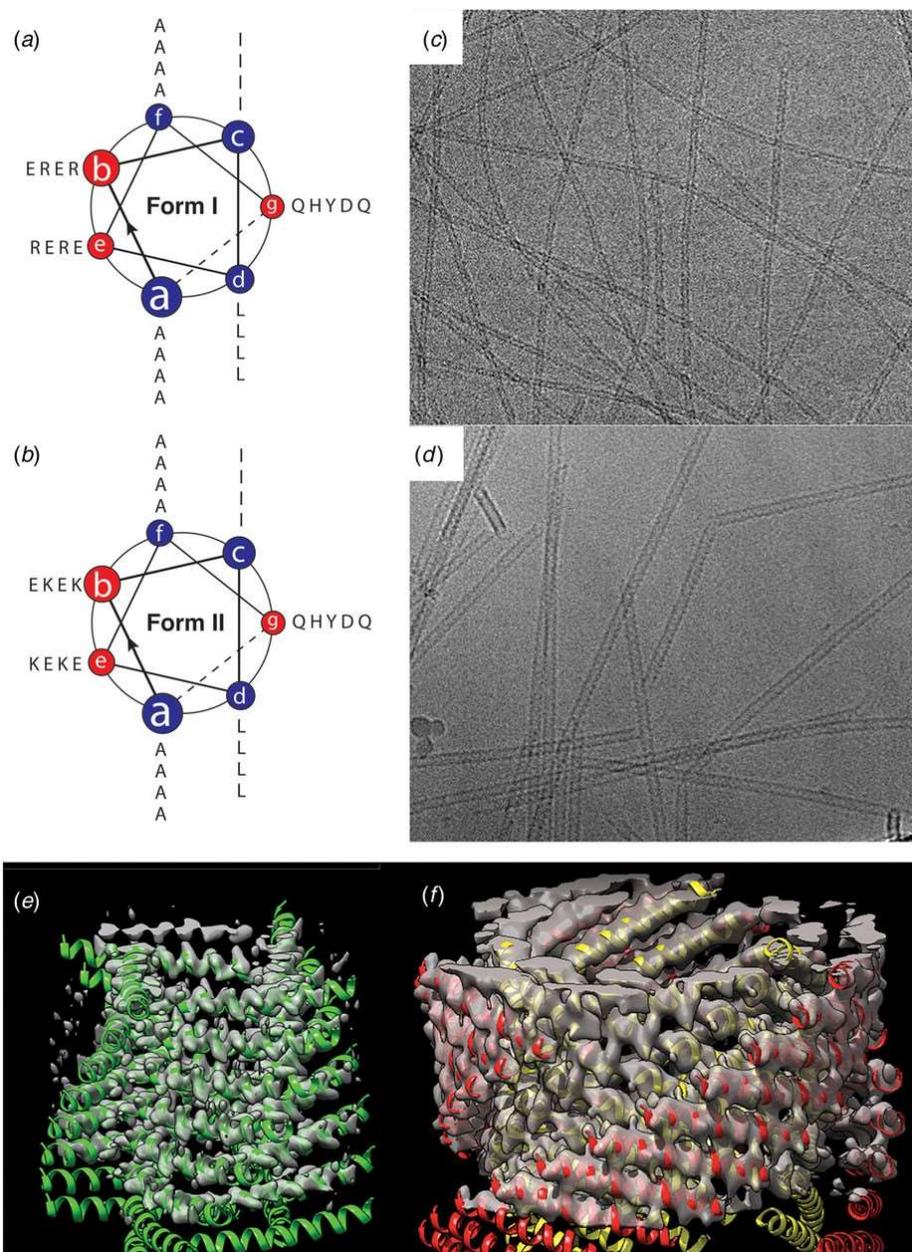


Figure 1.19c and d).

Helical reconstruction afforded distinct atomic models for the Form I and Form II filaments (

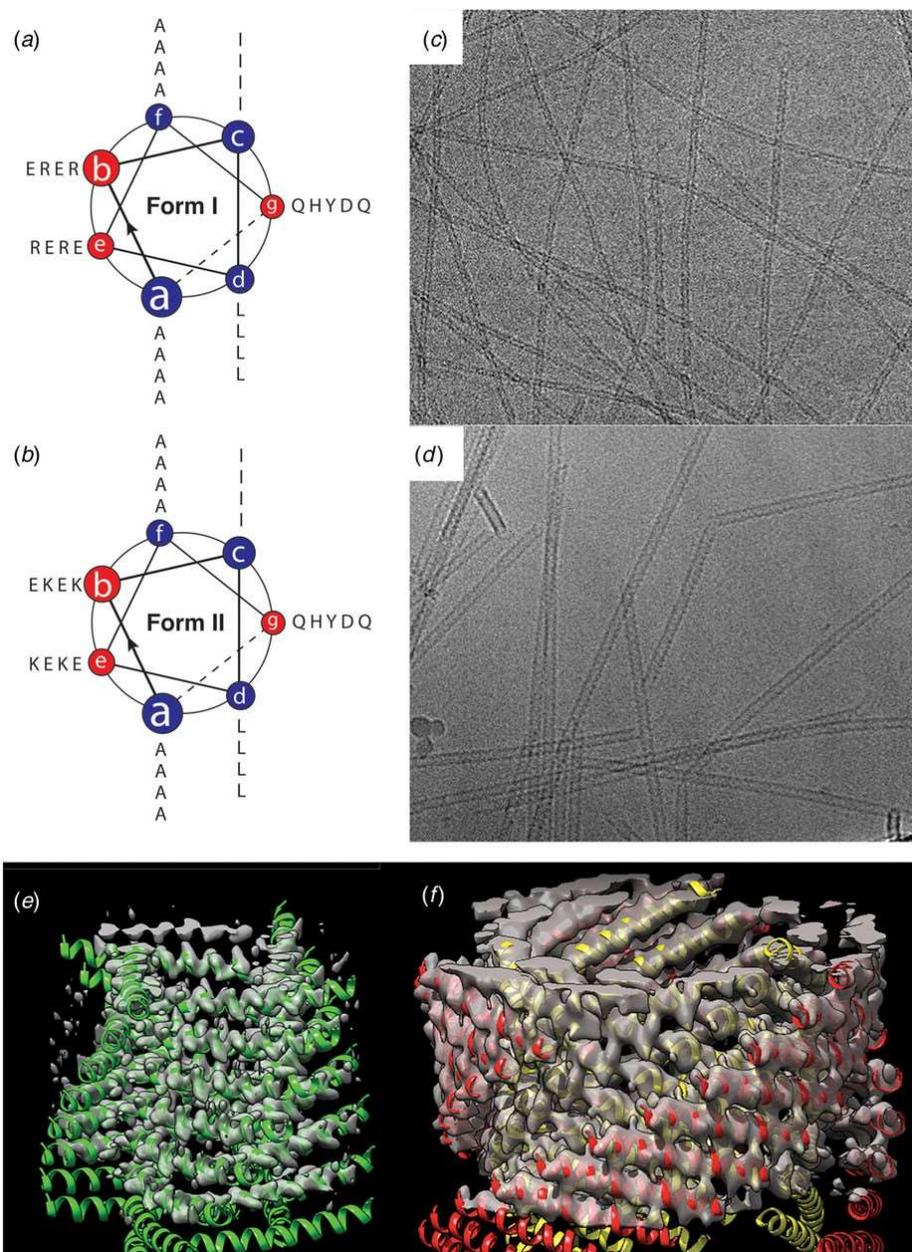
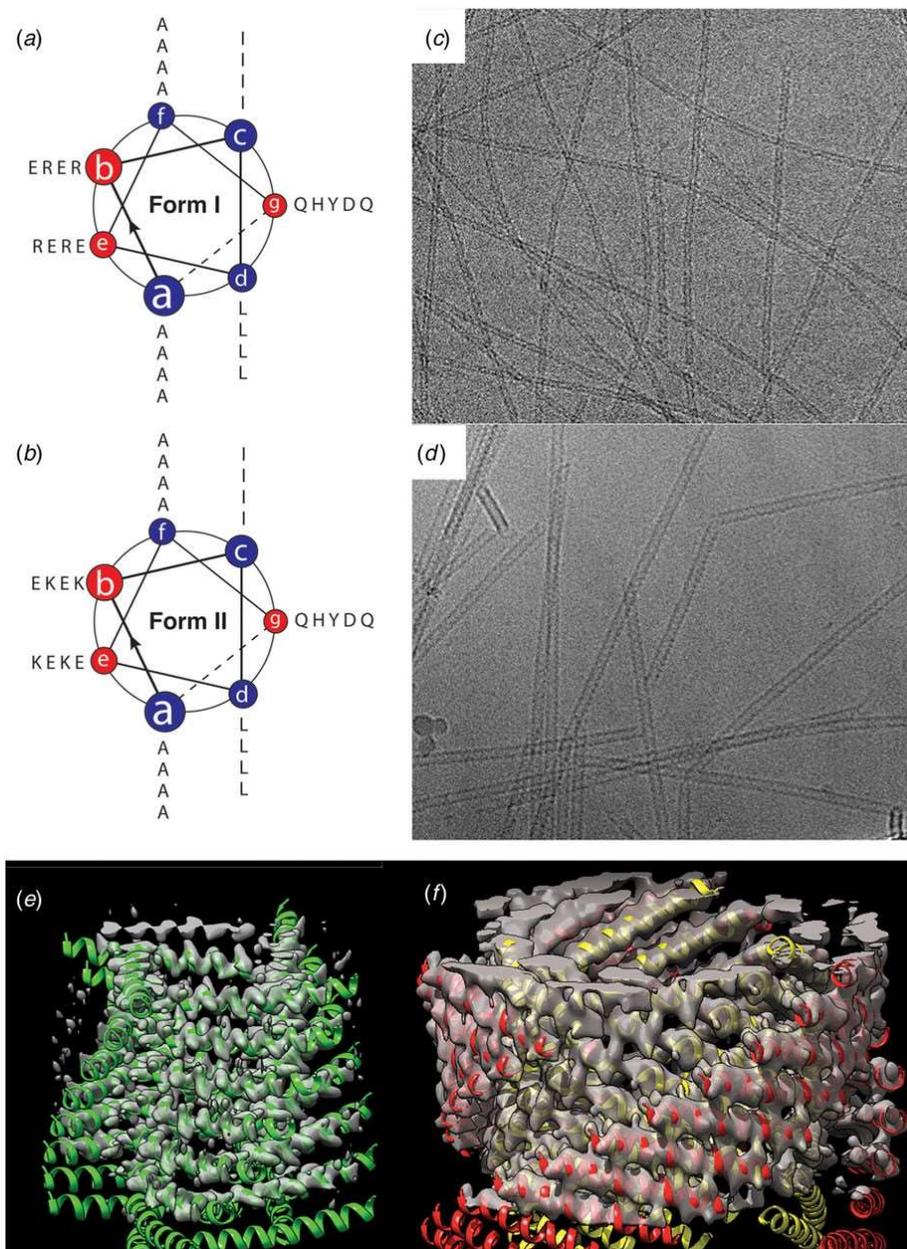


Figure 1.19*e* and *f*). This work represented an early success for singleparticle cryo-EM structural analysis of designed peptide filaments. While similar in sequence and identical in apparent secondary structure, Forms I and II assembled into



Figure

1.19. (a, b) Helical wheel diagrams indicating the heptad register of the Form I and Form II peptide sequences. (c, d) Representative cryo-EM images of the Form I (c) and Form II (d) nanotubes. (e, f) Atomic models fit into the 3D reconstructions of the Form I (e) and the Form II (f) nanotubes derived from cryo-EM analysis. Reprinted with permission from Egelman *et al.* (2015), *Structure*, 23 (2), 280–289.

different quaternary structures. The structure of the Form I nanotubes (PDB: 3J89) was solved at 3.6 Å resolution and was based on a left-handed, 1-start helix (~4.1 subunits per turn). The asymmetric unit corresponded to a single α -helix, which resulted in formation of a single-walled nanotube with a helical pitch of ca. 9.0 Å and a helical twist of ca. -88° . The Form I nanotube has a nearly square cross section in which the subunits were arranged into four protofilaments that coincided with the right-handed 4-start helices of the nanotube.

In contrast, the initial reconstruction of the Form II nanotube afforded a density map at relatively low resolution (~7 Å), but clearly displayed an asymmetric unit based on a pair of α -helices that resulted in a double-walled nanotube. Subsequent cryo-EM data collection on Form II filaments resulted in a higher-resolution (~4.2 Å) reconstruction, which resulted in a more reliable atomic model (PDB: 6WL8). The structure was based on a right-handed 1-start helix with a helical pitch of

5.6 Å and a helical twist of 124.4° . The individual protofilaments within the nanotube corresponded to the right-handed 3-start helices. The reconstruction had sufficient resolution to determine that the pair of helices in the asymmetric unit was oriented in parallel across the inner-outer interface of the bilayer.

The primary difference between the Form I and Form II assemblies resided in the nature of the cohesive interactions between protofilaments that correspond to the cross- α helical stacks. The stacking interactions along a protofilament were quite similar between the two assemblies and corresponded to the KIH interactions that are typically observed for coiled-coil motifs. In contrast, the interactions between protofilaments involved side-to-end helical association and end-to-end association for Form I and Form II, respectively. The Form I interactions involved a pair of arginine residues, R13 and R17, that capped the C-terminus of a helix on an adjacent protofilament through a network of hydrogen-bonding interactions. This unique

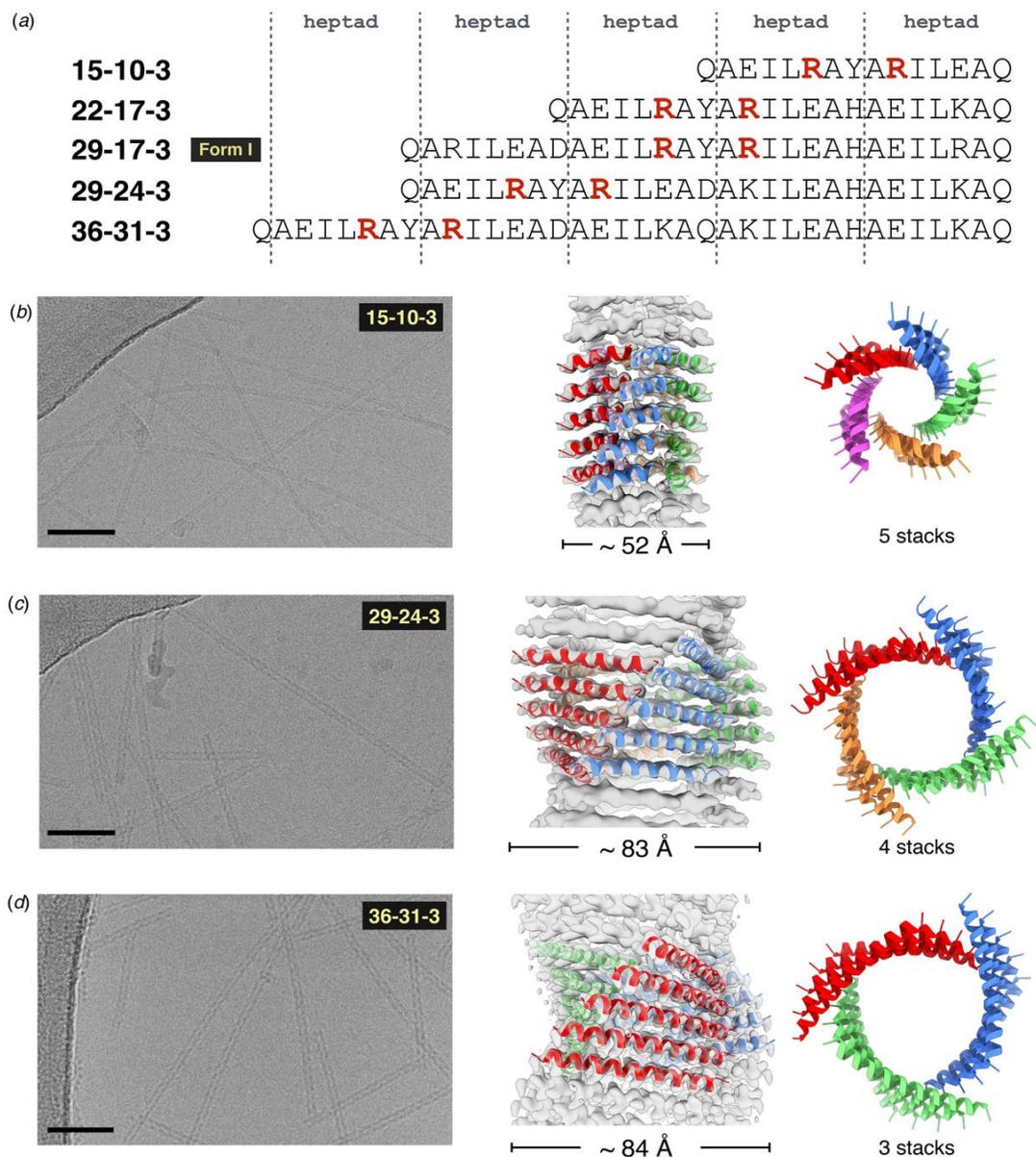


Figure 1.20. (a) Peptide sequences of Form I variants in which the position of the arginine clasp motif is highlighted in red. Structural analyses of the 15-10-3 (b), the 29-24-3 (c), and the 36-31-3 (d) filaments. Representative cryo-EM images are shown on the left (scale bar = 50 nm). Atomic models fit into the respective 3D reconstructions are in the middle. Top views of the respective assemblies are shown on the right. Reproduced under the Creative Commons License from Wang *et al.* (2021a). *Nature Communications*, 12 (1), 407.

interaction, which was designated as an arginine clasp, does not appear to have been evolutionarily

sampled as an interaction motif between helices in native protein structures (Wang *et al.*, 2021a). Lateral association between protofilaments in the Form II structure was based on a much weaker hydrogen bonding between Q29 and Q1 on structurally adjacent helices at the concave (inner) interface. The two structures can be interconverted through limited mutagenesis at these structurally critical positions within the respective peptide sequences (Egelman *et al.*, 2015).

The arginine clasp motif defined the number of protofilaments in the Form I structure and consequently its diameter. A series of peptides based on the Form I sequence was synthesized in which the length of the peptide was varied from two to five heptad repeats (Figure 1.20a) (Wang *et al.*, 2021a). Each sequence preserved a single arginine clasp motif within the most N-terminal heptad. This peptide series was observed to self-assemble into high aspect-ratio filaments. Cryo-EM imaging (Figure 1.20b–d) was employed to determine the structures of the two-heptad (15-10-3), four-heptad (29-24-3), and five-heptad (36-31-3) peptides at resolutions 4.2, 4.1, and 4.0 Å, respectively (PDB: 6WKX, 6WKY, and 6WL0). In combination with the Form I structure, these filaments represent a homologous series in which the arginine clasp interaction is conserved between laterally associating protofilaments within the respective structures. However, the number of protofilaments within the respective structures depended on the length of the peptide and decreased with an increasing number of heptad repeats (Figure 1.20).

Finally, alternative strategies can be conceived for the fabrication of nanotubes from self-association of discrete helical bundles. Zhang *et al.* reported the computational design of antiparallel four-helix bundles that could self-associate into crystalline 2D assemblies of defined layer symmetry through optimization of protein–protein interactions at the interfaces between the tetrameric protomers (Zhang *et al.*, 2016). However, in lower pH buffers (10 mM sodium acetate, pH 4.5), one of these peptides assembled into nanotubes of defined diameter (Tian *et al.*, 2018). Multiple analytical measurements (TEM, SAXS, AFM, cryo-EM, and STEM) supported the formation of apparently homogeneous nanotubes, although the nature of the cohesive interactions between tetramers remains speculative in the absence of high-resolution structural information. The tubes were observed to slowly convert into sheets

over extended periods of time. The sheet form was dominant at neutral pH, which suggested that the tubes represented a kinetic product. A similar pH-dependent transition between sheets and scrolled tubes was reported for a series of designed collagen-mimetic peptides (Merg *et al.*, 2020). These data suggested that preformed supercoiled peptide motifs could serve as protomers for the creation of shape-shifting nanomaterials that could toggle between sheetlike and tubular forms. Further structural characterization at high-resolution would enable a more detailed understanding regarding the supramolecular structural parameters that underlie these morphological transitions.

Tandem repeat assemblies

Tandem repeat proteins (TRPs) consist of concatenated sequences of relatively short secondary structure motifs, e.g. α -helical hairpins, that form extended folded structures displaying helical symmetry (Kobe and Kajava, 2000; Kajava, 2012). Within a given TRP class, sequence features that mediate interactions between the repeats are often highly conserved, while non-structurally integral amino acid positions are usually hypervariable. These protein families represent attractive substrates for the design of synthetic helical assemblies, especially for roles in substrate recognition and binding.

One of the most studied classes of TRPs is derived from the tetratricopeptide repeat (TPR) motifs, which are naturally occurring TRPs based on a helix–turn–helix structural motif. Regan and co-workers identified the consensus repeat sequence for a 34 amino acid TPR motif (Main *et al.*, 2003). Synthetic concatemers based on different lengths of this consensus TPR (CTPR) were prepared and structurally characterized. Single-crystal X-ray diffraction analysis of CTPR8 indicated that it crystallized in three different unit cells corresponding to different space groups with distinct crystallographic symmetry elements (PDB: 2AVP, 2FO7, and 2HYZ) (Kajander *et al.*, 2007). In each case, the CTPR units stacked in a polar, head-to-tail fashion to form a continuous right-handed superhelix. The main-chain atoms within the respective CTPR subunits of the three crystal structures were superimposable. The right-handed superhelix consisted of eight TPR motifs per helical turn and therefore a CTPR8 polypeptide corresponded to a single turn of the superhelix. The crystal structure of the longer

CTPR20 indicated that it adopted an identical helical structure with respect to individual superhelical filaments within stacked unit cells along the central axis of the assembly. SAXS measurements of CTPR8 in aqueous buffers did not provide evidence for the formation of persistent filamentous structures in solution, although the local helical conformation was conserved. However, Grove and coworkers demonstrated that a concatemer of different length, CTPR18, could spontaneously self-associate into oriented thin solid films that display multi-scale order and stimulus responsive behavior (Carter and Grove, 2015, 2018).

In an attempt to create synthetic helical filaments based on minimal-length tandem repeat motifs, Conticello and coworkers (Hughes *et al.*, 2019) designed two peptides based on sequences derived from a leucine-rich repeat variant (LRV) (Peters *et al.*, 1996) and a thermostable HEAT repeat (PBS_HEAT) (Urvoas *et al.*, 2010) sequences, respectively. While single-repeat motifs have been postulated to be thermodynamically unstable toward self-association at ambient temperature, the resulting sequences, LRV_M3 Δ 1 and HEAT_R1, assembled into high aspect-ratio helical filaments in aqueous buffers over a wide pH range (Figure 1.21). Synchrotron SAXS measurements on solutions of the respective peptides confirmed the presence of narrow diameter filamentous assemblies. Cryo-EM was employed to investigate the structures of the two filaments derived from these structurally related tandem repeat motifs.

Structural analysis of the LRV_M3 Δ 1 filaments (PDB: 6HQE) indicated that it formed a left-handed, 1-start helix (Figure 1.21a–c). The helical hand of the LRV filament differed from the native right-handed helical twist of a crystallographically characterized LRV repeat concatemer (PDB: 1LRV). In addition, while the asymmetric unit corresponded to a single-repeat motif, the repeat units adopted an α -helix–loop– α -helix conformation, which contrasted with the α -helix–loop– 3_{10} -helix conformation of the tandem repeat motifs in the crystal structure of a native LRV concatemer. The lateral interaction between adjacent LRV repeat motifs stabilized the assembly along the direction of the 1-start helix of the assembly and provided the primary driving force for polymerization. The formation of the closed nanotube involved the interaction between arginine residues on protomers at the interface between

adjacent turns of the helical assembly (Figure 1.21*b*). Notably, this axial interaction buried about twice as much surface area per subunit as the HEAT_R1 filaments (*vide infra*).

In contrast, the structure of the HEAT_R1 filament (PDB: 6MK1) displayed right-handed 1-start helical symmetry (Figure 1.21*d-f*) in which the periodicity was similar to the superhelical structure observed within the crystal structure of a concatemer based on a consensus PBS_HEAT repeat (PDB: 3LTJ). However, the arrangement of protomers in the HEAT_R1 filament differed from the crystal structure of the soluble concatemers. The asymmetric unit of the HEAT_R1 filament was based on non-covalent association between two peptide motifs (Figure 1.21*e*). One helical hairpin in the asymmetric unit could be aligned with repeat units derived from PBS_HEAT motifs in the crystal structure of the synthetic concatemer. The structure of the other peptide was distinct from that of the canonical PBS_HEAT repeats due to the need to accommodate a Trp–Trp interaction between the two peptides in the asymmetric unit. PISA (proteins, interfaces, surfaces, and assemblies) (Krissinel and Henrick, 2007) analysis showed that the buried surface area of the lateral interface was significantly larger than the corresponding value for axial interface. This difference arises because of the structural distortions induced by the Trp–Trp interaction between the two peptides in the asymmetric unit. TEM imaging indicated that the HEAT_R1 filaments would frequently unwind on the grid, presumably because of the weak axial interaction between adjacent helical turns of the assembly. While these results confirmed that stable helical filaments could be constructed from self-assembly of tandem repeat sequences, the structural analysis also highlighted the plasticity of helical symmetry within supramolecular assemblies. Slight differences in local interactions and peptide conformation led to differences in peptide packing at the lateral and axial interfaces within a filament, which altered the helical symmetry of the assembly vis-à-vis the corresponding soluble oligomers.

Shen *et al.* reported successful *de novo* design of helical filaments (DHF) derived from

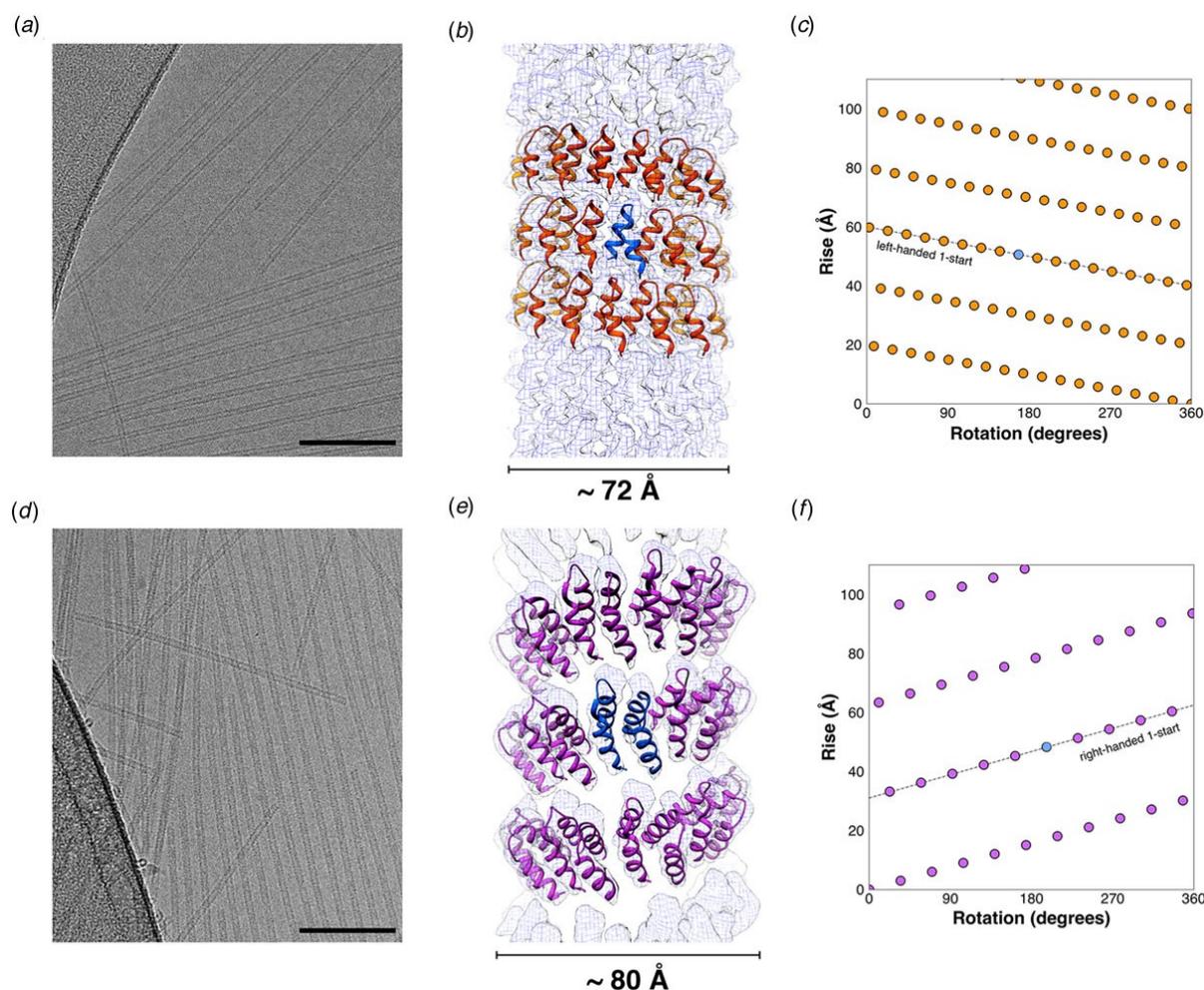


Figure 1.21. Single tandem repeat motifs self-assemble into stable nanotubes. (a, d) Representative cryo-EM images of LRV_M3Δ1 (a) and HEAT_R1 (d) filaments. (Scale bar = 100 nm.) (b, e) Atomic models fit into the 3D reconstructions of the LRV_M3Δ1 (b) and HEAT_R1 (e) nanotubes. The asymmetric units are highlighted for the respective filaments in blue, which, in the case of HEAT_R1, corresponds to a dimer of peptides. (c, f) Helical net diagrams for the LRV_M3Δ1 (c) and HEAT_R1 (f) nanotubes, in which the difference in helical hand between the respective 1-start helices is apparent. The helical nets show the unrolled surface lattice viewed from the outside of the filament. From Hughes *et al.* (2019), *Proceedings of the National Academy of Sciences of the United States of America*, 116 (29), 14456–14464.

computationally optimized synthetic tandem repeat oligomers based on helix–turn–helix motifs (Shen *et al.*, 2018). The design algorithm screened for self-assembling proteins by selecting an arbitrary asymmetric protomer and randomly sliding an identical protomer into contact. The screen employed 15 structurally distinct, computationally designed folded protomers, which resulted in 124 target sequences that were expressed in *Escherichia coli*. Filamentous nanomaterials were observed from expression of 34 of these synthetic constructs. Cryo-EM was employed to analyze the structure of the six most promising filaments. The helical reconstructions ranged in resolution from 3.4 to 7.8 Å (Figure 1.22). The geometry of the experimentally determined structural interfaces matched well to the computational designs for four of the six designed filaments, while the overall supramolecular architectures were retained for all of the designs. Despite these observations, even the most accurate computational predictions cannot recapture helical symmetry precisely since slight differences in interfacial interactions may lead to changes in helical symmetry (*vide supra*). The latter represents a known but unappreciated phenomenon in the design of helical peptide and protein assemblies (Egelman *et al.*, 2015; Lu *et al.*, 2015). Nevertheless, this observation does not diminish the impact of these results, which represent the most reliable and accurate predictive design of helical filaments that has been reported thus far.

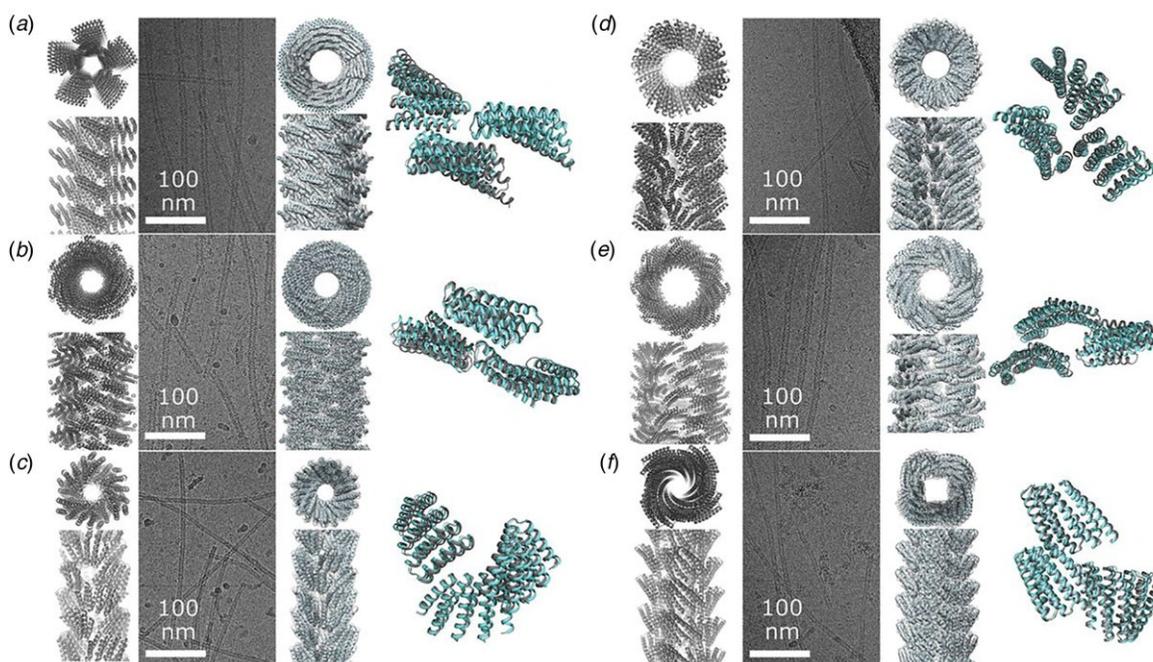


Figure 1.22. Computational design of helical filaments from designed TRPs. Left to right: Computational models, representative cryo-EM images, cryo-EM-derived atomic models, and structural overlays between the computational model and experimental structure for designed helical filaments DHF58 (a), DHF119 (b), DHF91 (c), DHF46 (d), DHF79 (e), and DHF38 (f). From Shen *et al.* (2018), *Science*, 362 (6415), 705–709. Reprinted with permission from AAAS.

The stability and persistent tertiary structures of the synthetic protomers enabled rational control of filament structure as well as the dynamics of filament assembly. For one designed helical filament, DHF58 (PDB: 6E9T), the number of repeats in the concatemer was employed to control filament width, while preserving the designed inter-protomer interfaces (Figure 1.23). Kinetic analysis of another designed filament, DHF119 (PDB: 6E9Z), indicated that the self-assembly process was concentration-dependent and reversible below a ceiling concentration. Filament disassembly resulted upon dilution below this critical concentration. Synthetic capping proteins were designed in which one interface in the protomer was prevented from self-association. Addition of these capping agents prevented further addition at the blocked terminus, but also initiated a concentration-dependent disassembly of the filaments due to the dynamically reversible association/dissociation of kinetically competent protomers.

Green fluorescent protein-tagged monomers of DHF58 could assemble into filaments in live and phagelysed *E. coli* expression strains, which suggested that these assemblies could be employed for *in cellulo* applications.

While TRPs based on α -helical motifs have been more extensively examined for the fabrication of synthetic peptide and protein filaments, β -sheet tandem repeat motifs commonly occur as structural elements in native proteins as well (Kajava and Steven, 2006). These β -solenoid or β -helix motifs consist of two, three, or four β -strands that are linked through short turn or β -arc segments. The individual repeat motifs often occur as extended concatemers within native proteins, which form the turns of a parallel cross- β helix. Stacking of the repeats is mediated through the formation of hydrogen-bonded β -sheet along the helical axis of the structure and is reinforced through hydrophobic interactions that occur within the core of the helix. In contrast to cross- β filaments such as amyloids, uncontrolled polymerization in native β -solenoid proteins is prevented through the presence of N and C-terminal globular domains that prevent end-to-end association.

Toney and co-workers (Peralta *et al.*, 2015; Peng *et al.*, 2017b, 2020), have employed native β -solenoid domains as the starting point for the design of synthetic β -helical filaments. Manual and computational designs were used separately or in combination to create discrete-length concatemeric sequences based on two-strand and three-strand solenoid repeats. The proteins were obtained from bacterial expression from synthetic genes. To promote assembly of extended filaments, the corresponding proteins lacked N and C-terminal capping groups. Remarkably, despite their strong aggregation potential, the β -helical proteins could be prepared and purified in reasonable yield (30–40 mg l⁻¹ of culture) using recombinant protein expression in an *E. coli* bacterial host.

Protein solutions were meta-stable in the unassociated state but could be driven to self-assemble through incubation at 37 °C. TEM and AFM analysis of the assembled proteins indicated the presence of filaments of similar diameter to the parent β -helices. The CD spectra were consistent with the presence of a high content of β -sheet conformation. The resulting filaments displayed remarkable resistance to elevated temperature, organic solvents, chemical denaturants, and were stable over a wide pH range. In

addition, nanomechanical measurements and molecular dynamics simulations provided evidence that mechanical properties of the synthetic filaments, e.g. ultimate tensile strength and Young's modulus, were similar to the corresponding values for other β -sheet materials including amyloids and spider silk

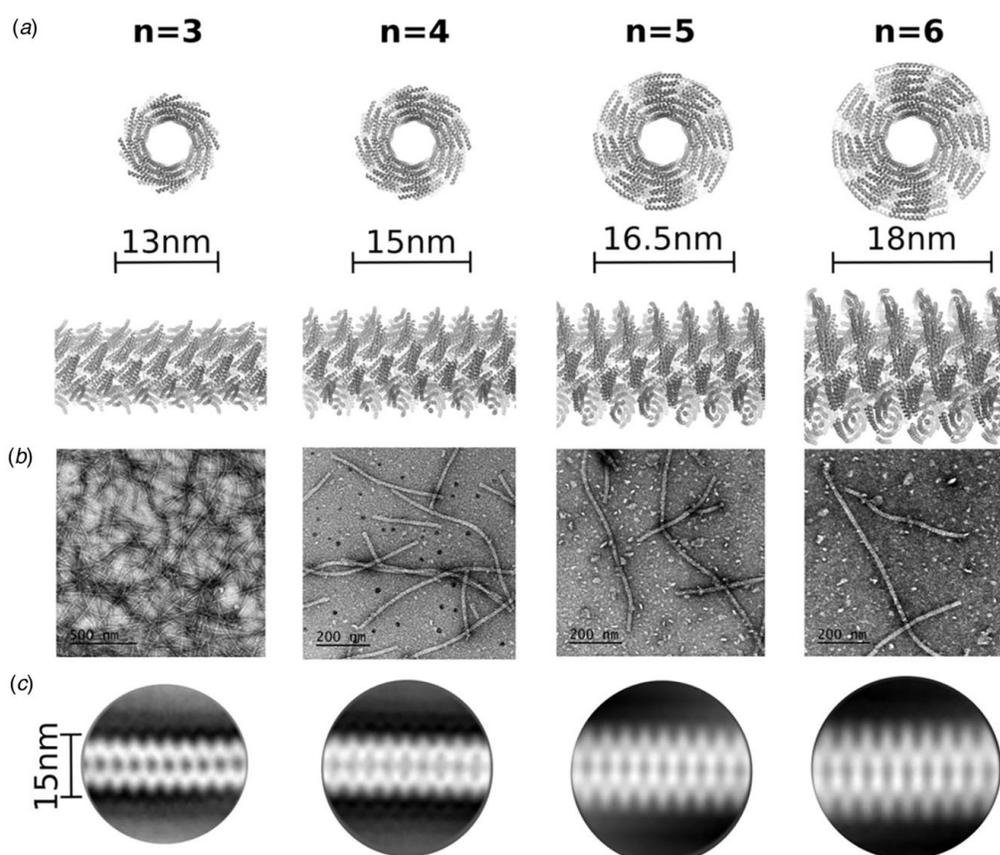


Figure 1.23. Diameter of designed helical filament DHF58 can be controlled through the number of tandem repeats in the protomer. (a) Cross sections and side views of computational models based on the four-repeat cryo-EM structure. The number of repeats (n) is shown at the top. (b) Representative negative stain electron micrographs of the corresponding helical filaments. (c) 2D-class averages derived from electron micrographs of the respective filaments. From Shen *et al.* (2018), *Science*, 362 (6415), 705–709. Reprinted with permission from AAAS.

(Peng *et al.*, 2017a). These results suggested that β -solenoid assemblies present attractive targets for rational design of novel nanomaterials.

Most of the designed filaments described thus far have been derived from relatively simple structural motifs. As protein engineering and computational design efforts advance, more structurally complex subunits may be considered as substrates for the design of synthetic protein filaments. Biological systems often employ globular protein domains as building blocks for the assembly of filaments including F-actin, microtubules, etc. Recent studies suggested that filamentation may represent a mechanism to regulate protein function, including enzymatic catalysis (Lynch *et al.*, 2017, 2020).

In this vein, Kaltofen *et al.* (2015) reported the computational design of a small self-assembling $\beta\alpha\beta$ domain (Figure 1.24). While technically neither a globular domain nor a tandem repeat motif, this small modular protein fold was constructed from a pair of β -strands that were connected via turns to an α -helix (Figure 1.24a). The sequence of the β -strand motifs was derived from a self-associating heptapeptide in the N-terminal prion domain of the *Saccharomyces cerevisiae* termination factor Sup35 (Balbirnie *et al.*, 2001; Nelson *et al.*, 2005). The self-assembly behavior of this oligopeptide sequence was associated with manifestation of the [PSI⁺] prion phenotype in *S. cerevisiae* (Cox *et al.*, 1988). The amino acid residues responsible for the polar zipper interface between β -sheets were retained, but the remainder of the structure was designed *de novo* using Rosetta (Huang *et al.*, 2011; Leaver-Fay *et al.*, 2011). Self-assembly was proposed to occur through dimerization of the computationally designed $\beta\alpha\beta$ motif at the β -sheet interface and subsequent elongation through end-to-end propagation of the β -sheet in a cross- β orientation (Figure 1.24b and c). The computationally designed peptide, $\beta\alpha\beta$ Zip, was synthesized using solid-phase synthesis (Figure 1.24d). TEM imaging indicated the presence of high aspect-ratio filaments of apparent width that compared well with the lateral dimensions of the computational model. FTIR spectroscopy and CD spectropolarimetry measurements agreed with a mixed α/β conformational population. While resolution structure was not available to compare to the atomistic model, X-ray fiber diffraction analysis performed on oriented filaments provided evidence for a cross- β structure that agreed well with the computational model.

Foldamer-based helical assemblies

Until now, peptide assemblies derived from discrete length oligomeric α -amino acid sequences, corresponding to native polyamide backbones, have been discussed exclusively. However, these sequences do not represent the sole plausible substrates for design of oligomers competent for self-assembly into structurally defined filaments. Recently, the design of peptido-mimetic foldamers (Goodman *et al.*, 2007) has been explored in order to evaluate for the potential formation of structurally defined supramolecular assemblies (Misra *et al.*, 2021). While sequence–structure correlations are less well-defined for covalent oligomers based on alternative backbone chemistries, foldamer assemblies offer several potential advantages as substrates for the fabrication of biomimetic helical filaments and nanotubes. The primary consideration that has driven investigation into foldamer assemblies is their potential resistance to proteolytic cleavage in a biological environment as well as a lowered propensity to elicit an immune response. Moreover, synthetic foldamers exhibit a rich structural chemistry that differs from that of conventional α -amino acid

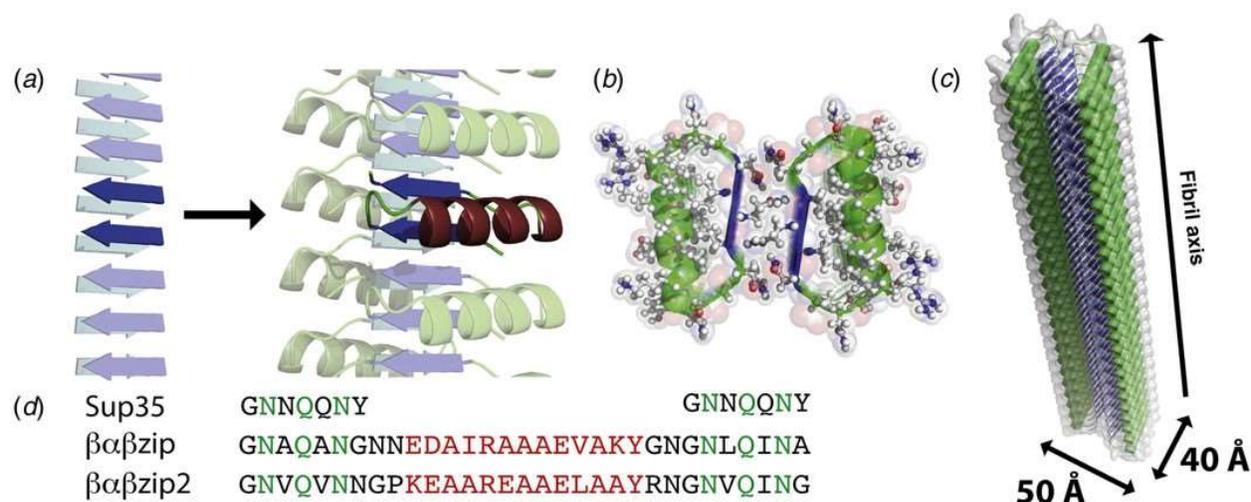


Figure 1.24 De novo design of self-assembling $\beta\alpha\beta$ peptides. (a) The computational design is based on the crystal structure of the amyloidogenic heptapeptide sequence GNNQQNY derived from the sup35 N-terminal prion domain. A loop–helix–loop segment constructed using flexible backbone

design connects two β -strands with backbone conformations taken from the sup35 peptide. (b) Cross section of the computational model of the fibril in a view perpendicular to the fiber axis. The interactions between the two layers of sheets are stabilized by a steric zipper interface formed by interdigitating side chains at the strand interfaces. (c) Structural model the peptide filament with estimated lateral dimensions for a cross section. (d) Sequences of computationally designed peptides $\beta\alpha\beta$ Zip and $\beta\alpha\beta$ Zip2. Residues highlighted in green correspond to the steric zipper interface, while residues highlighted in red correspond to the α -helix. From Kaltofen *et al.* (2015), *Journal of Molecular Biology*, 427 (2), 550–562. Reprinted with permission from Elsevier.

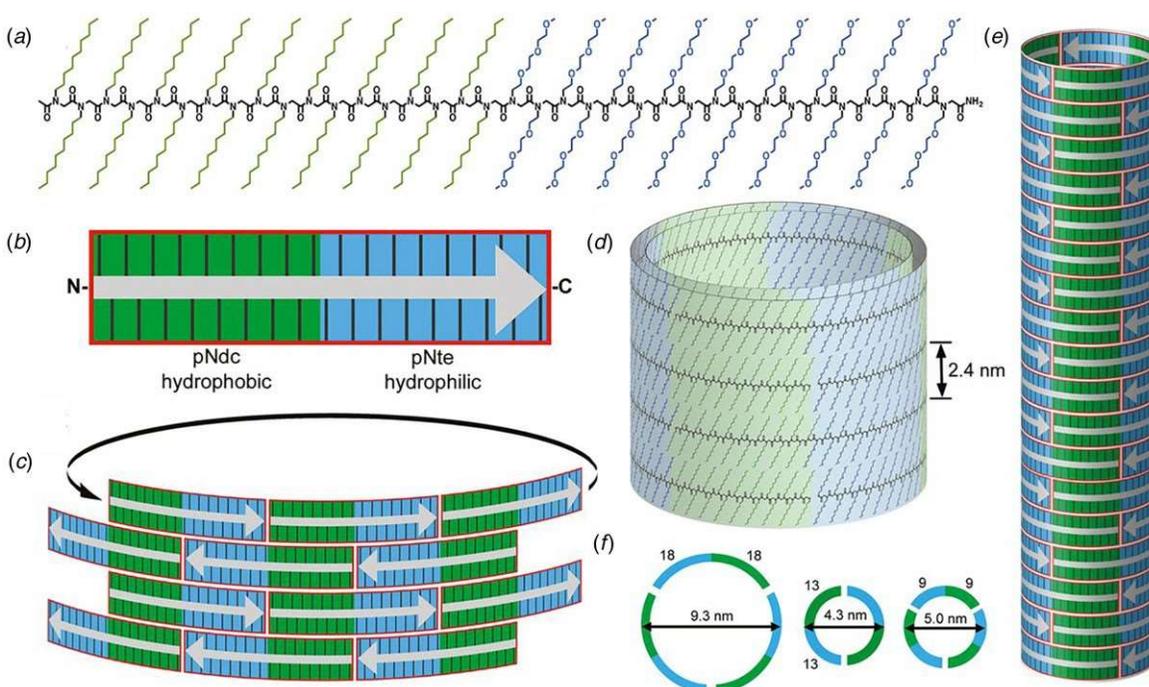


Figure 1.25. Amphiphilic peptoids block copolymer tiles self-assemble into hollow nanotubes. (a) Sequence and chemical structure of pNdc₁₈-b-pNte₁₈. (b) Schematic depiction of the structure of pNdc₁₈-b-pNte₁₈ tile in which the hydrophobic and hydrophilic domains are highlighted in green and blue,

respectively. The gray arrow indicates the *N* to *C* chain trajectory. (c) Schematic representation of the packing of four layers of peptoid tiles within the proposed nanotube structures. (d) Cross section of a four-layer nanotube segment with depiction of side-chain arrangement and the stacking repeat distance. (e) Schematic representation of the proposed nanotube structure with amphiphilic tile packing indicated. (f) Cross-sectional view of the proposed packing of amphiphilic tiles for three peptoid polymers with approximate tube diameter for each cross section. Reprinted with permission from Sun *et al.* (2016), *Proceedings of the National Academy of Sciences of the United States of America*, 113 (15), 3954–3959.

oligomers, which may lead to different supramolecular architectures that may be more suited for a particular downstream application.

Polypeptoids, or polymers of *N*-substituted glycines, have been extensively studied as building blocks for construction of supramolecular assemblies (Sun and Zuckermann, 2013). Unlike natural proteins they cannot be overexpressed in bacteria, but the chemical synthesis of peptoids is no more difficult than solid phase peptide synthesis (Culf and Ouellette, 2010). A diverse range of quaternary structures have been prepared from self assembly of sequence-specific peptoids (Nam *et al.*, 2010), but here we focus on the design of filamentous assemblies. Two research groups independently demonstrated that block-like peptoid oligomers (Xuan and Zuckermann, 2020) based on sequence-encoded contour-length amphiphilicity could self-assemble into filamentous nanomaterials. Similar self-assembly behavior was observed for block-like peptide sequences that display contour-length amphiphilicity (Van Rijt *et al.*, 2019).

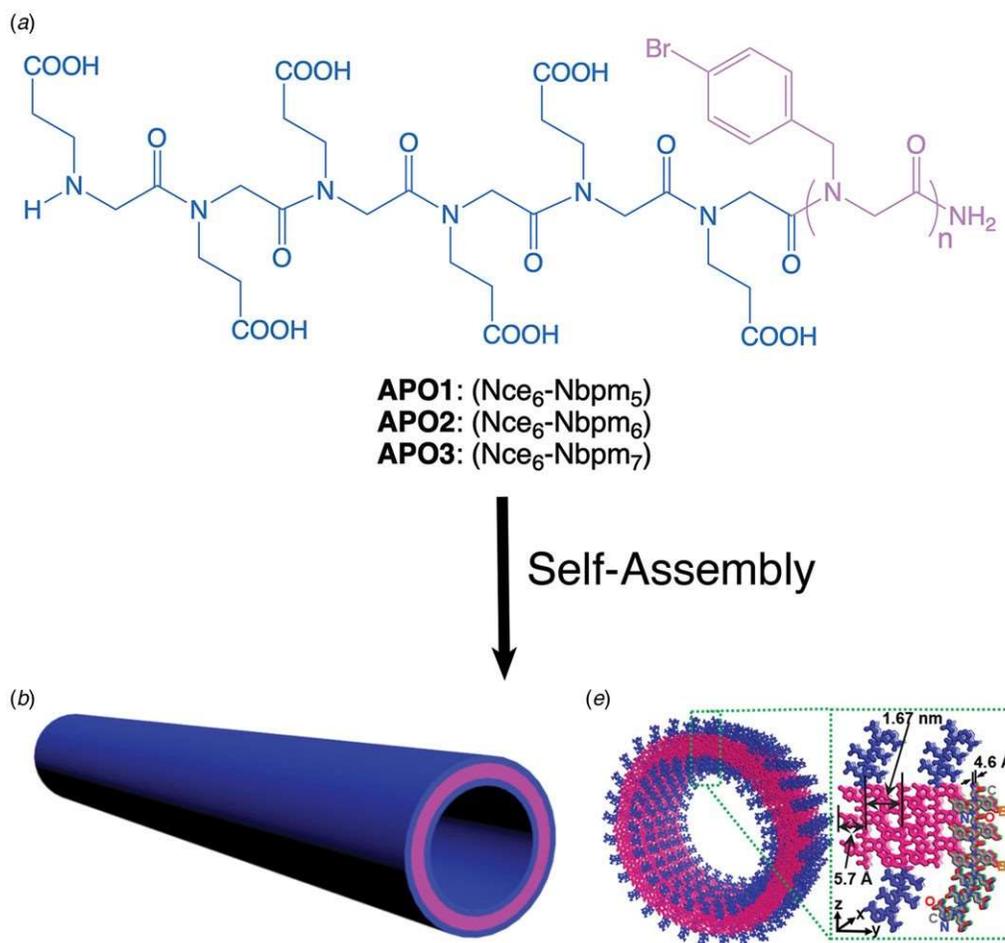


Figure 1.26. Amphiphilic peptoid oligomers (APOs) self-assemble into highly ordered, crystalline, single-walled peptoid nanotubes. (a) Sequence and chemical structure of the APOs. Hydrophilic and hydrophobic blocks are indicated in blue and magenta, respectively. (b) Schematic representation of the proposed structure of the monolayer nanotube with color-coding of the hydrophilic and hydrophobic blocks. (c) Cross-sectional representation of the proposed model for the APO2 nanotubes. The molecular packing of the peptoid protomers is depicted along with the critical distances that define the packing arrangement. Reprinted under Creative Commons License from Jin *et al.* (2018), *Nature Communications*, 9 (1), 270.

Zuckermann and coworkers described a series of peptoid polymers, $(\text{pNdc})_n\text{-}b\text{-}(\text{pNte})_n$ ($n = 9, 13, 18$), that displayed an AB-type amphiphilic block copolymer architecture (Sun *et al.*, 2016). The *N*-substituents of the respective blocks had identical length, i.e. number of non-hydrogen atoms, and linear architecture, but differed significantly in polarity (Figure 1.25*a* and *b*). The poly-*N*-decylglycine (pNdc) A-block was hydrophobic in character, while the poly-*N*-2-(2-(2-methoxyethoxy)ethyl)glycine (pNte) B-block was hydrophilic in nature. The amphiphilic block peptoids self-assembled from a mixed solvent system into structurally well-defined nanotubes. SAXS, TEM, cryo-EM, and cryo-electron tomography (cryo-ET) provided evidence that the nanotubes displayed an unusual supramolecular architecture. The peptoid backbone was arranged circumferentially around the periphery of the nanotubes in an orientation perpendicular to the long axis of the nanotube (Figure 1.25*c-e*). The side chains projected along the tube axis but segregated into separate hydrophilic and hydrophobic domains, which resulted in a regular pattern of stripes along the contour length of the nanotube. The packing of peptoids within the tubes resembled the stacking of tiles. The tiles were composed of side chains in which the uniform length of the hydrophilic and hydrophobic substituents was proposed to abet the ordered stacking. Three different peptoid block polymers were prepared in which the A and B block lengths were identical in the number of monomers ($n = 9, 13, \text{ and } 18$). In every case, the axial periodicity of 24 Å was conserved in the corresponding assemblies, although the diameter of the tubes varied depending on the peptoid sequence (Figure 1.25*f*).

In contrast, Chen and co-workers described a class of peptoid block copolymer amphiphiles, $(\text{pNce})_6\text{-}b\text{-}(\text{pNbpm})_n$ ($n = 5, 6, 7$) (Figure 1.26*a* and *b*) in which the length of the hydrophobic block was varied (Jin *et al.*, 2018). Self-assembly of these peptoid polymers from a mixed solvent system afforded mechanically robust, crystalline nanotubes. TEM analysis indicated that the self-assembly process involved a structural evolution from particles to ribbons to tubes. The wall thickness, diameter, and mechanical properties of the tubes could be controlled through the sequence of the peptoid oligomers, that is, the length of the

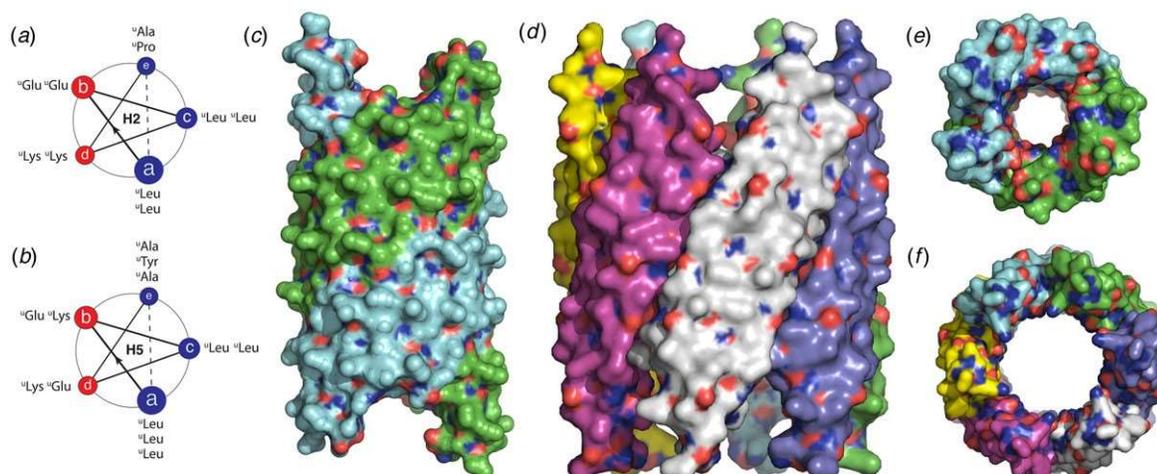


Figure 1.27. Oligoureas can adopt amphiphilic helical secondary structures that promote self-assembly into nanotubes. (a, b) Helical wheel diagrams of the decameric (H2) and dodecameric (H5) oligourea sequences indicating the pentad registry. The facial amphiphilicity can be detected from the pattern of polar residues (red) and non-polar residues (blue). (^uXaa indicates the urea analog of a canonical amino acid.) (c, e) Side and top views of the H2 nanotube derived from the corresponding crystal structure. (d, f) Side and top views of the H5 nanotube derived from the corresponding crystal structure. In each case, the individual protofilaments within the respective structures are color coded.

hydrophobic peptoid segment. Remarkably, the nanotubes were able to undergo a reversible pH-driven expansion and contraction of the tube diameter without any evidence for disassembly. Despite a similar contour-length amphiphilicity to the nanotubes described by Zuckermann and coworkers, these peptoid polymers were arranged in an antiparallel orientation within an amphiphilic monolayer. The peptoid backbones were proposed to be oriented radially within the nanotube rather than circumferentially as proposed for the previously described peptoid block copolymers (Figure 1.26c). The hydrophobic 4-bromophenylmethyl (pNbpm) B-block substituents were buried in the monolayer core, while the polar 2-carboxyethyl (pNce) A-block segments extended from either surface. TEM measurements of the wall thickness of the tubes supported this structural model, in that the wall thickness increased as the length of

the pNbpmn segment increased.

Similarly, foldamers based on aliphatic oligourea backbone (Violette *et al.*, 2005) have been demonstrated to form helical assemblies in which polar sequence patterning was observed to promote selective self-assembly (Collie *et al.*, 2015). In contrast to the peptoids described above, the sequences of the synthetic oligoureas were designed to display conformational, that is, facial, amphiphilicity. Oligoureas had been previously demonstrated to adopt local helical secondary structures based on a pentad repeat motif, which corresponded to five monomers per two helical turns within the conformation (

Figure 1.27*a* and *b*). Guichard and coworkers (Collie *et al.*, 2015) demonstrated that oligourea foldamers could assemble into either discrete hexameric helical bundles or extended nanotubes. The preference for the respective structures depended on the polar patterning of the urea side chains within the pentad-based sequence of the oligomer. Formation of an extended hydrophobic interface within a helical conformation promoted higher-order assembly of the oligourea protomers into helical assemblies.

Two oligourea oligomers, decamer H2 and dodecamer H5, were synthesized based on the same polar pattern although with slightly different sequences. Single-crystal X-ray diffraction analysis of these two oligourea assemblies revealed the presence of nanotubes in both cases. The structure of H2 was based on a helical assembly with apparent C_2 symmetry in which the two protofilaments corresponded to pairs of antiparallel helices stacked along the respective 2-start helices (

Figure 1.27*c* and *e*). The structurally related oligourea H5 assembled into a nanotube of apparent C_6 symmetry, in which the protofilaments corresponded to the 6-start helices of the assembly (

Figure 1.27*d* and *f*). Conventional TEM and cryo-EM confirmed the presence of isolated nanotubes of similar dimensions to those observed crystallographically. Notably, slight changes in oligourea sequence could result in dramatic changes in the structure of the assembly, in a manner that was reminiscent of the structural plasticity often observed in peptide assemblies (Yoo *et al.*, 2020). *N*-Methylation of the C-terminal amide resulted in a structural variant of the H2 decamer, H2', which crystallized into two distinctive helical assemblies under different sets of experimental conditions. One

form adopted a structure that was identical to that of H2 (

Figure 1.27*c* and *e*). However, under a separate set of conditions, a completely different structure was observed with a different helical periodicity although it retained C_2 symmetry.

Finally, higher-order filamentous assemblies have been observed for peptides based on non- α -amino acid sequences. Oligomers of β^3 -homoamino acids (Gopalan *et al.*, 2015) have been observed to self-assemble into filaments (Del Borgo *et al.*, 2013; Christofferson *et al.*, 2018), and form lyotropic mesophases based on filamentous assemblies (Pomerantz *et al.*, 2006, 2008, 2011; Pizzey *et al.*, 2008). The sequence determinants responsible for self-assembly have been investigated, but thus far are not wellunderstood due to the absence of NAR structural information. At least one example of an oligo- β^3 -homopeptide sequence has been reported in which formation of hollow cylindrical nanotubes was inferred on the basis of conventional TEM, low-resolution cryo-EM, and SAXS measurements (Pizzey *et al.*, 2008; Pomerantz *et al.*, 2008). The polar patterning of this peptide system does not fit a conventional model that can correlate sequence with higher-order structure for the 14-helix conformation (Cheng *et al.*, 2001) that is often observed for oligo- β^3 -homopeptides.

In the case of oligo- β^3 -homopeptide sequences, complex filamentous structures can arise even from tripeptides (Figure 1.28). Christofferson *et al.* reported a detailed structural analysis of assemblies derived from hydrophobic tripeptides composed of different sequence combinations of β^3 -homoLeu, β^3 -homolle, and β^3 -homoAla (Christofferson *et al.*, 2018). X-ray fiber diffraction of oriented samples derived from assemblies of tripeptide Ac- β^3 [hLhIhA] (hL, homoleucine; hI, homoisoleucine; hA, homoalanine) provided evidence for a supramolecular structure in which three helical protofilaments formed a trimeric superhelix through self-association (Figure 1.28*b-d*). X-ray diffraction data on Ac- β^3 [hLhIhA] were employed to benchmark molecular dynamics simulations, which were extended to the other tripeptide variants that were sequence permutants of Ac- β^3 [hLhIhA]. MD simulations were compared with AFM measurements and provided evidence that the supramolecular structural variations between the tripeptides originated in differences in the packing details at the helix-helix interface within the corresponding atomic models. Further confirmation of these sequence-structure correlations will await structural

determination at NAR.

While the structural analysis of foldamer assemblies is at an early stage in its development, opportunities abound for the creation of structurally defined helical assemblies from sequences based on alternative backbone and side-chain chemistries.

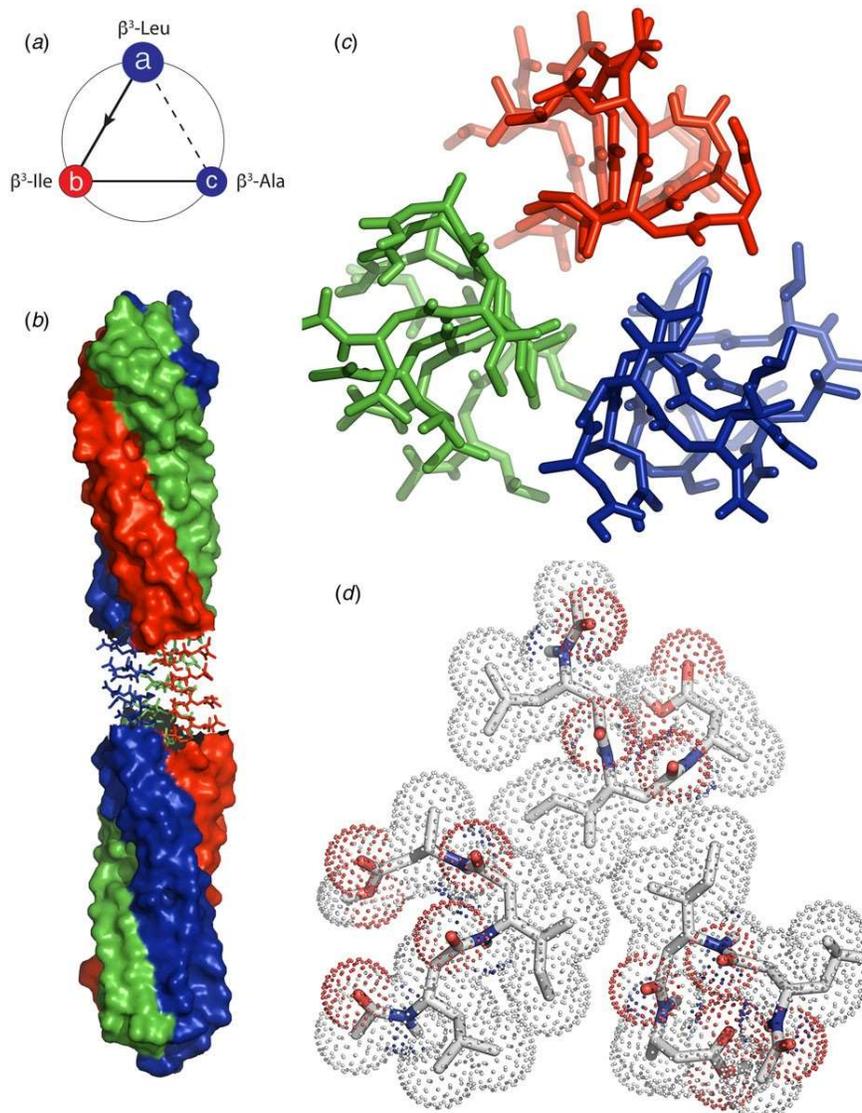


Figure 1.28 (a) Helical wheel diagram of tripeptide Ac- β^3 [hLhIhA] indicating the triad registry in a lefthanded 14-helix conformation. (b) Side view of the triple-helical ribbon model of the Ac- β^3 [hLhIhA] filament in water. (c) Expansion of a four-layer helical stack from (b) highlighting packing at the trimeric interface in a stick representation. (d) Trimeric packing interface of a single-layer of the triple helix in which inter-residue contacts are indicated using a mesh depiction.

High-resolution structural analysis, particularly cryo-EM, offers the ability to provide insight into the sequence determinants that underlie the formation of these chemically and structurally orthogonal helical assemblies, which would enable a better understanding of the local sequence and conformational factors that underlie self-assembly behavior and provide rubrics to better guide the design of such synthetic assemblies.

Conclusions

Helical peptide assemblies offer engineering opportunities that range from peptide design to assembly functionalization. However, the rational design of synthetic helical assemblies is complicated in that minor sequence modifications within the subunits can have a considerable influence on the quaternary structure of the resultant assemblies. These effects can be difficult to predict *a priori*, in part due to the limited availability of high-resolution structural information on peptide and peptido-mimetic helical filaments and tubes. While the latter situation is rapidly undergoing transformative growth, reliable *de novo* design of helical assemblies remains a significant challenge.

NAR methods for structural determination, i.e. cryo-EM, MicroED, and ssNMR, are revolutionizing the structural analysis of peptide and protein assemblies. As these resources become more widely available, we expect to see increased information from structural studies, which will enable a better understanding of the factors that influence higher-order structure within helical assemblies. Many significant questions remain to be answered, in particular, the role of amphiphilicity in conformational selection and in orientational preferences for packing of protomers within the filaments. A better understanding of structural principles is critical to the development of synthetic helical assemblies that emulate the functional properties of native biological filaments and can be tailored for specific applications.

Acknowledgments. The authors thank Edward Egelman, Fengbin Wang, James Nowick, Louise Serpell, Mibel Aguilar, Gilles Guichard, Anil Mehta, Chunlong Chen, and Ulf Olsson for useful discussions and providing data and graphics for figures.

Financial support. This work was supported by the National Science Foundation (NSF) (V. P. C., grant number DMR-1534317).

Conflict of interest. The authors declare no competing interests.

References

Adamcik J and Mezzenga R (2018) Amyloid polymorphism in the protein folding and aggregation energy landscape. *Angewandte Chemie International Edition* 57, 8370–8382.

Aggeli A, Bell M, Boden N, Keen JN, Knowles PF, Mclleish TCB, Pitkeathly M and Radford SE (1997) Responsive gels formed by the spontaneous self-assembly of peptides into polymeric β -sheet tapes. *Nature* 386, 259–262.

Aggeli A, Nyrkova IA, Bell M, Harding R, Carrick L, Mclleish TC, Semenov AN and Boden N (2001) Hierarchical self-assembly of chiral rod-like molecules as a model for peptide β -sheet tapes, ribbons, fibrils, and fibers. *Proceedings of the National Academy of Sciences of the United States of America* 98, 11857–11862.

Akey DL, Malashkevich VN and Kim PS (2001) Buried polar residues in coiled-coil interfaces. *Biochemistry* 40, 6352–6360.

Anzini P, Xu C, Hughes S, Magnotti E, Jiang T, Hemmingsen L, Demeler B and Conticello VP (2013) Controlling self-assembly of a peptide-based material via metal-Ion induced registry shift. *Journal of the American Chemical Society* 135, 10278–10281.

Baek M, Dimaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, Millan C, Park H, Adams C, Glassman CR, Degiovanni A, Pereira JH, Rodrigues AV, Van Dijk AA, Ebrecht AC, Opperman DJ, Sagmeister T, Buhlheller C, Pavkov-Keller T,

Rathinaswamy MK, Dalwadi U, Yip CK, Burke JE, Garcia KC, Grishin NV, Adams PD, Read RJ and Baker D (2021)

Accurate prediction of protein structures and interactions using a three-track neural network. *Science (New York, N.Y.)* 373, 871–876.

Balbirnie M, Grothe R and Eisenberg DS (2001) An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated beta-sheet structure for amyloid. *Proceedings of the National Academy of Sciences of the United States of America* 98, 2375–2380.

Beesley JL and Woolfson DN (2019) The *de novo* design of α -helical peptides for supramolecular self-assembly. *Current Opinion in Biotechnology* 58, 175–182.

Bera S, Mondal S, Xue B, Shimon LJW, Cao Y and Gazit E (2019) Rigid helical-like assemblies from a self-aggregating tripeptide. *Nature Materials* 18, 503–509.

Bong DT, Clark TD, Granja JR and Ghadiri MR (2001) Self-assembling organic nanotubes. *Angewandte Chemie International Edition* 40, 988–1011.

Bowerman CJ and Nilsson BL (2012) Review self-assembly of amphipathic β -sheet peptides: insights and applications. *Peptide Science* 98, 169–184.

Bucak S, Cenker C, Nasir I, Olsson U and Zackrisson M (2009) Peptide nanotube nematic phase. *Langmuir* 25, 4262–4265.

Burgess NC, Sharp TH, Thomas F, Wood CW, Thomson AR, Zaccai NR, Brady RL, Serpell LC and Woolfson DN (2015) Modular design of self-assembling peptide-based nanotubes. *Journal of the American Chemical Society* 137, 10554–10562.

Carter NA and Grove TZ (2015) Repeat-protein films exhibit hierarchical anisotropic mechanical properties. *Biomacromolecules* 16, 706–714.

Carter NA and Grove TZ (2018) Protein self-assemblies that can generate, hold, and discharge electric potential in response to changes in relative humidity. *Journal of the American Chemical Society* 140, 7144–7151.

Castelletto V, Nutt DR, Hamley IW, Bucak S, Cenkler Ç and Olsson U (2010) Structure of single-wall peptide nanotubes: *in situ* flow aligning X-ray diffraction. *Chemical Communications* 46, 6270–6272.

Cenkler ÇÇ, Bomans PHH, Friedrich H, Dedeoğlu B, Aviyente V, Olsson U,

Sommerdijk NAJM and Bucak S (2012) Peptide nanotube formation: a crystal growth process. *Soft Matter* 8, 7463–7470.

Chamberlain AK, Macphee CE, Zurdo J, Morozova-Roche LA, Hill HA, Dobson CM and Davis JJ (2000) Ultrastructural organization of amyloid fibrils by atomic force microscopy. *Biophysical Journal* 79, 3282–3293.

Chen KH, Corro KA, Le SP and Nowick JS (2017) X-ray crystallographic structure of a giant double-walled peptide nanotube formed by a macrocyclic β -sheet containing $A\beta_{16-22}$. *Journal of the American Chemical Society* 139, 8102–8105.

Chen YX, Ing NL, Wang F, Xu D, Sloan NB, Lam NT, Winter DL, Egelman EH, Hochbaum AI, Clark DS and Glover DJ (2020) Structural determination of a filamentous chaperone to fabricate electronically conductive metalloprotein nanowires. *ACS Nano* 14, 6559–6569.

Cheng RP, Gellman SH and Degrado WF (2001) Beta-peptides: from structure to function. *Chemical Reviews* 101, 3219–3232.

Chervy P, Petcut C, Rault D, Meriadec C, Bizien T, François K, Richard J, Chassaing C, Benamar N, Artzner F and Paternostre M (2019) Organic nanoscrolls from electrostatic interactions between peptides and lipids: assembly steps and structure. *Langmuir* 35, 10648–10657.

Chiu W, Holton J, Langan P, Sauter NK, Schlichting I, Terwilliger T, Martin JL, Read RJ and Wakatsuki S (2017) Responses to ‘atomic resolution’: a badly abused term in structural biology. *Acta Crystallographica Section D: Structural Biology* 73 (Pt 4), 381–383.

Chothia C (1973) Conformation of twisted beta-pleated sheets in proteins. *Journal of Molecular Biology* 75, 295–302.

Christofferson AJ, Al-Garawi ZS, Todorova N, Turner J, Del Borgo MP, Serpell LC, Aguilar M-I and Yarovsky I (2018) Identifying the coiled-coil triple helix structure of β -peptide nanofibers at atomic resolution. *ACS Nano* 12, 9101–9109.

Close W, Neumann M, Schmidt A, Hora M, Annamalai K, Schmidt M, Reif B, Schmidt V, Grigorieff N and Fändrich M (2018) Physical basis of amyloid fibril polymorphism. *Nature Communications* 9, 699.

Collie GW, Pulka-Ziach K, Lombardo CM, Fremaux J, Rosu F, Decossas M, Mauran L, Lambert O, Gabelica V, Mackereth CD and Guichard G (2015) Shaping quaternary assemblies of water-soluble non-peptide helical foldamers by sequence manipulation. *Nature Chemistry* 7, 871–878.

Colvin MT, Silvers R, Ni QZ, Can TV, Sergeyev I, Rosay M, Donovan KJ, Michael B, Wall J, Linse S and Griffin RG (2016) Atomic resolution structure of monomorphic A β ₄₂ amyloid fibrils. *Journal of the American Chemical Society* 138, 9663–9674.

Cormier AR, Pang X, Zimmerman MI, Zhou H-X and Paravastu AK (2013) Molecular structure of RADA16-I designer self-assembling peptide nanofibers. *ACS Nano* 7, 7562–7572.

Costa TRD, Felisberto-Rodrigues C, Meir A, Prevost MS, Redzej A, Trokter M and Waksman G (2015) Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nature Reviews Microbiology* 13, 343–359.

Costa TRD, Ilangovan A, Ukleja M, Redzej A, Santini JM, Smith TK, Egelman EH and Waksman G (2016) Structure of the bacterial sex F pilus reveals an assembly of a stoichiometric protein–phospholipid complex. *Cell* 166, 1436–1444.e1410.

Cox BS, Tuite MF and Mclaughlin CS (1988) The psi factor of yeast: a problem in inheritance. *Yeast* 4, 159–178.

Crick F (1953) The packing of α -helices: simple coiled-coils. *Acta Crystallographica* 6, 689–697.

Culf AS and Ouellette RJ (2010) Solid-phase synthesis of *N*-substituted glycine oligomers (α -peptoids) and derivatives. *Molecules* 15, 5282–5335.

Da Silva ER, Alves WA, Castelletto V, Reza M, Ruokolainen J, Hussain R and Hamley IW (2015a) Self-assembly pathway of peptide nanotubes formed by a glutamatic acid-based bolaamphiphile. *Chemical Communications* 51, 11634–11637.

Da Silva ER, Walter MNM, Reza M, Castelletto V, Ruokolainen J, Connon CJ, Alves WA and Hamley IW (2015b) Self-assembled arginine-capped peptide bolaamphiphile nanosheets for cell culture and controlled wettability surfaces. *Biomacromolecules* 16, 3180–3190.

Dauter Z and Jaskolski M (2018) On the helical arrangements of protein molecules. *Protein Science* 27, 643–652.

Dawson WM, Martin FJO, Rhys GG, Shelley KL, Brady RL and Woolfson DN (2021) Coiled coils 9-to-5: rational *de novo* design of alpha-helical barrels with tunable oligomeric states. *Chemical Science* 12, 6923–6928.

Del Borgo MP, Mechler AI, Traore D, Forsyth C, Wilce JA, Wilce MCJ, Aguilar M-I and Perlmutter P (2013) Supramolecular self-assembly of *N*-acetyl-capped β -peptides leads to nanoto macroscale fiber formation. *Angewandte Chemie International Edition* 52, 8266–8270.

Diaz-Avalos R, Long C, Fontano E, Balbirnie M, Grothe R, Eisenberg D and Caspar DL (2003) Cross-beta order and diversity in nanocrystals of an amyloid-forming peptide. *Journal of Molecular Biology* 330, 1165–1175. Dublin SN and Conticello VP (2008) Design of a selective metal Ion switch for self-assembly of peptide-based fibrils. *Journal of the American Chemical Society* 130, 49–51.

Dubochet J (2012) Cryo-EM – the first thirty years. *Journal of Microscopy* 245, 221–224.

Eanes ED and Glenner GG (1968) X-ray diffraction studies on amyloid filaments. *Journal of Histochemistry & Cytochemistry* 16, 673–677.

Egelman EH (2000) A robust algorithm for the reconstruction of helical filaments using single-particle methods. *Ultramicroscopy* 85, 225–234.

Egelman EH (2007) The iterative helical real space reconstruction method: surmounting the problems posed by real polymers. *Journal of Structural Biology* 157, 83–94.

Egelman EH (2010) Reconstruction of helical filaments and tubes. *Methods Enzymol* 482, 167–183.

- Egelman EH (2014) Ambiguities in helical reconstruction. *Elife* 3, e04969.
- Egelman EH (2017) Cryo-EM of bacterial pili and archaeal flagellar filaments. *Current Opinion in Structural Biology* 46, 31–37.
- Egelman EH, Francis N and Derosier DJ (1982) F-actin is a helix with a random variable twist. *Nature* 298, 131–135.
- Egelman EH, Xu C, Dimaio F, Magnotti E, Modlin C, Yu X, Wright E, Baker D and Conticello VP (2015) Structural plasticity of helical nanotubes based on coiled-coil assemblies. *Structure (London, England: 1993)* 23, 280–289.
- Eisenberg DS and Sawaya MR (2017) Structural studies of amyloid proteins at the molecular level. *Annual Review of Biochemistry* 86, 69–95.
- Feng Z, Wang H, Wang F, Oh Y, Berciu C, Cui Q, Egelman EH and Xu B (2020) Artificial intracellular filaments. *Cell Reports Physical Science* 1, 100085.
- Frederix PWJM, Ulijn RV, Hunt NT and Tuttle T (2011) Virtual screening for dipeptide aggregation: toward predictive tools for peptide self-assembly. *The Journal of Physical Chemistry Letters* 2, 2380–2384.
- Frederix PWJM, Scott GG, Abul-Haija YM, Kalafatovic D, Pappas CG, Javid N, Hunt NT, Ulijn RV and Tuttle T (2015) Exploring the sequence space for (tri-)peptide self-assembly to design and discover new hydrogels. *Nature Chemistry* 7, 30–37.
- Gallardo R, Ranson NA and Radford SE (2020) Amyloid structures: much more than just a cross- β fold. *Current Opinion in Structural Biology* 60, 7–16.
- Garcia-Seisdedos H, Empereur-Mot C, Elad N and Levy ED (2017) Proteins evolve on the edge of supramolecular self-assembly. *Nature* 548, 244–247. Garcia-Seisdedos H, Villegas JA and Levy ED (2019) Infinite assembly of folded proteins in evolution, disease, and engineering. *Angewandte Chemie International Edition* 58, 5514–5531.

Ge P and Zhou ZH (2011) Hydrogen-bonding networks and RNA bases revealed by cryo electron microscopy suggest a triggering mechanism for calcium switches. *Proceedings of the National Academy of Sciences of the United States of America* 108, 9637–9642.

Gelain F, Luo Z and Zhang S (2020) Self-assembling peptide EAK16 and RADA16 nanofiber scaffold hydrogel. *Chemical Reviews* 120, 13434–13460. Ghadiri MR, Granja JR, Milligan RA, Mcree DE and Khazanovich N (1993) Self-assembling organic nanotubes based on a cyclic peptide architecture.

Nature 366, 324–327.

Gobeaux F, Fay N, Tarabout C, Meriadec C, Meneau F, Ligeti M, Buisson DA, Cintrat JC, Nguyen KM, Perrin L, Valery C, Artzner F and Paternostre M (2012) Structural role of counterions adsorbed on self-assembled peptide nanotubes. *Journal of the American Chemical Society* 134, 723–733.

Goodman CM, Choi S, Shandler S and Degrado WF (2007) Foldamers as versatile frameworks for the design and evolution of function. *Nature Chemical Biology* 3, 252–262.

Gopalan RD, Del Borgo MP, Mechler AI, Perlmutter P and Aguilar M-I (2015) Geometrically precise building blocks: the self-assembly of β -peptides. *Chemistry & Biology* 22, 1417–1423.

Görbitz CH (2001) Nanotube formation by hydrophobic dipeptides.

Chemistry – A European Journal 7, 5153–5159.

Görbitz CH (2006) The structure of nanotubes formed by diphenylalanine, the core recognition motif of Alzheimer's β -amyloid polypeptide. *Chemical Communications* 22, 2332–2334.

Gremer L, Schölzel D, Schenk C, Reinartz E, Labahn J, Ravelli RBG, Tusche M, Lopez-Iglesias C, Hoyer W, Heise H, Willbold D and Schröder GF (2017) Fibril structure of amyloid- β (1–42) by cryo-electron microscopy. *Science (New York, N.Y.)* 358, 116–119.

Gribbon C, Channon KJ, Zhang W, Banwell EF, Bromley EHC, Chaudhuri

JB, Oreffo ROC and Woolfson DN (2008) MagicWand: a single, designed peptide that assembles to stable, ordered α -helical fibers. *Biochemistry* 47, 10365–10371.

Guenther EL, Ge P, Trinh H, Sawaya MR, Cascio D, Boyer DR, Gonen T, Zhou ZH and Eisenberg DS (2018) Atomic-level evidence for packing and positional amyloid polymorphism by segment from TDP-43 RRM2. *Nature Structural & Molecular Biology* 25, 311–319.

Guerrero-Ferreira R, Taylor NM, Mona D, Ringler P, Lauer ME, Riek R, Britschgi M and Stahlberg H (2018) Cryo-EM structure of alpha-synuclein fibrils. *Elife* 7, e36402.

Guerrero-Ferreira R, Taylor NM, Arteni A-A, Kumari P, Mona D, Ringler P, Britschgi M, Lauer ME, Makky A, Verasdonck J, Riek R, Melki R, Meier BH, Böckmann A, Bousset L and Stahlberg H (2019) Two new polymorphic structures of human full-length alpha-synuclein fibrils solved by cryo-electron microscopy. *Elife* 8, e48907.

Guilbaud J-B and Saiani A (2011) Using small angle scattering (SAS) to structurally characterise peptide and protein self-assembled materials. *Chemical Society Reviews* 40, 1200–1210.

Haines-Butterick L, Rajagopal K, Branco M, Salick D, Rughani R, Pilarz M, Lamm MS, Pochan DJ and Schneider JP (2007) Controlling hydrogelation kinetics by peptide design for three-dimensional encapsulation and injectable delivery of cells. *Proceedings of the National Academy of Sciences of the United States of America* 104, 7791–7796.

Hamley IW (2011) Self-assembly of amphiphilic peptides. *Soft Matter* 7, 4122–4138.

Hamley IW (2014) Peptide nanotubes. *Angewandte Chemie International Edition* 53, 6866–6881.

Hamley IW, Burholt S, Hutchinson J, Castelletto V, Da Silva ER, Alves W, Gutfreund P, Porcar L, Dattani R, Hermida-Merino D, Newby G, Reza M, Ruokolainen J and Stasiak J (2017) Shear alignment of bola-amphiphilic arginine-coated peptide nanotubes. *Biomacromolecules* 18, 141–149.

Harper JD, Lieber CM and Lansbury Jr PT (1997) Atomic force microscopic imaging of seeded fibril formation and fibril branching by the Alzheimer's disease amyloid-beta protein. *Chemistry & Biology* 4, 951–959.

He S and Scheres SHW (2017) Helical reconstruction in RELION. *Journal of Structural Biology* 198, 163–176.

Heise H, Hoyer W, Becker S, Andronesi OC, Riedel D and Baldus M (2005) Molecular-level secondary structure, polymorphism, and dynamics of full-length α -synuclein fibrils studied by solid-state NMR. *Proceedings of the National Academy of Sciences of the United States of America* 102, 15871–15876.

Huang PS, Ban YE, Richter F, Andre I, Vernon R, Schief WR and Baker D (2011) RosettaRemodel: a generalized framework for flexible backbone protein design. *PLoS One* 6, e24109.

Huang D, Hudson BC, Gao Y, Roberts EK and Paravastu AK (2018) Solid-state NMR structural characterization of self-assembled peptides with selective ^{13}C and ^{15}N isotopic labels. In Nilsson BL and Doran TM (eds), *Peptide Self-Assembly: Methods and Protocols*. New York, NY: Springer New York, pp. 23–68.

Hughes SA, Wang F, Wang S, Kreutzberger MAB, Osinski T, Orlova A, Wall JS, Zuo X, Egelman EH and Conticello VP (2019) Ambidextrous helical nanotubes from self-assembly of designed helical hairpin motifs. *Proceedings of the National Academy of Sciences of the United States of America* 116, 14456–14464.

Hwang W, Marini DM, Kamm RD and Zhang S (2003) Supramolecular structure of helical ribbons self-assembled from a β -sheet peptide. *The Journal of Chemical Physics* 118, 389–397.

Insua I and Montenegro J (2020) 1D to 2D self assembly of cyclic peptides. *Journal of the American Chemical Society* 142, 300–307.

Jiao W, Yang H, Wu Z, Liu J and Zhang W (2020) Self-assembled block polymer aggregates in selective solution: controllable morphology transitions and their applications in drug delivery. *Expert Opinion on Drug Delivery* 17, 947–961.

Jin H, Ding YH, Wang M, Song Y, Liao Z, Newcomb CJ, Wu X, Tang XQ, Li Z, Lin Y, Yan F, Jian T, Mu P and Chen CL (2018) Designable and dynamic single-walled stiff nanotubes assembled from sequence-defined peptoids. *Nature Communications* 9, 270.

Jumper J and Hassabis D (2022) Protein structure predictions to atomic accuracy with AlphaFold. *Nature Methods* 19, 11–12.

- Kajander T, Cortajarena AL, Mochrie S and Regan L (2007) Structure and stability of designed TPR protein superhelices: unusual crystal packing and implications for natural TPR proteins. *Acta Crystallographica Section D: Biological Crystallography* 63 (Pt 7), 800–811.
- Kajava AV (2012) Tandem repeats in proteins: from sequence to structure. *Journal of Structural Biology* 179, 279–288.
- Kajava AV and Steven AC (2006) Beta-rolls, beta-helices, and other betasolenoid proteins. *Advances in Protein Chemistry* 73, 55–96.
- Kaltofen S, Li C, Huang PS, Serpell LC, Barth A and Andre I (2015) Computational *de novo* design of a self-assembling peptide with predefined structure. *Journal of Molecular Biology* 427, 550–562.
- Ke PC, Zhou R, Serpell LC, Riek R, Knowles TPJ, Lashuel HA, Gazit E, Hamley IW, Davis TP, Fändrich M, Otzen DE, Chapman MR, Dobson CM, Eisenberg DS and Mezzenga R (2020) Half a century of amyloids: past, present and future. *Chemical Society Reviews* 49, 5473–5509.
- Kim J, Han TH, Kim Y-I, Park JS, Choi J, Churchill DG, Kim SO and Ihee H (2010) Role of water in directing diphenylalanine assembly into nanotubes and nanowires. *Advanced Materials* 22, 583–587.
- Knowles TP, Fitzpatrick AW, Meehan S, Mott HR, Vendruscolo M, Dobson CM and Welland ME (2007) Role of intermolecular forces in defining material properties of protein nanofibrils. *Science (New York, N.Y.)* 318, 1900–1903.
- Kobe B and Kajava AV (2000) When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends in Biochemical Sciences* 25, 509–515.
- Kollmer M, Close W, Funk L, Rasmussen J, Bsoul A, Schierhorn A, Schmidt M, Sigurdson CJ, Jucker M and Fändrich M (2019) Cryo-EM structure and polymorphism of abeta amyloid fibrils purified from Alzheimer's brain tissue. *Nature Communications* 10, 4760.
- Koronakis V, Sharff A, Koronakis E, Luisi B and Hughes C (2000) Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* 405, 914–919.
- Krissinel E and Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* 372, 774–797.

- Kühlbrandt W (2014) The resolution revolution. *Science (New York, N.Y.)* 343, 1443–1444.
- Lamm MS, Rajagopal K, Schneider JP and Pochan DJ (2005) Laminated morphology of nontwisting beta-sheet fibrils constructed via peptide self-assembly. *Journal of the American Chemical Society* 127, 16692–16700.
- Lanci CJ, Macdermaid CM, Kang SG, Acharya R, North B, Yang X, Qiu XJ, Degrado WF and Saven JG (2012) Computational design of a protein crystal. *Proceedings of the National Academy of Sciences of the United States of America* 109, 7304–7309.
- Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F, Ban YE, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popovic Z, Havranek JJ, Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D and Bradley P (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods in Enzymology* 487, 545–574.
- Lee M, Wang T, Makhlynets OV, Wu Y, Polizzi NF, Wu H, Gosavi PM, Stöhr J, Korendovych IV, Degrado WF and Hong M (2017) Zinc-binding structure of a catalytic amyloid from solid-state NMR. *Proceedings of the National Academy of Sciences of the United States of America* 114, 6191–6196.
- Lepault J, Booy FP and Dubochet J (1983) Electron microscopy of frozen biological suspensions. *Journal of Microscopy* 129, 89–102.
- Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, Agard DA and Cheng Y (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* 10, 584–590.
- Liu J, Zheng Q, Deng Y, Cheng C-S, Kallenbach NR and Lu M (2006) A seven-helix coiled coil. *Proceedings of the National Academy of Sciences of the United States of America* 103, 15457–15462.

- Loquet A, Sgourakis NG, Gupta R, Giller K, Riedel D, Goosmann C, Griesinger C, Kolbe M, Baker D, Becker S and Lange A (2012) Atomic model of the type III secretion system needle. *Nature* 486, 276–279.
- Lu K, Jacob J, Thiagarajan P, Conticello VP and Lynn DG (2003) Exploiting amyloid fibril lamination for nanotube self-assembly. *Journal of the American Chemical Society* 125, 6391–6393.
- Lu A, Li Y, Yin Q, Ruan J, Yu X, Egelman E and Wu H (2015) Plasticity in PYD assembly revealed by cryo-EM structure of the PYD filament of AIM2. *Cell Discovery* 1, 15013.
- Lupas AN and Bassler J (2017) Coiled coils – a model system for the 21st century. *Trends in Biochemical Sciences* 42, 130–140.
- Lynch EM, Hicks DR, Shepherd M, Endrizzi JA, Maker A, Hansen JM, Barry RM, Gitai Z, Baldwin EP and Kollman JM (2017) Human CTP synthase filament structure reveals the active enzyme conformation. *Nature Structural & Molecular Biology* 24, 507–514.
- Lynch EM, Kollman JM and Webb BA (2020) Filament formation by metabolic enzymes – a new twist on regulation. *Current Opinion in Cell Biology* 66, 28–33.
- Main ER, Xiong Y, Cocco MJ, D’Andrea L and Regan L (2003) Design of stable alpha-helical arrays from an idealized TPR motif. *Structure (London, England: 1993)* 11, 497–508.
- Marchesan S, Vargiu AV and Styan KE (2015) The Phe–Phe motif for peptide self-assembly in nanomedicine. *Molecules* 20, 19775–19788.
- Marini DM, Hwang W, Lauffenburger DA, Zhang S and Kamm RD (2002) Left-handed helical ribbon intermediates in the self-assembly of a β -sheet peptide. *Nano Letters* 2, 295–299.
- Marvin DA, Symmons MF and Straus SK (2014) Structure and assembly of filamentous bacteriophages. *Progress in Biophysics and Molecular Biology* 114, 80–122.
- Mehta AK, Shayo Y, Vankayalapati H, Hurley LH and Schaefer J (2004) Structure of a quinobenzoxazine–G-quadruplex complex by REDOR NMR. *Biochemistry* 43, 11953–11958.

Mehta AK, Lu K, Childers WS, Liang Y, Dublin SN, Dong J, Snyder JP, Pingali SV, Thiyagarajan P and Lynn DG (2008) Facial symmetry in protein self-assembly. *Journal of the American Chemical Society* 130, 9829–9835.

Merg AD, Touponse G, Genderen EV, Blum TB, Zuo X, Bazrafshan A, Siaw HMH, Mccanna A, Brian Dyer R, Salaita K, Abrahams JP and Conticello VP (2020) Shape-shifting peptide nanomaterials: surface asymmetry enables pH-dependent formation and interconversion of collagen tubes and sheets. *Journal of the American Chemical Society* 142, 19956–19968.

Middleton DA, Madine J, Castelletto V and Hamley IW (2013) Insights into the molecular architecture of a peptide nanotube using FTIR and solid-state NMR spectroscopic measurements on an aligned sample. *Angewandte Chemie International Edition* 52, 10537–10540.

Misra R, Rudnick-Glick S and Adler-Abramovich L (2021) From folding to assembly: functional supramolecular architectures of peptides comprised of non-canonical amino acids. *Macromolecular Bioscience* 21, e2100090.

Moore AN and Hartgerink JD (2017) Self-assembling multidomain peptide nanofibers for delivery of bioactive molecules and tissue regeneration. *Accounts of Chemical Research* 50, 714–722.

Morris RJ and Bricogne G (2003) Sheldrick's 1.2 Å rule and beyond. *Acta Crystallographica Section D: Biological Crystallography* 59 (Pt 3), 615–617.

Morris KL, Zibae S, Chen L, Goedert M, Sikorski P and Serpell LC (2013) The structure of cross- β tapes and tubes formed by an octapeptide, α S β 1. *Angewandte Chemie International Edition* 52, 2279–2283.

Nagy-Smith K, Moore E, Schneider J and Tycko R (2015) Molecular structure of monomorphic peptide fibrils within a kinetically trapped hydrogel network. *Proceedings of the National Academy of Sciences of the United States of America* 112, 9816–9821.

Nagy-Smith K, Beltramo PJ, Moore E, Tycko R, Furst EM and Schneider JP (2017) Molecular, local, and network-level basis for the enhanced stiffness of hydrogel networks formed from coassembled racemic peptides: predictions from Pauling and Corey. *ACS Central Science* 3, 586–597.

Nam KT, Shelby SA, Choi PH, Marciel AB, Chen R, Tan L, Chu TK, Mesch RA, Lee BC, Connolly MD, Kisielowski C and Zuckermann RN (2010) Free-floating ultrathin two-dimensional crystals from sequence-specific peptoid polymers. *Nature Materials* 9, 454–460.

Namba K, Pattanayek R and Stubbs G (1989) Visualization of protein– nucleic acid interactions in a virus: refined structure of intact tobacco mosaic virus at 2.9 Å resolution by X-ray fiber diffraction. *Journal of Molecular Biology* 208, 307–325.

Nelson R, Sawaya MR, Balbirnie M, Madsen AO, Riekel C, Grothe R and Eisenberg D (2005) Structure of the cross-beta spine of amyloid-like fibrils. *Nature* 435, 773–778.

Oakley MG and Kim PS (1998) A buried polar interaction can direct the relative orientation of helices in a coiled coil. *Biochemistry* 37, 12603–12610. Ogihara NL, Weiss MS, Degrado WF and Eisenberg D (1997) The crystal structure of the designed trimeric coiled coil coil-VaLd: implications for engineering crystals and supramolecular assemblies. *Protein Science* 6, 80–88.

Orlova A and Egelman EH (1993) A conformational change in the actin subunit can change the flexibility of the actin filament. *Journal of Molecular Biology* 232, 334–341.

Pandya MJ, Spooner GM, Sunde M, Thorpe JR, Rodger A and Woolfson DN (2000) Sticky-end assembly of a designed peptide fiber provides insight into protein fibrillogenesis. *Biochemistry* 39, 8728–8734.

Papapostolou D, Smith AM, Atkins EDT, Oliver SJ, Ryadnov MG, Serpell LC and Woolfson DN (2007) Engineering nanoscale order into a designed protein fiber. *Proceedings of the National Academy of Sciences of the United States of America* 104, 10853–10858.

Pauling L and Corey RB (1951) The pleated sheet, a new layer configuration of polypeptide chains. *Proceedings of the National Academy of Sciences of the United States of America* 37, 251–256.

Peng Z, Parker AS, Peralta MDR, Ravikumar KM, Cox DL and Toney MD (2017a) High tensile strength of engineered beta-solenoid fibrils via sonication and pulling. *Biophysical Journal* 113, 1945–1955.

Peng Z, Peralta MDR and Toney MD (2017b) Extraordinarily stable amyloid fibrils engineered from structurally defined beta-solenoid proteins. *Biochemistry* 56, 6041–6050.

Peng Z, Peralta MDR, Cox DL and Toney MD (2020) Bottom-up synthesis of protein-based nanomaterials from engineered beta-solenoid proteins. *PLoS One* 15, e0229319.

Peralta MD, Karsai A, Ngo A, Sierra C, Fong KT, Hayre NR, Mirzaee N, Ravikumar KM, Kluber AJ, Chen X, Liu GY, Toney MD, Singh RR and Cox DL (2015) Engineering amyloid fibrils from beta-solenoid proteins for biomaterials applications. *ACS Nano* 9, 449–463.

Peters JW, Stowell MHB and Rees DC (1996) A leucine-rich repeat variant with a novel repetitive protein structural motif. *Nature Structural Biology* 3, 991–994.

Pieri L, Wang F, Arteni AA, Vos M, Winter JM, Le Du MH, Artzner F, Gobeaux F, Legrand P, Boulard Y, Bressanelli S, Egelman EH and Paternostre M (2022) Atomic structure of lanreotide nanotubes revealed by cryo-EM. *Proceedings of the National Academy of Sciences of the United States of America* 119, 4.

Pizzey CL, Pomerantz WC, Sung B-J, Yuwono VM, Gellman SH, Hartgerink JD, Yethiraj A and Abbott NL (2008) Characterization of nanofibers formed by self-assembly of β -peptide oligomers using small angle X-ray scattering. *The Journal of Chemical Physics* 129, 095103.

Pomerantz WC, Abbott NL and Gellman SH (2006) Lyotropic liquid crystals from designed helical β -peptides. *Journal of the American Chemical Society* 128, 8730–8731.

Pomerantz WC, Yuwono VM, Pizzey CL, Hartgerink JD, Abbott NL and Gellman SH (2008) Nanofibers and lyotropic liquid crystals from a class of self-assembling β -peptides. *Angewandte Chemie International Edition* 47, 1241–1244.

Pomerantz WC, Yuwono VM, Drake R, Hartgerink JD, Abbott NL and Gellman SH (2011) Lyotropic liquid crystals formed from ACHC-rich β -peptides. *Journal of the American Chemical Society* 133, 13604–13613.

Potekhin SA, Melnik TN, Popov V, Lanina NF, Vazina AA, Rigler P, Verdini AS, Corradin G and Kajava AV (2001) *De novo* design of fibrils made of short alpha-helical coiled coil peptides. *Chemistry & Biology* 8, 1025–1032.

Pouget E, Fay N, Dujardin E, Jamin N, Berthault P, Perrin L, Pandit A, Rose T, Valéry C, Thomas D, Paternostre M and Artzner F (2010) Elucidation of the self-assembly pathway of lanreotide octapeptide into β -sheet nanotubes: role of two stable intermediates. *Journal of the American Chemical Society* 132, 4230–4241.

Poweleit N, Ge P, Nguyen HH, Loo RRO, Gunsalus RP and Zhou ZH (2016) CryoEM structure of the methanospirillum hungatei archaellum reveals structural features distinct from the bacterial flagellum and type IV pilus. *Nature Microbiology* 2, 16222.

Punjani A, Rubinstein JL, Fleet DJ and Brubaker MA (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* 14, 290–296.

Raeburn J, Zamith Cardoso A and Adams DJ (2013) The importance of the self-assembly process to control mechanical properties of low molecular weight hydrogels. *Chemical Society Reviews* 42, 5143–5156.

Reches M and Gazit E (2003) Casting metal nanowires within discrete selfassembled peptide nanotubes. *Science (New York, N.Y.)* 300, 625–627.

Reiersen H and Rees AR (2001) The hunchback and its neighbours: proline as an environmental modulator. *Trends in Biochemical Sciences* 26, 679–684. Rhys GG, Wood CW, Lang EJM, Mulholland AJ, Brady RL, Thomson AR and Woolfson DN (2018) Maintaining and breaking symmetry in homomeric coiled-coil assemblies. *Nature Communications* 9, 4132.

Rodriguez JA, Ivanova MI, Sawaya MR, Cascio D, Reyes FE, Shi D, Sangwan S, Guenther EL, Johnson LM, Zhang M, Jiang L, Arbing MA, Nannenga BL, Hattne J, Whitelegge J, Brewster AS, Messerschmidt M, Boutet S, Sauter NK, Gonen T and Eisenberg DS (2015) Structure of the toxic core of α -synuclein from invisible crystals. *Nature* 525, 486–490.

Rodriguez JA, Eisenberg DS and Gonen T (2017) Taking the measure of MicroED. *Current Opinion in Structural Biology* 46, 79–86.

Rüter A, Kuczera S, Stenhammar J, Zinn T, Narayanan T and Olsson U (2020) Tube to ribbon transition in a self-assembling model peptide system. *Physical Chemistry Chemical Physics* 22, 18320–18327.

Sachse C (2015) Single-particle based helical reconstruction – how to make the most of real and Fourier space. *AIMS Biophysics* 2, 219–244.

Sawaya MR, Rodriguez J, Cascio D, Collazo MJ, Shi D, Reyes FE, Hattne J, Gonen T and Eisenberg DS (2016) *Ab initio* structure determination from prion nanocrystals at atomic resolution by MicroED. *Proceedings of the National Academy of Sciences of the United States of America* 113, 11232–11236.

Scheres S (2020) Amyloid structure determination in RELION-3.1. *Acta Crystallographica Section D* 76, 94–101.

Scheres SHW, Zhang W, Falcon B and Goedert M (2020) Cryo-EM structures of tau filaments. *Current Opinion in Structural Biology* 64, 17–25. Schmidt A, Annamalai K, Schmidt M, Grigorieff N and Fandrich M (2016)

Cryo-EM reveals the steric zipper structure of a light chain-derived amyloid fibril. *Proceedings of the National Academy of Sciences of the United States of America* 113, 6200–6205.

Schneider JP, Pochan DJ, Ozbas B, Rajagopal K, Pakstis L and Kretsinger J (2002) Responsive hydrogels from the intramolecular folding and self-assembly of a designed peptide. *Journal of the American Chemical Society* 124, 15030–15037.

Selinger RL, Selinger JV, Malanoski AP and Schnur JM (2004) Shape selection in chiral self-assembly. *Physical Review Letters* 93, 158103.

Sharp TH, Bruning M, Mantell J, Sessions RB, Thomson AR, Zaccai NR, Brady RL, Verkade P and Woolfson DN (2012) Cryo-transmission electron microscopy structure of a gigadalton peptide fiber of *de novo* design. *Proceedings of the National Academy of Sciences of the United States of America* 109, 13266–13271.

Sheldrick G (1990) Phase annealing in SHELX-90: direct methods for larger structures. *Acta Crystallographica Section A* 46, 467–473.

- Shen H, Fallas JA, Lynch E, Sheffler W, Parry B, Jannetty N, Decarreau J, Wagenbach M, Vicente JJ, Chen J, Wang L, Dowling Q, Oberdorfer G, Stewart L, Wordeman L, De Yoreo J, Jacobs-Wagner C, Kollman J and Baker D (2018) *De novo* design of self-assembling helical protein filaments. *Science (New York, N.Y.)* 362, 705–709.
- Siemer AB, Ritter C, Ernst M, Riek R and Meier BH (2005) High-resolution solid-state NMR spectroscopy of the prion protein HET-s in its amyloid conformation. *Angewandte Chemie International Edition* 44, 2441–2444.
- Smith AM, Banwell EF, Edwards WR, Pandya MJ and Woolfson DN (2006) Engineering increased stability into self-assembled protein fibers. *Advanced Functional Materials* 16, 1022–1030.
- Song B, Lenhart J, Flegler VJ, Makbul C, Rasmussen T and Böttcher B (2019) Capabilities of the Falcon III detector for single-particle structure determination. *Ultramicroscopy* 203, 145–154.
- Song Q, Cheng Z, Kariuki M, Hall SCL, Hill SK, Rho JY and Perrier S (2021) Molecular self-assembly and supramolecular chemistry of cyclic peptides. *Chemical Reviews* 121, 13936–13995.
- Spaulding CN, Schreiber HLIV, Zheng W, Dodson KW, Hazen JE, Conover MS, Wang F, Svenmarker P, Luna-Rico A, Francetic O, Andersson M, Hultgren S and Egelman EH (2018) Functional role of the type 1 pilus rod structure in mediating host-pathogen interactions. *Elife* 7, e31662.
- Spencer R, Chen KH, Manuel G and Nowick JS (2013) Recipe for β -sheets: foldamers containing amyloidogenic peptide sequences. *European Journal of Organic Chemistry* 2013, 3523–3528.
- Spillantini MG, Schmidt ML, Lee VMY, Trojanowski JQ, Jakes R and Goedert M (1997) α -Synuclein in Lewy bodies. *Nature* 388, 839–840.
- Stubbs G and Kendall A (2012) Helical viruses. *Advances in Experimental Medicine and Biology* 726, 631–658.
- Subramaniam S, Earl LA, Falconieri V, Milne JLS and Egelman EH (2016) Resolution advances in cryo-EM enable application to drug discovery. *Current Opinion in Structural Biology* 41, 194–202.
- Sun J and Zuckermann RN (2013) Peptoid polymers: a highly designable bioinspired material. *ACS Nano* 7, 4715–4732.

Sun L, Young LN, Zhang X, Boudko SP, Fokine A, Zbornik E, Roznowski AP, Molineux IJ, Rossmann MG and Fane BA (2014) Icosahedral bacteriophage PhiX174 forms a tail for DNA transport during infection. *Nature* 505, 432–435.

Sun J, Jiang X, Lund R, Downing KH, Balsara NP and Zuckermann RN (2016) Self-assembly of crystalline nanotubes from monodisperse amphiphilic diblock copolypeptoid tiles. *Proceedings of the National Academy of Sciences of the United States of America* 113, 3954–3959.

Sunde M, Serpell LC, Bartlam M, Fraser PE, Pepys MB and Blake CC (1997) Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *Journal of Molecular Biology* 273, 729–739.

Tarabout C, Roux S, Gobeaux F, Fay N, Pouget E, Meriadec C, Ligeti M, Thomas D, IJsselstijn M, Besselievre F, Buisson DA, Verbavatz JM, Petitjean M, Valery C, Perrin L, Rousseau B, Artzner F, Paternostre M

and Cintrat JC (2011) Control of peptide nanotube diameter by chemical modifications of an aromatic residue involved in a single close contact. *Proceedings of the National Academy of Sciences of the United States of America* 108, 7679–7684.

Tarafder AK, Von K ugelgen A, Mellul AJ, Schulze U, Aarts DGAL and Bharat TAM (2020) Phage liquid crystalline droplets form occlusive sheaths that encapsulate and protect infectious rod-shaped bacteria. *Proceedings of the National Academy of Sciences* 117, 4724–4731.

Tayeb-Fligelman E, Tabachnikov O, Moshe A, Goldshmidt-Tran O, Sawaya MR, Coquelle N, Colletier JP and Landau M (2017) The cytotoxic *Staphylococcus aureus* PSM α 3 reveals a cross- α amyloid-like fibril. *Science (New York, N.Y.)* 355, 831–833.

Thomas F, Dawson WM, Lang EJM, Burton AJ, Bartlett GJ, Rhys GG, Mulholland AJ and Woolfson DN (2018) *De novo*-designed α -helical barrels as receptors for small molecules. *ACS Synthetic Biology* 7, 1808–1816.

Thomson AR, Wood CW, Burton AJ, Bartlett GJ, Sessions RB, Brady RL and Woolfson DN (2014) Computational design of water-soluble α -helical barrels. *Science (New York, N.Y.)* 346, 485–488.

- Tian Y, Polzer FB, Zhang HV, Kiick KL, Saven JG and Pochan DJ (2018) Nanotubes, plates, and needles: pathway-dependent self-assembly of computationally designed peptides. *Biomacromolecules* 19, 4286–4298.
- Urvoas A, Guellouz A, Valerio-Lepiniec M, Graille M, Durand D, Desravines DC, Van Tilbeurgh H, Desmadril M and Minard P (2010) Design, production and molecular structure of a new family of artificial alpha-helicoïdal repeat proteins (α Rep) based on thermostable HEAT-like repeats. *Journal of Molecular Biology* 404, 307–327.
- Usov I, Adamcik J and Mezzenga R (2013) Polymorphism complexity and handedness inversion in serum albumin amyloid fibrils. *ACS Nano* 7, 10465–10474.
- Valéry C, Paternostre M, Robert B, Gulik-Krzywicki T, Narayanan T, Dedieu J-C, Keller G, Torres M-L, Cherif-Cheikh R, Calvo P and Artzner F (2003) Biomimetic organization: octapeptide self-assembly into nanotubes of viral capsid-like dimension. *Proceedings of the National Academy of Sciences of the United States of America* 100, 10258–10262.
- Valéry C, Artzner F, Robert B, Gulick T, Keller G, Grabielle-Madelmont C, Torres ML, Cherif-Cheikh R and Paternostre M (2004) Self-association process of a peptide in solution: from β -sheet filaments to large embedded nanotubes. *Biophysical Journal* 86, 2484–2501.
- Valery C, Deville-Foillard S, Lefebvre C, Taberner N, Legrand P, Meneau F, Meriadec C, Delvaux C, Bizien T, Kasotakis E, Lopez-Iglesias C, Gall A, Bressanelli S, Le Du MH, Paternostre M and Artzner F (2015) Atomic view of the histidine environment stabilizing higher-pH conformations of pH-dependent proteins. *Nature Communications* 6, 7771.
- Van Rijt MMJ, Ciaffoni A, Ianiro A, Moradi MA, Boyle AL, Kros A, Friedrich H, Sommerdijk N and Patterson JP (2019) Designing stable, hierarchical peptide fibers from block co-polypeptide sequences. *Chemical Science* 10, 9001–9008.
- Vauthey S, Santoso S, Gong H, Watson N and Zhang S (2002) Molecular self-assembly of surfactant-like peptides to form nanotubes and nanovesicles. *Proceedings of the National Academy of Sciences of the United States of America* 99, 5355–5360.

Violette A, Averlant-Petit MC, Semetey V, Hemmerlin C, Casimir R, Graff R, Marraud M, Briand JP, Rognan D and Guichard G (2005) *N,N'*-linked oligoureas as foldamers: chain length requirements for helix formation in protic solvent investigated by circular dichroism, NMR spectroscopy, and molecular dynamics. *Journal of the American Chemical Society* 127, 2156–2164.

Walshaw J and Woolfson DN (2001) Open-and-shut cases in coiled-coil assembly: alpha-sheets and alpha-cylinders. *Protein Science* 10, 668–673.

Wang YA, Yu X, Yip C, Strynadka NC and Egelman EH (2006) Structural polymorphism in bacterial EspA filaments revealed by cryo-EM and an improved approach to helical reconstruction. *Structure (London, England: 1993)* 14, 1189–1196.

Wang F, Burrage AM, Postel S, Clark RE, Orlova A, Sundberg EJ, Kearns DB and Egelman EH (2017) A structural model of flagellar filament switching across multiple bacterial species. *Nature Communications* 8, 960. Wang F, Gu Y, O'Brien JP, Yi SM, Yalcin SE, Srikanth V, Shen C, Vu D, Ing NL, Hochbaum AI, Egelman EH and Malvankar NS (2019) Structure of microbial nanowires reveals stacked hemes that transport electrons over micrometers. *Cell* 177, 361–369.e310.

Wang F, Gnewou O, Modlin C, Beltran LC, Xu C, Su Z, Juneja P, Grigoryan G, Egelman EH and Conticello VP (2021a) Structural analysis of cross α -helical nanotubes provides insight into the designability of filamentous peptide nanomaterials. *Nature Communications* 12, 407.

Wang F, Gnewou O, Wang S, Osinski T, Zuo X, Egelman EH and Conticello VP (2021b) Deterministic chaos in the self-assembly of beta sheet nanotubes from an amphipathic oligopeptide. *Matter* 4, 3217–3231.

Wang F, Gnewou O, Solemanifar A, Conticello VP and Egelman EH (2022) Cryo-EM of helical polymers. *Chemical Reviews*. doi: 10.1021/acs.chemrev.1c00753.

Warmack RA, Boyer DR, Zee C-T, Richards LS, Sawaya MR, Cascio D, Gonen T, Eisenberg DS and Clarke SG (2019) Structure of amyloid- β (20–34) with Alzheimer's-associated isomerization at Asp23 reveals a distinct protofilament interface. *Nature Communications* 10, 3357.

- Wlodawer A and Dauter Z (2017) ‘Atomic resolution’: a badly abused term in structural biology. *Acta Crystallographica Section D: Structural Biology* 73 (Pt 4), 379–380.
- Wlodawer A, Li M and Dauter Z (2017) High-resolution cryo-EM maps and models: a crystallographer’s perspective. *Structure (London, England: 1993)* 25, 1589–1597 e1581.
- Woolfson DN (2017) Coiled-coil design: updated and upgraded. *Subcellular Biochemistry* 82, 35–61.
- Wu Y, Norberg PK, Reap EA, Congdon KL, Fries CN, Kelly SH, Sampson JH, Conticello VP and Collier JH (2017) A supramolecular vaccine platform based on alpha-helical peptide nanofibers. *ACS Biomaterials Science & Engineering* 3, 3128–3132.
- Wu X, Ma Y, Zhao K, Zhang J, Sun Y, Li Y, Dong X, Hu H, Liu J, Wang J, Zhang X, Li B, Wang H, Li D, Sun B, Lu J and Liu C (2021) The structure of a minimum amyloid fibril core formed by necroptosis-mediating RHIM of human RIPK3. *Proceedings of the National Academy of Sciences of the United States of America* 118, 14.
- Xu C, Liu R, Mehta AK, Guerrero-Ferreira RC, Wright ER, DuninHorkawicz S, Morris K, Serpell LC, Zuo X, Wall JS and Conticello VP (2013) Rational design of helical nanotubes from self-assembly of coiled-coil lock washers. *Journal of the American Chemical Society* 135, 15565–15578.
- Xu J, Dayan N, Goldbourn A and Xiang Y (2019) Cryo-electron microscopy structure of the filamentous bacteriophage IKE. *Proceedings of the National Academy of Sciences* 116, 5493–5498.
- Xuan S and Zuckermann RN (2020) Diblock copolypeptoids: a review of phase separation, crystallization, self-assembly and biological applications. *Journal of Materials Chemistry B* 8, 5380–5394.
- Yan H, Nykanen A, Ruokolainen J, Farrar D, Gough JE, Saiani A and Miller AF (2008) Thermo-reversible protein fibrillar hydrogels as cell scaffolds. *Faraday Discussions* 139, 71–84; discussion 105–128, 419–120.
- Yoo SH, Collie GW, Mauran L and Guichard G (2020) Formation and modulation of nanotubular assemblies of oligoureia foldamers in aqueous conditions using alcohol additives. *ChemPlusChem* 85, 2243–2250.

Zaccai NR, Chi B, Thomson AR, Boyle AL, Bartlett GJ, Bruning M, Linden N, Sessions RB, Booth PJ, Brady RL and Woolfson DN (2011) A *de novo* peptide hexamer with a mutable channel. *Nature Chemical Biology* 7, 935–941.

Zhang S, Holmes T, Lockshin C and Rich A (1993) Spontaneous assembly of a self-complementary oligopeptide to form a stable macroscopic membrane. *Proceedings of the National Academy of Sciences* 90, 3334–3338.

Zhang S, Lockshin C, Cook R and Rich A (1994) Unusually stable β -sheet formation in an ionic self-complementary oligopeptide. *Biopolymers* 34, 663–672.

Zhang HV, Polzer F, Haider MJ, Tian Y, Villegas JA, Kiick KL, Pochan DJ and Saven JG (2016) Computationally designed peptides for self-assembly of nanostructured lattices. *Science Advances* 2, e1600307.

Zhang SQ, Huang H, Yang J, Kratochvil HT, Lolicato M, Liu Y, Shu X, Liu L and Degrado WF (2018) Designed peptides that assemble into cross- α amyloid-like structures. *Nature Chemical Biology* 14, 870–875.

Zhang W, Falcon B, Murzin AG, Fan J, Crowther RA, Goedert M and Scheres SHW (2019) Heparin-induced tau filaments are polymorphic and differ from those in Alzheimer's and Pick's diseases. *Elife* 8, e43584.

Zhao Y, Wang J, Deng L, Zhou P, Wang S, Wang Y, Xu H and Lu JR (2013) Tuning the self-assembly of short peptides via sequence variations. *Langmuir* 29, 13457–13464.

Zhao Y, Deng L, Wang J, Xu H and Lu JR (2015) Solvent controlled structural transition of KI4K self-assemblies: from nanotubes to nanofibrils. *Langmuir* 31, 12975–12983.

Zhao Y, Yang W, Wang D, Wang J, Li Z, Hu X, King S, Rogers S, Lu JR and Xu H (2018) Controlling the diameters of nanotubes self-assembled from designed peptide bolaphiles. *Small* 14, 1703216.

Zimenkov Y, Conticello VP, Guo L and Thiagarajan P (2004) Rational design of a nanoscale helical scaffold derived from self-assembly of a dimeric coiled coil motif. *Tetrahedron* 60, 7237–7246.

Zimenkov Y, Dublin SN, Ni R, Tu RS, Breedveld V, Apkarian RP and Conticello VP (2006) Rational design of a reversible pH-responsive switch for peptide self-assembly. *Journal of the American Chemical Society* 128, 6770–6771.

Ziserman L, Lee HY, Raghavan SR, Mor A and Danino D (2011) Unraveling the mechanism of nanotube formation by chiral self-assembly of amphiphiles. *Journal of the American Chemical Society* 133, 2511–2517.

Chapter 2 |

Deciphering the rules governing synthetic protein assemblies based on repeat proteins

α Rep tandem repeat protein binders have been shown to assemble into helical filaments and spontaneously crystallize. Here we describe an improved and expanded series of α Reps that form filaments of different symmetries and attempts to stabilize the interior and exterior surfaces of the filament. We demonstrate that minute changes to the shape of the protein affect the unit cell of the crystal and the way the basic building block of the filament, the “Brick” protein, interacts with the accessory proteins. This set of proteins is an important step toward tailoring these remarkable versatile proteins toward nanometer-to-micron scale applications.

Background

Self-assembling helical protein assemblies such as microtubules, pili, and flagella are ubiquitous in living cells. These architectures enable complex functions such as intracellular transport, cell motility, and cell division. The utility and diversity of these structures in nature indicate that these proteins may be designed toward protein-templated functional nanomaterials synthesis (Voet and Tame, 2017), self-assembling nanostructures (Hamley, 2019; Hughes et al., 2019), self-healing or responsive material (Aida et al., 2012; Beloqui and Cortajarena, 2020; Meyers et al., 2008) spatially ordered multi-enzyme cascades (Lee et al., 2012), and biomolecular display for atomically-resolved structure determination by electron microscopy (Yeates et al., 2020). To date, protein assembly design tends to use natural proteins as a starting point. To date, design of ordered assemblies of new synthetic proteins mostly derives from the alteration of natural proteins (Kim et al., 2022; Kobayashi and Arai, 2017; Luo et al., 2016; Zhu et al., 2021). Synthetic protein cages, tubes, and filaments have been created by designing chimeric proteins from two natural proteins, (Kobayashi et al., 2015; Lai et al., 2013; Yeates, 2017) by embedding metal-binding sites (Churchfield and Tezcan, 2019) or small bi-specific molecules, (Carlson et al., 2006) or by using covalent self-splicing intein (Harvey et al., 2018). Importantly, computational protein design methods have reached such an accurate level of prediction that complex self-assembled architectures such

as protein filaments (Shen et al., 2018), arrays (Ben-Sasson et al., 2021), rings (Correnti et al., 2020) or cages (Cannon et al., 2020) can be accurately produced from existing designed protein components.

Nature often accomplishes structural and functional innovation by combining or repeating simple structural modules such as coiled coils, beta- α units, or propeller blades leading to functional protein architectures (Parmeggiani et al., 2015). In the same way, artificial tandem repeat proteins such as DARPins and HEAT repeats offer an ideal platform for protein design because of their modular structure and highly variable surfaces. The α Rep family of designed α -helical tandem repeat proteins has been designed to form dimers (Urvoas et al., 2010), host catalytic sites, and bind gold nanoparticles (Fernandez et al., 2020), quantum dots (Léger et al., 2020), larger proteins (Chevrel et al., 2015; Guellouz et al., 2013), and bind drug molecules. Through rational design and screening, they have recently been co-assembled into large, stable supramolecular assemblies of laterally-associated filaments (Moreaud et al., 2023).

Here, we greatly expand the repertoire of α Rep helical co-assembly proteins, with increased binding affinity and diversity. All proteins in this system use the Brick and Staple motifs described earlier but combine various geometries and linking strategies to produce a wide range of filament and superassembly morphologies. Bricks and staples are synthetic proteins engineered for mutual affinity by directed evolution from synthetic repeat protein libraries. We herein describe the rational design of these proteins and report rapid co-assembly, as determined by SDS-PAGE, cryo-EM, and spectrophotometric techniques.

The framework of the α Rep library was sourced from PBS-HEAT protein, a subset of small HEAT proteins. Two capping domains at each terminus prevent α Reps from binding each other to improve solubility during bacterial expression. The C-terminal cap is original to the PBS-HEAT family. The N-terminal cap was sourced from another TRP, protein A0B7C6. It features a variable surface like the repeats. The number of repeats was variable, with more repeats yielding a more stable protein. A diverse library was constructed using rolling-circle amplification with degenerate codons at variable sites

to yield 3×10^8 clones. This approach yielded proteins that were soluble and could bind other molecules on their concave or convex surfaces.

An α Rep nanotube was designed to organize gold nanoparticles within its lumen for laser fusion into a gold nanowire. Gold nanoparticles were to be bound in the concave surface of a “Brick B” α Rep, and the α Reps assembled nose-to-tail in a helix. Another α Rep dubbed “Backbinder” would “staple” each Brick to its neighbor, creating a robust helix or tube (Moreaud et al., 2023). To design these proteins, a three-repeat “bait” α Rep was designed with a tyrosine appointed to convex helices 2 and 3 based on the observation that tyrosine is enriched on protein recognition domains (Fellouse et al., 2004; Koide and Sidhu, 2009; Shiroishi et al., 2007). The concave face was chosen arbitrarily with charged residues for solubility. Bait was screened against the α Rep library. The hit from this screen, called Backbinder, bound the convex face of the bait ($K_d=68\text{nM}$). The proteins were co-crystallized and the structure was solved. The modularity of TRPs was instrumental in turning these paired proteins into a filament. The two repeats with tyrosines were transposed, and six nonspecific repeats were added between them, to produce “Brick B” protein.

The proteins assembled spontaneously at ambient temperature or 37°C , pH 7.5. After incubation, a visible precipitate forms, which will pellet when centrifuged at $10000 \times g$ for 30 minutes. This pelleting test was used to screen for favorable assembly conditions. Brick B/Backbinder E3 were found to assemble at concentrations as low as $4 \mu\text{M}$ each protein. Negative stain transmission electron microscopy yielded images of free filaments or crystalline super-assemblies, depending on the stain. Molybdate stain showed free filaments; uranyl acetate showed crystals. Under cryo-electron microscopy, Brick B/Backbinder E3 were confirmed to be a crystalline super-assembly.

A Backbinder/GFP fusion protein was designed as proof-of-concept for external functionalization of the filament but did not assemble with Brick B. Simple fusion was not the right strategy for external functionalization. This is likely due to the bulky GFP blocking the ridges-into-grooves interaction that stabilizes the filaments. Therefore this approach to functionalization was abandoned.

The principles used to design this new type of protein origami have proven to be effective but the experimental characterization of the resulting assemblies suggests a new set of questions:

- 1) The residues facing into the lumen of the assembly is not directly involved in the Brick/Backbinder interaction and has so far been considered neutral toward filament stability. Is it possible to design new binding partners on the assembly surface or to stabilize the Brick/Brick interactions from the luminal surface?
- 2) Is it possible to modulate the length of the assembly by using N- or C-terminal Cap proteins that can join the filament, but cannot be further extended?
- 3) Is it possible to functionalize the assembled origami by fusing new domains to the staple proteins?
- 4) What effect does modifying the periodicity of the Backbinder have on the helix?

Results

Stabilizing the assembly by optimizing Brick:Brick binding interface

Previously (Moreaud et al., 2023), Brick B and Backbinder E3 (Backbinder) were found to assemble at concentrations at or above 4 mM. Brick C was designed by improving the Brick-Brick interfaces of Brick B to include more nonpolar interactions (Figure 2.2). To verify Brick C/Backbinder binding and determine a lower concentration limit, the two proteins were incubated together at 8, 4, 2, and 1 mM each for 16 hours at 37°C. After incubation, a cloudy white sediment was observed that pelleted easily upon centrifugation at 10000×g for 30 minutes. Analysis of mix, supernatant, and pellet fractions by SDS-PAGE revealed that both proteins were present in the pellet, which indicates that pellets are composed of co-assembled protein (Figure 2.3a).

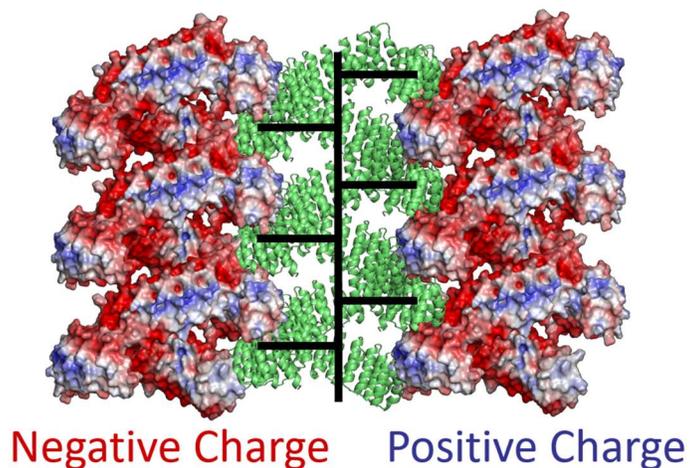


Figure 2.1. Three Brick B/Backbinder filaments in a crystalline arrangement. Surface charges of Brick B and Backbinder coassembly are thought to further stabilize ridges-into-grooves packing, leading to crystal formation. Negative charges (red) line the “groove” created by Backbinders and Bricks on successive turns of the helix. The positively-charged “ridge” along the convex surface of Backbinder fits into this groove and makes contact with the convex surface of the Brick. Figure is reused from Moreaud, et al., 2023.

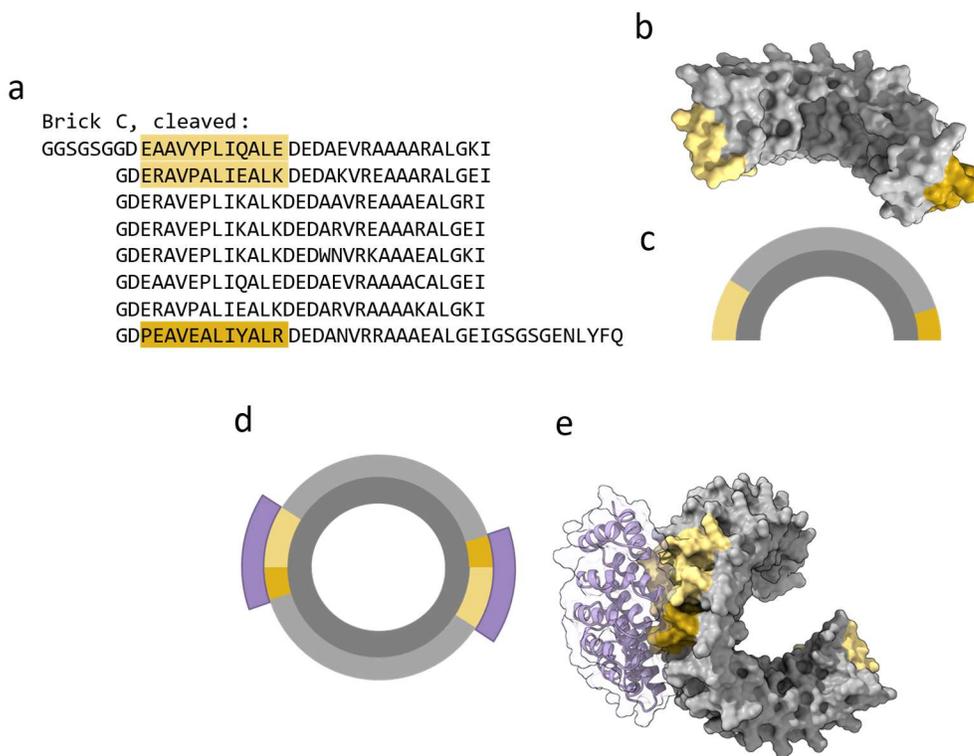


Figure 2.2. A. Sequence of Brick C, arranged to show repeats. I1 and I2 from the Bait protein are highlighted in gold, and I3 is highlighted in orange. B. Model of Brick C showing the highlighted repeats. C,D. Schematic model of Brick C showing how Backbinder (purple) recognizes the reconstituted binding site. E. Model of Backbinder bound to two Brick C proteins.

Under TEM, nanotubes from the 8 μ M pellet fraction appear to be free and packed into a crystalline bundle. Crystals have a jagged appearance at the edges (Figure 2.3b,c). Importantly, analysis of the supernatant fraction showed that it was composed of nonspecifically aggregated protein. This confirms that for this system, the presence of a pellet after centrifugation is an appropriate quick test for specific association.

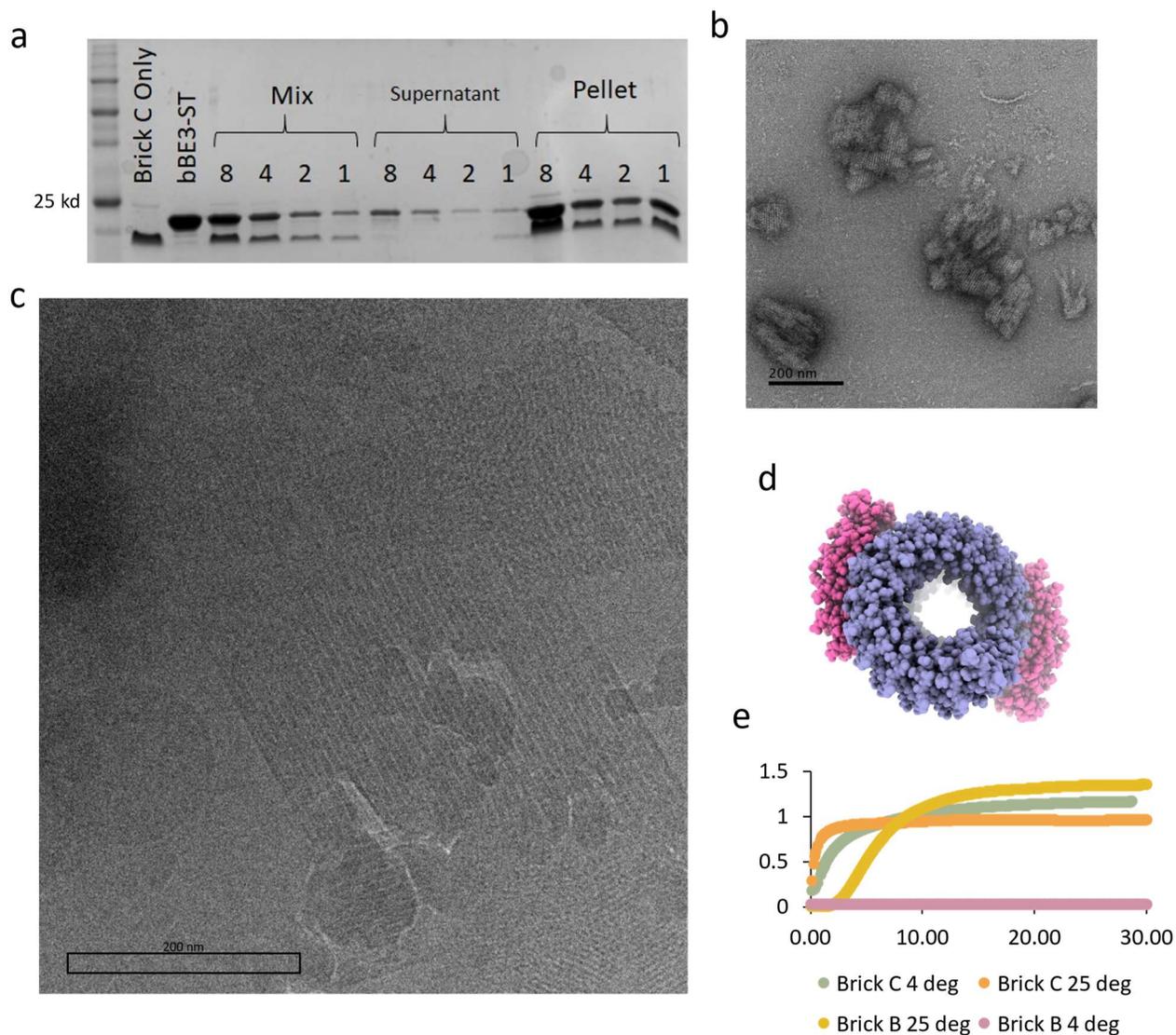


Figure 2.3 A. Brick C/BBST assembly gel showing both Brick C and back binder in the pellet fraction after assembly at concentrations of 8, 4, 2, or 1 μM each protein. B. Negative-stain TEM micrograph showing crystalline superassemblies of Brick C and BBST. C. Cryo-EM micrograph confirming crystalline arrangement of proteins in B. D. Model of Brick C (lavender) and BBST (pink) coassembly. E. Rate of crystallization monitored by absorbance at 350 nm for Brick B and Brick C, at two temperatures. Brick C is an improved version of Brick B with more nonpolar contacts at the Brick:Brick interface, and assembles faster than Brick B with less temperature dependence.

Direct assembly functionalization via SpyCatcher/SpyTag

A Backbinder/GFP fusion protein was designed as proof-of-concept for external functionalization of the filament but did not assemble with Brick B, indicating that simple fusion was not the right strategy for external functionalization. This is likely due to the bulky GFP blocking the ridges-into-grooves interaction that stabilizes the filaments. Therefore this approach to functionalization was abandoned.

A SpyTag/SpyCatcher system was designed to link GFP to the surface of formed nanotubes. The much smaller 13-residue SpyTag sequence was appended to the C terminus of Backbinder so that it would not interfere with nanotube formation. SpyCatcher was appended to the C terminus of GFP superfolder (sfGFP-SC).

Initially three experiments were designed to test conditions that would result in decorated nanotubes: In the first experiment, modeled after ST-HEAT nanotube decoration, nanotubes were formed first by incubating 8 μM each Brick C and Backbinder-ST overnight at 37°C, and then incubated with GFPsf-SC at RT. No pellet was observed. Combining all three components into a single tube at 5.3 μM and incubating overnight at 37°C likewise failed to produce a pellet and only nonspecifically aggregated protein was observed under TEM (SI Figure 2-7). Adding Brick C to linked Backbinder-ST-SC-GFPsf also failed to produce a pellet or nanotubes. Taken together, these experiments suggest that no combination of Brick C, Backbinder-ST, and SC-GFPsf will produce decorated nanotubes.

Table 2-1. Brick C/back binder assembly strategies.

Strategy	First Step	Second Step	Assembly Condition	Result
Add Brick C to decorated Backbinder	Incubate 8 μM Backbinder-ST, excess GFP-SC	Add Brick C to 8 μM	O/N, 37°C	No filaments or crystals observed
Add GFP to assembled nanotubes	Incubate 8 μM each Brick C and Backbinder-ST	Add excess GFP-SC	O/N, 37°C	No filaments or crystals observed

One-pot	Mix 5.3 μM Brick C with excess GFP-SC, add 5.3 μM Backbinder-ST	n/a	O/N, 37°C	No filaments or crystals observed
---------	---	-----	-----------	-----------------------------------

It is likely that adjacent bound GFP proteins abolish the assembly by clashing with each other. To space them farther away from each other along the nanotube, Backbinder-ST was mixed with “Backbinder classic,” which has no SpyTag. Backbinder-ST was doped into the assembly mixture at 1%, 10%, 20%, 30%, 40% of total Backbinder protein. 8 μM total Backbinder and 8 μM Brick C were incubated overnight at 37°C. The next day GFP was added to 16 μM total concentration. After 90 minutes incubation, samples were centrifuged 10000 $\times g$ for 30 minutes. Crystals did form, but GFP was not associated with them, and free filaments were not observed via TEM (SI Figure 2-7).

Kinetics of Brick C crystallization

Earlier work on αRep tubular assemblies was performed by Laureen Moreaud at the University of Toulouse in southern France. She observed that filament formation and precipitation were rapid at ambient conditions. Meanwhile, near Paris in the much colder north of the country, the same process was slow and unpredictable. It was determined that nucleation events, and possibly elongation, were temperature-dependent. Thereafter, αReps were assembled at 37°C. The nature of temperature dependence at different phases and binding affinities was still unknown.

Based on the observation that the large crystals scatter light, optical light transmittance was used to monitor the assembly in real time. The assembly was known to be highly sensitive to temperature, so a temperature-controlled spectrophotometer was used. Two temperature-controlled water baths are set to 10°C or 25°C. A pump in each bath pushes water through insulated tubes, which connect to a single tube at a switch, which toggles between each water source. The single tube runs into the spectrophotometer itself, where the water warms a quartz cuvette.

In preliminary experiments, 4 μM each chilled Brick C and Backbinder were placed in a cuvette chilled to 10°C. No signal change was observed for 10 minutes so the temperature was switched to 25°C.

Within several minutes the signal began to climb in a linear curve up to about Abs=0.1. This signal then seemed to switch to a curve that reached a maximum about Abs=0.24. The temperature was switched to 10°C to test for temperature-dependent reversibility. The equilibrium Abs appeared to be stable or only slightly diminished at 10°C. Thinking that the linear and curved phases of the graph may correspond to phases of filament or crystal formation, the temperature switch was repeated in the linear phase. The signal still progressed to the curved phase, indicating that linear-phase assembly is not reversible. However, the rate of increase was markedly slower at the cooler temperature. This is consistent with previous observations (*vide supra*).

Curves were analyzed again at 10 μ M, 4°C and 25°C. At 4°C, the signal rapidly increased, approaching a maximum signal of about 1.2. At 25°C, the signal increased even more rapidly, reaching a plateau at about 0.9 within three minutes. Although the concentrations of Brick C and Backbinder were both 10 μ M, the signal plateau occurs at a different point in the Abs/time graph. This may be related to the size of the crystals that form. When high temperature and concentration produce many nucleation events, high subunit consumption promotes small crystals, which may be less refractive than those that grow from less favorable conditions.

Changing morphology of the Brick to alter assembly morphology

The ridges-into-grooves packing and crystallization observed in the Brick B and C/ Backbinder E3 assembly was interesting but not desirable. For the filament surface to be functionalizable, filaments must be soluble without crystallizing or aggregating. The first attempt to prevent crystal formation was decoration itself. It was thought that GFP bound to the Backbinder would block the next filament from aligning as described above. Surprisingly, GFP decoration proved to be incompatible with filament formation in the first place.

A new design was needed to reduce filament aggregation. Considering that the Backbinder proteins are 180° from each other when bound to the 8-repeat Brick C, a 7-repeat protein named Brick D was

designed to stagger Backbinder around the filament 150° apart. The Brick D protein retained the tyrosine-rich C- and N-terminal repeats and a cysteine on the concave face for functionalization or Alexa dye.

Protein was expressed, cleaved, and purified as described above. Brick D was mixed with Backbinder at 8, 4, 2, and 1 μM and incubated overnight at 37°C. SDS-PAGE as above showed that the two proteins formed a sediment at concentrations as low as 2 μM . The capping motif at either terminus and the His tag were cleaved off via TEV proteolysis. After a single round of cleavage at 30°C, 2.5h, cleaved protein was purified by passing through NiNTA resin one time, which yielded cleaved Brick D with some contaminating capped Brick D. This was assembled with “Backbinder classic,” which does not include SpyTag. The two proteins were mixed at 8 μm , 4 μm , 2 μm , and 1 μm and incubated overnight at 37°C for assembly. Unlike Brick C, these mixtures did not become cloudy and did not pellet when centrifuged. TEM microscopy revealed short, poorly formed filaments.

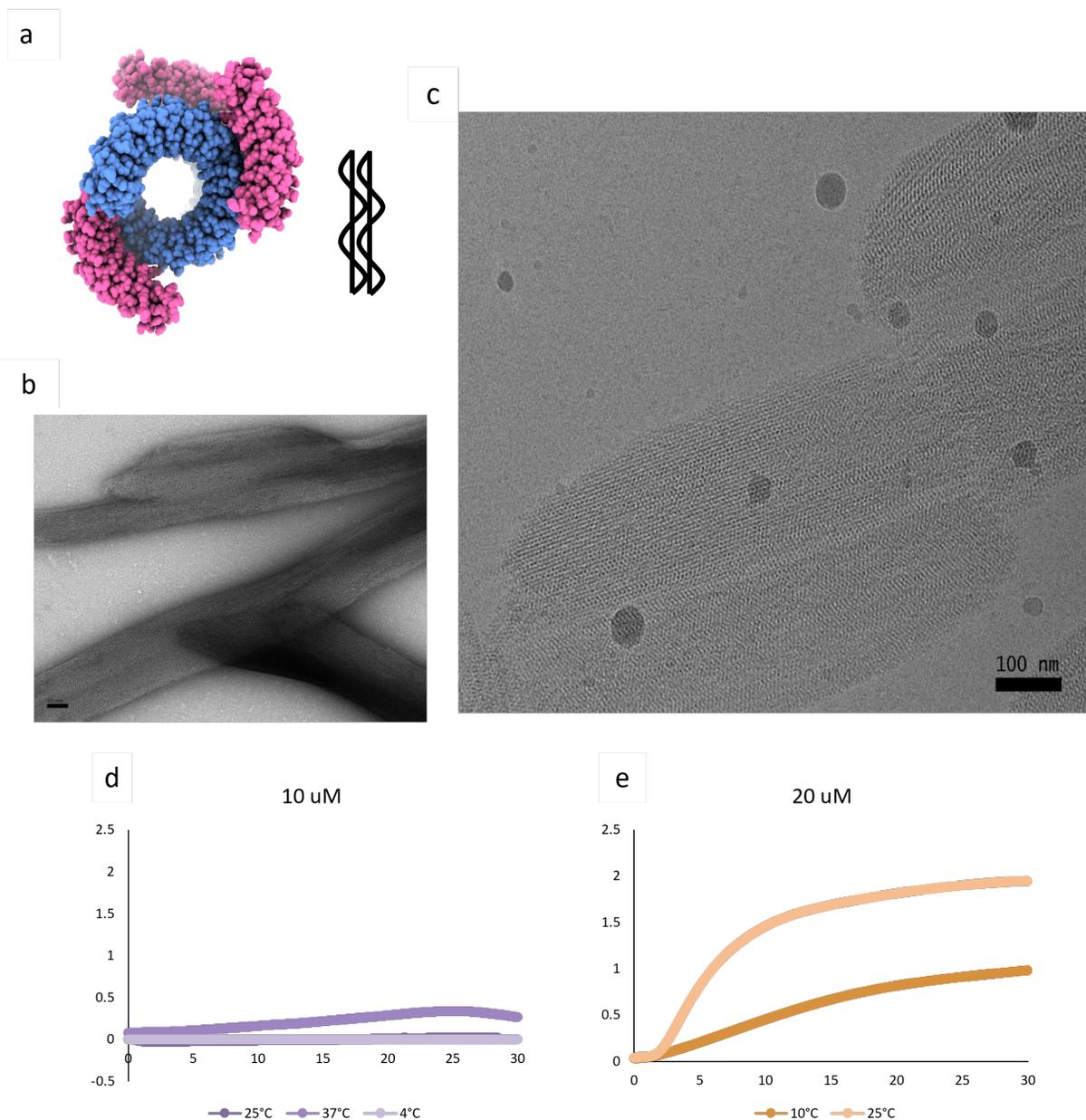


Figure 2.4 A. Model of Brick D (blue) assembled with Backbinder (pink) with cartoon showing screw symmetry. B. Negative-stain TEM image of Brick D/Backbinder crystals. Curving and cracking under the vacuum of the electron microscope is normal for nanoparticles that are flat in solution. C. Cryo-EM image of Brick D/back binder co-crystal with distinct moiré patterning. The crystal is flat in its hydrated state. D and E. Assembly curves of Brick D and Backbinder at 10 or 20 μM concentration showing temperature and concentration dependence of assembly rate.

I hypothesized that the poor quality of the filaments was consistent with incompatibility between the Backbinder and Brick D. I prepared to analyze them by analytical size exclusion chromatography to verify that both the Brick D and Backbinder proteins were present in the assembly. I concentrated each protein to 1.3 mM for the high-concentration experiment and assembled them overnight at 37°C. The next morning there was an abundant white sediment. The proteins had assembled into crystals at the extremely high concentration. Even as low as 30 μ M, a sediment formed within seconds at ambient temperature. According to SDS-PAGE, the two proteins assembled together. TEM revealed long, needle-like crystalline superassemblies: a completely different morphology than the Brick B and C crystals (Figure 2.4b,c). Although Brick D was designed to form soluble filaments, the filaments simply packed together differently. Under TEM, they appear as high aspect ratio curled sheets that collapse on the grid.

In the supernatant fraction of the assembly mixture, short filaments under 75 nm were visible, with no crystals. This result indicated that Brick D assembles as short, soluble filaments that must join a crystal after they reach a certain length, but cannot join if they never grow long enough.

Taking together that contamination with capped Brick D produces short soluble filaments, and that filaments pack into a crystalline super-assembly with a distinct morphology, a model of assembly emerges. Brick D self-associates into filaments “stapled” together by Backbinder. These filaments have a helical symmetry due to the offset of the Backbinder along the axis. At a certain length, the filaments begin to align with each other, probably wrapping around each other like a coiled-coil or double helix due to their helical symmetry. At a longer length, they irreversibly crystallize.

In real-time crystallization studies as above, Brick D reached peak absorbance in 25 minutes at 37°C, 10 μ M protein, but was very slow to assemble at 25°C or 4°C (Figure 2.4d). However, when the concentration was increased to 20 μ M, the assembly appeared to approach its peak turbidity by 30 minutes at 25°C (Figure 2.4e). Assembly was slower at 10°C. Taken together, these rate studies indicate that Brick D/Backbinder crystallization is more concentration- and temperature-dependent than Brick C, which is consistent with Brick D and Backbinder proteins forming short filaments that may join a crystal at a sufficient length, or remain soluble.

Decorating Brick C/Backbinder-ST nanotubes with GFP-SC was not achieved (*vide supra*). It was hypothesized that neighboring linked GFP proteins could not fit on the surface of the nanotube without clashing, thereby abolishing the assembly. As a remedy, the 7-repeat Brick D/Backbinder-ST assembly features staggered SpyTags that should allow each bound GFP more space, thus stabilizing the assembly.

Decorated assemblies were attempted in the same manner as Brick C but were not successful.

Strategy	First Step	Second Step	Assembly Condition	Result
Add Brick D to decorated Backbinder	Incubate 8 μ M Backbinder-ST, excess GFP-SC	Add Brick D to 8 μ M	RT, 2 hrs	Short, unstable filaments
Add GFP to assembled nanotubes	Incubate 8 μ M each Brick D and Backbinder-ST	Add excess GFP-SC	RT, overnight	Long (>1 μ M), thick filaments
One-pot	Mix 8 μ M Brick D with excess GFP-SC, add 8 μ M Backbinder-ST	n/a	RT, overnight	No or short filaments

Thermostability

The proteins were known to be thermostable because they are derived from thermostable archaea (Urvoas et al., 2010). Furthermore, unfolded α Reps cannot aggregate when buried hydrophobic residues are exposed because they are not globular proteins with a hydrophobic core. It was previously determined that a greater number of α Rep repeats is correlated with greater thermal stability (Urvoas et al., 2010). To determine whether assembly confers additional thermostability, circular dichroism spectropolarimetry was used to analyze thermal stability of single and co-assembled proteins Bricks C and D and Backbinder, as well as an optically clear assembly of 5% C, 95% D, and Backbinder.

As shown in Figure 2.5, the proteins described here follow the expected trend. At room temperature (Figure 2.5a), all proteins exhibit a pronounced α -helical signature. At 95°C, all signatures are degraded (Figure 2.5b). Bricks C and D did not show a substantial melting transition temperature, but Backbinder did, which was consistent with expected trends in thermostability (Figure 2.5c). An optically clear mixture of 5% C/95% D/BBST was also melted and showed a melt transition characteristic of

Backbinder. Brick C and D with Backbinder were incubated at 95°C then cooled 1°C per minute to 20°C to determine the effect of high heat on these proteins. Both samples produced filaments and crystals indistinguishable from those incubated at 37°C. Taken together, these results indicate that the less-thermostable Backbinder loses some secondary structure at high temperatures but does not aggregate and can refold upon cooling. Turbid samples of assembled protein have been observed to clarify when incubated at 99°C, then re-assemble and become turbid again after cooling to room temperature.

Stabilizing the interior of the filament

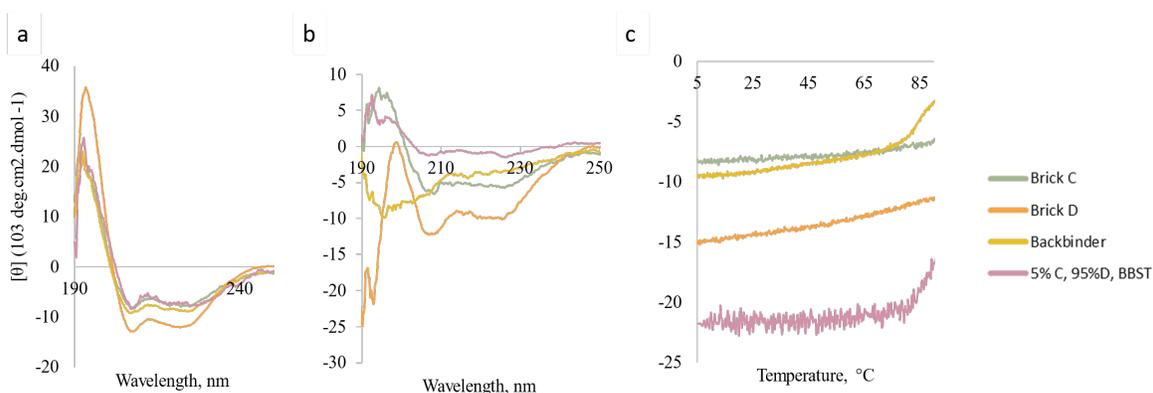


Figure 2.5. Thermostability monitored by circular dichroism spectropolarimetry. A. All α Reps have a distinctive α helical signature. B. At 95°C, there is some helicity but signatures are generally degraded. C Thermal melt curves from 4 to 95°C. Bricks C and D do not have a thermal transition, but Backbinder begins to melt around 70°C. Backbinder melts at the same temperature whether soluble or in an assembly.

Brick F¹ was designed by grafting the concave binding surface of A3 (Guellouz et al., 2013), which binds the small α Rep α 2, onto the termini of the concave face of Brick C, so that the two Brick F proteins together reconstitute the A3/ α 2 binding site. The convex face is identical to Brick C (Figure 2.8. SDS-

¹ Readers may notice that we skipped straight from Brick D to F. Brick E was an imagined 9-repeat protein, the gene for which was too onerous to construct for what probably would have behaved like Brick D, but with fewer stabilizing interactions per length.

PAGE gel showing all constituent proteins in pellet fractions. $\alpha 2$ is difficult to see in the Brick F/ $\alpha 2$ /Backbinder pellet sample. Figure 2.8). According to the interface measurement tool PISA (Krissinel and Henrick, 2007) and our model generated in Rosetta (Kleffner et al., 2017), each $\alpha 2$ contacts three Brick F proteins: The $\alpha 2$ shown in red makes contact with one Brick F shown in green in Figure 2.7, another shown in lavender, and the third shown in blue. The two $\alpha 2$ proteins also share an interface. The filament was further stabilized by Backbinders on the convex face of Brick F, with each Backbinder:Brick F interface expected to be very favorable. Because the filament included the protruding Backbinder, it was expected to form crystals like those seen in Brick C/BBST.

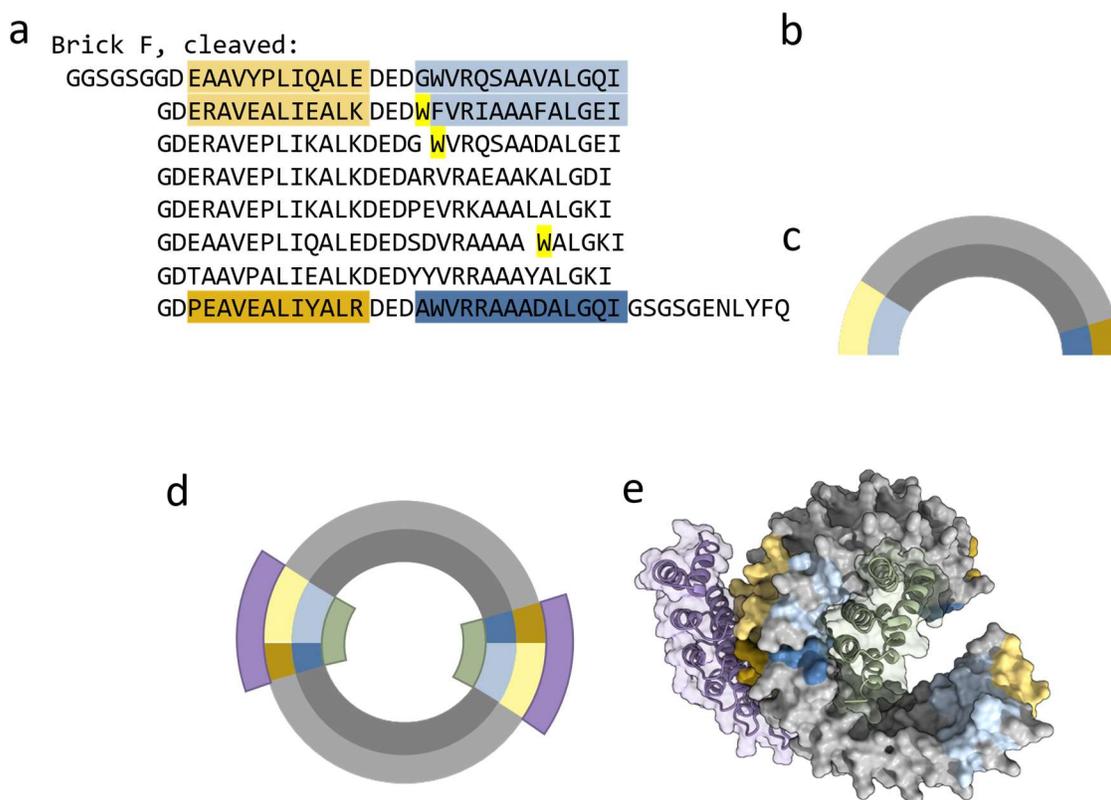


Figure 2.6. A. Sequence of Brick F, arranged to show repeats. I1 and I2 from the Bait protein are highlighted in gold, and I3 is highlighted in orange. Repeats from A3 protein are highlighted in blue. B. Model of Brick F showing the highlighted repeats on the concave and convex faces of the protein. C,D. Schematic models of Brick F showing how Backbinder (purple) and $\alpha 2$ recognize their reconstituted

binding sites. E. Model of Backbinder and $\alpha 2$ bound to two Brick F proteins.

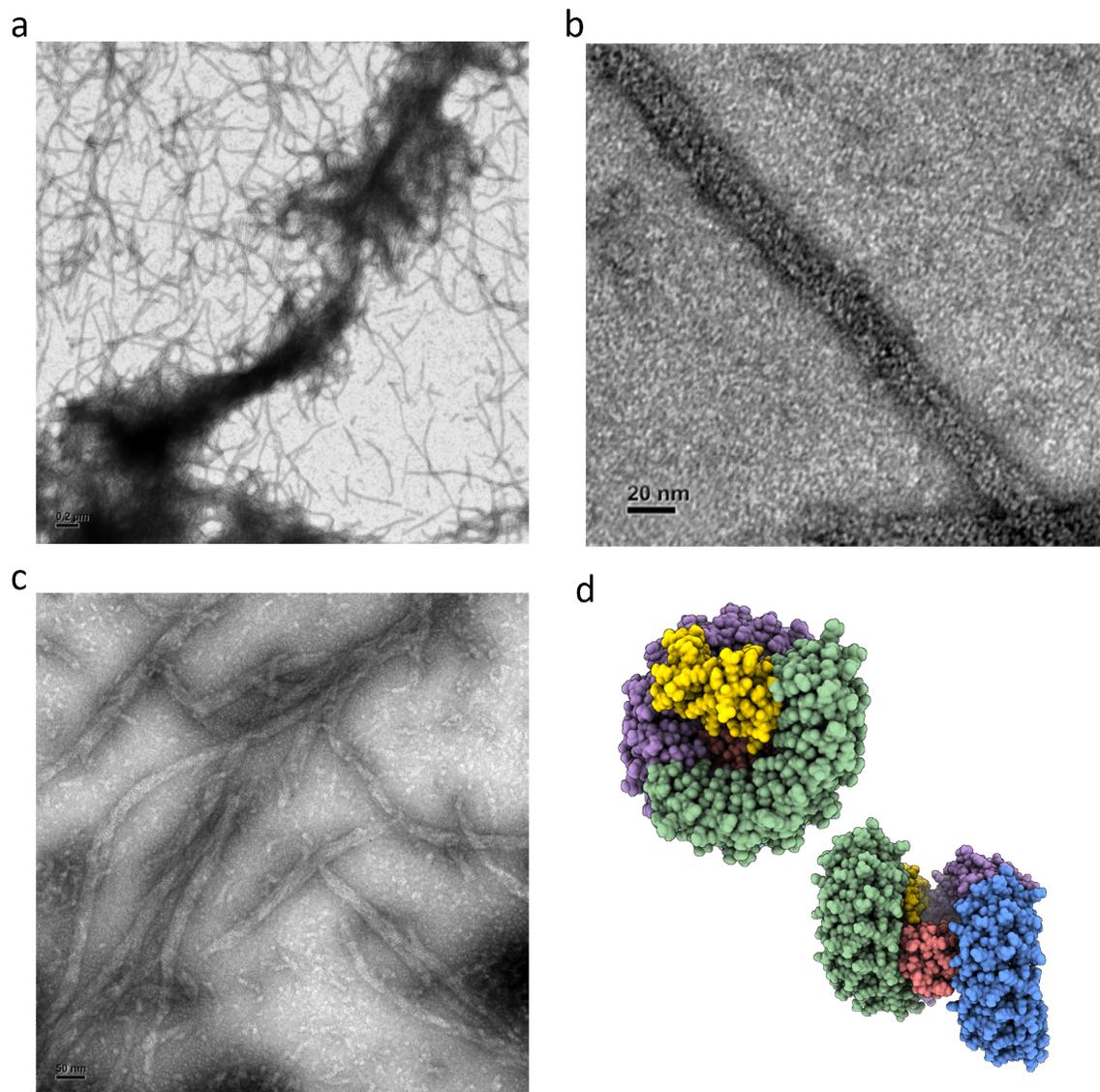


Figure 2.7 A. Very long, twisted filaments of Brick F and $\alpha 2$ only. B. Close-up from same sample as A showing two or three filaments in each bundle. C. Tape-like crystals of Brick F and $\alpha 2$. D. Model of Brick F (blue, violet, green) assembled with only $\alpha 2$ (yellow, red).

Tested combinations of proteins included Brick F, $\alpha 2$, and BBST; Brick F and $\alpha 2$ only; Brick F and BBST only; and Brick C, $\alpha 2$, and BBST as a control. All proteins were mixed at a final concentration of 8 μM and imaged at 4 μM . $\alpha 2$ and BBST were mixed before addition of Brick F. All combinations except

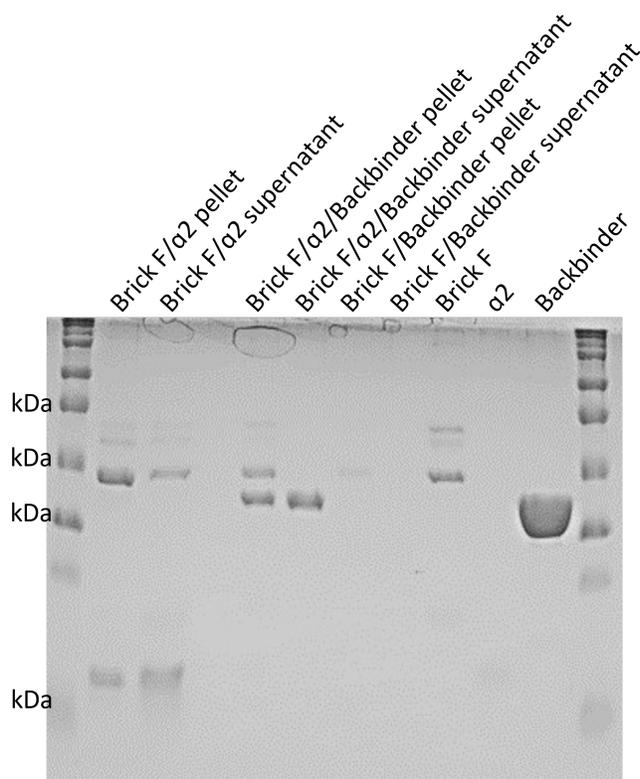


Figure 2.8. SDS-PAGE gel showing all constituent proteins in pellet fractions. $\alpha 2$ is difficult to see in the Brick F/ $\alpha 2$ /Backbinder pellet sample.

Brick F/ $\alpha 2$ /BBST became turbid as soon as Brick F was added. SDS-PAGE indicates that pellets from turbid samples contain all constituent proteins (Figure 2.8). In negative-stain microscopy, Brick F/ $\alpha 2$ /BBST assembles as free filaments roughly 100-200 nm long, or as shaggy aggregates of aligned filaments, but filaments do not form ordered crystals. Stain can disrupt crystals, so the sample was imaged by cryo-TEM. Again, no crystals are found, but the filaments occur as a condensate of flexible filaments (Figure 2.9). Brick F/BBST appeared as free filaments ranging from very short (<10 nm) assemblies to about 100 nm in length, or crystals identical to Brick C/Backbinder crystals (Figure 2.9a). Brick F/ $\alpha 2$ produced flexible free filaments with an average width of 16.6 nm. Each single Brick F/ $\alpha 2$ filament is

expected to be about 6.3 nm thick according to the model, so this corresponds to 2 or 3 filaments intertwined. High-aspect-ratio tapes over 1 μm long are also observed (Figure 2.7c).

Taken together, these results indicate that: 1) $\alpha 2$ incorporates into the Brick F/BBST assembly; 2) $\alpha 2$ and Brick F are sufficient to form a stable assembly, further stabilized by lateral interactions; 3) all laterally associated filaments are longer than their single counterparts; 4) Back binder does not need to be evenly distributed along the filament because $\alpha 2$ can stabilize independently, so the surface of the filament is not uniform enough to promote crystallization although the surface charges still permit lateral association. Most importantly, 5) both the concave and convex surfaces of αReps are highly variable and a single αRep can act as a Janus particle, binding one partner on each side.

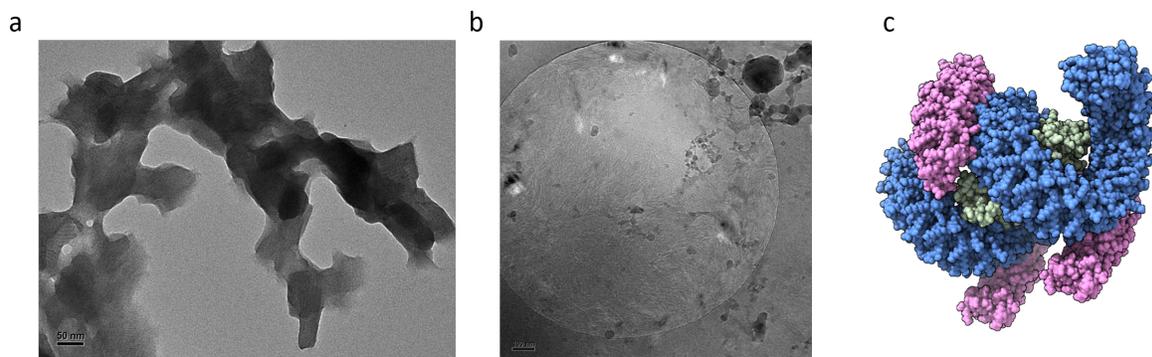


Figure 2.9. A. Negative stain TEM image of Brick F/Backbinder crystals, identical to Brick C/Backbinder. B. Cryo-EM image of Brick F/Backbinder/ $\alpha 2$ condensate with a few free filaments in the thin ice patch toward the top of the hole. C. Model of Brick F (blue), Backbinder (pink), and $\alpha 2$ (green) complex.

Stabilizing the exterior of the filament

The Minard lab had previously (Urvoas et al., 2010) developed an αRep named A3 into a multi-domain protein switch (Di Meo et al., 2017; Léger et al., 2019). A3 exists in solution as a dimer resembling a clamshell, with each curved monomer interacting via the terminal repeats, leaving the concave surface solvent-accessible. Two A3 proteins may be linked via a glycine-serine linker. To both

stabilize the Brick/Backbinder filament and prevent filament aggregation, Backbinder was linked to A3 so that once assembled with a Brick, A3 will dimerize to lock the Backbinder moieties together. The dimerized ligand may also be bulky and irregular enough to inhibit crystalline packing. These effects combined were expected to produce longer, more soluble filaments than Brick/Backbinder alone.

Brick C was incubated together with A3-E3 at a final concentration of 8 μM , 4 μM , 2 μM , and 1 μM each protein overnight at 37°C. No pellet was visible, indicating that any assembly was not packing into crystalline particles. At 8 μM , the TEM grid stained with 2% UA was highly occluded and the only discernable features short and shaggy in appearance. At 4 μM , filaments were visible and were seen to be loosely aggregated into bundles 400 μM long (Figure 2.10a). At 2 μM , filaments were not visible. Brick F was identical to Brick C (Figure 2.10**Error! Reference source not found.**b).

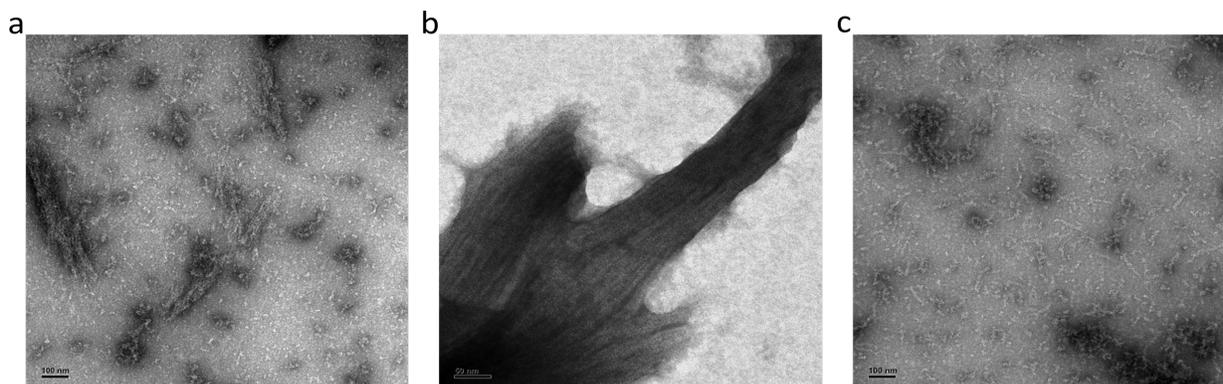


Figure 2.10. A. Brick C assembled with A3E3. Filaments form and join disordered, somewhat linear aggregates. B. Brick F and A3E3. Identical to Brick C. C. Short, irregular assemblies of Brick D and A3E3.

The same experiments were performed with Brick D (Figure 2.10c). While some filaments did form at 8 μM , imaging was poor due to crowding on the grid. At 2 μM , protein was aggregated but no filaments were visible. When A3-E3 was mixed 50/50 with Backbinder, filaments appeared to be crosslinked at 8 μM and were not seen at 2 μM .

Solubilizing filaments

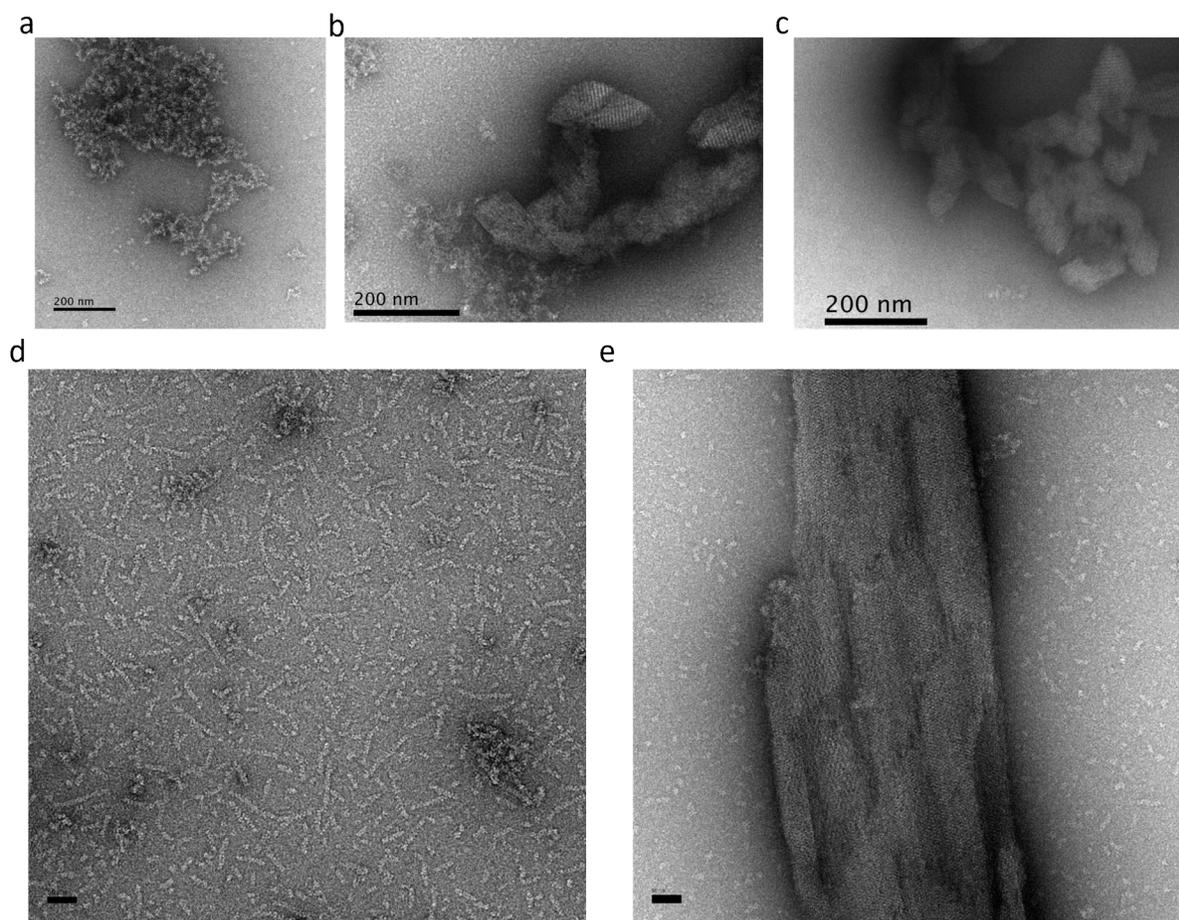


Figure 2.11. A. 1:4 Caps to Brick C ratio abolishes filament formation and produces disordered aggregates of free protein. 1:16 (B) and 1:80 (C) ratios of Caps to Brick C both produce crystals, indicating that crystals can form regardless of filament length. No stoichiometric relationship is observed. However, Brick D filaments were solubilized by a 1:16 Caps:Brick D ratio, with no crystals observed (D) (scale bar = 50 nm). 1:32 Caps:Brick D produced crystals (E) (scale bar = 50 nm).

Capping

Based on the observation that Brick D/Backbinder filaments under a certain length do not aggregate into a crystal, and only short filaments are seen in supernatant, several methods for shortening

the filaments were attempted: randomness, capping, and high heat. Each of these uses a critical point in the crystallization process to prevent crystals from forming: crystals require substituents to be highly regular; short Brick D/Backbinder crystals do not associate durably; high heat encourages many nucleation events, such that the monomer is consumed before the filaments are sufficiently long for crystalline arrangement.

0.1 μM (1:80 Caps:Bricks), 0.25 μM (1:32 Caps:Bricks), and 0.5 μM (1:80 Caps:Bricks), total Brick N End and Brick C End (*vide supra*) were combined with 8 μM Brick D/Backbinder and incubated overnight at 37°C. After centrifuging for 5 minutes the next day, there was no visible pellet in the 0.5 μM Caps sample, and pellets in the 0.25 μM Caps sample, and the 0.1 μM Caps sample, indicating that Brick D crystals were very small or nonexistent in the 0.5 μM Caps sample.

The 0.1 μM Caps and 0.25 μM Caps samples did have some large Brick D crystals identical to those observed in samples with no Caps. Large, high-aspect ratio crystals were visible along with loose protein. The 0.25 μM Caps crystals had a “layered” appearance (Figure 2.11e). However, the 0.5 μM Caps sample produced soluble filaments up to about 100 nm in length (Figure 2.11d). No crystals were visible on the grid.

Mixing Bricks

Crystal packing depends entirely on regularity and periodicity of the subunits. Subunits of random shapes cannot pack into a crystal. Based on these facts it was hypothesized that mixing the two Bricks C and D together would produce a nanotube with protruding Backbinder randomly scattered across the surface, frustrating crystallization and rendering the nanotubes soluble.

8 μM 50% Brick C/50% Brick D/Backbinder yielded crystals consistent with those of Brick C. This is consistent with the earlier observation that Brick C rapidly crystallizes, which could have been expected at 4 μM Brick C. Hypothesizing that the mixture was too C-rich to assemble the proteins together, I assembled mixes of 8 μM 5% Brick C/95% Brick D/Backbinder and 8 μM 20% Brick C/80% Brick D/Backbinder.

At 20% Brick C, many short filaments were observed, and one crystal was observed on the entire grid. At 5% Brick C, many free filaments were observed. There was too much protein to discern individual filaments so the solution was diluted to 4 μM (Figure 2.12) and 0.8 μM . Importantly, the 0.8 μM grid had very tiny clusters of protein but no filaments. These observations are consistent with the single or double filaments being dynamic and concentration-dependent.

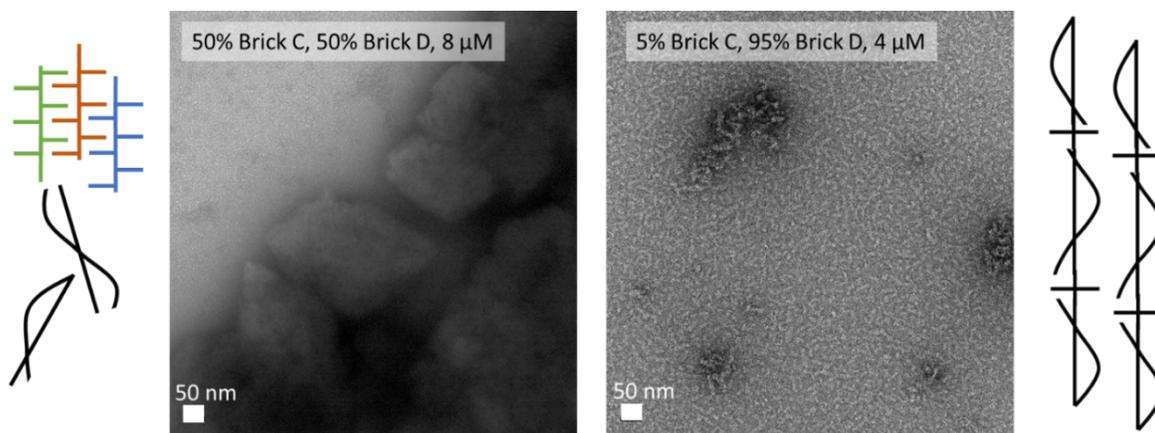


Figure 2.12. Mixing 7-repeat Brick D with 8-repeat Brick C to introduce randomness to the Brick/back binder filament and prevent crystallization. Mixing equal amounts Brick C and Brick D (left) failed to prevent crystallization because the Brick C/back binder crystal forms very rapidly. However, adding 5% Brick C to 95% Brick D (right) prevents crystals from forming and produces a high concentration of short, free filaments.

Discussion

The principle of deterministic chaos, i.e., that protein structure is determined by so many factors that exact structure prediction is impossible, is even more true of protein assemblies (Wang et al., 2021a). Here, an attempt to strengthen an interface created not soluble filaments, but stronger sheets. An attempt to solubilize the filament by changing filament geometry created a new crystal and illustrated the

stabilizing role of lateral interactions. Strengthening the filament core allowed new flexibility incompatible with extensive lateral packing, but instead invited condensation.

One of the most important differences between Brick C and Brick D is the role of lateral interactions in stabilizing the filament. Because Backbinder E3 proteins sit across from each other on the nanotube surface like teeth of a zipper, they can begin to intercalate with Backbinder proteins on an adjacent tube when each assembly is still very small, theoretically just two turns of two filaments. This ridges-into-grooves packing is further encouraged by the opposite electrostatic charges on the two faces (Figure 2.1). This process is temperature-dependent but rapid even at cold temperature. The other side of the tube forms the same interaction and a crystalline assembly rapidly grows. The crystal is very stable and can be rinsed without dissolving. However, Brick D filaments have a spiraling ridge of Backbinder along their surface. They may wrap around each other in pairs. These filaments have a much longer pitch than Brick C. Below a critical length, the filaments or pairs of filaments are soluble and dynamic. Their length is concentration-dependent. Once a critical length has been achieved (about 50 nm), and if there is sufficient thermal energy, the filaments rapidly pack into crystalline sheets. These sheets are irreversible and can be rinsed or chilled. Crystals persist when heated to 80°C for 30 minutes.

At the beginning of our investigation, precipitation served as a rapid test for protein assembly. For Bricks B and C, it was accurate to infer that Brick C precipitated at 1 μM and Brick B did not because the inter-Brick interface had been optimized. This was reflected in the speed of crystallization. However, the fact that Brick D does not precipitate below 4 μM is not due to this interface, which is identical to Brick C. At this low concentration, the dynamic filaments never grow long enough to achieve stabilizing lateral interactions. This is consistent with the observation that filaments solubilized by capping or randomness disintegrate when diluted. Therefore sedimentation is not necessarily indicative of assembly formation. Filaments solubilized by capping or randomization do not precipitate or sediment in a tabletop centrifuge but may easily be confirmed by negative-stain TEM.

Ridges-into-grooves packing is a defining feature of this group of protein co-assemblies and may be a challenge in protein assembly design in general. Whereas assemblies of peptides tend to smooth or have slight helical grooves and bulges, such as α stacks (Wang et al., 2021b), proteins tend to be globular and thus have domains that protrude from the assembly at regular intervals, encouraging lateral interactions. Assigning similarly charged residues to the packing interface can make packing less favorable but is not always conducive to the nanotube's application. In α Reps, packing is mediated by electrostatic interactions and further stabilized by entropic contributions.

Of the techniques intended to solubilize the Brick/Backbinder filaments, only capping and randomization were effective. Capping inhibits filament growth before a sufficient surface for ridges-into-grooves packing establishes. Because this minimum surface is so small for Brick C, the solubilized filaments are too small to detect by TEM and may not even have the aspect ratio to qualify as filaments. Brick D/BBST filaments, by contrast, have a curved surface that packs over about 50 nm. Soluble filaments of a somewhat tunable length can be produced by capping the filaments. Mixing Brick C and D together distorts the surface so that two adjacent filaments are very unlikely to find a site suitable for ridges-into-grooves packing. In each case, solubilized filaments are short and form only at relatively high concentrations (16 μ M total protein); dilution will cause these dynamic filaments to disintegrate. Brick F/ α 2 filaments are very flexible and pack into bundles of two or three filaments. Brick F/ α 2/BBST filaments achieve a longer length of about 100 nm due to their internally stabilizing α 2 protein, but condense instead of packing in a regular crystal. These results indicate that lateral interactions stabilize the filaments, which otherwise do not persist at lengths over 50 nm.

To further stabilize soluble filaments, the Backbinder was linked to a dimerized α Rep (A3-E3). Brick C/A3-E3 coassembly filaments laterally associate but do not crystallize, probably due to the flexible linker joining the A3 and E3 domains. These filaments achieve lengths of around 150 A, significantly longer than Brick C/Backbinder. However, Brick D was not observed to stably co-assemble with Brick D, even when mixed with Backbinder. This is likely due to spatial constraints: A3-E3 should exist as a dimer, so E3 domains are locked into position that is compatible with Brick C but not Brick D.

Without a means of stabilizing soluble filaments, a functionalized Brick/Backbinder co-assembly remains elusive. The lateral interactions between fibers seem to be highly stabilizing. If GFP or any other ligand frustrates this interaction, the soluble filament is unlikely to accrue and keep enough monomers to grow. This explanation is consistent with the observation that no crystals of any Brick/Backbinder-ST include GFP-SC, existing Brick D/Backbinder/GFP-SF filaments appear to be doubled. However, the terminus/i of the filament may be more amenable to decoration if the rest of the filament can be strengthened. The terminal Brick/Brick interface has already been optimized, but the Brick/Backbinder interfaces could theoretically be improved by optimizing sidechain contacts or simply by lengthening the Backbinder to increase its size. Backbinder was developed from a random screen. Another, better binder may be retrieved from another round of screening, or the existing Backbinder could be computationally optimized.

The assembled structure results directly from the Brick/Staple interactions and on this aspect is readily programmable as long as the structure of Brick/staple interactions is known. A more subtle set of assembly rules emerges from the importance of interfilament or “lateral” interactions that are clearly important in the final assembly, but are much less predictable than the axial interactions. The importance of lateral interactions is apparent in the following results. First, the highly regular and crystalline organization previously observed with Brick B is also observed, as expected, with Brick C, which has the same geometry of Backbinder distribution on the filament surface. However, a different but also highly regular interfilament organization is observed with Brick D, although the design of this Brick cannot allow the same type of interdigitation of the Backbinders without alteration of the geometry of the filament surface. Second, even when the protruding Backbinders are completely absent such as in the Brick F/a2 assembly, the structure rapidly formed is not an isolated filament but a “rope” like structure with 2 or 3 mutually twisted filaments, that slowly evolve to form a greater organized structure. This type of assembly is not easily predicted but emerges frequently from regular assemblies. As already mentioned, even a weak interaction between two isolated proteins becomes a predominant source of intermolecular interactions when repeated many times between two repeated filaments (Moreaud et al.,

2023). The difficulty of predicting the exact outcome of multimolecular repeated association of helical peptides has already been noted (Egelman et al., 2015). Low-energy, high-repeat interactions are similarly difficult to predict. These interfilament interactions play an important role in determining which structures will finally result from the associated proteins.

Brick C bears significant resemblance to HEAT_8 (Bartlett, 2018). The two proteins are both 8-repeat α Reps, but they assemble by different means. While Brick C was designed for high Brick-Brick affinity, HEAT_8R was simply concatenated from the single-repeat consensus sequence and only modified to avoid Trp stacking. HEAT_1R (Hughes et al., 2019) is thought to be stabilized along the axial interface by H-bonding between the side-chain hydroxide of Tyr18 and the backbone amide of Asp1. HEAT_8R lacks an analogous tyrosine but still shows a smooth-surfaced nanotube. Notably, the HEAT_1R nanotube has 20 proteins per turn for a total of 20 repeats, whereas Brick C has 16 repeats per turn. According to the AlphaFold model (Jumper et al., 2021; Mirdita et al., 2022), the axial interfaces of HEAT_8R are not quite as negatively charged as Brick C, and may be further neutralized at pH 5, the optimal assembly condition. Otherwise, the two proteins are expected to have the same size and curvature. Another obvious difference between the two filaments is the Backbinder, which partially invades the space between successive turns of the Brick C helix and may prevent an axial interaction between Brick proteins.

Due to their modularity, thermostability, small size, and two variable surfaces, α Rep filaments represent an exciting tool for nanofabrication and a starting point for designing nanorobotic systems. Overall conformation is robust in sequence space among proteins described here, but the fact that the α Rep-derived peptide HEATR1 assembles into a closed tube suggests that different geometries are possible (Hughes et al., 2019). Grafting a binding site onto the Backbinder would orient the bound ligand relative to the filament and may enable cryo-EM imaging or crystallization of small proteins. Overall, these diverse filament and sheet assembly systems will be useful to protein designers seeking a predictable, stable scaffold.

Finally, our results suggest that any functionalization of the assembled protein must be structurally compatible with their assembled superstructure. For example, the very strong linkage indicated by the Spytag/Spycatcher association does not allow decoration with a reporter protein like GFP but instead leads to a dissociation of the assembled structure. More subtle ways of functionalization, taking into account the structural compatibility of the graft protein onto the host superstructure, should be explored. Note that dynamic behavior of associated superstructures with a range of alternative partners also open routes towards dynamically controlled superstructure, and is a recurrent theme in the dynamics of natural protein superstructures such as actin filaments or microtubules (Knossow et al., 2020).

Materials and Methods

Expression plasmids The expression plasmids adapted to Golden Gate cloning were designed and kindly provided by R Sterner. Proteins were expressed in *E. coli* using synthetic genes cloned pUR22 and pUR23 expression plasmids (Rohweder et al., 2018). These plasmids allow simple and efficient cloning using Golden Gate assembly and efficient expression based on IPTG inducible T5 promoter.

Plasmid preparation. Plasmids were isolated with a NucleoSpin Plasmid mini plasmid prep kit (Macherey Nagel, Hoerdt, France) according to kit instructions. DNA was eluted into a clean microcentrifuge tube by placing 50 μ L elution buffer on membrane and centrifuging 2 minutes. Insert sequences were verified by Eurofins Genomics (Germany). After sequence verification, plasmids were transformed into BL21 (Gold) *E. coli*. Cells were plated onto LB agar plates supplemented with ampicillin and 1% glucose.

Protein expression and preparation. Two colonies were selected for an overnight biomass build in 50 mL 2YT+ampicillin at 37°C. 1L 2YT+ampicillin, divided into two 2L Erlenmeyer flasks, was inoculated from a single culture to a final OD₆₀₀ of 0.15. This culture was shaken at 37°C until OD₆₀₀=0.6-0.8, about 2.5 hours. When this OD was reached, IPTG was added to a final concentration of 1 mM and temperature was reduced to 30°C to encourage protein production. After 4 hours (16 hours for α 2) of induction, the

cells were pelleted in a centrifuge at $10000\times g$ for 30 minutes, 4°C . Whole culture and pellet samples were taken to verify protein production by SDS-PAGE. Pellet was resuspended in 25 mL Tris Buffered Saline (Tris HCl 50 mM, NaCl 150 mM, pH 7.5) with one tablet cOmplete protease inhibitor (Roche Boulogne-Billancourt, France) and stored at -20°C . 2 μL DNase (Thermo Scientific, Villebon-sur-Yvette, France) was added to the pellet as it thawed in a tepid water bath. Once thawed, the pellet was sonicated on ice 30 seconds on/30 seconds off for 5 minutes. The lysed pellet was clarified by centrifuging at $10000\times g$ 60-90 minutes at 4°C . Supernatant was carefully removed and passed through a $0.2\ \mu\text{M}$ syringe filter to remove cells. Protein was purified via nickel-NTA immobilized metal affinity column chromatography. For a gravity column, 4 mL resin stored in 20% ethanol (Protino® Ni-NTA Agarose, Machery-Nagel, Hoerd, France) was loaded onto a column. Ethanol was drained out. 5 CV wash buffer (NaCl 300 mM, NaH_2PO_4 50 mM, imidazole 20 mM) was passed through. Clarified lysate was passed through and collected as "flow-through." Resin was washed with 5 CV wash buffer, which was collected. Protein was eluted in elution buffer (NaCl 300 mM, NaH_2PO_4 50 mM, imidazole 300 mM). One CV elution buffer was applied to resin at a time until the OD_{280} of the eluent was below 0.05.

When possible, FPLC Ni-NTA IMAC was used instead of the gravity method. The AKTA Pur method used a 5 mL Ni-NTA column (His-Trap FF Crude, GE Healthcare Bio-Sciences AB, Uppsala, Sweden) and auto-collected flow-through and wash fractions in 50 mL conical tubes, and elution fractions in a 1.5 mL 96-well deep-well plate. Buffers were the same as above. As soon as possible, eluted protein was purified by size exclusion chromatography. Auto-collected fractions were analyzed by SDS-PAGE.

To cleave the 6x His tag and capping motifs, protein was incubated with tobacco etch virus protein (TEV) at 1:50 TEV:protein, OD/OD in TBS with 1 mM dithiothreitol, 2 mM EDTA. This cleavage cocktail was incubated with shaking at 4°C overnight or 30°C for 2.5 hours. Cleaved protein was isolated by passing through Ni-NTA resin so that cleaved His tag and TEV protein are retained on the column. Cleavage was verified by SDS-PAGE. When necessary, a second round of cleavage was performed. Pure cleaved protein was concentrated to about 4 mL, desalted to remove DTT and EDTA, and aliquoted for storage at -80°C .

Filament assembly. Unless specified otherwise, proteins were incubated at a 1:1 molar ratio in TBS buffer at 37°C for assembly, and stored at -20°C.

Crystallization rates. To monitor crystallization rates, OD₃₅₀ was monitored using a Varian Cary 50 Bio Spectrophotometer with Cary WinUV software. Two water baths heated or cooled water to a specified temperature. Water ran through plastic tubing to the quartz cuvette. The chamber was continuously purged with dry air to prevent condensation. After the desired water temperature was achieved, the cuvette temperature was equilibrated for several minutes. For an experiment with 4 μM each protein, 50 μL 8 μM Backbinder was placed in the cuvette and the absorbance reading at 350 nm was set to 0. As quickly as possible, 50 μL 8 μM Brick protein was added to the cuvette, pipetted up and down to mix, and the lid to the chamber was shut and the data collection begun. Data was collected every 0.2 seconds for at least 15 minutes. Dense crystals settle at the bottom of the cuvette so only the first 15-20 minutes of data collection was useful for analysis.

Circular dichroism. 40 μL of μM protein was applied to a 0.1 mm quartz High Precision Cell cuvette (Hellma Analytics, Jena, Germany), which was assembled and wrapped with Teflon plumber's tape and Parafilm to retard evaporation. After blanking, a spectrum from 190 to 250 nm was measured in a JASCO J-1500 CD Spectrometer, with SpectraManager software. Once a characteristic low point was determined, the molar ellipticity was read at 208 from 4 to 95°C, ramping up the temperature by 4°C per minute. After 95°C was reached, a second spectrum was measured.

Negative staining sample protocol. Specimens were prepared for electron microscopy using the conventional negative staining procedure. 5 μl of solutions was adsorbed on carbon-coated copper grids (Electron Microscopy Sciences, Hatfield, PA, USA) for 1 min, blotted, and negatively stained with uranyl acetate (2%) for 1 min. Grids were examined with a transmission electron microscope (Jeol JEM-1400, JEOL Inc, Peabody, MA, USA) at 80 kV. Images were acquired using a Gatan Orius digital camera (Gatan Inc, Pleasanton, CA, USA).

Cryo imaging protocol. 3 μ L of sample were deposited onto negative glow-discharged holey carbon grids (300 mesh, Ted Pella Inc, Redding, CA, USA) and frozen using a VitroBot Mark IV plunge freezer (Thermo Fisher, set at 20°C and 100% humidity. Excess solution was removed by blotting with Whatman n°1 filter paper, and the grids were immediately flash frozen in liquid ethane at -185°C. The frozen specimens were placed in a Gatan 626 cryo-holder, and cryo EM imaging was carried out on a JEOL 2100 microscope, equipped with a LaB6 cathode and operating at 200 kV, under low dose conditions, or an FEI Talos 120 KV microscope with a LaB6 cathode operating at 120 kV. Images were acquired with SerialEM software or Gatan Digital Micrograph, with defocus of 1–2 μ m, on a Gatan US4000 CCD camera.

References

- Aida, T., Meijer, E.W., Stupp, S.I., 2012. Functional Supramolecular Polymers. *Science* 335, 813-817.
- Bartlett, R., 2018. Engineering Advanced Self-Assembling Protein Biomaterials through the Extrapolation and Functionalization of Peptide-Based Systems.
- Beloqui, A., Cortajarena, A.L., 2020. Protein-based functional hybrid bionanomaterials by bottom-up approaches. *Curr Opin Struct Biol* 63, 74-81.
- Ben-Sasson, A.J., Watson, J.L., Sheffler, W., Johnson, M.C., Bittleston, A., Somasundaram, L., Decarreau, J., Jiao, F., Chen, J., Mela, I., Drabek, A.A., Jarrett, S.M., Blacklow, S.C., Kaminski, C.F., Hura, G.L., De Yoreo, J.J., Kollman, J.M., Ruohola-Baker, H., Derivery, E., Baker, D., 2021. Design of biologically active binary protein 2D materials. *Nature* 589, 468-473.
- Cannon, K.A., Park, R.U., Boyken, S.E., Nattermann, U., Yi, S., Baker, D., King, N.P., Yeates, T.O., 2020. Design and structure of two new protein cages illustrate successes and ongoing challenges in protein engineering. *Protein Sci* 29, 919-929.

- Carlson, J.C., Jena, S.S., Flenniken, M., Chou, T.F., Siegel, R.A., Wagner, C.R., 2006. Chemically controlled self-assembly of protein nanorings. *J Am Chem Soc* 128, 7630-7638.
- Chevrel, A., Urvoas, A., de la Sierra-Gallay, Ines L., Aumont-Nicaise, M., Moutel, S., Desmadril, M., Perez, F., Gautreau, A., van Tilbeurgh, H., Minard, P., Valerio-Lepiniec, M., 2015. Specific GFP-binding artificial proteins (α Rep): a new tool for in vitro to live cell applications. *Bioscience Reports* 35.
- Churchfield, L.A., Tezcan, F.A., 2019. Design and Construction of Functional Supramolecular Metalloprotein Assemblies. *Acc Chem Res* 52, 345-355.
- Correnti, C.E., Hallinan, J.P., Doyle, L.A., Ruff, R.O., Jaeger-Ruckstuhl, C.A., Xu, Y., Shen, B.W., Qu, A., Polkinghorn, C., Friend, D.J., Bandaranayake, A.D., Riddell, S.R., Kaiser, B.K., Stoddard, B.L., Bradley, P., 2020. Engineering and functionalization of large circular tandem repeat protein nanoparticles. *Nat Struct Mol Biol* 27, 342-350.
- Di Meo, T., Ghattas, W., Herrero, C., Velours, C., Minard, P., Mahy, J.P., Ricoux, R., Urvoas, A., 2017. α Rep A3: A Versatile Artificial Scaffold for Metalloenzyme Design. *Chemistry* 23, 10156-10166.
- Egelman, E.H., Xu, C., DiMaio, F., Magnotti, E., Modlin, C., Yu, X., Wright, E., Baker, D., Conticello, V.P., 2015. Structural plasticity of helical nanotubes based on coiled-coil assemblies. *Structure* 23, 280-289.
- Fellouse, F.A., Wiesmann, C., Sidhu, S.S., 2004. Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. *Proc Natl Acad Sci U S A* 101, 12467-12472.

- Fernandez, M., Urvoas, A., Even-Hernandez, P., Burel, A., Mériadec, C., Artzner, F., Bouceba, T., Minard, P., Dujardin, E., Marchi, V., 2020. Hybrid gold nanoparticle–quantum dot self-assembled nanostructures driven by complementary artificial proteins. *Nanoscale* 12, 4612-4621.
- Guellouz, A., Valerio-Lepiniec, M., Urvoas, A., Chevrel, A., Graille, M., Fourati-Kammoun, Z., Desmadril, M., van Tilbeurgh, H., Minard, P., 2013. Selection of specific protein binders for pre-defined targets from an optimized library of artificial helicoidal repeat proteins (alphaRep). *PLoS One* 8, e71512.
- Hamley, I.W., 2019. Protein assemblies: nature-inspired and designed nanostructures. *Biomacromolecules* 20, 1829-1848.
- Harvey, J.A., Itzhaki, L.S., Main, E.R.G., 2018. Programmed Protein Self-Assembly Driven by Genetically Encoded Intein-Mediated Native Chemical Ligation. *ACS Synth Biol* 7, 1067-1074.
- Hughes, S.A., Wang, F., Wang, S., Kreutzberger, M.A.B., Osinski, T., Orlova, A., Wall, J.S., Zuo, X., Egelman, E.H., Conticello, V.P., 2019. Ambidextrous helical nanotubes from self-assembly of designed helical hairpin motifs. *Proceedings of the National Academy of Sciences of the United States of America* 116, 14456-14464.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P.,

- Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583-589.
- Kim, N.H., Choi, H., Shahzad, Z.M., Ki, H., Lee, J., Chae, H., Kim, Y.H., 2022. Supramolecular assembly of protein building blocks: from folding to function. *Nano Convergence* 9, 4.
- Kleffner, R., Flatten, J., Leaver-Fay, A., Baker, D., Siegel, J.B., Khatib, F., Cooper, S., 2017. Foldit Standalone: a video game-derived protein structure manipulation interface using Rosetta. *Bioinformatics* 33, 2765-2767.
- Knossow, M., Campanacci, V., Khodja, L.A., Gigant, B., 2020. The Mechanism of Tubulin Assembly into Microtubules: Insights from Structural Studies. *iScience* 23.
- Kobayashi, N., Arai, R., 2017. Design and construction of self-assembling supramolecular protein complexes using artificial and fusion proteins as nanoscale building blocks. *Curr Opin Biotechnol* 46, 57-65.
- Kobayashi, N., Yanase, K., Sato, T., Unzai, S., Hecht, M.H., 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.
- Koide, S., Sidhu, S.S., 2009. The Importance of Being Tyrosine: Lessons in Molecular Recognition from Minimalist Synthetic Binding Proteins. *ACS Chemical Biology* 4, 325-334.
- Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* 372, 774-797.
- Lai, Y.T., Tsai, K.L., Sawaya, M.R., Asturias, F.J., Yeates, T.O., 2013. Structure and flexibility of nanoscale protein cages designed by symmetric self-assembly. *J Am Chem Soc* 135, 7738-7743.

- Lee, H., DeLoache, W.C., Dueber, J.E., 2012. Spatial organization of enzymes for metabolic engineering. *Metab Eng* 14, 242-251.
- Léger, C., Yahia-Ammar, A., Susumu, K., Medintz, I.L., Urvoas, A., Valerio-Lepiniec, M., Minard, P., Hildebrandt, N., 2020. Picomolar Biosensing and Conformational Analysis Using Artificial Bidomain Proteins and Terbium-to-Quantum Dot Förster Resonance Energy Transfer. *ACS Nano* 14, 5956-5967.
- Léger, C., Di Meo, T., Aumont-Nicaise, M., Velours, C., Durand, D., Li de la Sierra-Gallay, I., van Tilbeurgh, H., Hildebrandt, N., Desmadril, M., Urvoas, A., Valerio-Lepiniec, M., Minard, P., 2019. Ligand-induced conformational switch in an artificial bidomain protein scaffold. *Scientific Reports* 9, 1178.
- Luo, Q., Hou, C., Bai, Y., Wang, R., Liu, J., 2016. Protein Assembly: Versatile Approaches to Construct Highly Ordered Nanostructures. *Chem Rev* 116, 13571-13632.
- Meyers, M.A., Chen, P.Y., Lin, A.Y.M., Seki, Y., 2008. Biological materials: Structure and mechanical properties. *Prog. Mater. Sci.* 53, 1-206.
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M., 2022. ColabFold: making protein folding accessible to all. *Nature Methods* 19, 679-682.
- Moreaud, L., Viollet, S., Urvoas, A., Valerio-Lepiniec, M., Mesneau, A., Li de la Sierra-Gallay, I., Miller, J., Ouldali, M., Marcelot, C., Balor, S., Soldan, V., Meriadec, C., Artzner, F., Dujardin, E., Minard, P., 2023. Design, synthesis, and characterization of protein origami based on self-assembly of a brick and staple artificial protein pair. *Proceedings of the National Academy of Sciences* 120, e2218428120.

- Parmeggiani, F., Huang, P.S., Vorobiev, S., Xiao, R., Park, K., Caprari, S., Su, M., Seetharaman, J., Mao, L., Janjua, H., Montelione, G.T., Hunt, J., Baker, D., 2015. A general computational approach for repeat protein design. *J Mol Biol* 427, 563-575.
- Rohweder, B., Semmelmann, F., Endres, C., Sterner, R., 2018. Standardized cloning vectors for protein production and generation of large gene libraries in *Escherichia coli*. *BioTechniques* 64, 24-26.
- Shen, H., Fallas, J.A., Lynch, E., Sheffler, W., Parry, B., Jannetty, N., Decarreau, J., Wagenbach, M., Vicente, J.J., Chen, J., Wang, L., Dowling, Q., Oberdorfer, G., Stewart, L., Wordeman, L., De Yoreo, J., Jacobs-Wagner, C., Kollman, J., Baker, D., 2018. *De novo* design of self-assembling helical protein filaments. *Science* 362, 705-709.
- Shiroishi, M., Tsumoto, K., Tanaka, Y., Yokota, A., Nakanishi, T., Kondo, H., Kumagai, I., 2007. Structural Consequences of Mutations in Interfacial Tyr Residues of a Protein Antigen-Antibody Complex: *THE CASE OF HyHEL-10-HEL*. *Journal of Biological Chemistry* 282, 6783-6791.
- Urvoas, A., Guellouz, A., Valerio-Lepiniec, M., Graille, M., Durand, D., Desravines, D.C., van Tilbeurgh, H., Desmadril, M., Minard, P., 2010. Design, production and molecular structure of a new family of artificial alpha-helical repeat proteins (α Rep) based on thermostable HEAT-like repeats. *Journal of Molecular Biology* 404, 307-327.
- Voet, A.R., Tame, J.R., 2017. Protein-templated synthesis of metal-based nanomaterials. *Curr Opin Biotechnol* 46, 14-19.
- Wang, F., Gnewou, O., Wang, S., Osinski, T., Zuo, X., Egelman, E.H., Conticello, V.P., 2021a. Deterministic chaos in the self-assembly of β sheet nanotubes from an amphipathic oligopeptide. *Matter*.

- Wang, F., Gnewou, O., Modlin, C., Beltran, L.C., Xu, C., Su, Z., Juneja, P., Grigoryan, G., Egelman, E.H., Conticello, V.P., 2021b. Structural analysis of cross α -helical nanotubes provides insight into the designability of filamentous peptide nanomaterials. *Nature Communications* 12, 407.
- Yeates, T.O., 2017. Geometric Principles for Designing Highly Symmetric Self-Assembling Protein Nanomaterials. *Annu Rev Biophys* 46, 23-42.
- Yeates, T.O., Agdanowski, M.P., Liu, Y., 2020. Development of imaging scaffolds for cryo-electron microscopy. *Curr Opin Struct Biol* 60, 142-149.
- Zhu, J., Avakyan, N., Kakkis, A., Hoffnagle, A.M., Han, K., Li, Y., Zhang, Z., Choi, T.S., Na, Y., Yu, C.-J., Tezcan, F.A., 2021. Protein Assembly by Design. *Chemical Reviews* 121, 13701-13796.

GDERAVPALIEALKDEDARVRAAAAKALGKI
 GDPEAVEALIYALRDEDANVRRAAAALGEIGSGSGENLYFQ

Brick D, nascent:

SHMTDPEKVDMYIENLRDEDPEVRARAAEALGKIGSGSGENLYFQGGSGSG
 GDEAAVYPLIQALEDEDAEVRAAAARALGKI
 GDERAVPALIEALKDEDKVRAAAARALGEI
 GDERAVEPLIKALKDEDAAVREAAAALGRI
 GDERAVEPLIKALKDEDWNRKAAAALGKI
 GDEAAVEPLIQALEDEDAEVRAAAACALGEI
 GDERAVPALIEALKDEDARVRAAAAKALGKI
 GDPEAVEALIYALRDEDANVRRAAAALGEIGSGSGENLYFQGGSGSGDPRAEEALRRAREDEDPEVQKEAEKAEG
 EIGSGSGHHHHHKLIS

Brick D, cleaved:

GGSGSGDEAAVYPLIQALEDEDAEVRAAAARALGKI
 GDERAVPALIEALKDEDKVRAAAARALGEI
 GDERAVEPLIKALKDEDAAVREAAAALGRI
 GDERAVEPLIKALKDEDWNRKAAAALGKI
 GDEAAVEPLIQALEDEDAEVRAAAACALGEI
 GDERAVPALIEALKDEDARVRAAAAKALGKI
 GDPEAVEALIYALRDEDANVRRAAAALGEIGSGSGENLYFQ

Backbinder E3 **SpyTag**:

MHHHHHLDMENLYFQGTDPKVMYIKNLQDDSIWVRYSAASALGKI
 GDERAVEPLIKALKDEDDGYVRQAAALALGQI
 GDERAVEPLIKALKDEDSTVRIRAARALGKI
 GDERAVEPLIKALKDEDDWQVRLSAASALGKI
 GDERAVEPLIKALKDEDDPSVRMAAANALGQI
 GGERVRAAMEKLAETGTGFARKVAVNYLETHPSETRG**VPHIVMVDAYKRYK**

GFP superfolder SpyCatcher:

MHHHHHLDMENLYFQGPSETVTTLTSLSGEQGPGSDMTTEEDSATHIKFSKRDEDGRELATMELRDSSGKTIST
 WISDGHVKDFLYLPKGYTFVETAAPDGYEVATPIEFTVNEDGQVTVDEATEGDAHTASQTMKGEELFTGVVPILV
 ELDGDVNGHKFSVRGEGEGDATNGKLTLLKFICTTGKLPVPWPTLVTTLYTGVQCFSRYPDHMKRHDFFKSAMPEGYV
 QERTISFKDDGTYKTRAEVKFEEDTLVNRIELKIDFKEDGNILGHKLEYNFNHNVYITADKQKNGIKANFKIRHN
 VEDGSVQLADHYQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMLLEFVTAAGITHGMDELYK

Brick N End nascent:

MTDPEKVDMYIENLRDEDPEVRARAAEALGKI
 GDERAVEPLIKALKDEDKVRAAAARALGEI
 GDERAVEPLIKALKDEDAAVREAAAALGRI
 GDERAVEPLIKALKDEDARVRAAAARALGEI
 GDERAVEPLIKALKDEDWNRKAAAALGKI
 GDEAAVEPLIQALEDEDAEVRAAAACALGEI
 GDTAAVPALIEALKDEDARVRAAAAKALGKI
 GDPEAVEALIYALRDEDANVRRAAAALGEIGSGSGENLYFQGGSGSGDPRAEEALRRAREDEDPEVQKEAEKAEG
 EILEHHHHH

Brick N End cleaved:

MTDPEKVDMYIENLRDEDPEVRARAAEALGKI
 GDERAVEPLIKALKDEDAKVREAAARALGEI
 GDERAVEPLIKALKDEDAAVREAAAEALGRI
 GDERAVEPLIKALKDEEDARVREAAARALGEI
 GDERAVEPLIKALKDEEDWNRKAAAEALGKI
 GDEAAVEPLIQALEDEDAEVRAAAACALGEI
 GDAAVPALIEALKDEEDARVRAAAAKALGKI
 GDPEAVEALIIYALRDEDANVRRAAAEALGEIGSGSGENLYFQ

Brick C End nascent:

MHHHHHHLDMTDPEKVDMYIENLRDEDPEVRARAAEALGKIGSGSGENLYFQGGSGSGGDEAAVYPLIQALEDEDAE
 VRAAAARALGKI
 GDERAVEALIEALKDEDAKVREAAARALGEI
 GDERAVEPLIKALKDEDAAVREAAAEALGRI
 GDERAVEPLIKALKDEEDARVREAAARALGEI
 GDERAVEPLIKALKDEEDWNRKAAAEALGKI
 GDEAAVEPLIKALKDEDAEVRAAAACALGEI
 GDERAVEPLIKALKDEEDARVRAAAAKALGKI
 GDPRAEEALRRAREDEDPEVQKEAEKAEGEI

Brick C End cleaved:

GGSGSGGDEAAVYPLIQALEDEDAEVRAAAARALGKI
 GDERAVEALIEALKDEDAKVREAAARALGEI
 GDERAVEPLIKALKDEDAAVREAAAEALGRI
 GDERAVEPLIKALKDEEDARVREAAARALGEI
 GDERAVEPLIKALKDEEDWNRKAAAEALGKI
 GDEAAVEPLIKALKDEDAEVRAAAACALGEI
 GDERAVEPLIKALKDEEDARVRAAAAKALGKI
 GDPRAEEALRRAREDEDPEVQKEAEKAEGEI

E3-A3 construct:

MRGSHHHHHTDPEKVMYIKNLQDDSIWVRYSAASALGKI
 GDERAVEPLIKALKDEEDGYVRQAAALALGQI
 GDERAVEPLIKALKDEEDSTVRIRAARALGKI
 GDERAVEPLIKALKDEEDWQVRLSAASALGKI
 GDERAVEPLIKALKDEEDPSVRMAAANALGQI
 GGERVRAAMEKLAETGTGFARKVAVNYLETHKSLISGAGSGSGDPEKVMYIKNLQDDSYVRRAAAYALGKI
 GDERAVEPLIKALKDEEDAWVRRAAADALGQI
 GDERAVEPLIKALKDEEDGWVRQSAVALGQI
 GDERAVEPLIKALKDEEDWVRIAAAFALGEI
 GDERAVEPLIKALKDEEDGWVRQSAADALGEI
 GGERVRAAMEKLAETGTGFARKVAVNYLETH

Brick F, nascent:

MRGSHHHHHTDPEKVDMYIENLRDEDPEVRARAAEALGKIGSGSGENLYFQGGSGSG
 GDEAAVYPLIQALEDEEDGWVRQSAVALGQI
 GDERAVEALIEALKDEEDWVRIAAAFALGEI
 GDERAVEPLIKALKDEEDGWVRQSAADALGEI
 GDERAVEPLIKALKDEEDARVRAEAAKALGDI

GDERAVEPLIKALKDEDPEVRKAAALALGKI
 GDEAAVEPLIQALEDESDVRAAAAWALGKI
 GDAAVPALIEALKDEDYYVRRAAAYALGKI
 GDPEAVEALIYALRDEDAWVRRAAADALGQIGSGSGENLYFQGGSGSGDPRAEEALRRAREDEDPEVQKEAEKAEG
 EILEHHHHHH*

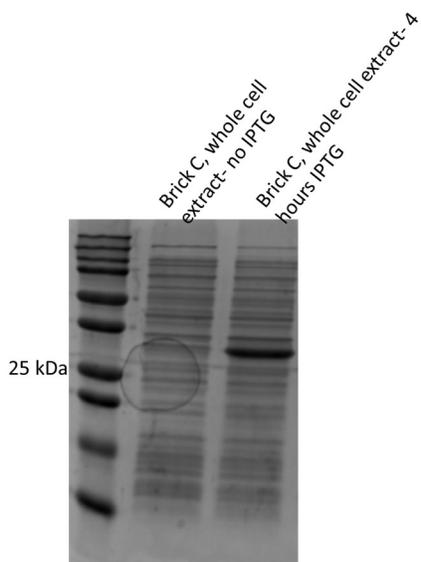
Brick F, cleaved:

GGSGSGDEAAVYPLIQALEDEDGWVRQSAVALGQI
 GDERAVEALIEALKDEDWVFRIAAAFALGEI
 GDERAVEPLIKALKDEDGWVRQSAADALGEI
 GDERAVEPLIKALKDEDARVRAEAAKALGDI
 GDERAVEPLIKALKDEDPEVRKAAALALGKI
 GDEAAVEPLIQALEDESDVRAAAAWALGKI
 GDAAVPALIEALKDEDYYVRRAAAYALGKI
 GDPEAVEALIYALRDEDAWVRRAAADALGQIGSGSGENLYFQ

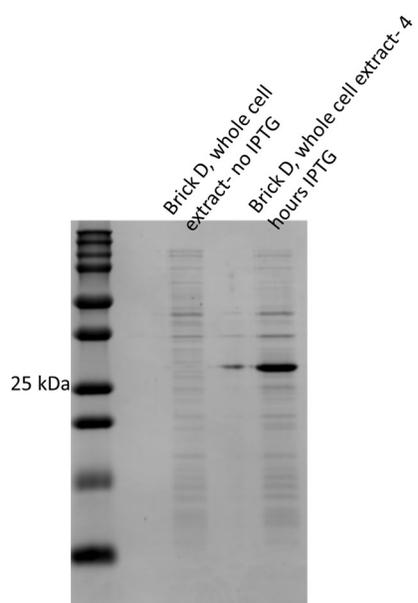
$\alpha 2$:

MRGSHHHHHHTDPEKVEYIKNLQDDSSVVRKAAVALGEI
 GDERAVEPLIKALKDEDQVFRIAAAWALGKIGGERVRAAMEKLAETGTGFARKVAVNYLETHKSLIS

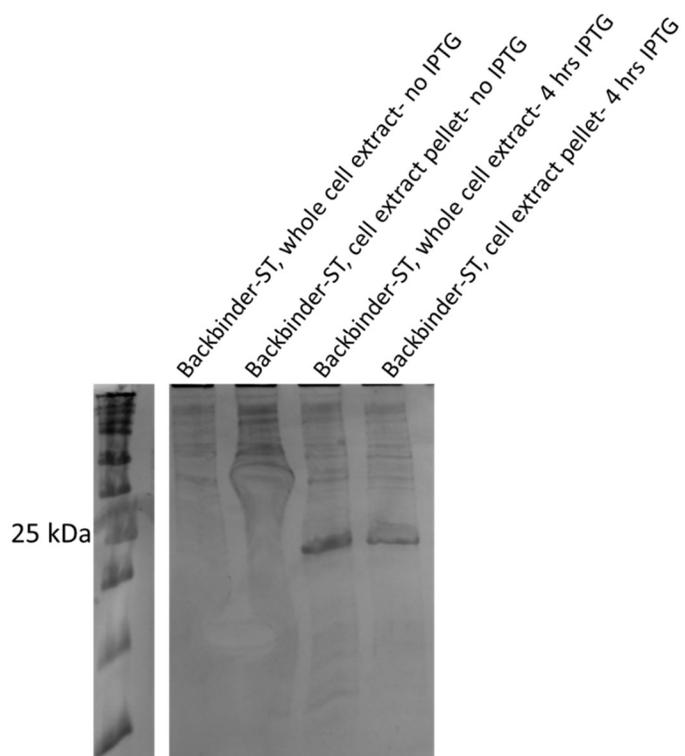
SI Figure 2-1. Sequences of proteins described in this chapter.



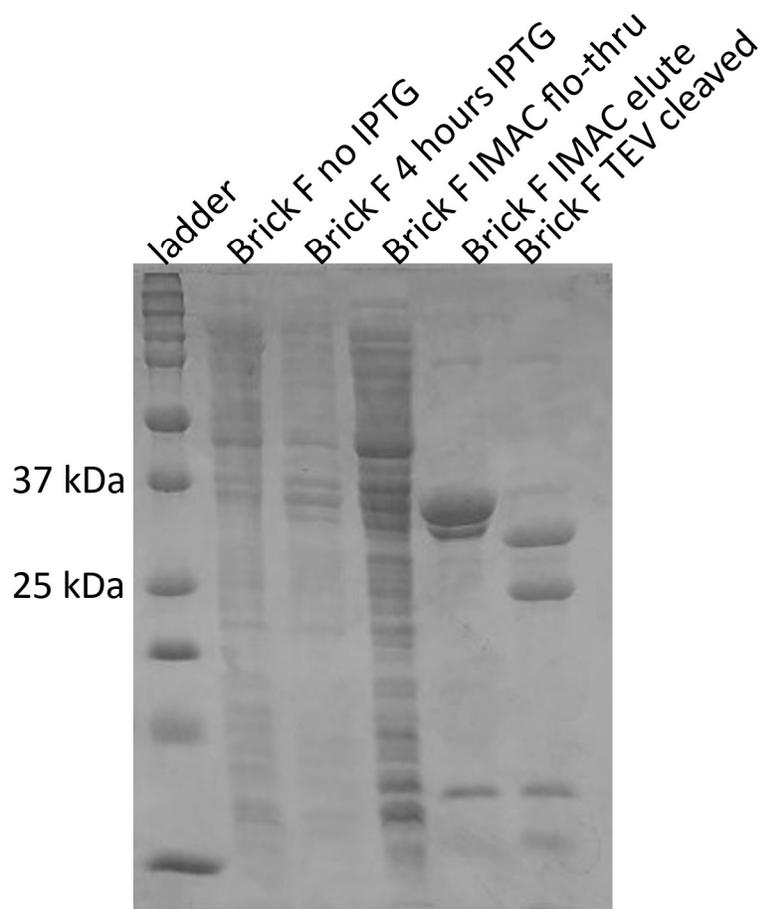
SI Figure 2-2. SDS-PAGE gel showing Brick C expression.



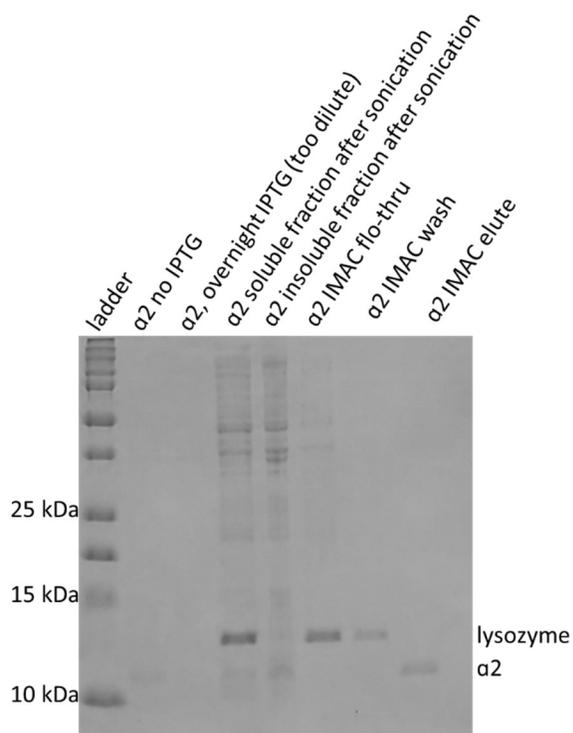
SI Figure 2-3. SDS-PAGE gel showing Brick D expression.



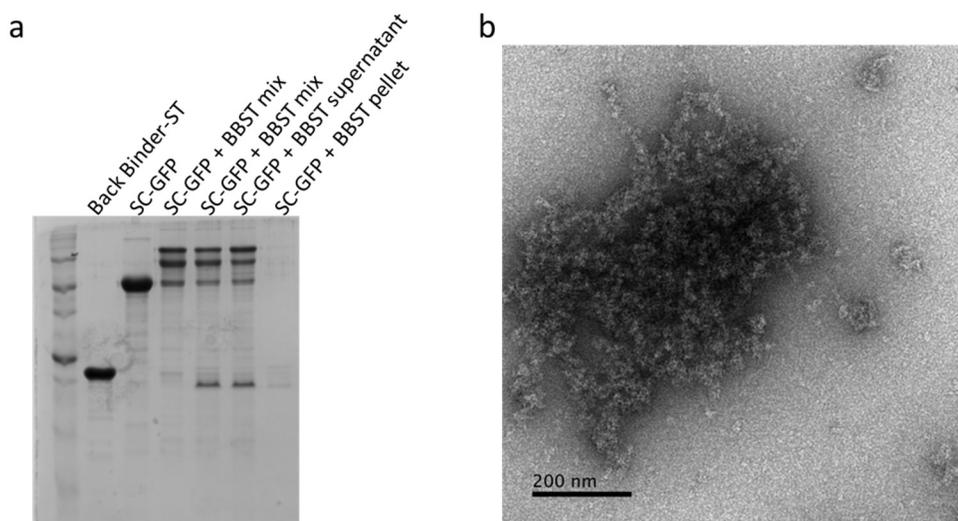
SI Figure 2-4. SDS-PAGE gel showing Backbinder-ST expression.



SI Figure 2-5. SDS-PAGE gel showing Brick F expression, purification, and TEV cleavage.



SI Figure 2-6. SDS-PAGE gel showing $\alpha 2$ overnight expression and purification.



SI Figure 2-7. A. SDS-PAGE gel showing that sfGFP-SC and Backbinder-ST do form a covalent bond. B. Brick C assembled with Backbinder-ST, then decorated with sfGFP-SC. No ordered assemblies were observed.

Chapter 3 |

Archaeal DNA-import apparatus is homologous to bacterial conjugation machinery

Note: This chapter is reformatted to follow Emory's dissertation guidelines. It is reprinted from Beltran, L.C., Cvirkaite-Krupovic, V., Miller, J. *et al.* Archaeal DNA-import apparatus is homologous to bacterial conjugation machinery. *Nat Commun* **14**, 666 (2023). <https://doi.org/10.1038/s41467-023-36349-8> under a Creative Commons CC BY license.



Article <https://doi.org/10.1038/s41467-023-36349-8>

Archaeal DNA-import apparatus is homologous to bacterial conjugation machinery
Received: 4 November 2022

Accepted: 27 January 2023

Published online: 07 February 2023

Leticia C. Beltran,¹ Virginija Cvirkaite-Krupovic,² Jessalyn Miller,³ Fengbin Wang,^{1,7} Mark A. B. Kreuzberger,¹ Jonasz B. Patkowski,⁴ Tiago R. D. Costa,⁴ Stefan Schouten,⁵ Ilya Levental,⁶ Vincent P. Conticello,³ Edward H. Egelman¹ & Mart Krupovic²

¹Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA 22903, USA. ²Institut Pasteur, Université Paris Cité, CNRS UMR6047, Archaeal Virology Unit, 75015 Paris, France. ³Department of Chemistry, Emory University, Atlanta, GA 30322, USA. ⁴MRC Centre for Molecular Bacteriology and Infection, Department of Life Sciences, Imperial College, London, UK. ⁵NIOZ Royal Netherlands Institute for Sea Research, Department of Marine Microbiology and Biogeochemistry, Texel, The Netherlands. ⁶Department of Molecular Physiology and Biological Physics, Center for Membrane and Cell Physiology, University of Virginia, Charlottesville, VA 22903, USA. ⁷Present address: Department of Biochemistry and Molecular Genetics, University of Alabama Birmingham, Birmingham, AL 35233, USA. ✉ e-mail: egelman@virginia.edu; mart.krupovic@pasteur.fr

Conjugation is a major mechanism of horizontal gene transfer promoting the spread of antibiotic resistance among human pathogens. It involves establishing a junction between a donor and a recipient cell via an extracellular appendage known as the mating pilus. In bacteria, the conjugation machinery is encoded by plasmids or transposons and typically mediates the transfer of cognate mobile genetic elements. Much less is known about conjugation in archaea. Here, we determine atomic structures by cryo-electron microscopy of three conjugative pili, two from hyperthermophilic archaea (*Aeropyrum pernix* and *Pyrobaculum calidifontis*) and one encoded by the Ti plasmid of the bacterium *Agrobacterium tumefaciens*, and show that the archaeal pili are homologous to bacterial mating pili. However, the archaeal conjugation machinery, known as Ced, has been ‘domesticated’, that is, the genes for the conjugation machinery are encoded on the chromosome rather than on mobile genetic elements, and mediates the transfer of cellular DNA.

Introduction

The importance of horizontal gene transfer (HGT) in microbial persistence and evolution cannot be overstated. Exchange of genetic information is essential for the survival of microbial populations that otherwise succumb to Muller’s ratchet, a process whereby the irreversible accumulation of deleterious mutations leads to extinction of an asexual population^{1,2}. Furthermore, HGT plays a major role during the adaptation of microbes to constantly changing environmental conditions by providing immediate access to beneficial traits and promoting cooperation within microbial communities³. Accordingly, bacteria and archaea have evolved dedicated mechanisms of HGT^{4,5}. Traditionally, three major routes of HGT are recognized, namely, natural transformation, transduction, and conjugation. Whereas transformation is a natural ability of cells to uptake exogenous DNA from the environment through a dedicated competence system^{6–8}, the other two HGT mechanisms rely on distinct types of mobile genetic elements (MGE), viruses and plasmids (or integrative and conjugative elements), respectively. An additional HGT route, which is gaining

increasing recognition, is intercellular DNA transfer through membrane-bound extracellular vesicles⁹⁻¹².

In bacteria, conjugation is one of the main mechanisms for the spread of antibiotic resistance and other adaptive traits^{13,14}. Conjugation requires a sophisticated MGE-encoded apparatus, which belongs to the type IV secretion system (T4SS) superfamily, and in diderm bacteria consists of four key components: (i) a conjugative pilus, a multimeric assembly of the major pilin protein, which connects the donor and recipient cells and serves as a conduit for DNA transfer; (ii) the type IV coupling protein, an AAA + ATPase essential for pilus biogenesis and substrate transfer; (iii) the T4SS membrane-spanning protein complex enabling DNA transfer across the membrane of the donor cell; and (iv) the relaxosome, which nicks the double-stranded DNA (dsDNA), yielding the single-stranded DNA (ssDNA) substrate for intercellular transfer¹⁵⁻¹⁹. Conjugative elements have been identified as extrachromosomal plasmids or as integrated elements in certain archaea, including hyperthermophilic archaea of the order Sulfolobales²⁰⁻²⁴ and ammonia-oxidizing archaea of the class Nitrososphaeria²⁵, but the mechanism of conjugation has not been investigated in detail. Notably, none of the archaeal conjugative plasmids encode recognizable homologs of the relaxase or pilus protein and it has been suggested that the mechanism of conjugation in archaea might be different from that operating in bacteria²³.

Hyperthermophilic archaea of the order Sulfolobales have evolved a distinct DNA transfer system, named crenarchaeal exchange of DNA (Ced), which is dependent on species-specific cell aggregation and is inducible upon UV irradiation²⁶. The Ced system operates in conjunction with the UV-inducible type IV pili operon of Sulfolobales (Ups) system²⁷. The Ups pili produced upon UV irradiation mediate cellular aggregation in a species-specific manner, ensured by specific glycosylation patterns on the Ups pili and the protein S-layer, which covers the cellular membrane²⁸. Both Ced and Ups systems are required for efficient DNA exchange, but the two do not have to be expressed in the same cell²⁶. Notably, the Ced system mediates unidirectional import of DNA, which is then used as a template for genome repair by homologous recombination; cells that cannot exchange DNA show

significantly lower survival rates upon DNA damage²⁹. Notably, some crenarchaeal species encode the Ced but not the Ups system, whereas others, such as members of the order Thermoproteales, have not been found to encode either²⁶, suggesting alternative mechanisms for DNA exchange.

The Ced system consists of four proteins, Ceda, Ceda1, Ceda2, and CedB²⁶. Ceda contains six or seven transmembrane domains and is believed to form a transmembrane channel for DNA import, whereas CedB is homologous to VirB4/HerA-like AAA + ATPases and appears to

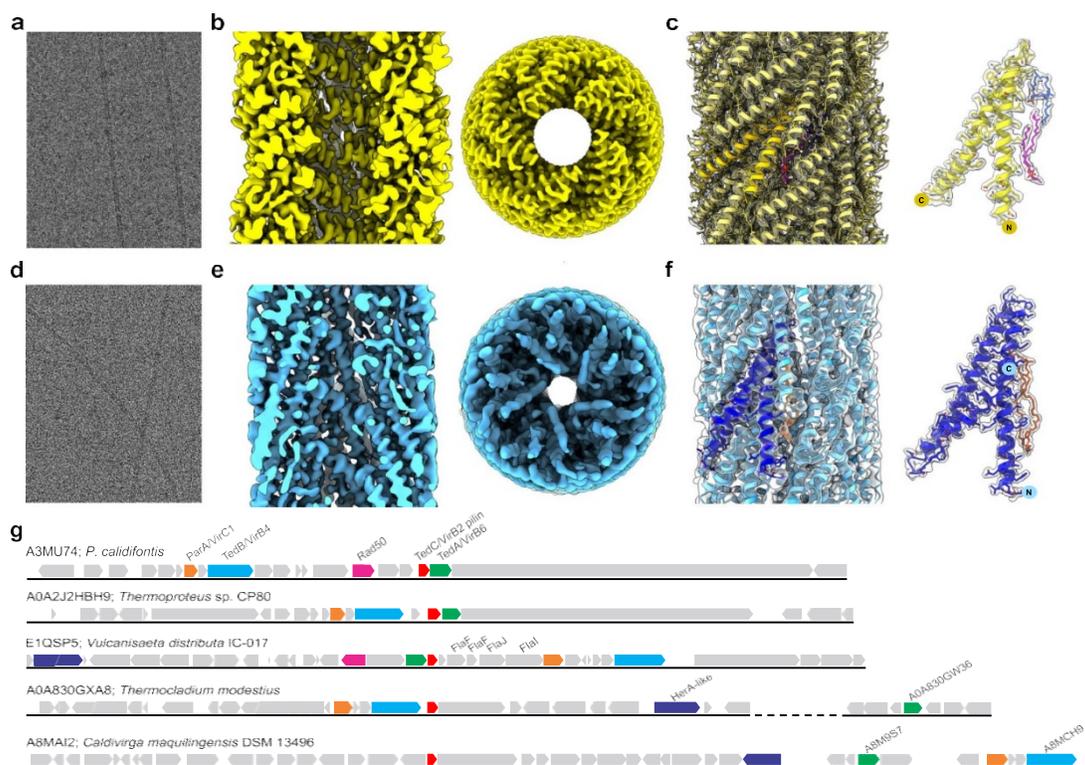


Figure 3.1|Archaeal conjugative pili. a, d Representative cryo-electron micrographs of *A. pernix* and *P. calidifontis* pili, respectively, from >8000 images collected for each. Scale bars, 50 nm. b, e Side and top views of the *A. pernix* (b, yellow) and *P. calidifontis* (e, blue) cryo-EM density maps at resolutions of 3.3 Å and 4.0 Å, respectively. The front half of the filament has been removed in the side views, so that

lumens are visible. c, f Atomic models in ribbon representation of the *A. pernix* (c, yellow) and *P. calidifontis* (f, blue) pili docked within their respective transparent cryo-EM density maps. The asymmetric unit of the *A. pernix* pilin CedA1 (c, bright yellow) shows two bound lipids (magenta and blue), while a single asymmetric unit of the *P. calidifontis* pilin TedC (f, dark blue) shows one bound tetraether lipid (orange). g Genomic loci encompassing the Ted system in different members of the order Thermoproteales. Each of the five genera within the Thermoproteales (genera *Caldivirga*, *Pyrobaculum*, *Thermocladium*, *Thermoproteus*, and *Vulcanisaeta*) is represented. Genes encoding the conserved VirB2/ CedA1-like pilin protein TedC, VirB6/CedA-like transmembrane channel TedA, and VirB4/CedB-like AAA + ATPase TedB are shown as red, green, and cyan arrows, respectively. Additional conserved genes are also indicated, including VirC1/ParAlike, Rad50, and HerA-like helicase which are shown as orange, magenta, and dark blue arrows, respectively. Other genes are shown in gray. Genomic loci are aligned using the TedC gene and indicated with the corresponding UniProt accession numbers, followed by the organism name. In some species, the components of the Ted systems are encoded within distal genomic loci which are separated from the TedC-encoding loci by dashed lines, and the corresponding genes are identified with their UniProt accession numbers.

power the DNA translocation across the membrane. The function of CedA1 and CedA2, each with two predicted transmembrane domains, is less clear, but they were shown to form a membrane-localized complex with CedA²⁶. The Ced system was considered to be unrelated to the bacterial conjugation system because of the opposite directionality of DNA transfer (import versus export, respectively) and the lack of homologs other than VirB4-like ATPase²⁶. However, how the DNA is transported between the cells within the Ups pili-mediated cellular aggregates and the nature of the channel connecting the donor and recipient cells remained unresolved. It also remained unclear whether the Ced system transfers dsDNA or ssDNA substrates.

Here, using cryo-electron microscopy (cryo-EM), we show that a protein from the hyperthermophilic archaeon *Aeropyrum pernix*, a homolog of CedA1, forms a pilus which is

structurally homologous to bacterial conjugative pili. We also discover that structurally similar pili, although with no sequence similarity, are produced by members of the Thermoproteales, which were previously not considered to encode the Ced-like system. We present high-resolution structures of two putative conjugative pili from hyperthermophilic archaea, *A. pernix* and *Pyrobaculum calidifontis*, and a bacterial conjugation pilus from a model system^{30–33}, encoded by the pTiC58 plasmid of *Agrobacterium tumefaciens*³⁴. It has been previously stated that the pilin subunit in the *A. tumefaciens* mating pilus is cyclic^{35,36}, and that this accounts for its robust stability³⁷. We show that it is not cyclic and is actually similar in fold to other bacterial and archaeal mating pili. Collectively, our results suggest that the archaeal Ced-like systems share a common ancestor with bacterial T4SS conjugation system. However, unlike in bacteria, where conjugation systems are proprietary to mobile genetic elements, in hyperthermophilic archaea the DNA transfer system has been domesticated, and we propose that this has evolved to ensure survival in extreme environments.

Results

Identification of putative DNA transfer pili in hyperthermophilic archaea in Sulfolobales, expression of the *ced* and *ups* genes is activated exclusively upon UV irradiation²⁶. We set out to study the behavior of the Ced system in *Aeropyrum pernix* (order Desulfurococcales), a hyperthermophilic archaeon that grows at temperatures up to 100 °C³⁸ and lacks the Ups system²⁶. Given that intercellular DNA transfer typically involves extracellular filaments, the extracellular fraction of *A. pernix* cells was analyzed using cryo-EM. In addition to the flagella³⁹, we identified a new type of filament (Figure 3.1a), not previously observed in archaea. The reconstruction of this pilus to 3.3 Å resolution allowed us to determine the pilin identity directly from the cryo-EM map. The pilin was identified as *A. pernix* protein APE_0220a (WP_010865579), an ortholog of the *S. acidocaldarius* protein CedA1 (WP_011277463), one of the conserved components of the Ced system previously thought to be an integral membrane protein²⁶.

We have previously shown that *Pyrobaculum calidifontis*, a hyperthermophilic archaeon of the

order Thermoproteales⁴⁰, which lacks both Ced and Ups systems, is prone to aggregation mediated by bundling pili related to TasA-like fibers, a major component of the biofilm matrix in many bacteria⁴¹. We thus explored whether *P. calidifontis* cells produce pili which could be involved in DNA transfer. Cryo-EM analysis of the *P. calidifontis* filament preparation revealed pili (Figure 3.1d), which following the reconstruction to 4.0 Å resolution (Figure 3.1e), proved to be structurally similar to the CedA1 pili of *A. pernix* (Figure 3.1b, SI Figure 3-1a). While in the *A. pernix* filament the helical rise and twist per subunit were 3.6 Å and 76.5°, respectively, in the *P. calidifontis* filament these parameters were 5.0 Å and 74.2°. From the secondary structure and side-chain information present, we were able to determine the pilin identity directly from the cryo-EM map using DeepTracer-ID³⁹ to be *P. calidifontis* protein Pcal_0765 (WP_011849449) (Figure 3.1f), which we name TedC (see below). Although TedC displays a similar fold to CedA1 from *A. pernix* (SI Figure 3-1a), the two pilins are processed differently, with *A. pernix* pilin not being processed and the *P. calidifontis* pilin, similar to bacterial plasmid conjugative pilins, undergoing proteolytic cleavage. Indeed, SignalP analysis⁴² predicts that TedC carries a cleavable signal peptide, with the predicted signal peptidase I cleavage site, 20AVA ↓ QA-24 (cleavage site probability of 0.97; SI Figure 3-1b). Since the atomic model built into the reconstruction starts at residue 38, rather than the predicted residue 23, we conclude that residues 23-37 are disordered and therefore not visualized in the density map.

Sequences homologous to TedC were identified using BLASTP searches in members of all five genera of the order Thermoproteales, namely, *Pyrobaculum*, *Thermoproteus*, *Caldivirga*, *Vulcanisaeta*, and *Thermocladium*. Genomic neighborhood analysis (Figure 3.1g) showed that the gene downstream of the *tedC* encodes a protein with seven predicted transmembrane domains, similar to CedA of *S. acidocaldarius* and *A. pernix*. Although BLASTP searches did not reveal the relationship between Pcal_0766 and CedA, sensitive profile-profile comparisons showed that the two proteins are indeed homologous (HHpred probability: 98.6), despite negligible pairwise sequence identity of 13% (SI Figure 3-2a). Notably, profile-profile comparisons of Pcal_0766 against the PDB database showed that it is distantly related to VirB6-like proteins encoded by bacterial conjugative

plasmids and involved in formation of the mating pore complex (SI Figure 3-2b). A gene encoding the VirB4-like ATPase was identified transcriptionally upstream of the *tedC* and *cedA*-like genes, separated by a few genes (Figure 3.1g). In *Thermocladium* species, the ortholog of *Pyrobaculum virB4*-like gene is adjacent to the *tedC* pilin gene, suggesting that the corresponding proteins function together. Given the high sequence divergence between the components of the Ced system of Sulfolobales and the related system of Thermoproteales, we termed the latter as Ted, for Thermoproteales exchange of DNA system, with the CedA-like, CedA1-like, and VirB4-like components as TedA, TedC, and TedB, respectively. CedA2 is not conserved in the Ced systems from different species²⁶, and homologs or even counterparts of this protein are not identifiable in the Ted system.

Genomic loci containing the Ted system also commonly include genes encoding homologs of HerA helicase and MinD/ParA family ATPases, whereas *Pyrobaculum* and *Vulcanisaeta* species in addition carry the *rad50* recombinase genes. HerA helicase and Rad50 recombinase play an essential role during homologous recombination in hyperthermophilic archaea^{43,44}. The co-localization of these genes with the Ted system suggests a coordination of the DNA import in Thermoproteales and DNA repair by homologous recombination. In the *A. tumefaciens* systems, MinD/ParA family ATPase, known as VirC1, spatially coordinates early conjugative DNA transfer reactions⁴⁵. The finding that archaeal Ced and Ted systems form pili suggests that DNA transfer through these systems might be more similar to bacterial plasmid-mediated conjugation than previously recognized.

Archaeal conjugation pili are stoichiometric complexes of pilins and lipids

With the atomic models for the pilin subunits docked within the respective cryo-EM maps, we observed unaccounted-for densities between each of the proteins in both maps. These densities were similar to the densities for lipid molecules found in each of the previously reported bacterial conjugation pili, where there is a stoichiometric 1:1 ratio of pilin:phospholipid^{46,47}. However, the putative lipid densities in the archaeal conjugation pili were larger. It has been impossible to do the

lipidomics analysis for the archaeal conjugation pili, due to our inability at this point to obtain a highly enriched preparation containing just the conjugation pili. However, given that *Aeropyrum* and *Pyrobaculum* contain only one membrane, the lipids in the pili must come from the archaeal cytoplasmic membrane.

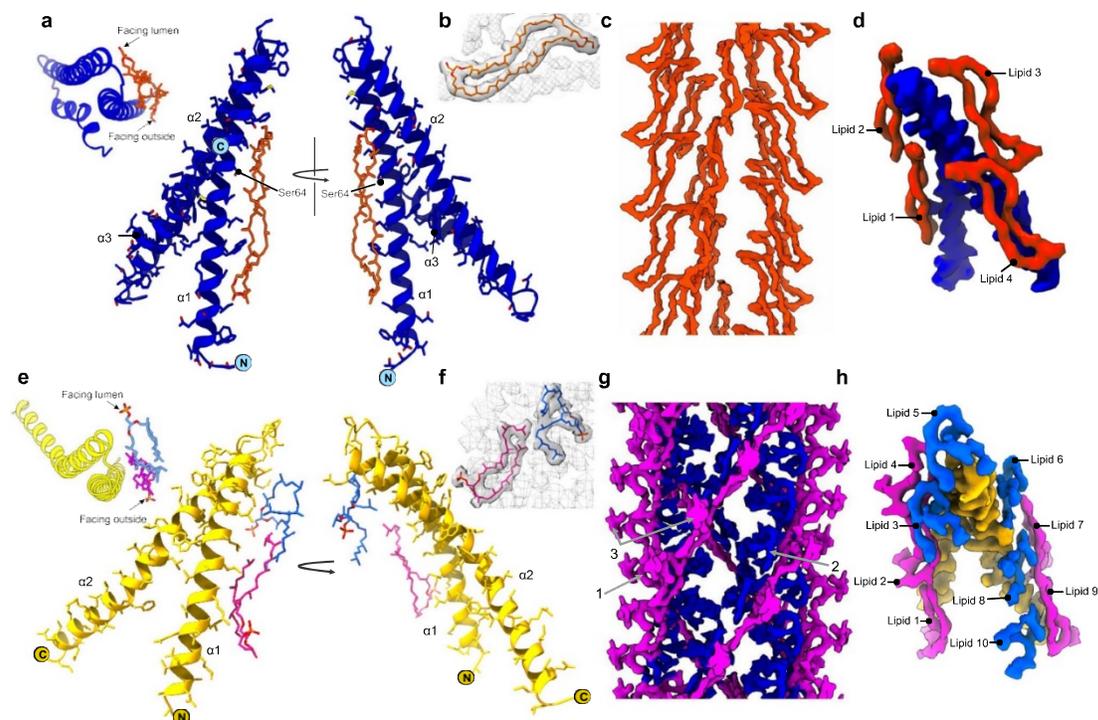


Figure 3.2| Intricate pilin-lipid interaction networks within archaeal conjugative pili. a Front and back views of a single asymmetric unit of the *P. calidifontis* pilus (blue) containing one pilin (TedC) and one phospholipid GDGT-0 (orange). One polar headgroup of the tetraether lipid faces the lumen and the other faces the outside of the pilus. A top view of a subunit, looking down the helical axis, is shown on the left for orientation. The isoprenoid chains of the GDGT-0 lipids are buried between hydrophobic helices and interact closely with helix $\alpha 1$ of the pilin shown but will also interact with $\alpha 2$ of neighboring pilins. b An atomic model of the cyclic GDGT-0 lipid docked within the lipid density. c Lipid density from *P. calidifontis* (blue). The density is very well resolved and shows that the GDGT-0 lipids have one head group facing the outside of the filament and the other head group facing the lumen. d

A single pilin (blue) contacts four surrounding GDGT-0 lipids (orange). e The front and back view of the CedA1 pilin (yellow) of *A. pernix*. The asymmetric unit contains two lipids and one pilin, with the lipid in two different conformations: one having a partially folded shape (blue) and the other a crescent-like shape (magenta). The crescent head group is facing the outside of the pilus while the isoprenoid chains are buried between the pilin subunits. The partially folded lipid's phosphate head group is facing the lumen of the pilus and the isoprenoid chains are bent and buried between the subunits. A top view of a subunit, looking down the helical axis, is shown on the left for orientation. For both lipids the contacts with the pilin are mediated by hydrophobic residues such as leucine, isoleucine and valine from helices $\alpha 1$ and $\alpha 2$. The crescent-like lipid has one hydrophobic interaction with the partially folded lipid. f Atomic models for the crescent-shaped and partially folded lipid docked into the lipid densities. g Lipid density from *A. pernix* (yellow). The lipid density is less resolved than in *P. calidifontis* but shows two C25-C25-diether lipids, one of which forms a crescent-like shape (arrow 1) and the other forms a partially folded shape (arrow 2). Both lipids are capped with extra density (arrow 3) which is likely a dihexose sugar attached to the phosphate head group. h A single protein subunit (yellow) makes contacts with ten lipids (five crescent-shaped (pink), and five partially folded (blue)).

In *P. calidifontis* the extended density could only be explained by a bipolar cyclic tetraether lipid (Figure 3.2a–d). Bipolar archaeal lipids were proposed almost 40 years ago⁴⁸, but, to the best of our knowledge, they have never been directly visualized. The cryo-EM map was good enough to identify the cyclic lipid as a glycerol dialkyl glycerol tetraether species (GDGT), a dominant membrane lipid in many hyperthermophilic archaea, including *Pyrobaculum* species⁴⁹, and we have modeled the simplest form, GDGT-0, into the cryo-EM density (Figure 3.2b). It is important to note the possibility that more complex forms of GDGT (GDGT-1 through 8) containing cyclopentane rings may fit into the density as well. Both head groups are found to be solvent-exposed, with one of the polar head groups facing the lumen and the other facing the outside of the pilus, while the acyl chains

are buried between the hydrophobic helices of the pilin subunits (Figure 3.2a). The GDGT-0 lipid is found positioned in the middle relative to helix $\alpha 1$ (Figure 3.2a). Most of the pilin-lipid interaction network in the *P. calidifontis* pilus relies extensively on contacts with helix $\alpha 1$ in the asymmetric unit (ASU), but the lipid is sandwiched between $\alpha 1$ and $\alpha 2$ of a subunit in the neighboring ASU. All contacts are primarily associated with hydrophobic residues apart from one polar electrically neutral serine residue, Ser64 (Figure 3.2a). There appears to be minimal contact with the charged head groups. Thus, the protein-lipid interaction stabilization relies heavily on hydrophobic interactions for *P. calidifontis*. While the ASU consists of one protein to one lipid, there is a total of four lipids that make contact with the protein subunit TedC (Figure 3.2d).

Interestingly, the lipid density was less resolved for *A. pernix* compared to the lipid density for *P. calidifontis*, even though the overall resolution for the *A. pernix* map was higher, 3.3 Å vs 4.0 Å, as determined by the map:map FSC (Table S3-1). In *A. pernix*, there are two lipids in every ASU, one of which adopts a partially folded conformation, with one of the isoprenoid chains folding back on itself, while the other has a crescent-like shape (Figure 3.2e–g). Unlike most other members of the phylum Thermoproteota, the membranes of *Aeropyrum* species contain only a small amount of GDGT lipids. Instead, an unusual C25, C25-diether with a phospho-dihexose head group was the main lipid species observed by mass spectrometry of a

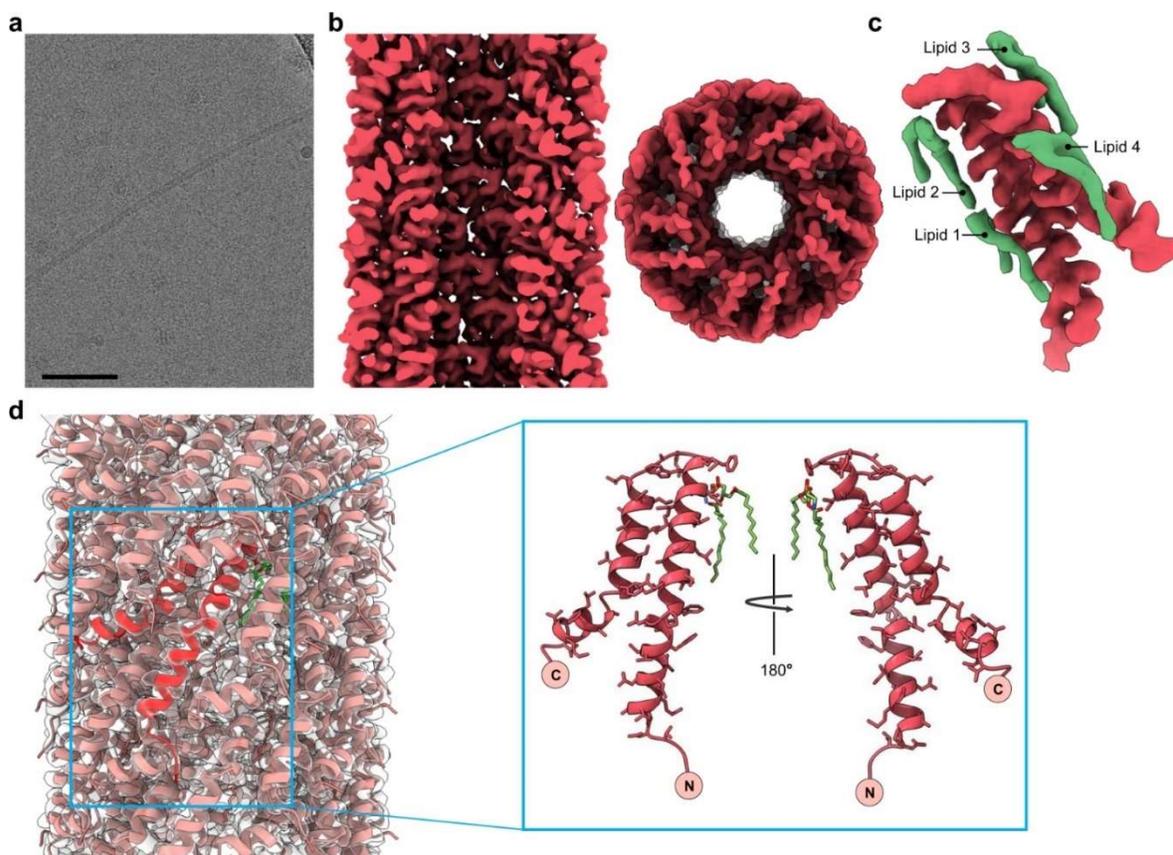


Figure 3.3 | Structure of the *A. tumefaciens* T-pilus. a Representative cryo-electron micrograph of *A. tumefaciens* T-pilus, from 8,363 images collected. Scale bar, 50 nm. b Side and top view of the T-pilus cryo-EM density map at a resolution of 3.5 Å. The front half of the filament has been removed in the side view, and we are looking at the lumen. c While there is a 1:1 stoichiometry of lipids to pilins in the filament, a pilin (red) makes contact with four lipids (green). d Atomic model of the T-pilus in ribbon representation docked within the transparent cryo-EM density map. A single subunit model is shown in red. An inset on the right shows front and back views of the VirB2 subunit.

cellular membrane preparation of *A. pernix* (SI Figure 3-3a–b), consistent with the previous identification of this lipid in *A. pernix*⁵⁰. Both non-protein densities in the *A. pernix* map could be fit with this lipid, suggesting that the same lipid is present in two different conformations (Figure 3.2f). The crescent-shaped lipid head group is directed toward the extracellular space with the isoprenoid

chains extended and buried between pilin subunits (Figure 3.2e). The crescent-shaped lipid primarily contacts helix $\alpha 1$ of the pilin, with one potential hydrophobic interaction with the partially folded lipid (Figure 3.2e). The contacts with helix $\alpha 1$ are mediated by leucine-rich hydrophobic interactions and one charged interaction between the phosphate head group and the positively charged Lys15. The second lipid density resulted in a model with bent isoprenoid chains and the phosphate head group directed toward the lumen (Figure 3.2e). The isoprenoid chains are buried between the subunits. Contacts with the pilin are made exclusively by the hydrophobic residues leucine, isoleucine, and valine from helices $\alpha 1$ and $\alpha 2$ (Figure 3.2e). There are no observed contacts between the pilin and the lumen-facing phosphate head group of the second lipid. There are two lipids and one protein subunit in the ASU of the *A. pernix* CedA1 pilus, generating a complex network of lipid contacts for each protein subunit (Figure 3.2g–h). In total, each CedA1 subunit will make contacts with 10 lipid molecules (Figure 3.2h).

In addition to the lipids, weak peripheral density was observed for both of the archaeal conjugation pili (SI Figure 3-4) which would be consistent with glycosylation. However, this density was diffuse, and we could not see clear additional density on specific residues such as serine, threonine, or glutamine that might be targets of such glycosylation⁵¹.

Conjugation pili of *Agrobacterium tumefaciens*

To extend the comparison between the archaeal mating pili and the existing structures of bacterial ones, we used cryo-EM to solve the structure of the *A. tumefaciens* T-pilus encoded by the pTiC58 plasmid to 3.5 Å resolution (Figure 3.3a, b; Table S1). The T-pilus has a fivefold rotational symmetry with a rise of 13.7 Å and a twist of 32.5° per subunit. Similar to other bacterial conjugation pili^{46,47}, there is a stoichiometry of one lipid molecule to each protein subunit. However, unlike in pED208⁴⁶ and pKpQIL⁴⁷, where each protein subunit contacts five lipid molecules, in T-pilus, each pilin subunit contacts four lipid molecules (Figure 3.3c), in a fashion similar to the TedC pilins (Figure 3.2d).

We conducted comprehensive shotgun lipidomics by electrospray ionization mass spectrometry

(ESI-MS) on isolated pili to identify and detail the lipid species tightly associated with fibrillar pili-forming proteins. The isolated pili were pre-treated with PLA2 to hydrolyze all (contaminating) phospholipids that were not stably associated with pili proteins. As a control, we analyzed isolated pili from pED208, whose associated lipids were previously reported as being two subspecies of the anionic lipid phosphatidyl-glycerol (PG) (Costa et al., 2016). Our analysis (SI Figure 3-5) confirmed and extended these results, showing that these two species (PG 16:0/16:1 and PG 16:0/ 18:1) are indeed the most abundant lipids associated with pED208, and that three other similar PG species are also present. Thus, ESI-MS allows quantitative analysis of lipid species associated with bacterial pili. The same analysis applied to VirB2 T-pili (SI Figure 3-5) revealed a strikingly different set of PLA2-resistant lipids, with >65% of the phospholipids being phosphatidyl-ethanolamine (PE). Interestingly, the acyl chains of T-pilus lipids (PE 16:0/16:1 and PE 16:0/18:1) were similar to those observed in pED208, despite having different chemistry of the headgroup. A minor fraction (<20%) of PLA2-resistant lipids were phosphatidyl-cholines (PC).

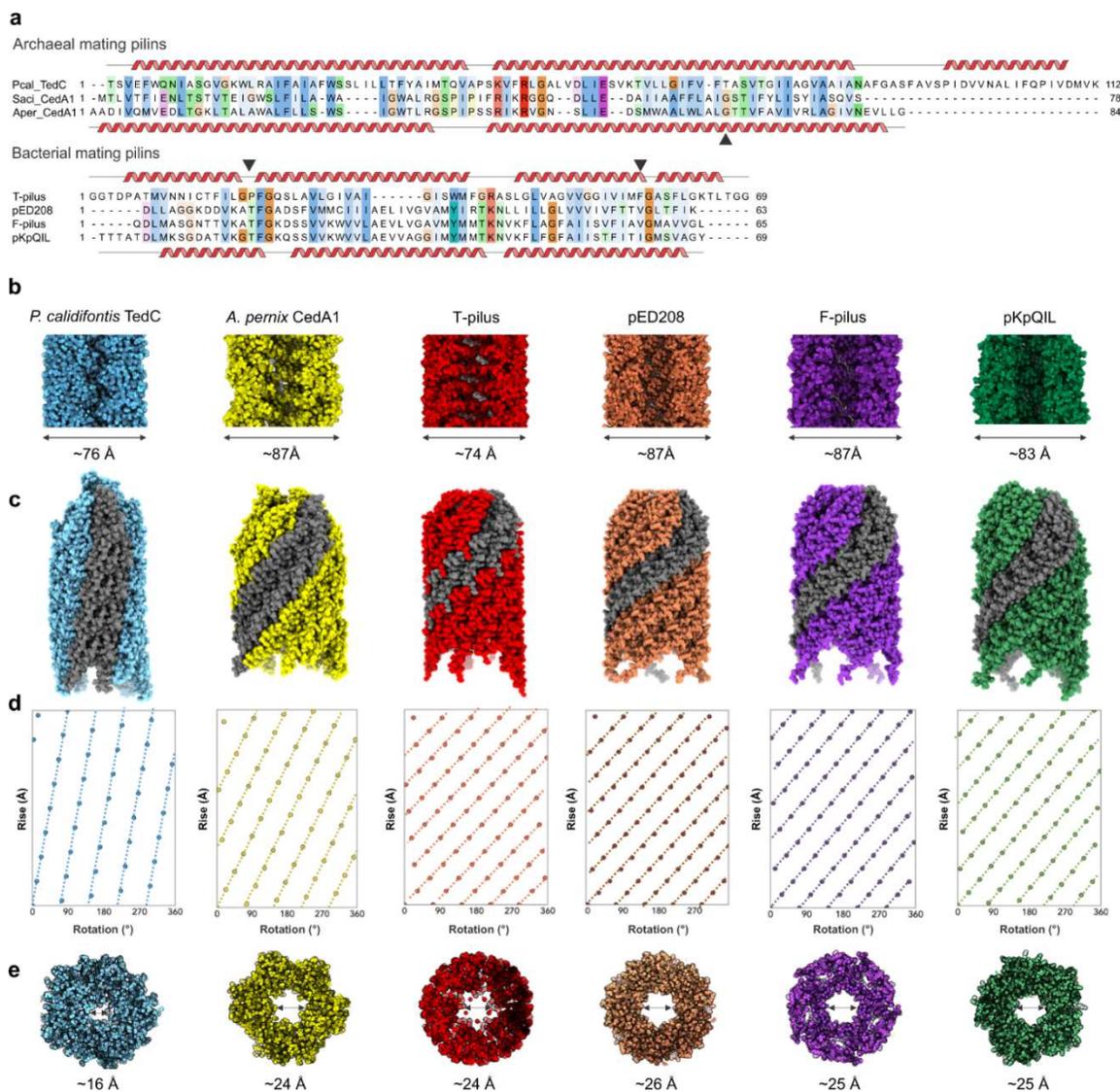


Figure 3.4 | Comparison of archaeal and bacterial conjugative pili. **a** Sequence alignments of archaeal (top) and bacterial (bottom) mating pili. Sequences of mature pilins are shown with the secondary structure elements determined from the structure for the pilins shown above and below the corresponding sequences. Kinks in the α -helices are indicated with triangles. **b**, **e** Comparison of the lumen and outer diameter of archaeal *P. calidifontis* (blue) and *A. pernix* (yellow) with bacterial pili: *A. tumefaciens* T-pilus (red), pED208 (mauve), F-pilus (purple) and pKpQIL (green). The outer diameters range between 74 Å to 87 Å. The lumen diameters range between 16 Å to 26 Å, but in some cases cannot be easily reduced to a single number. **c**, **d** Atomic models with a single strand, shown in gray, show connectivity between the subunits. This connectivity is also represented with helical nets for the archaeal and bacterial

pili. The helical nets show the unrolled surface lattice viewed from the outside of the filament. Each point represents a subunit, and the dotted lines are drawn to highlight the fact that all have right-handed 5-start helices. All of the pili have substantial connectivity between subunits along these 5-start helices.

It is important to emphasize the essential need for PLA2-catalyzed hydrolysis of contaminating phospholipids in these assays, as the pili-bound lipids are relatively unabundant, and even minor membrane contamination will overwhelm the signal. We, therefore, expect that in the absence of phospholipase treatment, any lipidomics analysis of the T-pilus will mainly find the lipids present in the contaminating membrane blebs and vesicles. This might explain the large difference between our results, where PE is the dominant lipid in the pili, and that found in two other very recent reports, where either PC⁵² or PG⁵³ were found as the dominant lipid in the T-pilus. In neither of these two studies was phospholipase treatment used to minimize the contribution of contaminant lipids.

The 3.5 Å resolution of the T-pilus map allowed for unambiguous model building and interpretation of the protein subunit, VirB2 (Figure 3.3d). The atomic model of VirB2 shows clear structural homology to the TraA subunit of the F-pilus (SI Figure 3-1c). One clear difference between VirB2 on the one hand, and TraA of pED208 on the other, is that there are kinks in helix 1 and 3 of the VirB2 subunit, produced by Pro23 and Phe61, respectively (SI Figure 3-1c). The electrostatic surface for the lumen of the T-pilus generated by the atomic model is overall quite positive. The headgroup of PE is zwitterionic, with a net neutral charge, and the addition of the PE lipid into the T-pilus model does not significantly change this electrostatic surface, which results in a more positively charged lumen compared to the lumen of pED208⁴⁶ (SI Figure 3-6). Interestingly, when viewed from the top there is an alternating positive to a negative charge arising from the arginine and the lipid, respectively (SI Figure 3-6). The T-pilus is known for its diverse substrates, enabling transport of both ssDNA-VirE2⁵⁴ as well as effector proteins⁵⁵ whose transport is independent of the DNA³⁰. The inclusion of VirE2 is thought to protect the T-strand from nuclease degradation in *A. tumefaciens* and

facilitate its transfer through the T4SS and T-pilus⁵⁴. An AlphaFold prediction for the structure of VirE2 suggests that it would be too large to pass through the lumen of the T-pilus in a natively-folded state. We, therefore, suggest that it must be partially unfolded to allow for such transport. While the lumen of other bacterial mating pili may have evolved to be optimal for DNA transport, the lumen of the T-pilus, also used for the transport of a diverse range of other substrates, has a positive electrostatic potential that would still allow DNA transfer, but not be optimal due to the greater friction resulting from DNA sticking to the walls. Obviously, this would suggest that the other substrates are likely to have overall positive electrostatic surfaces, and it has previously been noted that the effector proteins, where characterized, carry C-terminal domains that are positively charged³⁰.

The high resilience of the T-pilus to extreme chemical or physical conditions has been reported in a previous study³⁷, and its stability was attributed to the putative cyclic nature of the T-pilin^{35,36}. Surprisingly, the cryo-EM structure of VirB2 reveals no cyclization of the pilin. Similar to other bacterial conjugative pili and TedC of *P. calidifontis*, VirB2 is proteolytically processed by a signal peptidase⁵⁶. In the mature pilin, residues QSAG from the N-terminus and G from the C-terminus are not seen in the density map, most likely due to disorder, and have not been built into the atomic model. The two residues suggested to be covalently linked in the T-pilus³⁵ were Gln1 and Gln74 (using our numbering for the mature pilin), neither of which are in our atomic model. But the five missing residues would be unable to span the distance of ~34 Å between Gly5 and Gly73 (the first N-terminal residue and last C-terminal residue in the model), so it is not possible that the subunit really is cyclic. This leads us to the question of what is the basis for the resiliency of the T-pilus? To answer this question, we subjected the T-pilus to many of the same extreme chemical or physical conditions reported by Lai and Kado³⁷ (SI Figure 3-7Table S3-2. **Environmental effects on pilus structure**. Table S3-2) and compared the resiliency to the well-studied F-pilus which is known to be non-cyclic. Interestingly, the T-pili look intact under 50% glycerol and 4 M urea, flexuous and partially degraded under high temperatures (70 °C) and 0.1% SDS, and depolymerized under 1% Triton-X-100. In contrast, the F-pili is much more stable and remain intact under all conditions, including 1% Triton-X-100 (SI Figure 3-7Table S3-2. **Environmental effects**

on pilus structure. Table S3-2). These results suggest that the architecture of bacterial conjugation pili is generally very stable and not related to cyclization. Further, we see no evidence of extensive glycosylation for the T-pilus (SI Figure 3-4), nor was any potential glycosylation described for previous conjugation pili structures^{46,47}, so we can exclude extensive glycosylation which has been suggested as a mechanism for stabilizing extracellular archaeal filaments in the most extreme environments⁵¹.

The lumen of prokaryotic pili is too narrow to transfer dsDNA. With structures for four bacterial conjugation pili (pED208, F, pKpQIL, and pTiC58 from *A. tumefaciens*) and two archaeal ones (*A. pernix* and *P. calidifontis*), it is clear that there is a common architecture for all and obvious homology, despite negligible sequence similarity (Figure 3.4a). Each prokaryotic pilin subunit consists of two or three hydrophobic α -helices, with kinks appearing in some of the helices, such as in *A. pernix* and *A. tumefaciens* (Figure 3.2a,d, 4.3d and 4.4a). When a heat-map of global structural similarity (based upon the Dali server⁵⁷ Z-scores) between a single pilin from all known structures of prokaryotic conjugation pili was generated, unsurprisingly a clustering of the archaeal pilins was observed (SI Figure 3-8). Notably, however, the mature *P. calidifontis* pilin contains an additional hydrophobic α -helix (α 3) compared to pilins encoded by *A. pernix* and bacterial plasmids. Despite the *A. tumefaciens* pilin having high structural similarity with pED208 and pKpQIL, it was clustered with the archaeal pilins rather than bacterial pilins from the F-pilus, pED208, and pKpQIL, which formed a separate cluster. The clustering of *P. calidifontis*, *A. pernix*, and *A. tumefaciens* pilins is likely due to the helix kinks which are absent in the other bacterial pilins.

We compared the external diameters of the archaeal and bacterial (T-pilus, pED208, F-pilus, and pKpQIL) conjugation pili as well as the diameters of their central channels, which allow the transfer of DNA (Figure 3.4b, e). It must be noted that all such measurements are quite approximate, as discussed recently⁵⁸, and usually ignore the contribution of hydrogens and tightly bound water molecules. Further, structural varicosity in both archaeal pili complicates reducing the diameter to a single

number. Nonetheless, with unavoidable approximations, lumen diameters (Figure 3.4b, e), range from 16 Å to 26 Å. The outer diameter of the *A. pernix* pilus more closely resembles the outer diameters of the bacterial F and F-like pili, while *P. calidifontis* has an outer diameter that is approximately the same as the T-pilus. The lumen diameter of *A. pernix* pilus is very similar to that of the T-pilus, whereas the lumen diameter of *P. calidifontis* is considerably narrower (Figure 3.4e).

To determine whether dsDNA could pass through the lumen of the prokaryotic conjugative pili, we placed a model for B-form dsDNA within the lumen of the conjugation pili models from *P. calidifontis*, *A. pernix*, and *A. tumefaciens*. We observed extensive clashes between dsDNA and the atomic surface of all models (

SI Figure 3-9). These observations suggest that ssDNA, not dsDNA, passes through conjugative pili, although no experimental evidence exists to support this notion. Indeed, using synthetic nanopores in ultrathin silicon nitride membranes, it has been shown that ssDNA permeates pores with diameters as small as 10 Å⁵⁹. Hence, pore diameters of all prokaryotic conjugative pili are sufficiently wide for the passage of ssDNA, but not dsDNA. Previous structural studies have suggested that the lumen of the F-pilus has an overall negative charge which contributes to a repulsive force that will keep negatively charged ssDNA away from the wall of the lumen, effectively lowering friction^{46,47}. In *A. pernix* each pilin is bound to two diether lipids with phosphorylated dihexose (i.e., glucose-inositol) head groups. One of these lipids faces the lumen of the central channel where it might also provide a similar negative charge to facilitate DNA transfer (Figure 3.2e).

Different helical symmetries still allow quasi-equivalent interactions

The twist of the archaeal pili from both *P. calidifontis* (74.2°) and *A. pernix* (76.5°) is similar to that of the pKpQIL pilus from *K. pneumoniae* (77.6°), generating strong connectivity along 5-start

protofilaments (Figure 3.4c,d). Similar 5-start protofilaments are observed in other bacterial conjugation pili as well, including the *A. tumefaciens* T-pilus reported here, the pED208 F-like pilus⁴⁶, and the F-pilus⁴⁶. However, those pili have a C5 rotational symmetry rather than the 1-start

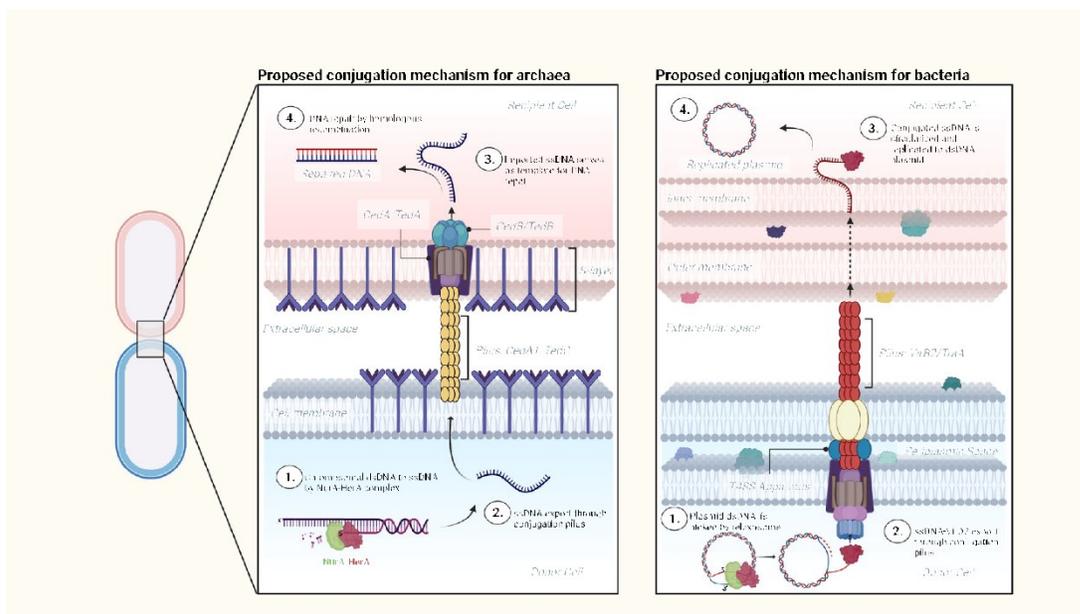


Figure 3.5| Proposed conjugation mechanisms between donor and recipient cells in archaea (left) and bacteria (right). The schematic shows how ssDNA substrates are generated by the HerA-NurA machinery in the donor archaeal cells and by the plasmid-encoded relaxosome in bacteria. Note that CedA and, potentially, Ted systems function as DNA importers rather than DNA exporters, contrary to the bacterial plasmid conjugative machinery.

helical symmetry observed in pKpQIL⁴⁷, *P. calidifontis*, and *A. pernix* (Figure 3.4c, d). The difference in helical symmetries results in very small differences in intermolecular interfaces, much like what is observed for pED208 and pKpQIL, which have C5 and C1 symmetries, respectively⁴⁷. This is similar to the quasi-equivalence phenomenon reported in other helical tubes made of helix-turn-helix subunits, such as in archaeal virus SMV1⁶⁰, where very similar interfaces can be preserved even though there are large apparent changes in symmetry (between C7 and C1 in SMV1).

Discussion

Bacterial conjugation machinery is nearly exclusively encoded by plasmids or integrating and conjugating elements and hence typically promote the transfer of the cognate mobile genetic elements, with the occasional transfer of the host DNA, e.g., when the plasmid origin of transfer (*oriT*) is recombined into the cellular chromosome⁶¹. An exception to this paradigm is the conjugation-like process, dubbed distributive conjugal transfer, described in *Mycobacterium* species, whereby large fragments of the chromosomal DNA are exchanged between the donor and recipient species⁶². However, although superficially this form of DNA transfer is reminiscent of the classical plasmid conjugation, it is mediated by a poorly understood mechanism involving the Type VII secretion system, rather than the T4SS⁶³. Here we show that the Ced system which imports DNA in hyperthermophilic archaea of the order Sulfolobales²⁶, and the related Ted system of Thermoproteales described herein, are domesticated derivatives of the T4SS. Both Ced and Ted systems encode homologs of the VirB4 ATPase (CedB/TedB), VirB6 membrane pore (CedA/TedA), and VirB2 conjugative pilin (CedA1/TedC). Nevertheless, the individual components display no recognizable sequence similarity between the Ced and Ted systems, and the assembly pathways might also differ. For instance, whereas the pilins in the Ced system are apparently secreted without processing, the N-terminal signal sequence of TedC appears to be cleaved by a signal peptidase I. In this respect, TedC is more similar to the bacterial plasmid VirB2-like pilins. It is interesting to note that the N-terminal region of CedA proteins is homologous to nearly entirety of the CedA1 pilin (SI Figure 3-2c), suggesting a common export pathway for CedA and CedA1 and that one has evolved from the other. Notably, the same result was obtained when VirB2 and VirB6 of *A. tumefaciens* were compared, albeit with a lower significance score (SI Figure 3-2d). It is not possible to claim with confidence which of the two proteins is ancestral. However, given the central role of the VirB6-like pore for conjugation and the absence of conjugative pili in monoderm bacteria, it is tempting to suggest that the pilin is a more recent addition to the T4SS apparatus to facilitate the DNA transfer between spatially

separated cells. According to this scenario, the stoichiometric incorporation of lipids, a unique feature of bacterial conjugation pili and their archaeal homologs, might be a vestige of the ancestral function of these pilins as bona fide membrane proteins. Notably, our results provide the first direct visualization of archaeal GDGT lipids in the *P. calidifontis* pilus and highlight the flexibility of diether lipids in the pilus of *A. pernix*. The pilins are surrounded by lipid molecules, with most interactions holding the pili together being between pilin subunits and lipids. Thus, in a way, the conjugative pili can be regarded as highly ordered extensions of the cytoplasmic membrane.

Given the direct visualization of horizontal gene transfer between spatially separated bacterial cells using fluorescence microscopy⁶⁴, it is clear that bacterial conjugation pili can act as conduits for DNA transfer. However, the possibility still exists that the main role of such pili is to depolymerize and bring two mating cells into physical juxtaposition, and that the bulk of DNA transfer only takes place when this conjugation junction is established. Our structural results cannot address this possibility. Many of the proteins involved in conjugation have been identified in bacteria, such as in *E. coli*^{65,66} and *A. tumefaciens*⁶⁷, but these proteins have remained elusive in archaea. Mutagenesis studies in bacteria have shown ssDNA to be the genetic material exported from the donor to the recipient cell via the conjugation pilus^{64,68}, where dsDNA is nicked by the enzyme relaxase. Relaxase in complex with several other proteins, known collectively as the relaxosome, is responsible for mediating the unwinding of ssDNA. To our knowledge, there are no apparent homologs of bacterial relaxases encoded in the archaeal genomes or conjugative plasmids²³. Nevertheless, our data indicate that archaeal pili, similar to their bacterial counterparts, most likely transfer ssDNA. The observation that Ted genes in some genera of Thermoproteales co-occur with genes encoding Rad50 and HerA-like helicase, might hold a clue to this conundrum. In hyperthermophilic archaea, *herA* and *rad50* usually form an operon with the genes encoding nuclease NurA and Mre11, and the four proteins function during DNA damage repair through homologous recombination⁶⁹. NurA, an RNase H-fold nuclease, is endowed with the endonuclease and exonuclease activities that are modulated by the HerA helicase^{70,71}. The integrated activity of NurA-HerA is responsible for DNA end-resection, a process that generates the 3' single-

stranded tails that are subsequently coated by the Rad50 recombinase to initiate strand invasion and DNA repair. In Sulfolobales, expression of the HerA operon, Ced, and Ups systems is activated by DNA damage and all three operons are coregulated by the transcription factor B3⁷². We hypothesize that the ssDNA substrate for the transfer through Ced and Ted systems is generated by the activity of NurA-HerA system, rather than by a dedicated relaxosome as in the case of bacterial plasmid conjugation systems (Figure 3.5). Notably, Ted system, but not Ced, apparently includes a ParA/MinD-family ATPase related to the VirC1 protein of *A. tumefaciens* plasmid Ti, which functions during the delivery of the relaxosome-bound ssDNA to the T4SS complex⁴⁵, and a similar role can be postulated for the homologous protein of the Ted system. The narrower pore in the *P. calidifontis* mating pilus, compared to *A. pernix* and the bacterial mating pili (Figure 3.4,

SI Figure 3-9), may reflect the fact that only ssDNA is transferred, rather than a relaxase-ssDNA complex.

Our current study provides new insights into the mechanism of horizontal DNA transfer in hyperthermophilic archaea through the domesticated conjugative T4SS apparatus. Such domestication is a remarkable example of the ‘guns-for-hire’ paradigm⁷³, whereby molecular machines evolving at the interface of the interaction between mobile genetic elements and their hosts are captured and repurposed by the competing parties. To our knowledge, the domestication of the conjugative apparatus for DNA transfer has not been reported in other organisms. Many questions remain unanswered, including the generation of ssDNA substrates for intercellular transfer as well as the mechanistic details of the biogenesis and full molecular complexity of the Ced and Ted systems. Notably, none of the species in the orders Thermoproteales and Desulfurococcales, including members of the genera *Pyrobaculum* and *Aeropyrum*, are genetically tractable. Thus, our present study further highlights the utility of cryo-EM in gaining important insights into the biology of non-model organisms.

Methods

Cultivation of archaeal cells and preparation of pili samples *Pyrobaculum calidifontis* DSM 21063⁴⁰ and *Aeropyrum pernix* K1 DSM 11879³⁸ cells were purchased from the DSMZ culture collection. *P. calidifontis* was grown in 1090 medium (1.0% tryptone, 0.1% yeast extract, 0.3% sodium thiosulfate, pH 7) at 90 °C without agitation. Preculture (30 mL) was started from a 200 µL cryo-stock, grown for 2 days and then diluted into 200 mL of fresh medium. When OD₆₀₀ reached ~0.2, the cells were collected by centrifugation (Sorval SLA1500 rotor, 7438 × g, 10 min, 20 °C). The resultant pellet was re-suspended in 10 mL of phosphate-buffered saline (PBS) buffer, and the cell suspension was vortexed for 15 min to shear off the extracellular filaments. The cells were removed by centrifugation (Eppendorf F-35-6-30 rotor, 7197 × g, 20 min, 20 °C). The supernatant was collected and the filaments were pelleted by ultracentrifugation (Beckman SW60Ti rotor, 194,038 × g, 2 h, 15 °C). After the run, the supernatant was removed and the pellet was re-suspended in 200 µL of PBS buffer.

A. pernix K1 cells were grown in 3ST medium (35 g/L Sea salts [Sigma], 0.1% tryptone, 0.1% yeast extract, 0.1% thiosulfate, pH 7) at 90 °C without agitation. Pre-culture (10 mL) was started from a 1 mL cryo-stock, grown for 3 days, and then diluted into 60 mL of fresh 3ST medium. After 3 days of growth 100 mL of fresh media was added to the culture and the growth was continued for another 3 days. Then the cells were removed by centrifugation (Sorval SLA1500 rotor, 7438 × g, 10 min, 20 °C), and the filaments were pelleted from the supernatant by ultracentrifugation (Beckman SW60Ti rotor, 194,038 × g, 2 h, 15 °C) and re-suspended in 200 µL of PBS.

pED208 pili purification The F-pili were purified as described in Costa et al., 2016 with some minor modifications. In short, *E. coli* JE2571 harboring the pED208-plasmid were grown on large (24 × 24 cm) LB agar plates overnight. Bacteria were gently collected from the plates with SSC buffer (15 mM sodium citrate pH 7.2 150 mM NaCl) and left to resuspend for 2 h at 4 °C with mixing, followed by two rounds

of centrifugations at $10,800 \times g$ for 20 min. The pili were precipitated from the supernatant by addition of 500 mM NaCl and 5% PEG 6000, followed by incubation for 2 h at 4 °C. Precipitate was rescued by centrifugation at $15,000 \times g$ for 20 min and re-suspended in 120 ml of water, followed by another round of precipitation as described above but this time the precipitate was re-suspended in 1 ml of PBS (pH 7.4) buffer. The suspension was layered on pre-formed CsCl step gradients (1.0–1.3 g/cm³) and separated by 17 h centrifugation at $192,000 \times g$ in 4 °C. The gradient was fractionated and the fraction containing the F-pili was dialyzed against PBS (pH 7.4). Purity of the F-pilus was assessed by SDS-PAGE and the presence of the F-pilin (TraA) was further confirmed by mass spectrometry.

A. tumefaciens pili purification Methods for *Agrobacterium tumefaciens* T-pilus isolation and concentration are adapted from Lai and Kado⁷⁴. Briefly, *A. tumefaciens* C58 Δ visR (flagella-knockout) was streaked from a stab or frozen culture on LB agar with no antibiotics at 28 °C. After colonies appeared, about 2 days, a single colony was cultured overnight in 5 mL 523 media in a covered culture tube with shaking at 19 °C. The next day the turbid culture was transferred to a centrifuge tube and pelleted by centrifuging 15 mins at $119.5 \times g$. Media was aspirated from the culture tube and the pellet was gently re-suspended in 25 mL AB/MES with 5% glucose, then incubated at 19 °C for 4 hours. The culture was then spread on 20 petri dishes or six screening trays containing AB/MES with 5% glucose, 200 μ M acetosyringone to induce pili growth, 1.2% agar. Agar plates were incubated at 19 °C for 6 days. After a lawn developed on each plate, 1 mL of cold 10 mM sodium phosphate buffer, pH 5.3 buffer was added to each petri dish (or 5 mL for screening trays) and layers of bacteria were scraped off with a cell spreader and transferred to a 50 ml Falcon tube on ice. This step was repeated with a second aliquot of buffer added to remove the remaining bacteria. The total volume of bacteria and buffer was about 50 mL. The bacterial suspension was pipetted up and down to break up the biofilm then gently pushed through 1 mL of glass wool in a 30 mL syringe to strain out agar gel. The strained bacterial suspension was then forced through a 26-gauge needle eight times total to shear the T-pili off of the bacteria. After shearing, the suspension was transferred to a 0.2 μ m disposable filter vacuum flask and

filtered on ice to separate sheared pili from whole cells, periodically removing build-up from the filter. The isolated pili were concentrated by a factor of 10 using a 100 MWCO centrifugal filter. The pili were rinsed twice with 10 mM Tris-HCl, 100 mM NaCl, pH 7.5 buffer, re-suspended to 10% of the starting volume, and frozen at -80°C for storage.

Cryo-EM sample preparation and data collection A $3\mu\text{L}$ aliquot of sample containing either *Pyrobaculum calidifontis*, *Aeropyrum pernix*, or *Agrobacterium tumefaciens* pili was applied to a plasma cleaned (Gatan Solarus) lacey carbon grid (Ted Pella, Inc.), blotted with automated blotting for 3 s at 90% humidity and flash frozen in liquid ethane using an EM GP Plunge Freezer (Leica). The dataset used for structure determination was collected at the Molecular Electron Microscopy Core at the University of Virginia on a Titan Krios EM operated at 300 keV, equipped with an energy filter and K3 direct electron detector (Gatan). An energy filter slit width of 10 eV was used during data collection and was aligned automatically every hour. All 8127 *P. calidifontis*, 598 *A. pernix*, and 8363 *A. tumefaciens* movies were collected in counting mode using EPU v2.4 (Thermo Fisher) at a magnification of 81 K, pixel size of 1.08 \AA , and a defocus range from -2.2 to $-1.2\text{ }\mu\text{m}$. Data were collected using a total dose of $50\text{ e}^{-}\text{ \AA}^{-2}$ across 40 frames with an exposure time of 2.98 s.

Data processing and helical reconstruction

Unless otherwise stated, all data processing was done using cryoSPARC v3.2.0⁷⁵. Movies were corrected for full-frame motion using patch motion correction followed by patch CTF Estimation⁷⁶. After CTF estimation, micrographs were sorted and selected based on estimated resolution (0 to 4 \AA), defocus (0.6 to $-2.6\text{ }\mu\text{m}$), ice thickness, and total full-frame motion. Initial particles were automatically picked using ‘Filament Tracer’ with a filament diameter of 100 – 160 \AA and an overlap fraction of 0.05 – 0.07 . Particles were extracted at a box size of 300 or 320 pixels, followed by 2D classification. Class averages containing filaments distinguishable from noise were selected for template-based particle

picking. A total of 549,015 and 427,344 filament segments were extracted using a box size of 320 Å, for *P. calidifontis* and *A. pernix*, respectively. A total of 197,531 T-pilus filament segments were extracted using a box size of 300 Å for *A. tumefaciens*. These particles were sorted using two iterative rounds of 2D classification with 50 classes each, number of online-EM iterations set to 20, and a batch size of 100 per class. The final iteration of 2D classification yielded a subset of 71,981, 44,262, and 49,308 filament segments for *P. calidifontis*, *A. pernix*, and *A. tumefaciens*, respectively. Reconstructions of archaeal and prokaryotic conjugation pili were generated using the following method: (1) an averaged power spectrum was generated using the raw images of aligned filament segments selected from 2D classification,

(2) layer lines were indexed to produce a list of possible helical symmetries, and (3) the correct helical symmetry was determined by trial and error by inspection of an output 3D map looking for obvious structural motifs (i.e., recognizable secondary structural and amino acid side-chain densities). Particles were further refined using local CTF refinement, and another round of helical refinement was performed to generate the final reconstruction. The final resolution achieved for *P. calidifontis*, *A. pernix*, and *A. tumefaciens*, were 4.0 Å, 3.3 Å, and 3.5 Å, respectively. The cryo-EM and refinement statistics for each conjugation pilus are listed in Table S3-1.

Model building and refinement The sequence identity of the subunit for *P. calidifontis* and *A. pernix* was unknown. Using AlphaFold⁷⁷ the sequence identity for *A. pernix* pilin was narrowed to the one that best fit the cryo-EM density map. Using DeepTracer-ID³⁹ we were able to determine the pilin identity of *P. calidifontis* directly from the cryo-EM map. With the sequences identified for their respective map the side chains of each ASU model were adjusted manually in COOT⁷⁸ and inspected using UCSF Chimera⁷⁹. For the *A. tumefaciens* structure the density was good enough to trace the entire backbone and localize most side chains. The cryo-EM structure of pKpQIL (PDB ID: [7JSV](#)) was used as a starting point for building the *A. tumefaciens* cryo-EM model. Following model completion, side chains of the model were manually adjusted in COOT⁷⁸ and inspected in UCSF Chimera

(Pettersen et al., 2004). All models were refined using PHENIX real-space refinement⁸⁰. Refinement included global minimization, B-factor optimization, and applied secondary structure and Ramachandran restraints. The final models were validated with the MolProbity⁸¹ implementation in PHENIX. Refinement statistics for each filament are listed in Table S1. Both cryo-EM maps and atomic coordinates have been deposited with the Electron Microscopy Data Bank and Protein Data Bank with the accession codes given in Table S1. Model-map correlation coefficients were also used to estimate the resolution of the reconstructions and are listed in Table S1.

Sequence analyses Multiple sequence alignment of bacterial and archaeal pilins was made using PROMALS3D and manually adjusted⁸². Genomic neighborhoods were analyzed using the enzyme function initiative-genome neighborhood tool (EFI-GNT)⁸³. Profile-profile comparisons and annotation of proteins encoded in the vicinity of TedC pilin in Thermoproteales were performed using HH-suite package v3⁸⁴. Profiles of the query sequences were constructed by running three iterations of HHblits against the UniRef90 database and the resultant profiles were compared against various profile databases, including the protein family (Pfam) database, the Protein Data Bank (PDB), and the NCBI Conserved Domain database, all available from http://ftp.tuebingen.mpg.de/pub/protevo/toolkit/databases/hhsuite_dbs/⁸⁵. Signal peptides were predicted using SignalP v6⁴².

Mass spectrometry and lipidomics of *A. pernix* The freeze-dried cell preparation was extracted using a modified Bligh-Dyer protocol⁸⁶. Part of the extract was acid hydrolyzed by refluxing with 5% HCl in methanol for 3 hours to release core ether lipids. Both the Bligh-Dyer extract (containing the intact polar lipids with head groups) as well the hydrolyzed extract was analyzed by ultra-high performance liquid chromatography-high-resolution mass spectrometry using a Q Exactive Orbitrap MS following Bale et al.⁸⁶.

Electron spray ionization and tandem MS-MS of pED208 F-pilus and pC58 T-pilus lipidomics
Lipidomics was performed at Lipotype, GmbH (Dresden, Germany) as described previously⁸⁷⁻⁹⁰.

Nomenclature. The following lipid names and abbreviations are used. Cer – Ceramide, Chol – Cholesterol, CL – cardiolipin, DAG – Diacylglycerol, HexCer – Glucosyl/Galactosyl Ceramide, PA – Phosphatidic Acid, PC – Phosphatidylcholine, PE – Phosphatidyl-ethanolamine, PG – Phosphatidyl-glycerol, PI – Phosphatidylinositol, PS – Phosphatidylserine, and their respective lysospecies: lysoPA, lysoPC, lysoPE, lysoPI and lysoPS; and their ether derivatives: PC O-, PE O-, LPC O-, LPE O-; SE – Sterol Ester, SM – Sphingomyelin, TAG – Triacylglycerol.

Lipid species were annotated according to their molecular composition as follows: [lipid class]-[sum of carbon atoms in the fatty acids]:[sum of double bonds in the fatty acids];[sum of hydroxyl groups in the long chain base and the fatty acid moiety] (e.g., SM32:2;1). Where available, individual fatty acid composition following the same rules is given in brackets (e.g., 18:1;0-24:2;0).

Lipid extraction. Samples were extracted and analyzed as described⁸⁷⁻⁹¹, which is a modification of a previously published method for shotgun lipidomics⁸⁸. Briefly, samples were suspended in 150 μL of 150 mM ammonium bicarbonate in water and spiked with 20 μL of internal standard lipid mixture, then extracted with 750 μL chloroform/methanol 10:1 (v:v) mixture for 2 hours at 4 °C with 1400 rpm shaking. After centrifugation (3 min, 3000 $\times g$) to facilitate phase partitioning, the lower, lipid-containing, organic phase was collected (1st step extract), and the remaining water phase was extracted further with 750 μL chloroform/methanol 2:1 (v:v) mixture under the same conditions. Again, the lower, organic phase was collected (2nd step extract). Extracts were dried in a speed vacuum concentrator. 120 μL of a dried 1st step extract underwent acetylation with 75 μL acetyl chloride/chloroform 1:2 (v:v) mixture for 1 h to derivatize cholesterol. After completing the reaction, the mixture was dried. 120 μL of a dried 1st step extract and a derivatized extract were re-suspended in an

acquisition mixture with 8 mM ammonium acetate (400 mM ammonium acetate in methanol:chloroform:methanol:propan-2-ol, 1:7:14:28, v-v:v:v). 120 μ L of the 2nd step extract was re-suspended in an acquisition mixture with 30 μ L 33% methylamine in methanol, in 60 mL methanol:chloroform 1:5 (v:v). All liquid handling steps were performed using a Hamilton STARlet robotic platform.

PLA2 treatment. 0.2 U PLA2 was added to Eppendorf tubes containing samples with a concentration of \sim 18 μ M of pili from either pED208 F-pilus or pC58 T-pilus. These samples were then incubated for 60 minutes at 37 $^{\circ}$ C and flash-frozen in liquid ethane.

Lipid standards. Synthetic lipid standards were purchased from Sigma-Aldrich (cholesterol D6), Larodan Fine Chemicals (DAG, TAG), and Avanti Polar Lipids (all remaining lipids). Standard lipid mixtures were chloroform/methanol 1:1 (v:v) solutions containing the lipids listed in Table S3.

Lipid spectrum acquisition. Extracts in acquisition mixtures were infused with a robotic nanoflow ion source (TriVersa NanoMate; Advion Biosciences) into a mass spectrometer instrument (Q Exactive, Thermo Scientific). Cer, DiHexCer, HexCer, lysolipids, and SM were monitored by negative ion mode FT MS. PA, PC, PE, PI, CL, PS, and ether species were monitored by negative ion mode FT MS-MS. Acetylated cholesterol was monitored by positive ion mode FT MS. SE, DAG, TAG, and species were monitored by positive ion mode FT MS-MS.

Lipid identification and quantification. Automated processing of acquired mass spectra, identification, and quantification of detected molecular lipid species was performed by LipidXplorer software⁹². Data post-processing and normalization were performed using an in-house developed data management system. Only lipid identifications with a signal-to-noise ratio >5 , an absolute abundance

of at least 1 pmol, and a signal intensity 5-fold higher than in corresponding blank samples were considered for further data analysis.

Chemical and physical treatments of pED208 F-pilus and pC58 T-pilus. The following chemical or physical treatments were introduced and incubated at room temperature (23 °C) unless otherwise stated for ~10 minutes to samples of either pED208 F-pilus or pC58 T-pilus: 50% glycerol, high temperature (70 °C), 0.1% SDS, 1% Triton-X-100, and 4M urea.

Negative stain transmission electron microscopy of treated pED208 F-pilus and pTiC58 T-pilus 2 µL of the sample either of pED208 F-pilus or pTiC58 T-pilus subjected to their respective chemical or physical treatment were applied to a carbon grid and stained with 2% uranyl acetate and examined by transmission electron microscopy using a Tecnai T12 at 80 kV.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability The atomic model for the *P. calidifontis* pilus was deposited in the Protein Data Bank with accession code [8DFT](#), and the corresponding map was deposited in the Electron Microscopy Data Bank with accession code [EMD-27413](#). The atomic model for the *A. pernix* pilus was deposited in the Protein Data Bank with accession code [8DFU](#), and the corresponding map was deposited in the Electron Microscopy Data Bank with accession code [EMD-27414](#). The atomic model for the *A. tumefaciens* pilus was deposited in the Protein Data Bank with accession code [8EXH](#), and the corresponding map was deposited in the Electron Microscopy Data Bank with accession code [EMD-28657](#).

References

1. Takeuchi, N., Kaneko, K. & Koonin, E. V. Horizontal gene transfer can rescue prokaryotes from Muller's ratchet: benefit of DNA from dead cells and population subdivision. *G3 (Bethesda)* 4, 325–339 (2014).
2. Muller, H. J. The relation of recombination to mutational advance. *Mutat. Res.* 106, 2–9 (1964).
3. Lee, I. P. A., Eldakar, O. T., Gogarten, J. P. & Andam, C. P. Bacterial cooperation through horizontal gene transfer. *Trends Ecol. Evol.* 37, 223–232 (2022).
4. Thomas, C. M. & Nielsen, K. M. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* 3, 711–721 (2005).
5. Gophna, U. & Altman-Price, N. Horizontal gene transfer in archaea from mechanisms to genome evolution. *Annu Rev. Microbiol.* 76, 481–502 (2022).
6. Blokesch, M. Natural competence for transformation. *Curr. Biol.* 26, R1126–R1130 (2016).
7. Hileman, T. H. & Santangelo, T. J. Genetics techniques for *Thermococcus kodakarensis*. *Front. Microbiol.* 3, 195 (2012).
8. Chen, I. & Dubnau, D. DNA uptake during bacterial transformation. *Nat. Rev. Microbiol.* 2, 241–249 (2004).
9. Gill, S., Catchpole, R. & Forterre, P. Extracellular membrane vesicles in the three domains of life and beyond. *FEMS Microbiol. Rev.* 43, 273–303 (2019).
10. Liu, J. et al. Archaeal extracellular vesicles are produced in an ESCRT-dependent

manner and promote gene transfer and nutrient cycling in extreme environments. *ISME J.* 15, 2892–2905 (2021).

11. Erdmann, S., Tschitschko, B., Zhong, L., Raftery, M. J. & Cavicchioli, R. A plasmid from an Antarctic haloarchaeon uses specialized membrane vesicles to disseminate and infect plasmid-free cells. *Nat. Microbiol.* 2, 1446–1455 (2017).

12. Soler, N. & Forterre, P. Vesiduction: the fourth way of HGT. *Environ. Microbiol.* 22, 2457–2460 (2020).

13. Bartke, K., Garoff, L., Huseby, D. L., Brandis, G. & Hughes, D. Genetic architecture and fitness of bacterial interspecies hybrids. *Mol. Biol. Evol.* 38, 1472–1481 (2021).

14. Llosa, M., Gomis-Ruth, F. X., Coll, M. & de la Cruz Fd, F. Bacterial conjugation: a two-step mechanism for DNA transport. *Mol. Microbiol.* 45, 1–8 (2002).

15. Delavat, F., Miyazaki, R., Carraro, N., Pradervand, N. & van der Meer, J. R. The hidden life of integrative and conjugative elements. *FEMS Microbiol. Rev.* 41, 512–537 (2017).

16. Botelho, J. & Schulenburg, H. The role of integrative and conjugative elements in antibiotic resistance evolution. *Trends Microbiol.* 29, 8–18 (2021).

17. Johnson, C. M. & Grossman, A. D. Integrative and conjugative elements (ICEs): what they do and how they work. *Annu. Rev. Genet.* 49, 577–601 (2015).

18. Coluzzi, C., Garcillan-Barcia, M. P., de la Cruz, F. & Rocha, E. P. C. Evolution of plasmid mobility: origin and fate of conjugative and nonconjugative plasmids. *Mol. Biol. Evol.* 39, msac115 (2022).

19. Cury, J., Abby, S. S., Doppelt-Azeroual, O., Neron, B. & Rocha, E. P.

C. Identifying conjugative plasmids and integrative conjugative elements with CONJscan. *Methods Mol. Biol.* 2075, 265–283 (2020).

20. Prangishvili, D. et al. Conjugation in archaea: frequent occurrence of conjugative plasmids in *Sulfolobus*. *Plasmid* 40, 190–202 (1998).

21. Medvedeva, S. et al. New insights into the diversity and evolution of the archaeal mobilome from three complete genomes of *Saccharolobus shibatae*. *Environ. Microbiol.* 23, 4612–4630 (2021).

22. Erauso, G., Stedman, K. M., van de Werken, H. J. G., Zillig, W. & van der Oost, J. Two novel conjugative plasmids from a single strain of *Sulfolobus*. *Microbiology* 152, 1951–1968 (2006).

23. Greve, B., Jensen, S., Brugger, K., Zillig, W. & Garrett, R. A. Genomic comparison of archaeal conjugative plasmids from *Sulfolobus*. *Archaea* 1, 231–239 (2004).

24. Stedman, K. M. et al. pING family of conjugative plasmids from the extremely thermophilic archaeon *Sulfolobus islandicus*: insights into recombination and conjugation in Crenarchaeota. *J. Bacteriol.* 182, 7014–7020 (2000).

25. Krupovic, M. et al. Integrated mobile genetic elements in Thaumarchaeota. *Environ. Microbiol.* 21, 2056–2078 (2019).

26. van Wolferen, M., Wagner, A., van der Does, C. & Albers, S. V. The archaeal Ced system imports DNA. *Proc. Natl Acad. Sci. USA* 113, 2496–2501 (2016).

27. Frols, S., White, M. F. & Schleper, C. Reactions to UV damage in the model archaeon *Sulfolobus solfataricus*. *Biochem. Soc. Trans.* 37, 36–41 (2009).

28. van Wolferen, M. et al. Species-specific recognition of sulfolobales mediated by UV-inducible pili and S-layer glycosylation Patterns. *mBio* 11, e03014–19 (2020).

29. Ajon, M. et al. UV-inducible DNA exchange in hyperthermophilic archaea mediated by type IV pili. *Mol. Microbiol.* 82, 807–817 (2011).
30. Li, Y. G. & Christie, P. J. The agrobacterium VirB/VirD4 T4SS: mechanism and architecture defined through in vivo mutagenesis and chimeric systems. *Curr. Top. Microbiol Immunol.* 418, 233–260 (2018).
31. Chou, L. et al. Modular evolution of secretion systems and virulence plasmids in a bacterial species complex. *BMC Biol.* 20, 16 (2022).
32. Costa, T. R. D. et al. Type IV secretion systems: advances in structure, function, and activation. *Mol. Microbiol* 115, 436–452 (2021).
33. Li, Y. G., Hu, B. & Christie, P. J. Biological and structural diversity of type IV secretion systems. *Microbiol. Spectr.* 7, 7.2.30 (2019).
34. Hagiya, M., Close, T. J., Tait, R. C. & Kado, C. I. Identification of pTiC58 plasmid-encoded proteins for virulence in *Agrobacterium tumefaciens*. *Proc. Natl. Acad. Sci. USA* 82, 2669–2673 (1985).
35. Eisenbrandt, R. et al. Conjugative Pili of IncP plasmids, and the Ti plasmid T pilus are composed of cyclic subunits*. *J. Biol. Chem.* 274, 22548–22555 (1999).
36. Lai, E.-M. & Kado, C. I. The T-pilus of *Agrobacterium tumefaciens*. *Trends Microbiol* 8, 361–369 (2000).
37. Lai, E.-M. & Kado, C. I. The *Agrobacterium tumefaciens* T pilus composed of cyclic T pilin is highly resilient to extreme environments. *FEMS Microbiol. Lett.* 210, 111–114 (2002).
38. Sako, Y. et al. *Aeropyrum pernix* gen. nov., sp. nov., a novel aerobic hyperthermophilic

archaeon growing at temperatures up to 100 degrees C. *Int. J. Syst. Bacteriol.* 46, 1070–1077 (1996).

39. Chang, L. et al. DeepTracer-ID: De novo protein identification from cryo-EM maps. *Biophys. J.* 121, 2840–2848 (2022).
40. Amo, T. et al. *Pyrobaculum calidifontis* sp. nov., a novel hyperthermophilic archaeon that grows in atmospheric air. *Archaea* 1, 113–121 (2002).
41. Wang, F., Cvirkaite-Krupovic, V., Krupovic, M. & Egelman, E. H. Archaeal bundling pili of *Pyrobaculum calidifontis* reveal similarities between archaeal and bacterial biofilms. *Proc. Natl Acad. Sci. USA* 119, e2207037119 (2022).
42. Almagro Armenteros, J. J. et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423 (2019).
43. Huang, Q. et al. Efficient 5'-3' DNA end resection by HerA and NurA is essential for cell viability in the crenarchaeon *Sulfolobus islandicus*. *BMC Mol. Biol.* 16,2 (2015).
44. Constantinesco, F., Forterre, P., Koonin, E. V., Aravind, L. & Elie, C. A bipolar DNA helicase gene, *herA*, clusters with *rad50*, *mre11* and *nurA* genes in thermophilic archaea. *Nucleic Acids Res.* 32, 1439–1447 (2004).
45. Atmakuri, K., Cascales, E., Burton, O. T., Banta, L. M. & Christie, P. J. *Agrobacterium* ParA/MinD-like VirC1 spatially coordinates early conjugative DNA transfer reactions. *EMBO J.* 26, 2540–2551 (2007).
46. Costa, T. R. D. et al. Structure of the bacterial sex F pilus reveals an assembly of a stoichiometric protein-phospholipid complex. *Cell* 166, 1436–1444.e1410 (2016).

47. Zheng, W. et al. Cryoelectron-microscopic structure of the pKpQIL conjugative pili from carbapenem-resistant *Klebsiella pneumoniae*. *Structure* 28, 1321–1328.e1322 (2020).
48. De Rosa, M., Gambacorta, A. & Gliozzi, A. Structure, biosynthesis, and physicochemical properties of archaeobacterial lipids. *Microbiol. Rev.* 50, 70–80 (1986).
49. Rensen, E. I. et al. A virus of hyperthermophilic archaea with a unique architecture among DNA viruses. *Proc. Natl Acad. Sci. USA* 113, 2478–2483 (2016).
50. Yoshida, R., Yoshimura, T. & Hemmi, H. Biosynthetic machinery for C25,C25-diether archaeal lipids from the hyperthermophilic archaeon *Aeropyrum pernix*. *Biochem. Biophys. Res. Commun.* 497, 87–92 (2018).
51. Wang, F. et al. An extensively glycosylated archaeal pilus survives extreme conditions. *Nat. Microbiol.* 4, 1401–1410 (2019).
52. Amro, J. et al. Cryo-EM structure of the *Agrobacterium tumefaciens* T-pilus reveals the importance of positive charges in the lumen. *Structure* S0969-2126, 00456–7 (2022).
53. Kreida, S., et al. Cryo-EM structure of the *Agrobacterium tumefaciens* type IV secretion system-associated T-pilus reveals stoichiometric protein-phospholipid assembly. *bioRxiv*, <https://www.biorxiv.org/content/10.1101/2022.09.25.509369v1> (2022).
54. Abu-Arish, A. et al. Three-dimensional reconstruction of *Agrobacterium* VirE2 protein with single-stranded DNA. *J. Biol. Chem.* 279, 25359–25363 (2004).
55. Winans, S. C., Burns, D. L. & Christie, P. J. Adaptation of a conjugal transfer system for the export of pathogenic macromolecules. *Trends Microbiol.* 4, 64–68 (1996).
56. Jones, A. L., Lai, E. M., Shirasu, K. & Kado, C. I. VirB2 is a processed pilin-like protein

encoded by the *Agrobacterium tumefaciens* Ti plasmid. *J. Bacteriol.* 178, 5706–5711 (1996).

57. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* 38, W545–549 (2010).

58. Wang, F. et al. Cryo-EM structure of an extracellular *Geobacter* OmcE cytochrome filament reveals tetrahaem packing. *Nat. Microbiol.* 7, 1291–1300 (2022).

59. Heng, J. B. et al. The electromechanics of DNA in a synthetic nanopore. *Biophys. J.* 90, 1098–1106 (2006).

60. Wang, F. et al. Spindle-shaped archaeal viruses evolved from rodshaped ancestors to package a larger genome. *Cell* 185, 1297–1307. e1211 (2022).

61. Wollman, E. L., Jacob, F. & Hayes, W. Conjugation and genetic recombination in *Escherichia coli* K-12. *Cold Spring Harb. Symp. Quant. Biol.* 21, 141–162 (1956).

62. Derbyshire, K. M. & Gray, T. A. Distributive conjugal transfer: new insights into horizontal gene transfer and genetic exchange in *Mycobacteria*. *Microbiol. Spectr.* 2, 04 (2014).

63. Coros, A., Callahan, B., Battaglioli, E. & Derbyshire, K. M. The specialized secretory apparatus ESX-1 is essential for DNA transfer in *Mycobacterium smegmatis*. *Mol. Microbiol.* 69, 794–808 (2008).

64. Babić, A., Lindner, A. B., Vulić, M., Stewart, E. J. & Radman, M. Direct visualization of horizontal gene transfer. *Science* 319, 1533 (2008).

65. Low, H. H. et al. Structure of a type IV secretion system. *Nature* 508, 550–553 (2014).

66. Liu, X., Khara, P., Baker, M. L., Christie, P. J. & Hu, B. Structure of a type IV secretion system core complex encoded by multi-drug resistance F plasmids. *Nat. Commun.* 13, 379

(2022).

67. Christie, P. J., Atmakuri, K., Krishnamoorthy, V., Jakubowski, S. & Cascales, E. Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annu. Rev. Microbiol.* 59, 451–485 (2005).
68. Moncalián, G. et al. Characterization of ATP and DNA binding activities of TrwB, the coupling protein essential in plasmid R388 conjugation*. *J. Biol. Chem.* 274, 36117–36124 (1999).
69. Hopkins, B. B. & Paull, T. T. The *P. furiosus* mre11/rad50 complex promotes 5' strand resection at a DNA double-strand break. *Cell* 135, 250–260 (2008).
70. Blackwood, J. K. et al. Structural and functional insights into DNA end processing by the archaeal HerA helicase-NurA nuclease complex. *Nucleic Acids Res.* 40, 3183–3196 (2012).
71. De Falco, M., Catalano, F., Rossi, M., Ciaramella, M. & De Felice, M. NurA is endowed with endoand exonuclease activities that are modulated by HerA: new insight into their role in DNA-end processing. *PLoS One* 10, e0142345 (2015).
72. Feng, X., Sun, M., Han, W., Liang, Y. X. & She, Q. A transcriptional factor B paralog functions as an activator to DNA damage responsive expression in archaea. *Nucleic Acids Res.* 46, 7085–7096 (2018).
73. Koonin, E. V., Makarova, K. S., Wolf, Y. I. & Krupovic, M. Evolutionary entanglement of mobile genetic elements and host defence systems: guns for hire. *Nat. Rev. Genet.* 21, 119–131 (2020).
74. Lai, E. M. & Kado, C. I. Processed VirB2 is the major subunit of the promiscuous pilus of *Agrobacterium tumefaciens*. *J. Bacteriol.* 180, 2711–2717 (1998).

75. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* 14, 290–296 (2017).
76. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* 193, 1–12 (2016).
77. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021).
78. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D., Biol. Crystallogr.* 66, 486–501 (2010).
79. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput Chem.* 25, 1605–1612 (2004).
80. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D., Biol. Crystallogr.* 66, 213–221 (2010).
81. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D., Biol. Crystallogr.* 66, 12–21 (2010).
82. Pei, J. & Grishin, N. V. PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods Mol. Biol.* 1079, 263–271 (2014).
83. Zallot, R., Oberg, N. & Gerlt, J. A. The EFI web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry* 58, 4169–4182 (2019).
84. Steinegger, M. et al. HH-suite3 for fast remote homology detection and deep protein

annotation. *BMC Bioinformatics* 20, 473 (2019).

85. Gabler, F. et al. Protein sequence analysis using the MPI bioinformatics toolkit. *Curr. Protoc. Bioinformatics*. 72, e108 (2020).
86. Bale, N. J. et al. New insights into the polar lipid composition of extremely halo (alkali) philic euryarchaea from hypersaline lakes. *Front. Microbiol.* 10, 377 (2019).
87. Sampaio, J. L. et al. Membrane lipidome of an epithelial cell line. *Proc. Natl Acad. Sci. USA* 108, 1903–1907 (2011).
88. Ejsing, C. S. et al. Global analysis of the yeast lipidome by quantitative shotgun mass spectrometry. *Proc. Natl Acad. Sci. USA* 106, 2136–2141 (2009).
89. Gerl, M. J. et al. Quantitative analysis of the lipidomes of the influenza virus envelope and MDCK cell apical membrane. *J. Cell Biol.* 196, 213–221 (2012).
90. Surma, M. A. et al. An automated shotgun lipidomics platform for high throughput, comprehensive, and quantitative analysis of blood plasma intact lipids. *Eur. J. Lipid Sci. Technol.* 117, 1540–1549 (2015).
91. Levental, K. R. et al. omega-3 polyunsaturated fatty acids direct differentiation of the membrane phenotype in mesenchymal stem cells to potentiate osteogenesis. *Sci. Adv.* 3, eaao1193 (2017).
92. Herzog, R. et al. LipidXplorer: a software for consensual crossplatform lipidomics. *PloS One* 7, e29851 (2012).

Acknowledgements

Cryo-EM imaging was conducted at the Molecular Electron Microscopy Core facility at the

University of Virginia, which is supported by the School of Medicine and built with the National Institutes of Health (NIH) grant G20-RR31199. In addition, the Titan Krios (SIG S10-RR025067) and K3/GIF (U24-GM116790) were purchased with the aid of the designated NIH grants. We thank Anhelique Mets and Ellen Hopmans (NIOZ) for technical support and mass spectral interpretations, respectively. Special thanks to Clay Fuqua for providing the *Agrobacterium tumefaciens* strain from which we obtained the T-pili. This work was supported by NIH Grant GM122510 (E.H.E.), GM138756 (F.W.) Wellcome Trust Grant 215164/ Z/18/Z (T.R.D.C), and l'Agence Nationale de la Recherche grant ANR-21CE11-0001-01 (M.K.).

Author contributions

L.B., V.C.-K., V.C., F.W., M.A.B.K., E.H.E., and M.K. designed the studies and experiments. V.C.-K. prepared the archaeal samples. J.B.P. and T.R.D.C. prepared the pED208 pili. J.M., V.C., and M.A.B.K. prepared the *A. tumefaciens* T-pili. S.S. analyzed the archaeal lipids, and I.L. analyzed the bacterial lipids. M.K. performed sequence analyses. L.B., E.H.E., and M.K. wrote the paper, with assistance from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-36349-8>.

Correspondence and requests for materials should be addressed to Edward H. Egelman or Mart Krupovic.

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

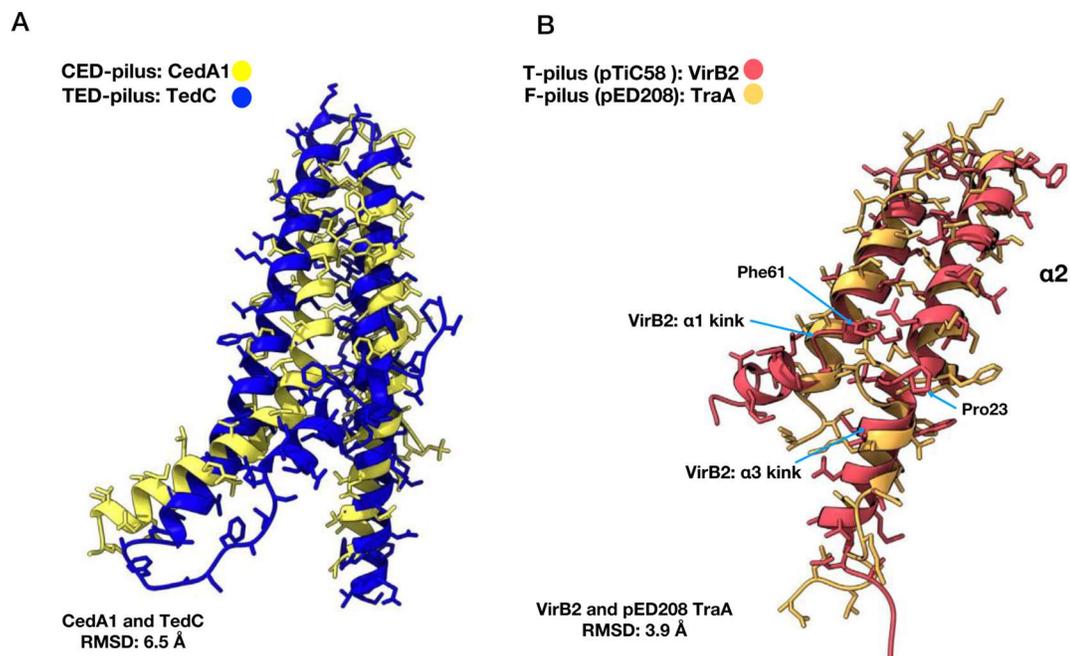
Reprints and permissions information is available at <http://www.nature.com/reprints>

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

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

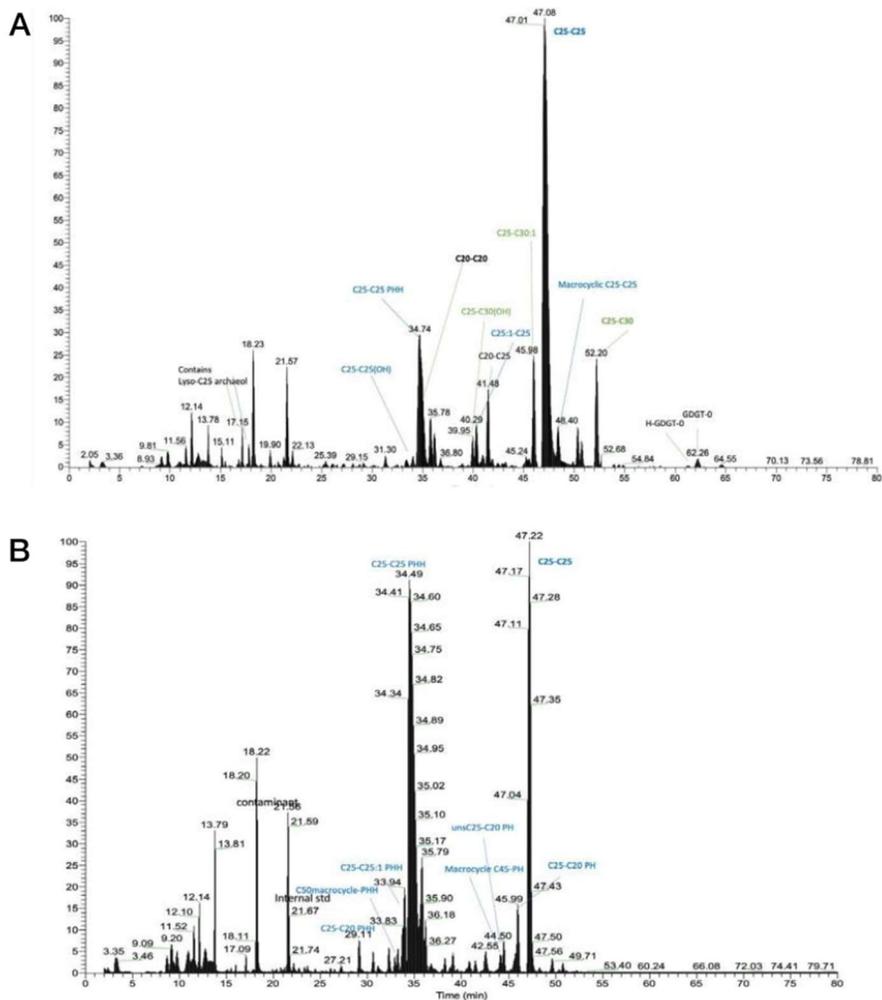
© The Author (s) 2023

Supplementary Information

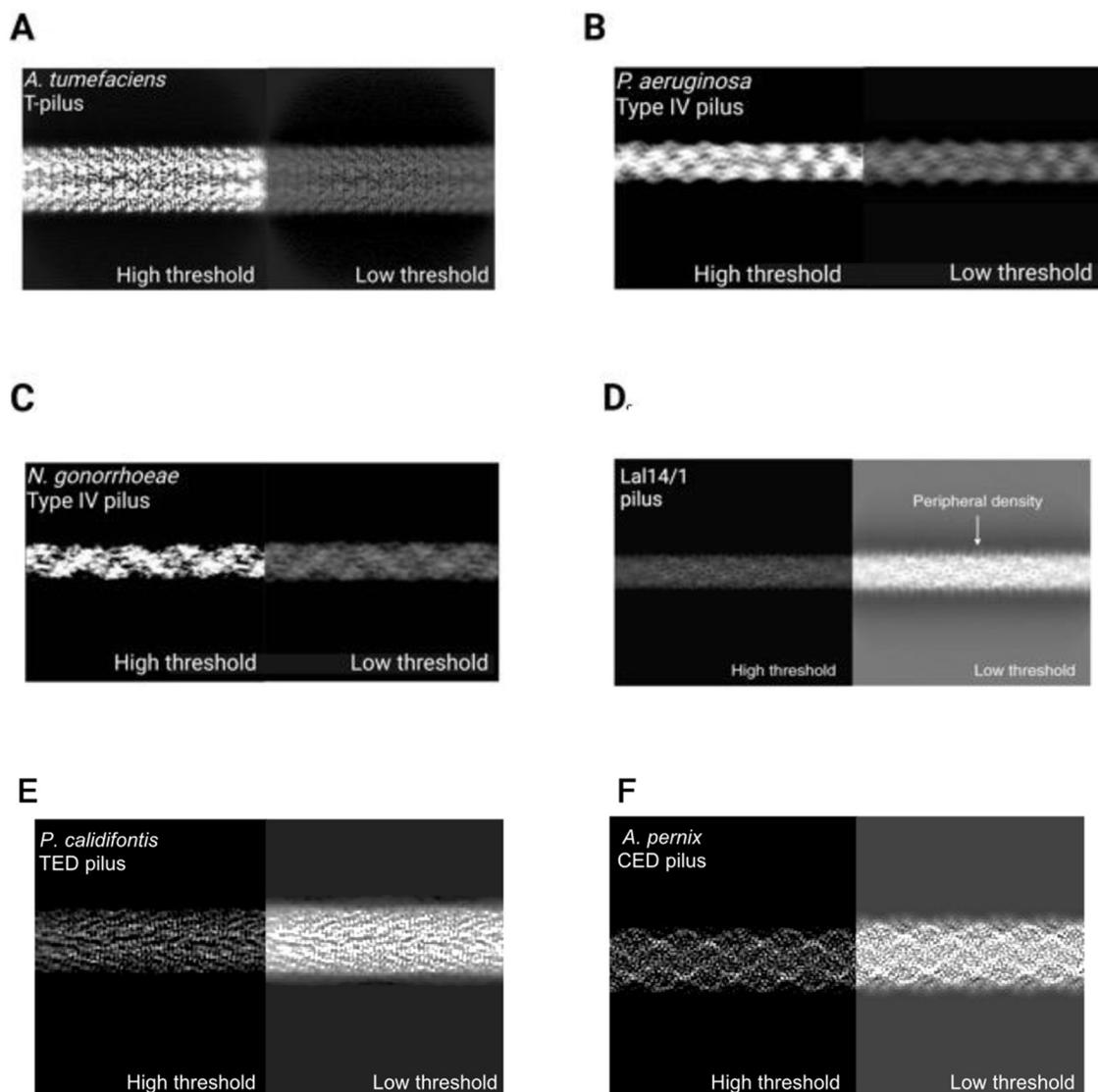


SI Figure 3-1. (A) An alignment of the CedA1 (yellow) with the TedC (blue) subunits shows an RMSD of 6.5 Å between the two subunits. A distinct kink in helix $\alpha 2$ of CedA1, while the kink in helix

$\alpha 2$ of TedC is present but less prominent. **(B)** Alignment of the VirB2 subunit (red) with the TraA subunits of pED208 (orange). There is a 3.9 Å RMSD between VirB2 and pED208 across 58 atom pairs. There are two distinct kinks in VirB2 not present in pED208. The kink in VirB2 helix $\alpha 1$ occurs at phenylalanine 61 and the kink in $\alpha 3$ occurs at proline 23.

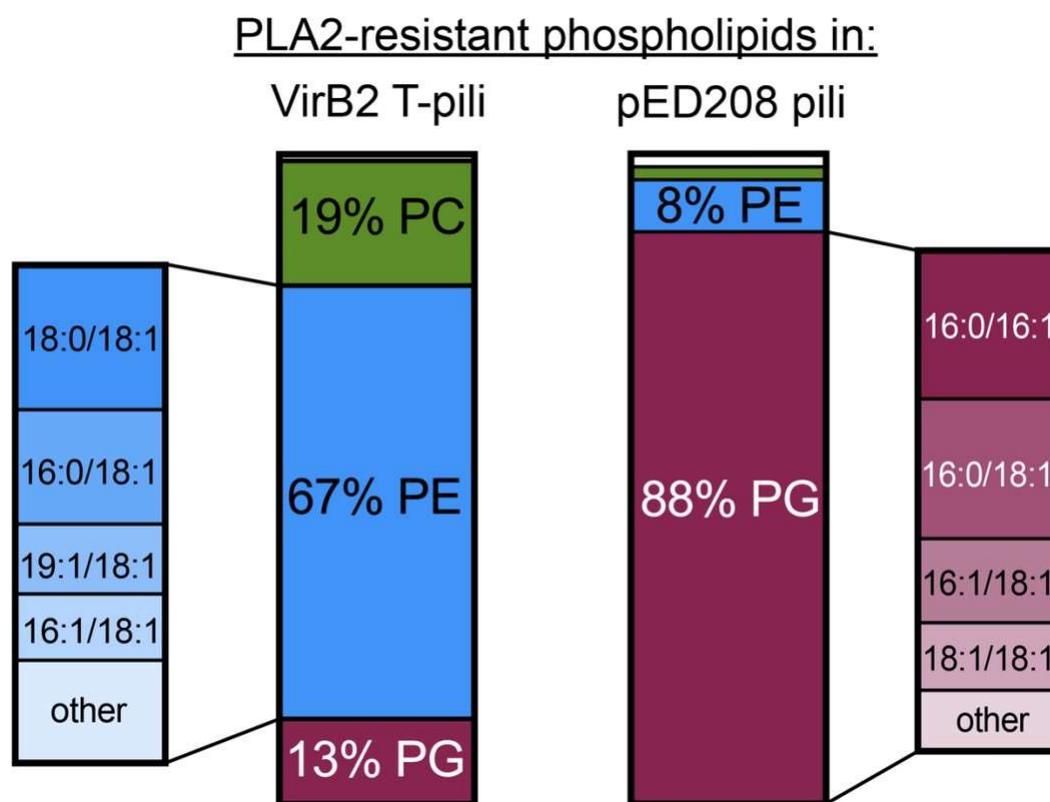


SI Figure 3-3. Mass spectrometry of the lipids extracted from *A. pernix* cell culture. **(A)** A chromatogram of the core lipids found in the sample showing C25-C25 diether lipids as the most abundant species. **(B)** A chromatogram of the most abundant lipid headgroups reveals phospho-dihexose, i.e., a dihexose sugar connected to the glycerol by a phosphate group, as the most prevalent lipid headgroup and phospho-hexose, one sugar connected via a phosphate-group, as the second most prevalent.

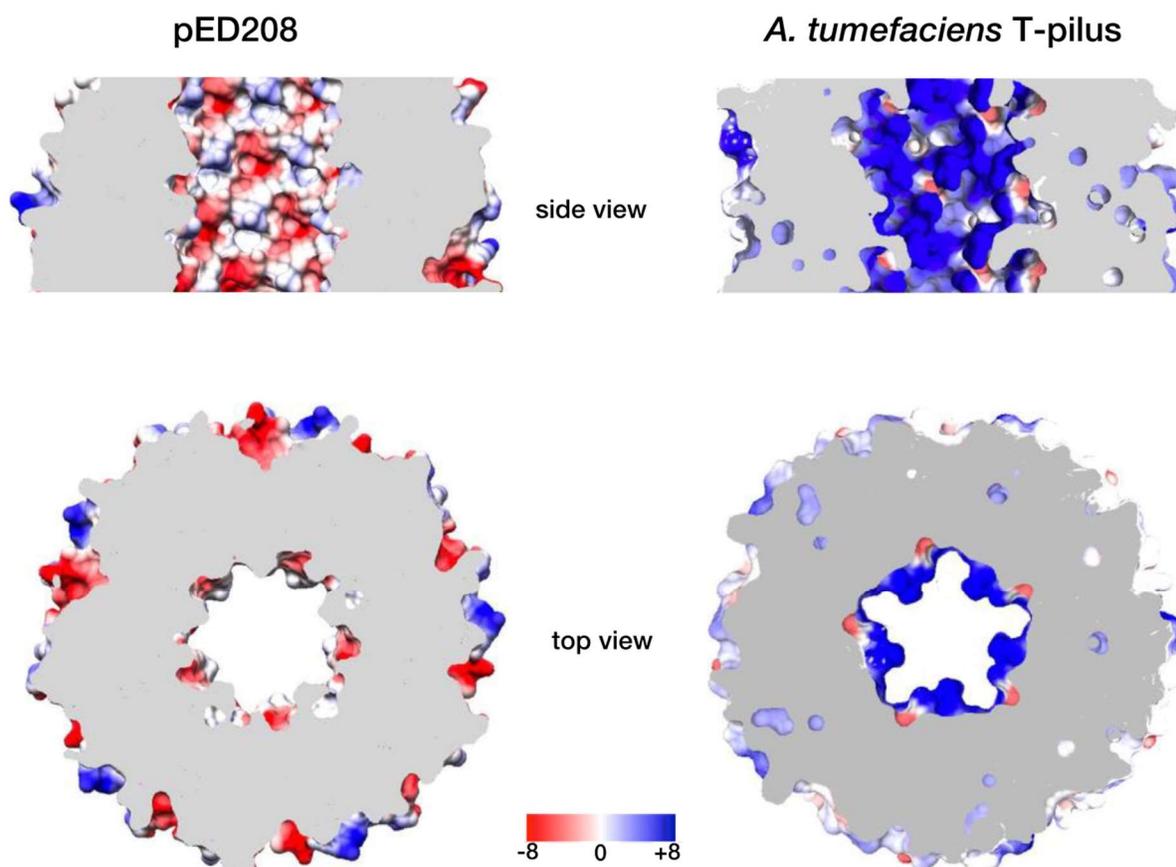


SI Figure 3-4. Projections of the 3D reconstructions at different thresholds reveal if pili are extensively glycosylated. T4P pili of *Neisseria gonorrhoeae* (A, EMDB deposition EMD-8739, [https://www.emdataresource.org/EMD-8739]) and *Pseudomonas aeruginosa* (B, EMDB deposition EMD-8740, [https://www.emdataresource.org/EMD-8740]) are shown as negative controls, whereas the highly glycosylated type IV pilus of *Saccharolobus islandicus* LAL14/1 (C, EMDB deposition EMD-0397, [https://www.emdataresource.org/EMD-0397]) shows peripheral density at low threshold, and serves as a positive control. Both the Ted pilus of *P. calidifontis* (D) and the Ced pilus of *A. pernix* (E)

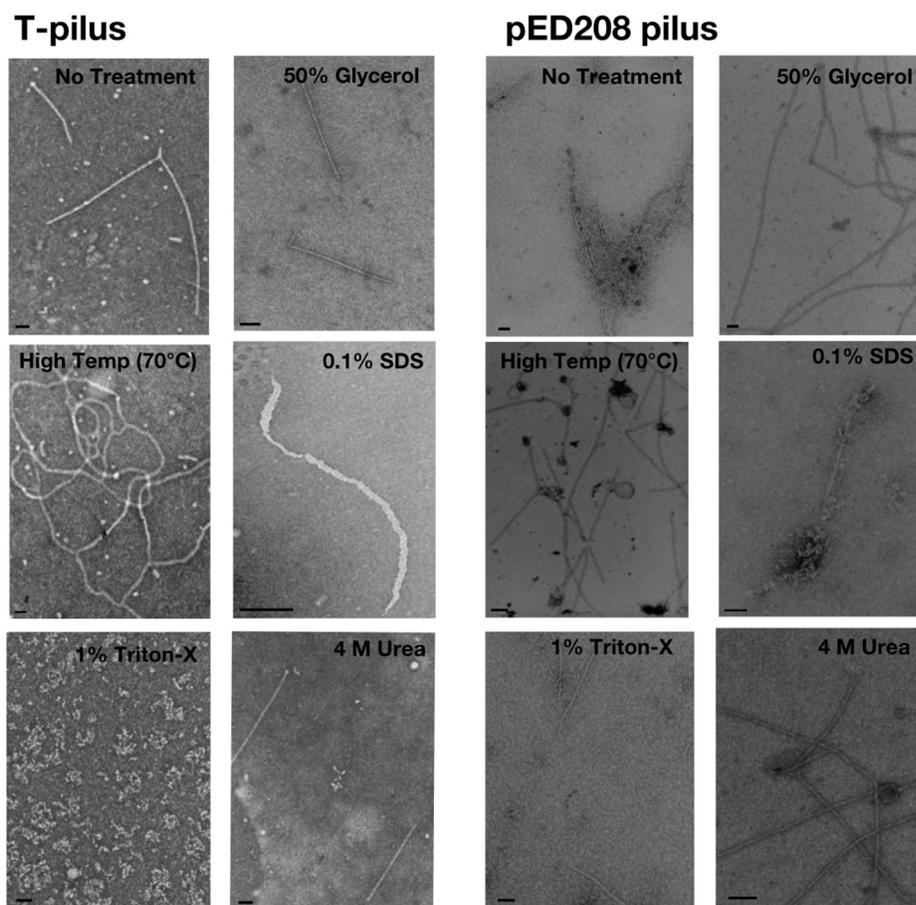
show some surrounding density, which could represent glycosylation. In contrast, the T-pilus of *A. tumefaciens* (F) shows no obvious surrounding density at low threshold and behaves like controls A and B.



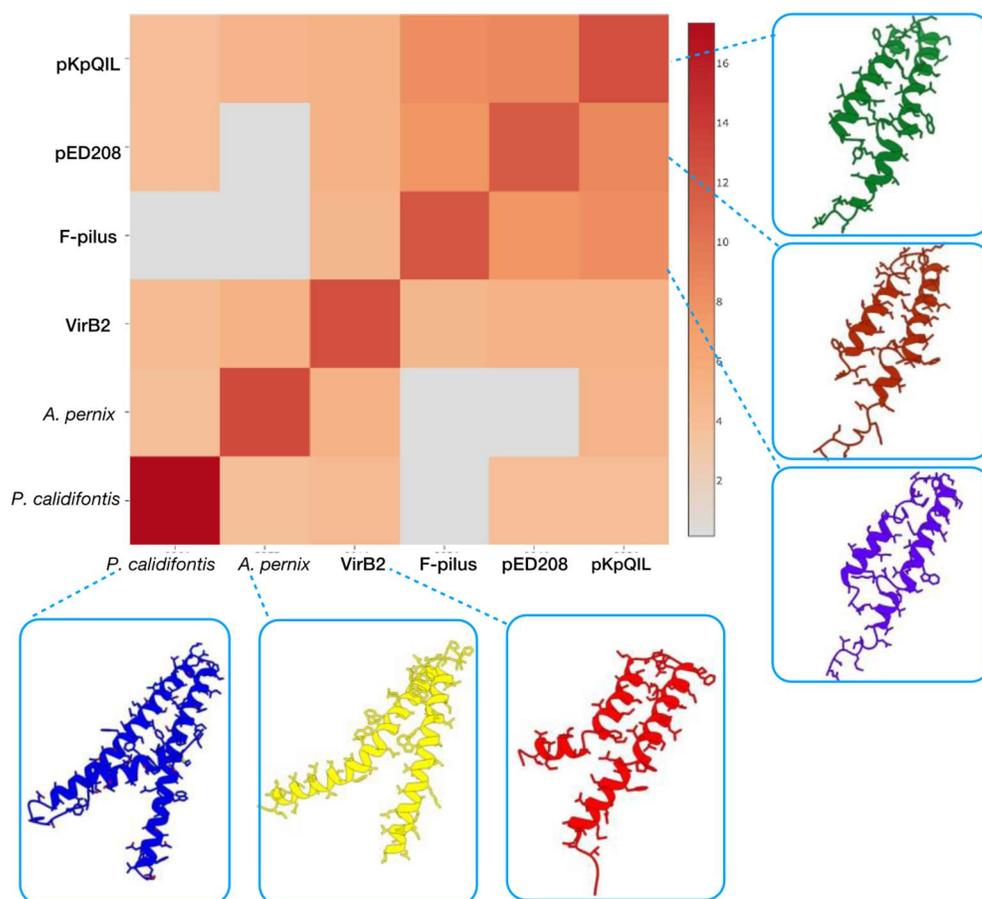
SI Figure 3-5. PLA2 was applied to purified pili from VirB2 and pED208 to digest any unprotected phospholipids. The remaining lipids were extracted, then identified by mass spectrometry and quantified using internal standards. The major phospholipid classes in the pili samples are shown by headgroup class, and the acyl chain compositions are shown for the major headgroup classes (PE for VirB2, PG for pED208). Unlabeled bars represent minor species.



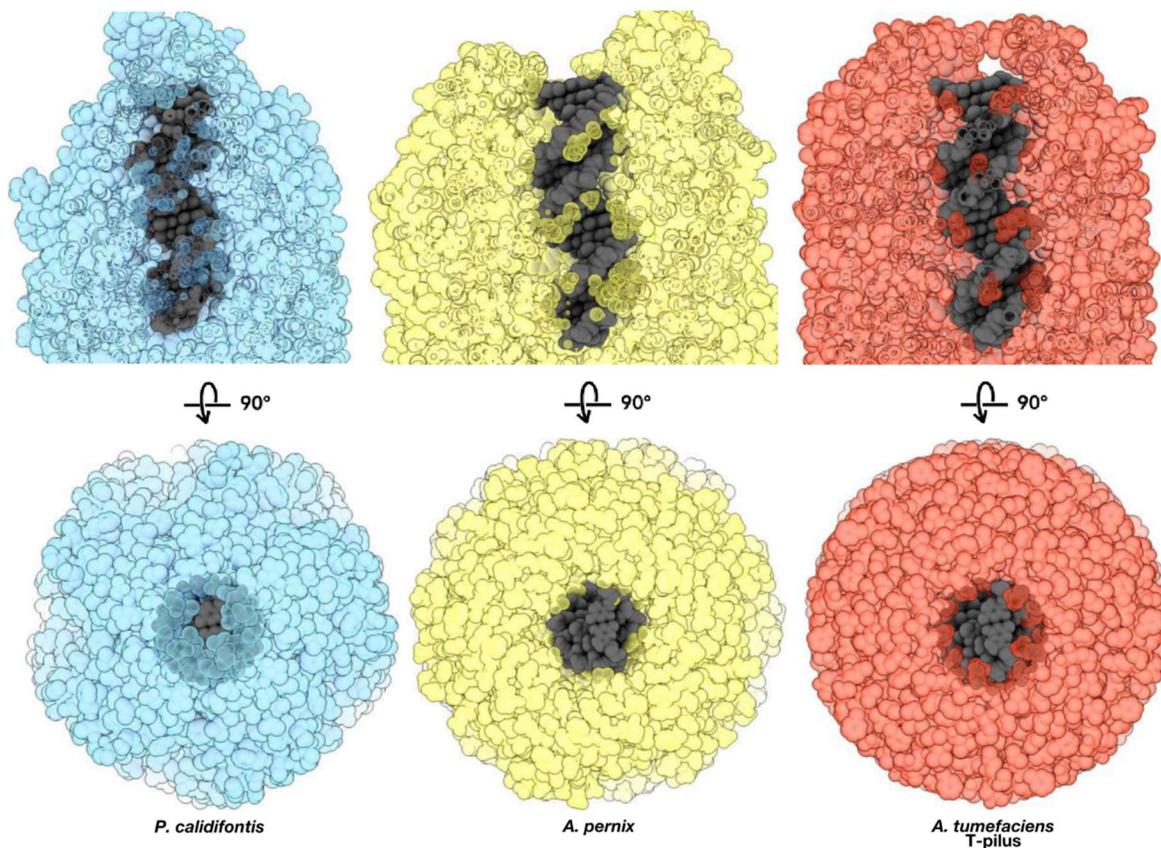
SI Figure 3-6. Electrostatic potential surface for the lumen of the pED208 (left) and T-pilus (right). The side and top view of the pED208 pilus shows a more electronegative lumen compared to the T-pilus. The top view of the T-pilus shows an alternating positive to negative charge arising from the arginine and the lipid, respectively. All electrostatic potentials and surfaces were calculated using UCSF Chimera. The units for the potential are kT/e .



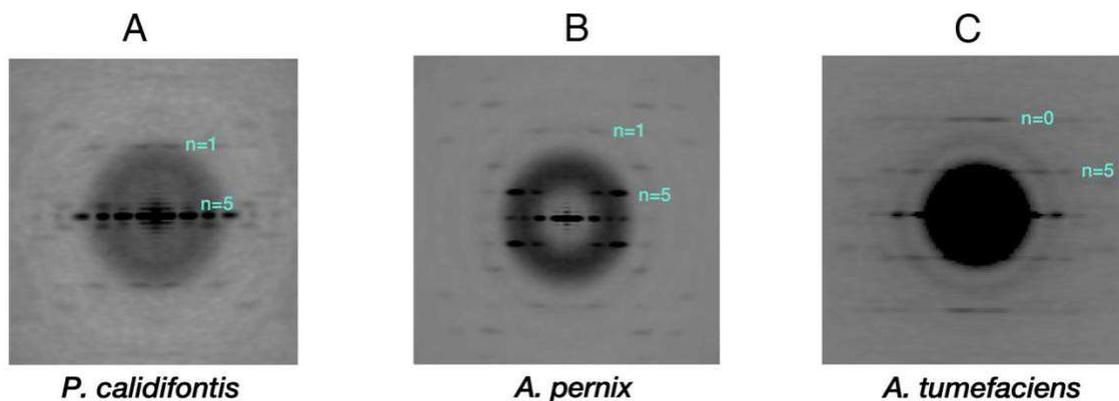
SI Figure 3-7. Negative stain electron micrographs showing the effects on pilus structure of different harsh environmental conditions for *A. tumefaciens* T-pilus (left) and *E. coli* pED208 (right), scale bars 50 nm. The morphology of the T-pilus under no treatment, glycerol (50%) and 4M urea appears fully intact. The high-temperature (70°C) and SDS (0.1%) partially degrades the pilus and results in flexuous filaments. Triton-X (1%) fully depolymerizes the filament and creates aggregated clusters. The pED208 pilus morphology is mostly in a flexuous state with partial bundling at the ends of the pilus under SDS (0.1%). Each condition/treatment was replicated three times for both the T-pilus and pED208. Over 15 grid squares were chosen at random and examined for each sample to ensure a broad representation of pilus morphology across the entire electron microscopy grid. See Table S2 for more details.



SI Figure 3-8. Heat-map of global structural similarity between pilins from all known structures of prokaryotic conjugation pili: *P. calidifontis* (blue), *A. pernix* (yellow), *A. tumefaciens* (red), *F-pilus* (purple), *pED208* (light brown) and *pKpQIL* (green). The color scale corresponds to the Dali Z-score values. The dark red indicates increased similarity, the lighter red indicates some similarity and grey is no to low structure similarity. Each pilin consists of two or three hydrophobic α -helices with kinks appearing in both of the archaeal pili and the *A. tumefaciens* T-pilin. All share a common helix-turn-helix architecture and have obvious structural homology.



SI Figure 3-9. **Conjugative pili cannot transport dsDNA.** Side views (top) and top views (bottom) of the atomic models of *P. calidifontis* (blue), *A. pernix* (yellow), and *A. tumefaciens* (red) with a model for B-form dsDNA placed within the lumen. The narrow lumen of all the prokaryotic conjugative pili would result in extensive clashes with dsDNA suggesting that only ssDNA could be accommodated within the lumen. The atomic model of *P. calidifontis* (blue) does not include a lipid model with phospho-dihexose headgroups. The addition of these sugar headgroups would decrease the pore size and result in even further clashes with dsDNA.



SI Figure 3-10. Averaged power spectra of the *P. calidifontis*, *A. pernix* and *A. tumefaciens* pili. (A.) Averaged power spectrum generated for *P. calidifontis* from 54,000 segments. Two layer lines are indexed with their Bessel orders: the $n=1$ and $n=5$. (B.) Averaged power spectrum generated for *A. pernix* from 44,000 segments. Two layer lines are indexed with their Bessel orders: the $n=1$ and $n=5$. (C.) Averaged power spectrum generated for *A. tumefaciens* from 49,000 segments. Two layer lines are indexed with their Bessel orders: the $n=0$ and $n=5$. To increase the dynamic range, the log of the intensities are shown for *P. calidifontis* and *A. tumefaciens*.

Table S3-1. Cryo-EM and refinement statistics.

Parameter	<i>P. calidifontis</i> TedC	<i>A. pernix</i> CedA1	<i>A. tumefaciens</i> VirB2
Data collection and processing			
Voltage (kV)	300	300	300
Electron exposure (e-Å ⁻²)	50	50	50
Pixel size (Å)	1.08	1.08	1.08
Segments (n)	71,981	44,262	49,308
Helical symmetry			
Point group	C1	C1	C5
Helical rise (Å)	5.00	3.61	13.68
Helical twist (°)	74.21	76.5	32.45
Map resolution (Å)			
Software (final reconstruction)	cryoSPARC	cryoSPARC	cryoSPARC
Model:map FSC (0.38)	4.2	3.5	3.7
Map:map FSC (0.143)	4.0	3.3	3.5
Refinement and Model validation			
Bond lengths rmsd (Å)	0.004	0.007	0.003
Bond angles rmsd (°)	0.796	0.612	0.632
Clashscore	16.28	3.45	4.3
Ramachandran Favored(%)	97.27	98.78	100
Ramachandran Outlier(%)	0	0	0
MolProbity score	1.85	1.14	1.66
Deposition ID			
PDB (model)	8DFT	8DFU	8EXH
EMDB (map)	EMD-27413	EMD-27414	EMD-28657

Table S3-2. Environmental effects on pilus structure.

Environmental effects on T-pilus and pED208 structure

Agent or condition	Temperature (°C)	Morphology of T-pilus filament ¹	Morphology of pED208 pilus filament ¹
No Treatment	23	Rigid/Fully intact	Flexuous/Fully intact
Glycerol (50%)	23	Rigid/Fully intact	Flexuous/Fully intact
High temperature	70	Flexuous/Partially degraded	Flexuous/Fully intact
SOS (0.1%)	23	Flexuous/Partially degraded	Bundling at the ends of the pilus /Fully intact
Triton X-100 (1%)	23	Aggregated clusters/Fully depolymerized	Flexuous/Fully intact
Urea (4 M)	23	Rigid/Fully intact	Flexuous/Fully intact

¹For each preparation the sample was negatively stained with 2% uranyl acetate and examined by transmission electron microscopy

Table S3-3. Standard Lipid Mixtures

Cer 35:1;2, (D18:1;2, 17:0;0)
Chol D6
DAG 34:0;0 (17:0;0, 17:0;0)
DiHexCer 30:1;2 (D18:1;2.12:0;0)
HexCer 30:1;2 (D18:1;2.12:0;0)
LPA 17:0;0 (17:0;0)
LPC 12:0;0 (12:0;0)
LPE 17:1;0 (17:1;0)
LPI 17:1;0 (17:1;0)
LPS 17:1;0 (17:1;0)
PA 34:0;0 (17:0;0, 17:0;0)

PC 34:0;0 (17:0;0, 17:0;0)
PE 34:0;0 (17:0;0, 17:0;0)
PG 34:0;0 (17:0;0, 17:0;0)
PI 32:0;0 (16:0;0, 16:0;0)
PS 34:0;0 (17:0;0, 17:0;0)
SE 20:0;0 (20:0;0)
SM 30:1;2 (18:1;2, 12:0;0)
TAG 51:0;0 (17:0;0, 17:0;0, 17:0;0)

Chapter 4 |

Stacked cyclic peptide assemblies mediated by pyrene stacking

The short sequences of peptides belie the remarkable diversity in their assembly strategies. Cyclization restricts their arrangements but also makes new ones possible. Here, we begin to investigate a pair of cyclic peptides with one or two pyrene moieties, which are expected to interleave with neighboring assemblies and mediate lateral interactions. Peptide assembly and lateral interactions are responsive to pH changes. Although they are a challenging subject for cryo-EM reconstruction, this tool will be key to determining the subtle structural differences that lead to large changes to the supramolecular morphology.

Introduction

This chapter is unique from the rest of this dissertation in two ways: the subunits are stacked instead of helical, and the subunit is a designed cyclic peptide instead of a small naturally-derived protein. However, the 1TP and 2TP filaments described in this chapter obey the same rules for assembly as all proteinaceous molecules, making them a simple model to isolate variables in protein assembly.

Stacked peptide assemblies are not well-represented in nature, probably because they lack the stability of helical symmetry. Cross- β stacks such as those composing amyloids have been used to study amyloid pathogenicity and formation (Fitzpatrick et al., 2013; Jahn et al., 2010; Nelson et al., 2005) and have served as the basis for several designed assemblies. In general, these assemblies are mediated by complementary electrostatic interactions on an internal interface parallel to the z axis of the filament. Nelson, et al also found hydrogen-bonded “amide stacks” between adjacent Asn or Gln side-chain amides, as well as π -stacking between Tyr rings in GNNQQNY and NNQQNY amyloid models. Amyloids assemble into more than one structure per peptide (Gallardo et al., 2020), an excellent example of deterministic chaos and a significant frustration to peptide designers.

Recently, several examples of synthetic stacked assemblies have been designed, which include canonical as well as non-canonical side chains. A water-soluble stacked assembly was derived from

bacterial Outer surface protein A (OspA) by modifying it to reconstitute the “hydrophobic ladder” motif observed among amyloids (Biancalana et al., 2010). Stacks along the z-axis of each protein consisting of all Tyr or alternating Leu/Glu, Lys/Glu, or Leu/Phe were all effective; Leu/Phe could also mediate dimerization between identical proteins. Aromatic side chains are somewhat of a fast-track to inducing stacked assemblies due to their ability to π -stack. Even diphenylalanine peptide (FF) can form nanotubes, which have attracted considerable attention for their piezoelectric properties, unidirectional polarization, and ability to translate axial forces into shear deformation, which generates a voltage (Görbitz, 2006; Kim et al., 2010; Lee et al., 2018; Nguyen et al., 2016). The simplicity of stacked peptide rings allows small functional moieties to be easily grafted onto the surface, such as the hydrophobic “arms” in Rho, et al, 2019.

Figure 4.2. Possible assembly strategy for 2TP and gel. Increasing the pH of the solution above 8 is thought to induce assembly into single or paired filaments. CPs, cyclic peptides.

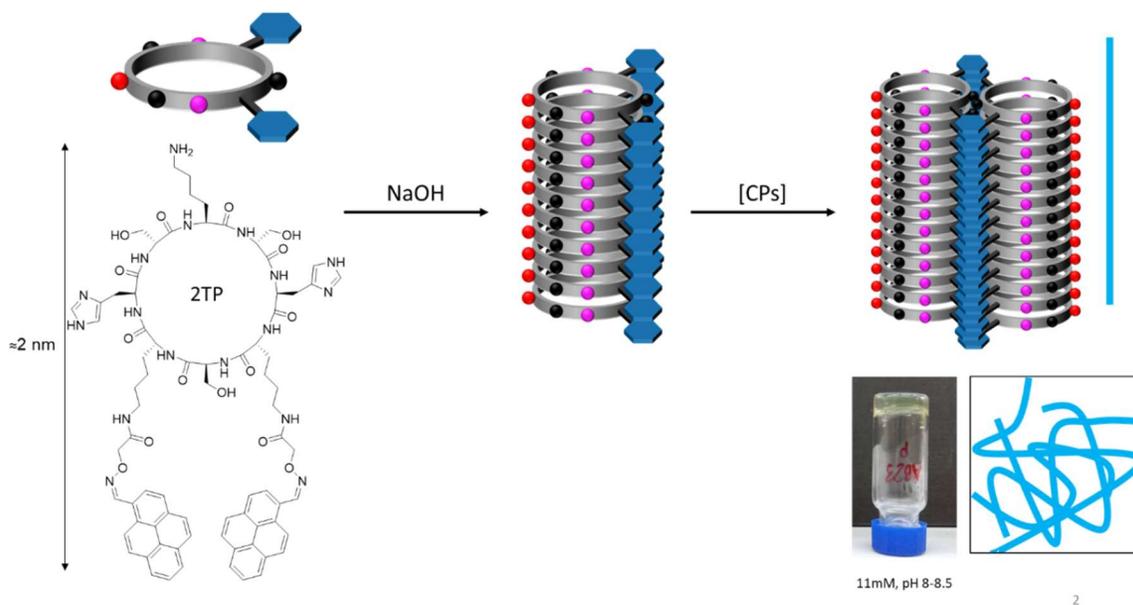


Figure 4.1. Model of 2TP assembly. 2TP is thought to assemble under basic conditions into stacked filaments, with pyrene appendages stacked in register. Filaments are expected to laterally associate by interleaving pyrenes into two stacks. 2TP forms a gel at high concentration (11 mM) and pH 8-8.5, inset.

Pyrene is widely used in organic chemistry as a fluorescent dye and probe. Its fluorescence is highly sensitive to its chemical environment and can supply information about polarity and proximity of other pyrenes (Bains et al., 2011). These features make it an attractive probe for studying protein folding and conformational changes. Pyrene also exhibits very strong pi-pi stacking, with a binding energy of about 15.0 kcal/mol per pair (Dharmarwardana et al., 2021). It is incorporated into the design of 1TP and 2TP to further stabilize the stacked CPNT and is thought to mediate lateral interactions of 2TP (**Error! Reference source not found.**).

Cyclic peptide nanotubes (CPNT) evade the problem of amyloid design: *Cyclizing the peptide pre-organizes it into a limited conformational range and should, in turn, limit possible conformations of the assembly.* Several have been designed, including transmembrane ion channels (Montenegro et al., 2013), viral-capsid-like tubes with the ability to deliver cargo to cells (Brendel et al., 2018), CPNT with a site for display of a side chain on the interior of the tube (Hourani et al., 2011), among others (Abdullah et al., 2020; Insua and Montenegro, 2020; Shaikh et al., 2018; Valéry et al., 2003). Due to their similarity to viral capsids, they are thought to be good vaccine candidates if immune response can be improved.

1TP and 2TP are designed cyclic peptides (Figure 4.3) of 8 residues. Assigning Position 1 to the lysine, 1TP has a single pyrene moiety at position 5 directly across from the lysine, whereas 2TP features two pyrenes at positions 4 and 6. Both peptides have two histidines positioned directly across from one another. The peptides are expected to stack into filaments in basic conditions, mediated by π -stacking of the pyrenes and further stabilized by the neutral charge of the histidines.

Under TEM and cryo-EM, both are observed to readily form filaments at high concentrations in acidic or basic conditions. 1TP forms long, thin filaments, whereas 2TP forms thicker, shorter filaments. At high concentration (11 mM), both peptide solutions form a yellow to orange gel when resuspended in water and pH adjusted to pH 8 with NaOH.

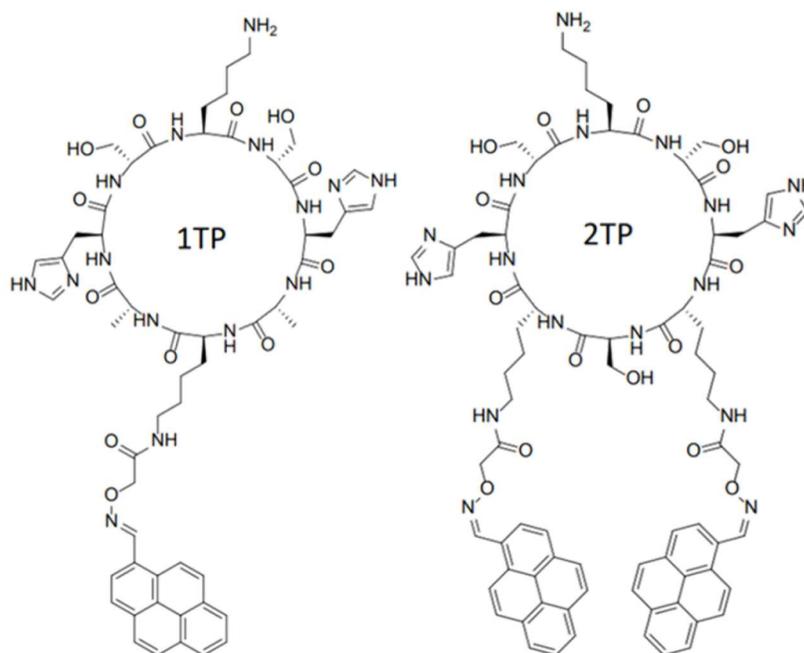


Figure 4.3. Designs of 1TP and 2TP cyclic peptides. 1TP features one pyrene moiety opposite lysine at position 5, whereas 2TP features two pyrenes at positions 4 and 6.

The aim of this study is to determine the structure of both 1TP and 2TP assemblies via cryo-EM in order to understand a) the position of pyrene in this assembly and 2) how the chemical environment affects the assembly. According to models, peptides are expected to be packed in an antiparallel orientation with the pyrenes directly stacked over each other. The stacked pyrenes of assembled 2TP are expected to interleave with those of the paired filament.

Results and Discussion

Under negative stain TEM, 1TP in water is clearly visible as long ($>1 \mu\text{m}$) filaments, usually in groups of two or three filaments. Because the solution is not buffered and the peptide was purified using trifluoroacetic acid, the pH had shifted to 3.8 over time. In HEPES buffer pH 8, 1TP filaments were much shorter ($\sim 50 \text{ nm}$) and were more likely to be bundled in groups of three or more. This change in the morphology and behavior of the nanotube is likely due to the protonation states of the histidines ($\text{pK}_a 6$). At pH 8 they should be deprotonated and therefore neutrally charged. Below pH 6, histidine is protonated and positively charged. The filament was not expected to persist at acidic pH because of these repulsive interactions between adjacent histidines. Instead, the filaments were abundant, long, and still bundled into groups of two or three filaments. In HEPES, lateral interaction networks formed as expected, probably mediated by neutrally charged histidines. Tubes appeared to be the same length as in water.

2TP was expected to form pairs of stacked filaments as in **Error! Reference source not found.**, with pairs of pyrenes π -stacking to form a stable interface. Under the microscope, the short filaments are arranged in lateral groups of two or more.

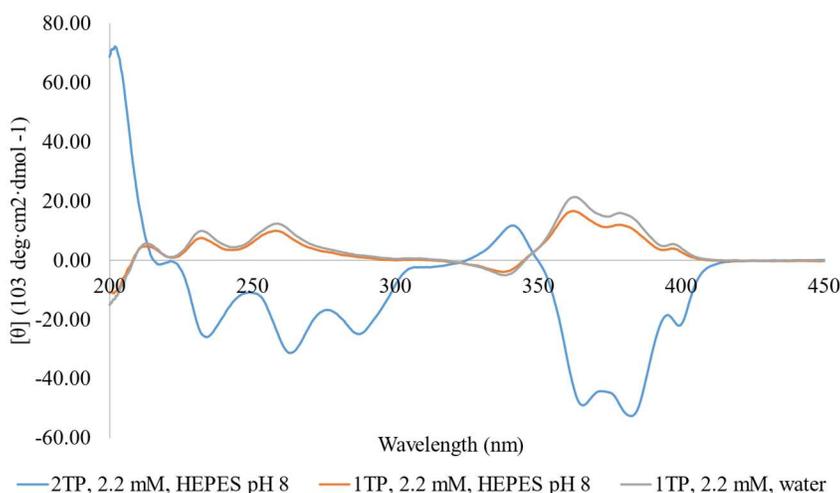


Figure 4.4. Circular dichroism spectra for 1TP and 2TP. The two peptides appear to polarize light at the same wavelengths, but in opposite directions. 2TP also shows a strong signal around 200 nm. Solving the structures of these filaments will also help identify these CD peaks.

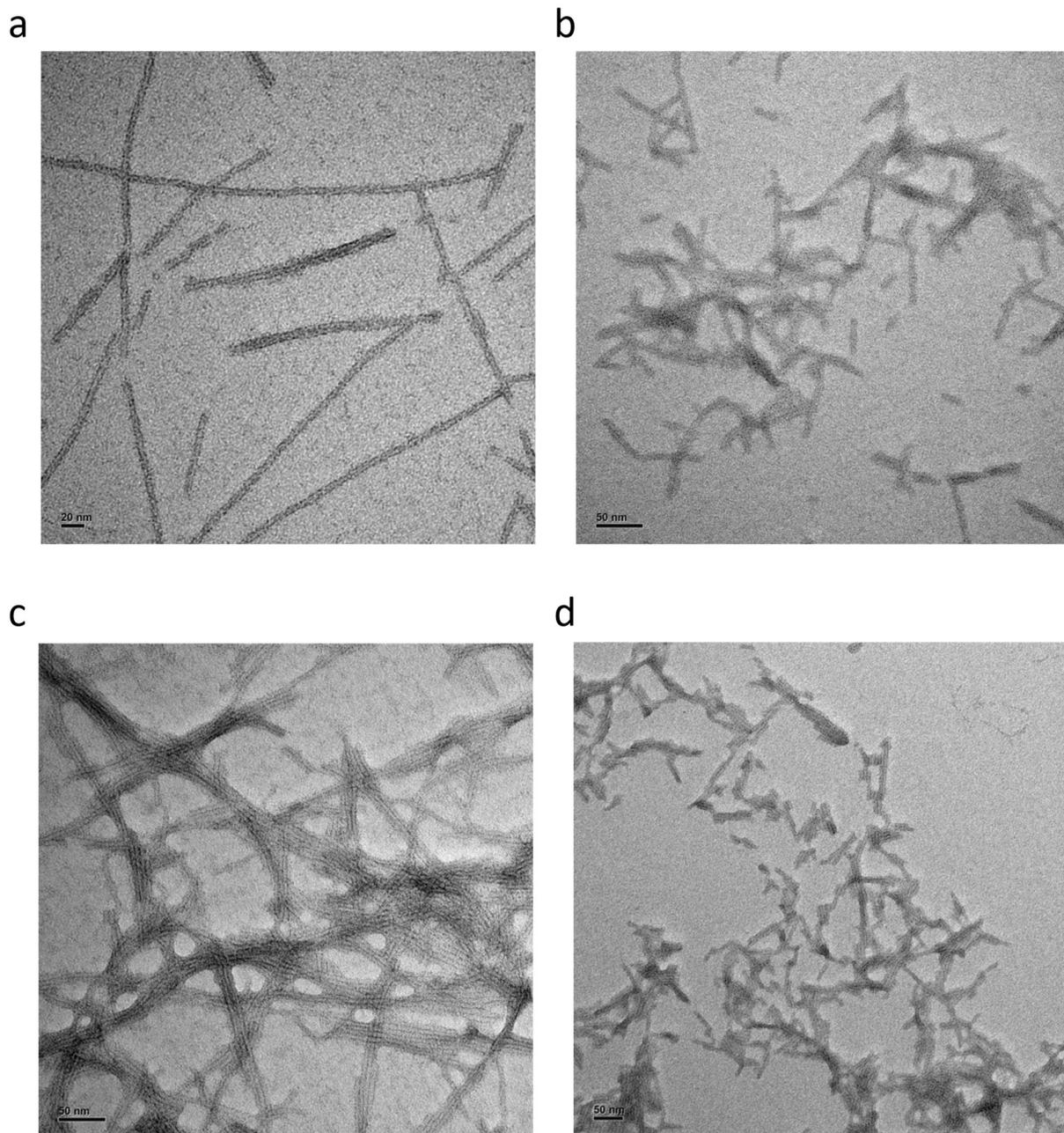


Figure 4.5. Negative-stain TEM images from 1TP and 2TP assembly. A. 1TP, 11 μM in water presents as very long, single filaments with excellent dispersion on the film. Some filaments are paired but most are single. B. 2TP, 11 μM in water. Filaments are much shorter than 1TP, about 50-100 nm long. Filaments are almost always in bundles which appear flat. C. 1TP, 11 μM in HEPES pH 8. Filaments are still very long, but are now bundled into groups of 3-5 filaments. The length and bundling behavior

combine to create a network of laterally associated filaments. D. 2TP 11 μM in HEPES pH 8. Filaments are still bundled and are too short to develop a network of filaments.

Identifying ideal conditions for cryo-EM freezing was very challenging. I started with 1TP in water because it appeared to assemble into long, single filaments with a good aspect ratio and dispersity for autopicking in RELION. However, without salt, water does not stay in the holes long enough to freeze, resulting in dry holes. The conditions that produced good 225 μM peptide, doubly applied (apply out of Vitrobot, wick, then apply in Vitrobot as usual), mixed with 300 mM NaCl immediately before application. The next batch of grids will be frozen as before, except with peptide concentration around 50 μM to avoid the background monolayer. When freezing peptide in HEPES, NaCl will be omitted. Screening will be performed at 200 kV.

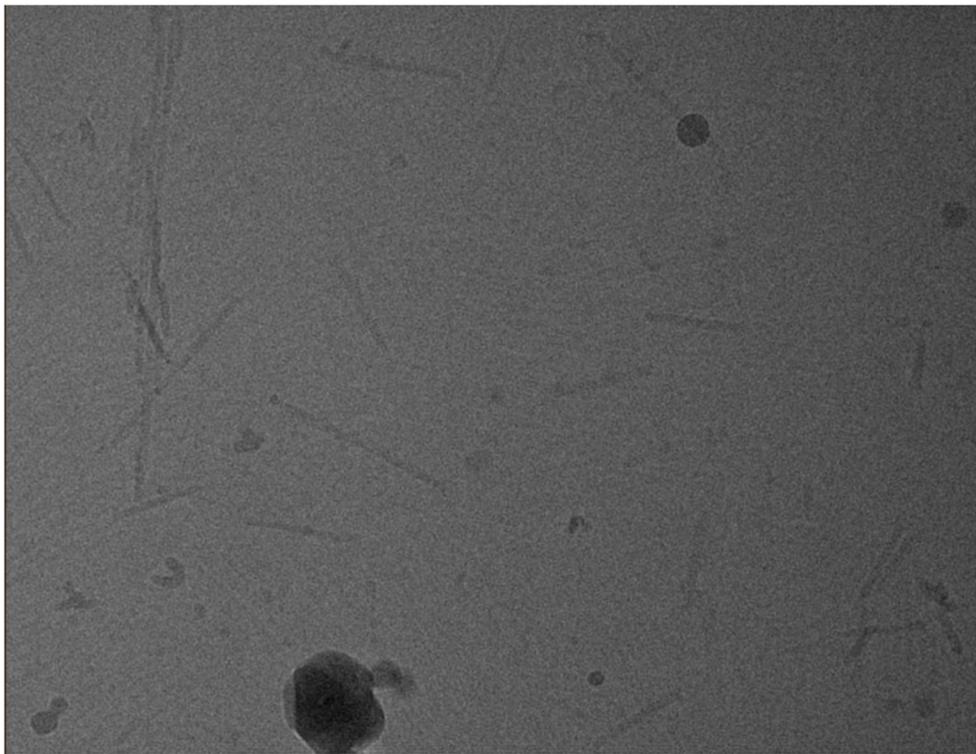


Figure 4.6. Cryo-EM micrograph of 225 μM 1TP in 300 mM NaCl. Background of the image is not clear ice, but a monolayer of filaments.

Typically, screening is performed on a lower-voltage (and cheaper to run) microscope, e.g. 80 or 120 kV, to confirm ice quality, dispersity, and sample quality before moving on to a higher-voltage, e.g., 200 to 300 kV, microscope for data collection. Replicating our collaborator's earlier work at 550 μM peptide produced a layer of filaments that was not visible at lower voltage but was very apparent at 200 kV (Figure 4.6), and made good 2D class average nearly impossible. If the filaments had been farther apart it would have produced an excellent dataset, albeit with low contrast. This problem is easily addressed by lowering the peptide concentration to a more typical 25-50 μM and is a good example of information loss at lower voltage.

The different lengths of 1TP and 2TP filaments is a surprising result. Comparing both in HEPES, they both feature strong lateral interactions which tend to stabilize the filament, as seen in Chapter 2. It is still early in the structural analysis of these peptides, but the answer may lie in the nucleation rate of each filament. 2TP is designed with more stabilizing interactions, and may nucleate more quickly or the lateral interaction, mediated by two π -stacked pyrenes, may be more likely than the axial interaction, mediated by π -stacking and/or H-bonds. All CPNTs in this study probably formed during the initial dissolution of the peptide powder at 11 mM. Tubes should form at this extremely high concentration even if the affinity was low. Adjusting pH to 8 should stabilize existing tubes.

Materials and Methods

Preparing filaments: Peptides were synthesized and kindly sent by Javier Montenegro at Universidade de Santiago de Compostela (Santiago de Compostela, Spain). Upon arrival, they were stored at -20°C . For assembly, peptides were dissolved in HPLC-purity water at 11 mM concentration to make a stock solution. This stock solution was sonicated in a bath for 10 minutes to fully dissolve the peptide. A 50 μL aliquot of each peptide was removed. pH of this aliquot was adjusted to pH 8 using 1 μL doses of 0.1 M NaOH. The pH-adjusted peptide solution turned into a thick yellow to orange gel. The gel equilibrated at

room temperature overnight. The gel was then diluted to 11 μM in water and vortexed until all gel was dissolved.

For preparation in HEPES pH 8, the stock solution was diluted with HPLC-purity water to 20 μM and sonicated 10 minutes. Afterwards, the solution was diluted 1:1 with HEPES 20mM pH8 and equilibrated at room temperature for 4 hours.

Electron Microscopy: TEM grids were prepared by incubating 5 μL of the sample on the grid for 10 minutes then blotted from the side. 5 μL NanoVan/NanoW was applied to each grid for one minute, then wicked from the side.

Cryo-EM grids were prepared by applying 3 μL 11 μM 1TP in water only to negatively glow-discharged Cu 300 Quantifoil grids (blot force 0, blot time 1 second, wait time 20 seconds) and then plunged into liquid ethane. This yielded a nearly completely dry grid, but the few holes that had ice also had a few visible filaments. Concentration was increased to 0.55 mM to replicate earlier findings by our collaborator. Neither double application nor applying 5 μL improved ice quality. Ice was significantly improved by mixing NaCl with the sample to a final concentration of 300 mM NaCl, 0.225 mM peptide. 2TP was prepared in the same way.

After screening at 80 kV (JEOL JEM-1400, Japan Electron Optics Co., Japan), a preliminary data set was collected at 200 kV on the FEI Arctica microscope (Thermo Fisher). At 200kV, it was apparent that the background was not clear ice, but instead a monolayer of filaments with single filaments resting on top (Figure 4.6). Grids were prepared again at 225 μM in 20 mM HEPES.

Structural determination: Images were imported into RELION 3.1. At the time of this writing, the background monolayer of filaments precluded structural determination. I am preparing to collect another data set.

Circular dichroism spectropolarimetry. 40 μL of 550 μM peptide, freshly diluted from 11 mM stock, was applied to a 0.1 mm quartz High Precision Cell cuvette (Hellma Analytics, Jena, Germany), which was assembled and wrapped with Teflon plumber's tape and Parafilm to retard evaporation. After

blanking with water or buffer, a spectrum from 190 to 250 nm was measured in a JASCO J-1500 CD Spectrometer, with SpectraManager software. Molar ellipticity was determined using the formula $\text{ellipticity} = ((100 * \text{raw ellipticity}) / (\text{concentration of protein in mM}) (\text{peptide bonds in protein}) (0.01 \text{ cm path length})) / 1000$ to compare proteins of different lengths.

Conclusions

Conclusions at this stage are necessarily preliminary. The difference in filament length is surprising. The 1TP filaments are longer when 2TP filament should be better stabilized by the pyrene stacking. This could be due to some sort of frustration, possibly caused by π stacking, that is tolerated over hundreds of interactions, but not thousands. Our models of filament assembly are educated human guesses but have not been informed by any kind of software tools, because no reliable tools exist (Chapter 1). Structural analysis will be key to understanding this phenomenon and is indeed the reason for this collaboration.

In the near future, we will examine the impacts of heat annealing on the sample, which has been shown to influence filament morphology of other peptide nanomaterials.

References

- Abdullah, T., Bhatt, K., Eggermont, L.J., O'Hare, N., Memic, A., Bencherif, S.A., 2020. Supramolecular Self-Assembled Peptide-Based Vaccines: Current State and Future Perspectives. *Frontiers in Chemistry* 8.
- Bains, G., Patel, A.B., Narayanaswami, V., 2011. Pyrene: a probe to study protein conformation and conformational changes. *Molecules (Basel, Switzerland)* 16, 7909-7935.
- Biancalana, M., Makabe, K., Koide, S., 2010. Minimalist design of water-soluble cross- β architecture. *Proceedings of the National Academy of Sciences* 107, 3469-3474.

- Brendel, J.C., Sanchis, J., Catrouillet, S., Czuba, E., Chen, M.Z., Long, B.M., Nowell, C., Johnston, A., Jolliffe, K.A., Perrier, S., 2018. Secondary Self-Assembly of Supramolecular Nanotubes into Tubisomes and Their Activity on Cells. *Angewandte Chemie International Edition* 57, 16678-16682.
- Dharmarwardana, M., Otten, B.M., Ghimire, M.M., Arimilli, B.S., Williams, C.M., Boateng, S., Lu, Z., McCandless, G.T., Gassensmith, J.J., Omary, M.A., 2021. Strong π -stacking causes unusually large anisotropic thermal expansion and thermochromism. *Proceedings of the National Academy of Sciences* 118, e2106572118.
- Fitzpatrick, A.W.P., Debelouchina, G.T., Bayro, M.J., Clare, D.K., Caporini, M.A., Bajaj, V.S., Jaroniec, C.P., Wang, L., Ladizhansky, V., Müller, S.A., MacPhee, C.E., Waudby, C.A., Mott, H.R., De Simone, A., Knowles, T.P.J., Saibil, H.R., Vendruscolo, M., Orlova, E.V., Griffin, R.G., Dobson, C.M., 2013. Atomic structure and hierarchical assembly of a cross- β amyloid fibril. *Proceedings of the National Academy of Sciences* 110, 5468-5473.
- Gallardo, R., Ranson, N.A., Radford, S.E., 2020. Amyloid structures: much more than just a cross- β fold. *Current Opinion in Structural Biology* 60, 7-16.
- Görbitz, C.H., 2006. The structure of nanotubes formed by diphenylalanine, the core recognition motif of Alzheimer's β -amyloid polypeptide. *Chemical Communications*, 2332-2334.
- Hourani, R., Zhang, C., van der Weegen, R., Ruiz, L., Li, C., Keten, S., Helms, B.A., Xu, T., 2011. Processable Cyclic Peptide Nanotubes with Tunable Interiors. *Journal of the American Chemical Society* 133, 15296-15299.
- Insua, I., Montenegro, J., 2020. 1D to 2D Self Assembly of Cyclic Peptides. *Journal of the American Chemical Society* 142, 300-307.

- Jahn, T.R., Makin, O.S., Morris, K.L., Marshall, K.E., Tian, P., Sikorski, P., Serpell, L.C., 2010. The common architecture of cross-beta amyloid. *J Mol Biol* 395, 717-727.
- Kim, J., Han, T.H., Kim, Y.-I., Park, J.S., Choi, J., Churchill, D.G., Kim, S.O., Ihee, H., 2010. Role of Water in Directing Diphenylalanine Assembly into Nanotubes and Nanowires. *Advanced Materials* 22, 583-587.
- Lee, J.-H., Heo, K., Schulz-Schönhagen, K., Lee, J.H., Desai, M.S., Jin, H.-E., Lee, S.-W., 2018. Diphenylalanine Peptide Nanotube Energy Harvesters. *ACS Nano* 12, 8138-8144.
- Montenegro, J., Ghadiri, M.R., Granja, J.R., 2013. Ion Channel Models Based on Self-Assembling Cyclic Peptide Nanotubes. *Accounts of Chemical Research* 46, 2955-2965.
- Nelson, R., Sawaya, M.R., Balbirnie, M., Madsen, A.Ø., Riekel, C., Grothe, R., Eisenberg, D., 2005. Structure of the cross-beta spine of amyloid-like fibrils. *Nature* 435, 773-778.
- Nguyen, V., Zhu, R., Jenkins, K., Yang, R., 2016. Self-assembly of diphenylalanine peptide with controlled polarization for power generation. *Nature Communications* 7, 13566.
- Rho, J.Y., Cox, H., Mansfield, E.D.H., Ellacott, S.H., Peltier, R., Brendel, J.C., Hartlieb, M., Waigh, T.A., Perrier, S., 2019. Dual self-assembly of supramolecular peptide nanotubes to provide stabilisation in water. *Nature Communications* 10, 4708.
- Shaikh, H., Rho, J.Y., Macdougall, L.J., Gurnani, P., Lunn, A.M., Yang, J., Huband, S., Mansfield, E.D., Peltier, R., Perrier, S., 2018. Hydrogel and Organogel Formation by Hierarchical Self-Assembly of Cyclic Peptides Nanotubes. *Chemistry—A European Journal* 24, 19066-19074.
- Valéry, C., Paternostre, M., Robert, B., Gulik-Krzywicki, T., Narayanan, T., Dedieu, J.-C., Keller, G., Torres, M.-L., Cherif-Cheikh, R., Calvo, P., Artzner, F., 2003. Biomimetic organization: Octapeptide self-assembly into nanotubes of viral capsid-like dimension.

Proceedings of the National Academy of Sciences of the United States of America 100,
10258-10262.

Chapter 5 |

Structure and characterization of self-assembling archaeal cannula from metagenomic data

Deep-sea hyperthermophilic archaea have been observed to grow cannulae, long glycoprotein assemblies whose function is unknown. Here we present the structures of two archaeal cannulae derived from metagenomic data. We show that both cannulae employ a donor strand complementation motif and divalent cations to rapidly assemble into a closed tube, with limited or multiple symmetries. This is the first example of a rigid assembly mediated by donor strand complementation. We also investigate the roles of cations and heat in the assembly. While the role of cannulae in the archaeon remains elusive, this new structural understanding will guide protein designers and microbiologists.

Introduction

Filamentous protein assemblies are common in biological systems in which they are involved in a wide range of functional roles that are critical to the survival and propagation of the organism. Many of these functions, e.g., locomotion, adhesion, tunable mechanical response, electrical conductivity, directional transport of substrates, etc., would be desirable to emulate in synthetic protein-based materials. However, this approach requires prior knowledge of sequence-structure-function relationships that are responsible for the observed behavior of the corresponding biologically derived assemblies. The structural principles that underlie the function of these biologically derived protein-based materials are gradually being revealed through structural analysis at near-atomic resolution, primarily via cryo-electron microscopy (cryo-EM). These studies have provided insights into the relationship between structure and biological function that may be useful for the *de novo* design of protein-based biomaterials.

Counterintuitively, hyperthermostable proteins do not differ wildly from their mesophilic counterparts. They serve metabolic and structural purposes in the organism, just like mesophilic proteins do. To perform their diverse functions, many enzymes must undergo conformational changes, which rules out rigidity as a strategy for thermophilicity. In fact, thermostable proteins may even be more flexible than mesophilic proteins, in that they sample more conformations without denaturing. For example, Hernández, et al measured the solvent accessibility of rubredoxin from the hyperthermophile *Pyrococcus*

furiosus using hydrogen exchange (Hernández et al., 2000). They found that it to be roughly equivalent to that of unstructured peptides, meaning that rubredoxin is just as flexible as mesophilic proteins and does not derive stability from being stapled into a rigid conformation.

The question of how proteins become hyperthermostable is still debated and there seems to be no single strategy. Hyperthermostable rubredoxin does have slightly more H-bonding within the protein than mesophilic rubredoxin, which is consistent with observations that H-bonds and electrostatic interactions between charged residues are better represented in thermostable proteins. However, all stabilization motifs from salt bridges to hydrophobic cores are found in thermostable proteins and do confer thermostability. Agnostic of the mechanism, what sets hyperthermostable proteins apart is the speed at which they fold. Briefly, thermophilic proteins fold more slowly than mesophilic proteins do at room temperature, and psychrophilic (cold-stable) proteins fold faster. At the optimal temperature for each protein, they all fold at about the same speed. This is achieved in the cell by chaperones, the chemical environment, and the intrinsic folding speed of the individual protein.

Here, we report the cryo-EM structural analysis and physical characterization of protein assemblies that mimic cannulae, which are tubular structures that form during the life cycle of the hyperthermophilic archaeon *P. abyssi* and related *Pyrodictium* species. In the native state of *P. abyssi*, cannulae embed in the periplasmic space and connect *Pyrodictium* cells, which grow and divide through binary fission (Nickell et al., 2003). This process results in a dense network of cells that are linked through the cannulae, which have been postulated to serve as a primitive extracellular matrix. Interestingly, cannulae production is concomitant with cell splitting; daughter cells remain attached to mother cells by cannulae. It is unknown what nutrients, nucleoproteins, etc. are transmitted through the hollow cannulae, i.e., whether they play the same role as bacterial pili. Nickell et al note that in 20 years of study, they have never seen a *Pyrodictium* mutant with no cannulae and speculated that it relies on the large, excreted protein structure for cell growth (Nickell et al., 2003). In cryo-ET experiments they find that cannulae appear to connect at any point of the cell (as opposed to T4SS pili, which extrude from one pole of the cell) and cross the periplasm but not the cytoplasm.

The biological function of the cannulae remains obscure, but it has been suggested that the tubular channels are involved in transport of nutrients, genetic information, or signal molecules between cells due to the large central lumen of the assemblies (Horn et al., 1999). The *P. abyssi* cannulae consist of at least three homologous glycoproteins, CanA, CanB, and CanC, which have molar masses in the range from 20-24 kDa. Self-assembly of a recombinant protein derived from secretion signal processed sequence of CanA (Figure S1) results in the formation of tubular assemblies having a similar diameter and morphology to native cannulae (Kreitner et al., 2020). Here, we report the cryo-EM structure of the CanA nanotubes, which reveals an unusual mode of interaction between protomers in the assembly. Donor-strand complementation, a form of non-covalent polymerization that has been observed in the structures of bacterial and archaeal pili, drives the self-assembly of CanA in the presence of metal ions. The resultant CanA nanotubes display an extensive network of non-covalent contacts that extend over the contour length of the assemblies. In contrast to Chaperone-Usher (CU) pili, in which donor strand complementation leads to the formation of filaments that display an elastic mechanical response, a structurally similar process of strand exchange in the CanA nanotubes results in high mechanical stiffness and thermal stability.

Results

CanA

Previous researchers have demonstrated that the CanA protein could be expressed and purified as a recombinant protein in bacterial host systems in reasonably good yields (2.1 g protein/250 g well cell weight) (Kreitner et al., 2020; Nelson R. Barton, 2008). Here, we expressed a version of CanA lacking the N-terminal signal peptide (Kreitner et al., 2020) from a codon-optimized genetic cassette as a transcriptional fusion under the control of a T7 promoter in plasmid pD451-SR. Following the literature procedure for purification, recombinant CanA could be obtained in good yield (>150 μ M protein without concentrating) from the bacterial cell lysate of *E. coli* strain BL21 (DE3) expression cultures after IPTG induction (see *Methods*). The purity and identity of the recombinant CanA protein were confirmed using

SDS-PAGE analysis and electro-spray ionization (ESI) mass spectrometry (Figures S2 and S3). Previously (Nelson R. Barton, 2008), recombinant CanA had been reported to self-assemble in a low-salt polymerization buffer (50 mM Tris-HCl, pH 7.5, 80 mM NaCl, 9% glycerol) after addition of CaCl_2 and MgCl_2 to a final concentration of 20 mM each, although the need for the presence of both alkaline earth metal cations was not further clarified. In our hands, negative-stain TEM images of recombinant CanA assembled under identical conditions revealed a structurally uniform population of high aspect-ratio tubular assemblies of ~ 25 nm in diameter (Figure 5.1), which was consistent with previous reports on natively occurring and recombinant cannulae. The CanA assemblies were observed to laterally associate along the contour length of the assemblies, which often led to the formation of large rafts of nanotubes. Similar behavior was observed for cannulae under native growth conditions in cultures of *P. abyssi* strain TAG11, in which freeze-etching revealed the presence of bundles containing up to 100 cannulae at the cell surface (Rieger et al., 1995).

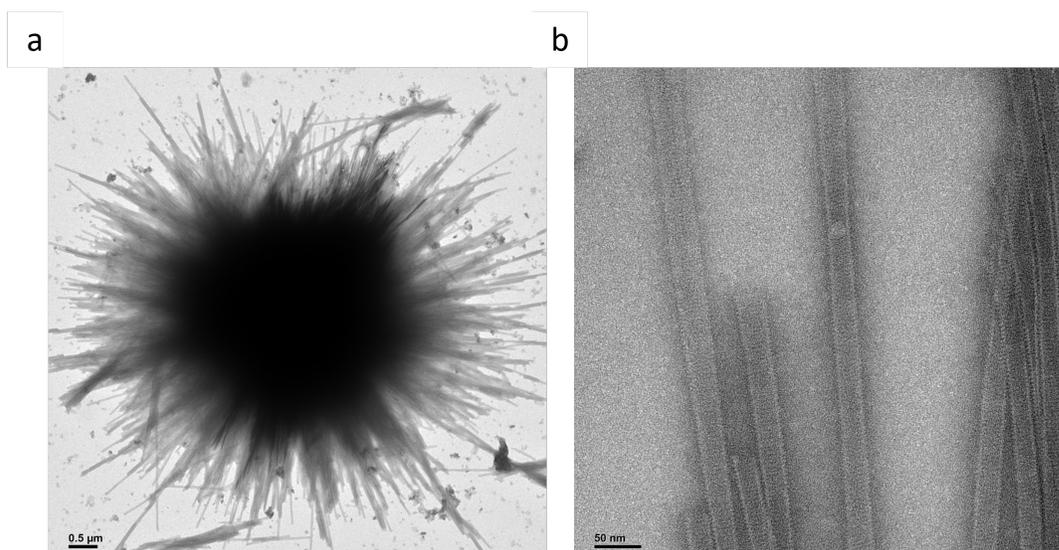


Figure 5.1. a) "Sea urchin" appearance of clusters of CanA tubes. b) Closer view of tubes showing ridged surface.

To investigate the role of the metal cation within the assembly, excess EDTA was added to CanA tubes assembled in either 40 mM Ca^{2+} or 40 mM Tb^{3+} . At room temperature, neither sample exhibited any changes. When heated to 60°C for one hour, both samples became optically clear. TEM showed that the

Ca^{2+} nanotubes dissolved completely, but the Tb^{3+} assemblies were damaged but still visible (Figure 5.3). To investigate lower concentrations of protein and metal ions, 50 μM CanA was incubated with 40mM (control), 30, 20, or 10 mM each Ca^{2+} , Zn^{2+} , and Tb^{3+} , as well as 100 μM Tb^{3+} . All Zn^{2+} and Tb^{3+} samples became turbid within seconds at room temperature. After incubation at room temperature overnight, all samples had visible white precipitate. 5 μM CanA with 40 mM Ca^{2+} did not produce enough precipitate to be visible to the eye, but nanotubes were clearly visible via TEM (SI figure 8).

When excess EDTA was added to the protein solution prior to cation addition, CanA failed to assemble. However, it did assemble with either cation in the presence of 0.2% octyl glucoside. When assemblies were boiled at 100°C for 30 minutes, the Ca^{2+} -mediated assemblies were damaged but still present. The Tb^{3+} assemblies appeared normal, except some of them had fluffy-looking debris on the surface (Figure 5.3c,d).

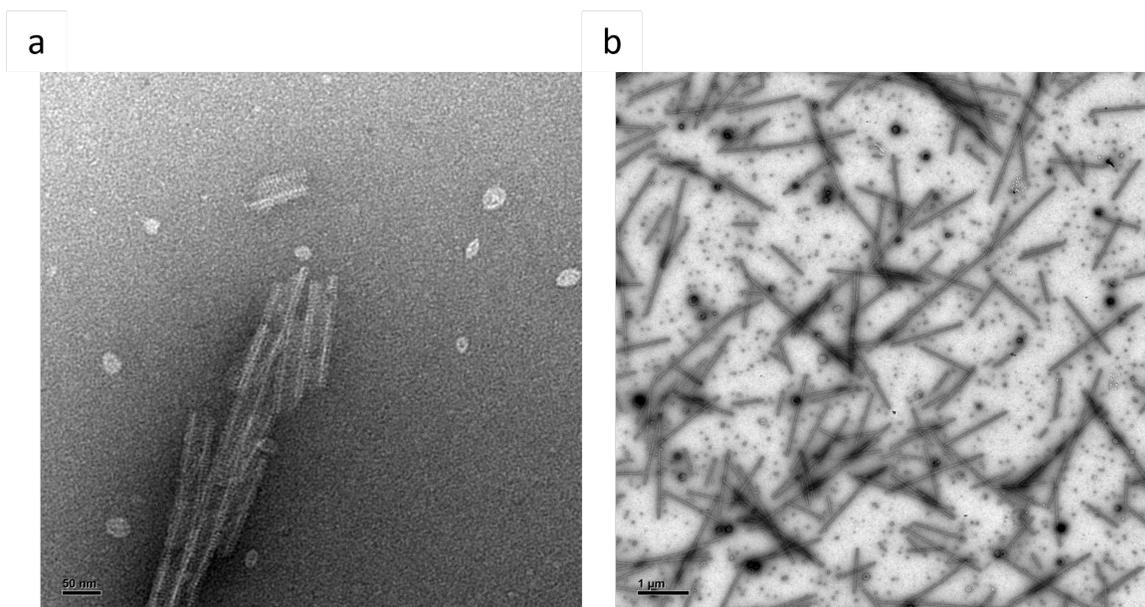


Figure 5.2. a) TEM micrograph showing poorly-formed CanA assemblies mediated by Zn^{2+} . b) TEM micrograph showing high density of long, well-dispersed CanA assemblies mediated by Tb^{3+} . Circles are not short rings of CanA, but other debris, possibly aggregates of a contaminating protein.

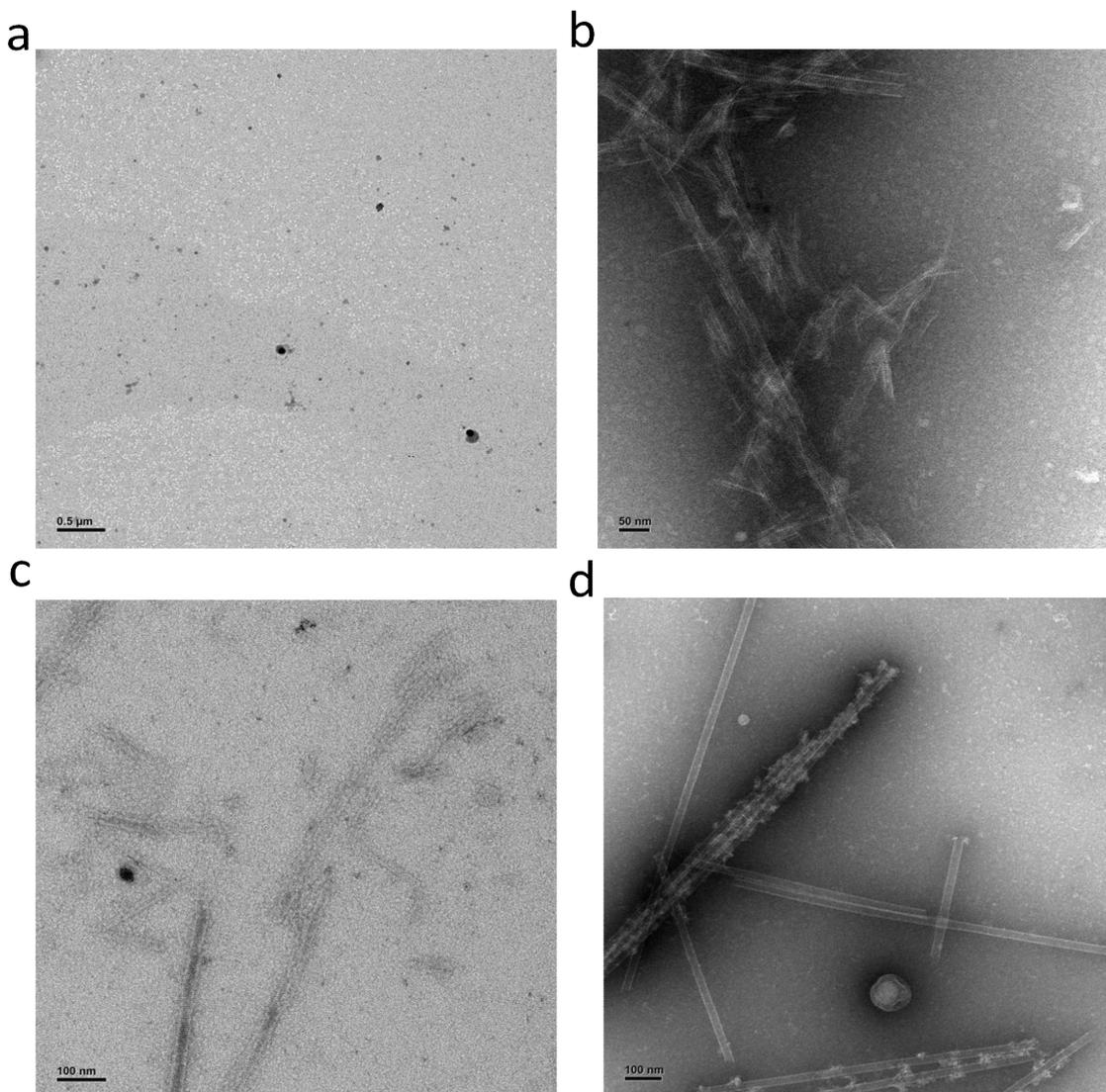


Figure 5.3. Heat and EDTA effects on CanA tubes. A) CanA tubes made with Ca^{2+} incubated at 60°C in excess EDTA. Tubes are completely dissolved; solution was optically clear. B) CanA tubes made with Tb^{3+} incubated at 60°C in excess EDTA. Tubes are badly damaged, especially along the azimuthal dimension. Solution is optically clear. C) CanA tubes incubated at 100°C for 30 minutes without EDTA. Tubes are badly damaged, especially along the azimuthal dimension. Solution is optically clear. D) CanA tubes made with Tb^{3+} incubated at 100°C for 30 minutes without EDTA. Tubes appear unaffected except for debris attached to some tubes, but not all.

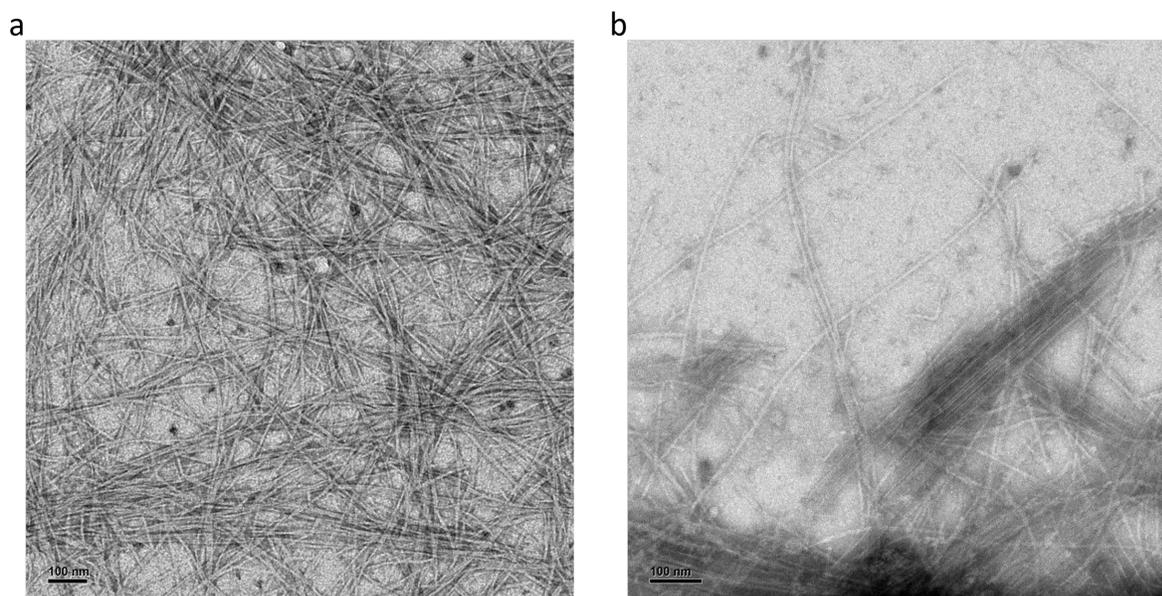


Figure 5.4. A) N20 peptide alone. β fibrils formed as soon as peptide was resuspended in low-salt buffer. B) N20 and terbium-assembled CanA tubes. The peptide failed to incorporate into the nanotube assembly.

The 20 N-terminal residues are predicted in AlphaFold (Jumper et al., 2021; Mirdita et al., 2022) to form a tail protruding from the protein, and were expected to participate in domain-swapping with adjacent proteins in the assembly. This peptide, called N20, was mixed with CanA prior to addition of the cation. N20 formed β -sheet fibrils as soon as it was dissolved for this experiment and was not incorporated into the nanotubes.

Cryo-EM analysis was employed to determine the structure of the CanA assemblies. Despite some lateral association, a sufficient number of isolated particles could be classified to serve as the basis for helical reconstruction (Figure 5.5a, inset). The 2D class averages confirmed that the nanotubes were uniform in inner and outer diameter, 18 nm and 27 nm, respectively, despite *in vitro* assembly conditions that differed significantly from the native environment of *P. abyssi*. A 3D volume was reconstructed from the 2D projection images using iterative helical real-space reconstruction (IHRSR) after determination of

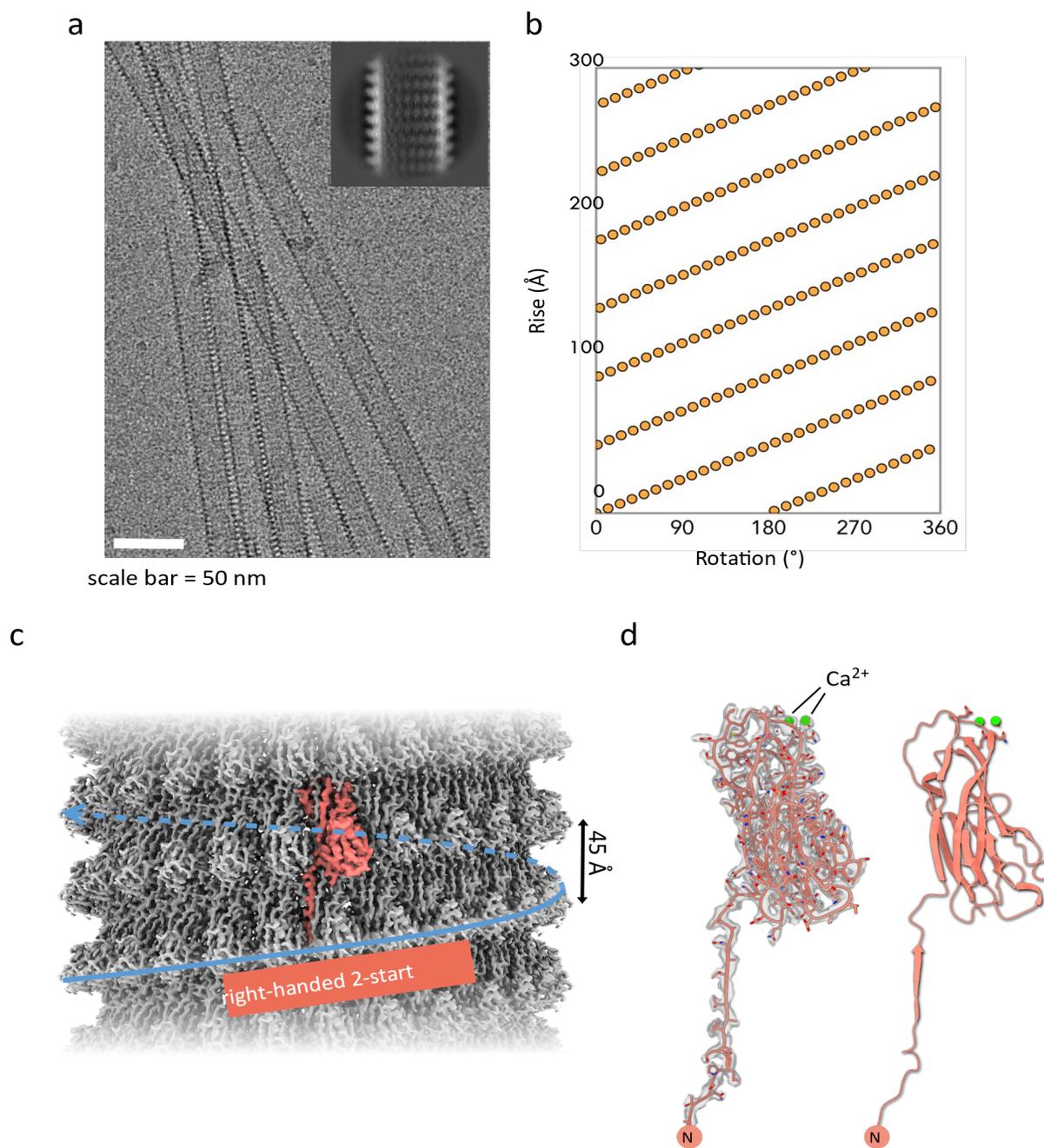


Figure 5.5. A) CanA tubes under 300 kV cryo-electron microscope. Ridged surface is clearly visible. Single proteins are visible in D class inset. B) Helical map diagram of CanA showing right-handed 2-start helical symmetry. C) Density map of CanA filament. D) Single CanA protein in map (L) and cartoon (R) showing beta-roll fold and calcium ions.

helical symmetry. An unambiguous atomic model could be built into the 3D density at a resolution of 2.6 Å using the map-model FSC as a criterion (Figure S5). The structure of the CanA tube was based on C1 symmetry derived from a left-handed 1-start helix with an axial rise of 1.59 Å and an angle of rotation of -173.8° (Table 5-1). The most prominent structural feature in the helical net diagram (Figure 5.5b) of the CanA assemblies was the presence of a pair of right-handed 2-start helices that corresponded to a 4.5 nm periodicity, i.e., half the pitch of the 2-start helices, which was easily observable in the 3D reconstruction (Figure 5.6a) as well as in TEM images of native cannulae and recombinant CanA tubules.

The structure of the CanA protomer (Figure 5.5d) was based on a nine-stranded β -sandwich fold in which the N-terminal strand participated in donor strand complementation with an acceptor groove located in a protomer that lay twenty-nine subunits away along the 1-start helix. AlphaFold2 (AF2) provided an accurate prediction of the core structure of the CanA protomer (based on pLDDT score, SI Figure 5-16) with the exception of N-terminal donor strand (SI Figure 5-15), despite the absence of sequence homology to other known proteins. Notably, ^1H NMR analysis of the chemical shifts of a truncated CanA protein, K1-CanA, in which the ten N-terminal residues were deleted (SI Figure 5-1), revealed a similar distribution of secondary structure along the backbone to that observed in the atomic model of the CanA protomers in the nanotubes (SI Figure 5-4). The atomic model of the CanA tubules suggested that donor strand complementation was the primary mechanism that drove formation of the cannulae and stabilized the resultant assemblies. The propagation of donor strand complementation occurred along the contour length of the cannula and coincided with the 29-start helices within the assembly. The relative orientation of protomers along the 29-start helices was defined by a rotation of 1.25° and an axial rise of 46 Å, which indicated that successive protomers along the 29-start helices were aligned nearly parallel to the central axis of the helical assembly. In addition, donor-strand complementation linked protomers located across the interface between the 2-start helices along the 29-start helices, which reinforced the structural interfaces between protomers in the CanA tubule.

The second noteworthy structural feature of the CanA assemblies was the presence of two calcium ions per protomer (Figure 5.6d). While both calcium ions and magnesium ions were present in the

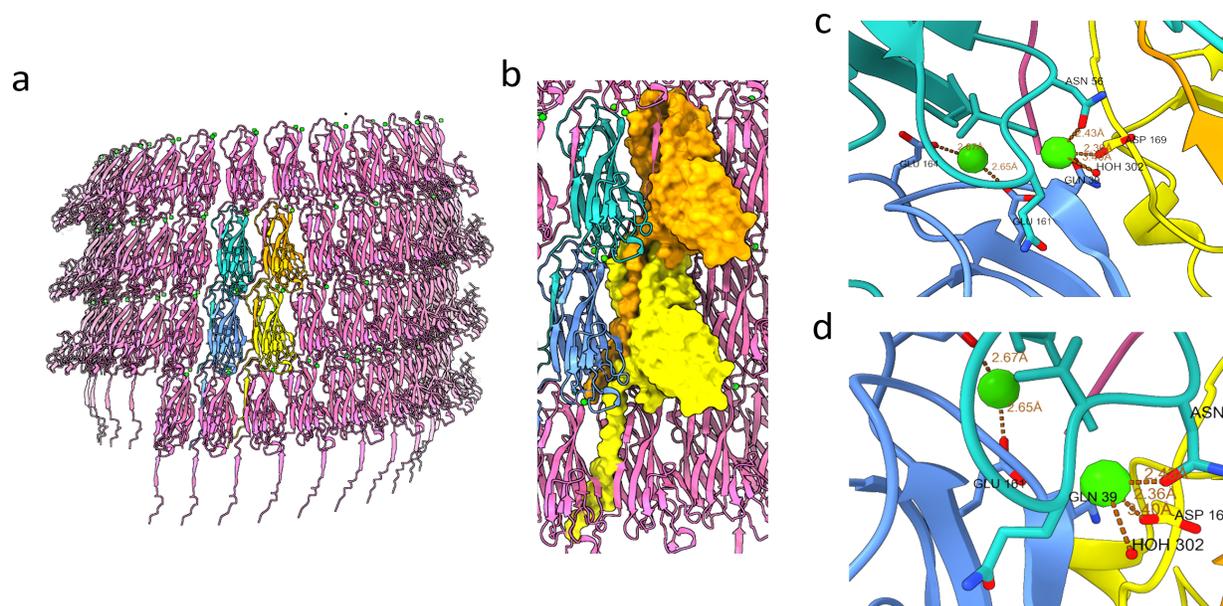


Figure 5.6. CanA intra-filament interactions. A) Front slice of entire filament. B) Close-up view of A with space-filling models of two proteins. The N-terminus of the orange CanA protein associates with the yellow protein via donor strand complementation. C and D) Two angles of view of the calcium ions (green) mediating interaction between blue, teal, and yellow proteins.

polymerization buffer, experimental refinement of the structure was more consistent with calcium ion occupancy of the binding sites. In addition, subsequent polymerization experiments performed solely in the presence of calcium or magnesium indicated that only the former ion could initiate polymerization under experimental conditions (*vide infra*). The calcium ion binding sites bridged protomers between structurally adjacent 2-start helices that were aligned along the 29-start helices. However, the structural analysis also revealed a more extensive network of inter-protomer contacts that involved subunits located as far as sixty units on either side of a central protomer (Figure 5.6a,b). PISA (Protein, Interfaces, Surfaces, and Assemblies) (Krissinel and Henrick, 2007) analysis indicated that while donor strand complementation between protomers along the 29-start helices buried the greatest amount of surface area, significant interactions also occurred between a more extensive network of protomers including those located within the same 2-start helix and were often mediated through calcium ion coordination.

The presence of the calcium ions was essential to initiate the formation of the CanA assemblies and,

presumably, to additionally stabilize the resultant structure. While donor strand complementation has been observed in bacterial pili, the proper formation of the assemblies often requires the assistance of specialized machinery such as that observed in the chaperone-usher system (Busch and Waksman, 2012; Geibel and Waksman, 2014; Werneburg and Thanassi, 2018). Although, in theory, donor-strand complementation could initiate polymerization on its own, no evidence for polymerization was observed in the absence of calcium ion. In the CanA assembly, the N-terminal donor strand formed a bridge between the second and ninth β -strands of a successive protomer along the 29-start helix (Figure 5.6b). However, the resulting β -sheet interactions involve only a few residues, with 21 H bonds forming the 1280.1 \AA^2 interface.

In contrast to other donor strand complemented systems, the hydrophobic surface area exposed on either the donor strand or the acceptor groove was not very substantial. Calcium ion binding may be essential to orient the structure in order to accommodate the donor strand-acceptor groove interaction as well as to stabilize the structural interactions at the inter-protomer interfaces. The formation of this extensive network of interactions results in the formation of a highly crosslinked, polar filament.

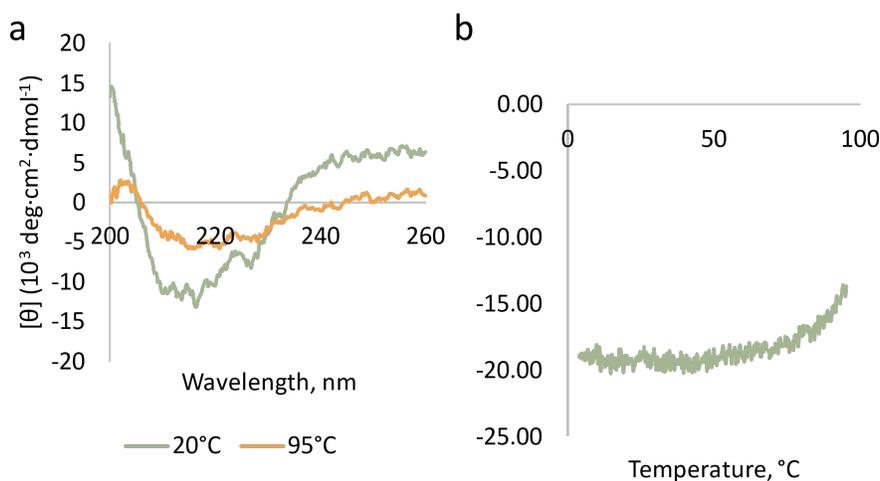


Figure 5.7. A. CD spectra for CanA at 20°C and at 95°C. The spectrum at 20°C appears to indicate β -sheet content, with a minimum at 217 nm. There is also some evidence of α -helical character around 200 nm. At 95°C, the signal degrades but shows more β -sheet features, especially in the 200-210 nm region. B. CanA melt between 25 and 95°C at 216 nm. The signal trends toward 0 at 80°C but is not indicative of complete unfolding. There is likely some conformation change at high temperatures.

CanA tubes assembled with calcium were found to be very brittle by AFM. They collapse on the mica when dried in air, consistent with literature results (Rieger et al., 1995). They were also very brittle when scanned by AFM in buffer at 500 pN force, so all scans show features with dimensions consistent with flattened cannulae (Figure 5.8). Quantitative nanomechanical mapping was unable to discriminate between the tubes and the mica behind them.

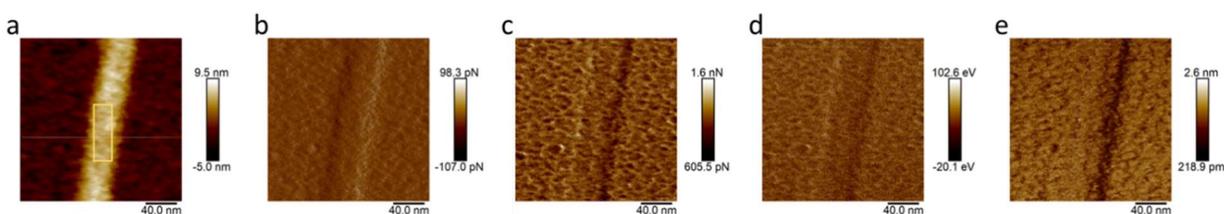
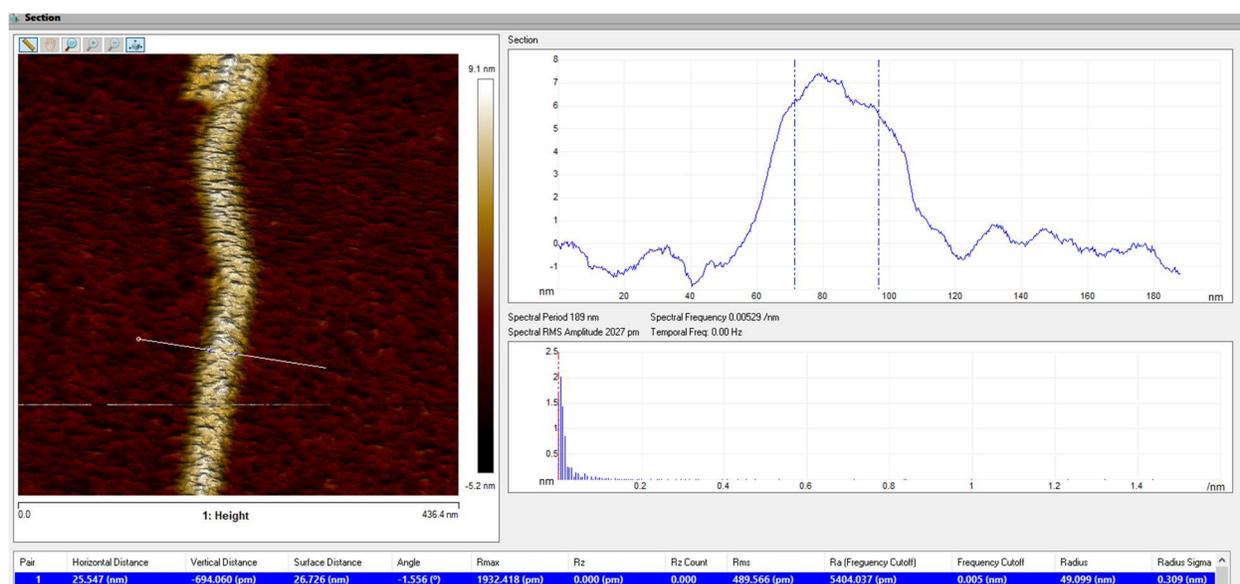


Figure 5.8. CanA under AFM in air. Images and data are consistent with filaments that collapse on the dry carbon-coated mica. A. Height of filament in boxed area average 2 nm. B. Peak force error, average 0.146 nN. C. Adhesion average 0.589 nN. D. Dissipation average 64.9 eV. E. Deformation average 1.21 nm.

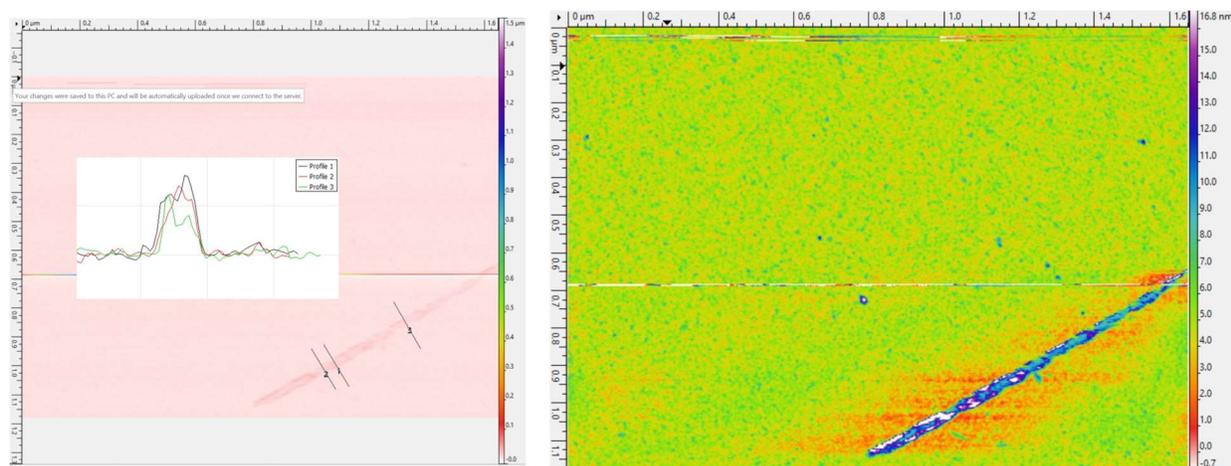


Figure 5.9. AFM in fluid showed that filament is extremely brittle. QNM cannot discriminate stiffness of assembly vs mica.

To investigate the role of the metal cation within the assembly, excess EDTA was added to CanA tubes assembled in either 40 mM Ca^{2+} or 40 mM Tb^{3+} . At room temperature, neither sample exhibited any changes. When heated to 60°C for one hour, both samples became optically clear. TEM showed that the Ca^{2+} nanotubes dissolved completely, but the Tb^{3+} assemblies were damaged but still visible (Figure 5.3b). To investigate lower concentrations of protein and metal ions, 50 μM CanA was incubated with 40mM (control), 30, 20, or 10 mM each Ca^{2+} , Zn^{2+} , and Tb^{3+} , as well as 100 μM Tb^{3+} . All Zn^{2+} and Tb^{3+} samples became turbid within seconds at room temperature. After incubation at room temperature overnight, all samples had visible white precipitate. 5 μM CanA with 40 mM Ca^{2+} did not produce enough precipitate to be visible to the eye, but nanotubes were clearly visible via TEM (SI Figure 5-10).

Assembled CanA tubes demonstrate a remarkable affinity for calcium and terbium ions. Typically, excess EDTA can strip calcium from metalloproteins. Tubes could not assemble in the presence of EDTA. However, adding excess EDTA to assembled CanA tubes had no effect on either calcium nor terbium-assembled tubes when incubated at room temperature. However, when heated to 60°C for one hour, the calcium tubes completely dissolve (Figure 5.3a), but the terbium tubes take on a shredded appearance, tearing along the azimuthal axis (Figure 5.3b). Unlike their naturally occurring counterparts,

CanA tubes assembled with calcium were not able to withstand boiling or 30 minutes (Figure 5.3c). The assemblies were mostly intact but clearly disrupted along the azimuthal axis. Boiled terbium tubes were unaffected (Figure 5.3d).

Analytical ultracentrifugation failed to determine the nature of the CanA dimer; i.e. The interaction that forms first when a tube nucleates. Both Ca^{2+} and Tb^{3+} were tested. However, despite multiple attempts, no dimers or trimers were observed. For Ca^{2+} -assembled tubes, low-concentration was mixed with 1 or 10 mM Ca^{2+} for the run at 20°C. Only 20 kDa particles, corresponding to the mass of single CanA proteins, were observed. Samples of CanA and Ca^{2+} were incubated at room temperature for two days but these also failed to assemble. The Ca^{2+} samples were insensitive to oligomerization, or mass action. Low and high concentration behaved identically. It is possible that a much higher calcium concentration could produce dimers during the experiment, but we were not able to access this condition. CanA with terbium, on the other hand, failed to sediment as a single particle. All samples with Tb^{3+} aggregated to very large particles which sedimented completely before the AUC run began. These samples could not be measured because these aggregates were too large to be detected at the speeds used in the run.

These results are somewhat consistent with previous observations. Typically, CanA is assembled at 40 mM CaCl_2 and at least 100 μM protein to produce abundant nanotubes. However, concentrations as low as 40 mM CaCl_2 and 5 μM protein have been demonstrated to produce nanotubes, though they are not as easy to confirm via TEM simply because they are few and far between (SI Figure 5-10). Terbium-assembled tubes assemble faster than calcium-assembled, and the binding interaction is much stronger. The observation that all particles were large aggregates is consistent with TEM observations. Terbium-assembled tubes assemble faster than calcium-assembled, and the binding interaction is much stronger. The observation that all particles were large aggregates is consistent with TEM observations.

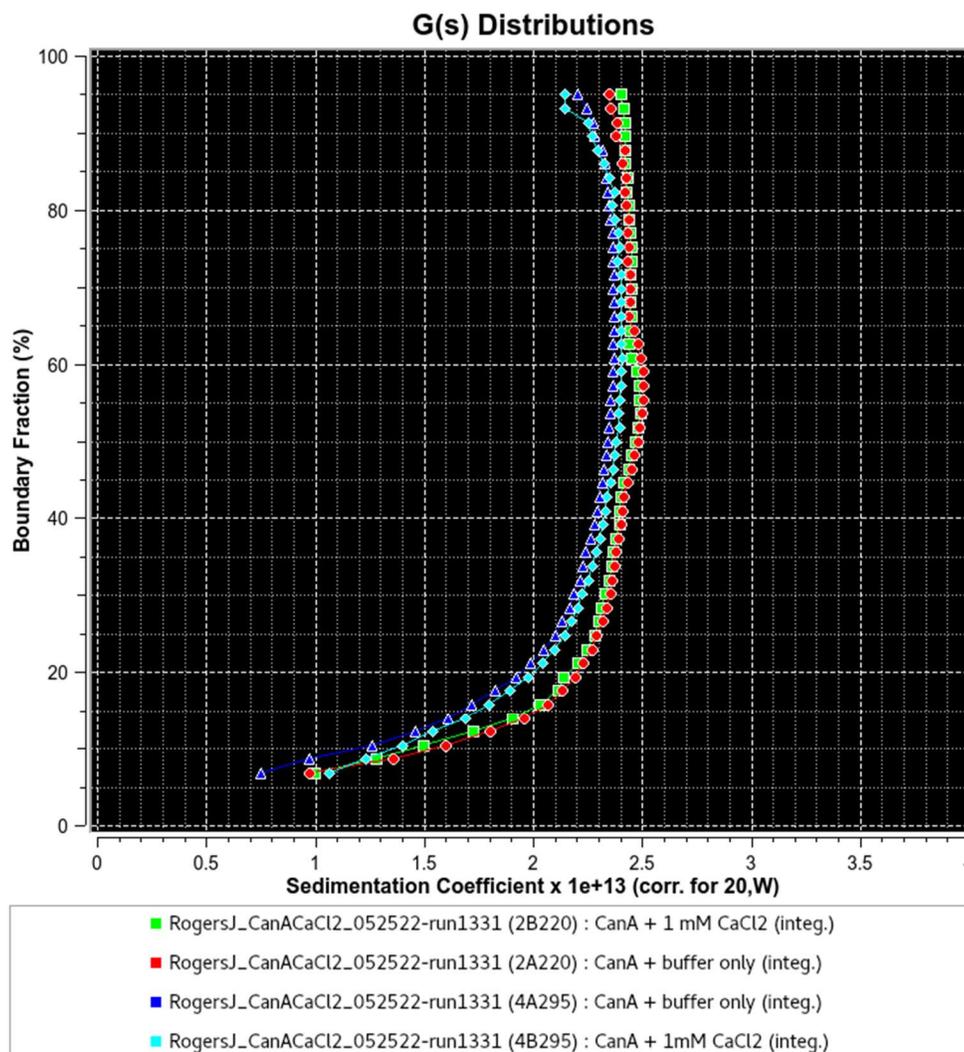


Figure 5.10. Sedimentation coefficients for CanA±Ca²⁺. There is no difference between the samples because no assemblies are observed under AUC conditions.

Hyper2

Based on these findings, it was hypothesized that proteins with a similar sequence to CanA might share its thermostable and self-assembling properties. An NCBI BLAST search was performed to find proteins homologous to CanA. Two proteins from a metagenomic analysis, attributed to *Hyperthermus* sp., called Hyper1 and Hyper2 (Trembath-Reichert et al., 2019), as well as a protein from *Pyrodictium occultum* named POC 386 (Stetter et al., 1983), shared at least 25% sequence identity with the CanA

sequence queried. Of these three proteins, only Hyper1 and Hyper2 expressed in *E. coli*; of these, only Hyper2 was soluble. Hyper2 was successfully purified and assembled in the same manner as CanA with calcium; however, terbium and zinc resulted in nonspecific aggregation of protein.

When incubated at 80°C, the Ca²⁺ and Mg²⁺ samples developed a white precipitate after an hour, which was much greater in the Ca²⁺ sample. Notably, the no-cation control sample also developed a small amount of white precipitate. Ca²⁺ and Mg²⁺ samples also became turbid at 60°C with identical results.

All assemblies appeared as tubes or high-aspect-ratio curled sheets, which could be broken or poorly formed tubes (Figure 5.11). Median width in representative micrographs was 595 nm, average was 754 nm. Strikingly, assemblies also appeared to be multi-layered or nested, with variable diameters from 23 to 43 nm. The surface was smooth for all assemblies, but a crosshatched appearance indicated periodicity. Assemblies were well-dispersed on the carbon film. TEM showed nonspecific aggregation in the zinc and terbium samples (not shown).

Surprisingly, precipitate from only heat was in fact nanotubes, identical to those formed with Ca²⁺ or Mg²⁺ (Figure 5.11c). No precipitate formed after several days at room temperature without added Ca²⁺ or Mg²⁺. 40 mM Ca²⁺, 80°C was selected as the best assembly condition (Figure 5.11).

In contrast to the relative ease of solving the structure of CanA, which had excellent filament distribution and ice quality on the first try and only a single symmetry, Hyper2 would only adhere to a positively glow discharged grid, and the nested multiple layers visible by negative stain correspond to multiple symmetries of the assembly. Once the cryo-EM dataset was collected, visual inspection of the 2D class averages suggests that the Hyper2 tubes are heterogeneous in diameter and packing. Several discrete diameter tubes (240 Å, 270 Å and 290 Å) and discrete packing, mono-, di- and triple-layers in the wall of tubes were observed (Figure 5.15). We focused on solving the structure of the most abundant population, the 270 Å diameter tube having a single-layer wall.

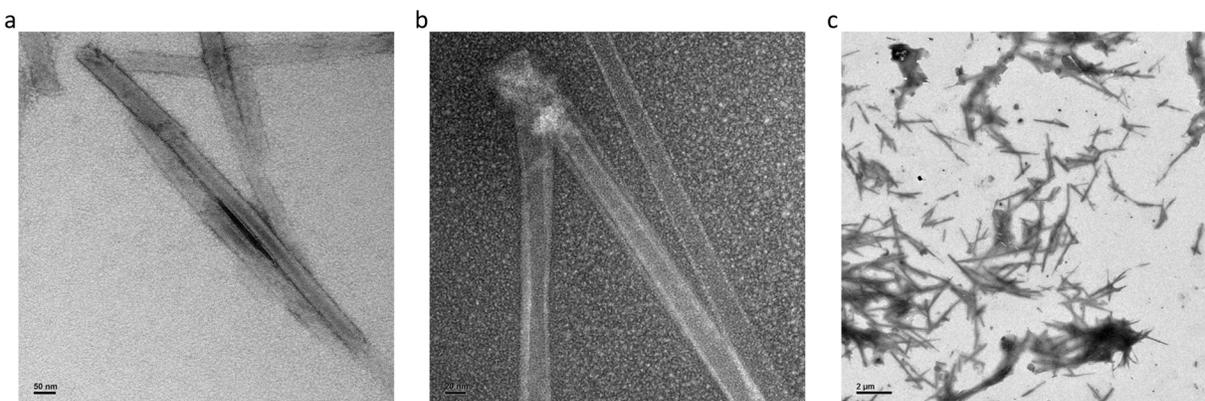


Figure 5.11. Negative-stain TEM micrographs of Hyper2 assembled in a) 40 mM CaCl_2 , b) 40 mM MgCl_2 , c) 80°C heat only with no added cation.

The 8-start Hyper2 helix was solved to a resolution of 3.5 \AA . The right-handed helix assembles via donor strand complementation like CanA, but the protomers are much more “tilted” back relative to the helical axis, unlike the nearly parallel CanA protomers. A density attributed to a single calcium per protein is in a different location than the CanA calcium coordination site (Figure 5.14a,b). The ion is coordinated by Glu112 of one protein, and Asn23 and Glu20 of the adjacent protein in the 8-start helix (Figure 5.14b). Importantly, the 7-, 8-, and 9-start helices all share similar power spectra (Figure 5.15d,h,l), indicating that they share the same rise of 11.15 \AA and twist of 9.58° (Table 5-1).

CD spectra at 20°C and 95°C after melting are nearly identical, with a distinct minimum at 216 nm, indicative of high β -sheet content at both temperatures. Upon melting from 4 to 95°C at 216 nm, there is no significant change to the signal, indicating that secondary structure is consistent throughout the temperature range.

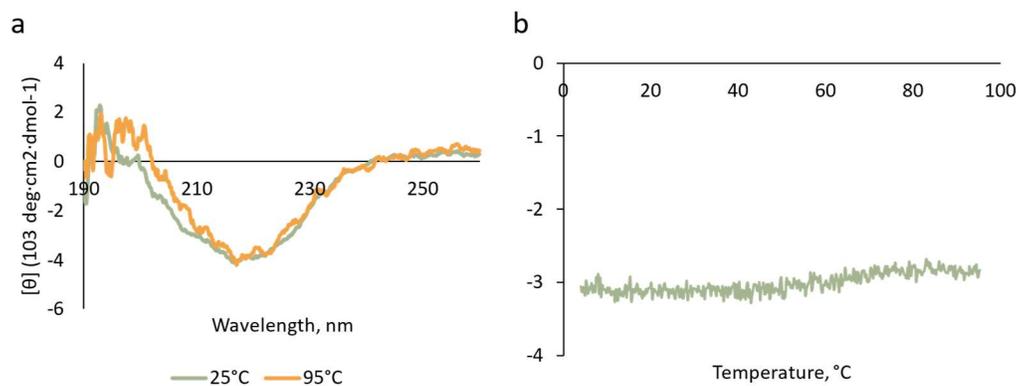


Figure 5.12. a) CD spectra for Hyper2 at 25°C and at 95°C after melting. b) Melt curve showing no change in ellipticity from 4°C to 95°C at 216 nm wavelength. Both graphs are consistent with no to negligible thermal conformational change in this range.



Figure 5.13. Cryo-EM micrograph showing Hyper2 assemblies.

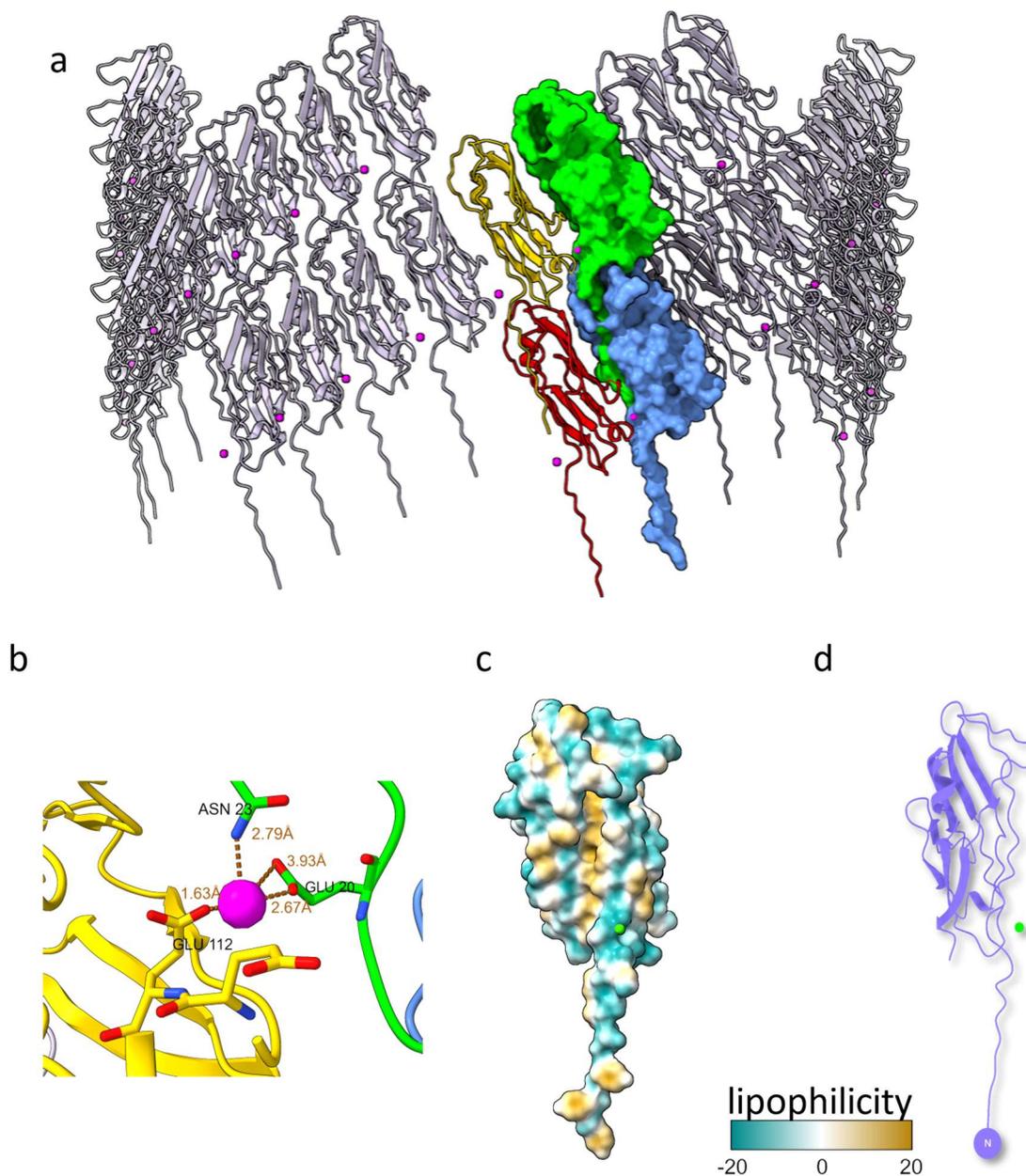


Figure 5.14 A. Model of 8-start Hyper2 assembly. Calcium ions shown in pink. Green and blue space-filling models demonstrate donor strand complementation mediating axial interactions. B. Close view of calcium ion coordinating residues on azimuthally adjacent green and yellow Hyper2 proteins. C. Lipophilicity diagram for Hyper2, showing overall low lipophilicity/high hydrophilicity and high tyrosine content of the N-terminus. D. Simplified model of Hyper2 showing high B-sheet content.

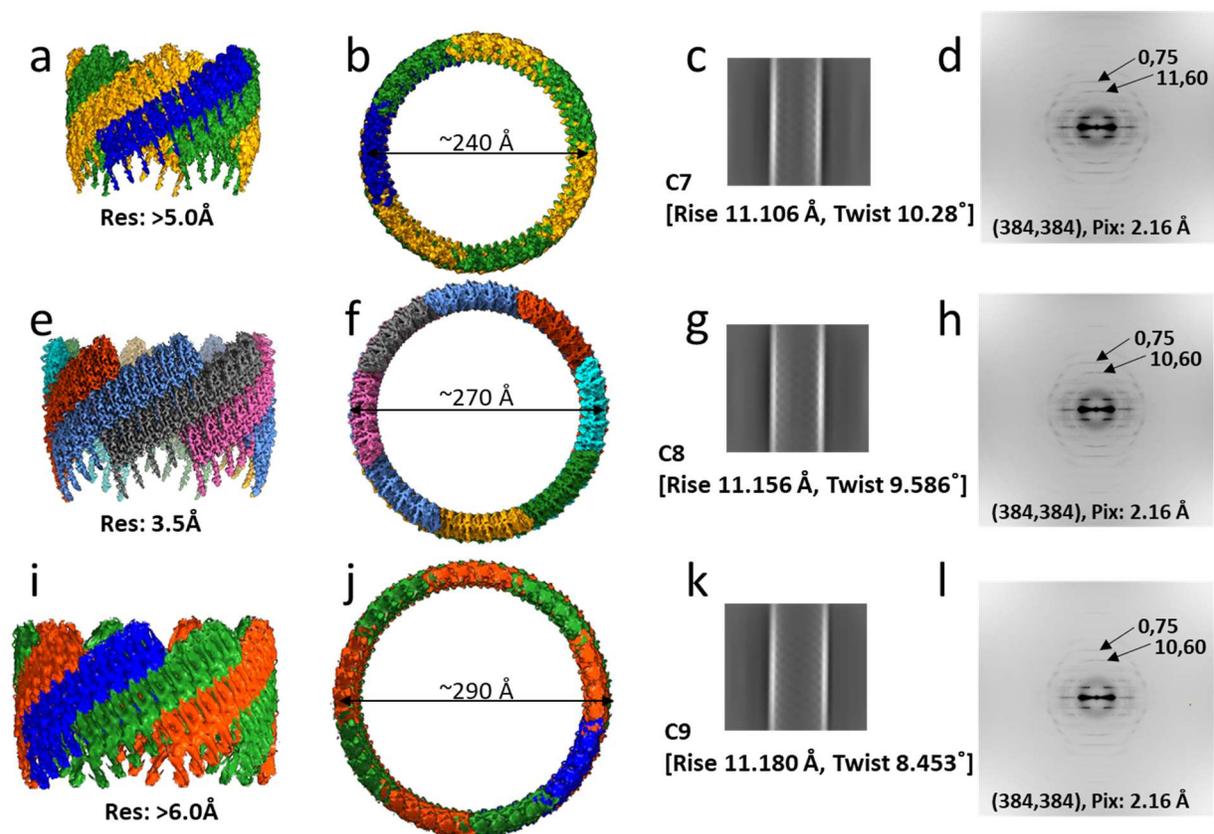


Figure 5.15. Cryo-EM maps of single-walled Hyper2 tubes. A-E) The C7 symmetry assembly has a diameter of about 240 Å and was solved to a resolution of >5Å. F-J) The C8 symmetry assembly is about 270 Å in diameter and was solved to 3.5Å resolution, the best out of the observed single-walled tubes. K-O) The C9 symmetry assembly is about 290 Å in diameter and was solved to >6 Å resolution. All assemblies are right-handed and have similar rise and differ slightly in twist. Layer lines are also very similar between all classes.

No change to filaments was observed under excess (200 mM) EDTA added to Hyper2 tubes assembled in Ca²⁺ or Mg. Filaments looked identical whether exposed to EDTA at room temperature or 80°C (Figure 5.16a,b). Excess EDTA mixed with protein before the cation was added prevented any assembly at room temperature, but tubes were observed after heating (Figure 5.16c). Assembled tubes heated to 100°C for 20 minutes had a degraded appearance, with kinked tubes and tattered single layers (Figure 5.16d).

Atomic force microscopy with quantitative nanomechanical mapping was attempted with Hyper2, but was not successful. Filaments adhered to the carbon-coated mica, but appeared crushed. Multi-layering was also visible (Figure 5.17)

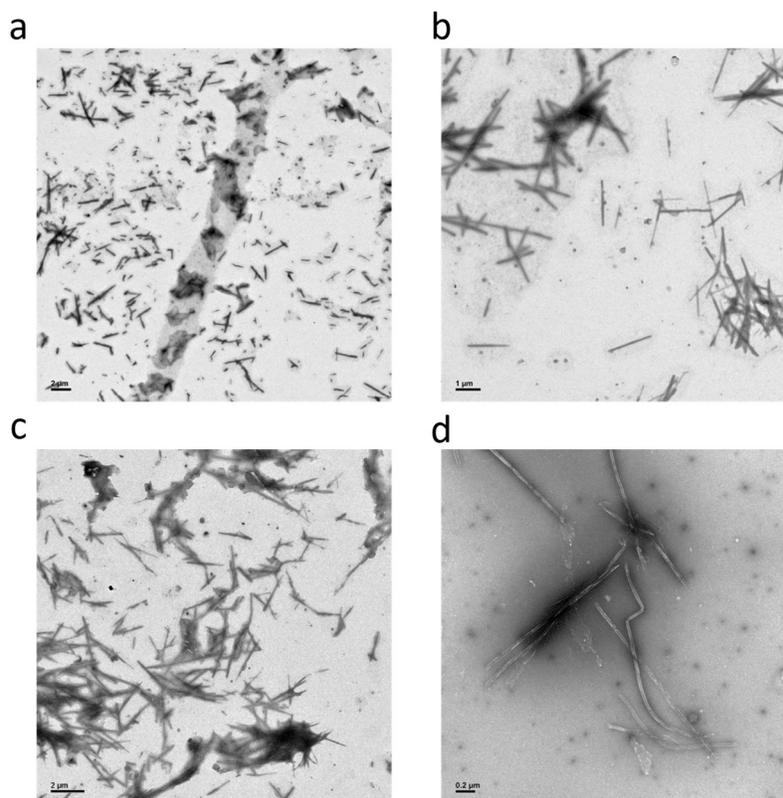


Figure 5.16. a) EDTA added to Hyper2 assemblies in Ca^{2+} , incubated at room temperature. b) EDTA added to Hyper2 assemblies in Ca^{2+} , incubated at 80°C for 20 minutes. In A and B tubes appear to be unaffected by EDTA. c) Hyper2 mixed with EDTA, then Ca^{2+} , incubated at 80°C for 20 minutes. d) Hyper2 assembled in Ca^{2+} , then boiled 20 minutes. Assemblies are present but appear to be degraded.

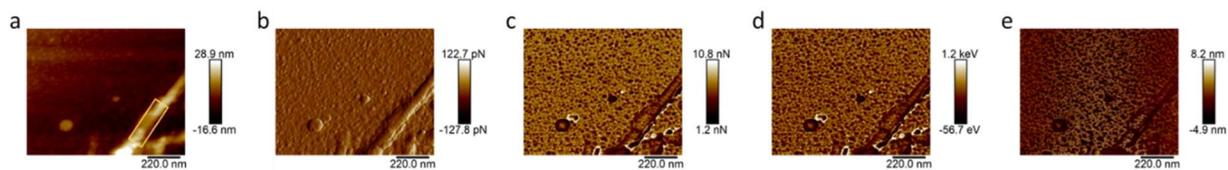


Figure 5.17. QNM for Hyper2 filament in air. Images and data are consistent with filaments that collapse on the dry carbon-coated mica and appear to be more brittle than CanA filaments. They show characteristic shape but variable thickness and height, consistent with their appearance in electron micrographs. A. Height of filament, 2.9-3.9 nm in boxed area. B. Peak force error, average 0.193 nN. C. Adhesion average 3.13 nN. D. Dissipation average 480 eV. E. Deformation average 8.76 nm.

Materials and Methods

Protein expression Genes were purchased as pD451-SR plasmids from ATUM (Newark, CA). Plasmids were transformed into chemically-competent Top 10 F' *E. coli* by heat-shock for plasmid amplification. One colony was selected for plasmid harvesting using a miniprep kit. Recovered plasmid was transformed into BL21 Gold (DE3) *E. coli*.

Protein was expressed via traditional IPTG expression. One colony of transformed BL21 Gold (DE3) *E. coli* was selected for overnight growth to terminal OD in LB medium with kanamycin. 4 mL culture was added to 196 mL LB+Kan and grown to $OD_{600}=0.6$. IPTG was added to a final concentration of 1 mM and protein expression was induced for six hours. The cell pellet was then harvested by centrifugation. Cells were gently resuspended to 2 mL in “low salt buffer” (Tris with enzymatic lysis cocktail (lysozyme final concentration 1.25 mg/mL, 1 μ L Benzonase nuclease (Merck KGaA, Darmstadt, Germany), $MgCl_2$ final concentration 1 mM, Protease Inhibitor Cocktail Set V, Animal-free (Millipore Corp., Billerica, MA) (final concentration 1X) and incubated with shaking 30°C for one hour. Cell lysate was clarified via centrifugation (Eppendorf centrifuge, 14000 rpm, 10 minutes). Clarified lysate was heated to 80°C in a heat block for 20 minutes to remove heat-labile proteins and purify the protein of interest. Heat-clarified protein was dialyzed overnight against low-salt buffer to remove other metabolites. Final protein concentration was $\sim 400 \mu M$ with excellent purity (SI Figure 5-3. SDS-PAGE gels from CanA and Hyper2 expression and purification. A. 19 kDa CanA is not expressed without IPTG, but is visible after 6 hours IPTG-induced expression. Neither the pellet from the enzyme lysis with lysozyme and Benzonase nor the pellet from incubating the enzyme supernatant at 80°C contain CanA. Supernatant from the heating step

contains CanA at high concentration and purity. B. Hyper2 expresses well under IPTG expression. Some Hyper2 partitions to the enzyme lysis pellet fraction, but most remains soluble. Some Hyper2 also crashes out of the heat purification, which was later determined to be heat-mediated nanotube assembly. Hyper2 is present in the heat-purified supernatant at high concentration and good purity. (SI Figure 5-3a). Protein samples were dialyzed against water for mass spectrometry, which confirmed identity (SI Figure 5-7, SI Figure 5-8).

Nanotube assembly Following the patent, 5 μL each 0.1M CaCl_2 and 0.1M MgCl_2 were added to 100 μL high-concentration ($\sim 100 \mu\text{M}$) CanA for a final concentration of 5 μM each cation and heated to 80°C for 20 minutes. The sample was spotted onto TEM grids but did not yield any visible structures under transmission electron microscopy (TEM). The concentrations of each salt were then increased to 20 mM and the procedure was repeated. A white, flocculent precipitate was visible after heating.

Negative Stain transmission electron microscopy. 5 μL of each assembly condition was incubated on CF200-CU grids (Electron Microscopy Sciences, Inc, Hatfield, PA) for one minute before wicking from the side. 5 μL of a mixture of 50% NanoVan/50% NanoW (Ted Pella, Inc, Redding, CA) was placed on the grid, incubated one minute, and then wicked from the side.

Circular dichroism spectropolarimetry. 40 μL of μM protein was applied to a 0.1 mm quartz High Precision Cell cuvette (Hellma Analytics, Jena, Germany), which was assembled and wrapped with Teflon plumber's tape and Parafilm to retard evaporation. After blanking, a spectrum from 190 to 250 nm was measured in a JASCO J-1500 CD Spectrometer, with SpectraManager software. Thermal denaturation was monitored at 217 nm from 4 to 95°C, ramping up the temperature by 2°C per minute. After 95°C was reached, a second spectrum was measured. Molar ellipticity was determined using the

formula ellipticity= ((100*raw ellipticity)/ (concentration of protein in mM) (peptide bonds in protein) (0.01 cm path length))/1000 to compare proteins of different lengths.

Cryo-electron microscopy and structural determination. 3 μ L CanA assembled with calcium was deposited onto a negative-glow-discharged C-flat grid (Electron Microscopy Sciences, Hatfield, PA) or a (Gatan Solarus) lacey carbon grid (Ted Pella, Inc.), with blot time 4s, wait time 10s, blot force 0 and vitrified in liquid ethane using a Vitrobot Mark IV cryoplunger (Thermo Fisher) or an EM GP Plunge Freezer (Leica). 3 μ L Hyper2 assembled with calcium was deposited onto grids as above, except grids were positive-glow-discharged using a GloQube plasma cleaner (Quorum Technologies) or in the presence of n-amylamine (Gatan Solarus) (Morris and da Fonseca, 2017). After screening at 80 kV, the dataset used for structure determination was collected at the Molecular Electron Microscopy Core at the University of Virginia on a Titan Krios EM operated at 300 keV, equipped with an energy filter and K3 direct electron detector (Gatan). An energy filter slit width of 10 eV was used during data collection and was aligned automatically every hour. Data were collected using a total dose of 50 $e^- \text{ \AA}^{-2}$ across 40 frames with an exposure time of 2.98s.

Image processing and reconstruction was conducted through cryoSPARC v3.2.0 (Punjani et al., 2017). Cryo-EM movies were motion corrected and CTF (contrast transfer function) estimated by ‘patch motion correction’ and ‘patch CTF estimation’ jobs in cryoSPARC, respectively. Initially, filaments were manually boxed using ‘manual picker’ to generate the templates which were further used for template-based automatic boxing by ‘filament tracer’. Boxed segments were subjected to 2D classification to remove bad picks. Selected segments of this tube were used to generate an averaged power spectrum. Potential helical symmetries were indexed from the power spectrum and tested by a ‘helical refinement’ job to find the correct symmetry. The helical symmetry parameters and the resolution of map were further refined by iterative cycles of ‘helical refinement’ and ‘local CTF refinement’ jobs. The hyper2 tube 3D volume map having 3.5 \AA (FSC, 0.143) map:map Fourier shell correlation has been reconstructed by

applying the following refined helical symmetry parameters: axial rise, 11.156 Å; twist, 9.586°; point group, C8 (Table 5-1).

The initial structure of the Hyper2 protomer was generated in alphaFold (Jumper et al., 2021; Mirdita et al., 2022) and manually rigid-body fitted in the map using Chimera (Goddard et al., 2018). The fitted model was refined against the density map by iterative cycles of interactive refinement in Coot (Emsley and Cowtan, 2004) and automatic refinement by PHENIX (Afonine et al., 2018). Density for eight (1-8) N-terminal residues and four (149-152) C-terminal residues was not resolved in helical map and thus these residues were left unmodeled. Tubes were observed to be multi-layered, in agreement with the negative-stain images. Even single-walled tubes have heterogeneous diameters. Three 2D classes have similar power spectra and only differ by cyclic symmetry (Figure 5.15). The class with 235 Å diameter has C7 symmetry (Figure 5.15a-d), 270 Å has C8 symmetry (e-h), 290 Å has C9 symmetry (i-l). The best resolution (3.5Å) was achieved with the 270 Å diameter, C8 symmetry class.

Reconstructions of nanotubes were generated using the following method: (1) an averaged power spectrum was generated using the raw images of aligned filament segments selected from 2D classification, (2) layer lines were indexed to produce a list of possible helical symmetries, and (3) the correct helical symmetry was determined by trial and error by inspection of an output 3D map looking for obvious structural motifs (i.e., recognizable secondary structural and amino acid side chain densities). Particles were further refined using local CTF refinement, and another round of helical refinement was performed to generate the final reconstruction. The cryo-EM and refinement statistics for each nanotube are listed in Table 5-1.

Parameter	CanA	Hyper2
-----------	------	--------

Data collection and processing		
Voltage (kV)	300	300
Electron exposure (e⁻ Å⁻²)	50	50
Pixel size (Å)	1.08	1.08
Particle images (n)	659,059	140,207
Shift (pixel)	13	24
Helical symmetry		
Point group	C1	C8
Helical rise (Å)	1.59	11.156
Helical twist (°)	-173.75	9.586
Map resolution (Å)		
Map:map FSC (0.143)	2.6	3.5
Model:map FSC (0.38)	2.7	3.55
d₉₉	2.9	3.2
Refinement and Model validation		
Ramachandran Favored (%)	94.9	83.33
Ramachandran Outliers (%)	0	0
RSCC	0.87	0.80
Clashscore	3.9	9.73
Bonds RMSD, length (Å)	0.002	0.007
Bonds RMSD, angles (°)	0.59	0.876
Deposition ID		
PDB (model)	7UII	Not yet deposited
EMDB (map)	EMD-26546	Not yet deposited

Table 5-1. Information about CanA and Hyper2 data collection and refinement.

Atomic force microscopy and quantitative nanomechanical mapping. Atomic Force Microscopy (AFM) with quantitative nanomechanical mapping (QNM) was used to measure the mechanical properties of CanA tubes assembled with calcium. For scanning in air, freshly cleaved mica was carbon-

coated with 20 μm carbon to improve adhesion. 5 μL assembled CanA or Hyper2 tubes in water were applied to the coated mica and incubated for 10 minutes, then carefully wicked from the side with filter paper. 30 μL of water was applied to the mica and then wicked and allowed to air dry. In liquid, 30 μL of 100 mM NiCl_2 was deposited onto freshly cleaved mica, then 10 μL high-concentration tubes (467 μM protein, 50 mM CaCl_2) were injected into the NiCl_2 drop. After 15 minutes the loose tubes were wicked away and the mica was rinsed 3x with HPLC water. Images were collected with a Bruker Multi-Mode 8 microscope (Bruker, Billerica, MA) with a Scan-Asyst Air tip (Bruker) in QNM in Air or Fluid mode.

Analytical ultracentrifugation. Oligomerization of CanA was monitored with and without calcium using a mass action analysis. Two 0.5 mL samples at 0.6 OD at 220 nm or 0.6 OD 295 nm were run as-is or mixed with 1 or 10 mM Ca^{2+} for the run at 20°C. Both Ca^{2+} and Tb^{3+} were used. Rotor speed was 45,000 rpm. Run time was 14 hours. Despite multiple attempts, no dimers or trimers were observed. Only 20 kDa particles, corresponding to the mass of single CanA proteins, were observed. Samples of CanA and Ca^{2+} were incubated at room temperature for two days but these also failed to assemble. Low and high concentration behaved identically. All samples with Tb^{3+} aggregated to very large particles which sedimented completely before the AUC run began. These samples could not be measured because these aggregates were too large to be detected at the speeds used in the run.

Discussion

Structural motifs

Both the CanA and Hyper2 proteins are primarily characterized by a classic β -barrel jelly-roll fold, with four long and four short antiparallel strands making a single barrel. CanA has a looped-out domain that creates the ridged surface visible in TEM micrographs. Both proteins also have a prominent hydrophobic free strand at the N terminus which participates in donor strand complementation (also known as donor strand exchange) in the β -barrel binding pocket. Both the Hyper2 and CanA filaments

rely on strand exchange, consistent with other extracellular nanotubes in archaea and bacteria, especially hyperstable filaments, for stability in the axial dimension. These structures demonstrate that donor strand complementation can be spontaneous and give rise to rigid assemblies, in contrast to previously known examples such as chaperone-usher pili. Despite the ubiquity of the motif, it is not predicted in AlphaFold Multimer (Evans et al., 2022). In fact, no dimers—mediated by strand exchange or any other interaction—were found in AUC (*vide supra*), so the dimer state may be very transient. In stability assays with heat and EDTA, damaged or malformed filaments of both proteins are primarily disrupted along the azimuthal axis, producing high-aspect ribbons (Figure 5.3). This is consistent with the axial dimension of the nanotube being better stabilized, ostensibly by strand exchange, than the azimuthal dimension.

To our knowledge, this is the first description of a rigid strand exchange-mediated self-assembly. The donor strand complementation motif here produces a very rigid, completely straight tube that forms spontaneously under the correct conditions, whereas previously described examples of the motif are flexible and often elastic (Pakharukova et al., 2022; Wang et al., 2022), and assembly is dependent on post-translational modifications and chaperone and usher proteins (Geibel and Waksman, 2014; Slonim et al., 1992; Thanassi et al., 1998; Werneburg and Thanassi, 2018). The donor strand complementation motif-mediated assembly is very common among archaea, especially hyperthermophiles, and may have been present in the common ancestor of bacteria and archaea. At 25.8% sequence identity and 39.8% sequence similarity, the two proteins are similar enough in sequence and structure to be a generic conserved protein but are probably not phylogenetically close. Both rigid and flexible structure alignments using FATCAT (Flexible structure Alignment by Chaining Aligned fragment pairs allowing Twists) (Li et al., 2020) yield an RMSD of 1.26 between the backbones of the two proteins (SI figure 4). Both RMSDs are calculated without twisting. Given how similar the protein structures are, the helical symmetries are surprisingly different. The metal ion binding sites may play a role in this discrepancy. The mix of differences and similarities between the two structures demonstrates the “deterministic chaos” inherent in generating structure from sequence. Judging from these two examples, the rigid donor strand complementation motif appears to be very robust in sequence space, i.e., it can withstand substantial

changes to its sequence. This observation invites study into consensus sequences, which may have to wait for more metagenomic data from these inaccessible archaea.

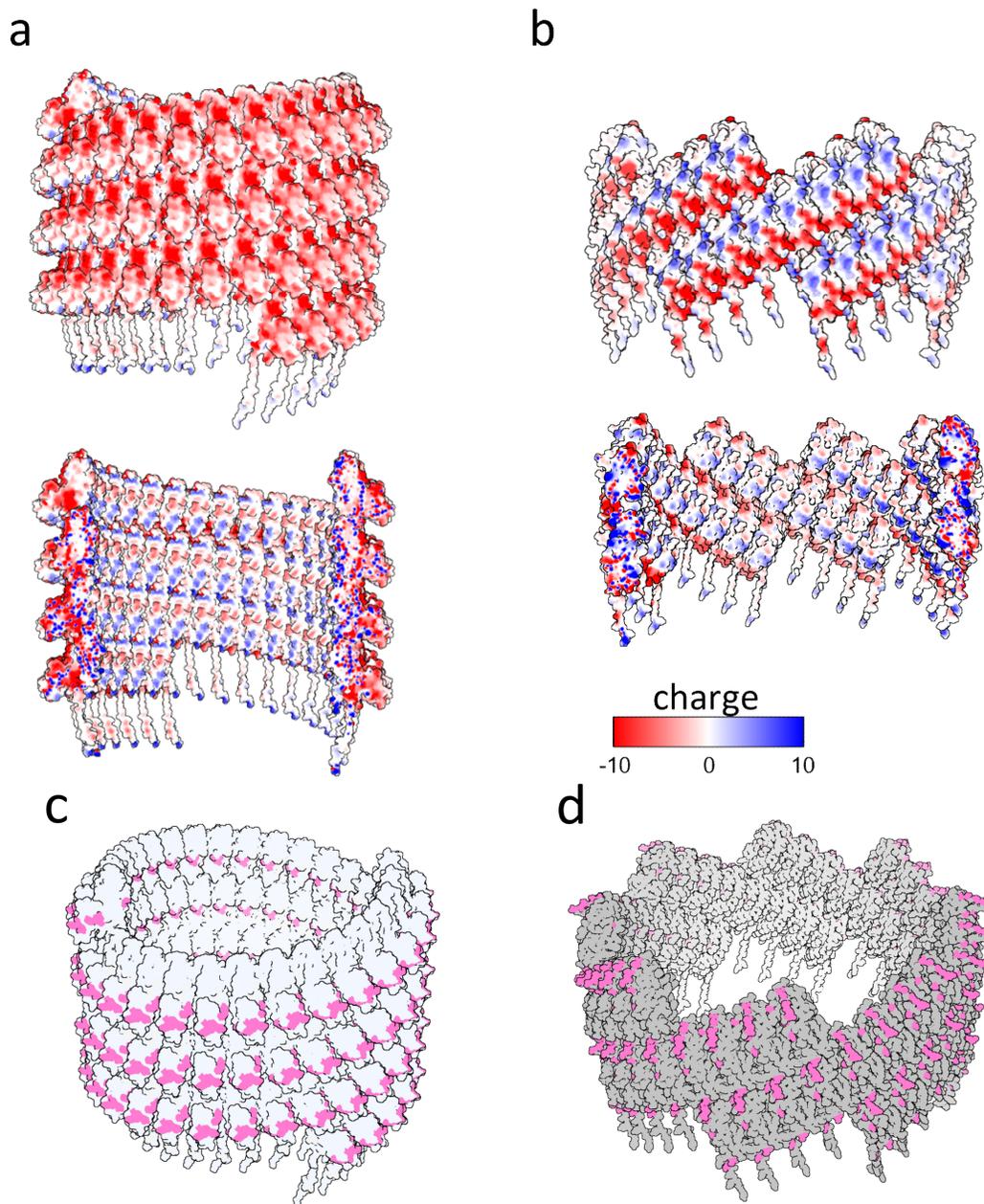


Figure 5.18. A,B) Surface charges of inner and outer surfaces of A) CanA and B) Hyper2 filaments. C,D) N-X-S/T glycosylation sequon shown in pink on surfaces of c) CanA and d) Hyper2 filaments.

The most notable feature of Hyper2 is the striking heterogeneity of the filaments as seen in micrographs. Because this protein sequence is recovered from a metagenomic screen and has no other biological data associated with it, we cannot know what conditions, if any, produce a homogenous

population of filaments. Furthermore, the sequence used here is only a part of the entire sequence extracted from the metagenome, and metagenomic sequences may not be 100% accurate, so the protein sequence itself may be different in the native context. Information about post-translational modifications, namely N-glycosylation which is widespread in archaea, are also not available. The natively expressed proteins are probably N-glycosylated like the cannulae of *Pyrodictium*, which could bias the assembly toward one or a few symmetries (Eichler, 2020; Jarrell et al., 2014; Nickell et al., 2003). N-glycosylation is very common in archaea and is thought to improve protein stability in extreme environments (Mengele and Sumper, 1992) (Eichler et al., 2013; Tamir and Eichler, 2017), in addition to other functions such as species recognition (van Wolferen et al., 2020). In our hands, the assembly appears to behave as a scrolled sheet that can heal to a tube at a few discrete symmetries. The multi-layered appearance may be due to complementary charged patches between the inner and outer surfaces (Figure 5.18b); stabilizing effects probably bias the nascent tube toward multilayers. The heterogeneity and multilayering are consistent among tubes formed by calcium, magnesium, or heat alone.

Role of ion or heat

Both these proteins coordinate Ca^{2+} ions, but the role of the ion differs. For CanA, there are two Ca^{2+} ions per protein: One coordinated by residues Q39, Q54, N56, and G183, and the other by residues G37, 161E, 164E, 182V (Figure 17). These ions share a binding site on the azimuthal interface. CanA assembly is completely dependent on the ion, whether Ca^{2+} , Tb^{3+} or Zn^{2+} . The ion plays a role in thermal stability of the assembly: Tb^{3+} -mediated tubes appear shredded after heating in EDTA, but Ca^{2+} -mediated tubes dissolve completely. Without EDTA, Ca^{2+} -mediated tubes are very damaged, but Tb^{3+} -mediated tubes remain entirely intact. Disruption of the nanotube occurs primarily on the azimuthal axis as seen in Figure 5.3. Heat and EDTA effects on CanA tubes. A) CanA tubes made with Ca^{2+} incubated at 60°C in excess EDTA. Tubes are completely dissolved; solution was optically clear. B) CanA tubes made with Tb^{3+} incubated at 60°C in excess EDTA. Tubes are badly damaged, especially along the azimuthal dimension. Solution is optically clear. C) CanA tubes incubated at 100°C for 30 minutes without EDTA. Tubes are badly damaged, especially along the azimuthal dimension. Solution is optically clear. D) CanA

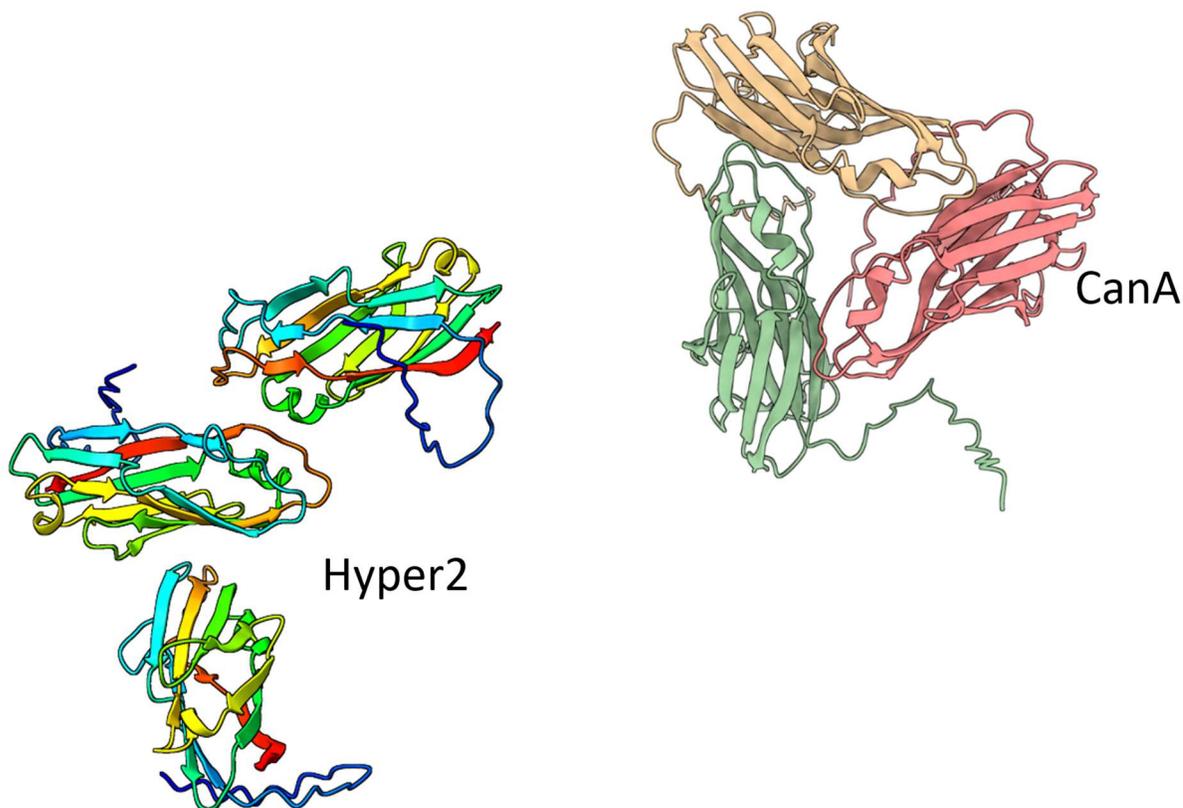
tubes made with Tb^{3+} incubated at $100^{\circ}C$ for 30 minutes without EDTA. Tubes appear unaffected except for debris attached to some tubes, but not all. Figure 5.3. This is consistent with the model showing stabilizing interactions (strand exchange and divalent ion coordination) stabilizing helical axial interactions.

Calmodulin Ca^{2+} binding sites will readily bind other metal ions including Zn^{2+} and Tb^{3+} (Wang et al., 1984; Warren et al., 2007). CanA formed tubes with both of these, but not Mg^{2+} , although the patent described mixing the protein with both Ca^{2+} and Mg^{2+} together. Tubes form overnight when CanA is mixed with Ca^{2+} ; assembly occurs within an hour when heated. Tb^{3+} apparently binds so tightly that the mixture becomes turbid upon mixing at room temperature. In AUC experiments conducted at very low CanA concentration, no single dimers were found for either Ca^{2+} or Tb^{3+} mixtures because the protein concentration was too low for Ca^{2+} to assemble tubes, but far above whatever equilibrium concentration would produce a CanA dimer from Tb^{3+} . Zn^{2+} did rapidly induce assembly at room temperature, but the tubes were noticeably shorter than those formed by Ca^{2+} or Tb^{3+} . Future spectroscopic studies (e.g., FTIR) may demonstrate a difference in the coordination of Tb^{3+} versus Ca^{2+} . For now, we can assume that the location, if not necessarily the residues or stoichiometry, of Tb^{3+} is the same because it induces the same assembly (Edington et al., 2018). In contrast, Hyper2 assembly can be mediated by Ca^{2+} and Mg^{2+} but not Tb^{3+} or Zn^{2+} , which induce nonspecific aggregation. There are probably charged residues on the surface of Hyper2 proteins that coordinate Tb^{3+} or Zn^{2+} without inducing a regular assembly.

The role of heat in this assembly mechanism probably comes down to simple diffusion for CanA, but not Hyper2. High-concentration Hyper2 protein is soluble for several weeks at room temperature but will assemble within minutes when heated to $60^{\circ}C$, even in EDTA (Figure 5.11). Hyper2 assembles fastest when mixed with Ca^{2+} and heated. The Ca^{2+} ion does coordinate adjacent proteins, and at room temperature may serve to pre-organize a larger strand-exchange surface, making the axial interaction more favorable.

Outlook

That CanA and Hyper2 can assemble spontaneously in vitro, without their native post-translational modifications, is an exciting new development for protein designers. These proteins express well in *E. coli* and can be left at room temperature for weeks without degrading, making them especially easy to work with. Self-assembling proteins have already been investigated for applications such as drug and vaccine delivery and as protein organization tools in the laboratory. Once assembled, they may be able to confer thermostability to proteins bound to their surfaces by increasing rigidity at high temperatures. The lumens of these nanotubes may be especially protective to sensitive drug molecules bound inside. As we have seen from these two examples, there may be high sequence variability in related proteins, especially on residues facing the lumen. AlphaFold accurately predicts the jelly-roll structure, but the N-terminal domain is predicted, with low confidence, to be flipped toward the beta-barrel; multimers are predicted to be symmetrical homomultimers without DSC (



SI Figure 5-9). Therefore designers should use this tool with caution.

Metagenomic research can be a rich source of candidate proteins, with the caveat that any information not encoded in DNA is lost. PTM clues such as cysteines and likely glycosylation sites can generate hypotheses; the other proteins in the operon may be necessary for assembly. This structural study cannot answer questions of cannula secretion or their role in the cell. However, other studies indicate that hyperthermophilic archaeal cannulae constitute a biofilm that anchors the archaea to the hot surface of the vent and protects them from the turbulent water and gases (König et al., 1988; Nickell et al., 2003; Rieger et al., 1995; van Wolferen et al., 2018; van Wolferen et al., 2020).

References

- Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A., Adams, P.D., 2018. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* 74, 531-544.
- Busch, A., Waksman, G., 2012. Chaperone–usher pathways: diversity and pilus assembly mechanism. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367, 1112-1122.
- Edington, S.C., Gonzalez, A., Middendorf, T.R., Halling, D.B., Aldrich, R.W., Baiz, C.R., 2018. Coordination to lanthanide ions distorts binding site conformation in calmodulin. *Proceedings of the National Academy of Sciences* 115, E3126-E3134.
- Eichler, J., 2020. N-glycosylation in Archaea—New roles for an ancient posttranslational modification. *Molecular Microbiology* 114, 735-741.
- Eichler, J., Arbiv, A., Cohen-Rosenzweig, C., Kaminski, L., Kandiba, L., Konrad, Z., 2013. N-glycosylation in *Haloferax volcanii*: adjusting the sweetness. *Front Microbiol* 4, 403.
- Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-2132.
- Evans, R., O’Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Židek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie,

- A., Tunyasuvunakool, K., Jain, R., Clancy, E., Kohli, P., Jumper, J., Hassabis, D., 2022. Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2021.2010.2004.463034.
- Geibel, S., Waksman, G., 2014. The molecular dissection of the chaperone-usher pathway. *Biochim Biophys Acta* 1843, 1559-1567.
- Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., Ferrin, T.E., 2018. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* 27, 14-25.
- Hernández, G., Jenney, F.E., Adams, M.W.W., LeMaster, D.M., 2000. Millisecond time scale conformational flexibility in a hyperthermophile protein at ambient temperature. *Proceedings of the National Academy of Sciences* 97, 3166-3170.
- Horn, C., Paulmann, B., Kerlen, G., Junker, N., Huber, H., 1999. In Vivo Observation of Cell Division of Anaerobic Hyperthermophiles by Using a High-Intensity Dark-Field Microscope. *Journal of Bacteriology* 181, 5114-5118.
- Jarrell, K.F., Ding, Y., Meyer, B.H., Albers, S.-V., Kaminski, L., Eichler, J., 2014. N-linked glycosylation in Archaea: a structural, functional, and genetic analysis. *Microbiology and molecular biology reviews* : MMBR 78, 304-341.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583-589.
- König, H., Messner, P., Stetter, K.O., 1988. The fine structure of the fibers of *Pyrodictium occultum*. *FEMS Microbiology Letters* 49, 207-212.
- Kreitner, R., Munte, C.E., Singer, K., Stetter, K.O., Horn, G., Kremer, W., Kalbitzer, H.R., 2020. Complete sequential assignment and secondary structure prediction of the cannulae forming

- protein CanA from the hyperthermophilic archaeon *Pyrodictium abyssi*. *Biomolecular NMR Assignments* 14, 141-146.
- Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* 372, 774-797.
- Li, Z., Jaroszewski, L., Iyer, M., Sedova, M., Godzik, A., 2020. FATCAT 2.0: towards a better understanding of the structural diversity of proteins. *Nucleic Acids Research* 48, W60-W64.
- Mengele, R., Sumper, M., 1992. Drastic differences in glycosylation of related S-layer glycoproteins from moderate and extreme halophiles. *J Biol Chem* 267, 8182-8185.
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M., 2022. ColabFold: making protein folding accessible to all. *Nature Methods* 19, 679-682.
- Morris, E.P., da Fonseca, P.C.A., 2017. High-resolution cryo-EM proteasome structures in drug development. *Acta Crystallogr D Struct Biol* 73, 522-533.
- Nelson R. Barton, E.O.D., Ryan Short, Gerhard Frey, David Weiner, Dan E. Robertson, Steven Briggs, Paul Zorner. 2008. CHIMERIC CANNULAE PROTEINS, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AND USING THEM, US.
- Nickell, S., Hegerl, R., Baumeister, W., Rachel, R., 2003. *Pyrodictium cannulae* enter the periplasmic space but do not enter the cytoplasm, as revealed by cryo-electron tomography. *Journal of Structural Biology* 141, 34-42.
- Pakharukova, N., Malmi, H., Tuittila, M., Dahlberg, T., Ghosal, D., Chang, Y.-W., Myint, S.L., Paavilainen, S., Knight, S.D., Lamminmäki, U., Uhlin, B.E., Andersson, M., Jensen, G., Zavialov, A.V., 2022. Archaic chaperone–usher pili self-secrete into superelastic zigzag springs. *Nature* 609, 335-340.
- Punjani, A., Rubinstein, J.L., Fleet, D.J., Brubaker, M.A., 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* 14, 290-296.
- Rieger, G., Rachel, R., Hermann, R., Stetter, K.O., 1995. Ultrastructure of the Hyperthermophilic Archaeon *Pyrodictium abyssi*. *Journal of Structural Biology* 115, 78-87.

- Slonim, L.N., Pinkner, J.S., Brändén, C.I., Hultgren, S.J., 1992. Interactive surface in the PapD chaperone cleft is conserved in pilus chaperone superfamily and essential in subunit recognition and assembly. *Embo j* 11, 4747-4756.
- Stetter, K.O., König, H., Stackebrandt, E., 1983. *Pyrodictium* gen. nov., a New Genus of Submarine Disc-Shaped Sulphur Reducing Archaeobacteria Growing Optimally at 105°C. *Syst Appl Microbiol* 4, 535-551.
- Tamir, A., Eichler, J., 2017. N-Glycosylation Is Important for Proper *Haloferax volcanii* S-Layer Stability and Function. *Appl Environ Microbiol* 83.
- Thanassi, D.G., Saulino, E.T., Hultgren, S.J., 1998. The chaperone/usher pathway: a major terminal branch of the general secretory pathway. *Current Opinion in Microbiology* 1, 223-231.
- Trembath-Reichert, E., Butterfield, D.A., Huber, J.A., 2019. Active seafloor microbial communities from Mariana back-arc venting fluids share metabolic strategies across different thermal niches and taxa. *The ISME Journal* 13, 2264-2279.
- van Wolferen, M., Orell, A., Albers, S.-V., 2018. Archaeal biofilm formation. *Nature Reviews Microbiology* 16, 699-713.
- van Wolferen, M., Shajahan, A., Heinrich, K., Brenzinger, S., Black, I.M., Wagner, A., Briegel, A., Azadi, P., Albers, S.V., 2020. Species-Specific Recognition of Sulfolobales Mediated by UV-Inducible Pili and S-Layer Glycosylation Patterns. *mBio* 11.
- Wang, C.L.A., Leavis, P.C., Gergely, J., 1984. Kinetic studies show that calcium and terbium have different binding preferences toward the four calcium-binding sites of calmodulin. *Biochemistry* 23, 6410-6415.
- Wang, F., Cvirkaite-Krupovic, V., Krupovic, M., Egelman, E.H., 2022. Archaeal bundling pili of *Pyrobaculum calidifontis* reveal similarities between archaeal and bacterial biofilms. *Proc Natl Acad Sci U S A* 119, e2207037119.

Warren, J.T., Guo, Q., Tang, W.-J., 2007. A 1.3-Å Structure of Zinc-bound N-terminal Domain of Calmodulin Elucidates Potential Early Ion-binding Step. *Journal of Molecular Biology* 374, 517-527.

Werneburg, G.T., Thanassi, D.G., 2018. Pili Assembled by the Chaperone/Usher Pathway in *Escherichia coli* and *Salmonella*. *EcoSal Plus* 8.

Supplemental Information

CanA:

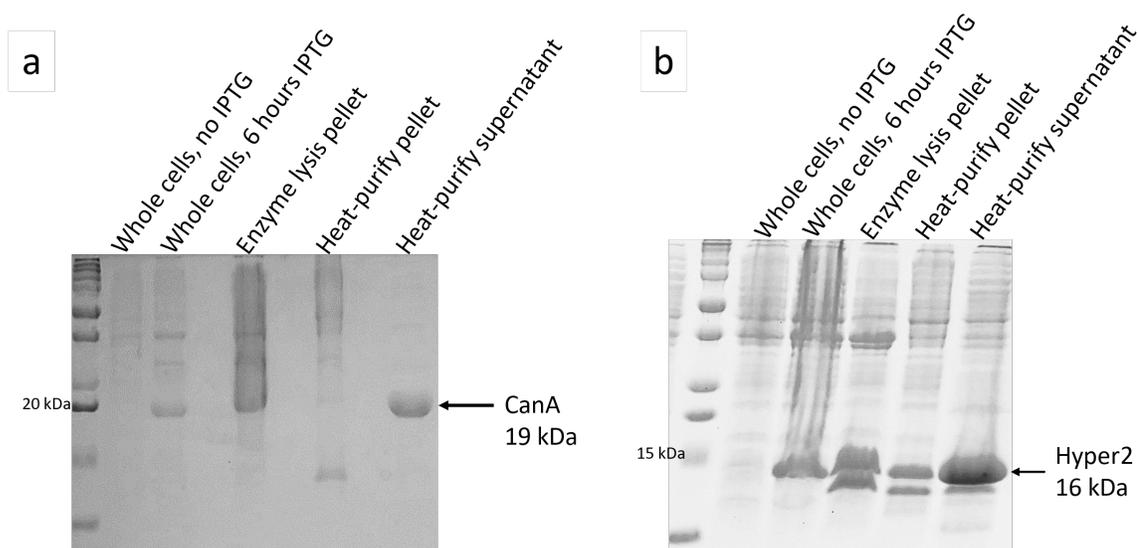
TTQSPLNSFYMTTQSPLNSFYATGTAQAVSEPIDVESHLSITPAAGAQQSDDIGYAIWIKDQVNDVKLKVTL
 ANAEQLKPYFKYLQIQITSGYETNSTALGNFSETKAVISLDNPSAVIVLDKEDIAVLYPDKTGYTNTSIWVPGE
 PDKIIVYNETKPVAILNFKAFYEAKEGMLFDSLPIVFNFQVLQVG

SI Figure 5-1. Sequence of CanA used in this work. The first 10 N-terminal residues, highlighted in red, were omitted from the sequence to improve solubility.

Hyper2:

MSVTSDRTYYGYGDVSVKNEPVMVVVSPFSFPGANASLSSGGQGVFKKPDWIRVVNQSDVENVKLEIDWV
 NANQAANYFDYARILVTGPNQVKGYSLSLQHGKAWITLDAEELREGAVLGAVMYEYEVKEGVLASRLPLVFKVRVVE
 TGSI

SI Figure 5-2. Sequence of Hyper2 used in this work, with signal sequence omitted.



SI Figure 5-3. SDS-PAGE gels from CanA and Hyper2 expression and purification. A. 19 kDa CanA is not expressed without IPTG, but is visible after 6 hours IPTG-induced expression. Neither the pellet from the enzyme lysis with lysozyme and Benzonase nor the pellet from incubating the enzyme supernatant at 80°C contain CanA. Supernatant from the heating step contains CanA at high concentration and purity. B. Hyper2 expresses well under IPTG expression. Some Hyper2 partitions to the enzyme lysis pellet fraction, but most remains soluble. Some Hyper2 also crashes out of the heat

Secondary structure	Amino acid	Helix	Strand	Coil
L1	11–21			×
E1	22–31		×	
L2	32–43			×
E2	44–50		×	
L3	51–55			×
E3	56–63		×	
L4	64–65			×
H1	66–72	×		
L5	73–74			×
E4	75–82		×	
L6	83–94			×
E5	95–102		×	
L7	103–106			×
E6	107–112		×	
L8	113			×
H2	114–116	×		
L9	117–130			×
E7	131–133		×	
L10	134–140			×
E8	141–144		×	
L11	145–148			×
E9	149–160		×	
L12	161–172			×
E10	173–181		×	
L13	182			×

SI Figure 5-4. The secondary structure propensities were obtained from the chemical shifts of C', C^α, C^β, N, H^N and H^α atoms by the program TALOS-N (Shen and Bax [2013](https://doi.org/10.1007/s12104-020-09934-x)). *Li* coil regions, *Ei* β-strand regions, *Hi* helix. Reused under Creative Commons CC BY license from Kreitner, R., Munte, C.E., Singer, K. *et al.* Complete sequential assignment and secondary structure prediction of the cannulae forming protein CanA from the hyperthermophilic archaeon *Pyrodicticum abyssi*. *Biomol NMR Assign* **14**, 141–146 (2020). <https://doi.org/10.1007/s12104-020-09934-x>.

Alignment:

Identity: 48/186 (25.8%)
 # Similarity: 74/186 (39.8%)
 # Gaps: 56/186 (30.1%)

```

CanA__ 9 SFYATGTAQAVSEPIDVESHLSITPAA--GAQGSDDIGYAIV-----WI 51
  :|..|.....:|::| .::|. : ||..|...|...| ||
Hyper2 8 TYYGYGDVSVKNEPQNV-----VWSPFSFPGANASLSSGGQGVFKKPDWI 52

CanA__ 52 K----DQVNDVKLKVTLANAEQLKPYFKYLQIQITSGYETNSTALGNFSE 97
  : ..|.::|||:.....||..|...||..|.:|. : |....:
Hyper2 53 RVVNQSDVENVKLEIDWVNNANQAANYFDYARILVT-----GPNGQ 92

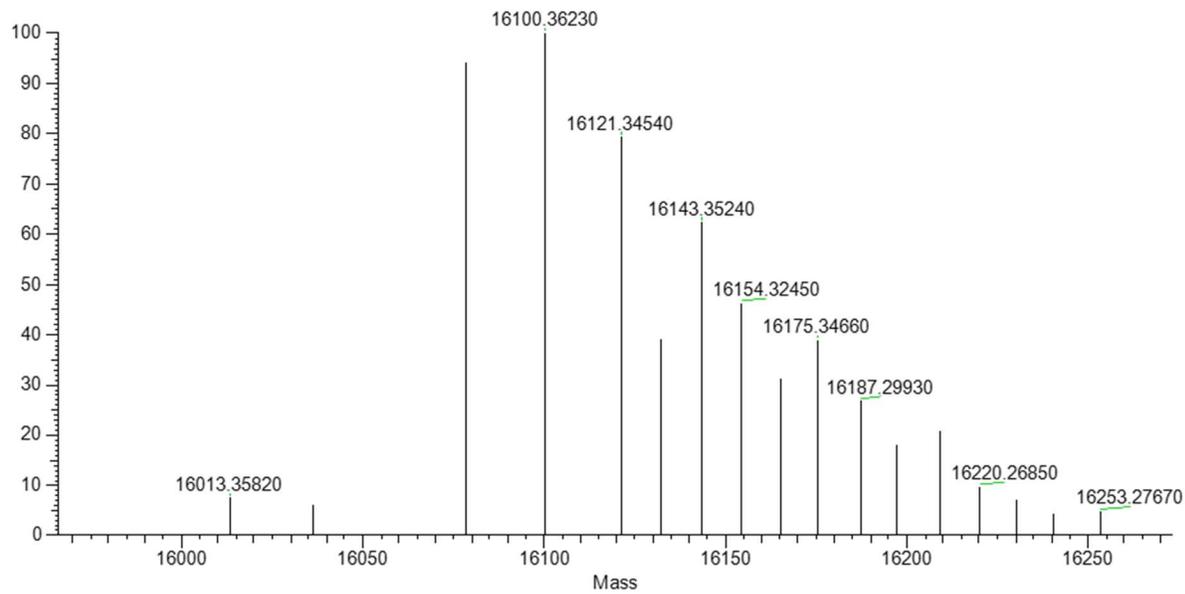
CanA__ 98 TKAVISLDNPSAVIVLDKEDIAVLYPKDTGYTNTSIWVPGEPDKIIVYNE 147
  .|.:||:..|.|.||. | |
Hyper2 93 VKGYLSLQHGKAWITLDAE-----E 112

CanA__ 148 TKPVAILNFKAFYEAKEGMLFD SLPVIFNFQVLQVG 183
  .:..|:|.....:|.|::|...|::|...:|:..:|:..:|
Hyper2 113 LREGAVLGAVMYEYVKEGVLASRLPLVFKVRVETG 148
  
```

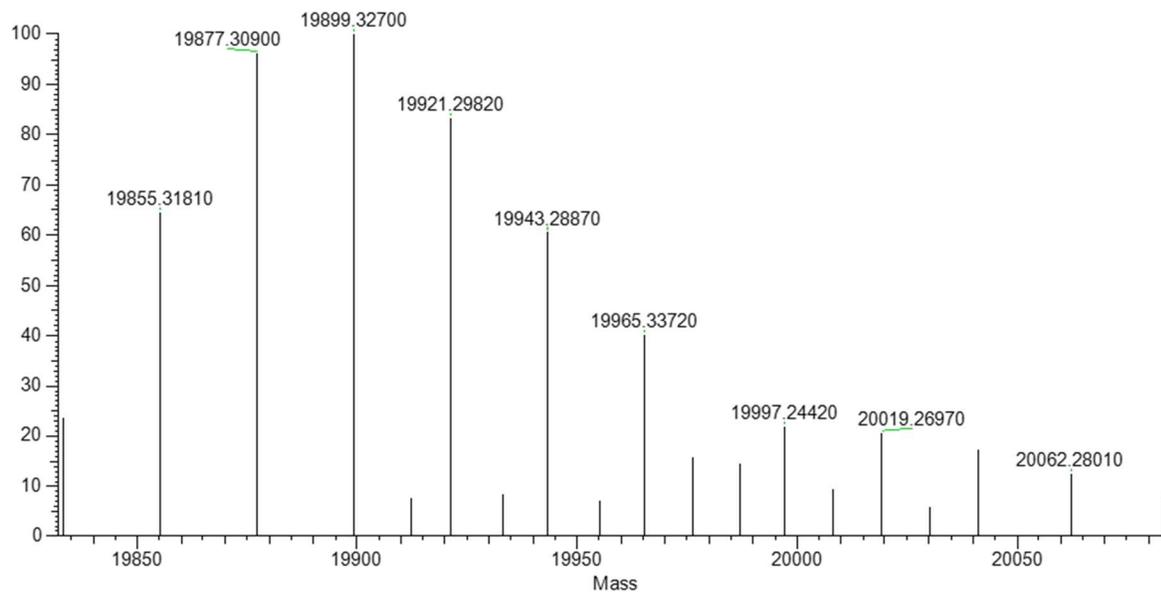
SI Figure 5-5. ClustalW alignment of CanA and Hyper2. Residues involved in metal ion binding are highlighted.



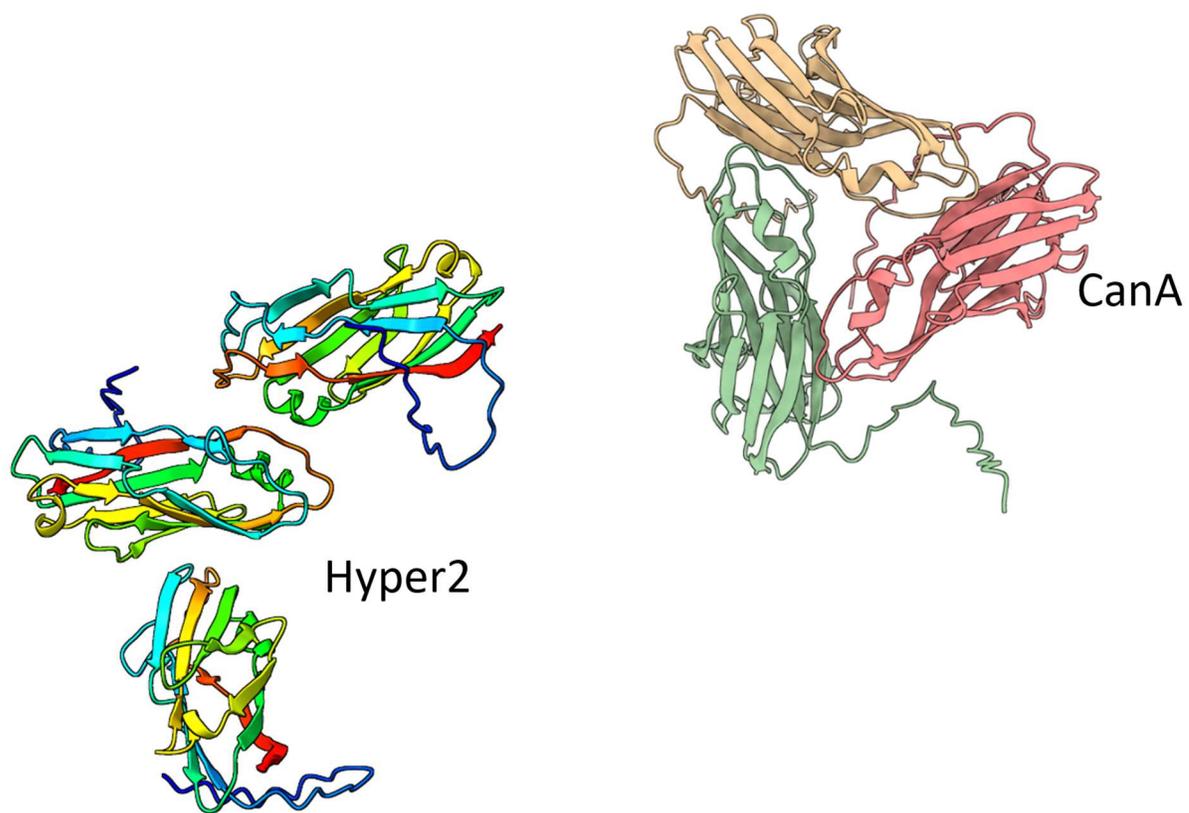
SI Figure 5-6. FATCAT rigid alignment of single Hyper2 (blue) and CanA (pink) proteins.



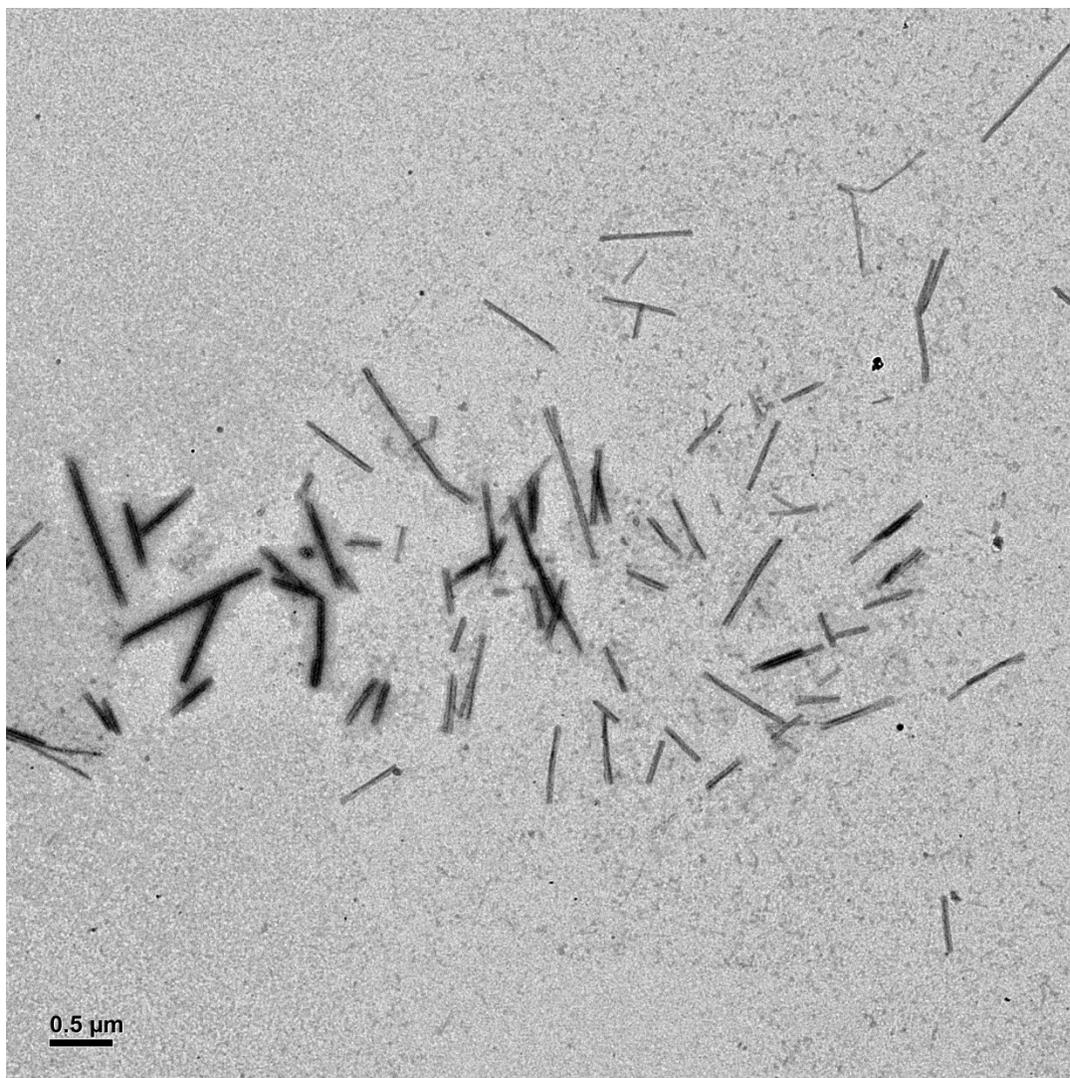
SI Figure 5-7. Hyper2 ESI-Mass spec readout confirming expected mass at 16.1 kDa.



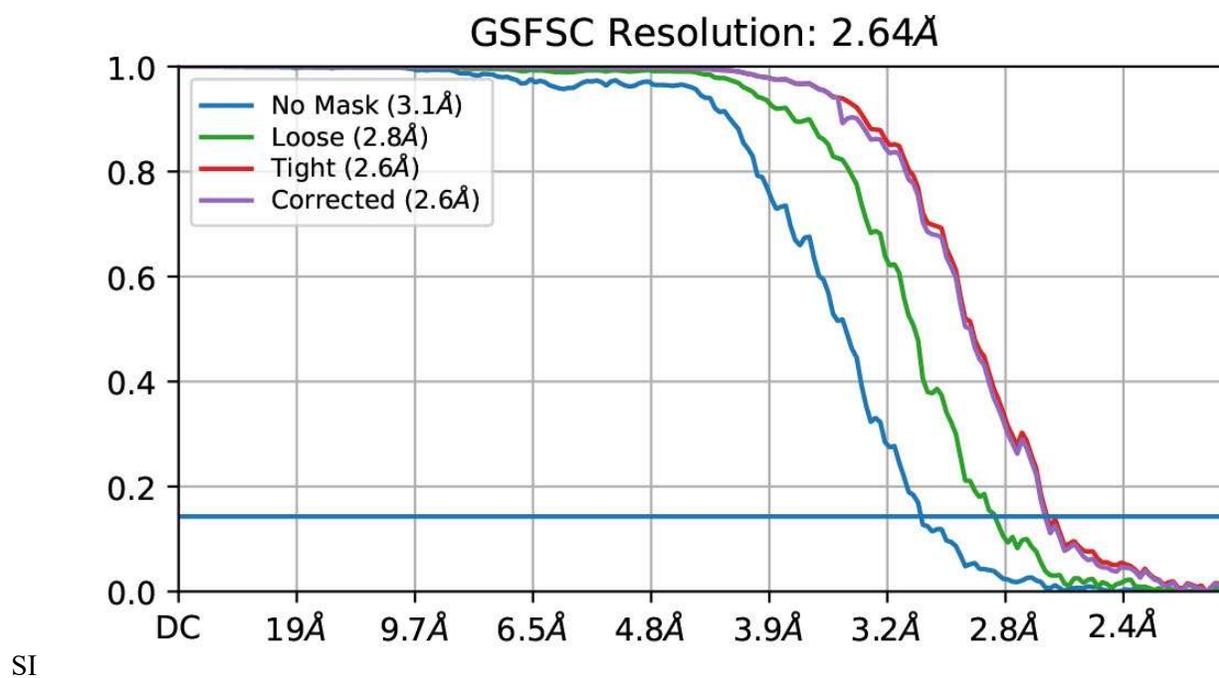
SI Figure 5-8. CanA ESI-Mass spec readout confirming expected mass at 19.9 kDa.



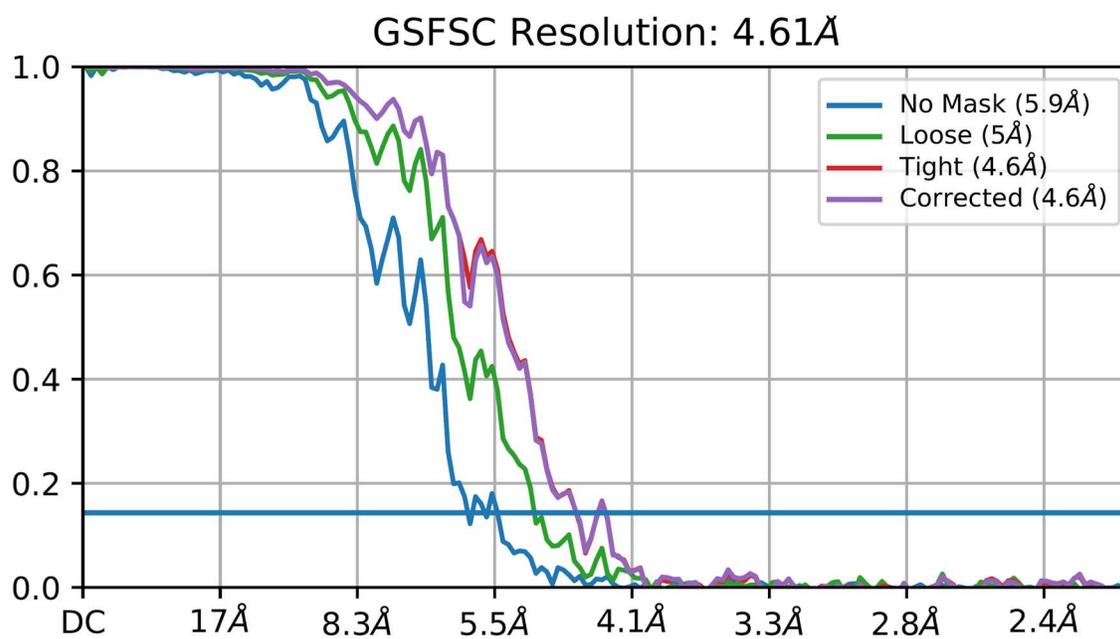
SI Figure 5-9. AlphaFold Multimer predictions for Hyper2 and CanA trimers.



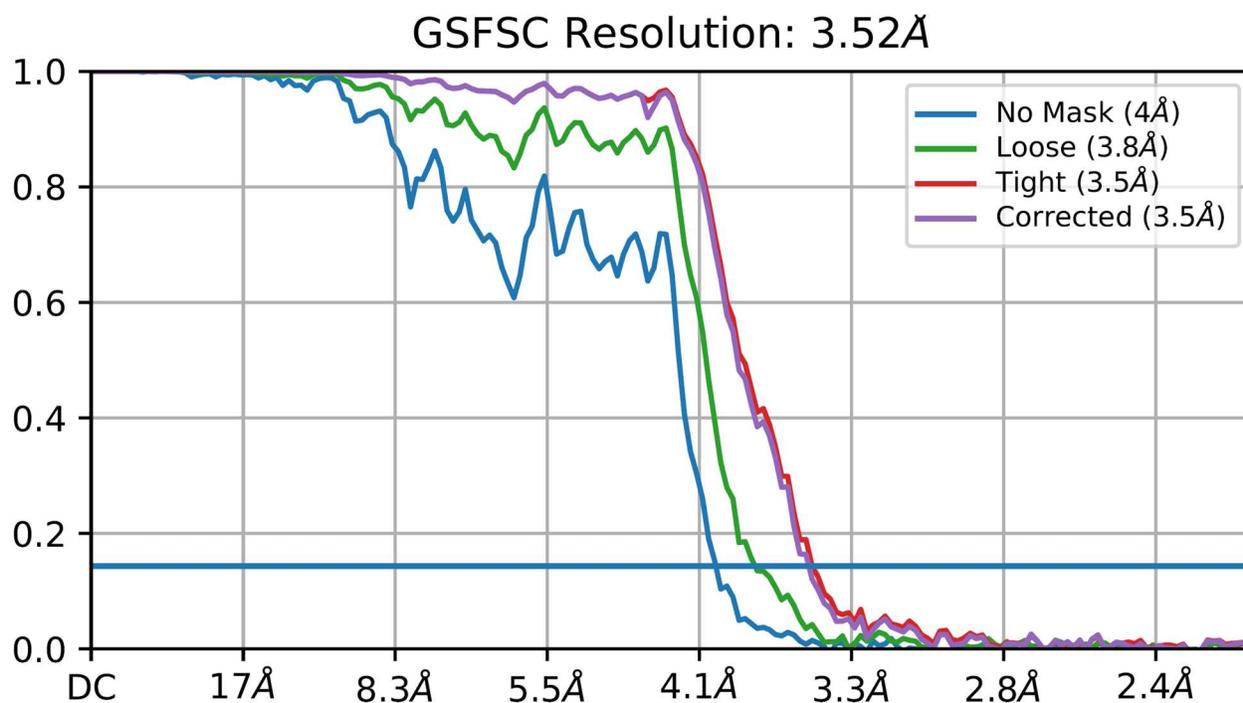
SI Figure 5-10. 40 mM CaCl_2 and 5 μM CanA have been demonstrated to produce nanotubes, though they are not as easy to confirm via TEM simply because they are few and far between.



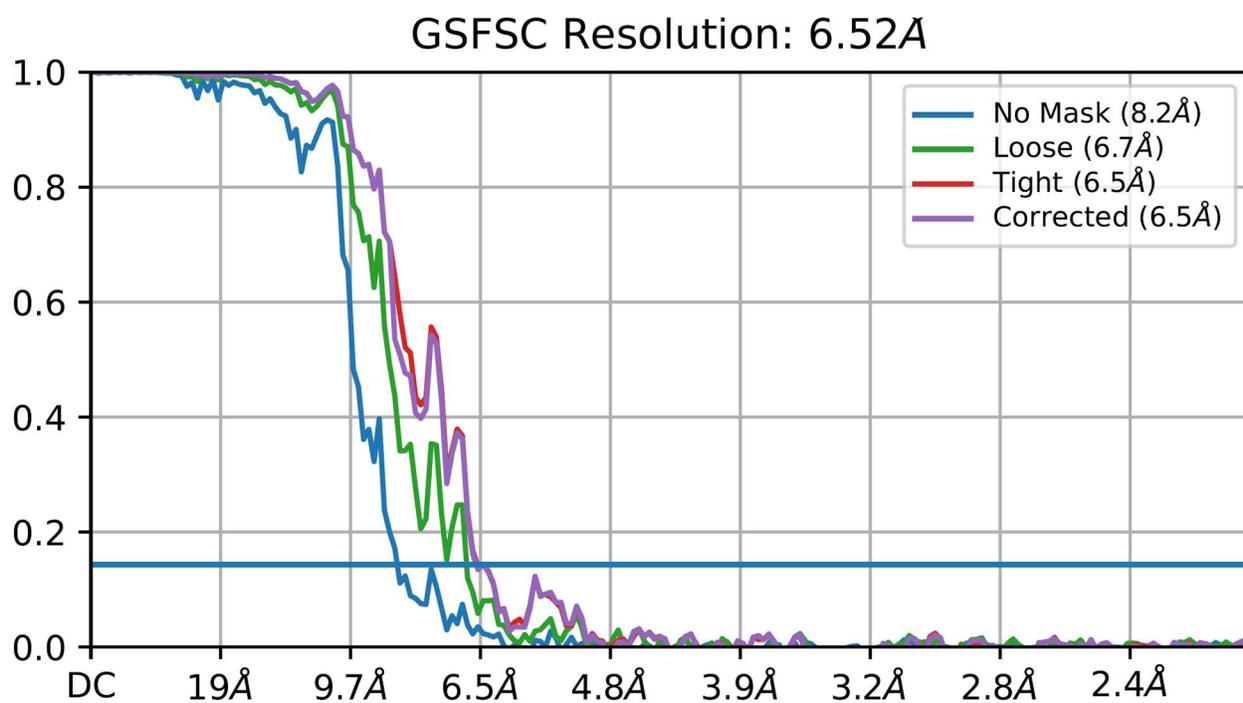
SI Figure 5-11. CanA FSC.



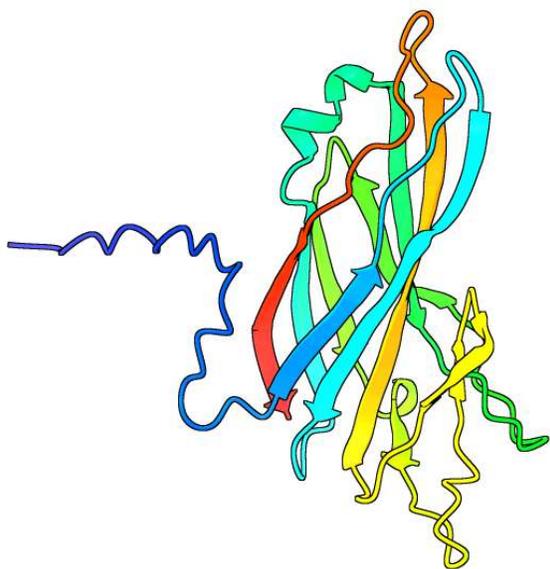
SI Figure 5-12. Hyper2 C7 FSC.



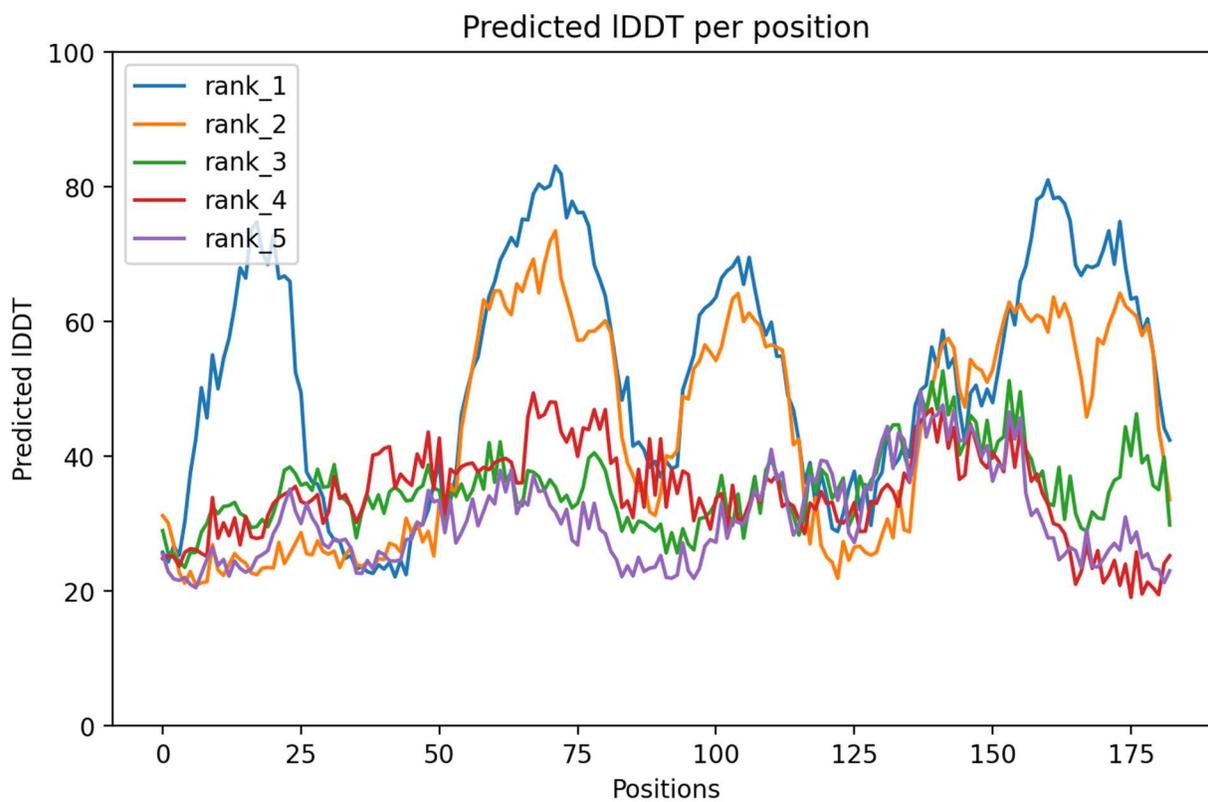
SI Figure 5-13. Hyper2 C8 FSC.



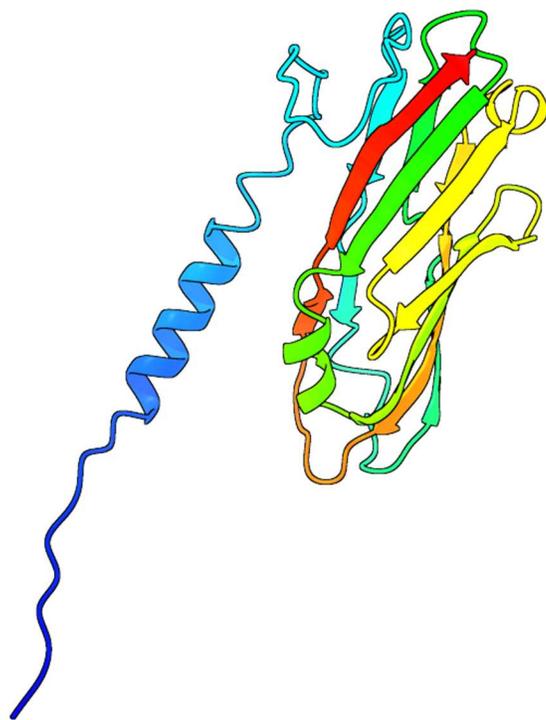
SI Figure 5-14. Hyper2 C9 FSC.



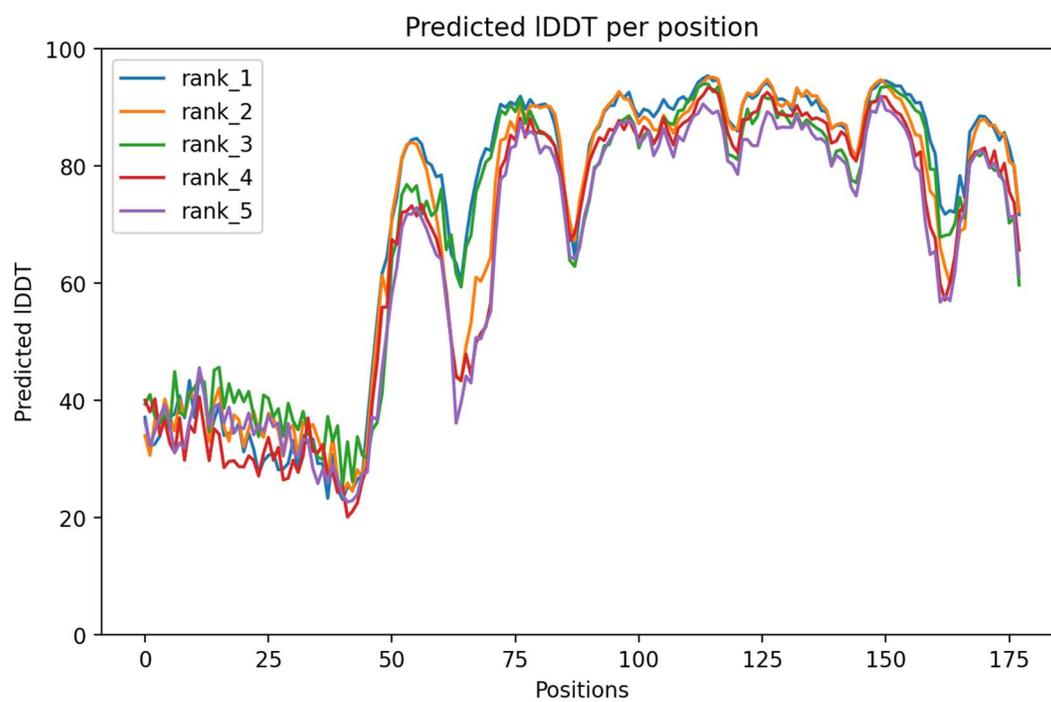
SI Figure 5-15. AlphaFold2.0 prediction for CanA, colored by residue.



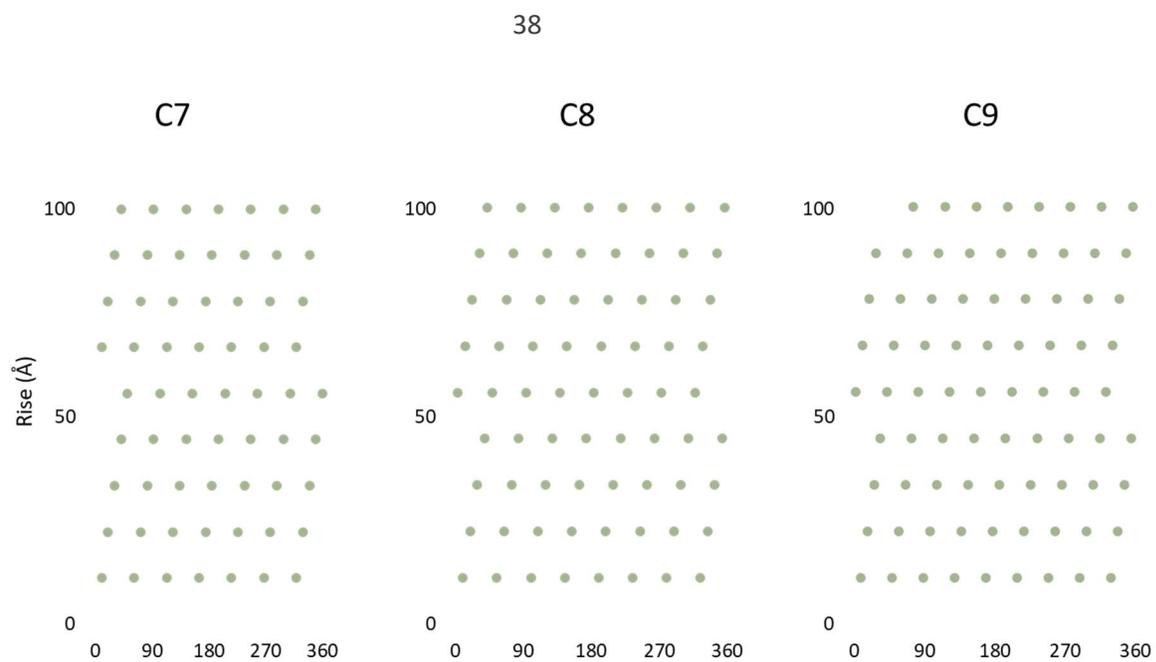
SI Figure 5-16. pLDDT score for single CanA protein prediction in AlphaFold.



SI Figure 5-17. AlphaFold2.0 prediction for Hyper2, colored by residue.



SI Figure 5-18. pLDDT score for single Hyper2 protein prediction in AlphaFold.



SI Figure 5-19. Helical net maps for the three best classes of Hyper2.