De Novo Design of Self-assembling Helical Protein Filaments

Hao Shen

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

David Baker, Chair

Jesse Zalatan

Justin Kollman

Frank DiMaio

Program Authorized to Offer Degree:

Molecular Engineering

Hao Shen

University of Washington

Abstract

De Novo Design of Self-assembling Helical Protein Filaments

Hao Shen

Chair of the Supervisory Committee:

David Baker

Department of Biochemistry

There has been some success in designing stable peptide filaments; however, mimicking the controllable and reversible assembly of many natural protein filaments is challenging. We devised a general computational approach to designing self-assembling helical filaments from monomeric proteins and use this approach to design proteins that assemble into micrometer-scale filaments with a wide range of geometries in vivo and in vitro. Cryo-electron microscopy structures of six designs are close to the computational design models. The filament diameter can be tuned by varying the number of repeats in the monomer. Anchor and capping units, built from monomers that lack an interaction interface, can be used to control assembly and disassembly. The filaments provide new phenomena through interactions with cells. We also extended the building blocks to multi-component, pH-responsive and potentially more functional proteins to design more controllable filaments. The ability to generate dynamic, highly ordered structures that span micrometers from protein monomers opens up possibilities for the fabrication of new multiscale metamaterials.

Table of Contents

Acknowledgements

I've been so fortunate to be at this place for my PhD study. Computational protein design is such an exciting field which brings computational modeling, molecular biology, biochemistry and biophysics from electron microscopy to fluorescent microscopy all together. The process of designing fine protein interactions, producing the actual protein and observing the amazing detail and dynamic behavior has been so intriguing. I remember the excitement when I saw my first designed filaments, first filament structure, first time seeing control of the assembly...all of this is enabled by the great people.

I would first like to thank my advisor David Baker, for his vision and effort to pioneer in this field, bring in all the talents, cultivate a great lab culture and guide me throughout the PhD journey. David is a real model in always being positive, creative, supportive, valuing communications and focusing on the important things.

I couldn't thank Jorge Fallas enough for being such a great postdoc mentoring me throughout my first project and keep providing excellent advice. He taught me the way to do science and think creatively.

I would like to thank all my great collaborators. It's been so much inspiring and fun to work with them and I've learned so much from them. Eric Lynch has spent so much time with me, collecting data in the cold cryoEM suite (in the old days), advising me on cryo and solves all my designed filament structures within the snap of a finger. William Sheffler built the basis for my computational protocol and inspired me a lot about programing and spatial imagination. I've been and will continue to learn from and enjoy working with Justin Decarreau, Michael Wagenbach, Juan Jesus Vicente and Ajit Joglekar on fluorescent imaging. Thank Joel Quispe for teaching me how to do cryoEM and everyone in the EM community.

Thank Bradley Parry, Nicholas Jannetty, Jiajun Chen, Lei Wang, Quinton Dowling, Gustav Oberdorfer, Lance Stewart, Linda Wordeman, James De Yoreo, Christine Jacobs-Wagner, Justin Kollman, TJ Brunette, Fabio Parmeggiani, Possu Huang, Zibo Chen, Scott Boyken, Issa Yousif and Jorgen Nelson for their contribution to my project.

I would also like to thank my wonderful colleagues and friends in the Baker Lab: Peilong Lu, Hua Bai, Zibo Chen, Jiayi Dou, Hanlun Jiang, Longxing Cao, Chunfu Xu, Qian Cong, Lei Shi, Yuan Liu, Yang Hsia, George Ueda, Una Nattermann, Chris Bahl, Sergey Ovchinnikov, Cassie Bryan, Ariel Ben-Sasson, Alexis Courbet, Sherry Bermeo, Marc Lajoie, Gabe Butterfield, Yu-Ru Lin, Brian Coventry, Cameron Chow, Lauren Carter, Michelle Matsunaga, Ruud van Deursen...Thank you all for your help and the great memories.

I would like to additionally thank my committee members, Ning Zheng, Justin Kollman and Frank DiMaio for their guidance and help.

I’m very grateful to be in the Molecular Engineering PhD program. Paul Neubert, along with Corin Shelley-Reuss, Renske Dyedov and Amanda McCracken, has been my strong support. I would also like to thank my molecular engineering cohorts for all the moments we've shared.

Special thanks to my fiancée Simin Wang. Our love story started the same time I started in the program, crossed the Pacific Ocean and is moving to the next step in our life. Thanks for her love and accompany.

Thank you Mom and Dad for bringing me to this wonderful world and care about me with all your love. Thanks to my family for all the support.

Chapter 1: *De novo* design of single component self-assembling helical filaments

Abstract

We describe a general computational approach to designing self-assembling helical filaments from monomeric proteins and use this approach to design proteins that assemble into micrometer-scale filaments with a wide range of geometries in vivo and in vitro. Cryo–electron microscopy structures of six designs are close to the computational design models. The filament building blocks are idealized repeat proteins, and thus the diameter of the filaments can be systematically tuned by varying the number of repeat units. The assembly and disassembly of the filaments can be controlled by engineered anchor and capping units built from monomers lacking one of the interaction surfaces. The ability to generate dynamic, highly ordered structures that span micrometers from protein monomers opens up possibilities for the fabrication of new multiscale metamaterials.

Introduction

Natural protein filaments differ considerably in their dynamic properties: some, like collagen, are relatively static with turnover rates in order of several weeks (*1–4*), while others, like cytoskeletal polymers, are dynamic--growing or disassembling in response to changing physiological conditions (*3*, *5–7*). The fraction of the total residue-residue interactions in the filament that are within (rather than between) the monomeric building blocks is generally higher for dynamic polymers; the monomers are usually independently folded structures rather than relatively extended polypeptides (Fig. 1A). While there has been success designing peptide filaments in the first class by staggering extended interaction motifs and generating end to end interactions between peptide coiled coils (*8–14*), the accurate computational design of reversibly assembling filaments built from folded protein monomers is an as yet unsolved challenge. Much of the progress in recent years in computational design of self-assembling nanomaterials has relied on building blocks with internal symmetry, which allows the generation of architectures with tetrahedral, octahedral and icosahedral point group (*15–17*) and 2D crystal space group (*18*) symmetry through the design of a single new protein-protein interface. In contrast, the building blocks in most reversibly assembling filaments have no internal symmetry, and hence multiple designed interfaces are required to drive formation of the desired structure. The reduced symmetry also makes the sampling problem more challenging, as the space of possible filament geometries is extremely large.

Computational Design Method

To tackle the challenge of *de novo* designing dynamic protein filaments, we devised a computational approach that exploits the requirement for multiple inter-monomer interfaces to reduce the size of the search space (Fig. 1C). Simple helical symmetry results from repeated application of a single rigid body transform; we also consider architectures in which multiple such simple helical filaments are arrayed with cyclic symmetry. Hence the search is over the 6 rigid body degrees of freedom, and the discrete degrees of freedom associated with the different cyclic symmetries. The approach starts from an arbitrary asymmetric protein monomer structure, and generates a second randomly oriented copy in physical contact by 1) applying a random rotation (3 degrees of freedom), 2) choosing a random direction (two degrees of freedom), and sliding the second copy towards the first until they come into contact (Fig. 1C, left; the sliding into contact effectively reduces the number of degrees of freedom from the six for an arbitrary rigid body transform to five). Successive monomers related by the filament defining rigid body transform need not themselves be in contact, and such arrangements are rare in biology. To go beyond this restriction, we consider not only filaments generated by the rigid body transform relating the two contacting monomers, but also those generated by the n-th root of this transform, where n ranges from 2 to 5--with a choice of n = 4, for example, the 1st monomer will be in contact with the 4th monomer (Fig. 1C, bottom, Fig. S1). We also consider filaments with cyclic symmetry generated by application of Cn symmetry operations around the superhelical axis, where n is between 2 and 5 (Fig. 1C, middle). In all cases, we then generate several repeating turns of the full filament by repeated application of the rigid body transformation and cyclic symmetry operations, eliminate geometries with clashing subunits, and require the existence of at least one additional interface beyond that generated in the initial sliding into contact step. Filament architectures with multiple interacting surfaces predicted to have low energy after design (*19*) are selected, and Rosetta combinatorial sequence optimization is carried out on a central monomer, propagating the sequence to all other monomers. The resulting designs, which span the range of helical parameters (diameter, rise, and rotation, table S1) of native filaments (Fig. 1B, blue dots), are filtered for high shape complementarity, low monomer-monomer interaction energy and few or no buried unsatisfied hydrogen bonds.

We chose as the monomeric building blocks a set of 15 *de novo* Designed Helical Repeat proteins (*20*) (DHRs) which span a wide range of geometries and hence can give rise to a wide range of filament architectures. In addition to shape diversity, the DHRs have the advantages of very high stability and solubility, and are likely to tolerate the substitutions needed to design the multiple interfaces required to drive filament formation. They can also be extended or shortened simply by addition or removal of one or more of the 30-60 residue repeat units, potentially allowing tuning of the diameter of designed filaments. Starting from both the computational design models and the x-ray crystal structures of the DHRs, we generated 230000 helical filament backbones as described above and selected 124 designs for experimental testing (we refer to these as *de novo* Designed Helical Filaments or DHFs throughout the text; for comparison with filaments generated from native backbones, see Fig. S2).

Biophysical characterization

The designs were expressed in *Escherichia coli* under the control of a T7 promoter and purified using immobilized metal affinity chromatography (IMAC). Eighty-five of the designs were recovered in the

IMAC eluate, while 22 were in the insoluble fraction (17 designs were not found in either fraction). IMAC eluates were concentrated, and filament formation was monitored by negative stain electron microscopy (EM); insoluble designs were characterized by EM either directly in the initial insoluble fraction, or after solubilization in guanidine hydrochloride, IMAC, and subsequent removal of denaturant. A total of 34 designs (15 soluble and 19 insoluble) were found to form one-dimensional nanostructures (Fig. S3 and S4). A subset of the designs were synthesized as SUMO fusions to prevent premature filament formation; the SUMO tag was removed using SUMO protease and the samples characterized by negative stain EM (Fig. S5).

Structural characterization

We chose six designs with a range of model architectures and highly ordered negative stain EM morphologies for higher resolution structure determination by cryo-electron microscopy (cryoEM). We determined the filament structures and refined helical symmetry parameters using iterative helical real space reconstruction in SPIDER (*21*, *22*), followed by further 3D refinement in Relion (*23*) and Frealign (*24*). In all six cases, the overall orientation and packing of the monomers in the filament were similar in the experimentally determined structures and design models, but there was considerable variation in the accuracy with which the details of the interacting interfaces were modeled (Fig. 2, Fig. S6). Subtle shifts in the interaction interfaces in several cases altered the designed symmetry; DHF119 for example, was designed to be C1 but the cryoEM structure has C3 symmetry (helical lattice plot comparisons are in Fig. S7). Four of the six designed filaments matched the computational models at near-atomic resolution: for DHF38 and DHF 91 the experimentally observed rigid body orientation was nearly identical to the design models (0.9 Å and 1.2 Å r.m.s.d. over three chains containing all unique interfaces), for DHF46 and DHF119 the r.m.s.d. over three chains was 2.3 Å, and for DHF91 and DHF58, 3.6 and 4 Å. The structure of DHF119 was solved to 3.4 Å resolution; the backbone and side chain conformations at the subunit interfaces are very similar to those in the design model (Fig. 2G).

Modular tuning of fiber diameter

To determine whether the filament diameter could be modulated by changing the number of repeat units in the monomer, we generated a series of DHF58 variants that retain the fiber interaction interfaces but have three, four, five or six repeats in the protomer. The designs were expressed, purified and characterized by negative stain EM: consistent with the computational models (Fig. 3A), the diameter of the filaments changes linearly with the number of repeat units (Fig. 3B, C).

Characterization of fiber growth and disassembly

We monitored assembly dynamics *in vitro* by solution scattering and in living cells using fluorescence microscopy with monomers fused to green fluorescent protein (GFP). The extent and kinetics of DHF119 filament formation *in vitro* was strongly concentration-dependent. Filament nucleation was too fast to observe by manual mixing; the rate of the observed elongation phase was linear with respect to monomer concentration, and extrapolation of the plateau values from progress curves back to zero yielded a critical concentration of 3 μM (Fig. S8). Upon dilution below the critical concentration, filaments disassembled in several hours (Fig. S9). In *E. coli* following induction of expression of DHF58-GFP, discrete punctae were first observed, which over time resolved into filaments up to microns in length (Fig. 4A, Movie S1; the punctae may simply be filaments below the resolution limit of the microscope, approximately 250 nm). The filaments formed *in vivo* have high stability: after lysis of bacteria expressing the monomer by lysogenic phages, the filaments largely retain their shape (Movie S2).

Natural systems achieve remarkable complexity and diversity of filament-based structures through modulating the nucleation, growth, and cellular location of the polymers. In some natural systems, nucleation and location are controlled by complexes that act as templates that initiate new growth and anchor filaments to specific locations, like the gamma-tubulin ring complex for microtubules and the Arp2/3 complex for actin. We sought to replicate this mechanism of control by designing multimeric anchor constructs, with multiple monomeric subunits held close to the relative orientations in the corresponding filaments by a fusion to designed homo-oligomers with the appropriate geometry (Fig. S10; one of the interaction interfaces is eliminated to restrict fiber growth in one direction). For example, anchor DHF119_C6 (Fig. 4B) is a hexamer in which each monomer consists of a designed oligomerization domain fused to the fiber monomer; the orientations of the monomers in the hexamer are close to those in the filament structure to promote both nucleation and fiber attachment. To study the kinetics of filament formation *in vitro* in more detail, we attached the anchors to glass slides, added monomers fused to yellow fluorescent protein (YFP) and monitored fiber formation by total internal reflection (TIRF) microscopy. The anchors seeded the rapid growth of multiple micron length fibers over 30 minutes (Fig. 4C, Movie S3; for analysis of growth kinetics of a second fiber see Fig. S11). Few or no fibers were observed to grow from the glass slide surface when it was coated with an anchor designed for a different fiber (Movie S4), or with no anchor at all (Movie S5). Attachment of biotinylated anchor to streptavidin-coated beads, followed by incubation with filament monomer resulted in an extensive network of filaments emanating from the beads (Fig. 4D, left panel); in contrast, very few filaments were observed around control beads that lacked the anchor protein (Fig. 4D, right panel).

To determine whether filament dissolution could also be modulated by designed accessory proteins, we produced monomeric capping units lacking one of the two designed interfaces in the DHF119 filament--these caps are expected to add to one end of the filament, but not the other, preventing further elongation (since the two ends of the filaments are distinct, there are two types of caps). Addition of increasing concentrations of the caps to already formed filaments resulted in shrinking and ultimately disappearance of the filaments (Fig. S12), suggesting that filaments are dynamically exchanging protomers at equilibrium. Monitoring of cap-induced disassembly by atomic force microscopy (AFM) showed that fibers incubated with equal concentrations of the protomers and single end caps disassemble primarily from one (presumably the uncapped) end, while in the presence of both caps disassembly occurs from both ends (Fig. 4E, Fig. S12). In the absence of caps, increasing the monomer concentration led to growth from both ends of the fibers at a rate (~15 nm/minute per end at 18µM monomer; Fig. S12) similar to that observed by fluorescence for anchored fiber growth (8.4 nm/minute at 18µM monomer, Fig. S11). The observed behavior can be understood as follows. At the critical monomer concentration where fibers neither grow or shrink, the (concentration dependent) rate of monomer addition to the ends is balanced by the (concentration independent) disassociation rate. Caps perturb this balance by complexing with monomers in solution (Fig. S13, bottom) effectively reducing the free monomer concentration, hence when both end caps are present, disassembly wins out over growth, leading to a net shrinking of the filaments. When one cap is present, the net rate of subunit addition is greater at the end where both free monomers and free caps can add (Fig. S13, top right) than at the other end where only monomers can add (Fig. S13, top left). Since the rate of monomer dissociation is the same at both ends, the fibers shrink primarily from one end as observed.

Conclusion

The ability to program micron scale order from Ångstrom scale designed interactions between asymmetric monomers is an advance for computational protein design. In contrast to previous nanomaterial design efforts relying on an already existing interface within symmetric building blocks, proper assembly requires the design of two independent interfaces. Introduction of a small number of hydrophobic substitutions near the periphery of dihedral complexes can promote stacking into extended filaments since each sequence change is replicated multiple times at the stacking interface (*25*); the filaments described here are instead built from monomeric building blocks and have a much wider range of geometries since only a small fraction of possible helical assemblies contain dihedral point group symmetry. Both designed interfaces were accurately recapitulated in four of the six structures solved by cryoEM; despite the deviations in the interfaces in the other two, the overall filament architecture was reasonably well recapitulated. The ability to program filament dynamics provides a baseline for understanding the much more complex regulation of the dynamic behavior of naturally occurring filaments. The repeat protein building blocks are hyperstable proteins robust to genetic fusion, and hence the designed filaments provide readily modifiable scaffolds to which binding sites for other proteins or metal nanoclusters can be added for applications ranging from cryoEM structure determination to nano-electronics.

Main Text Figures

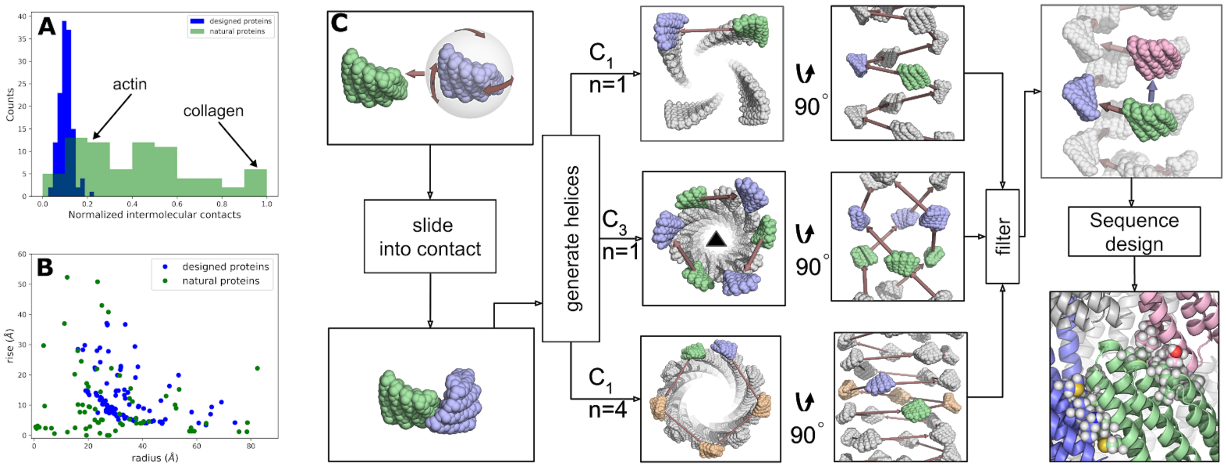


Fig. 1. Filament architectures and computational design protocol. (**A**) The fraction of total residue-residue interactions within (rather than between) monomers. (**B**) Super helical parameters. (**C**) Computational design protocol. In (A) and (B) the properties of filaments generated by the design protocol (blue) are compared to those of naturally occurring proteins (green).

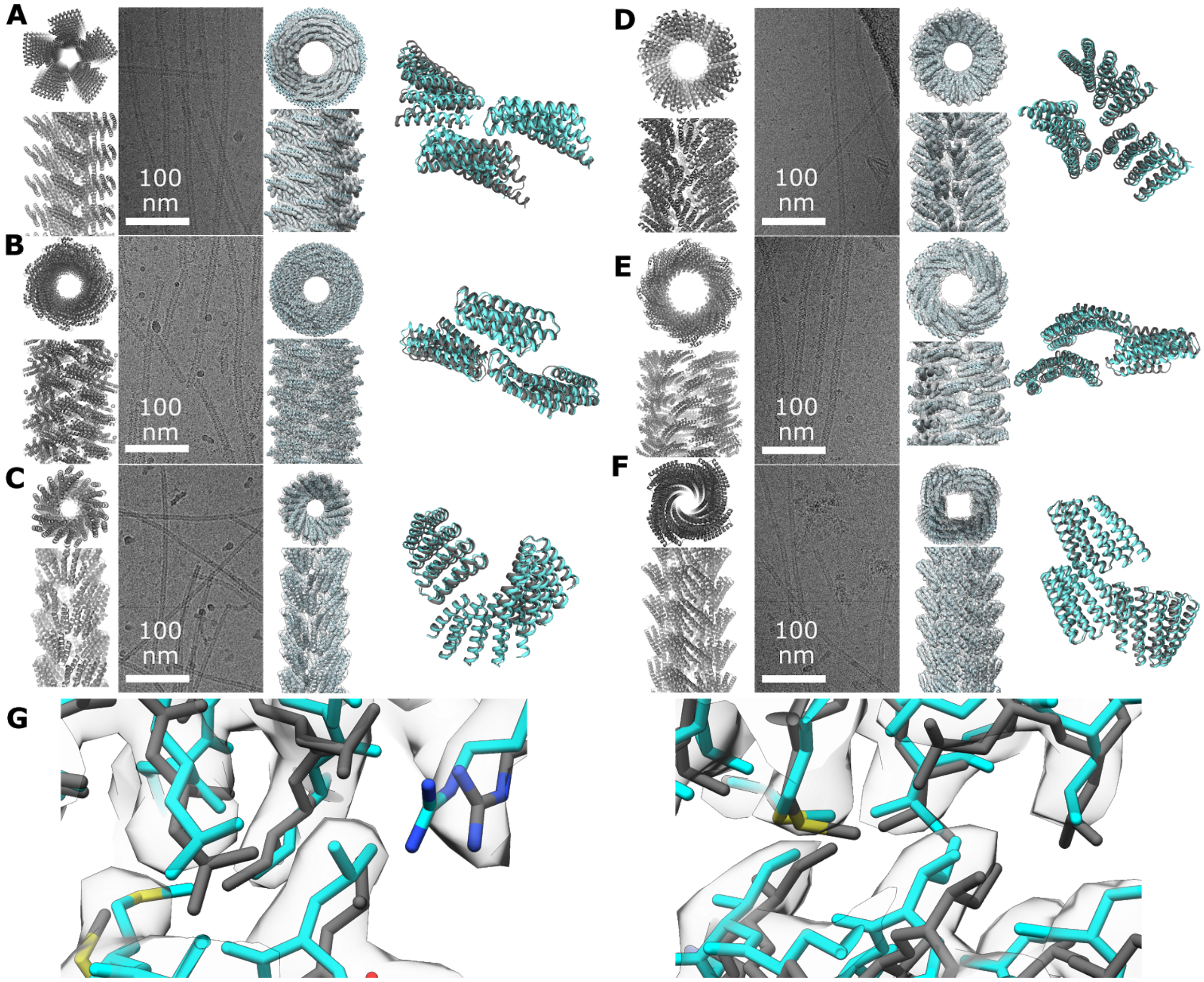


Fig. 2. CryoEM structure determination. Computational model (first panel), representative filaments in cryoEM micrographs (second panel), cryoEM structure (third panel) and overlay between model and structure (fourth panel) for (**A**) DHF58 - r.m.s.d. 3.3 Å (**B**) DHF119 - r.m.s.d. 2.3 Å (**C**) DHF91 - r.m.s.d. 1.2 Å (**D**) DHF46 - r.m.s.d. 2.2 Å (**E**) DHF79 - r.m.s.d. 4 Å (**F**) DHF38 - r.m.s.d. 0.9 Å (**G**) The high-resolution structure of design DHF119 is very close to the design model. Close up views of the two main intermonomer interfaces in the filament, with the computational model (grey) and cryoEM structure (cyan) in sticks in the helical reconstruction density (3.4 Å resolution).

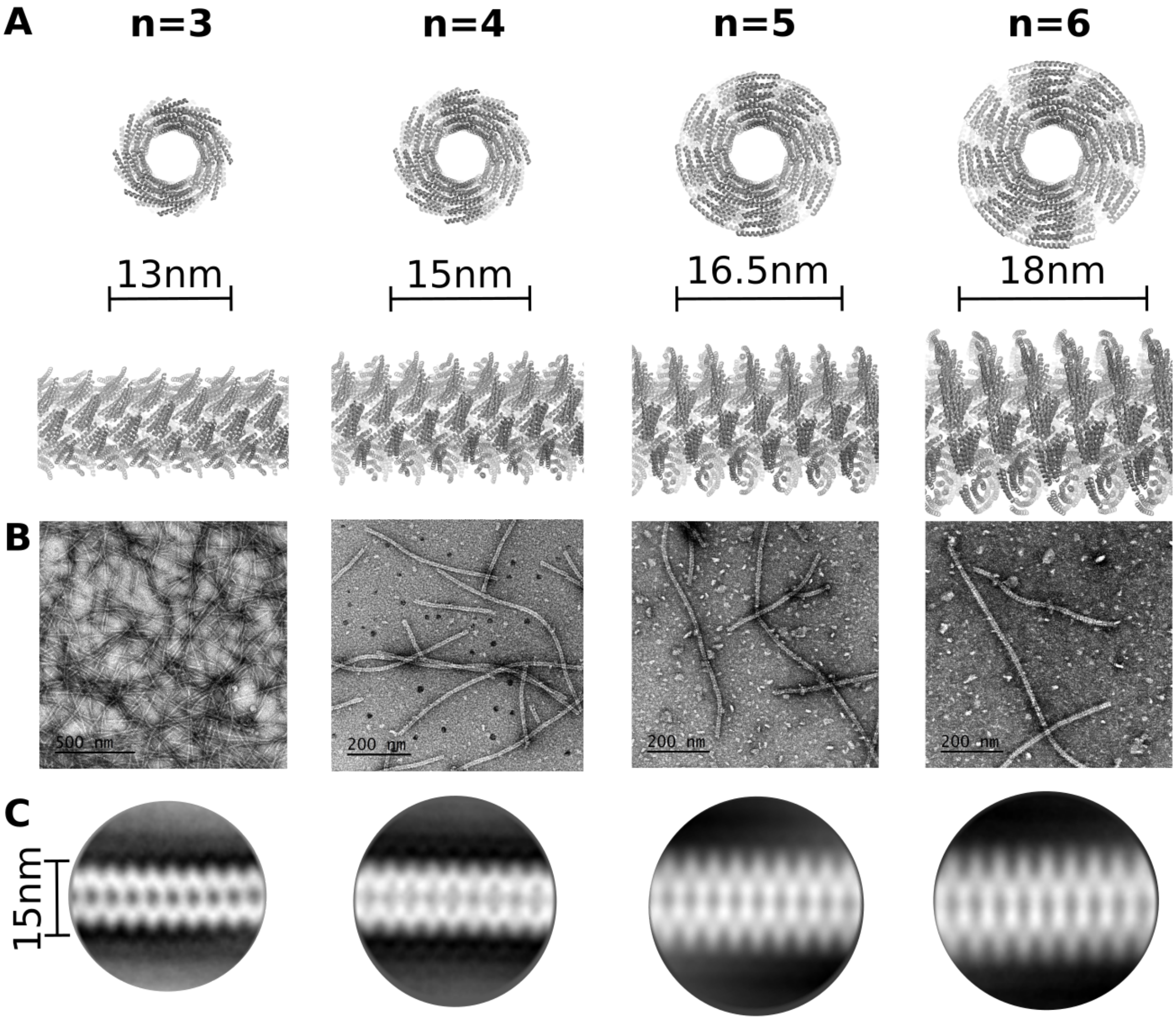


Fig. 3. Modular tuning of fiber diameter. DHF58 filament variants with different numbers of repeats were characterized by electron microscopy. (**A**) Top: number of repeats. Cross sections (middle) and side views of computational models based on the 4-repeat cryoEM structure. (**B**) Negative stain electron micrographs. (**C**) 2D class averages.

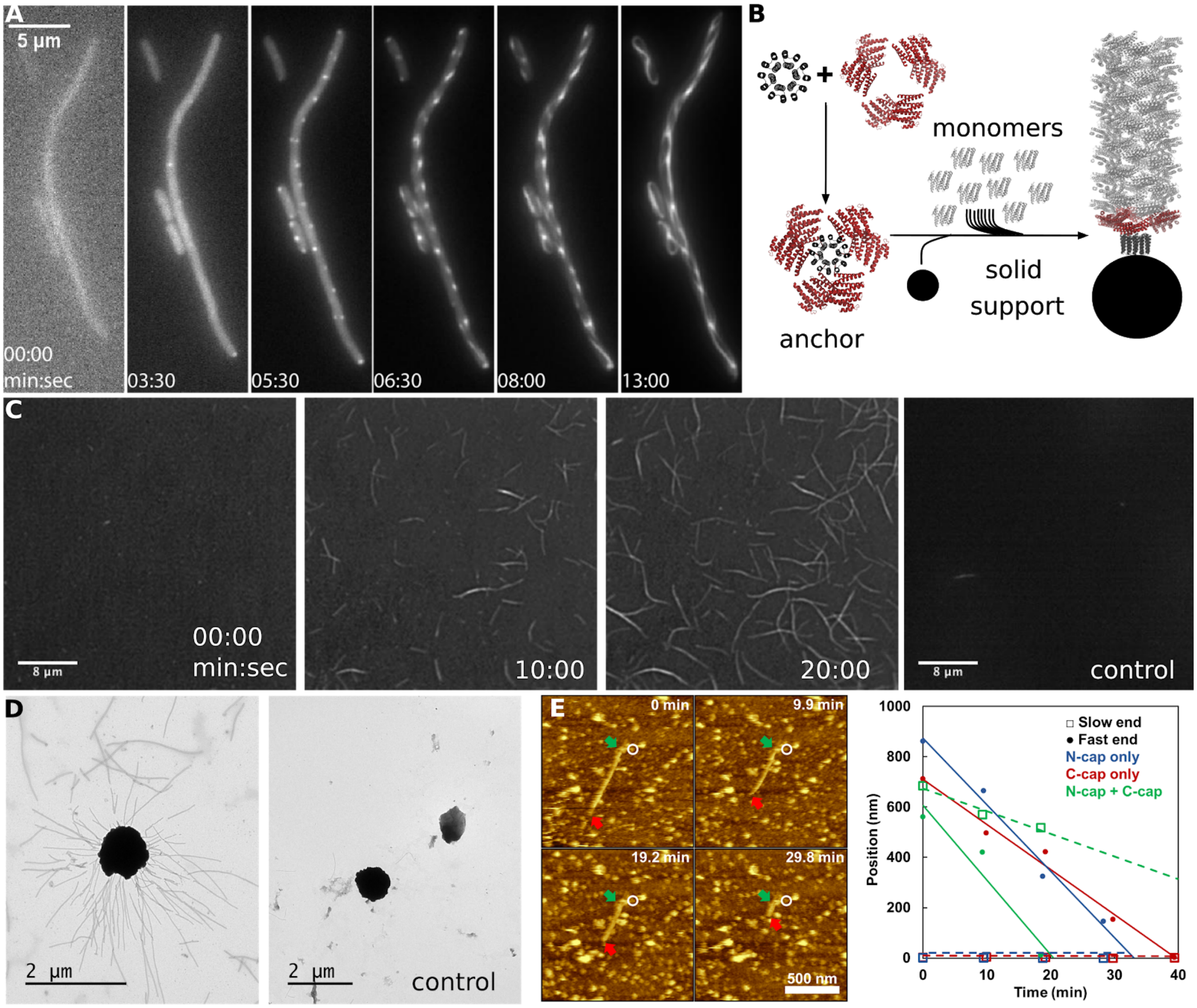


Figure 4. Characterization of fiber growth and disassembly. (**A**) Kinetics of assembly of DHF58-GFP filaments in *E. coli*. (**B**) Construction of fiber anchors holding monomers in rigid body arrangement found in the filament. (**C**) Kinetics of DHF119-YFP filament assembly *in vitro* on glass surface coated with DHF119_C6 anchor. In right panel, the glass surface was coated with the non-cognate DHF91 anchor. (**D**) DHF119 filaments emanating from DHF119_C6 anchor coated magnetic bead incubated with monomer; beads on right lack anchor. (**E**) Disassembly of DHF119 fibers in the presence of capping units monitored by *in situ* AFM. Left, image sequence showing disassembly in presence of N-caps. White circle marks fixed position in all images. Right, positions of fiber ends vs time in solutions with N-caps, C-caps, and N-caps + C-caps, respectively. In all cases, the DHF119 monomer concentration and the total cap concentration are each 3.8 µM (at this concentration of monomer, fibers neither dissolve nor grow in the absence of caps). Because they lack one of the filament interfaces, caps can only bind to one

end; disassembly from this end will be slower as the combined on rate of caps and monomers is greater than the on rate of monomers alone at the other end.

Materials and Methods

Generation of Filament Models from Monomeric Building Blocks

The goal of the computational helix docking procedure was to exhaustively sample, to within some specified resolution and acceptable interface quality, all possible ways to build a symmetric helix from a monomeric building block. We start by enumerating all possible head-to-tail dimeric arrangements of the monomer. The six-dimensional rigid body docking space of three rotations and three translations is reduced to five by requiring contact between the bodies; the three translational degrees of freedom are replaced by a two-dimensional space of normal vector directions and a slide into contact (Fig. 1C, left). Experience suggests this reduction works reasonably well for globular bodies (*19*). The resulting dimer interface is scored using the RPX method, and those below a cutoff are discarded. Given a head-to-tail homodimeric interface X, docking proceeds by generating possible helix geometries containing interface X along with at least one other interface Y. Two discrete parameters N and C determine what helices can be constructed by repeating interface X. Parameter C specifies the cyclic symmetry of the result (Fig. 1C middle), and parameter N specifies how many helix unit transforms are needed to produce interface X (Fig. 1C bottom). Given X, N, and C, a rapid check of the helical spacing is performed, as most combinations will result in clashing or overly extended helices without a second protein interface. If this check passes, the geometry is explicitly generated and checked for the presence of a second homomeric interface Y. Interface Y is then scored using RPX, and the score for the overall helix is the worst of X and Y.

Interface Design

Docks with appropriate and evenly-distributed interface sizes as well as good RPX scores *((19))* were selected to perform interface sequence design in RosettaScripts. In each design trajectory, the protomer was initially perturbed by a random rotation around its center of mass. A polymer with the specified helical symmetry was generated using the information stored in the symmetry definition file, which was generated from the initial docking configuration using tools distributed with the Rosetta Macromolecular Modeling suite. Amino acids at the interface were optimized using Monte Carlo simulated annealing protocol available in the Rosetta Macromolecular Modeling suite. An initial optimization step was executed with the rotamers available in the database of residue-pair motifs and a modified score function with a down-weighted repulsive term. Once a sequence was converged on, designable positions were allowed to minimize side-chain torsion angles. A subsequent round of minimization was conducted with the standard score function to obtain a conformation that corresponds to a local minimum of the energy function. Individual design trajectories were filtered by the following criteria: the difference between the Rosetta energy of the bound (polymeric) and unbound (monomeric) states less than −15.0 Rosetta Energy Units, interface surface area greater than 700 $Å^2$, Rosetta shape complementarity greater than 0.62 and unsatisfied polar residues less than 5. Designs that passed these criteria were manually inspected and refined by single-point reversions for mutations that were deemed not to contribute to stabilizing the bound state of the interface. The design with the best overall scores for each docked configuration was then added to a set of finalized proteins to be validated experimentally.

Accessory Protein Design

Capping units for DHF58 and DHF119 were designed by mutating the residue identities at the interfaces that drive filament growth to identities in the corresponding scaffold proteins. Capping proteins with reversions in primary sequence close to the N-terminus are referred to as N caps while proteins with reversions in the primary sequence at the C-terminal end are referred to as C-caps. The anchor protein DHF119_C6 was designed by fusing the monomer from designed hexamer 3H22 to the C cap of DHF119 with a (GGS)5 linker. An avi-tag (GLNDIFEAQKIEWHE) was added to the N terminus of 3H22 for biotinylation.

Protein Expression and Purification

Synthetic genes for 124 designs were optimized for *E. coli* expression and purchased from Gen9 and Genscript ligated in the multiple cloning site of the pET28b vector between NdeI and XhoI restriction sites or in vector pCDB24 (*26*). This vector contains SUMO protein Smt3 from *Saccharomyces cerevisiae* to prevent premature assembly in *E. coli* and improve solubility. These plasmids were cloned into BL21* (DE3) (Invitrogen) *E. coli* competent cells. Transformants were inoculated into 50 ml of TB medium with 200 mg L^{-1} kanamycin. Expression proceeded for 24 hours at 37 °C following the expression *via* Studier autoinduction (*27*) until the cultures were harvested by centrifugation. Cell pellets were resuspended in TBS and lysed using the Bugbuster detergent (Millipore). The soluble fraction upon lysate clarification by centrifugation was purified by Ni^{2+} immobilized metal affinity chromatography with Ni-NTA Superflow resin (Qiagen). Resin with bound cell lysate was washed with 10 column volumes of 40 mM imidazole and 500 mM NaCl and eluted with 400 mM imidazole and 75 mM NaCl. Both the soluble and insoluble fractions were run on an SDS-PAGE gel. Samples that showed protein bands at the correct molecular weight were selected for screening by electron microscopy. Proteins expressed in the pCDB24 vector were screened before and after cleavage of the fusion protein using the SUMO protease (Fig. S5). Selected designs were expressed at the 0.5 L scale to carry out further characterization. Expression proceeded for 24 hours at 37 °C following the expression *via* Studier autoinduction (*27*) until the cultures were harvested by centrifugation. Cell pellets were resuspended in TBS and lysed by microfluidization. Purification was carried out as described above.

Negative Stain Electron Microscopy

Soluble fractions were concentrated and insoluble fractions were resuspended in buffer (25mM Tris, 75 mM NaCl, pH 8) for EM screening. A drop of 6µL (1µl sample instantly diluted with 5µl of buffer) was applied on negatively glow discharged, carbon-coated 200-mesh copper grids (Ted Pella, Inc.), washed with Milli-Q Water and stained using 0.75% uranyl formate as described previously (*28*). The screening was performed on either a 120kV Tecnai Spirit T12 transmission electron microscope (FEI, Hillsboro, OR) or a 100kV Morgagni M268 transmission electron microscope (FEI, Hillsboro, OR). Images were recorded on a bottom mount Teitz CMOS 4k camera system. The contrast of the images was enhanced in the Fiji software (*29*) for clarity.

CryoEM Sample Preparation and Data Collection

CryoEM samples were prepared by applying protein to glow-discharged C-Flat holey-carbon grids (Protochips Inc.), blotting with a Vitrobot (FEI co.), and plunging into liquid ethane. For DHF58, DHF46, DHF79, and DHF91 samples, data was collected on a Tecnai G2 F20 (FEI co.) operating at 200 kV with a K-2 Summit Direct Detect camera (Gatan Inc.) with a pixel size of 1.26 Å/pixel. Movies were acquired in counting mode with 36 frames and a total dose of ~45 $e^-/Å^2$. For DHF119 and DHF38 samples, data was collected on a Titan Krios (FEI co.) operating at 300 kV, with a Quantum GIF energy filter (Gatan Inc.) operating in zero-loss mode with a 20 eV slit width, and a K-2 Summit Direct Detect camera with a pixel size of 0.525 Å/pixel. Movies were acquired in super-resolution mode with 50 frames and a total dose of ~90 $e^-/Å^2$. All data was collected with a defocus range between 1.0 and 2.5 µm, using Leginon (*29*, *30*) or EPU (FEI co.) software for automated data collection.

Image Processing, 3D Reconstruction, and Model Building

Movie frames were aligned and dose-weighted using MotionCor2 (*31*) and CTF values were determined using GCTF (*32*). Helices were picked manually using Appion (*32*, *33*) or Relion (*23*) software, and particles were extracted as overlapping segments along the length of each helix. Reference-free 2D classification of helical segments was then performed using Relion. For DHF119 and DHF38, selected 2D classes obtained from manual picking of a subset of images were used as templates for automated picking in Relion. Following 2D classification of all particles, particles from good classes were selected for subsequent 3D reconstructions. Initial 3D reconstructions were performed by iterative helical real space reconstruction (IHRSR) (*21*), (*34*) in SPIDER, using cylinders as starting models, and using hsearch_lorentz (*22*) to refine helical symmetry parameters. In cases where additional point group symmetry became apparent, this was enforced in subsequent rounds of refinement. Gold-standard refinement in SPIDER was performed with increasingly smaller angular sampling, with a minimum sampling of 1.5°. For DHF119, DHF38, DHF79, and DHF91, further 3D helical refinement was performed using Relion, using the values determined by hsearch_lorentz as initial helical symmetry parameters, and the SPIDER volumes (low-pass filtered to 30Å) as starting models. For DHF38, angles and shifts determined by Relion were further refined by local refinement in Frealign MODE 1 (*24*). For DHF58 and DHF46, volumes were amplitude corrected and low/high-pass filtered in SPIDER. For DHF119, DHF38, DHF79, and DHF91, volumes were B-factor sharpened and low-pass filtered using Relion post-processing. The gold-standard FSC=0.143 criterion was used for estimating resolution. Atomic models were fit into cryoEM density as rigid bodies. For DHF119 and DHF38, atomic models were further refined by real-space refinement in Phenix (*35*) and Coot (*36*).

Fluorescent Microscopy in Living Cells

Plasmid pET28b-DHF58-GFP was electroporated into competent cells of *E. coli* strain BL21(DE3) (Novagen/EMD Millipore cat#69450), yielding strain CJW6472. For the light microscopy experiments, CJW6472 was grown at 37° C in LB medium (10 g/L NaCl, 5 g/L yeast extract, 10 g/L tryptone) containing kanamycin (50 μg/mL) After overnight growth, the bacterial culture was diluted into fresh medium and allowed to grow for 4 to 5 h until reaching an optical density at 600 nm (OD_{600nm}) of 0.2 - 0.35. Cells were then spotted onto a 1% agarose-LB pad containing kanamycin (50 μg/mL) and IPTG (1 mM) to maintain the plasmid and to induce expression of the filament-GFP protein, respectively. This agarose pad was then imaged by time-lapse microscopy. For the phage-lysis experiment, IPTG was added to an exponentially growing culture of CJW6472 cells (OD_{600nm} ~ 0.3). After 30 min of incubation, 3 μL of cells mixed with 1 μL of a T7 phage preparation were spotted on a 1% agarose-LB pad containing kanamycin for imaging. Cell imaging was performed on Eclipse Ti-E microscopes (Nikon, Tokyo, Japan), equipped with a Perfect Focus System (Nikon), a phase-contrast Plan Apochromat 100X, 1.40 numerical aperture objective (Nikon), and an ORCA-Flash4.0 V2 Digital CMOS camera (Hamamatsu Photonics, Hamamatsu City, Japan). Still images and live streaming videos were acquired using NIS-Elements Ar software (Nikon Instruments) or

MetaMorph software (Molecular Devices, San Jose, CA, USA) and processed using the Fiji software (*29*). During imaging of live cells, the temperature was maintained at 37°C using a custom built environmental enclosure chamber designed by Precision Plastics Inc. (Beltsville, MD) or an objective heater (20/20 Technologies Temp Controller TC-500).

Kinetic Measurements

Filament formation kinetics were determined by turbidity due to light scattering, monitored by absorption at 330 nm wavelength, using an Agilent Technologies (Santa Clara, CA) 8454 UV-Visible spectrophotometer. DHF119 was concentrated at 720µM in 7M guanidine hydrochloride (GuHCl) to prevent self-assembling. To initiate filament formation, DHF119 in GuHCl was diluted with TBS such that the GuHCl concentration was low enough to no longer inhibit growth, typically less than 72 µM. Measurements were started immediately after dilution.

Filament Growth In Vitro

PEG-silane coated glass coverslips were attached to similarly-coated slides with strips of double-stick tape to make flow chambers. All incubations were at 25°C. Dry glass chambers were coated for 2 minutes with 8 mg/ml kappa-casein (Sigma C0406) 10:1 biotinylated casein in BRB80 (80 mM PIPES-KOH ph 6.85 + 1 mM $MgCl_2$ + 1 mM EGTA), washed twice with CK buffer (BRB80 + 1 mg/ml casein + 70 mM KCl), incubated 3 minutes with 0.5 mg/ml neutravidin (Molecular Probes A2666) in CK, then washed three times with CK. Prepared cells were washed once in IB (imaging buffer: 75 mM NaCl + 25 mM Tris-HCl pH 8.0 + 11 mM glucose + 2.5 mM DTT + 0.2 mg/ml glucose oxidase (Sigma G2133) + 0.04 mg/ml catalase (Sigma C40). Biotinylated anchor protein (DHF119_C6 with C-terminal GFP fusion) 36.6 nM in IB was incubated in chamber for 3 minutes, chamber washed twice with IB, and replaced with 1.16 µM DHF119-YFP in IB for observation of assembly. Imaging was carried out using a Personal Deltavision microscope (GE Healthcare) outfitted with 4-laser TIRF capabilities, Olympus 60×, 1.49 NA TIRF objective and Ultimate focus (Applied Precision) at room temperature.

For analysis of growth kinetics of DHF119-GFP fiber: Images were processed for subsequent analysis in CellProfiler. The background of the .tif movies was subtracted using the software Fiji

(*29*) with a ball radius of 5.0. The contrast was modified to reduce background noise to facilitate fiber identification. CellProfiler software (*37*) was used to identify and track fiber through the different time frames. The output files from CellProfiler provided the Major axis length for every fiber through all the movie.

In Situ Atomic Force Microscopy Imaging

Freshly cleaved mica (15 mm; Ted Pella) was incubated with 50 uL 0.01% wt poly-lysine solution (Ted Pella) for 2 min and then washed with water and dried with flowing N_2. The mica substrate was placed into the AFM liquid cell and protein solution was then injected into the cell. In situ images were captured using Silicon Nitride probes (MSCT, k: 0.1 N/m, tip radius: 10 nm; Bruker) under ScanAsyst mode with Nanoscope VIII (Bruker) at room temperature (imaging buffer: 21 mM Tris-HCl, 400 mM KCl, pH 8). Images were analyzed using Nanoscope Analysis v1.5 (Bruker) and Gwyddion SPM data analysis software.

Supplementary Figures

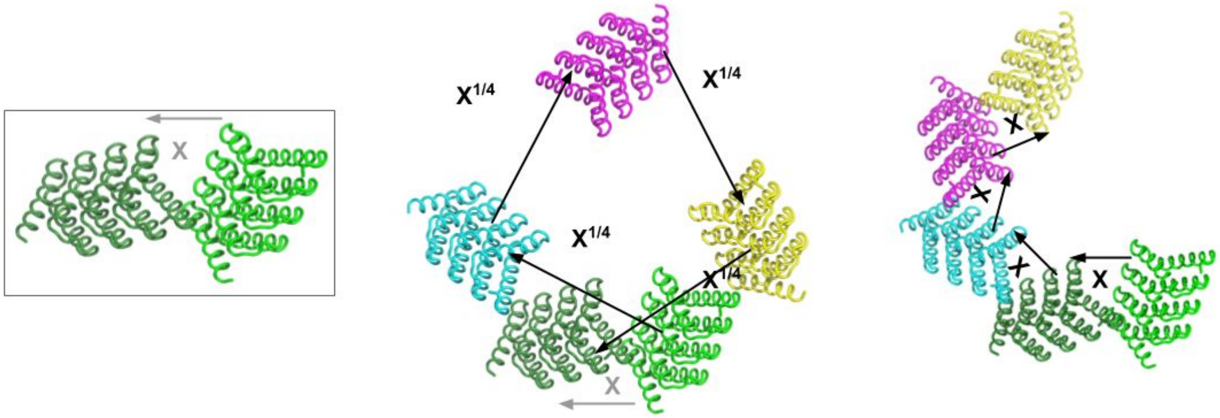


Fig. S1.

Computational Sampling of helical filaments containing a given protein dimer. Left, dimeric complex sampled during the slide into contact stage of the protocol. Middle, helical filament produced by repeatedly applying the fourth root or the transform defined in A (n=4 - the two subunits that define the helical symmetry are not in contact). Right, helical filament produced by repeatedly applying the transform defined in the first panel (n=1 - the two subunits that define the helical symmetry are in contact).

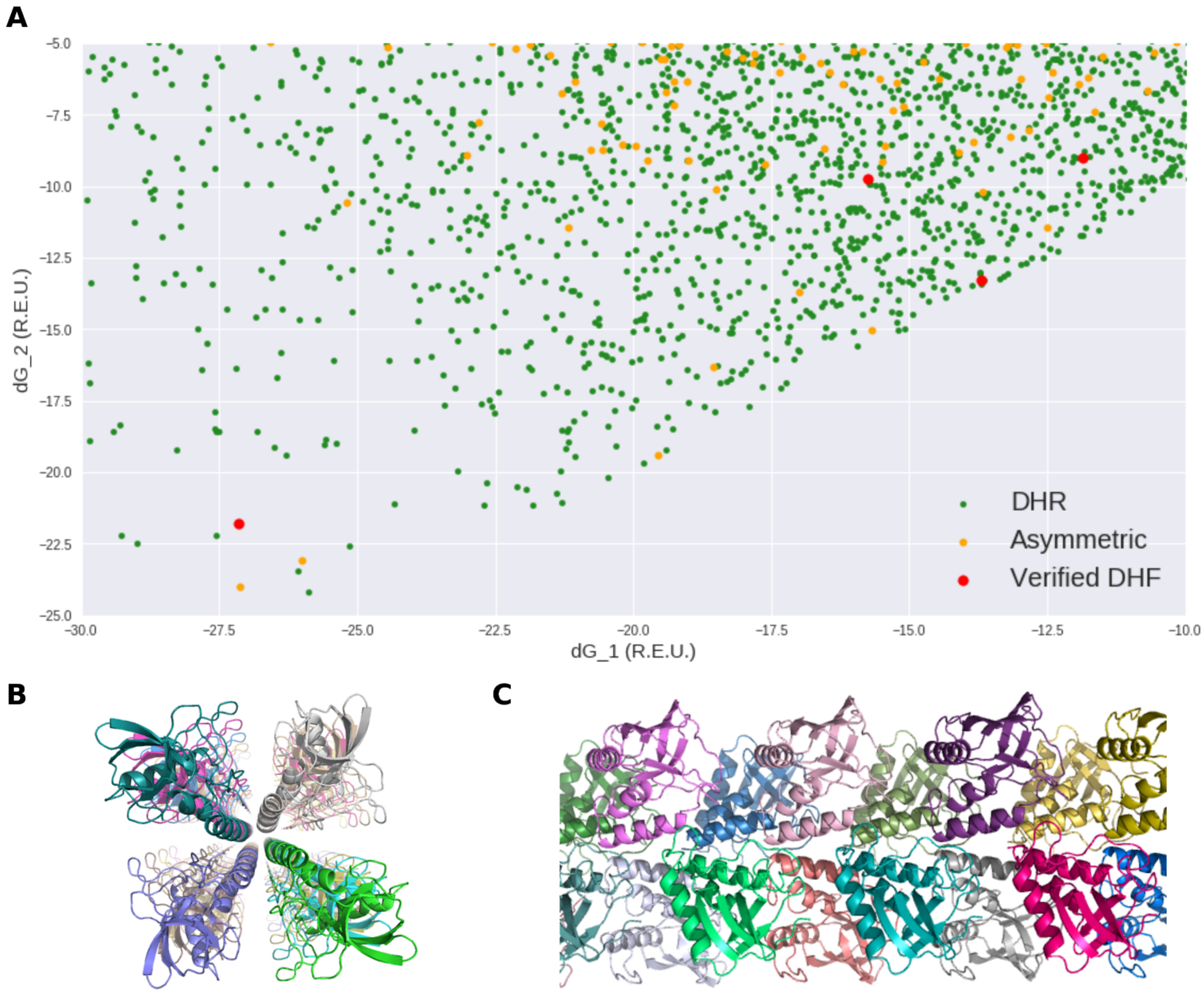


Fig. S2.

Comparison of designs generated from *de novo* Designed Helical Repeat proteins (DHRs) and natural asymmetric proteins. (**A**) Scatter plot of Rosetta binding energy for main and secondary interfaces for designs generated from DHRs, natural asymmetric proteins (PDB ID: 1stn, 2bk9 and 5ghl) and structurally verified *de novo* Designed Helical Filaments (DHFs). (**B**) Top and (**C**) side views for an example fiber design model generated from Staphylococcal nuclease (PDB ID: 1stn) colored by chains.

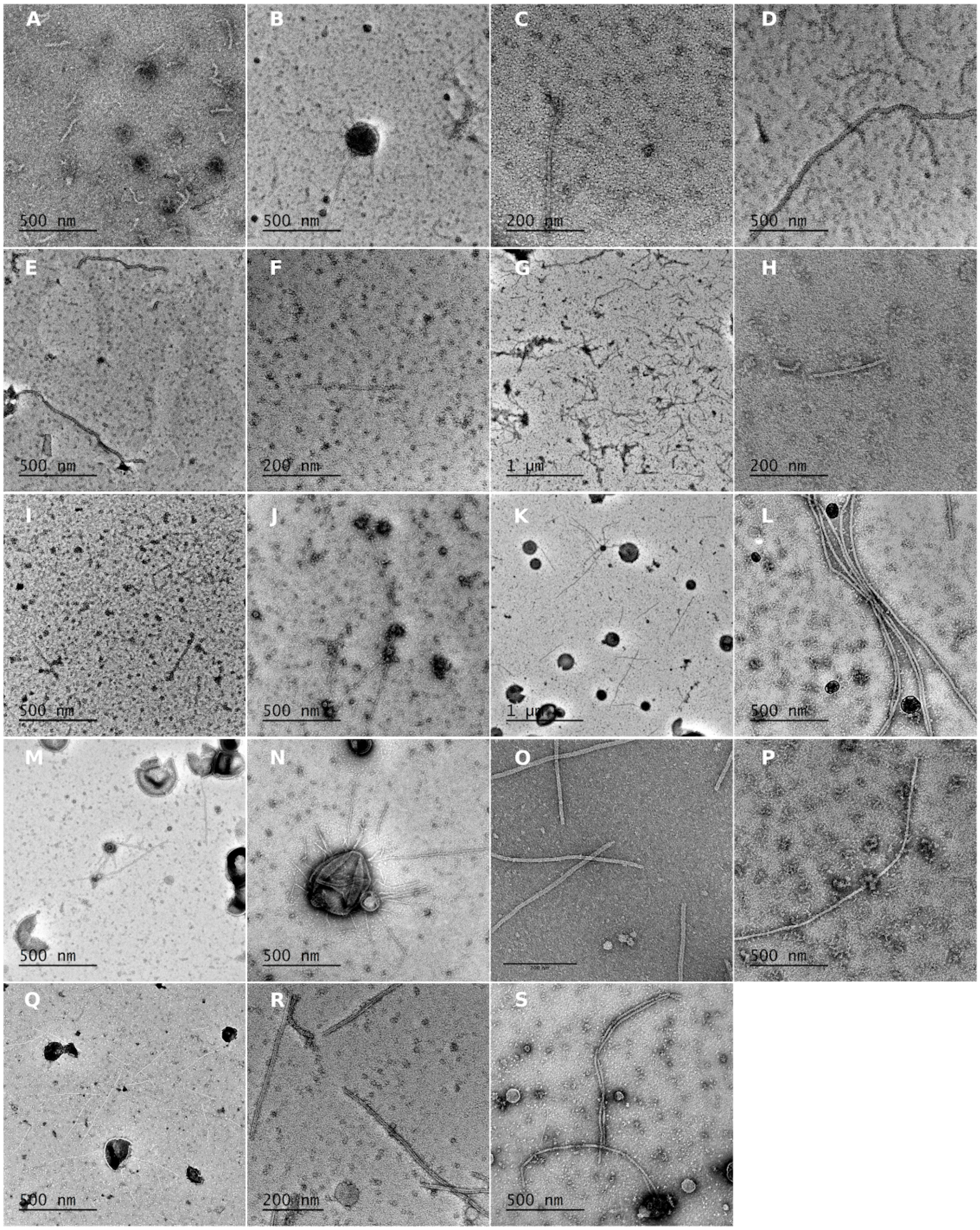


Fig. S3.

Negative stain EM of insoluble filaments. (**A**) DHF4 (**B**) DHF5 (**C**) DHF8 (**D**) DHF16 (**E**) DHF17 (**F**) DHF23 (**G**) DHF25 (**H**) DHF28 (**I**) DHF31 (**J**) DHF34 (**K**) DHF36 (**L**) DHF40 (**M**) DHF43 (**N**) DHF44 (**O**) DHF46 (**P**) DHF47 (**Q**) DHF49 (**R**) DHF50 (**S**) DHF77

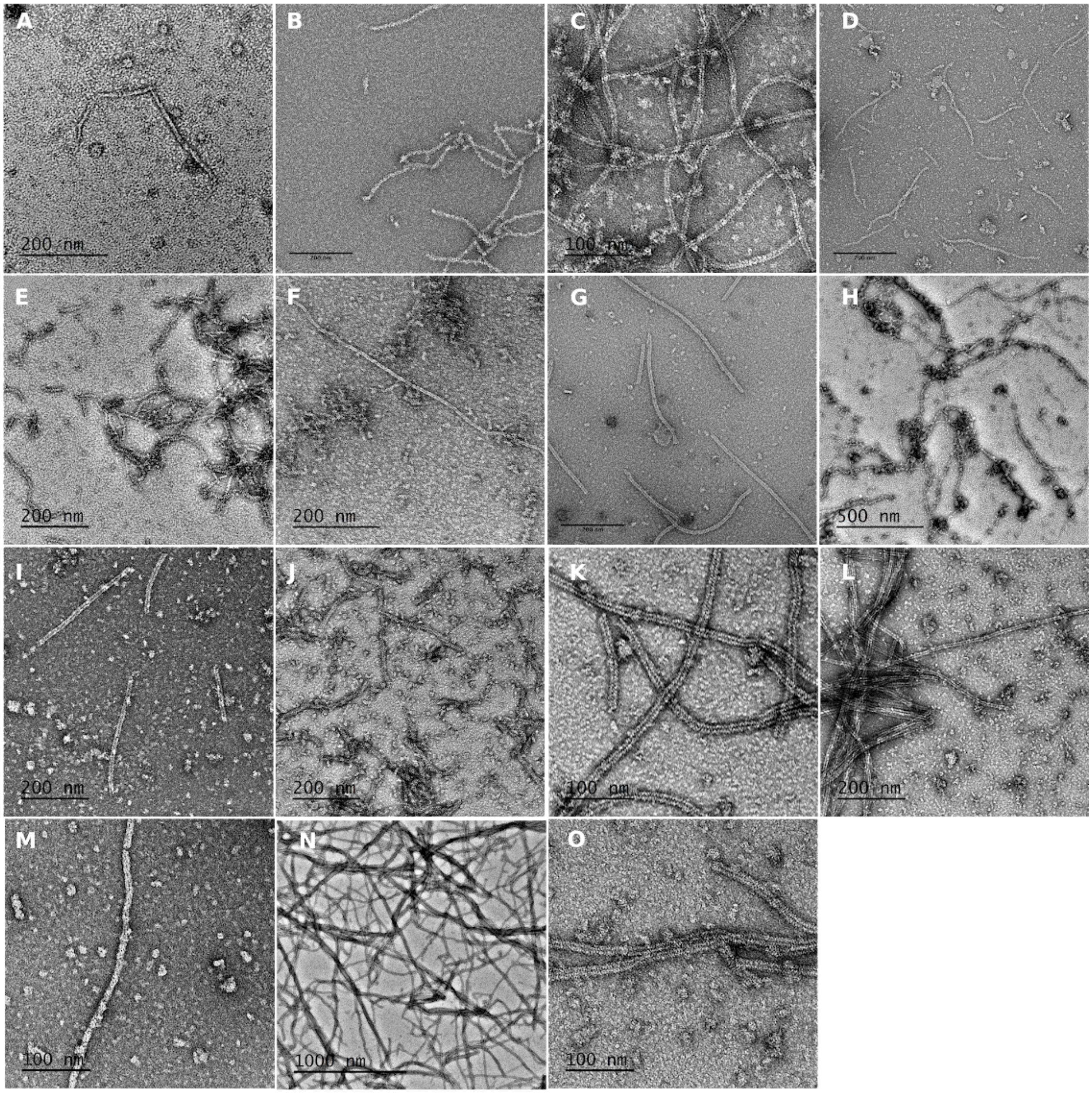


Fig. S4.

Negative stain EM of soluble filaments. (**A**) DHF9 (**B**) DHF20 (**C**) DHF38 (**D**) DHF48 (**E**) DHF51 (**F**) DHF52 (**G**) DHF58 (**H**) DHF62 (**I**) DHF76 (**J**) DHF78 (**K**) DHF79 (**L**) DHF82 **(M**) DHF91 (**N**) DHF107 (**O**) DHF119

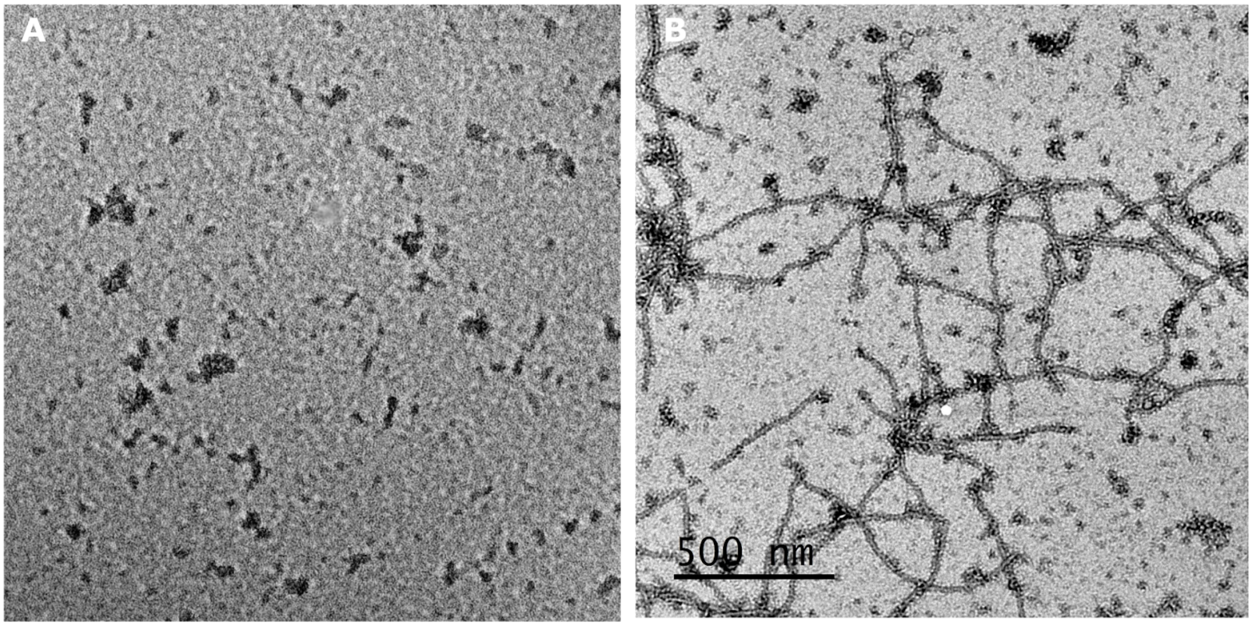


Fig. S5.

Protease induced filament assembly using SUMO fusion constructs. (**A**) Design DHF38 before treatment with SUMO protease. (**B**) Design DHF38 after treatment with SUMO protease.

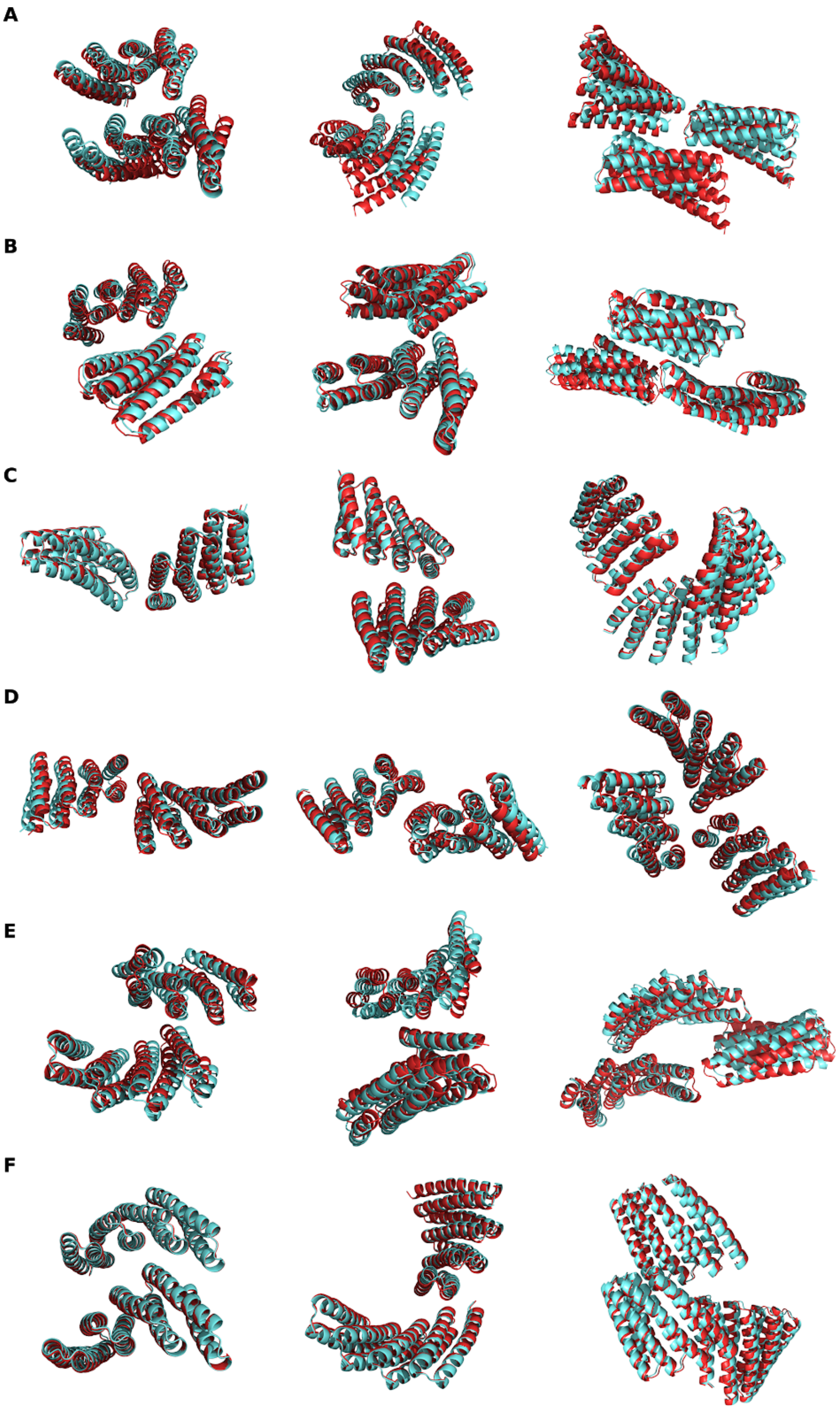


Fig. S6.
Overlay for each individual interface and both interfaces to highlight the differences/similarities between the models(red) and experimental structures(cyan). (**A**) DHF58 (**B**) DHF119 (**C**) DHF91 (**D**) DHF46 (**E**) DHF79 (**F**) DHF38

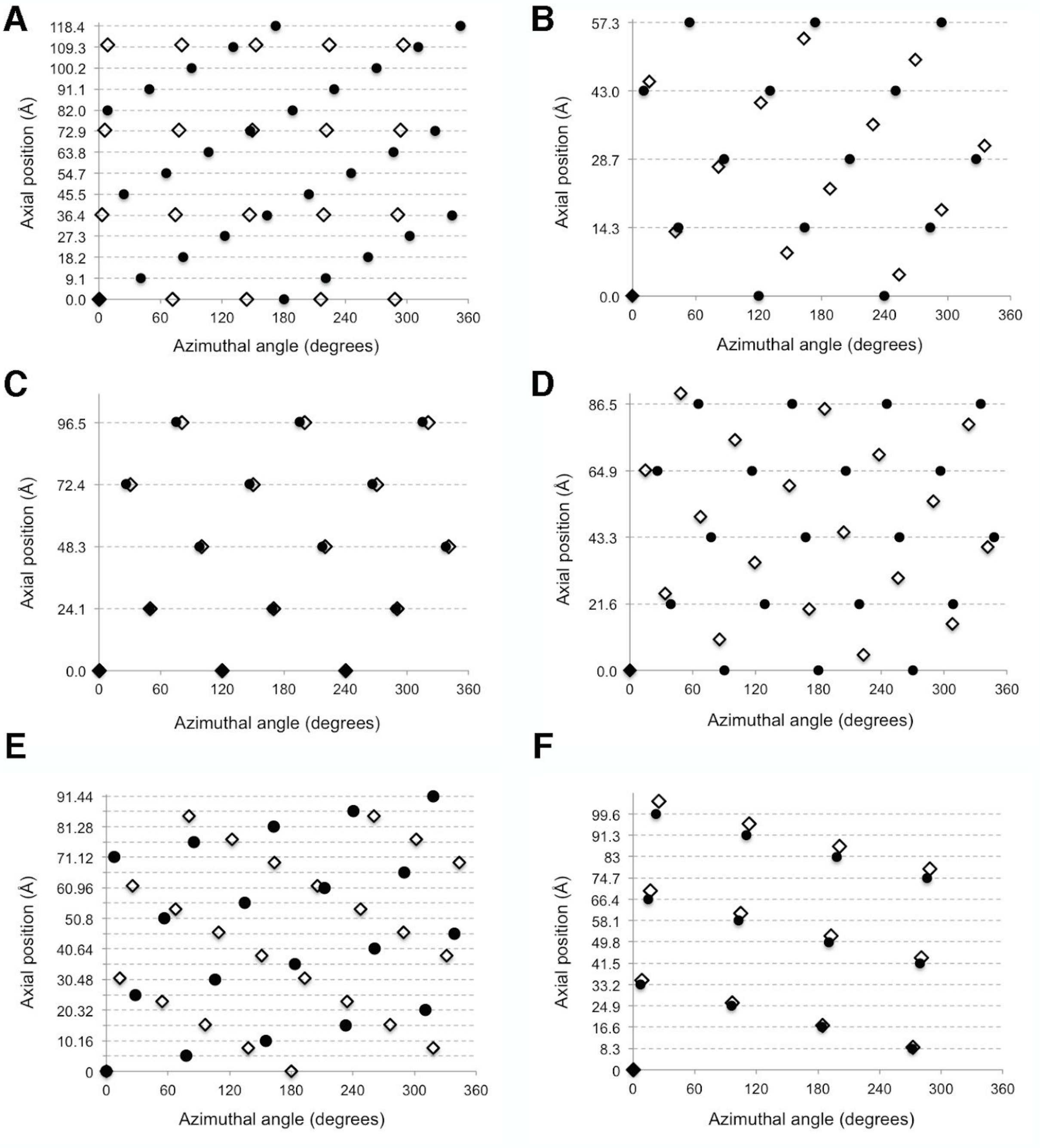


Fig. S7.

Helical lattice plots comparing designed helical symmetry (open diamonds) to experimentally determined helical symmetry (closed circles) for (**A**) DHF58 (**B**) DHF119 (**C**) DHF91 (**D**) DHF46 (**E**) DHF79 and (**F**) DHF38.

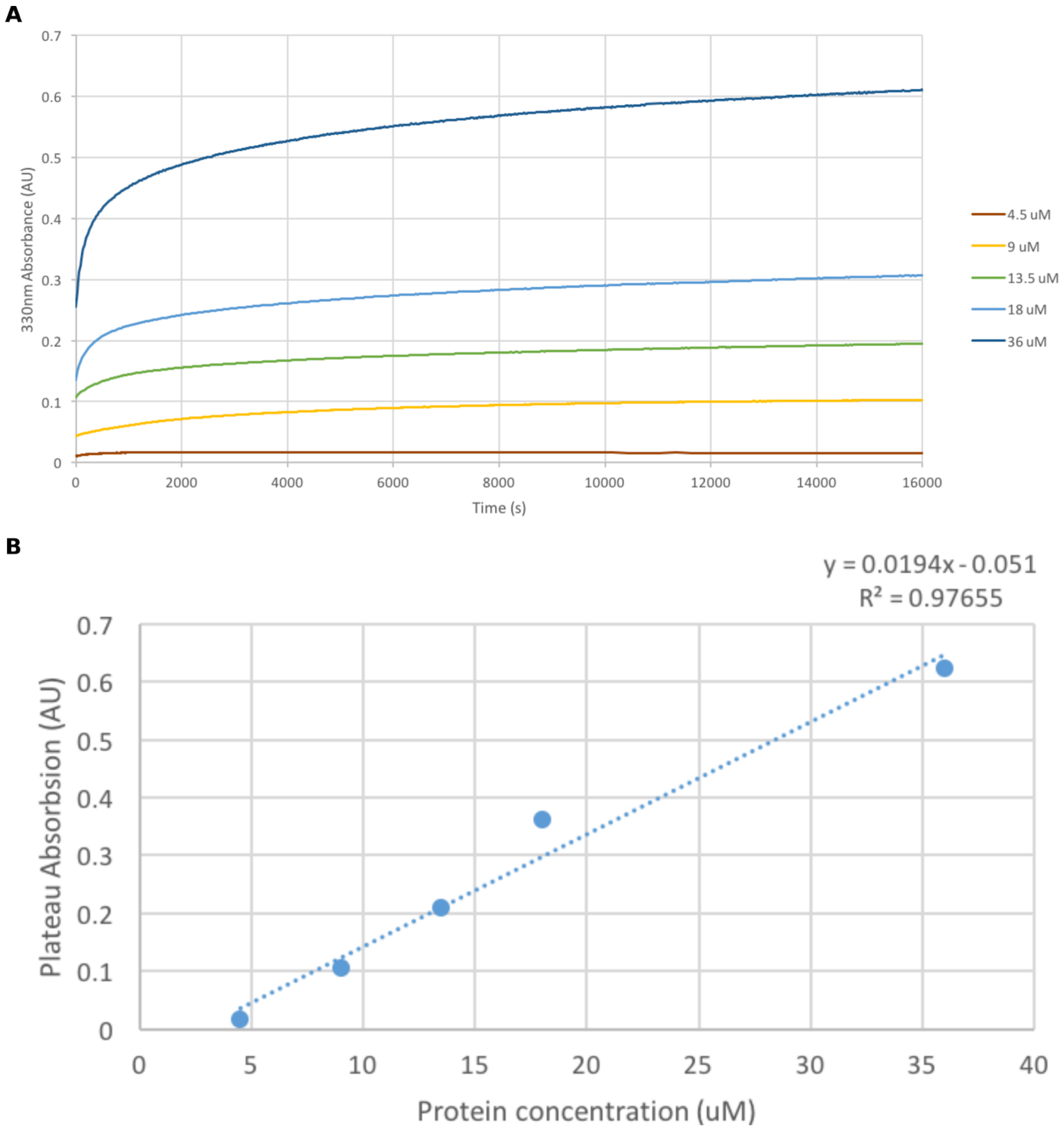
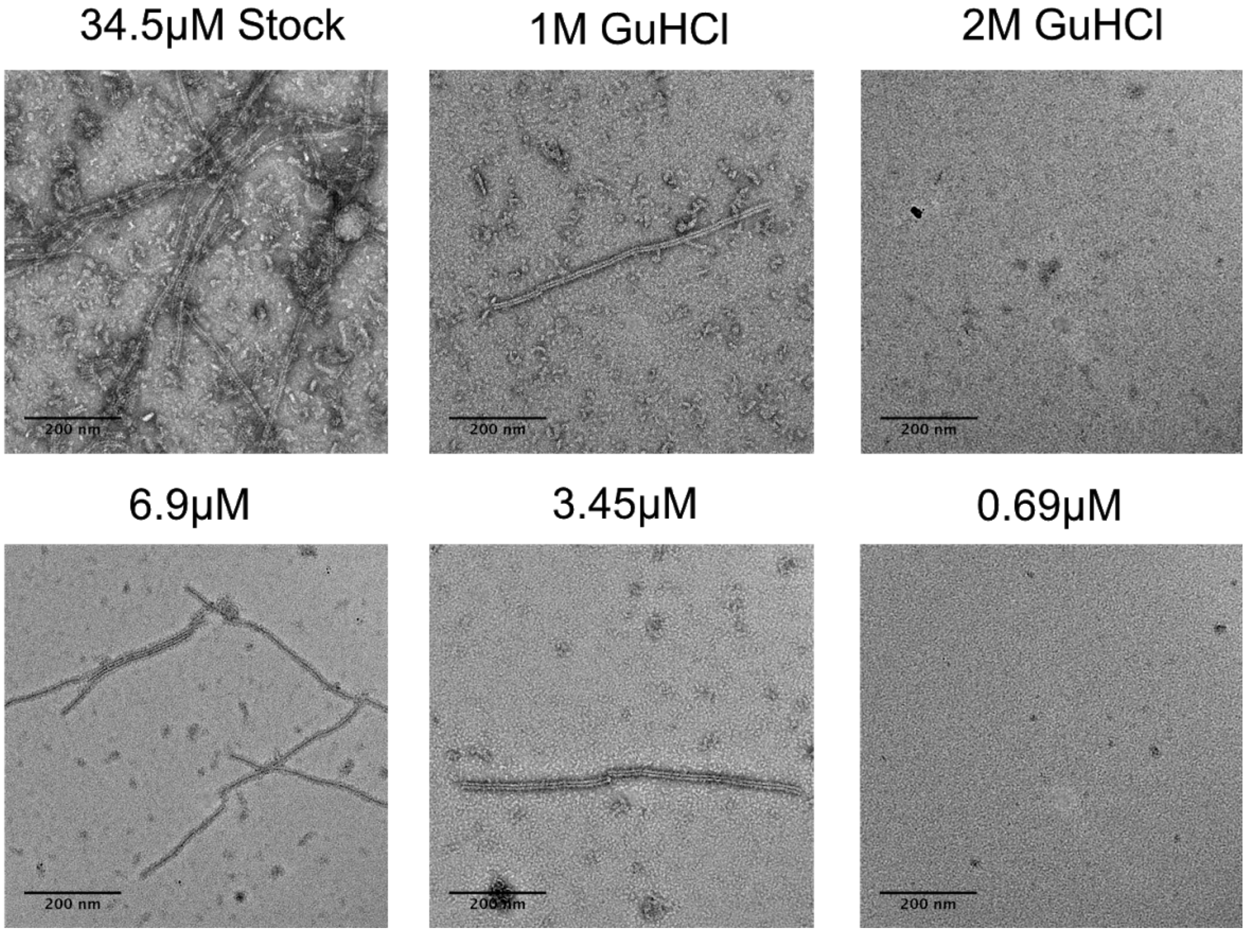


Fig. S8.

Filament assembly kinetics for DHF119. (**A**) Kinetic measurements of filament assembly by solution scattering. (**B**) Extrapolation of Critical concentration for assembly using the asymptotic values for the fits in (**A**).


34.5μM Stock

1M GuHCl

2M GuHCl

6.9μM

3.45μM

0.69μM

Fig. S9.

Concentration dependent assembly for DHF119. Top, Negative stain EM micrograph of DHF119 at 34.5 μM concentration in 25mM Tris and 75 mM NaCl (left), 1M GuHCl (middle), 2M GuHCl (right). Bottom, Negative stain EM micrograph of DHF119 at 6.9 μM (left), 3.5 μM (middle), 0.7 μM (left) concentration in 25mM Tris and 75 mM NaCl.

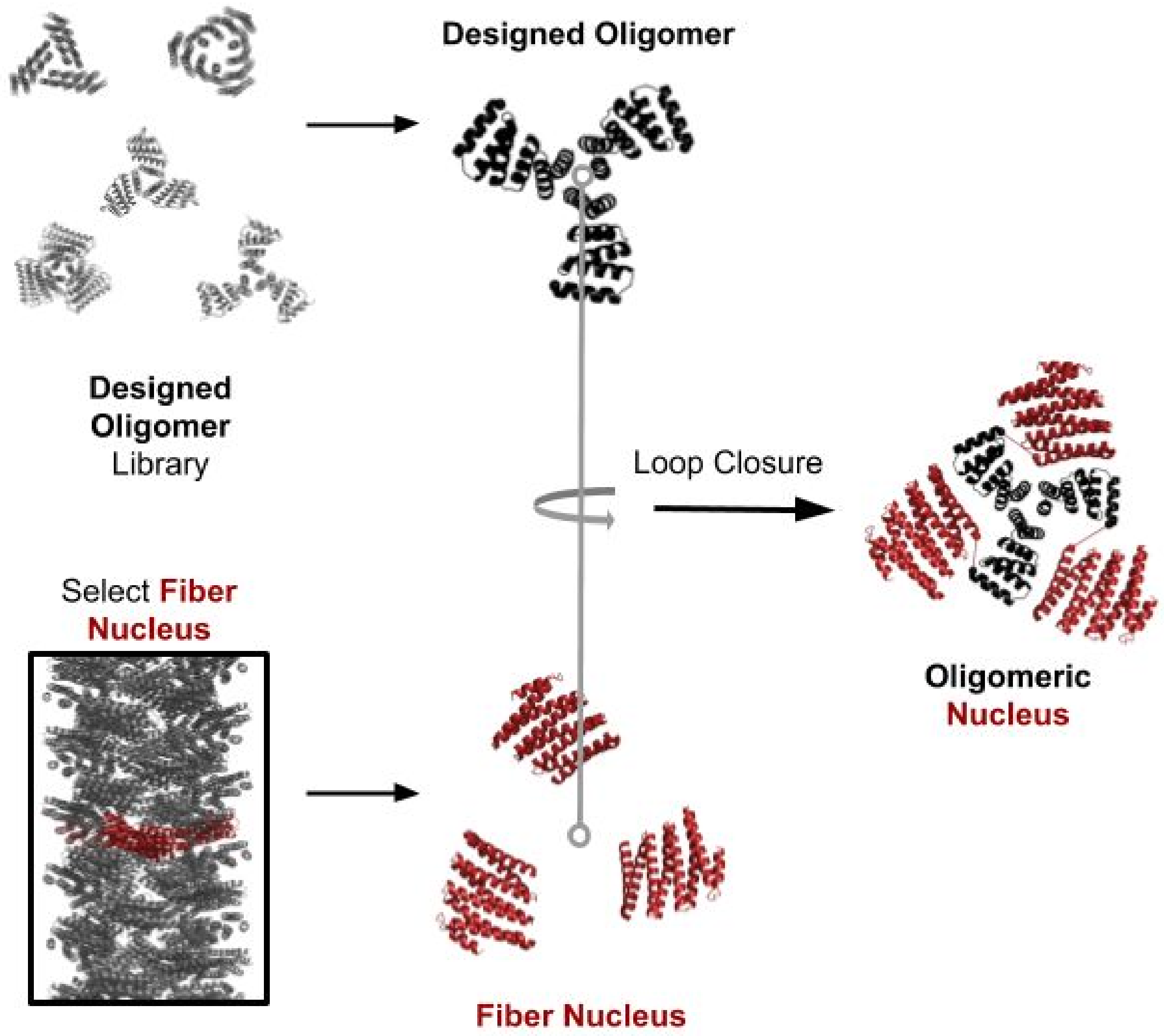


Fig. S10.

Design of Anchoring Proteins. A library of designed oligomers with cyclic symmetry around the Z axis is aligned with a layer of fiber components taken from the cryoEM structure, with the helical axis also aligned along Z. Translations and rotations around Z are applied to find the closest distance between the oligomer termini and the fiber components. These are then linked using a flexible linker and substituting the fiber component for the appropriate capping accessory protein.

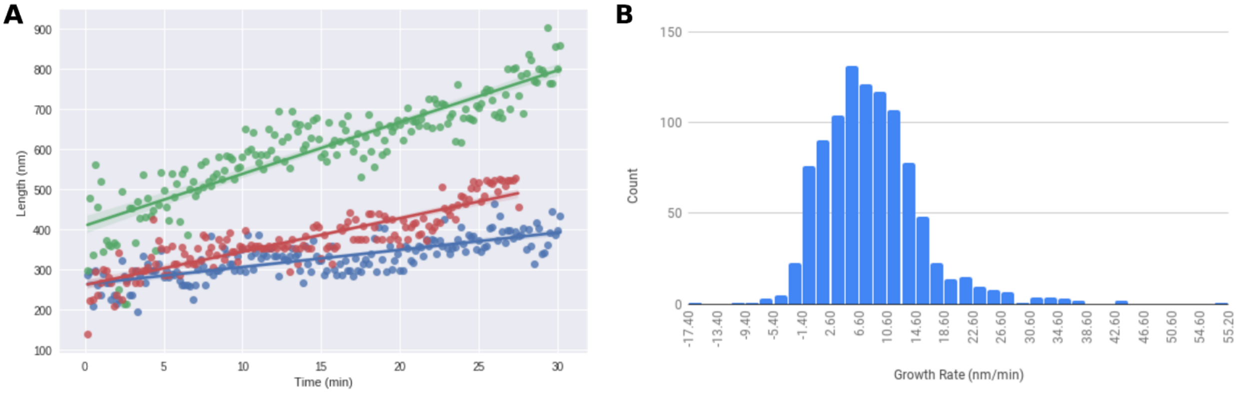


Fig. S11.

Analysis of growth kinetics of DHF119 GFP fiber at 18µM concentration from biotinylated anchor proteins DHF119_C6 immobilized on streptavidin-coated slides monitored by TIRF microscopy over 30 minutes. (**A**) Tracked length over time for 3 individual fibers. (**B**) Histogram of linear-fit growth rate for 1000 tracked fibers (8.4 nm/minute on average, with standard deviation of 7.2).

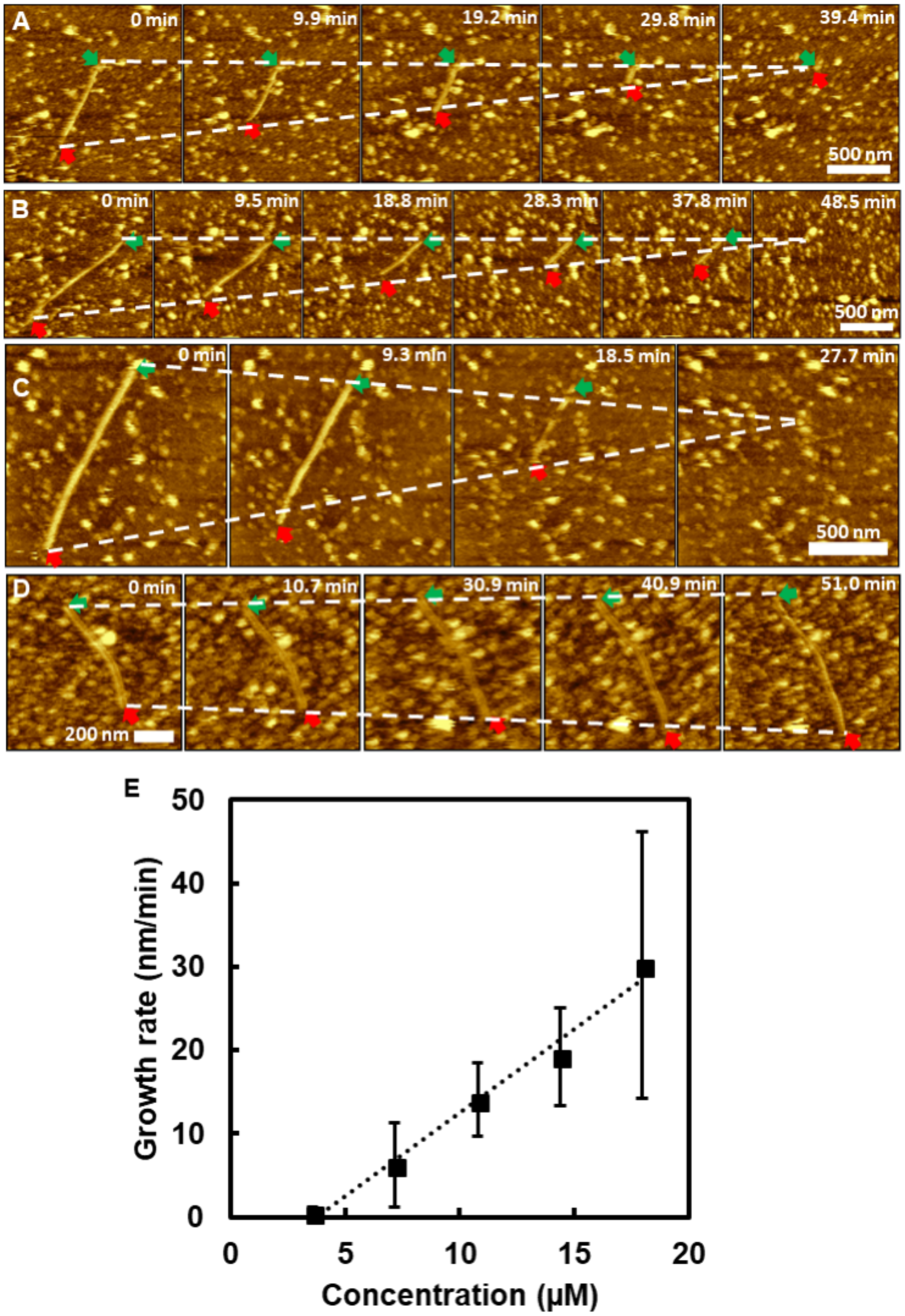


Fig. S12.

(**A**) *In situ* AFM images showing the dissolution of fiber (3.8 μM) from one end in the presence of N-caps (3.8 μM). (**B**) *In situ* AFM images showing the dissolution of fiber (3.8 μM) from one end in the presence of C-caps (3.8 μM). (**C**) *In situ* AFM images showing the dissolution of fiber (3.8 μM) from both ends in the presence of N-caps (1.9 μM) and C-caps (1.9 μM). (**D**) *In situ* AFM images showing the growth process of fibers with 7.2 μM DHF119. (**E**) Measurements of growth rates of DHF119 fiber at different concentrations shows that the growth rate changes linearly with protein concentration.

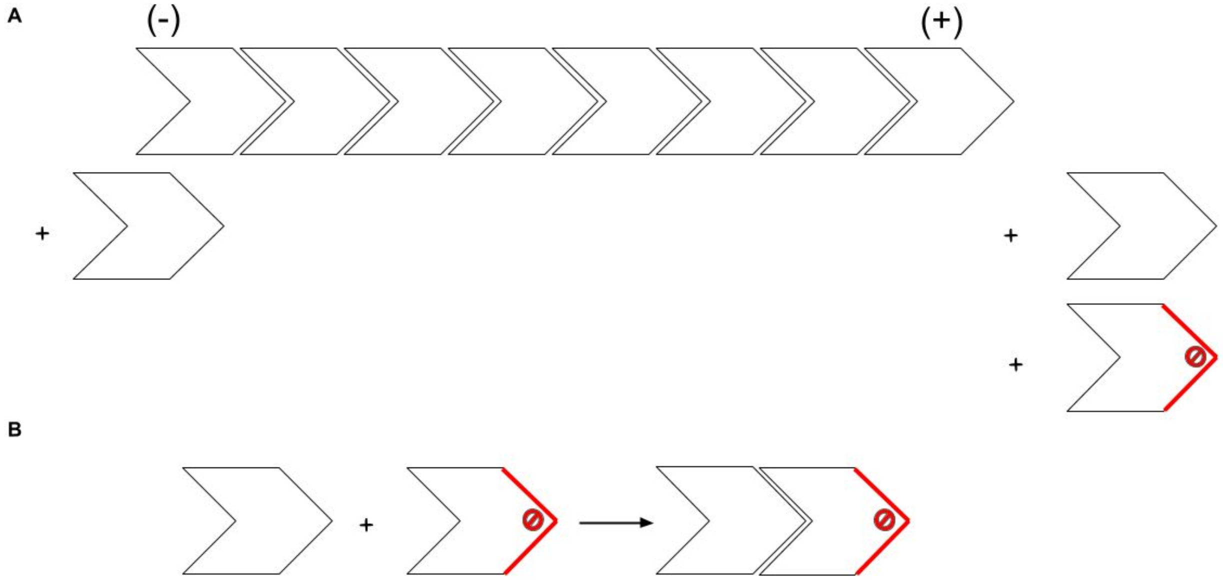


Fig. S13.

Simple model for disassembly. (**A)** At the (+) end both monomers and caps bind to the DHF while at the **(-**) end only monomers can be incorporated (**B**) Caps complex with monomers in solution effectively reducing the free monomer concentration that can bind to the (**-**) end.

Supplementary Tables

Table S1. Computational and experimental parameters for all designs tested.

Computation									Experiment		
Design	Cn	Rise(Å)	Rotation (°)	Radius (Å)	SC _1[¶]	SC_2[§]	dG_1 (R.E.U.)[†]	dG_2 (R.E.U.)[‡]	vector	fiber[*]	solubility[#]
DHF1	1	7.5	-82.7	23.5	0.9	0.7	-19.2	-10.0	pET28b	0	2
DHF2	1	7.8	130.6	27.0	0.7	0.7	-23.6	-19.3	pET28b	0	0
DHF3	1	9.4	131.1	28.6	0.9	0.7	-19.5	-10.5	pET28b	0	2
DHF4	1	7.7	-86	28.2	0.8	0.8	-25.6	-20.9	pCDB24	1	1
DHF5	1	9.0	131.9	27.0	0.7	0.7	-25.0	-22.5	pCDB24	1	1
DHF6	1	11.7	-151.4	24.3	0.8	0.7	-15.7	-7.3	pCDB24	0	2
DHF7	1	6.7	82.9	31.5	0.8	0.7	-18.9	-8.9	pCDB24	0	2
DHF8	1	14.6	94.1	19.8	0.7	0.6	-22.7	-1.6	pET28b	1	1
DHF9	1	12.1	-152	24.5	0.8	0.8	-8.7	-7.8	pET28b	1	2
DHF10	1	9.3	132.9	27.8	0.8	0.7	-18.7	-12.9	pET28b	0	2
DHF11	2	4.2	-31.5	74.1	0.6	0.6	-32.3	-30.5	pCDB24	0	0
DHF12	1	7.8	130.2	26.6	0.8	0.7	-17.7	-14.9	pCDB24	0	2
DHF13	1	11.8	-150.3	24.4	0.7	0.7	-16.8	-9.4	pCDB24	0	2
DHF14	1	11.5	84	26.3	0.7	0.7	-21.4	-14.7	pCDB24	0	2
DHF15	2	8.9	52.2	40.4	0.7	0.7	-24.8	-9.9	pET28b	0	0
DHF16	1	8.7	-109.1	28.8	0.8	0.6	-6.3	-5.6	pET28b	1	1
DHF17	1	4.1	102.3	47.3	0.7	0.6	-29.8	-21.1	pET28b	1	1
DHF18	2	10.2	-79.4	35.1	0.8	0.7	-10.9	-1.0	pCDB24	0	2
DHF19	1	7.9	130.6	26.9	0.7	0.7	-24.1	-16.3	pCDB24	0	2
DHF20	1	4.8	-139.9	44.1	0.7	0.7	-20.5	-13.9	pCDB24	1	2
DHF21	2	21.8	-102.9	22.6	0.7	0.6	-17.9	-12.6	pCDB24	0	2
DHF22	2	14.6	-71	30.2	0.7	0.7	-18.6	-15.2	pET28b	0	0
DHF23	1	6.5	83	33.1	0.8	0.8	-14.1	-12.8	pET28b	1	1
DHF24	1	3.4	48.8	63.4	0.9	0.7	-17.7	-10.6	pCDB24	0	2
DHF25	4	29.4	23.9	37.2	0.7	0.7	-19.2	-14.7	pCDB24	1	1
DHF26	1	7.9	130.1	26.7	0.9	0.7	-22.8	-15.0	pCDB24	0	1

DHF27	2	11.5	-43.2	46.8	0.7	0.7	-21.6	-18.3	pCDB24	0	2
DHF28	1	10.4	-108.6	25.5	0.7	0.6	-16.3	-7.8	pCDB24	1	1
DHF29	2	15.1	-76.3	33.3	0.8	0.7	-2.7	-2.3	pET28b	0	2
DHF30	1	8.0	131.9	27.0	0.8	0.8	-27.1	-11.9	pET28b	0	2
DHF31	1	14.2	99	19.8	0.7	0.5	-13.9	-2.7	pCDB24	1	1
DHF32	4	22.7	25.7	38.3	0.8	0.7	-12.1	-8.0	pCDB24	0	2
DHF33	3	19.9	28.6	30.7	0.7	0.7	-43.5	-8.7	pCDB24	0	2
DHF34	1	8.4	-110.3	27.2	0.8	0.7	-22.4	-19.7	pCDB24	1	1
DHF35	1	10.3	-108.9	25.2	0.7	0.7	-19.2	-16.9	pCDB24	0	1
DHF36	1	5.9	-147.2	33.7	0.8	0.7	-15.5	-6.3	pET28b	1	1
DHF37	2	14.8	-77.5	31.1	0.8	0.7	-18.2	-17.7	pET28b	0	1
DHF38	1	8.7	-87.9	30.0	0.8	0.7	-15.7	-9.7	pCDB24	1	2
DHF39	5	19.9	43.8	53.3	0.8	0.7	-18.2	-11.2	pCDB24	0	2
DHF40	1	5.4	139.9	37.6	0.7	0.7	-29.5	-6.6	pCDB24	1	1
DHF41	1	10.4	-107.9	25.9	0.7	0.7	-11.9	-9.1	pCDB24	0	2
DHF42	1	7.1	-135.6	29.6	0.8	0.8	-17.1	0.2	pET28b	0	0
DHF43	2	28.3	81.5	18.0	0.8	0.7	-17.2	-1.3	pET28b	1	1
DHF44	1	14.5	97.4	19.7	0.7	0.5	-23.6	-4.6	pCDB24	1	1
DHF45	1	8.4	-136	26.6	0.7	0.7	-13.7	-6.7	pCDB24	0	2
DHF46	1	5.0	-137.3	36.5	0.7	0.7	-13.7	-13.3	pET28b	1	1
DHF47	1	10.3	-109.1	25.4	0.8	0.8	-14.5	-12.2	pCDB24	1	1
DHF48	1	10.1	135.2	26.9	0.8	0.7	-18.6	-8.3	pET28b	1	2
DHF49	2	29.8	-99.2	18.7	0.7	0.5	-2.5	28.3	pET28b	1	1
DHF50	1	14.5	94.9	19.8	0.7	0.6	-24.7	-4.2	pCDB24	1	1
DHF51	1	10.0	135	26.7	0.8	0.7	-10.0	-3.9	pCDB24	1	2
DHF52	1	5.2	148.7	31.1	0.7	0.7	-9.0	-3.7	pCDB24	1	2
DHF53	1	7.6	-109.1	28.7	0.8	0.7	-15.9	-14.5	pCDB24	0	2
DHF54	1	14.4	94.6	19.9	0.8	0.7	-25.1	-2.4	pET28b	0	2
DHF55	2	14.8	-71	30.4	0.8	0.7	-21.2	-11.7	pET28b	0	0
DHF56	2	22.4	-108.8	22.4	0.7	0.6	-22.6	-16.9	pET28b	0	2
DHF57	2	28.5	83	16.2	0.8	0.7	-30.5	-16.6	pET28b	0	2
DHF58	5	36.7	3.1	33.6	0.8	0.7	-17.4	-5.1	pET28b	1	2
DHF59	1	7.1	-134.8	30.0	0.8	0.8	-23.2	-11.2	pET28b	0	2
DHF60	1	14.0	98.7	22.8	0.7	0.7	-29.6	-4.3	pET28b	0	0

DHF61	2	13.3	-67.3	33.3	0.8	0.7	-15.7	-13.4	pCDB24	0	2
DHF62	2	28.0	-100.8	23.2	0.8	0.7	-16.0	-4.9	pET28b	1	2
DHF63	2	9.4	51.3	38.4	0.8	0.7	-17.9	-17.8	pET28b	0	0
DHF64	3	27.8	-95.1	27.4	0.8	0.7	-11.2	-9.4	pET28b	0	2
DHF65	1	5.4	-151.4	35.6	0.8	0.7	-20.1	-19.1	pET28b	0	2
DHF66	1	13.4	-134.1	20.3	0.8	0.7	-14.6	-2.2	pET28b	0	2
DHF67	1	14.4	-148.7	19.8	0.8	0.6	-20.0	-6.1	pET28b	0	2
DHF68	1	13.2	-151.7	23.2	0.8	0.7	-19.0	-8.6	pET28b	0	2
DHF69	1	8.9	-107.3	28.8	0.8	0.7	-8.5	-8.0	pET28b	0	2
DHF70	1	6.8	-85.2	32.8	0.8	0.7	-14.6	-13.7	pET28b	0	2
DHF71	1	4.8	57.2	36.8	0.7	0.6	-27.0	-17.8	pET28b	0	0
DHF72	1	9.2	88.6	27.1	0.7	0.7	-19.8	-11.1	pET28b	0	2
DHF73	1	7.8	-85.8	28.0	0.8	0.7	-24.6	-12.8	pET28b	0	2
DHF74	2	17.2	60.8	26.8	0.7	0.7	-15.5	-0.7	pET28b	0	2
DHF75	4	16.3	35.5	49.9	0.7	0.7	-18.0	-15.2	pET28b	0	2
DHF76	1	5.3	-151.9	35.5	0.8	0.7	-19.6	-6.1	pET28b	1	2
DHF77	1	7.4	135.2	27.7	0.8	0.7	-13.2	-8.7	pET28b	1	1
DHF78	1	12.6	-147.5	24.6	0.8	0.7	-24.5	-17.8	pET28b	1	2
DHF79	2	7.7	-41.8	58.1	0.7	0.7	-16.2	-12.4	pET28b	1	2
DHF80	1	4.1	79.1	60.7	0.8	0.7	-15.0	-9.7	pET28b	0	2
DHF81	1	6.5	-86.7	30.4	0.8	0.7	-14.0	-7.0	pET28b	0	2
DHF82	1	4.5	-84	46.9	0.8	0.7	-21.2	-0.2	pET28b	1	2
DHF83	2	9.7	50.1	38.5	0.8	0.7	-22.4	-8.9	pET28b	0	0
DHF84	1	5.9	68	34.1	0.8	0.7	-19.4	-10.2	pET28b	0	0
DHF85	2	12.2	65.7	33.6	0.8	0.8	-20.5	-14.9	pET28b	0	0
DHF86	2	13.2	70.8	30.1	0.7	0.7	-11.8	-11.3	pET28b	0	0
DHF87	2	14.2	-109.6	34.6	0.7	0.7	-20.4	-8.6	pET28b	0	2
DHF88	2	7.0	31.5	65.1	0.8	0.7	-14.9	-13.0	pET28b	0	2
DHF89	3	9.4	35.6	65.5	0.7	0.7	-24.9	-12.7	pET28b	0	2
DHF90	3	29.0	-85.9	24.3	0.7	0.7	-13.1	-12.8	pET28b	0	2
DHF91	3	24.2	48.8	24.4	0.7	0.7	-27.2	-21.8	pET28b	1	2
DHF92	3	22.4	-84.6	32.0	0.8	0.7	-11.6	-7.9	pET28b	0	2
DHF93	3	22.9	-84.4	32.0	0.8	0.8	-14.7	-10.5	pET28b	0	2
DHF94	3	21.7	40.8	31.8	0.7	0.7	-16.8	-14.1	pET28b	0	2

DHF95	4	19.1	54.3	37.4	0.7	0.7	-15.8	-14.4	pET28b	0	2
DHF96	4	37.1	-4.8	26.8	0.8	0.8	-22.7	-10.8	pET28b	0	2
DHF97	4	36.5	85.8	27.0	0.7	0.7	-22.1	-21.1	pET28b	0	2
DHF98	4	36.7	-4.1	27.1	0.7	0.7	-22.8	-16.7	pET28b	0	2
DHF99	1	5.6	-149.4	33.4	0.7	0.7	-20.3	-13.6	pET28b	0	0
DHF100	1	4.1	-156.8	40.9	0.8	0.6	-17.1	-10.8	pET28b	0	2
DHF101	1	5.7	-149.3	33.5	0.8	0.7	-9.2	-4.4	pET28b	0	2
DHF102	1	4.1	-156.8	41.5	0.8	0.7	-17.6	-16.5	pET28b	0	2
DHF103	1	3.9	-157.1	47.2	0.8	0.7	-23.1	-12.0	pET28b	0	2
DHF104	1	8.3	131.5	27.5	0.8	0.6	-18.9	-9.7	pET28b	0	2
DHF105	1	8.4	132.7	27.5	0.7	0.7	-15.3	-1.6	pET28b	0	2
DHF106	1	8.4	131.9	29.4	0.8	0.7	-16.9	-12.9	pET28b	0	2
DHF107	3	11.0	40.9	69.1	0.8	0.7	-20.7	-17.7	pET28b	1	2
DHF108	1	12.0	-149.2	24.9	0.8	0.6	-25.8	-20.7	pET28b	0	2
DHF109	1	11.8	-150	24.8	0.8	0.8	-24.6	-21.5	pET28b	0	0
DHF110	1	13.3	-152.6	23.0	0.8	0.7	-14.0	-11.5	pET28b	0	2
DHF111	1	5.1	-138.5	36.5	0.8	0.7	-15.9	-0.9	pET28b	0	2
DHF112	1	6.4	-145.3	30.4	0.8	0.7	-17.3	-9.3	pET28b	0	2
DHF113	1	12.6	-148.3	24.5	0.8	0.8	-23.3	-15.7	pET28b	0	2
DHF114	1	12.4	-146.5	24.8	0.8	0.7	-25.8	-0.6	pET28b	0	2
DHF115	1	8.8	-150.3	29.9	0.7	0.7	-12.6	-6.6	pET28b	0	2
DHF116	1	14.8	-152.9	19.0	0.7	0.7	-31.7	-8.3	pET28b	0	2
DHF117	2	8.1	-42	42.6	0.7	0.7	-16.5	-7.8	pET28b	0	2
DHF118	1	5.3	107.3	48.4	0.8	0.7	-15.3	-9.9	pET28b	0	2
DHF119	1	4.5	-106.4	46.4	0.7	0.7	-11.9	-9.0	pET28b	1	2
DHF120	1	8.6	-110.6	27.5	0.7	0.7	-19.6	-16.3	pET28b	0	2
DHF121	1	9.2	-107	28.9	0.9	0.7	-15.6	-2.4	pET28b	0	2
DHF122	1	4.6	-81.3	43.4	0.8	0.7	-29.7	-9.3	pET28b	0	2
DHF123	2	9.6	50.1	38.7	0.8	0.8	-17.9	-13.7	pET28b	0	0
DHF124	1	12.3	-147.1	24.7	0.8	0.7	-30.4	-16.3	pET28b	0	0

[¶] Shape complementarity of main interface

[§] Shape complementarity of secondary interface

[†] Computed Binding energy of main interface

[‡] Computed Binding energy of secondary interface

[*] 0: no filaments observed by negative stain EM 1: filaments observed by EM

[#] 0: no expression, 1:insoluble expression 2: soluble expression

Table S2. CryoEM data collection and refinement

	DHF119 (EMD-9021, PDB 6E9Z)	**DHF38** (EMD-9020, PDB 6E9Y)	**DHF58** (EMD-9017, PDB 6E9T)	**DHF46** (EMD-9016, PDB 6E9R)	**DHF79** (EMD-9018, PDB 6E9V)	**DHF91** (EMD-9019, PDB 6E9X)
Data collection						
Microscope	Titan Krios	Titan Krios	TF20	TF20	TF20	TF20
Voltage	300 kV	300 kV	200 kV	200 kV	200 kV	200 kV
Electron detector	K2 Summit	K2 Summit	K2 Summit	K2 Summit	K2 Summit	K2 Summit
Electron dose (e^-/Å^2)	90	90	45	45	45	45
Pixel size (Å)	1.05	1.05	1.26	1.26	1.26	1.26
Reconstruction						
Point group symmetry	C3	C1	C2	C4	C1	C3
Refined helical symmetry (twist) (deg)	43.5023	-88.17	40.93287	-51.23024	77.6764	50.1109
Refined helical symmetry (rise) (Å)	14.3312	8.29531	9.10835	21.632	5.07952	24.1338
Particles	63,067	112,593	76,994	36,692	32,583	15,670
Resolution (0.143 fsc)(Å)	3.4	4.3	5.4	5.9	6.9	7.8
Reconstruction software	Relion	Relion, Frealign	SPIDER	SPIDER	Relion	Relion

Chapter 2: *De novo* design of multi-component self-assembling helical filaments

Abstract

We extend the computational approach previously developed for designing self-assembling helical filaments from monomeric to heterodimeric proteins and use it to design two-component proteins that assemble into micron scale helical filaments with a wide range of geometries. CryoEM structures of three designs are close to the computational design models. The assembly of the filaments can be initiated by mixing its two components and tuned by varying length of complimentary blocking units. The ability to generate multicomponent filaments opens up possibilities for the fabrication of controllable protein nanomaterials tailored to specific applications.

Introduction

Self-assembling protein filaments serve functions ranging from cellular transportation, cell division and change of shape. To achieve higher order of control of assembling and functionality, nature has evolved multicomponent microtubule consist of α- and β-tubulin dimers (*6*). Such controllable properties are of great nanomaterial application interest.

Although there has been progress in designing two component peptide-based filaments (*38*), the irreversible assembling process limits its applications. On the other hand, recent advances in computational protein design have generated reversible single components self-assembling protein filaments (*39*) and heterodimeric proteins that bind to each other specifically (*40*), which provide the basis for designing multicomponent high order assembly.

Here we apply the computational method to design self-assembling helical filament from monomeric protein to hetero-dimeric protein to generate multi-component reversible filament that allows precise control of the timing of assembly and enhanced modularity through independently addressable subunits.

Results

We use the design model for heterodimeric proteins (DHD131, DHD37_1:234, DHD127 and DHD15) (*40*)) as starting scaffold and applied computation method described in Shen *et al* (*39*) to design interface for self-assembling protein filaments.

We generate 195,000 helical filament backbones and selected 55 designs for experimental testing [we refer to these as de novo–designed 2-component helical filaments (D2HFs)].

The 2-component designs were co-expressed in Escherichia coli under the control of a T7 promoter and purified by immobilized metal affinity chromatography (IMAC, only one of the components contains hexahistidine tag). The non-histagged component for 45 of the 55 selected designs was fused to green fluorescent protein (GFP) to facilitate screening and kinetic measurement (Supplementary Table 1).

Of the 55 selected designs, 43 were well-expressed and co-recovered in the IMAC eluate. IMAC eluates were concentrated, and filament formation was monitored by negative stain electron microscopy (EM). A total of six designs were found to form 1D nanostructures over 200 nm (Fig. 1).

We chose three designs with a range of model architectures and highly ordered negative stain EM morphologies for higher resolution structure determination by cryo-electron microscopy (cryoEM). We determined the filament structures and refined helical symmetry parameters using iterative helical real space reconstruction in SPIDER (*21*, *22*), followed by further 3D refinement in Relion (*23*) and Frealign (*24*). The structures were solved to 3.5 Å resolution. In all three cases, the overall orientation and packing of the monomers in the filament were similar in the experimentally determined structures and design models, but there was considerable variation in the accuracy with which the details of the interacting interfaces were modeled (Fig. 2). Two of the three designed filaments matched the computational models at near-atomic resolution: for D2HF20 and D2HF30 the experimentally observed rigid body orientation was nearly identical to the design models (1.4 Å and 2.0 Å r.m.s.d. over three chains containing all unique interfaces, the backbone and side chain conformations at the subunit interfaces are very similar to those in the design model), for D2HF3 the r.m.s.d. over three chains was 6 Å. For D2HF3, the r.m.s.d over one designed interface was 1 Å while the other 5.5 Å. Higher resolution cryoEM map shows that the main interface deviation was due to the backbone deviation where the design model and crystal structure of the scaffold DHD131 disagree (Fig. S1). As our computational method assumes rigid backbone, deviation at backbone level would cause deviation at the interface level.

We expressed and purified D2HF30's A and B components separately, and then mix them to initiate fiber assembly. Confirmed by negative stain EM, neither component A or B formed filaments, and filaments started to form after mixing (Fig. 3).

To study the kinetics of D2HF30 filament formation *in vitro* in more detail, we attached some component B to glass slides, mix GFP labeled component A with unlabeled component B and monitored fiber formation by total internal reflection (TIRF) microscopy. At 1uM low concentration, fiber growth into one micron in length was observed over an hour (Fig. 3, Movie S1).

We sought to control the assembly process by expressing each component with its corresponding complementary wild-type pair without fiber interface (here referred to as caps). The caps could be truncated to different length to facilitate assembly for the correct fiber pair at different rates (Fig. 5 A and B). When capped component A and B were mixed, fiber formation was found to be temperature-dependent (Fig. 5C), as forming the correct pairs requires energy input to remove the cap.

Discussion

The ability to form accurate micron scale multicomponent filaments is an advance for computational protein design. Fiber assembly initiated by mixing different components as wells as temperature capped fiber allows more precise control of the assembly. The different fiber components ready for fusion to different targets. This highly controlled multicomponent protein assembly opens up application ranging from diagnostic to making smart nanomaterials.

Main Text Figures

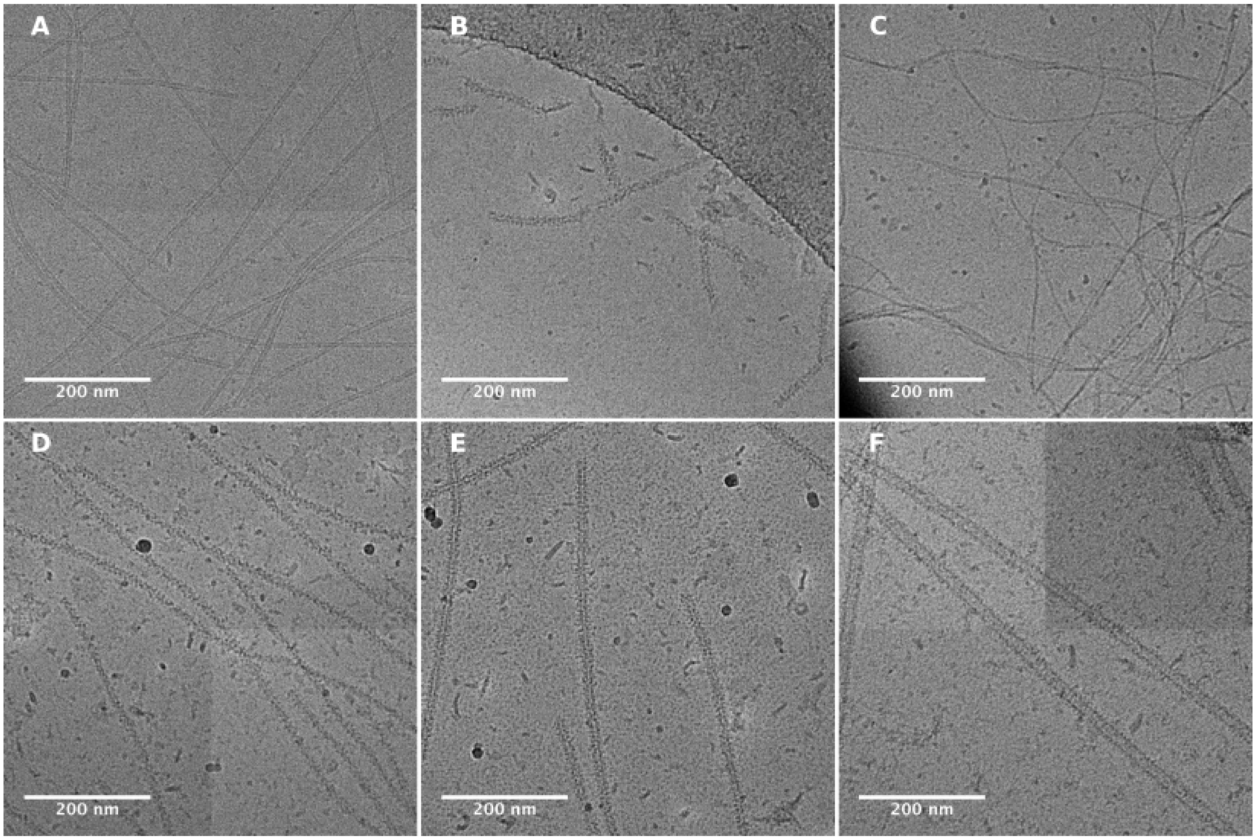


Fig. 1. Cryo Electron Micrograph of 2-component filaments. (**A**) D2HF3 (**B**) D2HF16 (**C**) D2HF18 (**D**) D2HF20 (**E**) D2HF30 (**F**) D2HF41

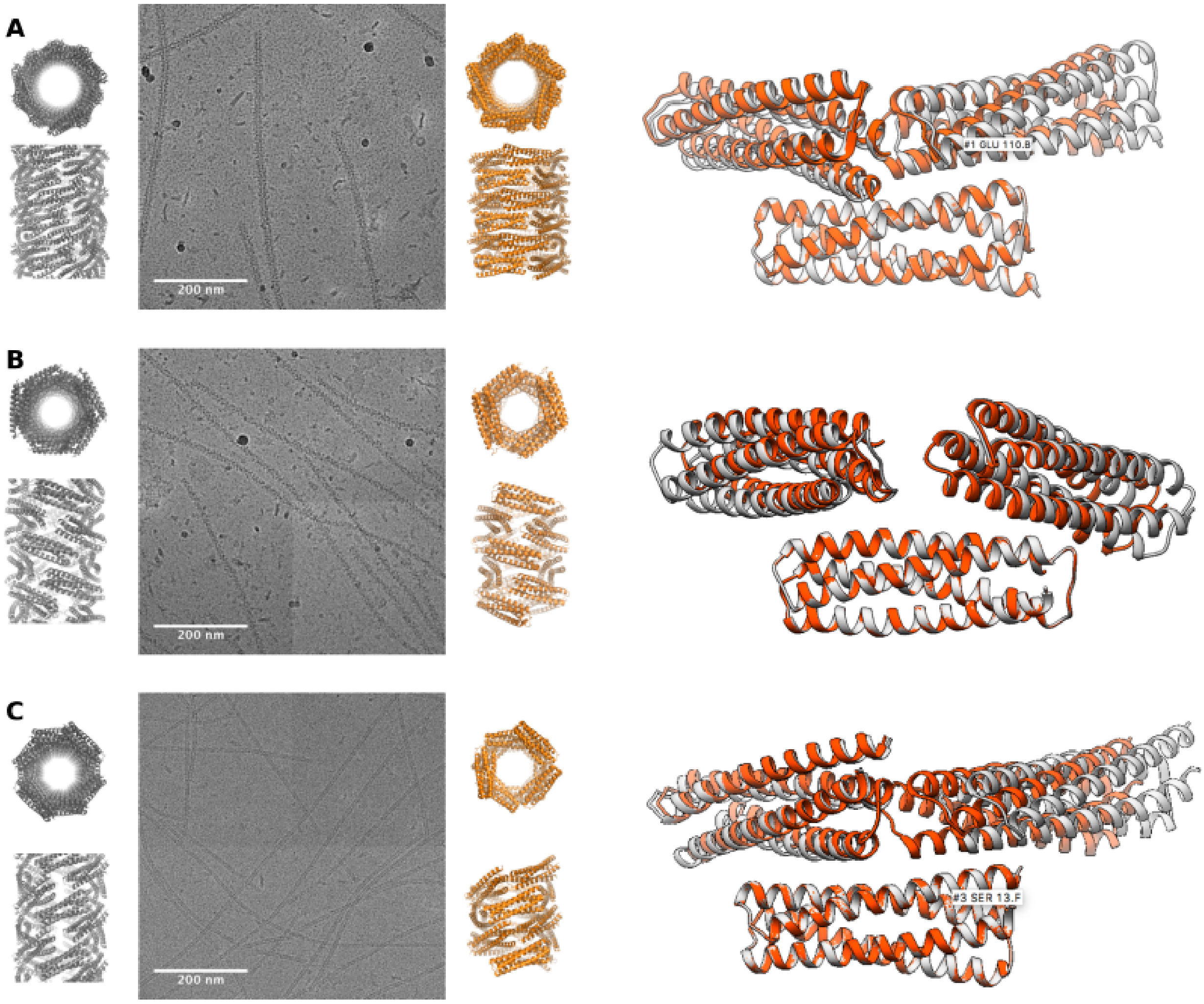


Fig. 2. CryoEM structure determination. Computational model (first panel), representative filaments in cryoEM micrographs (second panel), cryoEM structure (third panel) and overlay between model and structure (fourth panel) for (**A**) D2HF30 - r.m.s.d. 1.4 Å (**B**) D2HF20 - r.m.s.d. 2.0 Å (**C**) D2HF3 - r.m.s.d. 6.0 Å

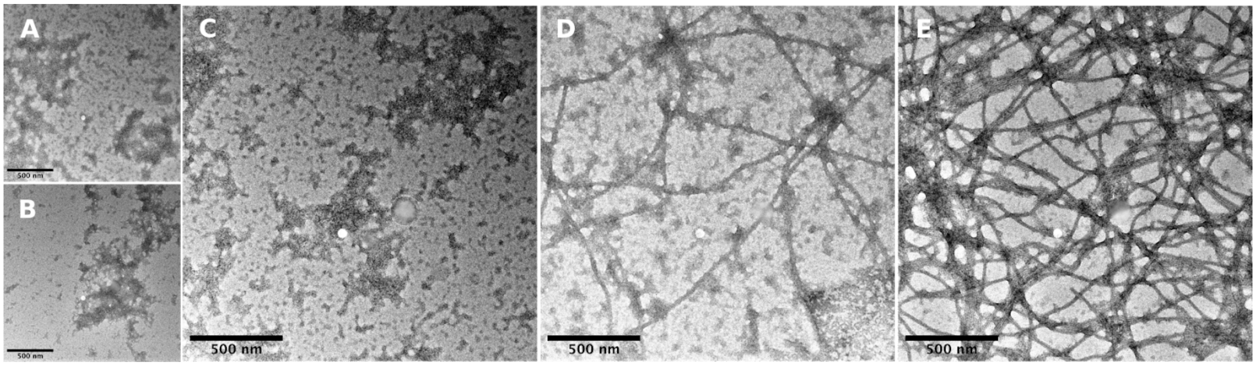


Fig. 3. Fiber formation by mixing components. Design D2HF30's component (**A**) and (**B**) were mix at 95uM for 10 minutes (**C**), 5 hours (**D**) and 1 day (**E**), characterized by negative stain electron micrograph.

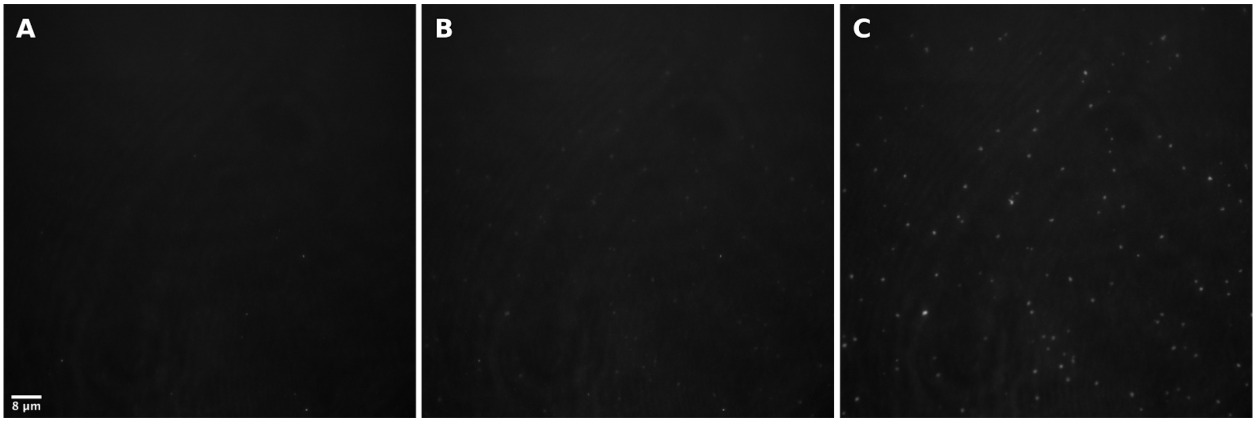


Fig. 4. Kinetics of filament assembly. D2HF30's A and B component were mixed for (**A**) 0, (**B**) 30 minutes and (**C**) 1 hour, monitored by TIRF.

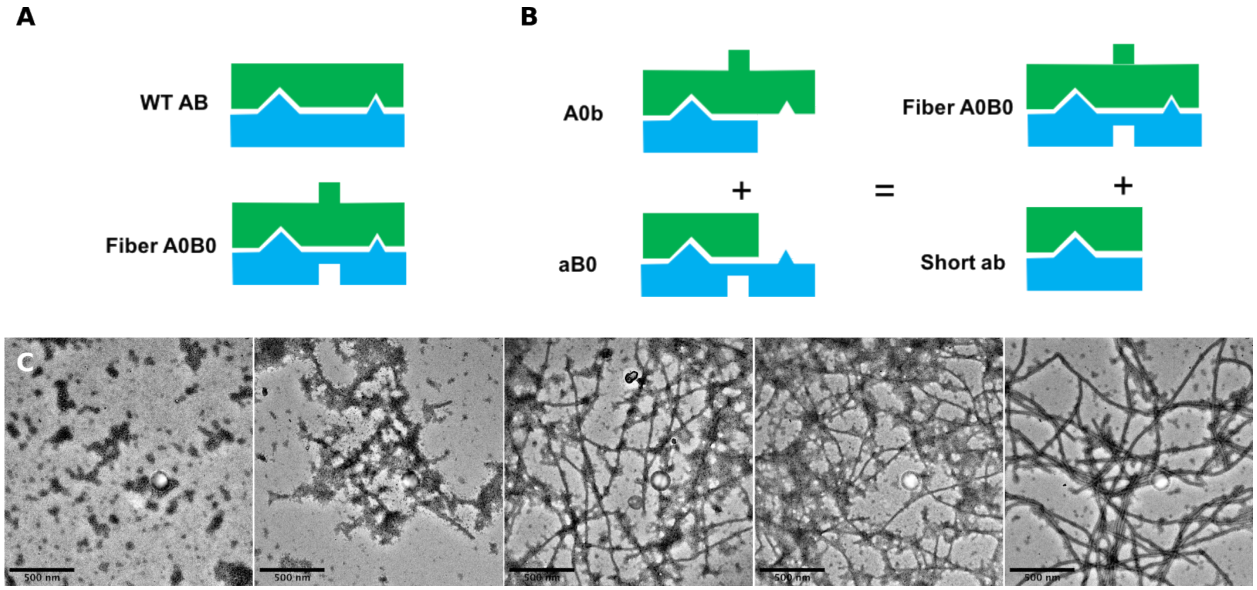


Fig. 5. Fiber formation by mixing components coexpressed with caps. (**A**) Schematic representation for the scaffold DHD (WT) compared with fiber design. (**B**) Schematic representation for mixing capped components. (**C**) Design D2HF30's component A and B with caps were mixed at 230uM for 20 minutes at room temperature, 37 °C, 42 °C, 50 °C and 65 °C, characterized by negative stain electron micrograph.

Materials and Methods

Protein Expression and Purification

For the expression of heterodimers, both monomers were encoded in the same plasmid, separated by a ribosome binding sequence (GAAGGAGATATCATC). Synthetic genes were ordered from Integrated DNA Technologies, Inc.(Coralville, IA., USA) and delivered in pET29b+ E. coli expression vector, inserted between the NdeI and XhoI sites. These plasmids were cloned into BL21* (DE3) (Invitrogen) E. coli competent cells. Transformants were inoculated into 50 ml of TB medium with 200 mg L-1 kanamycin. Expression proceeded for 24 hours at 37 °C following the expression via Studier autoinduction (*27*) until the cultures were harvested by centrifugation. Cell pellets were resuspended in TBS and lysed using the Bugbuster detergent (Millipore). The soluble fraction upon lysate clarification by centrifugation was purified by Ni2+ immobilized metal affinity chromatography with Ni-NTA Superflow resin (Qiagen). Resin with bound cell lysate was washed with 10 column volumes of 40 mM imidazole and 500 mM NaCl and eluted with 400 mM imidazole and 75 mM NaCl. Both the soluble and insoluble fractions were run on an SDS-PAGE gel. Samples that showed protein bands at the correct molecular weight were selected for screening by electron microscopy. Selected designs were expressed at the 0.5 L scale to carry out further characterization. Expression proceeded for 24 hours at 37 °C following the expression via Studier autoinduction (*27*) until the cultures were harvested by centrifugation. Cell pellets were resuspended in TBS and lysed by microfluidization. Purification was carried out as described above.

Accessory Protein Design

Capping units for D2HF30's component A and B were designed by truncating a heptad from the N and C terminus of the wild-type B and A component for scaffold DHD131 (*40*).

Filament Growth In Vitro

PEG-silane coated glass coverslips were attached to similarly-coated slides with strips of double-stick tape to make flow chambers. All incubations were at 25°C. Dry glass chambers were coated for 2 minutes with 8 mg/ml kappa-casein (Sigma C0406) 10:1 biotinylated casein in BRB80 (80 mM PIPES-KOH ph 6.85 + 1 mM $MgCl_2$ + 1 mM EGTA), washed twice with CK buffer (BRB80 + 1 mg/ml casein + 70 mM KCl), incubated 3 minutes with 0.5 mg/ml neutravidin (Molecular Probes A2666) in CK, then washed three times with CK. Prepared cells were washed once in IB (imaging buffer: 75 mM NaCl + 25 mM Tris-HCl pH 8.0 + 11 mM glucose + 2.5 mM DTT + 0.2 mg/ml glucose oxidase (Sigma G2133) + 0.04 mg/ml catalase (Sigma C40). Biotinylated D2HF component B protein 75 nM in IB was incubated in chamber for 3 minutes, chamber washed twice with IB, replaced with 32uM Preform D2HF fiber to help focus, chamber washed twice with IB again, and replaced with 1μM mixture of 80°C-preheated GFP labeled component A with unlabeled component B in IB for observation of assembly. Imaging was carried out using a Personal Deltavision microscope (GE Healthcare) outfitted with 4-laser TIRF capabilities, Olympus 60×, 1.49 NA TIRF objective and Ultimate focus (Applied Precision) at 37 °C.

Supplementary Figures

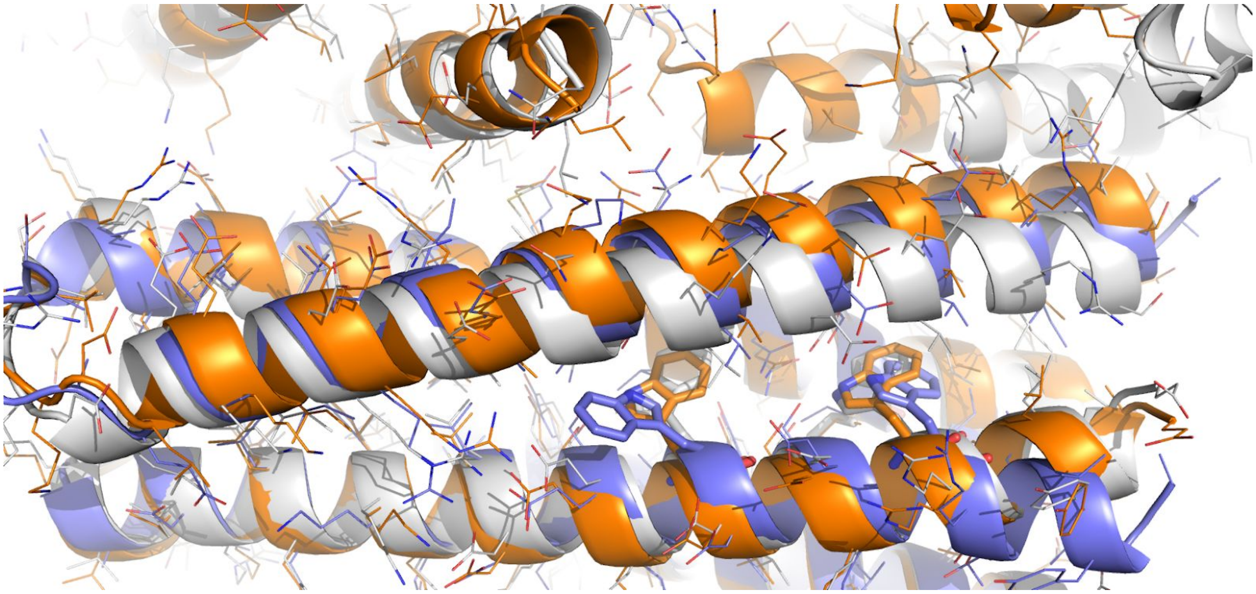


Fig. S1.
Comparison of the deviated interface for D2HF3. Aligned on one subunit are the D2HF3 design model (grey), cryoEM structure (orange) and crystal structure for its scaffold DHD131 (purple) (PDB id:6DKM)

Chapter 3: *De novo* design of pH-responsive self-assembling helical filaments

Abstract

We apply computational approach previously developed for designing self-assembling helical filaments to design pH-responsive proteins that assemble into micron scale helical filaments. CryoEM structures of two designs are close to the computational design models. The filaments can reversibly assemble at neutral pH and disassemble at acidic pH. The ability to generate pH-responsive filaments opens up possibilities for functional protein nanomaterials.

Introduction

PH is a very effective way to control protein assembly. For example, spider silk protein adopts a pH-sensitive relay to control the generation of solid silk (*41*). Designed protein-based pH-responsive polymers are a promising class of new polymer materials with potential applications in fields such as tissue engineering, drug- and gene delivery or self-healing biomaterials. There has been progress in making pH-responsive assembling nanomaterial by mimicking silk protein domain (*42*). However, the de novo design of pH-responsive filament with broader geometries and higher control is an unmet challenge. Recent advances in computational protein design have generated reversible single components self-assembling protein filaments (*39*) and pH-responsive proteins with designed histidine network buried at protein core (*43*), which provides the basis for designing pH-responsive high order assembly.

Here we apply the computational method to design self-assembling helical filament from monomeric protein to pH-responsive protein to generate reversible pH-responsive self-assembling protein filament that allows effective control of the assembly.

Results

We use the crystal structure for pH-responsive trimeric helical bundle pRO-2.3 (PDB ID: 6MSQ) as starting scaffold and applied computation method described in Shen *et al* (*39*) to design interface for self-assembling protein filaments. We generate 45,000 helical filament backbones and selected 18 designs for experimental testing [we refer to these as de novo–designed pH-responsive helical filaments (DpHFs)].

A linked single chain version of the homo-trimeric helical bundle was used for testing to display fiber interfaces at the designed orientation. The designs were expressed in Escherichia coli under the control of a T7 promoter and purified by immobilized metal affinity chromatography (IMAC).

All 18 selected designs were well-expressed and recovered in the IMAC eluate. IMAC eluates were concentrated, and filament formation was monitored by negative stain electron microscopy (EM) with Nano-W stain (pH 6.8). A total of two designs (DpHF7 and DpHF18) were found to form 1D nanostructures (Fig. 1).

We characterized the pH responsiveness of DpHF7 and DpHF18 with two methods.

For DpHF7, while filaments formation was observed when stained with neutral pH stain Nano-W (Fig. 2A), no filaments could be observed when stained with Uranyl Formate (pH 4.0) (Fig. 2B) and subsequence water and buffer washed follow by staining with Nano-W recover the filament formation (Fig. 2C). As the staining and wash process happened in a minute, DpHF7 shows fast and reversible pH response.

For DpHF18, after concentrated in TBS buffer at pH 8, filaments formation was observed when stained with neutral pH stain Nano-W (Fig. 2D). When DpHF18 was concentrated in NaCitrate buffer at pH 4, no fiber could be observed with negative stain EM (Fig. 2E). When the same solution was diluted 1:10 into TBS buffer at pH 8 for a minute and then stained with Nano-W, filaments were recovered (Fig. 2F). Compared with DpHF7, DpHF18 was slower at filament disassembly at low pH but recover fast at high pH.

Discussion

The ability to form micron scale reversibly pH-responsive filaments is an advance for computational protein design. Designed filaments can reversibly assemble at neutral pH and disassemble at acidic pH at different rates. The different pH-responsive properties of the different design filaments suit unique applications. The effective control over filaments assembly and disassembly opens up possibilities ranging from drug delivery to building responsive nanomaterials.

Main Text Figures

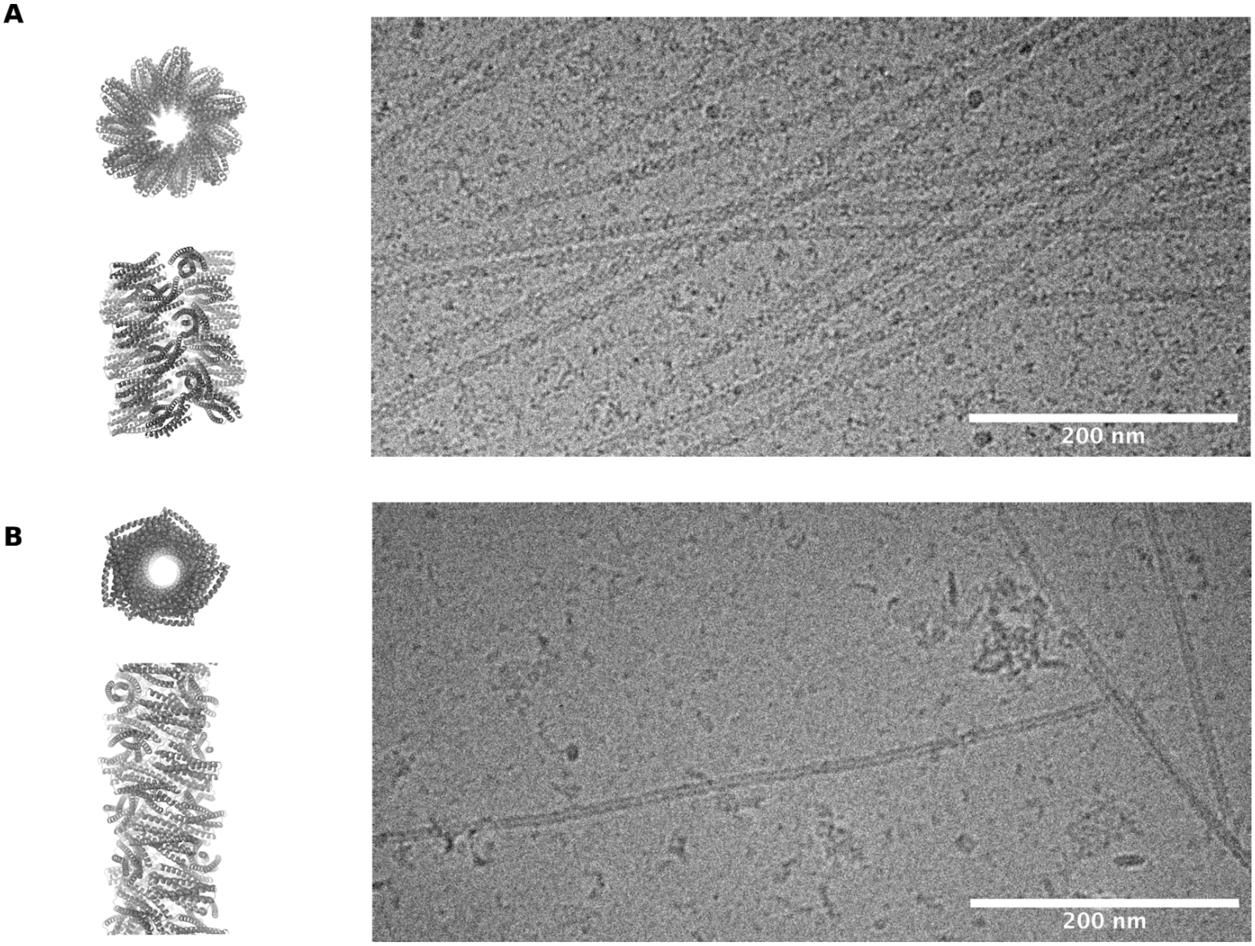


Fig. 1. Design model and cryoEM of filaments. Computational model (first panel) and representative filaments in cryoEM micrographs (second panel) for (**A**) DpHF7 (**B**) DpHF18

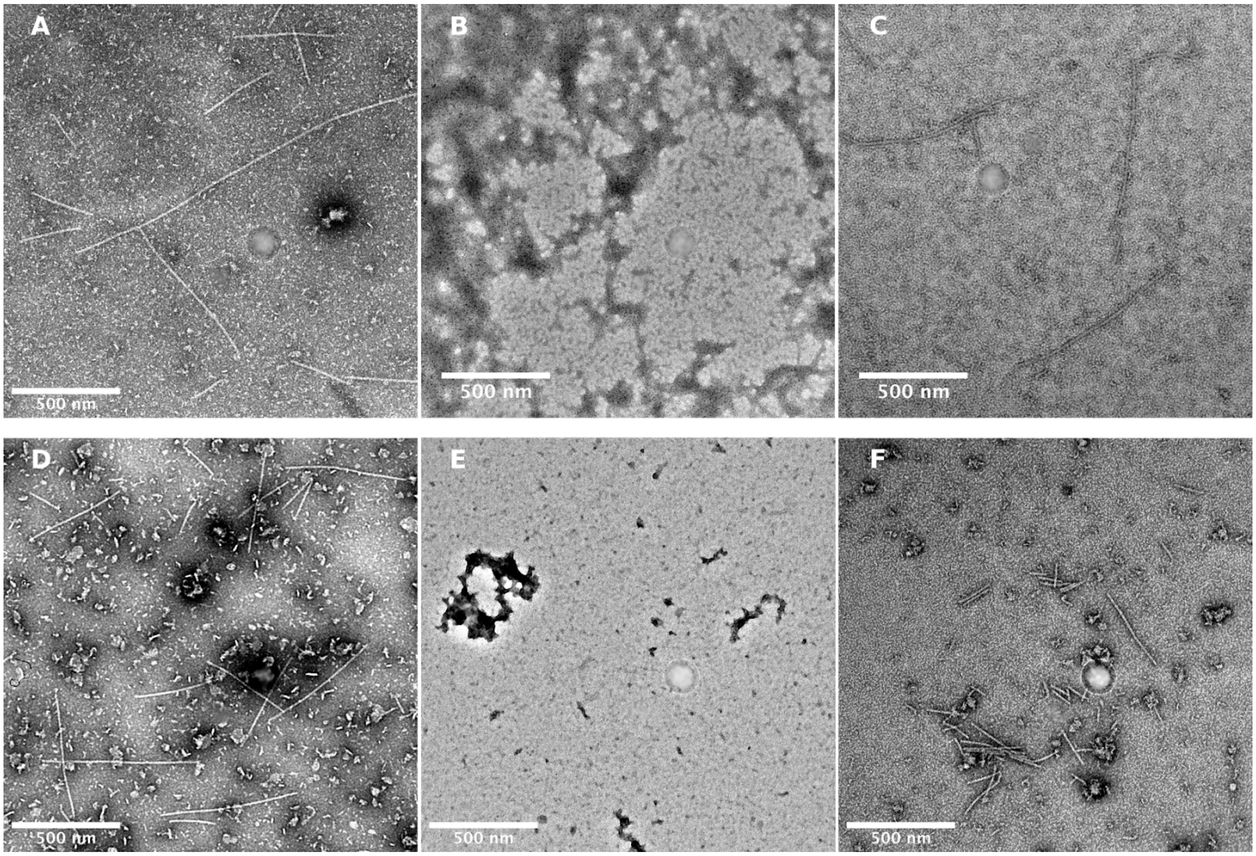


Fig. 2. Filaments' reversible response to pH change characterized by stained EM.

(**A**) Design DpHF7 stained with Nano-W (pH 6.8) (**B**) DpHF7 stained with Uranyl Formate stain (pH 4.0), washed with water, TBS buffer (pH 8) and then (**C**) stained with Nano-W

(**D**) Design DpHF18 stained with Nano-W (**E**) DpHF18 concentrated in NaCitrate (pH 4), then diluted 1:10 into TBS buffer (pH 8) for 1 minute and then (**F**) stained with Nano-W

Materials and Methods

Negative Stain Electron Microscopy

Soluble fractions were concentrated in buffer (25mM Tris, 75 mM NaCl, pH 8) for EM screening. A drop of 6µL (1µl sample instantly diluted with 5µl of buffer) was applied on negatively glow discharged, carbon-coated 200-mesh copper grids (Ted Pella, Inc.), washed with Milli-Q Water and stained either using 0.75% uranyl formate (pH 4.0) or Nano-W (pH 6.8) purchased from Nanoprobes, Inc (Yaphank, NY., USA) as described previously (*28*). The screening was performed on a 100kV Morgagni M268 transmission electron microscope (FEI, Hillsboro, OR). Images were recorded on a bottom mount Teitz CMOS 4k camera system. The contrast of the images was enhanced in the Fiji software (*29*) for clarity.

References

1. S. Ricard-Blum, F. Ruggiero, M. van der Rest, in *Collagen*, J. Brinckmann, H. Notbohm, P. K. Müller, Eds. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2005), vol. 247 of *Topics in Current Chemistry*, pp. 35–84.

2. L. C. Serpell, Alzheimer's amyloid fibrils: structure and assembly. *Biochim. Biophys. Acta*. **1502**, 16–30 (2000).

3. H. Herrmann, U. Aebi, Intermediate filaments: molecular structure, assembly mechanism, and integration into functionally distinct intracellular Scaffolds. *Annu. Rev. Biochem.* **73**, 749–789 (2004).

4. G. J. Rucklidge, G. Milne, B. A. McGaw, E. Milne, S. P. Robins, Turnover rates of different collagen types measured by isotope ratio mass spectrometry. *Biochim. Biophys. Acta*. **1156**, 57–61 (1992).

5. K. C. Holmes, D. Popp, W. Gebhard, W. Kabsch, Atomic model of the actin filament. *Nature*. **347**, 44–49 (1990).

6. E. Nogales, M. Whittaker, R. A. Milligan, K. H. Downing, High-Resolution Model of the Microtubule. *Cell*. **96**, 79–88 (1999).

7. B. Bhyravbhatla, S. J. Watowich, D. L. D. Caspar, Refined Atomic Model of the Four-Layer Aggregate of the Tobacco Mosaic Virus Coat Protein at 2.4-Å Resolution. *Biophys. J.* **74**, 604–615 (1998).

8. A. M. Smith *et al.*, Polar assembly in a designed protein fiber. *Angew. Chem. Int. Ed Engl.* **44**, 325–328 (2004).

9. L. E. R. O'Leary, J. A. Fallas, E. L. Bakota, M. K. Kang, J. D. Hartgerink, Multi-hierarchical self-assembly of a collagen mimetic peptide from triple helix to nanofibre and hydrogel. *Nat. Chem.* **3**, 821–828 (2011).

10. C. J. Bowerman, B. L. Nilsson, Self-assembly of amphipathic β-sheet peptides: insights and applications. *Biopolymers*. **98**, 169–184 (2012).

11. J. D. Hartgerink, J. R. Granja, R. A. Milligan, M. Reza Ghadiri, Self-Assembling Peptide Nanotubes. *J. Am. Chem. Soc.* **118**, 43–50 (1996).

12. E. H. Egelman *et al.*, Structural plasticity of helical nanotubes based on coiled-coil assemblies. *Structure*. **23**, 280–289 (2015).

13. N. C. Burgess *et al.*, Modular Design of Self-Assembling Peptide-Based Nanotubes. *J. Am. Chem. Soc.* **137**, 10554–10562 (2015).

14. C. Xu *et al.*, Rational design of helical nanotubes from self-assembly of coiled-coil lock washers. *J. Am. Chem. Soc.* **135**, 15565–15578 (2013).

15. F. A. Tezcan, F. Akif Tezcan, in *Coordination Chemistry in Protein Cages* (2013), pp. 149–174.

16. Y. Hsia *et al.*, Corrigendum: Design of a hyperstable 60-subunit protein icosahedron. *Nature*. **540**, 150 (2016).

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

18. S. Gonen, F. DiMaio, T. Gonen, D. Baker, Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. *Science*. **348**, 1365–1368 (2015).

19. J. A. Fallas *et al.*, Computational design of self-assembling cyclic protein homo-oligomers. *Nat. Chem.* **9**, 353–360 (2017).

20. T. J. Brunette *et al.*, Exploring the repeat protein universe through computational protein design. *Nature*. **528**, 580–584 (2015).

21. E. H. Egelman, The iterative helical real space reconstruction method: surmounting the problems posed by real polymers. *J. Struct. Biol.* **157**, 83–94 (2007).

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

23. S. H. W. Scheres, RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

24. N. Grigorieff, FREALIGN: high-resolution refinement of single particle structures. *J. Struct. Biol.* **157**, 117–125 (2007).

25. H. Garcia-Seisdedos, C. Empereur-Mot, N. Elad, E. D. Levy, Proteins evolve on the edge of supramolecular self-assembly. *Nature*. **548**, 244–247 (2017).

26. G. Bhardwaj *et al.*, Accurate de novo design of hyperstable constrained peptides. *Nature*. **538**, 329–335 (2016).

27. F. W. Studier, Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).

28. B. L. Nannenga, M. G. Iadanza, B. S. Vollmar, T. Gonen, Overview of Electron Crystallography of Membrane Proteins: Crystallization and Screening Strategies Using Negative Stain Electron Microscopy. *Curr. Protoc. Protein Sci.* **72**, 17.15.1–17.15.11 (2013).

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

30. C. Suloway *et al.*, Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).

31. S. Q. Zheng *et al.*, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods*. **14**, 331–332 (2017).

32. K. Zhang, Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).

33. G. C. Lander *et al.*, Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166**, 95–102 (2009).

34. C. Sachse *et al.*, High-resolution Electron Microscopy of Helical Specimens: A Fresh Look at Tobacco Mosaic Virus. *J. Mol. Biol.* **371**, 812–835 (2007).

35. P. D. Adams *et al.*, in *International Tables for Crystallography* (2012), pp. 539–547.

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

37. A. E. Carpenter *et al.*, CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).

38. D. Papapostolou *et al.*, Engineering nanoscale order into a designed protein fiber. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 10853–10858 (2007).

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

40. Z. Chen *et al.*, Programmable design of orthogonal protein heterodimers. *Nature*. **565**, 106–111 (2019).

41. G. Askarieh *et al.*, Self-assembly of spider silk proteins is controlled by a pH-sensitive relay. *Nature*. **465**, 236–238 (2010).

42. L. H. Beun *et al.*, From micelles to fibers: balancing self-assembling and random coiling domains in pH-responsive silk-collagen-like protein-based polymers. *Biomacromolecules*. **15**, 3349–3357 (2014).

43. S. E. Boyken *et al.*, De novo design of tunable, pH-driven conformational changes. *Science*. **364** (2019), pp. 658–664.