PROTEIN DESIGN

De novo design of self-assembling helical protein filaments

Hao Shen[1,2,3]*, Jorge A. Fallas[1,2]*†, Eric Lynch[2]*, William Sheffler[1,2], Bradley Parry[4,5], Nicholas Jannetty[5,6], Justin Decarreau[7], Michael Wagenbach[7], Juan Jesus Vicente[7], Jiajun Chen[8,9], Lei Wang[9,10], Quinton Dowling[2,11], Gustav Oberdorfer[1,2]‡, Lance Stewart[1], Linda Wordeman[7], James De Yoreo[8,9], Christine Jacobs-Wagner[5,6,12], Justin Kollman[2], David Baker[1,2,13]†

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.

Natural protein filaments differ considerably in their dynamic properties. Some, such as collagen, are relatively static, with turnover rates of several weeks [1–4], whereas others, such as 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). Although peptide filaments in the first class have been successfully designed 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 has remained an unsolved challenge. Much of the progress in recent years in the 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 two-dimensional (2D) crystal space group [18] symmetry through the design of a single new protein-protein interface. By contrast, the building blocks in most reversibly assembling filaments have no internal symmetry, and hence multiple designed interfaces are required to drive the formation of the desired structure. The reduced symmetry also makes the sampling problem more challenging, as the space of possible filament geometries is very large.

To tackle the challenge of de novo designing dynamic protein filaments, we devised a computational approach that exploits the requirement for multiple intermonomer 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. The search is therefore over the six 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 applying a random rotation (three degrees of freedom), choosing a random direction (two degrees of freedom), and sliding the second copy toward the first until they come into contact (Fig. 1C, left) (the sliding into contact effectively reduces the number of degrees of freedom from six for an arbitrary rigid-body transform to five). Successive monomers related by the filament-defining rigid-body trans-

form 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 nth root of this transform, where n ranges from two to five—with a choice of $n = 4$, for example, the first monomer will be in contact with the fourth monomer (Fig. 1C, bottom, and fig. S1). We also consider filaments with cyclic symmetry generated by the application of n-fold cyclic (C_n) symmetry operations around the superhelical axis, where n is between two and five (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 the addition or removal of one or more of the 30- to 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 230,000 helical filament backbones as described above and selected 124 designs for experimental testing [we refer to these as de novo–designed helical filaments (DHFs); for comparison with filaments generated from native backbones, see fig. S2].

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

[1]Institute for Protein Design, University of Washington, Seattle, WA 98195, USA. [2]Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. [3]Molecular Engineering Ph.D. Program, University of Washington, Seattle, WA 98195, USA. [4]Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06511, USA. [5]Microbial Sciences Institute, Yale University, West Haven, CT 06516, USA. [6]Howard Hughes Medical Institute, Yale University, West Haven, CT 06516, USA. [7]Department of Physiology and Biophysics, University of Washington, Seattle, WA 98195, USA. [8]Department of Materials Science and Engineering, University of Washington, Seattle, WA 98195, USA. [9]Physical Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA. [10]State Key Laboratory of Crystal Materials, Shandong University, Jinan 250100, China. [11]Department of Bioengineering, University of Washington, Seattle, WA 98195, USA. [12]Department of Microbial Pathogenesis, Yale School of Medicine, New Haven, CT 06510, USA. [13]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: jaf18@uw.edu (J.A.F.); dabaker@uw.edu (D.B.)
‡Present address: Institute of Biochemistry, Graz University of Technology, Graz, Austria.

(15 soluble and 19 insoluble) were found to form 1D nanostructures (figs. S3 and S4). A subset of the designs were synthesized as SUMO (small ubiquitin-like modifier) fusions to prevent premature filament formation; the SUMO tag was removed by using SUMO protease, and the samples were characterized by negative stain EM (fig. S5).

We chose six designs with a range of model architectures and longer persistence lengths for higher-resolution structure determination by cryo-EM. We determined the filament structures and refined helical symmetry parameters by using iterative helical real-space reconstruction in SPIDER (*21, 22*), followed by further 3D refinement in Relion (*23*) and Frealign
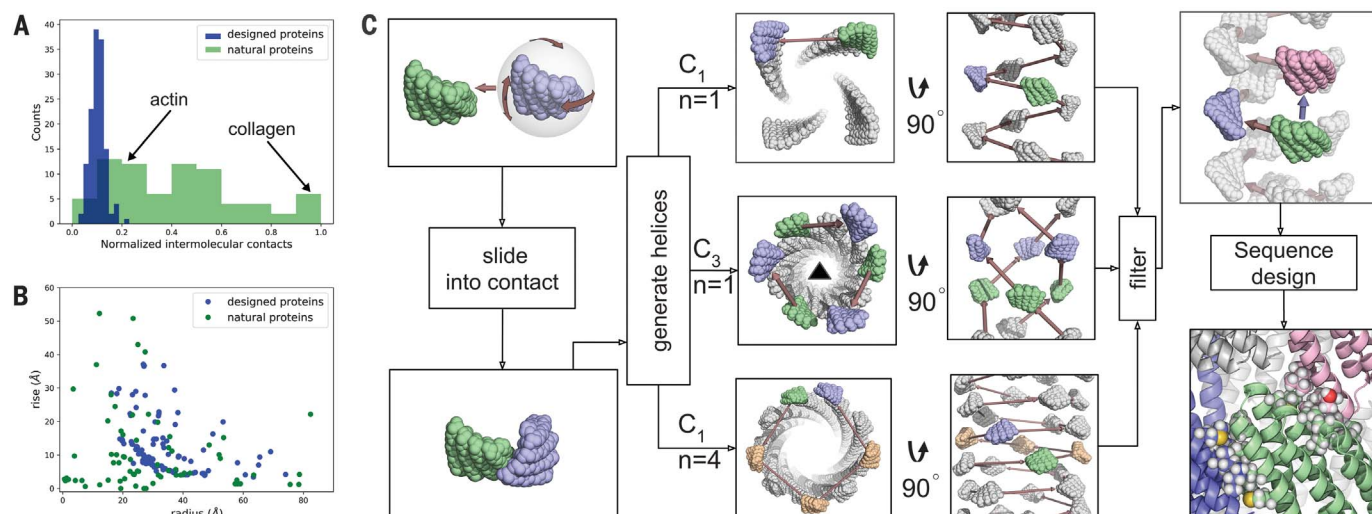
Fig. 1. Filament architectures and computational design protocol. (**A** and **B**) Comparison of properties of our designed filaments (blue) with those of native filaments (green). (A) The fraction of total residue-residue interactions between (rather than within) monomers. (B) Superhelical parameters. (**C**) Computational design protocol.

Fig. 2. Cryo-EM structure determination. (**A** to **F**) (Left to right) Computational model, representative filaments in cryo-EM micrographs, cryo-EM structure, and overlay between the model and structure for (A) DHF58 (RMSD, 3.3 Å), (B) DHF119 (RMSD, 2.3 Å), (C) DHF91 (RMSD, 1.2 Å), (D) DHF46 (RMSD, 2.3 Å), (E) DHF79 (RMSD, 4 Å), and (F) DHF38 (RMSD, 0.9 Å). (**G**) Close-up views of the two main intermonomer interfaces in the filament for DHF119, with the computational model (gray) and cryo-EM structure (cyan) in sticks in the helical reconstruction density (3.4-Å resolution). The high-resolution structure of design DHF119 is very close to the design model.
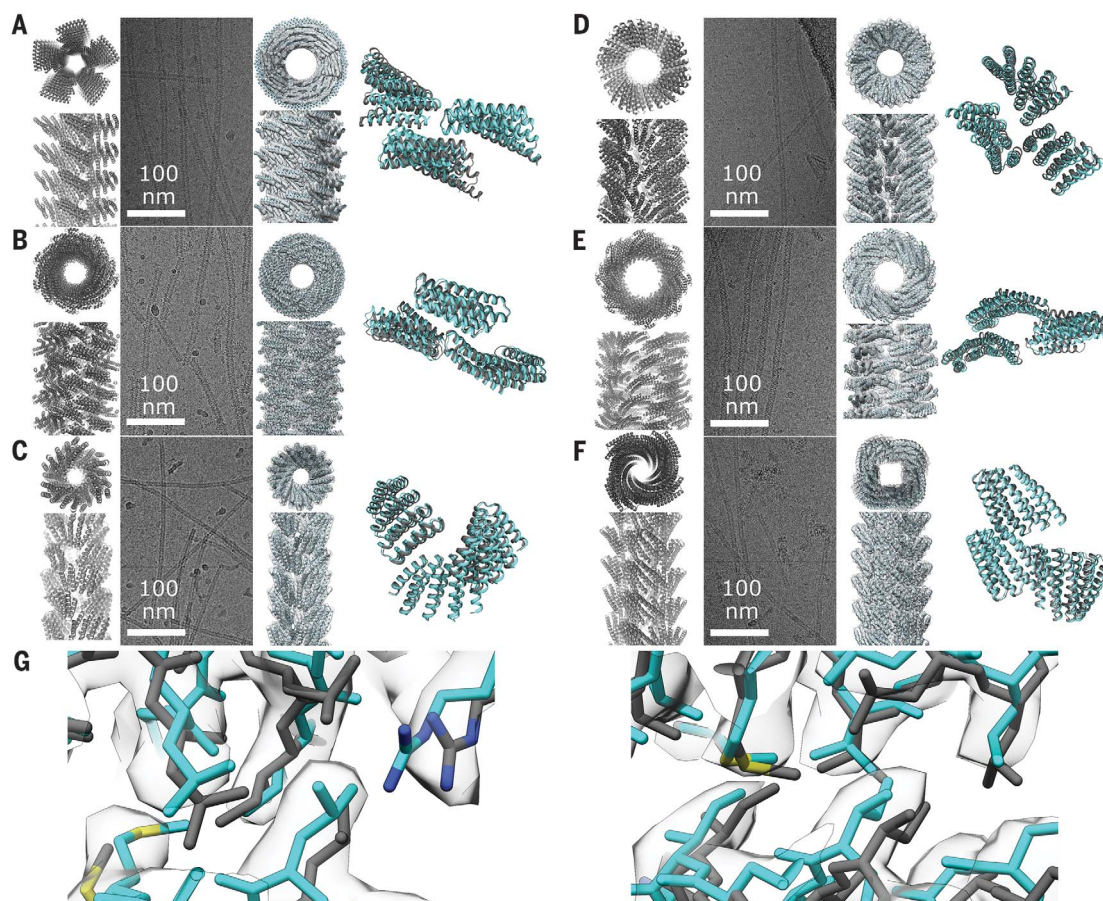
(24). In all six cases, the overall orientation and packing of the monomers in the filament were similar in the experimentally determined structures and design models, but the accuracy with which the details of the interacting interfaces were modeled varied considerably (Fig. 2 and fig. S6). Subtle shifts in the interaction interfaces in several cases altered the designed symmetry; DHF119, for example, was designed to be C_1, but the cryo-EM structure has C_3 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 DHF91, the experimentally observed rigid-body orientation was nearly identical to that of the design models [0.9- and 1.2-Å root mean square deviation (RMSD) over three chains containing all unique interfaces]; for DHF46 and DHF119, the RMSD 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).

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. 3, B and C).

We monitored assembly dynamics in vitro by solution scattering and in living cells by using fluorescence microscopy with monomers fused to green fluorescent protein (GFP). The extent and kinetics of DHF119 filament formation in vitro were 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 the 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* after the induction of expression of DHF58-GFP, discrete puncta were first observed, which over time resolved into filaments up to micrometers in length (Fig. 4A and movie S1) (the puncta may simply be filaments below the resolution limit of the microscope, ~250 nm). The filaments formed in vivo have high stability: After bacteria expressing the monomer were lysed by lysogenic phages, the filaments largely retained their shape (movie S2).

Natural systems achieve 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, such as the γ-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 to 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 fluorescence microscopy. The anchors seeded the rapid growth of multiple-micrometer-length fibers over 30 min (Fig. 4C and 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). The attachment of a biotinylated anchor to streptavidin-coated beads, followed by incubation with a filament monomer, resulted in an extensive network of filaments emanating from the beads (Fig. 4D, left); by contrast, very few filaments were observed around control beads that lacked the anchor protein (Fig. 4D, right).

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 (because the two ends of the filaments are distinct, there are two types of caps). The addition of increasing concentrations of the caps to already formed filaments resulted in shrinking and ultimate 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 whereas, in the presence of both caps, disassembly occurs from both ends (Fig. 4E and 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/min per end at 18 μM monomer) (fig. S12) similar to that observed by fluorescence for anchored fiber growth (8.4 nm/min at 18 μM monomer) (fig. S11). The observed behavior can be understood as follows. At the critical monomer concentration where fibers neither grow nor 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; thus, 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
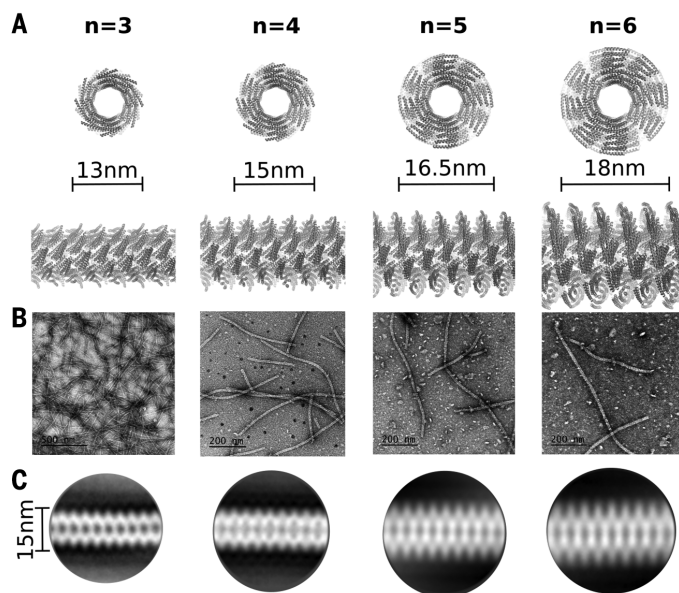

Fig. 3. Modular tuning of fiber diameter. DHF58 filament variants with different numbers of repeats were characterized by EM. (**A**) Cross sections and side views of computational models based on the four-repeat cryo-EM structure. The number of repeats (*n*) is shown at the top. (**B**) Negative stain electron micrographs. (**C**) 2D-class averages.
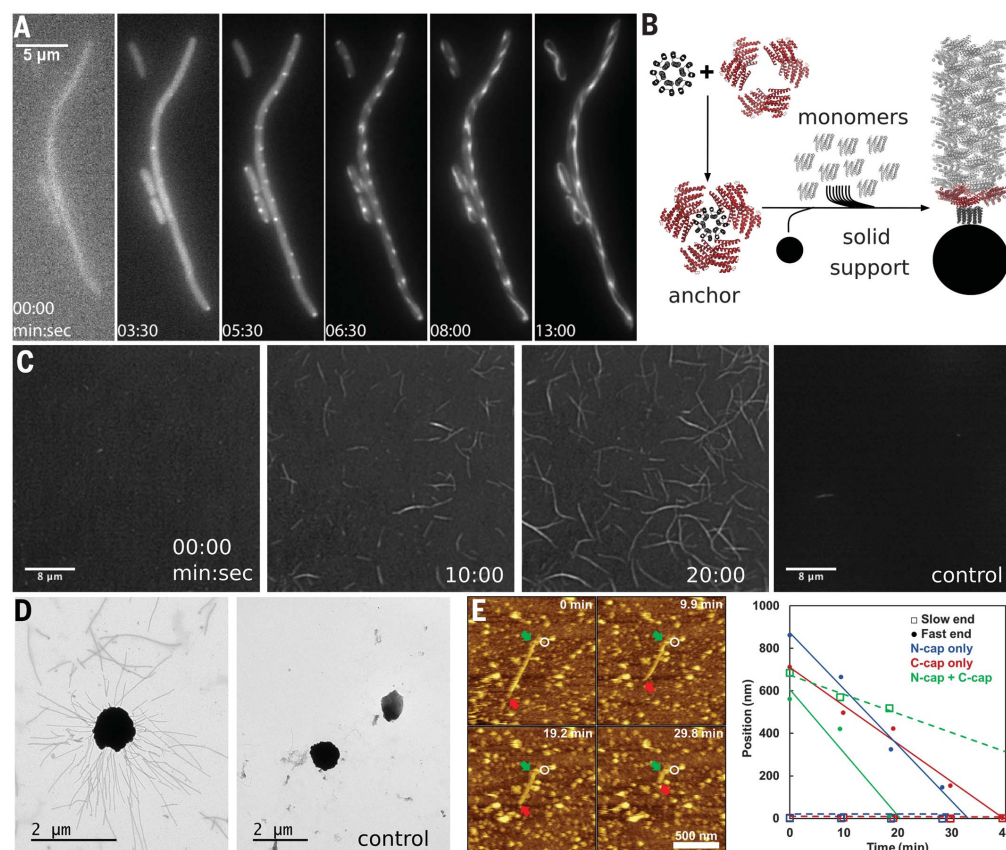
Fig. 4. Characterization of fiber growth and disassembly. (**A**) Kinetics of the assembly of DHF58-GFP filaments in *E. coli*. (**B**) Construction of fiber anchors holding monomers in the rigid-body arrangement found in the filament. (**C**) Kinetics of DHF119-YFP filament assembly in vitro on a glass surface coated with the DHF119_C6 anchor. In the control panel, the glass surface was coated with the noncognate DHF91 anchor. (**D**) DHF119 filaments emanating from DHF119_C6 anchor–coated magnetic beads incubated with the monomer. Beads on the right lack the anchor. (**E**) Disassembly of DHF119 fibers in the presence of capping units monitored by in situ AFM. (Left) Image sequence showing disassembly in the presence of N caps. The white circles mark a fixed position in all images. (Right) Positions of fiber ends versus time in solutions with N caps, C caps, and N caps plus C caps. 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 bind only 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.

and free caps can add (fig. S13, top right) than at the other end, where only monomers can add (fig. S13, top left). Because the rate of monomer dissociation is the same at both ends, the fibers shrink primarily from one end, as observed.

The ability to program micrometer-scale order from angstrom-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. The introduction of a small number of hydrophobic substitutions near the periphery of dihedral complexes can promote stacking into extended filaments because 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 because only a small fraction of possible helical assemblies contain dihedral point group symmetry. Both designed interfaces were ac-

curately recapitulated in four of the six structures solved by cryo-EM; 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 therefore, the designed filaments provide readily modifiable scaffolds to which binding sites for other proteins or metal nanoclusters can be added for applications ranging from cryo-EM structure determination to nanoelectronics.

REFERENCES AND NOTES

1. S. Ricard-Blum, F. Ruggiero, M. van der Rest, in *Topics in Current Chemistry*, vol. 247, *Collagen*, J. Brinckmann, H. Notbohm, P. K. Müller, Eds. (Springer, 2005), pp. 35–84.
2. L. C. Serpell, *Biochim. Biophys. Acta* **1502**, 16–30 (2000).
3. H. Herrmann, U. Aebi, *Annu. Rev. Biochem.* **73**, 749–789 (2004).
4. G. J. Rucklidge, G. Milne, B. A. McGaw, E. Milne, S. P. Robins, *Biochim. Biophys. Acta* **1156**, 57–61 (1992).
5. K. C. Holmes, D. Popp, W. Gebhard, W. Kabsch, *Nature* **347**, 44–49 (1990).
6. E. Nogales, M. Whittaker, R. A. Milligan, K. H. Downing, *Cell* **96**, 79–88 (1999).
7. B. Bhyravbhatla, S. J. Watowich, D. L. D. Caspar, *Biophys. J.* **74**, 604–615 (1998).
8. A. M. Smith et al., *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, *Nat. Chem.* **3**, 821–828 (2011).
10. C. J. Bowerman, B. L. Nilsson, *Biopolymers* **98**, 169–184 (2012).
11. J. D. Hartgerink, J. R. Granja, R. A. Milligan, M. R. Ghadiri, *J. Am. Chem. Soc.* **118**, 43–50 (1996).
12. E. H. Egelman et al., *Structure* **23**, 280–289 (2015).
13. N. C. Burgess et al., *J. Am. Chem. Soc.* **137**, 10554–10562 (2015).
14. C. Xu et al., *J. Am. Chem. Soc.* **135**, 15565–15578 (2013).
15. F. A. Tezcan, in *Coordination Chemistry in Protein Cages: Principles, Design and Applications*, T. Ueno, Y. Watanabe, Eds. (Wiley, 2013), pp. 149–174.
16. Y. Hsia et al., *Nature* **540**, 150 (2016).
17. N. P. King et al., *Nature* **510**, 103–108 (2014).
18. S. Gonen, F. DiMaio, T. Gonen, D. Baker, *Science* **348**, 1365–1368 (2015).
19. J. A. Fallas et al., *Nat. Chem.* **9**, 353–360 (2017).

20. T. J. Brunette *et al.*, *Nature* **528**, 580–584 (2015).
21. E. H. Egelman, *J. Struct. Biol.* **157**, 83–94 (2007).
22. E. H. Egelman, *Ultramicroscopy* **85**, 225–234 (2000).
23. S. H. W. Scheres, *J. Struct. Biol.* **180**, 519–530 (2012).
24. N. Grigorieff, *J. Struct. Biol.* **157**, 117–125 (2007).
25. H. Garcia-Seisdedos, C. Empereur-Mot, N. Elad, E. D. Levy, *Nature* **548**, 244–247 (2017).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6415/705/suppl/DC1
Materials and Methods
Figs. S1 to S13
Tables S1 and S2
References (*26–37*)
Movies S1 to S5
Data S1

Science

De novo design of self-assembling helical protein filaments

Hao Shen, Jorge A. Fallas, Eric Lynch, William Sheffler, Bradley Parry, Nicholas Jannetty, Justin Decarreau, Michael Wagenbach, Juan Jesus Vicente, Jiajun Chen, Lei Wang, Quinton Dowling, Gustav Oberdorfer, Lance Stewart, Linda Wordeman, James De Yoreo, Christine Jacobs-Wagner, Justin Kollman and David Baker

Built to be reversible

There has been some success in designing stable peptide filaments; however, mimicking the reversible assembly of many natural protein filaments is challenging. Dynamic filaments usually comprise independently folded and asymmetric proteins and using such building blocks requires the design of multiple intermonomer interfaces. Shen *et al.* report the design of self-assembling helical filaments based on previously designed stable repeat proteins. The filaments are micron scale, and their 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.

Science, this issue p. 705

ARTICLE TOOLS	http://science.sciencemag.org/content/362/6415/705
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2018/11/07/362.6415.705.DC1
REFERENCES	This article cites 34 articles, 1 of which you can access for free http://science.sciencemag.org/content/362/6415/705#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions