Atomic insights into the genesis of cellular filaments by globular proteins

Laura McPartland1,2, Danielle M. Heller1,2, David S. Eisenberg2,3,4,5,6, Ann Hochschild1,* and Michael R. Sawaya2,3,4,5,6,*

Self-assembly of proteins into filaments, such as actin and tubulin filaments, underlies essential cellular processes in all three domains of life. The early emergence of filaments in evolutionary history suggests that filament genesis might be a robust process. Here we describe the fortuitous construction of GFP fusion proteins that self-assemble as fluorescent polar filaments in Escherichia coli. Filament formation is achieved by appending as few as 12 residues to GFP. Crystal structures reveal that each protomer donates an appendage to fill a groove between the two following protomers along the filament. This exchange of appendages resembles runaway domain swapping but is distinguished by higher efficiency because monomers cannot competitively bind their own appendages. Ample evidence for this ‘runaway domain coupling’ mechanism in nature suggests it could facilitate the evolutionary pathway from globular protein to polar filament, requiring a minimal extension of protein sequence and no substantial refolding.

The self-association of proteins to form filaments is the basis for the formation of essential cytoskeletal elements such as actin and tubulin in all three domains of life. Cytoskeletal elements participate in a variety of critical cellular processes, including DNA segregation, cell division, morphogenesis, cell motility, maintenance of cell shape, and the transport of cellular cargoes. The identification of diverse families of actin and tubulin homologs in prokaryotes over the last 20 years has led to the suggestion that these are ancient proteins that predate the splits between bacteria, archaia and eukarya. Furthermore, the recent discovery that certain metabolic enzymes can polymerize as filaments has provoked the suggestion that an original rudimentary cytoskeleton may have evolved in an ancestral cell through the repurposing of enzyme polymerization.

The early emergence of filaments in biological history raises the questions of what structural mechanisms are capable of establishing filament-competent interfaces and how facile is their evolution. One mechanism requiring few specific structural features other than a globular fold is runaway three-dimensional domain swapping. By this mechanism, the intermolecular interface between successive protomers in the filament recapitulates an intramolecular interface that exists within the isolated monomer. Thus, a domain of the protein, usually residing at either the N or C terminus and connected to the remainder of the protein by a flexible linker (the so-called hinge loop), exchanges its intramolecular contacts for otherwise identical intermolecular contacts, an exchange that usually occurs only under non-native conditions. Runaway domain swapping occurs in a unidirectional fashion, with each subunit donating a domain to the next, leading to the formation of an open-ended protofilament. Examples include fibrils of a designed three-α-helix bundle, filaments of α-antitrypsin and ribbons of ‘cab’-type carbonic anhydrase. Runaway domain swapping has also been evidenced in some amyloid fibrils. (We consider fibrils and filaments to be equivalent in the assembly hierarchy—a level more complex than protofilaments—and use the term most closely associated with the particular system discussed.)

A simpler, more obvious mechanism for filament formation is the stacking of domains in an end-to-end fashion. Other than runaway domain swapping and end-to-end stacking, no other naturally occurring mechanisms for filament formation have been formally categorized, despite the recent rapid increase in the number of filament structures determined.

Here we describe a nested set of GFP fusion proteins with a shared ability to self-assemble as fluorescent filaments in bacterial cells. Remarkably, these filaments typically span the entire length of the cell. The smallest of these fusion proteins bears just 12 non-native amino acids at the C terminus of GFP. Structure determination by X-ray crystallography revealed that these appended residues form an α-helix that extends away from the GFP domain (a β-barrel) through a hinge loop and docks into a groove formed by the following two β-barrels in the protofilacon. We call this filament-forming mechanism ‘runaway domain coupling’ because the protomer–protomer interfaces involve complementation between distinct donor and acceptor domains. It is distinct from end-to-end domain stacking, wherein the protomer–protomer interfaces reside entirely within a single domain type. It is also distinct from runaway domain swapping, because the helix-in-groove complementation cannot be reproduced in the monomer. The fact that runaway domain coupling arose fortuitously suggests that the genesis of coupled interfaces is facile, an inference supported by our discovery of their occurrence in numerous natural filaments. We propose that the chances of such randomly generated surfaces aligning in a productive interface are enhanced by a flexible hinge loop that connects the two domains of each protomer. Our findings illustrate a pathway whereby a non-polymerizing protein might evolve the ability to self-assemble as protofilaments and filaments.

*Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA, USA. †Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA, USA. ‡Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA, USA. §Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, CA, USA. †Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA, USA. ‡UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, Los Angeles, CA, USA.

These authors contributed equally: Laura McPartland, Danielle M. Heller. e-mail: ann_hochschild@hms.harvard.edu; sawaya@mbi.ucla.edu
Results

GFP-RNase A fusion proteins form fluorescent rods in bacterial cells. Previous work from the Eisenberg group demonstrated that RNase A variants containing short amyloidogenic peptide insertions in the C-terminal hinge loop could form amyloid-like fibrils in vitro. We sought to investigate whether these RNase A variants could also access the fibrillar form when produced in E. coli cells.

To do this, we constructed a plasmid vector that directed the inducible synthesis of RNase A (with or without an inserted amyloidogenic peptide sequence; see Methods) fused to a monomeric variant of GFP (Fig. 1a). Cells containing these plasmids were examined by fluorescence microscopy at various times after the induction of fusion protein synthesis. Whereas most cells contained diffuse fluorescence, we observed fluorescent rod structures in a small minority (1–2%) of cells. Such rod-containing cells typically contained a single rod (Fig. 1b, left). These structures were observed regardless of whether or not the RNase A moiety contained the amyloidogenic peptide insertion. Furthermore, we found that refrigeration dramatically stimulated formation of the rod-like structures. When we grew cells overnight after the induction of fusion protein synthesis and then incubated the cell cultures for 2 h at 4 °C, we detected fluorescent rod structures in essentially all of the cells (Fig. 1b, right).

To define the sequence requirements for intracellular rod formation, we constructed a series of truncations from the C terminus of RNase A (Fig. 1a). We found that as few as eight N-terminal residues of RNase A (separated from the C terminus of GFP by a three-residue linker) sufficed to permit rod formation. When cell cultures containing this construct were induced overnight and incubated for 2 h at 4 °C, we detected fluorescent rod structures in essentially all of the cells (Fig. 1b, right).

Fig. 1 | Fluorescence images of E. coli cells containing GFP-RNase A fusion proteins. a, Fusion constructs. RNase A and truncated segments (gray boxes) were fused to the C terminus of GFP (green boxes) via a three-residue linker (white boxes). b, After overnight induction of GFP-AMA-RNaseA fusion protein, cells were imaged before (left) or after incubation for 2 h at 4 °C (right). c, Variations in the linker of GFP-AMA-RNaseA(1–14) cause variation in rod-forming propensities, as observed after overnight induction before (AMA, ALA (left)) or after incubation for 24 h at 4 °C (AQA, ALA (right), AEA, AKA, AAA). The graph indicates the energy of the highest propensity amyloid-forming hexapeptide predicted by ZipperDB for each of the six variants. Most notably, the AQA variant, predicted to have the strongest amyloid propensity, produced no rods, suggesting that the rods are not amyloid-like. d, Eight appended RNase A residues suffice for rod formation. e,Appending a hexahistidine tag does not interfere with rod formation. Constructs in d,e were imaged after overnight induction and incubation for 2 h at 4 °C. f, V219E, a mutation within the GFP moiety, disrupts filament formation. Cells were imaged after overnight induction and incubation for 120 h at 4 °C. For all panels, fluorescence images show representative fields, with illustrative portions enlarged (insets). g, Fluorescence and phase contrast images of dividing cells connected by continuous, cell-spanning rod structures (white arrowheads). A similar phenomenon was observed in E. coli cells overexpressing an actin homolog. h, Cells exhibiting rod-induced bulges (yellow arrowheads) at their poles. Cells in g,h contain GFP-RNaseA(1–14) fusion protein and were imaged after induction for 4 h and incubation for 24 h at 4 °C. Scale bars, 5 μm.
Rod formation does not depend on the fortuitous presence of an HP segment. We considered the possibility that the propensities of the constructed GFP fusion proteins to form fluorescent rod structures was due to the fortuitous presence of an amino acid segment with high amyloid fiber-forming propensity (termed an HP segment) either spanning the fusion junction or contained wholly within the appended sequence. We discovered an HP segment comprising the three-residue linker and the first three residues of RNase A (AMA) that was shared by our rod-forming constructs (Fig. 1a and Supplementary Fig. 1)19.

Results from mutagenesis experiments were inconsistent with the hypothesis that rod formation was due to the amyloid-forming propensity of the HP segment located at the N-terminal end of the appended peptide sequence. Using a fusion protein containing RNase A residues 1–14 as our reference construct, we compared the rod-forming behaviors of the reference fusion (GFP-AMA-RNase(1–14)) and several variants bearing replacements of the linker methionine residue (residue 240 in the fusion sequence) with alanine, glutamate, lysine, leucine or glutamine (Fig. 1c and Supplementary Fig. 1). Whereas the reference construct gave rise to rods in essentially all of the cells after overnight induction of fusion protein synthesis (Fig. 1c), the variant encoding AQAKET, which has a higher predicted amyloid-forming propensity than AMAKET, gave rise to no rod-containing cells even after induced overnight cultures were refrigerated for another 24 h (Fig. 1c). Among the remaining variants, only that encoding ALAKET gave rise to rods, which formed efficiently only after refrigeration (Fig. 1c), suggesting that a hydrophobic residue was required at the second linker position (see below). Overall, our analysis of this set of single amino acid substitution variants suggests that rod formation is not due to the amyloid-forming propensity of the appended peptide. Conclusive evidence that the rods lack amyloid architecture follows from diffraction analysis and crystal structure determination, presented below.

Crystal structures of GFP fused to eight RNase A residues. To elucidate the intermolecular interactions that guide assembly of GFP into intracellular fluorescent rods, we crystallized a fusion protein containing RNase A residues 1–14 as our reference construct, we compared the rod-forming behaviors of the reference fusion (GFP-AMA-RNase(1–14)) (Fig. 1d), the shortest fusion protein that efficiently formed rods. To facilitate purification of the fusion, we appended a C-terminal His tag (GFP-AMA-RNase(1–8)-His6) after determining that the His tag did not alter the ability of the fusion protein to form fluorescent rods (Fig. 1c).

We discovered crystals in dozens of conditions, all of which shared the same needle morphology, recapitulating the rods observed in cells (Supplementary Fig. 2a). The needles were typically over 100 μm long, but seldom greater than 5 μm thick. Among the different conditions, we discovered four different crystal forms (P21212 form1, P212121 form2, C2 and P21) (Supplementary Fig. 2b), collected diffraction data and solved their structures (Table 1).

Remarkably, the structures obtained from the four crystal forms revealed essentially identical protofilament architectures, suggesting that this architecture is likely to persist under many other conditions, including those within the cell. In each crystal form, the protofilament is assembled with 2, screw symmetry, meaning that each protomer is related to its closest neighbor protomer by a 180° rotation around the protofilament axis and translation along the axis by half the pitch length (i.e., rise per turn along the protofilament axis) (Fig. 2a,c). In each crystal form, this pitch length is approximately 51.4 Å, (less than 1 Å variation) and corresponds to one of the unit cell dimensions (a = 51.2 Å in P21212, form 1, a = 51.1 Å in P212121, form 2, b = 51.3 Å in C2, and b = 51.8 Å in P21) (Fig. 3b). There is no significant structural variation in protofilament architecture among crystal forms (Fig. 2b). Indeed, the root mean square deviation is less than 0.8 Å between protofilament segments from different crystals (composed of three consecutive GFP-AMA-RNase(1–8)-His6, protomers totaling 5,667 atoms). Thus, the robust nature of this protofilament assembly is evidenced by the high degree of similarity in protofilament structures (Fig. 2b) despite clear differences in packing between protofilaments (Fig. 3b).

A closer view of the protofilament revealed that one structural feature in particular stabilizes protofilament assembly: a new, fusion-created, C-terminal helix that fastens together neighboring pairs of protomers by binding in a groove between them (Fig. 2a,c,d). This new C-terminal helix comprises seventeen residues that span a sequence derived from three different origins: six residues from the natural GFP C terminus (233-MDELYK-238), a three-residue linker (239-AMA-241) and eight residues from RNase (242-KETAAKF-249) (Fig. 2d). The six residues contributed by GFP are normally disordered in crystal structures of natural-length GFP. The propensity of this segment to form a helix was evidently enhanced by its fusion with RNase residues 1–8, which also form a helix in native RNase.

The new, fusion-created helix extends up to 40 Å away from its attachment to the barrel’s last strand, enabling the helix to reach a binding groove that is formed by the following two protomers in the protofilament. This 40-Å reach is thus directed by the relative positions of three successive protomers in the protofilament; the helix originates from protomer i, and the groove is formed between protomers i + 1 and i + 2. The ability of the helix to reach this groove is facilitated by the extended conformation adopted by GFP residues 228-GITG-232, which connect the helix to strand 11, the last strand in the GFP barrel (Fig. 2d).

Protomer i + 1 forms one wall of the groove (Fig. 2d,e, brown surface). This wall comprises the surfaces of strands 3, 10 and 11 near the waist of the barrel, and the helix docks tangentially across these. The helix-in-groove interface is snug and defined almost entirely by side-chain-to-side-chain contacts. A stripe of hydrophobic side chains (M233, L236, Y237, M240 and T244) lines the length of the C-terminal helix, and these embed between side chains on the surface of the barrel (T43, K45, A206, S208, V219 and L221). Loss of this hydrophobic interface upon substituting M240 with polar, charged, or small residues, such as glutamine, glutamate, lysine or alanine, would likely explain our observation that these substitutions interfere with efficient rod formation (Fig. 1c and Supplementary Fig. 3).

This interface between helix i and protomer i + 1 buries a total of 727 Å² on the two surfaces (Supplementary Fig. 4).

Protomer i + 2 forms the opposite wall of the groove. This wall comprises the end of strand 5 and residues at one cap of the barrel (Fig. 2d,e, blue surface). Specifically, residues L7, P89 and F114 create a small hydrophobic pocket into which docks the side chain of A245 of the helix of protomer i. This contact is flanked on both sides by salt bridges (protomer i + 2 residues E90 and D117 with helix i residues K242 and K248, respectively). This interface between protomer i + 2 and helix i buries a total of 426 Å² on the two surfaces (Supplementary Fig. 4). Overall, the shape complementarity between the helix and groove (0.78) indicates that the fit is even tighter than typically observed among antibody–antigen complexes20.

Direct contact between the sides of neighboring GFP barrels strengthens and rigidifies the contiguous helix-in-groove interaction (Fig. 2f). The interface is approximately flat. Side chains from strands 1, 2, 5 and 6 of protomer i + 1 nestle with side chains from strands 3, 10 and 11 of protomer i + 2. The interactions are primarily polar rather than nonpolar. The area buried by the two surfaces of this interface is 901 Å² (Supplementary Fig. 4). This patch of direct contact between barrels appears to be adventitious; it does not correspond in any way to the natural dimer interface of wild-type GFP21.

The combined surface area buried by protomers i, i + 1 and i + 2 appears large enough to maintain protofilament architecture...
in solution, free of the crystal lattice. Coordination of neighboring protomers by the C-terminal helix buries a total of 1,153 Å². Direct contact between protomers buries an additional 901 Å² of surface area. Either interface by itself would likely be insufficient to support protofilament assembly. But, together, these total to 2,054 Å², which exceeds the 1,712 Å² threshold value that discriminates between biological and artificial dimers. The division of the protomer–protomer interface over three surfaces implies cooperativity in protofilament assembly.

Single amino acid substitution disrupts rod formation. As a further test of whether the intermolecular interface observed in the crystals is relevant to rod formation in the cell, we mutated a solvent-exposed residue on the GFP barrel that was predicted to contribute significantly to the interface formed by the binding of the C-terminal α-helix of the neighboring protomer. Specifically, we introduced a charged residue (glutamate) in place of V219, which should disfavor the docking of M233, L236 and M240 in the C-terminal α-helix (Supplementary Fig. 3). We then induced the synthesis of GFP-AMA-RNase(1–8)-His₆ with or without the V219E substitution in the GFP moiety. Whereas the unmutated fusion protein formed rod structures in essentially all of the cells after overnight induction followed by incubation for 2 h at 4 °C, the mutant fusion protein failed to form rods even after overnight cultures were refrigerated for several days, and instead gave rise to diffuse fluorescence (Fig. 1f and Supplementary Fig. 5). These findings provide strong support for the hypothesis that the intermolecular interface observed in the four crystal forms is also the basis for fluorescent rod formation in bacterial cells.

The pattern of protofilament bundling in intracellular rods resembles that observed in the C2 crystal form. We performed in cellulo diffraction experiments to evaluate the pattern of protofilament bundling in fluorescent rods produced in E. coli. X-ray diffraction patterns from E. coli producing GFP-AMA-RNaseA(1–8)-His₆ revealed six reflections with Bragg spacings between 55 Å and 25 Å (Fig. 3c). Remarkably, these spacings observed from intracellular filaments coincide with those of the crystal structure of purified GFP-AMA-RNaseA(1–8)-His₆ in space group C2 (Fig. 3e). Moreover, the relative intensities of the six observed reflections correlate between the observed and calculated patterns, as well as the geometric disposition with respect to the meridional and equatorial

### Table 1 | Crystallographic data collection and refinement statistics

|                      | GFP-RNase(1-8)-His₆ (PDB 5HGE) | GFP-RNase(1-8)-His₆ (PDB 6AS9) | GFP-RNase(1-8)-His₆ (PDB 5HBD) | GFP-RNase(1-8)-His₆ (PDB SHW9) |
|----------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| **Data collection**   |                                 |                                 |                                 |                                 |
| Space group          | P2,2,2, form 1                  | P2,2,2, form 2                  | C2                              | P2,                                           |
| Cell dimensions      |                                 |                                 |                                 |                                 |
| a, b, c (Å)          | 51.2, 62.9, 69.3                | 51.1, 54.5, 85.5                | 112.7, 51.3, 55.1               | 45.7, 51.8, 54.8                   |
| α, β, γ (°)          | 90.0, 90.0, 90.0                | 90.0, 90.0, 90.0                | 90.0, 90.6, 90.0               | 90.0, 96.0, 90.0                   |
| Resolution (Å)       | 46.0-1.86 (1.96-1.86)           | 46.0-1.75 (1.80-1.75)           | 46.0-1.65 (1.75-1.65)           | 46.0-3.0 (3.08-3.00)               |
| Rmerge               | 0.106 (0.556)                   | 0.117 (0.585)                   | 0.094 (0.534)                   | 0.189 (0.750)                      |
| I / σ(I)             | 9.4 (2.3)                       | 14.1 (3.8)                      | 13.7 (2.3)                      | 5.8 (1.2)                          |
| CC₁/₂                | 0.997 (0.836)                   | 0.996 (0.630)                   | 0.998 (0.765)                   | 0.975 (0.554)                      |
| Completeness (%)     | 96.7 (98.2)                     | 99.4 (98.5)                     | 91.2 (63.8)                     | 95.1 (84.4)                        |
| Redundancy           | 3.7 (3.7)                       | 7.1 (7.0)                       | 6.0 (2.9)                       | 3.2 (2.1)                          |
| **Refinement**       |                                 |                                 |                                 |                                 |
| Resolution (Å)       | 46.0-1.86                       | 46.0-1.75                       | 46.0-1.65                       | 46.0-3.00                          |
| No. of reflections   | 18,492                          | 24,416                          | 30,817                          | 4,963                             |
| Rwork / Rfree       | 0.169 / 0.211                   | 0.156 / 0.156                   | 0.174 / 0.206                   | 0.186 / 0.232                      |
| No. of atoms         | 1,937                           | 1,952                           | 1,925                           | 1,932                             |
| **Ligands**          |                                 |                                 |                                 |                                 |
| Protein              | 8                               | 16                              | 0                               | 0                                 |
| Acetate              | 0                               | 8                               | 0                               | 0                                 |
| Water                | 172                             | 237                             | 312                             | 6                                 |
| **B factors**        |                                 |                                 |                                 |                                 |
| Protein              | 24.4                            | 14.5                            | 19.1                            | 50.0                              |
| Ligands              | 46.7                            | 36.7                            | 24.3                            |                                   |
| 2-Methyl-2,4-pentanediol | 30.6 | 25.9 | 30.9 | 39.2 |
| Acetate              | 0                               | 8                               | 0                               | 0                                 |
| Water                | 172                             | 237                             | 312                             | 6                                 |
| R.m.s. deviations    |                                 |                                 |                                 |                                 |
| Bond lengths (Å)     | 0.010                           | 0.010                           | 0.013                           | 0.010                             |
| Bond angles (°)      | 1.1                             | 1.1                             | 1.7                             | 1.2                               |

*Values in parentheses are for highest-resolution shell.
Conversely, many functional protofilaments, such as actin and tubulin, have resisted crystallization because their symmetry is imprecise or does not coincide with common crystallographic symmetry elements.

Discussion

We found that appending as few as 12 residues to the C terminus of GFP enabled the fortuitous assembly of GFP filaments both in vivo and in vitro. These filaments were detectable as fluorescent rods that spanned the length of bacterial cells harboring the fusion construct. The purified fusion protein readily formed needle-like crystals, and structure determination revealed the basis for filament formation. Specifically, a fusion-created C-terminal α-helix extends away from the barrel to which it is attached and nestsles into a groove that is formed by the barrels of two successive protomers along the protofilament axis.

This helix-in-groove linkage between protomers resembles runway domain swapping, a mechanism proposed for the assembly of disease-related fibrils. By this mechanism, each protomer swaps a part of its structure into an identical protomer along the protofilament (Fig. 4e). The swapped part is termed the ‘swapping domain’ and may be as small as a single α-helix or β-strand; the receiving part is termed the ‘complementary domain’. The swapping event begins when a folded or ‘closed’ monomer (Fig. 4c) opens up to expose its own swapping domain to another protomer (Fig. 4d). The flexibility required to open the monomer and execute the swap is afforded by a
**Fig. 3 | Protofilament architecture and bundling in vivo.**

a. Four independent crystal structures of the GFP-RNase fusion reveal essentially identical protofilament assemblies. b. Remarkably, this consensus protofilament structure (one is highlighted in green, blue, brown) is unperturbed despite differences in its alignment with neighboring protofilaments (parallel in C2 and P2₁, and antiparallel in P2₂2₁2₁ forms 1 and 2) and different spacings between neighboring protofilaments. Only one of the unit cell dimensions is constant among the four crystal forms, and this corresponds to the filament pitch (brown scale bar). Crystallographic 2₁ symmetry axes are denoted by black symbols and half-arrows. c. An X-ray diffraction pattern from dried *E. coli* cells carrying GFP-RNase(1–8) filaments reveals six low-resolution reflections (labeled 55 to 25 Å). d. Diffraction radial profiles calculated for each crystal form (orange traces) are overlaid with the observed radial profile of *E. coli* cells (blue dashed traces). One of the calculated patterns (space group C2) matches the observed pattern, implying that protofilament bundling in vivo resembles the packing observed in the C2 crystal. The correlation coefficient (CC) between observed and simulated diffraction profiles is noted in the lower right corner of each subpanel. e, f. Calculated filament diffraction pattern of CC = 0.472. Overlay of observed and calculated (space group C2) diffraction patterns. Bands of dark green highlight the coincidence of the most intense reflections. Similarity is evident in the Bragg spacings, relative intensities of the reflections, and orientation of the reflections with respect to the filament axis (filament axis is vertical). Labels indicate Miller indices and Bragg spacings.
hinge loop that links the two domains. In GFP-RNase(1–8) protofilaments, the C-terminal helix acts as the swapping domain, the GFP barrel acts as the complementary domain, and intervening residues 228–232 form the hinge loop.

Other properties of the GFP-RNase filaments, such as facile nucleation and mechanical rigidity, seem incongruous with runaway domain swapping. Indeed, GFP-RNase(1–8) filaments assemble spontaneously in vivo within hours, unlike runaway domain swapped filaments, which may take years to assemble in vivo or require high concentrations, harsh conditions, proteolytic cleavage or multiple days to nucleate in vitro. Furthermore, GFP-RNase filaments are mechanically strong and rigid. In some instances, the rods connect two deeply constricted cells, implying that septum formation was incapable of severing a preexisting filament (Fig. 1g). In other instances, the ends of growing GFP-RNase rods appear to impose bulges at the cell poles (Fig. 1h). These distortions of the cell shape are reminiscent of those observed in red blood cells upon aggregation of E6V hemoglobin. This mutation causes hemoglobin tetramers to assemble into rigid fibrils that distort the normally disk-shaped cell to a sickle shape. In contrast, runaway domain-swapped filaments are connected by flexible hinge loops that typically confer a ‘beads-on-a-string’ appearance. These incongruities prompted us to examine the GFP-RNase protofilament structure for attributes that differ from those of filaments formed by runaway domain swapping and facilitate nucleation and mechanical rigidity.

Consideration of the monomeric state illuminates the defining attribute that distinguishes GFP-RNase filament formation from runaway domain swapping and explains why GFP-RNase filaments grow more robustly than runaway domain swapped filaments: the GFP-RNase monomer cannot competitively bind its own helical appendage. A defining feature of runaway domain swapping is the high affinity for binding the swapping domain, whether nested within its own protomer (monomeric state) or swapped into a neighboring protomer (filamentous state). The equally deep energetic stability of both the monomeric and filamentous states impedes filament nucleation and growth in two ways: (i) it imposes a large kinetic barrier to the conversion from monomer to filament because the intramolecular domain–domain interface in the monomer must be disrupted to form the equivalent intermolecular domain–domain interface in the filament, and (ii) it renders the monomeric state an energetically viable competitor of filament nucleation and growth because the two states contain equally stabilizing domain–domain interfaces.

In violation of a defining principle of runaway domain swapping, the helix-in-groove interface that stabilizes GFP-RNase protofilaments is unattainable within individual monomers. Indeed, the C-terminal helix of GFP-RNase (donor domain) cannot bind its complementary site within the same monomer (acceptor domain) because the acceptor site is out of reach; the hinge loop is too short (Fig. 4g). Self-complementation is further precluded by that fact that the acceptor site itself is not a single surface, but is assembled from two protomers (i+1 and i+2), which respectively contribute residues from opposite faces of the barrel (Figs. 2d and 4i). The unattainability of a self-complementary donor-acceptor interface within GFP-RNase monomers explains why this protein forms filaments more readily than those of proteins that undergo runaway domain swapping. Without a means of self-complementation, protomers in the monomeric state cannot attain the same low-energy state as do the protomers in the filament, thereby tilting the energy landscape in favor of filament growth (Fig. 4f–h). The absence of stable self-complementation also has the advantage of eliminating the kinetic barrier to filament formation: with no donor–acceptor interface in the monomer, there is no barrier imposed by its disruption, as there would be in runaway-domain-swapped filaments.

We discovered that many natural filaments display donor–acceptor linkages analogous to those observed in the GFP-RNase filaments (Supplementary Table 1). For example, in filaments of nucleocapsid protein from phlebovirus, each protomer donates a helix to an acceptor domain in the neighboring protomer. Similarly, in pilus rods of fimA protein, each protomer donates an N-terminal strand to its neighboring protomer. Moreover, GFP-RNase, nucleocapsid protein and fimA are all incapable of self-complementation. The mechanism of pilus rod formation has been named ‘donor-strand exchange’, but a more general name is needed to encompass those filaments assembled by donation of elements other than strands. We refrain from using the previously coined term ‘obligate domain swapping’ because we have shown the mechanism is not a true domain swap. Instead, we coin the term ‘runaway domain coupling’ to describe all filaments like GFP-RNase in which protomers interface through distinct donor and acceptor domains—interfaces that are unattainable in individual monomers.

Runaway domain coupling describes many protofilament assemblies, but it does not explain why some, like GFP-RNase, are mechanically rigid. Indeed, a flexible linker is characteristic of this mechanism. Rigidity, on the other hand, is characteristic of domain stacking, in which copies of a single rigid domain stack together in uninterrupted succession (Fig. 4a). In GFP-RNase protofilaments, protomers stack through patches on the rigid barrel domains, thus rigidifying the otherwise flexible domain coupled linkages (Fig. 2f). This combination of domain coupling and stacking is also evident in naturally occurring recA nucleoprotein filaments; even in the absence of nucleic acid, recA forms filaments with a persistence length of about 100 nm (Fig. 4i and Supplementary Table 1). In contrast, the runaway-domain-coupled filaments of phlebovirus nucleocapsid protein have no supplementary connection through domain stacking (Fig. 4h). Accordingly, the filaments give the appearance of a random walk in electron micrographs. Hence, domain stacking imparts rigidity to a protofilament, acting alone or in combination with other mechanisms.

Bridging is yet another mechanism of filament formation that contributes to the mechanical rigidity of GFP-RNase protofilaments. Bridging interfaces join nonconsecutive protomers in the protofilament, creating a mutually supportive closed network that includes the intervening ‘bridged’ protomer(s) (Fig. 4b). In GFP-RNase protofilaments, a bridging interface exists between protomers i and i+2, as part of the helix-in-groove interaction (Fig. 4k). This interface, in turn, bolsters contacts between these protomers and the intervening protomer i+1. Bridging is also observed in tropomyosin, which forms a coiled-coil bridge across protomers in actin filaments, thereby reinforcing stacking between actin domains and doubling the persistence length (Supplementary Table 1). Bridging also helps assemble soluble E6V hemoglobin tetramers into cell-deforming rods and tubulin protofilaments into microtubules. Cooperativity is implied by bridging because assembly of two protomers enhances binding of a third protomer.

Genesis of protofilaments by either runaway domain coupling or domain stacking appears relatively difficult to achieve through a single, de novo interface alone. The interface would have to be sufficiently large and snug for the protofilament to have a significant lifetime, presumably requiring many iterations of genetic selection. Circular permutation may offer a shortcut by exploiting a preexisting intramolecular interface. For example, circular permutation may shift an N-terminal element to the C terminus. From its new location, the element may not reach its own complementary site near the N terminus, but it might easily reach the complementary site of another protomer—a perfect setup for runaway domain coupling (Supplementary Fig. 7). Circular permutation occurs with sufficient frequency in nature to be a plausible step in protofilament genesis; according to one study, as many as 47% of all domains examined have a circularly permuted homolog. Simpler still, runaway-domain-coupled protofilaments might also arise from insertion of a stop codon in a multidomain oligomeric protein, as engineered
in the T7 primase-helicase gene. Evidently, the removal of some structural elements disrupted cyclic oligomerization in favor of a runaway polymerization.

Surprisingly, protofilaments might be generated faster when protomer–protomer interfaces are divided over multiple surfaces, as observed in GFP-RNase protofilaments (Fig. 4k), rather than collected in a single surface. Contrary to the elevated level of complexity inferred by the presence of three distinct interfaces in GFP–RNase protofilaments (runaway domain coupling, stacking and bridging), this combination allows each of the interfaces to be smaller and therefore relatively easier to achieve. In fact, we found that a majority of crystal structures of monomeric proteins demonstrate crystal packing interfaces as large as the individual interfaces observed in GFP–RNase protofilaments (Supplementary Note 2).

Flexible linkage between domains, such as occurs in runaway domain coupling, seems advantageous for consolidating these smaller interfaces into a network stable enough to support filament genesis. Consider an early stage of protofilament evolution in which domains can stack but the stacking interface is too small to be stable. Suppose that this stacking domain is linked to an accessory domain with the potential to evolve a second, stabilizing interface in parallel with the first. If the linker between domains is rigid, only a small portion of the accessory domain may be in reach of establishing a new interface without disturbing the first. However, if the linker is flexible, as is the case with domain coupling, more residues of the accessory domain may be within reach of the neighboring protomer, improving chances of encountering a snug interface. Indeed, one-third of all the naturally occurring filament structures that we surveyed involve combinations of runaway domain coupling with stacking (Supplementary Table 1). Notably, runaway domain coupling, stacking and bridging were revealed recently in CTP synthase, a representative of a class of metabolic enzymes that evolved filamentation as a survival response to certain cellular conditions.

The advantage of flexibility for consolidating stacking interfaces comes with an entropic penalty—the immobilization of the interdomain loop. This penalty partially diminishes the stability of the protofilament, perhaps explaining the temperature sensitivity of filament formation by some of the GFP–RNase constructs (Fig. 1b,c) and the sensitivity of CTP synthase filamentation to cellular conditions. These two filaments also score lower values of interface area and solvation free energy than the majority of naturally occurring filamentations. These two filaments also score lower values of interface area and solvation free energy than the majority of naturally occurring filamentations. These two filaments also score lower values of interface area and solvation free energy than the majority of naturally occurring filamentations.

A strategy of combining runaway domain coupling with stacking may offer new advantages for filament design. Notably, flexible
linkers were not in fact included in the few examples where globular or helical proteins were successfully engineered to form filaments\textsuperscript{1,4,46}. In all these cases, alignment between protomers was designed computationally, thereby eliminating the need for flexibility to attain alignment. However, the success of most of these designs depended on the presence of a naturally evolved dimerization interface. We propose that a design strategy of combining a runaway domain coupling with stacking could open up filamentation to the majority of monomeric proteins, employing interfaces as small as the average crystal packing interface.

**Methods**

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41594-018-0096-7.

Received: 8 December 2017; Accepted: 21 June 2018; Published online: 3 August 2018

**References**

1. Shih, Y.-L. & Rothfield, L. The bacterial cytoskeleton. *Microbiol. Mol. Biol. Rev.* 70, 729–754 (2006).

2. Jockusch, B.M. & Graumann, PL. The long journey: actin on the road to pro- and eukaryotic cells, in *Reviews of Physiology, Biochemistry and Pharmacology* vol. 161 (eds. Amara, S. G. et al.) 67–85 https://doi.org/10.1007/112_2011_1 (Springer, Berlin, 2011).

3. Bernard, R., Lind, A. E. & Eiteman, T. J. G. An archaeal origin for the actin cytoskeleton: implications for eukaryogenesis. *Commun. Integr. Biol.* 4, 664–667 (2011).

4. Wickstead, B. & Gull, K. The evolution of the cytoskeleton. *J. Cell Biol.* 194, 513–525 (2011).

5. O’Connell, J. D., Zhao, A., Ellington, A. D. & Marcotte, E. M. Dynamic reorganization of metabolic enzymes into intracellular bodies. *Annu. Rev. Cell Dev. Biol.* 28, 1–21 (2012).

6. Petrovskaya, I. et al. Filament formation by metabolic enzymes is a specific adaptation to an advanced state of cellular starvation. Elife https://doi.org/10.7554/eLife.02409 (2014).

7. Aughey, G. N. & Liu, J.-L. Metabolic regulation via enzyme filamentation. *Crit. Rev. Biochem. Mol. Biol.* 51, 282–293 (2015).

8. Barry, R. M. et al. Large-scale filament formation inhibits the activity of CTP synthetase. *Elife* 3, e06368 (2014).

9. Liu, Y. & Eisenberg, D. 3D domain swapping: as domains continue to swap. *Protein Sci.* 11, 1285–1299 (2002).

10. Bennett, M. J., Sawaya, M. R. & Eisenberg, D. Deposition diseases and 3D domain-swapping, *Protein Sci.* 14, 811–824 (2005).

11. Ogihara, N. L. et al. Design of three-dimensional domain-swapped dimers depends on the presence of a naturally evolved dimerization or helical proteins were successfully engineered to form filamentation. *Protein Sci.* 14, 811–824 (2005).

12. O’Connell, J. D., Zhao, A., Ellington, A. D. & Marcotte, E. M. Dynamic reorganization of metabolic enzymes into intracellular bodies. *Annu. Rev. Cell Dev. Biol.* 28, 1–21 (2012).

13. Petrovskaya, I. et al. Filament formation by metabolic enzymes is a specific adaptation to an advanced state of cellular starvation. *Elife* 3, e06368 (2014).

14. Liu, Y. & Eisenberg, D. 3D domain swapping: as domains continue to swap. *Protein Sci.* 11, 1285–1299 (2002).

15. Bennett, M. J., Sawaya, M. R. & Eisenberg, D. Deposition diseases and 3D domain swapping, *Protein Sci.* 14, 811–824 (2005).

16. Ogihara, N. L. et al. Design of three-dimensional domain-swapped dimers and fibrous oligomers. *Proc. Natl Acad. Sci. USA* 98, 1404–1409 (2001).

17. Huntington, J. A. et al. A 2.6 Å structure of a serpin polymer and implications for conformational disease. *J. Mol. Biol.* 293, 449–455 (1999).

18. Strop, P., Smith, K. S., Iverson, T. M., Ferry, J. G. & Rees, D. C. Crystal structure of the “cab”-type β class carbonic anhydrase from the archaeon *Methanobacterium thermoautotrophicum*. *J. Biol. Chem.* 276, 10299–10305 (2001).

19. Guo, Z., & Eisenberg, D. Runaway domain swapping in amyloid-like fibrils of T7 endonuclease I. *Proc. Natl Acad. Sci. USA* 103, 8042–8047 (2006).

20. Wahlbom, M. et al. Fibrolignocytic oligomers of human cystatin C are formed by propagated domain swapping, *J. Biol. Chem.* 282, 18318–18326 (2007).

21. Jaksöoki, M. 3D domain swapping, protein oligomerization, and amyloid formation. *Acta Biochim. Pol.* 48, 807–827 (2001).

22. Sambashivan, S., Liu, Y., Sawaya, M. R., Gingery, M. & Eisenberg, D. Amyloid-like fibrils of ribonuclease A with three-dimensional domain-swapped and native-like structure. *Nature* 437, 266–269 (2005).

23. Teng, P. K. & Eisenberg, D. Short protein segments can drive a non-fibrillizing protein into the amyloid state. *Protein Eng. Des. Sel.* 22, 531–536 (2009).

24. Goldschmidt, L., Teng, P. K., Rek, R., & Eisenberg, D. Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc. Natl Acad. Sci. USA* 107, 3487–3492 (2010).

25. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* 234, 946–950 (1993).

26. Yang, F., Moss, L. G. & Phillips, G. N. Jr. The molecular structure of green fluorescent protein. *Nat. Biotechnol.* 14, 1246–1251 (1996).

27. Poonsting, H., Henriek, K. & Thornton, J. M. Discriminating between homodimeric and monomeric proteins in the crystalline state. *Proteins* 41, 47–57 (2000).

28. Wellem, T. E. & Josephs, R. Crystallization of deoxyhemoglobin S by fiber alignment and fusion. *J. Mol. Biol.* 135, 651–674 (1979).

29. Harrington, D. J., Adachi, K. & Royer, W. E. Jr. The high resolution crystal structure of deoxyhemoglobin S. *J. Mol. Biol.* 272, 398–407 (1997).

30. Liu, Y., Gotte, G., Libonati, M. & Eisenberg, D. Structures of the two 3D domain-swapped RNase A trimers. *Protein Sci.* 11, 371–380 (2002).

31. Wang, Z., Kishchenko, G., Chen, Y. & Josephs, R. Polymerization of deoxy-sickle cell hemoglobin in high-phosphate buffer. *J. Struct. Biol.* 131, 197–209 (2000).

32. Lomas, D. A., Evans, D. L., Finch, J. T. & Carroll, R. W. The mechanism of α,-antitrypsin accumulation in the liver. *Nature* 357, 605–607 (1992).

33. Bennett, M. J., Choe, S., & Eisenberg, D. Domain swapping: entangling alliances between proteins. *Proc. Natl Acad. Sci. USA* 91, 3127–3131 (1994).

34. Raymond, D. D., Pipper, M. E., Gerrard, S. R., Skiniotis, G., & Smith, J. L. Phleboviruses encapsidate their genomes by sequestering RNA bases. *Proc. Natl Acad. Sci. USA* 109, 1928–1931 (2012).
experiments and protein purifications, and contributed to writing the manuscript. D.S.E. discussed plans and results with A.H. and M.R.S. A.H. designed the study and contributed to writing the manuscript. M.R.S. determined and analyzed the crystal structures, performed in cellulo diffraction experiments and contributed to writing the manuscript.

**Competing interests**
The authors declare no competing interests.

**Additional information**
Supplementary information is available for this paper at https://doi.org/10.1038/s41594-018-0096-7.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.H. or M.R.S.

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

Plasmids, strains and cell growth. GFP fusions were produced from pBR322-derived plasmids under the control of the arabinose-inducible promoter pBAD. Specifically, the constructs depicted in Fig. 1 (top to bottom) were produced from plasmids pLM163, pLM168, pLM180, pLM179, pLM177, pLM178, pLM176, pLM187, pLM192 and pLM198. Plasmid pLM136 encodes the full-length RNase A fusion with the amyloidogenic peptide GGGVIA (derived from Aβ) inserted between RNase A residues 113 and 114.

Constructs with only eight RNase A residues included an additional aspartate residue with or without a C-terminal hexahistidine tag. Overnight cultures of *E. coli* BW27785 were transformed with the appropriate GFP-fusion construct were diluted to an OD$_{600}$ of 0.02 in 50 ml LB supplemented with carbencillin (100 µg/ml), grown for 30 min at 30°C, and induced with t-arabinose (0.2% w/v). For experiments in which early stationary-phase cultures were examined, cells were harvested after 4 h induction at 30°C. Otherwise, cultures were induced overnight (~20 h) at 30°C. Where indicated, both early stationary-phase and overnight cultures were refrigerated at 4°C for 2 h or 24 h to enhance formation of GFP rods.

Fluorescence microscopy. Cells were imaged after overnight induction of fusion protein synthesis either without any subsequent incubation or following incubation for the time specified at 4°C. Cells were harvested at 3,000 g, resuspended in PBS, and spotted onto agarose pads (1% w/v in PBS; Seakem LE Agarose, Lonza) for visualization with a UplanFLN 10X phase contrast objective on an Olympus BX61 microscope as described elsewhere. Images were captured with a CoolSnapHQ camera (Photometrics) and the Metamorph software package, version 6.1 (Universal Imaging). Exposures were typically 50–100 ms. Images were cropped and adjusted using Metamorph or ImageJ.

Bacterial extract preparation and western blotting. Cultures (50 ml) were grown as described above, and cell densities were recorded (OD$_{600}$). The cultures were harvested at 3,000 g, washed once in cold PBS, centrifuged 10 min at 3,000 × g, and the supernatant mixture was added to a 1.5 ml Eppendorf tube and drained. The column was washed with 1 ml PBS and the supernatant fraction was incubated with 3 ml prewashed Ni-NTA resin (Qiagen) at 4°C for 1 h with rocking. After this incubation step, the resin/supernatant mixture was added to a gravity column and drained. The column was washed three times with a buffer containing 50 mM HEPES, pH 7.9, 300 mM NaCl, 5 mM BME and increasing concentrations of imidazole (0, 20 mM and 60 mM, respectively). Bound protein was eluted in 8 ml elution buffer (50 mM HEPES, pH 7.9, 300 mM NaCl, 5 mM BME, 200 mM imidazole), and 1 ml aliquots were frozen on dry ice. The purification protocol was adapted from that of Ormø et al.

Crystallization. GFP-RNaseA (1–8) in space group P2$_1_2_1_2_1$, form 1. The reservoir contained 20% ethanol and 30% (v/v) 2-methyl-2,4-pentanediol (Qiagen MPD Suite condition 90 (H6)). The crystal grew at 20°C.

GFP-RNaseA (1–8) in space group P2$_2_1_2_1$, form 1. The reservoir contained 20% sodium acetate, pH 5.0, and 65% (v/v) 2-methyl-2,4-pentanediol (Qiagen MPD Suite condition 88 (F8)). The crystal grew at 4°C.

GFP-RNaseA (1–8) in space group P2$_2_2_2$. The reservoir contained 0.1 M sodium citrate, 0.1 M HEPES, pH 7.5, and 10% (v/v) 2-methyl-2,4-pentanediol (Qiagen MPD Suite condition 73 (G1)). The crystal grew at 4°C.

GFP-RNaseA (1–8) in space group P2. The reservoir contained 0.2 M KCl, 0.005 M MgCl$_2$, 0.05 M sodium cacodylate, pH 6.5, and 10% (w/v) 1.6 hexamidog (Qiagen Nucleix Suite condition 21 (B9)). The crystal grew at 4°C.

Data collection and processing. Diffraction data from GFP-RNaseA (1–8) crystals in space group P2$_2_1_2_1$, form 1, C2 and P2, were collected at the Advanced Photon Source (APS) 24-ID-E using the IO-Forelix detector. The reservoir from crystal form 2 of space group P2,2,2, was collected at the beamline 24-ID-E using an ADSC Quantum 315 CCD detector. Each dataset was collected from a single crystal at a temperature of 100 K. The following X-ray wavelengths were used for crystal forms P2,2,2, form 1, P2,2,2, form 2, C2 and P2, respectively: 0.979 Å, 0.979 Å, 0.979 Å and 1.4760 Å. Data were processed using the XDS package for all crystals. In addition, autoPROC was used for processing datasets in P2,2,2, form 1 and C2; aimless was used for scaling in P2,2,2, form 1. Data collection statistics are reported in Table 1.

Structure determination and refinement. The first crystal form we obtained of GFP-RNaseA (1–8), form 1, was isomorphous with a previously determined GFP structure, PDB 4P1Q. We used these coordinates as a starting model in the refinement of the GFP-RNaseA (1–8) with BUSTER. Difference density maps revealed the position of the N-terminal fusion, and these atoms were built using the graphics program COOT. The remaining structures were determined by molecular replacement using the program Phaser. The search model for molecular replacement in space group C2 was the refined structure from space group P2,2,2, form 1. The search model for molecular replacement in space group P2, and P2,2,2, form 2 was the refined structure from space group C2. All models were built and refined using the same software as described above for P2,2,2, form 1. Structures were illustrated using the program PyMOL.

Fiber diffraction simulation. Crystal diffraction intensities were cylindrically averaged around the filament axis to simulate the radial disorder of a filament using custom-written Fortran code.

Identification and analysis of potential protofilament-forming interfaces in the Protein Data Bank. To identify how many globular, monomeric proteins are poised to form protofilaments, we began by screening 130,446 PDB entry headers to identify monomers. We defined ‘monomers’ as entries that contained only a single protein chain, were determined by crystallography, had a PDBsum score of greater than 0.30, and had no biological assembly operators other than the identity matrix. We also excluded redundant entries defined as having the same MOLECULE record, unit cell lengths identical to ±3 Å and unit cell angles identical to ±5°. We found 12,933 entries that passed these criteria for monomers.

To determine how many of these monomers were poised to form screw-related protofilaments, we performed the following steps. First, we applied crystallographic operations to bring the molecule as close as possible to the origin, generating a reference molecule. Then, we applied crystallographic screw operations to the reference molecules, allowing up to three unit cell translations in any combination along positive and negative a, b, and c. Screw operators were accepted as potential protofilament generators only if the reference molecule and the operated molecule were close enough to touch. We defined touching distance as any distance less than twice the distance of the furthest atom of the reference molecule from its center of mass. These criteria produced 73,968 screw-related pairs. Surface areas buried by these interfaces were calculated using the CCPM program areaimol. If these 26,375 screw-related pairs meet the criterion of a quasi-stable interface as defined by equaling or surpassing the 901 Å² of buried surface area per square observation of barrel domains of our GFP fusion protein protofilament. That is, 9,119 PDB entries, or 74% of the monomer entries, meet our criteria for interface size and symmetry. Including a constraint that one of the termini be located within 12 Å of the intermolecular interface (a threshold defined by our GFP fusion protein) leaves 53% of monomeric PDB entries meeting these combined criteria. That is, over half the monomer entries in the PDB (6,816 entries) are poised for filament formation.

Simulation of helix-forming propensity for randomly generated 13-mers. We used a random number generator and probability-weighted amino acid frequencies in vertebrates to calculate 1 million peptide sequences. We scored the sequences for helix propensity using Chou & Fasman rules. Using two different seeds for the random number generator, we found 127,952 and 128,171 sequences that passed the criteria for helix. That is approximately 13%.
Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Code availability. The custom-written Fortran program used to calculate the filament diffraction pattern displayed in Fig. 3e is available upon request.

Data availability. Atomic coordinates of the X-ray structures of GFP-RNase(1−8) in different space groups were deposited in the Protein Data Bank with accession codes 5HGE (P2₁2₁2₁, form 1), 6AS9 (P2₁2₁2₁, form 2), 5HBD (C2) and 5HW9 (P2₁).

References
49. Garrity, S. J., Sivanathan, V., Dong, J., Lindquist, S., & Hochschild, A. Conversion of a yeast prion protein to an infectious form in bacteria. Proc. Natl Acad. Sci. USA 107, 10596–10601 (2010).

50. Khlebnikov, A., Datsenko, K. A., Skaug, T., Wanner, B. L. & Keasling, J. D. Homogeneous expression of the pBAD promoter in Escherichia coli by constitutive expression of the low-affinity high-capacity AraE transporter. Microbiology 147, 3241–3247 (2001).

51. Doan, T., Marquis, K. A. & Rudner, D. Z. Subcellular localization of a sporulation membrane protein is achieved through a network of interactions along and across the septum. Mol. Microbiol. 55, 1767–1781 (2005).

52. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9, 671–675 (2012).

53. Kabsch, W. XDS. Acta Crystallogr. D Biol. Crystallogr. 66, 125–132 (2010).

54. Vonrhein, C. et al. Data processing and analysis with the autoPROC toolbox. Acta Crystallogr. D Biol. Crystallogr. 67, 293–302 (2011).

55. Winna, M. D. et al. Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. 67, 235–242 (2011).

56. Auerbach, D. et al. Replacement of highly conserved E222 by the photostable non-photoconvertible histidine in GFP. ChemBioChem 15, 1404–1408 (2014).

57. Bricogne, G. et al. Buster version 1.10.0 (Global Phasing Ltd., Cambridge, 2016).

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

59. McCoy, A. J. et al. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674 (2007).

60. Delano, W. The PyMOL Molecular Graphics System (Schrödinger LLC, New York, 2017).

61. Chou, P. Y. & Fasman, G. D. Empirical predictions of protein conformation. Annu. Rev. Biochem. 47, 251–276 (1978).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted. Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
- Clearly defined error bars
  - State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

- We used custom written software from the Advanced Photon Source beamlines 24-ID-C and 24-ID-E. The software is available to all beamline users.

Data analysis

- We used CCP4, and XDS software, as cited in the manuscript. The custom written Fortran program used to calculate fibril diffraction patterns is available upon request.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Our crystallographic diffraction data were deposited in the Protein Data Bank, and freely available.
Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☑ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | The structures were determined from single crystals, which is a standard crystallographic practice. |
|-------------|---------------------------------------------------------------------------------------------------|
| Data exclusions | No data were excluded. |
| Replication | Procedures are described for preparing the protein, growing the crystals, and collecting the data. |
| Randomization | No experiments were performed that required allocation into experimental groups. |
| Blinding | Standard crystallographic practices use a Free R test set, which is a set of 10% of the reflections chosen at random and withheld from the refinement process. |

Reporting for specific materials, systems and methods

| Materials & experimental systems | Methods |
|----------------------------------|---------|
| n/a Involved in the study | n/a Involved in the study |
| ☑ Unique biological materials | ☑ ChIP-seq |
| ☑ Antibodies | ☑ Flow cytometry |
| ☑ Eukaryotic cell lines | ☑ MRI-based neuroimaging |
| ☑ Palaeontology | |
| ☑ Animals and other organisms | |
| ☑ Human research participants | |