nature microbiology

ARTICLES

https://doi.org/10.1038/s41564-019-0544-0

The structure of bactofilin filaments reveals their mode of membrane binding and lack of polarity

Xian Deng[1,2], Andres Gonzalez Llamazares[1,2], James M. Wagstaff[1], Victoria L. Hale[1], Giuseppe Cannone[1], Stephen H. McLaughlin[1], Danguole Kureisaite-Ciziene[1] and Jan Löwe[1*]

Bactofilins are small β-helical proteins that form cytoskeletal filaments in a range of bacteria. Bactofilins have diverse functions, from cell stalk formation in *Caulobacter crescentus* to chromosome segregation and motility in *Myxococcus xanthus*. However, the precise molecular architecture of bactofilin filaments has remained unclear. Here, sequence analysis and electron microscopy results reveal that, in addition to being widely distributed across bacteria and archaea, bactofilins are also present in a few eukaryotic lineages such as the Oomycetes. Electron cryomicroscopy analysis demonstrated that the sole bactofilin from *Thermus thermophilus* (TtBac) forms constitutive filaments that polymerize through end-to-end association of the β-helical domains. Using a nanobody, we determined the near-atomic filament structure, showing that the filaments are non-polar. A polymerization-impairing mutation enabled crystallization and structure determination, while reaffirming the lack of polarity and the strength of the β-stacking interface. To confirm the generality of the lack of polarity, we performed coevolutionary analysis on a large set of sequences. Finally, we determined that the widely conserved N-terminal disordered tail of TtBac is responsible for direct binding to lipid membranes, both on liposomes and in *Escherichia coli* cells. Membrane binding is probably a common feature of these widespread but only recently discovered filaments of the prokaryotic cytoskeleton.

Most bacteria and archaea contain functional protein filaments that have been collectively termed prokaryotic cytoskeletons[1]. Among those, cytomotive filaments, for example of the actin and tubulin types, are characterized by complex dynamics. Most prokaryotes, however, also encode protein filaments that do not belong to the actin and tubulin families and that instead act as molecular scaffolds[2]. In 2010, Thanbichler and co-workers recognised the existence of a conserved and widespread family of bacterial proteins that form constitutive and very stable filaments in vitro and perform scaffolding roles in cells; they named these proteins bactofilins[3].

Bactofilins contain a small domain of about 110 amino acids (Pfam domain PF04519), which has been found in many bacterial genomes, often within multiple genes[2]. Most sequences consist of one bactofilin domain flanked by presumably disordered proline-rich tails at the N and C termini, although this is not universal; enterobacterial bactofilins, including *Proteus mirabilis* CcmA, have predicted transmembrane helices in the N-terminal tail[3,4]. Bactofilin filaments are very stable, as they are largely insensitive to pH, salt concentration and chelating agents; hence they are always filamentous when purified from source or expressed heterologously[3,5].

The structure of the conserved bactofilin domain was solved by solid state NMR[6], with supporting evidence from sequence-based modelling and electron microscopy of the filaments[7–9]. The bactofilin domain has a right-handed β-helical fold, with six windings of ~17 amino acid residues producing triangular-shaped monomers that measure roughly 3 nm along the β-helical axis. However, it has not proved possible to determine the filament structure of bactofilins.

The understanding of bactofilin function was limited to a small number of examples. *Caulobacter crescentus* bactofilins BacA and BacB were identified in a localization screen for proteins involved in stalk formation[3]. BacAB are expressed throughout the cell cycle and condense at the stalk-forming site within the cell at the onset of the S phase. These bactofilins have been found to directly interact with the cell wall synthesis enzyme PbpC; deletion of either the filaments or PbpC produces much shorter stalks. In cells, BacAB have been reported to form filaments or sheets close to the cell's inner membrane and overexpression deforms the cells, making it possible that the filaments have intrinsic curvature and bind membranes. Some biochemical evidence suggested that BacAB are peripheral membrane proteins[3].

Other well-investigated bactofilins are BacM, N, O and P from *Myxococcus xanthus*. A *bacM* gene knockout formed 'crooked' cells that have increased sensitivity to antibiotics[5]. BacNOP, in contrast, copolymerize into filaments that recruit the ParABS chromosome segregation machinery to subpolar regions of the cells[10]. BacP was also reported to be involved in type IV pilus localization together with the GTPase SofG, both of which are important for the direction of motility of *M. xanthus* cells[11].

Helicobacter pylori contains a single bactofilin called CcmA. A *ccmA* gene knockout completely abolished the characteristic helical shape of the cells and a model was put forward in which CcmA bactofilin filaments position lytic endopeptidases Csd1–3 in the periplasm to remodel the shape-giving cell wall[12,13]. Bactofilins have also been described in *Leptospira biflexa*, with one of its bactofilin paralogues controlling the helical pitch of cells[14].

Although not described as such then, an early sighting of a bactofilin was CcmA from *P. mirabilis*, where it is involved in cellular motility[4], and a similar functional context was described for CcmA/bactofilin proteins in *Vibrio parahaemolyticus*[15] and *Bacillus subtilis*[16,17].

More functional investigations are needed and we hope to be able to facilitate this with the elucidation of the electron cryomicroscopy (cryo-EM) structure of the bactofilin filament from *Thermus thermophilus*. The filament is composed of domains that are stacked

[1]MRC Laboratory of Molecular Biology, Cambridge, UK. [2]These authors contributed equally: Xian Deng, Andres Gonzalez Llamazares.
*e-mail: jyl@mrc-lmb.cam.ac.uk

so that a continuous β-helical filament results. The subunits are arranged head-to-head, resulting in a filament that lacks polarity. This finding was confirmed by crystallography and coevolutionary analysis. We show that the filaments bind directly to membranes in vitro and when heterologously expressed in *Escherichia coli* cells, and that this interaction is enabled by a short conserved and hydrophobic motif in the N-terminal tail. Finally, we show that polymerizing bactofilins are not restricted to bacteria, with the visualization of bactofilin filaments from *Phytophthora infestans*, an Oomycete eukaryote.

Results

Bactofilin domains are found across the tree of life. The presence of bactofilins was reported and experimentally verified in several bacterial clades (Fig. 1a,b). To further assess the distribution of the conserved bactofilin domain we searched for the presence of the conserved PF04519/DUF 583 domain in a curated set of prokaryotic genomes[18], annotated with a standardized phylogenomic taxonomy (GTDB v86)[19] (resulting tree: Supplementary Dataset 1). We found that bactofilin domains were present in 82 of the 114 phylum-level clades (Fig. 1a). Thus, we conclude that bactofilins are very widely distributed in bacteria. We investigated the presence of bactofilin domains in archaea in the same way and found that many archaeal genomes harbour bactofilin domains. Within the phylum Halobacterota, more than 80% of genomes contained at least one bactofilin domain hit. It remains to be experimentally verified that these archaeal sequences encode polymerizing bactofilin.

While browsing the Pfam entry for the bactofilin domain PF04519 we noticed that several hits are listed within the eukarya[20]. We found convincing bactofilin-like sequences in two taxonomic clusters, one within the Stramenopiles (a deeply rooted eukaryotic clade, Fig. 1c) and another within the Ascomycete fungi (Supplementary Dataset 2). We recombinantly expressed a putative bactofilin gene (UniProt D0N980, PITG_07992, herein 'PiBac') from the Stramenopile plant pathogen *P. infestans*, a member of the Oomycete group, and found that it indeed forms bactofilin-like filaments (Fig. 1d). Published data show that the messenger RNA encoding PiBac is expressed before and during spore formation (Supplementary Fig. 1) and that the gene is conserved in context across the spore-forming Oomycetes (Supplementary Fig. 2)[21]. Therefore, bactofilin domains retaining the ability to polymerize are conserved in eukaryotic genomes, are expressed and probably play functional roles. The widespread distribution of polymerizing bactofilin domains underscored the need for a fuller structural understanding of bactofilin polymers.

Bactofilin from *T. thermophilus* (TtBac) forms completely β-helical filaments. TtBac seemed to be a promising candidate for structural determination due to its short N- and C-terminal tails and the biological tractability of the source organism. Purification of recombinantly expressed His_6-tagged TtBac (H_6-TtBac, Supplementary Table 3) under denaturing conditions and subsequent refolding allowed the visualization of TtBac filaments by negative stain electron microscopy and confirmed the presence of a polymerizing bactofilin in *T. thermophilus* (Fig. 2a). The H_6-TtBac filaments also recapitulated previously reported two-dimensional (2D) sheets that seem to be an intrinsic consequence of fibrillar bactofilin assemblies.

Visualization by cryo-EM of both H_6-TtBac and natively purified wild-type TtBac (TtBac-WT) filaments showed that the fibrils were prone to bundling and persisted unchanged over very wide ranges of pH and salt concentrations. Preparing the filaments at pH 11 allowed the production of cryo-EM grids with largely unbundled filaments that were amenable to structure determination but otherwise looked like filaments at more physiological pH values (Fig. 2b,c).

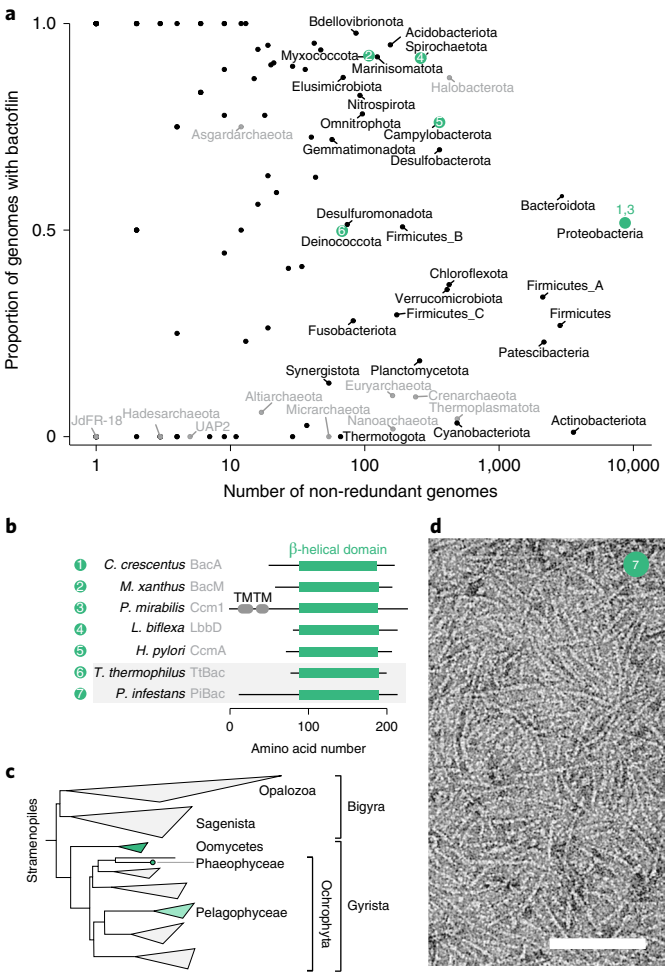

Fig. 1 | Bactofilins are highly conserved and widespread across bacteria and archaea and also occur in some eukaryotic organisms. a, Phylum-level clades from the GTDB taxonomy v86 (ref. [19]) of bacteria (black) and archaea (grey) are plotted as points with the number of non-redundant genomes in the clade against the proportion of genomes with at least one bactofilin hit. Bactofilin hits were HMMSEARCH results with Pfam PF04519, using an E-value cut-off of 1×10^{-10} for bacteria and 1×10^{-6} for archaea, chosen to retain clusters of similar, convincing sequences in each case. Precomputed HMMSEARCH results were retrieved from the AnnoTree server[18]. Clades containing the validated bactofilins in **b** are shown in green and numbered to correspond to **b**. **b**, A domain schematic of selected experimentally investigated bactofilins. The β-helical domains (green) are shown as regions aligning to the β-helical regions in *C. crescentus* BacA (ref. [6]) and *T. thermophilus* TtBac structures. UniProt accession codes for the proteins are shown: 1, Q9A753; 2, Q1CVJ5; 3, B4F0H9; 4, B0SPX3; 5 corresponds to GenBank CP001173.1 1607606:1607196 (UniProt shows an incorrect start codon, see Sycuro et al.[13]); 6, Q72HS6; 7, D0N980. TM, transmembrane domain. **c**, A phylogenomic tree of the eukaryotic Stramenopile clade[54]. The subclades with genomes containing putative bactofilins are highlighted in green; the lighter shade of the Pelagophycaeae corresponds to a patchier distribution (see text). **d**, A negative stain electron micrograph showing filaments formed by recombinant bactofilin PiBac (PITG_07992) from the Oomycete *P. infestans*. Scale bar, 100 nm.

Many 2D class averages showed two lines or protofilaments with twice as much signal in the top protofilament, suggesting that the majority of assemblies present in the grids were in fact made out of three protofilaments (Fig. 2d). The class averages all showed strong

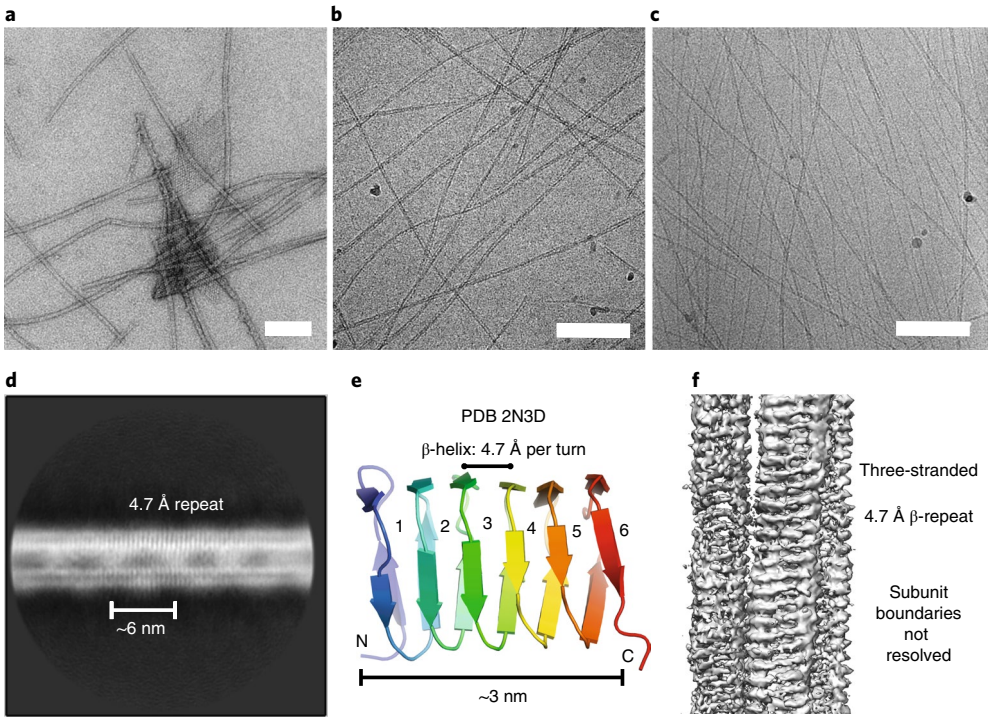

Fig. 2 | TtBac forms filaments that show a continuous β-stacking repeat of 4.7 Å. a, Hexahistidine-tagged and refolded TtBac (H_6-TtBac) forms filaments and 2D sheets as determined by negative stain electron microscopy. **b**, The same protein shown using cryo-EM. Differing filament thicknesses are caused by varying protofilament numbers. **c**, Untagged TtBac-WT protein expressed in *E. coli* and purified using a centrifugation protocol, imaged by cryo-EM, also showing protofilament numbers of 2–4. Scale bars in **a**–**c**, 100 nm. **d**, The 2D class average for TtBac-WT (imaged using cryo-EM), calculated using RELION v.3.0 (ref. [39]), showing a continuous repeat of 4.7 Å perpendicular to the filament axis. The upper strand shows stronger density, probably because many of the filaments averaged have three protofilaments. There are fuzzy bridges visible between the protofilaments, roughly 6 nm apart. **e**, A ribbon representation of the *C. crescentus* BacA bactofilin structure determined by solid state NMR[6]. Taking **d** into account, the structure suggests that bactofilin protofilaments are made by stacking the β-helical domains head-to-head and tail-to-tail. The 6-nm fuzzy bridges suggest that the structure might not be polar as it is difficult to produce those repeat lengths from a 3-nm-long β-helical domain otherwise. **f**, A preliminary RELION v.3.0 helical 3D reconstruction using the cryo-EM data[40], revealing the structure of a three-stranded TtBac-WT filament. As the individual domains show no features indicating their ends, the alignment does not converge and resolution remains low because the subunits cannot be registered correctly along the filament axis.

vertical striations within the protofilaments, spaced 4.7 Å apart, which is the distance between β-strands in a sheet. This indicates that the bactofilin subunits arrange into continuous β-helical protofilaments, as suggested previously[8,9]. The filaments also showed a weak ~6 nm repetitive feature visible between and in the protofilaments (Fig. 2d), roughly twice the size of a BacA monomer[6] (Fig. 2e). This indicates that the repeating unit of the filament might be a head-to-head dimer making polarity within the filament impossible.

Subsequent three-dimensional (3D) helical reconstruction of TtBac-WT filaments yielded a preliminary low-resolution structure that confirmed both the three-strandedness and the β-helical nature of the subunits, but did not allow any atomistic description of the monomer (Fig. 2f). The inability to reach a higher resolution was mainly caused by the similarity between each β-helical winding, which impeded the correct positioning of the subunits along each protofilament.

A near-atomic cryo-EM structure of the TtBac filament bound to a nanobody shows it to be non-polar. In an attempt to provide a solution to the computational problem of the subunit register along each protofilament, we raised *Lama glama* nanobodies against TtBac-WT. After selection of the suitable antibody clone nanobody 4 (NB4), cryo-EM of the wild-type filaments incubated with the anti-TtBac-WT nanobody showed fully decorated fibrils that formed superhelical structures unfit for structure determination (Fig. 3a). As a result we designed four mutations at the opposite end from the complementarity-determining region loops of the nanobody to counteract the bundling (L13S, Q15D, K45D and K66D on NB4-mut2), which almost completely abolished superhelicity (Fig. 3b–d). Images were clear enough to directly derive approximate helical parameters for image reconstruction (Fig. 3d). The 2D class averages of the decorated filaments almost exclusively showed filaments with two protofilaments, a clear signal for each nanobody at a spacing of 5.7 nm and the two β-helical protofilaments (Fig. 3e). The formation of higher order protofilament bundles is probably prevented by the bulky presence of the nanobodies (compare Figs. 2f and 3f).

Thus, by averaging ~346,000 helical segments and applying helical symmetry, the structure of TtBac-WT:NB4-mut2 filaments was solved to a nominal resolution of 3.4 Å (4.2 Å against the atomic model, Fig. 3f,g, Supplementary Table 1, Supplementary Fig. 3, Supplementary Video 1). Fitting the previously determined structure of BacA from *C. crescentus*[6] revealed the close similarity between the two (Fig. 3g). As in BacA, TtBac is made up of six right-handed windings of triangular parallel β-sheet structures with an exclusively hydrophobic core (a major contributor to the extreme stability of bactofilins) and two disordered terminal tails. Similar to the highly conserved glutamine and asparagine hydrogen bonds that give an increased sturdiness to amyloid fibres, bactofilins contain an outside bonding network comprising glutamates and aspartates placed diametrically above or below lysines and arginines. The structural simplicity of the TtBac filaments means that amino

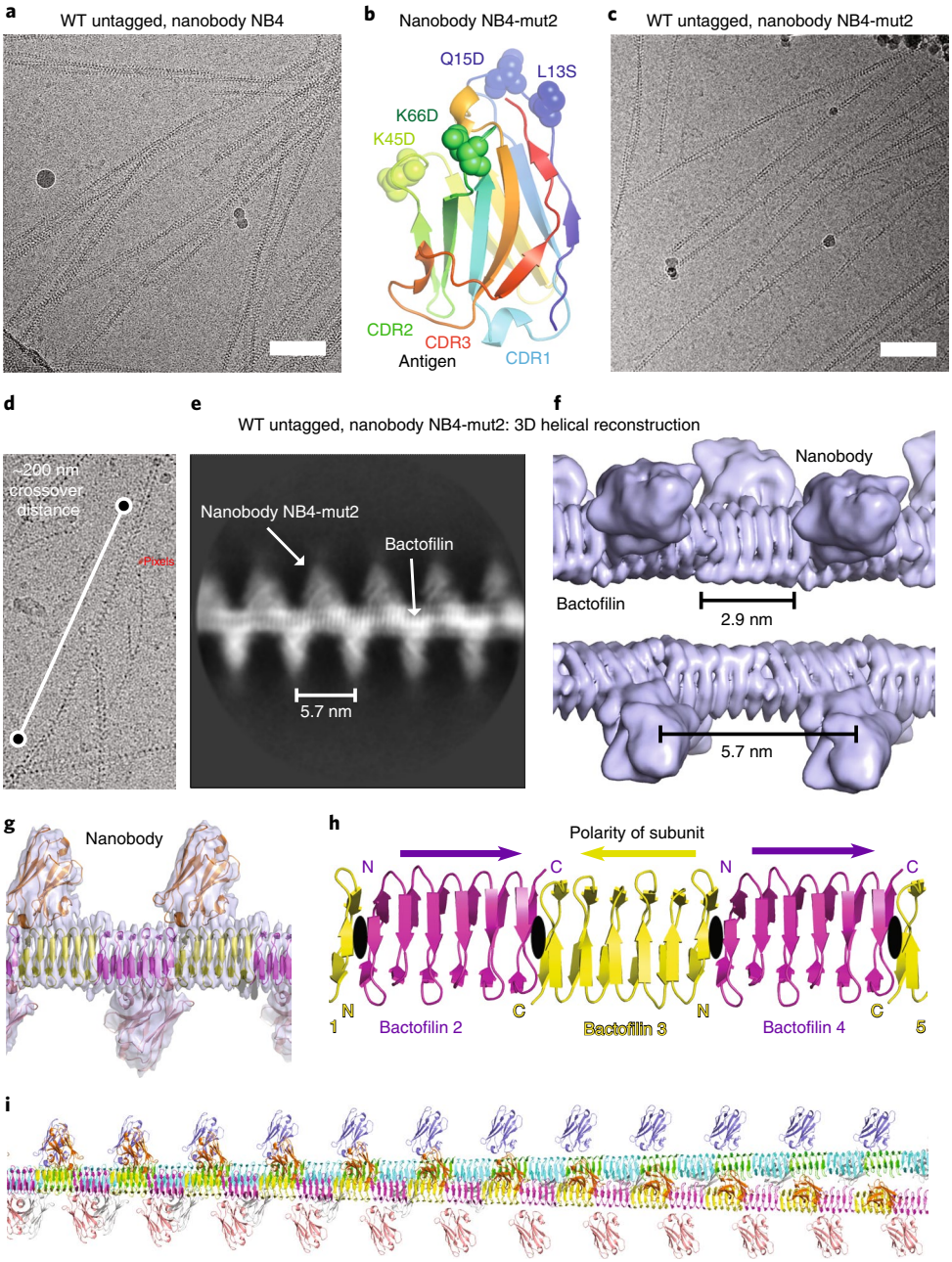

Fig. 3 | Cryo-EM structure of TtBac shows a head-to-head non-polar filament. a, To overcome the register problems highlighted in Fig. 2d,e, a nanobody against TtBac was raised, purified and added to the filaments before cryo-EM imaging. The nanobody clearly binds but causes severe bundling. Scale bar, 100 nm. **b**, On the assumption that bundling is caused by the antigen-distal surface of the nanobody, four mutations were introduced, yielding NB4-mut2. **c**, Using the modified nanobody, NB4-mut2 bundling is substantially reduced and cryo-EM images show many single filaments. Scale bar, 100 nm. **d**, An estimation of the crossover distance (corresponding to a 180° twist of the filament) to over 2,000 Å, indicating only very slowly twisting filaments of about 3–4° per 6 nm helical rise. **e**, The reference-free 2D class calculated using RELION v.3.0, showing the altered appearance of the TtBac filaments after addition of the NB4-mut2 nanobody (compare to Fig. 2d). The nanobody densities appear at a repeat distance of 5.7 nm, clearly indicating a non-polar arrangement of the subunits in each protofilament. **f**, A preliminary low-resolution (5–6 Å) 3D reconstruction using RELION v.3.0, clearly showing the continuous β-helical TtBac filament pair and the nanobodies attached at 5.7 nm distances to each other. **g**, The final density with fitted (but not refined) atomic models at around 4 Å resolution. Correctness of the handedness of the structure was verified using the crystal structure fit of the nanobody atomic model. **h**, A ribbon representation of the arrangement of individual TtBac subunits in each protofilament. Subunits are arranged in a head-to-head and tail-to-tail manner, with N and C termini coming together at alternating interfaces (N–N and C–C, leading to the 5.7 nm repeat of the nanobody binding). The two-fold axes at the N–N and C–C interfaces are nearly aligned, leading to a slowly twisting, overall helical filament. The black ovals indicate two-fold rotational symmetry axes. **i**, An atomic model of a longer stretch of the double-helical filament (see also Supplementary Video 2). The entire double filament has additional two-fold symmetry axes between the two protofilaments.

acid composition alone reflects their three main characteristics: a hydrophobic core consisting of leucine, valine or alanine (34% of residues); a strong interaction network between residues with surface charges, together comprising 28% of amino acids; and a triangular arrangement of β-sheets with glycines (13%) at each sharp corner.

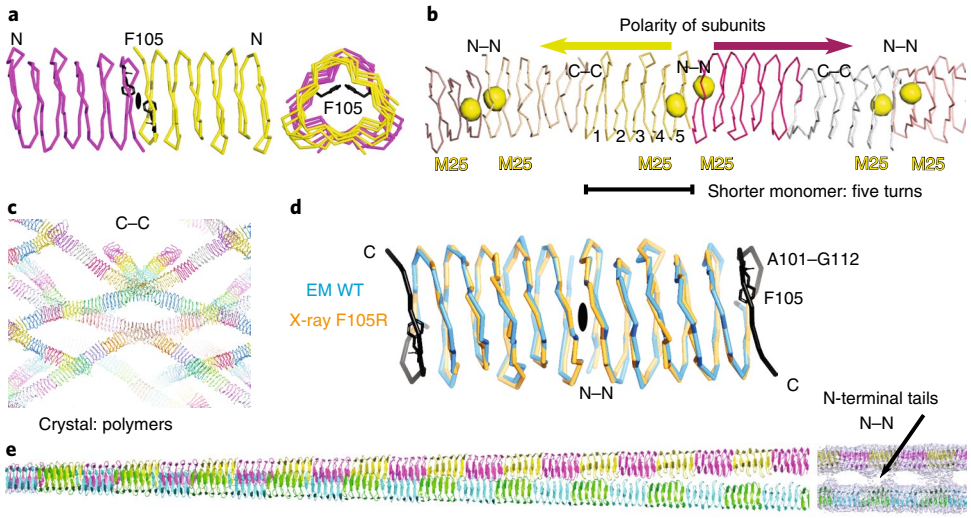

Fig. 4 | A polymerization-impaired TtBac mutant crystallizes as non-polar filaments with reduced subunit length. a, Based on previous studies (see ref. [9]), F105 in TtBac was mutated to R to produce a version that does not polymerize under normal conditions. Normally, F105 interacts with itself across the C–C interface along the TtBac protofilament and changing this residue to charged arginine impairs normal filament formation. TtBac(F105R) can be purified easily, especially when the disordered N-terminal region is also removed. **b**, The crystal structure of ΔN-TtBac(F105R). The structure was solved by SeMet single-wavelength anomalous diffraction (SAD) phasing and the resulting signals of the single Met25 residue all lie very close to each other, signalling the N–N interface and indicating a non-polar head-to-head and tail-to-tail arrangement. **c**, The very large unit cell of the TtBac crystals contains several strands of filaments, formed from 32 molecules per asymmetric unit. **d**, The mutant crystallized as a distorted filament because the F105R mutation caused residues 101–112 to become disordered, allowing an entirely new C–C β-helical interface to form. This rearrangement indicates the power of β-stacking interfaces, which also occur in amyloid structures, for example, and lead to the extraordinary stability of bactofilin filaments. The black oval depicts the point symmetry. **e**, The final atomic model of the double-helical TtBac filament. Using subtraction[15], the nanobody density was removed; the atomic model refined against the crystal structure was used as guidance to produce a reliable atomic model that was real-space refined against the cryo-EM density. Note the position of the interprotofilament density bridges, occurring at the locations of N–N interfaces, indicating that it is the N-terminal tails of TtBac that hold the two protofilaments together in double-helical filaments.

Primarily, the cryo-EM structure provides insight into the filamentous arrangement of bactofilins, showing that the monomers are ordered in a non-polar fashion with alternating head-to-head (N terminal to N terminal, N–N) and tail-to-tail (C terminal to C terminal, C–C) interfaces (Fig. 3h). This in turn produces two-fold symmetry at each homoterminal interface. The interfaces themselves have a very small misalignment of their two-fold axes, translating into a slight helical twist of 4.89° per dimer (Fig. 3i, Supplementary Video 2, Supplementary Fig. 4).

Crystal structure of a TtBac polymerization-impaired mutant confirms non-polar filament architecture. Although the filamentous structure converged to near-atomic resolution, the presence of anisotropy in the cryo-EM map, caused by the strong 4.7 Å repeat in a single direction, made it difficult to confidently assert the exact conformation of amino acids and other characteristics of the structure. Therefore, to validate the cryo-EM reconstruction, we aimed to solve the structure of non-polymerizing TtBac versions through X-ray crystallography. By introducing the mutation F105R (ref. [9]) into the C-terminal interface of TtBac we aimed to obstruct polymerization at this interface while preserving the head-to-head N–N contacts and thus producing a dimer of TtBac(F105R) (Fig. 4a). To further aid purification and crystallization we added a C-terminal His_6-tag and removed the first ten N-terminal residues that, according to the unresolved density in the 2D classes, seemed to be implicated in bundling (Fig. 2d). Indeed, the easy handling of ΔN-TtBac(F105R) was in stark contrast to the difficult purification of TtBac-WT. Moreover, visualization of the polymerization-impaired and non-bundling mutant by negative stain electron microscopy showed a lack of filaments when tested at fairly high concentrations (~2 mg ml^{-1}; not shown).

The crystal structure of ΔN-TtBac(F105R) was solved by making use of the single N-terminal Met25 for selenomethionine-substituted protein (SeMet) SAD phasing (Fig. 4b, Supplementary Table 2). In the solved structure, the unmutated N–N interface was intact and head-to-head, as unequivocally shown by the pairs of anomalous selenium peaks. The map also showed that crystallization occurred through formation of a C–C interface, leading to filaments in the crystals with 32 bactofilin subunits per asymmetric unit of the crystals (Fig. 4c). This unexpected polymer was formed by the unwinding of the sixth β-helical turn, essentially flipping out the R105 residue that would otherwise have impaired polymerization (Fig. 4d). This not only suggests that F105R impaired normal polymerization but also shows the outstanding propensity of β-stacks to form, as for example in amyloid fibres.

Aside from confirming the lack of polarity within the protofilaments, the structure of ΔN-TtBac(F105R) also suggested that the formation of higher order structures may well be caused by the N-terminal residues 1–10. The very large unit cell of the crystals contains well-dispersed filaments that do not form any type of doublets or bundles (Fig. 4c).

The refined crystallographic model (Supplementary Table 2) was fitted into a new cryo-EM map generated by signal subtraction of the nanobody density[22]. We added the remaining residues and refined in reciprocal and real space against the cryo-EM map to obtain a complete and reliable atomic model of the TtBac-WT bactofilin double filament (Fig. 4e, Supplementary Table 1). The resulting map reaffirmed that the bridges between the protofilaments that make the double filaments are formed by N-terminal residues (Fig. 4e, right).

Bactofilins form non-polar filaments as demonstrated by evolutionary coupling analysis. The high level of conservation between the monomer structures of TtBac and BacA (Protein Data Bank (PDB) ID 3N3D, root mean squared deviation of 1.7 Å over

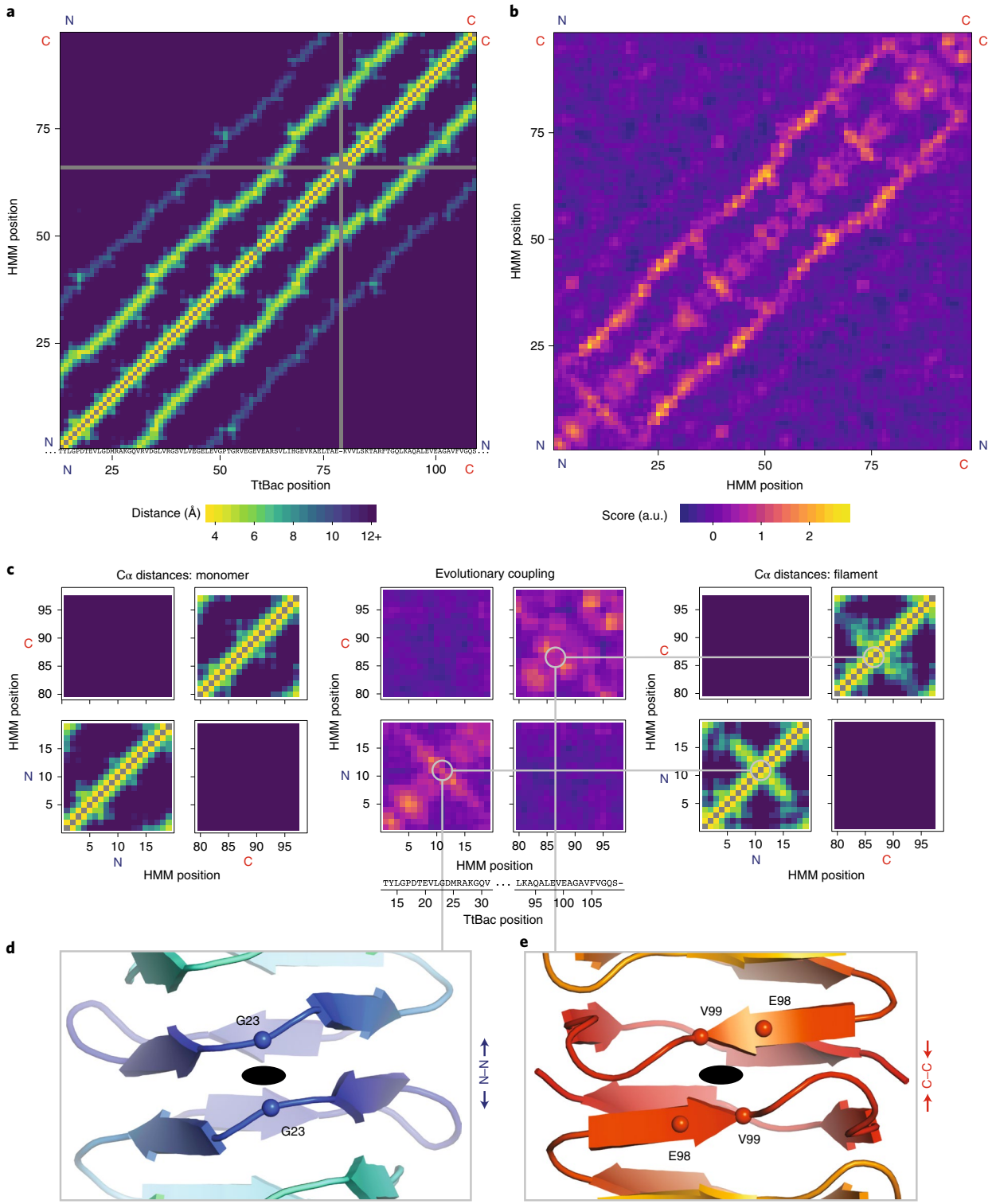

Fig. 5 | Evolutionary sequence coupling analysis reveals the conserved non-polar architecture of bactofilin filaments. a, The Cα distance matrix for the TtBac monomer structure within the cryo-EM filament structure. The ~17 amino acid repeat is obvious. The TtBac sequence was aligned to Pfam Hidden Markov Model (HMM) PF04519 using HMMALIGN; insertions and gaps relative to HMM are indicated below the *x*-axis. **b**, Evolutionary coupling scores calculated for an alignment of 12,646 bactofilin sequences. Scores in arbitrary units (a.u.) were smoothed using a 3×3 Gaussian kernel. The ~17 amino acid repeat is clear. **c**, Zoomed plots showing N–N, C–C and N–C interaction regions in the distance matrix plot and evolutionary coupling matrix. Left, magnifications of the corners of **a**. Middle, magnifications of the corners of **b** with the TtBac sequence added beneath. Right, minimum distances between Cα values in the magnified corners of the filament model. Intersubunit interactions appear in both the filament distance matrix and coupling matrix as diagonals going through the residues closest to the C_2 axis. **d**, The view along the C_2 axis at the N–N intersubunit interface of the TtBac filament model. Chains are displayed as a cartoon representation, coloured blue–red from the N terminus to the C terminus. The C_2 axis is shown as a black oval. Cα values for the residues closest to the axes are shown as spheres and labelled. Grey lines connect the panel to the relevant positions in **c**. **e**, As **d** but for the C_2 axis at the C–C intersubunit interface.

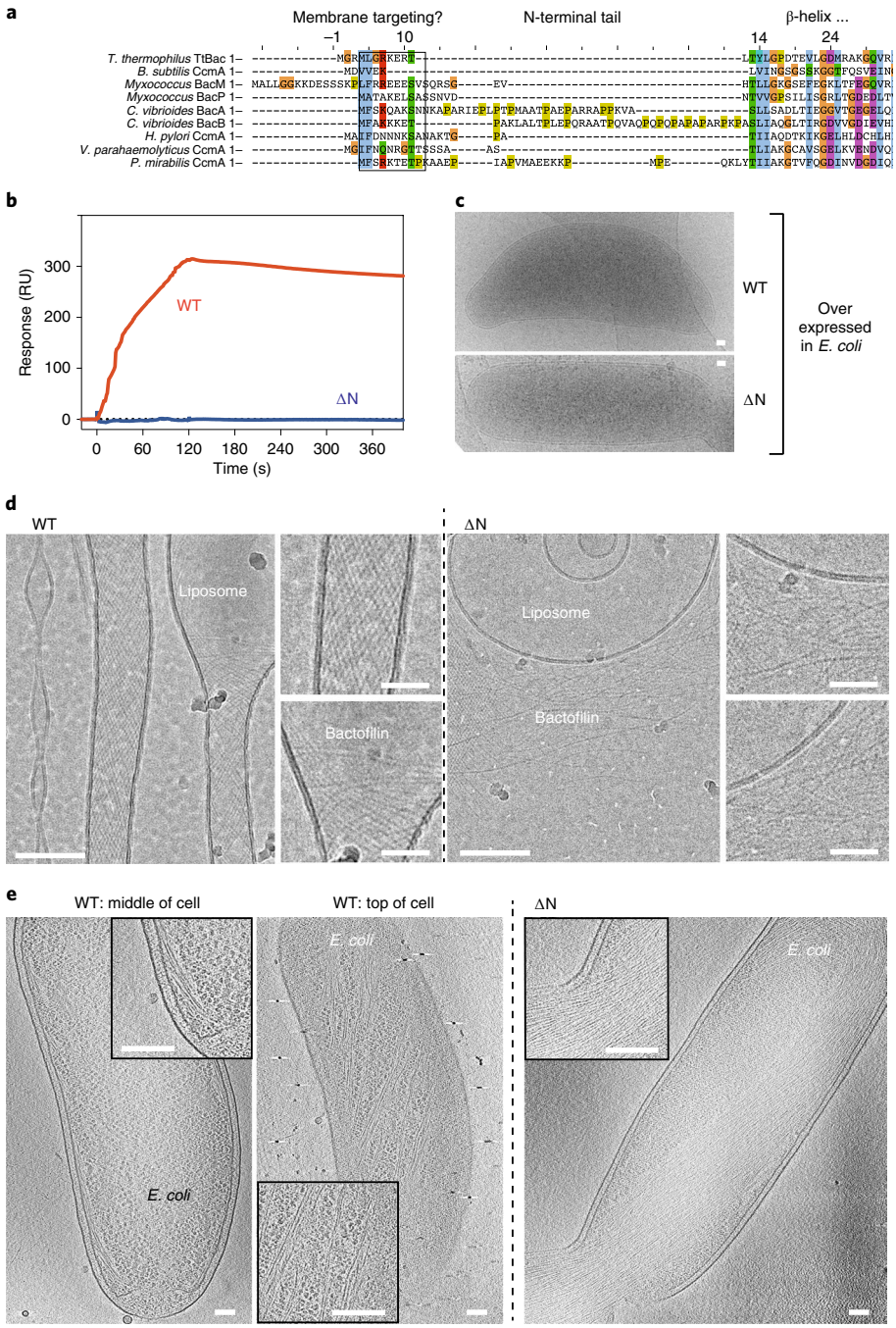

Fig. 6 | TtBac binds to lipid membranes through its N-terminal tail. a, A multiple sequence alignment showing the N-terminal tails of bactofilins, including TtBac. A doublet of hydrophobic residues, often followed by a positively charged residue, is conserved across species. The black box highlights residues for comparison, residues are colour-coded according to their chemical properties and the elipses depict the lack of amino acids for alignment. The fact that the N-terminal tails interact to form double-helical filaments in our study (Fig. 4e) and the occurrence of unusual hydrophobicity in a disordered tail region prompted us to investigate the membrane binding of TtBac in vitro and in situ. **b**, SPR traces of TtBac-WT and ΔN-TtBac proteins when added to liposome-coated L1 sensor chips (GE Healthcare Life Sciences). The unmodified protein clearly shows binding, whereas the mutant protein lacking the N-terminal tails shows no signal. The response is shown in response units (RU). **c**, When overexpressed in *E. coli*, TtBac leads to a morphological phenotype: it causes the cells to become crescent-shaped, and this effect disappears when overexpressing ΔN-TtBac under the same conditions. Scale bars, 100 nm. **d**, When adding TtBac protein to preformed liposomes made from *E. coli* lipids, the protein was shown by subsequent cryo-EM imaging to deform and tubulate the liposomes. Single protofilaments appear to deform the liposomes from the outside (insets). When ΔN-TtBac was added to the same liposomes, no binding was observed and the liposomes were not deformed, consistent with the idea that the N-terminal tails of TtBac, and presumably most bactofilins, are the membrane-targeting domains of these proteins. Scale bars, 100 nm for the two large images; 25 nm for the four smaller ones. **e**, An investigation of the overexpression system depicted in Fig. 6c by cryo-ET. TtBac-WT (left two images, see also Supplementary Video 3) clearly formed long bundles going around the cell close to the membrane, indicative of direct binding to the inner membrane of *E. coli* as no putative accessory proteins were expected to be present. When performing the same experiment with ΔN-TtBac (right image, see also Supplementary Video 4), a very large bundle of filaments was observed going through the centre of the cell, not close to the membrane, and impeding cell division. These results again indicated that it is the N-terminal tails of bactofilins that facilitate direct membrane binding. Scale bars, 100 nm.

96 residues, sequence identity 35%) suggested that the non-polar architecture seen for TtBac filaments may also be widely conserved. To investigate this, we aligned 12,646 putative bactofilin domain sequences and derived an evolutionary coupling score for each pair of residues in the alignment[23]. Visualization of these scores as a heat map produced an overall view of coevolution within the bactofilin sequence and was compared to calculated Cα distances in the TtBac monomer structure (Fig. 5a,b). As shown by the second diagonal in Fig. 5b, there is a high level of coevolution, and high coupling scores, between amino acids ~17 residues apart, that is one turn of the β-helix, and a weaker line ~34 residues apart (two turns) is also visible. In other words, the β-helical structure of bactofilins, which enforces interactions between the residues exactly above and below, is conserved among a very large proportion of the aligned sequences.

As the strength of evolutionary coupling between amino acids is largely dictated by their physical distance, it is reassuring that the heat map of Cα distances for residues in the TtBac monomer in Fig. 5a almost perfectly recapitulates the coupling scores observed in Fig. 5b. However, the distances between TtBac residues in the monomer do not explain the coevolution between homoterminal amino acids observed in the lower left and upper right corners of the coupling heat map (Fig. 5b and magnified in Fig. 5c, middle). These scores can be explained, though, by taking into account the architecture of the TtBac filament as determined here (Fig. 5c, right). In the TtBac filament it is precisely those residues with surprisingly high coupling scores that are brought closer together within the head-to-head and tail-to-tail polymer architecture. There is no sign of coevolution corresponding to interaction between the C and N termini (top left/bottom right corners) as would be expected for a polar head-to-tail arrangement.

As the strong coupling scores exclusively indicate interactions between residues within each terminus (N−N and C−C, not N−C) it is clear that the non-polar arrangement of the TtBac filament must be highly conserved within bactofilins. The generality of the TtBac filament architecture extends to fine details such as minimal protofilament twist, as the position of the orthogonal diagonal in the coupling heat map describing the homoterminal interaction exactly matches the location of the two-fold symmetry seen in the cryo-EM structure (Fig. 5d,e).

TtBac binds to membranes in vitro and in situ, an activity facilitated by its N-terminal tail. It has previously been shown that bactofilins are located close to membranes in cells in vivo[3]. Other filaments of prokaryotic cytoskeletons have been shown to directly interact with membranes, either through amphipathic helices or other small membrane-targeting signals[24–26]. As the N-terminal tail of TtBac was shown here to mediate interactions between the two protofilaments (Fig. 4e), we hypothesized that it might contain hydrophobicity to interact directly with membranes. Indeed, when aligning a subset of well-characterized homologous bactofilin sequences, a conserved hydrophobic motif emerged at the very N terminus of these sequences (Fig. 6a).

First we needed to show that TtBac does indeed bind to membranes directly. To investigate this hypothesis we used surface plasmon resonance (SPR) with liposomes (Fig. 6b). TtBac-WT showed a strong signal, indicating liposome binding. Equally, when overexpressing TtBac in *E. coli* cells, we noted a strong phenotype of bent cells, which is in our experience a hallmark of filament formation directly on the inner membrane on one side of the cell's body (Fig. 6c).

To demonstrate that the N terminus of TtBac is necessary for membrane binding we produced an N-terminal deletion mutant ΔN-TtBac containing residues 11–123 (Supplementary Table 3). Both overexpression of the mutant in *E. coli* and incubation of the protein with liposomes and SPR showed that membrane binding and morphological changes were completely abolished (Fig. 6b,c).

The binding of TtBac to membranes appears to include a demand for curvature, as seen by the morphological phenotype caused by overexpression of the full-length filamentous protein in *E. coli* (Fig. 6c). Similarly, its interaction with liposomes in vitro is characterized by strong deformations or even tubulation (Fig. 6d).

Finally, we also visualized TtBac-WT filaments in *E. coli* cells by electron cryotomography (cryo-ET), where they were visible as bundles circling the cell at an angle on or close to the inner membrane. The helicity confirms TtBac's preference for curvature. When performing the same experiment with ΔN-TtBac, all membrane proximity was abolished and the filaments formed one large bundle in the cytoplasm (that was so large that it also inhibited cell division) (Fig. 6e, left and right, Supplementary Videos 3 and 4).

Discussion

We found bactofilins in eukaryotes although, as far as we could tell, this was restricted to Stramenopiles and Ascomycete fungi. We believe that this may be the result of horizontal gene transfer from prokaryotes into these organisms[27,28]. It will be interesting to determine the function of bactofilins in eukaryotes and in archaea where they appear to be more widespread.

Our structural work showed TtBac to be closely related to BacA from *C. crescentus*[6]; it polymerized into completely β-helical protofilaments that have N−N and C−C interfaces. This arrangement produces non-polar protofilaments and because the two-fold axes at the N−N and C−C interface are nearly aligned, the filament twists only very slightly, ~5° per 6-nm-long dimer. The unanticipated lack of polarity[8,9] makes the two ends of the filament equal and hence excludes all cytomotive mechanisms[1]. Bactofilins should therefore be classified as cytoskeletal, along with MreB, DivIVA, SepF and MinCD, all of which form non-polar filaments.

It is striking that the above list of cytoskeletal filaments contains exclusively cooperative filaments[29] that bind directly to cell membranes from the inside. We showed here that bactofilins are no exception as they contain a conserved membrane-targeting sequence within their N-terminal tail.

From Fig. 6d it is clear that the TtBac filaments form mostly single protofilaments when bound to membranes, indicating and confirming that the N-terminal tails function to bind the membrane and not each other. We propose that the bona fide bactofilin filament is an almost straight, non-helical single protofilament, and that it is constitutively membrane bound (Supplementary Fig. 6).

Both tomography and microscopy (Fig. 6c–e) demonstrated that TtBac preferred curvature, another feature that it shares with MreB[30], DivIVA[31] and SepF[32]. Curvature preference seems to stem from the fact that they are not completely straight and this could also be a general mechanism of filament length restriction on flat, non-deformable membranes and/or curvature induction/sensing.

We firmly believe that bactofilins deserve greater attention as components of prokaryotic cytoskeletons. Armed with the analysis presented here, which demonstrates the widespread occurrence, biochemical stability, direct membrane binding and non-polar filament architecture of these extraordinary proteins, we hope to promote studies into the roles of bactofilins in cell biology. We anticipate further roles in morphogenesis and in processes where very stable membrane attachment is needed that cannot overlap with any of the other known systems. It also remains unclear how polymerization, and with it membrane attachment, is regulated in cells and what role, if any, the C-terminal tails have.

Methods

Identification of bactofilins outside bacteria. We became interested in the possibility that bactofilins are present outside the kingdom of Bacteria after noting that Pfam (http://pfam.xfam.org) annotates several members of the Pfam PF04519 (DUF 583, bactofilin) family in eukaryotes and archaea. This was investigated further by running a HMMSEARCH (HMMER3)[33] with Pfam PF04519 against the eukaryotic UniProtKB. The results included two taxonomic clusters of hits,

one consisting of proteins found among the Stramenopiles and the other of proteins within Ascomycete fungi. There were also hits in other eukaryotes, but these were either non-bactofilin repetitive sequences or isolated hits within well-sequenced clades (in all these cases the best BLASTP matches were bacterial sequences). Clusters of hits within the Stramenopiles and Ascomycetes were aligned, HMMs were built and additional examples within UniProtKB were searched for. In both cases, further examples were recovered from the relevant clades, but not outside them. To better understand the relationship between these putative eukaryotic bactofilins and the bacterial sequences, an alignment (using HMMALIGN) was made of putative bactofilin domains. It comprised bactofilins that are representative of sequence diversity within bacteria and archaea (found using the proGenomes set of representative prokaryotic genomes[34] and a HMMSEARCH with Pfam PF04519), all of the putative eukaryotic bactofilins, including the taxonomic singletons, and the additional sequences from the putative clusters. From the alignment a phylogeny was constructed using the FastTree v.2.1 (ref. [35])algorithm, after selecting informative columns using GBlocks v.0.91b[36] (Supplementary Dataset 1). The phylogeny did not imply the existence of a common ancestor of eukaryotic bactofilins within eukaryotes. After inspecting the alignments and trees, three putative eukaryotic bactofilins were chosen for cloning and expression: one fungal protein (Q7RX79 from *Neurospora crassa*) and two Stramenopile proteins (D0N980 from *P. infestans* and D7G0L1 from *Ectocarpus siliculosis*). Only D0N980 (PITG_07992) was successfully expressed and confirmed to be a bactofilin. It remains to be seen how many of the putative bactofilins outside the Oomycetes are polymerizing bactofilins.

Protein cloning, expression and purification. The amino acid sequences of all proteins used in this study are listed in Supplementary Table 3.

PiBac Oomycetes bactofilin. To obtain a plasmid that codes for full-length bactofilin protein from *P. infestans* (PiBac), the gene PITG_07992 (UniProtKB, D0N980) was synthesized (GenScript), amplified and cloned into the pHis17 plasmid using Gibson assembly (New England Biolabs), with a stop codon before the C-terminal hexahistidine tag on the plasmid. C41(DE3) *E. coli* cells (Lucigen) were transformed with the resulting plasmid by electroporation. Next, 40 ml 2xTY media supplemented with 100 μg ml^{-1} ampicillin was inoculated with a single colony from the plate, which was grown at 200 r.p.m., 37 °C overnight. The preculture was then used to inoculate 4 l 2xTY media with 100 μg ml^{-1} ampicillin. After reaching an optical density $(OD)_{600}$ of 0.6–1.0 at 200 r.p.m., 37 °C, the expression was induced with 1 mM isopropyl β-D-thiogalactoside for 4 h at the same temperature and the cells were harvested by centrifugation. For purification, the entire pellet was resuspended in 150 ml Buffer D (50 mM Tris, 200 mM NaCl, 1 mM TCEP, pH 8) supplemented with DNase I, RNase A (Sigma) and EDTA-free protease inhibitor tablets (Roche). Cells were lysed by sonication and the lysate was cleared by centrifugation at 10,000 r.p.m. in a 45 Ti rotor (Beckman) for 30 min at 4 °C, followed by another centrifugation at 20,000 r.p.m. in a 45 Ti rotor for 30 min at 4 °C. The supernatant was supplemented with 2% weight/volume PEG 8000 before further centrifugation at 40,000 r.p.m. in the 45 Ti rotor for 30 min at 4 °C. Each pellet was resuspended in 5 ml Buffer D supplemented with 0.3 g ml^{-1} caesium chloride and 1% volume/volume Triton X-100, and the pellet was solubilized overnight at 4 °C. The solubilized mixture was centrifuged in a TLA 100.3 rotor (Beckman) at 50,000 r.p.m. for 15 min at 4 °C, the pellet was discarded and the supernatant was centrifuged in a TLA 100.3 rotor at 80,000 r.p.m. for 5 h at 4 °C. The white layer near the bottom of the centrifuge tube was extracted and resuspended in 3 ml Buffer D supplemented with 0.3 g ml^{-1} caesium chloride before further centrifugation in a TLA 100.3 rotor at 80,000 r.p.m. overnight at 4 °C. The white layer containing filaments near the bottom of the centrifuge tube was extracted again and examined using negative stain electron microscopy (see below). Identity of the protein was confirmed by SDS–polyacrylamide gel electrophoresis (SDS–PAGE).

H_6-TtBac 1–123. To obtain an *E. coli* expression plasmid for N-terminally His_6-tagged full-length TtBac, MGSSHHHHHH-1–123, the gene coding for protein WP_011173792.1 (NCBI) was amplified from genomic DNA and cloned into vector pET15b using Gibson assembly (New England Biolabs). C41(DE3) *E. coli* cells (Lucigen) were transformed by electroporation. Next, 60 ml 2xTY media supplemented with 100 μg ml^{-1} ampicillin were inoculated with a single colony from the plate, which was grown at 200 r.p.m., 37 °C overnight. The culture was then used to inoculate 6 l 2xTY media with 100 μg ml^{-1} ampicillin. After reaching an OD_{600} of 0.6–1.0 at 200 r.p.m., 37 °C, protein expression was induced with 1 mM isopropyl β-D-thiogalactoside for 4 h at the same temperature and the cells were harvested by centrifugation. For purification, the entire pellet was resuspended in 200 ml Buffer A (50 mM Tris, 6 M guanidinium chloride, 1 mM TCEP, pH 7). Cells were disrupted at 25 kPSI in a cell disruptor (Constant Systems) and the lysate was cleared by centrifugation at 40,000 r.p.m. in a 45 Ti rotor (Beckman) for 30 min at 25 °C. The cleared lysate was loaded onto a 5 ml HisTrap HP column (GE Healthcare) and washed with stepwise increases of imidazole in Buffer A: 0, 20, 50, 200, 500 and 1,000 mM. Eluted fractions were analysed by SDS–PAGE and those containing TtBac (mostly at 50 mM imidazole) were pooled and concentrated in Centriprep concentrators (10 kDa molecular weight cut-off (MWCO), Millipore) to 10 mg ml^{-1}. The H_6-TtBac protein was refolded by a single-step dialysis process into Buffer B (50 mM CAPS, 200 mM NaCl, 1 mM TCEP, pH 11) overnight at room temperature. The protein was stored at 4 °C.

TtBac-WT 1–123. To obtain an *E. coli* expression plasmid for untagged, full-length TtBac 1–123, the gene coding for protein WP_011173792.1 (NCBI) was amplified from the above H_6-TtBac pET15b plasmid and cloned into plasmid pHis17 using Gibson assembly (New England Biolabs) with a stop codon before the C-terminal tag on the plasmid. The plasmid encoding untagged TtBac was used to transform C41(DE3) *E. coli* cells (Lucigen) by electroporation. Next, 40 ml 2xTY media supplemented with 100 μg ml^{-1} ampicillin was inoculated with a single colony from the plate, which was grown at 200 r.p.m., 37 °C overnight. The culture was then used to inoculate 4 l 2xTY media with 100 μg ml^{-1} ampicillin. After reaching an OD_{600} of 0.6–1.0 at 200 r.p.m., 37 °C, the expression was induced with 1 mM isopropyl β-D-thiogalactoside for 4 h at the same temperature and the cells were harvested by centrifugation.

To isolate full-length, unmodified TtBac (TtBac-WT) we developed a protocol based on a series of gradient centrifugation steps adapted from previous work on *M. xanthus* BacM[5]. For purification, the entire pellet was resuspended in 150 ml Buffer B (50 mM CAPS, 200 mM NaCl, 1 mM TCEP, pH 11) supplemented with DNase I, RNase A (Sigma) and EDTA-free protease inhibitor tablets (Roche). Cells were lysed by sonication and the lysate was cleared by centrifugation at 10,000 r.p.m. in a 45 Ti rotor (Beckman) for 30 min at 4 °C, followed by another centrifugation at 20,000 r.p.m. in the 45 Ti rotor for 30 min at 4 °C. The supernatant was supplemented with 2% volume/weight PEG 8000 before centrifugation at 40,000 r.p.m. in the 45 Ti rotor for 30 min at 4 °C. Each pellet was resuspended in 5 ml Buffer B supplemented with 0.3 g ml^{-1} caesium chloride and 1% volume/volume Triton X-100, and the pellet was solubilized overnight at 4 °C. The mixture was centrifuged in a TLA 100.3 rotor (Beckman) at 50,000 r.p.m. for 15 min at 4 °C, the pellet was discarded and the supernatant was centrifuged in the TLA 100.3 rotor at 80,000 r.p.m. for 5 h at 4 °C. The white layer containing TtBac filaments near the bottom of the tube was extracted, resuspended in 3 ml Buffer B supplemented with 0.3 g ml^{-1} caesium chloride before centrifugation in the TLA 100.3 rotor at 80,000 r.p.m. overnight at 4 °C. The layer containing TtBac filaments near the bottom of the centrifuge tube was extracted and diluted with an equal volume of Buffer C (50 mM CAPS, 100 mM NaCl, 1 mM TCEP, 1 mM EDTA, pH 11) before being loaded onto a Superose 6 10/300 size exclusion column (GE Healthcare) equilibrated in Buffer C. The fractions containing TtBac filaments (in the void volume of the column) were pooled and centrifuged in a TLA 100.3 rotor (Beckman) at 80,000 r.p.m. for 30 min at 4 °C. The TtBac filaments formed a transparent pellet at the bottom of the tube and were resuspended in Buffer C to 1–2 mg ml^{-1} and stored at 4 °C.

NB4-mut2. Purified and refolded H_6-TtBac filaments were sent to the Vlaams Instituut voor Biotechnologie's Nanobody Core (VIB) for commercial camelid nanobody generation and selection. Frozen bacterial cultures containing individual nanobody genes in the pMECS vector were delivered by VIB. Genes encoding each nanobody were amplified using colony PCR and cloned into the pHEN6c expression vector (VIB) using Gibson assembly (New England Biolabs). The resulting constructs using pHEN6c encoded the gene for PelB(leader)-nanobody-SSHHHHHH proteins.

For expression, the pHEN6c plasmids encoding the nanobody proteins were used to transform WK6 *E. coli* cells (VIB, essential for this to work) by electroporation. Next, 20 ml 2xTY media supplemented with 100 μg ml^{-1} ampicillin was inoculated with a single colony from the plate, which was grown at 200 r.p.m., 37 °C overnight. The culture was then used to inoculate 2 l 2xTY media with 100 μg ml^{-1} ampicillin. After reaching an OD_{600} of 0.6–1.0 at 200 r.p.m., 37 °C, the expression was induced with 1 mM isopropyl β-D-thiogalactoside for 4 h at the same temperature and the cells were harvested by centrifugation.

The purification protocol followed recommendations from VIB. Since the plasmids encoded the gene for the PelB leader sequence, nanobody proteins were secreted into the periplasm. For purification, the entire pellet from 2 l culture was resuspended in 24 ml TES buffer (200 mM Tris, 500 mM sucrose, 0.5 mM EDTA, pH 8), with shaking at 4 °C for 1 h. The mixture was supplemented with 36 ml TES/4 buffer (TES buffer diluted 4 times in water) and was mixed at 4 °C for 1 h. The suspension was centrifuged in a 45 Ti rotor (Beckman) for 30 min at 40,000 r.p.m. at 4 °C. The supernatant containing the periplasmic extract was sonicated before being loaded onto a 5 ml HisTrap HP column (GE Healthcare). The column was washed with stepwise increases of imidazole in buffer E (50 mM Tris, 200 mM NaCl, 1 mM TCEP, pH 8): 3, 200, 500 and 1,000 mM. Elutions were analysed by SDS–PAGE and the fractions containing nanobodies (mostly at 200 mM imidazole) were pooled and concentrated in Centriprep concentrators (10 kDa MWCO, Millipore) to around 6 mg ml^{-1}. The concentrated proteins were buffer exchanged using a PD-10 desalting column (GE Healthcare) equilibrated with Buffer C (50 mM CAPS, 100 mM NaCl, 1 mM TCEP, 1 mM EDTA, pH 11) using the spin protocol.

To solve the bundling issue when nanobody proteins were added to bactofilin filaments, mutants were made to try to change selected surface amino acid residues into aspartic acids using NB4 as a template, which had been selected by negative stain electron microscopy of TtBac and nanobody equimolar mixtures. The gene

encoding NB4-mut2 (L13S, Q15D, K45D, K66D) was synthesized (GenScript), amplified and cloned into the pHEN6c expression vector. Nanobody mutant proteins were expressed and purified using the same protocol as the non-mutated nanobody proteins. The final concentration of NB4-mut2 was 6.8 mg ml^{-1}.

ΔN-TtBac(F105R)-H$_6$. The required coding region was amplified from the H$_6$-TtBac construct by PCR and was cloned into the plasmid pHis17 using Gibson assembly (New England Biolabs), resulting in a C-terminal tag: GSHHHHHH. The point mutation was introduced by Q5 mutagenesis (NEB). The resulting plasmid was used to transform C41(DE3) *E. coli* cells (Lucigen) by electroporation. Next, 12 l cultures in 2xTY were inoculated and induced with 1 mM IPTG at an OD$_{600}$ of 0.6 and further grown for 6 h at 37 °C. After harvesting of the cells, the pellet was dissolved in Buffer F (50 mM Tris/HCl, 200 mM NaCl, pH 7.5) and lysed by cell disruption at 35 kPSI (Constant Systems). The lysate was cleared by ultracentrifugation (2 h at 35,000 r.p.m. in a 45 Ti rotor) and loaded onto a 5 ml HisTrap column. The bound fraction was eluted with 0.6 M imidazole, pH 7.0 and further purified by size exclusion on a Sephacryl S300 16/60 column (GE Healthcare) in Buffer G (20 mM CHES/NaOH, 250 mM NaCl, pH 9.5). Fractions were analysed by SDS–PAGE and the fractions containing ΔN-TtBac(F105R)-H$_6$ were pooled and concentrated in Centriprep concentrators (10 kDa MWCO, Millipore) to around 12 mg ml^{-1}.

ΔN-TtBac 11–123. The required coding region was amplified from the H$_6$-TtBac construct by PCR and was cloned into the plasmid pHis17 using Gibson assembly (New England Biolabs). The plasmid was used to transform C41(DE3) *E. coli* cells (Lucigen) by electroporation. Next, 12 l cultures in 2xTY were induced with 1 mM IPTG at an OD$_{600}$ of 0.6 and further grown for 4 h at 37 °C. After harvesting of the cells, the cells were dissolved in the same buffer as for TtBac-WT 1–123 and spun for 30 min at 9,000 r.p.m. in a Ti 45 rotor (Beckman). The pellet was dissolved in the same buffer and from here onwards the same protocol as for TtBac-WT 1–123 was used.

Electron microscopy. *Negative stain.* Continuous carbon grids were purchased from Electron Microscopy Sciences. After glow discharging, 3 µl of sample was applied, blotted and stained with fresh 2% uranyl acetate solution. Uranyl acetate was applied 1–3 times with wait times of up to 60 s. After air drying, grids were imaged in a Thermo Fisher F20 electron microscope equipped with a Falcon 2 detector, operated at room temperature.

Cryo-EM, helical reconstruction of TtBac-WT and with NB4-mut2. Initially, refolded and His$_6$-tagged TtBac protein (H$_6$-TtBac) was investigated using cryo-EM with the aim to obtain an atomic model of the filament. For this, 2 µl 60 µM H$_6$-TtBac in Buffer C was added onto freshly glow-discharged Quantifoil Cu/Rh R2/2 holey carbon 200 mesh grids (Quantifoil). The grids were blotted for 2.5 s with a blotting force of −15 and a drain time of 0.5 s, and were flash frozen in liquid-nitrogen-cooled liquid ethane using a Thermo Fisher Vitrobot Mark IV. The Vitrobot chamber was set to 10 °C and 100% humidity.

Grids were imaged at 300 kV using a Thermo Fisher Tecnai Polara G2 or Titan Krios microscope using a Falcon III detector (one dataset was collected at eBIC), using pixel sizes between 1.0 and 1.4 Å and average total doses of about 40 e$^-$ per Å^2, distributed over 40–70 frames. Images were motion corrected using MOTIONCOR2 (ref. [37]) and contrast transfer function (CTF) corrected using GCTF v.1.06 (ref. [38]). 2D classification analysis in RELION v.3.0 (ref. [39]) revealed that the filaments had varying widths, probably caused by varying protofilament numbers that ranged from 2–4. We then switched to untagged and natively purified TtBac-WT material, using the same vitrification and imaging conditions as above. The data showed very similar variations in protofilament number but it was possible to discern the 4.7 Å axial repeat caused by the β-stacking of the subunits in 2D classes obtained by processing in RELION v.3.0. However, it proved impossible to go beyond a 5 Å resolution when performing reconstructions in RELION v.3.0, probably because the filaments are smooth and the beginning and end of each subunit could not be determined during the RELION v.3.0 3D refinement procedure. To overcome this issue, a nanobody was obtained, NB4, that showed very clear binding to the filaments as it changed their appearance and also led to a strong reduction in filaments with more than two protofilaments. Unfortunately, the filaments coated with NB4 tended to bundle heavily, impeding further analysis. Hence four mutations were introduced, L13S, Q15D, K45D and K66S, yielding NB4-mut2 nanobody that substantially reduced bundling and enabled image analysis and helical reconstruction of TtBac-WT to near-atomic resolution. For the cryo-EM grid preparation with the nanobody, purified nanobody protein was added to TtBac filaments at a molar ratio of 1.2:1 (nanobody:TtBac). The final dataset used is summarized in Supplementary Table 1. It was collected on a Thermo Fisher Titan Krios microscope at 300 kV, using a Falcon III detector in electron counting mode with an estimated pixel size of 1.07 Å per pixel. Next, 2,130 good images were selected after MOTIONCOR2 and GCTF and from those around 456,000 helical segments were picked in RELION v.3.0, 57 Å apart along the filament axis. 2D classification in RELION v.3.0 selected 346,000 good helical segments and these were used for helical reconstruction in RELION with a helical parameters twist = 4.73° and rise = 57.48 Å (ref. [40]). Particle polishing and post-processing followed standard RELION v.3.0 procedures and led to a final map with a resolution determined from gold-standard two halves Fourier shell correlation (FSC) calculations (FSC = 0.143) of 3.6 Å (ref. [41]). When inspecting the map, this value hides the fact that the map has very anisotropic resolution, probably caused by the dominant 4.7 Å β-stacking repeat along the filament axis, and poor resolution perpendicular to the filament axis. The previously determined *B. subtilis* BacA solid state NMR structure (PDB 2N3D)[6] was homology modelled into TtBac using the SWISSMODEL server[42] and placed in the cryo-EM map, as was a SWISSMODEL of the NB4-mut2 nanobody. Placing these models produced excellent fit but the anisotropy of the map made atomic refinement difficult.

After the crystal structure of ΔN-TtBac(F105R)-H$_6$ was solved during the course of this study, a much better atomic model became available. After performing signal subtraction of the nanobody density from all 346,000 helical segment images in RELION v.3.0 (ref. [22]), an improved cryo-EM map of the two TtBac protofilaments was obtained using RELION v.3.0 with a helical parameters twist = 4.89° and rise = 57.46 Å. Several cycles of manual model adjustment in MAIN v.2019 and refinement with PHENIX.real_space_refine v.1.16 (with additional secondary structure restraints)[43] produced the final model (Supplementary Table 1). Resolution was estimated to be 3.4 Å from half-map FSC analysis in RELION but FSC analysis against the atomic model yielded a lower value of 4.2 Å, caused by the strong resolution anisotropy of the map. The atomic coordinates were deposited in PDB with accession number 6RIB and the subtracted cryo-EM map was deposited in EMDB with accession code EMD-4887.

Evolutionary sequence coupling analysis of filament formation. A large set of bactofilin sequences was obtained by searching the UniProtKB with the Pfam HMM PF04519 using HMMSEARCH with defaults. These 20,746 sequences were aligned to PF04519 using HMMALIGN. Inserts relative to the HMM were removed and sequences with more than 5% gaps relative to the HMM were discarded using the dcaTools package (https://gitlab.com/ducciomalinverni/dcaTools/), leaving 12,646 sequences. This alignment was used to infer a direct coupling analysis (DCA) model of the bactofilin family using lbsDCA (https://gitlab.com/ducciomalinverni/lbsDCA/), an implementation of the asymmetric pseudo-likelihood method for DCA[23]. lbsDCA reported an effective number of sequences after reweighting with a 90% identity cut-off of 5,770.4. The average-product corrected Frobenius norm DCA scores reported were scaled arbitrarily for plotting after smoothing of the heatmap with a 3 × 3 Gaussian kernel.

Crystal structure determination of ΔN-TtBac(F105R)-H$_6$. ΔN-TtBac(F105R)-H$_6$ protein was produced as described above and SeMet was obtained using published protocols[44] with the same subsequent purification protocol as for the native protein, with the sole change that all buffers contained reducing agent TCEP at 1 mM. Initial crystallization conditions were obtained using the LMB in-house high-throughput crystallization facility[45]. All crystals were produced in MRC two-drop crystallization plates using 100 + 100 nl sitting drop set-ups; crystallization experiments were performed at 19 °C. SeMet ΔN-TtBac(F105R)-H$_6$ was crystallized using reservoir solution containing 6.7–7.3% volume/volume 2-propanol, 0.17–0.19 M lithium sulphate, 0.1 M phosphate citrate, pH 3.7 and 0.4 M ammonium acetate as an additive. The crystals were cryocooled using 30% glycerol in reservoir solution. The crystals belonged to space group $I2_12_12_1$ and showed a very large unit cell (Supplementary Table 1). Three 360° datasets were collected at beamline I03 (Diamond Light Source) and merged together using CCP4 programs[46] resulting in very high multiplicity and extending to about 4.0 Å resolution. 20–30 selenium sites were readily identified using SHELXD v.7 (ref. [47]) and subsequent phasing with PHASER v.7 (ref. [48]) and non-crystallographic symmetry averaging (operators deduced from SeMet sites within PHENIX v.1.16) using density modification resulted in interpretable electron density maps. It was determined that the number of recognisable subunits was 32 and both phasing and non-crystallographic symmetry averaging were adjusted to take this into account to obtain a final electron density map. The preliminary atomic model obtained from cryo-EM maps at 4–5 Å resolution was fitted manually, guided by the position of the single SeMet25 residue and its anomalous signal, and it was recognised that residues A101–G112 had become disordered, presumably due to the F105R mutation used in the construct. Analysis then switched to the native protein. Native ΔN-TtBac(F105R)-H$_6$ was crystallized using reservoir solutions as above and the crystals were cryocooled as above. The crystals again belonged to space group $I2_12_12_1$ with only slightly different cell constants. The crystals were isomorphous enough to enable simple rigid body refinement of all 32 chains since molecular replacement was not possible, presumably due to the smooth appearance of the protofilaments that made it difficult to discern the beginning and end of each subunit; this is very similar to the initial problems with cryo-EM image reconstruction. The structure was rebuilt manually in MAIN v.2019 (ref. [49]) and refined with Phenix.refine v.1.16 (ref. [50]) (using non-crystallographic symmetry restraints and real-space refinement) for several cycles. Final statistics are summarized in Supplementary Table 1 and the resulting coordinates haven been deposited in PDB with accession code 6RIA.

SPR. SPR data were collected using a BIAcore T200 instrument using a L1 Sensor Chip (GE Healthcare). Both reference and ligand channels were equilibrated in 50 mM CAPS, pH 8.0, 100 mM NaCl, 4 mM TCEP at 25 °C. Liposomes

were prepared with *E. coli* total lipid extract (Avanti Polar Lipids) using three freeze–thaw cycles. Each cycle consisted of two freeze–thaw events with liquid nitrogen followed by a single 10 min water bath sonication. Liposomes at a lipid concentration of 3 mg ml^{-1} were captured onto the ligand surface at 10 µl min^{-1} to a level of ~1,200 RU. To prevent non-specific binding, the surfaces were passivated by injections of 0.2 mg ml^{-1} BSA with 1 mg ml^{-1} NSB (GE Healthcare Life Sciences) for 120 s before a 120 s injection at 30 µl min^{-1} of TtBac-WT, ΔN-TtBac at 10 µM or buffer followed by dissociation for 300 s. After each measurement the surfaces were regenerated with a 30 s injection of 20 mM CHAPS. Data were doubly referenced by subtraction of the reference channel data and from injections of buffer alone.

Bactofilin binding to liposomes. Liposomes were prepared with *E. coli* total lipid extract (Avanti Polar Lipids) using freeze–thaw cycles followed by sonication in buffer D (50 mM CAPS, 100 mM NaCl, 4 mM TCEP, pH 8.0). Images of liposomes together with native TtBac-WT and ΔN-TtBac were obtained by mixing 120 µM protein in buffer C with 2 mg ml^{-1} liposomes in buffer D. Of this mixture, 3 µl was applied onto freshly glow-discharged Quantifoil Cu/Rh R2/2 holey carbon 200 mesh grids (Quantifoil). The grids were blotted for 4 s with a blotting force of −15 and a drain time of 0.5 s, and were flash frozen in liquid-nitrogen-cooled liquid ethane using a Vitrobot Mark IV (Thermo Fisher). The Vitrobot chamber was set to 10 °C and 100% humidity. Grids were imaged in a Thermo Fisher F20 electron microscope equipped with a Falcon 2 detector, operated at cryogenic temperature.

Bactofilin overexpression in *E. coli* and electron tomography. Cells expressing TtBac-WT or ΔN-TtBac were mixed with 10 nm protein A gold fiducials and plunge frozen on Quantifoil R2/2 holey carbon grids using a Vitrobot Mark IV (Thermo Fisher). Tomography data were collected on a Titan Krios electron microscope equipped with a Quantum imaging filter and K2 direct detector (both Gatan). Tilt series were acquired using SerialEM[51] from 0° to ±60° using a grouped dose symmetric tilt scheme[52] with a 2° increment and a total dose of 160 e$^-$ per Å^2. The pixel size was 5.44 Å, the target defocus 8 µm and the slit width of the energy filter 20 eV. Tomograms were reconstructed in IMOD v.4.9 (ref. [53]), using the SIRT algorithm.

Statistics and reproducibility. The filaments shown in the micrograph displayed in Fig. 1d were observed on all 23 images from two different grids. Similarly, the TtBac filaments shown in Fig. 2a–c were observed in almost all of the images in large datasets of more than 1,000 images in total, from several grids. The nanobody-bound filaments were also seen in five sessions, with 2,130 images collected for final 3D reconstruction. The SPR experiment shown in Fig. 6b was run independently twice with similar results. The cells displayed in Fig. 6c,e stem from a dataset with eight tomographic reconstructions each for the ΔN-TtBac and TtBac-WT filaments and all show the same localization of the filaments, either off or on the cell membrane, respectively. The binding and lack thereof of liposomes by TtBac and ΔN-TtBac filaments in Fig. 6d was observed in more than 24 images and across three grids each. Regarding the data shown in Supplementary Fig. 1b, as per the original publication[21], expression and differential expression calls were made with edgeR using trimmed mean of M-values normalization, a generalized linear model and false discovery rate calculations based on the Benjamini–Hochberg method. The generalized linear model is two-sided.

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

Data availability

The data that support the findings of this study are available from the corresponding author, who will also provide all expression plasmids generated for this study on request. Atomic coordinates have been deposited in PDB with accession codes 6RIA and 6RIB. The cryo-EM volume has been deposited in EMDB with accession code EMD-4887.

Received: 13 April 2019; Accepted: 19 July 2019;
Published online: 09 September 2019

References

1. Wagstaff, J. & Löwe, J. Prokaryotic cytoskeletons: protein filaments organizing small cells. *Nat. Rev. Microbiol.* **16**, 187–201 (2018).
2. Lin, L. & Thanbichler, M. Nucleotide-independent cytoskeletal scaffolds in bacteria. *Cytoskeleton* **70**, 409–423 (2013).
3. Kühn, J. et al. Bactofilins, a ubiquitous class of cytoskeletal proteins mediating polar localization of a cell wall synthase in *Caulobacter crescentus*. *EMBO J.* **29**, 327–339 (2010).
4. Hay, N. A., Tipper, D. J., Gygi, D. & Hughes, C. A novel membrane protein influencing cell shape and multicellular swarming of *Proteus mirabilis*. *J. Bacteriol.* **181**, 2008–2016 (1999).
5. Koch, M. K., McHugh, C. A. & Hoiczyk, E. BacM, an N-terminally processed bactofilin of *Myxococcus xanthus*, is crucial for proper cell shape. *Mol. Microbiol.* **80**, 1031–1051 (2011).
6. Shi, C. et al. Atomic-resolution structure of cytoskeletal bactofilin by solid-state NMR. *Sci. Adv.* **1**, e1501087 (2015).
7. Kassem, M. M., Wang, Y., Boomsma, W. & Lindorff-Larsen, K. Structure of the bacterial cytoskeleton protein bactofilin by NMR chemical shifts and sequence variation. *Biophys. J.* **110**, 2342–2348 (2016).
8. Vasa, S. et al. β-Helical architecture of cytoskeletal bactofilin filaments revealed by solid-state NMR. *Proc. Natl Acad. Sci. USA* **112**, E127–E136 (2015).
9. Zuckerman, D. M. et al. The bactofilin cytoskeleton protein BacM of *Myxococcus xanthus* forms an extended β-sheet structure likely mediated by hydrophobic interactions. *PLoS ONE* **10**, e0121074 (2015).
10. Lin, L., Osorio Valeriano, M., Harms, A., Søgaard-Andersen, L. & Thanbichler, M. Bactofilin-mediated organization of the ParABS chromosome segregation system in *Myxococcus xanthus*. *Nat. Commun.* **8**, 1817 (2017).
11. Bulyha, I. et al. Two small GTPases act in concert with the bactofilin cytoskeleton to regulate dynamic bacterial cell polarity. *Dev. Cell* **25**, 119–131 (2013).
12. Blair, K. M. et al. The *Helicobacter pylori* cell shape promoting protein Csd5 interacts with the cell wall, MurF, and the bacterial cytoskeleton. *Mol. Microbiol.* **110**, 114–127 (2018).
13. Sycuro, L. K. et al. Peptidoglycan crosslinking relaxation promotes *Helicobacter pylori*'s helical shape and stomach colonization. *Cell* **141**, 822–833 (2010).
14. Jackson, K. M., Schwartz, C., Wachter, J., Rosa, P. A. & Stewart, P. E. A widely conserved bacterial cytoskeletal component influences unique helical shape and motility of the spirochete *Leptospira biflexa*. *Mol. Microbiol.* **108**, 77–89 (2018).
15. Gode-Potratz, C. J., Kustusch, R. J., Breheny, P. J., Weiss, D. S. & McCarter, L. L. Surface sensing in *Vibrio parahaemolyticus* triggers a programme of gene expression that promotes colonization and virulence. *Mol. Microbiol.* **79**, 240–263 (2011).
16. El Andari, J., Altegoer, F., Bange, G. & Graumann, P. L. *Bacillus subtilis* bactofilins are essential for flagellar hook- and filament assembly and dynamically localize into structures of less than 100 nm diameter underneath the cell membrane. *PLoS ONE* **10**, e0141546 (2015).
17. Rajagopala, S. V. et al. The protein network of bacterial motility. *Mol. Syst. Biol.* **3**, 128 (2007).
18. Mendler, K., Chen, H., Parks, D. H., Hug, L. A. & Doxey, A. AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res.* **47**, 4442–4448 (2019).
19. Parks, D. H. et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
20. El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
21. Ah-Fong, A. M., Kim, K. S. & Judelson, H. S. RNA-seq of life stages of the oomycete *Phytophthora infestans* reveals dynamic changes in metabolic, signal transduction, and pathogenesis genes and a major role for calcium signaling in development. *BMC Genom.* **18**, 198 (2017).
22. Bai, X. C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. Sampling the conformational space of the catalytic subunit of human γ-secretase. *eLife* **4**, e11182 (2015).
23. Ekeberg, M., Lövkvist, C., Lan, Y., Weigt, M. & Aurell, E. Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models. *Phys. Rev. E* **87**, 012707 (2013).
24. Pichoff, S. & Lutkenhaus, J. Tethering the Z ring to the membrane through a conserved membrane targeting sequence in FtsA. *Mol. Microbiol.* **55**, 1722–1734 (2005).
25. Salje, J., van den Ent, F., de Boer, P. & Löwe, J. Direct membrane binding by bacterial actin MreB. *Mol. Cell* **43**, 478–487 (2011).
26. Szeto, T. H., Rowland, S. L., Rothfield, L. I. & King, G. F. Membrane localization of MinD is mediated by a C-terminal motif that is conserved across eubacteria, archaea, and chloroplasts. *Proc. Natl Acad. Sci. USA* **99**, 15693–15698 (2002).
27. McCarthy, C. G. & Fitzpatrick, D. A. Systematic search for evidence of interdomain horizontal gene transfer from prokaryotes to oomycete lineages. *mSphere* **1**, e00195–16 (2016).
28. Richards, T. A., Dacks, J. B., Jenkinson, J. M., Thornton, C. R. & Talbot, N. J. Evolution of filamentous plant pathogens: gene exchange across eukaryotic kingdoms. *Curr. Biol.* **16**, 1857–1864 (2006).
29. Ghosal, D. & Löwe, J. Collaborative protein filaments. *EMBO J.* **34**, 2312–2320 (2015).
30. Hussain, S. et al. MreB filaments align along greatest principal membrane curvature to orient cell wall synthesis. *eLife* **7**, e32471 (2018).
31. Lenarcic, R. et al. Localisation of DivIVA by targeting to negatively curved membranes. *EMBO J.* **28**, 2272–2282 (2009).
32. Gündoğdu, M. E. et al. Large ring polymers align FtsZ polymers for normal septum formation. *EMBO J.* **30**, 617–626 (2011).
33. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).

34. Mende, D. R. et al. proGenomes: a resource for consistent functional and taxonomic annotations of prokaryotic genomes. *Nucleic Acids Res.* **45**, D529–D534 (2017).
35. Price, M. N., Paramvir, D. S., Arkin, P. A. & Poon, A. F. Y. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
36. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).
37. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
38. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
39. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
40. He, S. & Scheres, S. H. W. Helical reconstruction in RELION. *J. Struct. Biol.* **198**, 163–176 (2017).
41. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
42. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
43. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).
44. van den Ent, F., Lockhart, A., Kendrick-Jones, J. & Löwe, J. Crystal structure of the N-terminal domain of MukB: a protein involved in chromosome partitioning. *Structure* **7**, 1181–1187 (1999).
45. Stock, D., Perisic, O. & Löwe, J. Robotic nanolitre protein crystallisation at the MRC Laboratory of Molecular Biology. *Prog. Biophys. Mol. Biol.* **88**, 311–327 (2005).
46. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
47. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
48. Read, R. J. & McCoy, A. J. Using SAD data in Phaser. *Acta Crystallogr. D* **67**, 338–344 (2011).
49. Turk, D. MAIN software for density averaging, model building, structure refinement and validation. *Acta Crystallogr. D* **69**, 1342–1357 (2013).
50. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
51. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
52. Hagen, W. J. H., Wan, W. & Briggs, J. A. G. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* **197**, 191–198 (2017).
53. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
54. Derelle, R., Purificación, L. -G., Timpano, H. & Moreira, D. A phylogenomic framework to study the diversity and evolution of stramenopiles (=Heterokonts). *Mol. Biol. Evol.* **33**, 2890–2898 (2016).

Acknowledgements

We thank H. S. Judelson (UC Riverside) for discussions regarding Oomycetes. We acknowledge Diamond Light Source for the cryo-EM facilities at eBIC. We thank M. Yu (MRC–LMB) for help with synchrotron data collection and the staff at beamline I03 (Diamond Light Source). This work was funded by the MRC (grant no. U105184326 to J.L.) and the Wellcome Trust (grant no. 202754/Z/16/Z to J.L.). J.W. and X.D. were also supported by the Boehringer Ingelheim Fonds.

Author contributions

X.D., A.G.L. D.K.-C. and J.L. performed the protein purification experiments. X.D., A.G.L., G.C. and J.L. collected the cryo-EM data and processed the images. A.G.L. and J.L. performed the crystallization and crystallography. J.M.W. performed the phylogeny and coupling analyses. V.L.H. performed the cryo-ET experiments. S.H.M. and A.G.L. performed the SPR experiments and analyses. A.G.L., J.M.W. and J.L. wrote 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/s41564-019-0544-0.

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

Correspondence and requests for materials should be addressed to J.L.

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

nature research

Corresponding author(s): Jan Löwe

Last updated by author(s): 11th July 2019

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.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on 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 statistical parameters 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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	cryo-EM: FEI EPU 2.3 , Serial EM 3.7 xtallography: beamline software (Diamond GDA version 8.3) Annotree server v1.1, http://annotree.uwaterloo.ca/
Data analysis	cryoEM: MotionCor2 2.1, Gctf 0.50, RELION 3.0, Phenix 1.4, REFMAC 5, MAIN 2017 xtallography: XDS 2018, CCP4 6.4, Phaser 2.83, Phenix 1.4, REFMAC 5, MAIN 2017 HMMER3 v3.1b2 hmmer.org/ Gblocks v0.91b http://molevol.cmima.csic.es/castresana/Gblocks.html FastTree v2.1.10 http://www.microbesonline.org/fasttree/ dcaTools package commit f4c0bd7df5b3949961991e480b0c8034dc71df3d gitlab.com/ducciomalinverni/dcaTools/ lbsDCA commit 47cfe110d0537865352c02f4ace153450a321849 gitlab.com/ducciomalinverni/lbsDCA/

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. 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

Accession codes listed for both xtallography data ad cryo-EM (PDB and EMDB): 6RIA, 6RIB and EMD-4887
No restrictions on availability of any data in the manuscript.

Field-specific reporting

Please select the one below that is 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/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	Electron micrographs were selected from at least two different sessions on two different grids. Each session generated at least 23 images, for cryo-EM always more than 1,000. Representative images for publication were chosen by eye. Details are given the the section 'Statistics and reproducibility'
Data exclusions	Chrystallography: no reflections were excluded Cryo-EM: particle sorting and exclusion in RELION followed standard procedures and the remaining particle numbers are mentioned in the methods section
Replication	Microscopy: at least two replicates for negative staining, at least 5 for cryo-EM and 8 for electron tomography. Replicates means at least different grid preparation for EM.
Randomization	No randomisation was performed.
Blinding	No blinding was performed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	We generated camelid nanobodies against TtBac in this study. These were cloned for bacterial expression and DNA sequenced. The nanobody protein sequence is made available in the manuscript (Table T3).
Validation	We validated the activity of the nanobody by binding it to TtBac filaments and solving the structure of the resulting filaments at molecular resolution.