A Generic Self-Assembly Process in Microcompartments and Synthetic Protein Nanotubes

*Ismail Uddin, Stefanie Frank, Martin J. Warren, Richard W. Pickersgill**

Ismail Uddin, Prof. Richard W. Pickersgill
School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London, E1 4AA, UK
E-mail: r.w.pickersgill@qmul.ac.uk

Dr. Stefanie Frank
Department of Biochemical Engineering, University College London, Gordon Street, London, WC1E 6BT, UK

Prof. Martin. J. Warren
School of Biosciences, University of Kent, Giles Lane, Canterbury, Kent, CT2 7NJ, UK

We reveal the orientation of protein subunits in microcompartments and synthetic nanotubes formed by the bacterial microcompartment shell proteins PduA and PduB.

Microcompartments are pleomorphic structures in bacterial cells that comprise a protein shell harbouring a biochemical pathway and containing a reactive intermediate. PduA and PduB are hexameric and trimeric (pseudo-hexameric) assemblies respectively, which contribute to the shell of the propanediol-utilization (Pdu) bacterial microcompartment. These protein oligomers when produced alone tessellate to form flat sheets in the crystal, and self-assemble to form protein nanotubes in solution. Modelling the arrangement of PduA hexamers in the 20 nm diameter nanotube so as to preserve shape complementarity and key interactions seen in the crystal lattice reveals that the concave surface of the PduA hexameric disk faces out. This subunit orientation is confirmed by experimental measurements of the binding and release of nanotubes from an affinity matrix and by nanoparticle labelling. We also present evidence that shows the concave surface of PduA faces out in the intact bacterial microcompartment

produced *in vivo*. We describe PduB nanotubes for the first time; they have a larger diameter, 63 nm, again with the concave surface of the pseudo-hexameric again facing out. The conserved concave surface out characteristic of these nano-structures reveals a generic assembly process that causes the interface between adjacent subunits to bend in a common direction that optimizes shape complementarity and minimizes steric clashes. This understanding underpins engineering strategies for the biotechnological application of protein nanotubes.

Bacterial microcompartments are proteinaceous organelles found in a wide range of bacterial phyla, they comprise a protein shell encapsulating an enzymatic core and contain a reactive intermediate. The propanediol-utilisation (Pdu) microcompartment is comprised of eight shell proteins forming a closed capsule 100 - 150 nm across which encapsulates a number of enzymes including: diol-dehydratase and propanol dehydrogenase. The eight shell proteins fall into three families: hexamers, trimers and pentamers, which form a pleomorphic pseudo-icosahedral shell around the condensed enzymes[1–3]. PduA, a major shell protein, tessellates in the crystal lattice and when produced on its own, self-assembles into nanotubes[4–6].

Nanotubes, comprising the hexameric bacterial microcompartment protein PduA, assemble spontaneously *in vitro* once the salt concentration is reduced below 0.05 M (**Figure 1a**). These nanotubes, of fixed diameter 20 nm and variable length, are thought to assemble in the same way as the flat sheets of tessellated hexamer subunits seen in the crystal. To form a tube, rather than a flat sheet, either the hexamer is distorted or the interface between hexamers is bent. The rather fragile interface, dominated by electrostatic rather than hydrophobic interactions, appears much easier to bend than the hexamers themselves (**Figure 1b**), so our models are based on making a bend at the interface in such a way as to preserve the characteristic hydrogen–bonding of the antiparallel lysine pair of PduA shown to be essential

for nanotube formation (**Figure 1c**). We have previously proposed two basic models of nanotubes: zigzag with the flat edge of the hexamer subunit approximately parallel to the tube axis and with 10 hexamers per turn with a bend angle of 36° (**Figure 1d**) and armchair with the flat edge perpendicular to the tube axis 12 hexamers per turn with a bend angle of 30° (**Figure 1e**)[5]. Another possibility, close in architecture to the zigzag is a single-start helical model with 10 hexamers per turn, a bend angle of 37.5° and pitch 61 nm (**Figure 1f**). However, these models do not provide an answer to the question of which face of the hexamer is accessible to the lumen of the nanotube and which side is out (**Figure 1g**)? Here, we used several criteria to assess the direction of the bend *in silico* including monitoring the lysine interaction, measuring hydrophobic burial, and counting protein clashes, it is the latter that gave the clearest prediction of preferred bend angle (**Figure 1h,i**; also see **Figure 1b**)[7–10]. Both models preserve the lysine interaction across the two distinct interfaces generated when the hexamers are bent to form a ring and when rings stack to form a tube, but the zigzag model with the concave surface facing out gave minimal clashes compared to the opposite bend angle and either bend angle with the armchair model (**Figure 1h,j**). It is for this reason that we favour the zigzag model or the closely related single start helix model with concave surface facing out. Of these two, the helical model has the bend axis most closely parallel to the edge of the hexamer. Forming a tube, rather than a flat sheet, reduces the edge-effects present in the nano-structure.

The cleavable N-terminal hexa-histidine-tag of PduA used for purification of the hexamer is located on its concave surface. If the orientation of the hexamer in the proposed model is correct then the tagged PduA nanotubes should have the hexa-histidine-tag available for binding to Ni-NTA beads[11], whereas tubes formed from PduA with the tag removed (non-tagged nanotubes) will be unable to bind with the same affinity. We assembled nanotubes from both samples, added Ni-NTA magnetic beads and assessed the release of the nanotubes

from the Ni-NTA beads after unbound nanotubes (due to heavy loading) had been washed off. The tagged PduA tubes were bound as judged from their release at 0.15 M and higher imidazolium concentrations (**Figure 2a),** but the non-tagged nanotubes failed to bind to the Ni-NTA beads (**Figure 2b**). This provides clear evidence that the N-terminus of the protein is available for binding, thereby supporting the calculations showing that the concave surface of PduA faces out. The eluted tubes were subsequently imaged using electron microscopy. This result was substantiated by imaging small Ni-NTA gold nanoparticles (AuNPs) bound to the eluted nanotubes with hexa-histidine tag revealing binding to the exterior of the nanotubes (**Figure 2c**). Far-UV circular dichroism spectroscopy reveals both the hexamers and nanotubes have α-helical content, providing a direct measure of native protein structure in the protein nanotubes (**Figure S1**).

Remarkably, when the hexa-histidine tag was cleaved from the PduA hexamers the nanoparticles still bound to the nanotubes, but now the binding was more regular showing that a better-ordered binding-site had either been created or revealed (**Figure 2d**). Generation of the new N-terminus, sequence GSH (residues -3 to -1), on thrombin cleavage of the histidine tag would create a high-affinity binding site where the amino-terminus [-3] and imidazole side chain of His[-1] bind the nanoparticle[12]. An alternative is that more ordered binding to an accessible histidine was revealed when the hexa-histidine tag was removed. His81 is the only other accessible histidine in PduA and it is in appropriate proximity with Arg79 from the adjacent hexamer in the nanotube to form a high-affinity nanoparticle binding-site. Mutating His[-1] to Ala and His81 to Ala reveals that neither residue alone is responsible for nanoparticle binding in the absence of the hexa-histidine tag, so we conclude that both His[-1] and His81 form high-affinity nanoparticle-binding sites (**Figure 2e,f**). Note that these high affinity binding-sites are not accessible to the affinity matrix used in the binding and release experiments and that mutation of Arg79 prevents the formation of nanotubes.

It has previously been demonstrated that overexpression of the *Citrobacter freundii* propanediol utilization microcompartment shell protein genes, *pduABJKNU*, in *E. coli* produces empty bacterial microcompartments[13] and that cargo proteins can be targeted to the capsule using a specific N-terminal sequence[14,15]. Here we generated a TEV protease cleavable N-terminal mCherry fluorescently tagged PduA (mCherry-TEV-PduA) and co-produced it together with PduB, PduJ, PduK, PduN and PduU to allow the formation of a shell complex[13]. By employing a rhamnose inducible TEV protease it was possible to explore the accessibility of the TEV cleavage site[16], results which strongly indicate that the concave surface of PduA is pointing to the bacterial cytoplasm (see Supporting Information **Figure S2** and Materials for more details). In the absence of induced protease, the fluorescent tag is clearly punctate and associated with the microcompartments, but when the TEV protease is induced cleavage occurs and the fluorescence is no longer discrete, but spreads throughout the bacterial cytoplasm (**Figure 3**). In the absence of the TEV cleavable sequence, expression of the protease has no effect on the distribution of the fluorescence. Production of microcompartments was confirmed using TEM (Supporting Information **Figure S3**). These data provide compelling evidence that the concave surface of PduA is external in the Pdu microcompartment. The interactions involved in the tiling of PduA hexamers seen in the crystal lattice makes it clear that alternating orientations are not possible, a result underlined by AFM studies[3,17].

We also discovered that the trimeric, pseudo-hexameric protein[4,18–20], PduB[21] is capable of forming protein nanotubes when the salt concentration is lowered, but these nanotubes are larger in diameter, typically around 63 nm (**Figure 4a**, with statistical summary in **G**). In contrast to PduA, PduB nanotubes are not labelled by AuNPs (**Figure 4b,c**); unlike PduA the N-terminus of PduB is on the convex side, so this result is consistent with the convex side of

the PduB hexamer inside, reducing accessibility of the His-tag. The algorithms used to generate PduA nanotube models can be used to generate PduB nanotube structures with bending such that the concave surface is external, giving fewer clashes for both zigzag and armchair models. (**Figure 4d-f, i, j**; further details of the modelling are given in Supporting Information). It is remarkable that in the crystal lattice sheets of PduB molecules do not show the precise hydrogen-bonding of antiparallel lysine pairs seen in PduA, the lysine residues are conserved, but the interaction is not, rather the interface is characterised by general shape and electrostatic complementarity. Clearly, the bend angle is shallower in these tubes and this allows for bending along the length of the tubes as a less-severe restraint is placed on the positioning of trimers along the z-axis of the tube.

The evidence from *in silico* and experimental studies suggest these structures have a dimpled surface appearance with the concave surface of the hexamers and trimers facing out of nanotubes and microcompartments. There is a consistent orientation and the walls of both nanotube and microcompartment structures are anticipated to be a single molecule thick. The preferred model of the nanotube is the zigzag or the closely related helical model. That the larger diameter PduB tubes can accommodate bend may be anticipated from the less stringent z-axis (nanotube axis) packing due to the shallower bend angle in the plane of the ring and lack of precise lysine-lysine interaction. Although we provide clear evidence that the concave surface of the hexamers and trimers faces out, the precise organisation of hexameric (PduA) and trimeric (PduB) subunits will likely be revealed by cryo-electron microscopy studies. The subunit orientation demonstrated for PduA nanotubes and the PduA hexamer in the propane-diol utilization microcompartment shell agrees with the recent crystal structure of the *Haliangium ochraceum* microcompartment shell[22]. This remarkable crystal structure of the small microcompartment shell has concave surface of the subunits facing out and the bend

angles between hexamers are consistent with those predicted from our *in silico* modelling of PduA nanotubes.

This knowledge can be used to engineer specific protein-protein interactions to either target proteins more accurately to the lumen of microcompartments or to the outside of the filament scaffolds. To present nanobodies and enzymes on the outside, we now know to label the concave surface of these structures; and to place biochemical pathways within the lumen, they should be introduced on the convex side of the subunits forming these nanostructures.

Experimental Section

Protein expression and purification: The expression of *Citrobacter freundii* PduA used the *pduA** gene, which contains a mutation that knocks out the stop codon and gives an additional 23 C-terminal residues from the pET14b vector (LVKDPAANKARKEAELAAATAEQ). The presence of these additional residues makes PduA* more soluble and easier to work with *in vitro.* PduA*, referred to simply as PduA in this paper, was expressed from Bioline BL21 (DE3) cell lines. The recombinant protein was purified via immobilized metal affinity chromatography (IMAC), using GE Chelating sepharose media (charged with nickel sulphate). *Lactobacillus reuteri* DSM20016 PduB was expressed from a pET-14b plasmid, in BL21 (DE3) cell lines. PduB was subsequently purified via size exclusion chromatography (SEC) using a GE Sephadex S200 HR 10/30 GL column, eluting at around 14 mL, corresponding to its expected molecular weight. Cleavage of the His-tag is described in Supplementary Information. Strains and plasmids used are presented in Supplementary Information **Table 2** and **S3**, respectively.

Preparation of nanotubes: Nanotubes assembled from PduA and PduB were prepared by dialyzing ~100 μL of a purified solution of the proteins overnight at 4°C, into 0.05 M Tris-HCl, 0.05 M NaCl, pH 8.0.

Binding and release of nanotubes to nickel-NTA magnetic beads: 150 μL of PduA nanotube solution (~5 mg mL^{-1}) was diluted in 400 μL final volume of binding buffer (0.05 M Tris-HCl, 0.05 M NaCl, 0.02 M imidazole, pH 8.0). 50 μL of Qiagen Ni-NTA magnetic agarose beads was added to the nanotube solution, and incubated at 4°C on a rocker-shaker at slow speed, for a minimum of 30 minutes. Following incubation, tubes were transferred into a Qiagen single tube magnet for 1 minute, and supernatant transferred to a separate tube. Beads were then washed with 200 μL of 0.02 M imidazole buffer initially, and 100 μL subsequently for buffers of 0.05, 0.15, 0.25 and 0.6 M imidazole. For each wash, supernatant was recovered and a 20 μL sample was taken for loading on to an SDS PAGE gel. A final sample of the magnetic beads after washing with 0.6 M imidazole was also taken for loading on the SDS PAGE gel.

Transmission electron microscopy: Nanotubes were imaged using a JEOL JEM 1230 electron microscope. Samples (5 μL) were prepared by depositing onto carbon coated copper grids (Structure Probe, Inc.) for 90 seconds, followed by 70 seconds of negative staining using 5 μL of 2.0% PTA (phosphotungstic acid). 5 seconds of washing using 5 μL of deionized H_2O was carried out after sample deposition and staining. PTA solution was prepared by dissolving 20 mg of phosphotungstic acid hydrate in 1 mL of deionized H_2O, followed by incubation at 37°C overnight. Immediately prior to use, PTA solution was adjusted to pH 7.4 with 10 M NaOH, and filtered through a 0.2 μm micro-syringe filter.

Nickel-NTA gold nanoparticle binding: Nanoprobes 5 nm Ni-NTA-Nanogold (gold nanoparticles) solution (AuNPs) was diluted 1:50 in 0.05 M Tris-HCl, 0.05 M NaCl, 0.01 – 0.5 M imidazole, pH 8.0. TEM grids were prepared with samples of nanotubes (90 s incubation on grid), and subsequently incubated on a 60 µL droplet of the AuNPs for 30 minutes. Following incubation, the grids were washed twice for 30 s in a droplet of 0.05 M Tris-HCl, 0.05 M NaCl, pH 8.0, before a final wash with ddH_2O for 10 s. Finally, the grids were stained with filtered 2.0% PTA, pH 7.4 stain for 70 s.

In silico modelling and algorithms: Models for nanotubes were generated using custom Python scripts, using the crystal structure of *S. enterica typhimurium* PduA (PDB code: 3NGK) as the building block. For the arrangements, zigzag (1) and armchair (2), the following equations describe the calculation of the bend angle to give rise to a nanotube of the specified radius, using a hexamer/trimer of the specified width:

$$\alpha = \arctan\left(\frac{\frac{w}{2}}{r_z}\right) \times 2 \quad (1)$$

$$\alpha = \arctan\left(\frac{w \times \frac{3}{8}}{r_a}\right) \times 2 \quad (2)$$

The width of the hexamer/trimer in the orientation required for the arrangement is denoted by, w, and the radius of the resulting nanotube is denoted by r.

The helical model was derived by correcting a z-axis rotational offset and subsequently adjusting each successive hexamer/trimer's horizontal offset, such that the horizontal displacement of a full turn of hexamers/trimers corresponded to the pitch of the helix.

Cloning of TEV protease and shell protein construct with TEV cleavage site: Tobacco Edge Virus (TEV) protease was cloned with an N-terminal hexa-histidine tag using Addgene plasmid pRK793 as template (no 8827, S219V mutant). The forward primer included an *Ase*I site (underlined) 5'GACATTAATATGCATCATCATCATCATCATCATGG3' and reverse primer included a BamHI site 5'GACGGATCCTTAGCGACGGCGACGACGATTC3'. The PCR product was digested *Ase*I and *BamH*I and ligated into the NdeI and BamHI sites of the pETpRha vector to form pETpRha-His-TEV. To insert a TEV cleavage site between mCherry and PduA, PduAB from *C. freundii* was amplified using a PduA forward primer including a NdeI site and TEV cleavage sequence 5'CACCATATGGAAAATCTTTATTTTCAAGGTATGCAACAAGAAGCGTTAGGAATG G3' and a PduB reverse primer including a *Spe*I site 5'GGCACTAGTTCAGATGTAGGACGGAC3' using pLysS-mCherry-PduAB as the template. The TEVsite-PduAB PCR product was cloned via *Nde*I and *Spe*I into pLysS-mCherry-PduAB. The mCherry-TEVsite-PduAB fragment was transferred via SbfI and PmlI into the pLysS-TBAD-mCherry-PduA-BJKNU plasmid to form pLysS-TBAD-mCherry-TEVsite-PduA-BJJKNU.

Culturing conditions: DH10β competent cells were co-transformed either with pLysS-TBAD-mCherry-TEVsite-PduA-BJKNU or as a control pLysS-TBAD-mCherry-PduA-BJKNU and pETpRha-His-TEV. To produce the Pdu shell first (pLysS-TBAD-mCherry-TEVsite-PduA-BJKNU) and then the TEV protease (pET-pRha-His-TEV), gene expression was induced with 0.15% w/v L-arabinose for 90 min. To remove L-arabinose cells were washed once in LB medium and re-suspended in 25 mL of fresh LB supplemented with 0.2% w/v L-rhamnose for the induction of the TEV protease and cultured for a further 140 min. In positive control samples TEV protease was induced first for 140 min and then shell proteins were produced for 90 min using the same inducer concentrations as above. The cells were not washed as

TEV protease was supposed to be co-produced with the shell proteins in this control. Washing cells added 30 min to the total induction time which is why final harvesting took place after 4 hr 20 min. Cells that weren't washed were removed from the shaking incubator for 30 min for consistency in sample treatment.

Fluorescence microscopy: Using wide field fluorescence microscopy, whole cells were observed for presence and localization of mCherry signal at the various time points after induction of gene expression. 1 mL cell cultures were harvested, the cells were collected by centrifugation and placed on a 1% agarose-LB pad prepared on a microscope slide and immediately imaged as described before.

TEM analysis of whole cells: 20 mL whole cell samples were collected at 4 hr 20 min for TEM sample preparation. Samples were ultra-thin sectioned on a RMC MT-XL ultra-microtome, collected on un-coated 300 mesh copper grids and stained by incubation in 4.5% uranyl acetate in 1% acetic acid solution for 45 minutes followed by staining with Reynolds lead citrate for 7 minutes. Electron microscopy was performed using a JEOL-1230 transmission electron microscope equipped with a Gatan multiscan digital camera operated at an accelerating voltage of 80 kV.

Western blot analysis of TEV protease and mCherry-PduA expression: Samples were collected at 90 min, 140 min and 4 hr 20 min, adjusted to OD_{600}=1 and loaded onto 15 % SDS PAGE gels for blotting onto nitrocellulose membranes. Protein production of His-TEV protease and cleavage of mCherry-PduA was probed by incubation of membranes with (i) primary monoclonal anti-poly-histidine antibody (Sigma, H1029) at a dilution of 1:3000 followed by secondary anti mouse IgG, alkaline phosphatase (AP) conjugated at a dilution of 1:5000 and (ii) primary anti-mCherry (Abcam, 1 mg mL^{-1}) antibody at 1:2500 followed by

secondary anti rabbit IgG antibody, AP-conjugated at 1:5000. Bands were visualised by incubation in substrate 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium (BCIP/NBT).

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work was supported by the Biotechnology and Biological Sciences Research Council of the UK strategic LoLa Award (BB/M002969/1 to MJW and RWP), BBSRC KTN (BB/M503149/1 to RWP) and by the Leverhulme Trust (ECF-2013-341 to SF). The construct used to produce *L. reuteri* PduB was kindly supplied by Mike B. Prentice (University College Cork) and Mingzhi Liang (University of Kent).

Received: ((will be filled in by the editorial staff))

Revised: ((will be filled in by the editorial staff))

Published online: ((will be filled in by the editorial staff))

References

[1] T. A. Bobik, G. D. Havemann, R. J. Busch, D. S. Williams, H. C. Aldrich, *J. Bacteriol.* **1999**, *181*, 5967.

[2] D. Walter, M. Ailion, J. Roth, D. Walter, M. Ailion, J. Roth, **1997**, *179*.

[3] T. O. Yeates, M. C. Thompson, T. a Bobik, *Curr. Opin. Struct. Biol.* **2011**, *21*, 223.

[4] C. S. Crowley, D. Cascio, M. R. Sawaya, J. S. Kopstein, T. a Bobik, T. O. Yeates, *J. Biol. Chem.* **2010**, *285*, 37838.

[5] A. Pang, S. Frank, I. Brown, M. J. Warren, R. W. Pickersgill, *J. Biol. Chem.* **2014**, *289*,

22377.

[6] C. R. Noël, F. Cai, C. A. Kerfeld, **2015**, DOI 10.1002/admi.201500295.

[7] J. D. Hunter, *Comput. Sci. Eng.* **2007**, *9*, 90.

[8] S. van der Walt, S. C. Colbert, G. Varoquaux, *Comput. Sci. Eng.* **2011**, *13*, 22.

[9] P. J. A. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, M. J. L. de Hoon, *Bioinformatics* **2009**, *25*, 1422.

[10] Schrödinger, LLC, *The {PyMOL} Molecular Graphics System, Version~1.8*, **2015**.

[11] V. Reddy, E. Lymar, M. Hu, J. F. Hainfeld, *Microsc. Microanal.* **2005**, *11*, 1118.

[12] P. F. Predki, C. Harford, P. Brar, B. Sarkar, *Biochem. J.* **1992**, *287 (Pt 1*, 211.

[13] J. B. Parsons, S. Frank, D. Bhella, M. Liang, M. B. Prentice, D. P. Mulvihill, M. J. Warren, *Mol. Cell* **2010**, *38*, 305.

[14] C. Fan, S. Cheng, Y. Liu, C. M. Escobar, C. S. Crowley, R. E. Jefferson, **2010**, *107*, 7509.

[15] J. N. Kinney, A. Salmeen, F. Cai, C. A. Kerfeld, *J. Biol. Chem.* **2012**, *287*, 17729.

[16] T. A. Eastwood, K. Baker, H. R. Brooker, S. Frank, D. P. Mulvihill, *FEBS Lett.* **2017**, *591*, 833.

[17] M. Sutter, M. Faulkner, C. Aussignargues, B. Paasch, S. Barrett, C. A. Kerfeld, L. Liu, *Nano Lett.* **2015**, acs. nanolett.5b04259.

[18] D. Heldt, S. Frank, A. Seyedarabi, D. Ladikis, J. B. Parsons, M. J. Warren, R. W. Pickersgill, **2009**, *207*, 199.

[19] A. Pang, M. J. Warren, R. W. Pickersgill, *Acta Crystallogr. D. Biol. Crystallogr.* **2011**, *67*, 91.

[20] M. G. Klein, P. Zwart, S. C. Bagby, F. Cai, S. W. Chisholm, S. Heinhorst, G. C. Cannon, C. A. Kerfeld, *J. Mol. Biol.* **2009**, *392*, 319.

[21] A. Pang, M. Liang, M. B. Prentice, R. W. Pickersgill, *Acta Crystallogr. D. Biol.*

Crystallogr. **2012**, *68*, 1642.

[22] M. Sutter, B. Greber, C. Aussignargues, C. A. Kerfeld, *Science* **2017**, *356*, 1293.

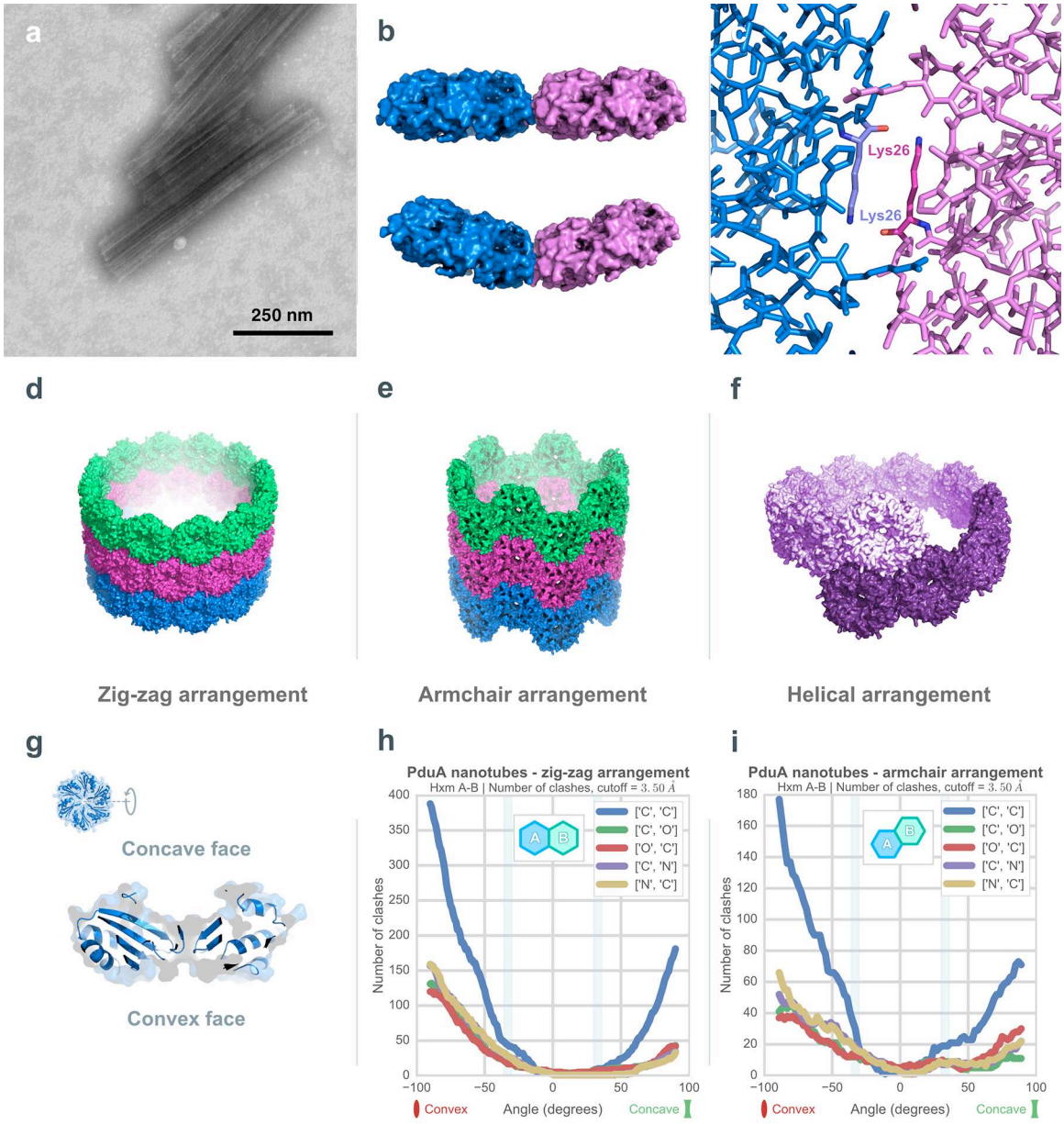


Figure 1. Modelling the architecture of PduA protein nanotubes based on the packing of PduA seen in the crystal. **a** Transmission electron micrograph of PduA nanotubes, consistently 20 nm in diameter. Scale bar 250 nm. **b** Top-down view of two adjacent PduA hexamers, illustrating the fragile hexamer-hexamer interface, at both 0° (top) and 36° (bottom) bend angle. **c** The anti-parallel lysine arrangement between adjacent Lys26 residues, held together by hydrogen bonding. **d, e & f** Zig-zag, armchair and helical arrangement model of PduA, respectively. **g** Top-down view of a PduA hexamer and cross-section (expanded), illustrating the concave and convex faces of the hexamer. The N-terminal hexa-histidine tag is on the concave surface. **h & i** Number of protein clashes counted at the PduA hexamer-hexamer interface, as a function of bend angle, for the zig-zag and armchair arrangement, respectively. The zig-zag or closely related helical model with concave surface facing out is preferred from the *in silico* modelling.

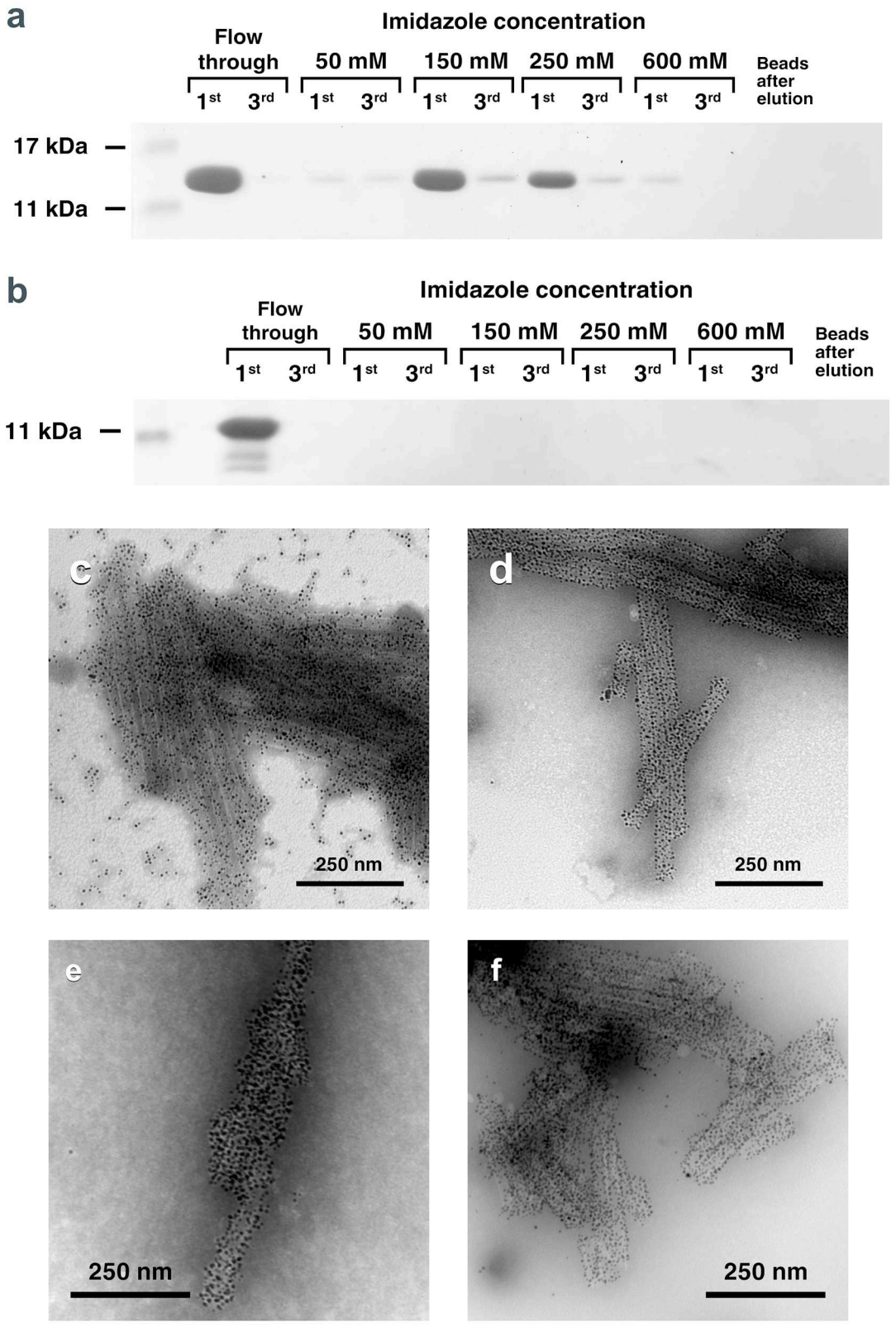


Figure 2. Evidence concerning the orientation of PduA hexamers in protein nanotubes**. a** SDS PAGE gel demonstrating the binding and release of hexa-histidine tagged PduA nanotubes from Ni-NTA magnetic beads. The beads were heavily loaded with nanotubes and the first

pair of lanes shows the flow-through, a considerable quantity of tubes were washed off in the first wash, none in the third (final) wash of the beads. Subsequent pairs of lanes show the elution in increasing imidazole concentrations. Again, the first and final washes are presented. There was no significant elution at 0.05 M imidazole. The nanotubes started to elute at 0.15 M imidazole. **b** SDS PAGE gel for the control experiment, demonstrating the lack of binding of non-tagged PduA nanotubes. The nanotubes all appear in the flow-through, none are eluted or remain on the beads. **c** Transmission electron micrograph of tagged PduA nanotubes binding gold nanoparticles in the presence of 0.25 M imidazole. **d** Transmission electron micrograph of non-tagged PduA nanotubes binding gold nanoparticles in the presence of 0.25 M imidazole. The binding of nanoparticles is tighter and more precise to the less-flexible N-terminal sequence GSH formed after cleavage of the external histidine-tag than to the longer and more flexible hexa-histidine tag. The much larger Ni-NTA beads used in the binding and release experiment using cleaved PduA do not have access to the GSH sequence. The binding and release experiments and the nanoparticle-binding images show the concave side of PduA faces out. **e** PduA H81A nanotubes with AuNPs bound under 0.25 M imidazole concentration. **f** PduA H[-1]A nanotubes with AuNPs bound under 0.25 M imidazole concentration.

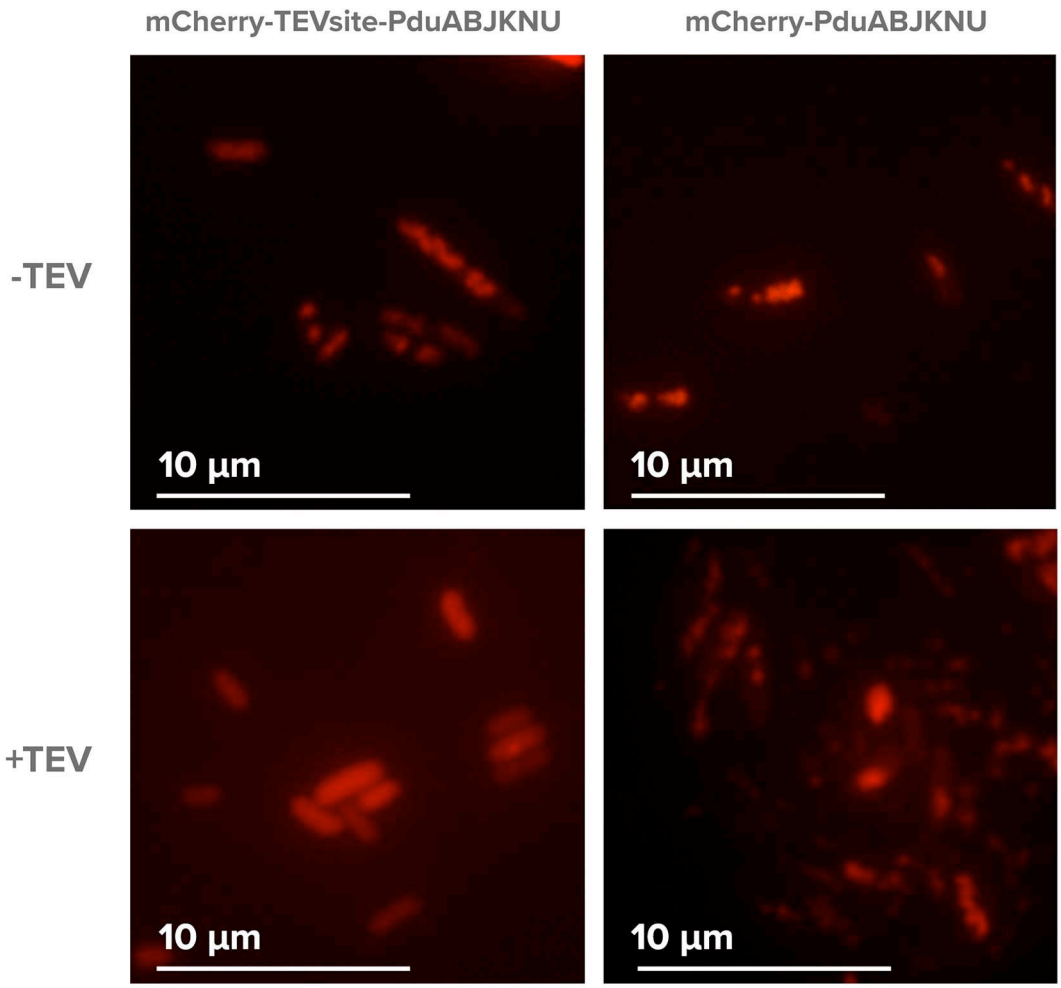


Figure 3. Evidence on the orientation of PduA hexamers in Pdu protein microcompartments. mCherry localization in live E. coli cells before (-TEV) and after (+TEV) induction of TEV protease. The strains expressing mCherry-TEVsite-PduABJKNU and the negative control mCherry-PduABJKNU were induced for 90 min, the cells were washed with fresh medium to remove inducer and stop production of more shell proteins and then the TEV protease was induced. In both strains, the mCherry signal is punctate before the production of TEV protease indicating that bacterial microcompartments are formed and mCherry is co-localized with these structures. After TEV protease has been induced for 140 min the mCherry signal is cytoplasmic in the mCherry-TEVsite-PduABJKNU-producing strain (bottom left) but appears as foci in the negative control strain lacking the TEV-cleavage site between mCherry and PduA (mCherry-PduABJKNU). These data provide clear evidence that the TEV cleavage site is on the outer surface of the microcompartment and therefore the concave surface faces out.

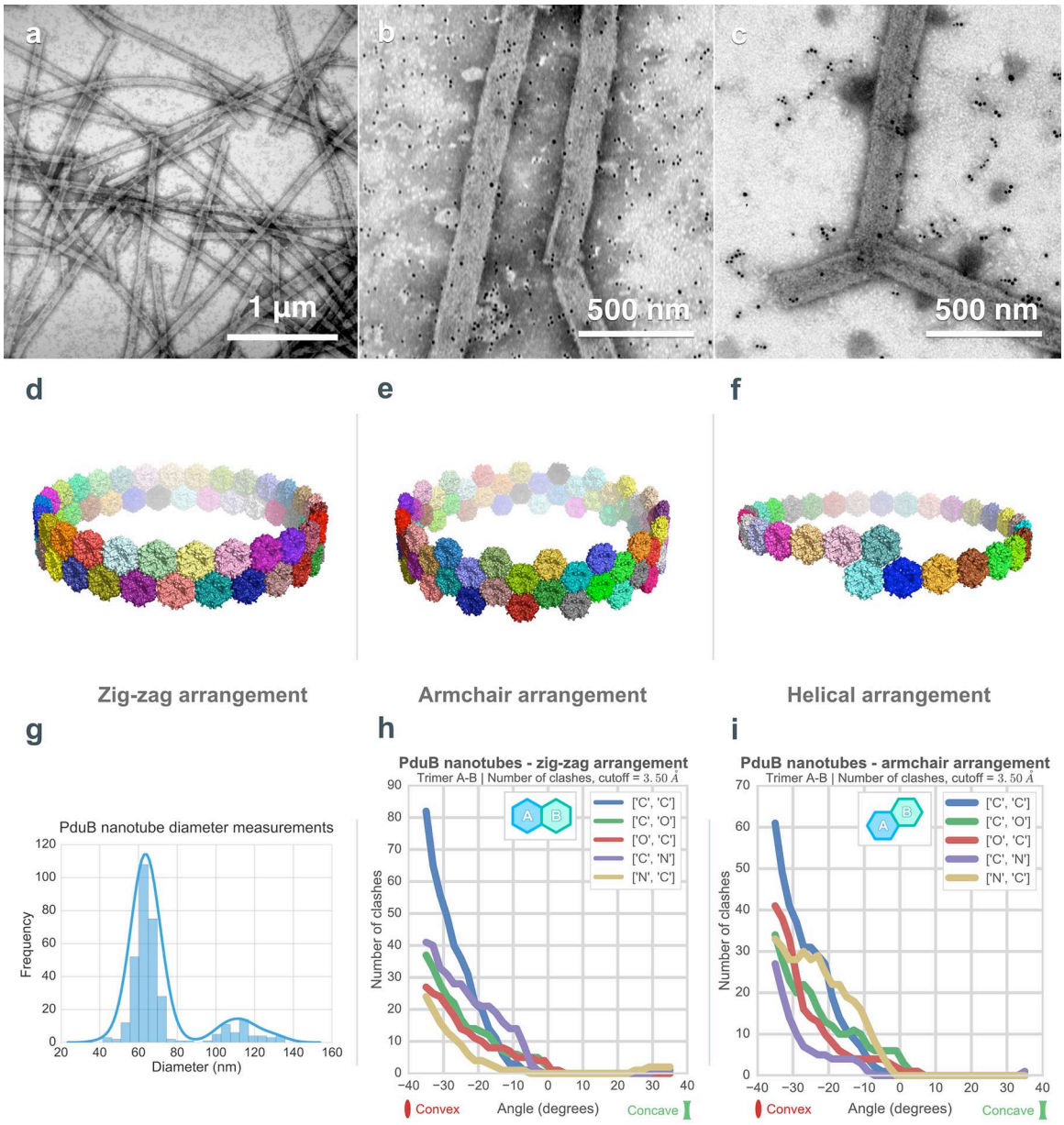


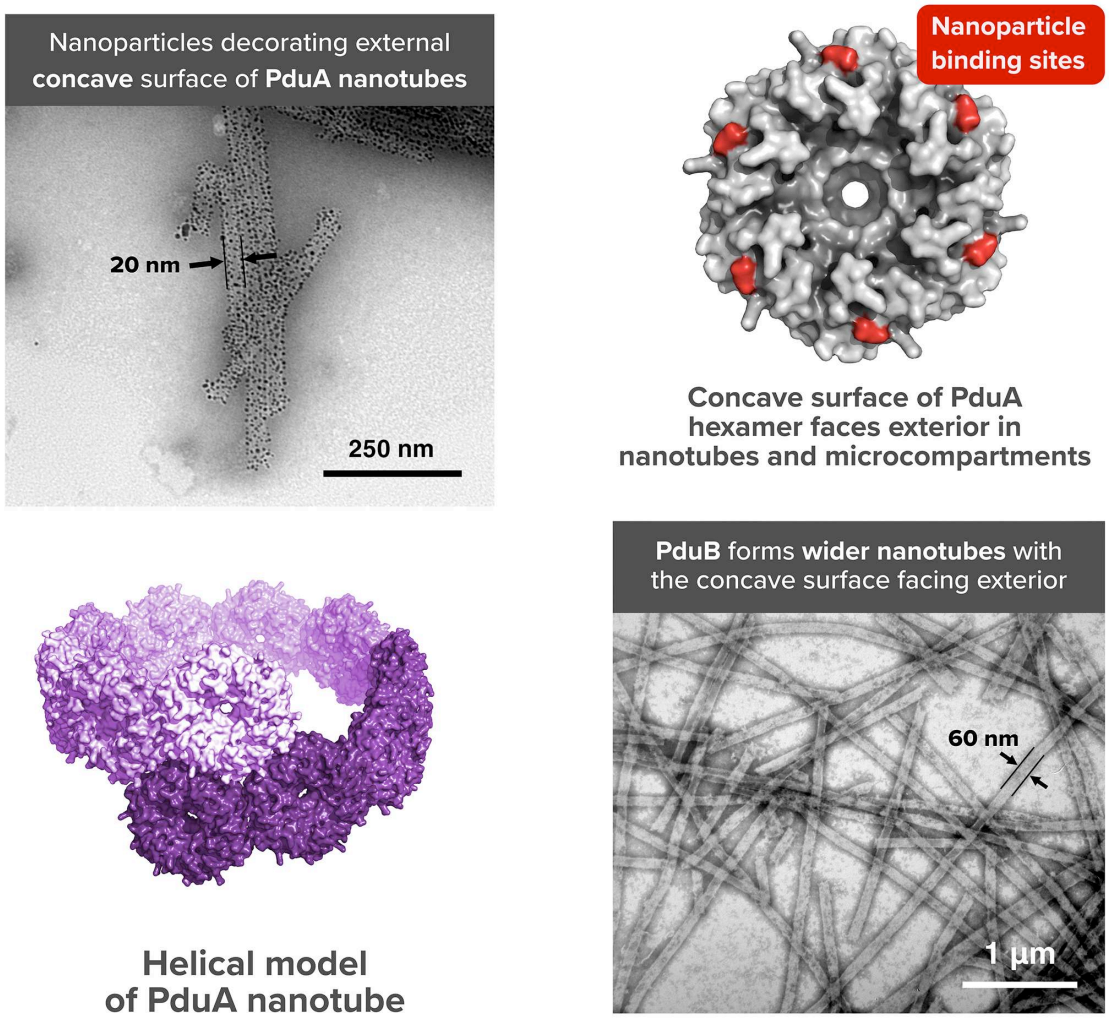
Figure 4. PduB nanotubes are larger and show more structural diversity but models suggest the concave surface of PduB also faces out. **a** Transmission electron micrographs of negatively stained PduB *L. reuteri* nanotubes, depicting the large variability in diameter and length, as well as extent of curvature along their lengths. Scale bar 1 µm. **b & c** PduB nanotubes labelled with gold nanoparticles in 0.08 and 0.25 M imidazole buffer respectively **d-f** *In silico* generated models for PduB nanotubes, depicting the zig-zag, armchair and helical arrangement respectively. **g** Histogram of PduB nanotubes diameters measured (n=336), illustrating their bimodal size distribution. **h & i** Number of protein clashes counted at the PduB trimer-trimer interface, as a function of bend angle, for the zig-zag and armchair arrangement, respectively. The concave surface out is again preferred as fewer clashes are seen in this structure.

The table of contents entry should be 50−60 words long, and the first phrase should be bold.

A generic assembly process is revealed in PduA and PduB nanotubes and in the bacterial microcompartment which results in the dimpled (concave) surface of the protein oligomers external. This is due to the inherent tendency of the hexamers and pseudo-hexamers to associate and bend in one direction, minimizing steric conflict, optimizing shape complementarity and reducing edge-effects.

Keywords
Bacterial microcompartment shell proteins, protein nanotubes, transmission electron microscopy, self-assembly

A generic self-assembly process in synthetic protein nanotubes



ToC figure

Supporting Information

A generic assembly process in synthetic protein nanotubes

Ismail Uddin, Stefanie Frank, Martin J. Warren, Richard W. Pickersgill*

Additional experimental details

Cleavage of hexa-histidine tag: The hexa-histidine tag of PduA was cleaved off using the N-terminal thrombin cleavage site provided by the pET-14b vector. Purified solution of PduA was buffer exchanged using a PD-10 column into PBS buffer, pH 7.4 and supplemented with 15 U of 1 U/μL of GE thrombin. Cleavage was carried out overnight at 4°C on a rocker/shaker. Following cleavage, the reaction was halted using 0.001 M final concentration of PMSF. The protein sample was then incubated for 10 minutes with 3 mL of GE chelating sepharose slurry (pre-charged with nickel sulphate) and 5 mL of binding buffer (0.05 M Tris-HCl, 0.5 M NaCl, 0.01 M imidazole, pH 8.0). The mixture was loaded onto an empty PD-10 column, and the flow through collected with an additional 15 mL of binding buffer passed through the column. The flow through was then concentrated in a Vivaspin 30 kDa MWCO concentrator prior to nanotube preparation.

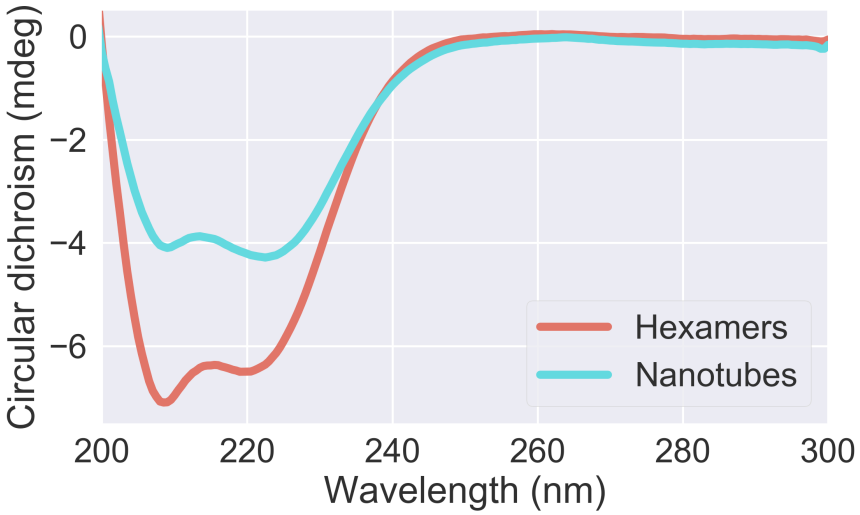


Figure S1 Circular dichroism spectroscopy with characteristic bands at 208 and 222 nm showing the α-helical content of PduA hexamers and nanotubes confirming α-helical structure remains after the nanotubes are formed. There is no evidence to suggest conversion of α-helix to β-sheet (β-sheet band is at 214 nm).

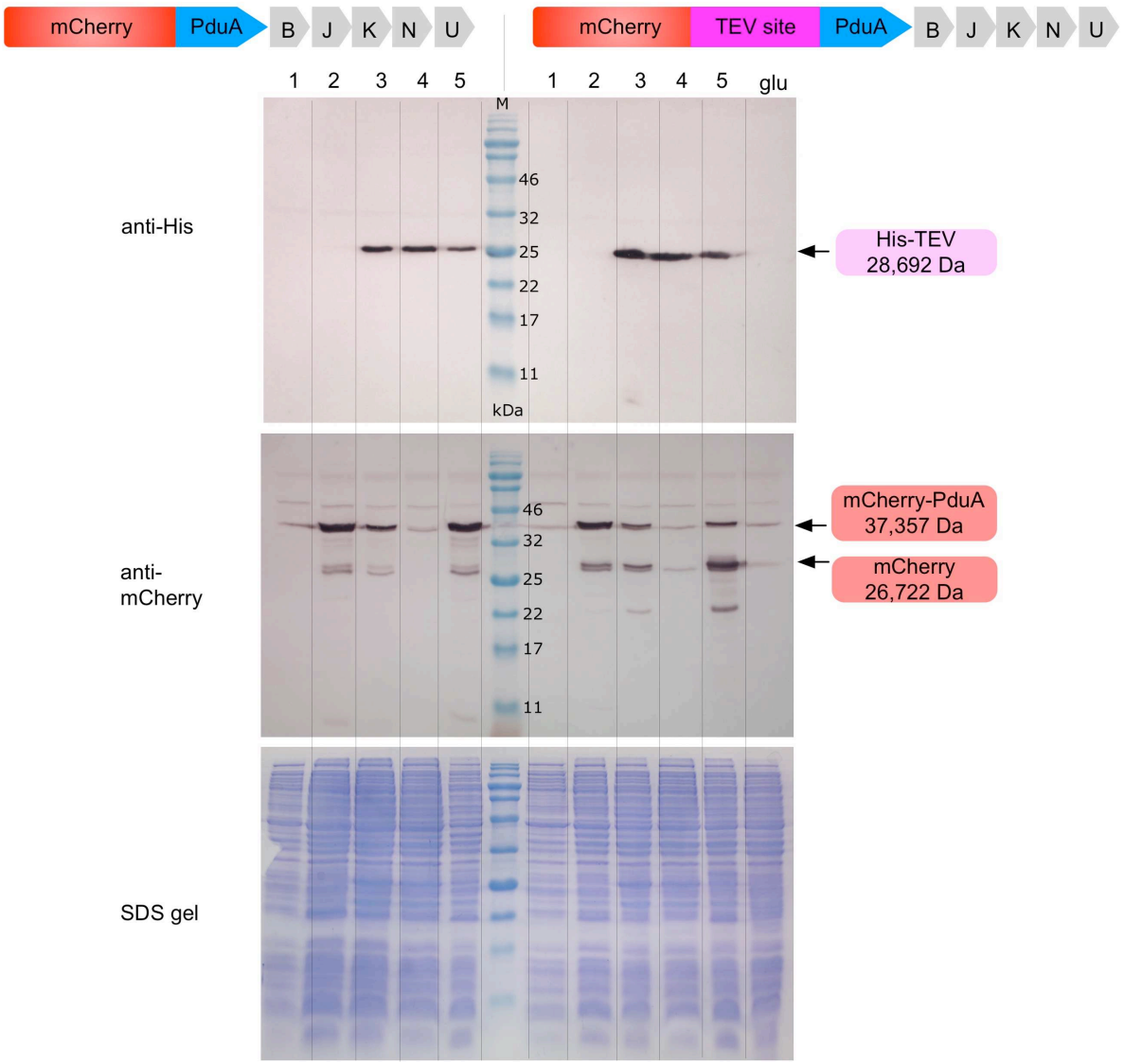


Figure S2 Western blot showing expression levels of His-TEV (anti-his antibody) and mCherry (anti-mCherry antibody). Cleavage of mCherry from PduA is accompanied by the dis-appearance of the larger mCherry-PduA band (37.4 kDa) and the appearance of a mCherry-sized band (26.7 kDa). All samples were adjusted to OD_{600}=1. Lane 1: no induction for 4 hrs. 20 min; lane 2: sample after induction of shell proteins for 90 min (no TEV protease); lane 3: sample after induction of shell proteins for 90 min, washing of cells to stop shell proteins being produced and subsequent induction of TEV protease for 140 min; lane 4: after production of TEV protease for 140 min (no shell proteins); lane 5: after production of TEV protease for 140 min and subsequent production of shell proteins (cells were not washed in between to allow TEV protease to be still produced during shell formation); lane 'glu': growth medium contained 0.2 % glucose to repress background gene expression, no induction for 4 hrs. 20 min. Lanes 5 represent positive controls were TEV protease is produced before the shell proteins assemble, gaining access to the TEV cleavage site. Cleavage of mCherry-PduA can be observed in the lysate of the strain producing mCherry-TEVsite-PduABJKNU but as expected not in the lysate of the negative control lacking the TEV cleavage site (mCherry-PduABJKNU). There is minimal background cleavage observed in all samples.

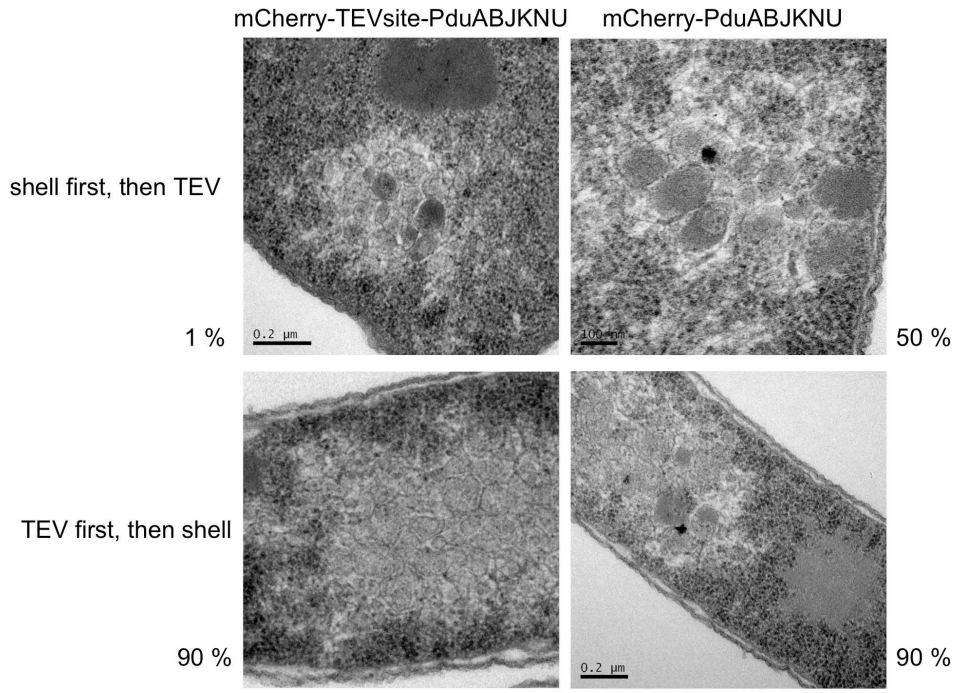


Figure S3 TEM analysis of sections through whole cells of E. coli DH10β showing empty microcompartment formation. The percentage of cells that contained microcompartments is shown next to the micrographs. 100 cells were examined per sample. Strain producing mCherry-TEVsite-PduA: top left: cells harvested after induction of shell proteins for 90 min, washing of cells to stop shell proteins being produced and subsequent induction of TEV protease for 140 min; bottom left: TEV protease for 140 min and subsequent production of shell proteins (cells were not washed in between to allow TEV protease to be still produced during shell formation). Strain producing mCherry-PduA (no TEV cleavage site): top right: cells harvested after induction of shell proteins for 90 min, washing of cells to stop shell proteins being produced an subsequent induction of TEV protease for 140 min; bottom right: TEV protease for 140 min and subsequent production of shell proteins (cells were not washed in between to allow TEV protease to be still produced during shell formation).

Table S1. Measurements of PduB nanotube diameters.

	Overall population	Smaller diameter population	Larger diameter population
Mean (nm)	71	63	113
Standard deviation	19	5	11
Sample number	366	282	54
Bend angle range for zig-zag arrangement	20 – 6°	20 – 10°	10 – 6°
Bend angle range for armchair arrangement	15 – 4°	15 – 7.5°	7.5 – 4°

Table S2. Strains used in this study

Strain	Genotype	Source
DH10β	F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 endA1 recA1 deoR Δ(ara,leu)7697 araD139 galU galK nupG rpsL λ-	Promega
BL21 (DE3)	F– ompT hsdSB (rB– mB–) gal dcm (DE3)	Bioline

Table S3. Plasmids used in this study.

Strain	Genotype	Source
pET-14b *pdu*A*	pET-14b containing *C. freundii pduA*.pduA** has a mutation in the natural stop codon, causing a read-through into the pET-14b vector and incorporating an additional 23 residues into the translated sequence.	Parsons *et al.* 2010
pET-14b *pdu*B *Cf*	pET-14b vector containing *C. freundii pduB*	Parsons *et al.* 2010
pET-14b *pdu*B *Lr*	pET-14b vector containing *L. reuteri pdu*B	Pang *et al.* 2011
pLysS-TBAD-mCherry-PduA-BJKNU	pET-TBAD-cobA (T7 promoter of pET3a swapped with BAD promoter, gift from Warren lab) and pET3a-mCherry-PduABJKNU both digested with XbaI and HindIII and ligated to give pET-TBAD-mCherry-PduABJKNU. Plasmid was digested with BglII and SpeI and TBAD-mCherry-PduABJKNU fragment ligated into pLysS containing a BglII followed by a pET3a multiple cloning site containing a SpeI just after the BamHI site. All genes derived from C. freundii.	This study
pLysS-mCherry-PduAB	Overexpression vector containing T7 promoter and C. freundii Pdu shell protein genes mCherry-pduA and pduB.	Parsons *et al.* 2010
pLysS-TBAD-mCherry-TEVsite-PduA-BJKNU	TEVsite-pduAB PCR product was cloned via NdeI and SpeI into pLysS-mCherry-pduAB. mCherry-TEVsite-pduAB fragment was transferred via SbfI and PmlI into pLysS-TBAD-mCherry-PduA-BJKNU.	This study
pET-pRha	T7 promoter of pET3a was swapped for pRha promoter.	Eastwood *et al.* 2007
pET-pRha-His-TEV	PCR product of His-TEV was digested with *Ase*I and *BamH*I and ligated into the *Nde*I and *BamH*I sites of pET-pRha.	This study