



Cite this: *Chem. Commun.*, 2016, 52, 5041

Received 27th January 2016,
Accepted 9th March 2016

DOI: 10.1039/c6cc00851h

www.rsc.org/chemcomm

Structure of a novel 13 nm dodecahedral nanocage assembled from a redesigned bacterial microcompartment shell protein†

J. Jorda,^a D. J. Leibly,^{ab} M. C. Thompson^b and T. O. Yeates^{*ab}

We report the crystal structure of a novel 60-subunit dodecahedral cage that results from self-assembly of a re-engineered version of a natural protein (PduA) from the Pdu microcompartment shell. Biophysical data illustrate the dependence of assembly on solution conditions, opening up new applications in microcompartment studies and nanotechnology.

Nature has evolved exquisite symmetrical structures across a range of length scales. Examples of such structures based on self-assembling proteins include virus capsids, clathrin, ferritin and bacterial microcompartments (MCPs).^{1–6} These assemblies encapsulate nucleic acids, endocytic cargo, iron and metabolic enzymes, respectively. Two common features of such structures are (1) their assembly from many copies of one or a small number of distinct protein subunit types, and (2) highly symmetric arrangements of the subunits, typically having cubic or icosahedral (or related dodecahedral) forms.

Inspired by nature, an emerging emphasis within the field of bionanotechnology is the design and production of novel three dimensional protein assemblies that might serve as molecular containers. Hollow, nanoscale structures have attracted considerable interest because they have the potential to be engineered for the targeted biological delivery of cargo, including drug molecules and imaging reagents such as dyes and nanoparticles.^{7,8} Molecular cages or shells built from protein subunits are privileged platforms for bionanotechnology applications because their properties can be modulated easily by changes to their amino acid sequences, and they can be produced using recombinant overexpression technologies. Efforts aiming at engineering proteins to self-assemble into complex polyhedral cages have led to a series of recent successes.^{9–15} Strategies aimed at engineering proteins to form geometrically regular architectures have focused on

targets obeying the symmetries of the Platonic solids: tetrahedral, cubic/octahedral and icosahedral/dodecahedral.¹⁶ A series of designed cages based on tetrahedral symmetry (12 copies of one or two distinct subunit types) and cubic/octahedral symmetry (24 copies of one or two distinct subunit types) have been validated in detail with crystallographic studies.^{10–12,17,18} This leaves icosahedral/dodecahedral architectures with 60 equivalent subunits – the highest possible cubic point symmetry in three-dimensions – as the ultimate target for designing novel protein cages.

Here we report a dodecahedral cage that self-assembles from 60 copies of a redesigned protein building block from the bacterial microcompartment shell (or BMC) family. The protein PduA is a major component of the shell of the propanediol utilization (Pdu) microcompartment in *Salmonella typhimurium*. As part of its natural function, PduA forms a 6-fold symmetric cyclic hexamer.¹⁹ This hexamer exhibits a shape and chemical self-complementarity at its perimeter that promotes further side-by-side assembly of hexamers to form a tightly packed molecular layer about 2 nm thick, which comprises (along with other related proteins) the outer shell of the bacterial microcompartment.^{20,21} In total, roughly five to 15 thousand protein subunits make up the entire shell structure, whose diameter can range from about 100 to 200 nm. We were motivated to radically redesign the PduA protein based on the observation that certain unusual members of the BMC protein family have undergone cyclic permutations during evolution, giving three-dimensional structures that are built from the same arrangement of secondary structure elements, but in a different linear order in the protein sequence.²² An interesting aspect of the permuted type of BMC protein is that one such protein (EutS) was revealed earlier to be unusually flexible, assembling into a bent rather than a flat hexamer.²³ We therefore began with the PduA protein sequence and converted its topology to match that of the most closely related BMC protein known to have a permuted topology, PduU.²²

The circular permutation of a protein is a topological reorganization of its sequence whereby the initial termini are linked (sometimes requiring a short intervening polypeptide), while new termini are created by a disconnection elsewhere in the sequence.

^a UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, CA 90095, USA. E-mail: yeates@mbi.ucla.edu

^b Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c6cc00851h

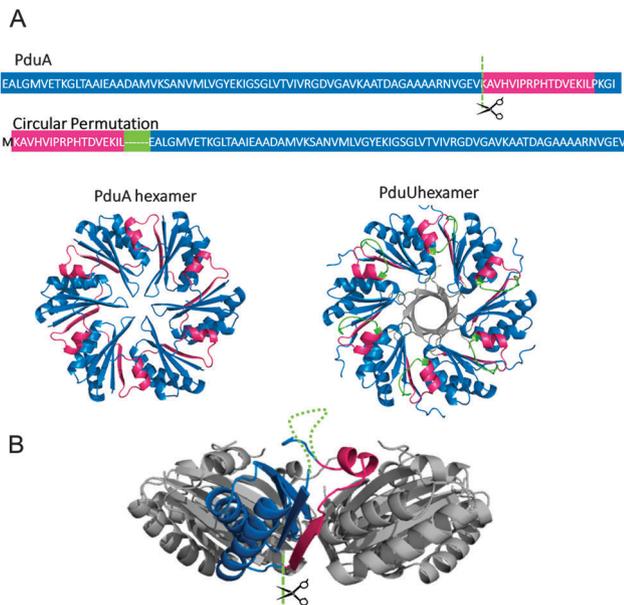


Fig. 1 A redesign of the PduA sequence by circular permutation. The C-terminal segment of the BMC domain (pink) was appended to the N-terminus (blue) with a linking loop extracted from the corresponding PduU sequence, depicted in green (A). A rotation by 90° of the PduA hexameric crystal structure, illustrating in green dots where the new linker was introduced to join the termini of native PduA, and with a scissor symbol where the new termini were created (B).

Our designed circular permutation of PduA borrowed from the PduU topology, but kept the native PduA sequence wherever possible. The PduA termini were connected with the same linker sequence as seen in PduU, while new chain termini were introduced in the location where they occur in PduU (Fig. 1). The feasibility of creating a circularly-permuted PduA construct was first evaluated computationally. We created a permuted version of the protein coordinates by removing a C-terminal segment of PduA, and appending it to the N-terminus, grafting a peptide linker taken from PduU. The structure of this chimeric protein was subsequently refined with the program Modeller.²⁴ Out of 100 independent computer runs, the model with the best energy (DOPE) score²⁵ was defined as the starting model (named P1). Three additional variations on the design were considered and evaluated computationally. Design variations P2 and P3 contained amino acid sequence changes suggested by the Rosetta Matdes program.¹¹ A final design variation, P4, featured a short linker sequence, GGSGGS, chosen for high flexibility. Full protein sequences are given in ESI,† Table S1. To assess the quality of these models, the Rosetta energy scores of the designed hexamers were calculated after relaxing the strict symmetry constraints in a custom protocol¹¹ followed by a geometry validation step by the ramalyze and rotalyze routines in the program PHENIX.²⁶ Final Rosetta scores for each model were -1225 , -1636 , -1658 and -1850 Rosetta Energy Units for P1, P2, P3 and P4, respectively. Unexpectedly, the P4 design with the empirically chosen glycine and serine linker was predicted to be the most stable design.

The P1–P4 protein constructs were created *via* gene synthesis and the amplicons were cloned into the pET-22b expression

vector *via* Gibson assembly.²⁷ Protein expression was carried out in *Escherichia coli* BL21 cells, and recombinant proteins were purified using metal affinity chromatography, facilitated by inclusion of a hexahistidine tag in the protein sequence, which was subsequently removed by treatment with TEV protease. A final gel filtration step resulted in pure protein samples (ESI,† Fig. S1). A major peak with an estimated molecular weight of 45 kDa was collected and concentrated to $\sim 10 \text{ mg ml}^{-1}$ in 50 mM tris pH 9, 50 mM NaCl for each protein.

Crystallization trials were conducted by hanging drop vapor diffusion on the four variations on the designed protein. In each instance initial screening was performed at a protein concentration of 5 mg ml^{-1} with up to five commercially available sparse matrix screens. Design variation P4 was the only case that gave crystals readily. Conditions for crystal growth were subsequently optimized, and high-quality crystals were obtained by diluting the protein to 2.5 mg ml^{-1} and crystallizing by vapor diffusion against a well solution of 0.1 M tris pH 8.5, 1.8 M ammonium sulphate, and 1.25% w/v PEG-10 000. Crystals grew in 1 week, after which they were soaked in 25% 1,2-propanediol as a cryoprotectant, and flash frozen in liquid nitrogen. X-ray diffraction data were collected at the Advanced Photon Source (NE-CAT beamline 24-ID-C). X-ray diffraction extended to a resolution of 2.5 Å, and the data were reduced in space group $P4_232$ using the program XDS/XSCALE.²⁸ The program PHASER²⁹ was used to obtain phases by molecular replacement with a PduA monomer (PDB 4PPD²⁰) serving as the search model. The molecular replacement solution revealed five monomers in the asymmetric unit, and a high solvent content of 56%. Structural refinement was performed by iterative rounds of model adjustment and refinement using Coot³⁰ and PHENIX²⁶ respectively. Structure validation was performed with PHENIX. The refined atomic coordinates and structure factors were deposited in the PDB under accession code 5HPN.

Surprisingly, the crystal structure of the redesigned, circularly-permuted PduA revealed a dodecahedral cage made of 60 copies of the protein subunit (Fig. 2A). The asymmetric unit of the crystal contains a single pentamer. The crystallographic symmetry operators then produce a protein assembly with icosahedral point group symmetry. Each pentameric unit constitutes one of the 12 faces of a dodecahedron (Fig. 2A). This polyhedral assembly is 13 nm in diameter and encloses an inner space with a diameter of about 7 nm and a volume of approximately 180 nm^3 . This 60 subunit cage is the first reported structure of a novel synthetic protein complex with icosahedral symmetry, though a number of new icosahedral protein architectures created by design have been obtained in recent work (Jacob Bale, Neil King, and David Baker, unpublished data).

The formation of this dodecahedral protein cage results from two structural changes introduced by the circular permutation. First, there is a decisive alteration in the primary oligomerization state of the protein. A switch from the typical cyclic 6-fold hexameric arrangement observed for native BMC-family proteins to a 5-fold pentameric arrangement is critical to the architecture observed; icosahedral point group symmetry requires pentagonal units with 5-fold symmetry. Interestingly, this switch occurs in

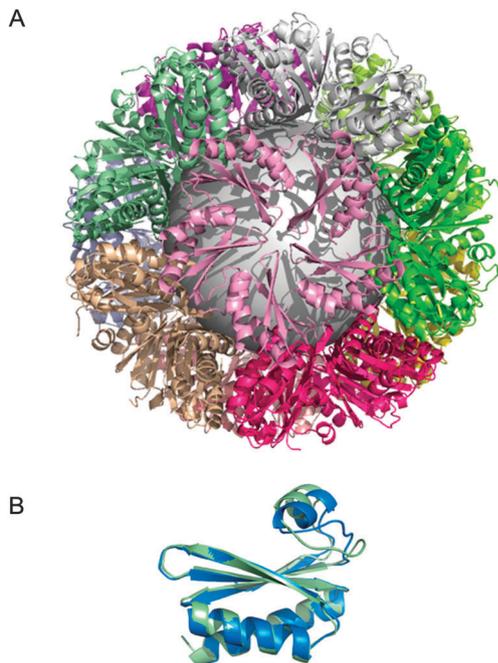


Fig. 2 Crystal structure of the circularly permuted PduA protein (design variant P4). (A) The unit cell of the crystal contains a cage of 12 self-assembling pentamers in the approximate shape of a dodecahedron with icosahedral symmetry, similar to those found in simple viral capsids. (B) A detailed look at the structure of a single protein subunit shows that the computational design (pale green) is in close agreement with the crystal structure (blue), but with notable differences in a loop region and the position of a short alpha helical segment.

the absence of mutations in the protein–protein interface that is responsible for the cyclic oligomerization; mutations near the subunit interface were avoided during the design stage. With the pentagonal shape forbidding a flat tessellation, but with the altered oligomeric unit still evidently exhibiting a tendency to self-associate further, the result is a novel dodecahedrally shaped structure with icosahedral symmetry.

Despite the dramatic architectural alterations evident in the dodecahedral cage, many of the features believed to be characteristic of bacterial microcompartment shells are recapitulated. The self-assembled pentamers form a tight, almost seamless interface. Likewise, the pore at the center of the pentamer is still present, although its diameter is reduced from roughly 6 Å in wild-type PduA to about 3.5 Å as a result of forming a smaller cyclic oligomer. Finally, a structural alignment of the monomer observed in the crystal structure with the computationally designed model reveals an overall difference of only 1.1 Å (rmsd). The agreement in the core of the protein domain is even closer; the glycine and serine loop region and a short alpha-helical segment account for most of the deviation (Fig. 2B).

The unexpected observation of the dodecahedral cage in the crystal led us to investigate whether or not that assembled form of the protein was well-populated in solution, a requisite property if the cage is to find utility in various applications in solution. The potentially reversible solution-dependence for

such an assembly is also a useful property for cargo delivery and other nanotechnology applications.

In order to analyze the solution behavior of the designed protein, we carried out dynamic light scattering (DLS) experiments at varying pH and salt concentrations to assess a dependence of cage formation on these two factors (Fig. 3). All samples were prepared from the same protein stock solution used for crystallization after exchanging buffers into a 10 mM CHES pH 9, 50 mM NaCl solution. Protein samples were subsequently diluted into DLS buffers to a final concentration of 3.5 mg ml⁻¹. By combining incremental pH values from 6 to 9 and three NaCl concentrations between 50 mM and 500 mM, we obtained 12 different buffer conditions for testing. Results from the DLS experiments (Fig. 3) indicate that self-assembly of pentamers into a dodecahedral species occurs at high yield in solution under specific conditions. The cage dominates (about 95% by mass) around pH 8 and 50 mM NaCl. Conversion appears to occur between the pentameric and dodecahedral states under other conditions, suggesting reversibility of the oligomerization. This is consistent with our earlier observation that the protein elutes at low concentration from a gel filtration column as a pentamer, but assembles into the dodecahedron under specific conditions evaluated by DLS and by crystallization.

Although this study involved a deliberate and dramatic redesign of a protein molecule, the discovery of its highly unusual assembly state was serendipitous. Despite the unexpected route by which this novel protein architecture was obtained, we expect that

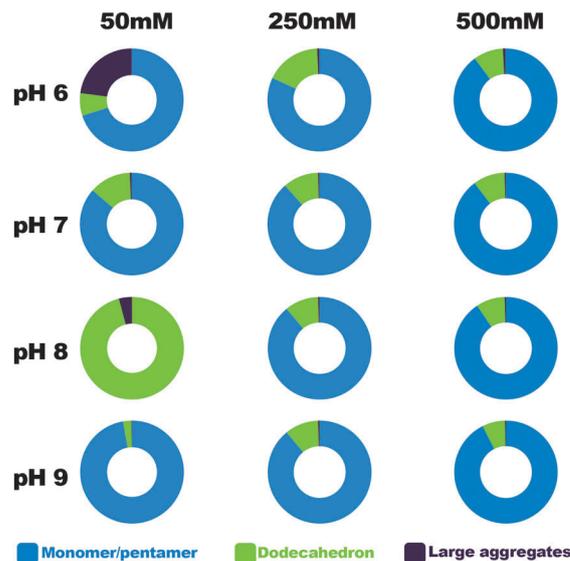


Fig. 3 Distribution of oligomeric forms of the designed protein according to dynamic light scattering experiments at varying pH and NaCl concentrations. Twelve different combinations of pH and salt concentrations are shown. For each condition, the mass percentages of the different oligomeric states are reported. An oligomeric state was defined based on the estimated mass of a detected particle. Mass peaks within the 5–30 kDa range were taken to be either monomers or pentamers, whereas mass peaks between 500 kDa and 700 kDa along with reported radii consistent with the dodecahedral cage structure were taken to be the dodecahedral cage, whose calculated mass and radius are 528 kDa and ~7 nm. Peaks of 40 MDa and above were considered as aggregates.

it could have applications in the field of bionanotechnology as platform for encapsulation and targeting of various cargos. We note that a particularly distinct property of the cage obtained is the overall tightness of the protein packing and the narrowness of the pores that run between the interior and exterior regions, which would be relevant for delivery applications. This very tight packing evidently derives from the natural shape properties of bacterial microcompartment shell proteins, which in their naturally assembled states form flat, tightly packed layers.

Future experiments aimed at modifying and advancing the utility of this protein cage include structure-based redesigns of the pentamer–pentamer interface to enhance the stability of the dodecahedron. Assessing the stability of these designs over a wider array of buffer conditions would provide finer control of the assembly process. Mutations to the interior and exterior surfaces of the cage could be explored for the purposes of encapsulation and targeting, respectively. Additionally, this protein assembly offers insight into the evolution of symmetric proteins in general, demonstrating that new symmetries can arise unexpectedly. In the present case, the changes in assembly state resulted from sequence permutation rather than from the more familiar scenario of mutations to interfacial regions of the protein subunit.

This work was supported by NSF grant CHE-1332907 and NIH grant AI081146. D. J. L. was supported by Ruth L. Kirschstein National Research Service Award T32GM007185. The authors thank Michael R. Sawaya, Duilio Cascio, and Dan McNamara for X-ray data collection at APS beamline 24-ID-C, and Joshua Laniado for helpful suggestions. X-ray core facilities at UCLA are supported by DOE Grant DE-FC03-02ER63421. The NECAT beamlines of the Advanced Photon Source are supported by NIH Grant RR-15301 (NCRR). Use of the Advanced Photon Source is supported by the DOE, Office of Basic Energy Sciences, under Contract DE-AC02 06CH11357.

Notes and references

- 1 J. E. Johnson and J. A. Speir, *J. Mol. Biol.*, 1997, **269**(5), 665–675.
- 2 D. S. Goodsell and A. J. Olsen, *Annu. Rev. Biophys. Biomol. Struct.*, 2000, **29**, 105–153.
- 3 E. D. Levy and S. Teichmann, *Prog. Mol. Biol. Transl. Sci.*, 2013, **117**, 25–51.
- 4 C. Chowdhury, S. Sinha, S. Chun, T. O. Yeates and T. A. Bobik, *Microbiol. Mol. Biol. Rev.*, 2014, **78**(30), 438–468.
- 5 T. O. Yeates, C. S. Crowley and S. Tanaka, *Annu. Rev. Biophys.*, 2010, **39**, 185–205.
- 6 T. O. Yeates, C. A. Kerfeld, S. Heinhorst, G. C. Cannon and J. M. Shively, *Nat. Rev. Microbiol.*, 2008, **6**(9), 681–691.
- 7 S. Howorka, *J. Mater. Chem.*, 2007, **17**, 2049–2053.
- 8 S. J. Tsai and T. O. Yeates, *Prog. Mol. Biol. Transl. Sci.*, 2011, **103**, 1–20.
- 9 Y. T. Lai, D. Cascio and T. O. Yeates, *Science*, 2012, **336**(6085), 1129.
- 10 N. P. King, W. Sheffler, M. R. Sawaya, B. S. Vollmar, J. P. Sumida, I. Andre, T. Gonen, T. O. Yeates and D. Baker, *Science*, 2012, **336**(6085), 1171–1174.
- 11 N. P. King, J. B. Bale, W. Sheffler, D. E. McNamara, S. Gonen, T. Gonen, T. O. Yeates and D. Baker, *Nature*, 2014, **510**, 103–108.
- 12 Y. T. Lai, E. Reading, G. L. Hura, K. L. Tsai, A. Lagonawsky, F. J. Asturias, J. A. Tainer, C. V. Robinson and T. O. Yeates, *Nat. Chem.*, 2014, **6**, 1065–1071.
- 13 H. Gradisar, S. Bozic, T. Doles, D. Vengust, I. Hafener-Bratkovic, A. Merteji, B. Webb, A. Sali, S. Klavzar and R. Jerala, *Nat. Chem. Biol.*, 2013, **9**(6), 362–366.
- 14 D. J. Huard, K. M. Kane and F. A. Tezcan, *Nat. Chem. Biol.*, 2013, **9**(3), 169–176.
- 15 B. Worsdorfer, K. J. Woycechowsky and D. Hilvert, *Science*, 2011, **331**(6017), 589–592.
- 16 J. E. Padilla, C. Colovos and T. O. Yeates, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**(5), 2217–2221.
- 17 Y. T. Lai, L. Jiang, W. Chen and T. O. Yeates, *Protein Eng., Des. Sel.*, 2015, **28**(11), 491–499.
- 18 Y. T. Lai, E. Reading, G. L. Hura, K. L. Tsai, A. Lagonawsky, F. J. Asturias, J. A. Tainer, C. V. Robinson and T. O. Yeates, *Nat. Chem.*, 2014, **6**(12), 1065–1071.
- 19 C. S. Crowley, D. Cascio, M. R. Sawaya, J. S. Kopstein, T. A. Bobik and T. O. Yeates, *J. Biol. Chem.*, 2010, **285**(48), 37838–37846.
- 20 S. Sinha, S. Cheng, Y. W. Sung, D. E. McNamara, M. R. Sawaya, T. O. Yeates and T. A. Bobik, *J. Mol. Biol.*, 2014, **426**(12), 2328–2345.
- 21 A. Pang, S. Frank, I. Brown, M. J. Warren and R. W. Pickersgill, *J. Biol. Chem.*, 2014, **289**(32), 22377–22384.
- 22 C. S. Crowley, M. R. Sawaya, T. A. Bobik and T. O. Yeates, *Structure*, 2008, **16**(9), 1324–1332.
- 23 S. Tanaka, M. R. Sawaya and T. O. Yeates, *Science*, 2010, **327**(5861), 81–84.
- 24 B. Webb and A. Sali, *Curr. Protoc. Bioinf.*, 2014, **47**, 6.6.1–5.6.32.
- 25 A. Colubri, A. K. Jha, M. Y. Shen, A. Sali, R. A. Berry, T. R. Sosnick and K. F. Freed, *J. Mol. Biol.*, 2006, **363**(4), 835–857.
- 26 P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral and R. W. Grosse-Kunstleve, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 213–221.
- 27 D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchinson and H. O. Smith, *Nat. Methods*, 2009, **6**, 343–345.
- 28 W. Kabsch, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 125–132.
- 29 A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J. Appl. Crystallogr.*, 2007, **40**, 658–674.
- 30 P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 486–501.