Intrinsically disordered CsoS2 acts as a general molecular thread for α-carboxysome shell assembly

Tao Ni[1,6#*], Qiuyao Jiang[2#], Pei Cing Ng[2], Juan Shen[1], Hao Dou[1], Yanan Zhu[1], Julika Radecke[3], Gregory F. Dykes[2], Fang Huang[2], Lu-Ning Liu[2,4*], Peijun Zhang[1,3,5*]

[1]Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK

[2]Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool, L69 7ZB, UK.

[3]Diamond Light Source, Harwell Science and Innovation Campus, Didcot OX11 0DE, UK

[4]College of Marine Life Sciences, and Frontiers Science Center for Deep Ocean Multispheres and Earth System, Ocean University of China, Qingdao, 266003, China.

[5]Chinese Academy of Medical Sciences Oxford Institute, University of Oxford, Oxford, OX3 7BN, UK

[6]School of Biomedical Sciences, LKS Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong SAR, China.

#Equal contribution

*Corresponding authors. Email: peijun.zhang@strubi.ox.ac.uk, Luning.Liu@liverpool.ac.uk, taoni@hku.hk

Abstract

Carboxysomes are a paradigm of self-assembling proteinaceous organelles found in nature, offering compartmentalisation of enzymes and pathways to enhance carbon fixation. In α-carboxysomes, the disordered linker protein CsoS2 plays an essential role in carboxysome assembly and Rubisco encapsulation. Its mechanism of action, however, is not fully understood. Here we synthetically engineered α-carboxysome shells using minimal shell components and determined cryoEM structures of these to decipher the principle of shell assembly and encapsulation. The structures reveal that the intrinsically disordered CsoS2 C-terminus is well-structured and acts as a universal "molecular thread" stitching through multiple shell protein interfaces. We further uncovered in CsoS2 a remarkable highly conserved repetitive key interaction motif, [IV]TG, which is critical to the shell assembly and architecture. Our study provides a general mechanism for the CsoS2-govern carboxysome shell assembly and cargo encapsulation and further advances synthetic engineering of carboxysomes for diverse biotechnological applications.

Introduction

Organelles confine specific biochemical pathways within the cell to enhance metabolic efficiency, alleviate metabolic crosstalk, and facilitate spatiotemporal regulation of sequestered pathways [1]. Apart from eukaryotes, in the past decades, advances in bioinformatics, biochemistry, imaging, and cell physiology have demonstrated that bacteria have also evolved subcellular organelles, including bacterial microcompartments (BMCs) which is composed entirely of proteins, to compartmentalize metabolism [2].

Carboxysomes are anabolic BMCs for autotrophic CO_2 fixation found in all cyanobacteria and many proteobacteria [3-7]. The carboxysome is composed of a polyhedral shell that encapsulates the key CO_2-fixation enzyme, ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), and carbonic anhydrase (CA), which dehydrates HCO_3^- to CO_2, the substrate for Rubisco carboxylation

[6,8-17]. The carboxysome shell acts as a selectively permeable barrier, allowing the influx of HCO_3^- and ribulose 1,5-bisphosphate (RuBP) while presumably precluding O_2 influx and CO_2 leakage [18-22]. The intriguing structural features of carboxysomes are fundamental for maximizing CO_2 assimilation and reducing the unproductive Rubisco oxygenation, thereby allowing carboxysomes to make substantial contributions to the global carbon fixation and primary production [13].

The α-carboxysome shell comprises predominantly CsoS1 hexamers that form the shell facets and CsoS4 pentamers that occupy the vertices of the polyhedral shell, both hexamers and pentamers containing multiple paralogous proteins [23-25]. An intrinsically disordered protein CsoS2 in high abundance [23] serves as a linker protein to bind both Rubisco and the shell through its N-terminal (CsoS2-N) and C-terminal domains (CsoS2-C), respectively [11,26-29]. It is presumed that α-carboxysome biogenesis adopts the 'Partial shell first' or 'Concomitant shell–core' assembly pathways [6,30,31], and CsoS2 is essential for α-carboxysome biogenesis and assembly of intact α-carboxysome shell [32]. However, how CsoS2 interacts with the shell and governs shell assembly remains enigmatic.

Synthetic BMC shells provide a means for investigating the assembly mechanisms and pairwise interactions that drive shell formation [32-38]. They also hold promises for generating new caging nanomaterials in new contexts, such as enzyme encapsulation, molecule scaffolding and delivery [39,40]. Here, we synthetically engineered α-carboxysome shells using minimal shell components derived from a chemoautotroph *Halothiobacillus* (*H.*) *neapolitanus*, a model system in the fundamental studies and synthetic engineering [11,20,23,41-48], and determined high-resolution cryoEM structures of the shells of variable constituents. Surprisingly, the structures show that the intrinsically disordered CsoS2-C makes well-defined multivalent contacts with shell proteins, functioning as a "molecular thread" to stitch the assembly interfaces and mediate shell assembly. Moreover, we uncovered a remarkable key repeating motif critical to the assembly and architecture of the shell. These findings provide unprecedented insight into the CsoS2-mediated assembly principles of α-carboxysome shell.

Results and Discussion

Assembly and structures of recombinant α-carboxysome shells

Expressing all shell components encoded by genes in the *cso* operon (Fig. 1a, coloured red, purple, and blue) results in formation of native-like α-carboxysome shells [32]. To investigate the molecular principles of α-carboxysome shell formation and the role of CsoS2 in α-carboxysome shell assembly, we took a reductionist approach and designed two shell constructs using minimal shell components, CsoS1A (a CsoS1 paralog) and CsoS4A (a CsoS4 paralog) for mini-shell 1, and CsoS1A and CsoS4A with additional CsoS2 (encoded by the natural *Halothiobacillus csoS2* gene that contain a ribosomal frameshifting region [27,49]) for mini-shell 2 (Fig. 1b, fig. S1a). Expression of either construct in *Escherichia coli* (*E. coli*) produces assembled shell architectures (figs. S2a and S2c), while expressing of mini-shell 3 (CsoS1A and CsoS2 without CsoS4A) does not yield any assembly (figs. S1a and S1b). These results indicate that CsoS1A and CsoS4A together are sufficient to form shell assemblies. However, dynamic light scattering (DLS) reveals that, while mini-shell 1 produces shells of ~23 nm in diameter, mini-shell 2 generates predominately larger shells of ~35 nm in diameter, in addition to the shells of ~23 nm in diameter (fig. S2b). CryoEM analysis further establishes an icosahedral architecture for the shell assemblies and reveals that mini-shell 1 assemblies contain mainly small shells (~21 nm, $T = 3$; see definition of T number is described in Methods), whereas mini-shell 2 products contain mostly large shells (~35 nm, $T = 9$) (Fig. 1C, figs. S3a-b). A medium sized shell (~23 nm, $T = 4$) is present in both constructs in low abundance. It is intriguing that the large $T = 9$ shells are only present in CsoS2-containing mini-shell 2.

We determined the structures of these *T* = 9, *T* = 4, and *T* = 3 shell assemblies at 1.86 Å, 3.54 Å and 2.79 Å resolution, respectively (Fig. 1d, Table S1, fig. S3c-f). The two *T* = 4 shell structures from both mini-shell constructs (Fig. 1c) appear largely identical, except that the β4-α2 loop of CsoS1A from mini-shell 1 is disordered (fig. S4a). All the shell proteins possess the same concave-out orientation, consistent with other shell assemblies [33,36,37]. Intriguingly, only *T* = 9 shells generated from mini-shell 2 contain extra densities not accounted for by CsoS1A and CsoS4A at the inner surface of the shell, which are absent in *T* = 4 and *T* = 3 shells. The high-quality map of the *T* = 9 shell (1.86 Å resolution) allows for accurate modelling of CsoS2-C that tightly associates with shell proteins, resulting in an atomic model of the *T* = 9 shell comprising 80 CsoS1A hexamers, 12 CsoS4A pentamers and 60 CsoS2-C (Fig. 1d).

Structural plasticity of shell proteins and protein-protein interactions

Three different sized shells with *T* = 3, *T* = 4 and *T* = 9 icosahedral symmetries are built of essentially the same hexameric CsoS1A and pentameric CsoS4A (Fig. 1d), which are highly conserved across α-cyanobacteria and many proteobacteria (fig. S5a). The RMSDs of the basic assembly units (a.k.a. capsomeres), CsoS1A hexamer and CsoS4A pentamer, in three icosahedral symmetries range 0.180-0.231 Å and 0.240-0.251 Å (calculated from the pairwise comparison of Cα), respectively. Superimposing the cryoEM structures of these capsomeres with their X-ray crystal structures (PDB: 2EWH, PDB: 2RCF) reveals subtle deviations of curvature within hexamer (the crystal structure closely resembles the least curved *T* = 9), and largely identical pentamer (fig. S5b), with the overall RMSD range of 0.321-0.597 and 0.204-0.417, respectively.

There are four inter-capsomere assembly interfaces in α-carboxysome shells (Fig. 2a, interface 1-4). Of those, interfaces 3 and 4 are unique to the *T* = 9 shell, whereas interfaces 1 and 2 are present in all three shells as previously reported [37]. The angles between capsomeres in interfaces 1 and 2 vary slightly among three shells, from 30° to 35° and 30° to 43°, respectively (Fig. 2b-c), leading to small curvature changes in the shell assembly, thus the size differences. In contrast, the angles between two hexamers in interfaces 3 and 4 are arranged in a planar fashion, markedly different from those in interfaces 1 and 2 (Fig. 2d), and also different from those measured in other BMC mini-shells [36]. Despite a large deviation of the tilt angle (from 0° to 43°), the main interactions at the interface between adjacent hexamers, which is mediated by the hydrogen bond network involving Lys29, Ala30, and Arg83, remain largely conserved (Fig. 2e). Collectively, the results suggest that the plasticity of inter-capsomere interfaces contributes to the curvature and thus, the structural polymorphism of carboxysome shells.

CsoS2 is a molecular thread linking multiple capsomeres

CsoS2 protein is present in α-cyanobacteria, α-proteobacteria, β-proteobacteria, and γ-proteobacteria (fig. S6). It functions as the linker protein connecting cargo enzymes to the shell and is a vital component in α-carboxysome assembly [23,26,32]. Genetic deletion of *csoS2* resulted in loss of carboxysomes in the *H. neapolitanus* cells and high CO_2-requiring phenotypes [27], and recombinant intact shells could not be formed in the absence of CsoS2 [32]. CsoS2 is a large polypeptide (~900 residues) composed of three regions: a N-terminal region (CsoS2-N), a middle region (CsoS2-M), and a C-terminal region (CsoS2-C) [27,28,32] (Fig. 3a). Repetitive short linear motifs have been identified in CsoS2, which vary in numbers among species; for example, the *H. neapolitanus* CsoS2 contains 4 N-repeats, 6 M-repeats, and 3 C-repeats [5] (Fig. 3a, fig. S1c). Structure prediction by AlphaFold2 revealed that CsoS2 represents a largely disordered protein, especially in the C-terminal region (fig. S7b), consistent with previous analyses [26,27]. Recent studies have shown that the CsoS2 N-terminal domain binds with Rubisco, playing roles in mediating Rubisco encapsulation [11,26], whereas the C-terminus of CsoS2 binds with the shell and could serve as an encapsulation anchor for cargo recruitment [27,32]. However, how CsoS2 anchors to the shell and whether it plays a role in governing shell assembly have remained enigmatic.

The mini-shell 2 vector comprises the native *csoS2* gene *Halothiobacillus*, which contains a ribosomal frameshifting region thereby resulting in the production of two CsoS2 isoforms, the full-length CsoS2B and the C-terminus-truncated CsoS2A. Our immunoblot analysis revealed that both CsoS2A and CsoS2B isoforms were expressed in the *E. coli* mini-shell construct (fig. S2b), and the ratio of CsoS2A and CsoS2B is compared to that found in the native α-carboxysome from *Halothiobacillus* [23]. By contrast, only CsoS2B was incorporated into the isolated mini-shells, confirming the importance of the CsoS2 C-terminus in attaching to the shell.

The cryoEM structure of the *T* = 9 shell at 1.86 Å resolution enables the first characterization of CsoS2 and its interactions with the shell at atomic details (Fig. 3b, fig. S7a, S7b). It is clearly well-resolved for the most of C-terminus in the cryoEM map (fig. S7c), allowing unambiguous assignment of amino acid residues of three fragments of CsoS2-C: F1, Arg712−Arg731; F2, Leu773−Gly823; F3, Glu829−Gly869 (Fig. 3b, fig. S7c). The rest of the CsoS2 regions were not resolved. The cryoEM structure revealed that the F1, F2, F3 fragments adopt completely different conformations from the AlphaFold2-predicted structures (fig. S7d). They form extensive hydrogen bonds and salt bridges with both CsoS1A and CsoS4A at three distinct binding sites (Fig. 3c-e), with the surface contact area of 767.1 Å^2, 3333.0 Å^2, and 1851.7 Å^2, respectively. Such extensive, multivalent interactions strongly anchor CsoS2 to the shell inner surface (Fig. 3b).

Interestingly, all three CsoS2-C fragments, F1 to F3, bind specifically to the tri-capsomere interfaces (Fig. 3b). While CsoS2-F1 and F2 interact with three CsoS1A hexamers (Fig. 3b-d), CsoS2-F3 acts on the interface between one CsoS4A pentamer and two CsoS1A hexamers, with the extreme C-terminus buried inside the cavity formed by the interface (Fig. 3e). It is explicit that the CsoS2 C-terminus is completely encapsulated inside the shell (Figs. 3b-e, fig. S7a), rather than being exposed to the exterior as previously proposed [27]. The three CsoS2-C F1-F3 fragments associate tightly with shell proteins simultaneously and essentially “stitches” the capsomeres together at the inner surface of the $T = 9$ shell. Therefore, CsoS2-C functions as a unique "molecular thread" reinforcing the connections between capsomeres thus potentially stabilizing the assembled α-carboxysome shell.

At the first glance, the three CsoS2 binding sites appear random (Fig. 3b-e). However, a close inspection of the interfaces uncovers a common interaction motif among these interactions: alignment of all the CsoS1A-CsoS2 interfaces identifies a repetitive Ile(Val)-Thr-Gly ([IV]TG) motif [27,28] that interacts with the β-strand of highly conserved CsoS1A through His79 and hydrogen bonds in the main chain (Fig. 3f, fig. S8). More intriguingly, the [IV]TG motif repeats itself about every 8 residues in all three CsoS2-C fragments (fig. S8b). Sequence alignment further reveals that the repeating [IV]TG motif is highly conserved in the CsoS2-family proteins (Fig. 3g, red asterisks), suggesting that the "molecular threading" of CsoS2-C is likely a widespread feature mediating α-carboxysome shell assembly. In native α-carboxysomes, if CsoS2 must bind to the shell with F3 located to pentamer-hexamer interface, then the maximum CsoS2 occupancy would be 60 (12 pentamers). However, our recent study revealed that the native α-carboxysome from *Halothiobacillus* contains 863 hexamers and 192 CsoS2B as the full-length CsoS2 [23], which indicates that CsoS2B does not have to restrict to all pentamer-hexamer interfaces. The local structures of CsoS2 on the authentic carboxysome shell could be more dynamic. Our structure therefore represents one of the local snapshots recapitulating the binding modes in native environments.

CsoS2-C dictates α-carboxysome shell assembly and architecture

Since all the three CsoS2-C F1-F3 fragments form contacts with the shell proteins and contain multiple [IV]TG motifs (Fig. 3, fig. S8b-d), we further investigated the importance of individual CsoS2-C fragments and [IV]TG motifs to $T = 9$ shell formation. To this end, we designed four mini-shell constructs, mini-shell 4-6, with variations in the CsoS2-C (Fig. S1a-b). Mini-shell 4-6 contain CsoS2-C fragment F1-F3 (S2-C1), F2-F3 (S2-C2) and F3 only (S2-C3), respectively (Fig. 4a, fig. S1a-b). In addition, a 7th mini-shell construct was generated where all CsoS2-C [IV]TG motifs were replaced by AAA (S2-Cm) in the background of mini-shell 4 (Fig. 4a). We measured ratio of assembled shell proteins to unassembled shell proteins by comparing the abundance of shell proteins in the soluble and the assembled forms. Interestingly, in the absence of CsoS2, mini-shell composed of CsoS4A and CsoS1A (mini-shell 1) assembles more efficiently than the mini-shell containing CsoS2-F3 fragment (mini-shell 6), but similar to the [IV]TG motif mutant (Cm). One plausible explanation is that the CsoS2-F3 may interfere shell assembly. It is also possible that CsoS4A-CsoS1A and S2-Cm assemble efficiently into different shell morphologies, such as $T = 4$ and $T = 3$ shells observed in mini-shell 1.

To systematically dissect the effects of CsoS2-C fragments on the morphology of shells, we further carried out cryoEM structural characterization of these four mini-shell variants containing S2-C1, C2, C3 and Cm. While CsoS2 (mini-shell 2) promotes assembly of large $T = 9$ shell (Fig. 1c), the mini-shell containing S2-C1 fragment, interestingly, leads to formation of shell with various sizes and symmetries: $T = 9$, $T = 7$, $T = 7$ $Q = 6$, $T = 4$, $T = 4$ $Q = 6$, and $T = 3$ (Fig. 4c-d, fig. S10, Table S2). CsoS2-C fragments were unambiguously identified in all the larger shells albeit with a relatively lower occupancy (fig. S11) compared with the full-length CsoS2 construct (fig. S10a-d), but not in the $T = 4$ and $T = 3$ shells. In contrast, the S2-C2 construct produces exclusively $T = 4$ shells (Fig. 4d), with little residual densities corresponding to the CsoS2 [IV]TG motifs (fig. S10e-f). The S2-C3 construct produced only $T = 4$ and $T = 3$ shells, both having a subpopulation missing pentamers (Fig. 4c-d). Notably, mutation of [IV]TG results in formation of only $T = 4$ and $T = 3$ shells (Fig. 4d), with the population ratio similar to the mini-shell 1 construct lacking CsoS2. These results illustrate the essential role of CsoS2-C and [IV]TG motif in controlling α-carboxysome shell assembly and overall architecture.

Carboxysomes are a paradigm of self-assembling proteinaceous organelles found in nature, offering compartmentalisation of enzymes and pathways to enhance carbon fixation. Given their significance in the global carbon cycle, carboxysomes are gaining increasing attention from fundamental studies and synthetic engineering, with the intent of generating metabolic factories for sustainably turbocharging carbon fixation and primary production. We devised a minimal system encompassing shell proteins and the linker CsoS2 to decipher the molecular principles of shell assembly and encapsulation. The distinctive multivalent interactions between CsoS2 C-terminus and shell proteins and between CsoS2 N-terminus and Rubisco are vital for governing the architectures of shell assemblies and encapsulation of Rubisco, respectively (Fig. 4e), while the actual role of CsoS2 middle region remains to be determined. Since the middle region also contains repetitive [IV]TG motif, it is likely that the middle region may also contribute to its interaction with shell components. CsoS2 may prefer to bind to these curved capsomer interfaces but would retain a certain affinity to flat interfaces, which is the case in the native carboxysomes facets. Advanced knowledge of carboxysome assembly will offer new strategies for design and engineering of carboxysome shell-based nanobioreactors and new cages in diverse biotechnological applications, such as enhancement of biocatalysis, food and energy production, molecule delivery, and therapeutics.

Acknowledgements

We acknowledge Diamond for access and support of the cryo-EM facilities at the U.K. national eBIC (proposals NT29812 and BI28713), funded by the Wellcome Trust, MRC, and BBSRC. We are grateful to Yuewen Sheng and Yun Song for their technical support in cryo-EM data collection at eBIC (Electron Bio-imaging Centre), Diamond Light Source, and Prof. Ian Prior and Mrs. Alison Beckett at the Liverpool Biomedical Electron Microscopy Unit for technical support of electron microscopy, and the Materials Innovation Factory (MIF) for provision of analytical equipment. The computational aspects of this research were supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z and the NIHR Oxford BRC.

Funding

This research was funded by the National Key R&D Program of China (2021YFA0909600), the UK Wellcome Trust Investigator Award (206422/Z/17/Z), the National Natural Science Foundation of China (32070109), the UK Biotechnology and Biological Sciences Research Council grants (BB/S003339/1, BB/V009729/1 and BB/R003890/1), the ERC AdG grant (101021133), the Royal Society (URF\R\180030, RGF\EA\181061 and RGF\EA\180233), and the Leverhulme Trust (RPG-2021-286). The research was supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z with additional support from the NIHR Oxford BRC. Q.J. was supported by a Liverpool-Chinese Scholarship Council PhD Studentship. P.C.N. was supported by a BBSRC DTP PhD Studentship.

Author contributions

T.N., Q.J., L-N.L., and P.Z. conceived the project and designed the experiments; Q.J., P.C.N. G.F.D., and F.H. performed shell construction and characterization, as well as negative-staining EM; T.N. prepared cryoEM samples; T.N. J.R. and Y.Z. collected cryoEM data; T.N. processed cryoEM

data and refined the structures with the assistance of H.D. and J.S.; T.N., L-N.L. and P.Z. analysed the structures. T.N., L-N.L., and P.Z. wrote the manuscript with support from all authors.

Declaration of interests

The authors declare no competing interests.

Data and material availability

The cryo-EM density maps and corresponding atomic models have been deposited in the EMDB and PDB, respectively. The accession codes are listed as follows: (1) CsoS4A and CsoS1A (mini-shell-1 construct): EMD-15798 and PDB 8B0Y for *T* = 4 shell, EMD-15792 for *T* = 3 shell; (2) Full-length CsoS2 with CsoS4A and CsoS1A (mini-shell-2 construct): EMD-15801 and PDB 8B12 for *T* = 9 shell, EMD-15799 and PDB 8B11 for *T* = 4 shell; (3) Truncated CsoS2 with F1-F3 fragments (mini-shell-4 construct): EMD-15722 (*T* = 9), EMD-15720 (*T* = 7), EMD-15595 (*T* = 7, *Q* = 6), EMD-15723 (*T* = 4, *Q* = 6 class 1), EMD-15724 (*T* = 4, *Q* = 6 class 2), EMD-15719 (*T* = 4); (4) Truncated CosS2 with F2-F3 fragments (mini-shell-5 construct): EMD-15611 for *T* = 4 shell; (5) Truncated CosS2 with F3 fragment (mini-shell-6 construct): EMD-15758 (*T* = 4 with pentamer), EMD-15759 (*T* = 4 without pentamer), EMD-15760 (*T* = 4 with pentamer) and EMD-15761 (*T* = 4 without pentamer); (6) CsoS2-C with [IV]TG mutant: EMD-15834 (*T* = 4) and EMD-15762 (*T* = 3).

Materials and Methods

Generation of constructs

Primers (Table S3) and the mutant *csoS2* sequence were ordered from Integrated DNA Technologies. The pHnCBS1D plasmid (Addgene, UK) was used as the template for amplification of native *csoS2*, *csoS4A*, and *csoS1A* genes from the *Halothiobacillus neapolitanus* genome and cloned into NcoI/XhoI and EcoRI linearized pBAD by Gibson assembly (New England Biolabs, UK) [50] to produce the synthetic mini-shell operons used in this study (Fig. 1b; fig. S1). Expression constructs were sequence verified by Eurofins Genomics and transformed into *E. coli* TOP10 for plasmid storage and expression.

Isolation of α-carboxysome mini-shells

E. coli TOP10 cells containing the desired mini-shell construct were cultured in Lysogeny Broth (LB) supplemented with ampicillin (100 µgmL^{-1}) at 37 ℃ to an OD at 600 nm (OD$_{600}$) between 0.6 and 0.8. Mini-shell expression was then induced with 1 mM arabinose (Melford) at 22 ℃ for 20 hr. Cells were harvested at 5,000 x *g* for 10 min and pellets resuspended in TEMB buffer (10 mM Tris-HCl pH = 8, 1 mM EDTA, 10 mM $MgCl_2$, 20 mM $NaHCO_3$) supplemented with 10% (v/v) CelLytic™ B cell Lysis Reagent (Sigma-Aldrich) and 0.1% Protease Inhibitor Cocktail (Sigma-Aldrich). Cell lysis was performed by sonication (MSE 8-75 MK2 sonicator, 6 cycles of 30 s ON/OFF) and cell debris removed by centrifugation at 27,000 x *g*, 30 min, 4 ℃. The supernatant was subsequently loaded on top of 5 mL 30% (w/v) sucrose and mini-shells pelleted by ultracentrifugation at 250,000 x *g*, 16 hr, 4 ℃. By using a soft brush, the pelleted mini-shells were either resuspended in 1 mL TEMB for further isolation, or in TEMB to the same volume as the supernatant to assess assembly ratio of the mini-shells.

For further purification of the mini-shells, resuspended pellets were centrifuged at 21100 x *g*, 2 min, 4 ℃. The supernatant was loaded onto a 10 mL step sucrose gradient consisting of 10%, 20%, 30%, 40% and 50% (w/v) sucrose and centrifuged at 70,000 x *g*, 16 hr, 4 □. Fractions were collected and analysed by SDS-PAGE. The fractions enriched with mini-shells were pooled and applied onto a HiTrap Q FF anion exchange chromatography column (Cytiva Life Sciences) equilibrated with buffer A (TEMB plus 50 mM NaCl). Mini-shells were eluted with a 0-40% linear gradient of buffer B (TEMB plus 1 M NaCl) and were found to be typically eluted at 30-35% buffer B. Purified samples were stored at 4 □ for further analysis.

SDS-PAGE and immunoblot analysis

Protein samples mixed with SDS-PAGE loading buffer were heated at 99 □ for 10 min and electrophoresed on 15 % SDS-PAGE gels. Gels were stained with Coomassie Brilliant Blue for SDS-PAGE analysis. For immunoblot analysis, gel electrophoresis was performed at 100 V 30 min, then 150 V 40 min, and blotted onto polyvinylidene fluoride membrane (Cytiva Life Sciences). Proteins were probed with the primary antibody anti-CsoS1A/B/C (Agrisera, Cat No. AS14 2760, dilution 1: 5,000), anti-CsoS2-N (1:10,000 dilution; synthesized by GenScript, NJ, USA with the RGTRAVPPKPQSQG peptide), and anti-rabbit IgG secondary antibody (Agrisera, Cat No. AS09 602, dilution 1:10,000). Protein signals were analyzed using a Bio-Rad chemiluminescence kit (Bio-Rad, UK) and images were captured using ImageQuant LAS 4000 (GE Healthcare, USA).

Dynamic Light Scattering (DLS) analysis

To measure the size distribution of isolated mini-shells and obtain an average size of their diameters, 1 mL mini-shell fractions (5-10 mg mL^{-1}) were analyzed by DLS using a ZetaSizer Nano ZS instrument (Malvern Panalytical Ltd, UK). All DLS measurements were performed in triplicate and Fig. S2B was plotted with GraphPad Prism 9 (GraphPad Software, San Diego, California, USA).

Negative-staining transmission electron microscopy

5 µL of isolated mini-shells were mounted on carbon grids (Carbon Films on 300 Mesh Grids Copper, Agar Scientific, UK) for 40 s, stained and washed with 60 µL of 2% uranyl acetate (Sigma-Aldrich), and excess stain wicked away with filter paper. Grids were left to air dry for at least 1 min. Images were recorded by the FEI Tecnai G2 Spirit Bio TWIN transmission electron microscope equipped with a Gatan Rio 16 camera. Images were visualized and analyzed by Fiji software [51]. Graphs were created using GraphPad Prism 9 (GraphPad Software, San Diego, California USA, www.graphpad.com).

CryoEM data collection

The cryoEM sample grids were prepared using Vitrobot. The grids (Quantifoil™ R 2/1 on 300 copper mesh) were glow-discharged using a plasma cleaner (Harrick Plasma) with the plasma level set to High position for 45 seconds using ambient air. 3 µL of mini-shell samples were applied to the grids and blotted with filter paper for 3 seconds before plunge freezing with liquid nitrogen-cooled ethane. The temperature was set to 20 □ and humidity at 100% during plunge freezing. The micrographs were taken using Thermo Scientific Titan Krios G3 microscope equipped with a Gatan K3 direct electron camera and BioQuantum energy filter (slit width 20 eV) or Falcon 4 with Selectris X energy filter (slit width 10 eV). Details of data collection parameters are listed in Table S1 and S2.

CryoEM data processing

The data processing was performed using RELION (v3.1) [52] and cryoSPARC (v3) [53]. The micrograph movies were gain normalized and motion corrected with MotionCor2 (v1.4.0) [54]. Contrast transfer function (CTF) was estimated using CtfFind4 (v4.1.14). In the first two dataset (Table S1), three different sized shells (T = 9, T = 4 and T = 3) were observed on the raw micrographs and processed independently. The triangulation number (T) is the number of quasisimilar subunits per icosahedral asymmetric unit with possible value of T, and Q is an elongation number to describe the prolate/oblate icosahedra.

For the $T = 9$ particles, a subset of particles was picked manually in RELION to generate initial 2D class averages for auto-picking. Two rounds of 2D classification were performed, resulting a final dataset with 143,769 particles. 2D classification did not reveal ordered densities within the shell. An *ab initio* model was generated with I1 symmetry in RELION. 3D auto refine was carried out with the initial model reconstructed with I1 symmetry in RELION, which resulted in a density map with a mixed handedness. The resulted refined particles dataset was 3D classified into 10 classes, skipping alignment, which revealed two major classes, with opposite handedness. These two classes of particles were refined separately with per-particle CTF refinement and polishing. To combine the two classes, particles with opposite handedness were inverted by changing the refined Euler angle in RELION star file (phi and tilt). The two half maps were reconstructed using relion_reconstruct, with CTF and Eward sphere correction. The final combined density maps were masked, and B-factor sharpened (-47.66, automatically determined by RELION) with relion_postprocess, which resulted in a final map at 1.86 Å resolution (fig. S3 and Table S1). Symmetry expansion and focused classification was carried out, which did not reveal alternative CsoS2 conformations binding to the shell proteins.

The small shells ($T = 4$ and $T = 3$) were processed in a similar way to $T = 9$ shell, except in particle picking step. A small number of small shell particles were manually picked in EMAN2.3 to train neural network, which was subsequently used to pick against the whole dataset. The coordinates of particles (EMAN2 box files) were imported into Relion for further processing in the same way as large shell. Similarly, a small portion of particles were found in opposite handedness after 3d refinement and classification, which was then corrected by updating the Euler angles in the Relion star file as above. The final map after per-particle CTF refinement and polishing is at 2.5 Å resolution.

The rest of datasets (S2-C1/4A-1A, S2-C2/4A-1A, S2-C3/4A-1A and S2-Cm/4A-1A) were processed in cryoSPARC (v3) [53]. Particles were picked by a combination of blob picking and template-based picking to ensure all the mini-shells on the micrographs are picked; the duplicated particles were removed after 2D classification. For S2-C1/4A-1A dataset, several new symmetries were identified, in addition to the $T = 9$, $T = 4$ and $T = 3$ shells. Initial 2D classification revealed particles with diameter smaller than $T = 9$ shell but larger than $T = 4$ shell. These particles were divided into 4 different classes ($T = 7$, $T = 7$ $Q = 6$, $T = 4$ $Q = 6$, $T = 4$ $Q = 6$ class 2) after several rounds of 2D and 3D classification with C1 symmetry. Further 3D refinement with icosahedral symmetry ($T = 7$), D3 (T = T $Q = 6$) and D5 ($T = 4$ $Q = 6$ and $T = 4$ $Q = 6$ class 2) symmetry were performed to obtain the final density map, respectively. The other three datasets were processed similarly with icosahedral symmetry applied during 3D classification and refinement. Data collection and classification results are summarized in Table S2.

Model building and refinement

Initial models were obtained from crystal structures of hexamer (PDB 2EWH) and pentamer (PDB 2RCF). For the $T = 9$ shell, the CsoS2 was traced manually into the density map in Coot (v0.8.9.2). At 1.86 Å resolution, the side chains of CsoS2 can be unambiguously placed (fig. S7c). One asymmetric unit of the icosahedral shell with additional surrounding subunits were further refined in Phenix.refine [55]. Water molecules were placed into density manually. The final icosahedral models were reconstructed in Chimera with symmetry command with I1 symmetry. Details of model geometry statistics are listed in Supplementary Table S1. Model alignment and comparison were performed in Chimera (v1.15) [56]. Figures were rendered in Chimera (v1.15) [56], ChimeraX (v1.5) [57] and Pymol (PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

Bioinformatics and structural prediction

Protein sequences assigned CsoS2, CsoS1A, and CsoS4A were blasted against the NCBI database of non-redundant protein sequences. A total of 395 sequences for CsoS2, 990 sequences for CsoS1A, and 970 sequences for CsoS4A were selected and aligned using Clustal Omega [58]. The resulting multiple sequence alignment files were used to determine conservation, visualised with WebLogo 3 [59]. Phylogenetic tree was built using IQ-TREE web server [60] and visualized using iTOL 6.4.3 [61]. The structures of the CsoS2 N-terminus (1-260), middle region (261-604), and C-terminus (605-869) were predicted by AlphaFold2 [62], accessed via ColabFold [63].

References

1. Kerfeld CA, Heinhorst S, Cannon GC. Bacterial microcompartments. *Annu Rev Microbiol* **64**, 391-408 (2010).
2. Greening C, Lithgow T. Formation and function of bacterial organelles. *Nat Reviews Microbiology* **18**, 677-689 (2020).
3. Hennacy JH, Jonikas MC. Prospects for Engineering Biophysical CO_2 Concentrating Mechanisms into Land Plants to Enhance Yields. *Annu Rev Plant Biol* **71**, 461-485 (2020).
4. Liu LN, Yang M, Sun Y, Yang J. Protein stoichiometry, structural plasticity and regulation of bacterial microcompartments. *Curr Opin Microbiol* **63**, 133-141 (2021).
5. Liu LN. Bacterial metabolosomes: new insights into their structure and bioengineering. *Microbial Biotechnology* **14**, 88-93 (2021).
6. Liu LN. Advances in the bacterial organelles for CO_2 fixation. *Trends in Microbiology* **30**, 567-580 (2022).
7. Hill NC, Tay JW, Altus S, Bortz DM, Cameron JC. Life cycle of a cyanobacterial carboxysome. *Sci Adv* **6**, eaba1269 (2020).
8. Price GD, Badger MR, Woodger FJ, Long BM. Advances in understanding the cyanobacterial CO_2-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. *J Exp Bot* **59**, 1441-1461 (2008).
9. Long BM, Forster B, Pulsford SB, Price GD, Badger MR. Rubisco proton production can drive the elevation of CO_2 within condensates and carboxysomes. *Proc Natl Acad Sci U S A* **118**, e2014406118 (2021).
10. Faulkner M, *et al.* Direct characterization of the native structure and mechanics of cyanobacterial carboxysomes. *Nanoscale* **9**, 10662–10673 (2017).
11. Ni T, *et al.* Structure and assembly of cargo Rubisco in two native alpha-carboxysomes. *Nat Commun* **13**, 4299 (2022).
12. Sun Y, Wollman AJM, Huang F, Leake MC, Liu LN. Single-organelle quantification reveals the stoichiometric and structural variability of carboxysomes dependent on the environment. *Plant Cell* **31**, 1648-1664 (2019).
13. Rae BD, Long BM, Badger MR, Price GD. Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO_2 fixation in cyanobacteria and some proteobacteria. *Microbiol Mol Biol Rev* **77**, 357-379 (2013).
14. Evans SL, *et al.* Single-particle cryo-EM analysis of the shell architecture and internal organization of an intact α-carboxysome. *Structure* **31**, 1-12 (2023).

15. Iancu CV, Morris DM, Dou Z, Heinhorst S, Cannon GC, Jensen GJ. Organization, structure, and assembly of alpha-carboxysomes determined by electron cryotomography of intact cells. *J Mol Biol* **396**, 105-117 (2010).

16. Cameron JC, Wilson SC, Bernstein SL, Kerfeld CA. Biogenesis of a bacterial organelle: the carboxysome assembly pathway. *Cell* **155**, 1131-1140 (2013).

17. Huang F, *et al.* Rubisco accumulation factor 1 (Raf1) plays essential roles in mediating Rubisco assembly and carboxysome biogenesis. *Proc Natl Acad Sci USA* **117**, 17418-17428 (2020).

18. Mahinthichaichan P, Morris DM, Wang Y, Jensen GJ, Tajkhorshid E. Selective permeability of carboxysome shell pores to anionic molecules. *J Phys Chem B* **122**, 9110-9118 (2018).

19. Faulkner M, *et al.* Molecular simulations unravel the molecular principles that mediate selective permeability of carboxysome shell protein. *Scientific Reports* **10**, 17501 (2020).

20. Huang J, *et al.* Probing the internal pH and permeability of a carboxysome shell. *Biomacromolecules* **23**, 4339-4348 (2022).

21. Kerfeld CA, *et al.* Protein structures forming the shell of primitive bacterial organelles. *Science* **309**, 936-938 (2005).

22. Cai F, Menon BB, Cannon GC, Curry KJ, Shively JM, Heinhorst S. The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a CO_2 leakage barrier. *PLoS One* **4**, e7521 (2009).

23. Sun Y, *et al.* Decoding the absolute stoichiometric composition and structural plasticity of α-carboxysomes. *mBio* **13**, e0362921 (2022).

24. Tsai Y, *et al.* Structural analysis of CsoS1A and the protein shell of the *Halothiobacillus neapolitanus* carboxysome. *PLoS Biol* **5**, e144 (2007).

25. Tanaka S, *et al.* Atomic-level models of the bacterial carboxysome shell. *Science* **319**, 1083-1086 (2008).

26. Oltrogge LM, Chaijarasphong T, Chen AW, Bolin ER, Marqusee S, Savage DF. Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation. *Nat Struct Mol Biol* **27**, 281-287 (2020).

27. Cai F, *et al.* Advances in understanding carboxysome assembly in *Prochlorococcus* and *Synechococcus* implicate CsoS2 as a critical component. *Life (Basel)* **5**, 1141-1171 (2015).

28. Chaijarasphong T, *et al.* Programmed Ribosomal Frameshifting Mediates Expression of the alpha-Carboxysome. *J Mol Biol* **428**, 153-164 (2016).

29. Baker SH, Lorbach SC, Rodriguez-Buey M, Williams DS, Aldrich HC, Shively JM. The correlation of the gene csoS2 of the carboxysome operon with two polypeptides of the carboxysome in thiobacillus neapolitanus. *Arch Microbiol* **172**, 233-239 (1999).

30. Kerfeld CA, Aussignargues C, Zarzycki J, Cai F, Sutter M. Bacterial microcompartments. *Nat Rev Microbiol* **16**, 277-290 (2018).

31. Dai W, *et al.* Visualizing individual RuBisCO and its assembly into carboxysomes in marine cyanobacteria by cryo-electron tomography. *J Mol Biol* **430**, 4156-4167 (2018).

32. Li T, *et al.* Reprogramming bacterial protein organelles as a nanoreactor for hydrogen production. *Nature Commun* **11**, 5448 (2020).

33. Kalnins G, Cesle EE, Jansons J, Liepins J, Filimonenko A, Tars K. Encapsulation mechanisms and structural studies of GRM2 bacterial microcompartment particles. *Nat Commun* **11**, 388 (2020).

34. Sutter M, Laughlin TG, Sloan NB, Serwas D, Davies KM, Kerfeld CA. Structure of a Synthetic beta-Carboxysome Shell. *Plant Physiol* **181**, 1050-1058 (2019).

35. Greber BJ, Sutter M, Kerfeld CA. The Plasticity of Molecular Interactions Governs Bacterial Microcompartment Shell Assembly. *Structure* **27**, 749-763 e744 (2019).

36. Sutter M, Greber B, Aussignargues C, Kerfeld CA. Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science* **356**, 1293-1297 (2017).

37. Tan YQ, *et al.* Structure of a minimal α-carboxysome-derived shell and its utility in enzyme stabilization. *Biomacromolecules* **22**, 4095-4109 (2021).

38. Parsons JB, *et al.* Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. *Mol Cell* **38**, 305-315 (2010).

39. Plegaria JS, Kerfeld CA. Engineering nanoreactors using bacterial microcompartment architectures. *Curr Opin Biotechnol* **51**, 1-7 (2018).

40. Jiang Q, *et al.* Synthetic engineering of a new biocatalyst encapsulating [NiFe]-hydrogenases for enhanced hydrogen production. *J Mater Chem B* **11**, 2684-2692 (2023).

41. Chen T, *et al.* Incorporation of functional Rubisco activases into engineered carboxysomes to enhance carbon fixation. *ACS Synth Biol* **11**, 154-161 (2022).

42. Bonacci W, *et al.* Modularity of a carbon-fixing protein organelle. *Proc Natl Acad Sci U S A* **109**, 478-483 (2012).

43. Chen T, *et al.* Producing fast and active Rubisco in tobacco to enhance photosynthesis. *Plant Cell* **35**, 795-807 (2023).

44. Chen T, *et al.* Engineering α-carboxysomes into plant chloroplasts to support autotrophic photosynthesis. *Nature Communications* **14**, 2118 (2023).

45. Baumgart M, Huber I, Abdollahzadeh I, Gensch T, Frunzke J. Heterologous expression of the *Halothiobacillus neapolitanus* carboxysomal gene cluster in *Corynebacterium glutamicum*. *J Biotechnol* **258**, 126-135 (2017).

46. Liu Y, *et al.* Deciphering molecular details in the assembly of alpha-type carboxysome. *Sci Rep* **8**, 15062 (2018).

47. Dou Z, Heinhorst S, Williams EB, Murin CD, Shively JM, Cannon GC. CO_2 fixation kinetics of *Halothiobacillus neapolitanus* mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO_2. *J Biol Chem* **283**, 10377-10384 (2008).

48. MacCready JS, Tran L, Basalla JL, Hakim P, Vecchiarelli AG. The McdAB system positions α-carboxysomes in proteobacteria. *Mol Microbiol* **16**, 277-297 (2021).

49. Chaijarasphong T, *et al.* Programmed ribosomal frameshifting mediates expression of the α-carboxysome. *J Mol Biol* **428**, 153-164 (2016).

50. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-345 (2009).

51. Schindelin J, *et al.* Fiji: an open-source platform for biological-image analysis. *Nature methods* **9**, 676 (2012).

52. Scheres SH. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-530 (2012).

53. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296 (2017).

54. Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, Agard DA. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017).

55. Liebschner D, *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877 (2019).

56. Pettersen EF, *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).

57. Pettersen EF, *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82 (2021).

58. Madeira F, *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47**, W636-W641 (2019).

59. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190 (2004).

60. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res* **44**, W232-235 (2016).

61. Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**, 1283-1287 (2006).

62. Senior AW, *et al.* Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706-710 (2020).

63. Mirdita M, Schutze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679-682 (2022).

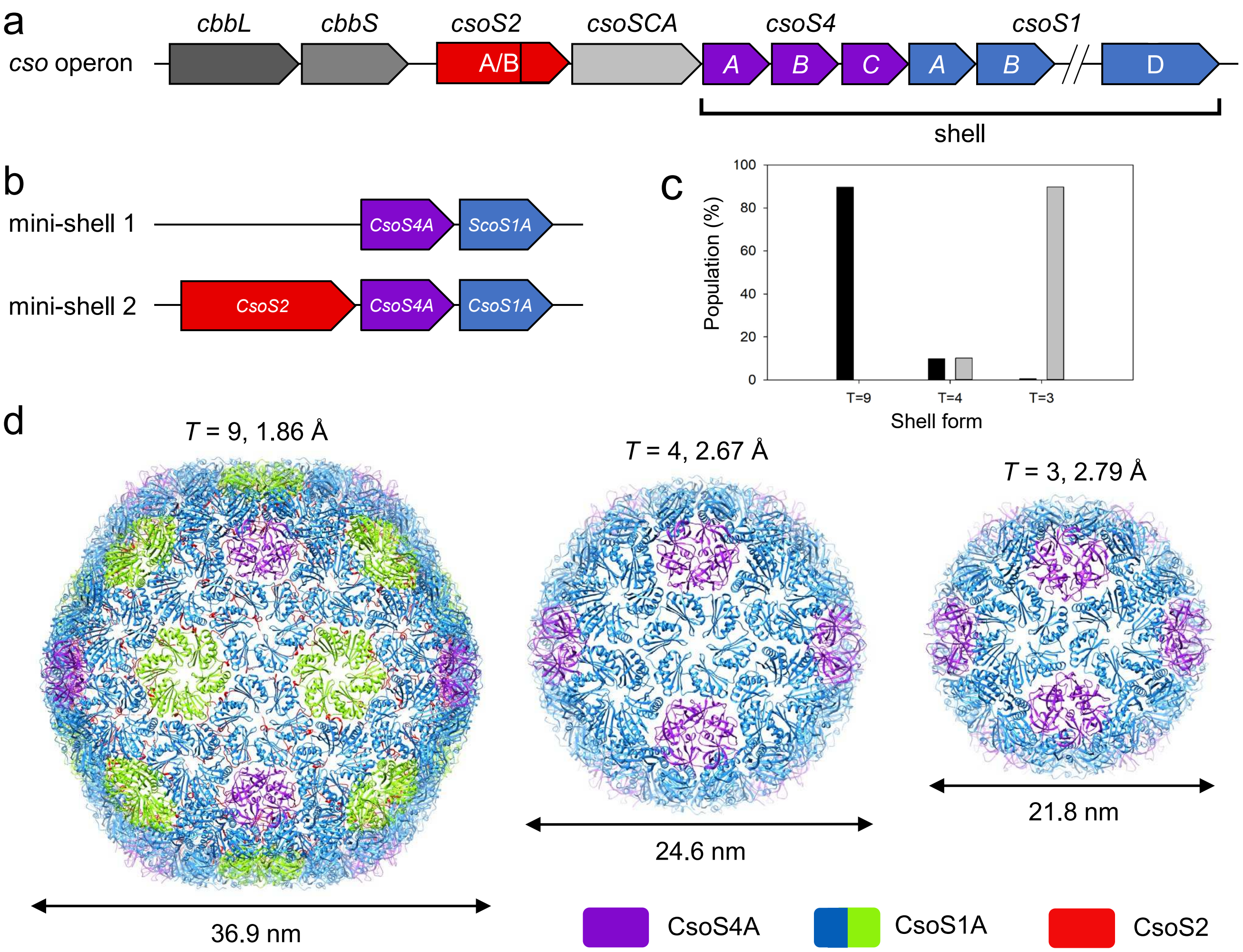

Figure 1 | Design and overall cryoEM structure of α-carboxysome mini-shells. (a) Schematics of α-carboxysome *cso* operon. (b) Schematics of two mini-shell constructs used to assemble shells. (c) Distribution of shell forms assembled with mini-shell construct 1 (gray, total 177,237 mini-shells) and mini-shell construct 2 (black, total 137,690 mini-shells). (d) CryoEM structures of three different shell forms with icosahedral symmetry of *T* = 9, *T* = 4 and *T* = 3, at the resolution of 1.86 Å, 2.67 Å, and 2.79 Å, respectively. The diameter of shells are indicated. Shell components are coloured in purple (CsoS4A pentamer), blue/green (quasi-equivalent CsoS1A hexamer) and red (CsoS2) which is only present in the *T* = 9 shell.

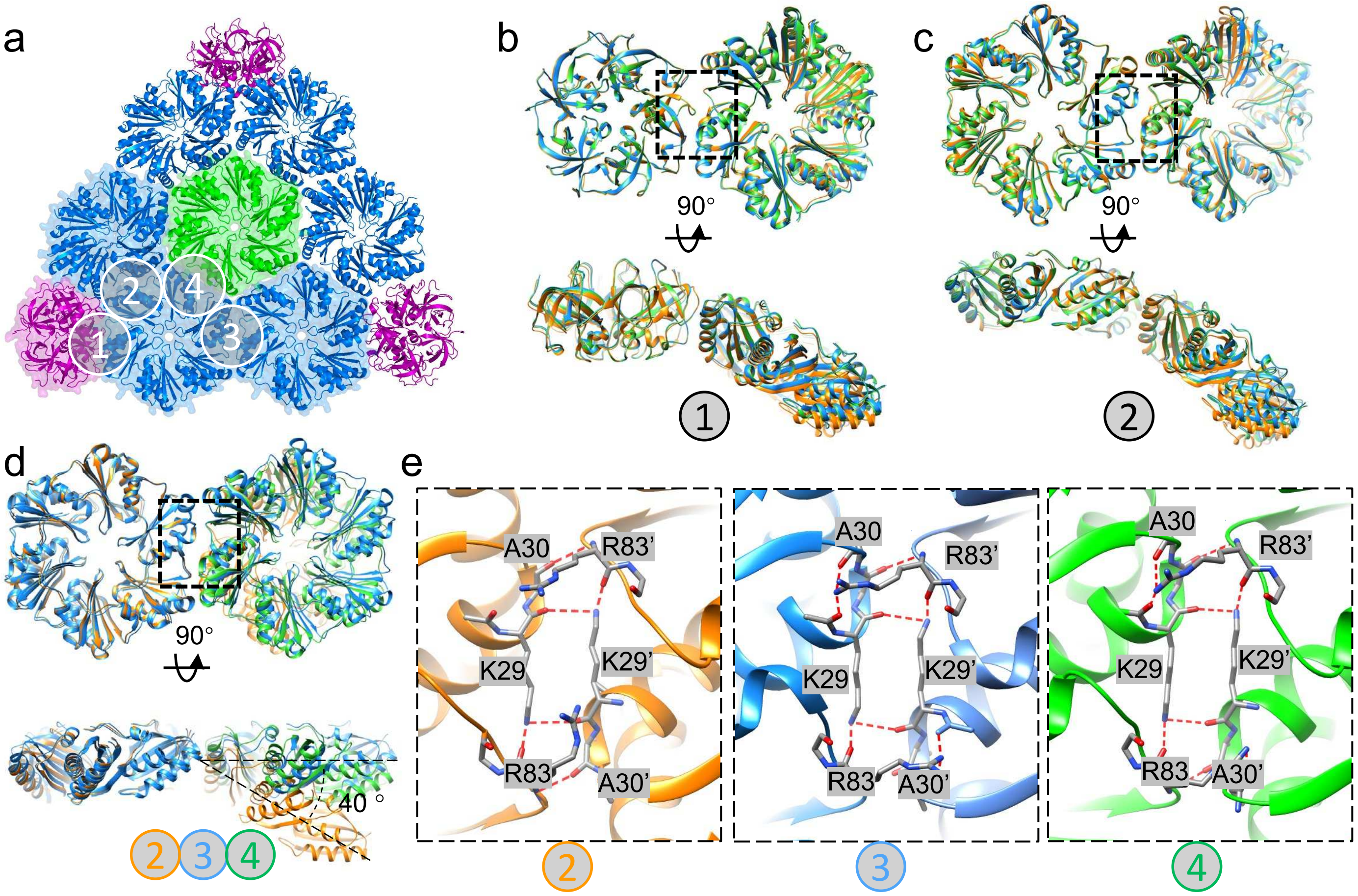

Figure 2 | Structurally conserved shell proteins with plastic assembly interfaces. (a) The overall organization *T* = 9 shell, with assembly interfaces 1 to 4 between capsomeres labelled. Interfaces 3 and 4 are unique to *T* = 9. (b-c) Overlay of interface 1 (b) and interface 2 (c) from *T* = 9 (blue), *T* = 4 (green) and *T* = 3 (orange) shells, viewed from top (left) and side (right). (d) Overlay of interfaces 2 (orange), 3 (blue) and 4 (green) from *T* = 9 shell only, aligned to the shared hexamer. There is a ~40 ° difference in curvature. (e) Details of interacting residues in the dimer interfaces 2, 3 and 4 in the *T* = 9 shell (dashed box in d). The hydrogen bond network is labelled with red dashed line.

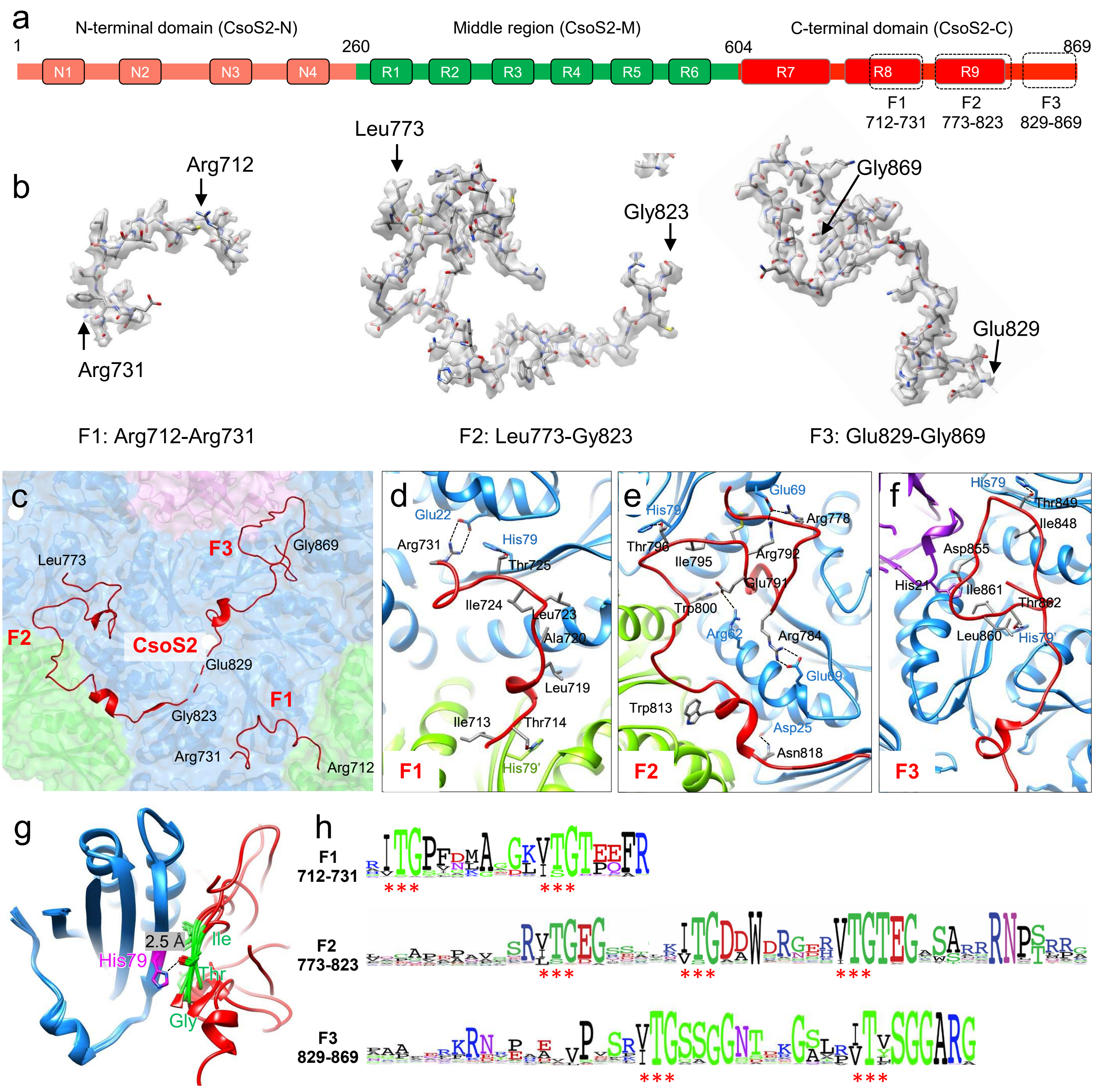

Figure 3 | CsoS2 binds to the shell through multivalent interactions with shell proteins and highly conserved interfaces via novel [IV]TG repeats. (a) The domain arrangement of CsoS2. The N-terminal, Middle and C-terminal domains are colored in pink, green and red, respectively. Three dashed boxes indicate the structured fragments resolved in *T* = 9 shell. (b) CryoEM densities of F1-F3 with atomic models. (c) CsoS2 interactions with shell components, viewed from inside. Three structured fragments in the C-terminal domain, F1, F2 and F3, are identified and labelled. (d-f) Interaction interfaces between CsoS2 F1(d), F2 (e) and F3 (f) fragments with shell components, CsoS1A (blue/green) and CsoS4A (purple). (g) Alignment of CsoS2-CsoS1A interacting motifs, showing the CsoS2 [IV]TG motif (green) in contact with CsoS1A His79. (h) Consensus sequences of CsoS2 C-terminal F1, F2 and F3 fragments from 100 CsoS2 sequences, plotted with Weblog. Asterisks indicate the conserved repeating [IV]TG motif present in each fragment.

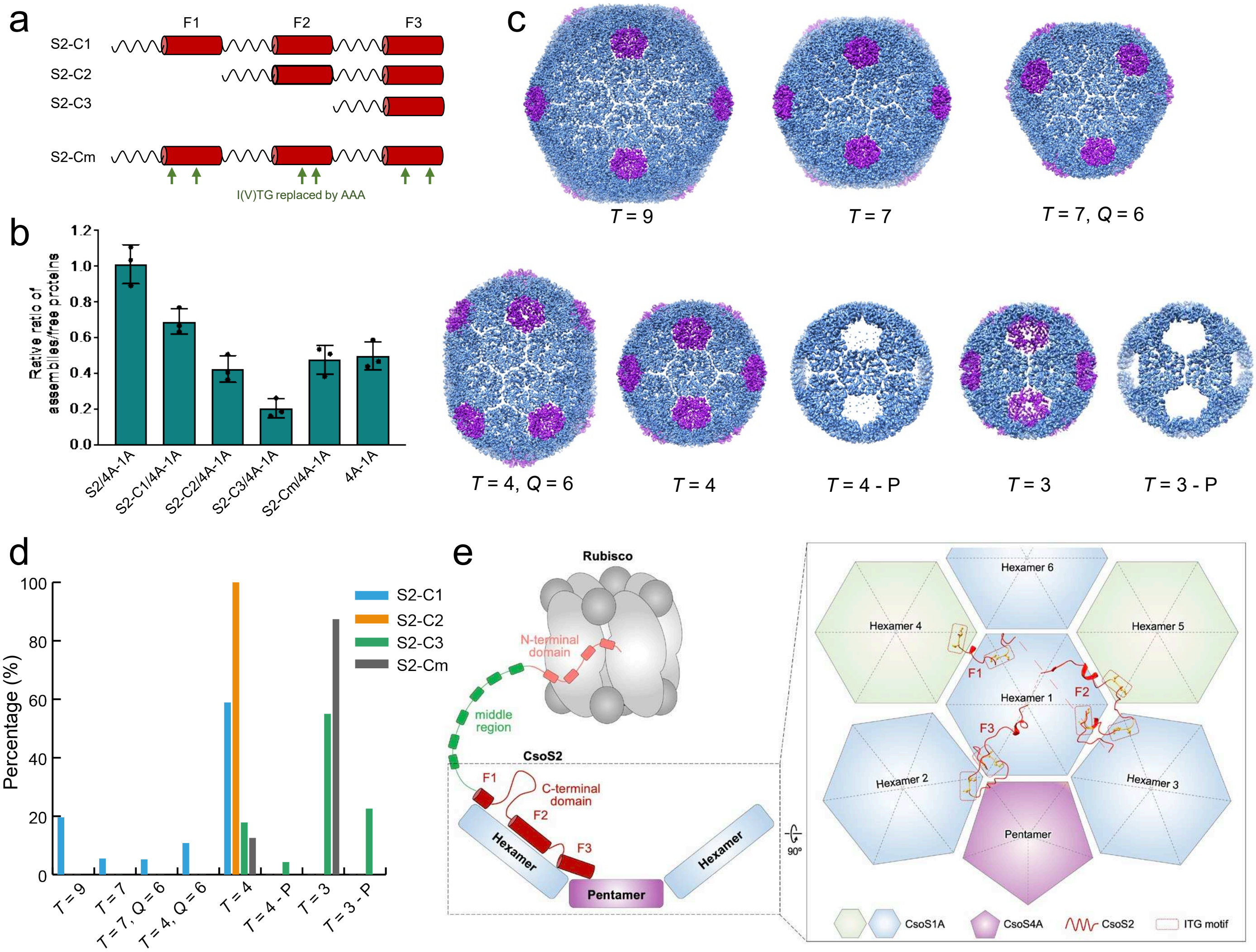

Figure 4 | CsoS2 C-terminal fragments interacts with shell proteins. (a) Schematic of four CsoS2 constructs. (b) Quantification of shell assembly from different CsoS2 constructs. The ratios of assembled shell and free shell proteins are quantified from western blot experiments (*n* = 3, see also fig. S9). 4A-1A denotes the CsoS4A-CsoS1A mini-shell. (c) Gallery of different shells formed in these constructs with hexamer in blue and pentamer in magenta. (d) Distribution of different shells in the four CsoS2 constructs as calculated from cryo-EM analysis; the number of particles for each shell is indicated in the Supplementary Table 1 and 2. (e) A schematic model of CsoS2 interacting with shell proteins and Rubisco using the C-terminal domain and N-terminal domain, respectively. Enlarged view shows the interactions of each fragment in CsoS2 C-terminal domain , F1, F2 and F3, with shell capsomers in the assembly.