nature plants

Article

https://doi.org/10.1038/s41477-024-01660-9

Structure and assembly of the α-carboxysome in the marine cyanobacterium *Prochlorococcus*

Received: 8 June 2023

Accepted: 29 February 2024

Published online: 08 April 2024

Check for updates

Rui-Qian Zhou[1,2,3], Yong-Liang Jiang[2,3] ✉, Haofu Li[1], Pu Hou[2], Wen-Wen Kong[2], Jia-Xin Deng[2], Yuxing Chen[2], Cong-Zhao Zhou[2] ✉ & Qinglu Zeng[1] ✉

Carboxysomes are bacterial microcompartments that encapsulate the enzymes RuBisCO and carbonic anhydrase in a proteinaceous shell to enhance the efficiency of photosynthetic carbon fixation. The self-assembly principles of the intact carboxysome remain elusive. Here we purified α-carboxysomes from *Prochlorococcus* and examined their intact structures using single-particle cryo-electron microscopy to solve the basic principles of their shell construction and internal RuBisCO organization. The 4.2 Å icosahedral-like shell structure reveals 24 CsoS1 hexamers on each facet and one CsoS4A pentamer at each vertex. RuBisCOs are organized into three concentric layers within the shell, consisting of 72, 32 and up to 4 RuBisCOs at the outer, middle and inner layers, respectively. We uniquely show how full-length and shorter forms of the scaffolding protein CsoS2 bind to the inner surface of the shell via repetitive motifs in the middle and C-terminal regions. Combined with previous reports, we propose a concomitant 'outside-in' assembly principle of α-carboxysomes: the inner surface of the self-assembled shell is reinforced by the middle and C-terminal motifs of the scaffolding protein, while the free N-terminal motifs cluster to recruit RuBisCO in concentric, three-layered spherical arrangements. These new insights into the coordinated assembly of α-carboxysomes may guide the rational design and repurposing of carboxysome structures for improving plant photosynthetic efficiency.

In photosynthetic organisms such as plants, algae and cyanobacteria, inorganic carbon is incorporated into biomass through photosynthetic carbon assimilation via the Calvin–Benson–Bassham (CBB) cycle[1]. As the often rate-limiting enzyme in the CBB cycle, ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) is the most abundant protein on Earth[2], accounting for a total mass of ~0.7×10^{15} g and is responsible for ~90% of atmospheric CO_2 fixation into biomass annually[3]. Despite its biological importance and global abundance, RuBisCO is known to be an enzyme of low efficiency, owing to its slow catalytic rate and poor capability to discriminate between the competing substrates, CO_2 and O_2 (ref. 4).

Cyanobacteria and many chemoautotrophs have evolved a CO_2-concentrating mechanism (CCM) to enhance the carbon fixation efficiency of RuBisCO[5–8]. A core component of their CCMs is

[1]Department of Ocean Science, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China. [2]School of Life Sciences, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China. [3]These authors contributed equally: Rui-Qian Zhou, Yong-Liang Jiang. ✉e-mail: jyl@ustc.edu.cn; zcz@ustc.edu.cn; zeng@ust.hk

carboxysome, a well-investigated bacterial microcompartment (BMC) of ~100 to 400 nm diameter[9,10]. Carboxysomes comprise an icosahedral-like proteinaceous shell that encapsulates the cargo enzymes RuBisCO and carbonic anhydrase that maintains the interconversion equilibrium between bicarbonate (HCO_3^-) and CO_2 (refs. 6,11). The shell is proposed to be selectively permeable to anions such as HCO_3^-, but prevents the passage of nonpolar compounds such as CO_2 (refs. 12–14) and potentially O_2 (refs. 13,14). The extracellular HCO_3^- is first transported into the cytosol via several membrane-bound HCO_3^- transporters[8,15,16]. Afterwards, the enriched intracellular HCO_3^- diffuses through the semipermeable shell into the carboxysome where HCO_3^- is converted to CO_2 via co-localized carbonic anhydrase. A mathematical model shows that RuBisCO-derived protons within carboxysomes help drive the conversion of HCO_3^- to CO_2, enhancing both CO_2 condensation and the RuBisCO carboxylation rate[17].

Carboxysomes are present in two distinct evolutionary lineages, designated as α-carboxysomes that globally dominate most aquatic habitats[18], and β-carboxysomes that are distributed mainly in freshwater/estuarine cyanobacteria[11,19]. They are similar in morphology and function, but differ in their RuBisCO compatibility and in the scaffolding proteins that mediate the crosslinking of various carboxysome components[11]. In both α- and β-carboxysomes, both RuBisCO and carbonic anhydrase are densely packed inside an icosahedral-like shell comprising hexameric and pentameric shell proteins. While the scaffolding proteins CcmM and CcmN are present in β-carboxysomes[20–22], only the scaffolding protein CsoS2 is found in α-carboxysomes[23,24]. In β-carboxysomes, RuBisCO is packed into paracrystalline arrays[25,26] with CcmM and CcmN, forming a heterotrimer[22] that acts as an adaptor to crosslink RuBisCO and other cargo enzymes with shell proteins to form liquid-like condensates[27] that trigger shell encapsulation into an intact β-carboxysome[28]. In contrast, the assembly of smaller α-carboxysomes is less well understood[11], with their components encoded by a cluster of genes within a *cso* operon[29]. The ~900-residue scaffolding protein CsoS2 contains no recognizable domains and is predicted to be an intrinsically disordered protein. Recently, it was reported that the N-terminal repetitive motifs of CsoS2 bind to internal RuBisCO and mediate α-carboxysome formation[24], whereas its repetitive motifs in the middle and C-terminal regions bind to the shell[30,31]. In addition, the cryo-electron microscopy (cryo-EM) structure of the recombinant α-carboxysome mini-shell from *Halothiobacillus neapolitanus* revealed that the C terminus of CsoS2 acts as a molecular thread that binds to the shell via conserved I/VTG motifs[32]. Moreover, cryo-EM studies of naturally purified α-carboxysomes from the marine α-cyanobacterium *Cyanobium* sp. PCC 7001 revealed that RuBisCO is organized in four concentric layers[33,34], whereas RuBisCO in the chemoautotrophic bacterium *H. neapolitanus* forms intertwining spirals that further pack into a lattice[33,35].

Carboxysome biogenesis requires the concerted action and ordered organization of thousands of individual proteins[28]. Although the structures of most protein components of the carboxysome are available, little is known about the structure and assembly pattern of intact carboxysomes, mainly due to the high structural heterogeneity and diversity of carboxysomes. In addition, although the structure and function of RuBisCO have been extensively studied for many years, the detailed chaperone-assisted assembly process of cyanobacterial RuBisCO has only recently been elucidated[36,37]. So far, the high-resolution structural information available is limited to the recombinant shell structures of *H. neapolitanus* α-carboxysome[32,38], *Halothece* sp. PCC 7418 β-carboxysome[39] and *Haliangium ochraceum* BMC[40] produced in *Escherichia coli*, which revealed the general principles of shell assembly. However, the process of natural carboxysome assembly, particularly the fine interaction pattern between the shell and scaffolding proteins, remains largely unknown.

Here we successfully isolated and purified intact α-carboxysomes from the marine cyanobacterium *Prochlorococcus* MED4 (ref. 41). Empowered by advances in the block-based reconstruction of single-particle cryo-EM analysis[42], we determined the ~86-nm-diameter intact α-carboxysome icosahedral-like structure, which revealed the architecture and assembly pattern of the shell as well as the internal three-layered organization of RuBisCO. Furthermore, we observed the multivalent interactions between the middle and C-terminal regions of the scaffolding protein CsoS2 and the shell capsomers in the native α-carboxysome. Our findings provide understanding of the structure and assembly of native α-carboxysomes by resolving how the CsoS2 scaffolding protein serves as a central hub to crosslink shell capsomers and recruit RuBisCOs.

Results

Prochlorococcus harbours the simplest carboxysomal operon

The components of α-carboxysome are mainly encoded by the *cso* operon among proteobacteria and cyanobacteria[6]. Phylogenetic analysis revealed that all *cso* operons comprise conserved gene sets that encode the essential components of the α-carboxysome: the large and small RuBisCO subunits CbbL and CbbS, respectively, the carbonic anhydrase CsoSCA, the scaffolding protein CsoS2, and the hexameric (CsoS1) and pentameric shell proteins CsoS4A/B (Fig. 1a,b). Notably, the genetic organization of *cso* operons varies greatly, with the total number of genes ranging from 10 to 25 (Fig. 1b).

Prochlorococcus is the most abundant photosynthetic picoplankton in the oceans, contributing to ~50% of marine carbon fixation[43,44]. The *cso* operon of *Prochlorococcus* strain MED4 consists of ten genes[29], representing the minimum number of genes in the *cso* operon (Fig. 1b). It contains four genes (*csoS1*, *csoS1D*, *csoS4A* and *csoS4B*) encoding the shell proteins, three genes encoding the cargo enzymes RuBisCO (*cbbL* and *cbbS*) and carbonic anhydrase (*csoSCA*), one gene encoding the scaffolding protein CsoS2, and two additional genes *pmm0548* (*ham1*) and *pmm0556* (*acRAF*). The protein CsoS1 contains a single Pfam00936 domain that assembles into BMC shell hexamer (BMC-H), which constitutes the majority of the α-carboxysome shell[29]. The CsoS1D protein, which contains a tandem repeat of two Pfam00936 domains[29], forms trimers (BMC-T) or pseudohexamers and presumably enables the passage of large molecules into and/or out of the carboxysome[45]. CsoS4A and CsoS4B, which share a sequence identity of 41%, have one Pfam03319 domain and form the BMC shell pentamers (BMC-P) that cap the 12 vertices of the shell[46]. The *pmm0548* gene encodes a putative histone acetyltransferase (HAM1) of the MYST family and is inserted between *csoS1* and *csoS1D*. Adjoining *csoS4B* is the *pmm0556* gene that encodes a conserved pterin dehydratase-like protein that functions as a RuBisCO assembly factor[47]. Overall, *Prochlorococcus* MED4 has the simplest α-carboxysome encoding the *cso* operon, thus making it an ideal model for investigating the structure and assembly principles of this CO_2-fixing BMC.

Overall architecture of *Prochlorococcus* α-carboxysome

The α-carboxysomes were purified from *Prochlorococcus* MED4 and analysed by electron microscopy, SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and mass spectrometry. Electron microscopy showed that the purified α-carboxysomes were mostly intact and had a canonical polyhedral shape, with an average diameter of ~90 nm (Fig. 2a). SDS–PAGE analysis of the purified α-carboxysome samples revealed four major protein bands (Fig. 2b), which were individually subjected to mass spectrometry analysis. Three of the four protein bands, including CbbL, CsoS1D and CsoS1/CbbS, were identified as the major components of the α-carboxysome (Fig. 2b). However, no protein information could be obtained for the remaining band. In addition to these major α-carboxysome components, liquid chromatography mass spectrometry analysis of the α-carboxysome samples detected other components, including CsoS2, CsoS4A, CsoS4B and CsoSCA (Supplementary Table 1). The ratio of the normalized protein abundance of each α-carboxysome component to that of CsoS4B ranged from

https://doi.org/10.1038/s41477-024-01660-9

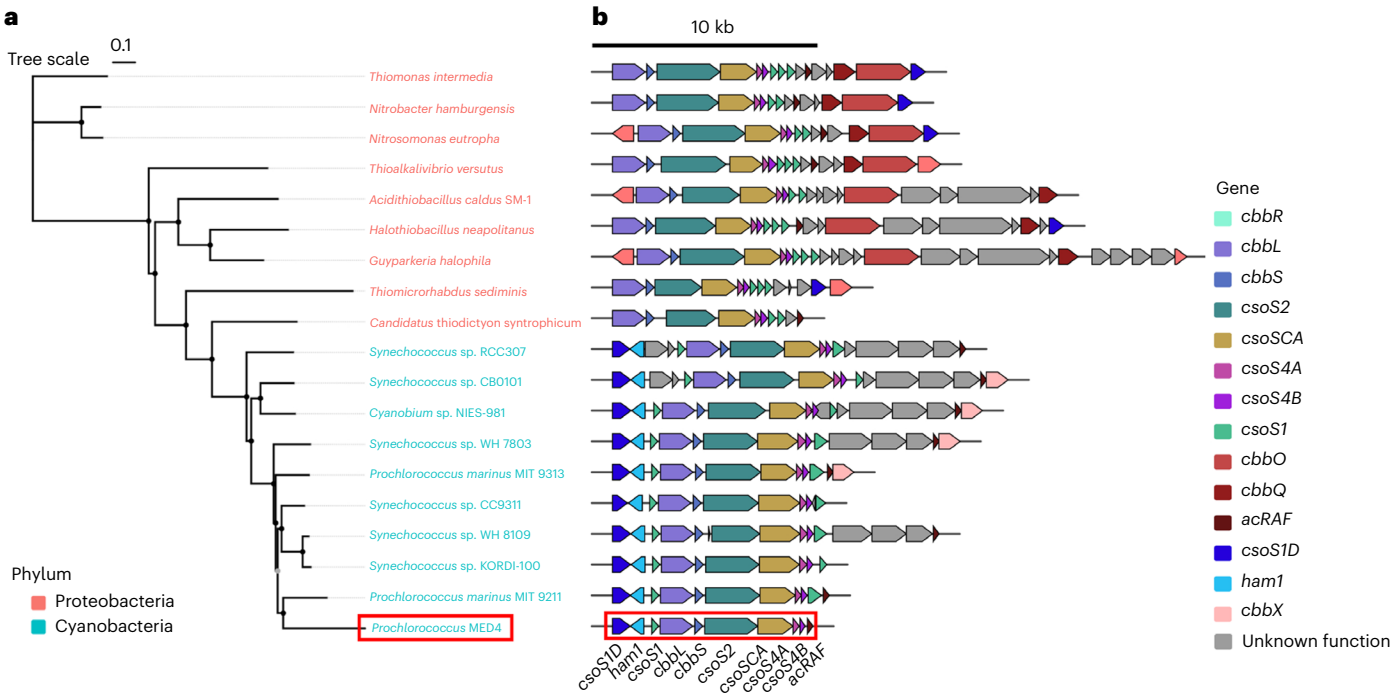


Fig. 1 | CsoSCA phylogenetic tree and gene organization of the *cso* operons among proteobacteria and α-cyanobacteria. a, Phylogenetic tree showing 19 representative *cso* operons from proteobacteria (red) and cyanobacteria (blue). The protein sequences were retrieved by BLASTP in the KEGG database using the CsoSCA sequence from *Prochlorococcus* MED4. The phylogenetic tree is midpoint rooted and was plotted by semiautomated workflow of NGphylogeny. The tree scale represents the amount of genetic substitution per site. **b**, Gene organizations of the 19 *cso* operons that were used to build the phylogenetic tree in **a**. The colour scheme for different genes is presented on the right. Arrows indicate the direction of the genes. The *cso* operon from *Prochlorococcus* MED4 is highlighted by a red box, with each gene labelled on the bottom.

700:1 to 4:1, consistent with previous analysis of the *H. neapolitanus* α-carboxysome[48]. Notably, several other proteins, including MinD (which is involved in cell division site positioning[49]) and HAM1 (which has a putative histone acetyltransferase domain), were also identified in the purified α-carboxysome samples (Supplementary Table 1), suggesting that these proteins might interact with α-carboxysome or play a role in the assembly and function of α-carboxysome.

Owing to the relatively small size and rigidity of *Prochlorococcus* α-carboxysomes, we were able to conduct a single-particle cryo-EM study to investigate their structure and assembly. We collected ~23,400 cryo-EM images and manually picked ~32,000 intact α-carboxysome particles for further analysis. Subsequent two-dimensional (2D) classification revealed an obvious pattern of overall architecture for the shell and cargo enzymes (Supplementary Fig. 1). However, the 3D classification and refinement attempts with different symmetries failed to converge to interpretable maps, indicating the structural flexibility of intact α-carboxysomes. Applying icosahedral symmetry to the best class of 3,634 intact α-carboxysome particles yielded a cryo-EM map at 7.5 Å resolution (Supplementary Fig. 2a,b), which was further improved to 6.2 Å with the shell mask (Supplementary Fig. 2a). It suggested that the native α-carboxysome shell generally adopts an icosahedral-like architecture, whereas the arrangement of internal enzymes displays plasticity at some level, consistent with previous reports[33–35].

To further push the resolution of the shell, we applied the block-based reconstruction method and segmented the intact α-carboxysome into 12 shell vertices[42]. After several rounds of 3D classification and refinement using *C5* symmetry, we finally obtained a 4.2 Å map for the shell vertex (Supplementary Fig. 2a,c), which allowed us to build an atomic model of the shell vertex. Notably, this structure was determined from the average of relatively homogeneous blocks of the α-carboxysome shell, which may not fully represent the authentic assembly pattern of all protein components in the intact α-carboxysome shell. Moreover, we also conducted masked 3D classification and refinement of the internal density with icosahedral symmetry, yielding an overall 15.0 Å cryo-EM map of the internal layers of α-carboxysomes (Supplementary Fig. 2a,d). In the cryo-EM map, we discerned three concentric layers of cargo proteins, which we named the outer, middle and inner layers (Fig. 2c). The icosahedral-like shell is ~860 Å in diameter and has a thickness of ~40 Å. The three layers of the internal core consist of tightly and regularly packed RuBisCO. These encapsulated spheres have diameters of ~720, 480 and 240 Å, respectively (Fig. 2c). Overall, the *Prochlorococcus* α-carboxysome has a well-organized four-layered architecture, which differs from previously reported architectures of α-carboxysomes in terms of size and internal RuBisCO organization[33,35].

Fine structure of the icosahedral-like shell

The *Prochlorococcus* α-carboxysome shell generally adopts an icosahedral shape of ~860 Å diameter and ~40 Å thickness (Fig. 3a). It has a triangulation number *T* of 49, with an asymmetric unit composed of eight BMC-H hexamers (48 CsoS1 subunits in total) and one BMC-P pentamer subunit (CsoS4A/4B) (Fig. 3b). The 20 shell facets of an α-carboxysome are formed by 480 BMC-H hexamers, totalling 2,880 CsoS1 subunits, and 12 vertices comprising BMC-P pentamers consisting of 60 CsoS4A and/or CsoS4B proteins (Fig. 3a). The 48:1 BMC-H:BMC-P stoichiometry matches the mass spectrometry results of the purified *Prochlorococcus* α-carboxysomes (Supplementary Table 1) and previous analyses of *H. neapolitanus* α-carboxysomes[48]. Compared with CsoS4B, CsoS4A has an N-terminal extension of ten residues that form a β-strand, which was clearly observed in our cryo-EM map (Extended Data Fig. 1). Therefore, the 12 vertices in the shell map are composed primarily of CsoS4A, with some vertices occupied by CsoS4B, which was detectable at low abundance by mass spectrometry (Supplementary Table 1). Previous studies of recombinant BMC have shown that BMC-T proteins form

https://doi.org/10.1038/s41477-024-01660-9

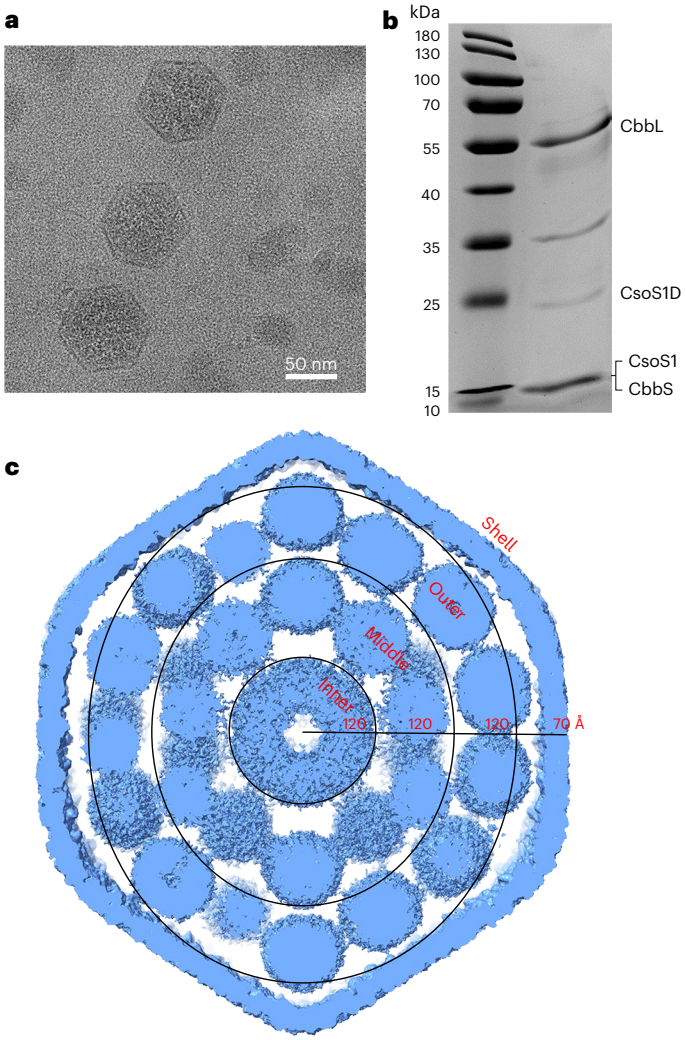


Fig. 2 | Biochemical characterization of the α-carboxysome from *Prochlorococcus* MED4. a, A representative cryo-EM image of α-carboxysome samples. The micrograph is a representative of ~11,700 cryo-EM images that contain α-carboxysome particles. **b**, SDS–PAGE analysis of the purified α-carboxysomes. The SDS–PAGE gel is a representative of >3 independent experiments that showed similar results. In total, four protein bands could be identified, three of which corresponded to the proteins CbbL, CsoS1D and CsoS1/CbbS, as indicated on the right of the SDS–PAGE gel. **c**, A sliced view of the overall architecture of the four-layered structure of *Prochlorococcus* α-carboxysome. The distances between each layer are labelled.

double-layered dimers of trimers that protrude from the shell surface[40,50]. However, in our shell structure we could not locate the BMC-T protein CsoS1D, presumably due to its low abundance (4 for every 700 CsoS1) (Supplementary Table 1).

Structural analysis of the native α-carboxysome shell showed that the BMC-P pentamers were arranged as a truncated pyramid, with the bottom facing outward and the concave sides of the BMC-H hexamer facets facing outward (Fig. 3a). In the shell structure, there are three distinct interfaces: two different hexamer–hexamer interactions and one hexamer–pentamer interaction (Fig. 3c,d). The hexamers along the longitudinal line connecting the pentamers between two vertices of the shell are in a side-by-side planar orientation, whereas the neighbouring hexamers surrounding the latitudinal line are tilted by 36° to 10°, with the tilt angle decreasing gradually from the pentamer vertex towards the periphery (Extended Data Fig. 2). Compared with the synthetic BMC shell of *Haliangium ochraceum*[40], our shell structure reveals greater variability in the tilt angles between hexamers along the latitudinal lines, implying that size variations in the BMC shell are achieved by adjusting these tilt angles.

For the two hexamer–hexamer interfaces, the K^{25}AA motif of one subunit and the A^{78}RPH motif of the symmetric subunit of the same hexamer interact with the counterparts of the neighbouring hexamer, which contributes the majority of the interface area (Fig. 3c). The Lys25 residues of the KAA motif are arranged in an antiparallel manner, providing interactions with Lys25 and Arg79 of the neighbouring hexamer via hydrogen bonds. For the hexamer–pentamer interface, the K^{25}AA and A^{78}RPH motifs of the hexamer also play a central role in forming an interface with the pentamer (Fig. 3d). In each subunit of the pentamer, the residues in the protruding loops on the lateral side participate in interactions with the surrounding hexamers via nonpolar interactions and several hydrogen bonds (Fig. 3d). Overall, the hexamer–hexamer and hexamer–pentamer interfaces are well complementary in shape

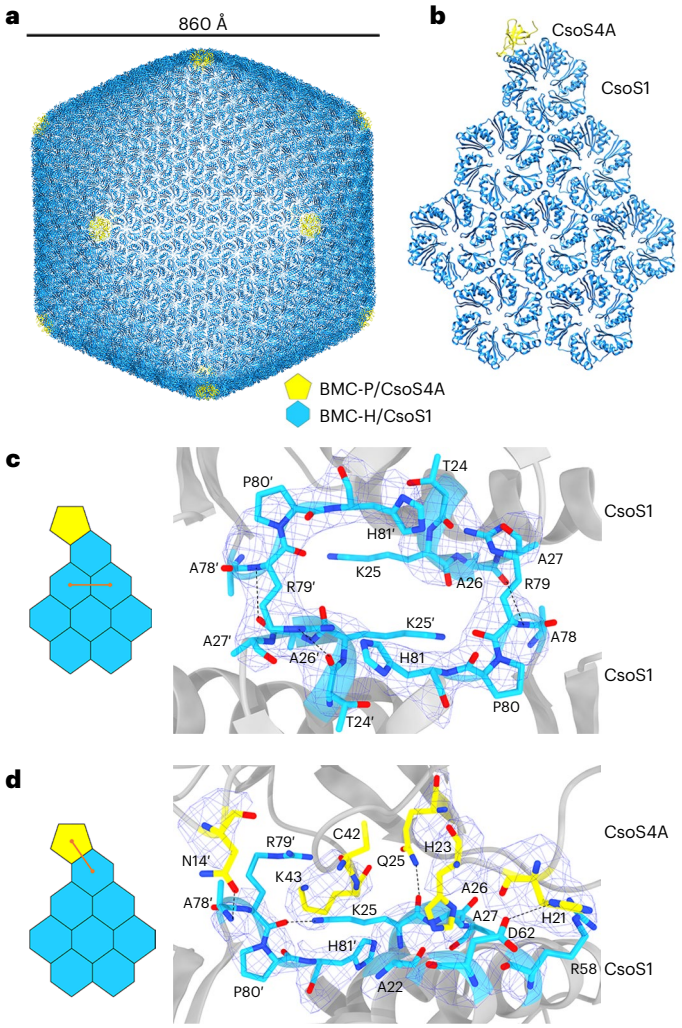


Fig. 3 | Structure and assembly pattern of the icosahedral-like shell of the *Prochlorococcus* α-carboxysome. a, Structural model of the intact shell. The CsoS1 hexamers of 480 copies are coloured blue whereas the CsoS4A pentamers of 12 copies are coloured yellow. **b**, Structure of an asymmetric unit of the shell. **c**,**d**, Detailed view of the hexamer–hexamer (**c**) and hexamer–pentamer (**d**) interfaces. The structures of the CsoS1 hexamer and CsoS4A pentamer are shown in cartoon and coloured grey. The interface residues are shown as sticks and coloured according to the atom types; cyan and yellow for the carbon atoms in CsoS1 and CsoS4A, whereas red and blue for the oxygen and nitrogen atoms, respectively. The cryo-EM maps corresponding to the interface residues are shown in blue mesh. The hydrogen bonds are indicated by dashed lines. Residues from the symmetric subunit of the BMC-H hexamer or BMC-P pentamer are labelled with a prime.

https://doi.org/10.1038/s41477-024-01660-9

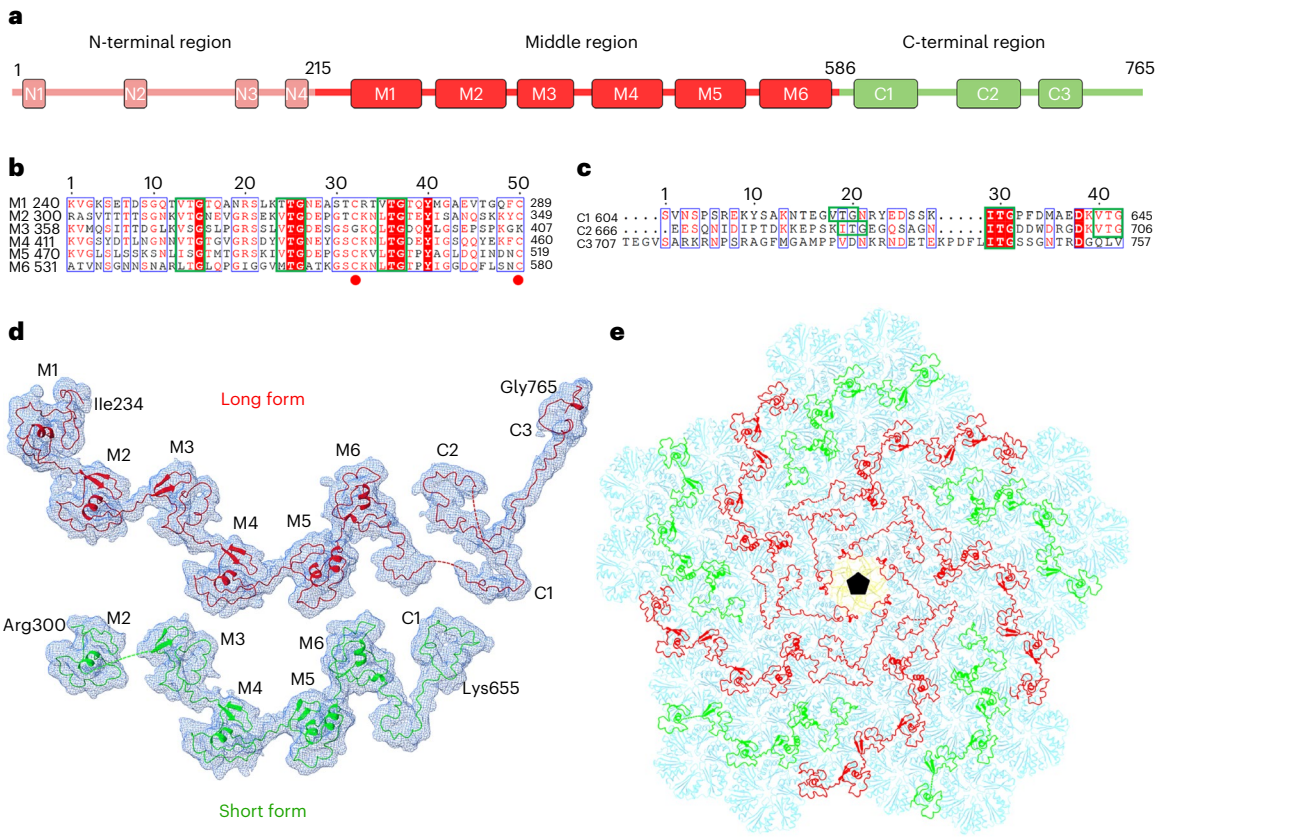


Fig. 4 | Structure of the scaffolding protein CsoS2 and its pattern of binding to the inner surface of the shell. a, CsoS2 is composed of the N-terminal, middle and C-terminal regions. The repetitive units among the three regions of CsoS2 are labelled sequentially. **b**,**c**, Sequence alignment of the repetitive units in the middle region (**b**) and C-terminal region (**c**) of CsoS2. The conserved V/L/ITG motifs are marked with green boxes. The conserved cysteine residues in the M repeats are indicated by red dots. **d**, Structure of CsoS2 in two forms, with the cryo-EM maps shown in blue mesh. The long and short forms are coloured red and green, respectively, with the M and C repeats labelled sequentially. The linkers of weaker densities that are not modelled in the final CsoS2 structures are indicated as dashed lines. **e**, The overall pattern of CsoS2 binding to the shell, as viewed from the inside of the shell vertex. The structures of 40 BMC-H hexamers and one BMC-P pentamer are shown in semitransparent cartoons and are coloured cyan and yellow, respectively. Ten subunits of CsoS2 are shown as cartoons, with the long and short forms coloured red and green, respectively. The long and short forms of CsoS2 are arranged in an alternative manner around the five-fold axis at the pentamer vertex, which is indicated by a black regular pentagon.

(Fig. 3b), which facilitates the formation of a regular shell. Considering the high sequence and structural conservation among all BMC-H and BMC-P proteins (Extended Data Fig. 3a,b), these interaction patterns and assembly principles are probably universal among BMCs, as observed in the shell structure of the *H. ochraceum* BMC[40].

CsoS2 binds multivalently to the inner surface of the shell

The scaffolding protein CsoS2 from *Prochlorococcus* MED4 possesses 765 residues comprising an N-terminal, middle and C-terminal region (Fig. 4a)[23,24]. The N-terminal region contains four repetitive motifs (N1–N4) that comprise an α-helix that interacts with RuBisCO to facilitate its encapsulation in carboxysomes[24]. The middle region has six repeats (M1–M6) of ~50 residues in length that contain three conserved V/L/ITG motifs separated by eight residues, each sharing a sequence identity of between 32% and 45% (Fig. 4b). Compared with the highly conserved M repeats among CsoS2 homologues, the C-terminal region is variable and has three repeats with <25% sequence identity, but it also contains V/L/ITG motifs (Fig. 4c).

Previous studies have shown that the middle and C-terminal regions of CsoS2 bind to shell proteins[23,30,31]. The recently reported cryo-EM structure of the *H. neapolitanus* α-carboxysome mini-shell revealed that the C terminus of CsoS2 specifically binds to the tricapsomer interface of the shell via conserved I/VTG motifs[32]. However, it remains unclear how the full-length CsoS2, especially the middle region harbouring six repeats, anchors to the native carboxysome shell. In the cryo-EM map of the shell vertex, we observed additional densities of ~4.5 Å resolution that ran along the inner surface of the shell (Fig. 4d and Supplementary Fig. 2c). By fitting the predicted middle and C-terminal region structures of CsoS2, we constructed a model of the middle and C-terminal regions of CsoS2 binding to the shell inner surface. In the final model, CsoS2 shows a long and a short form, but both forms share a similar overall conformation (Fig. 4d). The long form of CsoS2 covers the complete middle and C-terminal regions harbouring residues from Ile234 in the middle region to the most C-terminal Gly765, whereas the short form contains only M2–M6 and C1 with residues from Arg300 to Lys655 (Fig. 4d). Residues in some of the linkers spacing the M and/or C repeats had weaker densities and were not built in the final CsoS2 structures. Around each five-fold axis at the pentamer vertex, five long and five short forms of CsoS2 are arranged in an alternating manner, with a total of ten CsoS2 subunits binding to the inner surface of the shell (Fig. 4e). Each CsoS2 subunit binds to multiple BMC-H hexamers, five long forms extending to the vertex and five short forms terminating at the second ring of hexamers around the vertex (Fig. 4e). Notably, the middle and C-terminal regions of CsoS2 are absent from the central three-fold axis in each of the facets (Extended Data Fig. 4a). At the two-fold axis of the shell, two short CsoS2 subunits are aligned in an antiparallel manner, forming an extended binding surface to the shell (Extended Data Fig. 4b). Overall, the inner surface of the intact shell is regularly covered with 120 CsoS2 subunits (Extended Data Fig. 5a,b).

https://doi.org/10.1038/s41477-024-01660-9

Each M repeat adopts an α + β fold consisting of one or two α-helices and a two-stranded antiparallel β-sheet (Fig. 4d). The repeats M1–M6 share a similar overall structure, with a root-mean-square deviation of 1.3–2.6 Å (Extended Data Fig. 6). Two cysteines are conserved in all M repeats, except for M3 (Fig. 4b), and each cysteine pair forms a disulfide bond that further stabilizes the local conformation (Extended Data Fig. 6). Each M repeat binds to three neighbouring BMC-H hexamers at the pseudo-three-fold axis of the shell (Fig. 5a). The three conserved V/L/ITG motifs of each M repeat interact with the C-terminal β-strand (β4) centred at residue His75 of BMC-H, forming a hydrogen-bond network (Fig. 5b and Extended Data Fig. 7a–c). The six subunits of each BMC-H hexamer interact with at least two CsoS2 subunits (Extended Data Fig. 8). Therefore, each M repeat functions as a rivet to further reinforce the assembly of three neighbouring hexamers at the pseudo-three-fold axis. Overall, each long and short CsoS2 subunit runs along the inner surface of the shell, crosslinking ten and nine BMC-H hexamers, respectively (Fig. 4e).

In contrast to the middle region of CsoS2, the C-terminal region does not possess any notable secondary structure (Fig. 4d), and it exhibits greater diversity in interacting with the shell. Similar to M repeats, both C1 and C2 bind to three adjacent BMC-H hexamers at the pseudo-three-fold axis via three conserved V/L/ITG motifs via a hydrogen-bond network (Fig. 5c). The C3 of long-form CsoS2 stretches across the central pore of the adjacent BMC-H hexamer towards the vertex and forms a kink at the interface between two hexamers and a pentamer (Fig. 5a). Notably, in addition to binding to the CsoS1 hexamer via an ITG motif, the most C-terminal region of CsoS2 interacts with the loop connecting β1 and β2 of the CsoS4A pentamer (Extended Data Fig. 9). Compared to the structure of the α-carboxysome mini-shell from *H. neapolitanus*[32], the C-terminal region of *Prochlorococcus* CsoS2 adopts a similar pattern of binding to the shell through conserved V/L/ITG motifs, suggesting the general binding mode of the scaffolding proteins, which act as molecular threads that crosslink multiple shell capsomers.

RuBisCOs form three concentric layers within α-carboxysome

The focused refinements of the internal spheres obtained an overall 15.0 Å cryo-EM map, which enabled us to assign the overall pattern of RuBisCO arrangements within *Prochlorococcus* α-carboxysomes. We clearly assigned a three-layered structure within the α-carboxysome shell (Fig. 6a). The diameters of the three layers are 720, 480 and 240 Å, respectively (Figs. 2c and 6a). By manually placing individual RuBisCO enzymes at the corresponding density in the outer and middle layers, we constructed a model of the internal organization of RuBisCO within the α-carboxysome (Fig. 6b). A total of 72 and 32 RuBisCOs could be clearly modelled in the outer and middle layers, respectively (Fig. 6b). However, modelling RuBisCO with high certainty at the inner layer is difficult due to the low-quality cryo-EM map. Given that the diameter of the inner sphere is ~240 Å and the size of RuBisCO is 110 Å, the inner layer could theoretically accommodate only up to four RuBisCOs. The RuBisCOs in the outer and middle layers are arranged concentrically, with the fourfold axis of each RuBisCO roughly pointing to the centre of the carboxysome. The overall organization and number of RuBisCOs in each of the three layers of *Prochlorococcus* α-carboxysomes are consistent with those in the inner three layers of *Cyanobium* α-carboxysomes, which have four concentric layers of RuBisCO and a diameter of ~120 nm[34]. Previous studies have shown that CsoS2 N-terminal region[24] and the N-terminal domain of CsoSCA[51] bind to RuBisCO at an overlapped binding site on CbbL, but with different binding modes. However, due to high heterogeneity, we cannot model the CsoS2 N-terminal region or the low-abundance CsoSCA in the present low-resolution maps of α-carboxysomes.

Our model identified two distinct interfaces between RuBisCOs (Extended Data Fig. 10a). The first interface is formed by contacts between RuBisCOs within the same layer and involves interactions on the lateral side of RuBisCO (Extended Data Fig. 10b). The second interface is established by the interaction between RuBisCOs across concentric layers in a top-to-bottom configuration (Extended Data Fig. 10c).

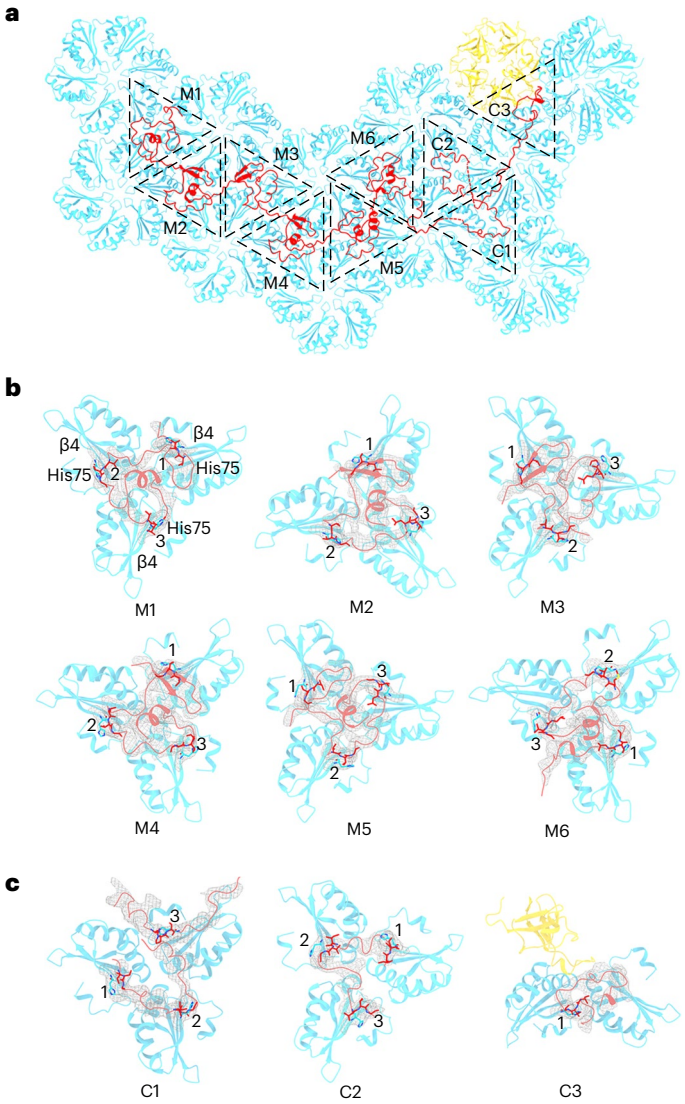


Fig. 5 | Multivalent interactions between CsoS2 and the shell. a, Overall view of the long form of CsoS2 binding to the inner surface of the shell. The structures of BMC-H, BMC-P and CsoS2 are shown as cartoons and are coloured cyan, yellow and red, respectively. The M and C repeats are marked with dashed triangles. **b**, The detailed interactions of six M repeats binding to three adjacent CsoS1 hexamers. The CsoS2 and CsoS1 structures are shown as cartoons and are coloured red and cyan, respectively. Three conserved V/L/ITG motifs in each M repeat are shown as sticks and are labelled sequentially. The interface residue His75 from β4 of CsoS1 is shown as sticks. The cryo-EM maps corresponding to CsoS2 and His75 of CsoS1 are shown in blue mesh. **c**, The detailed interactions of three C repeats of CsoS2 binding to adjacent CsoS1 hexamers and the CsoS4A pentamer (yellow).

Discussion

Unravelling the assembly mechanisms of carboxysomes is critical to understanding the biogenesis and functions of bacterial metabolic organelles. Owing to the compositional heterogeneity and structural plasticity of α-carboxysomes, there remains great challenges in solving their high-resolution structures. We present a single-particle cryo-EM analysis of the most structurally simple α-carboxysome purified from

https://doi.org/10.1038/s41477-024-01660-9

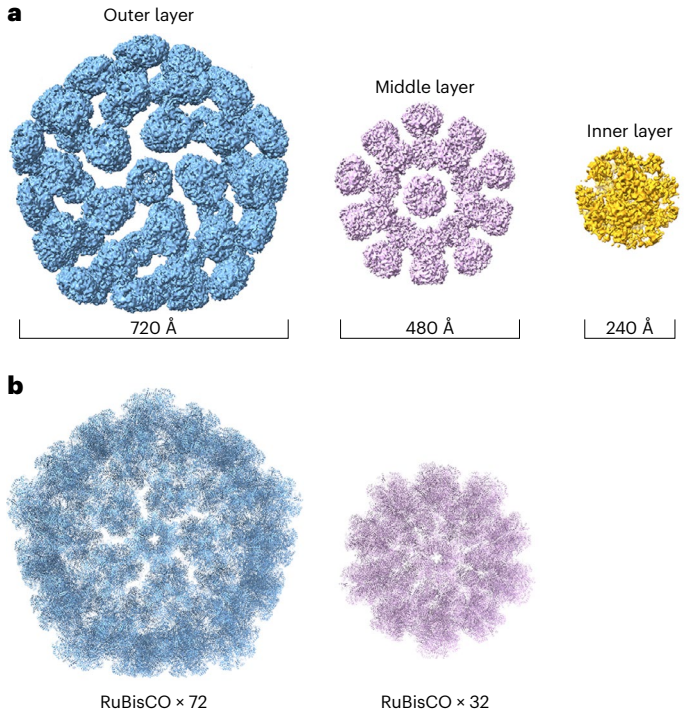


Fig. 6 | Internal organization of RuBisCO within the *Prochlorococcus* α-carboxysome. a, Cryo-EM maps of the three internal layers. The diameters of the layers are indicated at the bottom. **b**, Models illustrating the arrangement of RuBisCO in the outer and middle layers. In total, 72 and 32 copies of RuBisCO were fitted at the corresponding densities at the outer and middle layers, respectively.

Prochlorococcus MED4. Using block-based reconstruction and localized refinement, we obtained a 4.2-Å-resolution map of the shell vertex, as well as the overall organization of its internal RuBisCO. While our block-based reconstruction obtained the structure of the shell vertex, this structure represents only the relatively homogeneous components within the intact carboxysome shell, the bona fide structure of which could be more dynamic. In addition, low-abundance and/or heterogeneous components, such as CsoS1D, CsoS4B and CsoSCA, were averaged during the block-based calculations, making their roles in α-carboxysome biogenesis unclear.

Assembly and maturation of α-carboxysomes require the coordinated action of numerous protein components, wherein the multivalent scaffolding protein CsoS2 acts as a central hub in orchestrating multiple assembly reactions through its three distinct regions. Each region of CsoS2 has several repeats that form multivalent interactions with the shell and encapsulated RuBisCO[23,24,52]. Our 4.2 Å structure of the shell vertex elucidated the fine interaction pattern of CsoS2 middle and C-terminal regions binding to the shell, representing a snapshot of the binding mode in native carboxysomes. The middle and C-terminal regions of CsoS2, which harbour conserved V/L/ITG motifs, specifically bind to the joint of three adjacent BMC-H hexamers of the shell (Fig. 5a). Notably, the middle regions spans along the flat shell facets, whereas the C-terminal region clusters at the shell vertices that exhibit greater local curvature (Extended Data Fig. 5a,b). Consistent with a previous study[30], the number of M repeats probably determines the size of the α-carboxysome, and the C-terminal region is required for shell encapsulation and carboxysome formation. Given the high conservation of BMC-H proteins (Extended Data Figs. 3a and 7) and the shared consensus motifs of CsoS2 among α-carboxysomes[23], our model of the scaffolding protein CsoS2 binding to the inner surface of the shell could be a paradigm applicable to all α-carboxysomes. The distinctive multivalent interactions of CsoS2–shell and CsoS2–RuBisCO are vital for governing the architectures of shell assembly and concomitant encapsulation of RuBisCO.

In the *Prochlorococcus* carboxysome, RuBisCOs are neatly organized into three concentric layers (Fig. 6a,b). The outer layer has 72 RuBisCOs: 1 at each of the 12 vertices and 3 on each of the 20 facets. The middle layer has 32 RuBisCOs, 1 at each vertex and 1 on each facet. In a recent cryo-EM study of *Cyanobium* α-carboxysomes, four internal layers were proposed to contain 192, 72, 32 and 4 RuBisCOs[34]. For the outer layer of the *Cyanobium* α-carboxysome, the 192 RuBisCOs are distributed as 1 at each of the 12 vertices and 9 on each of the 20 facets. Therefore, we deduced a general formula to predict the total number (N) of RuBisCO molecules in each concentric layer of α-carboxysomes: $N = 12 + 20 \times 3^{n-2}$ ($n \geq 2$), where n represents the layer number counted from the innermost layer outwards.

Our results revealed that CsoS1, RuBisCO and CsoS2 are present at a molecular stoichiometry of 480:108:120, with the long and short forms of CsoS2 containing 25 and 18 motifs, respectively. Thus, in total there are 2,580 binding motifs from the 120 CsoS2 subunits that occupy 2,520 BMC-H subunits, excluding 360 BMC-H subunits at the three-fold axes of the shell (Extended Data Fig. 5a,b). In addition, each CsoS2 N-terminal region has 4 binding motifs for RuBisCO, resulting in a total of 480 binding motifs for 120 CsoS2 subunits. Ideally, these motifs could accommodate 60 RuBisCOs, assuming that each RuBisCO has 8 CsoS2-binding sites. Therefore, we propose that RuBisCOs in the outer layer are mainly crosslinked by the N-terminal regions of CsoS2, whereas the majority of the binding sites on RuBisCO in the middle and inner layers are most probably vacant. In addition, as recently reported, RuBisCO functions as an interaction hub to recruit CsoSCA to the carboxysome by specifically binding to the N-terminal peptide of CsoSCA[51]. Given that only a dozen CsoSCA dimers are present in an α-carboxysome (Supplementary Table 1), it is likely that a minor portion of the binding sites on RuBisCO could be occupied by CsoSCA dimers, which might be scattered in the space between RuBisCOs in the middle and inner layers of the α-carboxysome. However, further high-resolution structures are needed to elucidate the fine interaction patterns among RuBisCO, CsoS2 and CsoSCA in α-carboxysomes.

Previous studies have proposed distinct mechanisms for the assembly of α- and β-carboxysomes. The α-carboxysomes were proposed to co-assemble the shells concomitantly with the aggregation of cargo enzymes[6,53], whereas the β-carboxysomes were proposed to assemble in an 'inside-out' manner, initiating with the condensation of the inner core[28]. Together with previous findings[24,33–35,51], our structure enables us to update a concomitant 'outside-in' model for the assembly of α-carboxysomes (Fig. 7). In this model, dozens of BMC-H hexamers gather together to form the nascent shell patch. Multiple copies of the scaffolding protein CsoS2 reinforce the shell structure via binding of the middle and C-terminal regions to the inner surface of the nascent shell patch. Consequently, the N-terminal regions of CsoS2 are clustered together and further recruit RuBisCO, which forms condensates in the vicinity of the shell. Along with the coalescence of more shell patches and cargo condensates, an icosahedral-like structure of α-carboxysome is eventually formed upon the closure of 12 vertices with BMC-P pentamers.

In summary, carboxysomes are self-assembled proteinaceous organelles found in nature that encapsulate photosynthetic enzymes to enhance carbon fixation. This work advances our understanding of how the concomitant assembly of shell and cargo enzymes is coordinated in the α-carboxysome. Our results established the precise stoichiometry and fine interaction patterns among essential components of α-carboxysomes, especially the multivalent interactions between the scaffolding protein and the shell. These findings elucidate the fine mechanism of α-carboxysome assembly and architecture, tuned by scaffolding proteins, and will guide the design and engineering of carboxysome-based nanobioreactors in diverse

https://doi.org/10.1038/s41477-024-01660-9

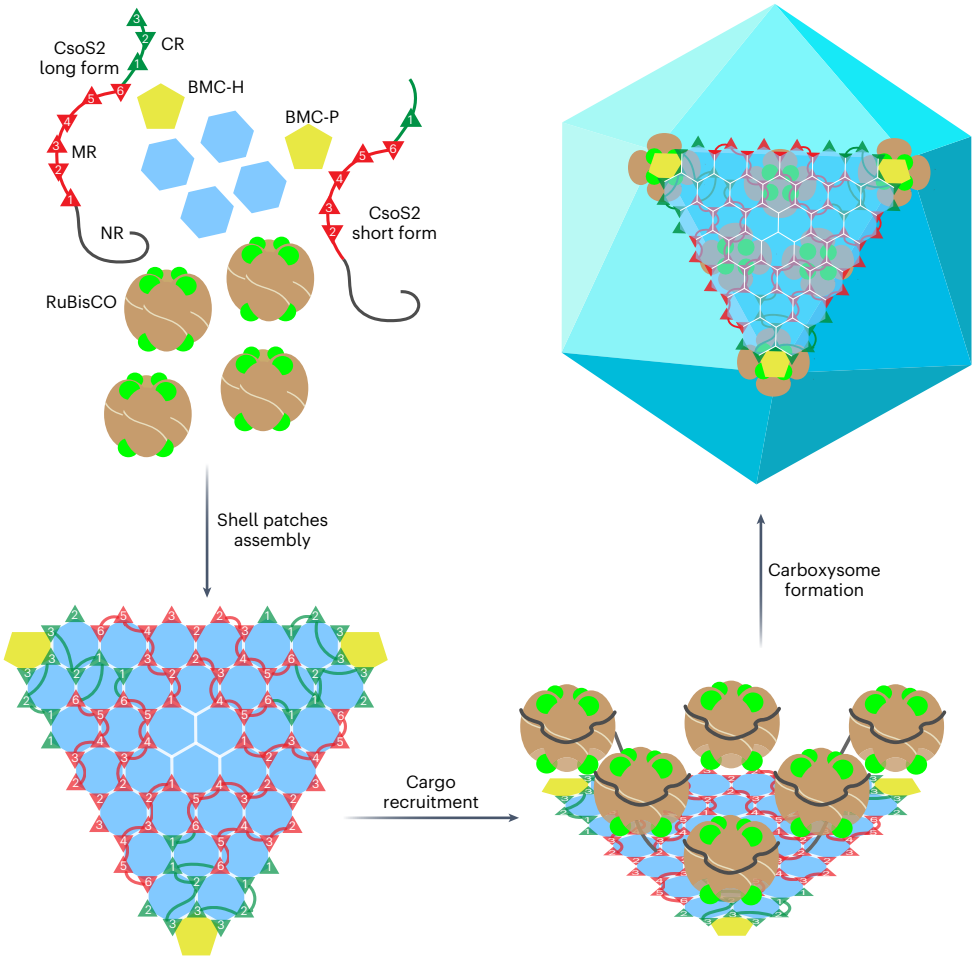


Fig. 7 | An updated concomitant model for α-carboxysome assembly. The BMC-H hexamers, BMC-P pentamers and RuBisCO are shown as blue hexagons, yellow pentagons and brown/green spheres, respectively. The M and C repeats of the long and short forms of CsoS2 are sequentially numbered and shown as red and green triangles, respectively. Dozens of BMC-H hexamers self-assemble into the nascent shell patches. The middle and C-terminal regions (MRs and CRs, respectively) of the scaffolding protein CsoS2 bind to the interface of three adjacent BMC-H hexamers to reinforce the structure of the shell. The N-terminal regions (NRs) of CsoS2 are clustered together to recruit RuBisCO to the vicinity of the shell. Accompanied by the accumulation of more shell patches and cargo enzymes, intact α-carboxysomes are eventually formed upon the closure of 12 vertices with BMC-P pentamers.

biotechnological approaches, such as improving photosynthesis[54] and other biocatalysis[55].

Methods

Phylogenetic and conservation analysis

The consecutive *Prochlorococcus* MED4 genes *pmm0550*, *pmm0551*, *pmm0552* and *pmm0553* were blasted using KEGG synteny for conserved gene order. The *cso* operons from 76 organisms were aligned against *Prochlorococcus* MED4 CsoSCA, with matrix settings of BLOSUM 80 and an *E*-value of 0.05. A phylogenetic tree was generated using the CsoSCA protein, which is uniquely found in chemoautotrophic proteobacteria and α-cyanobacteria that harbour α-carboxysomes. Therefore, we used the CsoSCA protein sequence to perform sequence alignments and construct a phylogenetic tree. The output tree was drawn by advanced NGphylogeny with 1,000 bootstraps, and the *cso* gene clusters were plotted using ChiPlot. The CsoS1 and CsoS2 sequences from the 19 *cso* operons used for building the phylogenetic tree were aligned at the amino acid level with CLUSTALW. The entropy-based conservation maps for CsoS2 binding to CsoS1 hexamers were generated by the AL2CO programme implemented in ChimeraX[56]. The sequence logo of alignments for the β4 of CsoS1 and the V/L/ITG motifs of CsoS2 were analysed using Weblogo 2.8.2.

Prochlorococcus MED4 growth and α-carboxysome purification

The α-carboxysomes were isolated from *Prochlorococcus* MED4 cells, which were grown in Pro99 medium[57] supplemented with 5 mM $NaHCO_3$ and 6 mM HEPES (pH 7.5) at 23 °C and 30 μmol photon m^{-2} s^{-1}. Pro99 medium was prepared with natural sea water supplemented with 800 μM NH_4Cl, 50 μM Na_2HPO_4 and trace metals (0.008 μM $ZnSO_4{\cdot}7H_2O$, 0.005 μM $CoCl_2{\cdot}6H_2O$, 0.090 μM $MnCl_2{\cdot}4H_2O$, 0.003 μM $Na_2MoO_4{\cdot}2H_2O$, 0.010 μM Na_2SeO_3 and 0.010 μM $NiCl_2{\cdot}6H_2O$). Cell numbers were measured using flow cytometry (BD Accuri C6 Plus). The methods used to purify α-carboxysomes from *Prochlorococcus* MED4 cells were adapted from a previous study[29]. In brief, the cells in a 12 l culture were collected by centrifugation at 10,000*g*, resuspended in 40 ml BEMB-SW buffer (10 mM bicine, 1 mM EDTA, 10 mM $MgCl_2{\cdot}6H_2O$, and dissolved in natural sea water, pH 8.5, supplemented with 0.6 mM PMSF and 20 mM $NaHCO_3$). The collected cells were ruptured using a bead beater (mini-beadbeater-16, bioSpec) with glass beads (acid-washed, Sigma Aldrich, 0.1 mm diameter) at the oscillation frequency of 3,450 rpm for 12 min. The supernatant was collected by centrifugation at 3,500*g* for 2 min at 4 °C and rotated at a speed of 16*g* at 4 °C with 80 μl of DNaseI (10 mg ml^{-1}, Sigma Aldrich) for 30 min. The proteinaceous supernatant was collected by centrifugation at 20,000*g* at 4 °C for 20 min, and 1% Triton X-100 was added and stirred for 30 min

at 4 °C to clean lipids. Crude α-carboxysomes were pelleted at 48,000 *g* for 30 min at 4 °C, resuspended in 1 ml of BEME-SW buffer and further purified by sucrose density gradient centrifugation (5, 10, 20, 40 and 60%) using a swing-rotor SW41Ti at a speed of 91,000*g* for 1 h. A yellowish opaque band between 40 and 60% was pipetted up, diluted in 15 ml of BEMB-SW buffer and pelleted by ultracentrifugation at 125,000*g* for 1 h at 4 °C. The concentrated α-carboxysomes were resuspended in 100 µl of BEMB-SW buffer and homogenized with 25 µl of L-lysine/$NaHCO_3$ (1 M) by rotation for 1 h at 4 °C. The final purified *Prochlorococcus* MED4 α-carboxysome samples in the supernatant were collected by centrifugation at 12,000*g* for 10 min at 4 °C.

SDS–PAGE analysis

Purified *Prochlorococcus* MED4 α-carboxysomes were mixed with 3x SDS sample buffer and heated at 95 °C for 10 min. In each well of a 12% SDS–PAGE gel, a 10 µl mixture was loaded for gel electrophoresis at 120 V for ~2 h. SDS–PAGE gels were stained with Coomassie blue and images were acquired using a Gel Dock system (Bio-Rad). Each protein band was excised from the gel with a scalpel and subjected to in-gel digestion with 20 ng μl^{-1} MS sequencing-grade trypsin (Sigma) in 50 mM NH_4HCO_3. Prepared samples were processed by the Hong Kong University of Science and Technology, Bioscience Research Facility (HKUST BioCRF) with a MALDI TOF/TOF mass spectrometer incorporating mass fingerprinting (PMF) and MS/MS peptide fragmentation. Proteins were identified by matching the theoretical fragmentation patterns in the MASCOT database with those in Uniprot. The hits with Mascot peptide scores greater than 16 were recorded.

Liquid chromatography mass spectrometry analysis

Freshly purified *Prochlorococcus* MED4 α-carboxysomes were subsequently sent to HKUST BioCRF for proteomic analysis using an Orbitrap Eclipse Tribrid mass spectrometer (Thermo Fisher). The samples were subjected to in-solution tryptic digestion, which was performed at 83 °C using a PreOmics iST Sample Preparation Kit 96× on an automated PreON machine. Proteome Discoverer software was used for database searching by HKUST BioCRF. The Score sequest HT displays the fitness of the detected proteins from purified α-carboxysomes to the theoretical spectral setting for the Orbitrap Eclipse mass spectrometer. Proteins with a score >2 were considered to indicate a 'good fit'. Abundance was calculated using the average of the three most intense peptides. The relative abundance of each protein from the isolated α-carboxysomes was normalized to that of the least abundant protein CsoS4B. Detailed information on the identified proteins is listed in Supplementary Table 1.

Cryo-EM sample preparation and data collection

A continuous carbon-film Cu 300-mesh grid (Quantifoil R2/1) was glow-discharged using PELCO easiGlow and subsequently coated with 0.1 mg ml^{-1} polylysine for 5 min. Purified α-carboxysome samples (3.5 µl each) were loaded on the grid and incubated for 1 min, and the process was repeated twice. The grids were blotted for 4 s with a blot force of 0 and a wait time of 60 s, and then plunged into liquid ethane using a Vitrobot IV system (Thermo Fisher). The cryo-EM datasets of α-carboxysomes were collected using a 300 kV Titan Krios electron microscope equipped with a Gatan K3 direct electron detector (Thermo Fisher) at the Biological Cryo-EM Centre of HKUST. Automated data acquisition was performed with EPU software (Thermo Fisher) at a magnification of 81,000, yielding a pixel size of 1.06 Å. A total of 23,400 movies (40 frames, each 0.11 s, total dose 50 $e^- Å^{-2}$) were collected, with defocus values ranging from −1.0 to −2.0 µm. Approximately half of the images have no carboxysome particles.

Cryo-EM data processing

All cryo-EM data processing steps were performed in Relion (3.1)[58]. All movies were corrected for beam-induced drift by aligning and averaging the individual frames using MOTIONCORR[59]. The contrast transfer function (CTF) parameters for drift-corrected micrographs were estimated using CTFFIND4 (ref. 60). In total, ~33,828 intact α-carboxysome particles, all of which displayed an icosahedral-like shape, were manually picked from 23,400 micrographs. The particles were extracted with a box size of 576 pixels and binned two-fold, yielding a pixel size of 2.12 Å. After several rounds of 2D classification, 32,044 good particles were selected for further 3D classification. Initial 3D classification and refinement without symmetry could only obtain a blurred and discontinuous cryo-EM map of only ~36 Å resolution. This cryo-EM map has no distinguishable features, indicating the apparent heterogeneity of the intact α-carboxysomes, which possess varying diameters ranging from 82 to 92 nm. Several rounds of 3D classification with icosahedral symmetry yielded the best class of 3,634 particles, yielding a reconstruction of the intact α-carboxysome at an overall 7.5 Å resolution. Applying the shell mask and icosahedral symmetry further improved the resolution of the intact shell to 6.2 Å.

To further improve the resolution of the shell, we applied the block-based reconstruction method as reported previously[42], in which the intact carboxysome is segmented into small blocks centred at the 12 shell vertices. In brief, we calculated the *x*/*y*/*z* positions of the 12 vertices relative to the centre of the cryo-EM map of intact α-carboxysomes in ChimeraX. Afterwards, we extracted the 12 vertices from a class of 11,424 particles at a box size of 686 pixels with the expansion of the icosahedral symmetry using the Relion command 'relion_star_handler'. The resulting 136,547 particles corresponding to the 12 vertices were subjected to several rounds of 3D classification by imposing different symmetries. Most of the classes yielded blurred maps of rather low resolution, which cannot be applied to build atomic models. The results also indicate that even if the intact carboxysome shell is segmented into small blocks, there is still some heterogeneity in these blocks. Nevertheless, via extensive trials of 3D classification, a cryo-EM map of the best class harbouring 28,093 particles showed clear and discernible features at the shell vertex, which were applied to 3D refinement by imposing *C5* symmetry, yielding a 4.2-Å-resolution map for the shell vertex as estimated on the basis of the gold-standard Fourier shell correlation (FSC) with the 0.143 criterion[61]. This map was calculated from the average of relatively homogeneous blocks of the α-carboxysome shell, which might not fully represent the authentic assembly pattern of all protein components in the intact α-carboxysome shell.

For the internal spheres of α-carboxysomes, we also applied the methods of block-based reconstruction and focused classification, but failed to improve the resolution, indicating the heterogeneity of the arrangements of internal cargo enzymes. Instead, we applied masked classification and refinement for the internal spheres with different symmetries. Finally a reconstruction of the 15.0-Å-resolution map of the internal spheres was obtained by imposing the icosahedral symmetry. All local resolution maps were generated using Relion 3.1. The cryo-EM parameters, data collection and refinement statistics are summarized in Supplementary Table 2.

Model building and refinement

All atomic models for the CsoS1 hexamer, the CsoS4A and CsoS4B pentamers, the scaffolding protein CsoS2, and large and small subunits of RuBisCO were generated using AlphaFold2 (ref. 62) and are listed in Supplementary Fig. 3, with the confidence scores (pLDDTs) labelled. To construct the shell model, CsoS1 hexamers and CsoS4A pentamers were automatically fitted in an asymmetric unit of the shell using ChimeraX. Afterwards, the cryo-EM map of an asymmetric unit was extracted in ChimeraX on the basis of the fitted initial model. The model of an asymmetric unit was then manually refined in Coot[63] and subjected to real-space refinement in Phenix[64].

Although the AlphaFold prediction of the full-length disordered CsoS2 protein showed an overall pLDDT confidence score of 65 (Supplementary Fig. 3), the models of the core structures of each M and

C repeat, especially the α-helices and β-sheets, have better confidence scores. Therefore, the predicted models of each individual M and C repeat could be clearly fitted to the additional densities of the full-length CsoS2 on the inner surface of the shell. After several rounds of manual adjustment in Coot with the assistance of the conserved V/L/ITG motifs binding to His75 of BMC-H, as identified recently[30–32], we eventually assigned the residues from each M and C repeat of CsoS2. The eight-residue linkers that separated the M and C repeats were manually built into the cryo-EM density at a relatively poor resolution but with a reasonable main-chain length. Residues in some of the linkers that had weaker densities were not included in the final CsoS2 models. All side chains were manually adjusted to avoid steric clashes with neighbouring residues. The short-form CsoS2 model was built by docking the model of full-length CsoS2, except for the M1, C2 and C3 regions, into the corresponding map. Finally, we successfully built an atomic model of CsoS2 in two forms. The initial model was subsequently applied to real-space refinement in Phenix[64]. The geometries of all final models were evaluated using Molprobity[65]. Details of the model geometry statistics are listed in Supplementary Table 2. The model of the intact α-carboxysome shell was reconstructed in ChimeraX by combining 60 copies of an asymmetric unit and imposing icosahedral symmetry.

For the internal density, copies of the predicted RuBisCO structure were placed in the corresponding maps of the different layers and fitted automatically in ChimeraX. We manually adjusted the position of RuBisCO if major clashes were observed between adjacent RuBisCO molecules.

All structural figures were generated in PyMoL (www.pymol.org/pymol), Chimera[66] or ChimeraX.

Reporting summary

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

Data availability

All cryo-EM maps have been deposited at the Electron Microscopy Data Bank (EMDB). The accession codes are as follows: shell vertex with *C5* symmetry, EMD-37902; intact shell with icosahedral symmetry, EMD-38544; internal RuBisCOs with icosahedral symmetry, EMD-38543; and intact α-carboxysome with icosahedral symmetry, EMD-37903. The model of the shell vertex has been deposited to the Protein Data Bank under the accession code 8WXB. Mass spectroscopy data are deposited on Figshare at https://doi.org/10.6084/m9.figshare.25239463.v2 (ref. 67). The consecutive genes were blasted and retrieved from the KEGG database (https://www.kegg.jp/kegg/). Source data are provided with this paper.

References

1. Fuchs, G. Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? *Annu. Rev. Microbiol.* **65**, 631–658 (2011).
2. Bar-On, Y. M. & Milo, R. The global mass and average rate of rubisco. *Proc. Natl Acad. Sci. USA* **116**, 4738–4743 (2019).
3. Erb, T. J. & Zarzycki, J. A short history of RubisCO: the rise and fall (?) of Nature's predominant CO_2 fixing enzyme. *Curr.Opin. Biotechnol.* **49**, 100–107 (2018).
4. Whitney, S. M., Houtz, R. L. & Alonso, H. Advancing our understanding and capacity to engineer nature's CO_2-sequestering enzyme, Rubisco. *Plant Physiol.* **155**, 27–35 (2011).
5. Badger, M. R. & Price, G. D. CO_2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *J. Exp. Bot.* **54**, 609–622 (2003).
6. Kerfeld, C. A. & Melnicki, M. R. Assembly, function and evolution of cyanobacterial carboxysomes. *Curr. Opin. Plant Biol.* **31**, 66–75 (2016).
7. Liu, L. N. Advances in the bacterial organelles for CO_2 fixation. *Trends Microbiol.* **30**, 567–580 (2022).
8. Price, G. D., Badger, M. R., Woodger, F. J. & Long, B. M. 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. Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F. & Sutter, M. Bacterial microcompartments. *Nat. Rev. Microbiol.* **16**, 277–290 (2018).
10. Sutter, M., Melnicki, M. R., Schulz, F., Woyke, T. & Kerfeld, C. A. A catalog of the diversity and ubiquity of bacterial microcompartments. *Nat. Commun.* **12**, 3809 (2021).
11. Rae, B. D., Long, B. M., Badger, M. R. & Price, G. D. 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).
12. Dou, Z. et al. 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).
13. Faulkner, M. et al. Molecular simulations unravel the molecular principles that mediate selective permeability of carboxysome shell protein. *Sci. Rep.* **10**, 17501 (2020).
14. Kinney, J. N., Axen, S. D. & Kerfeld, C. A. Comparative analysis of carboxysome shell proteins. *Photosynth. Res.* **109**, 21–32 (2011).
15. Liu, X. Y. et al. Structures of cyanobacterial bicarbonate transporter SbtA and its complex with PII-like SbtB. *Cell Discov.* **7**, 63 (2021).
16. Wang, C. et al. Structural mechanism of the active bicarbonate transporter from cyanobacteria. *Nat. Plants* **5**, 1184–1193 (2019).
17. Long, B. M., Forster, B., Pulsford, S. B., Price, G. D. & Badger, M. R. Rubisco proton production can drive the elevation of CO_2 within condensates and carboxysomes. *Proc. Natl Acad. Sci. USA* **118**, e2014406118 (2021).
18. Cabello-Yeves, P. J. et al. α-cyanobacteria possessing form IA RuBisCO globally dominate aquatic habitats. *ISME J.* **16**, 2421–2432 (2022).
19. Whitehead, L., Long, B. M., Price, G. D. & Badger, M. R. Comparing the in vivo function of alpha-carboxysomes and beta-carboxysomes in two model cyanobacteria. *Plant Physiol.* **165**, 398–411 (2014).
20. Long, B. M., Badger, M. R., Whitney, S. M. & Price, G. D. Analysis of carboxysomes from *Synechococcus* PCC7942 reveals multiple Rubisco complexes with carboxysomal proteins CcmM and CcaA. *J. Biol. Chem.* **282**, 29323–29335 (2007).
21. Kinney, J. N., Salmeen, A., Cai, F. & Kerfeld, C. A. Elucidating essential role of conserved carboxysomal protein CcmN reveals common feature of bacterial microcompartment assembly. *J. Biol. Chem.* **287**, 17729–17736 (2012).
22. Sun, H. et al. Complex structure reveals CcmM and CcmN form a heterotrimeric adaptor in beta-carboxysome. *Protein Sci.* **30**, 1566–1576 (2021).
23. Cai, F. et al. Advances in understanding carboxysome assembly in *Prochlorococcus* and *Synechococcus* implicate CsoS2 as a critical component. *Life* **5**, 1141–1171 (2015).
24. Oltrogge, L. M. et al. Multivalent interactions between CsoS2 and Rubisco mediate alpha-carboxysome formation. *Nat. Struct. Mol. Biol.* **27**, 281–287 (2020).
25. Faulkner, M. et al. Direct characterization of the native structure and mechanics of cyanobacterial carboxysomes. *Nanoscale* **9**, 10662–10673 (2017).
26. 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).

27. Wang, H. et al. Rubisco condensate formation by CcmM in beta-carboxysome biogenesis. *Nature* **566**, 131–135 (2019).
28. Cameron, J. C., Wilson, S. C., Bernstein, S. L. & Kerfeld, C. A. Biogenesis of a bacterial organelle: the carboxysome assembly pathway. *Cell* **155**, 1131–1140 (2013).
29. Roberts, E. W., Cai, F., Kerfeld, C. A., Cannon, G. C. & Heinhorst, S. Isolation and characterization of the *Prochlorococcus* carboxysome reveal the presence of the novel shell protein CsoS1D. *J. Bacteriol.* **194**, 787–795 (2012).
30. Oltrogge, L. M., Chen, A. W., Chaijarasphone, T., Turnšek, J. B. & Savage, D. F. α-carboxysome size is controlled by the disordered scaffold protein CsoS2. *Biochemistry* **63**, 219–229 (2024).
31. Turnšek, J. B., Oltrogge, L. M. & Savage, D. F. Conserved and repetitive motifs in an intrinsically disordered protein drive α-carboxysome assembly. Preprint at *bioRxiv* https://doi.org/10.1101/2023.07.08.548221 (2023).
32. Ni, T. et al. Intrinsically disordered CsoS2 acts as a general molecular thread for alpha-carboxysome shell assembly. *Nat. Commun.* **14**, 5512 (2023).
33. Ni, T. et al. Structure and assembly of cargo Rubisco in two native alpha-carboxysomes. *Nat. Commun.* **13**, 4299 (2022).
34. Evans, S. L. et al. Single-particle cryo-EM analysis of the shell architecture and internal organization of an intact alpha-carboxysome. *Structure* **31**, 677–688 (2023).
35. Metskas, L. A. et al. Rubisco forms a lattice inside alpha-carboxysomes. *Nat. Commun.* **13**, 4863 (2022).
36. Xia, L. Y. et al. Molecular basis for the assembly of RuBisCO assisted by the chaperone Raf1. *Nat. Plants* **6**, 708–717 (2020).
37. Li, Q., Jiang, Y. L., Xia, L. Y., Chen, Y. & Zhou, C. Z. Structural insights into cyanobacterial RuBisCO assembly coordinated by two chaperones Raf1 and RbcX. *Cell Discov.* **8**, 93 (2022).
38. Tan, Y. Q. et al. Structure of a minimal alpha-carboxysome-derived shell and its utility in enzyme stabilization. *Biomacromolecules* **22**, 4095–4109 (2021).
39. Sutter, M. et al. Structure of a synthetic beta-carboxysome shell. *Plant Physiol.* **181**, 1050–1058 (2019).
40. Sutter, M., Greber, B., Aussignargues, C. & Kerfeld, C. A. Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science* **356**, 1293–1297 (2017).
41. Ulloa, O. et al. The cyanobacterium *Prochlorococcus* has divergent light-harvesting antennae and may have evolved in a low-oxygen ocean. *Proc. Natl Acad. Sci. USA* **118**, e2025638118 (2021).
42. Zhu, D. et al. Pushing the resolution limit by correcting the Ewald sphere effect in single-particle Cryo-EM reconstructions. *Nat. Commun.* **9**, 1552 (2018).
43. Partensky, F., Hess, W. R. & Vaulot, D. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol. Mol. Biol. Rev.* **63**, 106–127 (1999).
44. Cai, L., Li, H., Deng, J., Zhou, R. Q. & Zeng, Q. Biological interactions with *Prochlorococcus*: implications for the marine carbon cycle. *Trends Microbiol.* **32**, 280–291 (2024).
45. Cai, F. et al. The structure of CcmP, a tandem bacterial microcompartment domain protein from the beta-carboxysome, forms a subcompartment within a microcompartment. *J. Biol. Chem.* **288**, 16055–16063 (2013).
46. Tanaka, S. et al. Atomic-level models of the bacterial carboxysome shell. *Science* **319**, 1083–1086 (2008).
47. Wheatley, N. M., Sundberg, C. D., Gidaniyan, S. D., Cascio, D. & Yeates, T. O. Structure and identification of a pterin dehydratase-like protein as a ribulose-bisphosphate carboxylase/oxygenase (RuBisCO) assembly factor in the α-carboxysome. *J. Biol. Chem.* **289**, 7973–7981 (2014).
48. Sun, Y. et al. Decoding the absolute stoichiometric composition and structural plasticity of alpha-carboxysomes. *mBio* **13**, e0362921 (2022).
49. Liu, J., Xing, W. Y., Liu, B. & Zhang, C. C. Three-dimensional coordination of cell-division site positioning in a filamentous cyanobacterium. *PNAS Nexus* **2**, pgac307 (2023).
50. Greber, B. J., Sutter, M. & Kerfeld, C. A. The plasticity of molecular interactions governs bacterial microcompartment shell assembly. *Structure* **27**, 749–763 (2019).
51. Blikstad, C. et al. Identification of a carbonic anhydrase-Rubisco complex within the alpha-carboxysome. *Proc. Natl Acad. Sci. USA* **120**, e2308600120 (2023).
52. Chaijarasphong, T. et al. Programmed ribosomal frameshifting mediates expression of the alpha-carboxysome. *J. Mol. Biol.* **428**, 153–164 (2016).
53. Iancu, C. V. et al. Organization, structure, and assembly of alpha-carboxysomes determined by electron cryotomography of intact cells. *J. Mol. Biol.* **396**, 105–117 (2010).
54. Chen, T. et al. Engineering α-carboxysomes into plant chloroplasts to support autotrophic photosynthesis. *Nat. Commun.* **14**, 2118 (2023).
55. Li, T. et al. Reprogramming bacterial protein organelles as a nanoreactor for hydrogen production. *Nat. Commun.* **11**, 5448 (2020).
56. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
57. Moore, L. R. et al. Culturing the marine cyanobacterium *Prochlorococcus*. *Limnol. Oceanogr. Methods* **5**, 353–362 (2007).
58. Scheres, S. H. W. Amyloid structure determination in RELION-3.1. *Acta Crystallogr. D* **76**, 94–101 (2020).
59. Li, X. et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).
60. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
61. Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853–854 (2012).
62. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
63. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
64. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
65. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
66. Pettersen, E. F. et al. UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
67. Zhou, R. Proteomic analysis for freshly purified *Prochlorococcus* MED4 carboxysome was performed via Orbitrap Eclipse Tribrid Mass Spectrometry (Thermo Scientific). *Figshare* https://doi.org/10.6084/m9.figshare.25239463.v2 (2024).

Acknowledgements

All cryo-EM datasets were collected at the Biological Cryo-EM Center at HKUST. This work was supported by grants from the Strategic Priority Research Program of the Chinese Academy of Sciences (http://www.cas.cn; Precision Seed Design and Breeding, grant no. XDA24020302, Y.-L.J.), the Research Grants Council of the Hong Kong Special Administrative Region, China (https://www.ugc.edu.hk; grant no. HKUST C6012-22GF, Q.Z.), the Environment and Conservation Fund (https://www.ugc.edu.hk; grant ECF project 128/2020, Q.Z.), the Strategic Priority Research Program of the Chinese Academy of Sciences (http://www.cas.cn; grant XDB37020301, C.-Z.Z.), and

the National Natural Science Foundation of China (http://www.nsfc.gov.cn; grants 32241025 to C.-Z.Z. and 32171198 to Y.-L.J.), the Anhui Provincial Natural Science Foundation (http://kjt.ah.gov.cn; 2108085J14, Y.-L.J.) and the Key R&D Projects of Anhui Province (http://kjt.ah.gov.cn; 2022l07020034, Y.-L.J.). Y.-L.J. thanks the Youth Innovation Promotion Association of the Chinese Academy of Sciences for support (membership no. 2020452).

Author contributions

Y.-L.J., Q.Z, C.-Z.Z. and Y.C. conceived, designed and supervised this project. R.-Q.Z., H.L., W.-W.K. and J.-X.D. performed the *Prochlorococcus* growth experiment. R.-Q.Z. performed α-carboxysome purification, biochemical assays, cryo-EM sample preparation and data collection. Y.-L.J. and P.H. conducted cryo-EM data processing, structure determination and model building. Y.-L.J., R.-Q.Z., C.-Z.Z. and Q.Z. wrote and revised the manuscript with input from all authors. All authors discussed the data and read the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

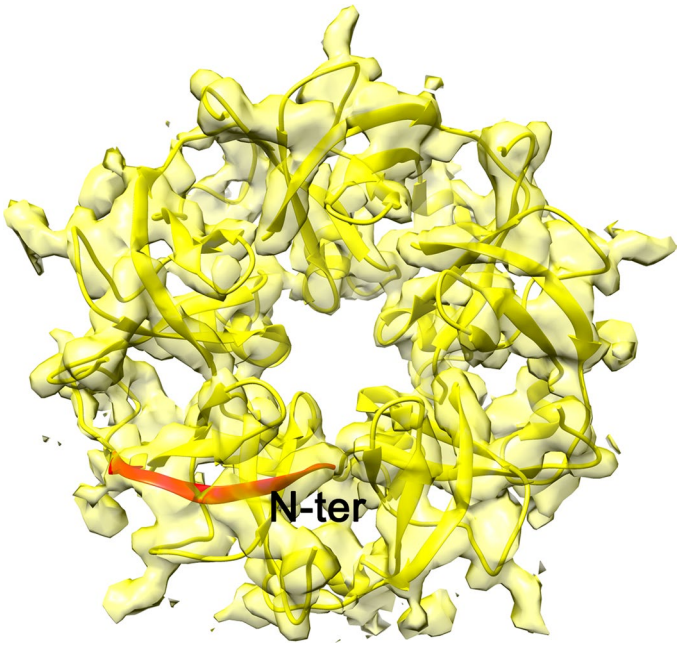
Extended data is available for this paper at https://doi.org/10.1038/s41477-024-01660-9.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41477-024-01660-9.

Correspondence and requests for materials should be addressed to Yong-Liang Jiang, Cong-Zhao Zhou or Qinglu Zeng.

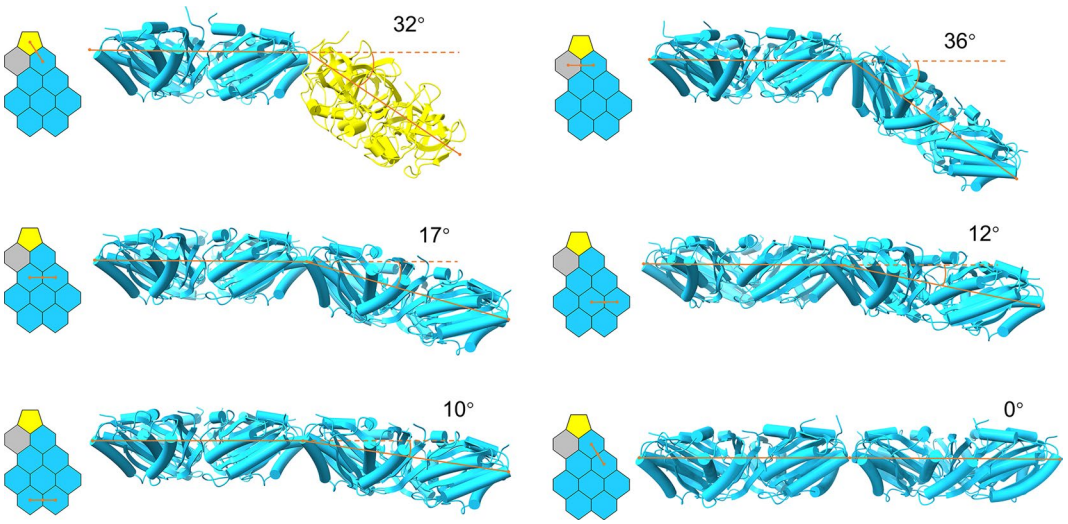
Peer review information *Nature Plants* thanks Spencer Whitney and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s41477-024-01660-9



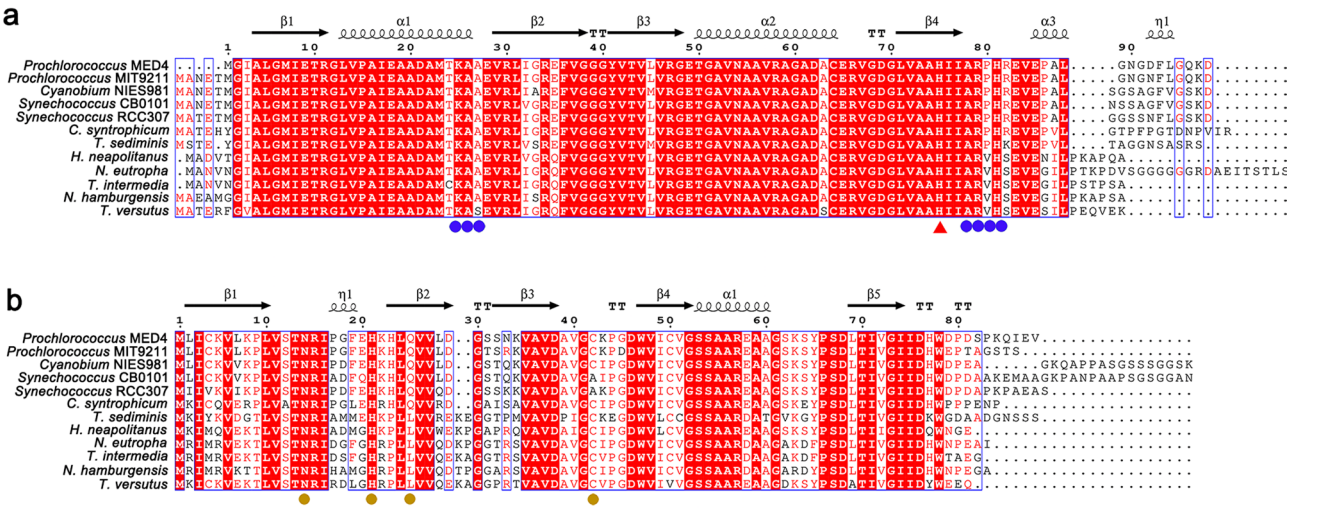
Extended Data Fig. 1 | Fitting of the CsoS4A pentamer to the Cryo-EM map at the shell vertex. The structure of CsoS4A pentamer is shown as a yellow cartoon with the N-terminal ten residues (labeled N-ter) colored red.

https://doi.org/10.1038/s41477-024-01660-9



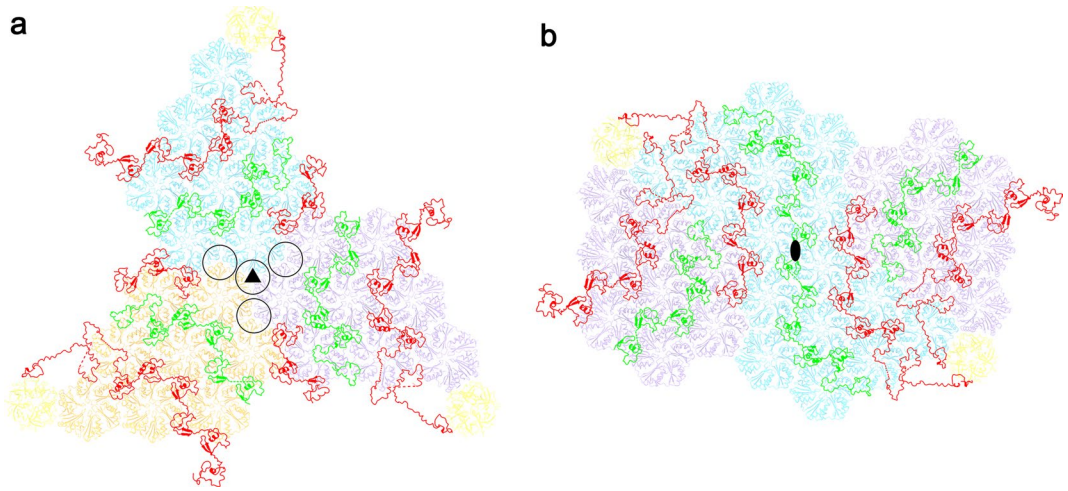
Extended Data Fig. 2 | Overview of the distinct interfaces among hexamers and pentamers. The structures of the hexamers (blue) and pentamers (yellow) are shown as cartoons, with a pictogram showing their location on the shell. The tilt angles of the hexamer-hexamer or hexamer-pentamer interfaces are labeled.

https://doi.org/10.1038/s41477-024-01660-9



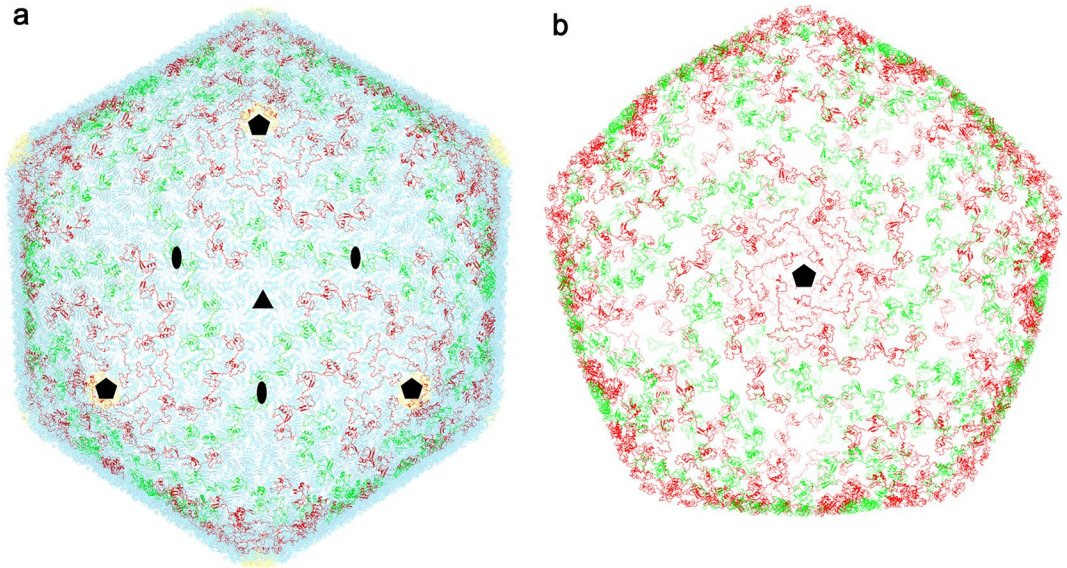
Extended Data Fig. 3 | Multiple sequence alignment of the α-carboxysome (a) BMC-H and (b) BMC-P proteins. The BMC-H and BMC-P sequences of twelve organisms that harbor α-carboxysomes were retrieved from the UniProt protein database. The KAA and ARPH motifs from BMC-H proteins and the interface residues from BMC-P proteins are labeled with blue and brown solid circles, respectively. The conserved His75 of BMC-H that binds to the V/L/ITG motif of CsoS2 is marked with a red triangle.

https://doi.org/10.1038/s41477-024-01660-9



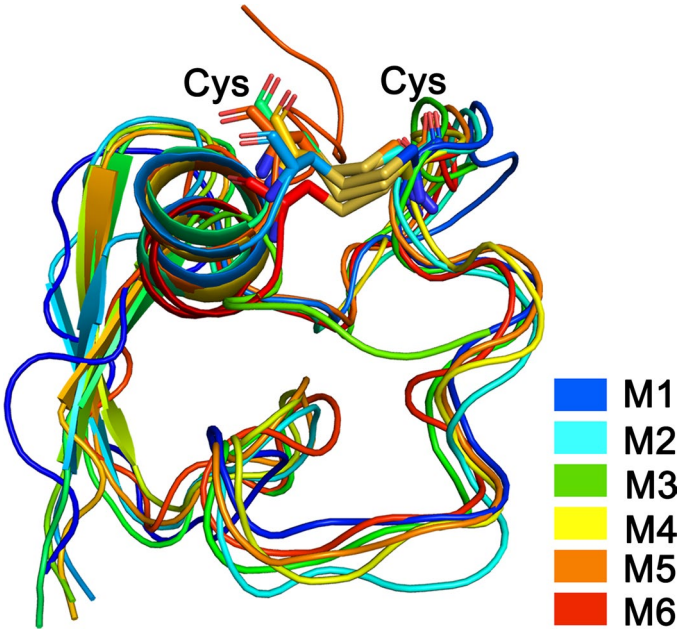
Extended Data Fig. 4 | Local arrangement of CsoS2 subunits on the inner surface of the shell viewed from the (a) threefold and (b) twofold axes of the α-carboxysome shell. The structures of the shell patches are shown as semitransparent cartoons, with each asymmetric unit colored differently. The structures of the long and short forms of CsoS2 are colored red and green, respectively. The three- and twofold axes of the α-carboxysome shell are indicated by black solid triangles and ovals, respectively. The positions that are vacant for CsoS2 binding are marked with four black circles in **a**.

https://doi.org/10.1038/s41477-024-01660-9



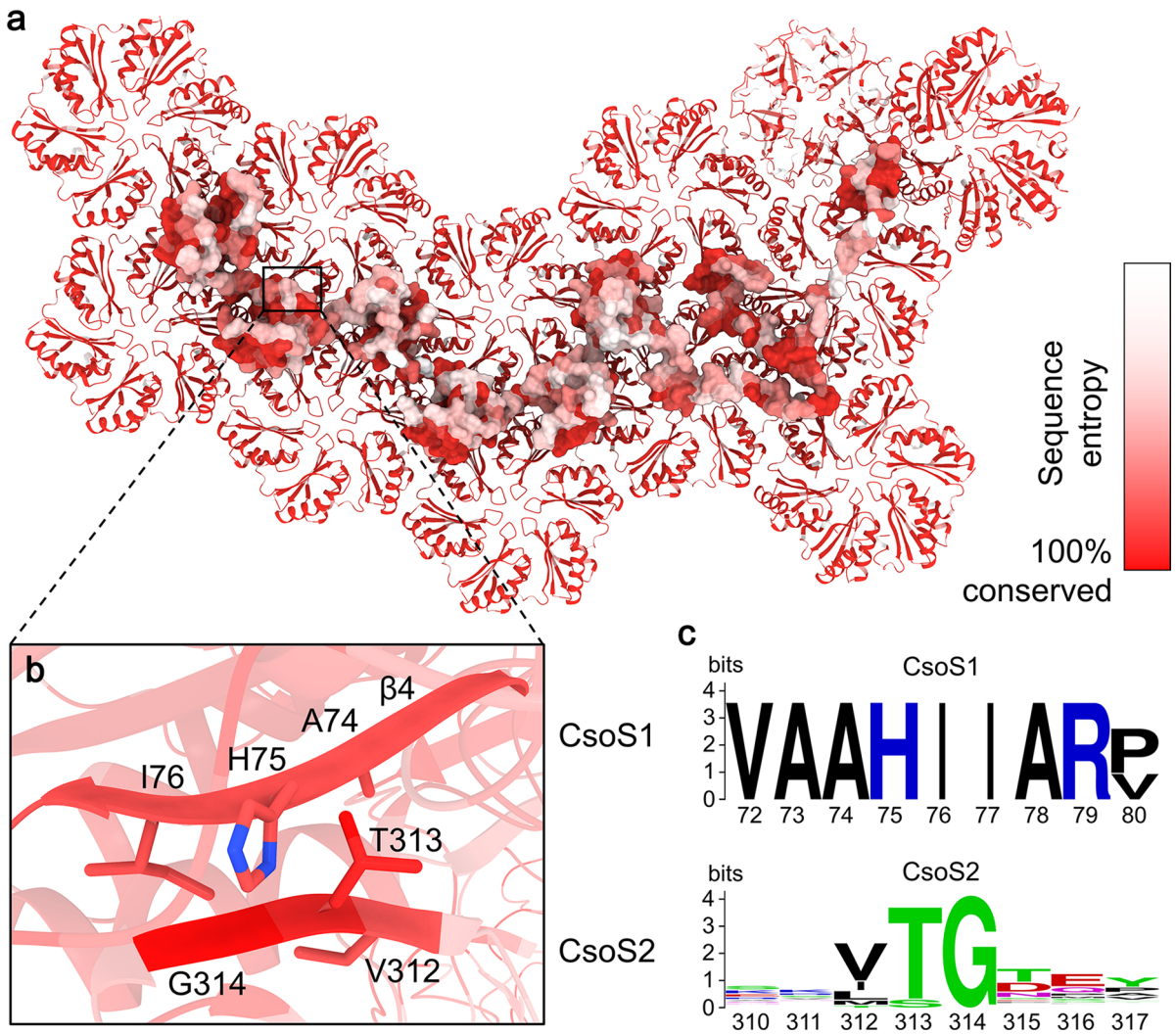
Extended Data Fig. 5 | A global view of the arrangement of CsoS2 subunits on the inner surface of the intact icosahedral-like shell viewed from the (a) threefold and (b) fivefold axes of the shell. The structure of the icosahedral-like shell is shown as a semitransparent cartoon, with BMC-H hexamers and BMC-P pentamers colored blue and yellow, respectively. The structures of the long and short forms of CsoS2 are colored red and green, respectively. The twofold, threefold and fivefold axes of the α-carboxysome shell are indicated by black solid ovals, triangles and pentagons, respectively.

https://doi.org/10.1038/s41477-024-01660-9



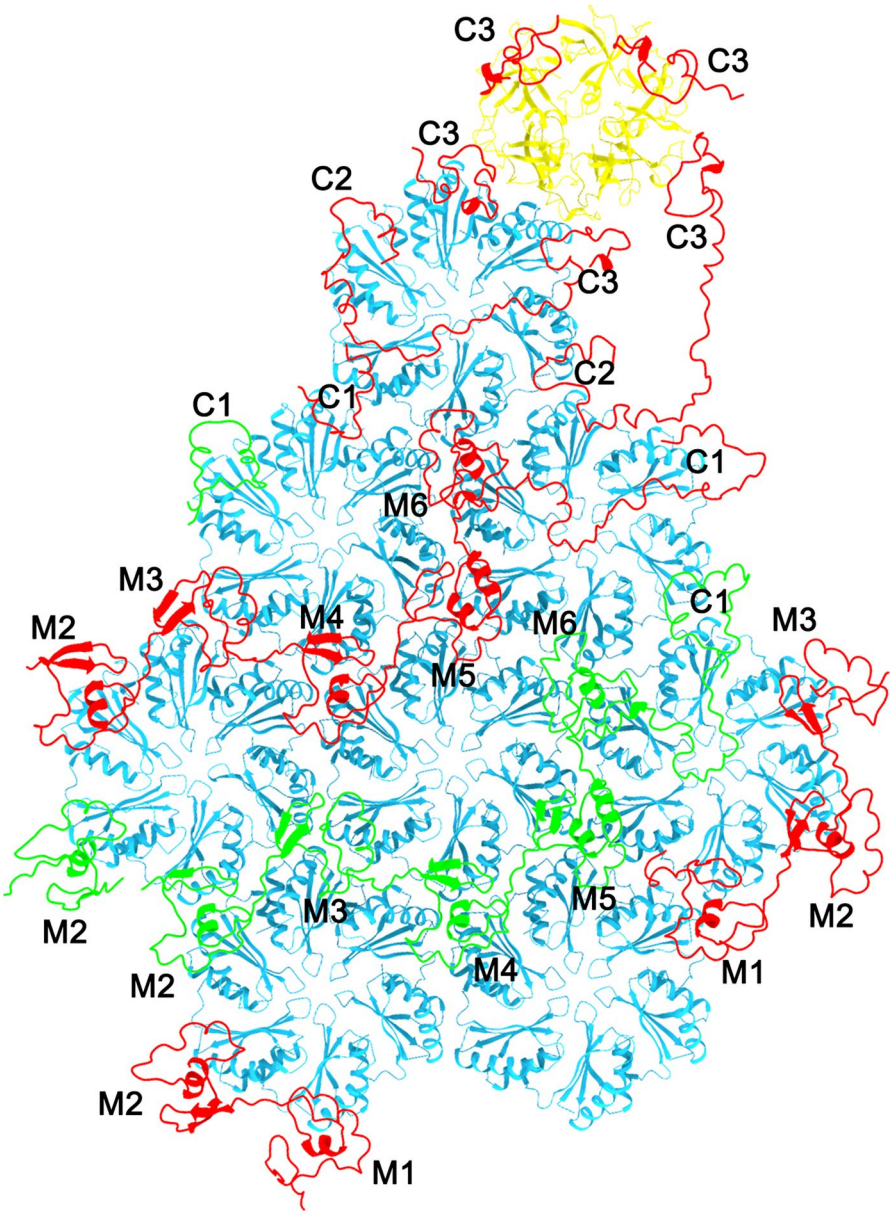
Extended Data Fig. 6 | Structural comparison of the six M repeats of CsoS2. The six M repeats are shown as cartoons and are colored differently, as shown on the right. A pair of cysteine residues in each of the M repeats is shown as sticks.

https://doi.org/10.1038/s41477-024-01660-9



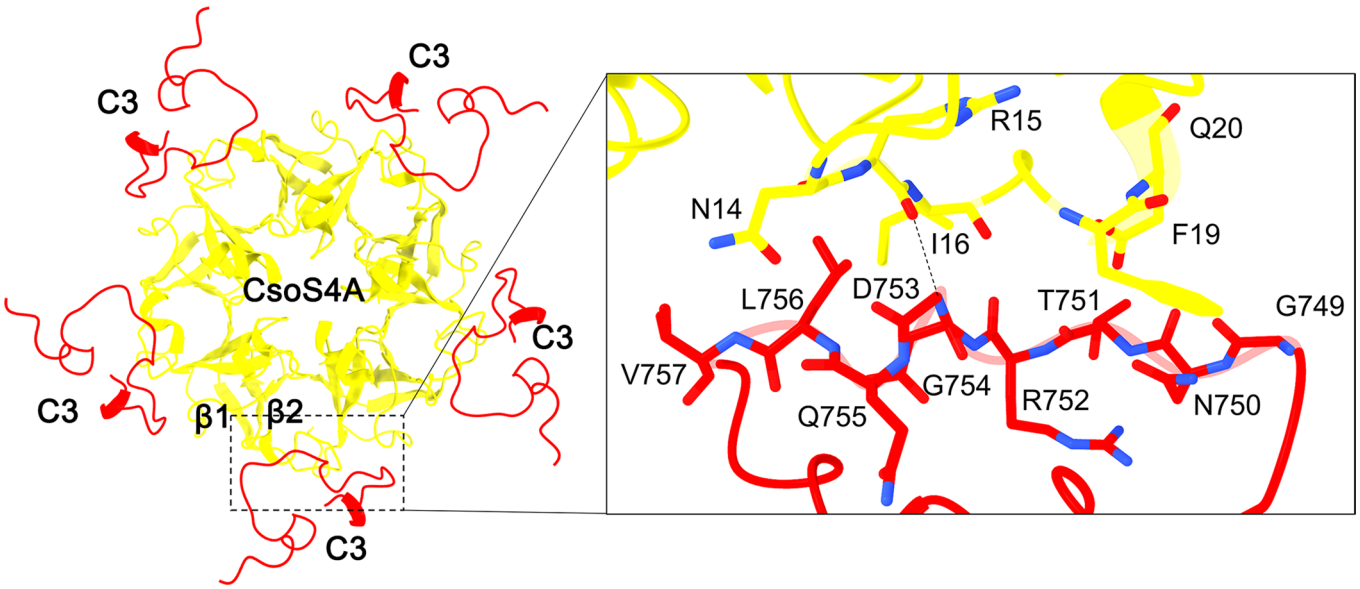
Extended Data Fig. 7 | Surface representation of the CsoS1-CsoS2 interface. **a**, Overall view of the conservation profile of the interface between the shell protein CsoS1 and the scaffolding protein CsoS2. The structures of CsoS1 and CsoS2 are shown in cartoon and surface, respectively. Coloring is by calculated amino acid conservation entropy (red, 100% conservation). **b**, Zoomed in view of the interface between β4 of the shell protein CsoS1 and the V/L/ITG motif of the scaffolding protein CsoS2. **c**, The sequence alignments were depicted using Weblogo. The sequence logo shows the conservation of the residues at the interface between CsoS1 and CsoS2.

https://doi.org/10.1038/s41477-024-01660-9



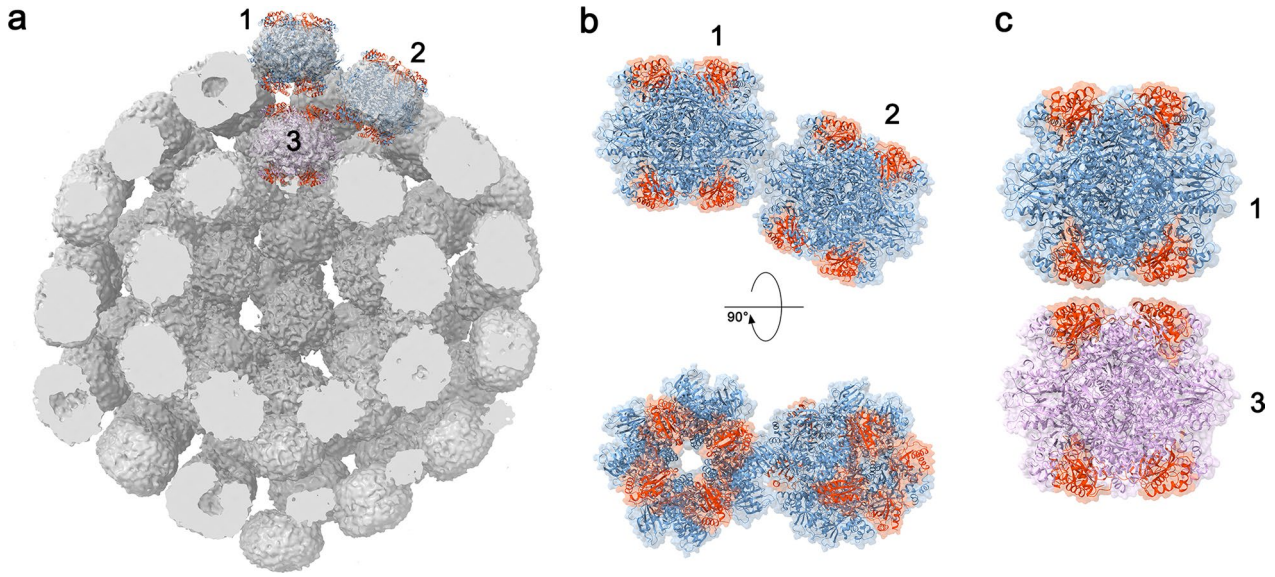
Extended Data Fig. 8 | Cartoon representation of CsoS2 subunits binding to an asymmetric unit of the icosahedral-like shell. The CsoS1 hexamers, the CsoS4A pentamer, and the CsoS2 subunits are shown as cartoons and are colored blue, yellow, and red/green, respectively. The individual M and C repeats of CsoS2 are numbered sequentially.

https://doi.org/10.1038/s41477-024-01660-9



Extended Data Fig. 9 | The interaction of the C-terminus of CsoS2 binding to the CsoS4A pentamer. The C3 region of CsoS2 and CsoS4A are shown as cartoons and are colored red and yellow, respectively. The detailed interaction networks are shown in the inset, with the interface residues shown as sticks.

https://doi.org/10.1038/s41477-024-01660-9



Extended Data Fig. 10 | Internal arrangement of RuBisCO within the outer and middle layers of a-carboxysomes. a, Slab section of the overall density in the outer and middle layers of the α-carboxysome. Three RuBisCOs were fitted to the cryo-EM density and are labeled 1, 2, and 3. Two RuBisCOs from the outer layer are colored blue, whereas the third one from the middle layer is colored pink. **b, c**, Cartoon representations of two adjacent RuBisCOs forming lateral **(b)** or longitudinal **(c)** interfaces, shown as cartoons through the transparent surface.

nature portfolio

Corresponding author(s): Qinglu Zeng, Cong-Zhao Zhou, Yong-Liang Jiang

Last updated by author(s): 02/20/2024

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	Proteomic data were acquired by Orbitrap Eclipse Tribrid Mass Spectrometer. Automated Cryo-EM data acquisition was performed by a 300 kV Titan Krios electron microscopy equipped with a Gatan K3 electron detector and EPU software.
Data analysis	Cryo-EM data were processed using MotionCor2, CTFFIND 4, and RELION 3.1. The proteomic data were analyzed by Proteome DiscovererTM. Consecutive genes were blasted using KEGG synteny and phylogeny tree was drawn by the NGphylogeny. The gene clusters were plotted using ChiPlot. The atomic models were built and refined by COOT 0.8.9.2 and PHENIX 1.18.2. All structures were analyzed by Chimera 1.14, ChimerX 1.5, and Pymol 2.3.4. The protein structures were predicted by AlphaFold2. The sequence logo were generated by Weblogo 2.8.2.

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All cryo-EM maps have been deposited at the Electron Microscopy Data Bank. The accession codes are as follows: shell vertex with C5 symmetry, EMD-37902; intact shell with icosahedral symmetry, EMD-38544; internal RuBisCOs with icosahedral symmetry, EMD-38543; and intact α-carboxysome with icosahedral symmetry, EMD-37903. The model of the shell vertex has been deposited at the Protein Data Bank under the accession code of 8WXB. Mass spectroscopy data are deposited on Figshare (https://doi.org/10.6084/m9.figshare.25239463.v2). The consecutive genes were blasted and retrieved from the KEGG database (https://www.kegg.jp/kegg/). Source data are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	The sample size has been described in the methods and figure legends.
Data exclusions	No data was excluded.
Replication	All biochemical experiments were repeated for at least three times that showed similar results.
Randomization	n/a (The current work does not involve population study)
Blinding	n/a (The current work does not involve experimental subjects with opinions)

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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