The Structure of CcmP, a Tandem Bacterial Microcompartment Domain Protein from the β-Carboxysome, Forms a Subcompartment Within a Microcompartment

Fei Cai, Markus Sutter, Jeffrey C. Cameron, Desiree N. Stanley, James N. Kinney, and Cheryl A. Kerfeld

Background: CcmP is a hypothetical protein conserved among all β-cyanobacteria.

Results: CcmP is a β-carboxysome component; it forms a bilayered shell protein.

Conclusion: CcmP may facilitate flux of larger metabolites across the carboxysome shell.

Significance: It is the first structure of a β-carboxysome tandem BMC domain protein; phylogenetically, it represents a new type of microcompartment building block.

Cyanobacteria use solar energy to fix CO₂. Like many other autotrophic organisms, they use 1,5-ribulose bisphosphate carboxylase/oxygenase (Rubisco) to catalyze the first rate-limiting step of the Calvin-Benson-Bassham cycle. To survive under ambient CO₂ concentrations with this notoriously inefficient enzyme, cyanobacteria have evolved a CO₂-concentrating mechanism to enhance CO₂ fixation. The CO₂-concentrating mechanism is composed of multiple inorganic carbon transporters and a proteinaceous organelle, the carboxysome.

Carboxysomes encapsulate Rubisco with carbonic anhydrase in a protein shell. Based on gene arrangement, the type of encapsulated Rubisco, and associated proteins, carboxysomes are divided into two types, the cso- or α-carboxysomes and the ccm- or β-carboxysomes. α-Carboxysomes are found in the α-cyanobacteria (namely marine Synechococcus and Prochlorococcus), and β-carboxysomes are found in most other cyanobacteria, which grow in a much wider range of ecological niches.

The ccm genes encoding β-carboxysome proteins are dispersed in several locations in β-cyanobacterial genomes (Fig. 1A). The α- and β-carboxysome share two types of conserved shell proteins. The major shell constituents are the CsoS1 (α-carboxysome) and CcmK (β-carboxysome) proteins; these contain an ~80 amino acid domain called the bacterial microcompartment domain (BMC domain; pfam00936). CsoS1 and CcmK proteins form hexamers. They self-assemble into layers that are proposed to constitute the facets (1–5) of the apparent icosahedral carboxysome shell (6–8). The shell has been proposed to be composed of a single layer (2, 4, 5) or a double layer (9) of hexameric BMC proteins. Based on electron cryotomography studies, in a single layer model, ~40 hexamers would be required per facet to form the complete icosahedral shell of the average sized α-carboxysome of Synechococcus strain WH8102 (6). This number is estimated to be ~30 for the

Received for publication, January 25, 2013, and in revised form, April 4, 2013. Published, JBC Papers in Press, April 9, 2013, DOI 10.1074/jbc.M113.456897

Supported by National Science Foundation Grant EF1105897.

Supported by a postdoctoral fellowship from the Swiss National Science Foundation.

Supported by National Science Foundation Grant EF1105897.

Present address: Dir. of Biochemistry, University of California, San Francisco, CA 94158.

To whom correspondence should be addressed: United States Department of Energy Joint Genome Institute, 2800 Mitchell Dr., Walnut Creek, CA 94598. Tel.: 925-296-5691; Fax: 925-296-5752; E-mail: ckerfeld@lbl.gov.

The carboxysome is a bacterial organelle found in all cyanobacteria; it encapsulates CO₂ fixation enzymes within a protein shell. The most abundant carboxysome shell protein contains a single bacterial microcompartment (BMC) domain. We present in vivo evidence that a hypothetical protein (dubbed CcmP) encoded in all β-cyanobacterial genomes is part of the carboxysome. We show that CcmP is a tandem BMC domain protein, the first to be structurally characterized from a β-carboxysome. CcmP forms a dimer of tightly stacked trimers, resulting in a nanocompartment-containing shell protein that may weakly bind 3-phosphoglycerate, the product of CO₂ fixation. The trimers have a large central pore through which metabolites presumably pass into the carboxysome. Conserved residues surrounding the pore have alternate side-chain conformations suggesting that it can be open or closed. Furthermore, CcmP and its orthologs in β-cyanobacterial genomes form a distinct clade of shell proteins. Members of this subgroup are also found in numerous heterotrophic BMC-associated gene clusters encoding functionally diverse bacterial organelles, suggesting that the potential to form a nanocompartment within a microcompartment shell is widespread. Given that carboxysomes and architecturally related bacterial organelles are the subject of intense interest for applications in synthetic biology/metabolic engineering, our results describe a new type of building block with which to functionalize BMC shells.
Subcompartment within a Microcompartment Shell

α-carboxysome of Halothiobacillus neapolitanus (10). Both α- and β-types of carboxysome shells contain a minor component, CsoS4 or CcmL, respectively, that belongs to the EutN domain family (pfam03319); CsoS4 and CcmL form pentamers proposed to serve as the vertices of the icosahedron (1, 3).

The co-occurrence of pfam00936 and pfam03319 shell protein families is not limited to the genomes of cyanobacteria. A recent bioinformatic survey revealed that ~20% of sequenced bacterial genomes, from 13 bacterial phyla, have genes coding for both protein families, suggesting that the potential for forming compartments enclosed by carboxysome-like shells is widespread in the bacterial domain (11). The functions of some of these have been experimentally determined (12–15), but for most the only available functional information is derived from the annotations of the genes clustering with shell protein genes (3, 11). Collectively, carboxysomes and these (presumably) architecturally similar organelles are known as BMCs.

Encapsulation of enzymes within a protein shell necessitates flux of substrates and products across the shell. A key feature of the current model for carboxysome shell structure and function are the positively charged pores formed at the 6-fold axis of symmetry of the hexameric shell proteins. Their diameters range between 4 and 7 Å, sufficiently large for diffusion of bicarbonate into the carboxysome (2). However, current models do not explain how the bulkier substrate, 1,5-D-ribulose bisphosphate (RuBP), and the product of CO2 fixation, 3-phosphoglycerate (3PGA), enter the current model for carboxysome shell structure and function (3, 11). Collectively, carboxysomes and these (presumably) architecturally similar organelles are known as BMCs.

The first structure of a protein containing a fusion of BMC domains provided a clue (16). This protein, CsoS1D, is encoded just upstream of the cso operon in most α-cyanobacteria. CsoS1D forms trimers with a relatively large central pore (14 Å in diameter) of sufficient size for both RuBP and 3PGA to pass. Two conserved residues (glutamate and arginine) at the 3-fold symmetry of the hexameric shell proteins. Their diameters range between 4 and 7 Å, sufficiently large for diffusion of bicarbonate into the carboxysome (2). However, current models do not explain how the bulkier substrate, 1,5-D-ribulose bisphosphate (RuBP), and the product of CO2 fixation, 3-phosphoglycerate (3PGA), enter the current model for carboxysome shell structure and function (3, 11). Collectively, carboxysomes and these (presumably) architecturally similar organelles are known as BMCs.

The first structure of a protein containing a fusion of BMC domains provided a clue (16). This protein, CsoS1D, is encoded just upstream of the cso operon in most α-cyanobacteria. CsoS1D forms trimers with a relatively large central pore (14 Å in diameter) of sufficient size for both RuBP and 3PGA to pass. Two conserved residues (glutamate and arginine) at the 3-fold symmetry of the hexameric shell proteins. Their diameters range between 4 and 7 Å, sufficiently large for diffusion of bicarbonate into the carboxysome (2). However, current models do not explain how the bulkier substrate, 1,5-D-ribulose bisphosphate (RuBP), and the product of CO2 fixation, 3-phosphoglycerate (3PGA), enter the current model for carboxysome shell structure and function (3, 11). Collectively, carboxysomes and these (presumably) architecturally similar organelles are known as BMCs.

The first structure of a protein containing a fusion of BMC domains provided a clue (16). This protein, CsoS1D, is encoded just upstream of the cso operon in most α-cyanobacteria. CsoS1D forms trimers with a relatively large central pore (14 Å in diameter) of sufficient size for both RuBP and 3PGA to pass. Two conserved residues (glutamate and arginine) at the 3-fold symmetry of the hexameric shell proteins. Their diameters range between 4 and 7 Å, sufficiently large for diffusion of bicarbonate into the carboxysome (2). However, current models do not explain how the bulkier substrate, 1,5-D-ribulose bisphosphate (RuBP), and the product of CO2 fixation, 3-phosphoglycerate (3PGA), enter the current model for carboxysome shell structure and function (3, 11). Collectively, carboxysomes and these (presumably) architecturally similar organelles are known as BMCs.

Bioinformatically, an ortholog of csoS1D was identified outside of the main ccm cluster in the genomes of all β-cyanobacteria (16, 19). Here, we show that the product of this gene, now called CcmP, is indeed a constituent of the β-carboxysome of Synechococcus elongatus PCC 7942 (Syn7942) in vivo. We also report two crystal structures of CcmP at resolutions of 2.5 and 3.3 Å, from two different crystallization conditions; this is the first structure of a tandem BMC domain protein from a β-carboxysome. Evidence from structural, biophysical, and computational analyses suggests CcmP functions to form a gated, discrete nanocompartment within the β-carboxysome shell, likely to conduct larger metabolites. Phylogenetic analysis indicates that the BMC domains of CcmP and its α-carboxysome ortholog, CsoS1D, belong to a lineage of BMC domain proteins. Members of this subclass are also found in several distinct bacterial microcompartment gene clusters of heterotrophic organisms; residues for pore gating and formation of a robust dimer of trimers are conserved in these orthologs suggesting the capacity to form gated nanocompartments within microcompartment shells is widespread.

EXPERIMENTAL PROCEDURES

Strains and Plasmid Construction for Generating Syn7942 Mutants—All DNA fragments were amplified from Syn7942 genomic DNA using primers indicated in supplemental Table S4 and cloned using a BglBrick strategy (EcoRI-BglII-part-BamHI) (20). The cerulean fluorescent protein was first fused to the C terminus of CcmP before the entire fusion construct was transferred into a BglBrick-modified neutral site I (pAM2991:SpR) vector containing an isopropyl 1-thio-β-d-galactopyranoside-inducible P_{trc} promoter (21) to generate P_{trc}-CcmP-cerulean. To generate P_{ccmK2}-RbcL-GFP, the ccmK2 promoter (P_{ccmK2}) was inserted in front of the RbcL-YFP fusion to generate a functional transcriptional unit, P_{trc}-CcmP-cerulean, which was subsequently transferred to a BglBrick-modified neutral site II vector (pAM1573:CmR) (22). Transformation of Sy7942 was accomplished as described previously (23).

Cloning, Expression, and Purification—The ccmP gene from Syn7942 ( locus tag Synpcc7942_0520) was obtained by PCR amplification using Syn7942 genomic DNA as template and the primers listed in supplemental Table S4. Then it was cloned in BamHI and Ascl sites of expression vector pCOLADuet-1 (Novagen, Madison, WI) via standard digestion and ligation procedure to generate pCOLA-ccmP. The sequence of the protein coding region was confirmed by DNA sequencing (University of California at Berkeley DNA sequencing facility). Protein expression and affinity purification of C-terminal His-tagged CcmP were performed following procedures previously described (16).

Fluorescence Microscopy—Syn7942 strains harboring P_{trc}-CcmP-cerulean and P_{trc}-RbcL-GFP or P_{trc}-cerulean alone were grown on solid BG11 (24) media in an environmental chamber with an atmosphere of 3% CO2 and a light intensity of ~100 microeinsteins/m² s. 24 h prior to imaging, cells were suspended in BG11 media and spotted onto BG11-agarose (1% Navagen, Madison, WI) via standard digestion and ligation procedure to generate pCOLA-ccmP. The sequence of the protein coding region was confirmed by DNA sequencing (University of California at Berkeley DNA sequencing facility). Protein expression and affinity purification of C-terminal His-tagged CcmP were performed following procedures previously described (16).

Crystallization, Data Collection, and Structure Determination—CcmP was crystallized at room temperature using the hanging drop vapor diffusion method. The protein was in 10 mM Tris-HCl, pH 8.0, at a concentration of 4.4 mg/ml, and 2 mM of 3PGA was added to the protein 12 h before the crystal was set up. Form 1 crystals were obtained by mixing 3 μl of CcmP (with 2 mM 3PGA) with 3 μl of reservoir solution of condition 1 (100 mM sodium acetate, pH 4.6, 1.2 M sodium formate, and 20% PEG 400). Form 2 crystals were obtained by mixing 4 μl of CcmP (with 2 mM 3PGA) with 2 μl of reservoir solution of condition 2 (17% tacsimate, pH 4.1).

Form 1 or form 2 crystals were rapidly transferred into mother liquor supplemented with 30% ethylene glycol or 100% tacsimate, pH 4.1, respectively, mounted in nylon loops, and frozen by placing them directly into a liquid nitrogen cryostream. Diffraction data were collected at the Advanced Light Source at Lawrence Berkeley National Laboratory beamline.
Subcompartment within a Microcompartment Shell

FIGURE 1. CcmP is a component of the β-carboxysome. 
A, organization of carboxysomal genes in S. elongatus PCC 7942 (Syn7942). Genes that contain BMC domain(s) are shown in red. B, wild-type cells harboring CcmP-cerulean and RbcL-GFP were visualized using laser scanning confocal microscopy following induction of PchL-CcmP-cerulean with 1 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG) for 24 h. RbcL-GFP is constitutively expressed using the endogenous promoter upstream of ccmK2. The chlorophyll-a (chl-a) channel was used as a marker to identify the cell outline. Scale bar, 1 μm.

RESULTS

CcmP Is a Component of the β-Carboxysome—A survey of all available cyanobacterial genome sequences (129 publicly available genomes at Integrated Microbial Genomes) (34) shows that an ortholog of CsoS1D, annotated as a hypothetical protein, is conserved among all β-cyanobacteria. Based on sequence homology (reciprocal best BLAST hit), its gene product is predicted to be the β-carboxysome counterpart of CsoS1D, recently experimentally confirmed to be a shell protein of the α-carboxysome (17). The percentage identity and similarity between the deduced amino acid sequences of gene Synpcc7942_0520 from Syn7942 (which contains β-carboxysomes) and CsoS1D from Prochlorococcus marinus strain MED4 CsoS1D is 46 and 69%, respectively. However, the β-carboxysome homolog lacks a 50 amino acid N-terminal extension prior to the docking step. Then a Rosetta-Docking protocol (33) was performed for incorporating CsoS1D/CcmP into CsoS1/CcmK2 layer. For each docking partner pair, Rosetta-Docking was performed for more than 5000 independent runs. The best solution was then used as a starting model in the Rosetta-Docking protocol for another 1000 independent runs, and the results were plotted to make sure an energy funnel is present (supplemental Fig. S6). A detailed description can be found in the supplemental material.

Transmission Electron Microscopy—A 300-mesh Formvar/carbon-coated copper grid (EMS, Fort Washington, PA) was floated on a drop of purified CcmP protein for 4 min. The grid was allowed to air dry and then stained with 2% aqueous uranyl acetate for 5 min. The negatively stained protein sample was imaged on a Tecnai 12 microscope at 120 kV, at the Electron Microscope Laboratory of the University of California, Berkeley.

Docking of CcmP or CsoS1D into Carboxysome Shell—The docking of CsoS1D/CcmP into the α- or β-carboxysome shell was performed with structural models of shell proteins for both model organisms, Prochlorococcus marinus strains MED4 and Syn7942. First, homology models of CsoS1 and CcmK2 were built using Swiss-Model (32) using CsoS1A from H. neapolitanus (2G13) and CcmK2 from Synechocystis sp. PCC 6803 (2A1B, for single layer) or from Thermosynechococcus elongatus (3SSQ, for double layer) as templates. Subsequently energy minimization using the Rosetta-Relax protocol was applied to both the CsoS1D/CcmP hexamer and the CsoS1/CcmK2 layer.
cients using thresholding are 0.608 and 0.522 for the cerulean and the GFP channels, respectively; this substantiates the co-localization of CcmP-cerulean and RbcL-GFP. Similar values were obtained for co-localization of internal carboxysome components (23). A control strain expressing cerulean fluorescent protein alone under the same conditions resulted in diffuse signal and no discrete puncta (supplemental Fig. S1). These results demonstrate that CcmP is a component of the β-carboxysome.

Structure of CcmP—Syn7942 CcmP was expressed with His tag in and purified from Escherichia coli. It was mixed with 3PGA, a product of Rubisco-catalyzed CO2 fixation, prior to crystallization. The protein crystallized in two different conditions with two different orthorhombic space groups (Table 1). One crystal (form 1) crystallized in space group P2₁,2,2 with six monomers in the asymmetric unit and diffracted to 2.5 Å. The second crystal (form 2) crystallized in space group P2₁,2,2 with 12 monomers in the asymmetric unit and diffracted to 3.3 Å. Both structures were solved by molecular replacement using CsoS1D (3F56) as the search model. The majority of the CcmP amino acid sequence, residues 3–205 or 206 (of a total of 213 residues), could be modeled into the calculated electron density for both structures. No residues were observed in the disallowed regions of the Ramachandran plot; 99% and more than 97% of residues are in the most favored region in form 1 and form 2, respectively (Table 1). The form 1 structure was refined to final R and Rfree values of 25.7 and 29.8%, respectively. Hexamers (dimers of trimers, see below) of CcmP (Fig. 2B) form layers in the form 1 crystal lattice, separated by a space that could accommodate an additional layer. This layer seems to be occupied by CcmP in different positions; the same phenomenon was previously observed for a different carboxysome shell protein, CsoS1C (4). This accounts for the relatively high R values for this resolution. The form 2 structure was refined to final R and Rfree values of 19.5 and 22.2, respectively; it does not have a layered crystal packing.

The CcmP monomer is composed of a fusion of two BMC domains (referred to as N- and C-BMC; Fig. 2B) that share less than 18% sequence identity to one another but have essentially identical folds (root mean square difference of 0.77 Å over 345 atoms). Both the N- and C-BMC domains of CcmP are a circular permutation of the typical BMC fold as described previously.

Table 1

|                      | CcmP form 1 | CcmP form 2 |
|----------------------|-------------|-------------|
| Space group          | P2₁2₁2₁     | P2₁2₁2₁     |
| Cell dimensions      |             |             |
| a, b, c              | 119.35, 137.36, 152.69 Å | 175.34, 176.20, 200.54 Å |
| α, β, γ              | 90.0, 90.0, 90.0° | 90.0, 90.0, 90.0° |
| Resolution           | 39.27 to 2.51 Å | 39.41 to 3.30 Å |
| No. of hexamers/ asymmetric unit | 1 | 12 |
| R_work               | 25.7%       | 19.5%       |
| R_free               | 29.8%       | 22.2%       |
| Ramachandran plot    |             |             |
| Most favored region  | 99.0%       | 97.5%       |
| Allowed region       | 1.0%        | 2.5%        |
| Outlier region       | 0.0%        | 0.0%        |

FIGURE 2. CcmP contains two BMC domains and forms a hexamer. A, CcmP hexamer shown in ribbons, colored by chain. B, monomer outlined in black was rotated 90° and rainbow-colored (blue, N terminus; red, C terminus) with 3PGA modeled in and shown as sticks. C, slab view of a CcmP hexamer, with one trimer (pseudo-hexamer) colored in green and one in cyan. The concave side of each trimer is indicated. D, when a map (σ of 1.3) was calculated without the presence of 3PGA molecule in the model, extra electron density was observed in the intra-domain pocket of a CcmP monomer near the conserved histidine residue (His-18). A 3PGA molecule (shown as sticks) can be credibly fit into this density.
A distinct difference between the CcmP trimer and the CcmK hexamers formed by single BMC domain proteins is that the size of the aperture at the 3-fold symmetry axis of CcmP is much larger (~13 Å). However, alternate conformations of the side chains of two absolutely conserved residues near the 3-fold symmetry axis, Glu-69 and Arg-70, result in CcmP trimers in which this pore is closed. Interestingly, the open and closed trimers in the form 1 structure are conformationally distinct, superimposing with an root mean square difference of 1.2 Å over 609 C-α atoms. A similarly sized gated pore was also observed in the CsoS1D structure (16). The observation that the conformations of absolutely conserved residues affect the size of the pore in multiple structures of both CcmP and CsoS1D suggest that transport into and out of the carboxysome through these proteins is controlled by the side-chain conformations of the conserved residues converging at the pore.

**Nanocompartment Formed by Dimerization of Two CcmP Trimers**—We observed tightly associated dimers of trimers in both crystal forms of CcmP (Table 2). The two trimers have a concave-to-concave orientation (Fig. 2C); residues on the concave side of the trimer are more conserved than on the opposite side (Fig. 3A). The residues forming the trimer-trimer interface are primarily found in the permuted region of the primary structure as follows: helix A and the loop connecting helix A to β-strand 2 of both N- and C-BMC domains (Fig. 2B). The two trimers are not stacked directly upon one another, but one trimer is rotated ~46° (θ_CcmP = 14°; see below and Table 2, footnote c, and supplemental Fig. S2B) with respect to the other (Fig. 2A and supplemental Fig. S2B). This results in a staggered interaction across the interface, so that the N-terminal domain of a monomer in the upper trimer interacts with the N- and the C-terminal domain of two different monomers in the lower trimer (Fig. 2A and supplemental Fig. S2B). The calculated shape complementarity (Sc) value for the dimer interface of CcmP is 0.677 (Table 2), which is in the range of Sc value for antibody/antigen interactions (31). Furthermore, CcmP elutes as a hexamer in gel filtration (supplemental Fig. S3), and its thermodynamically stable form in solution is also predicted by PDBePISA to be a hexamer; the calculated solvation free energy gain upon dimerization of CcmP trimers is ~51.2 kcal/mol, not including the effect of satisfied hydrogen bonds and salt bridges across the assembly’s interfaces. Moreover, the concave-to-concave interface is stabilized by 48 hydrogen bonds (Table 2), of which 36 are contributed by a conserved sequence motif: NR(X)1–2R(R/K)(G/A)(S/N/Q)(M/L) (residues 176–183 of CcmP, supplemental Fig. S5). Comparable values are obtained for its α-carboxysome ortholog CsoS1D (Table 2), and, interestingly, 36 of a total of 42 hydrogen bonds stabilizing CsoS1D hexamer formation are also contributed by the same sequence motif. Collectively, these data indicate that the stacked trimer assembly in both CcmP and CsoS1D is biologically relevant.

**Evidence for Metabolite Binding in the Nanocompartment of CcmP**—The dimerization of CcmP trimers across the concave surface results in the formation of a relatively large internal compartment, with a volume of ~12000 Å³. Access to the inte-
rior is restricted by the pores at the 3-fold axis of the two trimers; in the form 1 structure, the pore of one trimer is open, and the second is closed; in the form 2 structure, both pores are open. The nanocompartment has six identical pockets that may bind metabolites. The pockets are found within each CcmP monomer, located at the N- and C-BMC domain interface (Figs. 2 and 3B), between helix A of the N-BMC and β-strand 3 of the C-BMC. Elongated electron density was observed in each pocket, close to a strongly conserved histidine residue (His-18). All of the components in the mother liquor as well as cryoprotectant of the form 1 crystallization condition, with the exception of 3PGA, are too small to account for this density. A 3PGA molecule can be fit with its carboxyl group facing into the pocket of CcmP in the form 1 structure with an omit map calculated without the presence of the 3PGA molecule in the model (Fig. 2D). The refined occupancy of the 3PGA was 0.58. The distance between the carboxyl group of 3PGA and the nitrogen atom of the His-18 side chain ranges from 2.4 to 2.8 Å, which suggests a possible hydrogen bond. We also observed extra electron density in the same region in the form 2 structure; however, because the crystallization conditions contained several chemicals with 2–4 carbon chains, it was not possible to unambiguously identify the bound molecule. Isothermal titration calorimetry was performed in an attempt to quantify binding affinities between CcmP and 3PGA as well as RuBP. Unfortunately, the measured binding affinities are low and close to the detection limit of isothermal titration calorimetry (data not shown); therefore, no conclusive results were obtained. However, from a functional point of view, weak binding and partial occupancy in the crystal structure is consistent with CcmP’s proposed function as a conduit for metabolites (addressed under “Discussion”).

CcmP Incorporation into the Carboxysome Shell—We next considered how a CcmP hexamer could be fit into models of the β-carboxysome shell. There are two possible orientations for incorporation of a dimer of CcmP trimers into a single layer of uniformly oriented CcmK2 proteins. First, the CcmP hexamer can be incorporated into a single layer of CcmK2 by inserting one trimer in the same orientation as the CcmK2 hexamer (e.g. incorporated trimer concave side up in a concave-up layer of hexamers) (Table 3, models β1 and β3). Alternatively, CcmP hexamers can be incorporated into a single layer of CcmK2 by inserting one trimer in the orientation opposite that of the hexamers (e.g. incorporated trimer concave down in a concave-up layer of hexamers) (Table 3, models β2 and β4). Taking into consideration that the open and closed trimers in the form 1 structure are conformationally distinct, modeling was performed with either open or closed trimers for both possible orientations. Recently, a concave-to-concave dimerization of CcmK2 hexamers was reported, leading to the proposal that the β-carboxysome shell is double layered (9). We examined the fit of a CcmP hexamer into a layer of CcmK2 dodecamers. Because the CcmK2 dodecamers did not form layers in the crystal, it was first necessary to build a layer composed of CcmK2 dodecamers based on the single layers of hexamers. When the CcmP hexamer is docked into a layer of CcmK2 dodecamers (Table 3, model β5), only one-half of the CcmK2/CcmP interaction is closely packed; on the opposite side, there are obvious gaps between the CcmP trimer and the adjacent CcmK hexamer (supplemental Fig. S4). The inability to incorporate CcmP into a layer of dodecameric CcmK2 without creating gaps on one surface can be explained by the difference in the angle θ, the deviation from perfect superposition of two (pseudo)hexamers (supplemental Fig. S2). Whereas θCcmP is ~14°, θCcmK2 is only 5° (Table 2). This 5° angle was sufficient to bring complementarily charged surfaces to stabilize the CcmK2 dodecamer (9), but the difference between θCcmK2 and θCcmP precludes tight interactions between both layers of a CcmK2 dodecamer and a CcmP hexamer simultaneously. This is also reflected in the resulting buried surface area between the CcmK2 dodecamer and the CcmP hexamers, which is less than double of the area buried when the CcmP is fitted into the single layer of CcmK2 (Table 3).

A new model for packing of CcmP hexamers within the carboxysome shell emerges from the form 1 crystal packing, and CcmP alone forms layers. The surface area buried and shape complementarity between adjacent CcmP pseudohexamers in the layer are in the range reported for CcmK2/CcmK2 and CcmK4/CcmK4 interactions (supplemental Table S1) previously suggested to constitute the facets of the carboxysome shell (2, 39). Close inspection of the CcmP layer shows that it is formed by alternating strips of open and closed trimers (Fig. 4A). Self-assembly of CcmP particles is also evident in transmission electron micrographs (Fig. 4B).

**BMC Domains of CcmP and CsoS1D Form a Distinct Class**—Although the N-BMC and C-BMC domains of CcmP are structurally similar, they share less than 18% sequence identity.

### Table 3: Comparison of interface between two adjacent building blocks in models for the β-carboxysome shells and in CcmP crystal packing

| Interaction between two adjacent building blocks | Schematic Diagrama | Sc | Total Area Buried per Edge (Å²) |
|-------------------------------------------------|-----------------|---|-----------------------------|
| CcmK2 hexamer (single layer) b| ![Schematic Diagram](image)
| β1 | 0.453 | 453 : 468 |
| β2 | 0.416 | 524 : 528 |
| β3 | 0.503 | 484 : 521 |
| β4 | 0.475 | 467 : 470 |
| CcmK2 dodecamer (double layer) c | ![Schematic Diagram](image)
| β5 | 0.479 | 605 : 609 |
| CcmP hexamer from crystal packing | ![Schematic Diagram](image)
| | 0.518 | 801 : 801 |

a Key symbols for the schematic diagram. CcmP is shown in orange, and CcmK2 is shown in light blue. In the table, both proteins are shown in side view.

b Rosetta-Docking was used to find the best solution for each model.

c Homology model of CcmK2 hexamer was built based on CcmK2 from Synechocystis sp. PCC 6803 (2A1B). See supplemental material for details.

d Homology model of CcmK2 dodecamer was built based on CcmK2 from T. elongatus (355Q). See supplemental material for details.

e Adjacent hexamers were observed in CcmP crystal packing (form 1).
DISCUSSION

Previously, we suggested that a hypothetical protein found in cyanobacterial genomes was a component of the β-carboxysome (16, 19). In this work, we show that it is indeed a part of the β-carboxysome (Fig. 1). Structural (Fig. 2) and biophysical characterization (Table 2; supplemental Fig. S3) suggests that CcmP trimers dimerize to form a bilayered building block of the carboxysome shell. The residues at the trimer-trimer interface are primarily found in the permuted region of the primary structure, but permutation does not strictly correlate with dimerization of tandem BMC-domain trimers. For example, in the structures of noncarboxysomal tandem BMC domain proteins EutL (3GFH) (38), EutB (3IO0) (15), and PduB (4FAY) (40), permutation, but not dimerization, is observed. Notably, these three proteins lack a conserved amino acid pattern involved in hydrogen bonding that stabilizes dimerization of CcmP. In contrast, CsoS1D, which was recently shown to be an important component of the α-carboxysome (17, 18), also contains this motif. CsoS1D, like CcmP, crystallized as a dimer of trimers with comparable biophysical properties (16) suggesting that the dimer of trimers of both CcmP and CsoS1D is a building block of both the α- and β-carboxysome.

Dimerization of the trimers of CcmP and CsoS1D results in the formation of the nanocompartment, accessible only through a relatively large pore that is gated by the conformation of absolutely conserved residues. Co-crystallization of CcpM with 3PGA resulted in identification of a potential binding pocket, containing a conserved histidine (supplemental Fig. S5) for 3PGA within the interior of the nanocompartment (Figs. 2D and 3B). The low occupancy of the metabolite in the structure as well as lack of evidence for tight binding (using isothermal titration calorimetry) suggest the metabolite is only transiently retained, which is consistent with the function of the pore and nanocompartment as a passageway into the carboxysome. A similarly sized depression containing a conserved histidine residue (His-172) can also be identified in CsoS1D (supplemental Fig. S5). However, it is not located at the interface of N-BMC and C-BMC domains, as in CcmP, but at the interface of two adjacent monomers.

Unlike the pores of the single BMC domain hexameric shell proteins, the central pore of CcmP or CsoS1D is large enough for RuBP and 3PGA to readily enter. However, if such large pores were constitutively open, some of the advantages of the diffusive resistance to loss of substrate provided by the carboxysome shell would be negated. Perhaps the presence of a nanocompartment gated on both sides functions analogously to an airlock; the compartment served as a foyer where the metabolite enters through the open pore facing the cytosol and is temporarily retained in the compartment. Subsequently, the pore open to the cytosol closes, and the other pore facing the carboxysome interior opens. Interestingly, in one of the CcmP structures reported here, as well as in structures of CsoS1D (16), each hexamer contained one trimer with a pore in the open conformation, whereas the pore of the second trimer was closed.

In this model, CcmP is a specialized shell protein for the passage of larger metabolites. The bulk of the carboxysome shell is thought to be composed of single BMC domain protein hexamers (2, 5). The CcmP dimer of trimers can be fit into single layer models of the carboxysome shell (Table 3); however, one trimer of the pair is left out of the plane of the layer. Comparable values for fitting CsoS1D into a single layer of CsoS1 in the α-carboxysome shell are obtained (supplemental Table S3).
Recently, it has been suggested that the entire carboxysome shell is composed of a double layer of hexameric shell proteins (9). The major building block of these layers is proposed to be a CcmK2 dodecamer. A comparison of the CcmK2 dodecamer with the CcmP hexamer shows that they differ in the amount of surface area buried and the number of hydrogen bonds in the interface (Table 2), suggesting that the dimerization of hexamers of CcmK2 is less robust than the dimerization of trimers in CcmP. A key difference between these two building blocks, with implications for shell assembly, is how symmetrically the trimers of CcmP or hexamers of CcmK2 are appressed. Because of the difference in /H9258 and /H9252, a CcmP hexamer fits closely with only half of an adjacent CcmK2 dodecamer. In contrast, crystal packing (Fig. 4A) and transmission electron micrography (Fig. 4B) suggest that CcmP hexamers are capable of higher level assembly. In the observed layers and strips, the stacked trimers of CcmP create isolated nanocompartments (Fig. 3B). In contrast, because of the relatively loose association in the hexamer-hexamer interface in dodecameric CcmK2, the space formed between the two layers is continuous (Fig. 3 in Ref. 9).

Based on phylogenetic analyses, the primary structures of the N- or C-BMC domains of CcmP and CsoS1D form distinct clades of BMC shell proteins distal to their single BMC domain counterparts (e.g., CsoS1s and CcmKs; Fig. 5). Accordingly, the emergence of CcmP or CsoS1D is likely to have been an ancient event, before the separation of the α- and β-carboxysomes, and it supports our conclusion based on their similar structural properties (Table 2) that they are functionally equivalent.

Phylogenetic analysis also suggests the ability to form gated nanocompartment shell building blocks may extend to other bacterial microcompartments besides carboxysomes, where they may likewise be necessary for conducting relatively large metabolites across microcompartment shells without compromising the diffusive resistance afforded by a shell. Several subsets of the heterotrophic microcompartment groups (supplemental Table S2; and see Table 1 in Kerfeld et al. (11) and supplemental Table S3 in Kinney et al. (23)) also contain genes encoding tandem BMC domain proteins that cluster with CcmP and CsoS1D (Fig. 5). Many of the residues involved in the dimerization of trimers are conserved, and strikingly, the glutamate and arginine residues involved in gating the pore are also absolutely conserved (supplemental Fig. S5). This suggests that these BMCs of unknown function have shell proteins that also form nanocompartments with a large, gated central pore. Given the growing interest in BMCs for synthetic biology, an understanding of the structural and biophysical properties of the distinct shell protein building block represented by CcmP should be useful in informing the design of subcellular enzymatic reactors based on BMC architectures.

**Acknowledgments**—We thank the staff at the Advanced Light Source, Lawrence Berkeley National Laboratory, which is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the United States Department of Energy under Contract No. DE-AC02-05CH11231. We thank Annette Salmeen for help in synchrotron data collection and Kent McDonald and Reena Zalpuri at the Electron Microscope Laboratory of the University of California, Berkeley, for help with electron microscopy. We thank Seth D. Axen and Kirsten Fagnan for help with Rosetta-Docking and running modeling jobs on the supercomputing cluster Genepool at The National Energy Research Scientific Computing Center. We also thank Seth D. Axen, Jonathan Lassila, and Beth Wurzburg for critical reading of the manuscript.
