Production and Characterization of Synthetic Carboxysome Shells with Incorporated Luminal Proteins

AUTHORS: Fei Cai\textsuperscript{1,2}, Susan L. Bernstein\textsuperscript{1,2}, Steven C. Wilson\textsuperscript{1,2} and Cheryl A. Kerfeld\textsuperscript{*1,2,3}

\textsuperscript{1}Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720
\textsuperscript{2}Molecular Biophysics & Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720
\textsuperscript{3}MSU-DOE Plant Research Laboratory and Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824

* Corresponding author; e-mail: ckerfeld@lbl.gov

SHORT TITLE:

Synthetic carboxysome shells with cargo

ONE SENTENCE SUMMARY:

Synthetic carboxysome shells, containing all of the key component proteins provide a platform for fundamental studies of shell permeability and metabolic modeling of organelle function.

AUTHOR CONTRIBUTIONS

FC and CAK conceived and designed the experiments and analyzed the data; FC and SLB performed the experiments; SW designed the synthetic pHalo-1 construct and performed initial Halo-1 shell purifications. FC and CAK wrote the paper.
**FUNDING INFORMATION**

This work was supported by the National Science Foundation Emerging Frontiers Program (EF1105897) and Basic Energy Sciences, US Department of Energy (DE-FG02-91ER20021) with infrastructure support from MSU AgBioResearch.

**ABSTRACT**

Spatial segregation of metabolism, such as cellular-localized CO₂ fixation in C4 plants or in the cyanobacterial carboxysome, enhances the activity of inefficient enzymes by selectively concentrating them with their substrates. The carboxysome and other bacterial microcompartments (BMCs) have drawn particular attention for bioengineering of nanoreactors because they are self-assembling proteinaceous organelles. All BMCs share an architecturally similar, selectively permeable shell that encapsulates enzymes. Fundamental to engineering carboxysomes and other BMCs for applications in plant synthetic biology and metabolic engineering is understanding the structural determinants of cargo packaging and shell permeability. Here we describe the expression of a synthetic operon in *Escherichia coli* that produces carboxysome shells. Protein domains native to the carboxysome core were used to encapsulate foreign cargo (GFP) into the synthetic shells. These synthetic shells can be purified to homogeneity with or without luminal proteins. Our results not only further the understanding of protein-protein interactions governing carboxysome assembly, but also establish a platform to study shell permeability and the structural basis of the function of intact BMC shells both *in vivo* and *in vitro*. This system will be especially useful for developing synthetic carboxysomes for plant engineering.
INTRODUCTION

A key enzyme in photosynthesis is the CO₂ fixation enzyme ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO). RuBisCO not only fixes CO₂, resulting in carbon assimilation, but it can also fix O₂, leading to photorespiration. Suppressing the unwanted oxygenase activity of RuBisCO by sequestering RuBisCO with a source of CO₂ is Nature’s solution to this substrate discrimination problem. While C₄ plants compartmentalize CO₂ fixation in specific cells (Hibberd et al., 2008; Parry et al., 2010), cyanobacteria have evolved a specialized organelle composed entirely of protein to encapsulate RuBisCO-- the carboxysome.

The carboxysome is just one type of bacterial microcompartment (BMC), widespread, functionally diverse bacterial organelles (Axen et al., 2014). All BMCs consist of an enzymatic core surrounded by a selectively permeable protein shell (Kerfeld et al., 2005; Tanaka et al., 2008; Chowdhury et al., 2015; Kerfeld and Erbilgin, 2015). While the encapsulated enzymes differ among functionally distinct BMCs, they share an architecturally similar shell composed of three types of proteins: BMC-H, BMC-T, and BMC-P forming hexamers, pseudohexamers and pentamers, respectively (Kerfeld and Erbilgin, 2015). These constitute the building blocks of a self-assembling, apparently icosahedral shell with a diameter ranging from 40-400 nm (Shively et al., 1973; Shively et al., 1973; Price and Badger, 1991; Shively et al., 1998; Bobik et al., 1999; Iancu et al., 2007; Iancu et al., 2010; Petit et al., 2013; Erbilgin et al., 2014). Recent studies have also shown that in the biogenesis of BMCs an encapsulation peptide (EP) (Fan and Bobik, 2011; Kinney et al., 2012; Aussignargues et al., 2015; Jakobson et al., 2015), a short (~18 residues) amphipathic α-helix, mediates interactions between the core protein and the shell (Fan and Bobik, 2011; Choudhary et al., 2012; Kinney et al., 2012; Lawrence et al., 2014; Lin et al., 2014; Aussignargues et al., 2015). Indeed, because they are self-assembling organelles composed entirely of protein, BMCs hold great promise for diverse applications in bioengineering and development of bionanomaterials (Frank et al.,...
the key features of BMCs include selective permeability, spatial co-localization of enzymes, the establishment of private cofactor pools and the potentially beneficial effects of confinement on protein stability. For example, introducing carboxysomes into plants could provide a saltational enhancement of crop photosynthesis (Price et al., 2013; Zarzycki et al., 2013; Lin et al., 2014; McGrath and Long, 2014).

The β-carboxysome, which sequesters form 1B RuBisCO, has been an important model system for the study of the structural basis of carboxysome function, assembly and engineering (Kerfeld et al., 2005; Tanaka et al., 2008; Cameron et al., 2013; Aussignargues et al., 2015; Cai et al., 2015). Beta-carboxysomes assemble from the inside-out (Cameron et al., 2013; Gonzalez-Esquer et al., 2015); two proteins that are absolutely conserved and unique to β-carboxysomes, CcmM and CcmN, play essential roles in this process: CcmM crosslinks RuBisCO through its C-terminal RuBisCO small subunit-like domains (SSLDs; pfam00101); CcmM and CcmN interact through their N-terminal domains and C-terminal EP of CcmN interacts with the carboxysome shell.

Here we describe a system for producing synthetic β-carboxysome shells and encapsulating non-native cargo. We constructed a synthetic operon composed of ccmK1, ccmK2, ccmL and ccmO, genes encoding, respectively, two BMC-H, a BMC-P and a BMC-T proteins of the carboxysome shell of the halotolerant cyanobacterium, Halothece sp. PCC 7418 (Halo hereafter). Recombinant shells composed of all four proteins were produced and purified. We also demonstrated that the terminal α-helices of CcmK1 and CcmK2 are not, as had been proposed (Samborska and Kimber, 2012), required for the shell formation, and that the synthetic shell is a single-layered protein membrane. Cargo could be targeted to the interior of the synthetic shells using either the EP of CcmN or the N-terminal domain of CcmM; the latter observation provides new insight into the organization of the β-carboxysome. Our results not only further the understanding of protein-protein interactions governing carboxysome
assembly but also provide a platform to study carboxysome shell permeability.

These results will be useful in guiding the design and optimization of carboxysomes and other BMCs for introduction into plants.
RESULTS

Expression of a synthetic operon results in recombinant carboxysome shells that can be purified

Halo is a halotolerant cyanobacterium, with optimal growth under laboratory conditions with up to 20% salinity at 38 °C. (Garcia-Pichel et al., 1998). Enzymes and protein complexes encoded by halophiles often exhibit enhanced stability relative to their non-halophilic counterparts (Jaenicke and Bohm, 1998) and are thus valuable for applications in biotechnology (Margesin and Schinner, 2001). Therefore, we reasoned that carboxysome components of Halo may be good candidates for heterologous production of synthetic shells. The carboxysome genes of Halo are found in the main ccm locus (ccmK1, ccmK2, ccmL, ccmM, and ccmN) and three satellite loci. As in many other β-cyanobacteria, the essential ccmO gene is not encoded in the main ccm locus but elsewhere in the genome. A synthetic operon, Halo-1, was designed in an attempt to produce synthetic carboxysome shells heterologously (Figure 1A). In order to mimic shell protein ratios consistent with the current model of the β-carboxysome shell (Tanaka et al., 2008; Cameron et al., 2013) and currently available transcriptomic data for ccm genes expression (Schwarz et al., 2011; Vijayan et al., 2011; Billis et al., 2014), a strong ribosomal binding site (RBS) was added preceding the coding regions of ccmK1 and ccmK2, with a medium and a low strength RBS preceding ccmO and ccmL, respectively.

The synthetic operon Halo-1 was cloned into an E. coli vector under a T7 promoter, and expression was induced by addition of IPTG. Synthetic Halo carboxysome shells could be purified after detergent lysis of E. coli cells via differential centrifugations and anion exchange chromatography (see Materials and Methods for details). The purified shells were negatively stained and imaged using transmission electron microscopy (TEM). The shells measured 24.70±1.43 nm in diameter (n=1507), and many displayed polyhedral profiles consistent with icosahedral symmetry (Figure 1B, red arrows). The purified shells were resolved into five bands on an SDS-PAGE (Figure 1C). The presence of CcmO, CcmK1,
CcmK2 or CcmL proteins was confirmed by colorimetric immunoblots (Figure 1D), the latter two co-migrating as a single band (the calculated molecular masses of CcmK2 and CcmL are 10.8 kDa and 10.3 kDa, respectively) (Figure 1C, band 5). The identity of each protein band was also verified unambiguously by Mass Spectrometry (MS) fingerprinting analysis: band 3 and band 4 correspond to CcmO and CcmK1, respectively. Band 5 was a mixture of CcmK2 and CcmL. Two other bands (bands 1 & 2) were consistently resolved on SDS-PAGE from
synthetic shell preparations. Band 1 was identified as the LacZ protein of *E. coli*. A closer look at TEM images reveals that not all purified *Halo* shells appear empty (Figure 1E), possibly due to the presence of impurities (e.g. LacZ) captured during assembly. Band 2, interestingly, was identified as a mixture of CcmO, CcmK1, and CcmL, this may represent a subcomplex formed for the construction of the vertices of the shell.

Fluorescent proteins can be targeted to synthetic carboxysome shells using the EP of CcmN

One of the requirements for adapting carboxysomes and other BMC architectures for applications in bioengineering is the ability to target protein(s) of choice for compartmentalization. CcmN orthologs contain an EP on the C-terminus, following a poorly conserved linker region (Kinney et al., 2012; Aussignargues et al., 2015). The 18 amino acid long EP from *Synechococcus elongatus* PCC 7942 (*Syn* hereafter) (Figure S1, blue box) is essential for the interaction between CcmN and CcmK2 both *in vivo* and *in vitro* (Kinney et al., 2012). Compared to EPs from many other cyanobacteria, the predicted EP of *Halo* CcmN (residues 215-258) is unusual (Figure S1, red box): it is 26 residues longer than the experimentally characterized CcmN EP of *Syn* (Figure S1, blue box). Residues 215-232 resemble the canonical EP, followed by a C-terminal extension that is found among some CcmN orthologs (Kinney et al., 2012). The predicted secondary structure for the EP of *Halo* CcmN is unusual in that it consists of three α-helices (Figures 2A & S2).

To test the feasibility for using the *Halo* CcmN or its EP for targeting protein into the synthetic *Halo* shells, green fluorescent protein (GFP) was used as cargo. Full-length CcmN was fused to the C-terminus of GFP, but the resulting fusion protein was insoluble and formed inclusion bodies *in vivo*. The observed insolvency may be attributable to the predilection of recombinant CcmN to aggregate when expressed without its interaction partner, CcmM (data not
shown). In contrast, when CcmN211-258 was fused to the C-terminus of GFP, we observed diffuse fluorescent signal throughout the E. coli cytosol, indicating the fusion protein was soluble (Figure 2B). When the GFP-EP was co-expressed with the Halo shells, the signal from GFP-EP is no longer diffuse, but localized to the cell poles (Figure 2B).

To further confirm the co-localization of GFP-EP and Halo shells, an open reading frame encoding a cerulean fluorescent protein (CFP) was fused in-frame to the ccmK1 gene and inserted into the Halo-1 construct with a low-strength RBS following the ccmL gene (Figure 2C). For imaging, the resulting synthetic operon (Halo-6) was expressed alone or co-expressed with the GFP-EP (Figure
2D). CFP-labeled synthetic Halo shells appeared as fluorescent puncta in E. coli cells (Figure 2D). When GFP-EP was co-expressed with the CFP-labeled shell, although the puncta were frequently small, the co-localization of both signals was unambiguous (Figure 2D).

Synthetic carboxysome shells with cargo can be purified

We purified Halo shells from the strain co-expressing Halo-1 and GFP-EP or GFP-CcmN. Purified Halo shells with cargo were negatively stained and imaged with TEM (Figure 3F and 3G). They are homogenous and similar in size and appearance to Halo shells without cargo (Φ23.58±1.80 nm, n=2315 for Halo shells with GFP-EP; Φ24.50±2.27 nm, n=1663 for Halo shells with GFP-CcmN). To confirm the presence of cargo, purified shells were analyzed by SDS-PAGE and immunoblotting using chemiluminescent detection. As expected, separation patterns of both samples on SDS-PAGE were very similar to that of empty Halo shells (Figure 3A). Immunoblots confirmed the presence of all four shell proteins (Figure 3B-D). Although GFP fusion proteins were not clearly observed on the SDS-PAGE, immunoblots developed with anti-GFP antibodies indicate the presence of GFP fusions at the expected molecular weights (Figure 3E). The band previously identified as a mixture of CcmK1, CcmO, and CcmL in MS analysis was also confirmed by anti-SynCcmK2 and anti-SynCcmO antibodies (Figure 3B and 3D, black arrows). Interestingly, a band with an observed molecular mass slightly less than 75 kDa was also recognized by anti-SynCcmO antibodies and was present in all three samples regardless of absence or presence of cargo (Figure 3D, asterisk). This band is hardly visible on a coomassie blue stained SDS-PAGE but obvious in immunoblots. It may be an oligomer of CcmO.

The synthetic carboxysome shell is composed of single-layer
Based on measurements from micrographs of *Halo* shells, the thickness of the shell is 2.8±0.5 nm \( (n=192) \). This suggests that the synthetic carboxysome shells are composed of only a monolayer of shell proteins. However, because of the possible errors in TEM-based measurements, we examined the impact of truncation of the C-terminal α-helix (αD-helix; Figure S3B, shown in red) which has been described as crucial for the dimerization of shell proteins that is proposed to lead to the formation of a double layered facet in the carboxysome of *Thermosynechococcus elongatus* (Samborska and Kimber, 2012). Truncated variants of *Halo* CcmK1 and CcmK2 were designed based on sequence alignments with *T. elongatus* CcmK1 and CcmK2 (Figure S3A). Deletion of residues beyond Pro90, which precedes the αD-helix in both CcmK1 and CcmK2, should completely abolish dimerization (Figure S3C). The truncated ccmK1 and ccmK2.
genes were synthesized and used for replacement of full-length \( ccmK1 \) and \( ccmK2 \) in the synthetic operons Halo-1 and Halo-6. The resultant operons were named Halo-1T and Halo-6T (T for truncated), respectively (Figure 4A).

Expression of Halo-1T produced shells and TEM on an enrichment fraction showed that the thickness of the shells with truncated shell proteins were similar to wild type: 2.46±0.45 nm (\( n=57 \)) (Figure 4B), implying that the synthetic carboxysome shells are single-layered. Co-expression of Halo-6T and GFP-EP resulted in co-localization of fluorescently labeled shells and cargo (Figure 4C), suggesting the truncated shells retain the ability to encapsulate cargo and the \( \alpha \)D-helix is not involved in the interaction between the EP and the Halo shell proteins CcmK1 and CcmK2.

The γ-CA domain of CcmM interacts with the synthetic carboxysome shell

The absolutely conserved β-carboxysome protein, CcmM, is composed of an N-terminal γ-CA domain and three to five copies of SSLDs. Previous studies

\[ \text{ccmK2} \]

\[ \text{Halo-1T} \]

\[ \text{Halo-6T} \]

\[ \text{C} \]

\[ \text{Halo-6T} \text{ GFP-EP} \]

\[ \text{Cerulean} \]

\[ \text{GFP} \]

\[ \text{DIC} \]

Figure 4. Synthetic carboxysome shell formation with CcmK1 and CcmK2 truncations.

(A) Synthetic operons Halo-1T and Halo-6T are variants of Halo-1 and Halo-6, respectively in which the full-length \( ccmK1 \) or \( ccmK2 \) gene was replaced by truncated versions. (B) Visualization of cerulean-labeled mutant Halo shells (cyan) without or with encapsulated GFP proteins (green-fire-blue). Bars indicate 2 μm. (C) An enrichment of mutant Halo shells (red arrows) was negatively stained and visualized by TEM. The bar indicates 20 nm. Inlet: an enlarged mutant shell.
indicated that the SSLDs of CcmM are required for the nucleation of RuBisCO and, therefore, crucial for the β-carboxysome formation (Long et al., 2007; Long et al., 2010; Cameron et al., 2013). However, the potential structural role of the N-terminal γ-CA domain in the interaction between the carboxysome core and the shell has not been investigated. The synthetic Halo shell system provides a background to probe for a direct interaction between CcmM and the shell without the influence of RuBisCO. A super-folding GFP (sfGFP) variant was used as a reporter and fused to CcmM between the γ-CA domain and four SSLDs to

Figure 5. CcmM interacts with synthetic carboxysome shells through its γ-CA domain.

(A) Halo CcmM contains a γ-CA domain followed by four SSLD domains. A super folding-GFP was inserted between the γ-CA domain and the first SSLD domain to generate the fusion protein IM-GFP. Deletion of four SSLD domains of IM-GFP results in γCA-GFP, and deletion of the γ-CA domain results in GFP-SSLDs. (B) Visualization of sfGFP labeled full-length or partial CcmM (green-fire-blue) without or with the presence of cerulean-labeled Halo shells (cyan). Bars indicate 2 μm.
generate a recombinant protein, iM-GFP (Figure 5A). The sfGFP was also fused to the γ-CA domain or the four SSLDs of Halo CcmM to generate γCA-GFP or GFP-SSLDs, respectively (Figure 5A). Interestingly, different localization patterns were observed, even when these fusion proteins were expressed in isolation (Figure 5B, top row). When both the γ-CA domain and SSLDs were present, multiple fluorescent puncta per cell were observed; in contrast, no or at most one polar punctum per cell was apparent when only the γ-CA domain was present. The sfGFP labeled SSLDs resulted in diffuse fluorescent signal, similar to what was observed in the negative control (sfGFP alone). Furthermore, a similar trend was observed when these fusion proteins were co-expressed with fluorescently labeled Halo shells (Halo-6) (Figure 5B). Although the expression level of fluorescently labeled Halo shells seemed to vary, CFP-puncta were observed in all cases, indicative of the formation of Halo shells. Co-localization of GFP signal with the CFP signal was observed in the case of iM-GFP or γCA-GFP, with multiple and single puncta per cell, respectively. Collectively, in vivo labeling results suggested that there is a direct protein-protein interaction between CcmM and the shell proteins, in the absence of CcmN, and that this interaction is mediated via the N-terminal γ-CA domain of CcmM.

To further test if iM-GFP and γCA-GFP are sequestered within Halo shells, we purified shells from strains expressing both types of cargo. Isolated Halo shells with iM-GFP or γCA-GFP have the expected composition on SDS-PAGE: CcmK1, CcmK2, CcmL and CcmO were present as well as the bands for the mixture of CcmK1/CcmL/CcmO and the putative oligomer of CcmO (Figure 6A-D). The fusion proteins iM-GFP and γCA-GFP were also detected by immunoblots at their expected molecular weights (Figure 6E), suggesting iM-GFP and γCA-GFP are indeed sequestered and co-isolated with Halo shells. Purified samples were also imaged with TEM, and similar shell structures were evident in both cases (Figure 6F & 6G).
DISCUSSION

The synthetic Halo carboxysome shells reported here demonstrate that β-carboxysome shells can be formed with the four key shell proteins in the absence of cargo in E. coli. Empty β-carboxysome shells, to our knowledge, have not been observed in vivo. We demonstrate targeting of cargo to the synthetic carboxysome shells via an EP, albeit requiring detection by immunoblotting. This method of detection has also been required in other, (non-carboxysome) synthetic shell systems (Parsons et al., 2010; Choudhary et al., 2012; Lassila et al., 2014).
The inability to directly visualize cargo proteins on SDS-PAGE may hint at the importance of cross-linking among luminal proteins for filling the cores of shells (Cameron et al., 2013; Gonzalez-Esquer et al., 2015) in the natural encapsulation process.

Our data also provide new insights into the details of β-carboxysome assembly. The highly purified Halo shells can be resolved into at least four discernible bands that are identified by both immunoblotting and MS. Characterization by TEM indicates that the facets are formed by a single layer of shell proteins. Although which side of the BMC-H hexamer faces the cytosol is still an open question, our data suggests the αD-helix of shell proteins CcmK1 and CcmK2, which is located on the concave side of CcmK1 or CcmK2 hexamer, is not involved in the EP-mediated encapsulation. Interestingly, immunoblots also consistently revealed two bands that are SDS-resistant complexes of shell proteins in all of the shell preparations. One appears to be an oligomer of CcmO and the other a complex of CcmO/CcmK1/CcmL. Notably, structural information on CcmO is still unavailable, although this absolutely essential carboxysome protein (Marco et al., 1994; Martinez et al., 1997; Rae et al., 2012) is assumed to be a trimer (pseudo-hexamer). The observation of a robust complex of CcmO/CcmK1/CcmL may represent a shell assembly intermediate. In any case, these data indicate that our model for the carboxysome shell, which is based on the structures of CcmL and CcmK proteins and the assumption of icosahedral symmetry (Kerfeld et al., 2005; Tanaka et al., 2008) are perhaps too simplistic.

Furthermore, we demonstrate two different types of interaction between domains of carboxysomal luminal proteins and the synthetic shells. These include an unusual variant of the canonical EP. In contrast to previously experimentally characterized EPs that consist of a single amphipathic α-helix (Fan et al., 2012; Kinney et al., 2012; Lawrence et al., 2014), the EP of Halo is predicted to comprise at least two α-helices (Figure 2A). Similar observations have been reported for the EPs of some glycyl radical enzyme-associated BMCs (GRM) (Zarzychi et al., 2015). We also demonstrate interaction between the γ-
CA domain of CcmM and shell components of β-carboxysome. These observations not only provide strategies for encapsulation into the shell but provide new details on the internal organization of the carboxysome. CcmM is essential to β-carboxysome assembly as it is required for nucleating RuBisCO in pro-carboxysome formation; its multiple SSLDs crosslink RuBisCO molecules (Long et al., 2007; Long et al., 2010; Cameron et al., 2013). On the other hand, the γ-CA domain of CcmM interacts with CcmN, which also interacts with the shell through CcmN’s C-terminal EP (Kinney et al., 2012). Here we showed that CcmM, in the absence of both RuBisCO and CcmN, associates with the shell. This allows further refinement of the model for the interior organization of the carboxysome core and its interactions with the shell. Collectively, observations from the synthetic carboxysome shell system reported here provide new insight into the organization of the interior and of the shell of the β-carboxysome.

The synthetic carboxysome shells, containing all of the key component proteins established a platform for fundamental studies of shell permeability; this is essential for metabolic modeling of organelle function (McGrath and Long, 2014). For carboxysomes, the shell functions as the interface between the first step of CO₂ fixation and the rest of cyanobacterial metabolism. More generally, the successful production and purification of synthetic Halo shells provides a new system for repurposing BMCs for applications in biotechnology applications and development of nanomaterials.

MATERIALS AND METHODS

Constructions of Expression Vectors

The synthetic operon Halo-1 was constructed as following (Figure 1A): the ccmK1, ccmK2, ccmO and ccmL genes of Halo genome (GenBank ID NC_019779.1) were placed after the promoter sequence. A non-native, 79 bp linker region was added between any two adjacent genes, each including a unique restriction site and an RBS. A strong RBS was added preceding the
coding region of *ccmK1* or *ccmK2*, and there are a medium and a low strength RBS preceding *ccmO* and *ccmL*, respectively. Each coding sequence was codon optimized for expression in *E. coli* and the gene synthesized (GenScript). The Halo-1 operon was cloned into a Bgl-brick compatible vector pETBb3 at *EcoRI* and *BamHI* sites. Another DNA fragment containing the truncated *ccmK1* and *ccmK2* genes with *EcoRI* and *NcoI* flanking regions on either side was also synthesized. This fragment was sub-cloned into pHalo-1 at *EcoRI* and *NcoI* sites to replace the non-truncated *ccmK1* and *ccmK2* via standard digestion and ligation procedures, and the resulting operon was named Halo-2. pHalo-6 and pHalo-7 were constructed based on pHalo-1 and pHalo-2, respectively, using a ligation-free method (In-Fusion Cloning kit, Clontech) following the manufacturer’s protocol. DNA fragments encoding GFP-CcmN and iM-GFP fusion were also synthesized at GenScript with codon optimization for *E. coli* and sub-cloned into a low-copy number Bgl-brick compatible vector pMCLBb to generated pFC227 and pFC243. pFC223 was generated by a PCR-based approach using pFC227 as a template. pFC244, pFC245, and pFC246 were generated using PCR-based approach using pFC243 as templates. All the constructs and primers used are listed in Table S1 and Table S2, respectively.

**Strains and Growth Conditions**

Recombinant protein and synthetic operon expression were carried out in *E. coli* BL21 (DE3) strains (Invitrogen, Carlsbad, CA). Pre-cultures were grown overnight in LB broth (EMD Millipore, Billerica, MA) media at 30 °C and shaking at 200 rpm with appropriate antibiotics (100 μg/ml Ampicillin or 34 μg/ml Chloramphenicol). Then a 1:100 dilution was made in fresh LB, and cultures were grown at 37 °C and shaking at 180 rpm. To induce protein expression, IPTG (Gold Biotechnology, INC., St. Louis, MO) was added to a final concentration of 0.5 mM when cultures had grown to OD600 ≈ 0.7-0.8. Cultures were continued grown at 37 °C for 4 hours before harvesting. Following 15 min
centrifugation at 5,000 g, cell pellets were weighed and stored in -20 °C until purification.

**Synthetic Carboxysome Shell Purification**

The frozen cell pellet was thawed in a room temperature water bath. Then 2.5 ml/g of B-PER II reagent (Pierce Protein Biology) was added to the cell pellet and the pellet was resuspended by pipetting. Then 200 µl of RNase A at 10 mg/ml was added per 6-8 g of cells. rLysome (Novagen; at 30 kU/ul) was added at 1.4 µl/g cell. The cell lysate was incubated on a rocker vigorously for 30 min at room temperature. Cell debris was removed with centrifugation at 27,000 g for 18 min at 4 °C. Then Benzonase nuclease (Novagen; 25 U/ul) was added to the clear cell lysate at 7.1 µl/g cell, and the cell lysate was incubated on a rocker for another 45 min at room temperature. Extracts were then ultracentrifuged on 7 ml of 30% sucrose cushion made in TBS 20/50 pH 7.4 (20 mM Tris-HCl, pH 7.4; 50 mM NaCl) at 42,000 rpm in a Ti-70 rotor for 4 hours at 4 °C. The supernatant was carefully removed, and the soft glassy pellet was resuspended in 1 ml of ice-cold TBS 20/50 pH 7.4. The resuspended sample containing shells was briefly centrifuged at 1,200 g for 5 min at 4 °C, and the supernatant was loaded on a 20-70% sucrose gradient made in TBS 20/50 pH 7.4 buffer. The gradient was run in an SW-28 rotor at 23,000 rpm at 4 °C for 16 hours. After the ultracentrifugation, ten fractions (4 ml each) were recovered from the gradient from top to bottom. The last fraction was used to resuspend the pellet at the bottom of the tube. A sample of each fraction was run on SDS-PAGE and fractions with shell proteins were pooled. The pooled sample was the shell enrichment sample and can be used for TEM. The enriched sample was then loaded on a Mono Q 10/100 column (GE Healthcare); fractions were collected from the TBS 20/0 to TBS20/1000 pH 7.4 gradient. Based on SDS-PAGE of Mono Q fractions, samples containing shells were pooled. A final clean-up step included dilution of pooled sample in TBS 20/350 pH 7.4 to 40 ml total followed by a
ultracentrifugation run at 42,000 rpm in a Ti-70 rotor for 4 hours at 4 °C. The final pellet was resuspended in 200 µl of TBS 20/50 pH 7.4.

SDS-PAGE and Immunoblots

Protein samples were separated on pre-cast 4-20% or 10-20% sodium dodecyl sulfate polyacrylamide gradient gels (BioRad) to analyze their composition. Polypeptide bands were visualized by staining with Gel Code Blue (Pierce Protein Science). For immunoblotting, the proteins were transferred onto a 0.45 µm pore size nitrocellulose membrane in a Mini Trans-Blot electrophoretic transfer cell (Bio-Rad). The blot was blocked with immunoblot blocking buffer (5% nonfat dry milk in PBS pH 7.4 with 0.1% Triton X-100) for 45 minutes. The appropriate primary antibody (raised in rabbits to recombinant protein) was incubated with the blot for 1 hour at room temperature to probe the presence of the target antigen. After rinsing with PBS buffer, immunoblot blocking buffer and PBS buffer for 15 minutes each, the blot was incubated with goat anti-rabbit IgG antibody conjugated with alkaline phosphatase (AP) or horseradish peroxidase (HRP) at 1:10,000 dilution for 1 hour at room temperature. The blot was developed with one step NBT-BCIP solution for colorimetric detection of AP activity or SuperSignal® West Pico Chemiluminescent Substrate for HRP detection (Pierce Protein Science). Images of stained gels and immunoblots were captured and documented using a ChemiDoc imaging system (Bio-Rad). Densitometry analysis was performed using the ImageLab program (Bio-Rad).

Sequence Alignment and Bioinformatics

Multiple sequence alignment was performed using Clustal X 2.1 (Larkin et al., 2007). Pairwise alignment was performed using the online global alignment program, Needle, at http://www.ebi.ac.uk/Tools/emboss/ (Li et al., 2015). The secondary motif prediction was performed using Quick2D at
http://toolkit.tuebingen.mpg.de/quick2_d with the PSIPRED algorithm (Jones, 1999). Protein homology models were built using the Swiss-Model interface (http://swissmodel.expasy.org/SWISS-MODEL.html) (Arnold et al., 2006). Ab initio protein structure prediction for the EP region (CcmN_{211-258}) was performed at QUARK server (http://zhanglab.ccmb.med.umich.edu/QUARK/) (Xu and Zhang, 2012, 2013). All the structure figures were prepared with PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.3 Schrödinger, LLC.).

Fluorescence Microscopy and Transmission Electron Microscopy

Induced cells prior to harvesting were used for fluorescence microscopy imaging. Two microliters of cells were spotted on a thin agar pad and air-dried before imaging with a Zeiss Axioplan 2 microscope or Zeiss LSM710 using a 100× oil immersion objective. Images were visualized and analyzed with ImageJ 1.4.8 (Schneider et al., 2012). Purified synthetic shells were spotted on formvar/carbon-coated copper grids (Electron Microscopy Sciences, No. FCF300-Cu) and negatively stained with 2% uranyl acetate for 60 seconds. Images were taken on a JEOL 1200 EX TEM. Particle analysis was done using ImageJ 1.4.8.

SUPPLEMENTAL MATERIALS

Table S1. E. coli plasmids used in this study.
Table S2. Oligonucleotides used in cloning.
Figure S1. Multiple sequence alignment of CcmN from Halo with 50 most closely related orthologs.
Figure S2. Fold prediction for the Halo EP.
Figure S3. Homology model of Halo CcmK1 and CcmK2.
ACKNOWLEDGMENTS

We thank the UC Berkeley Electron Microscopy Lab for assistance with TEM. Thanks to Dr. Denise Schichnes and Dr. Steven Ruzin at UC Berkeley Biological Imaging Facility for advice and assistance with fluorescent microscopy. We also thank Dr. Markus Sutter for advice during the construction of the Halo-1 operon and development of the Halo shell purification. Thanks to Dr. Manuel Sommer for critical reading of the manuscript. This work was supported by the National Science Foundation Emerging Frontiers Program (EF1105897) and Basic Energy Sciences, US Department of Energy (DE-FG02-91ER20021) with infrastructure support from MSU AgBioResearch.

AUTHOR CONTRIBUTIONS

FC and CAK conceived and designed the experiments and analyzed the data; FC and SLB performed the experiments; SW designed the synthetic pHalo-1 construct and performed initial Halo-1 shell purifications. FC and CAK wrote the paper.

FIGURE LEGENDS

Figure 1. Synthetic operon design and purification of synthetic carboxysome shells.

(A) The synthetic operon for expression of the four Halo carboxysome shell genes. Each gene has its own RBS, shown as bent green arrows. The dark to faint color represents the strong to the weak strength of the RBS. (B) Synthetic carboxysome shells purified from E. coli and negatively stained for TEM. Shells with obvious hexagonal profiles were marked with red arrows. (C) Separation of purified shell components on SDS-PAGE. Bands 1-5 were analyzed using MS
fingerprint analysis. (D) Immunoblots on purified shell components separated by SDS-PAGE, developed with antibodies raised against *Synechococcus elongatus* PCC7942 (anti-syn) CcmO, CcmK2, and CcmL. (E) Examples of the different appearances of purified *Halo* carboxysome shells. Bars indicate 20 nm.

Figure 2. Fluorescent proteins can be targeted to the synthetic carboxysome shells.

(A) The amino acid sequence of the extended encapsulation peptide region from *Halo* CcmN. Residues in the region corresponding to the experimentally-characterized EP of CcmN are shown in red. The blue cylinders denote regions predicted to form α-helices. (B) GFP was fused to the EP sequence and visualized without or without co-expression of the *Halo* shells. Pseudo-color (green-fire-blue in the ImageJ lookup table (Schneider et al., 2012)) was applied for GFP signal. (C) The synthetic operon *Halo*-6 contains a fusion gene, *ccmK1*-cerulean, downstream from *ccmL*. (D) Visualization of cerulean-labeled *Halo* carboxysome shells (cyan) and encapsulated GFP proteins (green-fire-blue). Bars indicate 2 µm.

Figure 3. Purification of synthetic carboxysome shells with cargo.

SDS-PAGE (A) and immunoblots (B-E) of purified synthetic carboxysome shells with GFP-CcmN (lane 2) or GFP-EP (lane 3) are compared to Halo shells without cargo (lane 1). Antibodies used in (B-E) are α-SynCcmK2, α-SynCcmL, α-SynCcmO and α-GFP, respectively. The band indicated with an arrow in (A) was also recognized by α-SynCcmK2 and α-SynCcmO antibodies. Anti-SynCcmO antibodies recognized a band slightly smaller than the 75 kDa marker (indicated by asterisks in A and D). Visualization of negatively stained *Halo* shells with GFP-CcmN (F) and GFP-EP (G) under TEM. Bars indicate 20 nm.
Figure 4. Synthetic carboxysome shell formation with CcmK1 and CcmK2 truncations.

(A) Synthetic operons Halo-1T and Halo-6T are variants of Halo-1 and Halo-6, respectively in which the full-length ccmK1 or ccmK2 gene was replaced by truncated versions. (B) Visualization of cerulean-labeled mutant Halo shells (cyan) without or with encapsulated GFP proteins (green-fire-blue). Bars indicate 2 µm. (C) An enrichment of mutant Halo shells (red arrows) was negatively stained and visualized by TEM. The bar indicates 20 nm. Inlet: an enlarged mutant shell.

Figure 5. CcmM interacts with synthetic carboxysome shells through its γ-CA domain.

(A) Halo CcmM contains a γ-CA domain followed by four SSLD domains. A super folding-GFP was inserted between the γ-CA domain and the first SSLD domain to generate the fusion protein iM-GFP. Deletion of four SSLD domains of iM-GFP results in γCA-GFP, and deletion of the γ-CA domain results in GFP-SSLDs. (B) Visualization of sfGFP labeled full-length or partial CcmM (green-fire-blue) without or with the presence of cerulean-labeled Halo shells (cyan). Bars indicate 2 µm.

Figure 6. Purification of synthetic carboxysome shells with full-length and truncated CcmM.

SDS-PAGE (A) and immunoblots (B-E) of purified Halo shells with iM-GFP (lane 1) or γCA-GFP (lane 2). Antibodies used in B-E are α-SynCcmK2, α-SynCcmL, α-SynCcmO and α-GFP, respectively. The band indicated with an arrow in (A) was also recognized by α-SynCcmK2 and α-SynCcmO antibodies. Anti-SynCcmO antibodies recognized a band slightly smaller than the 75 kDa marker (indicated by asterisks in A and D). Visualization of negatively stained synthetic
carboxysome shells with iM-GFP (F) and γCA-GFP (G) under TEM. Bars indicate 20 nm.


Common Feature of Bacterial Microcompartment Assembly. J Biol Chem 287: 17729-17736


Martinez I, Orus MI, Marco E (1997) Carboxysome structure and function in a mutant of Synechococcus that requires high levels of CO2 for growth. Plant Physiology and Biochemistry 35: 137-146


