Assembly, Function and Evolution of Cyanobacterial Carboxysomes

Cheryl A. Kerfeld\textsuperscript{1,2,3,4} and Matthew R. Melnicki\textsuperscript{3,4}

\textsuperscript{1}MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, MI 48824, USA
\textsuperscript{2}Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA
\textsuperscript{3}Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
\textsuperscript{4}Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA 94720, USA

SHORT TITLE:
Assembly, Function and Evolution of Carboxysomes

ABSTRACT
All cyanobacteria contain carboxysomes, RuBisCO-encapsulating bacterial microcompartments that function as prokaryotic organelles. The two carboxysome types, alpha and beta, differ fundamentally in components, assembly, and species distribution. Alpha carboxysomes share a highly-conserved gene organization, with evidence of horizontal gene transfer from chemoautotrophic proteobacteria to the picocyanobacteria, and seem to co-assemble shells concomitantly with aggregation of cargo enzymes. In contrast, beta carboxysomes assemble an enzymatic core first, with an encapsulation peptide playing a critical role in formation of the surrounding shell. Based on similarities in assembly, and phylogenetic analysis of the pentameric shell protein conserved across all bacterial microcompartments, beta carboxysomes appear to be more closely related to the microcompartments of heterotrophic bacteria (metabolosomes) than to alpha carboxysomes, which appear deeply divergent. Beta carboxysomes can be found in the basal cyanobacterial clades that diverged before the ancestor of the chloroplast and have recently been shown to be able to encapsulate functional RuBisCO enzymes resurrected from ancestrally-reconstructed sequences, consistent with an ancient origin. Alpha and beta carboxysomes are not only distinct units of evolution, but are now emerging as genetic/metabolic modules for synthetic biology; heterologous expression and redesign of both the shell and the enzymatic core have recently been achieved.
INTRODUCTION

Carboxysomes: metabolic modules for CO₂ fixation

More than 50 years ago, unusual polyhedral bodies were discovered in cyanobacteria by electron microscopy [1] (Figure 1A-C). They were subsequently identified as proteinaceous shells that were packed with the CO₂-fixation enzyme ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) and, accordingly, named carboxysomes [2]. Through identification of homologs of the delimiting shell proteins in microbial genomes, it is now known that carboxysomes are members of a class of architecturally-related structures, bacterial microcompartments (BMCs). The potential to form BMCs has been found in at least 23 bacterial phyla, where they play diverse roles in metabolism dictated by the function of the encapsulated enzymes [3,4]. The widespread distribution of BMCs in bacteria suggests frequent horizontal gene transfer of these genetic and metabolic modules [3-5]. However, despite their functional diversity, only two types of BMC are known to be involved in autotrophic metabolism (specifically, the Calvin-Benson Cycle): the alpha and beta carboxysomes, which are found in many chemoautotrophic bacteria, some purple-sulfur phototrophs, and all cyanobacteria. Among the myriad of BMC types, only carboxysomes have been identified in the Cyanobacteria.

The carboxysome is the core component of the cyanobacterial Carbon Concentrating Mechanism (CCM) (Figure 1E) [7-9] that also involves the active uptake of inorganic carbon into the cell and its intracellular accumulation primarily as HCO₃⁻ [10,11]. HCO₃⁻ diffuses through the carboxysome shell; within the lumen, an encapsulated carbonic anhydrase (CA) converts it into CO₂, promoting the carboxylation reaction of RuBisCO. Studies of carboxysomal proteins provided the first structural examples of the ubiquitous BMC shell proteins, the hexamers (BMC-H), pseudohexamers (BMC-T), and pentamers (BMC-P) that assemble into an apparently icosahedral shell [12-14] (Figure 2). Hexamers constitute the predominant building blocks of the shell; these proteins are typically perforated by a small pore at the symmetry axis that presumably allows polar molecules such as bicarbonate to pass [15]. Experimentally, it has been shown that the shell provides diffusional resistance to CO₂ [16], preventing its loss from inside
the carboxysome [17], and possibly provides a barrier to $O_2$, minimizing the unproductive oxygenation reaction of RuBisCO.

**Figure 2. Carboxysome assembly.**

A,B,C. Alpha carboxysome shells begin assembling concomitantly with aggregation of RuBisCO. D,E,F. Beta carboxysome shells begin assembling after RuBisCO aggregation has formed a densely-packed paracrystalline array. Symbols for individual components are defined in Table 1.

**Alpha and beta carboxysomes are distinct BMCs**

Although alpha and beta carboxysomes share similar general architectural and functional principles – a selectively permeable protein shell that encapsulates RuBisCO with CA – they differ in many details, including the type of RuBisCO (Form 1A in alpha, Form 1B in beta). The “core” operons (Table 1) for alpha (cso) and beta (ccm) carboxysomes differ in the form of carbonic anhydrase encapsulated (CsoSCA, a beta-class CA; CcmM, a gamma-class CA). In addition to CcmM, a subset of cyanobacterial species also package an additional beta-class CA encoded at a “satellite” locus outside of the ccm operon [18]. This protein, CcaA, is evolutionarily distant from CsoSCA; the only residues conserved between their primary structures are the Zn-coordinating residues of the active sites [13]. Finally, the alpha and beta core operons each contain an additional gene for a protein of unknown function: the ~850 amino acid protein CsoS2 and the ~230 amino acid protein CcmN, respectively. These two proteins have no sequence homology and their functions, other than being critical for carboxysome assembly, remain enigmatic [19-23].

In freshwater and terrestrial beta cyanobacterial species in which beta carboxysomes are prevalent, the ccm genes are encoded in the ccmKLMN operon, but genes for other components, some essential, are scattered throughout the genome in “satellite” loci (Table 1; recently reviewed in [3]). For example, in many organisms the genes for RuBisCO are encoded remotely from the ccm operon. Beta carboxysomes have not yet been purified to homogeneity, and other structural components may yet be identified; for example, comparative genomic evidence suggests that RuBisCO activase may be localized
to beta carboxysomes [18]. In contrast, the alpha cyanobacteria, which includes marine Synechococcus and Prochlorococcus clades, contain the canonical cso operon, which has recently been shown to be nested in a “super locus” which contains other structural components of the alpha carboxysome [3,24]. For example, csoS1D, which encodes a member of a distinct clade of BMC-T proteins that form stacked trimers with gated apertures [12,25], is encoded upstream of the cso operon in all alpha cyanobacteria. Computational analysis using all available Cyanobacteria genomes indicates that both the ccm and cyanobacterial cso superloci also contain widely conserved genes for ancillary or co-regulated functions, such as components of membrane-bound inorganic carbon uptake systems (Table 1) [3,24].

Assembly of alpha and beta carboxysomes

The distinctive structural components of the alpha and beta carboxysomes are likely related to fundamental differences in their modes of assembly (Figure 2). Cryo-electron tomography studies on Halothiobacillus neapolitanus, the chemolithotrophic model organism for alpha carboxysome research [26], revealed partially assembled shells with attached RuBisCO molecules [27], suggesting a co-assembly pathway for shell and cargo proteins. Indeed some of the internal, core proteins appear to be integrated into the alpha carboxysome shell, including CsoSCA [28-30]. Alpha carboxysomes are the only type of BMC for which it has been established that an intrinsically disordered protein (IDP) – CsoS2 – plays a key role in assembly and function [20]. IDPs are known to alter their conformation, including acquiring secondary and tertiary structure, in the presence of their binding partners [31,32]. Small Angle X-ray Scattering analysis has shown that, in the presence of RuBisCO, CsoS2 adopts a more compact fold and interacts with both RuBisCO and the shell [20]. CsoS2 is the third most abundant protein in the alpha carboxysome and is essential to carboxysome formation, as deletion mutants lack carboxysomes [20]. In contrast, in RuBisCO deletion mutants of alpha cyanobacteria, shell structures are still observed, resembling wildtype alpha carboxysomes in shape and size [26].

In beta cyanobacteria there is no evidence for formation of empty carboxysomes/shells when core components of the carboxysome are deleted. However when the shell proteins are expressed in E. coli, empty polyhedral shells are formed [6], however they are considerably smaller (~20 nm) than their packed counterparts (Figures 1C & 1D). Beta carboxysomes have been shown to assemble from the inside out; initially, soluble RuBisCO is nucleated by domains of the CcmM protein that resemble the small subunit of RuBisCO [33-36] (Figure 2D). The N-terminal gamma-class carbonic anhydrase domain of CcmM interacts with both the shell [6] and with the N-terminal domain of CcmN. In turn, CcmN contains a C-terminal encapsulation peptide, which is typically 15-20 amino acids [19]. This region, which is predicted to form an amphipathic alpha helix that interacts with the shell, is essential for the assembly of the shell around the catalytic core (Figure 2E) [19]. CcmN mutants lacking this extension do not form carboxysomes [19]. Notably, encapsulation peptides have been identified in all of the experimentally
characterized types of BMCs, including heterotrophic “metabolosomes”, with the notable exception of the alpha carboxysome [19,37,38].

Evolution of carboxysomes

For large multiprotein complexes, assembly pathways frequently reflect their evolution [39,40]. Given the similarities among the structural features that are essential for assembly [4], we suggest that beta carboxysomes are more closely related to the metabolosomes (heterotrophic BMCs) than to the alpha carboxysomes. This inference is supported by a phylogenetic analysis of all BMC-P shell protein sequences (Figure 3). The alpha cyanobacterial sequences (for both the conserved CsoS4A and CsoS4B paralogs) are nested among the BMC-P sequences from the alpha carboxysome loci of chemoautotrophs. Moreover, it is clear that the alpha CsoS4 proteins are evolutionarily distant from all other BMC-P sequences. In contrast, the CcmL proteins of beta carboxysomes form a group clustered among the BMC-P proteins from diverse metabolosomes. This provides evidence of a deep divergence between the two types of carboxysomes.

Figure 3. Evolution of alpha and beta carboxysomal BMC-Ps.

All BMC-P amino acid sequences were taken from the RP75 database scored with the PF03319 Pfam HMM for BMC-P domains, aligned in MUSCLE, made non-redundant to 95%, and used to build trees in PhyML.

These observations raise the question of how the two types evolved. Among all BMCs, sequence homology can be detected for both classes of shell proteins (BMC-P & BMC-H/T), implicating a common ancestry for the shell. Previously, it had been speculated that beta carboxysomes arose by shell recruitment (to encapsulate already-aggregating RuBisCO-CA complexes) whereas alpha carboxysomes arose by "invasion" of a heterotrophic BMC by RuBisCO [41]. It is clear that the alpha carboxysomes of Cyanobacteria share ancestry with their chemoautotrophic counterparts (which are predominantly found in Proteobacteria), as both groups maintain the core gene order of the cso locus (cbbL, cbbS, csoS2, csoS3, csoS4A, csoS4B) [3,20]. The extent of this conservation between two evolutionarily distant phyla implies horizontal gene transfer. Notably, genetic markers showing strong evidence for horizontal gene
transfer between proteobacteria and alpha cyanobacteria have recently been found, although the
directionality of this transfer was unclear [42]. From the BMC-P phylogeny (Figure 3) as well as that of
CsoS2 [20], it appears that the alpha cyanobacteria most likely were the recipients of a transfer event
which included the CsoS4A/B and CsoS2 genes. Given the conservation of proximal components, this
would suggest that the whole cyanobacterial alpha carboxysome operon was derived in this manner.

As with studies of the early events of cyanobacterial evolution, it is difficult to predict how
beta carboxysomes arose in cyanobacteria. The incorporation of RuBisCO into carboxysomes appears to be
sensitive to its specific form; for example, Form II RuBisCO cannot be encapsulated in either alpha
[26,43] or beta carboxysomes [44,45]. Even more strikingly, only one of the two Form 1A RuBisCOs, the
one associated with the carboxysome locus, encoded in the chemoautotroph Thiomicrospira crunogena,
can be encapsulated in alpha carboxysomes [26]. However, computationally-reconstructed ancestral
RuBisCO genes have recently been expressed and shown to assemble into functional L8S8 complexes
with similar kinetics [46]. Moreover, the large subunits for the resurrected forms (ancestral Form 1A,
ancestral Form 1B) could be encapsulated by extant beta carboxysomes, suggesting an ancient origin for
both types of carboxysomes.

While even the basal cyanobacterial genus, Gloeobacter, contains beta carboxysomes, genes for BMC
shell proteins cannot be detected in genomes of the Melainabacteria, a new class of aphotic bacteria
which are basal but monophyletic with the cyanobacteria [47] (although these were sequenced with
culture-independent methods and, accordingly, may have assembly errors or incomplete coverage). In
contrast, genes for shell proteins (presumably components of metabolosomes) have been found in all
phyla that include photosynthetic prokaryotes (Chlorobi, Proteobacteria, Chloroflexi, Firmicutes,
Acidobacteria, and the recently described Gemmatimonadetes) [3]. It is clear that the photosynthetic
machinery (particularly for the two photosystems) must have been transferred horizontally at least once
from these organisms to the cyanobacteria [48]. As metabolosome-associated BMC genes tend to be
frequently transferred horizontally, often as entire operons [3-5], it is plausible that carboxysome shell
proteins were acquired in an analogous (or even the same) transfer event.

There are several arguments which suggest that carboxysomes might have originated as recently as the
Phanerozoic (~0.4 giga-annum (Ga)), when global CO₂ concentrations are predicted to have dropped and
O₂ concentrations had begun to approach present levels [46]. This would place the development of
carboxysomes after the primary endosymbiosis event which led to the development of the chloroplast
[49], and could explain the absence of carboxysomes from plants and algae. Likewise, the presence of
homologous photorespiratory salvage pathways in cyanobacteria and chloroplasts has been taken to
suggest a relatively late development of the CCM. However, it is puzzling that some subclades of extant
cyanobacteria, such as Gloeobacter (subclade G as defined in [50]) and Pleurocapsa (subclade F as
defined in [50]), contain beta carboxysomes despite having branched off before the ancestor of the plant/algal chloroplast [49,50]. Were these genes acquired afterwards? Or perhaps carboxysomes were already present in some cyanobacterial lineages, but not in the endosymbiont. Because CCMs are able to promote calcium mineralization as a consequence of bicarbonate uptake, it has been speculated that carboxysomes could have facilitated cyanobacterial stromatolite formation, fossils dating from as far back as the early Proterozoic (2.5 Ga) [51]. Fossilized akinetes from specimens resembling those derived from filamentous heterocystous bacteria have recently been observed in cherts as old as 2.1 Ga [52], thereby providing a calibration point for a late step in the evolution of cyanobacteria. If carboxysomes had arisen this far back in geologic history, perhaps there was another functional advantage besides facilitation of CO₂ fixation and avoidance of photorespiration.

**Prospects**

BMCs can be considered units of evolution, with their broad species distribution, conserved locus structure, and considerable evidence for their frequent horizontal gene transfer [3-5]. These features likewise are establishing BMC loci as units of engineering, when viewed as genetically-encoded metabolic modules which can be transferred for plug-and-play applications in synthetic biology [53]. For example, the canonical cso operon for the alpha carboxysome has been expressed in *E. coli*, yielding structures with demonstrable RuBisCO activity [54]. Likewise, given the speculation that RuBisCO has already been maximally optimized by evolution [55,56], efforts are underway to install carboxysomes into chloroplasts in order to enhance crop productivity and nitrogen utilization efficiency [57,58]. Parallel efforts to re-engineer the carboxysome based on knowledge of protein-protein interactions involved in carboxysome assembly, first simplifying it genetically for facile transfer into diverse organisms, provides proof-of-concept of a domain fusion approach for engineering designed BMCs with novel functions [59]. These achievements not only substantiate our current understanding of carboxysome structure, function, and assembly, they also portend a new era in which industrially-relevant pathways could be encapsulated in proteinaceous organelles, recapitulating the ancient evolutionary innovation of compartmentalization of photosynthesis by endosymbiosis.

**ACKNOWLEDGEMENTS**

We thank members of the Kerfeld lab for engaging discussions, Dr. C. Raul Gonzalez for electron micrographs of cyanobacteria, and Dr. Fei Cai for micrographs of purified carboxysome shells and for a critical reading and assistance in revising this manuscript. This work was supported by the Office of Science of the U.S. Department of Energy DE-FG02-91ER20021 and with infrastructure support from MSU AgBio Research.
REFERENCES AND RECOMMENDED READING

* of special interest
** of outstanding interest

** This paper provides a comprehensive study of all BMC types, and helped to establish which accessory genes are conserved across carboxysome loci as well as identified satellite loci, and strengthened the concept of a cso "superlocus".


* First demonstration that a small peptide segment, the encapsulation peptide, is essential for carboxysome formation, specifically for interaction with the shell, and that peptides with similar properties are widespread among BMCs.

* This paper includes a wealth of structural, functional, and evolutionary findings about an intrinsically-disordered protein which has been difficult to characterize, discusses carboxysome assembly, and includes a phylogenetic tree supporting the alpha cyanobacteria as recipients of horizontally-acquired alpha carboxysome genes from chemoautotrophic proteobacteria.


* This paper not only revealed a new subtype of shell protein for alpha carboxysomes, but established the concept of a carboxysome “superlocus” and describes some of the accessory genes that are conserved in the context of ecotype.


35. Ludwig M, Sültemeyer D, Price GD: Isolation of ccmKLMN genes from the marine cyanobacterium, *Synechococcus* sp. PCC7002 (Cyanophyceae), and evidence that CcmM is essential for carboxysome assembly. Journal of Phycology 2000, 36:1109-1119.


43. Baker SH, Jin S, Aldrich HC, Howard GT, Shively JM: Insertion Mutation of the Form I cbbL Gene Encoding Ribulose Bisphosphate Carboxylase/Oxygenase (RuBisCO)
in *Thiobacillus neapolitanus* Results in Expression of Form II RuBisCO, Loss of Carboxysomes, and an Increased CO2 Requirement for Growth. *Journal of Bacteriology* 1998, **180**:4133-4139.


46. Shih PM, Occhialini A, Cameron JC, Andralojc PJ, Parry MAJ, Kerfeld CA: Biochemical characterization of predicted Precambrian RuBisCO. *Nature Communications* 2016, **7**:10382.


49. Criscuolo A, & Gribaldo, S.: Large-scale phylogenomic analyses indicate a deep origin of primary plastids within cyanobacteria. *Molecular Biology and Evolution* 2011, **28**:3019-3032.


**Assembly of shell components and demonstration of targeting using the encapsulation peptide in chloroplasts.


* This paper demonstrates that domains mimicking the small subunit of RuBisCO are essential for carboxysome formation.


** This paper presents growth phenotypes and electron micrographs of mutants deficient in several different components, and proposed the minimal structural determinants for the outer shell of beta carboxysomes.


<table>
<thead>
<tr>
<th>Symbol in Fig. 2</th>
<th>Name</th>
<th>Description</th>
<th>PFAM</th>
<th>Function</th>
<th>Locus Type</th>
<th>Distribution</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Alpha Carboxysome:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CsoS1D BMC-T Protein: stacked and gated shell protein</td>
<td>PF00936</td>
<td>Shell</td>
<td>Superlocus</td>
<td>All alpha cyanobacteria</td>
<td>[3,12,24]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>Ham1 HAM1-family Protein (alpha-NTP-Pyrophosphatase family)</td>
<td>PF01725</td>
<td>Unknown</td>
<td>Superlocus</td>
<td>All alpha cyanobacteria</td>
<td>[3,24]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>SbtA Sodium-dependent Bicarbonate Transporter</td>
<td>PF05982</td>
<td>C\textsubscript{2} uptake?</td>
<td>Superlocus</td>
<td>CB4 &amp; CB5 clades, <em>C. gracile</em> sub-cluster</td>
<td>[24,60]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>SbtB Bicarbonate Transport Regulatory Component: PII-like</td>
<td>PF00543</td>
<td>C\textsubscript{2} uptake?</td>
<td>Superlocus</td>
<td>CB4 &amp; CB5 clades, <em>C. gracile</em> sub-cluster</td>
<td>[60]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>CsoS1 BMC-H Protein: CsoS1</td>
<td>PF00936</td>
<td>Major shell protein</td>
<td>Core</td>
<td>All alpha cyanobacteria</td>
<td>[61,62]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>CbbL Form 1A RuBisCO Large Subunit</td>
<td>PF02788/ PF00016</td>
<td>CO\textsubscript{2} fixation</td>
<td>Core</td>
<td>All alpha cyanobacteria</td>
<td>[14]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>CbbS Form 1A RuBisCO Small Subunit</td>
<td>PF00101</td>
<td>CO\textsubscript{2} fixation</td>
<td>Core</td>
<td>All alpha cyanobacteria</td>
<td>[14]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>CsoS2 Conserved Carboxysome Protein: CsoS2</td>
<td>PF12288</td>
<td>Carboxysome assembly</td>
<td>Core</td>
<td>All alpha cyanobacteria</td>
<td>[20,21,23]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>CsoSCA Beta-class Carbonic Anhydrase</td>
<td>PF08936</td>
<td>CO\textsubscript{2} generation</td>
<td>Core</td>
<td>All alpha cyanobacteria</td>
<td>[13]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>CsoS4A BMC-P Protein</td>
<td>PF03319</td>
<td>Shell vertex</td>
<td>Core</td>
<td>All alpha cyanobacteria</td>
<td>[24]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>CsoS4B BMC-P Protein</td>
<td>PF03319</td>
<td>Shell vertex</td>
<td>Core</td>
<td>All alpha cyanobacteria</td>
<td>[24]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>CsoS1E BMC-H Protein, with extension</td>
<td>PF00936</td>
<td>Shell (BMC-H)?</td>
<td>Core</td>
<td>All alpha cyanobacteria except high light-adapted Prochlorococcus</td>
<td>[24]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>Bfr Bacterioferritin homolog</td>
<td>PF00210</td>
<td>Unknown</td>
<td>Superlocus</td>
<td>Sub-clusters 5.2, 5.3, &amp; 6b</td>
<td>[24]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>NdhF NDH-1 Complex, Subunit 5 (chain L)</td>
<td>PF00361/ PF00662</td>
<td>CO\textsubscript{2} uptake?</td>
<td>Superlocus</td>
<td>Sub-clusters 5.1 (A/B), 5.2, 5.3, &amp; 6b</td>
<td>[24]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>NdhD NDH-1 Complex, Subunit 4 (chain M)</td>
<td>PF00361</td>
<td>CO\textsubscript{2} uptake?</td>
<td>Superlocus</td>
<td>Sub-clusters 5.1 (A/B), 5.2, 5.3, &amp; 6b</td>
<td>[24,63]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>ChpXY CO\textsubscript{2} Hydration Protein (CupB)</td>
<td>PF10216</td>
<td>CO\textsubscript{2} uptake?</td>
<td>Superlocus</td>
<td>Sub-clusters 5.1 (A/B), 5.2, 5.3, &amp; 6b</td>
<td>[24,64]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>PCD-like Pterin-4-alpha-Carbinolamine Dehydratase homolog / acRAF RuBisCO Activase</td>
<td>PF01329</td>
<td>RuBisCO assembly?</td>
<td>Superlocus</td>
<td>All alpha cyanobacteria and chemoautotrophs</td>
<td>[3,24,65,66]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>CbbX RuBisCO Large Subunit</td>
<td>PF00004</td>
<td>RuBisCO assembly</td>
<td>Superlocus</td>
<td>Sub-clusters 5.1 (A/B), 5.2, 5.3, &amp; 6b</td>
<td>[24,67]</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Beta Carboxysome:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>ChpXY CO\textsubscript{2} Hydration Protein (CupB)</td>
<td>PF10216</td>
<td>CO\textsubscript{2} uptake?</td>
<td>Superlocus</td>
<td>Majority of beta cyanobacteria</td>
<td>[3,68]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>NdhM NDH-1 Complex Subunit</td>
<td>PF01059/ PF00361</td>
<td>CO\textsubscript{2} uptake?</td>
<td>Superlocus</td>
<td>Majority of beta cyanobacteria</td>
<td>[3,63,68]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>NdhF3 NAD(P)H Dehydrogenase Subunit (NdhF3 family)</td>
<td>PF00361/ PF00662</td>
<td>CO\textsubscript{2} uptake?</td>
<td>Superlocus</td>
<td>Majority of beta cyanobacteria</td>
<td>[3,68]</td>
<td></td>
</tr>
<tr>
<td>CcmK1</td>
<td>BMC-H Protein</td>
<td>PF00936</td>
<td>Shell</td>
<td>Core</td>
<td>Majority of beta cyanobacteria</td>
<td>[14,15,19,62,69,70]</td>
<td></td>
</tr>
<tr>
<td>CcmK2</td>
<td>BMC-H Protein</td>
<td>PF00936</td>
<td>Shell</td>
<td>Core</td>
<td>All beta cyanobacteria</td>
<td>[14,15,19,62,69,70]</td>
<td></td>
</tr>
<tr>
<td>CcmL</td>
<td>BMC-P Protein</td>
<td>PF003319</td>
<td>Shell vertex</td>
<td>Core</td>
<td>All beta cyanobacteria</td>
<td>[14,17,72]</td>
<td></td>
</tr>
<tr>
<td>CcmM</td>
<td>Gamma-class Carbonic Anhydrase / RuBisCO Nucleation Protein</td>
<td>PF00132/ PF14602/ PF00101/ PF00132</td>
<td>CO\textsubscript{2} generation/ carboxysome assembly</td>
<td>Core</td>
<td>All beta cyanobacteria</td>
<td>[34,35,73,74]</td>
<td></td>
</tr>
<tr>
<td>CcmN</td>
<td>Hexapeptide Repeat Protein with Encapsulation Peptide</td>
<td>PF00936</td>
<td>Carboxysome assembly</td>
<td>Core</td>
<td>All beta cyanobacteria</td>
<td>[19]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>CcmO BMC-T Protein</td>
<td>PF00936</td>
<td>Shell</td>
<td>Core/Satellite</td>
<td>All beta cyanobacteria</td>
<td>[75]</td>
<td></td>
</tr>
<tr>
<td>CcmK3</td>
<td>BMC-H Protein</td>
<td>PF00936</td>
<td>Shell</td>
<td>Satellite</td>
<td>Beta cyanobacteria</td>
<td>[76]</td>
<td></td>
</tr>
<tr>
<td>CcmK4</td>
<td>BMC-H Protein</td>
<td>PF00936</td>
<td>Shell</td>
<td>Satellite</td>
<td>Beta cyanobacteria</td>
<td>[15,62,76]</td>
<td></td>
</tr>
<tr>
<td>CcmP</td>
<td>BMC-T Protein: stacked and gated shell protein</td>
<td>PF00936</td>
<td>Shell</td>
<td>Satellite</td>
<td>Beta cyanobacteria</td>
<td>[25]</td>
<td></td>
</tr>
<tr>
<td>RbcL</td>
<td>Form 1B RuBisCO Large Subunit</td>
<td>PF02788/ PF00016</td>
<td>CO\textsubscript{2} fixation</td>
<td>Core/Satellite</td>
<td>All beta cyanobacteria</td>
<td>[77]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>RbcX Chaperonin Family Protein: RbcX</td>
<td>PF02341</td>
<td>RuBisCO assembly</td>
<td>Core/Satellite</td>
<td>All beta cyanobacteria</td>
<td>[78,79]</td>
<td></td>
</tr>
<tr>
<td>RbcS</td>
<td>Form 1B RuBisCO Small Subunit</td>
<td>PF00101</td>
<td>CO\textsubscript{2} fixation</td>
<td>Core/Satellite</td>
<td>All beta cyanobacteria</td>
<td>[77]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>CcaA Beta-class Carbonic Anhydrase</td>
<td>PF00484</td>
<td>CO\textsubscript{2} generation</td>
<td>Satellite</td>
<td>Some beta cyanobacteria</td>
<td>[80,81]</td>
<td></td>
</tr>
</tbody>
</table>