Functions, Compositions, and Evolution of the Two Types of Carboxysomes: Polyhedral Microcompartments That Facilitate CO₂ Fixation in Cyanobacteria and Some Proteobacteria

Benjamin D. Rae,* Benedict M. Long, Murray R. Badger, G. Dean Price
Division of Plant Science, Research School of Biology, The Australian National University, Canberra, ACT, Australia

SUMMARY
Cyanobacteria are the globally dominant photoautotrophic lineage. Their success is dependent on a set of adaptations collectively termed the CO₂-concentrating mechanism (CCM). The purpose of the CCM is to support effective CO₂ fixation by enhancing the chemical conditions in the vicinity of the primary CO₂-fixing enzyme, d-ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO), to promote the carboxylase reaction and suppress the oxygenase reaction. In cyanobacteria and some proteobacteria, this is achieved by encapsulation of RubisCO within carboxysomes, which are examples of a group of proteinaceous bodies called bacterial microcompartments. Carboxysomes encapsulate the CO₂-fixing enzyme within the selectively permeable protein shell and simultaneously encapsulate a carbonic anhydrase enzyme for CO₂ supply from a cytoplasmic bicarbonate pool. These bodies appear to have arisen twice and undergone a process of convergent evolution. While the gross structures of all known carboxysomes are ostensibly very similar, with shared gross features such as a selectively permeable shell layer, each type of carboxysome encapsulates a phyletically distinct form of RubisCO enzyme. Furthermore, the specific proteins forming structures such as the protein shell or the inner RubisCO matrix are not identical between carboxysome types. Each type has evolutionarily distinct forms of the same proteins, as well as proteins that are entirely unrelated to one another. In light of recent developments in the study of carboxysome structure and function, we present this review to summarize the knowledge of the structure and function of both types of carboxysome. We also endeavor to cast light on differing evolutionary trajectories which may have led to the differences observed in extant carboxysomes.

Address correspondence to G. Dean Price, Dean.Price@anu.edu.au.
*Present address: Benjamin D. Rae, University of Oxford, Department of Earth Sciences, Oxford, Oxon, United Kingdom.
Copyright © 2013, American Society for Microbiology. All Rights Reserved.
doi:10.1128/MMBR.00061-12
INTRODUCTION

Carboxysomes are specialized protein microcompartments composed of a polyhedral protein shell within which cyanobacteria concentrate CO$_2$ around their primary carboxylating enzyme, d-ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO). All photosynthetically competent cyanobacteria, and some autotrophic bacteria, possess carboxysomes. These structures form part of what is known as the CO$_2$-concentrating mechanism (CCM), which operates together with active CO$_2$- and HCO$_3^-$-uptake transporters which accumulate HCO$_3^-$ in the cytoplasm of the cell. Cyanobacteria (also known as blue-green algae) are predominantly aquatic organisms and are remarkably productive on a global scale, especially those from deep-sea niches. This productivity would not be possible without the CCM, which is dependent on carboxysomes. Some autotrophic bacteria are also conditionally dependent on carboxysomes and may have been the origin of some of the smaller shell proteins during evolution or may have obtained them from more distantly related bacterial microcompartments (BMC) used for unrelated biochemistry. Remarkably, two types of carboxysomes have arisen, possibly by convergent evolution, with similar designs and functions but different protein makeups. The types of carboxysomes are predominantly aquatic organisms and are remarkably productive on a global scale, especially those from deep-sea niches. This productivity would not be possible without the CCM, which is dependent on carboxysomes. Some autotrophic bacteria are also conditionally dependent on carboxysomes and may have been the origin of some of the smaller shell proteins during evolution or may have obtained them from more distantly related bacterial microcompartments (BMC) used for unrelated biochemistry. Remarkably, two types of carboxysomes have arisen, possibly by convergent evolution, with similar designs and functions but different protein makeups. The types of carboxysomes are α-carboxysomes, found predominantly in oceanic cyanobacteria (α-cyanobacteria), and β-carboxysomes, found mainly in freshwater/estuarine cyanobacteria (β-cyanobacteria), based on their RubisCO phylogeny. Considerable progress has now been made toward understanding the structure, physiology, and evolution of carboxysomes, and these subjects are explored in this review.

Evolutionary Pressures Due to RubisCO and Altered Atmospheric Conditions

Cyanobacteria can be traced back at least 2.4 billion years ago (Gya), and possibly as far back as 3.5 Gya, but the current thinking is that they did not always have carboxysomes or a CCM. Although an ancestral cyanobacterial endosymbiont gave rise to chloroplasts in algae and plants, the evolution of the CCM appears to postdate this pivotal event. In fact, the cyanobacterial CCM, with fully functional carboxysomes, may not have evolved fully until as late as about 350 million years ago (1–3), although the adoption of carboxysomes may have predated this point to some extent (2). The inefficiencies of the RubisCO enzyme have been the driving force for these changes. As oxygen has risen, since the advent of oxygenic photosynthesis by cyanobacteria, and CO$_2$ levels have fallen, evolutionary pressure has driven either the development of more efficient forms of RubisCO or the development of various CCMs (4). The inefficiencies of RubisCO include a low affinity for CO$_2$, a low catalytic rate, and the ability of O$_2$ to act as an alternative substrate. The advent of a carboxysome-based CCM in cyanobacteria ameliorated these problems with RubisCO and resulted in a high catalytic rate for carboxylation and a low oxygenation rate in vivo. As the full name of the enzyme conveys, RubisCO has the unfortunate ability to undertake the wasteful fixation of O$_2$ into phosphoglycolate. This product is toxic and must be removed quickly from the cell or recycled via an energy-requiring process known as photorespiration (5, 6). The ability of O$_2$ to react as a secondary substrate of RubisCO under the appropriate conditions is largely because CO$_2$ and O$_2$, as enzyme substrates, are remarkably similar in molecular size and lack of charge (7, 8).

When cyanobacteria first evolved, up to 3.5 Gya, the oxygenase activity of RubisCO was irrelevant, largely because early atmospheric conditions were very different, featuring high CO$_2$ levels (up to 50-fold higher than present levels) and almost nonexistent O$_2$ levels (9). These conditions presumably meant that RubisCO operated efficiently in its primary role as a carboxylase, fixing CO$_2$ into phosphoglycolate (PGA), the first 3-carbon sugar product of the Calvin cycle. Paradoxically, it was the early cyanobacteria that gradually decreased CO$_2$ and elevated O$_2$ in the atmosphere, due to their ability to perform oxygenic photosynthesis. Initially, most O$_2$ produced by cyanobacteria was removed by vast amounts of marine reductant, such as Fe$^{2+}$, such that atmospheric changes were minimal. Approximately 1.5 Gya, O$_2$ levels rose dramatically, to levels similar to today’s atmospheric level (~21%), and CO$_2$ levels dropped, although they were still well above today’s level of 0.04% (9, 10). This change in atmospheric conditions would have exposed RubisCO to limiting levels of CO$_2$ and inhibitory levels of O$_2$ and created an increase of evolutionary pressures to evolve better kinetic properties of RubisCO or different types of CCMs. These pressures were more acute in aquatic phototrophs because of the slow diffusion of CO$_2$ in water (some 10,000 times slower than in air) and other factors, such as pH, poor mixing, and temperature, that also affect CO$_2$ availability.

In the algal lineages that finally led to terrestrial plants known as C$_3$ plants, the RubisCO enzymes evolved an increased affinity for CO$_2$ and a better selectivity between CO$_2$ and O$_2$, but the trade-off appears to have resulted in a reduction in the catalytic turnover rate per enzyme (8, 11). This evolutionary response worked, but nevertheless, most C$_3$ plants still deal with rates of photorespiration of some 30% of the maximum potential photosynthetic rate, and as much as 25% of leaf nitrogen is allocated to this relatively slow enzyme (5, 6). Some land plants, such as the C$_4$ plants, eventually evolved several types of anatomically and biochemically compartmentalized CCMs that function to elevate CO$_2$ around RubisCO, thus overcoming its inefficiencies (12).

However, in cyanobacteria and some eukaryotic microalgae, diatoms, and coccoliths, evolutionary pressure on RubisCO resulted in the evolution of other types of CCMs that are single cell based and feature active accumulation of HCO$_3^-$ before delivering elevated CO$_2$ levels to RubisCO (1, 2, 4, 10, 13, 14). While some lineages, such as the green algae, evolved RubisCO enzymes with a higher specificity for CO$_2$ and more moderate abilities to accumulate HCO$_3^-$, cyanobacteria and some proteobacteria were able to maintain a RubisCO form that may resemble the form that was present in cyanobacteria >2.4 Gya. Cyanobacteria have retained a form 1 RubisCO with a high carboxylation rate (typical carboxylation reaction $k_{cat} [k_{cat}^{-1}]$ of 12 to 13 s$^{-1}$), a low affinity for CO$_2$ as a substrate (typical $K_m$ (CO$_2$) of 250 to 330 µM), and a poor selectivity for CO$_2$ over O$_2$ (typical $S_{O2}$ of 43 to 53) (4, 8). The latter two attributes are hardly a problem if the organism possesses a potent CCM, and accordingly, cyanobacteria have one that is able to reach inorganic carbon (Ci) accumulation factors of up to 1,000-fold over the level in bulk medium (4, 14, 15). Among the few alpha-, beta-, and gammaplannoebacteria that possess carboxysomes, it is inferred from what is known about their RubisCO kinetics (e.g., a low affinity for CO$_2$), such as those of Halothiobacillus species (16, 17), that they also feature high HCO$_3^-$ accumulation factors and a CCM. One of the few studies of carboxysome-

Downloaded from http://mmbr.asm.org on September 4, 2013 by Australian National Univ.
containing bacteria, which examined the deep-sea vent bacterium *Thiomicrospira crunogena*, showed that a 100-fold internal overaccumulation can be achieved with moderately high affinities for CO$_2$ and HCO$_3^-$ as substrates (18).

There are two key parts of the CCM in cyanobacteria and those bacteria that possess carboxysomes. First, it is essential to transport CO$_2$ and/or HCO$_3^-$ as a substrate, enabling accumulation of the relatively membrane-impermeative HCO$_3^-$ species in the cytoplasm. To prevent rapid leakage of accumulated HCO$_3^-$ from the cytoplasm, the enzyme carbonic anhydrase (CA) must be absent from this compartment (19, 20) to restrict CO$_2$ production (though this may not necessarily be true of some structurally complex cyanobacteria, such as *Chlorogloeopsis fritschii* and *Anabaena variabilis*, whose cells contain some cytoplasmic CA activity [21, 22]). Nonetheless, the accumulated HCO$_3^-$ needs to be converted to CO$_2$ and utilized by RubisCO in carboxysome microcompartments where CO$_2$ is elevated by localized production of CO$_2$ (1, 2, 4, 14, 23–26). The requirement of carboxysomes is known to be essential for proper CCM function, largely as a result of the analysis of a large number of characterized mutants in which carboxysome functionality is destroyed (see Table 2), leading to an inability to elevate CO$_2$ levels in the carboxysome, even though the cells are still able to hyperaccumulate HCO$_3^-$ in the cytosol (see below). In a functional carboxysome, RubisCO is substrate saturated for CO$_2$, and the oxygenase reaction is largely eliminated under optimal conditions (i.e., sufficient CO$_2$/HCO$_3^-$ availability and light). However, it is now known that cyanobacteria possess three interconnected pathways for metabolizing phosphoglycolate (27), and thus it is likely that cyanobacteria do, from time to time, encounter photorespiratory conditions which allow some RubisCO oxygenase activity, such as with dim light (morning and evening), low environmental CO$_2$/HCO$_3^-$ levels, or acclimation to limiting CO$_2$ levels. Photorespiration is not generally a concern when the carboxysomal CO$_2$ level is very low or when cells are in darkness, as cyanobacterial RubisCO has a low affinity for O$_2$, is rapidly inactivated under such conditions, and does not reactivate until sufficient CO$_2$, Mg$^{2+}$, and light are available (28, 29).

The presence of a functional CCM is requisite for the survival of cyanobacteria living in most natural habitats, although less so for carboxysome-containing bacteria, but the advantages are considerable. Being able to operate RubisCO at substrate saturation results in a lower allocation of cellular nitrogen, to as low as 3 to 5% of total cellular nitrogen in the case of cyanobacteria (19, 30), whereas C$_3$ plants (no CCM) allocate up to 25% of cellular nitrogen to RubisCO (31). Combined with the aforementioned highly reduced rates of oxygenase activity and photorespiration, cyanobacterial cells possess elements of high photosynthetic efficiency and environmental competitiveness. Hence, oceanic cyanobacteria and other phytoplankton contribute almost 50% of global primary productivity (as kg C fixed per m$^2$ per year) on an annual basis (32), with the oceanic α-cyanobacteria being one of the most productive groups of organisms, contributing as much as half of this productivity (33, 34), suggesting that >25% of all global CO$_2$ fixation occurs within carboxysomes.

**BASIC FEATURES OF THE CCM**

**Cyanobacterial Ci Uptake**

The two most biologically relevant species of dissolved inorganic carbon (Ci) for photosynthesis in aquatic environments are dissolved CO$_2$ and HCO$_3^-$, with HCO$_3^-$ being the more abundant species above a pH of approximately 6.5. However, both species serve as substrates for as many as three identified HCO$_3^-$ transporters and two CO$_2$-uptake systems (CO$_2$-to-HCO$_3^-$ converters) (Fig. 1). Physiological information linking genes to proteins for CO$_2$- and HCO$_3^-$-uptake systems is most complete for a few model strains of the β-cyanobacteria, while for the α-cyanobacteria and bacteria that possess carboxysomes, the linkage is fragmentary. However, the availability of >120 completely sequenced cyanobacterial genomes and thousands of bacterial genomes does make it possible to track the presence of the known Ci transporter gene homologs throughout the prokaryotes (1, 2, 35, 36). Not all cyanobacteria have genes for the five known transporter types, but in general, the β-cyanobacteria have more transporter homologs than the α-cyanobacteria (1, 36).

Cyanobacterial CO$_2$ fixation is supported by up to three types of plasma membrane-associated bicarbonate transporters: SbtA, a high-affinity Na$^+$/HCO$_3^-$ symporter from the TC2.A.83 family of Na$^+$/solute symporters (37–39); BicA, a medium- to low-affinity Na$^+$-dependent bicarbonate transporter from the SulP/SLC26 anion transporter family (40–42); and BCT1, a multimeric high-affinity bicarbonate ABC transporter (43, 44). In addition to these transporters, the homologous sbtA2 and bicA2 genes have been identified as candidates for bicarbonate transport in α-cyanobacteria (14, 35–37), though it may be the case that the SbtA2 protein has altered transport specificity with respect to the canonical SbtA transporter (45). The roles of SbtA2 and BicA2 in cyanobacterial CCMs remain unknown; however, these are the only known candidates for bicarbonate transport in the ecologically important genus *Prochlorococcus*, and hence, they could be the most abundant bicarbonate transporters in the biosphere (14).

In addition to bicarbonate transport, vectorial CO$_2$ uptake and intracellular CO$_2$ scavenging are achieved through two types of CO$_2$-uptake complexes, namely, NDH-I$_4$ and NDH-I$_6$, which are modifications of the NDH-I respiratory complex (46–49). These thylakoid-bound complexes achieve CO$_2$ uptake by vectorial CA activity, contributing to the intracellular bicarbonate pool by hydration of passively accumulated CO$_2$ (50–52). The two types of CO$_2$-uptake complex have different affinities for CO$_2$ (53), with transcription of the higher-affinity NDH-I$_4$ complex being induced in response to Ci limitation by model β-cyanobacteria (54–58). As speculated previously (20, 59–62), CO$_2$ which has escaped the carboxysome is recycled into the bicarbonate pool by these complexes, enhancing the efficiency of the cyanobacterial CCM (49).

It is likely that cyanobacterial Ci uptake is subject to posttranslational control resulting in rapid activation of HCO$_3^-$ transport (63–65). We direct readers interested in these and other aspects of cyanobacterial Ci uptake to recent and thorough reviews (14, 15).

**Proteobacterial Ci Uptake**

In contrast to the case for cyanobacteria, the Ci-uptake properties of carboxysome-containing proteobacteria are poorly understood (66). Despite this, it is clear that at least *Thiomicrospira crunogena* and *Halothiobacillus neapolitanus* (gammaproteobacteria) possess CCMs (18, 67), whereas *Thiobacillus versutus* (*Paracoccus versutus*; an alphaproteobacterium) does not (68). Similarly, different species appear to preferentially utilize different carbon species (CO$_2$ and HCO$_3^-$) for Ci uptake (18, 67). The general observation that many α-carboxysome-containing proteobacteria are acidophilic
(69–71), or even acidogenic, through metabolic oxidation of inorganic sulfur, suggests that the primary carbon species utilized by Ci-uptake systems for these CCMs is CO$_2$. Indeed, CO$_2$ is the predominant carbon species utilized by _H. neapolitanus_ (67), whereas _T. crunogena_ can utilize CO$_2$ or HCO$_3^-$ (18).

It has been speculated that extracellular or cytoplasmic CA enzymes may play a role in CO$_2$ fixation by facilitating the interconversion of carbon species. However, in _T. crunogena_, noncarboxysomal _/H9251_ and _/H9252_ CA enzymes (note that the _/H9251_, _/H9252_, and _/H9253_ classes of CA enzymes mentioned in this review are a separate classification system from that for carboxysomes and cyanobacteria) were recently shown not to have a direct role in CO$_2$ fixation (72). Similarly, the EcaB _/H9252_ CA enzyme from _Synechocystis_ strain PCC 6803 is probably not involved in CO$_2$ fixation, though its periplasmic targeting suggests that it may play some role in external exchange of Ci species (73).

**Critical Role for Carboxysomes**

Carboxysomes were first identified as polyhedral cell inclusion bodies of unknown function in cyanobacteria and chemotrophic gammaproteobacteria (74–78). Later work in the cyanobacterium _Anabaena cylindrica_ and the chemautotroph _Halothiobacillus neapolitanus_ showed that a cell’s RubisCO content and activity were associated with these polyhedral bodies, and the term “carboxysome” was coined (16, 79–86). Subsequently, it was shown in the model organism _Synechococcus elongatus_ PCC 7942 that virtually all of the cellular CA enzymatic activity was also associated with the carboxysome (19, 30, 87) (though this is not always the case in cyanobacteria [21, 22]), an observation that is highly consistent with this being the site of carbon fixation in _S. elongatus_ PCC 7942 and other cyanobacteria. A putative mechanism/model for effective carboxysomal Ci fixation emerged from these data, which we summarize briefly here.

The carboxysome encapsulates the RubisCO enzyme, with a selectively permeable shell layer that provides a diffusion barrier to CO$_2$ efflux and O$_2$ influx yet permits transit of ribulose-1,5-bisphosphate (RuBP), 3-PGA, Mg$^{2+}$, and HCO$_3^-$ between the carboxysomal and cytoplasmic pools. Within the lumen of the carboxysome, a carboxysomal CA enzyme dehydrates HCO$_3^-$ to CO$_2$, where it is fixed into 3-carbon sugars by RubisCO. Basic concepts inherent in early models of carboxysome organization and function (88–90) have largely been borne out by structure-function analyses. Reinhold and coworkers’ most developed model specifically placed the CO$_2$ diffusion resistance function at the carboxysome shell (88), and this function was subsequently confirmed in both _/H9251_ and _/H9252_ carboxysomes (19, 91–94). Similarly, current models of carboxysomal CA localization, where CA nuclei are present throughout the shell structure (95–99), are consistent with one model version (88) in which numerous “CA sites” are surrounded by “RubisCO zones” rather than carboxysomes which contain a monolithic CA core.

**FIG 1** Overview of the general CCM characteristics of _/H9251_ and _/H9252_ cyanobacteria, showing the types of Ci transporters typically present in each cyanobacterial type. Typical _/H9252_ cyanobacteria possess up to two types of CO$_2$ pumps (green) and up to three types of HCO$_3^-$ transporters (orange) and make use of _/H9252_ carboxysomes (form 1B RubisCO plus _ccm_ gene products), while typical _/H9251_ cyanobacteria possess only two or three identifiable Ci transporters and make use of _/H9251_ carboxysomes (form 1A RubisCO plus _cso_ gene products). In the case of _Prochlorococcus_ species (oceanic _/H9251_ cyanobacteria), the NDH-1-based CO$_2$ pump genes are entirely missing, and the only candidates for HCO$_3^-$ transport are unproven (blue; BicA2 and SbtA2).
TWO TYPES OF CARBOXYSOMES WITH DISTINCT COMPONENTS AND EVOLUTION

Tabita (100) showed that Rubisco form 1 sequences cluster into four distinct evolutionary groups, with oceanic cyanobacteria possessing Rubisco genes with phylogenetic affinity to those of proteobacterial species (Rubisco form 1A) and other cyanobacteria containing Rubisco genes with phylogenetic affinity to those of higher plants (Rubisco form 1B). Hence, the cyanobacteria came to be known as \( \alpha \)-cyanobacteria if they contained form 1A Rubisco and \( \beta \)-cyanobacteria if they contained form 1B Rubisco (Fig. 3). This concept was extended by Badger et al. (1), who showed that the protein/gene components of the carboxysome also typify this phylogenetic divide. Thus, cyanobacteria have come to be known as \( \alpha \)-cyanobacteria if they have form 1A Rubisco within \( \alpha \)-type carboxysomes (encoded by the cso operon) and as \( \beta \)-cyanobacteria if they contain form 1B Rubisco within \( \beta \)-type carboxysomes (encoded primarily by the ccm operon) (Table 1; Fig. 3 to 5). The names of the carboxysome types

\[ \text{FIG 2 Carboxysomes and their subcellular context (arrowheads in panels A and B indicate the positions of carboxysomes). (A) } \beta \text{-Carboxysomes present in } \text{Synechococcus elongatus PCC 7942. (B) } \alpha \text{-Carboxysomes present in Cyanobium PCC 7001. (Courtesy of Lynne Whitehead.) (C) Close-up of a } \beta \text{-carboxysome from } S. \text{ elongatus PCC 7942. (D) Close-up of a } \beta \text{-carboxysome from } \text{Anabaena variabilis M3. Note the size differences between the different types of } \beta \text{-carboxysomes in panels C and D.} \]

\[ \text{FIG 3 Phylogenies of representative sets of cyanobacteria, based on Rubisco large-subunit proteins (RbcL and CbbL proteins). Phylogenies were constructed as detailed previously (40). Note that oceanic cyanobacteria, i.e., } \alpha \text{-cyanobacteria (Prochlorococcus and Synechococcus species), have form 1A Rubisco enzymes, fall into two minor groups, and are readily distinguished from the more diverse } \beta \text{-cyanobacterial group (form 1B Rubisco).} \]
are shortened to simply “α-carboxysomes” and “β-carboxysomes.”

The phyletic distributions of the two types of carboxysomes are quite well understood. β-Carboxysomes are found in all subsections of cyanobacteria sensu Rippka et al. (101) and, to date, have not been found in any other bacterial lineage. Recently, it emerged that a notable diazotrophic strain, “cyanobacterium UCYN-A,” and those for H. neapolitanus C2 are from reference 66. Other references for data are indicated in parentheses. n, gross number of proteins per carboxysome; n_m, number of protein multimers per carboxysome.

* a Calculated as the difference between the total number of hexagonal units required to cover beta-Cbx (2,168) (147).

TABLE 1 Relative abundances of α- and β-carboxysomal proteins in model carboxysomes from Synechococcus elongatus PCC 7942 (β-cyanobacteria), Prochlorococcus marinus MED4 (α-cyanobacteria), and Halothiobacillus neapolitanus C2 (Gammaproteobacteria)*

<table>
<thead>
<tr>
<th>Protein category</th>
<th>Protein</th>
<th>S. elongatus PCC 7942 (2% CO2) n</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>β-Carboxysomes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CcmK2-4</td>
<td>6,863</td>
<td>1,143</td>
<td>Hexamer (147)</td>
</tr>
<tr>
<td>CcmO</td>
<td>3,075</td>
<td>1,025</td>
<td>Trimer (159)</td>
</tr>
<tr>
<td>CcmP</td>
<td>?</td>
<td>?</td>
<td>Trimer (161)</td>
</tr>
<tr>
<td>CcmL</td>
<td>60</td>
<td>12</td>
<td>Pentamer (123)</td>
</tr>
<tr>
<td><strong>CA enzymes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CcaA</td>
<td>1,058</td>
<td>529</td>
<td>Dimer (173)</td>
</tr>
<tr>
<td>CcmM-58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Structural proteins</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CcmN</td>
<td>?</td>
<td>?</td>
<td>Monomer? (143)</td>
</tr>
<tr>
<td>CcmM-58</td>
<td>2,177</td>
<td>725</td>
<td>Trimer (98)</td>
</tr>
<tr>
<td>CcmM-35</td>
<td>3,829</td>
<td>3,829</td>
<td>Monomer</td>
</tr>
<tr>
<td><strong>Rubisco enzyme</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RbcL</td>
<td>8,960</td>
<td>1,120</td>
<td>L_S6CcmM-35, (96)</td>
</tr>
<tr>
<td>RbcS</td>
<td>6,073</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>α-Carboxysomes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CsoS1AB</td>
<td>3,232</td>
<td>539</td>
<td>540</td>
</tr>
<tr>
<td>CsoS1C</td>
<td>0</td>
<td>2,970</td>
<td>495</td>
</tr>
<tr>
<td>CsoS1D</td>
<td>38</td>
<td>13</td>
<td>(112, 120)</td>
</tr>
<tr>
<td>CsoS1E</td>
<td>0</td>
<td>?</td>
<td>(112)</td>
</tr>
<tr>
<td>CsoS4AB</td>
<td>60</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td><strong>Prochlorococcus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MED4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Halothiobacillus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The protein interactions, sizes, and putative structures of carboxysomes are illustrated in Fig. 7 and 8. Data for S. elongatus PCC 7942 are from references 96 and 159; those for P. marinus MED4 are from reference 112, and those for H. neapolitanus C2 are from reference 66. Other references for data are indicated in parentheses. n, gross number of proteins per carboxysome; n_m, number of protein multimers per carboxysome.

a Calculated as the difference between the total number of hexagonal units required to cover beta-Cbx (2,168) (96) and the apparent number of CcmKx hexamers (1,143) (159).

b CsoS2 from MED4 is presumed to be equivalent to the unglycosylated form of CsoS2B from H. neapolitanus (112).
c CcmM is an active γ-CA enzyme in T. elongatus BP-1, and probably also in other species (138).
d From the crystal structure deposited in the Protein Data Bank (PDB entry 1SVD).

It has emerged that α-cyanobacteria are a monophyletic clade which diverged from planktonic, unicellular, freshwater β-cyanobacteria approximately 1.0 Gya (36, 106) (Fig. 4) and that they probably gained their cso operon, encoding the components of α-carboxysomes, by horizontal gene transfer from a gammaproteobacterial genus such as Nitrococcus (2, 107). Indeed, numerous species of alpha-, beta-, and gammaproteobacteria contain α-carboxysomes that support chemolithoautotrophic and mixotrophic growth, and thus their phylogic distribution is paraphyletic, probably involving horizontal gene transfer (1, 2, 11). The two types of carboxysome, while having distinct phylic distributions, also have distinct sets of protein components, which are discussed below. Given these observations and the apparent spread of the α-carboxysomal cso operon by horizontal gene transfer into disparate cyanobacterial and proteobacterial lineages (107), we view it as possible that the two types of cyanobacterial carboxysomes arose by convergent evolution, perhaps after the divergence 1.0 Gya of α-cyanobacteria and β-cyanobacteria, since, so far at least, none of the α-cyanobacteria phylogenetically definable by core gene sequences possess β-carboxysome genes, and vice versa (36). On the other hand, it is very likely that throughout history many cyanobacterial lineages have occupied (and, indeed, still occupy) environmental niches in which intra- or extracellular O₂ concentrations may limit CO₂ fixation, such as microbial mats (35); thus, the requirement for a carboxysomal CCM is likely an ancient one.

Further support for the hypothesis of convergent evolution is gained by the observation that the primary sequences of proteins comprising the shell structure of each type of carboxysome are more similar to the shell proteins of other types of bacterial microcompartments than to each other (108). Thus, despite their conserved functional role and their almost identical outward appearances, α- and β-carboxysomes have strikingly different internal structures. Nonetheless, both carboxysome types share several key features that underlie their conserved function. Below, we describe how convergent evolution resulted in these functions being achieved in remarkably different ways and through distinct structural paradigms.

Certainly all photosynthetically competent cyanobacteria, whether α-cyanobacteria or β-cyanobacteria, contain functional carboxysomes, or at least have the necessary genes. However, among other bacteria, the presence of carboxysomes is much more sporadic. α-Carboxysome-containing organisms can be found in the following groups:

- α-cyanobacteria (1), including α-cyanobacterial symbionts (109);
- sulfur-oxidizing bacteria of the genera Thiobacillus Kelly and Harrison (110) (now named Acidiphilium) (alphaproteobacteria), Thiobacillus and Thiomonas (both betaproteo-
bacteria), and *Halothiobacillus* and *Acidithiobacillus* (both gammaproteobacteria) (69, 111);

- *Acidimicrobium* (actinobacteria);
- nitrifying bacteria of the genera *Bradyrhizobium* (alphaproteobacteria), *Nitrosomonas* (betaproteobacteria), and *Nitrococcus* (gammaproteobacteria); and
- the α-cyanobacterial “chromatophore” of *Paulinella chromatophora* (107).

Among the autotrophic proteobacteria that possess Rubisco genes, it is possible to detect those that have carboxysomes and those that do not based on sequence differences in their Rubisco protein sequences (11). Thus, form 1A Rubisco sequences can be subdivided further into Rubisco form 1Ac, which is found associated with carboxysomes, and form 1Aq, with the putative Rubisco chaperonin gene *cbbQ* and without an obvious carboxysome operon.

**Two Distinct Rubisco Types**

As previously mentioned, /H9251- and /H9252-carboxysomes encapsulate Rubisco form 1A and form 1B, respectively (1). There is growing evidence that these are packaged into the carboxysome in remarkably different ways. As we describe in greater depth here, β-carboxysomes have significant internal structure, with Rubisco holoenzymes being scaffolded in three dimensions by products of the *ccmM* gene, whereas the interior of α-carboxysomes appears to be less ordered or possesses one or more ordered layers bound to the inner surface of the shell.

**Structure of α-Carboxysomes**

Study of the α-carboxysomes has revealed a complex structure composed of as few as 8 polypeptides in *Prochlorococcus marinus* MED4 (112) to as many as 11 polypeptides in *H. neapolitanus* C2 or 10 polypeptides reported by Heinhorst et al. (66), as well as CsoS1D (112). Apart from CbbL and CbbS, the large and small subunits of Rubisco form 1A, the remaining proteins fall into two categories: shell proteins of small size (CsoS1A to -E and CsoS4A and -B) and larger shell-associated proteins (CsoS2A and -B and CsoSCA).

The shell of α-carboxysomes is formed from lineage-specific subsets of the CsoS1 protein type, as well as the CsoS4A and CsoS4B proteins common to all α-carboxysomes (113, 114). All are small, 10- to 11-kDa proteins. The protein structures forming the outer shells of carboxysomes have been reviewed extensively (113–117). Briefly, crystal structures of the CsoS1A, CsoS1C, and CsoS1D proteins have been elucidated, showing that these pro-
teins contain the characteristic bacterial microcompartment (BMC) domain and form flattened, regularly hexagonal hexamers (CsoS1A and -C) or trimers (CsoS1D) (118–120), as BMC proteins from ethanolamine (EUT) and propanediol (PDU) microcompartments do (113). An important observation is that each hexamer type has a central pore of a size and charge distribution that may allow entry of specific substrates (HCO\textsuperscript{-3}, RuBP, and Mg\textsuperscript{2+}) or exit of products such as PGA, while retarding entry of O\textsubscript{2} and leakage of CO\textsubscript{2} (113, 118, 119). These BMC oligomers form sheets in crystallographic studies, which strongly suggests that the sheet structure represents the behavior of CsoS1 proteins \textit{in vivo}. Indeed, CsoS1A hexamers probably interact with one another very closely, more closely than other BMC proteins such as CcmK2 do, forming sheets which could be impermeable to CO\textsubscript{2} or O\textsubscript{2} (118). However, multiple CsoS1 homologues participate in the outer shell of all α-carboxysomes, potentially modifying these tight interactions.

The proposed mechanism for selective permeability of the shell relates to the presence of a conserved charged pore at the 6-fold axis of symmetry in the CsoS1x oligomers studied to date (113, 118, 119), which may be sufficient to allow preferential transit of charged molecules. Kinney et al. (113) suggested that the positively charged pores may promote passage or binding of negatively charged molecules, such as bicarbonate, while remaining indifferent to uncharged molecules, such as O\textsubscript{2} or CO\textsubscript{2}. Furthermore, the crystal structure of the CsoS1D protein, with each protomer consisting of tandem BMC domains (120), suggests a unique, dual-layer hexameric unit. The hexameric CsoS1D structure consists of two trimeric subunits, with distinct “open” and “closed” conformations, and Klein et al. (120) proposed that interchange between these conformations could gate the entry and exit of the comparatively large RubisCO substrate RuBP. Indeed, biochemical studies support these models, with the α-carboxysome shell being a barrier to CO\textsubscript{2} and RuBP transit (91, 92, 121, 122), but without compelling evidence for an RuBP transit role for CsoS1D (122).

The nature of icosahedral geometry requires a specialized shell protein to fit in the gaps left at the 6-fold axes of symmetry, i.e., the vertices. In α-carboxysomes, the CsoS4A and CsoS4B proteins achieve this function (123), as they were shown to close the shell, thus preventing CO\textsubscript{2} escape (92). The CsoS4A protein formed pyramidal pentamers in crystal structure, which comfortably fit into the 12 vertices in idealized models of α-carboxysomes (123).

Empty α-carboxysome shells can form due to mutations \textit{in vivo} (121, 124), and largely or partially empty shells are observed regularly in wild-type cells or α-carboxysome preparations (80, 125–129). The shell of α-carboxysomes appears to be a scaffold to which as many as two isoforms of CsoS2 attach, as well as the α-carboxysomal CA enzyme CsoSCA/CsoS3 (95, 130–133). In bacterial two-hybrid studies, the CsoS2 protein was shown to interact with many α-carboxysome proteins (131); however, whether its role is essential is currently unknown. The protein–protein interactions reported by Gonzales et al. (131) could be spurious; for instance, molecular models of the α-carboxysome shell suggest that there is little potential for interaction of the vertex protein CsoS4A or CsoS4B with luminal proteins (123). There is evidence that the csoS2 gene codes for a full-length protein (130 kDa) and a shorter protein (85 kDa) in \textit{H. neapolitanus}; however, the roles of these two forms are not known (130), and only the shorter, unglycosylated form is observed in α-carboxysomes from \textit{P. marinus} (112). Nonetheless, an interesting hypothesis is that CsoS2 organizes the subshell structure of α-carboxysomes in some way, although evidence for significant internal structure in α-carboxysomes is scarce (see below).

Given that there appear to be two distinct protein domains within CsoS2 (134), much as there are two distinct domains in CcmM (see below), it is appealing to apply the structural role of CcmM to CsoS2. Potentially, as RubisCO does appear to be organized into defined layers, at least in the immediate subshell layer, CsoS2 links RubisCO enzymes to the inner shell. This could allow RubisCO to take up most of the carboxysome interior, depending on the size of the carboxysome; indeed, CsoS2A and CsoS2B are present in quantities roughly equal to those of RubisCO holoenzymes (66), indicating a 1:1 stoichiometry within the α-carboxysome (Table 1). We suggest that this structure may in fact be the basis for the consistently small diameters of α-carboxysomes with respect to β-carboxysomes. In other words, if RubisCO is attached to the inner shell of α-carboxysomes by CsoS2, then this protein, with a fixed size, must thus operate over a fixed distance from the shell, especially since csoS2 does not appear to produce the same types of domain-specific isoforms as ccmM (66, 112). This structural feature would ensure that the ratio of inner surface area (CA-rich zone) to stromal volume (RubisCO-rich zone) remained high. Recent electrostatic potential maps of the form 1A RubisCO enzyme showed that this enzyme has a much lower surface charge than that of form 1B enzymes (135). Whether this is a chemistrustructural feature allowing electrostatic agglomeration of RubisCO enzymes within α-carboxysomes remains debatable, but we see this as a promising avenue for investigation. In fact, Holthuizen et al. (127) showed in \textit{H. neapolitanus} that the RubisCO large subunit was preferentially liberated from urea-disrupted α-carboxysomes—though this phenomenon is debated (121)—thus, they posited that the small subunit was strongly bound to the shell structure. This is supported by the work of Badger and Bek (11), who showed that synapomorphies in the RubisCO small subunit are the principal distinguishing features of a carboxysomal form 1A RubisCO enzyme with respect to noncarboxysomal form 1A RubisCO. Thus, it appears that RubisCO enzyme incorporation into α-carboxysomes acts through the small subunit of RubisCO form 1A, although Menon et al. (121) have produced evidence which suggests that the large subunit may also have important assembly information.

Structure of β-Carboxysomes

To date, the majority of the work on β-carboxysomes has been undertaken in the freshwater strains \textit{S. elongatus} PCC 7942 and \textit{Synechocystis} PCC 6803. In the former species, two CcmM proteins (58 and 35 kDa) are expressed from the \textit{ccmM} gene, and these are undoubtedly the most important structural proteins in the β-carboxysome (136, 137). The protein structures of CcmM-58 and CcmM-35 inform us of their specific roles within the β-carboxysome. The full-length CcmM-58 isoform has an N terminus on the size of the carboxysome; indeed, CsoS2A and CsoS2B are present in quantities roughly equal to those of RubisCO holoenzymes (66), indicating a 1:1 stoichiometry within the α-carboxysome (Table 1). We suggest that this structure may in fact be the basis for the consistently small diameters of α-carboxysomes with respect to β-carboxysomes. In other words, if RubisCO is attached to the inner shell of α-carboxysomes by CsoS2, then this protein, with a fixed size, must thus operate over a fixed distance from the shell, especially since csoS2 does not appear to produce the same types of domain-specific isoforms as ccmM (66, 112). This structural feature would ensure that the ratio of inner surface area (CA-rich zone) to stromal volume (RubisCO-rich zone) remained high. Recent electrostatic potential maps of the form 1A RubisCO enzyme showed that this enzyme has a much lower surface charge than that of form 1B enzymes (135). Whether this is a chemistrustructural feature allowing electrostatic agglomeration of RubisCO enzymes within α-carboxysomes remains debatable, but we see this as a promising avenue for investigation. In fact, Holthuizen et al. (127) showed in \textit{H. neapolitanus} that the RubisCO large subunit was preferentially liberated from urea-disrupted α-carboxysomes—though this phenomenon is debated (121)—thus, they posited that the small subunit was strongly bound to the shell structure. This is supported by the work of Badger and Bek (11), who showed that synapomorphies in the RubisCO small subunit are the principal distinguishing features of a carboxysomal form 1A RubisCO enzyme with respect to noncarboxysomal form 1A RubisCO. Thus, it appears that RubisCO enzyme incorporation into α-carboxysomes acts through the small subunit of RubisCO form 1A, although Menon et al. (121) have produced evidence which suggests that the large subunit may also have important assembly information.

Structure of β-Carboxysomes

To date, the majority of the work on β-carboxysomes has been undertaken in the freshwater strains \textit{S. elongatus} PCC 7942 and \textit{Synechocystis} PCC 6803. In the former species, two CcmM proteins (58 and 35 kDa) are expressed from the \textit{ccmM} gene, and these are undoubtedly the most important structural proteins in the β-carboxysome (136, 137). The protein structures of CcmM-58 and CcmM-35 inform us of their specific roles within the β-carboxysome. The full-length CcmM-58 isoform has an N terminus with structural and primary sequence similarity to γ-carbonic anhydrase enzymes such as Cam from \textit{Methanosarcina thermophila} (108, 138, 139). The C terminus of CcmM-58 contains three tandem copies of a protein domain with similarity to the RubisCO small subunit (SSU-like domains) (140). The SSU-like domains share up to 30% sequence identity with their cognate RbcS sequences, although the ~85-amino-acid SSU-like domain does not align with the N terminus of RbcS (141). Despite this, the SSU-like domains of the proteins from \textit{S. elongatus} PCC 7942, \textit{Synechococ-
PCC 7002, and *Synechocystis* PCC 6803 have a translocated sequence motif with similarity to the absent N-terminal region (141). As many as five SSU-like repeats are evident in predicted CcmM proteins from /H9252-cyanobacteria, though the extra SSU copy number variation has an unknown function, if any (96, 140–142).

The CcmM-35 isoform is translated from an internal ribosome entry site and consists of only the SSU-like domains (98, 141). This scheme of multiple CcmM isoforms expressed from a single gene has been observed in *Synechocystis* PCC 6803 (99, 141), *Synechococcus* PCC 7002 (141), and *Acaryochloris marina* MBIC11017, and all ccmM genes are expected to behave similarly (96).

In current models of /H9252-carboxysome organization and stoichiometry, at least 11 polypeptides form the β-carboxysomes from *S. elongatus* PCC 7942. These fall into two categories: proteins forming the β-carboxysome shell layers (inner shell bicarbonate dehydration/RubisCO-organizing layer, formed by CcaA, CcmM-58, CcmN, RubisCO CcmK2-4, CcmL, and CcmO; and outer shell BMC layer, formed by CcmK2-4, CcmL, and CcmO) and proteins forming the carboxysome lumen (CcmM-35, RbcL, and RbcS) (Fig. 5). CcmM-58 is restricted to the inner shell (96, 97). There, it simultaneously interlinks adjacent RubisCO molecules, recruits the carboxysomal carbonic anhydrase CcaA, and recruits the outer shell (97–99). The interaction between CcmM-58 and the outer shell is due to direct interaction with the outer shell protein CcmK2 (97, 99) and indirect interaction through the CcmN protein, which may link the outer and inner shell structures (143). In contrast, the CcmM-35 isoform is predicted to be confined to the carboxysome lumen, interlinking adjacent RubisCO enzymes in two planes (96), thereby organizing RubisCO into the paracrystalline array evident in some electron microscopic studies of β-carboxysomes (144) (Fig. 6A; see Fig. 8).

---

CUS PCC 7002, and *Synechocystis* PCC 6803 have a translocated sequence motif with similarity to the absent N-terminal region (141). As many as five SSU-like repeats are evident in predicted CcmM proteins from β-cyanobacteria, though the extra SSU copy number variation has an unknown function, if any (96, 140–142).

The CcmM-35 isoform is translated from an internal ribosome entry site and consists of only the SSU-like domains (98, 141). This scheme of multiple CcmM isoforms expressed from a single gene has been observed in *Synechocystis* PCC 6803 (99, 141), *Synechococcus* PCC 7002 (141), and *Acaryochloris marina* MBIC11017, and all ccmM genes are expected to behave similarly (96).

In current models of β-carboxysomes organization and stoichiometry, at least 11 polypeptides form the β-carboxysomes from *S. elongatus* PCC 7942. These fall into two categories: proteins forming the β-carboxysome shell layers (inner shell bicarbonate dehydration/RubisCO-organizing layer, formed by CcaA, CcmM-58, CcmN, RubisCO CcmK2-4, CcmL, and CcmO; and outer shell BMC layer, formed by CcmK2-4, CcmL, and CcmO) and proteins forming the carboxysome lumen (CcmM-35, RbcL, and RbcS) (Fig. 5). CcmM-58 is restricted to the inner shell (96, 97). There, it simultaneously interlinks adjacent RubisCO molecules, recruits the carboxysomal carbonic anhydrase CcaA, and recruits the outer shell (97–99). The interaction between CcmM-58 and the outer shell is due to direct interaction with the outer shell protein CcmK2 (97, 99) and indirect interaction through the CcmN protein, which may link the outer and inner shell structures (143). In contrast, the CcmM-35 isoform is predicted to be confined to the carboxysome lumen, interlinking adjacent RubisCO enzymes in two planes (96), thereby organizing RubisCO into the paracrystalline array evident in some electron microscopic studies of β-carboxysomes (144) (Fig. 6A; see Fig. 8).

CcmM-mediated organization of β-carboxysomes implicitly suggests a strong interaction between the SSU-like domains of CcmM and the RubisCO enzyme. It is intuitive that these occur through occupation by the SSU-like domains of the RbcS-binding site of the RubisCO large subunit. The SSU-like domains do appear to retain many of the hydrophobic residues thought to underlie RbcL binding (141), but it is well known that perturbation
of the RbcL-RbcS interaction or stoichiometry has kinetic consequences for the enzyme (121, 145). We have previously shown that the RbcL-RbcS subunit stoichiometry is greater in vivo than that reported for the L_{S0} RubisCO holoenzyme, that is, the subunit stoichiometry appears to be about L_{S0} in vivo (96). This was supported by our data showing that the RbcL-RbcS stoichiometry was rather more fluid than previously supposed, potentially supporting the hypothesis that the SSU-like domain occupies a varying proportion of the RbcS-binding sites. It is important that the RubisCO standard used for this work was a crystal structure-verified L_{S0} stoichiometry produced by using an Escherichia coli ectopic expression system (146). The exact interaction underlying the scaffolding of RubisCO-CcmM interactions, especially the possibility that SSU-like domains in CcmM displace RubisCO SSU or could catalytically mimic the role of the SSU, is therefore worthy of further examination.

In current models, a complex outer shell structure is attached to this scaffolded array of RubisCO and CA enzymes. The outer shell is composed of oligomeric protein complexes consisting of CcmK2-4 and CcmO, which contain the BMC domain (pfam number PF00936) (113, 114, 123, 147). These proteins were identified primarily by mutant analyses of S. elongatus PCC 7942 which revealed that knockout of any of the genes in the operon upstream of rbcL, ckmK-ccmO (ccmK-O) resulted in high-CO_{2}-requiring mutants (93, 94, 140, 148-154). Significantly, a homologous relationship between the CcmK and CcmO genes was established, with CcmO consisting of tandem repeats of a CcmK-like unit (150, 155)—the BMC domain. With the advent of genomic sequencing, two more ccmK homologues became apparent in the cyanobacterial genome: ccmK3 and -4 (136, 156, 157). The products of these genes came to be known as CcmK3 and CcmK4, and although CcmK4 has been implicated in functioning of the CCM in Synechocystis (158), the secondary functions of these proteins in the β-carboxysome only recently became apparent (159). We showed that CcmK3 and CcmK4 are individually required for maximum β-carboxyosome function in S. elongatus PCC 7942; however, the carboxysome is not entirely crippled in their absence, much in contrast to the case for other carboxysome shell proteins. Thus, the ΔccmK3-4 mutant is capable of slow growth in ambient CO_{2} (159).

As previously mentioned for α-carboxysomes, the icosahedral geometry of β-carboxysomes requires a unique structural element to completely close the carboxysome shell at the 5-fold axes of symmetry—the vertices. The CcmL protein, containing the PF03319 protein domain, forms pentagonal, pyramidal pentamers in crystal studies (123). These fit well into models of the outer shell of idealized β-carboxysomes, where they occupy the five-way interface between adjacent β-carboxyosome facets (117, 123). Early work showing characteristic rod-like carboxysomes in CcmL mutants generally supports this hypothesis, with rod-like carboxysomes being formed due to the inability of the cell to close the vertices (93, 94, 140).

Recent work suggests that a double layer of BMC proteins makes up the outer β-carboxysome shell (160). The hypothesis is intriguing, as the reported width of a double layer of CcmK2 hexamers (4 to 7 nm) (147, 160) is generally supported by electron microscopic observation of the shell width (5 to 6 nm) (144). Moreover, the presence of two outer BMC shell layers would be supported by the potential hexameric, double-layer conformation of CcmP reported by Cai et al. (161). At this stage, however, no direct evidence for the existence of a dual shell layer, or the presence of CcmP in it, has been gathered. A criticism of this hypothesis is that there are more protein-protein interactions at play in the vicinity of CcmK2 in vivo than in protein crystals. Does the presence of CcmN and CcmM-58, which are known CcmK2 interactors (97, 99, 143), alter the self-interactivity of CcmK2 or provide a directionality to the CcmK2 pore? It is also unclear why a double shell layer is a likely structure given that, assuming a single shell layer, there is in fact a reported ~30% shortfall of CcmK2 for β-carboxysome surface coverage (96) (Table 1). Indeed, further work is required to validate this structure. We previously speculated that CcmO could account for this shortfall (159), but the exact composition of the outer shell in our models remains speculative.

Structural insights into BMC proteins have hinted at the underlying basis for the observed diffusion resistances of the β-carboxysome shell to certain metabolites (19, 123, 147, 162). Pioneering crystal structures of CcmK1, CcmK2, and CcmK4 from Synechocystis PCC 6803 revealed their ability to form hexamers which are then capable of oligomerization, such that they form sheets or linear strips of hexamers (117, 147), the corollary being that sheets and strips probably represent valid β-carboxysomal structures. Each hexameric unit has a positively charged pore which is thought to mediate charged-solute transit, through an unknown mechanism (147). Significantly, no crystal structure has been published for the tandem BMC domain protein CcmO; however, its role in the outer shell of the β-carboxysome is clearly established (159, 163, 164). CcmO is expected to form a trimeric unit with properties and a structure similar to those of CcmK hexamers (Fig. 7) and tandem BMC proteins from other types of microcompartment, such as CsoSID (120), PduT (165), and EutL (166). Recently, a gene designated ccmP was identified as an orthologue of the α-carboxysomal csoSID gene (161). CcmP has not previously been identified as an essential β-carboxysome shell protein, but it is expected to recapitulate the curious crystal structure of CsoSID, providing a potential pore for large metabolite transit.

The ultrastructure of β-carboxysome shell mutants in S. elongatus PCC 7942 is consistent with major roles for CcmK2, CcmO, CcmL, and CcmN in the outer shell (93, 140, 143, 159). We have previously shown that CcmK3 and CcmK4 are not essential for carboxysome biogenesis; however, at least one of these proteins is required for maximum carboxyosome function and correct subcellular localization (159). Given that the amount of CcmK2 present in β-cyanobacterial cells would result in a shortfall of carboxyosome surface coverage (96), the protein-protein interactions between CcmO and CcmK2 (159), and the minor role taken by the CcmK3 and CcmK4 proteins, CcmO was proposed as a second major outer shell protein in β-carboxysomes (159). These observations led to speculative models for facets of the outer β-carboxysome shell (Tables 1 and 2; Fig. 6 to 8), with the bulk facet being made up of CcmK2, with various roles for the CcmO protein, but typically at the vertex, the facet-facet interface (icosahedral edge), or both (159).

An α-Carboxysomal Carbonic Anhydrase

Carboxysomal CO_{2} accumulation beyond the spontaneous supply rate is achieved through the action of carboxysome-specific CA enzymes. These convert HCO_{3}^{-} to CO_{2} in the vicinity of RubisCO, allowing saturation of the RubisCO active site with
As a functional dimer, CcaA is recruited to the carboxysome by CcmM-58 in Synechococcus elongatus PCC 7942. CcmK2, -K3, and -K4 produce flattened hexamers, and CcmK2 is by far the most abundant form. CcmO is postulated to form flattened trimers that could potentially interface with the triangular facets. The rarer CcmL pentamers would close the 5-fold vertices. The carboxysome model is not drawn to scale. Protein structure images were generated using Jmol (240), based on structures of the following proteins: CcmK (Protein Data Bank [PDB] entry 2A1B) (147), CcmO (represented by the structure of CsoS1D [PDB entry 3F56]) (120), and CcmL (PDB entry 2QW7) (147).

FIG 7 Model for the outer shell structure of β-carboxysomes from Synechococcus elongatus PCC 7942. CcmK2, -K3, and -K4 produce flattened hexamers, and CcmK2 is by far the most abundant form. CcmO is postulated to form flattened trimers that could potentially interface with the triangular facets. The rarer CcmL pentamers would close the 5-fold vertices. The carboxysome model is not drawn to scale. Protein structure images were generated using Jmol (240), based on structures of the following proteins: CcmK (Protein Data Bank [PDB] entry 2A1B) (147), CcmO (represented by the structure of CsoS1D [PDB entry 3F56]) (120), and CcmL (PDB entry 2QW7) (147).

TABLE 2 Carboxysome morphologies in α- and β-carboxysome mutants

<table>
<thead>
<tr>
<th>Carboxysome morphology</th>
<th>α-Carboxysomes</th>
<th>β-Carboxysomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type</td>
<td>cosS4Aβ (92)</td>
<td>ccmK3 (159)</td>
</tr>
<tr>
<td></td>
<td>cosS1A (177)</td>
<td>ccmK4 (159)</td>
</tr>
<tr>
<td></td>
<td>cosS3 (91)</td>
<td>ccmK3-ccmK4 (159)</td>
</tr>
<tr>
<td></td>
<td>cosS1D (122)</td>
<td>ccaA (73, 94, 140)</td>
</tr>
<tr>
<td>Elongated (vertex protein deficit)</td>
<td>cosS4Aβ (92)</td>
<td>ccmL (93, 94, 140)</td>
</tr>
<tr>
<td>Polar body (shell recruitment deficit)</td>
<td>ccmK2 (140, 159)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ccmO (159)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ccmN (143, 150, 151)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rhsSβ (153)</td>
<td></td>
</tr>
<tr>
<td>Aggregated (carboxysome localization deficit)</td>
<td>ccmK3K4 (159)</td>
<td>parA (181)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mreB (181)</td>
</tr>
<tr>
<td>Empty shell</td>
<td>rbcL (121)</td>
<td></td>
</tr>
</tbody>
</table>

*The α-carboxysome mutants were largely generated in H. neapolitanus, and the β-carboxysome mutants in S. elongatus PCC 7942, though Synechocystis PCC 6803 and Synechococcus PCC 7002 mutants are included in this table.

*Most carboxysomes in Halothiobacillus solis4Aβ:Km' are ultrastructurally normal, but the proportion that are elongated occurs that of the wild (92).

*When the H. neapolitanus cso operon was expressed without cosS1D, the resulting α-carboxysomes were slightly aberrant, though essentially wild type (122).

*Extended rhs reading frame mutant (153).

*Mutant 28 (151) is probably an insertion mutant of ccmM.

*The cyanobium mutant replaced the native rbcLS genes with the type II RubisCO rbcM gene, abolishing carboxysomes in Synechocystis PCC 6803 (176).

CO₂. The CA enzymes of both types of carboxysomes are associated with the shell structure (95–97, 99, 138).

The α-carboxysomal CA enzyme, CsoSCA (formerly known as CsoS3), was identified as the product of cosS3 in H. neapolitanus (95, 132, 167). Initially identified as the type enzyme of the novel ε class of CA enzymes (132), X-ray crystallographic investigation revealed that CsoSCA was a structural analogue of the previously characterized β class of CA enzymes (133). CsoSCA, while forming a 57-kDa homodimer, contains only one active site where other β-CA enzymes contain two (133). The consequence of this unusual structure on CA activity is minimal, with the activity of CsoSCA being sufficient to saturate CO₂ fixation (91, 168). CsoSCA is extremely tightly bound to the shell structure (168), but the cosS3:Kn' mutant has apparently normal carboxysomes (91), leading Cannon et al. (169) to conclude that CsoSCA is probably not an intrinsic shell component but a shell-associated component. Interestingly, CsoSCA (CsoS3) is redox inactivated under reducing conditions (132), like CcaA (see the next section), but it is not known if it is also Mg²⁺ dependent.

Two Types of β-Carboxysomal Carbonic Anhydrases

The canonical β-carboxysomal CA enzyme is also a member of the β-CA class. The β-carboxysomal carbonic anhydrase CcaA (ccaA; formerly known as icfA) (170, 171) is a component of the inner shell bicarbonate dehydration/RubisCO scaffolding complex of β-carboxysomes in S. elongatus PCC 7942, Synechocystis PCC 6803, and a number of other β-cyanobacterial species (30, 96–99, 170–172). As a functional dimer, CcaA is recruited to the β-carboxysome by CcmM-58 in S. elongatus PCC 7942 (98). Dimerization and enzyme activity were shown to be dependent on the extended C terminus of CcaA, which is the major point of difference between it and canonical noncarboxysomal β-CA enzymes (173). The enzyme activity of CcaA is restricted to the β-carboxysome (30, 172), probably because the presence of CA enzymes in the...
cytoplasm is injurious to the CCM (19). The enzyme activity of CcaA is low, however, though still sufficient to saturate β-carboxysomal CO₂ fixation (30). In fact, cytoplasmic CcaA is rapidly degraded, such that Ci uptake is not short-circuited by noncarboxysomal CA activity (97, 136). A further regulatory function of CcaA is that it is inactivated under reducing conditions, suggesting that any enzyme en route to the carboxysome is effectively inactive (30), and also implying that the carboxysome interior is a thio-redoxin-inaccessible compartment. CcaA is also Mg²⁺ dependent in a way similar to that of Rubisco (30). Since the free Mg²⁺ concentration may be downregulated in the dark, similar to the situation in plant chloroplasts, this suggested that CcaA is predominantly active in the light, like Rubisco (30). With respect to its role in β-carboxysome structure, inactivation mutants of ccaA possess β-carboxysomes with a wild-type appearance; however, they are physiologically impaired (73, 94, 174). This leads to the question of carboxysomal CA activity in species that lack an obvious ccaA gene (96, 99, 169, 175). As we discuss below, it appears that CcaA is the alternative β-carboxysomal carbonic anhydrase, with the γ-CA-like domain present in CcmM being most likely to be the original CA enzyme in evolutionary terms.

It has long been presumed that the N-terminal γ-CA-like domain of CcmM could provide β-carboxysomal CA activity in species that lack CcaA (99, 169, 175). However, no CA activity could be detected in the CcmM proteins from a number of species which also possess ccaA (99, 167). Recently, however, it emerged that the N terminus of CcmM from *Thermosynechococcus elongatus* BP-1 (which lacks ccaA) is a catalytically active γ-CA enzyme (138). The γ-CA-like domain is highly redox sensitive, suggesting once again that the carboxysome is a compartment that is inaccessible to redox equivalents—much like the way in which eukaryotic organelles operate (108). With the cytoplasm being a highly reducing environment in the light, this redox inhibition of γ-CA activity is probably an effective adaptation, like CcaA being redox inhibited and rapidly degraded in the cytoplasm, protecting the CCM from cytoplasmic CA activity (108, 138). The crystal structure of the γ-CA-like CcmM domain forms a homotrimer (138), which is consistent with observations of trimeric CcmM behavior in carboxysomes (98). Because CcmM is an integral protein to β-carboxysomes, it is present in all β-cyanobacteria. In comparison, CcaA is found in only a few β-cyanobacterial lineages (96). The phyletic distribution of CcaA broadly supports the hypothesis that ccaA arose in the freshwater genus *Cyanothece*, and its current distribution reflects horizontal gene transfer.

What, then, are the individual roles of CcaA and CcmM in the numerous β-cyanobacterial species that appear to possess both the ccaA gene and a CA-active CcmM enzyme (96, 108, 138)? It is perhaps likely that the CcmM protein, which appears to be the original β-carboxysomal CA enzyme, is primarily structural in these species, as it is in *Synechocystis PCC 6803* and *S. elongatus PCC 7942* (108). However, an identified peptide motif that correlates with γ-CA activity in the CcmM protein (138) suggests that many species contain two active carboxysomal CA enzymes. There is no obvious reason why CcaA or CcmM would be preferred over the other in β-carboxysomes. However, one possibility is that adoption of CcaA allows greater flexibility toward the fine adjustment of the CA/Rubisco CO₂ activity ratio, whereas a CA-active version of CcmM would tend to be far less flexible in this ratio. Previous modeling has shown that optimal CO₂ levels in the carboxysome are quite sensitive to the CA/Rubisco ratio, with too...
much CA activity leading to excessive leakage and too little leading to restriction of the fixation rate (94, 148).

Distinct Lumen Structures in α- and β-Carboxysomes

The current state of knowledge of the protein structures within α- and β-carboxysomes suggests a tantalizing possibility, namely, that α-carboxysomes are shell-centric structures, while β-carboxysomes are structurally lumen- or RubisCO-centric (Fig. 6A). Electron microscopic investigations support these structural hypotheses for β-carboxysomes. Kaneko et al. (144) showed that the lumen of β-carboxysomes is composed of a paracrystalline array of RubisCO molecules. This is consistent with current models of β-carboxysome structure and stoichiometry, in which RubisCO is regularly scaffolded in three dimensions by the small isoform of CcmM (96–98). Moreover, the enzymatic and structural components of β-carboxysomes appear to be part of a greater RubisCO scaffolding complex, where the lumen of the carboxysome is composed of CcmM-35 and RubisCO, with the large isoform of CcmM (CcmM-58) recruiting the carboxysomal carboxy anhydride to the inner shell bicarbonate dehy dration complex at the periphery of this structure (96–99). An important finding is that the French press/Triton-Percoll process for purifying β-carboxysomes removes about 70 to 80% of the principal shell protein CcmK2 (97, 98), yet carboxysomes from these preparations retain their icosahedral structure (30, 137). Thus, the selectively permeable outer shell appears to be associated weakly with the regularly ordered, RubisCO-centric β-carboxysome lumen; in fact, the importance of CcmM to this structure is demonstrated by the carboxysomeless phenotype of all reported ccmM mutants (97, 140–142), in which no traces of empty shell structures are apparent.

On the other hand, while RubisCO knockouts cannot be investigated because they are lethal to β-cyanobacteria, replacement of RubisCO form 1B with a noncarboxosomal form II RubisCO enzyme abolished all carboxysome structure in Synechocystis PCC 6803 (176). The primacy of the CcmM-RubisCO structure was detailed by Long et al. (97), who showed that by varying the stoichiometry of two CcmM isoforms in vivo, relevant β-cyanobacterial subcomplexes could be produced. The presence of the shell protein CcmK2 in carboxysome-rich fractions was dependent upon the presence of M58, indicating that the β-carboxysome shell is recruited to the inner RubisCO-CcmM structure (96, 97), probably via CcmN (143). Similarly, in β-cyanobacterial mutants deficient in the shell protein CcmK2, CcmO, or CcmN, carboxysome-like polar bodies in which the entire β-carboxysome shell is absent form, leaving a naked RubisCO-CcmM-CcaA complex (159). These phenotypes seem to be consistent with the RubisCO-centric hypothesis. In other words, the polar bodies arise through CcmM scaffolding of RubisCO and CcaA ad infinitum.

This regular and ordered luminal structure is not consistently apparent in α-carboxysomes. While it seems possible that CsoS2 is involved in the interior organization of the α-carboxysome (131), Iancu et al. (128) suggested that the interior of the α-carboxysome from H. neapolitanus is for the most part disordered and that much of the RubisCO is free-floating within the shell rather than being specifically anchored. Furthermore, it appears that only about half of the α-carboxysomes from Synechococcus WH 8102 and H. neapolitanus have some level of internal order; when ordered structures are apparent, these are usually within the RubisCO layers immediately below the shell (125, 129). Thus, the center of most α-carboxysomes studied by Iancu et al. (129), in addition to the entire α-carboxysome in almost half of those studied, was relatively disordered. These results match earlier studies of α-carboxysome structure in which α-carboxysomes were estimated to contain only one layer of RubisCO molecules, against the inside of the shell, and many were seen to be electron transparent beyond the immediate subshell layer (126). It is conceivable that this first RubisCO layer, representing a significant fraction of the luminal volume (especially in smaller carboxysomes), is the ordered RubisCO signal recently detected by small angle X-ray scattering (SAXS) analysis of whole Halothiobacillus carboxysomes (C. A. Kerfeld, personal communication).

These transmission electron microscopy (TEM) data are largely consistent with our knowledge of the less organized protein interactions within the α-carboxysome, in contrast to the highly ordered structure observed in β-carboxysomes (144). It appears that all of the non-RubisCO components of the α-carboxysome interact primarily with the carboxysome shell. The CsoS2 protein, which is presumed by some to organize the α-carboxysome lumen through its numerous protein-protein interactions (131), is bound to the inner surface of the shell (130), as is the α-carboxysomal CA enzyme CsoSCA (95). It is well established that empty α-carboxysome shells can form in the absence of form 1A RubisCO (121, 124) and that knockout of the csoS1A gene, encoding an α-carboxysomal shell protein, results in fewer carboxysomes per cell and reduces the proportion of cellular RubisCO localized to the carboxysome (177) rather than forming polar bodies as β-carboxysomal shell mutants do. These results, as well as the fact that crystallized sheets of the α-carboxysomal shell subunit CsoS1A are more closely bound than sheets of β-carboxysomal CcmK2 (118), are consistent with the shell-centric hypothesis of α-carboxysome structure.

Biogenesis of the Two Types of Carboxysomes

The putative differences in internal organization of the two types of carboxysome led us to propose two separate but related processes for the biogenesis of carboxysomes. The ability of β-cyanobacteria to express mutant carboxysomes with nonstandard subunit stoichiometries (96, 97), as well as shell-free polar bodies (29, 143, 150, 159), showed that the β-carboxysomal RubisCO-organizing complex, while being the principal underlying structure of the β-carboxysome, requires the simultaneous addition of the carboxysome shell for carboxysome biogenesis. That is, without the shell, a single β-carboxysome-like polar body is formed, and with the shell, about 4 or 5 carboxysomes are formed under Ci sufficiency. It is likely that the CcmN protein provides a key link between the RubisCO-organizing inner shell layer and the selectively permeable outer shell (143), but CcmM is also known to interact with the primary outer shell protein CcmK2 (97, 99) (Fig. 8). This scheme of simultaneous encapsulation of the nascent β-carboxysome is consistent with early electron microscopic studies of β-carboxysome formation (178), that is, subcarboxysomal RubisCO arrays, presumed here to be scaffolded by CcmM, form simultaneously into β-carboxysomes with a defined shell structure.

Targeting of proteins to the β-carboxysome is thus an intuitive process. The long and short forms of CcmM interact with most β-carboxysomal proteins (97–99, 143), and the only proteins that are not currently known to do so, CcmO (159) and CcmK3 (99), are expected to interact through other outer shell proteins (114,
Interestingly, recent work proposed that a short, C-terminal peptide sequence in CcmN is the underlying basis for its interaction with the outer shell of β-carboxysomes (143). In fact, the contribution of the 18 C-terminal residues of CcmN to outer shell recruitment appeared to exceed that of the CcmM-58 protein (143). This is similar to the schemes evident in PDU metabolosomes whereby the PduP enzyme is targeted to the interior of the PDU shell by a short N-terminal peptide sequence which interacts with the PduA and PduJ shell proteins (179, 180). It seems that this targeting scheme may be common to all types of micro compartment (143, 180).

In contrast to β-carboxysomes, it is more likely that α-carboxysomes form by infilling of the α-carboxysome shell with the enzymatic components (121, 124, 128). Menon et al. (121) are right to emphasize that this process is nonsequential and that the simultaneous lumen-infilling/shell construction model of Price and Badger (93) is probably more realistic for α-carboxysomes. It is possible that the CsoS2 protein is involved in this process, perhaps by direct interaction with RubisCO, CsoSCA, and the outer shell, as seen by Gonzales et al. (131). A specific RbsC synapomorphy identified by Badger and Bek (11) in form 1A RubisCO led to RbsC being proposed as the factor underlying specific incorporation of form 1Ac RubisCO into α-carboxysomes, a proposal supported by the work of Holthuijzen et al. (127). However, Menon et al. (121) proposed that it is the RubisCO large subunit that determines incorporation of RubisCO into the α-carboxysome in those species that have multiple types of form 1A RubisCO. The precise mechanism by which α-carboxysome-specific RubisCO enzymes are incorporated into carboxysomes remains unclear. Naturally, it is expected that a process similar to those elucidated for β-carboxysomes and PDU metabolosomes may be at work, though its molecular mechanism is still unknown.

**Intracellular Localization and Partitioning of Carboxysomes**

Recent work by Savage et al. (181) showed that β-carboxysomes are spatially distributed throughout the cell. Microfilaments and microtubule structures have long been observed with β-carboxysomes (182). This interaction with the bacterial cytoskeleton resolves previous questions of how carboxysomes are apportioned to daughter cells at mitosis and is reminiscent of the partitioning systems of bacterial chromosomes, plasmids, magnetosomes, chromosomes, etc. (183, 184). While a definitive link has been established between the presence of ParA, and perhaps MinD, and correct subcellular localization and mitotic partitioning of β-carboxysomes in *S. elongatus* PCC7942, it was only recently proposed that the outer shell protein homologues CcmK3 and CcmK4 could also be required for this function (159). These proteins are individually required for the correct subcellular localization and mitotic partitioning of β-carboxysomes, but the ΔccmK3-4 mutant phenotype was beyond the expected deficit due to aberrant carboxysome partitioning; thus, the CcmK3 and CcmK4 proteins probably have further, as yet unidentified roles.

There is scant evidence supporting the same mechanisms for localization and mitotic partitioning of α-carboxysomes. In some genera, such as *Nitrooccus* and *Halothiobacillus*, the α-carboxysomes can seem to be distributed randomly within the cell in electron micrographs (77, 79, 80, 185), although the same can occasionally be said of the β-cyanobacterium *S. elongatus* PCC 7942, which has unambiguously well-ordered β-carboxysomes when imaged by fluorescence confocal microscopy (181). Nonetheless, there are some instances where α-carboxysomes are observed with filamentous structures (128), though the functional relationship between these filaments and the α-carboxysome has not been established. It should be noted that although other types of microcompartment appear to be subject to cytoskeleton-dependent subcellular localization and are associated with filament structures (186, 187), ectopic expression of the *H. neapolitanus* carboxysome in *E. coli* resulted in disordered α-carboxysomes (122). Intriguingly, though, most noncyanobacterial *cso* operons have a homologue of *parA* in the extended operon. These are perhaps most similar to the *parA*/parF homologues carried by mobile plasmids and prophages, and homologues of the *cso*-associated *parA* gene are generally found in prophages or integrated conjugative elements. Whether or not these contribute to partitioning or subcellular localization of α-carboxysomes remains to be seen and should provide a focus for future investigation.

**GENE ORGANIZATION OF TYPICAL α- AND β-CARBOXYSOMES**

The *cso* Operon Encodes Components of the α-Carboxysome

To date, all species with α-carboxysomes encode these components in a single operon. Depending on the species, the extended α-carboxysome (*cso*) operon may also contain other CCM genes, such as the CO2-uptake operon *ndhF4D4-chpX* in many α-cyanobacteria (36, 143) and genes for putative form 1A RubisCO chaperonins (11). In addition, the *cso* operons of many species contain other, seemingly extraneous genes (143); for instance, a gene with homology to the *parA*/parF family of bacterial actin homologues is found within some noncyanobacterial *cso* operons.

As a specific example, the canonical *cso* operon from *H. neapolitanus* C2 (Fig. 5) carries the genes for the large and small subunits of RubisCO form 1A (*cbbL-cbbS*), a protein thought to organize the interior of α-carboxysomes (*csoS2*), and the α-carboxysomal carbonic anhydrase enzyme (*csoS3*). These are followed by five short genes encoding small proteins with functions in the shell of α-carboxysomes: *csoS4AB* and *csoS1A* to -C. The products of the *csoS1A* to -C genes share the BMC domain, whereas the *csoS4AB* genes encode proteins with a distinct but functionally related domain (pfam number PF03319). Recently, a gene near the canonical *cso* operon in α-cyanobacteria and proteobacteria, termed *csoSID*, was found to encode a protein with tandem BMC domains and that forms a quaternary structure similar to that of carboxysome shell proteins (120). In addition, the *csoS1E* gene, downstream of the *cso* operon in most α-cyanobacteria, encodes a protein that probably contains tandem BMC domains (112).

It is important that while all α-carboxysomes are thought to be transcribed from a single operon, these operons vary from species to species (Fig. 5). Roberts et al. (112) identified six types of α-cyanobacterial *cso* operons which had specific CCM-related and extraneous genes; one of these operon types lacked the *csoS1E* gene and was limited to the high-light-adapted *Prochlorococcus* clade. Interestingly, the six α-carboxysomal *cso* operon types closely match published phylogenies of α-cyanobacteria (36, 105), suggesting that the α-carboxysome was present in the ancestor of the α-cyanobacteria and has been distributed vertically by horizontal gene transfer since its initial introgression.
The ccm and Other Operons Encode the Components of the β-Carboxysome

In contrast to the conserved nature of the α-carboxysomal cso operon, the genes encoding components of the β-carboxysome are distributed throughout β-cyanobacterial genomes, although the core carboxysome component genes are most often found in the canonical ccm operon. Interestingly, in the cyanobacterial strains Synechocystis PCC 6803 and Synechococcus PCC 7942 (M. aeruginosa PCC 7806), the corresponding genes for ccaA, ccaC, and ccaD are found in a single genomic position downstream of the core ccm operon and in the assembly chaperones RbcX, with the core ccm genes. Significantly, the β-carboxysome can contain different types of CA enzyme depending on the β-cyanobacterial lineage, with some species utilizing the carboxysomal carbonic anhydrase (CcaA) enzyme (170, 172) and some using the γ-CA-like domain in CcmM for this purpose (138). This isolated genomic position of ccaA is consistent with the view that ccaA was a late incorporation into β-cyanobacterial lineages.

Regulation of CCM/Carboxysome Genes

In general, the set of genes that make up the β-carboxysome are constitutively expressed with regard to a range of CO2 levels. However, fine adjustments in the relative expression of specific carboxysome gene transcripts have been observed, particularly at low CO2 levels. Studies using quantitative PCR analysis of transcript changes upon transfer of cells from 2 to 5% CO2 to air (Ci limitation) have found little change or transient small changes (2- to 5-fold increases) in the ccmKLMO, ccaA, and rbcLXS genes (56, 188, 189). Changes found using microarray analysis have generally been small (190, 191). In comparison, low-CO2-responsive genes (coding for CO2- and HCO3 uptake systems such as bicA except in PCC 6803), sbtAB, cmpABCDS, porB (in PCC 7002), and ndhF3-ndhD3-clpY undergo large changes in transcript abundance, although sometimes these changes are transiently maximal over the first 2 h from the onset of induction (56, 188–191). An electron microscopic investigation found that carboxysome numbers per cell for cells grown under severe Ci limitation can be up to 4-fold higher than those for high-CO2-grown cells of Synechococcus UTEX 625 (equivalent to PCC 6301) (192). However, much of this extra carboxysome accumulation could occur due to a markedly increased time between cell divisions while maintaining a relatively constant synthesis of carboxysome components. Other studies have shown that there is a degree of coregulation between ccmM and ccaA expression and that there is permissible flexibility in the relative abundances of carboxysome components before reductions in carboxysome function become apparent (96).

In comparison, there is little information on the regulation of carboxysomal genes in α-cyanobacteria. This has, however, been studied in some proteobacterial species, which for the most part contain distinct carboxysomal and noncarboxysomal Rubisco operons (11). Expression of the carboxysomal Rubisco operons in these organisms may not necessarily be responsive to changing Ci concentrations (66), though it has been observed that some species prefer noncarboxysomal Rubisco types during Ci-replete periods and that some repress all CO2 fixation during heterotrophic growth (193, 194). Thus, these species are in a constant process of metabolic streamlining, allowing them to forego the costly processes of α-carboxysome biogenesis and CO2 fixation.

REGULATORY ROLE FOR MICROCYSTINS IN THE β-CARBOXYSOME?

While carboxysomal Rubisco binds some proteins as a structural necessity, a number of other proteins are known to associate with Rubisco or carboxysomes. Notably, thioredoxin was observed to bind CcmM and Rubisco in Synechocystis PCC 6803 (195, 196), and the toxic nonribosomal cyclic peptide microcystin was shown to bind Rubisco in Microcystis aeruginosa PCC 7806 (197). The role of thioredoxin in the former example, or even whether this was an artifact of artificial cell fragmentation, was not discussed; however, further work in M. aeruginosa suggested that microcystin-bound Rubisco is less susceptible to proteolytic degradation. These results are generally supported by electron microscopic investigations of microcystin localization in which β-carboxysomes from M. aeruginosa PCC 7806 were hot spots of microcystin localization (198). The role of microcystins in the carboxysome is unclear; however, their production is strongly linked to the transition from high to low Ci in M. aeruginosa PCC 7806, with the microcystin maximum preceding peak Ci uptake (199). Some data suggest a relationship between microcystin production and photosynthesis (199–202). Perhaps, then, microcystins play a role in adaptation of cyanobacteria to limiting Ci in some environmental niches.

We perceive similar functional roles for microcystins and thioredoxins in cyanobacteria. These molecules could act as thioprotectants, reducing redox damage to carboxysome components during carboxysome biogenesis and maturation. Similarly, these moieties may alter carboxysomal CA enzyme activities depending on their subcellular location, i.e., as they transit the cytoplasm before incorporation into the nascent carboxysome. Thus, thioredoxins and microcystins may enhance carboxysome biogenesis and also protect the cytoplasmic HCO3 pool from CA activity during CA enzyme production, but more work is needed to determine a definitive link.

THE PYRENOID: A EUKARYOTIC ANALOGUE?

Similar to the CCMs of prokaryotes, most eukaryotic algae have structural and enzymatic features which enhance CO2 fixation: intracellular Ci accumulation by Ci pumps (sometimes assisted by extracellular carbonic anhydrase enzymes) and subsequent enzymatic conversion to CO2 in a Rubisco-rich compartment (24). While the carboxysome provides an ultrastructural focal point of
the many prokaryotic CCMs, most eukaryotic algae and some basal bryophytes possess an alternative macromolecular structure that is thought to be of similar importance, namely, the pyrenoid (203–205). The pyrenoid is the region of chloroplast stroma which contains the majority of plastidic RubisCO; as such, it is the site of algal CO₂ fixation (206, 207). To our current knowledge, a prerequisite for pyrenoids is the presence of an enzymatic CCM; however, pyrenoid-free eukaryotic CCMs do exist (208). Like the case for carboxysomes, multiple distinct morphotypes of pyrenoids are evident (4, 209); however, the structural components and function of these structures are currently poorly understood. It is known that some pyrenoids contain pyrenoid-localized CA enzymes (210), whereas others contain thylakoid invaginations/strands which may contain thylakoid-localized CA enzymes (211). Indeed, Markelova et al. (212) showed that the CAH3 CA enzyme was localized to both the thylakoid lumen and the pyrenoid body in *Chlamydomonas reinhardtii*, with this CA enzyme supporting CO₂ fixation only when correctly localized to the pyrenoid-traversing thylakoids (213).

While the presence of CA activity in pyrenoids is long resolved, the structures of a potential pyrenoid shell or internal organizing proteins are unknown. Genkov et al. (214) recently established that assembly of the *C. reinhardtii* pyrenoid is dependent on the presence of the native RubisCO small subunit, as strains in which this had been replaced by the SSU from *Helianthus annus* lacked this body. Similarly, the CIA6 protein appears to play a vital, if unclear, role in pyrenoid formation, with RubisCO being distributed throughout the chloroplast stroma in a *ciat* mutant (215). The latter study showed for the first time that the pyrenoid is required for proper function of the *C. reinhardtii* CCM, but the essential role of CIA6 remains unclear, with the intriguing possibilities being that this protein promotes pyrenoid formation through posttranslational modification of RubisCO or through linking of adjacent RubisCO holoenzymes, perhaps through their SSU as seen in α-carboxysomes (121, 215). Thus, there appear to be mechanistic similarities between carboxysomes and pyrenoids, but whether pyrenoids achieve a carboxysome-like level of structural sophistication remains an open question. It is nonetheless clear that the pyrenoid is paracrystalline at the ultrafine scale, with regular ordering of individual units apparent in early electron microscopic studies (we now interpret these units as RubisCO holoenzymes) (216–219).

The uncertainty over pyrenoid structure extends to the possibility of a pyrenoid shell structure, a concept which has been discussed extensively and seems unlikely (220). No candidate shell proteins have been identified, and early speculations suggesting a shell-like role for the starch sheaths commonly surrounding pyrenoids have proven to be unfruitful (221). An important consideration in assessing the necessity of a pyrenoid shell structure is the kinetics of RubisCO enzymes from pyrenoid-containing organisms, which may account for the lack of a shell through altered substrate specificity (4). There is a possibility that a semipermeable shell structure may be unnecessary for effective pyrenoid function, especially given that the CO₂/O₂ specificity of most algal RubisCO enzymes is higher than that of the cyanobacterial enzymes (4, 222). Indeed, an evolutionary comparison of carboxysomes, pyrenoids, and pyrenoid-free chloroplasts with their respective RubisCO enzymes showed that the more complex structure is inversely correlated with the specificity of the cognate RubisCO enzymes (4).

The concept of pyrenoids as “skinless carboxysomes” is not an alien one. Reinhold et al. (88–90) proposed that a carboxysome (or any macromolecular RubisCO complex) need not have an exclusive shell structure if it has sufficient internal CA activity, whether it be focused at the center of the body or at multiple foci as current models of β-carboxysomes contend (96–99). Further work to investigate this hypothesis would increase our knowledge of the largely unknown structure of the pyrenoid.

**RELATEDNESS OF CARBOXYSOMES TO OTHER BACTERIAL MICROCOMPARTMENTS**

While it is tempting to view carboxysomes in isolation, BMC diversity is not limited to carboxysomes. Rather than a unique innovation of the CO₂ fixation pathway, bacterial microcompartments are widespread in bacteria (113, 114, 155). The best studied of these are the propanediol (PDU) (223, 224) and ethanolamine (EUT) (225, 226) detoxification organelles, called metabolosomes or enterosomes. Additionally, comparative genomic approaches have revealed many putative operons encoding novel bacterial microcompartments with a wide range of functions, with some estimates suggesting that one-fifth of bacterial species possess the genetic capacity for microcompartment formation (114, 143). A common feature of these bacterial microcompartments is the sequestration of problematic metabolic pathways. For instance, while carboxysomes are thought to enhance RubBP carboxylation and to minimize oxygenation, PDU and EUT metabolosomes protect the cell from damage by sequestering reactive metabolic intermediates (227, 228).

The evolutionary relationships between different types of microcompartments are unclear, but numerous parallels are evident between them. They all contain a conserved outer shell structure, encapsulating problematic enzymatic reactions. The PDU and EUT microcompartments are reviewed in detail elsewhere (114–116).

**POSSIBLE EVOLUTIONARY SEQUENCES THAT GAVE RISE TO FUNCTIONAL CARBOXYSOMES**

Various mutant carboxysome phenotypes have been described which suggest an intriguing dichotomy in the evolution of the two types of carboxysome. As discussed above, β-carboxysomes appear to have high intrinsic order in three dimensions, their internal enzymatic structure could be largely self-assembled, and the outer semipermeable shell would be attached loosely to this central lumen. On the other hand, α-carboxysomes undoubtedly have a strong outer shell structure, which was shown to form even in the absence of the RubisCO enzyme (121). This could suggest the hypothesis that α-carboxysomes evolved by recruitment of the CO₂ fixation machinery to existing BMC protein shells in primordial cells with the genetic ability to produce both.

In contrast, the available evidence tends to support an opposite evolutionary scheme for β-carboxysomes—the underlying bicarbonate dehydration/RubisCO-organizing complex (CcmM–58–CcaA–RbcLS–CcmM–35) has immediate appeal as an interim functional unit in the absence of the outer shell.

**Simple Model for Evolution of β-Carboxysomes from a Naïve Cell**

The β-carboxysome may have evolved from a primordial photosynthetic CO₂ fixation structure made from RubisCO and γ-CA (proto-CcmM) enzymes, while the α-carboxysome may have
evolved by capture of a preexisting BMC shell structure. This highly ordered structure led to the speculative model outlining the evolution of the β-carboxysome as presented below.

A set of possible steps in the evolution of a β-carboxysome, as we see them, is shown in Fig. 9. The first step may have involved a loose arrangement of RubisCO and CA, possibly giving a small advantage to the kinetics of RubisCO by maintaining the local supply rate of CO$_2$ from ambient cellular Ci (Fig. 9, step 1). A second embellishment may have allowed specific colocalization of RubisCO and CA, possibly involving a third organizing protein. A contemporary example of this arrangement is in the Δ$ccmM ccmM-SD$ strain (97), and intriguing $ccmM$-like genes have been observed in the genera Acaryochloris and Leptolyngbya, where the γ-Ca domain is fused to a single RubisCO SSU-like domain (96); once again, the kinetic advantage may have been quite small. A third development may have allowed a trimeric γ-Ca to be fused to a RubisCO-organizing protein, producing an ordered RubisCO-CA aggregation (Fig. 9, step 2), or essentially a naked carboxysome looking like a Δ$ccmK2,$ Δ$ccmO,$ or Δ$ccmN$ mutant (143, 159). The most important step in the evolution of carboxysomes would have been the acquisition of the small BMC shell proteins from an existing organism with a BMC used for another purpose (Fig. 9, step 3). Although this step may have led to a significant elevation of CO$_2$ around RubisCO, the full advantage of the system would not have become apparent until systems for active HCO$_3^−$ uptake, and then CO$_2$ uptake, had evolved (2) so that the concentration of HCO$_3^−$ entering the carboxysome compartment could lead to maximal elevation of CO$_2$ around RubisCO. Other refinements of these hypothetical primitive carboxysomes would have involved molecular evolution of the pores in CcmK-type hexamers to control HCO$_3^−$ and RuBP entry (and PGA exit) while also allowing retardation of CO$_2$ loss and O$_2$ entry. Other refinements may have followed, such as abandoning the γ-Ca function of CcmM (while retaining the RubisCO-organizing function) after the recruitment of the alternative CA enzyme CcaA (Fig. 9, step 4), potentially allowing better control over the CA/RubisCO ratio. Modeling (94, 148) has shown that too little CA or too much CA activity in the carboxysomes can have negative effects on the efficiency of CO$_2$ fixation in the structures (too little CO$_2$ supply in the first case and too much leakage in the second case). The possibly older strategy of employing γ-Ca activity associated with CcmM trimers seems to be present in some of the filamentous cyanobacteria with large carboxysomes, such as Nostoc PCC 7120, whereas many unicellular strains (with smaller carboxysomes), such as S. elongatus PCC 7942, seem to have adopted the CcaA strategy (96) (Fig. 6B). The latter strategy might also have evolved a fine alteration of CcmN for the flexible recruitment of CcaA and the outer shell, as described previously (143).

**Simple Model for Evolution of α-Carboxysomes from Preexisting Shell Structures**

Whereas possible insights into β-carboxysome evolution have arisen from the use of β-cyanobacterial mutants with carboxysome phenotypes ranging from absent to wild-type appearance, the evolution of the α-carboxysome system is less clear, and a putative scheme is briefly described in Fig. 9. The incorporation of enzyme subunits through connection to the shell structure is a common feature of α-carboxysomes and PDU microcompartments (180), so it is conceivable that incorporation of form 1A RubisCO into the α-carboxysome may have arisen by invasion of an alternative form of microcompartment by divergent RubisCO and CA enzymes which were capable of incorporation into the forming microcompartment by using this targeting method (Fig. 9, step A) (2). This hypothesis relies on the coexistence of RubisCO and one of the alternative BMC types in the same bacterial strain, a scenario that is readily observed in database searches. Refinement of the BMC structure would then have followed (Fig. 9, step B). Interestingly, there are numerous examples of bacterial strains which contain both the $pdu$ and $eut$ operons simultaneously, such as *Salmonella enterica* serovar Typhimurium (229), and many strains, such as *Rhodospirillum rubrum*, which contain a RubisCO gene and a noncarboxysomal BMC operon (230). However, there is no evidence of a bacterial strain simultaneously containing an α- or β-carboxysomal operon and any other kind of BMC operon.

**CONCLUSIONS**

Carboxysomes form an essential part of the cyanobacterial CCM by acting as microcompartments which encapsulate the RubisCO enzyme and allow CO$_2$ levels to be elevated around this relatively sluggish enzyme. Partly as a result of these unique polyhedral structures, cyanobacteria are remarkably productive on a global scale. Two types of carboxysomes arose by convergent evolution, namely, the α-carboxysomes, found predominantly in oceanic cyanobacteria, and the β-carboxysomes, found mainly in freshwater/estuarine cyanobacteria. Impressive progress has now been made in understanding the structure, physiology, and evolution of carboxysomes, but considerable gaps are identifiable.

The solved crystal structures of the small hexameric (CcmK1-4 and CsoS1ABC forms) and rarer pentameric (CcmL and CsoS4AB forms) shell proteins have been immensely helpful for the development of models that describe the likely structures of the outer-
most part of the icosahedral shells of both carboxysome types. However, provision of crystal structures for the RuBisCO-organizing proteins (CcmM-58 and -35 forms), the potential “facet zipper” protein CcmO (or possibly CsoS1E in alpha types), and the large CsoS2 shell-associated protein CcmN (shell recruitment) has proved recalcitrant at this stage. Although technically difficult, a critical phase of crystal structure analyses is also to attempt to determine the fine structural basis of the functional interactions of two or more proteins, such as CcmM-58–RuBisCO, CcmM-35–RuBisCO, CcmK2–CcmO, CcmN–CcmM, CcmK2–CcmM–58, CsoS2–CsoSCA, etc., to allow the further refinement and validation of existing models.

Some of the questions yet to be answered are as follows:

- Does the smaller RuBisCO-organizing protein, CcmM-35, interact with RuBisCO to the center of Synechococcus PCC 7942 β-carboxysomes as postulated?
- Does the larger RuBisCO-organizing protein, CcmM-58, interact with RuBisCO at the inner shell layer of β-carboxysomes, as indicated by protein-protein interaction analysis?
- What direct approaches can be used to confirm that models describing the icosahedral shell structure as being composed of triangular sheets composed of tiled small hexamers are accurate?
- What direct approaches can be used to ascertain if the concave or convex face of CcmK2/CsoS1 hexamers face the cytoplasm?
- Do specific charged and/or gated pores in the different forms of CcmK or CsoS1 hexamers really allow specificity toward specific solutes, such as RuBP, PGA, HCO\textsubscript{3}-- or Mg\textsuperscript{2+}?
- Can a presumed early evolutionary event, such as a naked RuBisCO-CA aggregate structure, provide some advantage toward more efficient performance of RuBisCO?

New Directions

There may be possible biotech applications involving the generation of new ion-selective membranes for battery applications and purification of water by osmosis. Other applications may involve the introduction of carboxysomes into the chloroplasts of plants in an effort to improve crop yields through the introduction of a minimal form of the cyanobacterial CCM (231).

ACKNOWLEDGMENTS

B.D.R. was supported by an ANU Ph.D. scholarship. G.D.P. and M.R.B. thank the Australian Research Council for support via the Discovery grant program. We thank Britta Forster for helpful comments, Loraine Tucker for excellent technical assistance, and Lynne Whitehead for EM images of Cyanobium.

REFERENCES


34. \textit{Yeates T}, \textit{Crowley C}, \textit{Tanaka S}. 2010. Bacterial microcompartment...


