Self-assembly in the carboxysome: a viral capsid-like protein shell in bacterial cells

T.O. Yeates*†‡[1], Y. Tsai†, S. Tanaka*, M.R. Sawaya‡ and C.A. Kerfeld§

*UCLA Department of Chemistry and Biochemistry, University of California Los Angeles, 607 Charles E. Young Drive East, Los Angeles, CA 90095-1569, U.S.A., †UCLA Molecular Biology Institute, University of California Los Angeles, 611 Charles E. Young Drive East, Los Angeles, CA 90095-1570, U.S.A., ‡UCLA-DOE Institute for Genomics and Proteomics, University of California Los Angeles, 611 Charles E. Young Drive East, Los Angeles, CA 90095-1570, U.S.A., and §UCLA Life Sciences Core, University of California Los Angeles, Box 160606, Los Angeles, CA 90095-1570, U.S.A.

Abstract

Many proteins self-assemble to form large supramolecular complexes. Numerous examples of these structures have been characterized, ranging from spherical viruses to tubular protein assemblies. Some new kinds of supramolecular structures are just coming to light, while it is likely there are others that have not yet been discovered. The carboxysome is a subcellular structure that has been known for more than 40 years, but whose structural and functional details are just now emerging. This giant polyhedral body is constructed as a closed shell assembled from several thousand protein subunits. Within this protein shell, the carboxysome encapsulates the CO_2-fixing enzymes, Rubisco (ribulose-1,5-bisphosphate carboxylase/oxygenase) and carbonic anhydrase; this arrangement enhances the efficiency of cellular CO_2 fixation. The carboxysome is present in many photosynthetic and chemoautotrophic bacteria, and so plays an important role in the global carbon cycle. It also serves as the prototypical member of what appears to be a large class of primitive protein-based organelles in bacteria. A series of crystal structures is beginning to reveal the secrets of how the carboxysome is assembled and how it enhances the efficiency of CO_2 fixation. Some of the assembly principles revealed in the carboxysome are reminiscent of those seen in icosahedral viral capsids. In addition, the shell appears to be perforated by pores for metabolite transport into and out of the carboxysome, suggesting comparisons to the pores through oligomeric transmembrane proteins, which serve to transport small molecules across the membrane bilayers of cells and eukaryotic organelles.

Introduction

Until relatively recently, bacterial cells were generally believed to lack a high degree of internal organization, as they lack the typical membrane-bound organelles of eukaryotic cells. However, it is becoming clearer that many bacteria benefit from well-organized interiors [1–6]. For example, it is now understood that many bacteria have primitive cytoskeletons: ancient homologues of the well-known eukaryotic cytoskeletal proteins actin and tubulin are widely distributed across the Bacteria [7,8]. Furthermore, although bacteria lack the typical membrane-bound organelles of eukaryotes (e.g. mitochondria and chloroplasts), it is becoming evident that some have developed special mechanisms for achieving spatial localization inside the cell. One such mechanism arises in the carboxysome, a bacterial subcellular microcompartment whose details are just beginning to emerge.

Carboxysomes were first observed by electron microscopy more than 40 years ago inside the cells of cyanobacteria [9], and then in chemoautotrophic bacteria [10]. They appeared as polyhedral bodies having varying degrees of geometric regularity and diameters of 1000 Å (1 Å = 0.1 nm) or more (Figure 1). Their similarity to viral capsids was evident, but the carboxysome was found not to contain nucleic acid [11]. It is instead filled with enzymes, mainly Rubisco (ribulose-1,5-bisphosphate carboxylase/oxygenase), as well as carbonic anhydrase. This creates a subcellular enclosure for highly efficient CO_2 fixation [10,12–14]. CO_2 is generated by carbonic anhydrase and consumed by Rubisco, so the co-localization of the two enzymes results in substrate channelling. This accounts at least in part for the enzymatic advantage provided by the carboxysome. The arrangement is particularly critical in view of Rubisco's naturally low enzymatic efficiency, and the undesirable competition between molecular O_2 and CO_2 in reaction with Rubisco. Until recently, further details on how the carboxysome is constructed and how it functions have been scarce.

Some of the most useful information has come from genomic sequence data. Carboxysomes, and the genomic organization of their components, have been studied in the greatest detail from two organisms: *Synechocystis* PCC6803 (a cyanobacterium) and *Halothiobacillus neapolitanus* (a chemoautotroph) [15,16]. In *H. neapolitanus*, the genes for the small and large subunits of Rubisco occur in an operon encoding several other proteins. These include genes for three homologous proteins, CsoS1A, CsoS1B and CsoS1C, which were found to be major components in cell preparations enriched in carboxysomes [17,18]. Multiple homologues from

Key words: carbon dioxide, carbon fixation, carboxysome, microcompartment, organelle, protein assembly.

Abbreviation used: Rubisco, ribulose-1,5-bisphosphate carboxylase/oxygenase..

[1]To whom correspondence should be addressed (email yeates@mbi.ucla.edu).

Figure 1 | Electron micrograph of carboxysomes

Left-hand panel: a thin section electron micrograph of *H. neapolitanus* cells. Several carboxysomes are visible as polyhedral bodies inside each cell. Image courtesy of Gordon Cannon. Right-hand panel: a negatively stained electron micrograph of isolated carboxysomes. Scale bar, 1000 Å. Image courtesy of Mark Yeager and Kelly Dryden.

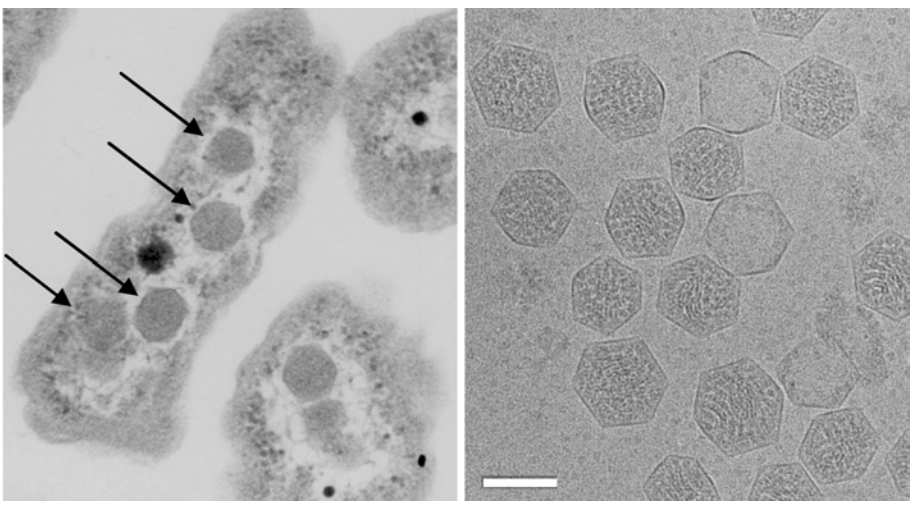
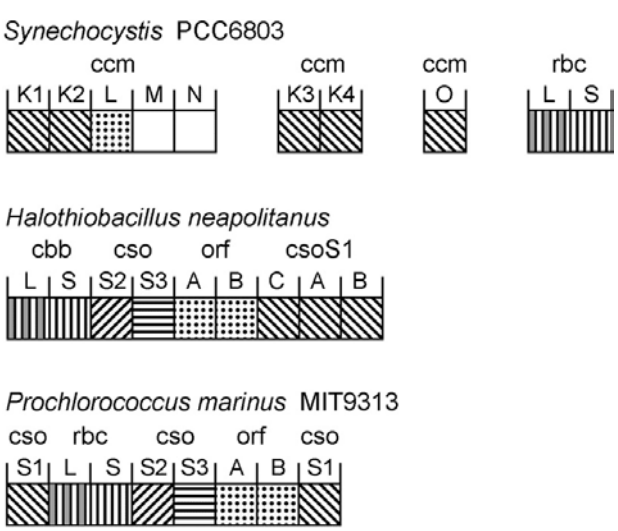


Figure 2 | Illustration of the operons encoding carboxysome proteins in three different organisms

Synechocystis PCC6803 and *Prochlorococcus marinus* are cyanobacteria. *H. neapolitanus* is a chemoautotrophic bacterium. Homologous proteins are shaded similarly. The CcmK/CsoS1 proteins are the major constituents of the outer shell of the carboxysome.



this protein family are also present in *Synechocystis* PCC6803 (the homologous proteins in that organism are named CcmK1, CcmK2, CcmK3 and CcmK4) and in other cyanobacteria containing carboxysomes (Figure 2) [2,16,18,19]. The carboxysome operons also encode other proteins, including some whose sequences identify them as carbonic anhydrases, and some that are apparently novel. In addition, purified preparations of carboxysomes from *H. neapolitanus* led to the identification of as many as nine distinct proteins. Despite the potential complexity of the carboxysome indicated by such studies, the presence of multiple homologues of a small protein (i.e. the CsoS1/CcmK proteins) suggested similarities to icosahedral viral capsids, some of which are constructed from many copies of a few distinct, but homologous, protein subunits [20,21]. This perspective led us to focus on the CcmK and CsoS1 proteins as the best starting point for understanding the structure of the carboxysome shell.

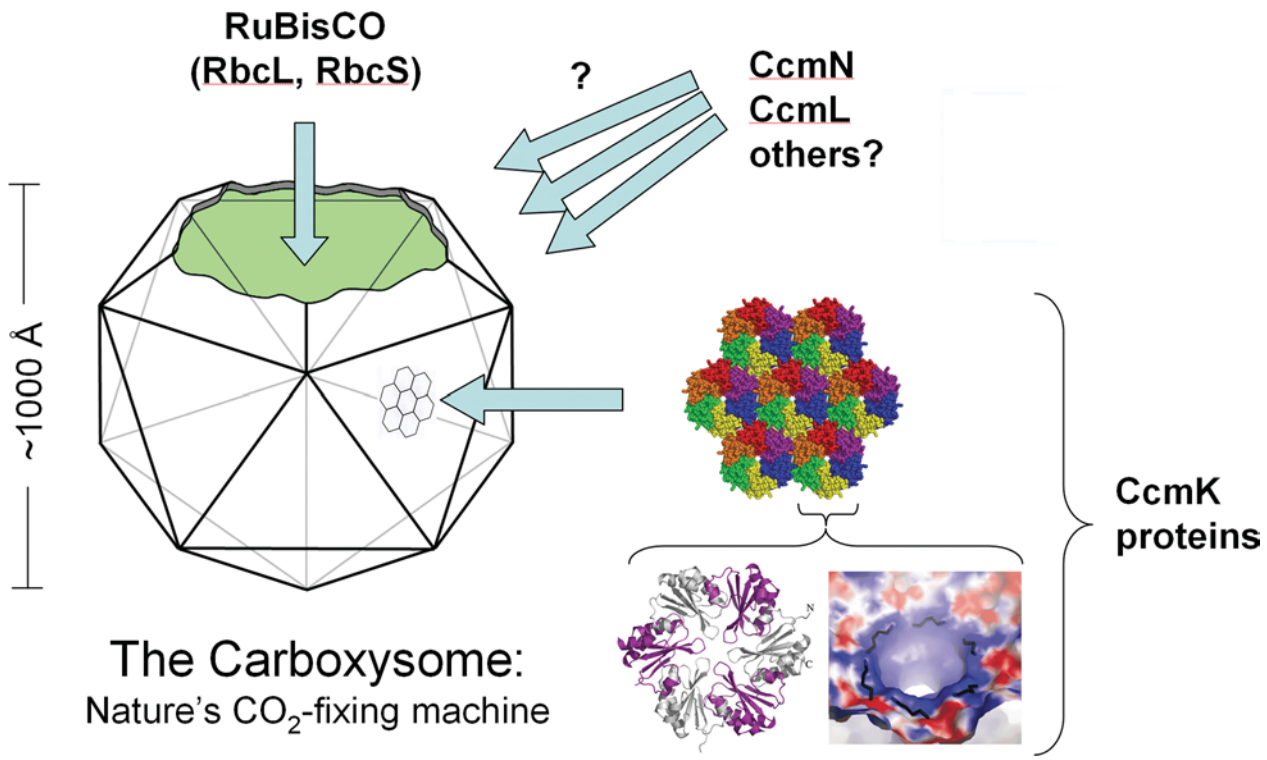
Insights from crystal structures of the major shell subunits

Crystal structures have now been determined for four different homologues of the major carboxysome shell protein (i.e. CcmK/CsoS1 proteins). The structures of CcmK2 and CcmK4 (in two different crystal forms) were reported by Kerfeld et al. [22], while the structure of CsoS1A has also been determined [23]. Informative diffraction data have also been obtained for CcmK1 (S. Tanaka and T.O. Yeates, unpublished work). The common features observed in these multiple structures have led to a number of conclusions about the outer shell of the carboxysome. Some of the minor variations between the structures have provided more subtle clues about assembly, mechanism and cellular function (Figure 3).

A central finding, consistent throughout all the structures, is that the major shell proteins assemble to form hexameric building blocks [22,23]. In addition, as seen in three different proteins representing two different organisms (CcmK2 and CcmK1 from *Synechocystis* PCC6803, and CsoS1A from *H. neapolitanus*), the hexameric building blocks show an ability to assemble further to form two-dimensional molecular layers of protein molecules [22,23]. This roughly 18 Å thick molecular layer appears to represent the flat facets of the

Figure 3 | Cartoon representation of the carboxysome, indicating some of the protein components involved, and showing the recently determined structures of the major shell proteins

The CcmK/Csos1 proteins (lower right) form hexamers that pack into a two-dimensional molecular layer. Narrow electrostatically positive pores through the layer may serve to transport negatively charged small molecules (such as bicarbonate) into and out of the carboxysome.



carboxysome shell, which a recent electron-microscopy study suggests is roughly icosahedral in shape [24]. Interestingly, not all of the subunits studied so far appear to be capable of forming an indefinite layer by a packing of hexagons in two dimensions. The CcmK4 subunit formed only individual hexamers in one crystal form, and linear strips of hexagons in a second crystal form [22]. A detailed comparison of the atomic structures reveals that the backbones of the different proteins adopt distinct conformations in their C-terminal regions. These structural differences appear to affect the differing assembly properties of the distinct proteins. The purpose of these differences in structure and assembly properties is not clear at this point. However, one possibility is that the different proteins fulfil different roles in the architecture of the complete carboxysome; a protein that cannot form an extended sheet in two dimensions might serve a role at the edges of the flat facets of the shell, for instance. Further studies will be required to test such ideas.

The packing between the hexamers is very tight; a computational analysis shows that the shape complimentarity between adjacent hexamers is as high as for other natural protein–protein interfaces that have been characterized [23,25]. Also, the six subunits of each hexamer fit together so that only a small pore remains down the central axis of symmetry. The molecular layer is therefore nearly solid, bearing only small pores whose diameters vary from approx. 4 Å to 7 Å depending on the particular protein. The tightness of the packing strongly suggests that the major shell proteins of the carboxysome have evolved with the purpose of limiting or controlling transport into and out of the carboxysome.

Our current hypothesis is that the metabolic intermediates that serve as the substrates and products of CO_2 fixation, bicarbonate and the C_3 and C_5 sugars (3-phosphoglycerate and ribulose 1,5-bisphosphate respectively), are able to cross into and out of the carboxysome by way of the pores noted above. The pores through the carboxysome shell are created at the axes of symmetry through the hexameric oligomers. This is reminiscent of the pores formed in transmembrane protein channels by oligomer formation; the tetrameric potassium channel [26] and the pentameric pressure-sensitive water channel [27] provide well-known examples. The carboxysome shell is considerably thinner than a typical lipid bilayer, so the pores through the carboxysome are shorter than the pores through transmembrane pores [23]. The shortness of the carboxysome pores may limit the selectivity that can be achieved, but the structures do suggest some degree of specificity. The pores visualized so far in the carboxysome hexamers bear a positive electrostatic potential. This could reflect a preference for the passage of negatively

charged molecules such as bicarbonate; the neutral O_2 molecule that presents a problematic competition with CO_2 would not enjoy any advantage from a positively charged pore. Experiments to date have failed to identify bicarbonate molecules bound in the pores, but bound sulfate ions have recently been observed [23]; their double-negative charge may enhance their tendency to be tightly bound in the pore.

Questions and future directions

The structures of the major shell proteins provide a number of early insights into the structure and function of the carboxysome. They also provide an atomic level framework for addressing a number of new questions. Some of those questions and future areas of investigation are noted here. (i) If the hexagonal pores through the major shell proteins serve for molecular transport, then structure-based mutagenesis experiments should be able to test this. (ii) Flat hexagonal layers have been visualized so far, prompting the question of how such a layer can bend or fold up to form a closed shell. Preliminary data suggest that other proteins in the carboxysome operon provide this curvature (S. Tanaka, C.A. Kerfeld, and T.O. Yeates, unpublished work). (iii) Essentially nothing is known about how the shell is connected to its contents. How does the carboxysome get assembled properly in the cell? What molecular interactions and recognition events are involved in enclosing virtually all of a cell's Rubisco molecules inside a protein shell? (iv) Is the carboxysome regulated, either in its assembly or its composition? (v) How did the carboxysome evolve? (vi) Homologues of the major shell protein can be found widely distributed across the bacterial kingdom, including in microbes that do not perform CO_2 fixation; those organisms appear to sequester other metabolic pathways inside the microcompartment [19]. What are the mechanistic and evolutionary relationships between these diverse microcompartments? (vii) Nature has evolved these microcompartments in order to enhance particular enzymatic pathways. Can novel microcompartments be engineered in order to enhance other useful pathways and processes? (viii) Is it possible that the known microcompartments, constructed from shell proteins homologous with those described here, represent just one of several kinds of microcompartments present in Nature? Continuing structural, biochemical and genomic studies should begin to answer these questions.

Summary

The emerging structure of the carboxysome illustrates how much remains to be learned about cells and the molecular mechanisms they have evolved to carry out their functions. Working from the bottom-up, structural studies on components of the carboxysome are revealing key insights about molecular evolution, mechanism and function. In particular, architectural similarities to viral capsids are evident. Whether this reflects a presently unknown evolutionary relationship or a case of convergent evolution remains to be seen. The structures also reveal pores that may serve for molecular transport into and out of the carboxysome, opening up comparisons with other molecular transport systems, such as transmembrane channels. The findings summarized here only scratch the surface of the carboxysome. Further studies will be required to reveal the remaining mysteries contained in the carboxysome and other subcellular microcompartments.

We are indebted to Gordon Cannon for materials and helpful advice. This work was supported by the CSREES (Cooperative State Research, Education, and Extension Service) NRI (National Research Intiative) programme of the U.S. Department of Agriculture.

References

1 Bazylinski, D.A. and Frankel, R.B. (2004) Nat. Rev. Microbiol. **2**, 217–230
2 Cannon, G.C., Bradburne, C.E., Aldrich, H.C., Baker, S.H., Heinhorst, S. and Shively, J.M. (2001) Appl. Environ. Microbiol. **67**, 5351–5361
3 Fuerst, J.A. (2005) Annu. Rev. Microbiol. **59**, 299–328
4 Gitai, Z. (2005) Cell **120**, 577–586
5 Mayer, F. (1993) FEMS Microbiol. Rev. **10**, 327–345
6 Seufferheld, M., Vieira, M.C., Ruiz, F.A., Rodrigues, C.O., Moreno, S.N. and Docampo, R. (2003) J. Biol. Chem. **278**, 29971–29978
7 Lowe, J. and Amos, L.A. (1998) Nature **391**, 203–206
8 van den Ent, F., Amos, L.A. and Lowe, J. (2001) Nature **413**, 39–44
9 Jensen, T.E. (1961) Proc. Iowa Acad. Sci. **68**, 86–89
10 Shively, J.M., Ball, F., Brown, D.H. and Saunders, R.E. (1973) Science **182**, 584–586
11 Vakeria, D., Codd, G.A., Marsden, W.J.N. and Stewart, W.D.P. (1984) FEMS Microbiol. Lett. **25**, 149–152
12 Badger, M.R. and Price, G.D. (2003) J. Exp. Bot. **54**, 609–622
13 Kaplan, A. and Reinhold, L. (1999) Annu. Rev. Plant Physiol. Plant Mol. Biol. **50**, 539–570
14 Price, G.D., Sültemeyer, D., Klughammer, B., Ludwig, M. and Badger, M.R. (1998) Can. J. Bot. **76**, 973–1002
15 Cannon, G.C., Baker, S.H., Soyer, F., Johnson, D.R., Bradburne, C.E., Mehlman, J.L., Davies, P.S., Jiang, Q.L., Heinhorst, S. and Shively, J.M. (2003) Curr. Microbiol. **46**, 115–119
16 Shively, J.M., Bradburne, C.E., Aldrich, H.C., Bobik, T.A., Mehlman, J.L., Jin, S. and Baker, S.H. (1998) Can. J. Bot. **76**, 906–916
17 English, R.S., Lorbach, S.C., Qin, X. and Shively, J.M. (1994) Mol. Microbiol. **12**, 647–654
18 Heinhorst, S., Cannon, G.C. and Shively, J.M. (2006) in Microbiology Monographs, pp. 141–166, Springer-Verlag, Berlin
19 Bobik, T.A. (2006) Appl. Microbiol. Biotechnol. **70**, 517–525
20 Hogle, J.M., Chow, M. and Filman, D.J. (1985) Science **229**, 1358–1365
21 Rossmann, M.G., Arnold, E., Erickson, J.W., Frankenberger, E.A., Griffith, J.P., Hecht, H.J., Johnson, J.E., Kamer, G., Luo, M., Mosser, A.G. et al. (1985) Nature **317**, 145–153
22 Kerfeld, C.A., Sawaya, M.R., Tanaka, S., Nguyen, C.V., Phillips, M., Beeby, M. and Yeates, T.O. (2005) Science **309**, 936–938
23 Tsai, Y., Sawaya, M.R., Cannon, G.C., Cai, F., Williams, E.B., Heinhorst, S., Kerfold, C.A. and Yeates, T.O. (2007) PLoS Biol., in the press
24 Schmid, M.F., Paredes, A.M., Khant, H.A., Soyer, F., Aldrich, H.C., Chiu, W. and Shively, J.M. (2006) J. Mol. Biol. **364**, 526–535
25 Lawrence, M.C. and Colman, P.M. (1993) J. Mol. Biol. **234**, 946–950
26 MacKinnon, R. (2003) FEBS Lett. **555**, 62–65
27 Chang, G., Spencer, R.H., Lee, A.T., Barclay, M.T. and Rees, D.C. (1998) Science **282**, 2220–2226

Received 28 February 2007