Protein Structures Forming the Shell of Primitive Bacterial Organelles

Cheryl A. Kerfeld, *et al.*

Science **309**, 936 (2005);
DOI: 10.1126/science.1113397

Protein Structures Forming the Shell of Primitive Bacterial Organelles

Cheryl A. Kerfeld,[1,2,4] Michael R. Sawaya,[1,3,4] Shiho Tanaka,[1] Chau V. Nguyen,[1] Martin Phillips,[1,3] Morgan Beeby,[1,3] Todd O. Yeates[1,3,4]*

Bacterial microcompartments are primitive organelles composed entirely of protein subunits. Genomic sequence databases reveal the widespread occurrence of microcompartments across diverse microbes. The prototypical bacterial microcompartment is the carboxysome, a protein shell for sequestering carbon fixation reactions. We report three-dimensional crystal structures of multiple carboxysome shell proteins, revealing a hexameric unit as the basic microcompartment building block and showing how these hexamers assemble to form flat facets of the polyhedral shell. The structures suggest how molecular transport across the shell may be controlled and how structural variations might govern the assembly and architecture of these subcellular compartments.

Intracellular polyhedral bodies (Fig. 1, A and B) were first observed in bacteria by electron microscopy more than 40 years ago (*1*–*3*). These large protein shells, generally ranging in size from 100 to 200 nm, were initially thought to be viral particles (*4*), but the first such structures isolated from chemoautotrophic bacteria (*5*) were filled with the enzyme ribulose bisphosphate carboxylase oxygenase (RuBisCO) and were therefore named carboxysomes. In organisms where they occur, carboxysomes compartmentalize most, if not all, of the cellular RuBisCO. Subsequent biochemical characterization demonstrated that carboxysomes also have associated carbonic anhydrase activity, providing localized conversion of bicarbonate to CO_2 (*6*–*8*), the substrate for RuBisCO. Carbon fixation may be optimized by providing a microenvironment rich in CO_2 within a shell that could be less permeable to RuBisCO's potent competitive inhibitor, molecular oxygen (*9*–*11*).

Comparative biochemical and genomic analysis of carboxysomes has revealed that several small (~10 kD) highly similar homologous proteins (Fig. 2) are the key components of the carboxysome shell (*11*, *12*). Evolutionarily related proteins have been identified in several obligately heterotrophic enteric bacteria (which do not fix CO_2), encoded within operons associated with oxygen-sensitive enzymatic processes (*11*, *13*–*15*). In those organisms, ultrastructural studies confirm the presence of carboxysome-like inclusions under growth conditions that induce those operons. Collectively, the shell proteins of bacterial microcompartments contain a conserved sequence referred to as the bacterial microcompartment (BMC) domain. Querying the sequence databases for similarity to the BMC domain reveals that the potential to form proteinaceous compartments is widespread among the bacteria (fig. S1). Typically, several genes coding for BMC domain–containing proteins cluster with genes coding for putative enzymes or proteins of unknown function; presumably, they form specialized compartments for as-yet uncharacterized metabolic processes. Thus, the microcompartment structure typified by the carboxysome can be viewed as a form of primitive organelle, organizing reactions that require special conditions for optimization, such as the sequestration of substrates, cofactors, or toxic intermediates.

Carboxysomes can be grouped into two classes, α and β, based on sequence analysis and gene organization. In β-carboxysomes, the shell proteins are the *ccmK* gene products (named for carbon-concentrating mechanism) (*11*, *12*, *16*). Several species of cyanobacteria including *Synechocystis* sp. PCC 6803 contain four similar genes coding for the CcmK proteins, CcmK1 to 4 (Fig. 2). We have determined the crystal structures of CcmK2 and CcmK4 (in two crystal forms), to reveal the structural fold of the widely distributed BMC domain. The overall three-dimensional folds of CcmK4 and CcmK2 are practically identical (Fig. 3, A and B). Searches of the databases of known three-dimensional structures (*17*, *18*) reveal that the N-terminal 80 amino acids of the BMC domain adopt an α/β fold, observed in ferredoxin and numerous apparently unrelated proteins. No direct evolutionary link could be established either by sequence or structure comparisons, between the BMC domain and known viral capsid proteins.

In all three structures visualized, the protein subunits are arranged in hexameric units about a central six-fold axis of symmetry (Fig. 3B). The structure of the BMC domain monomer is notably wedge-shaped, so that six subunits fit together to leave only a small central hole. Monomers of a more spherical shape would assemble to leave a circular hole about the size

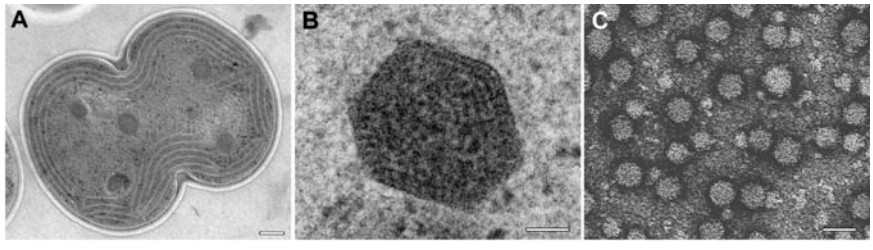

Fig. 1. (**A** and **B**) Transmission electron micrographs (EMs) of cyanobacterium *Synechocystis* sp. PCC 6803 showing (A) a single cell dividing (scale bar, 200 nm). The polyhedral shape and crystalline structure distinguishes the carboxysomes from other cytoplasmic inclusions. (B) A single carboxysome (scale bar, 50 nm). (**C**) Transmission EMs of purified recombinant BMC domain protein CcmK2, after precipitation and sonication (scale bar, 50 nm). EMs in (A) and (B) provided by A. M. L. van de Meene, W. F. J. Vermaas, and R. W. Roberson, School of Life Sciences, Arizona State University.

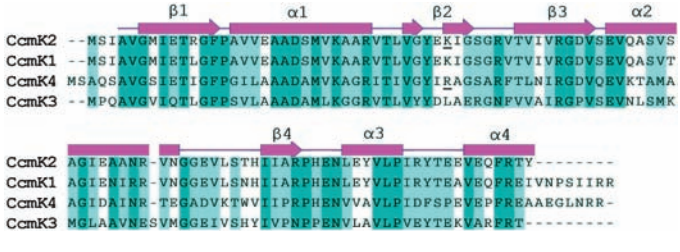

Fig. 2. A sequence alignment of the four BMC domain proteins in *Syn* 6803, CcmK1 to 4. Regions of high conservation are shaded, and the protein's secondary structure elements as determined here are shown (rectangles, α helices; arrows, β strands). Charged amino acids found in the hexamer pore (Fig. 3, C and D) are underlined.

[1]Molecular Biology Institute, University of California, Los Angeles (UCLA), Box 951570, [2]Life Sciences Core, UCLA, Box 160606, [3]Department of Chemistry and Biochemistry, UCLA, Box 951569, Los Angeles, CA 90095, USA. [4]UCLA–U.S. Department of Energy (DOE) Institute for Genomics and Proteomics, Box 951570, Los Angeles, CA 90095–1570, USA.

*To whom correspondence should be addressed. E-mail: yeates@mbi.ucla.edu

of one subunit. The pore at the center of the CcmK2 hexamer (taking into account atomic van der Waals radii) is ~7 Å in diameter (Fig. 3C). The CcmK4 pore is narrower, ~4 Å in diameter (Fig. 3D). The narrowness of the hole in the BMC domain hexamer apparently reflects a functional constraint. A further unexpected finding is a large net positive electrostatic potential within the central pore. Each of the six subunits contributes a conserved positively charged amino acid residue (Arg^{38} in CcmK4; Lys^{36} in CcmK2) to the pore (Fig. 3, C and D). The energetic cost of bringing together such a large number of like charges is considerable (SOM text). We conclude, from the cost of creating such a configuration and from its conservation in numerous carboxysome shell proteins from different organisms, that the charged pore serves a functional role in the carboxysome, most likely by regulating metabolite flux.

In the crystal structure of CcmK2 and in the second crystal form of CcmK4, the next level of subunit organization is evident (Fig. 4). Owing to their nearly flat edges, the hexameric units fit together readily. For CcmK2, the hexameric units form a nearly solid molecular sheet (Fig. 4A). In the second crystal form of CcmK4, we observe linear strips of hexamers (Fig. 4B). Two adjacent CcmK4 hexamers can be superimposed on two adjacent CcmK2 hexamers with only a 1.4 Å root mean square difference over 551 Cα atoms. The arrangement of hexamers in the CcmK2 sheet leaves only small gaps at the two-fold and three-fold axes of symmetry in the layer, where two and three hexameric units come together. The close packing of the hexamers in sheets is in marked contrast to other nonviral hexameric protein assemblies that have been observed (*19*). On the basis of the recurrence of the same interface in different crystal forms, involving distinct subunit types and of the unusually tight packing achieved, we argue that the observed packing arrangement is biologically relevant. The sheet structure suggests that the carboxysome shell is roughly 2 to 3 nm thick. According to the spacing of subunits in the hexagonal sheet, a microcompartment with a diameter of 150 nm (Fig. 1B) would contain some 10,000 BMC domain proteins in its outer shell.

In addition to the hexameric pore, the gaps between hexamers may serve as conduits for metabolites. The gap at the three-fold axis is ~6 Å in diameter (if one takes into account atomic radii), whereas the gap at the two-fold axis is elongated, but only ~4 Å wide. Charged amino acid residues, conserved across CcmK proteins, are also notable at these gaps (SOM text). Especially in view of possible minor side chain rearrangements, the pores and gaps appear to be large enough to allow transit of the negatively charged metabolites (namely, bicarbonate, ribulose bis-phosphate, phosphoglycerate, and possibly hydroxyl ions) that have to cross the carboxysome shell. In contrast, uncharged molecules such as CO_2 and O_2 would not be attracted to the charged pores and gaps. The possibility that pores within and gaps between hexamers might serve as portals for small molecules is reminiscent of certain viruses in which it has been determined that the interstitial spaces between subunits mediate the flow of metabolites into and out of the capsids (*20*, *21*).

The structural findings suggest that the carboxysome shell is involved in controlling metabolite flow, rather than serving merely as a storage device for excess RuBisCO or simply to localize the requisite enzymes for carbon fixation. Such selective permeability is fundamental to subcellular organelles. Although bacterial microcompartments do not meet the traditional criteria (for eukaryotic organelles) of being membrane-bound, the view emerging here is that they meet the requirements of organelle function in other ways, for example by controlling flow through a nearly solid protein shell rather than through proteins embedded in a continuous membrane.

When they were first observed in bacterial cells, the resemblance of carboxysomes to viral particles was so striking that it prompted efforts to induce viral-mediated lysis of the cells (*4*). Not surprisingly, our structural studies reveal a striking number of parallels between the basic features of bacterial microcompartments and viruses (*22*–*27*). Both are highly symmetrical structures sharing underlying design features. In some viruses, the subunit organization is best described by combinations of hexameric and pentameric capsomeres that self-assemble. According to the rules of solid geometry, forming a closed shell generally requires 12 pentameric capsomeres to

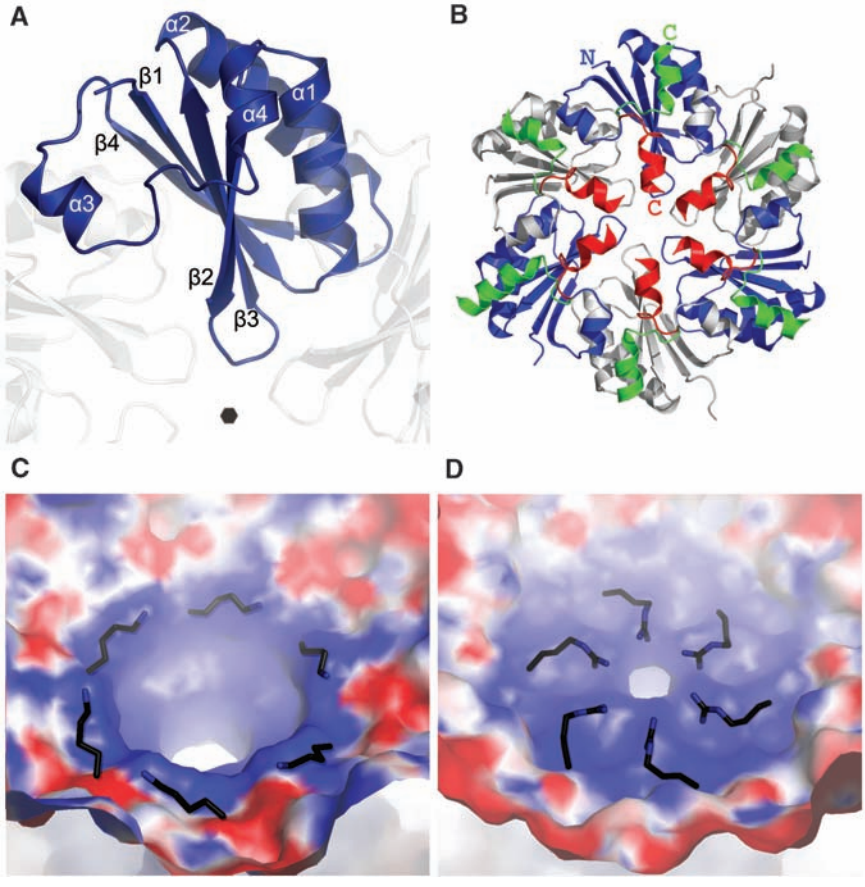

Fig. 3. (**A**) The three-dimensional crystal structure of BMC domain protein, CcmK4, determined at a resolution of 1.8 Å. (**B**) The concave surface of the CcmK4 hexamer. CcmK4 monomers are colored alternately blue and gray, with all six C termini in green. Also shown is the CcmK2 C terminus (red) in its corresponding position after a superposition of CcmK2 and CcmK4 hexamers. The superposition (not shown) indicates that the backbones of the two hexamers are nearly identical up to residues Pro^{97} (CcmK4)/Glu^{95} (CcmK2). The structural differences at the C termini (labeled) suggest one reason why CcmK4 packing is limited to chains of hexamers. In CcmK4, the C termini extend outward, toward the corners of the hexamer. These are the positions where three hexamers meet in the CcmK2 sheet but not in the CcmK4 structures. (**C** and **D**) A close-up view of the pores (concave side) formed at the six-fold axis of symmetry in the CcmK2 (C) and CcmK4 (D) hexamers. The surfaces are colored according to electrostatic potential, with blue positive and red negative. Positively charged, conserved amino acid side chains lining the pore are highlighted (Lys^{36} in CcmK2, Arg^{38} in CcmK4). This figure and Fig. 4 were illustrated with PyMOL (*32*).

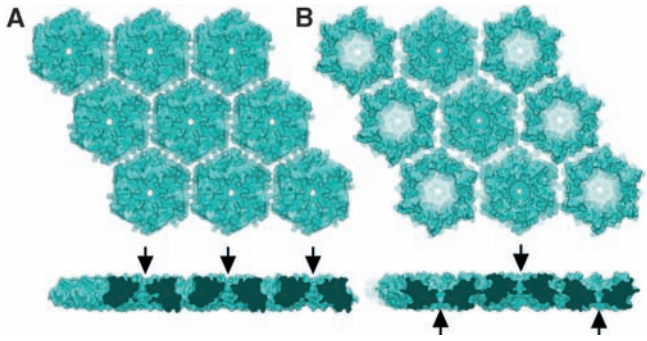

Fig. 4. Crystal packing of BMC domain proteins in molecular layers. (**A**) CcmK2 hexamers packed in uniform orientations (convex face shown). (**B**) CcmK4 hexamers (crystal form 2) packed in strips of alternating orientation. In the side view of each sheet, arrows mark the positions of the pores.

be present among a greater number of hexameric capsomeres (SOM text). These ideas are consistent with our studies of the BMC domain proteins. The first three crystal structures have revealed hexameric assemblies; sedimentation and native gel electrophoresis studies indicate that some of the subunits form pentamers in addition to hexamers (fig. S2, SOM text).

Likewise, similar but distinct protein subunits existing in quasi-equivalent environments or forms are required to construct virus capsids (*22*–*27*). Among the CcmK proteins, CcmK2 alone appears capable of forming closed shells under certain conditions, but these structures are much smaller than native carboxysomes and lack their polyhedral regularity (Fig. 1C). Apparently, as with many viruses, multiple distinct carboxysome subunit types appear to be required to achieve the correct architecture. The generally high conservation of amino acid residues among the CcmK paralogs at the interhexamer interfaces is consistent with the idea that hexameric (or possibly pentameric) units of different types could assemble together. The subtle differences in primary and tertiary structure could then govern the quasi-equivalent interactions necessary to create the native shell. The crystal structures of CcmK2 and CcmK4 give preliminary clues as to the origins of distinct assembly behavior, such as their disparate tendency to form sheets (Fig. 4). In CcmK4, clashes between C termini from adjacent hexamers could preclude the formation of flat sheets of the type seen with the CcmK2 hexamer. Interactions between the C-terminal tails of the BMC domain proteins could influence microcompartment domain assembly in the same way that the flexible termini of certain viral capsid proteins often participate as switches for distinct types of interactions in the mature viral capsid (*25*, *26*, *28*, *29*). The shape and assembly of the carboxysome could also be affected by other proteins that may be present in the shell (*30*) whose structures are not yet known.

As was the case for other large molecular machines such as viruses and ribosomes (*31*), fully elucidating the structure and function of bacterial microcompartments will require combining electron microscopy, biophysical experiments, and crystallographic studies on the multiple components in order to attain an understanding of the whole.

References and Notes

1. T. E. Jensen, C. C. Bowen, *Proc. Iowa Acad. Sci.* **68**, 89 (1961).
2. E. Gantt, S. F. Conti, *J. Bacteriol.* **97**, 1486 (1969).
3. J. M. Shively, G. L. Decker, J. W. Greenawalt, *J. Bacteriol.* **96**, 2138 (1970).
4. J. M. Shively, R. S. English, *Can. J. Bot.* **69**, 957 (1991).
5. J. M. Shively, F. Ball, D. H. Brown, R. E. Saunders, *Science* **182**, 584 (1973).
6. G. D. Price, J. R. Coleman, M. R. Badger, *Plant Physiol.* **100**, 784 (1992).
7. A. K.-C. So, G. S. Espie, *Plant Mol. Biol.* **37**, 205 (1998).
8. A. K.-C. So *et al.*, *J. Bacteriol.* **186**, 623 (2004).
9. G. A. Codd, W. J. N. Marsden, *Biol. Rev. Camb. Philos. Soc.* **59**, 389 (1984).
10. G. D. Price, M. R. Badger, *Can. J. Bot.* **69**, 963 (1990).
11. G. C. Cannon *et al.*, *Appl. Environ. Microbiol.* **67**, 5351 (2001).
12. G. C. Cannon, S. Heinhorst, C. E. Bradburne, J. M. Shively, *Funct. Plant Biol.* **29**, 175 (2002).
13. J. M. Shively *et al.*, *Can. J. Bot.* **76**, 906 (1998).
14. E. Kofoid, C. Rappleye, I. Stojiljkovic, J. Roth, *J. Bacteriol.* **181**, 5317 (1999).
15. G. D. Havemann, T. A. Bobik, *J. Bacteriol.* **185**, 5086 (2003).
16. M. R. Badger, D. Hanson, G. D. Price, *Funct. Plant Biol.* **29**, 161 (2002).
17. L. Holm, C. Sander, *J. Mol. Biol.* **233**, 123 (1993).
18. J.-F. Gibrat, T. Madej, S. H. Bryant, *Curr. Opin. Struct. Biol.* **6**, 377 (1996).
19. P. Yuan *et al.*, *Nat. Struct. Biol.* **10**, 264 (2003).
20. J. M. Diprose *et al.*, *EMBO J.* **20**, 7229 (2001).
21. H. Naitow, J. Tang, M. Canady, R. B. Wickner, J. E. Johnson, *Nat. Struct. Biol.* **9**, 725 (2002).
22. D. L. D. Caspar, A. Klug, *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1 (1962).
23. J. M. Hogle, M. Chow, D. J. Filman, *Sci. Am.* **256**, 42 (March 1987).
24. M. G. Rossmann, J. E. Johnson, *Annu. Rev. Biochem.* **58**, 533 (1989).
25. S. C. Harrison, *Curr. Opin. Struct. Biol.* **5**, 157 (1995).
26. J. E. Johnson, J. A. Speir, *J. Mol. Biol.* **269**, 665 (1997).
27. S. C. Harrison, *Curr. Opin. Struct. Biol.* **11**, 195 (2001).
28. T. Stehle, S. J. Gamblin, Y. Yan, S. C. Harrison, *Structure* **4**, 165 (1996).
29. C. San Martin *et al.*, *Nat. Struct. Biol.* **9**, 756 (2002).
30. A. K. So *et al.*, *J. Bacteriol.* **186**, 623 (2004).
31. N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz, *Science* **289**, 905 (2000).
32. W. L. DeLano, www.pymol.org (2002).

33. We thank G. Cannon and M. Yeager for helpful discussions, D. Krogmann for the gift of *Synechocystis* sp. PCC 6803 genomic DNA, and J. Laidman and H. Adesiteyo for technical assistance. We also thank R. W. Roberson, W. F. J. Vermass, and A. van de Meene for providing transmission electron micrographs of *Syn* 6803 cells. C.A.K and T.O.Y acknowledge the U.S. Department of Agriculture, the National Institutes of Health, and the Office of Science Biological and Environmental Research Program, the U.S. Department of Energy. The coordinates and structure factors have been deposited in the Protein Data Bank with accession numbers 2A10 (CcmK4, Form 1); 2A18 (CcmK4, Form 2); and 2A1B (CcmK2).

Supporting Online Material

www.sciencemag.org/cgi/content/full/309/5736/936/DC1
Materials and Methods
SOM Text
Figs. S1 to S3
References and Notes

11 April 2005; accepted 22 June 2005
10.1126/science.1113397

Rewiring of the Yeast Transcriptional Network Through the Evolution of Motif Usage

Jan Ihmels,[1] Sven Bergmann,[1] Maryam Gerami-Nejad,[2] Itai Yanai,[1] Mark McClellan,[2] Judith Berman,[2] Naama Barkai[1]*

Recent experiments revealed large-scale differences in the transcription programs of related species, yet little is known about the genetic basis underlying the evolution of gene expression and its contribution to phenotypic diversity. Here we describe a large-scale modulation of the yeast transcription program that is connected to the emergence of the capacity for rapid anaerobic growth. Genes coding for mitochondrial and cytoplasmic ribosomal proteins display a strongly correlated expression pattern in *Candida albicans*, but this correlation is lost in the fermentative yeast *Saccharomyces cerevisiae*. We provide evidence that this change in gene expression is connected to the loss of a specific cis-regulatory element from dozens of genes following the apparent whole-genome duplication event. Our results shed new light on the genetic mechanisms underlying the large-scale evolution of transcriptional networks.

Evolution of gene expression plays a prominent role in generating phenotypic diversity (*1*–*3*), but little is known about the genetic basis underlying broad modulations of the genome-wide transcription program. Here we describe a rewiring of the yeast transcrip-