Structural basis of the oxidative activation of the carboxysomal γ -carbonic anhydrase, CcmM

Kerry L. Peña^a, Stephane E. Castel^a, Charlotte de Araujo^b, George S. Espie^{b,c}, and Matthew S. Kimber^{a,1}

^aDepartment of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada, N1G2W1^bDepartments of Cell and Systems Biology, and ^cBiology, University of Toronto, Mississauga, ON, Canada L5L1C6

Edited by Gregory A. Petsko, Brandeis University, Waltham, MA, and approved December 15, 2009 (received for review September 21, 2009)

Cyanobacterial RuBisCO is sequestered in large, icosahedral, protein-bounded microcompartments called carboxysomes. Bicarbonate is pumped into the cytosol, diffuses into the carboxysome through small pores in its shell, and is then converted to CO₂ by carbonic anhydrase (CA) prior to fixation. Paradoxically, many β -cyanobacteria, including *Thermosynechococcus elongatus* BP-1, lack the conventional carboxysomal β -CA, *ccaA*. The N-terminal domain of the carboxysomal protein CcmM is homologous to γ -CA from *Methanosarcina thermophila* (Cam) but recombinant CcmM derived from *ccaA*-containing cyanobacteria show no CA activity. We demonstrate here that either full length CcmM from *T. elongatus*, or a construct truncated after 209 residues (CcmM209), is active as a CA—the first catalytically active bacterial γ -CA reported. The 2.0 Å structure of CcmM209 reveals a trimeric, left-handed β -helix structure that closely resembles Cam, except that residues 198–207 form a third α -helix stabilized by an essential Cys194–Cys200 disulfide bond. Deleting residues 194–209 (CcmM193) results in an inactive protein whose 1.1 Å structure shows disordering of the N- and C-termini, and reorganization of the trimeric interface and active site. Under reducing conditions, CcmM209 is similarly partially disordered and inactive as a CA. CcmM protein in fresh *E. coli* cell extracts is inactive, implying that the cellular reducing machinery can reduce and inactivate CcmM, while diamide, a thiol oxidizing agent, activates the enzyme. Thus, like membrane-bound eukaryotic cellular compartments, the β -carboxysome appears to be able to maintain an oxidizing interior by precluding the entry of thioredoxin and other endogenous reducing agents.

CO₂ concentrating mechanism | cyanobacteria | disulfide bond

The key step in photosynthetic carbon fixation is catalyzed by ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO, EC 4.1.1.39)—a problematic enzyme with an intrinsic low affinity for its second substrate, CO₂, and poor discrimination against the competing, more abundant, and ultimately counterproductive substrate, O₂ (1). In cyanobacteria, these catalytic limitations are minimized by a set of adaptations collectively known as the CO₂ concentrating mechanism (CCM), which enhances carbon fixation by increasing the intracellular CO₂/O₂ ratio (2). The CCM has two major functional components. First, a variety of light-dependent CO₂ and HCO₃[−] transporters actively pump inorganic carbon across the cellular membrane where it accumulates in the cytosol as HCO₃[−] at up to 1,000 times its external abundance. Second, structurally distinct protein microcompartments that sequester the cellular RuBisCO—the carboxysomes—are required to mediate CO₂ fixation. Because inorganic carbon is accumulated in the cytosol as HCO₃[−], but RuBisCO requires CO₂, carbonic anhydrase (CA; EC 4.2.1.1) is also required. Catalyzed dehydration of HCO₃[−] by a CA specifically localized within the carboxysome results in the accumulation of CO₂ in the vicinity of RuBisCO at concentrations high enough to promote efficient and near-maximum rates of fixation. Carboxysomes occur as two distinct varieties— α and β (the focus of this work). α - and β -carboxysomes are built predominantly from component proteins that are homologous, but divergent (including RuBisCO), though each also contains unique proteins that are type specific.

Whole genome phylogenetic analysis suggests that α -carboxysome containing cyanobacteria form a derived clade within the larger cyanobacterial family, implying that β -carboxysomes were present in the last common ancestor of extant cyanobacteria (3).

Carboxysomes are large (100–200 nm) structures with a regular icosahedral shape. They are organized into a RuBisCO rich core surrounded by a thin (3–4 nm) protein shell, which separates this core from bulk cytoplasm (4, 5). The shell of β -carboxysomes is comprised of proteins of the hexameric bacterial microcompartment family, CcmK[1–4], that tile into flat sheets, whereas the vertices are occupied by CcmL pentamers (6). Small pores in the center of each CcmK hexamer allow essential small metabolites access to the carboxysomal lumen. Recent experiments show that intact α -carboxysomes enhance RuBisCO kinetics relative to disrupted ones, supporting suggestions that the intact shell acts as a diffusional barrier for CO₂ (7).

The majority of the interior space of the β -carboxysome is occupied by RuBisCO, whereas the essential, type-specific gene products of CcmM, CcmN, and CcaA form minor components (8–12). These components all interact with one another, and with the shell, through a network of protein–protein interactions centered on CcmM (11). CcmM is comprised of an N-terminal domain, with homology to trimeric left-handed β -helical proteins, and a C-terminal domain, which is comprised of three or four copies of an $\sim$ 85 amino acid subdomain homologous to RuBisCO's small subunit, RbcS. These domains are separated by $\sim$ 40 amino acid, hydrophilic linker regions (9, 12). The CcmM transcript possesses an interior ribosome binding site, and is therefore produced in at least two forms—smaller quantities of the full length protein, and a more abundant, smaller form that comprises only the RbcS-like subdomains (13), both of which strongly bind to RuBisCO (11, 14). The N-terminal subdomain mediates additional protein–protein interactions, binding itself, CcaA, and CcmN, as well as the abundant shell protein CcmK1 (11, 14) and possibly also to CcmK2, CcmK4, and CcmL (11) positioning the HCO₃[−] dehydration complex on the inner face of the shell (11).

Most β -cyanobacteria contain a carboxysomally localized β -CA, CcaA, which in *Synechococcus elongatus* PCC 7942 and *Synechocystis* sp. PCC 6803 is essential for CO₂ concentrating mechanism (CCM) function (15, 16). Surprisingly, a subset of β -cyanobacteria lack *ccaA* homologs. In these species, the role of carboxysomal CA has been suggested to be played by the N-terminal subdomain of CcmM, which is 35% identical to the canonical γ -class CA, Cam from *Methanosarcina*

Author contributions: M.S.K. designed research; K.L.P., S.E.C., C.d.A., G.S.E., and M.S.K. performed research; K.L.P., S.E.C., and C.d.A. contributed new reagents/analytic tools; K.L.P., C.d.A., G.S.E., and M.S.K. analyzed data; G.S.E. and M.S.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates for CcmM193 and CcmM209 have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 3KWC for CcmM209, 3KWD for CcmM193, and 3KWE for the supplemental CcmM209 structure).

¹To whom correspondence should be addressed. E-mail: mkimber@uoguelph.ca.

This article contains supporting information online at www.pnas.org/cgi/content/full/0910866107/DCSupplemental.

thermophilus, and retains the critical metal binding and catalytic residues (17, 18). However, experiments with recombinant CcmM from *Synechococcus elongatus* PCC 7942 and *Synechocystis* sp. PCC 6803 have failed to detect any CA activity (11, 19). To obtain a better understanding of the function of the γ -CA-like domain of CcmM, and evaluate its potential as a functioning CA, we investigated the enzymatic activities and structure of CcmM from *Thermosynechococcus elongatus* BP-1, a thermophilic β -cyanobacterium whose only candidate CA is CcmM.

Results and Discussion

CcmM from *Thermosynechococcus elongatus* BP-1 Encodes an Active γ -CA. Recombinantly expressed full-length CcmM (CcmM652) catalyzed ^{18}O exchange between $^{13}\text{C}^{18}\text{O}_2$ and H_2O (Fig. 1A and Fig. S14). Boiled or denatured CcmM failed to catalyze ^{18}O exchange, whereas 2 mM ethoxzylamide (EZA), an inhibitor of all CAs including Cam (20) reduced the rate of ^{18}O exchange to near the uncatalyzed rate. This enzyme is also active in a CO_2 hydration assay, with 13% of the activity of bovine α -CAII (Table S1). Using an established protocol to fractionate *T. elongatus* cell lysate (11, 13, 14) we also obtained and tested carboxysome-enriched fractions for CA activity (Fig. 1B and C). A low level of CA activity was observed; this activity was almost completely abolished by EZA treatment and fully abolished by boiling the sample. CA activity was not observed in crude cell lysate or in the soluble fraction from which the carboxysomes had been precipitated. A similar pattern of distribution in CA activity has been observed in several other β -cyanobacteria (19) with known carboxysomal CAs. We therefore conclude that CcmM from *T. elongatus* BP-1 is an enzymatically active carboxysomally associated carbonic anhydrase.

Since the discovery of CA activity in Cam in 1994 (17), all reported tests for CA activity in Cam homologs have proven negative. Examples include the two cyanobacterial CcmM orthologs (11, 19), three γ -CA homologs that participate in *Arabidopsis* mitochondrial complex I (21), the *E. coli* proteins YrdA, CaiE, and PaaY (22), and γ -CA-like proteins from *Corynebacterium glutamicum* (23) and *Chlamydomonas reinhardtii* (24). Our finding CA activity for *T. elongatus* CcmM is significant because it not only establishes the identity of the missing carboxysomal CA, but also verifies the hypothesis that functional γ -CAs are found in bacteria.

Fig. 1. CcmM from *T. elongatus* is a carboxysomally associated γ -carbonic anhydrase. (A) CA activity assays illustrating the catalytic competency of the constructs. Assay condition: 100 mM *N*-(2-hydroxyethyl)piperazine-*N'*-(3-propanesulfonic acid) (EPPS) /NaOH, pH 8.0, 20 mM MgSO_4 , 600 μM $\text{K}_2^{13}\text{C}^{18}\text{O}_3$ at 30 $^\circ\text{C}$; CcmM193, 0.34 $\text{mg} \cdot \text{mL}^{-1}$; CcmM652 (0.22 $\text{mg} \cdot \text{mL}^{-1}$); CcmM209 (0.11 $\text{mg} \cdot \text{mL}^{-1}$); bovine CAII (Sigma) (0.03 $\text{mg} \cdot \text{mL}^{-1}$). EZA inhibits CcmM652 activity, but not completely. The uncatalyzed rate of ^{18}O exchange was nearly identical to the CcmM193 trace. (B) CA activity catalyzed by partially purified *T. elongatus* carboxysomes (cbx) at 4.2 $\text{mg} \cdot \text{mL}^{-1}$; this activity is inhibited by EZA and suppressed by boiling. (C) Negatively stained electron micrograph, showing a *T. elongatus* carboxysome.

We also constructed variants of the CcmM γ -CA-like domain for more detailed structural and biochemical characterization (Fig. S2B). CcmM193 (residues 1–193) terminates at approximately the same position as the Cam C terminus. Surprisingly, this construct lacked detectable CA activity. The CcmM209 construct includes an additional 16 C-terminal amino acids conserved among *ccaA*-less CcmM homologs, but not in *ccaA* containing homologs or Cam. This construct is, on a per mol basis, approximately 1.3 times more active than the wild-type protein (Fig. 1A and Table S1); the activity of this construct implies that at least some of the additional 16 conditionally conserved residues are essential for the CA function of CcmM.

Structure of the γ -CA Domain of CcmM Includes a Unique, Disulfide Containing C-Terminal Motif.

The structure for CcmM209 was determined at 2.0 $\text{\AA}$ from an orthorhombic crystal form. The protomer of the CcmM209 structure is organized around a seven-turn left-handed β -helix, which is packed into a trimer with the axes of the β -helices all parallel (Fig. 2A–C). Each turn of the β -helix consists of three short β -strands of equal length that, in cross-section, resemble a rounded equilateral triangle. All β -stranded cross-overs of the unusual left-handed variety and the faces of the β -helix are remarkably flat. Two corners of this triangle are always formed by tight β -turns with a stereotyped geometry, allowing them to stack closely on the turns above and below them in the β -helix. The third corner is structurally more variable, allows longer insertions, which contribute several putative catalytic residues. The β -helix is capped at the C terminus by a short α -helix, α A. Residues Glu172–Glu196 form a long α -helix, α B, which packs along one face of the β -helix, crossing at a shallow angle so that its C terminus is packed against the adjacent protomer. The C terminus of α B extends beyond the N terminus of the β -helix, followed by a third helix, α C, which packs against the β 1– β 2 loop of the adjacent protomer of the trimer. This C-terminal motif includes a disulfide bond between residues Cys194 and Cys200, which connects the C terminus of α B to the N terminus of α C. A zinc ion is coordinated by two histidine residues extending from equivalent positions of adjacent turns on the β helix, His75 and His107, and His102 contributed by the β -helix of the neighboring monomer (Fig. 2I; detailed metal ion geometry is shown in Fig. S3). The center of the trimer forms a partial channel, which is blocked at the N-terminal side by Phe31 and at the center of the β -helix by the close packing of three symmetry related copies of Arg121; the positive charge on this residue is stabilized by a chloride ion that forms a hydrogen bond with all three guanidinium groups.

This structure closely resembles the canonical γ -CA, Cam, with an rmsd of 0.57 $\text{\AA}$ over 159 $\text{C}\alpha$ atoms for the monomer and 0.8 $\text{\AA}$ for 494 $\text{C}\alpha$ atoms for the trimer (Fig. 2E). Significant differences between the two structures are localized to three areas: Cam has residues 85–95 inserted onto the β 10– β 11 loop, including two acidic residues that appear to enhance the overall catalytic rate of the enzyme by acting as proton shuttles (25, 26). In CcmM, α B has two additional turns of helix relative to Cam, where the equivalent residues (186–194) form an extended loop structure with two additional residues inserted. Finally, the residues that form α C and the α B– α C loop are not present in Cam. The catalytic sites of these two structures overlay closely (Fig. 2L).

Conserved Surface on CcmM is a Potential Protein-Binding Surface.

When the sequence conservation of all CcmM orthologs is mapped onto the CcmM structure (Fig. S5), a large conserved surface patch is found centered on Asn126, Asp143, and Glu145, which are all absolutely conserved. These residues are not conserved in other γ -CA homologs, suggesting the role of this conserved patch is specific to CcmM and not limited to either the *CcaA*-containing or *CcaA*-less species. We suggest that this region mediates protein–protein interactions. Recent studies

Fig. 2. The structure of CcmM. (A) Structure of the CcmM209 monomer, with secondary structure labeled. (B) CcmM209 trimer viewed down the 3-fold axis. The zinc binding histidine residues, as well as R121 (which binds a structural chloride ion), are shown in sticks. The zinc ion is shown as a pink sphere. (C) The CcmM209 trimer viewed orthogonal to the 3-fold axis. Individual monomers are arranged with the β -helical axis parallel to one another, while α C interacts with the adjacent protomer. (D) Details of the inset area, with α A weighted $2mF_o - DF_c$ electron density contoured at 1.0σ (blue) shown for residues E171–L208 of chain A (orange sticks), and for residues P9–L17 of chain B (yellow sticks). Electron density in this region is well defined, with temperature factors comparable to elsewhere in the structure. Density is also contoured at 3.0σ (green surface) for C194 and C200, showing the density associated with the sulfur atoms participating in the disulfide bond. For comparison, V182 is the last residue ordered in the CcmM193 structure. (E) Superposition of apo zinc Cam (1qrg; white and yellow) on CcmM209 (blue, α C in orange). Aside from the highlighted localized differences, the two structures are overall very similar. (F) Ribbon diagram of CcmM193 (white) superimposed on CcmM209 (blue) demonstrating that structural differences are localized, but substantial. Residues 4–16 (orange in CcmM209) are disordered in the CcmM193 structure. Of residues 172–208, which in CcmM209 comprise α B and α C (brown), only residues 172–182 are partially ordered in the CcmM193 structure (red), and these residues are displaced from the position seen in CcmM209. (G) Inset showing electron density for the CcmM193 structure. Density is contoured at 1σ (light blue) and 4σ (dark blue). The electron density for α B is considerably less defined than for the rest of the structure. (H) The oligomeric organization differs between CcmM209 (blue) and CcmM193 (white) structures. The structures were superimposed with reference to the protomer on the right only. The motion the CcmM193 and CcmM209 structures can be described as the rotation of each protomer approximately 6° away from the 3-fold axis, with the fulcrum located near R121. (I) Details of the CcmM209 catalytic site. The α B helix, which covers the catalytic site, is shown in cyan in transparent cartoon representation. (J) Details of the CcmM193 putative catalytic site. Residues labeled “b” are contributed by a symmetry related molecule. (K) Superposition of the CcmM209 active site (blue) on CcmM193 (white). Note that many of the residues contributed by the “b” side of the pocket are misplaced in CcmM193 and fail to form hydrogen bonds important for catalytic activity. (L) Overlay of the CcmM209 (blue) and Cam (yellow) catalytic sites. Note that, despite the low overall sequence identity ($\sim 35\%$), all catalytic site residues are conserved and adopt identical conformations. See Fig. S3 for details of the metal ion geometry, and Fig. S4 for further details of the catalytic site.

investigating the network of protein–protein interactions inherent in the β -carboxysome have implicated CcmM as a central organizing scaffold. The C-terminal domain portion binds RuBisCO, while the N-terminal domain binds CcaA, CcmN, CcmK1, and possibly several of the shell proteins (11, 14). Potentially, CcmK1 may interact with the β -helix through α A at the C terminus of the β -helix, where the 3-fold symmetry of CcmM matches the CcmK shell facet’s either 6-fold (at the center of a given hexamer) or 3-fold (at the meeting point of three hexameric “tiles”) symmetry (27). The details of this interaction are of interest as this is the only known specific interaction between the core and the shell of the β -carboxysome.

Structure of CcmM193 Shows Disordered N and C termini. The structure of the CcmM193 construct (determined at $1.1 \text{ \AA}$ from a rhombohedral crystal) maintains the core β -helix motif seen in the CcmM209 structure, but differs in several important details. The residues corresponding to α C in the CcmM209 structure are not present in this construct, so this motif is obviously not observed. However, the residues comprising α B are also represented by relatively weak and incomplete electron density (Fig. 2F and G). This density is interpretable as an α -helix with occupancy of 0.5 – 0.6 , and elevated B-factors; the helix density is clearly interpretable for three and a half turns, and becomes too diffuse to reliably trace beyond Val182. The positioning of α B in the

CcmM193 structure diverges noticeably from the equivalent motif in the CcmM209 structure, pointing toward, rather than covering, the metal binding site. Residues 1–16, which form $\beta 1$ and the $\beta 1$ – $\beta 2$ loop in the CcmM209 structure, are also wholly disordered in the CcmM193 structure. In CcmM209, the $\beta 1$ – $\beta 2$ loop packs against αC and the C terminus of αB of the adjacent monomer (Fig. 2C and D), so the disordering of these residues is likely a secondary consequence of the changes in helical structure. The αB and αC helices, as well as the $\beta 1$ – $\beta 2$ loop mediate extensive interactions between protomers, and the overall organization of the oligomer is also altered. The N terminus of each β -helix is rotated approximately 6° outward from the opposite protomer, around a point near Arg121 (Fig. 2H). This is accompanied by significant repacking of several residues that mediate interprotomer interactions, including several residues that stabilize the catalytic site.

CcmM193 Active Site Disorganization Explains the Lack of Catalytic Activity. The catalytic site of γ -CA is located at the protomer–protomer interface. Although the catalytic sites of CcmM209 and Cam superimpose closely (Fig. 2L), the relative motions of adjacent protomers in CcmM193 shifts the two halves of the active site relative to one another, rearranging most of the known catalytically important motifs (Fig. 2I–K and Fig. S4). His102 is contributed by a different protomer than the other two zinc ligands, His75 and His107, resulting in the His75–Zn–His102 angle being 6° wider in CcmM193. Instead of hydrogen bonding with Asp70b's carboxylate group, His102 N81 hydrogen bonds with its carbonyl oxygen. In mutagenesis experiments in α -CAII, analogous second shell ligand substitutions alter the pK_a of the zinc-bound water by as much as 0.9-pH units (28). The Asp70b carboxylate makes one, instead of two H bonds, with Arg53, which adopts two conformations in CcmM193—one displaced approximately 1.5 Å away from the catalytic site, and the other wholly displaced into solvent. In Cam, the Arg59Lys mutant shows a 20-fold reduction in k_{cat} , indicating that the interactions this residue mediates are critical for activity (29). In Cam, Gln75, positioned by Asn73, hydrogen bonds to the metal-bound substrate; mutations in these two residues show a 4–50-fold reduction in k_{cat} (30). In CcmM209, the equivalent residues, Gln69b and Asn67b, are displaced, with the latter adopting a different rotamer. The essential residue Glu56 (Glu62 in Cam) (25) also adopts an alternative rotamer in CcmM193, but still forms a direct H bond to the zinc-bound water; this rotamer is also seen in some CcmM209 active sites. In Cam Asn202, mutants result in 10-fold reduction in k_{cat} (30). In CcmM209, the equivalent residue, Asn184, is wholly disordered, along with most of αB , which otherwise forms one wall of the catalytic site. In summary, almost all functionally important residues in the CcmM193 catalytic site are significantly disrupted; given that enzyme catalysis can require control of substrate–enzyme interactions with 0.1 Å precision (31), it is unsurprising that this construct is inactive.

CcmM with Its αB Substituted with Cam αB Is Active as a CA. The inactivity of the CcmM193 construct appears to arise from the deletion of the helix αC . However, Cam stops at the equivalent of residue 196, and therefore entirely lacks this motif. We suggest that, in Cam, but not CcmM, the residues of αB alone make sufficient interactions to stabilize the protein in an active conformation. To test this idea, we made three constructs that replaced all or part of the C-terminal helix, αB , from CcmM with the equivalent residues from Cam (Fig. S2B). One of these constructs, CcmM χ 173, which replaces the entire C-terminal helix with the equivalent Cam sequence, has 11% of the activity of CcmM209, validating this idea (Fig. S1B and Table S1). Interestingly, a second construct, CcmM χ 182, is also weakly but measurably active, despite changing only 10 residues relative to

CcmM193, of which only four make contact with the rest of the protein in the Cam structure. This result implies that the requirement for $\alpha 3$ arises because $\alpha 2$ in CcmM, unlike in Cam, does not offer sufficient interaction energy to lock the protein into an active conformation, and furthermore, that the required amount of stabilization energy is likely small.

Reducing Conditions Inactivate and Disrupt the Structure of CcmM.

To test whether the Cys194–Cys200 disulfide bond is required for CA activity, we constructed a Cys200Ser mutant (CcmM209C200S). This protein proved to be inactive as a CA. We also tested for CA activity in the CcmM209 and CcmM652 constructs in the presence of reducing agents. CA activity was found to be largely abolished in the presence of 1 or 10 mM tris(2-carboxyethyl)phosphine (TCEP) (Fig. 3A), 15 mM β -mercaptoethanol, or 13 mM DTT (Fig. S2C). In contrast, the catalytic activity of the CcmM χ 173 construct was not attenuated in the presence of 10 mM TCEP (Fig. 3A), implying that TCEP does not directly impact the active site. These results indicate that the catalytic activity of CcmM is critically dependent upon the oxidation of the disulfide bond. CcmM209 protein screened in the presence of 1 mM TCEP fails to crystallize, suggesting that protein reduction is accompanied by significant structural rearrangement. This idea is reinforced by tryptophan fluorescence measurements. Trp13 is located on the $\beta 1$ – $\beta 2$ loop and in the CcmM209 structure, interacts with Ile184, Ala187, Leu188 from $\alpha 2$, as well as Thr11, Gln78, and Asp55, burying approximately 78% of its surface. In the CcmM193 structure, this residue is wholly disordered and therefore presumably highly exposed and fluorescently quenched. The inactive CcmM193 construct has 30% less intense fluorescence than the constitutively active CcmM χ 173 construct (Fig. 3B). Under reducing conditions, the fluorescence of the CcmM209 construct is attenuated from CcmM χ 173-like levels to CcmM193-like levels. This implies that the structure of CcmM209 resembles CcmM193 under reducing conditions and undergoes a concerted structural transition from an inactive to an active state when conditions become oxidizing. CcmM193 therefore represents a “preactivation” state, where the αB helix is poised to lock the structure into a fully activated state, but cannot derive sufficient energy from the available interactions to do so.

In the *E. coli* cytosol, disulfide bonds are reduced through the action of thioredoxin and glutaredoxin. Consistent with this, the CA activity of recombinant CcmM209 in freshly prepared *E. coli* cell lysate was remarkably low (Fig. 3C). Treatment with diamide, a thiol oxidizing agent, resulted in a dramatic increase in lysate activity, whereas diamide failed to activate fresh lysate containing CcmM209C200S, indicating that this effect is specific to disulfide bond formation.

CcmM and CcaA are Complementary CAs in β -cyanobacterial Carboxysomes.

Of 26 sequenced β -cyanobacterial genomes, nine lack a putative CcaA. In all of these species, the γ -CA domain of CcmM is at least 70% identical. Among those that contain a *ccaA* ortholog, many show a similar level of conservation, but a subset of *cmmM* sequences is much more divergent, with as little as 39% sequence identity in the γ -CA like domain. The core β -helix is conserved, with differences preferentially mapping to the N- and C-terminal motifs disordered in the CcmM193 structure. Seven species fail to conserve the PPTPW motif from the $\beta 1$ – $\beta 2$ loop, as well αB , αC , and the two cysteines that form the disulfide bond (Fig. 3D and E, and Fig. S5A). CcmM from *Synechocystis* sp. PCC 6803 and *Synechococcus elongatus* PCC 7942, which have been confirmed to be inactive as CAs (11, 19), are among this group. Given the absence of the potential to form stabilizing $\beta 1$ – $\beta 2$ loop/ αB / αC interactions, the structures of these constitutively inactive proteins likely resemble the CcmM193 structure. CcmM from *Synechococcus* sp. PCC 7335 is further divergent, because it also lacks most of the catalytic machinery including

Fig. 3. Reduction of the Cys194-Cys200 disulfide bond inactivates and partially disorders CcmM209. (A) Effect of 10 mM TCEP on the catalytic activity of CcmM209 (0.02 mg · mL⁻¹) and CcmMx173 (0.01 mg · mL⁻¹). CcmMx173 is constitutively active under both oxidizing and reducing (10 mM TCEP) conditions, whereas CcmM209 is active under oxidizing conditions, but activity is reduced to uncatalyzed rates in the presence of 10 mM TCEP. The CcmM209C200S variant is inactive under both oxidizing and reducing conditions. Assay condition: 200 mM EPPS/NaOH, pH 7.8, 1 mM NaCl, 600 μM K₂¹³C¹⁸O₃ at 30 °C. (B) Tryptophan fluorescence of CcmM193, CcmM209, and CcmMx173 under oxidizing and reducing (10 mM DTT) conditions. The constitutively active CcmMx173 chimera is strongly fluorescent under both conditions, the constitutively inactive construct CcmM193 is more weakly fluorescent under both conditions, while the CcmM209 switches from strongly fluorescent under oxidizing, to weakly fluorescent under reducing conditions. This is consistent with the quenching of W13 caused by the unstructuring of the β1-β2 loop and the α2 and α3 helices upon reduction of the C194-C200 disulfide bond. (C) *E. coli*'s cytosolic reducing machinery keeps CcmM209 in an inactive state. Activity traces for fresh cell lysate containing over expressed CcmM209 in the absence and presence (25 mM) of the thiol oxidizing agent diamide. Activation of CcmM209 also occurs during the purification process due to oxidation of the protein by molecular O₂. Cell lysate containing the CcmM209C200S variant remains inactive in the presence of diamide, indicating that the activation mechanism is dependent upon the potential to form a C194-C200 disulfide bond. (D) Details showing the regions of the β1-β2 loop and α2-α3 helices that are necessary for stabilizing the protein in an active conformation. (E) Excerpt of an alignment of CcmM sequences. Species are abbreviated as Te, *Thermosynechococcus elongatus*-BP1; Gv, *Gloeobacter violaceus* PCC 7421; Tric, *Trichodesmium erythraeum* IMS101; Np, *Nostoc punctiforme* PCC 73102; Cw, *Crocospaera watsonii* WH 8501; 6803, *Synechocystis* sp. PCC 6803; 7942, *Synechococcus elongatus* PCC 7942; 7335, *Synechococcus* sp. PCC 7335. Among the shown species, Te, Gv, and Tric contain no ccaA homolog; the other species shown do. The sequences for 6803, 7942, and 7335 do not conserve C194, C200 (green ellipses), the β1-β2 loop (cyan box with W13 marked with a cyan ellipse), or critical elements in the α2 and α3 helices (orange box, N184 magenta ellipse). Consequently, they (along with six other species not depicted) are unlikely to show CA activity.

two of the three metal-binding histidine residues. *Synechococcus* sp. PCC 7002 does not conserve αC or the Cys194-Cys200 pair, though it does have the β1-β2 loop motif. The remaining nine sequenced β-cyanobacterial genomes contain ccaA as well as a CcmM protein that appears possibly functional as a CA, though the presence of more subtle inactivating mutations cannot be excluded; however, it is possible that CcaA and CcmM are both functioning as CAs in these species, with presumably complementary roles.

The most basal cyanobacterial species, including *Gloeobacter violaceus* PCC 7421 and *Thermosynechococcus elongatus* BP-1 (3), have only CcmM. This suggests that the last common ancestor of cyanobacteria included only CcmM as a carboxysomal CA, and that CcaA was obtained independently in different lineages by horizontal gene transfer (HGT), followed, in some instances, with loss of CA activity from CcmM through the accumulation of deleterious mutations. A clear instance of this can be found in the *Nostocaceae* cyanobacteria. *Nodularia spumigena* CCY9414, *Nostoc* sp. PCC 7120, *Anabaena variabilis* ATCC 29413, and *Nostoc punctiforme* ATCC 29133 are closely related cyanobacteria whose CcmM N-terminal domains are 80–99% identical and have the hallmarks of a functional γ-CAs. CcaA is found only in the latter two species. However, the CcaA proteins in ATCC 29413 and ATCC 29133 share only 62% sequence identity and are located on different parts of the genome; each protein is significantly more similar to a separate ccaA paralog of *Cyanothece* sp. PCC 7425 (68 and 77%, respectively). This suggests that ccaA was acquired separately in these two species, rather than inherited from a common ancestor. In support of this idea, CcmM652

immobilized on an affinity column specifically binds *Synechocystis* PCC 6803 CcaA. This suggests that CcaA interacts with a CcmM motif that is universally conserved for other reasons, and will be appropriately assembled into the carboxysome if the ccaA gene is acquired by HGT (Fig. S6). In summary, it is clear that β-cyanobacteria include species that utilize only CcaA, only CcmM, or utilize both as their carboxysomal CA. The two primary laboratory model cyanobacteria—*Synechocystis* sp. PCC 6803 and *Synechococcus elongatus* PCC 7942—both belong to the first group. Our findings here suggest that a fuller appreciation of the diversity of cyanobacterial photosynthetic physiology may necessitate also studying examples of the latter two groups.

Carboxysomes Function as a Redox Privileged Compartment of the Cell. Eubacteria, like eukaryotes, have a cytosol that is markedly reducing, and, to the best of our knowledge, structural disulfide bonds required for protein activity have never been demonstrated for bacterial cytosolic proteins. This is not to say that disulfide bonds are not observed in the structures of cytosolic proteins; however, these either involve reversible disulfide bond formation as part of their catalytic cycle, or involve nonconserved residues and are generally thought to be artifacts of the oxidizing conditions that arise during handling. The exception to this rule is hyperthermophilic bacteria such as *Thermotoga maritima*, whose cytosolic proteins routinely contain structural disulfide bonds. In these species, as well as in hyperthermophilic archaea, cytosolic disulfide bond formation is mediated by protein disulfide oxidoreductase (32); this protein is absent from cyanobacteria. We have demonstrated here that the Cys194-Cys200 disulfide

bond in CcmM is required for this protein to be properly structured and to function as a CA. Given that Cam functions without such a disulfide bond, and that it is possible to make chimeric variants of CcmM that alter very few amino acids but are constitutively active, we conclude that oxidative activation of CcmM is functionally selected for. Ectopic expression of an active CA in the cyanobacterial cytosol short circuits the CCM (33), suggesting a clear adaptive rationale for having CcmM's CA inactivated by the cytosol's reducing machinery. This regulatory mechanism is only feasible if CcmM's ultimate cellular locality—the interior of the carboxysome—is sufficiently oxidizing to promote disulfide bond formation. The small (~40 Å in diameter) proteins thioredoxin and glutaredoxin reduce oxidized cytosolic proteins, maintaining the cytosol's overall reducing potential (34). The shell of the carboxysome presents a formidable barrier to the passage of these proteins as the pores that perforate it [generally ~6 Å in diameter, though recent work suggests that gated pores as large as 14 Å may also be present (35)] allow passage of small metabolites, but not proteins. In the absence of these agents, molecular oxygen (atmospheric and as a byproduct of photosynthesis) should rapidly establish oxidizing conditions. Carboxysomes and related microcompartments are commonly described as “bacterial organelles” (36): the finding that the carboxysome interior can maintain a redox state distinct from the bulk cytosol further extends and strengthens this analogy.

Materials and Methods

Molecular Biology, Protein Expression, and Purification. All expression constructs were made in pET-28a with N-terminal 6xHis tags, using standard molecular biology and protein purification techniques. Details are given in *SI Materials and Methods*. Primers used are described in *Fig. S2A*.

Culturing *T. elongatus* and Carboxysome Enrichment. *T. elongatus* was grown using standard culturing conditions. Carboxysomes were purified using the Mg^{2+} /Percoll precipitation method as described in refs. 13, 14. Details are given in *SI Materials and Methods*.

Carbonic Anhydrase Activity Assays. CcmM constructs were assayed for CA activity using the ^{18}O exchange method as described in ref. 11. Details are given in *SI Materials and Methods*.

Crystallization and Structure Determination. Large prismatic crystals of CcmM193 were grown from 14 mg/mL protein in a solution of 1.0 M ammonium sulphate, 0.1 M [bis(2-hydroxyethyl)amino]tris(hydroxymethyl)methane, 1% PEG 3350, pH 5.5. The CcmM209 crystals were grown from 58 mg/mL protein, 20% isopropanol, 30% PEG 4000, 0.1 M Na Citrate pH 5.5. Both crystals were flash-frozen using paratone-N oil as a cryoprotectant. Diffraction data for the CcmM193 crystals were collected to 1.1 Å at the Advanced Photon Source on beamline 19ID. Data for the CcmM209 crystals were collected to 2.0 Å on a Rigaku MicroMax-007 HF (40 kV, 30 mA) with Osmic multilayer optics and a Mar345 image plate detector. Both structures were determined using molecular replacement and refined using standard methods, as described in *SI Materials and Methods*. Data collection and structure determination statistics are given in *Table S2*.

ACKNOWLEDGMENTS. We would like to thank Dr. J.G. Ferry for the Cam construct, Prof. M. Ikeuchi and the Kazusa DNA Research Institute for the gift of *T. elongatus* cells and genomic DNA, respectively, Dr. Y. Tadesse and D. Arefeen for assistance with carboxysome isolation, the staff of APS 19ID for assistance during collection of the CcmM193 data, Dr. P. Grochowski for performing x-ray fluorescence analysis, Dr. C. Khursigara for collecting the electron micrographs, and J.P. Julien for assistance in collecting the CcmM209 dataset. This research was supported by a Canada Graduate scholarship (C.d.A.) and Discovery Grants from the Natural Sciences and Engineering Research Council of Canada [G.S.E. and to MSK (327280-06)].

- Andersson I, Backlund A (2008) Structure and function of Rubisco. *Plant Physiol Biochem*, 46:275–291.
- Badger MR, Price GD (2003) CO₂ concentrating mechanisms in cyanobacteria: Molecular components, their diversity and evolution. *J Exp Bot*, 54:609–622.
- Shi T, Falkowski PG (2008) Genome evolution in cyanobacteria: The stable core and the variable shell. *Proc Natl Acad Sci USA*, 105:2510–2515.
- Iancu CV, et al. (2007) The structure of isolated *Synechococcus* strain WH8102 carboxysomes as revealed by electron cryotomography. *J Mol Biol*, 372:764–773.
- Schmid MF, et al. (2006) Structure of *Halothiobacillus neapolitanus* carboxysomes by cryo-electron tomography. *J Mol Biol*, 364:526–535.
- Tsai Y, et al. (2007) Structural analysis of Cso51A and the protein shell of the *Halothiobacillus neapolitanus* carboxysome. *PLoS Biol*, 5:1345–1354.
- Dou Z, et al. (2008) CO₂ fixation kinetics of *Halothiobacillus neapolitanus* mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO₂. *J Biol Chem*, 283:10377–10384.
- Price GD, et al. (1998) The functioning of the CO₂ concentrating mechanism in several cyanobacterial strains: A review of general physiological characteristics, genes, proteins, and recent advances. *Can J Bot*, 76:973–1002.
- Ludwig M, Sultemeyer D, Price GD (2000) Isolation of ccmKLMN genes from the marine cyanobacterium, *Synechococcus* sp. PCC7002 (Cyanobacteria), and evidence that CcmM is essential for carboxysome assembly. *J Phycol*, 36:1109–1118.
- Price GD, Badger MR (1991) Evidence for the role of carboxysomes in the cyanobacterial CO₂-concentrating mechanism. *Can J Bot*, 69:963–973.
- Cot SS, So AK, Espie GS (2008) A multiprotein bicarbonate dehydration complex essential to carboxysome function in cyanobacteria. *J Bacteriol*, 190:936–945.
- Price GD, Howitt SM, Harrison K, Badger MR (1993) Analysis of a genomic DNA region from the cyanobacterium *Synechococcus* sp. strain PCC7942 involved in carboxysome assembly and function. *J Bacteriol*, 175:2871–2879.
- Long BM, Price GD, Badger MR (2005) Proteomic assessment of an established technique for carboxysome enrichment from *Synechococcus* PCC7942. *Can J Bot*, 83:746–757.
- Long BM, Badger MR, Whitney SM, Price GD (2007) Analysis of carboxysomes from *Synechococcus* PCC7942 reveals multiple Rubisco complexes with carboxysomal proteins CcmM and CcaA. *J Biol Chem*, 282:29323–29335.
- Fukuzawa H, Suzuki E, Komukai Y, Miyachi S (1992) A gene homologous to chloroplast carbonic anhydrase (icfa) is essential to photosynthetic carbon-dioxide fixation by *Synechococcus* PCC7942. *Proc Natl Acad Sci USA*, 89:4437–4441.
- So AK, John-McKay M, Espie GS (2002) Characterization of a mutant lacking carboxysomal carbonic anhydrase from the cyanobacterium *Synechocystis* PCC6803. *Planta*, 214:456–467.
- Alber BE, Ferry JG (1994) A carbonic anhydrase from the archaeon *Methanosarcina thermophila*. *Proc Natl Acad Sci USA*, 91:6909–6913.
- Kisker C, Schindelin H, Alber BE, Ferry JG, Rees DC (1996) A left-hand beta-helix revealed by the crystal structure of a carbonic anhydrase from the archaeon *Methanosarcina thermophila*. *EMBO J*, 15:2323–2330.
- So AK, Espie GS (2005) Cyanobacterial carbonic anhydrases. *Can J Bot*, 83:721–734.
- Zimmerman S, et al. (2004) Carbonic anhydrase inhibitors. Inhibition of the prokaryotic beta and gamma-class enzymes from Archaea with sulfonamides. *Bioorg Med Chem Lett*, 14:6001–6006.
- Parisi G, et al. (2004) Gamma carbonic anhydrases in plant mitochondria. *Plant Mol Biol*, 55:193–207.
- Merlin C, Masters M, McAteer S, Coulson A (2003) Why is carbonic anhydrase essential to *Escherichia coli*? *J Bacteriol*, 185:6415–6424.
- Mitsuhashi S, Ohnishi J, Hayashi M, Ikeda M (2004) A gene homologous to beta-type carbonic anhydrase is essential for the growth of *Corynebacterium glutamicum* under atmospheric conditions. *Appl Microbiol Biotechnol*, 63:592–601.
- Mitra M, Lato SM, Ynalvez RA, Xiao Y, Moroney JV (2004) Identification of a new chloroplast carbonic anhydrase in *Chlamydomonas reinhardtii*. *Plant Physiol*, 135:173–182.
- Tripp BC, Ferry JG (2000) A structure-function study of a proton transport pathway in the gamma-class carbonic anhydrase from *Methanosarcina thermophila*. *Biochemistry*, 39:9232–9240.
- Tu C, Rowlett RS, Tripp BC, Ferry JG, Silverman DN (2002) Chemical rescue of proton transfer in catalysis by carbonic anhydrases in the beta- and gamma-class. *Biochemistry*, 41:15429–15435.
- Kerfeld CA, et al. (2005) Protein structures forming the shell of primitive bacterial organelles. *Science*, 309:936–938.
- Christianson DW, Cox JD (1999) Catalysis by metal-activated hydroxide in zinc and manganese metalloenzymes. *Annu Rev Biochem*, 68:33–57.
- Tripp BC, Tu C, Ferry JG (2002) Role of arginine 59 in the gamma-class carbonic anhydrases. *Biochemistry*, 41:669–678.
- Zimmerman SA, Ferry JG (2006) Proposal for a hydrogen bond network in the active site of the prototypic gamma-class carbonic anhydrase. *Biochemistry*, 45:5149–5157.
- Signala PA, et al. (2008) Testing geometrical discrimination within an enzyme active site: Constrained hydrogen bonding in the ketosteroid isomerase oxyanion hole. *J Am Chem Soc*, 130:13696–13708.
- Ladenstein R, Ren B (2006) Protein disulfides and protein disulfide oxidoreductases in hyperthermophiles. *FEBS J*, 273:4170–4185.
- Price GD, Badger MR (1989) Expression of human carbonic anhydrase in the cyanobacterium *Synechococcus* PCC7942 creates a high CO₂-requiring phenotype: Evidence for a central role for carboxysomes in the CO₂ concentrating mechanism. *Plant Physiol*, 91:505–513.
- Carmel-Harel O, Storz G (2000) Roles of the glutathione- and thioredoxin-dependent reduction systems in the *Escherichia coli* and *Saccharomyces cerevisiae* responses to oxidative stress. *Annu Rev Microbiol*, 54:439–461.
- Klein MG, et al. (2009) Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. *J Mol Biol*, 392:319–333.
- Tanaka S, et al. (2008) Atomic-level models of the bacterial carboxysome shell. *Science*, 319:1083–1086.