The small RbcS-like domains of the β-carboxysome structural protein CcmM bind RubisCO at a site distinct from that binding the RbcS subunit

Received for publication, October 18, 2018, and in revised form, December 13, 2018 Published, Papers in Press, December 27, 2018, DOI 10.1074/jbc.RA118.006330

Patrick Ryan‡, Taylor J. B. Forrester‡, Charles Wroblewski‡, Tristan M. G. Kenney‡, Elena N. Kitova‡, John S. Klassen‡, and Matthew S. Kimber‡1

From the ‡Department of Molecular and Cellular Biology, University of Guelph, Guelph, Ontario N1G 2W1, Canada and §Alberta Glycomics Centre and Department of Chemistry, University of Alberta, Edmonton, Alberta T6G 2G2, Canada

Edited by Norma M. Allewell

Carboxysomes are compartments in bacterial cells that promote efficient carbon fixation by sequestering RubisCO and carbonic anhydrase within a protein shell that impedes CO2 escape. The key to assembling this protein complex is CcmM, a multidomain protein whose C-terminal region is required for RubisCO recruitment. This CcmM region is built as a series of copies (generally 3–5) of a small domain, CcmM5, joined by unstructured linkers. CcmM5 domains have weak, but significant, sequence identity to RubisCO’s small subunit, RbcS, suggesting that CcmM binds RubisCO by displacing RbcS. We report here the 1.35-Å structure of the first Thermosynechococcus elongatus CcmM5 domain, revealing that it adopts a compact, well-defined structure that resembles that of RbcS. CcmM5s, however, lacked key RbcS RubisCO-binding determinants, most notably an extended N-terminal loop. Nevertheless, individual CcmM5s domains are able to bind RubisCO in vitro with 1.16 μM affinity. Two or four linked CcmM5 domains did not exhibit dramatic increases in this affinity, implying that short, disordered linkers may frustrate successive CcmM5 domains attempting to simultaneously bind a single RubisCO oligomer. Size-exclusion chromatography-coupled right-angled light scattering (SEC-RALS) and native MS experiments indicated that multiple CcmM5 domains can bind a single RubisCO holoenzyme and, moreover, that RbcS is not released from these complexes. CcmM5 bound equally tightly to a RubisCO variant in which the α/β domain of RbcS was deleted, suggesting that CcmM5 binds RubisCO independently of its RbcS subunit. We propose that, instead, the electropositive CcmM5 may bind to an extended electronegative pocket between RbcL dimers.

Cyanobacteria are oxygenic photosynthetic bacteria that, like higher plants, fix carbon dioxide using the Calvin cycle with

ribulose-bisphosphate carboxylase/oxygenase (RubisCO); EC 4.1.1.39) catalyzing the key inorganic carbon fixation reaction (1). RubisCO catalyzes a chemically challenging reaction made more difficult by modern low ambient CO2 and high O2 concentrations, the latter acting as a competing substrate that results in an unwanted side product that requires energy to recycle (2). Cyanobacteria enhance the efficiency of this reaction by expending energy to concentrate intracellular inorganic carbon using a varied set of CO2 and HCO3− pumps (3). Because CO2 is lipophilic and readily escapes through cellular membranes, cyanobacteria accumulate only HCO3− in the cytosol and encapsulate RubisCO behind a secondary, (relatively) CO2-impermeable protein barrier to form a carboxysome (5–7). Carbonic anhydrase (EC 4.2.1.1), the enzyme that interconverts CO2 and HCO3−, is also encapsulated so that HCO3− pumped into the cell only evolves into CO2 once within the carboxysome shell (8). Carboxysomes are polyphyletic with two deeply divergent lineages, termed α- and β-carboxysomes. α-Carboxysomes, which contain form 1A RubisCO, likely originated in chemoheterotrophic α-proteobacteria and were then horizontally transferred to α-cyanobacteria (7). β-Carboxysomes, which encapsulate form 1B RubisCO and are found in all other cyanobacteria, are the focus of this study. In addition to their direct environmental import, researchers aiming to increase the carbon fixation efficiency of crops have started work on the daunting challenge of engineering aspects of cyanobacterial carbon physiology, including carboxysomes, into angiosperm hosts (4). Gaining a deeper understanding of carboxysome function and assembly is an important milestone in these efforts.

β-Carboxysomes are complex cellular bodies, generally at least 150 nm in diameter (and 300 MDa in molecular mass), built from 9–13 distinct proteins in tens of thousands of copies. The carboxysome shell is relatively thin (3–6 nm) and at least approximately icosahedral. The main component of the shell is small proteins of the Pfam0936 family (CcmK1, CcmK2, CcmK3, CcmK4, and CcmO), which in their basic organization form hexameric (or pseudohexameric) rings of small subunits, that can tile to form the continuous sheets that make up the

1 To whom correspondence should be addressed. Tel.: 519-824-4120 (ext. 52568); Fax: 519-837-1802; E-mail: mkimber@uoguelph.ca.
shell facets (9–12). A more distantly related protein of the same family, CcmP, forms a double ring of fused protomers and may allow passage of larger metabolites (13, 14). A second, unrelated small protein, CcmL (Pfam3319), forms pentameric species that plug the vertices of the shell (15–18). Together, these form a continuous protein barrier perforated only by small pores that appear optimized to preferentially allow passage of anionic species (19).

Ensuring that all required proteins are targeted to the carboxysome in appropriate quantities requires an extensive network of protein–protein interactions. In β-carboxysomes, CcmM seems to be the central actor in organizing the interior proteins. This large, modular protein has an N-terminal domain that is homologous to γ-class carbonic anhydrases. In many cyanobacteria, the γ-carbonic anhydrase–like domain is a functional carbonic anhydrase; in these species, the carbonic anhydrase activity requires formation of a disulfide bond, and the enzyme is only activated once it is encapsulated behind a mature shell and the environment becomes oxidizing (20, 21). A subset of cyanobacteria also have a separate β-class carbonic anhydrase, CcaA, that either supplements or supplants this matrix; this is then partitioned off behind a growing shell (22–24). The N-terminal domain of CcmM also recruits CcmN, another absolutely required protein that is believed to help recruit the protein shell (25). In addition to the N-terminal domain, CcmM’s C terminus contains three to five copies (depending on the strain) of an ~90-amino-acid domain, abbreviated here as CcmMs. Cyanobacterial (and plant) RubisCO forms a hexadecameric heterooligomeric complex with eight copies of a 55-kDa large, catalytic subunit, RbcL, and eight copies of a 14-kDa small subunit, RbcS (1); RbcS’s role seems to lie primarily in stabilizing the oligomer during assembly and fine-tuning RubisCO’s activity and specificity (26). CcmMs domains have appreciable sequence similarity to RbcS (27) and similarly bind RubisCO (28). Joining the CcmMs domains are short segments of poorly conserved, highly hydrophilic sequence that likely act as linkers or spacers. CcmM is also unusual in that it is made in two distinct isoforms: a full-length protein, encompassing all domains (generally referred to as CcmM58), and a shorter version, translated from a conserved internal ribosome entry site, that includes only the CcmMs domains and linkers (CcmM35) (29). Both isoforms are required to form a carboxysome.

Carboxysome assembly appears to be initiated by interactions of CcmM and RubisCO, which together form a dense matrix; this is then partitioned off behind a growing shell (30, 31). Both CcmM isoforms are recruited from the beginning of carboxysome formation and remain evenly distributed through the lumen of the mature carboxysome (32). If CcmM is absent, RubisCO remains diffusely distributed through the cytoplasm, and carboxysome formation is never initiated (30). In many cyanobacteria, a homolog of RubisCO activase (Rca) with a single CcmMs-like C-terminal domain is found; these domains are proposed to similarly drive interactions with RubisCO, ensuring that the activase is appropriately encapsulated within the carboxysome.

After the growth and maturation of a procarboxysome consisting primarily of RubisCO and CcmM, the final stages of maturation involve recruitment of a shell, which is proposed to be able to cleave through oversized RubisCO/CcmM procarboxysomes (31). Shell recruitment also correlates with changes to the interior of the carboxysome, notably reorganization of RubisCO into a paracrystalline array (30, 33) and a switch to an oxidizing environment (31) required to activate the carbonic anhydrase functionality of CcmM (20).

Although the importance of CcmM to the recruitment of RubisCO to the carboxysome is clear, the details of how this occurs remain unresolved. One complication is that RbcL recruits RbcS (which CcmM possibly displaces) through a complex process that generally entails at least one of two distinct dedicated chaperones, RbcX and Raf1, stabilizing the RbcL2 dimers (which themselves require GroEL/ES to fold) and then RbcL8 octamers in a soluble state that permits RbcS recruitment (34). In addition, analysis of protein abundance in cyanobacteria has shown RbcL and RbcS to be present in an approximate 8:5 ratio, leading to the suggestion that the remaining RbcS domains are displaced by CcmM (35). Because CcmMs domains are RbcS homologs, a reasonable hypothesis is then that they share RbcS’s requirement to bind RbcL in a chaperone-mediated assembly process (7, 28, 29, 36). Here, we show that CcmMs are structurally similar to RbcS but lack the key motifs associated with RbcL binding. We also show that these domains can bind the mature RubisCO with micromolar affinity but do not cause release of RbcS from the complex and, indeed, bind a RubisCO variant with most of RbcS deleted. Together, these findings indicate that CcmM binds RubisCO at a site distinct from the RbcS site.

**Results**

**CcmM RbcS-like subdomain forms a soluble, well-behaved protein**

We cloned the first (N-terminal–most) RbcS-like domain spanning residues 226–320 of *Thermosynechococcus elongatus* BP-1 CcmM (henceforth CcmM31) into an expression vector as both a His-tagged and tag-free variant. The resulting 94-amino-acid protein fragment was found to be expressed well, could be purified and concentrated up to 120 mg/ml in a simple buffer, and remained soluble indefinitely; this suggests that CcmMs forms a compact, well-folded structural subdomain that is stable independently of other CcmM motifs or any interaction with RubisCO. We also expressed and purified both WT RubisCO and an RbcL/RbcX/RbcSA18 construct (RbcS with a stop codon at position 18, which allows only the first 17 amino acids of RbcS to be translated; henceforth RbcLSA18). Both of these RubisCO variants were also successfully heterologously expressed and purified. For the RbcLSA18 complex, yield and solubility were similar to those of WT RubisCO, and both variants showed similar thermal denaturation profiles except that the RbcLSA18 profile was missing a denaturation peak around 75 °C.

**The structure of CcmMs resembles RbcS**

We determined the structure of CcmMs from a tagless construct that crystallized only in the presence of cobalt thiocyanate (which mediates important crystal contacts). We exploited the cobalt ions to perform cobalt single-wavelength anomalous
forms an additional, short very conserved face to the protein, spanning the exposed (rather than the green-like RubisCO characteristic of surprisingly, the closest matches are from red-like RubisCO domains. Surprisingly, the closest matches are from red-like RubisCO. CcmMS does, however, seem to have hydrophobic interactions that are estimated to contribute at least two-thirds of the estimated interaction energy (~5.6 kcal/mol; calculations using PISA (38)). The absence of this motif in CcmM would be consistent with the notion that it could possibly bind this motif in CcmM binds RubisCO independently of RbcS

The potential functional implications for the similarities and differences between RbcS and CcmM are best understood in the context of the interactions RbcS makes with the rest of Rubisco. RbcS primarily interacts with two adjacent RbcL protomers where most contacts are to different regions of the first two helices and the last helix (plus a helical bundle extension) of the RbcL β-barrel; additional minor interactions are also formed with a short helix that extends from an RbcL chain in the opposite ring of the hexadecamer and adjacent RbcS domains (Fig. 2A). RbcS’s N-terminal loop (residues 3–17) forms extensive interactions with one RbcL protomer (Fig. 2A and B), with the surface buried by this region comprising approximately a third (845 Å² of the 2600 Å²) of the buried surface between RbcS and the rest of Rubisco; these interactions are estimated to contribute at least two-thirds of the estimated interaction energy (~3.9 of ~5.6 kcal/mol; calculations using PISA (38)).

Searching the Protein Data Bank (PDB) with this structure using Dali (37) revealed that the closest structural homologs are, as predicted from sequence similarity, RbcS domains. Surprisingly, the closest matches are from red-like RubisCO (rather than the green-like RubisCO characteristic of β-cyano-bacteria and higher plants), including red alga Prochloron (e.g. Alcaligenes eutrophus PDB code 1bxn; Z-score, 11) and dia-
A second critical RbcS interaction motif is the highly conserved W56KLP/H9252F (where H9252 is a hydrophobic residue) encompassing H9252F.

This motif makes several contacts with RbcL, most notably an extended hydrophobic patch comprising Met55, Leu58, Pro59, Phe61, and Phe90, which together pack primarily on Trp70, Leu73, and Leu74 contributed by an RbcL protomer from the distal ring (Fig. 2C). The corresponding residues in CcmM are more strongly conserved, with a more regular β-strand (albeit with a broken hydrogen bond at Pro265).

**Figure 1. Structure of CcmM, compared with T. elongatus RbcS.** A, overall structural organization of CcmM. B, secondary structure of RbcS, shown in an equivalent orientation to A. C, surface conservation of CcmM homologs, mapped onto the CcmM structure. For C–F, the lower view is the same representation rotated 180° about the vertical axis. D, surface electrostatics for CcmM (calculated using Adaptive Poisson–Boltzmann Solver (APBS)). E, sequence conservation mapped onto the RbcS surface. F, surface electrostatics of RbcS.

**Figure 2.** A, organization of T. elongatus RubisCO (PDB code 2ybv). RbcL is shown in shades of green, and RbcS is in yellow with the N-terminal loop in orange. The trapezoidal boxes indicate the approximate location of the views shown in the subsequent panels. B, interactions between the N-terminal tail of RbcS and RubisCO. The dashed inset shows the interactions CcmM1 (blue) could potentially make with RbcS's N-terminal tail in this position. C, the interactions mediated by RbcS's conserved motif. The inset shows the detail of CcmM1 in the equivalent region; this protein lacks the characteristic β-bulge, and the corresponding residues are smaller, more polar, and weakly conserved.
with no bulge and smaller, more polar and less conserved residues; this region seems unlikely to interact favorably with the hydrophobic patch on RbcL (Fig. 2C, inset).

In summary, CcmM shows clear overall similarity to RbcS, but the two proteins have very few conserved surface residues in common and have different surface properties. Although many of the polar interactions made by RbcS could plausibly be recapitulated by roughly similar residues of a CcmM molecule placed analogously in a Rubisco complex, the absence of an N-terminal loop and the WKLP/H9021F hydrophobic bulge motif in CcmM suggests that CcmM is unlikely to outcompete RbcS for access to its binding site by forming more favorable interactions with RbcL.

CcmM subunit variants bind Rubisco with micromolar affinity

The association and dissociation behavior of Rubisco binding to CcmM was measured using localized surface plasmon resonance (LSPR). CcmM was immobilized on a gold nanoparticle nitrotriacetic acid chip, and an excess of Rubisco was then flowed over the chip as the analyte to determine binding kinetics. Global analysis of the sensorgram traces using a 1:1 binding model yielded an association rate constant, \( k_{on} \), of \( 8.37 \times 10^2 \) M\(^{-1}\) s\(^{-1}\) and a dissociation rate, \( k_{off} \), of \( 9.74 \times 10^{-4} \) s\(^{-1}\) (Fig. 3A and Table 2; note that Rubisco concentrations throughout the paper are reported in terms of RbcL–RbcS heterodimer concentrations; these concentrations and subsequent affinities should be divided by 8 to reflect the concentration of the hexadecamer). The calculated dissociation constant, \( K_D \), of this interaction is therefore \( 1.16 \times 10^{-2} \) M. The calculated half-life of the CcmM–Rubisco complex (the inverse of the \( k_{off} \)) in these experiments is \( \sim 17 \) min.

We also tested binding of Rubisco to constructs with multiple CcmM subdomains joined by linkers. CcmM\(_{51-2}\) (which contains the first two CcmM\(_5\) domains of CcmM and the linker...
CcmM binds RubisCO independently of RbcS

Table 2

| Ligand       | Analyte  | $k_{\text{on}}$ (± S.D.) | $k_{\text{off}}$ (± S.D.) | $K_D$ (± S.D.) |
|--------------|----------|--------------------------|---------------------------|---------------|
| CcmM$_{1}$   | RbcLS   | 8.37 $\times 10^4$ ± (4.83 $\times 10^3$) | 9.74 $\times 10^{-6}$ ± (7.34 $\times 10^{-7}$) | 1.16 $\times 10^{-6}$ ± (6.83 $\times 10^{-8}$) |
| CcmM$_{1,2}$ | RbcLS   | 4.2 $\times 10^3$ ± (1.2 $\times 10^2$) | 7.17 $\times 10^{-4}$ ± (1.25 $\times 10^{-5}$) | 1.71 $\times 10^{-5}$ ± (7.85 $\times 10^{-7}$) |
| CcmM$_{1,4}$ | RbcLS   | 3.32 $\times 10^2$ ± (4.6 $\times 10^1$) | 2.14 $\times 10^{-3}$ ± (1.8 $\times 10^{-4}$) | 6.44 $\times 10^{-4}$ ± (9.15 $\times 10^{-6}$) |
| CcmM$_{1}$   | RbcL$\Delta$18 | 5.72 $\times 10^2$ ± (1.89 $\times 10^1$) | 9.27 $\times 10^{-4}$ ± (4.44 $\times 10^{-5}$) | 1.62 $\times 10^{-5}$ ± (5.44 $\times 10^{-7}$) |

The native RubisCO spectrum showed a molecular mass of 544.74 kDa, 8.66 kDa larger than calculated from the measured component masses, suggesting that the complex binds multiple solute ions during the ESI process. A larger than calculated mass using nano-ESI under similar conditions was also previously observed for tobacco RubisCO (38). Spectra were then collected for samples where CcmM$_{1}$ at different concentrations (2, 4, and 6 $\mu$M) was preincubated with RubisCO at a fixed concentration of 0.1 $\mu$M. These showed the appearance of a series of new peaks in the spectrum (Fig. 4) with molecular masses calculated as 557.61, 570.44, 583.50, 596.25, and 609.09 kDa. These peak molecular masses differ from one another (and the native RubisCO) by the successive addition of 12.869 $\mu$M, allowing us to identify these species as RbcL$_8$S$_8$ complexes with between one and five CcmM$_{1}$ masses added (Fig. 4 and Table 3). The portion of higher molecular mass complexes increased at higher CcmM$_{1}$ concentrations; at the highest CcmM$_{1}$ concentrations, the appearance of additional high $m/z$ peaks makes spectra too complex to reliably interpret (Fig. 4D). Of note, all species had masses consistent with eight RbcL and eight RbcS subunits plus a varying number of CcmM$_{1}$ subunits, and no peak consistent with free RbcS was detected in any native experiment. These data therefore strongly argue that CcmM binds RubisCO without RbcS being released from the complex.

Discussion

CcmM’s interaction with RubisCO drives the early stages of β-carboxysome assembly, allowing the cross-linking of free RubisCO into a large, amorphous body known as the procarboxysome (30–32). The details of the stoichiometry, affinity, dynamics, and structure of the CcmM–RubisCO complex are therefore central to understanding carboxysome biogenesis. CcmM’s small subdomains are clearly homologous to RbcS, and this, along with cyanobacterial proteomics data that indicate that RbcS is depleted relative to RbcL (35), previously suggested a model where CcmM binds by replacing RbcS in some of its sites. Because cyanobacterial RubisCO assembly is complex and RbcS binding depends upon interventions by several chaperones (34), if CcmM is functioning primarily as an RbcS mimic, one might expect it to show a similar dependence upon the in vivo RubisCO assembly process. We found, in contrast, that the interaction between CcmM$_{1}$ domains and mature RubisCO occurs readily in solution, independently of any additional cellular factors, with CcmM$_{1}$ binding RubisCO with an $K_D$ of 1.16 $\mu$M. Binding RubisCO with a multi-CcmM$_{1}$ domain CcmM$_{1,2}$ or CcmM$_{1,4}$ construct results in only a small increase in affinity, suggesting that binding energy is not additive and that only one domain is able to bind RubisCO at a time. This is consistent with in vivo experiments showing that a
mixing this construct with RubisCO in solution results in rapid precipitation. These LSPR experiments also suggest a 17-min half-life for the CcmM31–RubisCO complex. A long half-life is similarly suggested by previously published fluorescence recovery after photobleaching experiments that show that individual RubisCO molecules within a growing procarboxysome are not free to move on a minute time scale (31). Conversely, Chen et al. propose (31) that shell formation may require the advancing shell edge to cleave through the mass of CcmM-cross-linked RubisCO during the last 2 h of carboxysome assembly; CcmM release from RubisCO may therefore be a necessary event in the later stages of carboxysome maturation with the observed half-life at least commensurable with the time scale of these events.

We investigated the stoichiometry and content of the complex using SEC-RALS and native MS experiments. Contrary to models where CcmM displaces RbcS from the RubisCO complex (35), these data show that binding of CcmM occurs with no release of RbcS. These SEC-RALS and native MS data also show that RubisCO can bind multiple CcmM subunits. The maximal stoichiometry is difficult to ascertain (the MS data become too complex to interpret as complex sizes increase, and the uncertainty in the SEC-RALS data is larger than the mass of a single CcmM1 protomer), but stoichiometry exceeds 5:8 CcmM:RbcL, and the SEC-RALS data suggest that the maximal binding stoichiometry is ~1:1. Characterization of whole-cell lysates of Synechococcus elongatus PCC7942 suggested that the total CcmM58 + CcmM35 concentration is roughly equal to the RbcL concentration (35). Given that these isofoms each contain three CcmM domains, more than enough CcmM domains would appear to be available in vivo to saturate available RbcL subunits.

We also determined the structure of CcmM31, showing the predicted overall resemblance to RbcS; however, detailed comparison with RbcS shows that CcmM conserves very few of the motifs RbcS uses to interact with RbcL. In particular, the N-terminal loop of RbcS (residues 3–17 in T. elongatus) is responsible for about one-third of the overall surface buried between RbcS and RbcL and is estimated to account for about a half of the interaction energy. Furthermore, single point mutations R8G and E11V (in Anabaena RbcS) prevent holoenzyme assembly of recombinant RubisCO in Escherichia coli, arguing that this N-terminal loop is likely indispensable for stabilizing RubisCO (26, 39). This motif is completely absent in CcmM where equivalent residues show negligible conservation and form part of the linker, although CcmM does retain a pair of hydrophobic pockets that could interact with a bound N-terminal loop. CcmM also lacks a very conserved RbcS motif around β2, which protrudes as a β-bulge to bury several bulky hydrophobic residues in an RbcL hydrophobic pocket. Overall, inspection of CcmM suggests that it does not closely mimic RbcS's binding surface, and it is not obvious how it might out-compete RbcS for its binding site.

One possible model to explain these findings is that CcmM binds by displacing RbcS's structural body (residues 18–end) but not the N-terminal loop (1–17) for which CcmM has no equivalent residues; this RbcS region might then help contribute to the CcmM-binding site while tethering the rest of RbcS to the complex (Fig. 5A). Although CcmM might not make especially strong interactions with RbcL, this could be offset by

CcmM binds RubisCO independently of RbcS

![Image](image-url)
CcmM binds Rubisco independently of RbcS

Figure 5. Modeling CcmM’s role in recruiting Rubisco to the carboxysome. A, a hypothetical model where CcmM displaces RbcS’s α/β domain from its binding site while the N-terminal tail of RbcS remains bound. This model would explain how CcmM might bind in RbcS’s site without causing release of RbcS but is unlikely in light of the observation that CcmMα, binding to Rubisco is not affected by RbcS truncation at residue 18. B, CcmMε occupies a site independent of RbcS. We propose that a large electronegative patch at the interface between RbcL dimer is a likely site for CcmMε (shown in yellow) interaction. C, CcmMε domains (orange) are linked together with a subset of chains having CcmMε domains. Each CcmMε chain has four CcmMε domains (in T. elongatus) with each successive domain preferring to bind a separate Rubisco (green, RbcL; yellow, RbcS) molecule. Binding of multiple CcmM molecules to different sites within a given Rubisco oligomer then cross-links Rubisco into a large aggregate, the procarboxysome.

increased conformational entropy for RbcS and/or RbcL in the CcmM-bound state and/or alternative interactions mediated elsewhere by RbcS on the Rubisco-CcmM complex surface. The observation that Rubisco has significantly higher atomic displacement parameters (B-factors) than RbcL in all β-cyanobacterial Rubisco structures makes this hypothesis at least plausible (average B values for RbcL/Rubisco are 13.5/33.5 and 28.8/54.6 Å² for PDB codes 2ybv and 3zxw, respectively). To test this model, we generated a construct that expressed RbcL and a truncated version of RbcS that retains only the N-terminal loop. This RbcL/SΔ18 protein could be expressed as a soluble holoenzyme in E. coli, indicating that the α/β domain of RbcS is not required for Rubisco to be stable. If CcmM competes with the RbcS body for its binding site, CcmMε should bind this Rubisco variant significantly more tightly. However, we observed only a slight decrease in affinity with this construct (1.62 versus 1.16 μM), suggesting that the RbcS α/β domain neither competes with nor contributes significantly to the CcmMε-binding site. Together, these data strongly indicate that CcmMε, despite its clear homology to RbcS, utilizes a binding site on Rubisco that is distinct from that occupied by RbcS.

If CcmM does not bind in the RbcS site, where does it bind? Although we have been unable to obtain a crystal of this complex, there are some constraints: in particular, CcmM likely does not significantly contact RbcS and is unlikely to impinge on the symmetry axes of the complex (or else binding sites would overlap, forcing substoichiometric binding). One strong candidate site is the highly electronegative groove that runs between RbcL dimers and widens out into a large pocket away from Rubisco’s equator. Given that the conserved surface of CcmM is markedly electropositive, binding in this area could be stabilized by multiple salt bridges; one possible configuration is depicted in Fig. 5B. Of note, the equivalent pocket in higher plant Rubisco is far less electronegative (e.g., PDB code 3axm).

The linker between successive CcmMε domains averages 32 ± 6 residues long across all CcmM homologs (using the first and last structured CcmMε residues as a benchmark). Two successive CcmMε domains would tend to average about 34 Å apart (assuming a perfectly flexible linker and a random walk configuration), reaching 100 Å if the linker is stretched taut. Because the holoenzyme forms a roughly 11-nm rounded cube, most candidate CcmMε-binding sites on Rubisco would be further apart from their symmetry copies than the linker would prefer to span. Our LSPR data indicate that the presence of additional binding domains in a CcmM molecule offers little advantage in binding Rubisco, with Kd values of 1.16, 0.171, and 0.644 μM for a single domain, two linked domains, and four linked domains, respectively. Although the longest variant would almost certainly allow two domains to bind a single Rubisco at once, the energetic cost in terms of loss of conformational entropy (possibly combined with residual structure of self-interaction by the linker) seems large enough to prevent linked CcmMε domains from binding favorably to the same Rubisco molecule. These biophysical properties likely serve to ensure that CcmMε domains prefer to bind multiple Rubisco molecules, cross-linking them into an extended aggregate, the procarboxysome (Fig. 5C).

Although the structure and possible interaction mode help clarify CcmM’s role in assembling Rubisco into a procarboxysome, there are hints that these CcmMε domains may have additional roles to play in carboxysome maturation. A growing β-cyanobacterial cell generally contains three to six carboxysomes with one more under construction (31). This requires that ~10–20% of cellular Rubisco is located in the cytosol where CO₂ levels are low and O₂ levels are high, conditions that promote photorespiration. This would seem to confer a strong selective pressure to maintain Rubisco in assembling procarboxysomes in an inactive state, using either the changes in conditions as the protein shell segregates Rubisco from the cytosol or the changes in protein interactions (such as release of an inhibitory CcmM) as an activation trigger. This model has a clear precedent in the behavior of CcmM’s N-terminal domain where the transition from reducing to oxidizing conditions upon shell formation of the carboxysome activates its carbonic
anhydride activity (20, 21, 31). The requirement for cooperative activation might explain the otherwise puzzling observation that the Rubisco–CcmM complex reorganizes from an amorphous state in immature procarboxysomes into paracrystalline layers in mature carboxysomes (33, 40). Given that CcmM tightly binds Rubisco during procarboxysome assembly and is present in near stoichiometric quantities, and the tightness of packing implied by the paracrystalline spacing (which, at 11 nm, is similar to the diameter of Rubisco), it is likely that CcmM either helps mediate the interactions that stabilize the more ordered paracrystalline state or releases Rubisco to allow this reorganization.

**Experimental procedures**

**Molecular biology**

DNA sequences were amplified from *T. elongatus* BP-1 genomic DNA (a generous gift from the Kazusa Research Institute, Japan) using pairs of primers as listed in Table 1 and iProof DNA polymerase (Bio-Rad). Successful amplicons were digested using the listed restriction endonucleases (New England Biolabs). pET28a vector (used for all constructs) was digested using the appropriate restriction endonuclease pair, treated with alkaline phosphatase, and purified using agarose gel electrophoresis. Insert and vector were then ligated using DNA ligase and transformed into chemically competent DH5α cells. Site-directed mutagenesis was performed using PCR using *Pfu*-X7 DNA polymerase (a gift from Dr. D. Christie, University of Toronto) with overlapping primers (Table 4; Thermo Fisher). All plasmids were verified by sequencing at the Advanced Analysis Center, University of Guelph. Chemically competent BL21(DE3) cells were transformed with verified plasmids for overexpression. For Rubisco construct variants, cells were also transformed with pACYC-GroEL/ES-TF plasmid (Addgene) that expresses GroEL/ES (as well as trigger factor) needed for Rubisco maturation.

**Protein expression and purification**

Transformed BL21(DE3) cells were grown overnight in 5 ml of LB at 37 °C and used to inoculate 1 liter of 2× yeast tryptone medium. Once the optical density at 600 nm reached 0.8, cultures were induced with 1 mM isopropyl β-d-1-thiogalactopyranoside and allowed to further incubate overnight. Cells were pelleted by centrifugation at 4,400 × g and resuspended in 35 ml of lysis buffer (20 mM Tris, pH 8, 150 mM NaCl). CcmM variant constructs could then be frozen at −20 °C for storage, whereas Rubisco preparations required working with fresh cells. Cell pellets were lysed using a Misonix XL-2020 sonicator, and cellular debris was pelleted by centrifugation at 48,384 × g for 30 min.

For N-terminally hexahistidine-tagged CcmM variants, the supernatant was then loaded onto a 1-ml HisTrap Fast Flow column (GE Healthcare) using an ÄKTA FPLC at 4 °C and then eluted using a gradient of 0–250 mM imidazole. Eluted protein was buffer-exchanged and concentrated using a 10-kDa molecular mass–cutoff Amicon filter unit, and the purity was assessed using SDS-PAGE. These CcmM variants were used for all experiments except crystallization. For the tagless CcmM5 variant used for crystallization, cell growth and purification proceeded as above except that pellets were resuspended in 20 mM MES, pH 5.5. Instead of metal-affinity chromatography, cell-free lysate was loaded onto a 5-ml HiTrap SP Sepharose Fast Flow column (GE Healthcare) in 20 mM MES, pH 5.5; washed; and then eluted across a 0–0.5 M NaCl gradient over 20 column volumes.

For Rubisco purifications, growth was scaled up to 4 liters; cell pellets were never frozen; and all cell handling, lysis protein purification, and protein storage steps were performed at room temperature or above (as the protein precipitates at 4 °C). Cell pellets were suspended in a buffer comprising 50 mM Tris-HCl, pH 7.6, 1 mM EDTA, 20 mM β-mercaptoethanol, 10 mM MgSO4, 10 mM NaHCO3 prior to lysis in a lukewarm waterbath. After centrifugation, the supernatant was slowly brought up to 20% (w/v) ammonium sulfate and stirred at room temperature for 1 h. Protein was centrifuged at 48,384 × g and 25 °C for 30 min, and ammonium sulfate was added to 30% (w/v) and stirred overnight at room temperature. After centrifugation as above, the supernatant was resuspended in lysis buffer and dialyzed against the same buffer overnight. After centrifugation, the supernatant was loaded onto a 5-ml HiTrap Q Sepharose Fast Flow column (GE Healthcare) in the same buffer and then eluted using a linear gradient from 0 to 0.5 M NaCl over 40 column volumes. Both Rubisco variants eluted at ∼0.25 M NaCl.

---

### Table 4

| Cloning and mutagenesis primer sequences | Restriction site |
|----------------------------------------|-----------------|
| Construct Primer sequence (5’→3’) | Restriction site |
| **Cloning primers** | | |
| CcmMll | GACGTAGATGACGTAG | Ndel |
| CcmMll-no tag | GAGCTAGATGACGTAG | Ndel |
| CcmMlt–2 | GAGCTAGATGACGTAG | Ndel |
| CcmMlt–4 | GAGCTAGATGACGTAG | Ndel |
| RbcLXS | GAGCTAGATGACGTAG | Ndel |
| RbcLXSΔ18 | GAGCTAGATGACGTAG | Ndel |
| **Mutagenesis primers** | | |
| RbcLXSΔ18 | GAGCTAGATGACGTAG | Ndel |

---

*J. Biol. Chem.* (2019) 294(8) 2593–2603
CcmM binds RubisCO independently of RbcS

SEC-RALS experiments were run on a Malvern OMNISEC instrument using PBS solution as the running buffer. The system was calibrated against a BSA standard immediately before the measurement runs. Protein samples for individual proteins were loaded at 5 mg/ml. For runs with RubisCO plus CcmM variants, each protein component was added at 1 mg/ml. Samples were separated over two tandem P3000 columns with a flow rate of 1 ml/min.

Mass spectrometry

Protein samples were buffer-exchanged into 100 mM aqueous ammonium acetate, pH 6.8, using either a 3-kDa-cutoff Amicon 0.5-ml microconcentrator (EMD, Millipore, Billerica, MA) for CcmM variants or a Bio-Rad 10DG desalting column for Rubisco. To determine accurate molecular mass values of the Rubisco and RbcL subunits, Rubisco aqueous solution was acidified to pH 3 by addition of acetic acid. Nano-ESI MS measurements were performed on a Synapt G2S quadrupole-ion mobility separation-TOF mass spectrometer (Waters) equipped with a nano-flow ESI source. Nano-ESI was performed by applying a voltage of ~1 kV to a platinum wire inserted into the nano-ESI tip, which was produced from a borosilicate glass capillary (1.0-mm outer diameter, 0.68-mm inner diameter) pulled to ~5-µm outer diameter using a P-1000 micropipette puller (Sutter Instruments, Novato, CA). The source temperature was 60 °C, and gas flow rate was 2 ml/min. The cone, trap, and transfer voltages were 30, 5, and 2 V, respectively. MassLynx software (version 4.1) was used for data acquisition and processing. Mass spectra were averaged over ~500 scans.

Localized surface plasmon resonance

Binding of CcmM variants to Rubisco variants was analyzed using an OpenSPR LSPR biosensor (Nicoya Life Sciences Inc., Kitchener, Ontario) according to the manufacturer’s instructions. All experiments utilized 20 mM Tris, pH 8.0, with 150 mM NaCl as the running buffer, and a new chip was used for each ligand. In all experiments, the CcmM construct variant serving as the ligand was immobilized to the nitrolotriacetic acid sensor chip through one injection of 100 µg/ml protein at 20 µl/min for 5 min. Once a stable baseline was achieved, analyte (RbcLS or RbcL/ΔS18) at 20 µl/min was injected for 4 min. Dissociation was then monitored for 26 min, and the signal was allowed to return to stable baseline before the next injection. Analyte injections were performed at 500, 750, and 1000 µg/ml protein, each in triplicate. Sensorgram traces of the CcmM–Rubisco interaction were recorded and analyzed using TraceDrawer software (Ridgeview Instruments).

Protein crystallization and structure determination

Crystals were grown from a solution of 1.8 M ammonium sulfate, 0.2 M sodium thiocyanate, 10 mM CoCl₂, 0.1 M MES, pH 6.5, against a 30 mg/ml protein solution in a sitting drop conformation. Spontaneously nucleating crystals typically grew as large, thin, purple plates that formed stacked columns. Diffraction quality crystals were obtained by microseeding using a seed stock prepared by vortexing initial crystals and then diluting the fragments 1,000-fold in crystallization well solution. Crystals were transferred into Paratone N oil to remove excess solution prior to flash freezing in liquid nitrogen. Data were collected at the Canadian Light Source, beamline 08ID-1. After the anomalous scattering peak was established using X-ray fluorescence, a data set was collected at a wavelength of 1.60522 Å, the cobalt anomalous scattering peak; the resolution of this data set was limited by detector geometry rather than crystal quality. A high-resolution native data set was also collected at a wavelength of 0.97949 Å. Data were processed in XDS and scaled in XSCALE (41). The peak data set was used to determine the structure by single anomalous diffraction using Autosolve in Phenix (42), resulting in a figure of merit of 0.58 for the experimental maps. This was followed by iterative manual rebuilding in Coot (43) and refinements in Phenix. Following refinement, molecular replacement of the native crystal data set was conducted using the single-wavelength anomalous diffraction–refined structure as a model followed by multiple refinement cycles in Coot. Both data sets show strong translational pseudosymmetry with a peak height of 28.4% of the origin at fractional cell coordinates 0.213, 0.500, 0.00. All structure figures were prepared in PyMOL v2.0.

Author contributions—P. R. and E. N. K. validation; P. R., T. J. F., C. W., T. M. K., E. N. K., and M. S. K. investigation; P. R., T. J. F., C. W., E. N. K., J. S. K., and M. S. K. methodology; P. R., T. J. F., E. N. K., and M. S. K. writing—original draft; E. N. K. and M. S. K. visualization; J. S. K. and M. S. K. supervision; M. S. K. conceptualization; M. S. K. formal analysis; M. S. K. funding acquisition; M. S. K. writing—review and editing.

Acknowledgments—We thank S. Castel for early cloning work, M. Pothecary (Malvern Instruments) for assistance with SEC-RALS experiments, and P. Grochulski and S. Labiuk at the Canadian Light Source for assistance in collecting diffraction data.

References

1. Andersson, I., and Backlund, A. (2008) Structure and function of Rubisco. Plant Physiol. Biochem. 46, 275–291 CrossRef Medline
2. Bathellier, C., Tcherkez, G., Lorimer, G. H., and Farquhar, G. D. (2018) Rubisco is not really so bad. Plant Cell Environ. 41, 705–716 CrossRef Medline
3. Price, G. D. (2011) Inorganic carbon transporters of the cyanobacterial CO₂ concentrating mechanism. Photosynth. Res. 109, 47–57 CrossRef Medline
4. Rae, B. D., Long, B. M., Förster, B., Nguyen, N. D., Velanis, C. N., Atkinson, N., Hee, W. Y., Mukherjee, B., Price, G. D., and McCormick, A. J. (2017) Progress and challenges of engineering a biophysical CO₂ concentrating mechanism into higher plants. J. Exp. Bot. 68, 3717–3737 CrossRef Medline
5. Rae, B. D., Long, B. M., Whitehead, L. F., Förster, B., Badger, M. R., and Price, G. D. (2013) Cyanobacterial carboxysomes: microcompartments that facilitate CO₂ fixation. J. Mol. Microbiol. Biotechnol. 23, 300–307 CrossRef Medline
6. Espie, G. S., and Kimber, M. S. (2011) Carboxysomes: cyanobacterial Rubisco comes in small packages. Photosynth. Res. 109, 7–20 CrossRef Medline
7. Kerfeld, C. A., and Melnicki, M. R. (2016) Assembly, function and evolution of cyanobacterial carboxysomes. Curr. Opin. Plant Biol. 31, 66–75 CrossRef Medline
8. Kimber, M. S. (2014) Carboxysomal carbonic anhydrases. Subcell. Biochem. 75, 89–103 CrossRef Medline
25. Kinney, J. N., Salmeen, A., Cai, F., and Kerfeld, C. A. (2012) Elucidating the
24. Fukuzawa, H., Suzuki, E., Komukai, Y., and Miyachi, S. (1992) A gene
21. de Araujo, C., Arefeen, D., Tadesse, Y., Long, B. M., Price, G. D., Rowlett,
19. Mahinthichaichan, P., Morris, D. M., Wang, Y., Jensen, G. J., and Tajkhor-
16. Keeling, T. J., Samborska, B., Demers, R. W., and Kimber, M. S. (2014)
15. Tanaka, S., Kerfeld, C. A., Sawaya, M. R., Cai, F., Heinhorst, S., Cannon,
14. Cai, F., Sutter, M., Cameron, J. C., Stanley, D. N., Kinney, J. N., and Kerfeld,
12. Rae, B. D., Long, B. M., Badger, M. R., and Price, G. D. (2012) Structural
sp. PCC 7120.
Nostoc
CcmL.
Photosynth. Res.
PCC 7942: roles for CcmK2, K3-K4, CcmO, and CcmL. PLoS One 7, e43871 CrossRef Medline
11. Tanaka, S., Sawaya, M. R., Phillips, M., and Yeates, T. O. (2009) Insights
from multiple structures of the shell proteins from the β-carboxysome.
Protein Sci. 18, 108–120 CrossRef Medline
10. Samborska, B., and Kimber, M. S. (2012) A dodecameric CcmK2 structure
suggests β-carboxysomal shell facets have a double-layered organization.
Structure 20, 1353–1362 CrossRef Medline
9. Kerfeld, C. A., Sawaya, M. R., Tanaka, S., Nguyen, C. V., Phillips, M., Beeby, M., and Yeates, T. O. (2005) Protein structures forming the shell of primitive bacterial organelles. Science 309, 936–938 CrossRef Medline
8. Kostov, R. V., Small, C. L., and McFadden, B. A. (1997) Mutations in a homologous to chloroplast carbonic anhydrase (icfA) is essential to photosynthesis. J. Mol. Biol. 274, 303–312 CrossRef Medline
7. Holm, L., and Rosenström, P. (2010) Dali server: conservation mapping in structural genomics. Acta Crystallogr. D Biol. Crystallogr. 66, 33–45 CrossRef Medline
6. Chethan, K., Bhangale, T., and Srinivasan, K. (2018) CcmM binds RubisCO independently of RbcS. J. Biol. Chem. 294(8), 2593–2603 PCCP, 10, 15–10 CrossRef Medline
5. Kostov, R. V., Small, C. L., and McFadden, B. A. (1997) Mutations in a homologous to chloroplast carbonic anhydrase (icfA) is essential to photosynthesis. J. Mol. Biol. 274, 303–312 CrossRef Medline
4. Cameron, J. C., Wilson, S. C., Bernstein, S. L., and Kerfeld, C. A. (2013) Biogenesis of a bacterial organelle: the carboxysome assembly pathway. Cell 155, 1131–1140 CrossRef Medline
3. Chen, A. H., Robinson-Mosher, A., Savage, D. F., Silver, P. A., and Polka, J. K. (2013) The bacterial carbon-fixing organelle is formed by shell envelopment of preassembled cargo. PLoS One 8, e76127 CrossRef Medline
2. Niederhuber, M. J., Lambert, T. I., Yapp, C. Silver, P. A., and Polka, J. K. (2017) Superresolution microscopy of the β-carboxysome reveals a homogeneous matrix. Mol. Biol. Cell. 28, 2734–2745 CrossRef Medline
1. Caneko, Y., Danee, R., Nagayama, K., and Nakamoto, H. (2006) Intact carboxysomes in a cyanobacterial cell visualized by Hilbert differential contrast transmission electron microscopy. J. Bacteriol. 188, 805–808 CrossRef Medline

CcmM binds RubisCO independently of RbcS