Title: Cyanobacterial α-carboxysome carbonic anhydrase is allosterically regulated by the Rubisco substrate RuBP

Authors: Sacha B. Pulsford[1,2], Megan A. Outram[3], Britta Förster[4], Timothy Rhodes[4], Simon J. Williams[4], Murray R. Badger[4], G. Dean Price[4], Colin J. Jackson[1,2], Benedict M. Long[1,5,*]

Affiliations:

[1]ARC Centre of Excellence in Synthetic Biology.

[2]Research School of Chemistry, Australian National University; Canberra, ACT, 2601 Australia

[3]Commonwealth Scientific and Industrial Research Organisation, Agriculture and Food, Canberra, ACT, 2601 Australia

[4]Research School of Biology, The Australian National University, Canberra, ACT, 2601 Australia

[5]School of Environmental and Life Sciences, University of Newcastle, Callaghan, NSW, 2308 Australia

*Corresponding author: ben.long@newcastle.edu.au

ABSTRACT

Cyanobacterial CO_2 concentrating mechanisms (CCMs) sequester a globally significant proportion of carbon into the biosphere. Proteinaceous microcompartments, called carboxysomes, play a critical role in CCM function, housing two enzymes to enhance CO_2 fixation: carbonic anhydrase (CA) and Rubisco. Despite its importance, our current understanding of the carboxysomal CAs found in α-cyanobacteria, CsoSCA, remains limited, particularly regarding the regulation of its activity. Here, we present the first structural and biochemical study of CsoSCA from the cyanobacterium *Cyanobium PCC7001*. Our results show that the *Cyanobium* CsoSCA is allosterically activated by the Rubisco substrate ribulose-1,5-bisphosphate (RuBP), and forms a hexameric trimer of dimers. Comprehensive phylogenetic and mutational analyses are consistent with this regulation appearing exclusively in cyanobacterial α-carboxysome CAs. These findings clarify the biologically relevant oligomeric state of α-carboxysomal CAs and advance our understanding of the regulation of photosynthesis in this globally dominant lineage.

One-Sentence Summary: The carboxysomal carbonic anhydrase, CsoSCA, is allosterically activated by the Rubisco substrate RuBP, revealing a novel mechanism controlling key enzyme activity in cyanobacterial α-carboxysomes**.**

INTRODUCTION

A myriad of CO_2 concentrating mechanisms (CCMs) have independently evolved to promote the rapid and efficient reduction of atmospheric CO_2 into organic compounds. CCMs work to increase the local concentration of CO_2 near ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), the primary carboxylase of the Calvin–Benson–Bassham (CBB) cycle, thereby increasing its substrate turnover and competitively inhibiting off-target oxygenation reactions[1–4]. These systems are an essential component of the global carbon cycle, catalysing about half of global photosynthesis[2,5]. The bacterial CCM, found in all cyanobacteria and some autotrophic bacteria, consists of two key elements: first, energy-coupled inorganic carbon (C_i; primarily HCO_3^- and CO_2) transporters actively establish a concentrated pool of HCO_3^- within a cytosol lacking free carbonic anhydrases (CAs); this HCO_3^- then diffuses into proteinaceous microcompartments called carboxysomes that house CA and Rubisco[2,3]. Here, the CA converts HCO_3^- to CO_2 to elevate luminal CO_2, promoting Rubisco-catalysed CO_2 reduction[6]. Bacterial CCMs have arisen in two distinct lineages: β-carboxysomes are found exclusively in β-cyanobacteria, containing Form1B Rubisco with component genes encoded by the *ccm* operon and satellite loci[7], whereas α-carboxysomes are found in photoautotrophic α-cyanobacteria and several bacterial chemoautotrophs, and are distinguished by the presence of FormIA Rubisco and the clustering of carboxysome-associated genes into a discrete *cso* carboxysome operon[8].

α-Cyanobacteria are a globally dominant photoautotrophic lineage across marine and freshwater systems, encompassing two of the most abundant photosynthetic taxa on Earth (*Synechococcus* and *Prochlorococcus*)[5,9]. Although their global mass is a fraction of that of plant systems, they are estimated to contribute around 25% of global primary production, i.e. carbon fixation[5,8]. Regulation of carbon fixation is essential for effective energy production. Indeed, the CCM is a striking example of how cells may induce changes in physiological state in response to environmental conditions. This adaptive capacity is a critical feature of these processes, involving regulation at the transcriptional and protein level, allowing the bacterial CCM to competitively support life in a range of ecological contexts[10,11]. For example, the CCM-related Ci transporter is regulated by gene expression and allosteric effectors[12–14]. Likewise, Rubisco content is transcriptionally regulated and its activity modulated by activases[15–19]. Finally, the carboxysome composition and morphology itself is responsive to environmental cues[20–24]. However, little is

known of how, or whether, carbonic anhydrase, the other enzymatic component of the carboxysome, is regulated[25].

The fundamental role of CAs in photosynthesis is well established[6]. This versatile protein superfamily catalyses the reversible hydration of CO_2 [$CO_2 + H_2O \rightleftharpoons HCO_3^- + H^+$], comprising eight reportedly evolutionarily distinct classes (α, β, γ, δ, θ, η, ζ and ι), distributed across the tree of life in a kingdom-nonspecific manner[26]. In many cases, the enzyme directly supplies Rubisco with CO_2, promoting its efficient reduction by ensuring reaction rate optimization through a constant, high concentration of the enzyme-substrate complex[6,27]. Controlling Rubisco activity through buffering a bicarbonate pool in this way optimizes carbon fixation, coordinating CO_2 assimilation rates with the generation of NADPH/ATP produced in light reactions[27]. Correspondingly, microbial and biochemical studies have established an absolute requirement for CA activity within the carboxysome[28,29]. The α-carboxysome contains a highly divergent β-CA known as CsoSCA, characterisation of which has occurred exclusively through the isoform from the chemoautotroph *H. neapolitanus*[30,31]. Indeed, structural and compositional studies have revealed sequence variation between cyanobacterial and proteobacterial carboxysome components, and distinct carboxysome organisation between these taxa[32–37]. Given the differences in underlying metabolism between these photo- and chemoautotrophs, this has restricted our understanding of the CA-Rubisco feedback in α-cyanobacterial carboxysomes.

Here, we present a detailed biochemical, structural, and evolutionary analysis of a CsoSCA from a photoautotrophic cyanobacterium, *Cyanobium* PCC7001 (*Cyanobium*), revealing previously unknown aspects of this isoform's activity and molecular structure that form the basis for carboxysome regulation and organisation. We found that, unlike the CA from the chemoautotrophic bacterium *H. neapolitanus* (*Hn*CsoSCA), the *Cyanobium* isoform (*Cy*CsoSCA) is regulated by the Rubisco substrate, ribulose-1,5-bisphosphate (RuBP) for activity, which constitutes a feedback loop in the metabolic pathway to optimize the use of carbon. Detailed evolutionary analysis extends this, revealing the sequence motifs for this regulation are not found in chemoautotrophic bacteria, and expanding our understanding of α-cyanobacterial photosynthesis and CCM diversification more broadly.

RESULTS

Cyanobium CsoSCA requires RuBP for activity

Despite homology to the constitutively active *Hn*CsoSCA[30], in our hands *Cy*CsoSCA did not show detectable HCO_3^- dehydration/CO_2 hydration activity under standard assay conditions[20]. This surprising result indicated a potential additional requirement for *Cy*CsoSCA function. Given the previous observation of *Cyanobium* carboxysome function *in vitro*[20], and the established reliance on CA activity[6,28,38], we assessed *Cy*CsoSCA function under Rubisco assay conditions, where RuBP and Mg^{2+} are the key additional components. The addition of RuBP resulted in the concentration-dependent activation of *Cy*CsoSCA, with a K_M for RuBP of 18 µM, in a similar range as the *Cyanobium* Rubisco K_M for RuBP (36 µM)[38]. Comparatively, *Hn*CsoSCA activity was unaffected by RuBP. Above 100µM RuBP *Cy*CsoSCA activity rates match those recorded for *Hn*CsoSCA (Fig. 1A). Notably, the *Cy*CsoSCA RuBP response curve is best described by a sigmoidal Hill function ($R^2 = 0.99$), typically indicative of an allosteric activation mechanism.

Cyanobium CsoSCA Structure reveals RuBP binding site and novel oligomeric state

To understand the structural basis for RuBP activation, CsoSCA was co-crystallised with RuBP. Diffracting crystals were obtained almost exclusively in saturating levels of RuBP with the final CsoSCA crystal structure solved through molecular replacement to a resolution of 2.3 Å (Table S1). The resulting structure showed a homohexameric trimer of dimers consistently arranged in the asymmetric unit with P212121 symmetry (Fig. 1B). Size-Exclusion Chromatography (SEC) corroborates both *Cy*CsoSCA and *Hn*CsoSCA are primarily hexameric in solution (Fig. S10), contrasting with previous observations[30]. While the *Cy*CsoSCA dimer interface is highly reminiscent of *Hn*CsoSCA[30], additional contacts at the N-terminal domain (NTD) of each monomer mediate further quaternary assembly, forming two apices of the final hexamer (Fig. 1B). A metal ion is evident at each apex, coordinated by a $His_3(H_2O)_3$ octahedral coordination sphere, comprising His155 donated by a distinct monomer (Fig. 1B). This residue sits within a helical bundle in the NTD denoted here as the 'hook motif'. The electron density and coordination geometry are consistent with a zinc ion. Density corresponding to a HCO_3^- ion was observed at one of the trimer apices. While the pH of crystallisation conditions favours bicarbonate, this species was not present at saturating levels in the crystallisation conditions, which could explain its absence at the opposing apex.

Density consistent with RuBP was observed in all monomers within a positively charged pocket near the dimer interface that extends into the protein core (Fig. 1C, S2 and S8; Table S2 and S3). While variations in omit map density at these positions were observed, given the consistency of this density across each chain, the dependence on RuBP for crystallization, and the observation RuBP was an allosteric activator of CA (Fig. 1), we were confident in the modelling of RuBP at this site. Two sulfate ions, likely from the crystallisation solvent, could be modelled with high confidence at the entrance of this site in all monomers. While slight variations in RuBP ligand conformation are evident in each chain, contacts at Arg266, Lys469 and Arg560 are consistently observed (Fig. S2 and S3). Most notably, RuBP curls around Lys469, mediating multiple H-bonds with the ligand. To confirm that this region is responsible for RuBP binding, we mutated Lys469 to an Asp, the amino acid at the corresponding position in the constitutively active *Hn*CsoSCA isoform. This results in a biphasic activity profile with detectable CA activity evident in the absence of RuBP and a minor increase in activity upon addition of the ligand (Fig. 1D). This directly implicates K469 in the RuBP-mediated activation mechanism and further supports this region as the RuBP binding pocket.

RuBP regulation of Cyanobium CsoSCA is allosteric

Given the sigmoidal *Cy*CsoSCA activation curve and RuBP binding site distinct from the active site (Fig.1), we hypothesised RuBP acts as an allosteric activator. The β-CA family is the only CA family known to exhibit allostery to date[40]. Alignments between *Cy*CsoSCA and a structure of a previously characterised Type II β-CA bound to the allosteric bicarbonate ion[40] show RuBP engages distinct residues and sits further from the active site, indicating a distinct regulatory mechanism (Fig. S4). The RuBP site overlays with the region of the CsoSCA C-terminal domain (CTD) that has lost the second symmetric catalytic zinc site seen in canonical β-CAs, likely following gene duplication and divergence of the catalytic domain[30]. A structural analysis of the *Cy*- and *Hn*CsoSCAs was conducted to identify potential allosteric networks within the cyanobacterial variant. *Cy*CsoSCA monomers align well with the canonical *Hn*CsoSCA (37.2% sequence identity, Cα RMSD of 1.5 Å, Fig. 2A) and the three domains (NTD, Catalytic domain and CTD)[30] are evident (Fig. 2A). The $Cys_2His(H_2O)$ tetrahedral coordination of the catalytic zinc ion typical of β-CAs is maintained and the overarching active site is highly homologous between each CsoSCA isoform. The Asp-Arg dyad between active site residues Asp246 and Arg248 precludes the inactive $Cys_2HisAsp$ coordination sphere across all *Cy*CsoSCA monomers,

reinforcing the classification of CsoSCA as Type I[30]. This type of β-CA has not previously been associated with allostery. Manual inspection of H-bonds within the protein identified a network linking the catalytic Asp246 backbone and the Arg266 sidechain that in turn binds RuBP, mediated by a water molecule and Leu249 backbone groups (Fig. 2B). In the corresponding region in the constitutively active *Hn*CsoSCA, Lys179 (Ala250 in *Cy*CsoSCA) occupies this space, coordinating a more extensive interaction network reinforced by multiple water molecules. Notably, Arg196 (Arg266 in *Cy*CsoSCA) binds Asp409 (Lys469 in *Cy*CsoSCA) identified above as a key determinant in RuBP-dependent activity (Fig. 2C).

Molecular dynamic simulations of the *Cy*CsoSCA structure (300 ns replicates) were conducted in the presence (holoprotein) and absence (apoprotein) of RuBP to assess for conformational changes upon ligand binding to evaluate the allosteric effect of RuBP. Analyses of replicate simulations are consistent with *Cy*CsoSCA accessing different conformational landscapes when RuBP is present or absent. A principal component analysis of the trajectories highlights differences between conformations sampled in apo- and holoprotein states, with holoprotein replicates converging on a distinct cluster as the simulations equilibrate (Fig. 2D, S6). On average, residues in the apoprotein trajectories across the structure had root mean square fluctuations (RMSF), indicating they are more mobile (Fig. 2E). While it is difficult to ascribe a molecular mechanism with confidence, these results support a model in which RuBP stabilises *Cy*CsoSCA by establishing an internal H-bond network, promoting access to the active conformation.

Sequence patterns indicate allosteric CsoSCA is limited to cyanobacteria

We sought to investigate the prevalence of RuBP allostery within the broader CsoSCA protein family by mapping the sequence diversity of the family to *Cy*CsoSCA functional variation. To examine CsoSCA divergence, a maximum likelihood phylogeny was inferred from a curated sequence database of the CsoSCA PFAM (PF08936) (Fig. 3A, S7). Cyanobacteria form a clear, tight cluster distinct from other bacterial species, supported by a high bootstrap value, as seen in related studies[33,41,42]. We hypothesised RuBP regulation may be specific to photoautotrophs, evolving as the *cso* operon adapted to these organisms' light-dependent metabolic requirements relative to ancestral chemoautotrophic α-carboxysomes[42].

To test this, concurrent approaches based on rational design and directed evolution were used to discern key residues involved in RuBP regulation. The conservation of these positions was then

assessed across the CsoSCA protein family (Fig. 3). Candidate residues for targeted mutagenesis were chosen through successive steps of analysing the sequence and structure of *Cy*CsoSCA and *Hn*CsoSCA to locate residues with distinct biophysical properties near the RuBP pocket. Final mutations were made at sites that differed between the two characterised isoforms and had varying levels of conserved difference between cyanobacterial taxa and other bacterial CsoSCA isoforms more broadly (K469D, H472Q, I466D, H436N). Manual sequence inspection revealed a loop region in *Hn*CsoSCA (position in *Hn*CsoSCA) that contained an insertion conserved across cyanobacterial species (position in *Cy*CsoSCA) but absent or non-conserved in other proteins. Mutants were created in the *Cy*CsoSCA background with either a deletion of this loop (Loop deletion) or with the corresponding *Hn*CsoSCA loop sequence substituted at this site (Loop insertion). Alongside this approach, *Cy*CsoSCA was also randomly mutagenized, facilitating a broader exploration of the sequence space involved in this activation mechanism.

Mutants were screened using an in-house CA knock-out *E. coli* strain[43] for variants with CA activity independent of RuBP. Activity assays of these mutants revealed that, in addition to K469D, a H472Q mutation (targeted approach) and T477A (random approach) also resulted in a biphasic activity profile with CA function independent of RuBP (Fig. 3D). Other mutations resulted in either reduced or undetectable CA activity, making their effects on RuBP dependence specifically, difficult to infer. Sequence-based analyses show residues with apparent involvement in RuBP-dependence are well conserved in cyanobacteria, but absent or non-conserved in other taxa (Fig. 3C). The conservation of sites underpinning RuBP-dependence is consistent with this regulation existing primarily, if not exclusively, in photoautotrophic CsoSCA variants.

The unique N-terminal oligomerisation domain is exclusive to carboxysomal CAs

Further bioinformatic analysis of the CsoSCA protein family revealed an orphan cluster within this β-CA clade (Fig. 4). Manual inspection of the gene neighbourhoods of these sequences demonstrates they are not associated with the *cso* operon, appearing instead within other metabolic gene clusters, often associated with NADH or [Fe-Ni] hydrogenases or permeases (Fig. S12). Structural modelling of 'non-*cso*' sequences and subsequent structure-based searches using Foldseek[44] and DALI[45] shows these 'non-*cso*' sequences align preferentially to the published *Hn*CsoSCA structure with high confidence relative to other β-CAs (Table S4). These sequences appear to have lost the typical β-CA two-fold symmetry, containing only one predicted Zinc

binding site per pseudo dimer, a defining feature of canonical CsoSCA sequences (Fig. 4). However, all 'non-*cso*' sequences are notably shorter than carboxysome associated variants. This is underlined by a consistent insertion in *cso*-associated sequences within the NTD which encodes a hook-like bundle of α-helices dubbed the 'hook' motif (Fig. 4) shown here to facilitate structural Zinc binding and oligomerization (Fig. 3). This is consistent with NTD presence, and thus hexamer formation being a more recent adaptation unique to carboxysome-encapsulated variants of this family.

Carboxysome functional modelling indicates an adaptive advantage for RuBP regulated CsoSCA

The results presented above support the hypothesis CsoSCA RuBP-dependence is a fixed trait unique to cyanobacterial α-carboxysome systems. To determine whether this feature emerged as an adaptive or neutral change in CsoSCA variants, we aimed to discern a functional benefit for RuBP regulation within the cyanobacterial system. An *in vivo* assessment of CsoSCA regulation is currently intractable, with no effective genetic transformation techniques reported to date. Instead, we modified our carboxysome steady state diffusion model[38], incorporating kinetic data presented in Figure 1A to compare the activity of a *Cyanobium* carboxysome with and without an RuBP-dependent CA (Fig. 5A). While second to physiological data, this approach permitted insights into the effects this regulation may have on core enzyme activity and metabolite flux of the *Cyanobium* α-carboxysome. No substantial changes in Rubisco carboxylation or oxygenation rates were observed between the standard model and one incorporating an RuBP-dependent CA (Fig. 5B, S9). However, we noted that the modelling of the unmodified *Cyanobium* carboxysome here had a slightly alkaline carboxysome pH compared with the observation of an acidic carboxysome in our previous modelling[38]. This is due to the use of *Cyanobium* Rubisco kinetics, the concentration of HCO_3^- tested here (20 mM), and a modification of the Rubisco active site concentrations compared with previous modelling (a reduction from 10 mM to 5.7 mM to fit recent estimates[32]). However, applying RuBP-dependence on CA activity in the model provides a shift to a more acidic condition that can support Rubisco carboxylation[38] and potentially ensures Rubisco is maintained within its optimum pH range for activity[46]. This implicates a regulated CA in maintaining an acidic carboxysome lumen under low RuBP conditions in the *Cyanobium* system, likely experienced as light levels fluctuate throughout the diurnal cycle[20], thereby contributing to efficient CCM function[47].

DISCUSSION

The α-carboxysomal CCM enables efficient C_i fixation across a diverse range of microorganisms, comprising a major component of the global biosphere[5,48]. Given the essential nature of CAs in this system[6], characterising the CsoSCA variant present in photosynthetic organisms, and understanding its evolutionary trajectory provides important insights into the emergence and function of bacterial CCMs. Data presented here are consistent with RuBP allosterically activating *Cy*CsoSCA and the likely confinement of this property to α-cyanobacteria. A hexameric 'trimer of dimers' quaternary state is also described, coordinated by NTD contacts with structural zinc ions that appear only in carboxysome-associated members of the CsoSCA protein family.

A distinct paradigm for carbonic anhydrase allosteric regulation

To our knowledge, this is the first case of allosteric activation reported for a carbonic anhydrase. The β-CA family is the only CA family known to exhibit allostery, with HCO_3^- acting as both a substrate and inhibitor in Type II members. In that case, HCO_3^- binding disrupts the 'gate-keeper' Asp-Arg dyad, leading to Asp-Zn bond that displaces the catalytic H_2O, leading to inhibition[40,49–51]. In addition to *activating* the enzyme, the RuBP binding pocket presented here appears distinct from these previously characterised sites, engaging different residues and sitting further from relative active sites (Fig. S4). Indeed, the RuBP site sits near the defunct active site of the CTD within the CsoSCA pseudo-dimer, previously identified as a highly divergent catalytic domain that has lost key zinc-binding sites and catalytic loops[30]. The duplication and divergence of domains, particularly at protein termini, are a common motif in protein evolution[52]. While seemingly acting as a regulatory domain in the cyanobacterial CsoSCA, it is unclear whether the CTD takes on alternative, more cryptic regulatory roles in RuBP-independent CsoSCA isoforms, or indeed in the smaller uncharacterised 'non-*cso*' isoforms.

We propose an allosteric activation mechanism in which RuBP binding establishes an internal H-bond network near the core of *Cy*CsoSCA that has a broadly stabilising effect on the protein, this in turn promotes access to the active conformation. RuBP binding engages Arg266 in H-bonding, establishing a H-bonding network that links to active site loops (Fig. 3). The equivalent network in HnCsoSCA is more extensive involving more residues, specifically negatively charged residues that are absent in the CyCsoSCA isoform. Specifically, in *Hn*CsoSCA, the Arg266 equivalent (Arg196) is stabilised by analogous interactions with Asp409 (Lys469 in *Cy*CsoSCA), thereby

negating the ligand-binding requirement for activation. This proposed mechanism is also consistent with the observed biphasic activity profile of mutants many of which introduce a relatively negative charge to this site, most notably K469D (Fig. 3). However, it remains difficult to rationalise the relatively large increase in RuBP-independent activity observed in the *Cy*CsoSCA^{T477A} mutant. While Thr residues are typically known to stabilise β-sheets, given this site is buried it may be that this exchange reduces secondary structure strain and enhances protein core packing while with more minimal disruption to the overarching protein fold that other mutants[53]. Consequently, T477A would result in increased apoprotein stability *and* a WT-like RuBP pocket, leading to high levels of ligand-independent activity and a notable increase with RuBP. Although the involvement of the Asp-Arg dyad destabilised by HCO_3^- does seem reasonable in the context of *Cy*CsoSCA given it's established role as an allosteric switch[40,49], we could not establish this as the clear mechanism for allosteric signal propagation.

An adaptive advantage for an RuBP-regulated CA in photoautotrophs

The conservation of residues central to RuBP-dependent activity corresponds with a clear divergence of the cyanobacterial clade in the CsoSCA phylogeny (Fig. 3). This suggests allosteric RuBP activation is either a neutral or adaptive change uniquely fixed in α-cyanobacteria relative to other α-carboxysomal taxa. RuBP allostery appears infrequently in the literature, primarily characterised in Rubisco adjacent proteins such as the AAA+ red-type activase Cbbx[54]. In these cases, it is proposed to synchronise the activity of such proteins with Rubisco, hinting at an intricate regulatory cycle to ensure efficient Rubisco function. Indeed, modelling and experimental data have established a similar functional link between CA and Rubisco, demonstrating Rubisco function in carboxysomes is intrinsically dependent upon CA activity[6,28,38]. This type of post-translational regulation would enable more rapid responses to transient metabolic signals, directly synchronising Calvin cycle fluxes with Rubisco-mediated carbon fixation.

We propose RuBP-dependent allostery may be linked to large fluctuation in cellular RuBP observed in photosynthetic α-cyanobacteria, but absent in proteobacterial systems such as *H. neapolitanus* that are likely under more constant environmental substrate supply[20,55]. Using a carboxysome reaction-diffusion model, we previously identified that carboxysomes may require molecular mechanisms that modulate internal pH as RuBP concentrations vary, to the detriment of Rubisco function[38]. Modification of this model to allow for allosteric activation of CA in

Cyanobium-like carboxysomes indicates such regulation ameliorates this effect, resulting in a modulation of carboxysome pH without any change in Rubisco carboxylation (Fig. 5, S9). Thus, the emergence of RuBP regulation in these systems may have been prompted by a requirement for more fine-tuned control over carboxysomal H^+ concentrations in photoautotrophic systems that are subject to more drastic cellular RuBP fluctuations across the diurnal cycle. That the protein can be switched from a constitutive to autoinhibitory isoform in one residue change highlights the evolutionary pliability of this protein, suggesting the evolution of allostery from a constitutively active isoform may have arisen from a relatively small sequence level change in ancestral photoautotrophic hosts in the presence of new fitness pressures. This raises the question as to why β-carboxysomes, present exclusively in photosynthetic cyanobacteria, does not appear to have the same requirement?[56,57] Our modelling has previously identified that the larger size of β-carboxysomes, and thereby surface area:volume ratio, relative to α-carboxysomes leads to diffusional differences between these structures[38]. In this way, β-carboxysomes may not be as vulnerable to the internal species fluctuations that underpin the proposed advantage of an RuBP-regulated CA. Given this, we conclude the emergence of an RuBP-dependent CA comprises a key molecular step in the adaptation of the α-carboxysome within cyanobacterial lineages.

The α-carboxysomal CA is hexameric in solution

The oligomeric state of enzyme cargo is an important detail of bacterial microcompartment systems, providing insights into cargo organisation and interaction networks within the shell. Here, we show that both *Cy*CsoSCA and *Hn*CsoSCA are both hexameric in solution, not dimeric as previously described[30], and that this quaternary structure is mediated by the previously unannotated NTD (Fig. 1). This hexamer likely eluded detection in the reported *Hn*CsoSCA structure due to an inadvertent mutation causing an artificial second zinc site within the NTD of each monomer that would likely inhibit the structural zinc interactions observed here[30] (Fig. S11). Given this mutant still exhibited high levels of CA activity, it is unlikely the two additional zinc sites detailed here are catalytically relevant, instead acting primarily to stabilise the quaternary structure. Moreover, the analysis of the CsoSCA protein family presented here suggests the NTD, and thus the ability to form this quaternary structure, is specific to carboxysome-associated CsoSCAs (Fig. 4). CsoSCA is now known to interact with Form IA Rubisco through interactions between the CsoSCA N-terminal disordered tail in a manner reminiscent of the α-carboxysome structural protein CsoS2[33,57]. Though the conservation of this binding motif is limited, often each

N-terminal disordered tail appears to be limited to a single motif. Comparatively, CsoS2 contains multiple Rubisco binding motifs, enhancing the multivalency and thus strength of the interaction. Considering this, a hexameric assembly would enhance the local concentration of Rubisco interaction motifs and thus multivalency of the system to promote CsoSCA-Rubisco binding. In this way, the emergence of the NTD may constitute a molecular marker for the early association of an ancestral CsoSCA-Rubisco complex, hypothesised as a likely carboxysome evolutionary precursor[38,58].

CONCLUSION

As we begin to resolve the evolutionary trajectories of bacterial CCMs, a detailed understanding of structural variation of core system components and how this relates to function will be essential for resolving plausible evolutionary routes. We have identified a distinct divergence between α-cyanobacteria and other α-carboxysome taxa, presenting a paradigm for CO_2 fixation in photoautotrophic α-cyanobacteria that hinges on the regulation of CsoSCA by RuBP. These results highlight the intrinsic role of CAs in photosynthesis, comprising an evolutionarily pliable enzyme at the bottleneck of key reactions in the carbon metabolism of many diverse organisms. This aspect of carboxysome function and evolution must be considered in future biotechnological applications that seek to adapt such systems, particularly efforts to transplant the bacterial CCM into photoautotrophic crop species.

Acknowledgements

The authors acknowledge use of the Australian Synchrotron MX facility and thank the staff for their support. This research was undertaken in part using the MX2 beamline at the Australian Synchrotron, part of ANSTO, and made use of the Australian Cancer Research Foundation (ACRF) detector.

Funding: Funding for some salary components (BF, SBP, BML), materials, DNA constructs, and sequencing work was supported by a grant from the Australian Research Council Centre of Excellence for translational Photosynthesis (CE140100015) to GDP. CJJ and SBP also acknowledge support from the Australian Research Council Centre of Excellence in Synthetic Biology (project number CE200100029). SBP acknowledges stipend funding from the Westpac Scholars Trust.

Author contributions: SBP and BML conceptualisation, experimental design, data collection and analysis, writing, editing; MAO Data collection and analysis; BF Data collection; SJW Data collection and analysis; CJJ Data collection and analysis, writing and editing, funding; MRB conceptualisation, experimental design; GDP conceptualisation, experimental design, editing, funding; TR experimental design, data collection.

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data are available in the main text or the supplementary materials. The *Cyanobium* CsoSCA structure is available at the PDB ID: 8THM.

References

1. Raven, J. A., Cockell, C. S. & De La Rocha, C. L. The evolution of inorganic carbon concentrating mechanisms in photosynthesis. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **363**, 2641–2650 (2008).
2. Badger, M. R. & Price, G. D. The CO2concentrating mechanism in cyanobactiria and microalgae. *Physiol. Plant.* **84**, 606–615 (1992).
3. Badger, M. R., Hanson, D. & Price, G. D. Evolution and diversity of CO2 concentrating mechanisms in cyanobacteria. *Funct. Plant Biol.* **29**, 161–173 (2002).
4. Tcherkez, G. G. B., Farquhar, G. D. & Andrews, T. J. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc. Natl. Acad. Sci.* **103**, 7246–7251 (2006).
5. Cabello-Yeves, P. J. *et al.* α-cyanobacteria possessing form IA RuBisCO globally dominate aquatic habitats. *ISME J.* **16**, 2421–2432 (2022).
6. Badger, M. The roles of carbonic anhydrases in photosynthetic CO2 concentrating mechanisms. *Photosynth. Res.* **77**, 83 (2003).
7. Sommer, M., Cai, F., Melnicki, M. & Kerfeld, C. A. β-Carboxysome bioinformatics: identification and evolution of new bacterial microcompartment protein gene classes and core locus constraints. *J. Exp. Bot.* **68**, 3841–3855 (2017).
8. Rae, B. D., Long, B. M., Badger, M. R. & Price, G. D. Functions, Compositions, and Evolution of the Two Types of Carboxysomes: Polyhedral Microcompartments That Facilitate CO2 Fixation in Cyanobacteria and Some Proteobacteria. *Microbiol. Mol. Biol. Rev. MMBR* **77**, 357–379 (2013).
9. Flombaum, P. *et al.* Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. *Proc. Natl. Acad. Sci.* **110**, 9824–9829 (2013).
10. Badger, M. R., Price, G. D., Long, B. M. & Woodger, F. J. The environmental plasticity and ecological genomics of the cyanobacterial CO2 concentrating mechanism. *J. Exp. Bot.* **57**, 249–265 (2006).

11. Raven, J. A. & Beardall, J. CO2 concentrating mechanisms and environmental change. *Aquat. Bot.* **118**, 24–37 (2014).
12. Fang, S. *et al.* Molecular mechanism underlying transport and allosteric inhibition of bicarbonate transporter SbtA. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2101632118 (2021).
13. Kaczmarski, J. A. *et al.* Structural Basis for the Allosteric Regulation of the SbtA Bicarbonate Transporter by the PII-like Protein, SbtB, from Cyanobium sp. PCC7001. *Biochemistry* **58**, 5030–5039 (2019).
14. Woodger, F. J., Badger, M. R. & Price, G. D. Inorganic Carbon Limitation Induces Transcripts Encoding Components of the CO2-Concentrating Mechanism in Synechococcus sp. PCC7942 through a Redox-Independent Pathway. *Plant Physiol.* **133**, 2069–2080 (2003).
15. Harano, K. *et al.* Regulation of the expression of ribulose-1,5-bisphosphate carboxylase/oxygenase (EC 4.1.1.39) in a cyanobacterium, Synechococcus PCC7942. *Photosynth. Res.* **78**, 59–65 (2003).
16. Huang, F. *et al.* Roles of RbcX in Carboxysome Biosynthesis in the Cyanobacterium *Synechococcus elongatus* PCC7942. *Plant Physiol.* **179**, 184–194 (2019).
17. Chen, T. *et al.* Incorporation of Functional Rubisco Activases into Engineered Carboxysomes to Enhance Carbon Fixation. *ACS Synth. Biol.* **11**, 154–161 (2022).
18. Tsai, Y.-C. C., Lapina, M. C., Bhushan, S. & Mueller-Cajar, O. Identification and characterization of multiple rubisco activases in chemoautotrophic bacteria. *Nat. Commun.* **6**, 8883 (2015).
19. Sutter, M. *et al.* Structural Characterization of a Newly Identified Component of α-Carboxysomes: The AAA+ Domain Protein CsoCbbQ. *Sci. Rep.* **5**, 16243 (2015).
20. Whitehead, L., Long, B. M., Price, G. D. & Badger, M. R. Comparing the in Vivo Function of α-Carboxysomes and β-Carboxysomes in Two Model Cyanobacteria. *Plant Physiol.* **165**, 398–411 (2014).
21. Sun, Y. *et al.* Light Modulates the Biosynthesis and Organization of Cyanobacterial Carbon Fixation Machinery through Photosynthetic Electron Flow. *Plant Physiol.* **171**, 530–541 (2016).

22. Rohnke Brandon A., Singh Shailendra P., Pattanaik Bagmi, & Montgomery Beronda L. RcaE-Dependent Regulation of Carboxysome Structural Proteins Has a Central Role in Environmental Determination of Carboxysome Morphology and Abundance in Fremyella diplosiphon. *mSphere* **3**, e00617-17 (2018).

23. Rillema, R., Hoang, Y., MacCready, J. S. & Vecchiarelli, A. G. Carboxysome Mispositioning Alters Growth, Morphology, and Rubisco Level of the Cyanobacterium Synechococcus elongatus PCC 7942. *mBio* **12**, e02696-20 (2021).

24. Sun, Y., Wollman, A. J. M., Huang, F., Leake, M. C. & Liu, L.-N. Single-Organelle Quantification Reveals Stoichiometric and Structural Variability of Carboxysomes Dependent on the Environment. *Plant Cell* **31**, 1648–1664 (2019).

25. Peña, K. L., Castel, S. E., de Araujo, C., Espie, G. S. & Kimber, M. S. Structural basis of the oxidative activation of the carboxysomal γ-carbonic anhydrase, CcmM. *Proc. Natl. Acad. Sci.* **107**, 2455–2460 (2010).

26. Smith, K. S., Jakubzick, C., Whittam, T. S. & Ferry, J. G. Carbonic anhydrase is an ancient enzyme widespread in prokaryotes. *Proc. Natl. Acad. Sci.* **96**, 15184–15189 (1999).

27. Igamberdiev, A. U. Control of Rubisco function via homeostatic equilibration of CO2 supply. *Front. Plant Sci.* **6**, (2015).

28. Price, G. D. & Badger, M. R. Expression of Human Carbonic Anhydrase in the Cyanobacterium Synechococcus PCC7942 Creates a High CO2-Requiring Phenotype 1. *Plant Physiol.* **91**, 505–513 (1989).

29. Badger, M. R. & Price, G. D. Carbonic Anhydrase Activity Associated with the Cyanobacterium Synechococcus PCC7942 1. *Plant Physiol.* **89**, 51–60 (1989).

30. Sawaya, M. R. *et al.* The Structure of β-Carbonic Anhydrase from the Carboxysomal Shell Reveals a Distinct Subclass with One Active Site for the Price of Two *. *J. Biol. Chem.* **281**, 7546–7555 (2006).

31. Heinhorst, S. *et al.* Characterization of the Carboxysomal Carbonic Anhydrase CsoSCA from Halothiobacillus neapolitanus. *J. Bacteriol.* **188**, 8087–8094 (2006).

32. Ni, T. *et al.* Structure and assembly of cargo Rubisco in two native α-carboxysomes. *Nat. Commun.* **13**, 4299 (2022).

33. Blikstad, C. *et al. Discovery of a carbonic anhydrase-Rubisco supercomplex within the alpha-carboxysome*. http://biorxiv.org/lookup/doi/10.1101/2021.11.05.467472 (2021) doi:10.1101/2021.11.05.467472.

34. Metskas, L. A. *et al.* Rubisco forms a lattice inside alpha-carboxysomes. *Nat. Commun.* **13**, 4863 (2022).

35. Evans, S. L. *et al.* Single-particle cryo-EM analysis of the shell architecture and internal organization of an intact α-carboxysome. 2022.02.18.481072 Preprint at https://doi.org/10.1101/2022.02.18.481072 (2022).

36. Sun, Y. *et al.* Decoding the Absolute Stoichiometric Composition and Structural Plasticity of α-Carboxysomes. *mBio* **13**, e0362921 (2022).

37. Chaijarasphong, T. *et al.* Programmed Ribosomal Frameshifting Mediates Expression of the α-Carboxysome. *J. Mol. Biol.* **428**, 153–164 (2016).

38. Long, B. M., Förster, B., Pulsford, S. B., Price, G. D. & Badger, M. R. Rubisco proton production can drive the elevation of CO_2 within condensates and carboxysomes. *Proc. Natl. Acad. Sci.* **118**, e2014406118 (2021).

39. Liebschner, D. *et al.* Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr. Sect. Struct. Biol.* **73**, 148–157 (2017).

40. Cronk, J. D. *et al.* Identification of a Novel Noncatalytic Bicarbonate Binding Site in Eubacterial β-Carbonic Anhydrase. *Biochemistry* **45**, 4351–4361 (2006).

41. Marin, B., Nowack, E. C., Glöckner, G. & Melkonian, M. The ancestor of the Paulinella chromatophore obtained a carboxysomal operon by horizontal gene transfer from a Nitrococcus-like γ-proteobacterium. *BMC Evol. Biol.* **7**, 85 (2007).

42. Rae, B., Forster, B., Badger, M. & Price, G. The CO2-concentrating mechanism of Synechococcus WH5701 is composed of native and horizontally-acquired components. *Photosynth. Res.* **109**, 59–72 (2011).

43. Förster, B. *et al.* The Chlamydomonas reinhardtii chloroplast envelope protein LCIA transports bicarbonate in planta. *J. Exp. Bot.* **74**, 3651–3666 (2023).

44. Kempen, M. van *et al.* Fast and accurate protein structure search with Foldseek. 2022.02.07.479398 Preprint at https://doi.org/10.1101/2022.02.07.479398 (2023).

45. Holm, L. Using Dali for Protein Structure Comparison. in *Structural Bioinformatics: Methods and Protocols* (ed. Gáspári, Z.) 29–42 (Springer US, 2020). doi:10.1007/978-1-0716-0270-6_3.

46. Badger, M. R. Kinetic properties of ribulose 1,5-bisphosphate carboxylase/oxygenase from Anabaena variabilis. *Arch. Biochem. Biophys.* **201**, 247–254 (1980).

47. Mangan, N. M., Flamholz, A., Hood, R. D., Milo, R. & Savage, D. F. pH determines the energetic efficiency of the cyanobacterial CO2 concentrating mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5354–E5362 (2016).

48. Flamholz, A. & Shih, P. M. Cell biology of photosynthesis over geologic time. *Curr. Biol.* **30**, R490–R494 (2020).

49. Rowlett, R. S. *et al.* Allosteric Site Variants of Haemophilus influenzae β-Carbonic Anhdyrase,. *Biochemistry* **48**, 6146–6156 (2009).

50. Kim, S. *et al.* Structural insights into novel mechanisms of inhibition of the major β-carbonic anhydrase CafB from the pathogenic fungus Aspergillus fumigatus. *J. Struct. Biol.* **213**, 107700 (2021).

51. Covarrubias, A. S., Bergfors, T., Jones, T. A. & Högbom, M. Structural Mechanics of the pH-dependent Activity of β-Carbonic Anhydrase from Mycobacterium tuberculosis*. *J. Biol. Chem.* **281**, 4993–4999 (2006).

52. Todd, A. E., Orengo, C. A. & Thornton, J. M. Evolution of function in protein superfamilies, from a structural perspective11Edited by A. R. Fersht. *J. Mol. Biol.* **307**, 1113–1143 (2001).

53. Fujiwara, K., Toda, H. & Ikeguchi, M. Dependence of α-helical and β-sheet amino acid propensities on the overall protein fold type. *BMC Struct. Biol.* **12**, 18 (2012).

54. Mueller-Cajar, O. *et al.* Structure and function of the AAA+ protein CbbX, a red-type Rubisco activase. *Nature* **479**, 194–199 (2011).

55. Dobrinski, K. P., Longo, D. L. & Scott, K. M. The Carbon-Concentrating Mechanism of the Hydrothermal Vent Chemolithoautotroph Thiomicrospira crunogena. *J. Bacteriol.* **187**, 5761–5766 (2005).

56. de Araujo, C. *et al.* Identification and characterization of a carboxysomal γ-carbonic anhydrase from the cyanobacterium Nostoc sp. PCC 7120. *Photosynth. Res.* **121**, 135–150 (2014).

57. McGurn, L. D. *et al.* The structure, kinetics and interactions of the β-carboxysomal β-carbonic anhydrase, CcaA. *Biochem. J.* **473**, 4559–4572 (2016).

58. Flamholz, A. I. *et al.* Trajectories for the evolution of bacterial CO2-concentrating mechanisms. *Proc. Natl. Acad. Sci.* **119**, e2210539119 (2022).

59. Catanzariti, A.-M., Soboleva, T. A., Jans, D. A., Board, P. G. & Baker, R. T. An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Sci. Publ. Protein Soc.* **13**, 1331–1339 (2004).

60. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

61. Wilkins, M. R. *et al.* Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol. Clifton NJ* **112**, 531–552 (1999).

62. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

63. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).

64. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

65. Agirre, J. *et al.* The CCP4 suite: integrative software for macromolecular crystallography. *Acta Crystallogr. Sect. Struct. Biol.* **79**, 449–461 (2023).

66. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. Sect. Struct. Biol.* **75**, 861–877 (2019).

67. McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **109**, 1528–1532 (2015).

68. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

69. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).

70. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

71. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

72. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

73. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

74. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

75. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

76. Markowitz, V. M. *et al.* The integrated microbial genomes (IMG) system. *Nucleic Acids Res.* **34**, D344–D348 (2006).

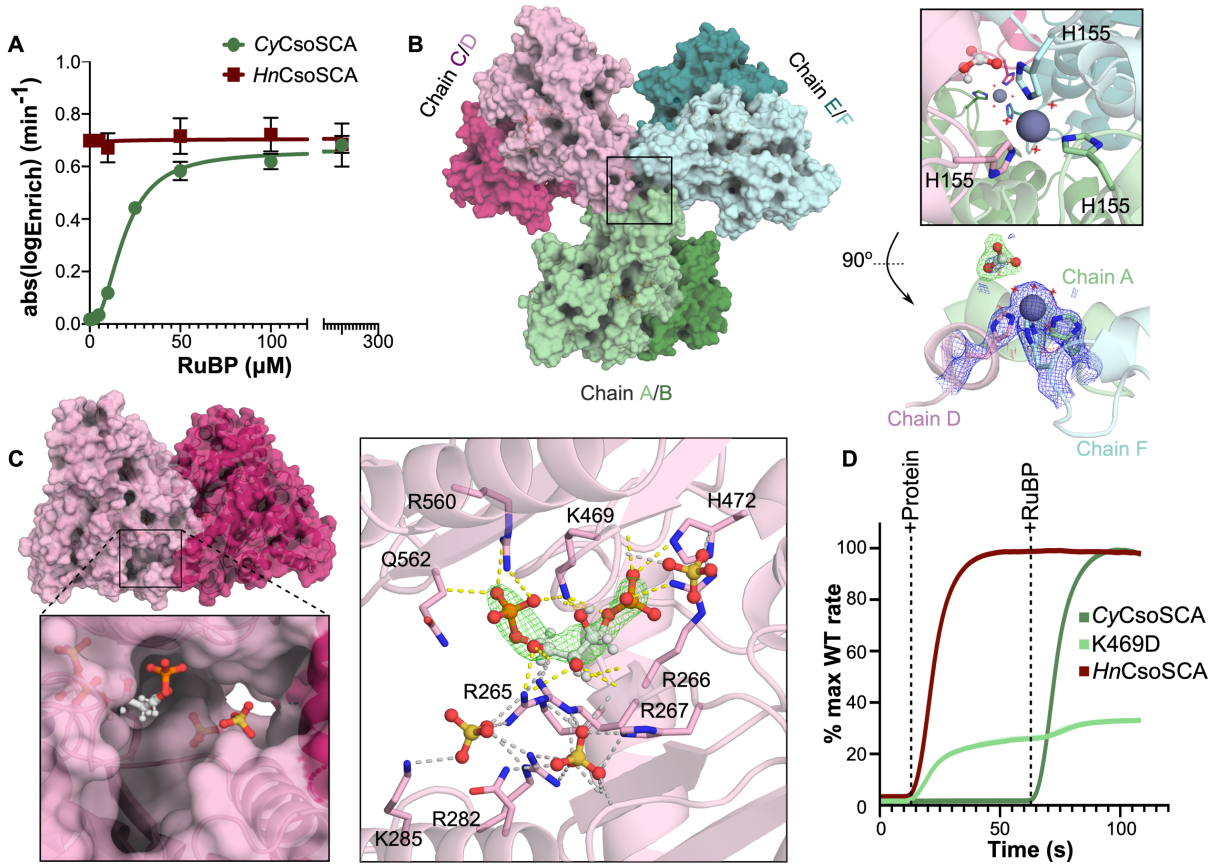

Figure 1. RuBP allosterically regulates CyCsoSCA. **A** CyCsoSCA and HnCsoSCA activity as a function of RuBP concentration, measured using Membrane Inlet Mass Spectrometry (MIMS). Measurements reported are an average of three technical replicates, error bars represent standard error. Curves fitted using GraphPad Prism. **B** The homohexameric CyCsoSCA structure solved to 2.3Å (PBD: 8THM). The 'trimer of dimers' arrangement is shown with dimers coloured pink, green or blue, monomers indicated as different shades, chains annotated. The black square denotes the apical structural zinc atoms with further detail shown in the right insert. Waters in the octahedral coordination sphere are shown as red crosses. Zinc ions are shown as grey spheres, an opportune bicarbonate ion is shown in ball and stick representation. Below: 2mF_o-dF_c density at key interacting residues is shown (1s). Polder omit map (green) is shown at a contour level of 5.0 σ to highlight bicarbonate ion density at the A/D/F apex. **C** The C/D dimer is shown with a box and insert highlighting the RuBP binding pocket of Chain D (light pink monomer). Right panel: The RuBP binding site is shown with Polder omit map density[39] of the ligand overlayed at a contour level of 7.0 σ. Sulfate ions and RuBP shown in ball and stick representation. Polar interactions are shown as dotted lines, those corresponding to interactions with RuBP directly are

565 coloured yellow, other secondary/SO_4 interactions shown in grey. **D** The relative rate of
566 CyCsoSCA, the K469D mutant of CyCsoSCA, and HnCsoSCA are shown over time as a
567 proportion of the maximum recorded reaction rate for each respective wild type enzyme. Time
568 points at which protein and 100µM RuBP were added to the MIMS cuvette are indicated. Curve
569 is the mean of three technical replicates with shaded area around the curve representative of
570 standard deviation from the mean (not visible due to scale).
571

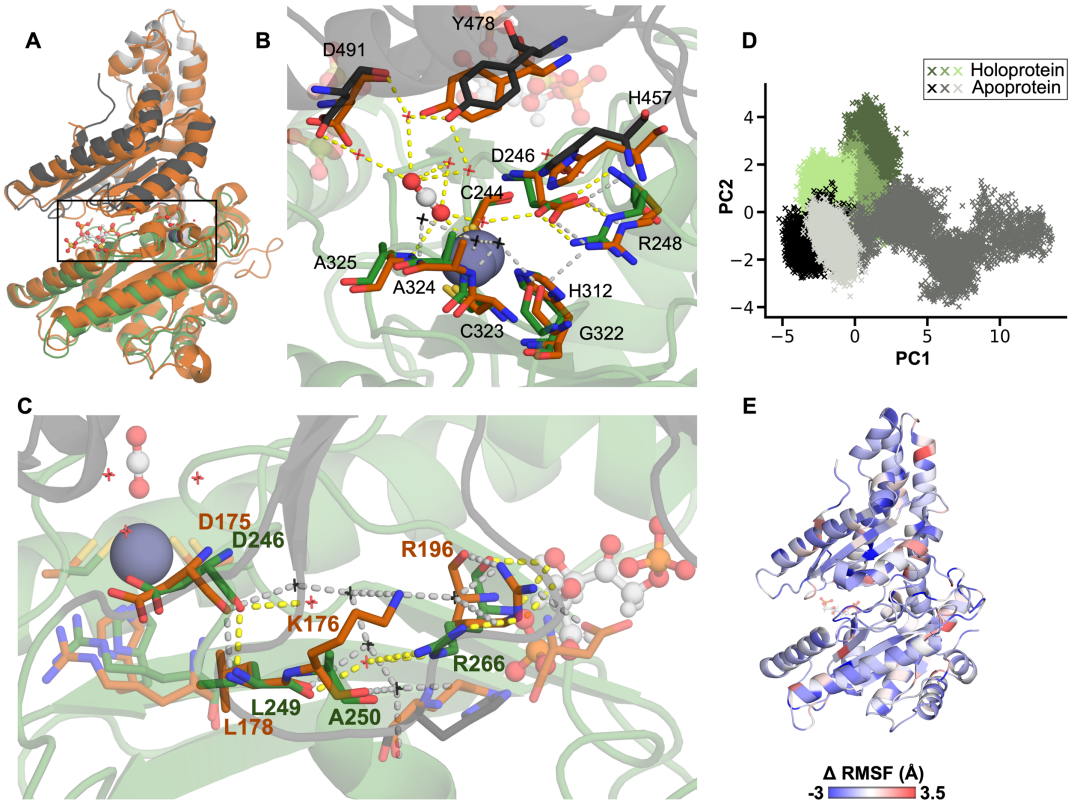

Figure 2. Active site differences from HnCsoSCA are key to allosteric activation of CyCsoSCA. **A** Structural alignment of CyCsoSCA and HnCsoSCA monomer (CA RMSD 1.5 Å). The CyCsoSCA variant is coloured by domain (white: N-terminal domain (NTD), green: Catalytic domain (Catalytic) and dark grey: C-terminal domain (CTD)). HnCsoSCA (PDB 2FGY) is shown in orange. A box denotes the active site and RuBP binding site region. **B** A close view of the CyCsoSCA active site is shown with key residues in stick representation, coloured as in A. Corresponding HnCsoSCA residues (orange) are overlaid. All ligands (CO_2, RuBP, SO_4) are shown in ball and stick representation. Catalytically relevant water molecules are shown as crosses (black waters are those in the HnCsoSCA structure, red indicates those in CyCsoSCA). Dashed lines indicate polar bonds (grey indicating those between HnCsoSCA molecules, yellow for CyCsoSCA interactions). Residue names are annotated according to the CyCsoSCA structure. **C** The proposed allosteric network in CyCsoSCA (top, green) and the corresponding region in HnCsoSCA aligned. Key residues for each structure are annotated in orange (HnCsoSCA) or green (CyCsoSCA). **D** Principal component analysis (PCA) comparing cartesian coordinates of the CyCsoSCA backbone in each MD simulation. Replicate simulations of CyCsoSCA with and

589 without RuBP are shown in green (Holoprotein) and grey (Apoprotein), respectively. **E** The
590 average difference in root mean square fluctuation (RMSF) values of holoprotein and apoprotein
591 simulations (Δ RMSF), where a negative value indicates a greater RMSF (and thus more mobile
592 residue) in the holoprotein. Values are mapped onto the CyCsoSCA monomer.
593

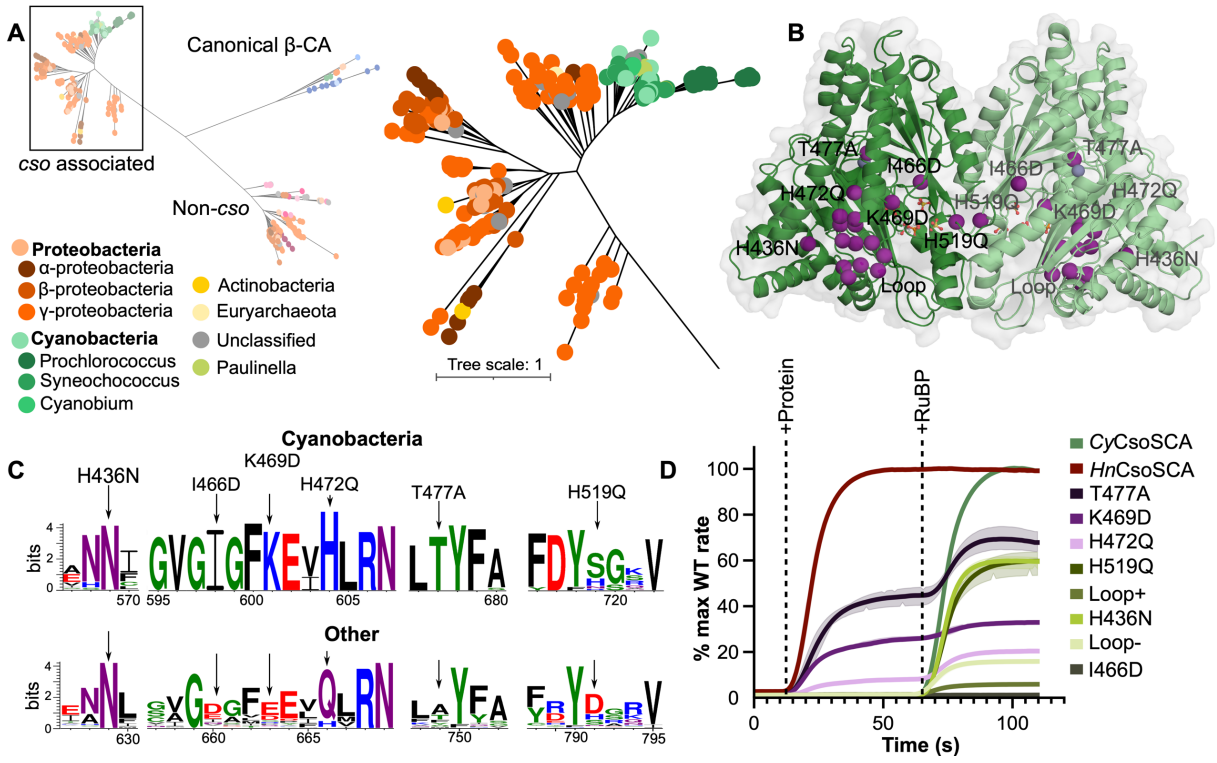

Figure 3 CsoSCA sequence analysis and mutagenesis. **A** Unrooted maximum likelihood phylogeny of 518 CsoSCA sequences produced by IQ-Tree with 18 canonical β-CAs as an outgroup (annotated as 'β-CA'). The cluster containing members associated with the *cso* operon is shown in detail. Tips are coloured by taxonomy according to the legend. Tree scale refers to the number of substitutions per site. See supplementary information for complete tree annotation and documentation. **B** Sites targeted for mutagenesis shown as purple spheres on the *Cy*CsoSCA dimer. **C** Sequence logos based on cyanobacterial *cso* associated CsoSCA sequences ('Cyano') or other bacterial *cso* associated CsoSCA sequences ('Other') of key sites targeted for mutation. Residues coloured by chemistry, logo generated through WebLogo3. Mutations are as represented above the logo **D** Activity assays of targeted mutants relative to the maximum rate recorded for wild type *Cy*CsoSCA. *Hn*CsoSCA activity is also shown for comparison as a proportion of its maximum recorded rate. HCO_3^- dehydration activity was recorded using MIMS after the addition of CsoSCA variants (+Protein) and upon addition of 100 μM RuBP (+RuBP). Three technical replicates were recorded for each variant, standard deviation is indicated by shading (error on some samples not visible due to scale).

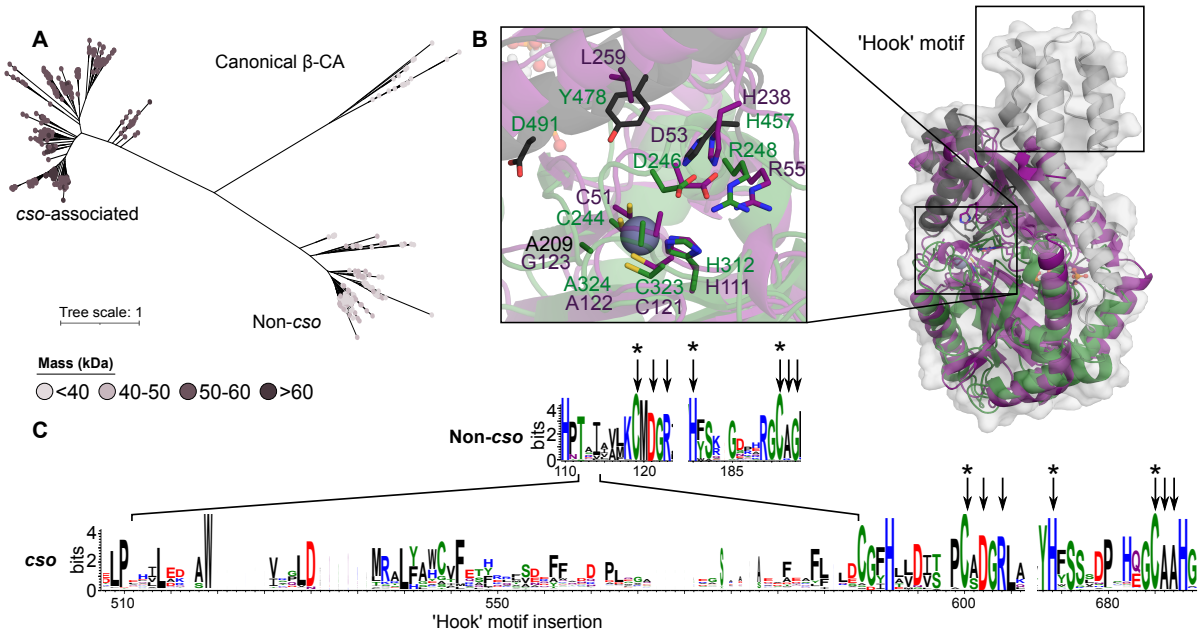

Figure 4. The N-terminal domain is unique to carboxysome operon-associated CsoSCA homologues. **A** Maximum likelihood tree of the CsoSCA PFAM (PF008936) and canonical β-CA with nodes coloured by predicted mass as per the legend. **B** Structural alignment of Cα backbones of *Cy*CsoSCA (NTD grey, Catalytic domain green, CTD black) to an AF2 generated model of a candidate 'non-*cso*' sequence (UniProt ID: A0A080M7C6, purple) with close-up view of key active site residues. The insertion exclusive to *cso*-associated sequences that encodes the NTD 'hook' motif is annotated. **C** Sequence logo of alignments of either CsoSCA members associated with *cso* operons (*cso)* or non-*cso* sequences. The *cso* insertion encoding the NTD hook motif is annotated. Key catalytic residues are indicated with arrows, zinc binding residues with asterisks.

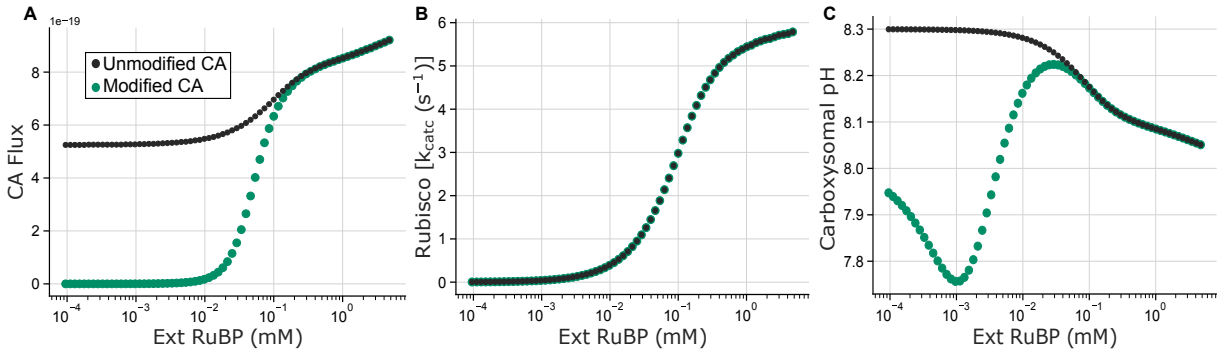

Figure 5 Results of a reaction-diffusion model[38] adapted to emulate *Cyanobium* α-carboxysome function with and RuBP d-dependent (Modified CA, green dots) or a constituently active (Unmodified CA, black dots) CA. **A** In the modified model, carboxysomal CA activity was altered based on data shown in Figure 1A to be dependent on carboxyosmal RuBP concentrations (Ext RuBP (mM)). 'CA flux' is indicative of CA activity, confirming RuBP dependence in the modified model. **B** Rubisco carboxylation turnover rates (Rubisco [k_{catc} (s^{-1})]) as a function of modelled cellular RuBP concentrations (Ext RuBP (mM)) in the modified and unmodified systems. **C** Modelled carboxysomal pH is plotted as a function of the modelled cellular RuBP content (Ext RuBP (mM)) indicates a decrease in carboxysomal pH when CA function is allosterically controlled by RuBP, with potential to aid Rubisco function as described previously[38].