***REVISED Manuscript (text UNmarked)**

Carboxysomal carbonic anhydrases:

Structure and role in microbial CO_2 fixation

Gordon C. Cannon*[1], Sabine Heinhorst[1] and Cheryl A. Kerfeld[2,3]

[1] Department of Chemistry and Biochemistry, The University of Southern Mississippi, 118 College Dr. #5043, Hattiesburg, MS 39406-0001.

2 US Department of Energy-Joint Genome Institute, 2800, Mitchell Dr., Walnut Creek, CA 94598

3 Department of Plant and Microbial Biology, University of California, 111 Koshland Hall, Berkeley, CA 94720

* Author to whom correspondence should be addressed.

Tel: 601-266-4221

Fax: 60126-6075

email: gordon.cannon@usm.edu

Keywords: carboxysome; carbonic anhydrase; CO_2-concentrating mechanism; cyanobacteria; chemoautotrophs.

Abstract

Cyanobacteria and some chemoautotrophic bacteria are able to grow in environments with limiting CO_2 concentrations by employing a $\underline{C}O_2$-$\underline{c}$oncentrating $\underline{m}$echanism (CCM) that allows them to accumulate inorganic carbon in their cytoplasm to concentrations several orders of magnitude higher than that on the outside. The final step of this process takes place in polyhedral protein microcompartments known as carboxysomes, which contain the majority of the CO_2-fixing enzyme, RubisCO. The efficiency of CO_2 fixation by the sequestered RubisCO is enhanced by co-localization with a specialized carbonic anhydrase that catalyzes dehydration of the cytoplasmic bicarbonate and ensures saturation of RubisCO with its substrate, CO_2. There are two genetically distinct carboxysome types that differ in their protein composition and in the carbonic anhydrase(s) they employ. Here we review the existing information concerning the genomics, structure and enzymology of these uniquely adapted carbonic anhydrases, which are of fundamental importance in the global carbon cycle.

1. **Introduction**

It has been the long-held view that the intracellular structure of bacteria is relatively simple compared to that of eukaryotes. This view has led to the notion that, while some co-localization of functionally related enzymes does occur in bacteria, specific or regulated compartmentalization of metabolic processes into distinct organelles is lacking. Recent genomic, biochemical, structural and imaging studies, however, have revealed a more complex picture of organization within the bacterial cell [1, 2]. Many bacteria compartmentalize metabolically related enzymes, presumably to increase their efficiency and/or control flow and concentrations of substrates and intermediates. Among these primitive organelles, the bacterial microcompartments (BMCs) provide a striking contrast to the lipid bilayer-bound organelles of eukaryotes in that their bounding shell is a thin (2-3 nm) protein monolayer.

The first BMC to be discovered was the carboxysome [3], which occurs in all cyanobacterial species thus far examined and in many non-photosynthetic autotrophic bacteria. Depending on species and growth condition, approximately 5 to 20 of these mainly icosahedral structures, which generally measure 100 to 200 nm in diameter, are visible per cell by transmission electron microscopy. The organelles contain most, if not all, of the cell's ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO), the central CO_2-fixing enzyme of the Calvin-Benson-Bassham cycle; hence the term carboxysome (reviewed in [4, 5]). Bacteria that contain carboxysomes are generally found in environments with CO_2 concentrations that are lower than the K_M for CO_2 of RubisCO [6]. To meet their need for carbon, these organisms rely on a CO_2-concentrating mechanism (CCM) that accumulates within the cytosol a pool of inorganic carbon (C_i), mainly in the form of bicarbonate, at many times the extracellular concentration. However, RubisCO requires CO_2 and cannot utilize bicarbonate. It is generally accepted that co-sequestration with a carbonic anhydrase (CA), which rapidly converts cytosolic bicarbonate to CO_2 within the carboxysome, effectively raises the local steady state CO_2 concentration to levels well above the K_M of RubisCO and allows the enzyme to function at a fully saturated state (Fig. 1). Indeed, all members of the two genetically different classes of carboxysomes studied so far have been reported to contain at least one species of CA [7, 8]. Here we review the current state of knowledge about structure and activity of the carboxysomal CAs and discuss models of their role in the carboxysome that have contributed to our current understanding of the way in which these enzymes perform their function within a BMC.

2. Two types of carboxysomes

The strong morphological resemblance all carboxysomes displayed in electron micrographs supported the casual assumption that microcompartments from different sources share similar protein composition and structure. When the genome sequences of a number of cyanobacteria and chemoautotrophs became available during the past decade, the existence of at least two types of carboxysomes, distinguished by their protein composition, became evident [9, 10]. The α-type is found in all carboxysome-forming chemoautotrophs and in cyanobacteria that contain form IA RubisCO, such as *Prochlorococcus* and oceanic species of *Synechococcus*. Cyanobacteria that contain form IB RubisCO, exemplified by well characterized species such as *Synechocystis* PCC6803 and *Synechococcus* PCC7002 and PCC7942, contain β-carboxysomes [9, 11].

The proteins that assemble into α-carboxysomes are encoded by a conserved *cso* operon of 6 to 9 genes. These include the genes for the large and small subunit of RubisCO, a gene (*csoS3*) specifying a CA unique to α-carboxysomes (see below), and three other genes (*csoS2, csoS4, csoS1*) -some of them existing in multiple paralogs- that encode carboxysome shell components (Fig. 2). The picture is not quite as neat for the β-carboxysomes in that the genes for its proteins are dispersed throughout the genome [11]. Difficulties in obtaining homogeneous preparations of β-carboxysomes have made it impossible to assemble a complete inventory of the protein components that comprise this carboxysome type [12]. As a result, the presumed β-carboxysome genes have been deduced primarily from genetic and comparative genomic studies [11].

In addition to the large and small subunit of RubisCO, α- and β-carboxysomes share two classes of proteins that are fundamental constituents of their shells [10]. The major shell proteins, which are always present as multiple paralogs, are the CsoS1 proteins of α-carboxysomes and the orthologous CcmK proteins of β-carboxysomes. Crystallographic structural analyses have revealed that these proteins form tile-like hexamers or pseudo-hexamers that tightly associate with each other to create the facets of the icosahedral shells [13-15]. The second class of shell protein genes common to both carboxysome types is represented by *csoS4A* and *csoS4B* in the α-carboxysome operon and their ortholog, *ccmL*, in β-carboxysome gene clusters. Both *csoS4* genes are expressed in actively growing cultures of *Halothiobacillus neapolitanus* [16]. However, the inability to visualize these proteins in SDS-PAGE gels of purified carboxysome preparations has led to the prediction that these proteins are of very low abundance in the shell. The crystal structures of CsoS4A from *H. neapolitanus* and CcmL from *Synechocystis* PCC6803 were recently solved and revealed that both proteins form pentamers whose shape and size fulfill the geometric requirements for the 12 vertex elements that are required to complete the icosahedral shape of the carboxysome shell [17]. The low abundance of these proteins in carboxysomes is consistent with their purported structural roles.

Both types of carboxysome contain one protein that is encoded by a single gene but exists in multiple molecular weight forms in the organelle. The *csoS2* gene encodes a protein with a predicted molecular mass of 92 KDa. On SDS-polyacrylamide gels of purified *H. neapolitanus* carboxysomes, however, two forms of this protein (apparent molecular mass 130 and 85 KDa, respectively) are resolved that share a common N-terminus [18]. CsoS2 was shown to be a shell constituent in *H. neapolitanus* carboxysomes, where the two polypeptides account for approximately 12% of the total protein mass of the organelle [4, 18]. The function of CsoS2 remains enigmatic; its strongly basic isoelectric point has led to the speculation that the protein(s) may interact with the negatively charged reactants and products of the RubisCO-catalyzed CO_2 fixation reaction, ribulose 1,5-bisphosphate and 3-phosphoglycerate, respectively, which must cross the carboxysome shell [19]. The CcmM protein of β-carboxysomes, which contains a γ-CA domain (discussed below) and shares no detectable primary structure homology with CsoS2, is also present in multiple forms [20, 21].

Structural and biochemical studies, combined with comparative genomic evidence have clearly established the existence of two types of carboxysomes. Although their shells are built from the same basic principles and both organelles appear to share a common function, they differ in composition. One crucial protein constituent that varies significantly between α- and β-carboxysomes is the associated CA, which supplies the encapsulated RubisCO with its substrate, CO_2, and contributes to the catalytic enhancement RubisCO derives from microcompartmentalization.

3. CO_2 concentrating mechanisms

Despite the central role of RubisCO in the Calvin-Benson-Bassham CO_2 fixation cycle, the enzyme is rather poorly equipped for its task in that is has a low affinity for CO_2 and a relatively low turnover number. Furthermore, RubisCO does not discriminate well between CO_2 and the competing substrate, O_2, the fixation of which feeds the apparently wasteful process of photorespiration [22]. Photosynthetic and chemoautotrophic organisms have evolved a number of mechanisms that compensate for these shortcomings, including specialized cells, pyrenoids, multiple RubisCO forms that have developed higher substrate affinities at the cost of catalytic efficiency, and in case of autotrophic bacteria, CCMs [6, 23, 24]. The molecular workings of CCMs vary greatly in autotrophic bacteria, reflecting the diverse environments in which the bacteria thrive and the levels of C_i available in their habitats [25]. In cyanobacteria, five distinct C_i uptake systems have been described that display a range of transport activities, include transporters for CO_2 and bicarbonate and can be either constitutively expressed or transcriptionally regulated based on the C_i concentration in the environment of the cells [11, 23, 25, 26]. Regardless of the nature of the transporters employed, the general strategy common to all cyanobacteria is to use energy for C_i uptake and to create a cytosolic pool of dissolved C_i that is estimated to be upwards of three orders of magnitude higher than that outside of the cell. The intracellular pH of around 8.0 favors an equilibrium ratio of HCO_3^- to CO_2 of near 100. Bicarbonate, the predominant C_i species, is charged and cannot escape by diffusion through the lipid bilayer of the plasma membrane.

The carboxysome is common to all cyanobacterial CCMs. Through the activity of its CA, the captured HCO_3^- is rapidly dehydrated to CO_2, which is subsequently fixed by the 250 to 500 RubisCO holoenzyme molecules estimated to be contained within each organelle. A wealth of existing genetic and physiological data supports this model of CCM function in cyanobacteria, gleaned primarily from studies conducted with the β-carboxysome-containing species *Synechococcus* PCC7942 and PCC7002, and *Synechocystis* PCC6803. Perhaps the most striking experimental evidence supporting CCM function was obtained by ectopically expressing human α-CA in the cytoplasm of *Synechococcus* PCC7942 [27]. Rapid equilibration of HCO_3^- to CO_2 by the recombinant enzyme allowed the gas to escape from the cell, dissipated the intracellular C_i pool and was responsible for the high CO_2 requiring (*hcr*) phenotype of the mutant.

Quantitative models of CCM function in a hypothetical cyanobacterial cell accurately predicted the carbon assimilation kinetics observed in both wild type *Synechocystis* PCC6803 cells and the mutant expressing human CA in its cytoplasm [28-31]. However, the models were unable to distinguish whether the thin carboxysome protein shell itself is a CO_2 leak barrier, or whether escape of CO_2 from the microcompartment is suppressed by burying the CA in the center of the carboxysome and effectively hindering passage of the gas through the densely packed RubisCO molecules.

Until recently, it was unclear whether chemoautotrophic bacteria possess a functional CCM, although many species contain well-defined carboxysomes [32]. Scott and co-workers [33] clearly demonstrated that the sulfur-oxidizing γ-proteobacterium *Thiomicrospira crunogena* is able to elevate its cytosolic C_i concentration through active import of CO_2 and bicarbonate. Since *T. crunogena* contains α-carboxysomes, these findings support a common role in the CCM for both α- and β-carboxysomes, regardless of differences in their composition, most significant of which is their CA complement.

4. Candidates for carboxysomal carbonic anhydrases

The earliest proposals for an inducible CCM in cyanobacteria assumed that bicarbonate would be the species of C_i accumulated in the cytoplasm and foresaw the necessity for its rapid conversion to CO_2 to allow subsequent fixation by RubisCO [28, 34]. Since it appears that at least a large portion of the cell's RubisCO is packaged in the carboxysome it seemed logical that a CA might be co-sequestered within the microcompartment, and efforts to identify and characterize the carboxysomal CA were initiated in both cyanobacteria and chemoautotrophs [35, 36]. This explanation was complicated by the fact that no CA activity was detected in homogeneous preparations of *H. neapolitanus* carboxysomes [36]. In addition, the problems encountered in isolating cyanobacterial carboxysomes free of contaminating cellular membranes made it difficult to state unambiguously that the low levels of CA activity measured in particulate fractions of cell-free extracts represented a bona fide carboxysomal CA [35, 37].

Direct evidence for a carboxysomal CA was established in *H. neapolitanus* by systematically overexpressing each gene of its carboxysome operon in *E. coli* and analyzing the recombinant protein products for CA activity. The CsoS3 protein, a low abundance component previously shown to be tightly associated with the α-carboxysomal shell [38], was demonstrated to be a novel CA [8]. Heterologous expression of the homologous genes from the cyanobacteria *Prochlorococcus* MED4 and MIT9313, and *Synechococcus* WH8102 confirmed the identity of CsoS3 as the CA of α-carboxysomes. Comparison of available microbial genome sequences has confirmed that CsoS3, which was renamed CsoSCA (for *c*arboxy*so*me *s*hell *c*arbonic *a*nhydrase), is unique to α-carboxysomes and its gene common to all known *cso* operons (Fig. 2) [10]. Improvements in the purification of α-carboxysomes have allowed direct detection of CA activity in homogeneous preparations of *H. neapolitanus* carboxysomes by the ^{18}O exchange assay [39], the standard potentiometric assay [40] and the stopped flow changing indicator method [41]. The inability to detect CA activity in *H. neapolitanus* carboxysomes previously has been attributed mainly to inclusion of dithiothreitol (DTT), now known to be a potent inhibitor of many CAs, in all carboxysome isolation buffers. The previously noted difficulty in isolating pure β-carboxysomes has made it difficult to ascertain the location of a CA within these organelles by direct biochemical assays.

To establish the contribution of the carboxysomal CA to CO_2 fixation by the encapsulated RubisCO requires detailed studies that address the number of CA molecules per carboxysome, the location of the enzyme within the microcompartment and its catalytic performance. Of the three carboxysomal CAs discussed in the following sections, to date such studies have been performed mainly with CsoSCA and, to a lesser extent, with CcaA. The fact that CcmM currently is the only potential carboxysome CA candidate in a number of cyanobacteria, although CA activity has yet to be shown, justifies its inclusion in this review.

4.1. CsoSCA, the carbonic anhydrase of α-carboxysomes

The sole CA associated with α-carboxysomes was originally demonstrated to be a carboxysome shell protein of unknown function in *H. neapolitanus* by immuno-electron microscopy of cell thin sections and isolated carboxysomes [38]. Densitometric analysis of the stained carboxysome polypeptides in SDS-polyacrylamide gels revealed the relative abundance of the protein to be approximately 3% of the total carboxysome protein by weight, which corresponds to approximately 80 molecules per carboxysome [4]. Upon discovery of its enzymatic activity, the lack of significant sequence homology to representatives of the other recognized CA classes (α, β, γ) led to the proposal of a new class designation (ε) [8]. Detailed CA activity measurements in carboxysomes disrupted by freeze-thawing and subsequent fractionation on density gradients confirmed the tight association of the enzyme with the shell [42]. Attempts to strip CsoSCA from the shell with mild detergents or high salt washes were unsuccessful. Partial disassembly of the shell by treatment with urea releases CsoSCA and led to the suggestion that the protein is an integral shell component and might traverse the protein layer. Heterologous expression and purification of recombinant CsoSCA (rCsoSCA) from *E. coli* resulted in the isolation of enzymatically active homodimers of approximately 120 KDa by size exclusion chromatography. rCsoSA was found to contain one zinc ion per monomer, which agrees with structural data. Removal of the zinc results in loss of activity that cannot be restored by addition of Co^{2+}, Fe^{2+} or Mg^{2+}. DTT is a potent inhibitor of CsoSCA ($IC_{50} = 2.6$ $X10^{-4}$ M), a characteristic the enzyme shares with β-CAs from cyanobacteria and many plants [42]. The moderate resistance of CsoSCA to inhibition by azide is reminiscent of that seen for the β- and γ-CA from *Methanosarcina thermophila* [43].

Measurement of the kinetic parameters for rCsoSA by the stopped flow changing indicator method [41] established k_{cat} values for the hydration and dehydration reaction that are pH-dependent and well within the range observed for other β-CAs [42]. In the pH range of 7 to 8, which is assumed to reflect the cytosolic pH of *H. neapolitanus*, the efficiency of HCO_3^- dehydration was measured to be 3 to 5 fold lower than that of CO_2 hydration. Nevertheless, even if the k_{cat} for bicarbonate dehydration by rCsoSCA is applied to 80 monomer or 40 dimer copies of the enzyme per carboxysome, its activity is likely to be more than adequate to supply the approximately 270 molecules of RubisCO holoenzyme molecules in the interior of the *H. neapolitanus* carboxysome with the CO_2 concentrations required for substrate saturation [42]. Comparison of turnover number for CO_2 hydration as a function of pH between intact and broken carboxysomes did not provide a basis for suggesting that the interior pH of the carboxysome differs from the external pH. However, more direct studies that examine the possibility of the carboxysome functioning as a proton “cage” are clearly required.

The effect of CsoSCA on the ability of carboxysomes to fix CO_2 was further examined by constructing a mutant in which the gene that encodes CsoSCA in *H. neapolitanus* (*csoS3*) is interrupted by insertion of a kanamycin cassette [44]. The resulting mutant requires elevated CO_2 to grow at wild type rates and was shown by immuno-blotting and enzymatic analysis to be devoid of CsoSCA protein, but produces carboxysomes that are morphologically indistinguishable from wild type microcompartments. Purified mutant carboxysomes contain all other proteins except CsoSCA in near-wild type ratios, suggesting that CsoSCA plays no major role in maintaining carboxysome structure or in the biogenesis of the organelle. The kinetic parameters of CO_2 fixation by purified mutant carboxysomes support the proposed role of CsoSCA in carboxysome function in that the V_{max} is not affected by the lack of CA activity. The bicarbonate concentration required to support half-maximal CO_2 fixation rates (K_c), on the other hand, is approximately three times higher than that measured for wild type carboxysomes. Addition of excess amounts of rCsoSCA to assays of intact mutant carboxysomes did not restore wild type K_c values, indicating that generation of CO_2 on the carboxysome exterior does not increase the CO_2 concentration in the interior. These results are consistent with the notion that the protein shell impedes the diffusion of CO_2.

Furthermore, mutant carboxysomes disrupted by freeze-thawing displayed K_c values similar to those measured for intact wild type microcompartments. In the absence of CsoSCA, the shell apparently must be rendered leaky for CO_2 to reach RubisCO inside the carboxysome rapidly enough and/or at local concentrations high enough to support wild type fixation rates [44].

The combined results from these studies are consistent with the prediction of the CCM model that the catalytic activity of the carboxysome-associated CA ensures the CO_2 fixation rates needed to support efficient cell growth. In addition, they demonstrate the importance of the association of CA with the inside of the carboxysomal shell and strongly imply a role for CsoSCA as a facilitator of C_i flux into the carboxysome across the CO_2 diffusional barrier established by the protein shell (Fig. 3).

At present, CsoSCA is the only carboxysomal CA that has been structurally characterized (Fig. 4A) [45]. The protein is composed of three domains (Figs. 4A, 5): the N-terminal domain (residues 38-144) is entirely α-helical, forming an up-down four-helix bundle that has no structural counterpart in the Protein Databank. The second domain, residues 151-397, is an α+β fold common to all β-CAs. It is the catalytically active domain, containing one His (242) and two Cys (173 and 253) residues for coordination of the Zn ion. The C-terminal domain (residues 398-514) is structurally similar to the catalytic domain (they superimpose with an rmsd of 1.45 Å over 61 Cα carbons) and is spatially related to it by a two-fold rotation axis. The similarity likely reflects an ancient gene duplication and fusion followed by divergence; the primary structures of the two domains have only 11% sequence identity. The C-terminal domain is smaller and lacks several elements of secondary structure found in the catalytic domain. It is missing all of the equivalent residues for coordination of the Zn ion; the structurally analogous positions in the C-terminal domain to Cys 173 and His 242 of the catalytic domain are occupied by Asp 406 and Arg 457. The loop in the catalytic domain that contains Cys 253 has been deleted from the C-terminal domain.

It was unexpected that CsoSCA proved to be a member of the β-CA family since primary structure homology to other β-CAs is undetectable. The fold of its catalytic domain is similar to those of other β-CAs; all contain a central five-stranded β-sheet and three or more flanking helices. The structure of CsoSCA superimposes on that of the β-CA from the red alga *Porphyridium purpureum* (PDB code 1DDZ) with an rmsd of 1.93 Å over 134 Cα atoms despite only 10.8% identity between their primary structures [45]. The differences between CsoSCA and representatives of the two previously described subclasses [46], the plant (*P. purpureum*, *E. coli*, *Pisum sativum*, *Mycobacterium tuberculosis* Rv3588c) and cab-type (*Methanobacterium thermoautotrophicum*, *Mycobacterium tuberculosis* Rv1284) suggest that CsoSCA represents a third subclass. The plant and cab β-CAs are more similar to one another than to CsoSCA; structures of members of the plant and the cab type superimpose with an average pairwise rmsd of 0.9Å for 132 Cα pairs while CsoSCA superimposes on these two with an average pairwise rmsd of 2.4Å over 81 Cα pairs [45].

In other β-CAs there is tendency to two-fold symmetric pairing, or multiples of this pairing, of active sites. Interestingly, despite the extensive divergence in primary structure among the β-CAs, all β-CAs dimerize with two-fold symmetry. This coincides with the pseudo-symmetry that relates the catalytic and C-terminal domains of CsoSCA. Thus not only is the tertiary structure of the catalytic domain conserved but the propensity to form this dimer has also been conserved.

At the level of primary structure, the cab, plant and carboxysomal β-CAs share only five absolutely conserved residues, each of which plays a key role in the catalytic mechanism (see Fig. 2 in [45]). The structural similarities among the β-CA active sites suggest that the mechanism for dehydration of HCO_3^- is analogous to the reverse reaction, CO_2 hydration. In general, β-CAs use a Zn^{2+}-mediated catalytic mechanism, although recently a cryptic β-CA found in diatoms was shown to be able to use Cd^{2+} as the catalytic metal and substitute for Zn^{2+} [47]. In all β-CA crystal structures the Zn^{2+} ion is coordinated by two Cys and one His (Fig. 5). The fourth ligand is more variable among β-CAs; in CsoSCA it is a water molecule. As in other CAs in which the fourth ligand is a water molecule, the water is positioned by a conserved Asp (175 in CsoSCA) that is oriented through hydrogen bonding to the sidechain of a conserved Arg residue (177 in CsoSCA). The displacement of the water molecule by Asp 175 to permit HCO_3^- binding is likely an initiation step in the catalytic cycle. Asp 175 also likely plays a role in orienting the HCO_3^- molecule so that the protonated oxygen atom is poised for coordination by the Zn^{2+}. The orientation of the Asp 175 sidechain by two hydrogen bonds to Arg 177 allows it to both displace the water ligand and orient the HCO_3^-.

The similarity in active site architecture shared by CsoSCA and the other β-CAs might be attributed to convergent evolution. However, the similarity in tertiary structure between CsoSCA and the other β-CAs is compelling evidence for its membership in this CA family; tertiary structure is more closely associated with function than primary structure and changes much more slowly over time than DNA or amino acid sequence. While convergent evolution leads to similar functions and is typically recognized by a similarity in active sites (e.g. subtilisin and γ-chymotrypsin) [48], the extent of the structural similarity between CsoSCA and the β-CAs extends beyond the catalytic site to the fold of the catalytic domain. The classic example of the use of structure to distinguish convergent from divergent evolution from a common ancestor was based on the structural similarity of bacteriophage T4 and hen egg white lysozyme [49]. The similarity between these functionally related structures was deemed to be the result of divergence from a common ancestor; convergent evolution was an implausible explanation for the similarity because it required positing that “the demands of function independently led to similarities in the mechanism of catalysis, in (substrate)-protein interactions and in three-dimensional fold of the polypeptide chain” [49]. Instead, divergence from a common ancestor provides a far more parsimonious explanation. Collectively, the structural evidence strongly suggests that CsoSCA shares a common ancestor with the other members of the β-CA family.

Most β-CAs are dimers or multiples of dimers. For example, the *Pisum sativum* β-CA is a novel dimer of a dimer of dimers (resulting in an octamer), and the red algal *P. purpureum* β-CA is a homodimer, in which each monomer is composed of two internally repeated structures, each with an active site. The notable exception to this tendency is CDCA1, the β-CA found in diatoms [47]. This protein is active as a monomer; however, interestingly, comparison of the three-dimensional structure of CDCA1 to β-CAs revealed that the CDCA1 monomer is structurally similar to a typical β-CA dimer. Based on its elution behavior in size exclusion chromatography, CsoSCA appears to be a dimer in solution [45]. The protein crystallizes with two molecules in the asymmetric unit. In the crystal, each CsoSCA molecule participates in two types of dimerization that, based on structural arguments, are plausibly relevant dimeric assemblies (Fig. 4A) [45]. The largest interface between adjacent CsoSCA molecules buries a total of 3304 $Å^2$ with substantial shape complementarity between the two molecules; the shape complementarity of this intermolecular interface (Sc = 0.717), is comparable to that observed in protease-inhibitor interfaces [50]. This interface is mediated by a pair of symmetry-related helices (residues 206-223) (Fig. 4A) within the catalytic domain and portions of the C-terminal domain. Nine of the 27 residues involved in this dimerization are absolutely conserved among the CsoSCA proteins in the sequence database. The interface is stabilized by ten intermolecular hydrogen bonds and a salt bridge between the absolutely conserved residues Glu 207 and Arg 468.

There is a second interface (Fig. 4A) between adjacent CsoSCA molecules in the crystal that, based on it structural properties, could be biologically relevant. The Sc of this interface is 0.684, which is similar to surface complementarity observed in antigen-antibody complexes [50]. Likewise, the amount of surface area buried in this interface, 2188 $Å^2$, exceeds the threshold value (1712$Å^2$) used to discriminate between a crystal contact and a biologically relevant dimerization interface [51]. However, relative to the other putative dimerization interface, it is formed by the interaction of only 18 residues in each monomer and only three hydrogen bonds. Hence, it is more likely that the larger dimerization interface (Fig. 4A) corresponds to the dimer in solution.

The juxtaposition of these two dimerization interfaces results in filaments of CsoSCA that are packed together to form layers of CsoSCA within the crystal (Fig. 4B). One side of the layer is relatively flat, whereas the other is punctuated by spikes formed by the N- and C-terminal domains (Fig. 4B). Despite the expectation that the contents of the carboxysome are symmetrically organized, this striking arrangement of filaments is probably not reflective of the organization of CsoSCA in the carboxysome. First, only a relatively small number of CsoSCA molecules are encapsulated and would presumably not be localized together. Moreover, the periodicity of the CsoSCA filaments does not match the spacings observed in the layer of the hexameric shell proteins that make up the facets of the carboxysome shell

[15]; it is likely that other proteins mediate its tight interaction with the shell and organize it within the interior of the carboxysome.

A comparison of the structure of CsoSCA with that of other β-CAs not encapsulated within BMCs suggests features of the structure that may be important for its organization and function in the carboxysome. The N-terminal domain is not found in any other β-CA (or any other known protein structure). The C-terminal, vestigial catalytic domain likely has evolved for a carboxysome-specific role. The N- and C-terminal domains are spatially close together (Fig. 4). In addition to the N- and C-terminal domains, CsoSCA also differs from other β-CAs in containing two insertions (residues 223-235 and 244-252) and a large (approximately 40 amino acid) deletion in its catalytic domain. Most of these features map to one side of the CsoSCA assembly (Fig. 4B).

4.2. CcaA and CcmM, two putative β-carboxysome carbonic anhydrases

4.2.1. CcaA

A β-CA, CcaA, has also been identified as a likely component of β-carboxysomes in *Synechococcus* PCC7942 and *Synechocystis* PCC6803 [52-54]; however, not all cyanobacteria contain a CcaA ortholog [7, 55]. The *ccaA* (formerly *icfA*) gene is not found, for example, in the genomes of *Nodularia spumigena*, *Nostoc* PCC7120, *Lyngbya* PCC8106, *Trichodesmium erythraeum* IMS101, *Gloebacter violaceus* PCC77421, *Synechococcus* sp. JA-3-3Ab. Furthermore, this gene is separated from the other carboxysome gene clusters, and its genomic context is not conserved. However, in *Synechococcus* PCC7942 and *Synechocystis* PCC6803 CcaA has been shown to be essential for efficient CO_2 fixation, and *ccaA* mutants require high CO_2 levels for growth [52-54].

The suggestion that CcaA is the CA of β-carboxysomes is based on the fact that certain *hcr* mutants of *Synechococcus* PCC7942 could be complemented by a fragment of genomic DNA that encodes an open reading frame (*icfA*) with significant sequence homology to a β-CA found in chloroplasts and *E. coli* [52, 54]. Subsequent biochemical analysis of carboxysome-enriched fractions obtained from *Synechococcus* PCC7942 provided evidence for the possible co-localization of a CA activity with the microcompartments [56], although characterization of the enzyme was somewhat complicated by the presence of multiple CA-like activities in the preparations. Nevertheless, a DTT-sensitive activity was characterized that was found to be dependent on the presence of 20 mM Mg^{2+}. As expected for all CAs, this activity is sensitive to the sulfonamide ethoxyzolamide. The activity of this enzyme, as estimated by its stimulation of the rate of ^{18}O exchange from $^{13}C^{18}O_2$ to $H_2{}^{16}O$, was suggested to be sufficient to satisfy the requirement of the then prevailing CCM models for a carboxysomal CA activity [56]. A similar analysis performed with the *hcr* mutants of *Synechococcus* PCC7942 that lack functional CcaA revealed only very low levels of the DTT-sensitive CA activity in analogous fractions [56].

The *ccaA* homolog of *Synechocystis* PCC6803 was overexpressed in *E. coli* and the CA activity of the recombinant CcaA protein (rCcaA) determined electrometrically in induced cell extracts [37]. A specific activity of 161 Wilbur-Anderson units per mg protein was reported for the soluble fraction after centrifugation. Antibodies raised against purified soluble rCcaA were used to localize the enzyme in subcellular fractions of *Synechocystis* PCC6803 that contain carboxysomes [37]. In a mutant whose *ccaA* gene was disrupted, fractions of cell-free extracts containing carboxysomes were found to be devoid of CA activity as defined by the isotope exchange assay [53]. CO_2 fixation assays performed with intact wild type and mutant cells revealed lower than wild type maximum catalytic rates and greatly reduced affinities for C_i in the mutant [53]. Since the ability of *Synechococcus* PCC7942 and *Synechocystis* PCC6803 *ccaA* mutants to accumulate C_i is apparently unaffected, the *ccaA* mutations were assumed to interfere with utilization of the cytoplasmic bicarbonate pool, as would be expected for a non-functional carboxysomal CA [52, 53].

In contrast to the α-carboxysomal CsoSCA, amino acid sequence homology clearly indicates that CcaA is a β-CA and most closely related to the plant type (the sequence identity between CcaA and the pea β-CA is 30.7%). The primary structure is highly conserved among CcaA orthologs, with the exception of the C-terminal (approximately 75 amino acid) portion of the protein (Fig. 7A). This region is not found in other β-CAs and appears to be important for catalytic activity and intermolecular interactions [57]. A structural model for residues 36-232 of the *Synechocystis* PCC6803 CcaA based on the structure of the *Pisum sativum* β-CA (PDB code 1EKJ) is shown in Fig. 6. In the *Pisum sativum* β-CA the active site is located in the interface between two monomers [46]. All of the key catalytic residues in the pea β-CA are conserved in CcaA (Figs. 6 and 7A); these include the amino acids for coordinating the Zn^{2+} (Cys 76, Cys 138 and His 135), the Asp in the active site (Asp 78) and the Arg residue that orients it (Arg 80), implying a very similar catalytic mechanism in the two proteins.

4.2.2. CcmM

In contrast to the genomic location of *ccaA*, which is not found in a consistent genomic context in cyanobacteria, the gene encoding the second putative β-carboxysomal CA, *ccmM*, is always found in a gene cluster that encodes several putative carboxysome shell proteins and an additional carboxysome protein, CcmN, that appears to be critical for proper carboxysome formation. CcmM mutants do not contain carboxysomes and have the *hcr* phenotype [20, 58-60], indicating that the protein is important for carboxysome structure and/or biogenesis.

The size of CcmM varies between 539 and 803 amino acids (*Synechococcus* PCC7942 and *Acaryochloris marina* MBIC 11017, respectively). The N-terminal domain is homologous to an archaeal γ-CA [43, 56]. The γ-CAs are found in all three domains of life [61] and use a variety of metals in catalysis. The structure of the γ-CA from *Methanosarcina thermophila* revealed that the fold is primarily a left-handed β-helix in which each turn of the β-helix is approximately 17 amino acids. The primary structure of the turn repeat has been dubbed the bacterial hexapeptide repeat (Pfam 00132). All of the CcmM orthologs contain four to five of these repeats within their N-terminal region. The primary structure of the N-terminal domain (residues 27-192) of CcmM is sufficiently similar to the structure of the *M. thermophila* γ-CA (37.7% identical) to build a homology model (Fig. 7B, C). Key residues for activity in the γ-CA are found in CcmM (Figs. 7B, C and 8).

To date, no direct evidence has been presented to suggest that CcmM possesses CA activity, despite its obvious homology with the γ-CA of *M. thermophila*. Examination of its predicted amino acid sequence tempts speculation that catalysis by its native form, if at all possible, may require complicated interaction(s) with multiple subunits and/or additional carboxysome components. Recombinant CcmM from *Synechocystis* PCC6803 is able to bind bicarbonate (or CO_2) [62], and a recent unpublished result (mentioned in [62]) suggests that carboxysome-enriched fractions from *Nostoc* PCC7120 have CA activity. Since this cyanobacterial species does not contain a *ccaA* homolog, perhaps CcmM provides the CA activity required for the function of the carboxysome as part of the CCM in that bacterium. Clearly, a more detailed analysis of the CA activity associated with β-carboxysomes is needed. Should CcmM act as the carboxysomal CA for *Nostoc* PCC7120, and perhaps other cyanobacteria lacking *ccaA*, the question of its function in *Synechocystis* PCC6803 and *Synechococcus* PCC7942 must be raised, since CcaA seems to be the only CA activity present in the carboxysome of these species [53].

The active form of the archaeal γ-CA is a trimer, and the conservation of the residues for trimer formation in CcmM (Arg 70, Asp 72 and Asp/Glu/His 87) suggests that the γ-CA domain of CcmM also forms a trimer [61]. A trimeric arrangement for CcmM has also been proposed for complexes of CcmM with other carboxysomal components [62, 63]. The active site in archaeal γ-CA is formed at the interface between two monomers in the trimer; the conservation of key residues for Zn^{2+} coordination (His 92, His 119 and His 124) in the equivalent positions in the *Synechocystis* PCC6803 CcmM model (Fig. 7B) suggests that it could likewise function similarly to the archaeal γ-CA. Two of the three additional active site residues in the γ-CA (Glu 62, Glu 84) [64] are strongly conserved in the primary structure of CcmM (Glu 73 and Glu 95) (Fig. 8).

Immuno-blot analysis of carboxysome-enriched fractions from *Synechococcus* PCC7942 and *Synechocystis* PCC6803 suggested that several mass variants of CcmM exist that may have different functions in the β-carboxysome [21, 62, 63]. The C-terminus of CcmM contains three to four repeated sequence elements of approximately 85 amino acids that are similar to the RubisCO small subunit (Pfam 00101) (Fig. 8); these correspond to the portion of the protein in the smaller mass variants. The repeats are separated from the N-terminal γ-CA domain by a large region that is poorly conserved among CcmM proteins (residues 193-245 in *Synechocystis* PCC6803 CcmM) (Fig. 8). The RubisCO small subunit domains themselves are separated from one another by segments of the polypeptide chain that are rich in Ser and Gly.

Recent experiments with a △*ccmM* mutant of *Synechococcus* PCC7942 employed plasmid-encoded, histidine-tagged versions of CcmM to isolate protein assemblages from partially denatured β-carboxysomes by affinity chromatography [63]. In this bacterium, several distinct complexes of RubisCO and full length CcmM (M58) were identified in which CcmM interacts with RubisCO through its C-terminal region. One of the complexes also contains CcaA, which binds to the N-terminal portion of CcmM. Since RubisCO is located in the carboxysome interior, the complexes were suggested to be associated with the internal surface of the β-carboxysome shell [63]. Yeast two-hybrid screens with putative β-carboxysome proteins of *Synechocystis* PCC6803 confirmed the interaction between the C-terminal domain of CcmM and CcaA observed in *Synechococcus* PCC7942 and, through in vitro binding experiments, further revealed complexes that also include CcmN and RubisCO [62]. In both studies the CA was proposed to be oriented so that external HCO_3^- is simultaneously converted to CO_2 as it enters the carboxysome [62, 63] in a manner similar to that proposed for the shell-associated CsoSCA of α-carboxysomes [42]. The stoichiometries of individual protein constituents detectable on immunoblots of these complexes led Cot et al. [62] to propose a model for a multiprotein bicarbonate dehydration complex that is associated with the β-carboxysome shell. CcmM presumably recruits other protein constituents to the side of the carboxysome shell facing the interior during complex assembly. In addition, through its bicarbonate binding ability, the protein may mediate and/or regulate access of its interaction partner, the β-carboxysomal CA CcaA, to the cytosolic C_i pool for subsequent dehydration to CO_2 in the carboxysome interior.

5. **Conclusions and Outlook**

The carboxysomal CAs so far described represent two, possibly three, distinctly different enzyme species that have been recruited to perform seemingly functionally identical roles in two classes of bacterial microcompartments. In the case of CsoSCA, shape conservation and homology of the active site with other β-CAs argue for divergence from a distant common ancestor to an extent that prevented identification of the α-carboxysomal CA from primary sequence comparisons. The proposed β-carboxysomal CA, CccA, on the other hand, while having accumulated some unique structural elements likely required to accommodate proper integration into the carboxysome, is easily identifiable as a β-CA with considerable similarity to plant enzymes. Given these differences between the two carboxysomal CAs, it seems reasonable to assume that their individual recruitment into carboxysomes occurred as two separate instances as opposed to evolution to their present forms after recruitment into an ancestral carboxysome that has since diverged into α- and β-types. This view is supported by the different genomic locations of genes encoding the two CAs: *csoS3* is always a part of the *cso* operon and therefore closely associated with genes encoding the α-carboxysomal RubisCO and shell proteins, while *ccaA* is not clustered with the rest of the β-carboxysome genes and absent altogether from some β-carboxysome-forming cyanobacterial species.

The β-carboxysome protein CcmM is likely functioning as CA in at least those β-carboxysomes that lack CcaA, although no direct evidence exists at present that suggests the protein is enzymatically active. The *ccmM* gene is always clustered with other carboxysomal genes in the genomes of all β-carboxysome forming cyanobacteria thus far studied, and genetic studies have demonstrated its crucial role in CCM function. Recent studies imply a role for CcmM as an important structural component that organizes the molecular position of carboxysome proteins. These studies tempt speculation that in β-carboxysomes lacking CcaA, CcmM has dual catalytic and structural functions, while in those cyanobacteria exemplified by *Synechocystis* PCC6803 and *Synechococcus* PCC7942 the CA function has been taken over by CcaA. However, if CcmM were playing purely an organizational role in these organisms, one would not expect the observed conservation of the active site residues. Elucidation of the advantage these scenarios might impart on β-carboxysome function awaits detailed structural and kinetic studies of these CAs in complex with other carboxysome proteins to fully understand their role in carboxysomal CO_2 fixation.

In all instances so far studied, carboxysomal CAs have been shown to be essential for CCM function. Their generally assumed role is to increase the local concentration of CO_2 around the carboxysomal RubisCO by accelerating the rate of bicarbonate equilibration to CO_2, which is subsequently trapped within the carboxysome interior. However, the dearth of direct experimental CO_2 fixation studies with isolated carboxysomes has made it difficult to precisely define the role of the CA in molecular terms. It has been suggested that other advantages of employing a directional carboxysome shell-bound CA as a “transporter system” may be realized in conjunction with the suggested differential resistance of the shell to the competitive RubisCO inhibitor, O_2. A molecular mechanism that would impart such permselectivity for CO_2 is not obvious in light of our current understanding of protein structure. On the other hand, the diffusional resistance of the carboxysome shell towards CO_2 would presumably also apply to O_2 and, in conjunction with the inward diffusion of bicarbonate and concerted dehydration to CO_2 by CA located within the shell, may effectively block competitive inhibition of the RubisCO-catalyzed carboxylation reaction by oxygen. In the sulfur bacterium *H. neapolitanus* the carboxysome shell has been shown directly to impede diffusion of CO_2 [44]. However, it is unlikely that cytosolic dissolved oxygen concentrations are high enough to significantly inhibit RubisCO in chemoautotrophs. This may not be the case in cyanobacteria undergoing oxygenic photosynthesis, and such differences in intra- and extracellular environments may require carboxysomes with different shell composition and CAs to ensure efficient CO_2 fixation by the carboxysomal RubisCO.

It is hoped that the account of carboxysomal CAs reviewed here will stimulate additional, more detailed and quantitative analyses of structure, catalysis and function of these enzymes to expand and elaborate upon their currently proposed role in the CCM of autotrophic bacteria.

Acknowledgements

We thank Seth Axen for assistance in preparing the figures. The authors' current carboxysome research is supported by National Science Foundation (NSF) grants MCB-0818680 (to GCC and SH) and MCB-0851070 (to GCC, SH and CAK). The authors' previous CA studies were funded by NSF grant MCB-0444568 (to GCC). CAK's work is performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract number DE-AC02-05CH11231, Lawrence Livermore National Laboratory under contract number DE-AC52-07NA27344.

Figure legends

Fig. 1. The CO_2 concentrating mechanism (CCM) of autotrophic bacteria. The model shows the two stages of the CCM: 1. Accumulation of an intracellular bicarbonate pool via energy-driven C_i import; 2. Dehydration of bicarbonate to the RubisCO substrate, CO_2, by the carboxysomal CA. Subsequent fixation of CO_2 onto ribulose-1,5-bisphosphate (RubP) by RubisCO yields two molecules of 3-phosphoglycerate (3-PGA).

Fig. 2. The organization of the genes encoding carboxysomal proteins. Shown are the α-carboxysome operons of *Halothiobacillus neapolitanus* and *Prochlorococcus marinus* and the β-carboxysome gene clusters of *Synechocystis* PCC6803 and *Synechococcus* PCC7942. The parentheses indicate the *csoS1* homolog that is not present in all *Prochlorococcus* strains.

Fig. 3. The proposed role of CsoSCA in α-carboxysomes. The CA is thought to interact with shell components in a way that positions the enzyme to facilitate entry of C_i into the BMC interior through conversion of cytoplasmic bicarbonate to CO_2. The model was adapted from [44].

Fig. 4. Structure of (A) three CsoSCA molecules (PDB code 2FGY) representing the two plausibly biologically relevant CsoSCA dimers observed in the crystal. The three domains of the central CsoSCA molecule are colored differently: The N-terminal domain is shown in red, the catalytic domain is shown in dark green and the C-terminal domain is blue. The interaction between the central CsoSCA molecule and the one in cyan (circled) is most likely

the biologically relevant dimer. The third molecule, shown in yellow, also has a substantial interface with the central CsoSCA molecule. The sidechains coordinating the Zn ion (shown as a black sphere) and of Asp 175 and Arg 177 are also shown in sticks. (B) The arrangement of filaments into layers of CsoSCA observed in the crystals. Eighteen molecules of CsoSCA are shown, in two views related by a 90° rotation in the plane of the page. Three CsoSCA molecules are colored as in (A); in the others, the N-terminal domain is shown in red, the C-terminal domain is colored forest green. The loops corresponding to insertions found only in the primary structure of carboxysomal β-CAs (see text) are shown in black. This Fig. and Fig. 7 were prepared with Pymol [65].

Fig. 5. Sequence logo for CsoSCA orthologs**.** Residues are numbered according to alignment with the *H. neapolitanus* CsoSCA. The height of each letter is proportional to the observed frequency of the amino acid in that position. The overall height of each stack of letters is proportional to sequence conservation as measured in bits [66]. A sample correction for underestimation of entropy was included in the generation of the Logo. The amino acids are colored according to their chemical properties: polar, green; basic, blue; acidic, red; hydrophobic, black. Regions of very low sequence conservation/indels are shown as gray blocks; the numbers indicate the average length of this region. The three domains are underlined and colored as in Fig. 4. Key residues discussed in the text are marked with asterisks. This Fig. and Figs. 6 and 8 were prepared with WEBLOGO [67].

Fig. 6. Sequence logo for CcaA orthologs**.** Residues are numbered according to alignment with the *Synechocystis* PCC6803 CcaA. Residues expected to be important for catalysis are

marked with asterisks. The portion underlined in green corresponds to the model shown in Fig. 7A. Logo parameters as in Fig. 5.

Fig. 7. (A) A model of the *Synechocystis* PCC 6803 CcaA dimer structure (residues 36-232) based on the β-CA from *Pisum sativum* (PDB code 1EKJ). The residues presumed to coordinate the Zn ion (shown as a red sphere) are shown in sticks and colored yellow. Asp 76 and Arg 78 are also shown in sticks. The N- and C-termini are labeled. Absolutely conserved residues are colored in blue. The dimer is completed by a second monomer from the *P. sativum* structure (shown in gray). (B) and (C) Model for the N-terminal γ-CA domain of the *Synechocystis* PCC 6803 CcmM (residues 27-192) based on the structure of the γ-CA from *Methanosarcina thermophila* (PDB Code 1THJ). Residues corresponding to the catalytic residues in the *M. thermophila* structure are shown in sticks. The trimer is completed by two additional molecules from the *M. thermophila* structure (shown in gray). The N- and C-termini are labeled. The view in (C) is rotated ~90° relative to that in (B).

Fig. 8. Sequence logo for CcmM orthologs. Residues are numbered according to alignment with the *Synechocystis* PCC6803 CcmM. Residues expected to be important for catalysis are marked with asterisks. The portion underlined in green corresponds to the model shown in Fig. 7B, C. Logo parameters as in Fig. 5.

References

[1] J.M. Shively (Ed.), Inclusions in Prokaryotes, Springer-Verlag, Berlin Heidelberg, 2006.

[2] J.M. Shively (Ed.), Complex Intracellular Structures in Prokaryotes, Springer, Berlin/Heidelberg, 2006.

[3] J.M. Shively, F. Ball, D.H. Brown, R.E. Saunders, Functional organelles in prokaryotes: Polyhedral inclusions (carboxysomes) of *Thiobacillus neapolitanus*, Science 182 (1973) 584-586.

[4] S. Heinhorst, G.C. Cannon, J.M. Shively, Carboxysomes and carboxysome-like inclusions, in: J.M. Shively (Ed.), Complex Intracellular Structures in Prokaryotes, vol. 2, Springer, Berlin/Heidelberg, 2006, pp. 141-164.

[5] T.O. Yeates, C.A. Kerfeld, S. Heinhorst, G.C. Cannon, J.M. Shively, Protein-based organelles in bacteria: carboxysomes and related microcompartments, Nat. Rev. Microbiol. 6 (2008) 681-691.

[6] G.D. Price, M.R. Badger, F.J. Woodger, B.M. Long, Advances in understanding the cyanobacterial CO_2-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants, J. Exp. Bot. 59 (2008) 1441-1461.

[7] A.K.-C. So, G.S. Espie, Cyanobacterial carbonic anhydrases, Can. J. Bot. 83 (2005) 721-734.

[8] A.K.-C. So, G.S. Espie, E.B. Williams, J.M. Shively, S. Heinhorst, G.C. Cannon, A novel evolutionary lineage of carbonic anhydrase (epsilon class) is a component of the carboxysome shell, J. Bacteriol. 186 (2004) 623-630.

[9] M.R. Badger, D. Hanson, G.D. Price, Evolution and diversity of CO_2 concentrating mechanisms in cyanobacteria, Funct. Plant Biol. 29 (2002) 161-173.

[10] G.C. Cannon, S. Heinhorst, C.E. Bradburne, J.M. Shively, Carboxysome genomics: a status report, Funct. Plant Biol. 29 (2002) 175-182.

[11] M.R. Badger, G.D. Price, CO_2 concentrating mechanisms in cyanobacteria: Molecular components, their diversity and evolution, J. Exp. Bot. 54 (2003) 609-622.

[12] B.M. Long, G.D. Price, M.R. Badger, Proteomic assessment of an established technique for carboxysome enrichment from *Synechococcus* PCC7942, Can J Bot 83 (2005) 746-757.

[13] C.A. Kerfeld, M.R. Sawaya, S. Tanaka, C.V. Nguyen, M. Phillips, M. Beeby, T.O. Yeates, Protein structures forming the shell of primitive bacterial organelles, Science 309 (2005) 936-938.

[14] M.G. Klein, P. Zwart, S.C. Bagby, F. Cai, S.W. Chisholm, S. Heinhorst, G.C. Cannon, C.A. Kerfeld, Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport, J. Mol. Biol. 392 (2009) 319-333.

[15] Y. Tsai, M.R. Sawaya, G.C. Cannon, F. Cai, E.B. Williams, S. Heinhorst, C.A. Kerfeld, T.O. Yeates, Structural analysis of CsoS1A and the protein shell of the *Halothiobacillus neapolitanus* carboxysome, PLoS Biol. 5 (2007) e144.

[16] F. Cai, S. Heinhorst, J. Shively, G. Cannon, Transcript analysis of the *Halothiobacillus neapolitanus cso* operon, Arch. Microbiol. 189 (2008) 141-150.

[17] S. Tanaka, C.A. Kerfeld, M.R. Sawaya, F. Cai, S. Heinhorst, G.C. Cannon, T.O. Yeates, Atomic-level models of the bacterial carboxysome shell, Science 319 (2008) 1083-1086.

[18] S.H. Baker, S.C. Lorbach, M. Rodriguez-Buey, D.S. Williams, H.C. Aldrich, J.M. Shively, The correlation of the gene *csoS2* of the carboxysome operon with two polypeptides of the carboxysome in *Thiobacillus neapolitanus*, Arch. Microbiol. 172 (1999) 233-239.

[19] G.C. Cannon, C.E. Bradburne, H.C. Aldrich, S.H. Baker, S. Heinhorst, J.M. Shively, Microcompartments in prokaryotes: carboxysomes and related polyhedra, Appl. Environ. Microbiol. 67 (2001) 5351-5361.

[20] M. Ludwig, D. Sültemeyer, G.D. Price, Isolation of *ccmKLMN* genes from the marine cyanobacterium *Synechococcus* sp. PCC7002 and evidence that CcmM is essential for carboxysome assembly, J. Phycol. 36 (2000) 1109-1118.

[21] G.D. Price, D. Sültemeyer, B. Klughammer, M. Ludwig, M.R. Badger, The functioning of the CO_2 concentrating mechanism in several cyanobacterial strains: a review of general physiological characteristics, genes, proteins and recent advances, Can. J. Bot. 76 (1998) 973-1002.

[22] F.R. Tabita, Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A different perspective, Photosynth. Res. 60 (1999) 1-28.

[23] M.R. Badger, G.D. Price, The CO_2 concentrating mechanism in cyanobacteria and microalgae, Physiol. Plant. 90 (1992) 529-536.

[24] J.A. Raven, W.J. Lucas, The energetics of carbon acquisition, in: W.J. Lucas, J.A. Berry (Eds.), Inorganic Carbon Uptake by Aquatic Photosynthetic Organisms, American Society of Plant Physiologists, Rockville, MD, 1985, pp. 305-324.

[25] M.R. Badger, G.D. Price, B.M. Long, F.J. Woodger, The environmental plasticity and ecological genomics of the cyanobacterial CO_2 concentrating mechanism, J. Exp. Bot.

57 (2006) 249-265.

[26] A. Kaplan, L. Reinhold, CO_2 concentrating mechanisms in photosynthetic microorganisms, Annu. Rev. Plant Physiol. Plant Mol. Biol. 50 (1999) 539-570.

[27] G.D. Price, M.R. Badger, Expression of human carbonic anhydrase in the cyanobacterium *Synechococcus* PCC7942 creates a high CO_2-requiring phenotype: Evidence for a central role for carboxysomes in the CO_2 concentrating mechanism, Plant Physiol. 91 (1989) 505-513.

[28] M.R. Badger, M. Bassett, H.N. Comins, A model for HCO_3^- accumulation and photosynthesis in the cyanobacterium *Synechococcus* sp: Theoretical predictions and experimental observations, Plant Physiol. 77 (1985) 465-471.

[29] L. Reinhold, R. Kosloff, A. Kaplan, A model for inorganic carbon fluxes and photosynthesis in cyanobacteria., Can. J. Bot. 69 (1991) 984-988.

[30] L. Reinhold, M. Zviman, A. Kaplan, Inorganic carbon fluxes and photosynthesis in cyanobacteria. A quantitative model, in: J. Biggens (Ed.), Progress in Photosynthetic Research, vol. 4, Martinus Nijhoff, Dordrecht, 1987, pp. 6.289-286.296.

[31] L. Reinhold, M. Zviman, A. Kaplan, A quantitative model for carbon fluxes and photosynthesis in cyanobacteria, Plant Physiol. Biochem. 27 (1989) 945-954.

[32] Y.A. Holthuijzen, F.F.M. Dissel-Emiliani, J.G. Kuenen, W.N. Konings, Energetic aspects of CO_2 uptake in *Thiobacillus neapolitanus*, Archives of Microbiology 147 (1987) 285-290.

[33] K.P. Dobrinski, D.L. Longo, K.M. Scott, The carbon-concentrating mechanism of the hydrothermal vent chemolithoautotroph *Thiomicrospira crunogena*, J. Bacteriol. 187 (2005) 5761-5766.

[34] A. Kaplan, M.R. Badger, J.A. Berry, Photosynthesis and the intracellular inorganic carbon pool in the bluegreen alga *Anabaena variabilis*: Response to external CO_2 concentration, Planta 149 (1980) 219-226.

[35] M.R. Badger, G.D. Price, Carbonic anhydrase activity associated with the cyanobacterium *Synechococcus* PCC 7942, Plant Physiol. 89 (1989) 51-60.

[36] G.C. Cannon, J.M. Shively, Characterization of a homogenous preparation of carboxysomes from *Thiobacillus neapolitanus*, Arch. Microbiol. 134 (1983) 52-59.

[37] A.K.-C. So, G.S. Espie, Cloning, characterization and expression of carbonic anhydrase from the cyanobacterium *Synechocystis* PCC6803, Plant Mol Biol 37 (1998) 205-215.

[38] S.H. Baker, D.S. Williams, H.C. Aldrich, A.C. Gambrell, J.M. Shively, Identification and localization of the carboxysome peptide CsoS3 and its corresponding gene in *Thiobacillus neapolitanus*, Arch. Microbiol. 173 (2000) 278-283.

[39] D.N. Silverman, Carbonic anhydrase: Oxygen-18 exchange catalyzed by an enzyme with rate-contributing proton-transfer steps, in: D.L. Purich (Ed.), Methods in Enzymology, vol. 87, 1982, pp. 732-752.

[40] K.M. Wilbur, N.G. Anderson, Electrometric and colorimetric determination of carbonic anhydrase, J. Biol. Chem. 176 (1948) 147-154.

[41] R.G. Khalifah, The carbon dioxide hydration activity of carbonic anhydrase. I. Stop-flow kinetic studies on the native human isoenzymes B and C, J. Biol. Chem. 246 (1971) 2561-2573.

[42] S. Heinhorst, E.B. Williams, F. Cai, C.D. Murin, J.M. Shively, G.C. Cannon, Characterization of the carboxysomal carbonic anhydrase CsoSCA from

Halothiobacillus neapolitanus., J. Bacteriol. 188 (2006) 8087-8094.

[43] B.E. Alber, J.G. Ferry, Characterization of heterologously produced carbonic anhydrase from *Methanosarcina thermophila*, J. Bacteriol. 178 (1996) 3270-3274.

[44] Z. Dou, S. Heinhorst, E.B. Williams, C.D. Murin, J.M. Shively, G.C. Cannon, CO_2 fixation kinetics of *Halothiobacillus neapolitanus* mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO_2, J. Biol. Chem. 283 (2008) 10377-10384.

[45] M.R. Sawaya, G.C. Cannon, S. Heinhorst, S. Tanaka, E.B. Williams, T.O. Yeates, C.A. Kerfeld, The structure of beta-carbonic anhydrase from the carboxysomal shell reveals a distinct subclass with one active site for the price of two, J. Biol. Chem. 281 (2006) 7546-7555.

[46] M.S. Kimber, E.F. Pai, The active site architecture of *Pisum sativum* beta-carbonic anhydrase is a mirror image of that of alpha-carbonic anhydrases, Embo J. 19 (2000) 1407-1418.

[47] Y. Xu, L. Feng, P.D. Jeffrey, Y. Shi, F.M.M. Morel, Structure and metal exchange in the cadmium carbonic anhydrase of marine diatoms, Nature 452 (2008) 56-61.

[48] J. Kraut, J.D. Robertus, J.J. Birktoft, R.A. Alden, P.E. Wilcox, J.C. Powers, The aromatic substrate binding site in subtilisin BPN' and its resemblance to chymotrypsin, Cold Spring Harbor Symp. Quant. Biol. 36 (1971) 117-123.

[49] B. Matthews, S. Remington, M. Gruter, W. Anderson, Relation between hen egg white lysozyme and bacteriophage T4 lysozyme: evolutionary implications, J. Mol. Biol. 147 (1981) 545-558.

[50] M.C. Lawrence, P.M. Colman, Shape complementarity at protein/protein interfaces, J.

Mol. Biol. 234 (1993) 946-950.

[51] H. Ponstingl, K. Henrick, J.M. Thornton, Discriminating between homodimeric and monomeric proteins in the crystalline state, Proteins 41 (2000) 47-57.

[52] H. Fukuzawa, E. Suzuki, Y. Komukai, S. Miyachi, A gene homologous to chloroplast carbonic anhydrase (*icfA*) is essential to photosynthetic carbon dioxide fixation by *Synechococcus* PCC7942, Proc. Natl. Acad. Sci. USA 89 (1992) 4437-4441.

[53] A.K.-C. So, M. John-McKay, G.S. Espie, Characterization of a mutant lacking carboxysomal carbonic anhydrase from the cyanobacterium *Synechocystis* PCC6803, Planta 214 (2002) 456-467.

[54] J.-W. Yu, G.D. Price, L. Song, M.R. Badger, Isolation of a putative carboxysomal carbonic anhydrase gene from the cyanobacterium *Synechococcus* PCC7942, Plant Physiol. 100 (1992) 794-800.

[55] M.R. Badger, The roles of carbonic anhydrases in photosynthetic CO_2 concentrating mechanisms, Photosynth. Res. 77 (2003) 83-94.

[56] G.D. Price, J.R. Coleman, M.R. Badger, Association of carbonic anhydrase activity with carboxysomes isolated from the cyanobacterium *Synechococcus* PCC7942, Plant Physiol. 100 (1992) 784-793.

[57] A.K.-C. So, S.S.-W. Cot, G.S. Espie, Characterization of the C-terminal extension of the carboxysomal carbonic anhydrase from *Synechocystis* sp. PCC6803, Funct. Plant Biol. 29 (2002) 183-194.

[58] S. Berry, J.H. Fischer, J. Kruip, M. Hauser, G.F. Wildner, Monitoring cytosolic pH of carboxysome-deficient cells of *Synechocystis* sp. PCC 6803 using fluorescence analysis, Plant Biol (Stuttg) 7 (2005) 342-347.

[59] D. Emlyn-Jones, F.J. Woodger, T.J. Andrews, G.D. Price, S.M. Whitney, A *Synechococcus* PCC 7942 Δ*ccmM* (cyanophyceae) mutant pseudoreverts to air growth without regaining carboxysomes, J. Phycol. 42 (2006) 769-777.

[60] F.J. Woodger, M.R. Badger, G.D. Price, Sensing of inorganic carbon limitation in *Synechococcus* PCC 7942 is correlated with the size of the internal inorganic carbon pool and involves oxygen, Plant Physiol. 139 (2005) 1959-1969.

[61] C. Kisker, H. Schindelin, B.E. Alber, J.G. Ferry, D.C. Rees, A left-hand beta-helix revealed by the crystal structure of a carbonic anhydrase from the archaeon *Methanosarcina thermophila*, Embo J. 15 (1996) 2323-2330.

[62] S.S.-W. Cot, A.K.-C. So, G.S. Espie, A multiprotein bicarbonate dehydration complex essential to carboxysome function in cyanobacteria, J. Bacteriol. 190 (2008) 936-945.

[63] B.M. Long, M.R. Badger, S.M. Whitney, G.D. Price, Analysis of carboxysomes from *Synechococcus* PCC7942 reveals multiple RubisCO complexes with carboxysomal proteins CcmM and CcaA, J. Biol. Chem. 282 (2007) 29323-29335.

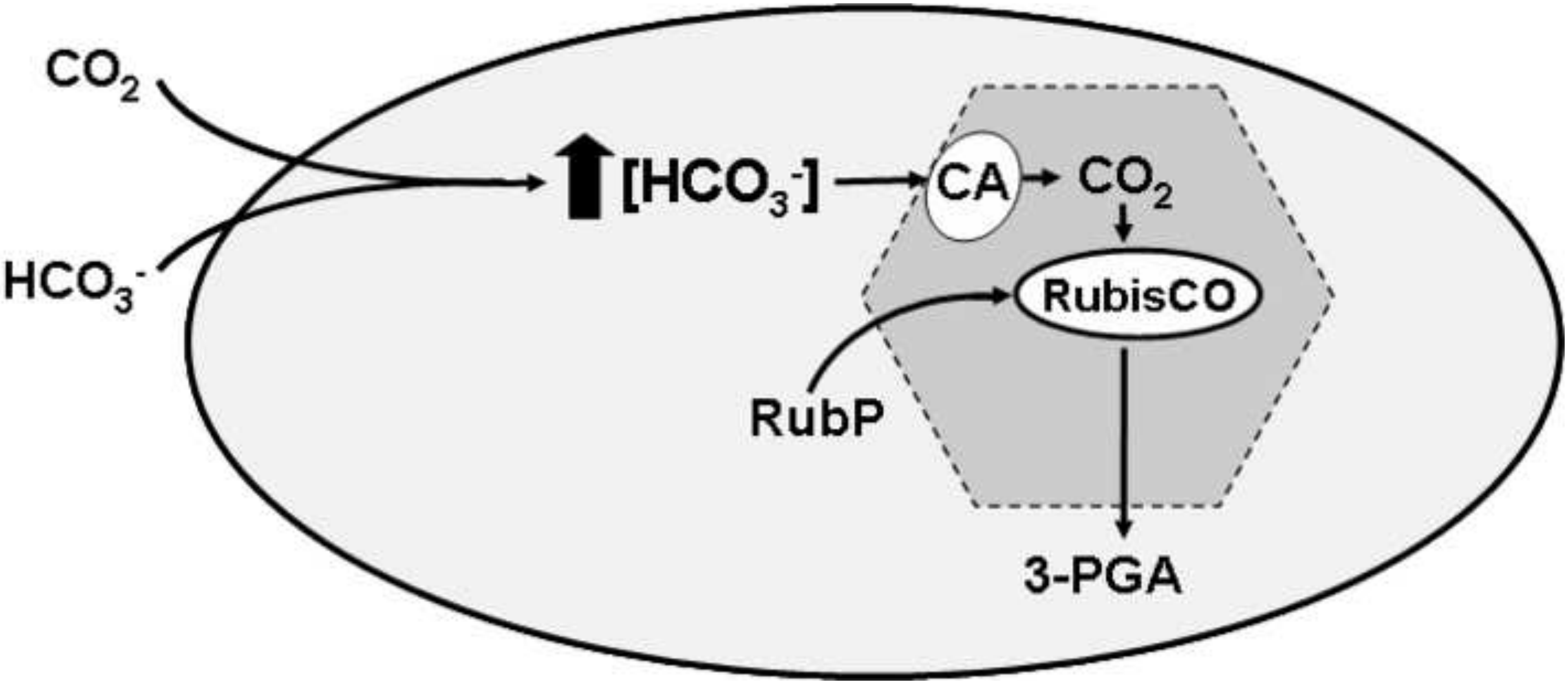
[64] T.M. Iverson, B.E. Alber, C. Kisker, J.G. Ferry, D.C. Rees, A closer look at the active site of γ-class carbonic anhydrases: high resolution crystallographic studies of the carbonic anhydrase from *Methanosarcina thermophila*, Biochemistry 39 (2000) 9222-9231.

[65] W.L. DeLano, DeLano Scientific, Palo Alto, CA, 2002.

[66] T.D. Schneider, R.M. Stephens, Sequence logos: a new way to display consensus sequences, Nucl. Acids Res. 18 (1990) 6097-6100.

[67] G.E. Crooks, G. Hon, J.-M. Chandonia, S.E. Brenner, WEBLOG: A sequence logo generator, Genome Res. 14 (2004) 1188-1190.

Figure
Click here to download high resolution image



CO_2
HCO_3^-
$[HCO_3^-]$
CA
CO_2
RubisCO
RubP
3-PGA

Figure
Click here to download high resolution image

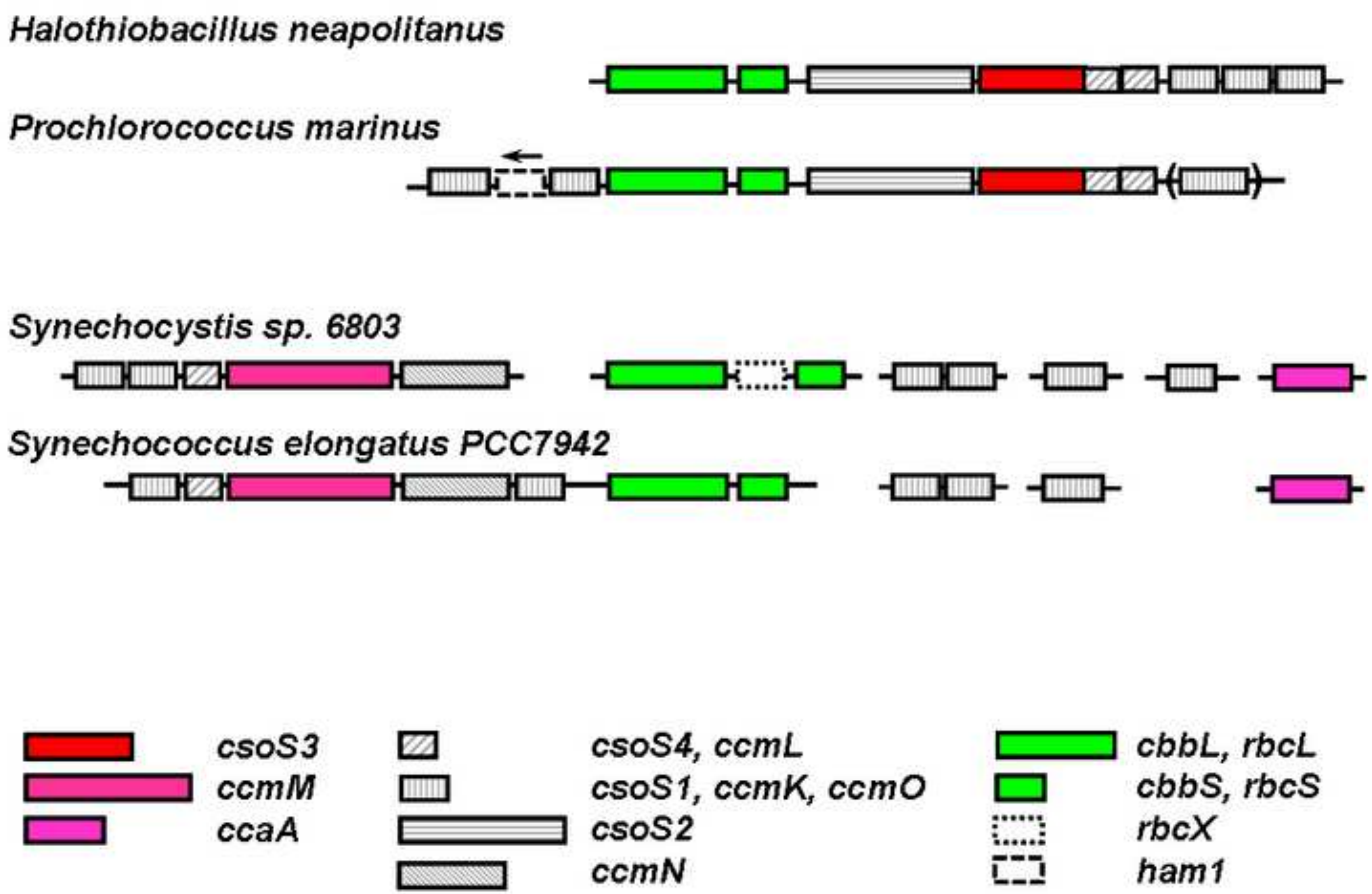

Halothiobacillus neapolitanus
Prochlorococcus marinus
Synechocystis sp. 6803
Synechococcus elongatus PCC7942
csoS3
ccmM
ccaA
csoS4, ccmL
csoS1, ccmK, ccmO
csoS2
ccmN
cbbL, rbcL
cbbS, rbcS
rbcX
ham1

Figure
Click here to download high resolution image

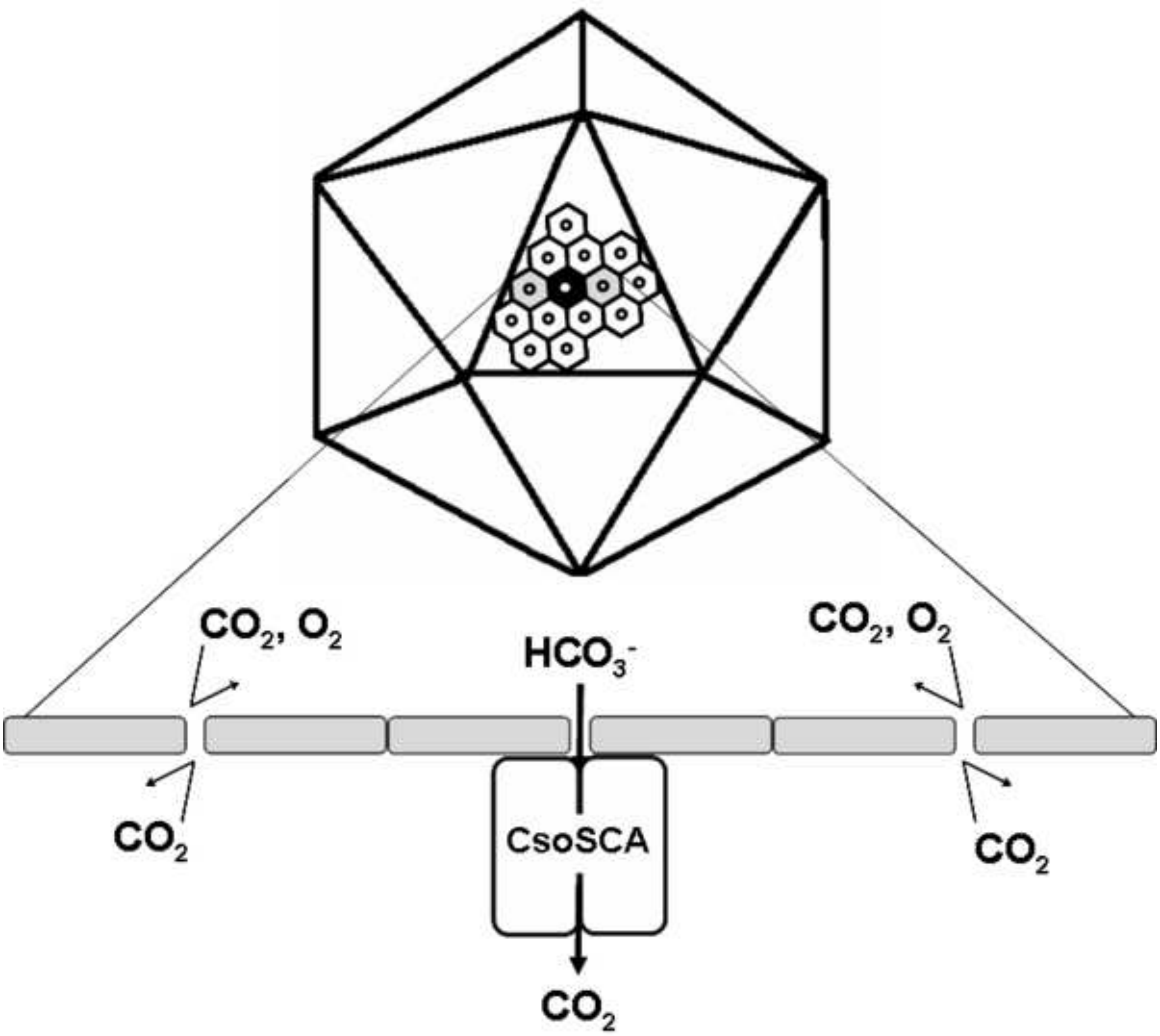

CO2, O2
CO2, O2
HCO3-
CO2
CO2
CsoSCA
CO2

Figure
Click here to download high resolution image

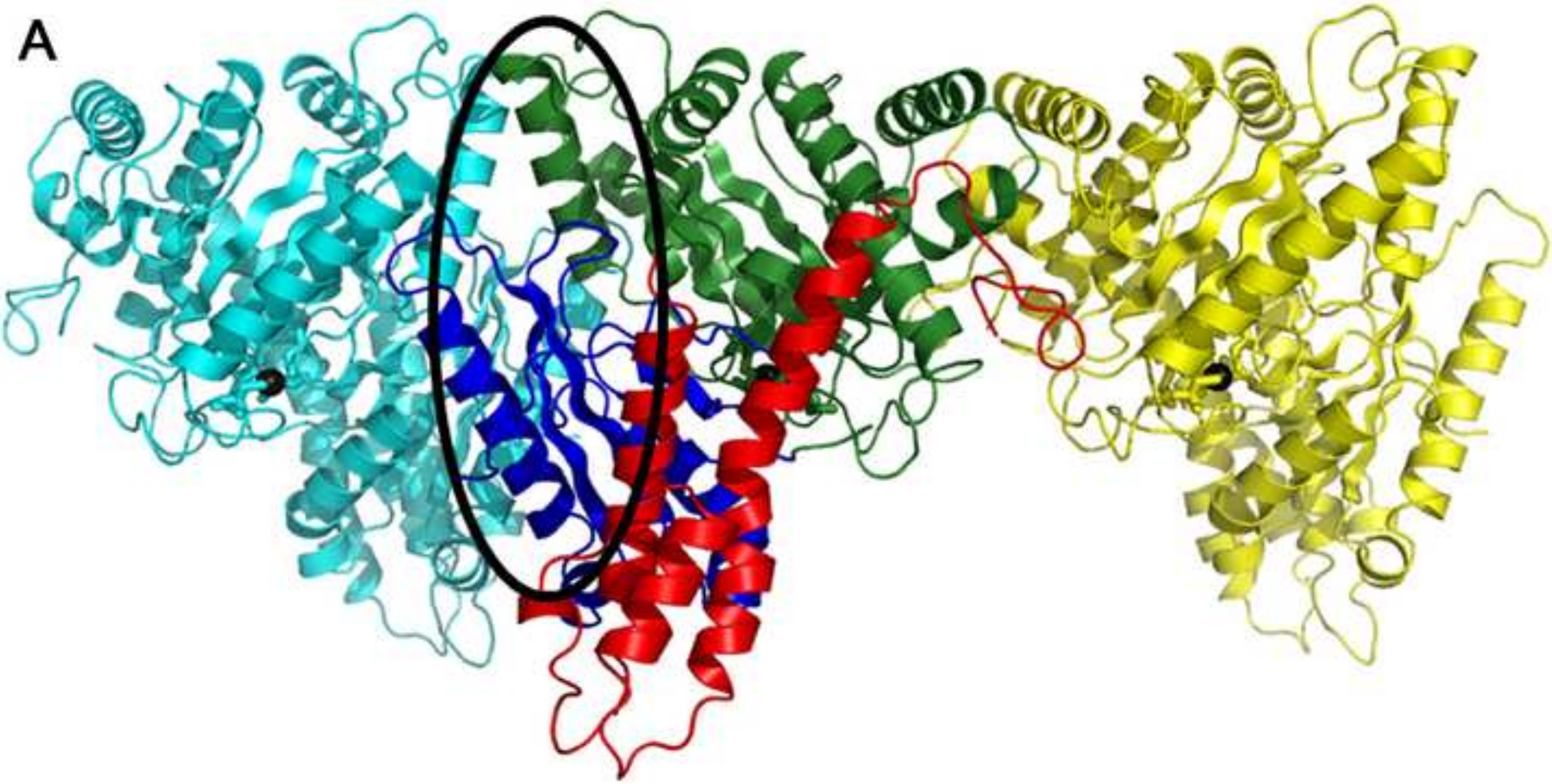
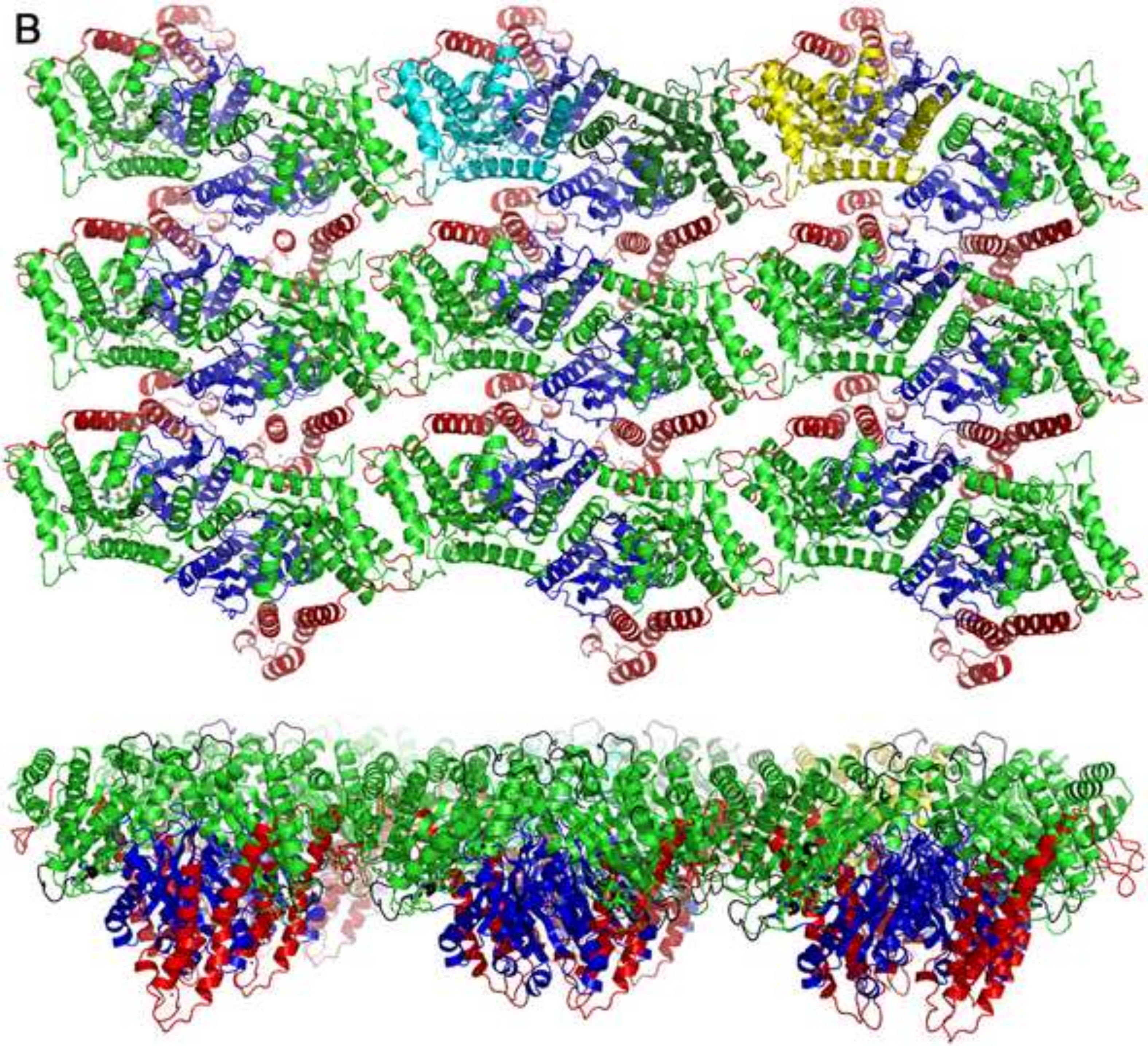
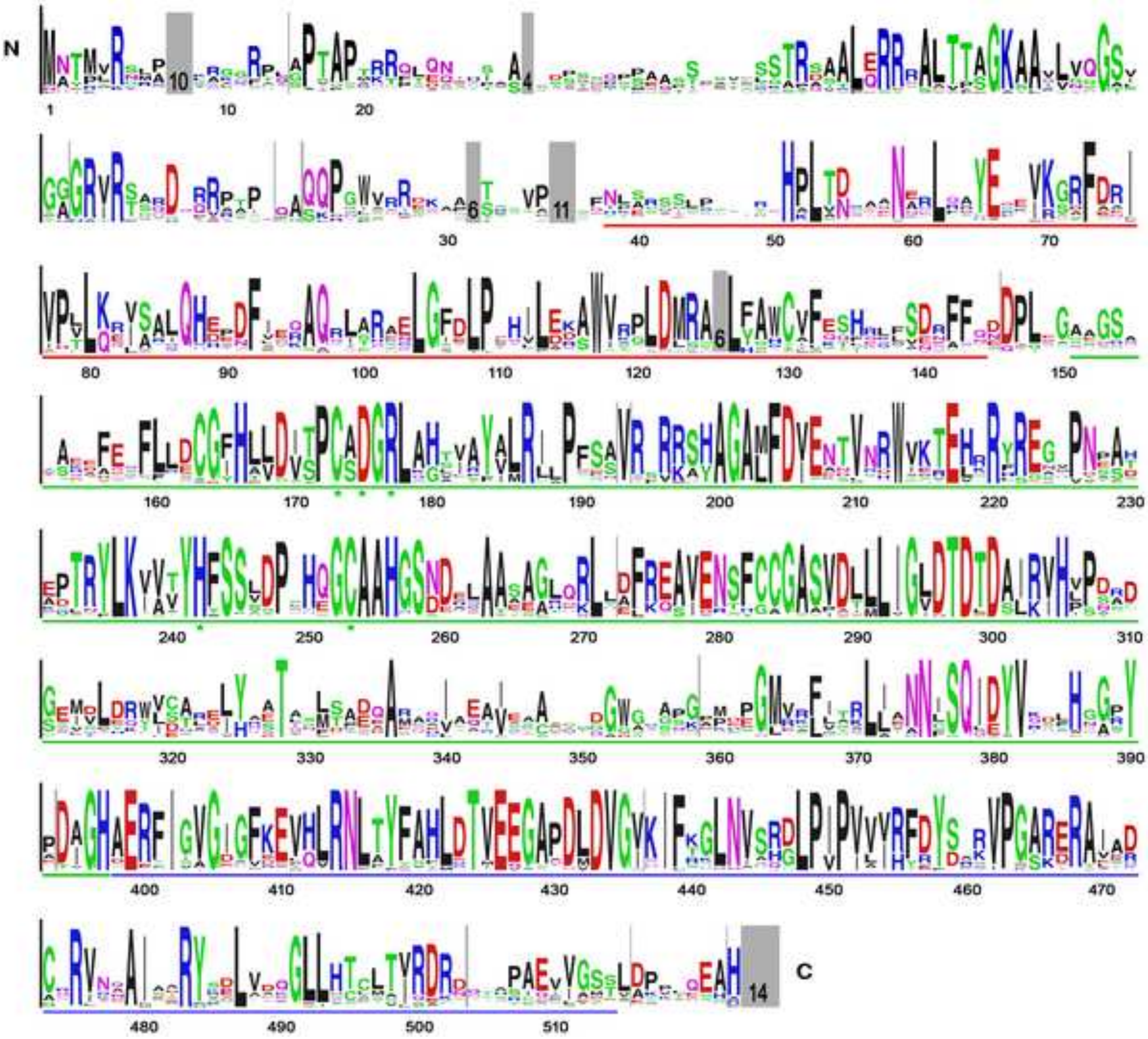


Figure
Click here to download high resolution image



Figure

Click here to download high resolution image


N
C
1
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
230
240
250
260
270
280
290
300
310
320
330
340
350
360
370
380
390
400
410
420
430
440
450
460
470
480
490
500
510
10
4
6
11
6
14

Figure

Click here to download high resolution image

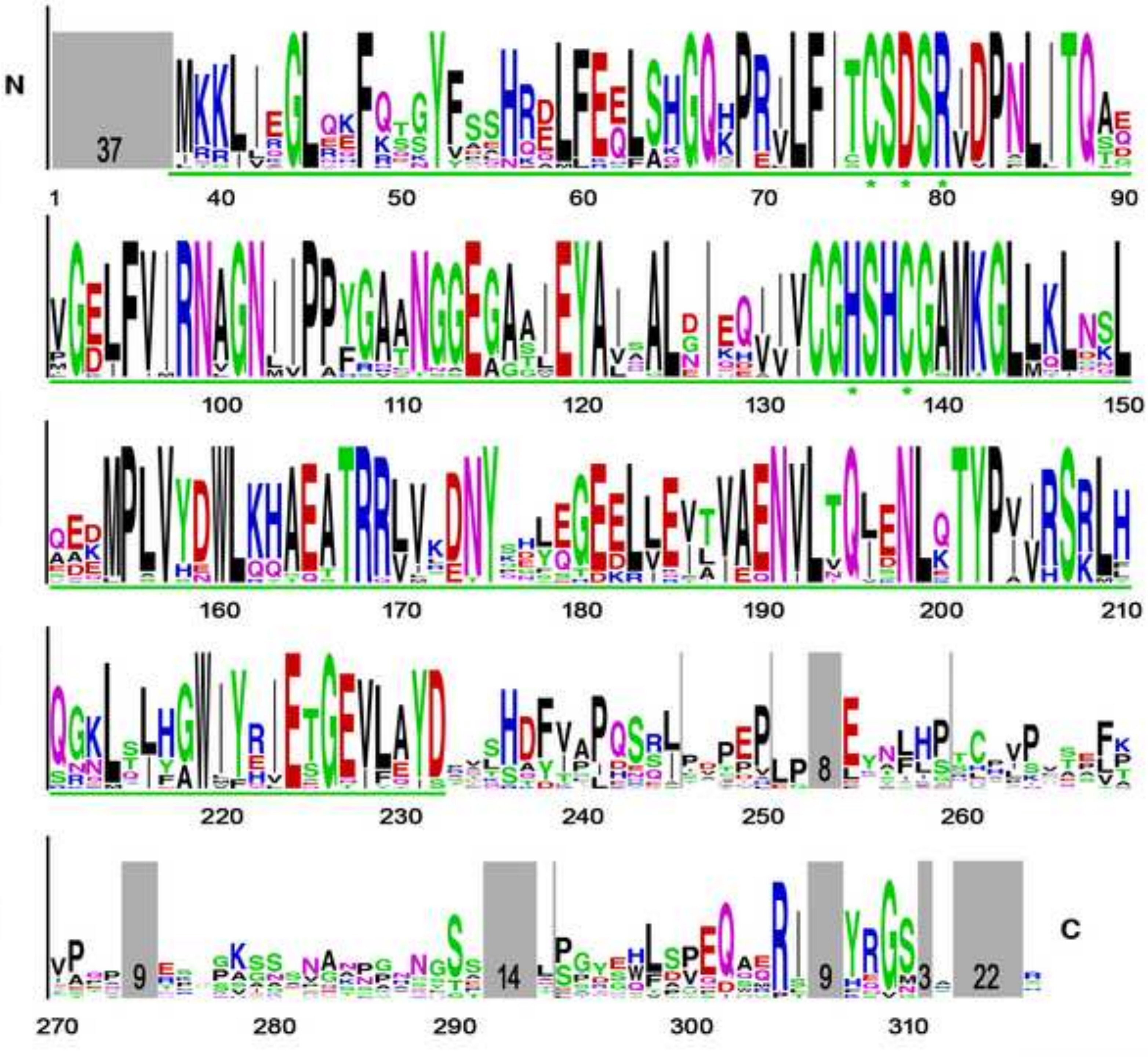
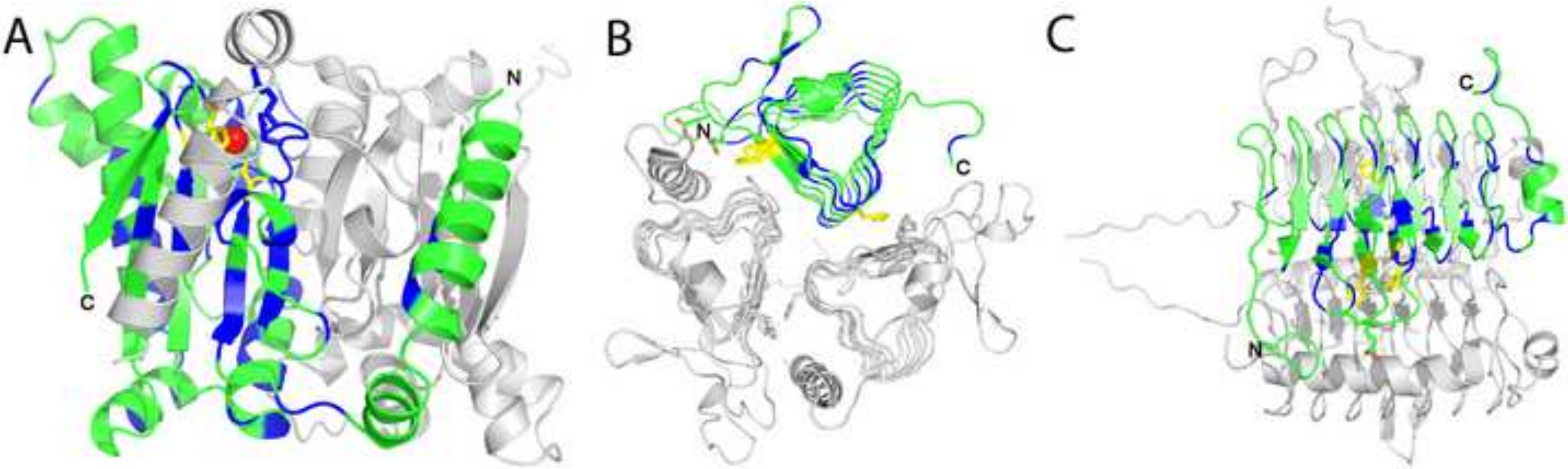
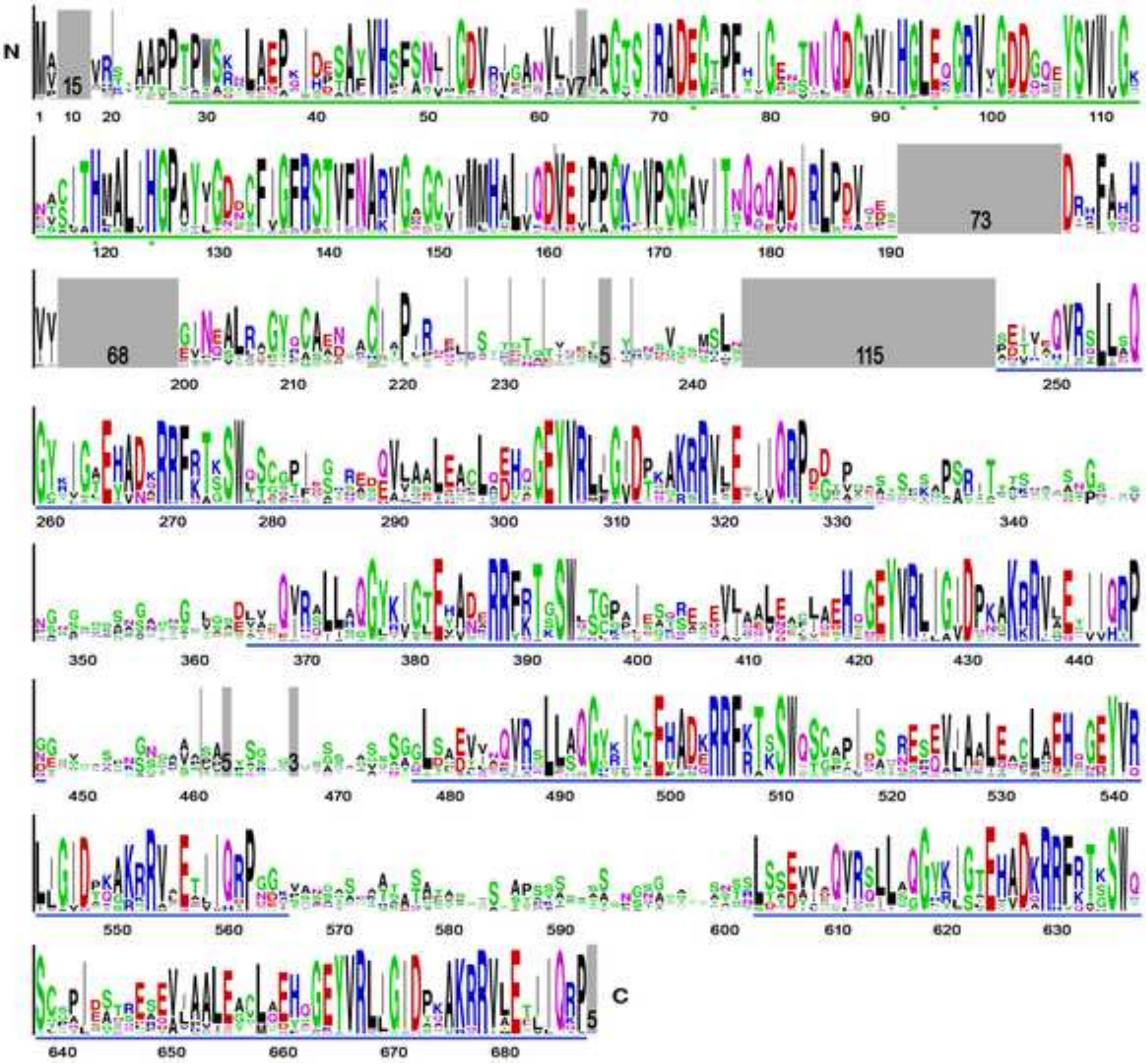


Figure
Click here to download high resolution image



Figure

Click here to download high resolution image



DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.