Many bacteria contain polyhedral protein microcompartments that are bounded by a 3- to 4-nm shell composed of several highly conserved polypeptides. However, the enzyme content of these structures varies depending on metabolic function and the bacterial strain in which they occur. For example, microcompartments found in a Salmonella sp. grown on propanediol contain propanediol dehydratase, while similar microcompartments appearing in Escherichia coli grown on ethanolamine contain ammonia lyase (12, 16). The widespread occurrence of such polyhedral bodies has led to the suggestion that their protein shell provides a catalytic advantage to the enzyme within.

The archetypal example of a prokaryotic microcompartment is the carboxysome, which is found in all cyanobacteria and many chemolithoautotrophic bacteria. Carboxysomes from the chemolithoautotroph Halothiobacillus neapolitanus (formerly Thiobacillus neapolitanus) measure 100 to 120 nm across and are filled with the CO₂-fixing enzyme of the Calvin-Benson-Bassham cycle, ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), most if not all of which is packaged in protein microcompartments called carboxysomes. These structures play an integral role in a cellular CO₂-concentrating mechanism and are essential components for autotrophic growth. Here we report that the carboxysomal shell protein, CsoS3, from Halothiobacillus neapolitanus is a novel carbonic anhydrase (ε-class CA) that has an evolutionary lineage distinct from those previously recognized in animals, plants, and other prokaryotes. Functional CAs encoded by csoS3 homologues were also identified in the cyanobacteria Prochlorococcus sp. and Synechococcus sp., which dominate the oligotrophic oceans and are major contributors to primary productivity. The location of the carboxysomal CA in the shell suggests that it could supply the active sites of RuBisCO in the carboxysome with the high concentrations of CO₂ necessary for optimal RuBisCO activity and efficient carbon fixation in these prokaryotes, which are important contributors to the global carbon cycle.
nobilateral strains (22, 31, 32). The enzyme has been shown to be an essential component of a cellular CO₂-concentrating mechanism (CCM) that is necessary for autotrophic growth under normal conditions (23, 32, 37), although its localization as an integral carboxysome component is still unclear. This CA (CcaA) is a member of the β class, one of three independent evolutionary lineages (α, β, and γ) that are widespread among the domains Bacteria, Archaea, and Eucarya (13). It was therefore surprising to find that the genomes of several prokaryotes encoding cso-carboxysomes lack genes for any identifiable member of the α, β, or γ class of CAs (10, 26).

In this study we demonstrate unequivocally that CsoS3, from the chemolithoautotroph H. neapolitanus, represents a new lineage of CA (ε class) and is a catalytically functional component of the shell of cso-carboxysomes. Homologues of CsoS3 from the marine cyanobacteria Prochlorococcus sp. and Synechococcus sp. are also ε-class CAs that likely reside in the carboxysome shell. The shell-localized CA is proposed to supply CO₂ to the active sites of RuBisCO in the carboxysome to support optimal CO₂ fixation activity.

MATERIALS AND METHODS

Carboxysome purification. H. neapolitanus was grown in a 3-liter chemostat at a dilution rate of 0.08 h⁻¹ as previously described (9). Cells (6 to 8 g) were resuspended in 50 ml of TEMB (5 mM Tris-HCl [pH 8.0], 1 mM EDTA, 10 mM MgCl₂, 20 mM NaHCO₃) before they were ruptured by sonication and centrifuged (12,000 × g, 20 min). The supernatant was collected and centrifuged at 40,000 × g for 30 min. The resulting pellet was resuspended in 20 ml of a 33% (vol/vol) solution of CelLytic B-II (Sigma) in TEMB and centrifuged (40,000 × g, 30 min). The carboxysome-enriched pellet was resuspended in 3 ml of TEMB, clarified by centrifugation (3,000 × g, 1 min), and loaded onto 10 to 50% (wt/vol) linear sucrose density gradients. After centrifugation (105,000 × g, 30 min) in a Beckman JS4.3 rotor, the gradients were fractionated and the prominent carboxysome band near the middle of the gradient was collected and recentrifuged at 100,000 × g for 90 min. The resulting pellet of carboxysomes was resuspended in 1 ml of TEMB before RuthBioCO (19) and CA (see below) activities were assayed. Carboxysomes were negatively stained in 2% (wt/vol) uranyl acetate on carbon-coated grids and were observed with a Philips EM 300 electron microscope. Thin-sections of H. neapolitanus cells were also prepared and viewed. Purified carboxysomes were disrupted by freezing pelleted carboxysomes at −20°C for 30 min. The pellet was subsequently thawed, resuspended in 100 to 200 μl of TEMB at 4°C, and subjected to centrifugation at 14,000 × g for 30 min. The resulting pellet consisted mainly of carboxysome shell “ghosts,” while the supernatant contained most of the RuBisCO that originally had been packaged in the intact particles.

CsoS3 expression in E. coli. Recombinant CsoS3 was produced in E. coli by using several different expression systems. An existing clone containing pcos33ProEx (4) was employed to express a histidine-tagged CsoS3 protein. All other constructs were generated by PCR amplification of the csoS3 gene using the primers shown in Table 1. The nucleotide sequences, orientations, and reading frames of all cloned fragments were verified prior to protein expression analyses. Clones producing T7-tagged CsoS3 were constructed by amplification of the csoS3 gene from H. neapolitanus genomic DNA, subsequent digestion with EcoRI, ligation into pET-21b (Novagen), and transformation of E. coli BL21(DE3) cells with the resulting expression construct. Homologues of csoS3 were similarly amplified from genomic DNA preparations of Prochlorococcus marinus strains MED4 and MIT9313 and Synechococcus strain WH8102, which were kindly provided by Sallie W. Chisholm (Massachusetts Institute of Technology) and Bianca Brambhasha ( Scripps Institution of Oceanography). The amplified fragments were restricted with HindIII and ligated into the corresponding site in pET-21b, and the resulting plasmids were used to transform E. coli. To provide recombinant CsoS3 without any purification tag, the intein-based IMPACT Protein Purification System (New England Biolabs) was employed. Briefly, the csoS3 gene from H. neapolitanus was amplified from plasmid pTr1, which contains the entire cso gene cluster from H. neapolitanus (2), and inserted into the Ndel and SapI sites of pETIY1 to generate plasmid pET32A. Recombinant CsoS3 protein was overexpressed as a fusion protein containing a C-terminal intein with a chitin binding domain and was purified by affinity chromatography on a chitin column. After thiol-induced intein self-cleavage, recombinant CsoS3 was eluted from the column.

Recombinant CsoS3 expression was typically induced by addition of 0.6 to 1.0 mM isopropyl-β-D-thiogalactoside to transformed cells, which were grown at 29°C with shaking (157 rpm) to an optical density at 600 nm of 0.5. After 4 h, cells were harvested by centrifugation (at 4,000 × g for 5 min) and washed with 20 ml of assay buffer (100 mM EPPS [N-[2-hydroxyethyl]piperazine-N′-3-propanesulfonic acid]-NaOH [pH 8.0]–20 mM MgSO₄). Cells were resuspended in 2 ml of assay buffer prior to disruption (four 20-s pulses) with an Ultrasonic Sonicator (model W-220F; Branson). Cell lysates were maintained on ice for a minimum of 6 h to facilitate proper folding of polyepitides prior to mass spectrometric (MS) analyses.

CA assays. CA activity in protein samples was detected as acceleration in the rate of 13C18O₂ appearance (m/z = 45), arising from the exchange of 13CO₂ from 13C18O18O (m/z = 49) to H₂O. Assay buffer (6 ml) was sealed in a cuvette and equilibrated to 30°C. K₂13C18O₃ (400 μM) was added, and the progress of the uncatalyzed exchange reaction was monitored by measuring m/z 49, 47, and 45 with an MS (MM14-80SC, VG Gas Analysis) (20). After 3 min, a protein sample (100 μl) was added, and the effect was recorded. Samples were incubated with classical CA inhibitors for 20 min at room temperature and 5 min at 30°C prior to the start of each assay (20). CA activity was also measured electrometrically. The protein sample was added to 800 μl of 20 mM Tris-HCl, pH 8.3, and the reaction was initiated by adding 600 μl of ice-cold water saturated with CO₂. The pH change resulting from the hydration of CO₂ was measured by a combination microelectrode (Microelectrodes Inc.). One Wilbur-Anderson (WA) unit of activity is defined as 2(Tₚ – Tₑ)⁻¹, where Tₚ and Tₑ are the times required for the pH to change from 8.3 to 6.3 for the uncatalyzed and catalyzed reactions, respectively (38). Inhibitors were incubated with protein samples for 30 min on ice before the start of the assay.

Western blot analysis. After protein transfer (at 100 V for 1 h) into nitrocellulose membranes, blots were probed with a 1:2,500 (vol/vol) dilution of a rabbit antiserum raised against CbbL from H. neapolitanus (designations begin with HNI for the IMPACT system and with HN for the pET system). P. marinus strains MED4 (MED) and MIT9313 (MIT), and Synechococcus strain WH8102 (WH). Nucleotide substitutions were made in primer sequences in order to introduce restriction sites (underlined) that would facilitate subsequent cloning and CsoS3 expression.

Results

CA activity is associated with purified cso-carboxysomes. Carboxysomes from H. neapolitanus (Fig. 1A) were purified to homogeneity (Fig. 1B) and assayed for CA activity by using a sensitive MS assay (Fig. 2A). The assay measures the depletion of 13O from 13C18O₂ (m/z = 49) to H₂O that occurs by repeated hydration-dehydration, yielding 13C18O18O (m/z = 47) as an intermediate and 13C18O₂ (m/z = 45) as the final product.
of this isotopic exchange reaction (20). CA activity was detected in cell lysates and all carboxysome-containing fractions obtained during purification (data not shown). When the purified carboxysomes (Fig. 2B) were added to the assay buffer containing K$_2^{13}$C$_{18}$O$_3$, a significant acceleration in the rate of loss of m/z 49 and a concomitant increase in the rate of appearance of m/z 45 was observed (Fig. 2A), consistent with the presence of CA activity in this fraction. CA activity assayed by the electrometric assay (36) was also found to comigrate with RuBisCO activity on sucrose density gradients (Fig. 2C), demonstrating a tight physical association between the two enzymes.

csoS3 encodes a CA. Purified H. neapolitanus carboxysomes are composed of nine polypeptides: CsoS1A, CsoS1B, CsoS1C, CsoS2, CsoS3, OrfA, OrfB, and the large and small subunits of RuBisCO (CbbL and CbbS) (Fig. 2B and 3A). These proteins were expressed individually in E. coli, and cell extracts of the bacteria were examined for CA activity. Only E. coli transformants expressing CsoS3 were found to possess CA activity (Fig. 3B). To determine if homologues of csoS3 from the globally important cyanobacteria P. marinus MED4, P. marinus MIT9313, and Synechococcus strain WH8102 encode a functional CA, the csoS3 genes from the three strains were cloned and expressed in E. coli. All three recombinant CsoS3 proteins also displayed substantial CA activity (Fig. 3C).

CsoS3 and CA activity reside in the carboxysome shell. A single freeze-thaw cycle releases much of the RuBisCO from purified carboxysomes and leaves the shells largely intact as empty “ghosts,” as demonstrated previously (28). The two components can be separated by centrifugation, yielding a pellet consisting mainly of carboxysome shells (Fig. 1C) and a soluble fraction enriched in RuBisCO. When these fractions were tested for CA activity, on a protein basis, 95 to 100% of the activity found in the intact carboxysomes was observed in the shell fraction (Fig. 4A). Immunoblot analysis of the super- natant and pelletable fractions of broken carboxysomes from H. neapolitanus (Fig. 4B) confirmed the findings of the activity assays. Most of the protein that is recognized by an antibody directed against the CsoS3 polypeptide was confined to the carboxysome shell fraction. Conversely, the majority of CbbL detected by anti-CbbL was found in the supernatant. These results clearly demonstrate that the CA activity of CsoS3 is located in the carboxysomal shell or tightly associated with it.
The localization of CsoS3 to the carboxysome periphery is also supported by immunogold electron microscopy (4).

Characterization of CA activity. To ensure that the observed CA activity was truly due to catalysis by the CsoS3 polypeptide, the \textit{H. neapolitanus} csoS3 gene was expressed as a fusion protein with a self-splicing intein (Impact protein purification system; New England Biolabs). Recombinant CsoS3 lacking any extraneous fusion tag was purified to homogeneity by affinity chromatography. This protein had a high level of CA activity (Fig. 5). By use of the electrometric assay (36), the specific activity of the recombinant CsoS3 was found to be approximately 30\% of that measured with commercially obtained purified bovine erythrocyte CA (Fig. 5B). CsoS3 CA activity was inhibited by dithiothreitol, suggesting the involvement of cysteine residues in catalysis. Ethozolamide (EZ), a classical inhibitor of CA that acts by interfering with the catalytically essential \( \text{Zn}^{2+} \) within the enzyme’s active site, also inhibited the activity of the carboxysomal CA. Likewise, divalent metal chelators such as dipicolinic acid and nitritolriactacid acid severely interfered with CA activity. These results are consistent with CsoS3 acting through a metal-mediated mechanism for the hydration of \( \text{CO}_2 \) and the dehydration of \( \text{HCO}_3^- \), as is the case for other known CAs (30, 35).

CsoS3: a CA of distinct evolutionary origin. By using \textit{H. neapolitanus} CsoS3 as a probe, searches of the National Center for Biotechnology Information nonredundant database (http://www.ncbi.nlm.nih.gov/), the Comprehensive Microbial Resource at The Institute for Genomic Research (http://www.tigr.org/), and unfinished genome sequences at the Joint Genome Institute (http://www.jgi.doe.gov/) revealed that the protein shares similarity with eight additional sequences, all found in organisms containing \textit{cso} gene clusters (1, 8). These include the chemolithoautotrophs \textit{Thiobacillus denitrificans}, \textit{Acidithiobacillus ferrooxidans}, and \textit{Thiomonas intermedia}, the three marine photolithoautotrophic cyanobacteria named earlier, \textit{P. marinus} \textit{SS120} (or CCMP1375) (10), and \textit{P. marinus} subsp. \textit{pastoris} CCMP1378.

Alignment of the CsoS3 sequences demonstrated a high degree of sequence conservation, with up to 99\% sequence identity observed between the MED4 and CCMP1378 polypeptides. Phylogenetic analysis (17) using nucleotide sequences indicated that \textit{csoS3} could be separated into two closely related clades (Fig. 6), reflecting the chemolithoau-
The trophic or photolithooautotrophic nature of the organisms from which the sequences were obtained. The high degree of relatedness of the \( \text{csoS3} \) genes from these otherwise phylogenetically distant autotrophic organisms is consistent with the notion that \( \text{csoS3} \) and the other members of the \( \text{cso} \) gene cluster have been laterally transferred as a genetic unit (1, 7).

BLAST searches using the most highly conserved regions of the CsoS3 amino acid sequence did not reveal other significant homologues. A search for remote homologues using the PROPSEARCH program (http://www.infobiosud.univ-montp1.fr/SERVEUR/PROPSEARCH/propsearch.html), which relies on analysis of protein properties, also failed to reveal any additional homologues, including any member of the \( \alpha \), \( \beta \), or \( \gamma \) class of CAs. The overall three-dimensional folding patterns of the three classical CA types, as determined by X-ray crystallography, are unique and emphasize their independent evolutionary origins (30, 35). A CA (TWCA1) from the diatom \( \text{Thalassiosira weissflogii} \) may represent a fourth lineage (\( \varepsilon \) class) (35) of the enzyme, although some evidence suggests that it may instead be a distant homologue of the \( \alpha \)-class CAs (18).

Using the 3D-PSSM program (http://www.sbg.bio.ic.ac.uk/servers/3dpssm/), we attempted to identify common folding patterns between CsoS3 and other known proteins. No structural homology between CsoS3 and any of the recognized CA classes was evident. The collective data indicate, therefore, that CsoS3 is a hitherto undiscovered evolutionary lineage of CA, which we propose to designate the \( \varepsilon \) class.

All of the \( \varepsilon \)-class CA genes identified to date are found within \( \text{cso} \) gene clusters, flanked by \( \text{csoS2} \) and \( \text{orfA} \) in an arrangement similar to that observed for \( \text{H. neapolitanus} \) (Fig. 3A). The genes range from 1,503 to 1,602 nucleotides, with the exception of the \( \text{Synechococcus} \) strain WH8102 \( \text{csoS3} \) gene, which is 1,728 nucleotides. The predicted molecular masses of the CsoS3 polypeptides range from 55.2 to 63.4 kDa and are considerably greater than the masses (17 to 32 kDa) of typical \( \alpha \)-, \( \beta \)-, and \( \gamma \)-class CA monomers. There are two potential translation start sites within the \( \text{Synechococcus} \) strain WH8102 \( \text{csoS3} \) gene. Only the larger of the two proteins (575 amino acids [\( M_r \), 63,400] as opposed to 482 amino acids [\( M_r \), 53,500]) was catalytically active (Fig. 3C). In all known instances, the
catalytically essential Zn$^{2+}$ is coordinated to the Ca apoprotein through three histidine residues or a combination of histidine, cysteine, and sometimes aspartate (Fig. 7) (35). There are six histidine, two cysteine, and six aspartate residues that are completely conserved across the nine CsoS3 polypeptides. Many (71%) of these candidate Zn$^{2+}$ binding residues are located in a conserved 140-amino-acid stretch in the midportion of the protein (Fig. 7).

**DISCUSSION**

Electrometric and MS assays demonstrate that purified CsoS3 from *H. neapolitanus* catalyzes the reversible hydration of CO$_2$ and the exchange of $^{18}$O between $^{13}$CO$_2$ and H$_2$O, reactions characteristic of the CAs (EC 4.2.1.1). Further support for this classification comes from the observation that catalysis may occur through a metal-mediated reaction mechanism that is inhibited by the classical CA inhibitor EZ. Recombinant CsoS3 from the marine cyanobacteria *Prochlorococcus* and *Synechococcus* also catalyze $^{18}$O exchange and are, therefore, functional homologues of the *H. neapolitanus* enzyme. The CsoS3 monomer is considerably larger (>50%) than typical α-, β- and γ-class CA monomers. Functional CAs with monomeric sizes comparable to that of CsoS3 have been identified from the salt-tolerant green alga *Dunaliella salina* (60 kDa) (11a) and the red alga *Porphyridium purpureum* (50 kDa) (20a). However, the increased size likely is the consequence of gene duplication, as the monomers contain either two α-class or two β-class CA domains. This is not the case for CsoS3. The larger size and mass of CsoS3 may reflect a dual role as a catalytically active enzyme and as a structural component of the carboxysome.

RuBisCO and CsoS3-associated CA activity colocalize to the carboxysome, with RuBisCO located within the core and CsoS3 tightly associated with the bounding shell. CsoS3 is a minor protein component of the carboxysome (Fig. 2B) and is therefore unlikely to form the major structural elements of the carboxysome shell. The close physical proximity of CsoS3-associated CA activity to RuBisCO within the cso-carboxysome interior, however, immediately suggests a direct role for this enzyme in the provision of CO$_2$ for autotrophic fixation. This suggestion is most compelling for *Prochlorococcus* spp. and *A. ferrooxidans*, since the genomes of these organisms do not contain recognizable genes for any member of the α, β, or γ class of CAs. *Synechococcus* strain WH8102 contains, in addition to *csoS3*, a single β-class CA gene whose product is most closely related to a noncarboxysomal β-CA thought to be associated with the cell surface in *Synechocystis* sp. strain PCC6803 (33). From a functional point of view, CsoS3 may be located at strategic positions throughout the carboxysome shell, where it can gain access to the intracellular HCO$_3^-$ pool and channel CO$_2$ to RuBisCO. Direct measurements have established the presence of a cytosolic HCO$_3^-$ pool in *H. neapolitanus* and many cyanobacteria (1, 14), but such mea-
measurements are lacking for the marine cyanobacteria considered here. The colocalization of RuBisCO with a direct enzymatic source of its substrate, CO₂, provides an obvious catalytic advantage to autotrophic CO₂ fixation, but whether additional benefits are afforded by the unique organization of the carboxysomes remains to be established. Knowledge of the spatial organization of CsoS3 and RuBisCO at the molecular level and of their kinetic characteristics is essential to furthering our understanding of the functioning of cso-carboxysomes.

Although CsoS3 catalyzes classic CA reactions, its primary function in the case of Prochlorococcus ε is currently unknown. However, due to the presence of CsoS3 in several chemolithoautotrophs and certain marine cyanobacteria, it is tempting to speculate that it represents a novel evolutionary lineage of the enzyme, the ε class. At present, ε-CA appears to be unique to the Bacteria and is narrowly distributed among several chemolithoautotrophs and certain marine cyanobacteria that contain cso-carboxysomes. Recent genome-wide analyses of P. marinus SS120 and MED4 indicate that these organisms may contain the minimal gene complement necessary for oxytrophic bacteria (10, 26). The perceived absence of a CA in these organisms has raised questions as to the necessity of this enzyme for autotrophic CO₂ fixation. Our work indicates that a catalytically functional ε-CA is encoded within these minimal genomes and that carboxysomal CA is an integral component of autotrophic metabolism.

From an ecological point of view, bacterial strains that contain ε-CA and cso-carboxysomes are globally distributed and, in the case of Prochlorococcus and Synechococcus spp., are abundant primary producers in the subtropical and tropical open oceans (21). Thus, through their photosynthetic and chemolithoautotrophic CO₂ fixation, carboxysomes and ε-CA influence the biogeochemical cycling of carbon in a globally significant way.

ACKNOWLEDGMENTS

This work was supported by the Ontario Graduate Scholarship Fund (to A.K.-C.S.), the Natural Sciences and Engineering Research Council of Canada (to G.C.C.), the National Science Foundation EPSCoR and MRSEC programs (to S.H. and G.C.C.), and the T. W. Bennett Distinguished Professorship Fund (to G.C.C.). We thank S. W. Chisholm and B. Brahamska for providing Prochlorococcus and Synechococcus genomic DNA.

REFERENCES


