Spring 5-2009

Functional Characterization and Assembly Studies of Carboxysomes in *Halothiobacillus neapolitanus*

Zhicheng Dou
University of Southern Mississippi

The University of Southern Mississippi

FUNCTIONAL CHARACTERIZATION AND ASSEMBLY STUDIES OF CARBOXYSOMES IN *HALOTHIOBACILLUS NEAPOLITANUS*

by

Zhicheng Dou

Abstract of a Dissertation
Submitted to the Graduate Studies Office
of The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

May 2009

The University of Southern Mississippi

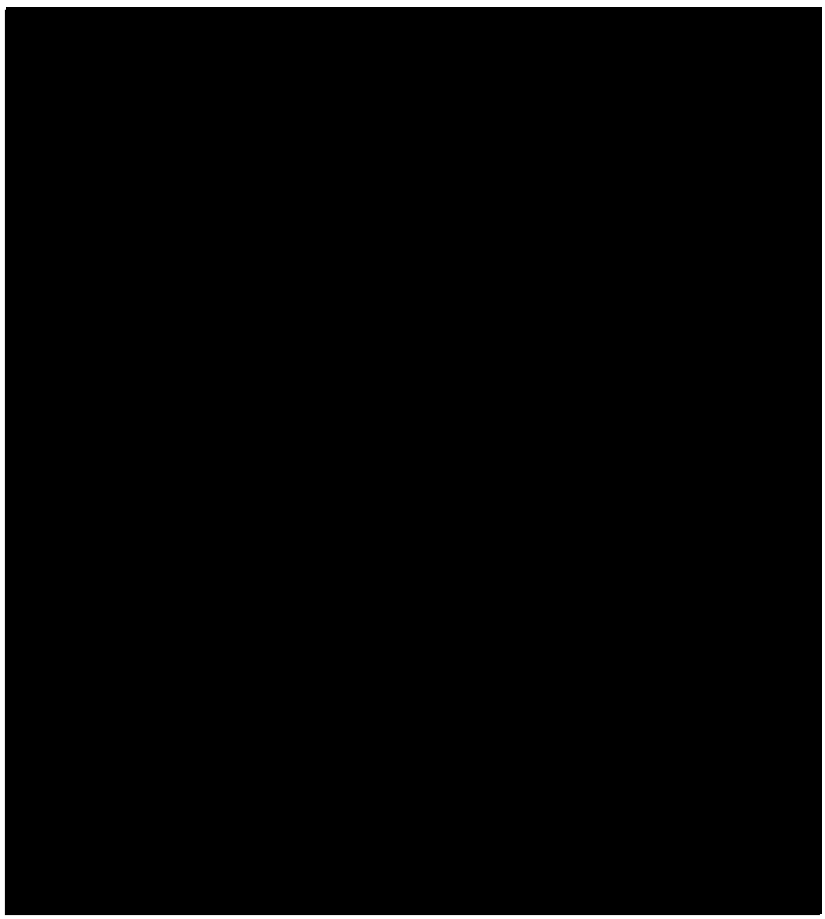
FUNCTIONAL CHARACTERIZATION AND ASSEMBLY STUDIES OF CARBOXYSOMES IN *HALOTHIOBACILLUS NEAPOLITANUS*

by

Zhicheng Dou

A Dissertation
Submitted to the Graduate Studies Office
of The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

Approved:



May 2009

ABSTRACT

FUNCTIONAL CHARACTERIZATION AND ASSEMBLY STUDIES OF CARBOXYSOMES IN *HALOTHIOBACILLUS NEAPOLITANUS*

by Zhicheng Dou

May 2009

Functional characterization and assembly studies of carboxysomes in *Halothiobacillus neapolitanus* were pursued in order to understand the roles of carboxysomes in the carbon metabolism in *H. neapolitanus* and the assembly of carboxysomes *in vitro* and *in vivo*. Previously, a low abundance *H. neapolitanus* carboxysomal protein, CsoSCA was identified as a novel carboxysomal shell-bound carbonic anhydrase. The enzyme is thought to dehydrate the cytosolic bicarbonate to CO_2, the substrate of the RuBisCO packaged within the carboxysome [14, 21]. In this study, the carboxysomal shell was identified as a diffusion barrier for CO_2 and O_2 molecules. The shell-bound CsoSCA protein facilitates the diffusion of CO_2 molecules into the carboxysomes and enhances the catalytic efficiency of the encapsulated RuBisCO. The discrimination between CO_2 and O_2 molecules by the shell makes the sequestered RuBisCO favor the carboxylation over the oxygenation reaction at low O_2 concentration. A second substrate, ribulose 1,5-bisphosphate (RuBP), and the carboxylation reaction product, 3-phosphoglycerate (PGA), are negatively charged molecules and need to be transported into and out of the carboxysomes. Bioinformatics analysis of CsoS2 protein revealed that CsoS2 protein carries a positive charge in the cytoplasm. The protein may interact with these negatively charged molecules due to its high isoelectric point (approximately 9.2). Full length CsoS2 and an N-terminally truncated CsoS2 protein were expressed in *E. coli* to permit three dimensional structure

determination. A sequential peptide affinity (SPA) tag was added at the C-terminus of CsoS2 protein to determine its location in the carboxysomes and trap the intermediates of *in vitro* carboxysome shell assembly. Finally, the *cso* operon was heterologously expressed in *E. coli* and some functional carboxysome-like structures were isolated and analyzed. The assembly of carboxysome-like structures in *E. coli* will guide the *in vitro* assembly of carboxysomes with recombinant carboxysomal proteins.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Sabine Heinhorst and co-advisor Dr. Gordon Cannon for their guidance, help, and support through my Ph.D program. I extend my sincere thanks for helpful discussion and suggestions to my committee members, Drs. Faqing Huang, Jeffery Evans, and Kenneth Curry.

I also want to thank my wife, Yanlin Gao for being supportive and understanding when I had to work late or on the weekends. Without her constant support and encouragement, I do not think that I could finish my Ph.D program smoothly. I really appreciate the huge help from my parents and parents-in-law who took care of my son so that I could have a lot of time and energy to focus on my research project.

I want to thank all the members of the Cannon/Heinhorst research group for their help for scientific discussions. I especially want to thank Balaraj Menon for his companionship during the last three years.

Finally, I would like to acknowledge the funding that supported this research: the Chemistry and Biochemistry Department at the University of Southern Mississippi, National Science Foundation MCB-0444568 and National Science Foundation Materials Research Science Engineering Center DMR 0213883.

TABLE OF CONTENTS

ABSTRACT ii

ACKNOWLEDGMENTS iv

LIST OF ILLUSTRATIONS vii

LIST OF TABLES x

LIST OFABBREVIATIONS xi

CHAPTER

I. INTRODUCTION 1

II. LITERATURE REVIEW 4

Carbon Dioxide Fixation in Prokaryotes
Carboxysome Function
Carboxysome Composition
Structural Insights into Carboxysomes
Biogenesis and Protein-Protein Interactions of Carboxysomes
Other Carboxysome-like Microcompartments
Objective of Study

III. EXPERIMENTAL PROCEDURES 29

Materials
Routine Apparatus
Bacterial Growth Media
Buffers
Bacterial Strains and Growth
Methods

IV. RESULTS 57

Characterization of csoS3::Km Mutant
Oxygenase Activity of the Carboxysomal RuBisCO
Characterization of CsoS2 Protein
Assembly Studies of *H. neapolitanus* Carboxysomes

V. DISCUSSION 122

Carboxysomes Are Essential in the CO_2-Concentrating Mechanism
The Carboxysomal Shell Is a Diffusion Barrier for Gas Molecules

CsoS2 Protein – A Potential Candidate to Transport RuBP and PGA Carboxysomal Shell Assembly

VI. CONCLUSION AND FUTURE WORK 147

REFERENCES 150

LIST OF ILLUSTRATIONS

Figure

1. A model for the two-stage CCM 6

2. Morphology and composition of wild type *H. neapolitanus* carboxysomes 12

3. Genomic organization of the genes of carboxysomes and related bacterial microcompartments in several different bacterial species 13

4. Diagram of the strategy used to generate the *H. neapolitanus csoS3::Km* mutant and immunoblot analysis of wild type and mutant cell extracts 58

5. Growth curves of *H. neapolitanus* wild type and *csoS3::Km* mutant cultures 60

6. Assimilation of inorganic carbon by *H. neapolitanus* wild type and *csoS3::Km* mutant cells 62

7. Transmission electron microscopy and dynamic light scattering analyses of wild type and *csoS3::Km* mutant carboxysomes 64

8. Composition of wild type and *csoS3::Km* mutant carboxysomes 66

9. Carbonic anhydrase activity of purified *H. neapolitanus* wild type and *csoS3::Km* mutant carboxysomes 67

10. Determination of K_M(RuBP) for *H. neapolitanus* wild type and *csoS3::Km* mutant carboxysomes 69

11. Effect of exogenously added rCsoSCA protein on CO_2 fixation activity of wild type and *csoS3::Km* mutant carboxysomes 72

12. CO_2 fixation by *H. neapolitanus* wild type and *csoS3::Km* mutant carboxysomes 74

13. pH dependence of the CO_2 fixation activity of carboxysomal and free RuBisCO 77

14. The isotope-based simultaneous assay of RuBisCO carboxylase and oxygenase activity 79

15. Purification of enzymatically synthesized 1-[^{3}H]RuBP 82

16. Specificity factor (SF) of carboxysomal and free RuBisCO 84

17. Separation of glycerate and glycolate by a HPLC carbohydrate analysis column....86

18. Carboxylase to oxygenase activity ratio for carboxysomal and free RuBisCO enzyme....88

19. O_2 inhibition constants for carboxysomal and free RuBisCO....91

20. The primary structure of CsoS2 protein from *Halothiobacillus neapolitanus*....94

21. Expression of recombinant CsoS2 protein in the pProEX® and the IMPACT® systems....96

22. Expression of the N-terminally truncated rΔ250CsoS2 protein in *E. coli*....99

23. Comparison of composition and morphology of carboxysomes from *Halothiobacillus neapolitanus*, *Thiomonas intermedia* and *Thiomicrospira crunogena*....101

24. A sequential peptide affinity (SPA) tag sequence added at the 3' end of the *csoS2* gene by homologous recombination in *H. neapolitanus*....102

25. Screening of HnSPAS2 mutant clones by PCR amplification....105

26. Morphology and composition of HnSPAS2 mutant carboxysomes....107

27. Mapping of CsoS2A and CsoS2B derived peptides by MALDI-ToF mass spectrometry....109

28. Immunoblot analysis of CsoS2 protein expressed in the IMPACT® system....110

29. Trapping of SPA-tagged carboxysomes on an anti-FLAG affinity column....112

30. Affinity trapping of SPA-tagged protein complexes during *in vitro* re-assembly of HnSPAS2 mutant carboxysomal shell structures....114

31. Transmission electron microscopy analysis of trapped intermediates formed during *in vitro* re-assembly of HnSPAS2 mutant carboxysomal shells....116

32. Expression of the entire *H. neapolitanus cso* operon in *E. coli*....118

33. Carboxysome-like structures assembled in *E. coli* have CO_2 fixation activity ...119

34. Transmission electron microscopy of carboxysome-like structures assembled in *E. coli*....121

35. Working model for the location and function of CsoSCA and CsoS2 in the carboxysome shell.....128

36. The potential frame shifting stop codon in *H. neapolitanus csoS2* gene.....139

LIST OF TABLES

Table

1. Primers 32

2. Kinetic constants for CO_2 fixation by RuBisCO 75

3. Short time carboxylase to oxygenase activity ratio at different concentrations of O_2 ... 89

4. Calculated molecular masses and isoelectric points (pIs) of CsoS2 proteins 93

LIST OF ABBREVIATIONS

β-ME	2-mercaptoethanol
APS	ammonium persulfate
BCA	bicinchoninic acid
BCIP	5-bromo-4-chloro-3-indolyl phosphate
Bicine	N,N-Bis(2-hydroxyethyl)glycine
BMC	bacterial microcompartment
bp	base pair
BSA	bovine serum albumin
CA	carbonic anhydrase
CABP	2-C-carboxyarabinitol-1,5-bisphosphate
CCM	carbon dioxide concentrating mechanism
CBB	Calvin-Benson-Bassham
C_i	inorganic carbon
CIAP	calf intestinal alkaline phosphatase
DIC	dissolved inorganic carbon
DLS	dynamic light scattering
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
eut	ethanolamine utilization
HCR	high CO_2 requiring
HEPES	4-(2-hydroxyethyl)-1-piperazine ethane sulfonic acid
HPLC	high pressure liquid chromatography
HRP	horseradish peroxidase
IgG	immunoglobulin G
IPTG	isopropyl-β-D-thiogalactoside
MALDI-ToF	Matrix Assisted Laser Desorption /Ionization - Time of Flight
MES	2-(N-morpholino)ethanesulfonic acid
MOPS	3-(N-morpholino)propanesulfonic acid
NADH	nicotinamide adenine dinucleotide
NADP	nicotinamide adenine dinucleotide phosphate
NBT	4-nitroblue tetrazolium
NTA	nitrilotriacetic acid
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PD	propanediol
pdu	propanediol utilization
PG	2-phosphoglycolate
PGA	3-phosphoglycerate
PMSF	phenylmethylsulfonyl fluoride
PTSF	p-toloenesulfonyl fluoride
RuBisCO	ribulose 1,5-bisphosphate carboxylase/oxygenase
RuBP	ribulose 1,5-bisphosphate

RuMP	ribulose 5-phosphate
SDS	sodium dodecyl sulfate
SPA	sequential peptide affinity
TEM	transmission electron microscopy
TEMED	N,N,N', N'-Tetramethylethylenediamine
TEV	tobacco etch virus
Tris	Tris-(hydroxymethyl)-aminomethane base

CHAPTER I

INTRODUCTION

Ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) is a key enzyme in the Calvin-Benson-Bassham (CBB) cycle. The enzyme catalyzes the first step of CO_2 fixation [43] and is widely distributed in green plants, chemo- and photoautotrophic bacteria. RuBisCO converts environmental inorganic carbon into 3-phosphoglycerate that is further metabolized and yields biomass [121, 122]. To adapt to the low concentration of inorganic carbon in the surrounding environment, many bacteria adopt an effective carbon dioxide concentrating mechanism (CCM) [9, 95]. In the first step, bicarbonate and CO_2 are taken up from the environment by energy-dependent transport into the cell, and their concentration is elevated several orders of magnitude compared to their environmental concentrations [23, 40, 49]. Bicarbonate constitutes the major species of the intracellular C_i-pool. The plasma membrane has nearly 1,000 fold less permeability for the negatively charged bicarbonate than for the uncharged CO_2 molecules [9]. Therefore, the existence of bicarbonate in the cytoplasm as the major species can help prevent the loss of inorganic carbon source [23, 40, 49]. Given that the substrate of RuBisCO is CO_2, rather than HCO_3^- [53, 115], the abundant cytosolic bicarbonate needs to be converted to CO_2 for efficient utilization by RuBisCO. Some chemoautotrophic bacteria and all cyanobacteria package virtually all of the cell's RuBisCO into polyhedral microcompartments, called carboxysomes [105-107]. Carboxysomes are the terminal component of the CCM. Their carbonic anhydrase accelerates the conversion of bicarbonate to CO_2 and provides the sequestered RuBisCO enzyme with sufficiently high concentration of CO_2 to enhance its catalytic efficiency. In addition, the shell is proposed

to restrict CO_2 leakage out of the interior, thereby contributing to increase the utilization efficiency of intracellular inorganic carbon [9, 10, 89, 95].

Homogeneous preparations of carboxysomes for the sulfur oxidizer *Halothiobacillus neapolitanus* are composed of ten polypeptides: RuBisCO large and small subunit, CsoS1A, CsoS1B, CsoS1C, CsoS2A, CsoS2B, CsoSCA, CsoS4A, and CsoS4B. RuBisCO is the most abundant protein in the carboxysome, accounting for 60 – 70% of the total carboxysome protein [25, 26]. The rest of the carboxysomal proteins are associated with the thin proteinaceous shell [28]. The major structural determinants of the carboxysome shell are CsoS1A, CsoS1B, and CsoS1C, highly conserved small proteins. Crystallographic studies of CsoS1A protein revealed that the protein forms hexamers, which can pack into sheets and likely form the facets of the icosahedral microcompartment [124, 125]. Another two conserved carboxysome components, CsoS4A and CsoS4B, crystallize as pentamers and may occupy the vertices of the icosahedral structure [124]. By a reverse genetics approach, the CsoSCA protein (formerly known as CsoS3) was identified as a carboxysome shell-bound carbonic anhydrase [57, 114]. Given that its primary structure lacks homology with known carbonic anhydrases, it was categorized into a novel class [114]. Subsequent crystallographic analysis revealed that CsoSCA protein forms a dimer and constitutes a subclass of β-carbonic anhydrases [100]. Although there are only 40 copies of CsoSCA dimers per carboxysome, the low abundance enzyme can supply a sufficient amount of CO_2 to the encapsulated RuBisCO to maintain the active sites of RuBisCO saturated with substrate and enhance RuBisCO's catalytic efficiency [57]. The CsoS2A and CsoS2B proteins together account for approximately 12% of the carboxysomal protein in mass

[56]. These two polypeptides are encoded by a single gene in the *cso* operon [13]. Their unusually high isoelectric points suggested that they might be involved in transferring the phosphorylated RuBisCO substrates and products across the carboxysome shell [30, 57].

Although the composition and structure of carboxysomes have been studied by biochemical, genetic and crystallographic approaches, the mechanisms by which RuBisCO substrates and products cross the shell and by which carboxysomes enhance the catalytic efficiency of the encapsulated RuBisCO are still unknown. This study aims at exploring the role played by the carboxysomal shell in enhancing the catalytic efficiency of the encapsulated RuBisCO and at gaining a better understanding of the carboxysome assembly pathway. The insights gained from the results obtained will help broaden the understanding of the roles of carboxysomes in carbon metabolism in chemo- and photo-autotrophic bacteria and provide information useful for the future *in vitro* reconstitution of carboxysomes from individual protein components. Elucidation of the molecular properties of biological protein sheets like the carboxysome shell also will facilitate the development of new materials for applications in medicine, biotechnology and materials science.

CHAPTER II

LITERATURE REVIEW

Carbon Dioxide Fixation in Prokaryotes

Ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO, EC 4.1.1.39) is the most abundant protein on Earth [43]. It is responsible for catalyzing the first step in the Calvin-Benson-Bassham (CBB) cycle by which inorganic carbon is converted into small organic compounds that subsequently yield biomass [121, 122]. Inorganic carbon fixation was thought to be mainly contributed by green plants. However, recent studies have revealed that autotrophic prokaryotes play a significant role in CO_2 fixation and in the global carbon cycling [111]. The environment in which the bacteria live has a very low atmospheric CO_2 concentration with an estimated 400 ppm [47]. To adapt to the low concentration of CO_2, cyanobacteria and some chemoautotrophic bacteria have adopted an efficient CO_2-concentrating mechanism (CCM). The CCM is composed of two stages. In the first step, the bacteria actively take up soluble inorganic carbon species from their environment and elevate their concentration as high as 1,000-fold within the cells via energy-dependent transmembrane transport [23, 40, 49]. In the second step, the accumulated cytosolic bicarbonate is converted into CO_2 molecules under the catalysis of carbonic anhydrase (CA). The CO_2 is fixed by RuBisCO and converted into organic cell building blocks [40, 95, 122].

Thus far, the most extensive investigations of inorganic carbon (C_i) transport in autotrophs have been performed in *Synechococcus* PCC 7942 and *Synechocystis* PCC 6803 [93, 104]. Five modes of C_i-uptake were documented in these cyanobacteria and categorized into two groups according to the inorganic species to be transported (CO_2 or

HCO_3^-): 1) a type I NAD(P)H dehydrogenase (NDH-1) complex, labeled as NDH-I_4 and residing in the cell membrane, constitutively takes up CO_2 [78, 102]; 2) a second NDH-1 complex, referred to as NDH-I_3, is induced under limiting C_i conditions and shows a higher affinity for CO_2 than NDH-I_4 [78, 102]; 3) a high-affinity HCO_3^- transporter (BCT1) belongs to the bacterial ATP binding cassette (ABC) transporter family and plays an important role under severe C_i limitation [85]; 4) an inducible medium affinity HCO_3^- transporter (SbtA) and 5) a BicA HCO_3^- transporter are Na^+-dependent Na^+/HCO_3^- symporters [94, 103]. Regardless of which form of C_i is transported into the cell cytoplasm, bicarbonate constitutes the majority of the intracellular C_i-pool. The plasma membrane is nearly 1,000-fold less permeable to HCO_3^- than to the uncharged CO_2 [9]. To maintain the high concentration of bicarbonate in the cytosol, it was hypothesized that there can not be a CA activity in the cytoplasm of cyanobacteria, and that the conversion between bicarbonate and CO_2 within the cytosol must constitute a very slow equilibrium to prevent the escape of CO_2 molecules from the cells [50]. The expression of human α-CA in the cytosol of *Synechococcus* PCC 7942 results in loss of the intracellular C_i-pool and in a high CO_2-requiring (HCR) phenotype in this mutant [88].

Given that the substrate of RuBisCO is CO_2, rather than HCO_3^-, and that RuBisCO is an inefficient catalyst with a low turnover rate of 2 to 5 s^{-1} [53, 115], the availability of CO_2 determines the rate of the carboxylation reaction. A RuBisCO-encapsulating microcompartment, the carboxysome, makes a significant contribution to carbon metabolism in carboxysome-containing bacteria [25, 28, 110, 111]. The presence of carboxysomal carbonic anhydrase accelerates the equilibration of bicarbonate and CO_2 [50]. In addition, the carboxysomal shell shows a diffusion barrier for CO_2 molecules

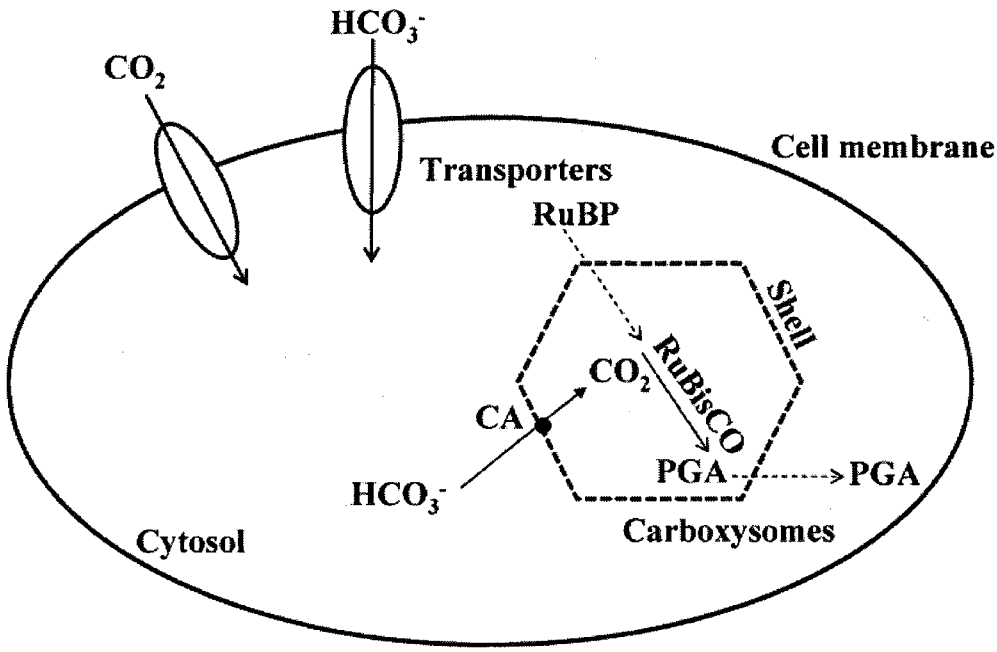


Figure 1. A model for the two-stage CCM.

In the first step, inorganic carbon species are assimilated and concentrated in the cytoplasm by energy-dependent inorganic carbon pumps. The intracellular bicarbonate is converted to CO_2 by the carboxysomal carbonic anhydrase. The carboxysomal shell is a diffusion barrier for CO_2 molecules, thereby restricting CO_2 leakage out of the carboxysomes [57]. The trapped CO_2 molecules within the carboxysomes are fixed by the encapsulated RuBisCO enzyme into 3-phosphoglycerate (PGA) that subsequently is metabolized into glucose and biomass. The transportation mechanism of RuBP and PGA across the shell (dashed line) is unknown. Redrawn from Yeates *et al.* [129].

[57], therefore, the shell may restrict CO_2 leakage out of the carboxysomes. Thus, the effective concentration of CO_2 is elevated around the active site of RuBisCO within the carboxysomes, thereby increasing the efficiency of CO_2 fixation by RuBisCO (Figure 1) [9].

Carboxysome Function

Since carboxysomes were identified in all cyanobacteria and some chemoautotrophic bacteria, two hypotheses have been proposed. 1) Carboxysomes are responsible for the storage of excess RuBisCO enzyme synthesized in response to a stress condition [108]; 2) Carboxysomes are active in CO_2 fixation and play an important role in bacterial carbon metabolism. The evidence that carboxysomes isolated from the chemoautotrophic bacterium *Halothiobacillus neapolitanus* show CO_2 fixation activity *in vitro* [25, 26] supports the latter hypothesis. Furthermore, in this bacterium, the number of carboxysomes is inversely correlated with the abundance of inorganic carbon in the environment [18, 25], suggesting that carboxysomes play an important role in helping the bacteria adapt to the inorganic carbon limitation. Similar results were obtained in *Thiomonas intermedia* (formerly *Thiobacillus intermedius*), which can grow autotrophically, mixotrophically and heterotrophically. In this bacterium, RuBisCO and carboxysome shell proteins are not expressed in the presence of an organic carbon source [96]. These findings strongly support the second hypothesis. Direct evidence of the formation of stable complexes between the transition state analog 2-C-carboxyarabinitol-1,5-bisphosphate (CABP) and carboxysomal RuBisCO activated with $^{14}CO_2$ and Mg^{2+} in purified intact carboxysomes revealed that the RuBisCO packaged inside the

microcompartments is kept in an activated state [27]. In addition, when the cell membrane of *H. neapolitanus* was treated with chloroform to increase its permeability, both carboxysomal and cytoplasmic RuBisCO were shown to bind CABP and the activating $^{14}CO_2$, demonstrating that both cytosolic and carboxysomal RuBisCO enzyme forms are active *in vivo* [27].

Several hypotheses have been formulated for the mechanism of carboxysome function. First, since type I RuBisCO enzymes from prokaryotes have a low affinity for CO_2 (K_M > 150 μM), the carboxysomal RuBisCO enzyme may be kept in a high-affinity conformation that increases the efficiency of its catalysis [28]. However, isolated carboxysomes do not show a significantly higher specific carboxylation activity than free RuBisCO enzyme *in vitro* [26, 60]. Second, RuBisCO is a bifunctional enzyme, and CO_2 and O_2 compete for the same active site [2, 53]. Nonproductive fixation of O_2 onto RuBP by RuBisCO therefore decreases the efficiency of carboxylation [53]. If the carboxysomal shell can selectively restrict the passage of O_2 into the interior of the microcompartment, the oxygenase activity would be inhibited, and CO_2 fixation would be enhanced [28]. Carboxysome-enriched fractions from the cyanobacterium *Coccochloris peniocystis* and *Synechocystis* PCC 6803 were found to be less sensitive to oxygen inhibition than free RuBisCO enzyme [35, 79], which is consistent with the hypothesis that the carboxysomal shell can differentiate between CO_2 and O_2 molecules. Oxygen electrode measurements were taken to assess the oxygenase activity of intact carboxysomes and free RuBisCO enzyme from *H. neapolitanus*. The two enzyme forms show approximately the same level of oxygenase activity [25, 26]. Given that the response time of the oxygen electrode is slow and the potentiometric determination of

oxygenase activity is not very sensitive, these assays were performed over only a limited range of oxygen concentrations. A more accurate, [^{3}H]RuBP-based simultaneous oxygenase assay will be used to determine the ratio of carboxylase to oxygenase for intact carboxysomes and free RuBisCO enzyme in this study. Third, the only form of inorganic carbon substrate that RuBisCO enzyme can utilize is CO_2 [122], so the concentration of CO_2 determines the rate of RuBP carboxylation by RuBisCO. Originally, it was thought that carbonic anhydrase within the cell can speed up the conversion of HCO_3^- to CO_2 and thereby increase the concentration of intracellular CO_2, thus allowing RuBisCO to work at near saturating concentrations of its substrate [7]. Contrary to this hypothesis, Reinhold *et al.* proposed a quantitative model that assumes carbonic anhydrase is located within the carboxysome [50]. Such an arrangement would supply carboxysomal RuBisCO enzymes with a high concentration of CO_2 locally to enhance the efficiency of the carboxylation reaction [28]. Subsequent experiments revealed that carbonic anhydrase activity co-localizes with a carboxysome-enriched fraction from the cyanobacterium *Synechococcus* PCC 7942 [90]. Expression of carbonic anhydrase in the cytosol accelerates the equilibration of bicarbonate and CO_2, thereby disturbing the cytoplasmic C_i-pool and diminishing CO_2 fixation by RuBisCO [88]. These findings strongly support this quantitative model. In purified α-carboxysomes of *H. neapolitanus*, the low abundance carboxysomal protein CsoSCA (formerly CsoS3) was the first one to be identified as a *bona fide* carboxysomal carbonic anhydrase [114]. Immunoelectron microscopy revealed that CsoSCA protein is localized at the periphery of carboxysome [14]. CsoSCA protein remains tightly associated with the shell after a freeze-thaw treatment of carboxysomes [57, 114]. The carbonic anhydrase activity in intact purified

carboxysomes is lower than that in broken carboxysomes, which suggested that the carboxysomal shell acts as a diffusion barrier for CO_2 molecules [50, 57]. Such an arrangement of carbonic anhydrase and RuBisCO enzyme is thought to maintain the active site of RuBisCO saturated with CO_2 molecules and to increase the efficiency of CO_2 fixation [57].

Besides the kinetic characterization of carboxysomes *in vitro*, mutational studies also have contributed experimental evidence for the function of carboxysomes. The *csoS1A* gene, which encodes a carboxysomal shell protein, was inactivated in *H. neapolitanus* via insertion of a kanamycin resistance gene. The resulting mutant showed a high CO_2-requiring phenotype and a higher ratio of carboxysomal (particulate) to free (soluble) RuBisCO than wild type cells [46], indicating that encapsulation of RuBisCO into microcompartments is essential for the chemoautotrophic bacteria to grow in a low inorganic carbon environment. In some cyanobacterial high CO_2-requiring mutants, bicarbonate is assimilated into the cytosol and maintained at high concentration, but can not be utilized efficiently. Some morphologically altered carboxysomes or no carboxysome-like structures at all are observed in many of these mutants by transmission electron microscopy [84, 92]. These findings demonstrated that carboxysomes play an essential role in the CCM and are crucial for efficient inorganic carbon utilization by the bacteria.

Carboxysome Composition

Over 60 years ago, polyhedral inclusion bodies were first documented in the cyanobacterium *Phormidium uncinatum* [42]. Subsequently, similar inclusion bodies

were found in many cyanobacteria and in some chemoautotrophic bacteria [107]. In 1973, the polyhedral bodies of the chemoautotrophic bacterium *Halothiobacillus neapolitanus* (formerly known as *Thiobacillus neapolitanus*) were purified (Figure 2A) [105]. Transmission electron microscopic studies showed that the purified inclusion bodies appear to be of regular hexagonal shape, approximately 100 nm in diameter (Figure 2B) [105, 106] and with a solid , granular interior that is surrounded by a 2 to 3 nm thick shell [26, 66, 106]. CO_2 fixation assay of purified *H. neapolitanus* carboxysomes, and SDS polyacrylamide gel electrophoresis (SDS-PAGE) and immunochemical analysis revealed that RuBisCO is packaged inside the carboxysomes [25, 26, 105, 106]. To reflect these findings, these inclusion bodies were named carboxysomes.

Carboxysomes are divided into two types according to their protein components and the organization of their genes [8, 29, 30]. The α-carboxysomes are found in α-cyanobacteria (*Prochlorococcus* species, marine *Synechococcus* WH 8102) and in chemolithoautotrophs. Form IA RuBisCO enzyme is sequestrated inside α-carboxysomes, and all carboxysomal genes are organized into a *cso* operon (Figure 3) [30, 122]. The β-carboxysomes contain Form IB RuBisCO molecules, and their carboxysome genes are localized in several gene clusters in the genome of β-cyanobacteria, such as *Synechocystis* and freshwater *Synechococcus* species (Figure 3) [28, 29, 97].

Highly purified *H. neapolitanus* carboxysomes are composed of ten polypeptides. The two most abundant ones are the large and small subunit of RuBisCO (Figure 2C), which together account for approximately 60% of the total carboxysomal protein

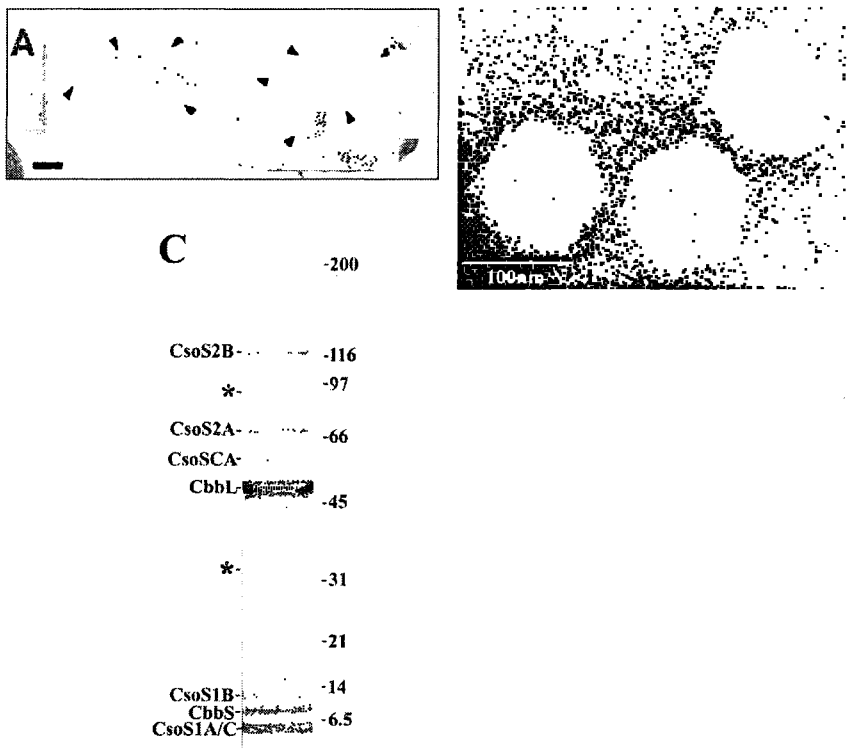


Figure 2. Morphology and composition of wild type *H. neapolitanus* carboxysomes.

A) Thin section image of a *H. neapolitanus* cell, carboxysomes pointed out by arrowheads [28]. Bar, 100 nm. B) Negatively stained homogenous preparation of carboxysomes from *H. neapolitanus*. Individual RuBisCO holoenzyme molecules can be seen inside [41]. C) Purified carboxysomes from *H. neapolitanus* separated by SDS-PAGE and stained with Coomassie Blue dye [41]. The protein bands labeled with asterisks are carboxysome protein aggregates that are resistant to disruption by SDS [56].

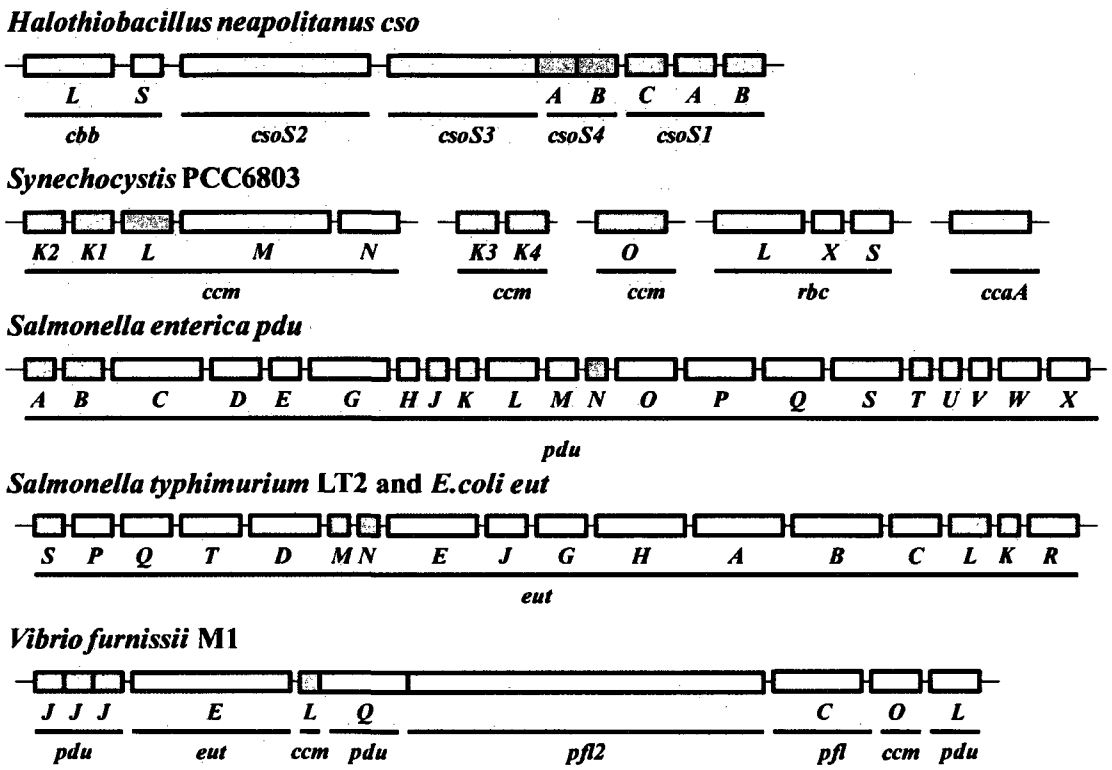


Figure 3. Genomic organization of the genes for carboxysomes and related bacterial microcompartments in several different bacterial species.

Homologs of the *csoS1* genes are labeled in red, and homologs of the *csoS4* genes are labeled in green. Homologs of the *csoS1* and *csoS4* gene family are widely distributed in the prokaryotes. Redrawn from Cannon *et al.* [30, 129].

in mass [26, 59]. In 1994, a 98 amino acid polypeptide, referred to as CsoS1A (carboxysomal shell protein 1A), was assigned to an open reading frame in the *cso* operon, and two genes that are highly homologous to *csoS1A* were also found in the same operon and named *csoS1B* and *csoS1C* [45]. There is only two amino acids difference between CsoS1A and CsoS1C proteins. CsoS1B protein carries a 12-amino-acid extension compared to CsoS1A and CsoS1C proteins, but the rest of protein sequence is highly homologous to the other two CsoS1 paralogs [109]. Two small overlapping open reading frames, *csoS4A* and *csoS4B* (formerly *orfA* and *orfB*, respectively), are located upstream from the *csoS1* gene family. For a long time, CsoS4A and CsoS4B proteins could not be identified in purified carboxysomes because the molecular weights of these two proteins are similar to those of CsoS1A and CsoS1C, and their abundance is extremely low. Recently, both proteins were shown to exist in carboxysomes by immunoblotting following two-dimensional SDS-PAGE of a carboxysome shell-enriched fraction (F. Cai, personal communication). Three additional carboxysomal proteins, CsoS2A, CsoS2B, and CsoSCA, have been identified in the microcompartments of *H. neapolitanus* through reverse genetics. The CsoS2 proteins are the largest two protein components in purified carboxysomes of *H. neapolitanus*. Both polypeptides are encoded by the *csoS2* gene. The apparent molecular weights of these two polypeptides (85 and 130 kDa) deduced from SDS-PAGE are not close to the calculated molecular weight of the predicted *csoS2* gene product (92 kDa). This discrepancy is attributed to differential post-translational glycosylation of CsoS2A and CsoS2B, since both polypeptides have the same 15 amino acids at their N-terminus [13]. Immunoelectron microscopic studies revealed that the CsoS2 proteins are localized at the periphery of the carboxysomes [13].

After a freeze-thaw treatment of purified carboxysomes, CsoS2 protein strongly associates with the carboxysomal shell [114, 127], indicating that this protein is a carboxysomal shell-associated protein. The CsoSCA protein (formerly named CsoS3) also associates with carboxysome shell tightly. This protein is the carboxysomal carbonic anhydrase [114].

Although cyanobacterial carboxysomes have not been purified to homogeneity, carboxysome-enriched fractions have been analyzed by SDS-PAGE, immunoblotting and mass spectrometry. In *Synechococcus* WH 8102, the large subunit of RuBisCO and CsoS2 protein were enriched in a particulate fraction that contains partially purified carboxysomes [51]. It is noteworthy that the CsoS2 protein also was identified as two polypeptides by immunoblotting analysis [51]. An anti-CcmK antibody recognized a polypeptide of an observed molecular weight approximately 10 kDa that is also enriched in the particulate fraction of *Synechococcus* WH 8102 [51]. That polypeptide may represent CsoS1 protein, given that CsoS1 and CcmK proteins are highly homologous [51, 68]. Therefore, RuBisCO, CsoS2 and CsoS1 proteins were likely associated with the carboxysomes of *Synechococcus* WH 8102. Numerous membrane-associated proteins were identified in the carboxysome-enriched fraction [51], therefore, the composition of *Synechococcus* WH 8102 carboxysomes and their stoichiometry can not be revealed.

In β-type cyanobacterial carboxysomes, CcmK protein is a putative carboxysomal shell protein because it is highly homologous with CsoS1 protein [68]. The *ccaA* (formerly as *icfA*) gene, which is located at a distance from the KLMN clusters (encoding putative β-carboxysomal shell proteins), encodes a carbonic anhydrase that has been proposed to be the β-carboxysomal equivalent of CsoSCA [90, 112, 113].

Immunoblotting analysis of subcellular fractions showed that CcaA activity co-purified with the carboxysome-enriched fractions from *Synechocystis* PCC 6803 [112] and *Synechococcus* PCC 7942 [92]. Bioinformatics analysis revealed an interesting characteristic of the putative shell protein CcmM: its N-terminus shows sequence similarity to γ-type carbonic anhydrases; the repeated motifs in its C-terminus are related to the small subunit of RuBisCO [91]. Like CsoS2 protein, CcmM protein migrates as two separate polypeptide bands of apparent molecular weights of 58 kDa and 35 kDa, respectively, on SDS-PAGE. Mass spectrometric studies indicated that the smaller protein of 35 kDa is an N-terminally truncated form of the full length 58 kDa CcmM protein. Further investigations found that the existence of the shorter CcmM polypeptide is independent of the presence of protease [17, 44, 128].

Yeast two-hybrid analysis revealed that CcmM protein interacts with CcmK, CcmL, CcmN, and CcaA proteins from *Synechocystis* PCC 6803, and that it does not show any interactions with either of the RuBisCO subunits, although its C-terminal end has several RbcS-like repeat motifs [129]. A multiprotein complex that constitutes CcaA, CcmM, the large and small subunit of RuBisCO was trapped on an affinity column conjugated with recombinant FLAG-tagged CcmM protein from the cell extracts [37]. I*n vitro* binding assays also revealed interactions between recombinant CcmM protein and bicarbonate [37]. The trapped protein complex was named bicarbonate dehydration complexes (BDC) and speculated to facilitate the dehydration of cytosolic HCO_3^- into CO_2, the substrate of the carboxysomal RuBisCO enzyme [37]. However, given that the cyanobacterial carboxysomes always associate with the internal membrane, the trapped protein complex could also be contaminated with the membrane. The protein components

identified in the trapped protein complex may nonspecifically bind to the membrane to give an artificial and inaccurate result.

In addition, a multiprotein complex from *Synechococcus* PCC 7942 that is composed of CcaA, CcmM, and multiple molecules of type IB RuBisCO was purified on the nickel-charged immobilized metal affinity chromatography via the His-tagged CcmM in the presence of 1 M urea [75]. The authors emphasized the importance of the urea in helping the added 6x His tag at the N- or C-terminus of CcmM protein bind to the Ni-NTA column [75], however, the low concentration of denaturant may disrupt the weak interactions among proteins or slightly change the conformation of some proteins to give an interaction pattern that is different from that *in vivo*. In addition, the CcmM protein was overexpressed in *Synechococcus* PCC 7942 CcmM deletion mutant, which may also result in an interaction pattern that is different from the *bona fide* interaction pattern under the physiological levels of CcmM protein.

Structural Insights into Carboxysomes

In recent years, X-ray crystallography and cryoelectron microscopy techniques have made significant progress towards increasing our understanding of the structure of carboxysomes and carboxysomal components. In 2005, Kerfeld *et al.* resolved the first three dimensional structure of two putative β-carboxysomal shell proteins (CcmK2 and CcmK4). Both proteins crystallize as hexamers, which are believed to assemble the facets of the carboxysome shell [68]. One of the three homologous proteins in α-carboxysomes, CsoS1A, also forms hexameric assemblies that fit together to form a tightly packed molecular layer [125]. Interestingly, the CsoS1A hexamers pack much more tightly than

CcmK hexamers. Interactions between the hexamers are stabilized by intermolecular hydrogen bonds and highly complementary van der Waals surfaces [125]. The two sides of CsoS1A and CcmK hexameric structures have opposite charge distributions. The concave side of the hexamers has a mostly positive electrostatic potential, whereas the convex side is mostly negatively charged [68, 125]. A Gly-Gly-Gly motif that is conserved in the α-carboxysomal CsoS1 paralogs and orthologs, but not in the β-carboxysomal CcmK proteins, was found to form the central pore of the CsoS1A hexamers [125]. Sulfate ions that co-crystallize with CsoS1A hexamers were found to reside in the pores [125]. The negatively charged sulfate oxygen atoms form hydrogen bonds with the backbone amide nitrogen and possibly with the C-alpha hydrogen atom of Gly43, which is the central amino acid in the Gly-Gly-Gly motif [125]. Although bicarbonate, PGA and RuBP could not be co-crystallized with CsoS1A, it was hypothesized that these small, negatively charged metabolites may be transported across the carboxysome shell through the pores [68, 125]. Although the bicarbonate ion is very similar to the sulfate ion in shape and size [125], RuBP and PGA molecules should be larger because they are five-membered sugar bisphosphate and three-membered acid monophosphate, respectively. Thus, it is possible that the central pores are too small to accommodate RuBP and PGA molecules. The sidedness of the CsoS1A hexamers in the carboxysome shell is still unknown. That piece of information can help elucidate the charge distribution of the outer surface of carboxysomes. If the positively charged side is exposed at the outer surface of the carboxysomes, it could attract the cytosolic bicarbonate ion and facilitate the delivery of HCO_3^- through the central positively charged pore.

Subsequently, CsoS4A (formerly OrfA) and its β-cyanobacterial homolog CcmL were determined to crystallize as pentametric structures with a narrow central pore. The pore diameter in CsoS4A pentamers is approximately 3.5 Å, which is only slightly smaller than that of the CsoS1A hexamer pore [125, 129]. However, given the low abundance of CsoS4A and CsoS4B in the carboxysomes of *H. neapolitanus* (F. Cai, personal communication), the CsoS4 pores may not make a major contribution to the transfer of metabolites across the shell.

The powerful cryoelectron tomography technique, coupled with scanning electron microscopy and image reconstruction revealed the three dimensional shape of the α-carboxysomes from *H. neapolitanus* and *Synechococcus* WH 8102 to be icosahedral [61, 101]. In *Synechococcus* WH 8102 carboxysomes, RuBisCO enzymes were found to be distributed into several concentric layers from the periphery to the interior. The holoenzyme molecules were predicted to move freely inside the microcompartment based on computational prediction, suggesting that there are no strong interactions between RuBisCO enzyme molecules and the carboxysome shell.

Combining evidence obtained from microscopic studies of carboxysomes and crystallographic structural information for individual carboxysome shell proteins, a first atomic level model of the α-carboxysome shell was constructed [124]. The CsoS1A hexameric structures were proposed to form sheets that constitute the facets of the icosahedral bodies and the CsoS4A pentameric structures are thought to occupy the vertices [124]. Clearly, this model oversimplifies the structure of the *H. neapolitanus* carboxysomal shell. The major shell building blocks are CsoS1A, CsoS1B, and CsoS1C. These three proteins have a similar abundance in the carboxysome (F. Cai, personal

communication). Because of their > 90% homology at the primary structure level, both CsoS1A and CsoS1C likely form similar hexameric structures and may contribute to the facets of the carboxysome shell.

CsoS1B protein has a unique C-terminal 12-amino acid tail that is not present in CsoS1A and CsoS1C [30]. Densitometric analysis of the polypeptides present in purified carboxysomes of *H. neapolitanus* revealed that the copy number of CsoS1B is approximately one sixth that of total CsoS1A and CsoS1C proteins [56, 127], and it is still unknown whether CsoS1B protein can form hexamers. The C-terminal tail may mediate the interactions between CsoSCA protein and the carboxysomal shell because yeast two-hybrid analysis revealed that CsoSCA protein only interacts with CsoS1B, not with CsoS1A and CsoS1C proteins [127].

CsoS4A and CsoS4B are both expressed in the purified *H. neapolitanus* carboxysomes but very low abundance. Although the crystal structure of CsoS4B is still unknown, this protein also likely forms pentamers because it is very similar in primary structure to CsoS4A. A *H. neapolitanus csoS4AB::Km* deletion mutant that is deficient in both *csoS4* genes forms some elongated carboxysomes. Remarkably, regular, apparently icosahedral carboxysomes of 100 to 120 nm in diameter were observed in the transmission electron microscope (F. Cai, personal communication). These findings emphasized the importance of CsoS4 protein in maintaining regular structure of carboxysomes, but called into question the model that places CsoS4 protein on the vertices of the icosahedral carboxysome structures.

CsoS2 and CsoSCA proteins are two carboxysome shell-associated components [13, 14, 114]. CsoSCA was first identified and categorized as a new class of carbonic

anhydrase due to the lack of homology with known carbonic anhydrases [114]. Recent crystallographic studies revealed that CsoSCA protein forms a homodimer and constitutes a subclass of β-carbonic anhydrases [100]. Each CsoSCA monomer is composed of three domains, an N-terminal domain, a catalytic middle domain and a C-terminal domain [100]. A zinc ion is an important cofactor for the carbonic anhydrase activity of CsoSCA protein [57, 100]. Most β-CAs utilize a pair of active sites for the catalysis, CsoSCA protein only has a single active site. The C-terminal domain is structurally similar to the catalytic domain, but does not bind zinc ion that is coordinated by three residues Cys-173, His-242 and Cys-253 [100]. The bound zinc ion coordinates only the protonated oxygen atom of HCO_3^- that forms a leaving group and interacts with two residues Asp-175 and Ala-254 by a pair of hydrogen bonds. The resulting CO_2 molecule diffuses from the active site as the substrate of RuBisCO enzyme [100]. Crystal packing analysis suggested that CsoSCA protein exists within the carboxysome shell either as an individual homodimer or as extended filaments [57, 100]. The structure of CsoS2 protein has not been resolved yet. Given its high isoelectric point, this protein carries a positive charge in the cytoplasm [30, 57] and is a potential candidate for mediating the transfer of the negatively charged RuBP and PGA across the carboxysome shell.

Biogenesis and Protein-Protein Interactions of Carboxysomes

It is still unknown how the individual protein components self-assemble to form carboxysomes. Thus far, there are two hypotheses for the biogenesis of carboxysomes. 1) RuBisCO molecules are first assembled into a core; carboxysome shell proteins are

recruited to the core to form the shell. Transmission electron microscopy revealed that the presumed β-carboxysome precursors contain regular arrays of RuBisCO in electron micrographs of *Synechococcus* sp. strain PCC 7942, which could be self-assembled RuBisCO core structures that represent intermediates of carboxysome biogenesis [86]. 2) Alternatively, it is possible that the carboxysome shell is formed prior to or concurrently with insertion of RuBisCO molecules. This process is not necessarily dependent on the presence of RuBisCO molecules. The fact that inactivation of the RuBisCO large subunit gene in *H. neapolitanus* results in a high CO_2 requiring mutant that contains "empty" carboxysome shell-like structures supports the latter hypothesis [12]. Given that the carboxysomes from *Synechococcus* PCC 7942 and *H. neapolitanus* belong to the α- and β-type, respectively, it is possible that both types of carboxysomes adopt different assembly pathways *in vivo*.

Elucidation of carboxysome biogenesis must await a combined biochemical, genetic and morphological analysis. Studying interactions among individual carboxysomal proteins is helpful in providing useful information for carboxysome assembly studies *in vitro*. In *H. neapolitanus*, a yeast two-hybrid screen was performed to detect protein-protein interactions among carboxysomal components [127]. CsoS2 protein showed strong interactions with the carboxysomal shell proteins CsoS4B and CsoS1C and with the small subunit of RuBisCO. CsoSCA protein was found to interact only with CsoS1B protein, but not with CsoS1A and CsoS1C. Bacterial two-hybrid approaches have identified that CsoS2 protein strongly interacts with CsoS4A protein, not with any of the CsoS1proteins, and that CsoS3 protein does not bind any carboxysomal shell protein in *Synechococcus* WH 8102 [51]. It is worthy to emphasize

that bacterial two-hybrid screen did not detect the well-documented interactions between large and small subunit of RuBisCO in *Synechococcus* WH 8102, and yeast two-hybrid analysis did not identify the interactions between large subunits of RuBisCO in *H. neapolitanus* [51]. Therefore, yeast and bacterial two-hybrid analyses have shortcomings to analyze all protein-protein interactions among carboxysomal proteins. Recently, a small sequential peptide affinity tag was adopted to tag the C-termini of the target proteins to purify multiprotein complexes in *E. coli* [131]. SDS-PAGE and mass spectrometry analyses were used to identify the interacting partners in the purified complexes. A similar approach will be adopted to analyze *bona fide* protein-protein interactions among *H. neapolitanus* carboxysomal proteins *in vivo*.

The elucidation of interactions between RuBisCO molecules and the carboxysome shell is helpful towards gaining an understanding of the pathway by which RuBisCO molecules are encapsulated into the interior of carboxysomes. In the α-carboxysomes of *H. neapolitanus*, most of the carboxysomal RuBisCO can be released by freeze-thaw treatment-induced breakage of purified microcompartments; however, a small amount of RuBisCO molecules always appears to remain tightly associated with the carboxysome shell [57, 127]. Similarly, in *Synechocystis* PCC 6803, the encapsulated RuBisCO molecules can be divided into two discrete populations after freeze-thaw treatment of carboxysome-enriched fractions. One RuBisCO population was found in the carboxysome shell-enriched pellet; another population is released into the resulting supernatant and was thought to be loosely associated with the shell [37]. Contrary to these findings, cryoelectron microscopy and computational simulation studies did not find specific contacts between the carboxysomal shell and RuBisCO molecules in

Synechococcus WH 8102. Recently, Menon *et al.* reported that carboxysome shell-like structures were purified from the *cbbLS::kan*^{r} mutant, which does not produce any Form I RuBisCO [80]. The purified shell-like structures resembled empty carboxysomes in the transmission electron microscope, and SDS-PAGE and immunoblotting analyses revealed that they are composed of all carboxysomal proteins except the large and small subunit of RuBisCO [80]. These findings indicated that shell assembly is independent of the presence of RuBisCO in α-carboxysomes and strongly supports the second hypothesis for the biogenesis of carboxysomes.

The biogenesis of carboxysomes needs to be further investigated. Identification of *bona fide* protein-protein interactions in purified carboxysomes and trapping and biochemical characterization of carboxysomal intermediates are needed to have an impact on the understanding of the *in vivo* and *in vitro* carboxysome assembly process.

Other Carboxysome-like Microcompartments

It was thought for many years that carboxysomes are the only polyhedral microcompartments. However, subsequently *Salmonella enterica* was found to form polyhedral bodies when grown in the presence of 1,2-propanediol (1,2-PD) [19, 110]. The *S. enterica* microcompartments are thought to aid in the catabolism of 1,2-PD. Later, more bacteria were identified to form carboxysome-like microcompartments, such as *Klebsiella*, *Shigella, Yersinia* and one strain of *Escherichia* [110]. Genetic analysis of 1,2-PD degradation in *S. enterica* identified a gene cluster that was named propanediol utilization *(pdu)* operon. The operon encodes 23 open reading frames. Among these, the PduA and PduJ proteins are homologs of carboxysomal shell proteins CsoS1, CcmK and

CcmO [20]. The PduK and PduT proteins are only approximately 25 to 30% identical to CsoS1 proteins and are putative shell proteins [19, 96]. The PduN protein is 46% identical to carboxysome protein CcmL from *Synechococcus elongates* PCC6301 and distantly related to CsoS4A and CsoS4B from *H. neapolitanus* [28, 129]. In 2003, Havemann *et al.* isolated the *pdu* microcompartments from *S. enterica*, which are very similar to carboxysomes in shape and size, but are more irregular than carboxysomes [54, 55]. SDS-PAGE and immunoelectron microscopy analyses revealed that the purified *pdu* microcompartments are composed of 15 different polypeptides, of which coenzyme B_{12}-dependent diol dehydratase is a major component [54, 55]. Later, PduO protein of *S. enterica* was identified as an ATP:cob(I)alamin adenosyltransferase that catalyzes the final step in the conversion of vitamin B_{12} to coenzyme B_{12} [63]. The PduL protein is a phosphotransacylase (PTAC) involved in 1,2-PD degradation [74]. Recently, the PduX protein of *Salmonella enterica* was shown to be a L-threonine kinase that participates in the *de novo* synthesis of coenzyme B_{12} [48]. Similar *pdu* microcompartments were also observed in *Lactobacillus reuteri* [116]. *S. enterica* grown in the presence of ethanolamine also forms microcompartments that are encoded by the ethanolamine utilization (*eut*) operon [21, 69]. Recently, recombinant EutT protein was overexpressed in *E. coli* and characterized as an ATP:cob(I)alamin adenosyltransferase [22]. Genetic and biochemical analyses revealed that EutD protein has phosphotransacetylase activity [117]. To date, *eut* microcompartments have not been isolated, and little information is available on their composition and function(s).

The precise role of these microcompartments in bacterial metabolism is still unknown. One suggestion is that the microcompartments can prevent the leakage of toxic

metabolic intermediates from the degradation of 1,2-PD and ethanolamine and thereby protect the cells. Alternatively, compartmentalization of volatile intermediates may prevent loss of fixed carbon [20, 21, 87, 118]. Another possible role of microcompartments is the sequestration of mutagenic metabolites, such as those generated during degradation of 1,2-PD, and thereby reduce the mispairing of DNA bases [99]. Given the existence of microcompartments in many autotrophic bacteria and some enteric bacteria, a bioinformatics study was performed to assess the distribution the shell protein genes of bacterial polyhedral microcompartments (Figure 3). The major shell proteins of bacterial microcompartments contain a conserved sequence known as the bacterial microcompartment (BMC) domain [129]. More than 189 bacterial species and strains have BMC gene clusters, among the Actinobacteria, Acidobacteria, Firmicutes, Planctomycetes and Fusobacteria, and all groups of the proteobacteria, except for the epsilon class [129]. The wide distribution of BMC domains suggests that packaging of different metabolic enzymes into microcompartments may represent a general strategy among prokaryotes that increases the efficiency of metabolic reactions, protects the cells from toxic metabolic intermediates, or harbors some novel metabolic functions [129]. For example, the *Vibrio furnissii* M1 strain contains a BMC gene cluster in its genome and can form microcompartments that metabolize pyruvate to ethanol [126]. A novel thermoacidophilic bacterium belonging to the Verrucomicrobia was observed to be filled with polyhedral bodies resembling carboxysomes under transmission electron microscopy. These inclusions may represent a novel subcellular bacterial "organelle" for methane oxidation [62].

The discovery of a smaller and simpler polyhedral organelle by Sutter and colleagues has broadened the scope of microcompartments [120]. The novel nanocompartment of approximately 24 nm diameter, only one fourth that of carboxysomes, was isolated and characterized from *Thermotoga maritime*. Its icosahedral shells are composed of twelve pentamers formed by the protein encapsulin. DyP, an iron-dependent peroxidase, or Flp, a homolog of the iron transporter ferritin, is packaged into the interior of the nanocompartment [120]. It was hypothesized that encapsulation of DyP proteins may metabolize harmful reactive oxygen species and protect the cells from oxidative damage [58, 120].

Objective of Study

Carboxysomes play an important role in bacterial CO_2 fixation as the terminal step of the CCM. The functions of carboxysomes have been characterized by CO_2 fixation kinetics *in vitro* and *in vivo*, structural and mutational analyses, however relatively little information is available about the transfer of RuBisCO metabolites across the carboxysomal shell and about carboxysome assembly. Recent investigations identified that CsoSCA protein is a novel type of carboxysomal shell-associated carbonic anhydrase and suggested that the carboxysomal shell is a diffusion barrier for CO_2 molecules [57, 114]. The goal of this study is to further investigate the permeability of the shell for the two competing RuBisCO substrates CO_2 and O_2 and to understand how packaging RuBisCO into microcompartments enhances its catalytic efficiency. CO_2 fixation analysis of carboxysomes from the *H. neapolitanus* CsoSCA insertion mutant was performed to assess the effect of the carbonic anhydrase on the catalytic efficiency of

the encapsulated RuBisCO. The permeability of the carboxysomal shell for O_2 molecules was measured to test whether RuBisCO that is sequestered into a microcompartment favors carboxylation over the oxygenation reaction.

In addition, this study was conducted to provide preliminary information to enhance the understanding of the structure of carboxysomes and *in vitro* carboxysome assembly. CsoS2 protein was expressed in *E. coli* as a prerequisite to structure determination. A small sequential peptide affinity (SPA) tag was added at the C-terminus of the protein to understand the formation of two CsoS2 proteins in *H. neapolitanus* and assess the location of CsoS2 protein in the carboxysome shell *in vivo*. Finally, intermediates of *in vitro* assembled carboxysome shell protein complexes were examined and the entire *cso* operon was expressed in *E. coli* to further probe the role of RuBisCO in carboxysome shell assembly.

CHAPTER III

EXPERIMENTAL PROCEDURES

Materials

Anti-RbcL (RuBisCO large subunit Form I and Form II) antibody produced in rabbit was obtained from Agrisera, Switzerland. Aminex® HPX-87H organic acid analytical HPLC column, agarose, 40% (29:1) acrylamide/bis, ammonium persulfate (APS), Bradford protein assay reagents, Criterion pre-cast polyacrylamide gels, glycine, Micro Bio-Spin™ 6 gel filtration column, nitrocellulose, Triton X-100, N,N,N',N'-Tetramethylethylenediamine (TEMED) were obtained from BioRad, Hercules, CA. Polymax-T and D-19 developer solutions were purchased from Eastman-Kodak, Rochester, NY. Dithiothreitol (DTT) and isopropyl-β-D-thiogalactoside (IPTG) was obtained from EMD Biosciences, San Diego, CA. Copper 300 mesh EMBRA finder grids, Formvar resin and 70mm Kodak electron microscopy film were purchased from Electron Microscopy Science, Ft. Washington, PA. Calmodulin Sepharose™ 4B resin, Mono Q 5/50 GL strong anion exchange column and aqueous sodium [^{14}C]bicarbonate solution were purchased from GE Healthcare, Piscataway, NJ. Oligonucleotide primers were ordered from Integrated DNA Technologies (IDT), Coralville, IA. One Shot TOP 10 Chemically Competent *E.coli* cells and Ni-NTA resin were obtained from Invitrogen, Carlsbad, CA. MICROCON® centrifugal filter devices were purchased from Millipore, Bedford, MA. Aqueous 2-[^{3}H]-D-Glucose solution was purchased from Moravek Biochemicals, Brea, CA. Calf Intestinal Phosphatase (CIP), 1kb DNA ladder, restriction enzymes and T4 DNA ligase were obtained from New England Biolabs, Beverly, MA. Bacterial Protein Extraction reagent II (B-PER II), BCA (bicinchoninic acid) protein

assay reagents, CL-Xposure film, Gel Code Blue gel stain reagents, 1-step 4-nitroblue tetrazolium chloride/5-bromo-4-chloro-3-indolyl phosphate (NBT/BCIP) substrate solution and SuperSignal® West Pico Chemiluminescent Substrate for Horseradish Peroxidase (HRP) detection were purchased from Pierce, Rockford, IL. QIAprep spin miniprep and QIAquick PCR purification kits were purchased from Qiagen Sciences, Valencia, CA. Anti-Rabbit IgG HRP conjugated antibody generated in goat was obtained from Santa Cruz Biotechnology, Santa Cruz, CA. Adenosine triphosphate (ATP), ammonium molybdate, ampicillin, anti-rabbit IgG (whole molecule)–Alkaline Phosphatase antibody produced in goat, anti-FLAG® monoclonal antibody generated in rabbit, anti-FLAG® M2 (monoclonal anti-FLAG antibodies from mouse)-Agarose beads, N,N-Bis(2-hydroxyethyl)glycine (BICINE), 1,3-bis-propane-HCl, dimethyl sulfoxide (DMSO), DNase I, 4-(2-hydroxyethyl)-1-piperazine ethane sulfonic acid (HEPES), imidazole, kanamycin, 2-mercaptoethanol (β-ME), 2-(N-morpholino)ethanesulfonic acid (MES), nicotinamide adenine dinucleotide (NADH), nicotinamide adenine dinucleotide phosphate (NADP), phenylmethylsulfonyl fluoride (PMSF), p-toloenesulfonyl fluoride (PTSF), phenol red, phosphocreatine, reactive green 19 agarose beads, D-ribulose 1,5-bisphosphate (RuBP) (Fluka brand), D-ribulose- 5-phosphate (RuMP), rabbit muscle creatine phosphokinase (EC 2.7.3.2), sodium azide, yeast glucose-6-phosphate dehydrogenase (EC 1.1.1.49), yeast hexokinase (EC 2.7.1.1), yeast 6-phosphogluconate dehydrogenase (EC 1.1.1.44) were obtained from Sigma, St. Louis, MO. Spectrapor dialysis membranes were obtained from Spectrum Laboratories, Rancho Dominguez, CA. Pfu Turbo DNA polymerase was obtained from Stratagene, La Jolla, CA. Deoxynucleotide triphosphate (dNTP) mixtures for PCR reactions were purchased from

TaKaRa Bio, Otsu, Japan. All other chemicals were obtained from Fisher Chemical Company, Fairlawn, NJ.

Table 1. Primers

Primer name	Primer sequence
Csos2FWD	GGTGGT***CATATG***CCTTCACAGTCAGG
Csos2REVInt	CATTTACAGAGGAGCGCACC
Csos2REV	GGTGGTT***GCTCTTCCGCA***ACCGCGCGCGCCGCCGGAGTAAGTGATCAGACTACC
S2SPAFWD	GTTGTT***AAGCTT***GACACCCAAGGTAGTCTGATCACTTACTCCGGCGGCGCGCGCGGTTCCA TGGAAAAGAGAAG
S2SPAREV	GTTGTT***GGTACC***TGCTTCGTGTGTTACGGGTGTTCATGATCGTTACACTTTACTTAACATAT GAATATCCTCCTTAG
S2f4018	ACACCAATCAGCCAGAAC
SPAf2212	GGAAAAGAGAAGATGGAAAAAG
S3r4851	CGGTTCTTGACGCTATCC
SPAr3204	GGATACTTTCTCGGCAGG

Underlined sequences are homologous with the 5' and 3' flanking regions of the stop codon of the *csoS2* gene in *Halothiobacillus neapolitanus*.

Italicized and bold sequences represent the Nde I, Sap I, Hind III or Kpn I restriction sites used for cloning purposes.

Routine Apparatus

Protein electrophoresis was performed using BioRad Criterion or Mini-Protean III cells. Protein transfer to nitrocellulose membrane was performed using the BioRad Mini-Transblot or Criterion blotter. Nucleic acid electrophoresis was performed using GibcoBRL Horizon 58 Model 200 and IBI Model MPH multipurpose gel electrophoresis cells. *E. coli* liquid cultures were grown in a New Brunswick Series 25 incubator shaker. *H. neapolitanus* liquid cultures were grown in an INFORS-HT Multitron 2 incubator shaker or INFORS-HT Minifors or Labfors bioreactors. *E. coli* and *H. neapolitanus* plates were grown in a Curtin Matheson Scientific Inc. Equatherm 1572 incubator and a Fisher Scientific Isotemp Plus incubator. Bacterial cells were disrupted using a Branson sonifier 450 sonicator or a SLM Aminco French Pressure cell. Centrifugations were performed in a Beckman Coulter Avanti J-30I, J26 XP, L7-65 Ultracentrifuge, Beckman Coulter Optima™ Ultracentrifuge, Eppendorf 5417C or 5417R centrifuge. PCR reactions were performed using a BioRad MyCycler™ thermal cycler. Cell lines and samples were stored in a Puffer Hubbard -80°C freezer. Spectrometric absorbance determination was performed with a Nanodrop ND-1000 or a Beckman Coulter DU800 spectrometer. Transmission electron microscopy was performed using a Zeiss EM109 electron microscope. Deionized water was obtained from a Barnstead NANOpure DIamond™ system. The pH of solutions was measured using an ORION benchtop 720A+ advanced pH/Ion meter. Labwares and solutions were sterilized in Steris Amsco Lab 250 and 3031-S autoclaves. The radioactivity of samples was measured in a Beckman Coulter LS-6500 Multi-Purpose scintillation counter. The sizes of proteins or protein complexes were measured with a Malvern Zetasizer nano series dynamic light scattering detector. Protein

and DNA sequence analyses and alignments were performed using OMIGA and ClustalW software, respectively. Images of gels and immunoblots were documented and analyzed with the Quantity One program supplied with the BioRad VersaDoc model 4000 MP imaging system.

Bacterial Growth Media

Antibotic stock solutions
100 mg/ml ampicilin in water (final concentration in LB medium is 100 μg/ml)
50 mg/ml kanamycin in water (final concentration in LB medium is 50 μg/ml)
34 mg/ml chloramphenicol in ethanol (final concentration in LB medium is 34 μg/ml)
Antibotic solutions were sterilized by filtering through a 0.2 μm filter (Millipore). They were added to the sterile medium after cooling to 40°C.

Luria-Bertani (LB) medium
10 g/L tryptone
5 g/L yeast extract
10 g/L NaCl
The medium was autoclaved for 20 minutes at 121°C and 15 psi.

Halothiobacillus neapolitanus liquid medium
4 g/L KH_2PO_4
4 g/L K_2HPO_4
0.4 g/L NH_4Cl
0.4 g/L $MgSO_4$
10 g/L $Na_2S_2O_3 \cdot 5H_2O$
10 ml/L trace element solution
4 ml/L 1% (m/v) phenol red solution if needed
The medium was adjusted to pH 6.8 with 1 M KOH if needed, prior to autoclaving for 20 minutes at 121°C and 15 psi. For 8 L media, the autoclave time was increased to 90 minutes.

H. neapolitanus solid medium
0.2 g/L KH_2PO_4
0.8 g/L K_2HPO_4
0.01 g/L $CaCl_2$
1 g/L NH_4Cl
0.24 g/L $MgSO_4$

10 g/L $Na_2S_2O_3 \cdot 5H_2O$
15 g/L agar
1 ml/L trace element solution
4 ml/L 1% (m/v) phenol red solution
The medium was adjusted to pH 6.8 with 1 M KOH if needed, prior to autoclaving for 20 minutes at 121°C and 15 psi.

H. neapolitanus trace element solution
50 g/L EDTA
5.44 g/L $CaCl_2$
1.61 g/L $CoCl_2$
1.57 g/L $CuSO_4 \cdot 5H_2O$
4.99 g/L $FeSO_4 \cdot 7H_2O$
5.06 g/L $MnCl_2 \cdot 4H_2O$
1.10 g/L $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$
2.20 g/L $ZnSO_4 \cdot 7H_2O$
EDTA was dissolved in water first and the pH was brought to 6.0 with 20% (w/v) KOH. Next, the remaining chemicals were added sequentially as listed and the pH was adjusted to 6.0 with 20% (w/v) KOH. Finally, the pH was brought to 6.8 with 20% (w/v) KOH and the solution was stored at 4°C prior to use.

SOC medium
20 g/L tryptone
5 g/L yeast extract
0.584 g/L NaCl
0.186 g/L KCl
0.952 g/L $MgCl_2$
1.204 g/L $MgSO_4$
3.603 g/L glucose
The medium was adjusted to pH 7.0 with 1 M KOH if needed, prior to autoclaving for 20 minutes at 121°C and 15 psi.

Buffers

6x DNA gel-loading solution
0.25% (w/v) bromophenol blue
40% sucrose in water

Immunoblot transfer buffer
25 mM Tris-HCl (pH 8.3)
250 mM glycine
20% (v/v) methanol

Phosphate-buffered saline (PBS)
10 mM Na_2HPO_4
2 mM KH_2PO_4
137 mM NaCl
2.7 mM KCl
pH 7.4

Immunoblot blocking buffer
5% nonfat milk in PBS with 0.1% Triton X-100

PEG-$MgCl_2$ solution
40% (w/v) polyethylene glycol (PEG 8,000)
30 mM $MgCl_2$

Phenylmethylsulfonylfluoride(PMSF)/p-toluenesulfonylfluoride(PTSF) stock solution
100 mM PMSF/PTSF dissolved in 95% ethanol

RuBisCO assay buffer
50 mM Bicine-NaOH (pH 8.0)
20 mM $MgCl_2$

1x SDS-PAGE running buffer
25 mM Tris-HCl (pH 8.3)
250 mM glycine
0.1% SDS

4x SDS gel-loading buffer
200 mM Tris-HCl (pH 6.8)
8% (w/v) SDS
0.4% (w/v) bromophenol blue
40% (v/v) glycerol
400 mM DTT or β-mercaptoethanol (β-ME)

1x TAE buffer
40 mM Tris-acetate (pH 8.3)
1 mM EDTA

0.5x TBE buffer
45 mM Tris-borate (pH 8.3)
1 mM EDTA

TE buffer
10 mM Tris-HCl (pH 8.0)
1 mM EDTA

TEMB buffer
10 mM Tris-HCl (pH 8.0)
1 mM EDTA
15 mM $MgCl_2$
20 mM $NaHCO_3$

Bacterial Strains and Growth

Halothiobacillus neapolitanus (ATCC 23641)

Halothiobacillus neapolitanus cultures were maintained on solid media and as liquid cultures. For biochemical characterization, *H. neapolitanus* cells were grown in the bioreactor at 0.08 h^{-1} dilution rate at pH 6.4 and 30°C. Wild type and the HnSPAS2 mutant of *H. neapolitanus* were grown at ambient CO_2 concentration. The *csoS3::Km* mutant was cultured CO_2-enriched air (5% CO_2). The concentration of O_2 is approximately 21% (v/v).

Escherichia coli

E. coli TOP10 and DH5α strains were maintained in liquid and on solid LB medium under the appropriate selection pressure at 37°C for cloning manipulation. The genetically engineered strain DY 330 was grown at 30°C for homologous recombination *in vivo*.

Genotype of *E. coli* strains used in this study [130]:

DH5α: *F- 80dlacZM15 (lacZYA-argF) U169 recA1 endA1 hsdR17(rk-, mk+) phoA supE44 - thi-1 gyrA96 relA1.*

TOP10: *F- mcrA (mrr-hsdRMS-mcrBC) 80lacZM15 lacX74 recA1 ara139 (ara-leu)7697 galU galK rpsL (StrR) endA1 nupG.*

DY330: *W3110 ΔlacU169 gal490 λcI857 Δ(cro-bioA)* (W3110 is a substrain of *E. coli* K-12).

Methods

Genomic DNA preparation from E. coli and H. neapolitanus [5]

A 5 ml overnight *E. coli* culture or a 40 ml exponential phase *H. neapolitanus* culture was harvested and resuspended into 567 µl of TE buffer supplemented with 0.5% SDS and 0.1 mg/ml proteinase K. The resuspension was incubated at 37°C for 1 hour and gently shaken several times during the incubation. Next, 100 µl of 5 M NaCl and 80 µl of CTAB/NaCl solution were added to the resuspension, mixed thoroughly and incubated at 65°C for 10 minutes. An equal volume of phenol/isoamyl alcohol (24:1) was added, mixed until an emulsion formed and spun for 5 minutes in a microcentrifuge at full speed. The aqueous phase was gently removed into a new tube and thoroughly mixed with an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1). After another centrifugation, the aqueous phase was removed, mixed with 0.6 volume of isopropanol and the genomic DNA was allowed to precipitate at room temperature for 10 minutes. The DNA was pelleted by centrifugation at maximum speed for 10 minutes. The resulting genomic DNA pellet was washed with 70% ethanol once and air-dried at room

temperature before being dissolved in 100 µl of TE buffer. The concentration and quality of the isolated genomic DNA were assessed by UV spectrometry and agarose gel electrophoresis.

Mini-preparation of plasmid DNA

A 3 ml *E. coli* overnight culture was pelleted by centrifugation in a microcentrifuge at maximum speed for 1 minute. Plasmid DNA was isolated as described in the manual for the Qiagen plasmid mini-preparation kit (Qiagen, CA). Usually, 3 to 5 µg of plasmid DNA were obtained from 3 ml of *E. coli* culture.

Midi-preparation of plasmid DNA [4]

A 250 ml saturated *E. coli* culture was harvested by centrifugation at 10,000 x *g* for 10 minutes and resuspended in 8 ml of GTE buffer (25 mM Tris-HCl, pH 8.0, 50 mM glucose and 10 mM EDTA). Lysozyme stock solution (25 mg/ml) was added to the bacterial suspension to 2.5 mg/ml and the mixture was incubated at room temperature for 10 minutes. Next, 20 ml of freshly prepared 0.2 M NaOH/1% SDS solution were added to the lysozyme-treated bacteria and gently mixed. At this stage, the solution became clear and viscous. After 10 minutes of incubation on ice, 15 ml of 3 M potassium acetate (pH 5.5) were added and the centrifuge bottle was shaken by hand until the solution became less viscous and white aggregates formed. After another 15 minutes of incubation on ice, the cell debris and protein aggregates were removed by spinning at 20,000 x *g* for 10 minutes at 4°C. The resulting supernatant was transferred to a new tube and 0.6 volume of isopropanol was added with gentle mixing. The mixture was incubated at room

temperature for 10 minutes before the plasmid DNA was pelleted at 15,000 x *g* for 10 minutes at room temperature and washed with 70% ethanol solution once. The resulting DNA pellet was air-dried and dissolved in 500 µl of TE buffer.

Plasmid DNA purification with polyethylene glycol [98]

A 500 µl plasmid DNA solution in TE buffer was transferred to a 1.5 ml microfuge tube and chilled in an ice bath. An equal volume of ice-cold 5 M LiCl was added to the crude plasmid preparation and mixed well. The sample was spun at 12,000 x *g* for 10 minutes at 4°C. The supernatant was removed to a new tube and an equal volume of isopropanol was added. After mixing well, the plasmid DNA was recovered by centrifugation at 12,000 x *g* for 10 minutes at room temperature. The plasmid DNA pellet was washed with 70% ethanol once, air-dried and dissolved in 500 µl of TE buffer. The DNA solution was treated with RNase A at a final concentration of 100 µg/ml for 30 minutes at room temperature. The RNase A was extracted once with phenol/chloroform (24:1) and once with chloroform. The final aqueous phase was transferred to a new tube and 2 volumes of absolute ethanol were added to precipitate the plasmid DNA. The resulting DNA pellet was dissolved in 1 ml of sterile water, and 0.5 ml of PEG-NaCl solution was added prior to a 10 minute incubation at room temperature. Plasmid DNA was pelleted by centrifugation at maximum speed for 20 minutes in a microcentrifuge. Trace amounts of PEG were removed by resuspending the nucleic acid pellet in 500 µl of 70% ethanol. The plasmid DNA was centrifuged at maximum speed for 5 minutes in a microcentrifuge and rinsed with 70% ethanol. Residual ethanol was removed from the resulting pellet by aspiration and the DNA was dissolved in 500 µl of TE buffer.

Concentration and purity of the DNA were assessed by spectrometric analysis and DNA gel electrophoresis.

Polymerase chain reaction (PCR)

PCR reactions were performed in 25 µl thin walled tubes. Each tube contained pfu polymerase buffer, 0.8 mM dNTP mix (0.2 mM dATP, dGTP, dCTP and dTTP each), 1 µM each of forward and reverse primer, 0.6 units pfu Turbo polymerase (Strategene) and 10 ng of plasmid or 100 ng of genomic DNA as template. If needed, additives such as 4-10% DMSO or betaine were included in the PCR reaction. The reaction was initiated by incubation at 95°C for 3 minutes to denature the DNA template completely. A subsequent chain amplification reaction was repeated for 30 cycles with the following setup: 95°C for 30 seconds, T_m (annealing temperature of primers) - 5°C for 30 seconds and 72°C for 1-3 minutes (1 minute per 1 kb DNA target). Finally, the reaction was incubated at 72°C for 10 minutes. The samples were stored at 4°C prior to gel electrophoresis.

Restriction digestion of DNA

Plasmid DNA and PCR products were digested with the target restriction endonuclease(s) in the appropriate buffer supplied by the manufacturer. Usually, for diagnostic enzymatic digestions, 100 to 150 ng of plasmid DNA was digested. For preparative recovery of DNA from agarose gels, 1.0 to 1.5 mg DNA was used for the digestion. The reactions were incubated at 37°C for 1.5 to 2 hours and the products were separated by agarose gel electrophoresis.

Agarose gel electrophoresis

Nucleic acids were separated in 0.6 to 2.0% agarose gels (agarose percentage depended on the size of the target nucleic acid fragment). After electrophoresis, the gel was stained with 0.5 μg/ml ethidium bromide solution for 10 minutes and destained with deionized water for 15 minutes before the nucleic acids were visualized under UV illumination.

Nucleic acid recovery from agarose gels

The stained agarose gel was illuminated with long wavelength UV light and the target band was cut out with a clean razor blade. Nucleic acids were recovered following the instruction for the Zymoclean DNA gel recovery kit (Zymo research corporation).

Ligations

DNA fragments recovered from agarose gels and 150 units of T4 DNA ligase were mixed in the appropriate buffer at different vector to insert ratios in a total volume of 10 μl. Ligation reactions were performed at 16°C overnight. A 1-2 μl aliquot of a ligation reaction was used to transform chemically competent *E. coli* cells.

Transformation in One Shot® TOP10 chemically competent E. coli cells

One 50 μl vial of One Shot® TOP10 chemically competent cells (Invitrogen) was thawed on ice. Ligation reaction products (1 to 2 μl) or intact control plasmid was added to the vial and incubated with competent cells on ice for 30 minutes. The cells were heat-shocked at 42°C for 30 seconds and immediately placed on ice. A 250 μl aliquot of SOC

medium was added to the vial and the cells were agitated at 37°C for 1 hour in a shaking incubator. Aliquots (20 to 200 µl) of the transformed culture were spread onto LB plates containing the appropriate antibiotic. The plates were incubated at 37°C overnight, and colonies were selected and analyzed by plasmid DNA mini preparation, diagnostic enzymatic digestion, PCR amplification and/or DNA sequencing.

Preparation and transformation of electrocompetent H. neapolitanus cells [46]

A 40 ml aliquot of mid-exponential phase *H. neapolitanus* culture from the bioreactor was transferred to a sterile 50 ml centrifuge tube and the cells were pelleted by centrifugation at 12,000 x *g* for 10 minutes at 4°C. The loose cell pellet was carefully resuspended in 40 ml of ice-cold sterile water and spun at 12,000 x *g* for 10 minutes at 4°C. This step was repeated at least three times. The final cell pellet was resuspended in 400 µl of chilled sterile water. Usually, 1 µg of purified plasmid DNA was mixed with 50 µl of freshly prepared electrocompetent *H. neapolitanus* cells and the mixture was loaded into a pre-cooled 0.1 cm gap electroporation cuvette (BioRad). The cuvette was pulsed at 19 kV/cm, 200 Ω and 25 µF. The usual pulse time was around 4.5 ms. The transformation mixture was transferred into 5 ml of ice-cold *H. neapolitanus* medium and placed on ice for 5 minutes before being incubated in a shaker at 30°C for 24 hours in 5% CO_2-enriched air. After 24 hours of recovery time, the cultures were switched to ambient CO_2 for another 24 hours of incubation. A 50 to 200 µl aliquot of transformed *H. neapolitanus* culture was spread onto agar plates with *H. neapolitanus* medium containing 50 µg/ml kanamycin and incubated for 4 to 5 days or longer in 5% CO_2-enriched air until single

colonies could be seen. Individual colonies were selected and analyzed by PCR amplification and genomic DNA sequencing.

Expression and purification of recombinant proteins from E. coli

Phosphoribulokinase (PRK) [31]. E. coli BL21 cells carrying the *Rhodobacter sphaeroides prk* gene in the pET3-d expression vector were grown at 37°C to OD_{600} 0.6 to 0.7. At this point, the culture was induced to produce the recombinant PRK by addition of 1 mM IPTG and incubated for 5 hours at 25°C. The cells were harvested by centrifugation at 12,000 x *g* for 10 minutes at 4°C. The cell pellet was resuspended in 50 ml of lysis buffer (25 mM Tris-HCl, pH 8.2, 1 mM EDTA and 3 mM β-ME) and the cells were broken by passage through a French Press cell twice at 20,000 psi. The lysate was spun at 100,000 x *g* for 1 hour to remove cell debris. The resulting supernatant was dialyzed against two changes of 4 liters of lysis buffer. The dialysate was mixed with 5 ml of Reactive Green 19 agarose beads (Sigma) that had been equilibrated with lysis buffer, and incubated overnight. The beads were loaded into an empty column and washed with 20 bed volumes of lysis buffer. The column was eluted with 40 ml of 10 mM ATP in lysis buffer. The protein concentration in the eluted fractions was estimated with the Bradford protein assay, and the polypeptide composition was estimated by polyacrylamide gel electrophoresis (SDS-PAGE). PRK enzyme activity was assessed with a coupled PRK/CO_2 fixation assay. Each PRK enzyme assay vial contained 80 µl of 50 mM Bicine-NaOH, pH 8.0, 20 mM $MgCl_2$, 20 mM $NaH[^{14}C]O_3$ (specific activity 0.1 µCi/µmol), 1 mM NADH, 5 mM ATP, 0.5 mM ribulose 5-monophosphate (RuMP) and 70 µg of RuBisCO enzyme. The reaction was initiated by adding 20 µl of PRK elution

fraction and quenched after 3 minutes by pipetting 300 µl of glacial acetic acid into each vial. The vials were heated to remove unreacted bicarbonate as CO_2. Radioactivity in acid-stable products was quantified by liquid scintillation counting in ScintiVerse II scintillant (Fisher). The PRK specific activity was converted into international units (µmole of phosphate·min^{-1}·mg $protein^{-1}$). Fractions containing PRK activity were stored at -20°C for the future use.

CsoS2 protein in the pProEX vector (native method). The *E. coli* cells carrying the pProEX-HTb-csoS2 plasmid were grown at 37°C to an OD_{600} of 0.6 to 0.7 and induced with 0.6 mM IPTG. After 4 hours at 37°C, one liter of *E. coli* cells were harvested by centrifugation at 10,000 x *g* for 10 minutes, the cell pellet was resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 5 mM β-ME, 1 mM PMSF/PTSF) and the cells were lysed by 20 30-second cycles of sonication in an ice bath. Cell debris was removed by centrifugation at 20,000 x *g* for 30 minutes. The supernatant was transferred to a new tube and mixed with 3 ml of Ni-NTA beads equilibrated with lysis buffer. The slurry was incubated for at least 3 hours. The beads were washed with 20 bed volumes of low salt washing buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 5 mM β-ME, 10% glycerol, 20 mM imidazole, 0.1 mM PMSF/PTSF), 10 bed volumes of high salt washing buffer (20 mM Tris-HCl, pH 8.0, 1 M NaCl, 5 mM β-ME, 10% glycerol, 0.1 mM PMSF/PTSF) and 5 bed volumes of low salt washing buffer sequentially. The His-tagged CsoS2 protein bound to the beads was eluted with 20 ml of elution buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 5 mM β-ME, 10% glycerol, 250 mM imidazole, 0.1 mM PMSF/PTSF). Ten 2-ml elution fractions were collected and analyzed by SDS-PAGE. All fractions

containing recombinant CsoS2 protein were pooled and dialyzed against two changes of 4 liters of dialysis buffer (10 mM Tris-HCl, pH 8.0, 20 mM NaCl and 10 mM β-ME).

CsoS2 protein in the pProEX vector (denaturation method). E. coli was grown and induced as described for the native method. The cells from one liter of culture were resuspended in 60 ml of lysis buffer (6M guanidine-HCl, 100 mM Na_2HPO_4, pH 8.0) and stirred for 1 hour at room temperature. Cell debris was removed by centrifugation at 20,000 x *g* for 30 minutes at room temperature. The supernatant was mixed with 3 ml of Ni-NTA beads equilibrated with lysis buffer and incubated at room temperature for 4 hours. The slurry was loaded into an empty column and washed with denaturing wash buffer (8 M urea, 100 mM Na_2HPO_4, pH 8.0) until the OD_{280} of the eluate was less than 0.01. Bound His-tagged CsoS2 protein was eluted from the column with denaturing wash buffer containing 250 mM imidazole. To induce protein refolding, the fractions containing CsoS2 protein were pooled and dialyzed against 4 liters of dialysis buffer (10 mM Tris-HCl, pH 8.0, 20 mM NaCl and the decreasing concentration of urea at 6 M, 4 M, 2 M and 0 M) for 6 hours at each concentration of urea. Any precipitates that had formed during dialysis were removed by centrifugation at maximum speed for 5 minutes in a microcentrifuge. The supernatant was transferred to a new tube and stored at -20°C until further use.

CsoS2 protein in the pTYB1 vector. E. coli ER 2566 cells carrying the pTYB1-csoS2 plasmid were grown at 37°C to an OD_{600} of 0.6 to 0.7. Protein expression was induced with 0.3 mM IPTG. After incubation for 5 hours at 30°C, one liter of *E. coli* cells

were harvested by centrifugation at 10,000 x *g* for 10 minutes and the cell pellet was resuspended in 50 ml of lysis buffer (20 mM Na_2HPO_4, pH 8.0, 500 mM NaCl, 1 mM EDTA and 1 mM PMSF/PTSF) and disrupted by passage through a French Press cells at 20,000 psi. The lysate was clarified by centrifugation at 20,000 x *g* for 30 minutes. The resulting supernatant was mixed with 20 ml of chitin beads equilibrated in lysis buffer and incubated overnight. The slurry was washed with 10 bed volumes of low salt wash buffer (20 mM Na_2HPO_4, pH 8.0, 500 mM NaCl, 1 mM EDTA and 0.1 mM PMSF/PTSF), 20 bed volumes of high salt wash buffer (20 mM Na_2HPO_4, pH 8.0, 2 M NaCl, 1 mM EDTA and 0.1 mM PMSF/PTSF), 10 bed volumes of low salt wash buffer and 3 bed volumes of cleavage buffer (20 mM Na_2HPO_4, pH 8.0, 500 mM NaCl, 1 mM EDTA, 50 mM DTT and 0.1 mM PMSF/PTSF). The fusion protein was allowed to self-cleave on the column over a 24-hour period at 4°C and eluted with 70 ml of low salt wash buffer. Ten 7-ml fractions were analyzed by SDS-PAGE. All fractions containing recombinant CsoS2 were pooled and placed into a 3,500 molecular weight cut off dialysis bag. The dialysis bag was covered with polyethyleneglycol (PEG) 8,000 powder to absorb any buffer leaving the dialysis bag, and dialysis was allowed to proceed until the final volume of the sample inside the bag had decreased to 1 ml. Residual PEG was rinsed off with deionized water prior to dialyzing the concentrated CsoS2 protein against 4 liters of dialysis buffer (10 mM Tris-HCl, pH 8.0, 20 mM NaCl and 10 mM β-ME) overnight.

Generation of polyclonal antiserum against recombinant CsoS2 protein

Approximately 1 mg of purified recombinant CsoS2 protein (pTYB1 system) was loaded onto a 7.5% SDS-polyacrylamide gel. After electrophoresis, the gel was stained with 0.05% aqueous Coomassie Brilliant Blue R250 for 20 minutes and destained with deionized water four times to remove the unbound dye. The target protein band was excised from the gel and sent to Cocalico Biologicals, Inc. for polyclonal antibody generation in rabbits. Different dilutions of the final serum sample were assessed for their ability to recognize the antigen and to optimize the conditions for its future use in immunoblotting and immunoprecipitation analyses.

Protein determination

The concentration of protein was estimated by the BCA (Pierce) or Bradford (BioRad) method following the manufacturers' instructions. Protein concentrations in bacterial cells were quantified by a modified Lowry assay (Pierce). Bovine serum albumin (BSA) was used as the standard.

SDS-PAGE and immunoblot analyses

Protein samples were separated on pre-cast 4-20% or 10-20% sodium dodecyl sulfate polyacryamide gradient gels (Criterion, BioRad) to analyze their composition. Polypeptide bands were visualized by staining with Gel Code Blue (Pierce). For immunoblotting, the proteins were transferred onto a 0.45 μm pore size nitrocellulose membrane in a Mini Trans-Blot electrophoretic transfer cell (BioRad). The blot was blocked with immunoblot blocking buffer (5% nonfat milk in PBS with 0.1% Triton X-

100) for 30 minutes. The appropriate primary antibody generated in rabbits was incubated with the blot for 1 hour to probe the presence of the target antigen. After rinsing with PBS buffer, immunoblot blocking buffer and PBS buffer for 15 minutes each, the blot was incubated with goat anti-rabbit IgG antibody conjugated with alkaline phosphatase (AP) or horseradish peroxidase (HRP) for 1 hour. The blot was developed with one step NBT-BCIP solution for chromatic detection of AP activity or SuperSignal® West Pico Chemiluminescent Substrate for HRP detection (Pierce). Images of stained gels and immunoblots were captured and documented using a VersaDoc imaging system model 4000MP (BioRad). Densitometry analysis was performed using the Quantity One program (BioRad).

Isolation of carboxysomes from H. neapolitanus

Wild type and SPAS2 mutant *H. neapolitanus* cells were grown in a bioreactor at 0.08 h^{-1} dilution rate at pH 6.4 and 30°C in an ambient CO_2 environment. The *csoS3::Km* mutant cells were grown in the bioreactor in 5% CO_2-enriched air and switched to ambient CO_2 for 18 to 24 hours prior to cell harvest. Eight liters of cells were concentrated to 500 ml with a Pellicon™-2 cassette filter (Millipore) and pelleted by centrifugation at 12,000 x *g* for 10 minutes. The cell pellet was resuspended in 20 ml of TEMB buffer supplemented with 1 mg/ml lysozyme and 5 mg/ml $MgSO_4$. The suspension was mixed with an equal volume of B-PER II detergent (Pierce) and sonicated for 10 to 15 seconds until the suspension became less viscous. The resulting cell lysate was mixed with 200 Kunitz units of DNase I (Sigma) and gently agitated for 30 minutes at room temperature. Cell debris was removed by centrifugation at 10,000 x *g*

for 10 minutes at 4°C. The resulting supernatant was spun at 48,000 x *g* for 30 minutes at 4°C to obtain a carboxysome-enriched pellet. This pellet was resuspended in 3 ml of TEMB buffer and loaded onto a linear 10-60% sucrose gradient prepared in TEMB buffer. Centrifugation at 100,000 x *g* for 30 min at 4°C in a Beckman JS-24.38 rotor yielded a narrow white band of carboxysomes in the middle of the gradient. This band was collected and the carboxysomes were pelleted by centrifugation at 126,000 x *g* for 2 hours at 4°C. The carboxysome pellet was resuspended in 0.5 ml of TEMB buffer and stored at 4°C.

Transmission electron microscopy

A suspension of purified carboxysomes was placed onto formvor/carbon coated copper grids and allowed to stand for 3 min. After drying the samples in air for 10 seconds, the samples were negatively stained with 1% (w/v) ammonium molybdate in 10 mM Tris-HCl, pH 8.0 for 40 seconds. The grids were air-dried for 10 minutes before being observed under a Zeiss EM-109 transmission electron microscope. Images of carboxysomes were taken by exposing Kodak EM film 4489 for 2 seconds.

Carboxysome breakage and preparation of free RuBisCO enzyme [57, 114]

Purified carboxysomes were spun at 21,000 x *g* for 30 minutes at 4°C, and the resulting pellet was frozen at -20°C for at least 30 minutes. Freshly prepared TEMB buffer at room temperature was used to resuspend the frozen pellet rapidly. Most of the carboxysomes were broken after resuspension, but retained their polyhedral shell structure when viewed by transmission electron microscopy. To obtain free RuBisCO

enzyme, broken carboxysomes were spun at 21,000 x *g* for 30 minutes at 4°C to pellet the broken shells and residual shell-associated RuBisCO. The majority of holoenzyme molecules remained in the supernatant.

Radiometric RuBisCO assay

Protein samples were desalted on a Micro Bio-Spin 6 column (Bio-Rad) and stored in 50 mM Bicine-NaOH, pH 8.0 prior to use. Radioactive $NaH[^{14}C]O_3$ solution was prepared by adding 10 µl of 2 µCi/µl $NaH[^{14}C]O_3$ stock solution (GE Healthcare) to 1 ml of 200 mM freshly prepared $NaHCO_3$ solution to a final specific radioactivity of 0.1 µCi/µmol. The desalted RuBisCO samples were activated in activation buffer (50 mM Bicine-NaOH, pH 8.0, 10 mM $MgCl_2$ and 10 mM $NaH[^{14}C]O_3$ (0.1 µCi/µmol)) for 10 minutes prior to the start of the assay. Each assay vial contained 0.5 ml of 50 mM Bicine-NaOH, pH 8.0, 20 mM $MgCl_2$, 0.5 mM ribulose 1,5-bisphosphate (RuBP, Fluka) and the appropriate amount of $NaHCO_3$ (1.25, 2.5, 5, 10, 20, 40 and 80 mM). A 10 µl aliquot of activated RuBisCO was added to initiate the reaction. The reaction was stopped by pipetting 100 µl aliquots of the assay mixture into 300 µl of glacial acetic acid every minute for a total of 3 minutes. The vials which contained the acidified samples were heated for 4 minutes to remove excess $[^{14}C]O_2$ and allowed to cool to room temperature before adding 4 ml of ScintiVerse II cocktail (Fisher) and mixing well. The radioactivity in acid-stable reaction products was quantified by scintillation counting (LS 6000SC, Beckman). The same assay mixture without added protein was treated identically to serve as a blank. Quench and counting efficiency were determined with the help of a standard curve that was constructed from various amounts of ^{14}C labeled n-hexadecane added to

0.3 ml glacial acetic acid followed by addition of 4 ml of ScintiVerse II cocktail. Similarly, K_M(RuBP) and V_{max} of RuBP were measured by keeping the concentration of $NaHCO_3$ constant at 60 mM and varying the concentration of RuBP (12.5, 25, 50, 100, 200, 400 and 800 μM). All RuBisCO assays were performed in triplicate and repeated at least three times with independent carboxysome preparations.

Carbon fixation by intact H. neapolitanus cells

A 25 ml aliquot of bioreactor grown *H. neapolitanus* culture was spun at 12,000 x *g* for 10 minutes at 4°C. The cell pellet was gently resuspended in 1 ml of freshly prepared *H. neapolitanus* growth medium. Each assay vial contained 460 μl of *H. neapolitanus* medium and 20 μl of $NaH[^{14}C]O_3$ solution to give a final concentration of 0.31, 0.62, 1.25, 2.5, 5.0 or 10 mM with a specific radioactivity of 1.0 μCi/μmol. A 20 μl aliquot of the concentrated bacteria was added to the vial to start the assay. To terminate the reaction, 100 μl of the assay solution was transferred to a vial containing 300 μl of glacial acetic acid every minute for 3 minutes. The acidified samples were heated for 5 minutes to expel heat-labile inorganic carbon species. After cooling, the samples were mixed with 4 ml of ScintiVerse II cocktail (Fisher) prior to measuring the amount of fixed inorganic carbon by scintillation counting.

Enzymatic synthesis of 1-[^{3}H]RuBP [65]

1-[^{3}H]RuBP was synthesized from 2-[^{3}H]glucose using a series of enzymatic reactions. The reaction was allowed to proceed at 25°C for 90 minutes in a 15 ml tube that contained 10 ml of 33 mM 1,3-bis-propane-HCl, pH 7.5, 7 mM $MgCl_2$, 1.5 mM

NADP, 0.5 mM ATP, 1 mM phosphocreatine, 10 mM DTT, 0.6 mM glucose, 200 international units of rabbit muscle phosphocreatine kinase, 80 international units of yeast glucose-6-phosphate dehydrogenase, 40 international units of yeast 6-phosphogluconate dehydrogenase and 20 international units of recombinant phosphoribulokinase. Yeast hexokinase (80 international units) was added to initiate the enzymatic reactions. After 90 minutes of incubation, the mixture was acidified with 3 M HCl solution to a pH of 2.8 to denature and precipitate the proteins, and 0.6 grams of activated charcoal was added to adsorb the proteins. Next, the mixture was incubated on ice for 15 minutes and filtered through a 0.45 μm pore size filter to remove the activated carbon. The filtrate was loaded onto a MonoQ anion exchange column equilibrated with 3 mM HCl to purify the radioactive RuBP. Reaction product 1-[^{3}H]RuBP was eluted from the column with 100 ml of a 0 to 1 M NaCl gradient in 3 mM HCl. A single radioactive peak eluted at approximately 0.25 M NaCl. To verify that the peak contained RuBP, all peak fractions were assessed with CO_2 fixation assays. The purified RuBP was frozen at -80°C for future use.

Radiometric simultaneous oxygenase assays

Freshly prepared carboxysomes and free RuBisCO enzyme were desalted in storage buffer (50 mM Bicine-NaOH, pH 8.0) as described before. The storage and assay buffers were acidified with 3 M HCl to a pH of 2.0 and boiled for 20 minutes. After cooling, the buffers were adjusted to pH 8.0 with freshly prepared 1 M NaOH solution and sealed with a rubber cap. Prior to the assay, the assay buffer was bubbled with pure N_2, pure O_2 or an N_2/O_2 gas mixture (6.25:1, 3.33:1, 2:1, 1:1 or 0.4:1 achieved by an

N_2/O_2 gas mixer) for 1 hour. Each assay vial contained 50 mM Bicine-NaOH, pH 8.0, 20 mM $MgCl_2$, 1.4 mM $NaHCO_3$, 0.5 mM radiolabeled RuBP (specific radioactivity 1.07 μCi/μmol) and various concentrations of oxygen from 50 to 1050 μM. The reaction was initiated by adding 20 μl of activated carboxysomes (60 to 70 μg) or free RuBisCO enzyme (20 to 30 μg). After 3 minutes of incubation at room temperature, the reaction was terminated by adding 50 μl of 20 mM freshly prepared $NaBH_4$ to a final concentration of 2 mM. After another 15 minutes of incubation at room temperature, 50 μl of 50 mM glucose solution was added to eliminate unconsumed $NaBH_4$. The CO_2 and O_2 fixation products 3-phosphoglycerate and 2-phosphoglycolate, respectively, were dephosphorylated with 60 units of calf intestinal alkaline phosphatase (CIAP) for 1 hour at room temperature to yield glycerate and glycolate. These two species were separated on an Aminex® HPX-87H organic acid analytical HPLC column (strong anion exchange column) with 0.0075 M H_2SO_4 solution. Before loading the readiolabeled glycerate and glycolate onto the HPLC column, a 100 μl of 2 mM nonradiolabeled glycerate or glycolate was loaded onto the HPLC column to measure their retention times. The two peaks corresponding to radiolabeled glycerate and glycolate, respectively, were collected and their radioactivities were counted to determine the ratio of carboxylase to oxygenase activity.

Determination of kinetic parameters

The initial velocity of inorganic carbon fixation was obtained by subjecting the measured amount of fixed carbon at three time points to linear regression using the Excel

2000 program. Initial velocities at various concentrations of $NaHCO_3$ were fitted to the Michaelis-Menten equation using GraphPad Prism 4 to obtain V_{max} and K_M values.

For oxygenase measurement, the ratio of the radioactivity of glycerate to glycolate was plotted over various $[CO_2]/[O_2]$ values using the GraphPad Prism 4 program. The plot was fitted by the method of linear least squares. The regression line was forced to go through the point (X=0, Y=0). The resulting slope represented the specificity factor. To determine the short time carboxylase to oxygenase activity ratio, the ratio of the radioactivity of glycerate to glycolate was calculated and plotted over various $[CO_2]/[O_2]$ values using the GraphPad Prism 4 program.

Identification of carboxysome proteins by mass spectrometry

Carboxysome protein samples were separated by SDS-PAGE. Gels were stained with 0.05% aqueous Coomassie Brilliant Blue R250 solution for 1 hour, and destained with deionized water several times until the target band was visible. The protein band of interests was excised and destained twice for 45 minutes each with 200 µl of 100 mM NH_4HCO_3/50% acetonitrile (ACN) at 37°C to remove trace amounts of remaining dye. The gel slice was dehydrated in 100 µl of 100% ACN for 5 minutes at room temperature and the ACN was removed by vacuum centrifugation in a SpeedVac for 10-15 minutes. The dry gel slice was rehydrated in 10-20 µl of a 20 µg/ml trypsin solution (Promega) in 40 mM NH_4HCO_3/10% ACN at room temperature for 1 hour. A volume of digestion buffer (40 mM NH_4HCO_3/10% ACN) sufficient to cover the gel slice was added and the sample was incubated at 37°C overnight. The gel slice was removed out and incubated in deionized water for 10 minutes with frequent vortex mixing. The liquid was transferred to

a new tube and the gel slice was extracted twice with 50 µl of 50% ACN/5% trifluoroacetic acid (TFA) at room temperature for 60 minutes each time. All extracts were pooled and dried in a SpeedVac for 2-4 hours at room temperature. The resulting dry white powder was reconstituted with 10 µl of 0.1% TFA and the peptide mixture was purified and concentrated with ZipTip C18 pipette tips (Millipore) prior to mass spectrometric analysis according to the manufacturer's instruction. The Mascot program was used to analyze the mass spectrometry data. All mass lists were searched against the NCBI nonredundant protein database.

CHAPTER IV

RESULTS

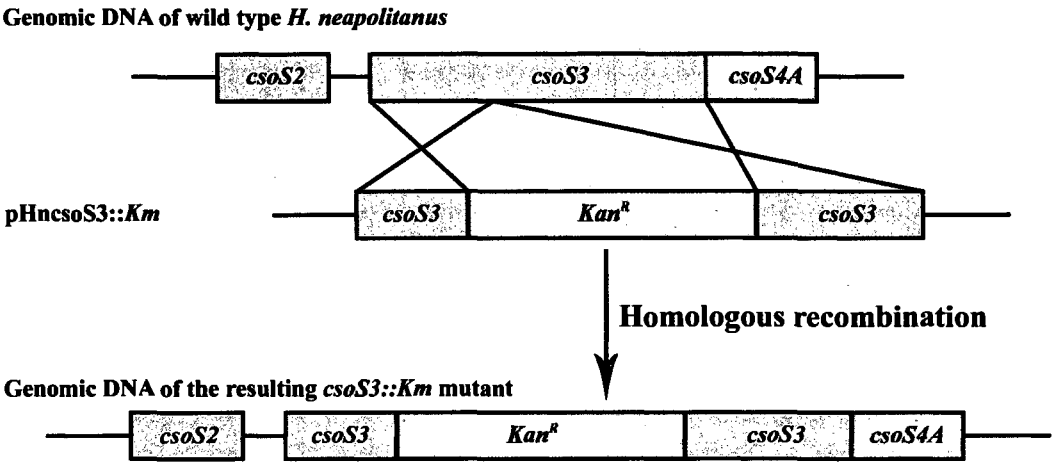
Characterization of the *csoS3::Km* Mutant

Carboxysomes of *H. neapolitanus* were purified to homogeneity nearly thirty years ago [25, 26]. Purified carboxysomes are composed of ten polypeptides: RuBisCO large and small subunit, CsoS1A, CsoS1B, CsoS1C, CsoS2A, CsoS2B, CsoSCA (formerly CsoS3), CsoS4A, and CsoS4B [13, 14, 25, 26, 28, 45, 129]. By a reverse genetics approach, the CsoSCA protein was identified as a novel type carboxysomal shell associated carbonic anhydrase [114]. Recent investigation has revealed that ruptured carboxysomes show similar carbonic anhydrase activity as does recombinant CsoSCA protein, however, the k_{cat} for CO_2 hydration in intact carboxysomes is at least three times lower than that observed with ruptured carboxysomes [57]. These findings indicated that the carboxysomal shell acts as a diffusion barrier for CO_2 molecules [57]. To better understand the role of CsoSCA in the carboxysome, a *csoS3* gene insertion mutant was generated.

csoS3::Km mutant generation

The wild type *H. neapolitanus csoS3* gene in the pHncsoS3 plasmid [14] was interrupted by insertion of a kanamycin resistance cassette between the Cla I sites in the *csoS3* coding sequence to generate a new plasmid, pHncsoS3*::Km*. The pHncsoS3*::Km* plasmid was introduced into mid-exponential phase *H. neapolitanus* cells via electroporation, and the kanamycin resistance cassette was integrated into the *csoS3* gene in *H. neapolitanus* genomic DNA by homologous recombination *in vivo* (Figure 4) (done

A



B

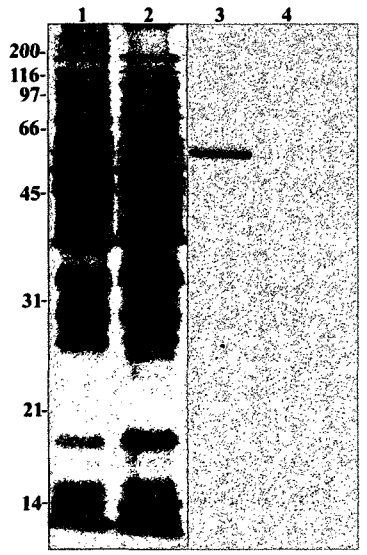


Figure 4. Diagram of the strategy used to generate the *H. neapolitanus csoS3::Km* mutant and immunoblot analysis of wild type and mutant cell extracts.

Panel A: A kanamycin resistance gene cassette was inserted into the coding sequence of the *csoS3* gene by homologous recombination to prevent expression of the CsoSCA protein. An undergraduate student, C. Daniel Murin, performed Southern blotting to verify the insertion in the *csoS3::Km* mutant and conducted complementation studies to prove the essential role of CsoSCA protein for efficient growth of wild type *H. neapolitanus* cells in air [41].

Panel B: The polypeptide composition of cell extracts (20 μg of protein) from wild type (lanes 1) and mutant (lane 2). Lanes 3 and 4: immunoblots of the gel shown in lanes 1 and 2, probed with polyclonal anti-CsoSCA antibody. CsoSCA protein could not be identified in mutant crude cell extracts (lane 4) [41]. Protein molecular weight standards are indicated on the left.

in Dr. Jessup Shively's lab). Given that the resulting mutant needed a high carbon dioxide environment to support efficient growth, the *H. neapolitanus* transformants were selected at elevated (5%) CO_2 in the presence of 50 μg/ml kanamycin. An undergraduate student, C. Daniel Murin, prepared the genomic DNA of transformants and screened the mutants by Southern blotting and sequence analysis [82]. Immunoblotting analysis of crude extracts of the resulting mutant with anti-CsoSCA antibodies revealed that there was no detectable CsoSCA protein in the mutants but a strong signal in the wild type (Figure 4B), suggesting that the *csoS3* gene was successfully inactivated in the *csoS3::Km* mutant. To test the essential role of CsoSCA protein for the normal growth of wild type *H. neapolitanus* cells, Murin electroporated *csoS3::Km H. neapolitanus* cells with the pcsoS3ProEx plasmid [127] that contains wild type *csoS3* gene. The resulting complementation clones regained the ability to survive at ambient CO_2 [82].

Phenotype studies of the csoS3::Km mutant

A single high CO_2-requiring, kanamycin-resistant colony was used as the inoculum to generate the working culture of the *csoS3::Km* mutant. Growth curves of wild type and mutant cells were established under ambient CO_2 and CO_2-enriched (5% CO_2) conditions. In air, *csoS3::Km* cells grew much more slowly than wild type cells, but in 5% CO_2-enriched air, the *csoS3::Km* mutant had a similar growth pattern as the wild type (Figure 5).

To understand the reason for the different growth patterns in air between wild type and the *csoS3::Km* mutant, a detailed kinetic analysis of carbon assimilation was performed, and the rates of carbon fixation by bioreactor-grown wild type and mutant

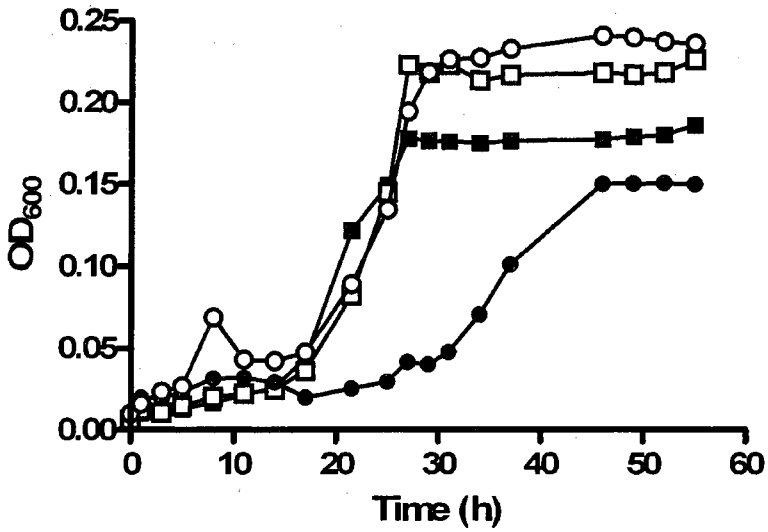


Figure 5. Growth curves of *H. neapolitanus* wild type and *csoS3::Km* mutant cultures.

Cell density was assessed by measuring optical density at 600 nm. ■, wild type in air; □, wild type in 5% CO_2-enriched air; ●, mutant in air; ○, mutant in 5% CO_2-enriched air. In air, the *csoS3::Km* mutant grew much more slowly than wild type, but in 5% CO_2-enriched air, the mutant had a similar growth pattern as the wild type [41].

cells at different dissolved inorganic carbon concentrations (DIC, includes bicarbonate, carbonate, and CO_2) were plotted. The plot was fitted to the Michaelis-Menten equation and showed that wild type and *csoS3::Km* mutant cells had comparable maximal carbon fixation rates (V_{max}) of 0.12 ± 0.01 and 0.13 ± 0.01 $\mu mol \cdot min^{-1} \cdot mg\ protein^{-1}$, respectively. The dissolved inorganic carbon species bicarbonate, carbonate and CO_2 are in equilibrium, according to the reactions $CO_2 + H_2O <=> HCO_3^- + H^+$ and $HCO_3^- <=> CO_3^{2-} + H^+$. Given that the intact cells can take up bicarbonate and CO_2 into their cytoplasm by energy-dependent inorganic carbon transport and utilize the inorganic carbon species for carbon fixation [93, 104], the kinetic parameter K_M that describes the concentration of the substrate required to achieve half-maximal catalytic rate is named here K_{DIC}. The K_{DIC} value for mutant cells (3.9 ± 1.0 mM) was more than three times higher than that of wild type cells (1.1 ± 0.2 mM) (Figure 6).

After obtaining the kinetic parameters V_{max} and K_{DIC} for DIC fixation by wild type and mutant intact cells, their carbon fixation rates could be calculated from the Michaelis-Menten equation. Under ambient CO_2 condition, the concentration of DIC is approximately 26 μM at pH 6.4; in 5% CO_2-enriched air, the concentration of DIC is elevated to 3,300 μM [39]. The ratio of wild type to mutant carbon fixation rates was 3.3 ± 1.0 in air, but only 1.5 ± 0.3 at elevated CO_2. This result was consistent with the growth patterns observed for wild type and mutant cells. In 5% CO_2-enriched air, wild type and mutant cells grew similarly well, presumably reflecting their comparable carbon fixation rates. However, in ambient CO_2, wild type cells fixed inorganic carbon approximately three times faster than mutant cells, and the mutant cells grew considerately more slowly than the wild type.

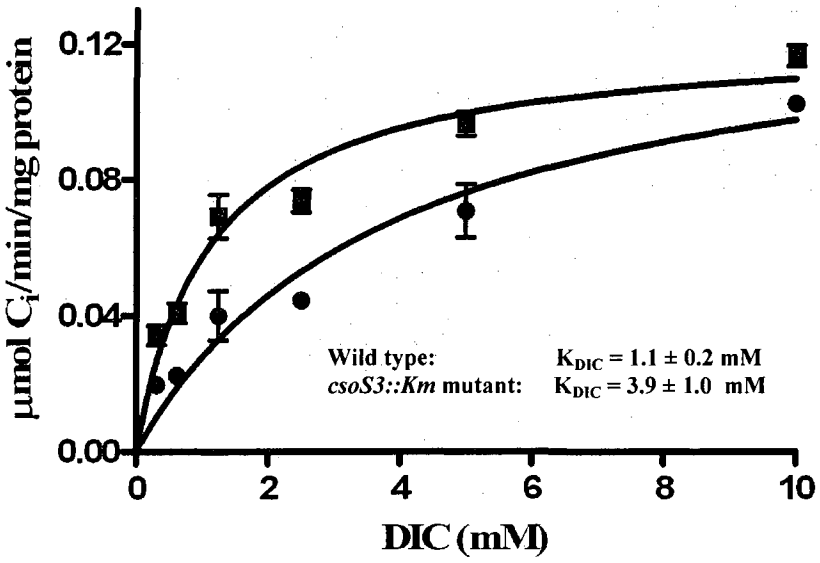
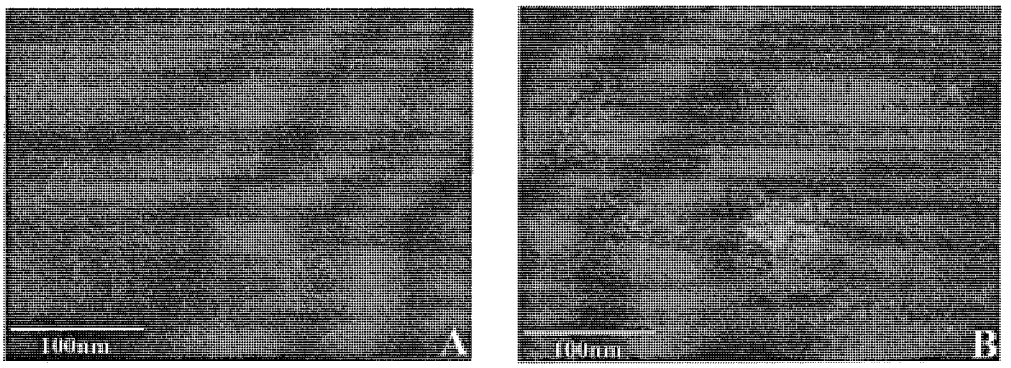


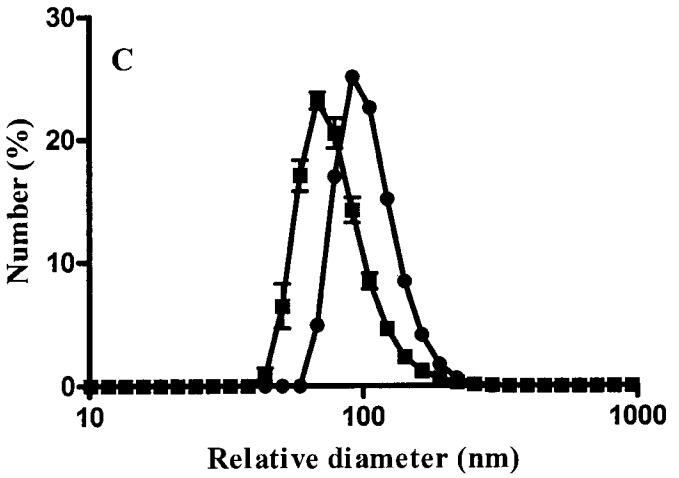
Figure 6. Assimilation of inorganic carbon by *H. neapolitanus* wild type and *csoS3::Km* mutant cells.

Bioreactor-grown cultures were kept in air (ambient CO_2) for 24 hours before being concentrated by centrifugation. The cell pellets were resuspended in fresh growth medium and assayed as described in Experimental Procedures. ■, wild type; ●, mutant. The K_{DIC} value for mutant cells was almost three times higher than that of wild type, which indicate that mutant cells assimilate dissolved inorganic carbon less efficiently than wild type cells in air [41]. The assays were performed in triplicates and repeated with three independent preparations. The error bars represent standard errors.

Preparation of csoS3::Km mutant carboxysomes

Wild type and *csoS3::Km* mutant *H. neapolitanus* cells were grown in a bioreactor under conditions described in Experimental Procedures. The mutant cells were aerated with air supplemented with 5% CO_2 until almost 6 to 7 liters of bioreactor efflux were collected in the receiving tank, then the culture was switched to ambient CO_2 for 18 to 24 hours prior to carboxysome purification; the wild type was always grown in air. The carbon downshift was done because the formation of carboxysomes is suppressed at high concentration of CO_2 [18]. Isolation of wild type and mutant carboxysomes was performed by differential and sucrose density gradient centrifugations as described in Experimental Procedures [57, 127]. The yield of carboxysomes from mutants was similar to that from wild type cells (approximately 10 mg of carboxysome protein per 8 g of wet cell pellet). The isolated mutant carboxysomes were indistinguishable morphologically from wild type carboxysomes by transmission electron microscopy (TEM) analysis (Figure 7A and B). To determine the sizes of wild type and mutant carboxysomes more accurately, dynamic light scattering analysis was performed and revealed that the mutant carboxysomes were 15 to 20% larger in diameter than wild type carboxysomes (Figure 7C). The polypeptide composition of wild type and mutant carboxysomes was similar by denaturing SDS-PAGE analysis, and CsoSCA protein was not detected in the purified mutant carboxysomes. Immunoblotting analysis using a polyclonal anti-CsoSCA antibody confirmed the absence of CsoSCA protein in purified mutant carboxysomes (Figure 8). Interestingly, the pattern of CsoS2 proteins in the wild type and mutant carboxysomes was different. The ratio of CsoS2B to CsoS2A in wild type carboxysomes





Wild type carboxysomes: **81 ± 2 nm**
***csoS3::Km* mutant carboxysomes:** **104 ± 4 nm**

Figure 7. Transmission electron microscopy and dynamic light scattering analyses of wild type and *csoS3::Km* mutant carboxysomes.

Panels A and B show transmission electron micrographs of purified wild type (A) and mutant (B) carboxysomes, negatively stained with 1% (v/v) ammonium molybdate. Both microcompartments appear as hexagonal structures of similar size [56]. Panel C: Wild type and *csoS3::Km* mutant carboxysomes were diluted in TEMB buffer to a concentration of approximately 10 μg/ml for dynamic light scattering analysis. The *csoS3::Km* mutant carboxysomes appeared to be 15 to 20% larger than wild type carboxysomes. ■, wild type carboxysomes; ●, the *csoS3::Km* carboxysomes.
The size distribution of carboxysomes was converted by the Zetasizer software (Malvern, UK) from the intensity distribution of particles detected by dynamic light scattering. Values are averages ± standard deviations for three measurements using three independent carboxysome preparations. The error bars represent standard deviations for each data point.

is close to 1; mutant carboxysomes contained more of the large form of CsoS2 protein, and the ratio was shifted closer to 2:1 (Figure 8). Therefore, the slight difference in size may indicate that the *csoS3::Km* carboxysomes have a different carboxysome assembly pattern compared with their wild type counterparts. The minor difference of the composition between wild type and mutant carboxysomes also could be responsible for the slight difference of carboxysomes in size.

The absence of carbonic anhydrase activity in the mutant carboxysomes is essential to assessing the inherent permeability of the carboxysomal shell for CO_2. Although CsoSCA protein was not found in the purified mutant carboxysomes, colorimetric stopped-flow assays were performed to measure the reaction progress of CO_2 hydration and to determine whether any additional carbonic anhydrase activity was associated with the mutant carboxysomes. Given that the hydration reaction of CO_2 is a spontaneous reaction, the assay was performed first in the absence of carbonic anhydrase to establish the background reaction rate. The purified mutant carboxysomes showed almost the same CO_2 hydration kinetics as the uncatalyzed reaction, which indicated that no carbonic anhydrase activity was present in purified mutant carboxysomes; the same amount of wild type particles showed significant enzyme activity as seen by the steeper slope of the absorbance change with time (Figure 9). These results confirmed that no additional carbonic anhydrase activity was associated with purified *csoS3::Km* carboxysomes.

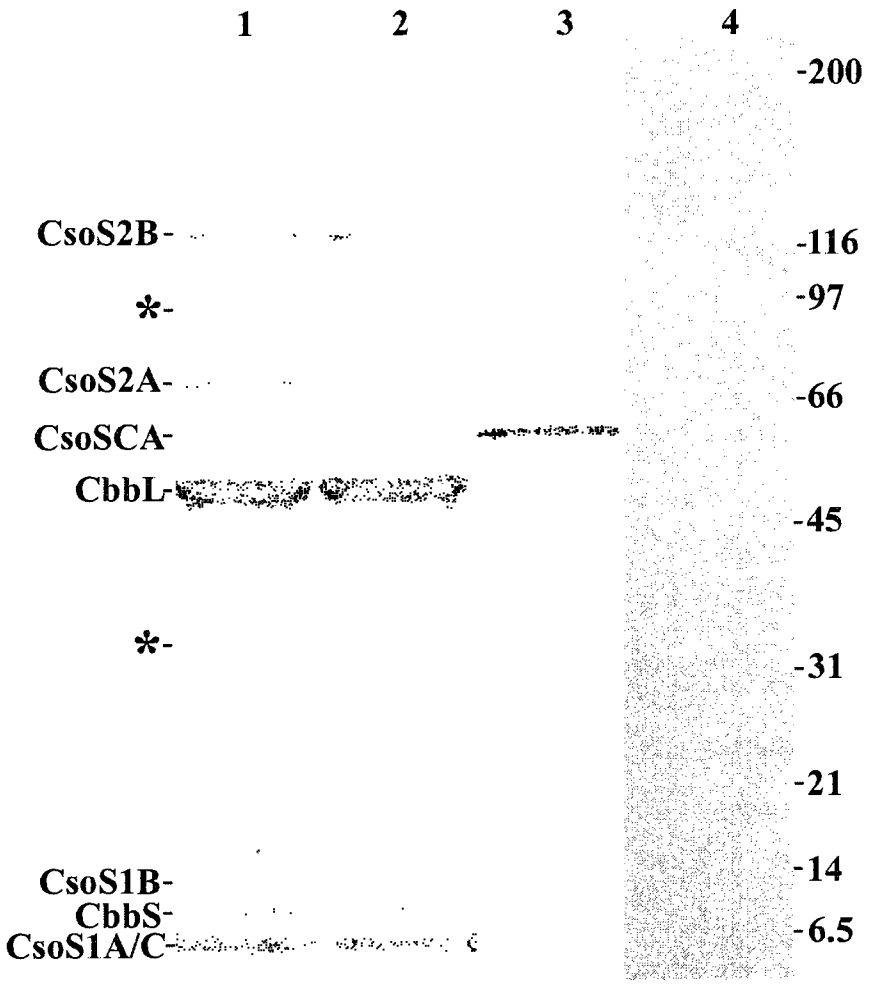


Figure 8. Composition of wild type and *csoS3::Km* mutant carboxysomes.

The polypeptide composition of purified carboxysomes (7.5 μg of protein) from wild type (lane 1) and mutant (lane 2). Lanes 1 and 2: stained polypeptides separated by SDS-PAGE; lanes 3 and 4: immunoblots probed with polyclonal anti-CsoSCA antibody. Protein molecular weight standards are indicated on the right.
CsoSCA protein could not be identified in purified carboxysomes of the mutant (lane 4). As determined by mass spectrometry, the polypeptide bands marked with asterisks are carboxysome protein aggregates that were not disrupted during electrophoresis [41, 56].

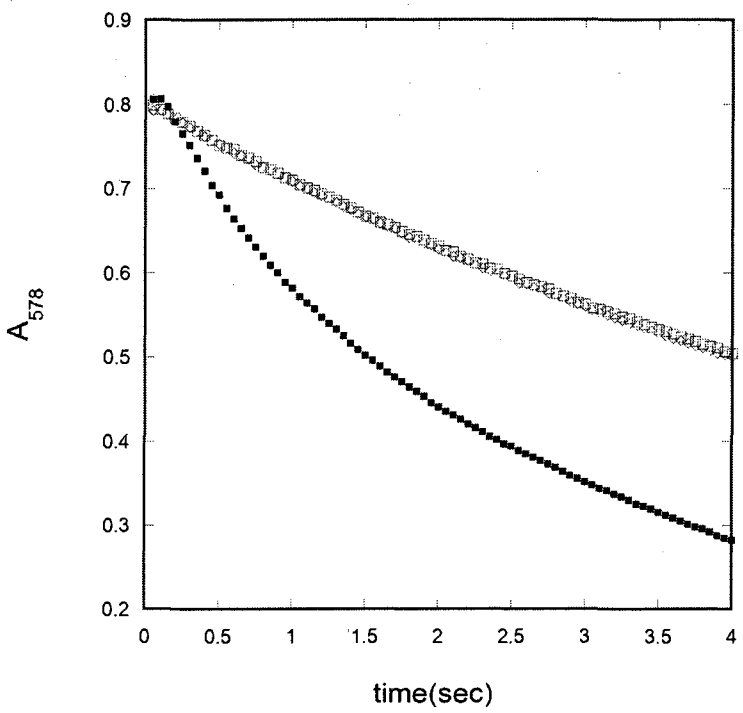


Figure 9. Carbonic anhydrase activity of purified *H. neapolitanus* wild type and *csoS3::Km* mutant carboxysomes.

The stopped flow changing indicator assays were performed as described previously [41, 57]. ○, background; ■, wild type; □, mutant. The assay measures changes in absorbance at 578 nm of *m*-cresol purple that is correlated with the appearance of protons, according to the following reaction, $CO_2 + H_2O <=> HCO_3^- + H^+$. This reaction is a spontaneous reaction, therefore, the assay first was performed in the absence of carbonic anhydrase. The mutant carboxysomes showed almost the same kinetics as the uncatalyzed reaction, while the parallel control experiment showed significant carbonic anhydrase activity in wild type carboxysomes. These results indicated that there is no additional carbonic anhydrase activity associated with *csoS3::Km* mutant carboxysomes.

CO_2 fixation kinetics in purified carboxysomes

To directly assess the effect of the carboxysomal carbonic anhydrase activity on CO_2 fixation by the encapsulated RuBisCO, the CO_2 fixation activity of purified mutant and wild type carboxysomes was determined radiometrically. Given that the substrate of RuBisCO enzyme is CO_2, not HCO_3^-, the concentration of CO_2 was derived from the concentration of HCO_3^-, according to the Henderson-Hasselbalch equation. Because the optimal pH for most RuBisCO enzymes is in the range of 7.0 to 8.5 [122, 123], the RuBisCO CO_2 fixation assay was performed at pH 8.0 as described in the literature [6]. At pH 8.0, the equilibrium concentration of CO_2 is 14 μM in a total DIC of 1 mM (CO_2 + H_2O <=> HCO_3^- + H^+; pKa = 6.12 [52]). The K_c value was defined as the concentration of CO_2 necessary to reach half-maximal velocity. This constant is different from the $K_{M(CO2)}$, which can be measured for free RuBisCO in the absence of a potential accessibility barrier for its substrate CO_2. Thus, the K_c values of intact and broken carboxysomes can be compared with the $K_{M(CO2)}$ value for free RuBisCO to estimate the permeability properties of the carboxysome shell for CO_2.

To determine the appropriate concentration of RuBP for the kinetic assay, CO_2 fixation reactions were set up with excess bicarbonate (60 mM) and various concentrations of RuBP in the range of 12.5 to 800 μM. The initial velocities of CO_2 fixation by wild type and mutant carboxysomes were plotted against the concentrations of RuBP, and the resulting plot was fitted to the Michaelis-Menten equation. The maximal velocity for mutant carboxysomes of 1.1 ± 0.1 μmol•min^{-1}•mg^{-1} protein was approximately 60% of the V_{max} of 1.6 ± 0.1 μmol•min^{-1}•mg^{-1} protein exhibited by wild type carboxysomes. The apparent K_M for RuBP was similar for wild type and mutant

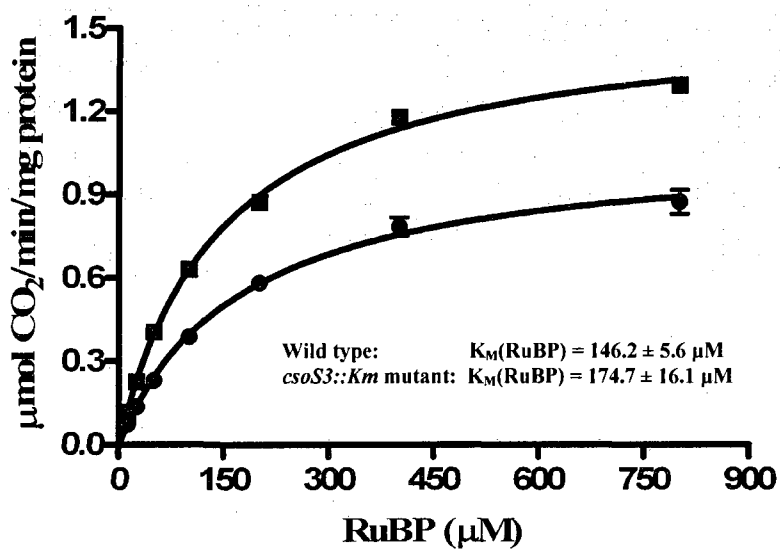


Figure 10. Determination of K_M(RuBP) for *H. neapolitanus* wild type and *csoS3::Km* mutant carboxysomes.

RuBisCO CO_2 fixation assays were performed at 60 mM bicarbonate and various concentrations of RuBP. ■, wild type; ●, mutant. Wild type and mutant carboxysomes showed comparable values for K_M(RuBP) at 146.2 ± 5.6 µM and 174.7 ± 16.1 µM, respectively. The assays were performed in triplicates and repeated with three independent preparations. The error bars represent standard errors.

carboxysomes with 146.2 ± 5.6 μM and 174.7 ± 16.1 μM, respectively (Figure 10). Based on these results, the concentration of RuBP was fixed at 0.5 mM for the subsequent K_c determinations. CO_2 fixation assays with wild type and mutant carboxysomes at 0.5 mM saturating RuBP concentration and increasing bicarbonate concentrations from 1.45 to 80.2 mM revealed a consistently higher V_{max} (1.7 ± 0.1 μmol•min^{-1}•mg^{-1} protein) for wild type carboxysomes than for mutant particles (1.3 ± 0.1 μmol•min^{-1}•mg^{-1} protein). Likewise, the K_c value (499.3 ± 47.1 μM) for mutant carboxysomes was almost threefold higher than that for their wild type counterparts (177.1 ± 16.3 μM) (Figure 12A and Table 2). These findings suggested that it is more difficult for the packaged RuBisCO molecules to gain access to CO_2 in the mutant than in wild type carboxysomes.

Effect of exogenously added carbonic anhydrase

Recombinant CsoSCA protein can accelerate the dehydration of soluble bicarbonate to CO_2. If the carboxysomal shell lacking carbonic anhydrase is freely permeable for CO_2 molecules, the carbonic anhydrase added to purified carboxysomes will equilibrate the concentration of CO_2 on both sides of the carboxysome shell quickly. The apparent K_c of mutant carboxysomes in the presence of rCsoSCA protein should therefore be lower than that in the absence of rCsoSCA protein. Purified rCsoSCA protein at an activity level that was five fold higher than that measured for wild type carboxysomes was mixed with isolated mutant carboxysomes to assess whether exogenously added carbonic anhydrase can functionally complement the missing carboxysome component. Results from CO_2 fixation assays showed very similar V_{max} values for mutant carboxysomes in the presence and absence of rCsoSCA (1.2 ± 0.1 and

1.3 ± 0.1 μmol•min^{-1}•mg^{-1} protein, respectively); the K_c values under these same conditions were slightly different with 430.6 ± 49.6 μM and 499.3 ± 47.1 μM, respectively (Figure 11). Control assays with wild type carboxysomes, likewise, revealed no significant difference in the kinetic constants regardless of the presence of rCsoSCA protein (Figure 11). These findings indicated that the carboxysomal shell-bound CsoSCA protein likely faces the interior of the carboxysomes and catalyzes the conversion of bicarbonate to CO_2, the substrate of the encapsulated RuBisCO enzyme. The more or less constant apparent K_c values of mutant carboxysomes regardless of the presence of exogenous rCsoSCA protein supported the conclusion that the carboxysome shell lacking carbonic anhydrase acts as a diffusion barrier for CO_2 molecules and the exogenously added carbonic anhydrase cannot functionally complement the missing carboxysome component of *csoS3::Km* carboxysomes.

CO_2 fixation kinetics of broken carboxysomes and freed RuBisCO

To assess how the presence and location of CsoSCA protein affect the function of carboxysomes, the carboxysomal shell was disrupted by the freeze-thaw method [57]. SDS-PAGE analysis showed that approximately 70% of the encapsulated RuBisCO is released into the supernatant, and transmission electron microscopy revealed that the remaining shell ghosts retain their hexagonal shape [57]. Breaking the carboxysomes increased their V_{max} values for CO_2 fixation for both wild type and mutant to 2.0 ± 0.1 μmol•min^{-1}•mg^{-1} protein, a value that is slightly higher than that of wild type intact carboxysomes (1.7 ± 0.1 μmol•min^{-1}•mg^{-1}) and significantly higher than that of mutant intact carboxysomes (1.3 ± 0.1 μmol•min^{-1}•mg^{-1}). By contrast, the K_c value calculated for

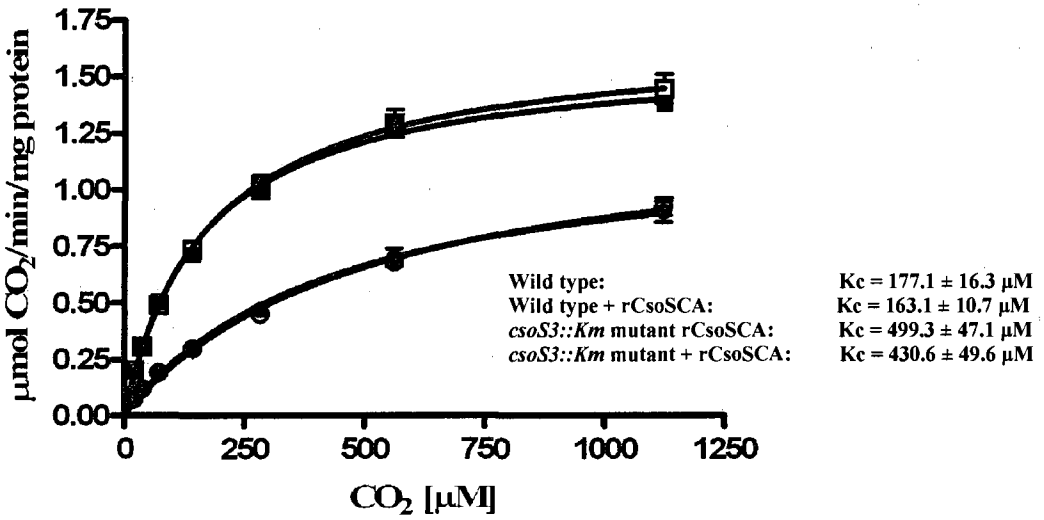


Figure 11. Effect of exogenously added rCsoSCA protein on CO_2 fixation activity of wild type and *csoS3::Km* mutant carboxysomes.

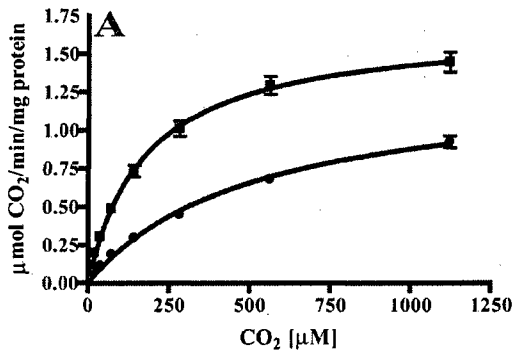
Recombinant CsoSCA protein was added to purified carboxysomes at a five fold higher carbonic anhydrase activity than that measured for wild type carboxysomes. ■, wild type carboxysomes with added rCsoSCA protein; □, wild type carboxysomes; ●, mutant carboxysomes with added rCsoSCA protein; ○, mutant carboxysomes. The exogenously added recombinant carbonic anhydrase did not significantly change the kinetic parameters of CO_2 fixation. These findings suggested that the carboxysomal shell-bound CsoSCA protein may face the interior of the carboxysomes and catalyze the conversion of bicarbonate to CO_2, the substrate of the encapsulated RuBisCO enzyme. The assays were performed in triplicates and repeated with three independent preparations. The error bars represent standard errors.

intact wild type carboxysomes is 177.1 ± 16.3 μM. These findings revealed that the carboxysomal shell-associated carbonic anhydrase facilitates the diffusion of CO_2 across the shell, thereby overcoming the diffusion barrier. Broken wild type carboxysomes, on the other hand, showed only a slightly lower K_c value than the intact particles (126.7 ± 10.8 μM and 177.1 ± 16.3 μM, respectively) (Figure 12B).

RuBisCO preparations released from wild type and mutant carboxysomes did not show significant differences in their kinetic parameters (Figure 12C and Table 2). These results supported the previous conclusion that the carboxysomal shell lacking the carboxysomal carbonic anhydrase limits the passage of CO_2 molecules to the active sites of the encapsulated RuBisCO enzyme and the shell-associated CsoSCA protein apparently increases the permeability of the carboxysomal shell for CO_2.

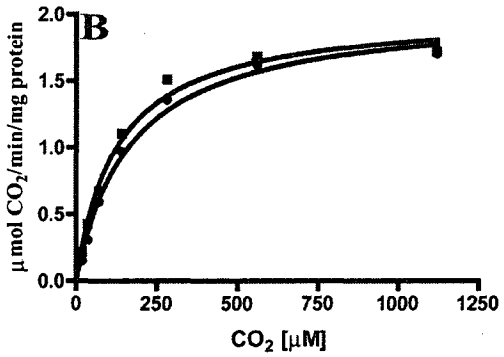
pH Effect on CO_2 fixation by carboxysomal RuBisCO

Given that the carboxysomal shell is a diffusion barrier for CO_2 molecules and the shell-bound carbonic anhydrase converts the bicarbonate to CO_2 molecules, the substrate of the encapsulated RuBisCO, the carboxysomes may utilize bicarbonate more efficiently than free RuBisCO. As the pH increases, bicarbonate becomes more abundant according to the Henderson-Hasselbalch equation, and the CO_2 fixation activity of carboxysomal RuBisCO should be higher than that of free RuBisCO. To assess the effect of pH, CO_2 fixation activities of encapsulated and free RuBisCO were measured over a broad pH range from 5.5 to 9.5 at 60 mM and 0.5 mM saturating bicarbonate and RuBP concentrations, respectively. The optimal pH for most RuBisCO enzymes is in the range



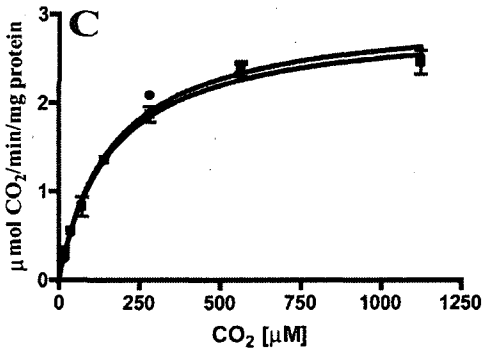
Wild type: V_{max} = 1.7 ± 0.1 µmol·min^{-1}·mg protein^{-1}
K_c = 177.1 ± 16.3 µM

csoS3::Km: V_{max} = 1.3 ± 0.1 µmol·min^{-1}·mg protein^{-1}
K_c = 499.3 ± 47.1 µM



Wild type: V_{max} = 2.0 ± 0.1 µmol·min^{-1}·mg protein^{-1}
K_c = 126.7 ± 10.8 µM

csoS3::Km: V_{max} = 2.0 ± 0.1 µmol·min^{-1}·mg protein^{-1}
K_c = 167.3 ± 13.2 µM



Wild type: V_{max} = 2.9 ± 0.1 µmol·min^{-1}·mg protein^{-1}
$K_{M(CO2)}$ = 163.7 ± 17.1 µM

csoS3::Km: V_{max} = 3.0 ± 0.1 µmol·min^{-1}·mg protein^{-1}
$K_{M(CO2)}$ = 165.8 ± 14.2 µM

Figure 12. CO_2 fixation by *H. neapolitanus* wild type and *csoS3::Km* mutant carboxysomes [41].

The figure shows carbon fixation by purified intact carboxysomes (*A*), broken carboxysomes (*B*), and freed RuBisCO (C). ■, wild type; ●, mutant. The V_{max} value of intact wild type carboxysomes is slightly higher than that of intact mutant carboxysomes, while the K_c value of intact mutant carboxysomes is nearly three times higher than that of intact wild type carboxysomes. Broken wild type and mutant carboxysomes and free RuBisCO enzyme show comparable kinetic parameters. The assays were performed in triplicates and repeated with three independent preparations. The error bars represent standard errors.

Table 2 Kinetic constants for CO_2 fixation by RuBisCO

		K_C [μM CO_2]	V_{max} [μmol · min^{-1} · mg^{-1}]
Intact Carboxysomes	Wild Type	177.1 ± 16.3	1.7 ± 0.1
	+ rCsoSCA	163.1 ± 10.7	1.6 ± 0.1
	Mutant	499.3 ± 47.1	1.3 ± 0.1
	+ rCsoSCA	430.6 ± 49.6	1.2 ± 0.1
Broken Carboxysomes	Wild Type	126.7 ± 10.8	2.0 ± 0.1
	Mutant	167.3 ± 13.2	2.0 ± 0.1
Freed RuBisCO	Wild Type	163.7 ± 17.1*	2.9 ± 0.1
	Mutant	165.8 ± 14.2*	3.0 ± 0.1

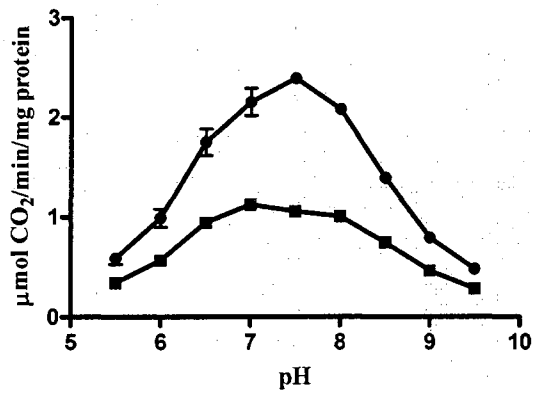
Recombinant CsoSCA protein was added to purified wild type and *csoS3::Km* mutant carboxysomes to assess the effect of the exogenously added carbonic anhydrase on CO_2 fixation activity of carboxysomes lacking carboxysomal carbonic anhydrase. All values were derived from the data shown in Figures 11 and 12. Values are averages ± standard errors for three measurements using three independent carboxysome preparations. The values marked with asterisks represent $K_{M(CO2)}$.

of 7.0 to 8.5, depending on the organism [122, 123]. The optimal pH values for maximal carboxylation activity were determined to be 7.0 for the carboxysomal enzyme and 7.5 for free RuBisCO (Figure 13A and B). The pH profiles of RuBisCO CO_2 fixation activity plotted in Figure 13B did not display significant differences in shape for carboxysomal and free RuBisCO and closely match those of the carboxysomal and recombinant carbonic anhydrase activity [57]. The similar pH profiles between two RuBisCO forms did not support the speculation that carboxysomes may utilize bicarbonate more efficiently than free RuBisCO. These results indicated that the internal pH of carboxysomes is likely very similar to, if not the same, as that of the surrounding environment. The bicarbonate dehydration reaction catalyzed by carbonic anhydrase is a fast reaction (k_{cat} is $(4.6 \pm 1.2) \times 10^4$) [57]. The same or a very similar pH on both sides of the carboxysome shell means that the CO_2 concentration on both sides of the shell is the same as well, and no significant difference of CO_2 fixation activity would be expected at different pH between carboxysomal and free RuBisCO.

Oxygenase Activity of the Carboxysomal RuBisCO

Kinetic characterization of the *csoS3::Km* mutant carboxysomes revealed that the carboxysomal shell lacking carbonic anhydrase is a diffusion barrier for CO_2 molecules. By converting external HCO_3^- to internal CO_2, the carboxysomal carbonic anhydrase apparently facilitates the diffusion of CO_2 molecules across the shell and enhances the catalytic efficiency of the encapsulated RuBisCO enzyme. Given that CO_2 and O_2 molecules have similar physical properties, such as shape and electrical potential distribution [67], the carboxysomal shell could also act as a diffusion barrier for O_2

A



B

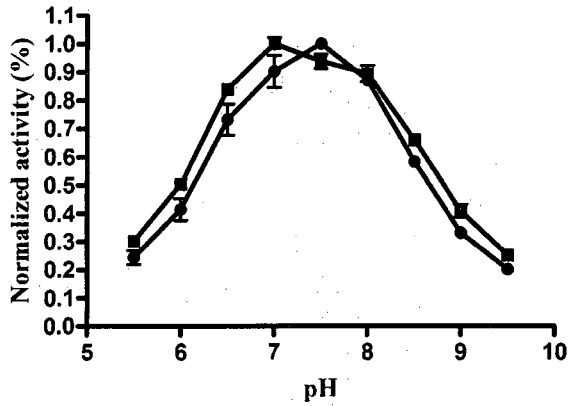


Figure 13. pH dependence of the CO_2 fixation activity of carboxysomal and free RuBisCO.

Panel A: Absolute specific CO_2 fixation activity. RuBisCO enzyme accounts for approximately 60% of total carboxysomal protein by weight. The absolute specific CO_2 fixation activity of carboxysomes (per total carboxysomal protein) is therefore much lower than that of free RuBisCO.

Panel B: Normalized CO_2 fixation activity. The absolute specific CO_2 fixation activities were normalized and expressed as a percentage of the maximal activity at pH 7.0 and 7.5, respectively. ■, carboxysomes; ●, free RuBisCO.

The similar pH profile between carboxysomal and free RuBisCO indicated that the internal pH of carboxysomes is likely very similar to, if not the same, as that of the surrounding environment. The assays were performed in triplicates and repeated with three independent preparations. The error bars represent standard errors.

molecules. The permeability of the carboxysomal shell for O_2 was therefore characterized in this study. If the shell can discriminate between these two molecular species, the ratio of CO_2 to O_2 inside the carboxysome should be higher than in the surrounding environment. Consequently, the carboxysomal RuBisCO enzyme should favor the carboxylation over the oxygenation reaction. The specificity factor (SF), also referred to as Ω or Γ in the literature, represents simply the ratio of carboxylase and oxygenase activities [72]. Its value can be determined by an assay that simultaneously assesses the conversion of 1-[^{3}H]RuBP to [^{3}H]PGA and [^{3}H]PG by reaction with CO_2 and O_2, respectively (Figure 14A). The ratio of radioactivity in [^{3}H]PGA and [^{3}H]PG reflects the molar ratio of these two species that are generated by the two competing reactions of RuBisCO (Figure 14A) according to the following relationship.

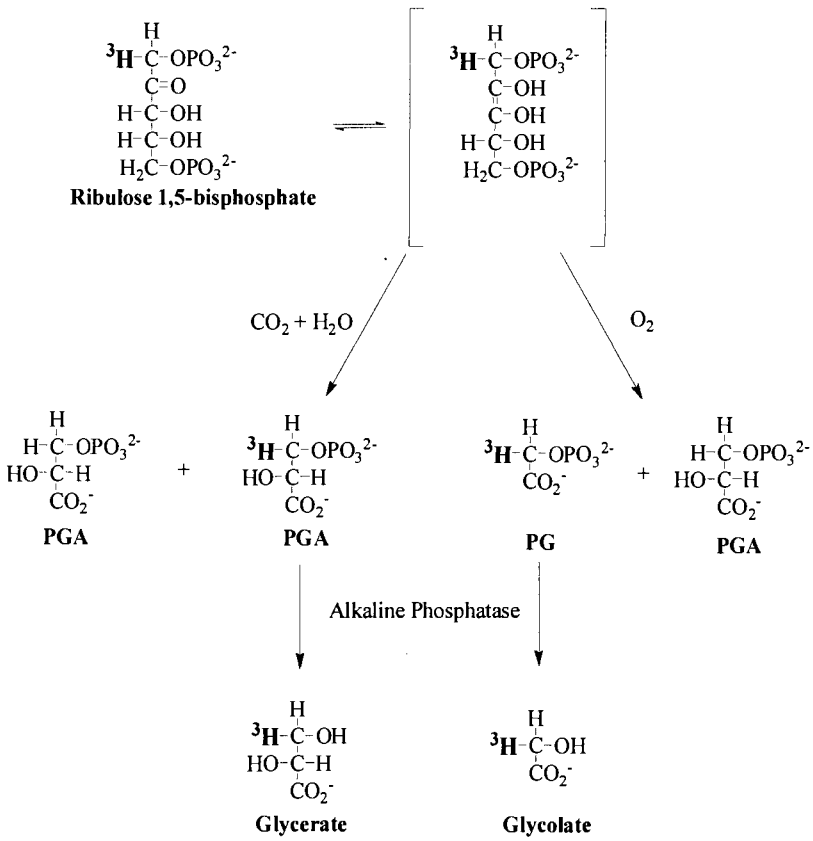
$v_c/v_o = SF \times [CO_2]/[O_2]$ [3] (v_c: carboxylation rate; v_o: oxygenation rate)

The 1-[^{3}H]RuBP is commercially unavailable and was therefore enzymatically synthesized from 2-[^{3}H]glucose through a four-step reaction developed by Kane *et al.* [65], as described in Experimental Procedures. Active recombinant phosphoribulokinase is the key component for the enzymatic synthesis of 1-[^{3}H]RuBP.

Expression of phosphoribulokinase (PRK) in E. coli and synthesis of 1-[^{3}H]RuBP

The wild type *prk* gene from *Rhodobacter sphaeroides* encoding phosphoribulokinase expressed in *E. coli* as described [32] (the clone was a gift from Dr. Jessup Shively's lab). An agarose-Reactive Green 19 column was used to purify the recombinant enzyme. The molecular weight of recombinant phosphoribulokinase

A:


Ribulose 1,5-bisphosphate
CO2 + H2O
O2
PGA
PGA
PG
PGA
Alkaline Phosphatase
Glycerate
Glycolate

B:

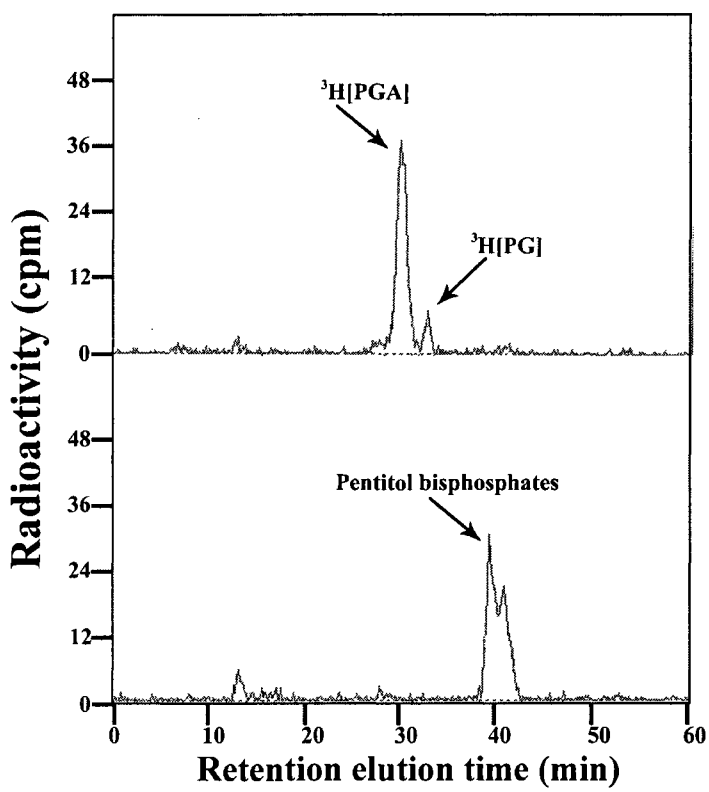

3H[PGA]
3H[PG]
Pentitol bisphosphates
Radioactivity (cpm)
Retention elution time (min)
0
12
24
36
48
0
10
20
30
40
50
60

Figure 14. The isotope-based simultaneous assay of RuBisCO carboxylase and oxygenase activity.

Panel A: Principle of the isotope-based simultaneous assay to measure the carboxylase and oxygenase activity of RuBisCO enzyme. The 1-[^{3}H]RuBP reacts with one CO_2 molecule to produce one ^{3}H-labeled PGA and one unlabeled PGA, or reacts with one O_2 molecule to yield one ^{3}H-labeled PG and one unlabeled PGA. Radiolabeled PGA and PG are separated and the radioactivity in these two species is quantified to calculate the carboxylase to oxygenase activity ratio.

Panel B: One example separation of [^{3}H]PGA and [^{3}H]PG on a MonoQ anion exchange column. The upper plot shows that radiolabeled PGA and PG are resolved as two neighboring peaks (elution times of PGA and PG: 28.5 and 31.5 minutes, respectively). For this graph, the assay time was 60 minutes and $[CO_2]/[O_2]$ was 0.27. The lower plot shows that pentitol bisphosphate derived from any remaining RuBP after reduction with NaBH4 has a longer elution time than PGA and PG. RuBP molecules are reduced to ribitol bisphosphate and arabinitol bisphosphate, yielding the two peaks visible in the lower plot. After a 60-minute incubation, there were only trace amounts of RuBP left (upper plot), indicating that most of the RuBP was converted to PGA and PG.

observed by denaturing SDS-PAGE analysis was approximately 33 kDa, a value that is very similar to the value reported in the literature [32]. The specific activity of PRK was determined by coupling the radiometric CO_2 fixation assay with the conversion of ribulose 5-monophosphate (RuMP) to RuBP as described in Experimental Procedures [32]. The specific activity of purified PRK was approximately 0.6 μmol RuBP·min^{-1}·mg protein^{-1}, which is comparable to values reported in the literature [32].

The enzymatic synthesis of 1-[^{3}H]RuBP was performed at pH 6.4 and 30 °C. The resulting final reaction mixture was loaded onto a strong anion exchange column (MonoQ column from GE Healthcare) and the 1-[^{3}H]RuBP eluted with a 0 to 1 M NaCl gradient in 3 mM HCl. The fractions were screened for the presence of RuBP with the CO_2 fixation assay. Most of the 1-[^{3}H]RuBP was eluted at 60 to 100 mM NaCl (Figure 15) and the specific radioactivity of the synthesized 1-[^{3}H]RuBP was approximately 8,000 cpm/μl in the pooled fraction.

RuBisCO specificity factor

Two isotope-based simultaneous assays were adopted to measure the RuBisCO specificity factor in this study. In 1993, Harpel *et al.* directly separated PGA and PG by high resolution chromatography on a strong anion-exchanger (MonoQ column from GE Healthcare) to calculate the ratio of PGA to PG [52]. Similarly, Kane *et al.* separated glycerate and glycolate derived from dephosphorylation of PGA and PG, respectively, on a HPLC carbohydrate analysis column [65]. The direct separation of PGA and PG is more convenient because no dephosphorylation step is necessary, while the HPLC

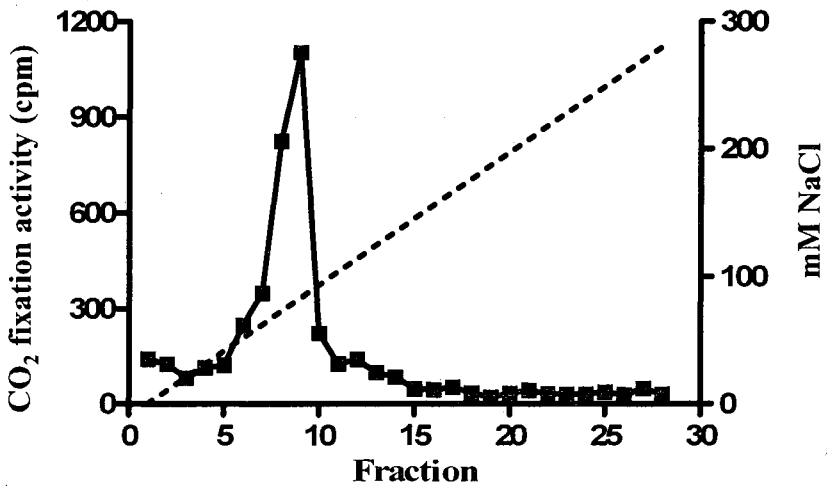


Figure 15. Purification of enzymatically synthesized 1-^{3}H[RuBP].

The 1-^{3}H[RuBP] was synthesized from 2-^{3}H[glucose] by a four-step enzymatic reaction as described in Experimental Procedures and purified by strong anion exchange chromatography. The abundance of RuBP in the first 28 fractions eluted with a 0 to 1 M NaCl gradient in 3 mM HCl was assessed by radiometric CO_2 fixation assays. -■-: CO_2 fixation activity (cpm); ---: sodium chloride gradient. Most of the radiolabeled RuBP was eluted between 60 and 100 mM NaCl.

column analysis is more sensitive when the dephosphorylated glycerate and glycolate have low levels of radioactivities.

Wild type carboxysomes and free RuBisCO enzyme were assayed at saturating 0.5 mM 1-[^{3}H]RuBP (total radioactivity approximately 150,000 cpm) and atmospheric O_2 tension (255 μM dissolved O_2, calculated from Henry's law) at room temperature. The concentrations of CO_2 were varied with exogenously added bicarbonate and calculated by the Henderson-Hasselbalch equation at pH 8.0. After the 60-minute incubation, unconsumed RuBP was reduced to the pentitol bisphosphate with $NaBH_4$ prior to separation of the radiolabeled PGA and PG. Almost all of the RuBP was converted to PGA and PG, as seen by the lack of a significant amount of pentitol bisphosphate (Figure 14B). The retention times of PGA and PG were approximately 28.5 and 31.5 minutes, respectively (Figure 14B), as documented previously in the literature [52]. The molar ratios of PGA and PG were plotted over four different $[CO_2]/[O_2]$ ratios ranging from 0.06 to 0.30. The data points were fitted to a straight line by the method of linear least squares, forcing the regression line to go through (0, 0). The slope of the resulting straight line represents the specificity factor. With this assay, the specificity factor of carboxysomal and free RuBisCO did not show significant difference (48.7 ± 1.8 and 48.6 ± 1.7, respectively) (Figure 16B).

To assess whether the specificity factor of carboxysomal and free RuBisCO enzyme in the narrow range of $[CO_2]/[O_2]$ was the same as that determined for a broader range of $[CO_2]/[O_2]$, a modified version of the simultaneous assay was performed at a series of narrow $[CO_2]/[O_2]$ ratios from 0.065 to 0.125 at room temperature. PGA and PG were dephosphorylated to glycerate and glycolate, respectively, by calf intestinal alkaline

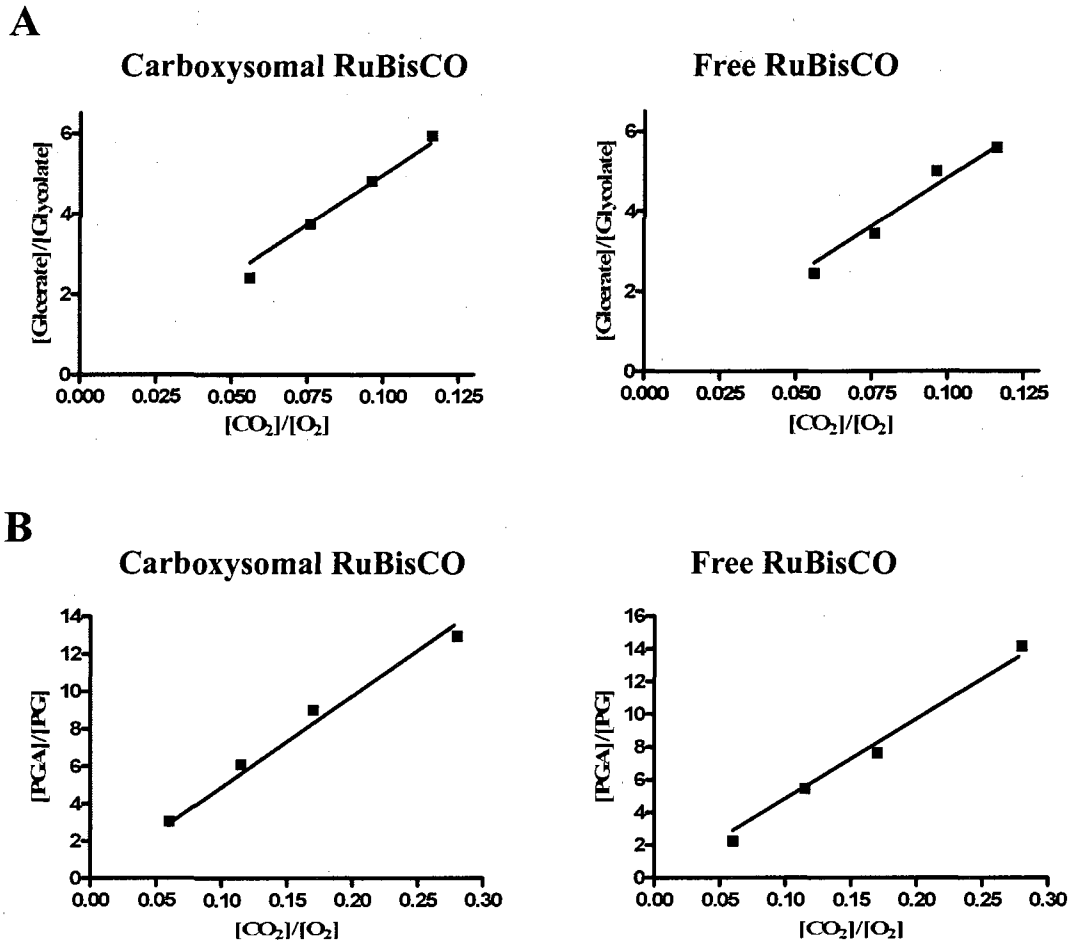


Figure 16. Specificity factor (SF) of carboxysomal and free RuBisCO.

Panel A: PGA and PG were dephosphorylated to glycerate and glycolate, respectively, as described in the literature [65], and the reaction products were separated on a HPLC carbohydrate analysis column. SF(carboxysomes): 49.8 ± 1.4; SF(RuBisCO): 48.4 ± 1.6. Panel B: PGA and PG were separated directly on a MonoQ strong anion exchange column [52]. SF(carboxysomes): 48.7 ± 1.8; SF(RuBisCO): 48.6 ± 1.7. The similar specificity factors of carboxysomal and free RuBisCO suggested that the carboxysomal shell does not limit the passage of O_2 over the 60-minute assay time.

phosphatase (CIAP) prior to separation and quantification. The retention times of glycerate and glycolate on an Aminex 87H HPLC carbohydrate analysis column (BioRad) were approximately 16.4 and 18.2 minutes, respectively (Figure 17) as described in the literature [65]. Likewise, with this assay variation, no significant difference between the specificity factor of carboxysomal and free RuBisCO was detected (49.8 ± 1.4 and 48.4 ± 1.6, respectively) (Figure 16A). These values were very similar as those measured by the MonoQ anion exchange column in the broad range of $[CO_2]/[O_2]$. The practically identical calculated specificity factors for RuBisCO suggested that the carboxysomal shell does not impede the diffusion of O_2 over the 60-minute assay time *in vitro*.

Short time radiometric simultaneous assay

The passage of O_2 molecules across the carboxysomal shell may be rapid, and may result in a fast equilibration of the concentration of O_2 inside and outside of the carboxysome. As a consequence, the ratio of CO_2 to O_2 would be the same on both sides of the carboxysome shell during the 60-minute incubation used in the previous assay, and would not allow distinguishing the permeability of the carboxysomal shell for CO_2 and O_2. To test the effect of incubation time on the ratio of [PGA]/[PG] produced by RuBisCO encapsulated in carboxysomes, the assay time was shortened considerably to 3 minutes. For this series of assays, the concentration of bicarbonate and RuBP was fixed at 1.2 mM and 0.5 mM, respectively; four different concentrations of oxygen over a range of 50 to 420 μM were tested. Any unconsumed RuBP was dephosphorylated to ribulose, which was separated from the dephosphorylated products glycerate and glycolate by a

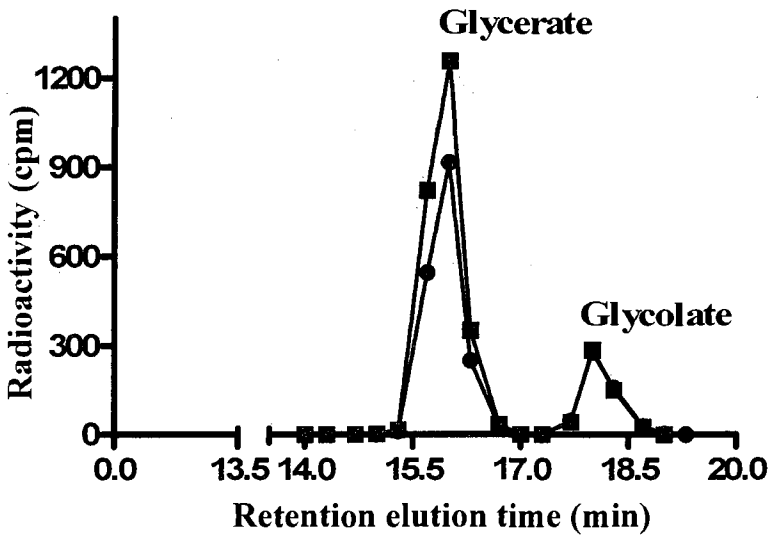


Figure 17. Separation of glycerate and glycolate by a HPLC carbohydrate analysis column.

The assay was performed at 1.2 mM bicarbonate and 330 μM O_2 at room temperature for 3 minutes. PGA and PG were dephosphorylated by calf intestinal alkaline phosphatase and the resulting glycerate and glycolate were separated on a strong anion exchange column. The elution times for glycerate and glycolate were 16.4 and 18.2 minutes, respectively. Carboxysomal RuBisCO synthesized slightly more PGA per PG than free RuBisCO, which suggested that in the carboxysomes carboxylation is slightly favored over oxygenation. ■, carboxysomal RuBisCO; ●, free RuBisCO.

strong anion exchange pre-column before the samples were loaded on the HPLC column. The uncharged ribulose does not bind to the pre-column; only glycerate and glycolate are eluted with 10% (v/v) H_2SO_4. The short assay time resulted in low conversion of 1-[^{3}H]RuBP to PGA and PG, so that the radioactivity of the collected eluate was only approximately 10% of the input radioactivity. Because of the low levels of radioactivity, the more sensitive HPLC separation was chosen to determine the ratio of RuBisCO carboxylation to oxygenation activity. Under the different O_2 concentrations tested from 50 to 420 μM, the ratio of [PGA]/[PG] was slightly higher for the carboxysomal RuBisCO than for free RuBisCO (Figure 18A and Table 3), indicating that carboxylation was favored 1.4 to 1.5 fold over oxygenation by the carboxysomal RuBisCO enzyme under these conditions (Figure 18B). The same assays were also performed at higher concentrations of oxygen (550 to 1050 μM). As the concentration of O_2 increased, the difference in carboxylase to oxygenase activity ratio for carboxysomal and free RuBisCO enzyme disappeared (Figure 18A and Table 3). Taken together, these findings suggested that the carboxylation reaction is slightly favored over the oxygenation reaction by the carboxysomal RuBisCO at low concentration of O_2 and over a short time.

O_2 inhibition constants

O_2 competes with CO_2 for the same active site on RuBisCO. Consequently, in the presence of O_2, few RuBisCO active sites are available for CO_2. According to the Michaelis-Menten equation, if the carboxysomal shell limits the passage of O_2 across the shell, the value of K_c should increase less for the carboxysomal RuBisCO than for free

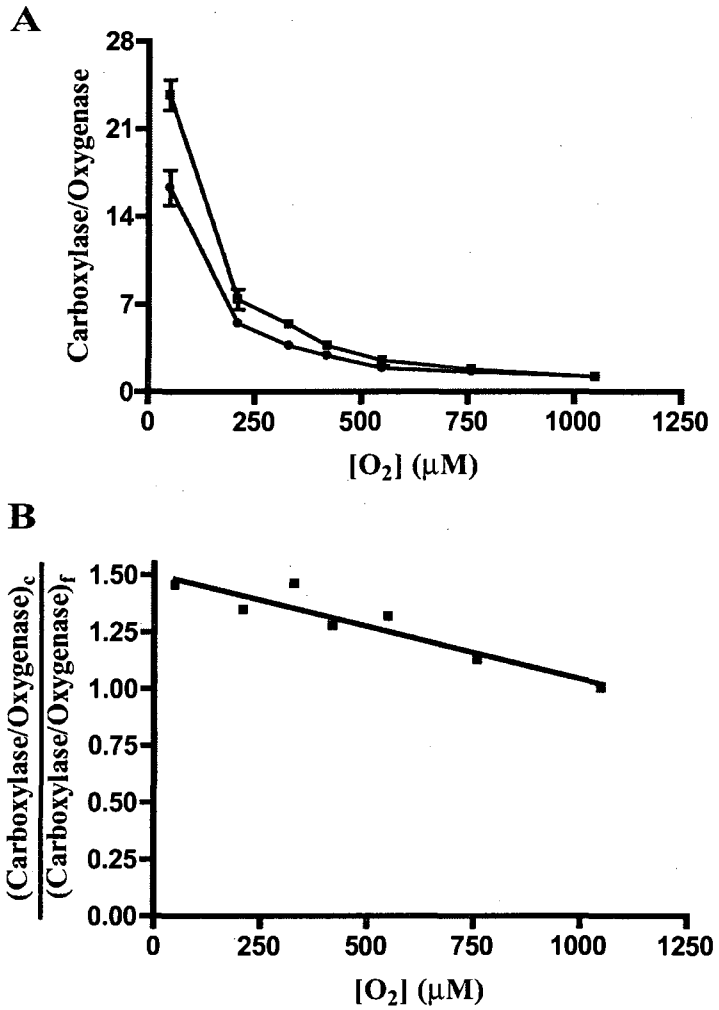


Figure 18. Carboxylase to oxygenase activity ratio for carboxysomal and free RuBisCO enzyme.

Panel A: The carboxylase to oxygenase activity ratios measured with the short time assay were plotted over the concentration of O_2. The difference between values for the carboxysomal and free RuBisCO disappeared as the concentration of O_2 increased. ■, carboxysomal RuBisCO; ●, free RuBisCO. The assays were repeated with three independent preparations. The error bars represent standard deviations.

Panel B: Comparison of carboxylase to oxygenase activity ratios of carboxysomes and free RuBisCO revealed that the carboxysome-encapsulated enzyme synthesized 1.4 to 1.5 times more PGA per PG than free RuBisCO at low O_2 concentration. This ability to discriminate between the two substrates disappeared at higher O_2 concentration. The data points were fitted to a straight line by the method of linear least squares.

Table 3. Short time carboxylase to oxygenase activity ratio at different concentrations of O_2.

$[O_2]$ (μM)	Carboxysomal RuBisCO	Free RuBisCO
50	23.7 ± 1.2	16.3 ± 1.4
210	7.4 ± 0.8	5.5 ± 0.2
330	5.4 ± 0.4	3.7 ± 0.3
420	3.7 ± 0.1	2.9 ± 0.3
550	2.5 ± 0.4	1.9 ± 0.2
760	1.8 ± 0.3	1.6 ± 0.2
1050	1.2 ± 0.1	1.2 ± 0.1

The assay was performed for 3 minutes at a fixed bicarbonate concentration of 1.2 mM. The products, PGA and PG, were dephosphorylated by alkaline phosphatase before being separated by HPLC. The ratio of glycerate to glycolate was calculated to determine the ratio of carboxylase to oxygenase activity of carboxysomal and free RuBisCO enzyme. At low O_2 concentration, the carboxysomal enzyme synthesized more PGA per PG than free RuBisCO. As the concentration of O_2 increases, the carboxylation to oxygenation ratios of these two forms of RuBisCO became more similar. Values are averages ± standard deviations for three measurements using three independent carboxysome preparations.

RuBisCO as the concentration of O_2 increases and should be accompanied by a higher O_2 inhibition constant for the carboxysomal enzyme. To determine the O_2 inhibition constant, CO_2 fixation rates by carboxysomes and free RuBisCO were plotted over seven different concentrations of CO_2 from 20.3 to 1123 μM, and the plot was fitted to the Michaelis-Menten equation to calculate V_{max} and K_c. The K_c values were plotted against the respective O_2 concentrations from 50 to 420 μM to determine the O_2 inhibition constant. The O_2 inhibition constant of carboxysomal RuBisCO was 1901 μM, a value that is almost threefold higher than that calculated for free RuBisCO (648 μM) (Figure 19). This result indicated that CO_2 fixation in carboxysomes is not as sensitive to inhibition by O_2 as is that by the free enzyme. It appears that the carboxysomal shell can discriminate between CO_2 and O_2, thereby inhibiting the oxygenation reaction and enhancing the carboxylation activity of the encapsulated RuBisCO enzyme. These assays were repeated with three independent preparations. The error bars of the K_c values of carboxysomal and free RuBisCO at some O_2 concentrations overlapped (Figure 19C), thereby the difference of estimated K_c values between two RuBisCO forms is not statistically significant. The difference of O_2 inhibition constants obtained from the linear regression analyses should be evaluated carefully.

To obtain more accurate O_2 inhibition constants for both RuBisCO forms, these assays may need to be replicated more times to decrease the error bar of the K_c value or a stopped-flow technique that can measure the rate of enzymatic reaction over a short time needs to be adopted.

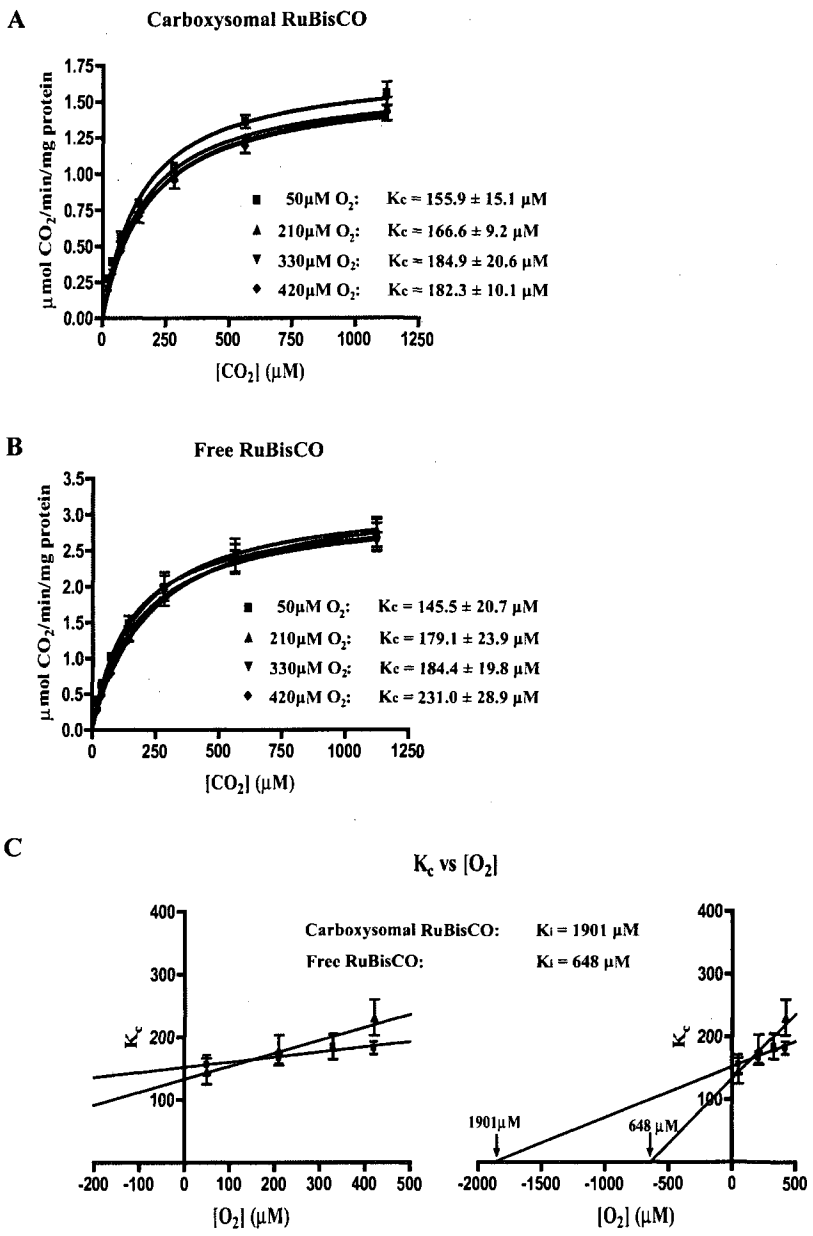


Figure 19. O_2 inhibition constants for carboxysomal and free RuBisCO.

Panel A: K_c values for carboxysomal enzyme.
Panel B: K_c values for free RuBisCO.
Panel C: The K_c values calculated for carboxysomal and free RuBisCO were plotted over O_2 concentrations. The left and right plots have different scales on X-axis. The data points were fitted to a straight line by the method of linear least squares. The X-intercept is $-K_i$ (indicated on right plot in Panel C). The assays were repeated with three independent preparations. The error bars represent standard errors in panels A, B and C.

Characterization of CsoS2 Protein

As the CO_2 fixation reaction proceeds within the carboxysome, RuBP, the second substrate of RuBisCO, and PGA, the product of its carboxylation reaction, are consumed and accumulated in the carboxysomes, respectively. Both species are negatively charged molecules. X-ray structural studies revealed that the CsoS1A proteins form hexamers with positively charged central pores [125]. The size of the central pore is approximately 4 Å and sulfate ion was found to reside within the pore [125]. Given the similar shape and size between sulfate and bicarbonate ions, the pore was hypothesized to facilitate the bicarbonate ion across the shell. However, it is likely to be too small to accommodate larger molecules, such as RuBP and PGA. It is still unknown how RuBP and PGA are transported into and out of the carboxysome. CsoS2 protein is the largest carboxysomal protein [13, 28], and its structure and function are still completely unknown. The theoretical isoelectric points of seven CsoS2 proteins from several chemo- and photo-autotrophic bacteria are calculated from their primary structure and they are all in the range of 9.1 to 10.0 (Table 4), which indicates that CsoS2 protein should carry a positive charge in the cytoplasm, in which the pH is estimated to be approximately 7.4 [30, 40]. The high pI value of CsoS2 protein suggested that this shell protein could be a potential candidate to interact with these negatively charged molecules [30] to transport them across the carboxysome shell.

The *csoS2* gene only exists in the *cso* operons of chemoautotrophs and α-cyanobacteria that encode α-carboxysomes [30, 110]. The primary structures of CsoS2 proteins contain many V/I-T/S-G repeat motifs (Figure 20) [30] that may be related an unusual three dimensional structure for their potential roles in the transportation of

Table 4. Calculated molecular masses and isoelectric points (pIs) of CsoS2 proteins

Organism	Expected mass (kDa)	pI
Halothiobacillus neapolitanus	92	9.06
Thiomicrospira crunogena XCL2	68	9.74
Thiobacillus denitrificans	68	9.66
Acidithiobacillus ferrooxidans	79	9.36
Thiomonas intermedia	94	9.99
Synechococcus sp. WH 8102	81	9.94
Prochlorococcus marinus str. MIT 9301	82	9.64

The expected molecular masses and isoelectric points of CsoS2 proteins were calculated from their amino acid sequences, using the MW/pI calculation program at http://www.expasy.ch/tools/pi_tool.html.

MPSQSGMNPADLSGLSGKELARARRAALSKQGKAAVSNKTASVNRSTKQAASSINTNQVRSSVNEVPTDYQMADQLCSTIDHADFGTESNRVRDLCRQRREALSTIGKKAAKTTG
KPSGRVRPQQSVVHNDAMIENAGDTNQSSSTSLNNELSEICSIADDMPERFGSQAKTVRDICRARRQALSERGTRAVPPKPQSQGGPGRNGYQIDGYLDTALHGRDAAKRHREML
CQYGRGTAPSCKPTGRVKNSVQSGNAAPKKVETGHTLSGGS
VTGTQVDRKSH**VTG**NEPGTCRA**VTG**TEYVGTEQFTSFCNTSPKPNATKVNVTTTARGRP
VSGTEVSRTEK**VTG**NESGVCRN**VTG**TEYMSNEAHFSLCGTAAKPSQADKVMFGATARTHQV
VSGSDEFRPSS**VTG**NESGAKRT**ITG**SQYADEGLARLTINGAPAKVARTHTFAGSD
VTGTEIGRSTR**VTG**DESGSCRS**ISG**TEYLSNEQFQSFCDTKPQRSPFKVGQDRTNKGQS
VTGNLVDRSEL**VTG**NEPGSCSR**VTG**SQYGQSKICGGGVGKVRSMRTLRGTS
VSGQQLDHAPK**MSG**DERGGCMP**VTG**NEYYGREHFEPFCTSTPEPEAQSTEQSLTCEGQI
ISGTSVDASDL**VTG**NEIGEQQL**ISG**DAYVGAQQTGCLPTSPRFNQTGNVQSMGFKNTNQPEQNFAPGEVMPTDFSIQTPARSAQNR
ITGNDIAPSGR**ITG**PGMLATGL**ITG**TPEFRHAARELVGSPQPMAMAMANRNKAAQAPVVQPEVVATQEK PELVCAPRSDQMDR
VSGEGKERCH**ITG**DDWSVNKH**ITG**TAGQWASGRNPSMRGNARVVETSAFANRNVPKPEKPGSK
ITGSSGNDTQGSLIT**YSG**GARG

Figure 20. The primary structure of CsoS2 protein from *Halothiobacillus neapolitanius*

The primary structure of CsoS2 protein contains multiple V/I-T/S-G unusual amino acid repeat motifs in the C-terminal two-thirds of the entire sequence. These repeat motifs are labeled in red.

negatively charged molecules. The *csoS2* gene is present in a single copy in the *H. neapolitanus cso* operon but is expressed as two distinct polypeptides – CsoS2A and CsoS2B [13]. N-terminal sequencing analysis revealed that the first 15 amino acids are identical in the two CsoS2 forms [13], and a polyclonal antibody against CsoS2A recognizes CsoS2B protein. Both polypeptides have the similar epitopes and the discrepancy between two CsoS2 proteins is attributed to differential post-translational glycosylation [13].

Expression of recombinant HnCsoS2 protein in E. coli

Given its unique primary structure, the CsoS2 protein may have an unusual three dimensional structure that allows the protein to transport RuBP and PGA across the carboxysome shell. To explore its three dimensional structure and understand the molecular difference between the two CsoS2 polypeptide forms, the *csoS2* gene from *H. neapolitanus* was expressed in *E. coli* using two different cloning strategies (Figure 21A).

First, the *csoS2* gene was cloned into a pProEX expression vector, which added a hexa-His tag to the N-terminus of the resulting recombinant CsoS2 protein (Figure 21A). Recombinant polypeptides of observed molecular weights corresponding to those of CsoS2A and CsoS2B, respectively, were expressed in *E. coli* (Figure 21B). The majority of the recombinant soluble CsoS2 protein was in the CsoS2A form. The concentration of soluble CsoS2A and CsoS2B protein that could be obtained was low, approximately 300 μg/ml. Most of the recombinant CsoS2 protein formed inclusion bodies [13]. To increase the yield of soluble CsoS2 protein, the proteins in a crude cell extract were solubilized

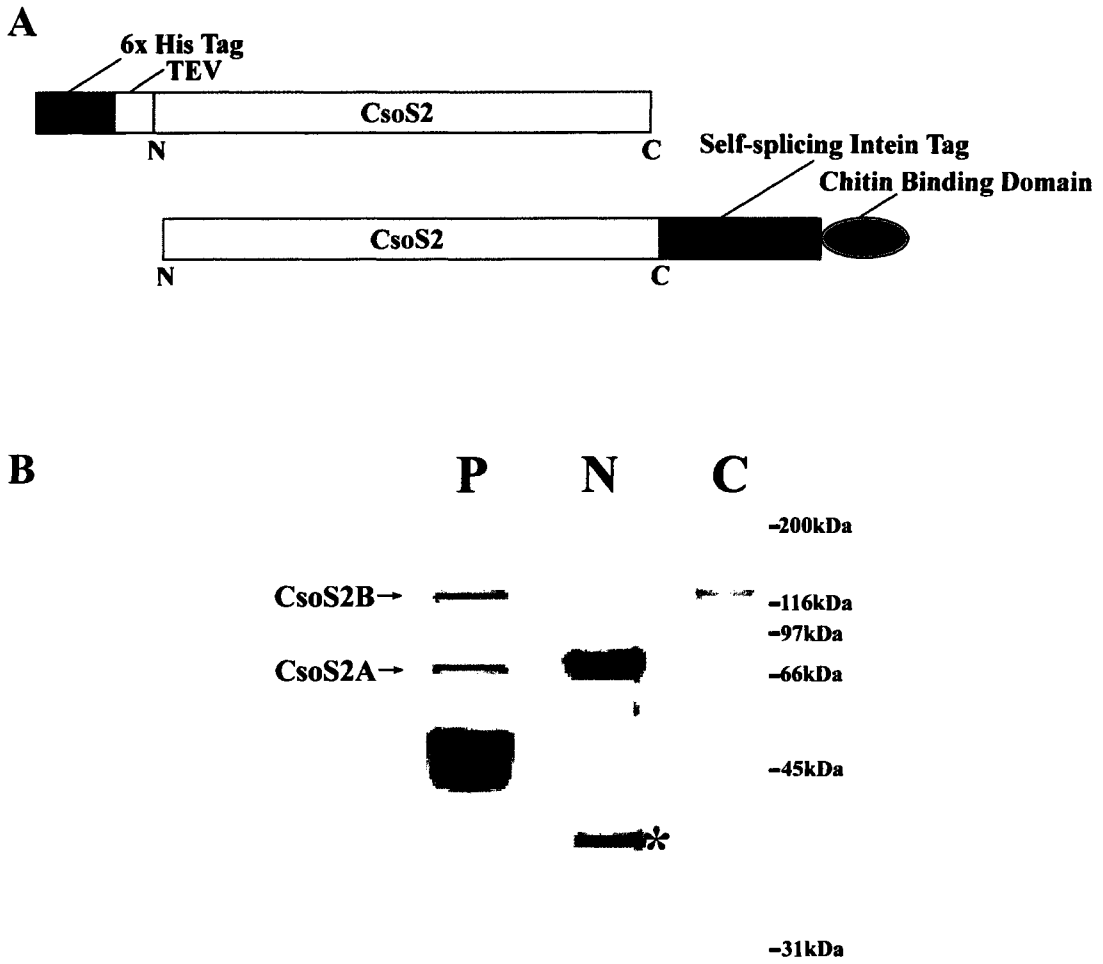


Figure 21. Expression of recombinant CsoS2 protein in the pProEX® and the IMPACT® systems.

Panel A: Cloning strategies adopted to express the csoS2 gene from H. neapolitanus in E. coli. Panel B: Expression patterns of recombinant CsoS2 protein in the pProEX® and the IMPACT® systems. Protein molecular weight standards are indicated on the right. Protein fractions were analyzed by SDS-PAGE (7.5% polyacrylamide gel under reducing conditions) and visualized by staining with Gel Code Blue. P: Purified H. neapolitanus carboxysomes, N: recombinant CsoS2 protein fused to a hexa-His tag at the N-terminus; C: recombinant CsoS2 protein fused to a self-splicing intein tag with chitin binding domain at the C-terminus.
In purified *H. neapolitanus* carboxysomes, the ratio of CsoS2A and CsoS2B is approximately 1. In the pProEX® system, two forms of CsoS2 protein could be purified by Ni-NTA chromatography. Most of the His-tagged recombinant CsoS2 was expressed in the CsoS2A form. The band with asterisks may be a contaminating E. coli protein or a CsoS2 degradation product. In the IMPACT® system, only CsoS2B protein was eluted from the affinity column.

with 6 M guanidine-HCl prior to loading on a Ni-NTA affinity column and eluting the denatured CsoS2A and CsoS2B proteins with 8 M urea. The denaturant is changed from guanidine-HCl in the lysis buffer to urea in the elution buffer because guanidine-HCl interferes with the downstream SDS-PAGE analysis of the eluate. Under these conditions, the yield of rCsoS2 protein could be increased to approximately 15 mg per liter of *E. coli* culture, with the majority representing the CsoS2A form. An unknown polypeptide of approximately 35kDa co-purified with CsoS2 on the Ni-NTA column. This protein is likely an *E. coli* contaminant, or a CsoS2 protein degradation product due to its strong binding capacity to the Ni-NTA resin. During the removal of urea by dialysis, the His-tagged rCsoS2 polypeptides formed an insoluble precipitate that resulted in a low concentration and yield of soluble recombinant protein.

To overcome the insolubility of rCsoS2, the protein was expressed in an IMPACT® expression system, which is designed to add a self-splicing intein tag and a chitin-binding domain at the C-terminus of the expressed protein (Figure 21A). These hydrophilic domains should increase the solubility of rCsoS2 protein in the *E. coli* cytosol. The intein tag with the chitin-binding domain self-cleaves in the presence of a reducing reagent, such as DTT or β-ME, and remains bound to the chitin affinity resin. Recombinant target protein can subsequently be eluted from the column. Interestingly, the rCsoS2 protein in the eluted fractions only consisted of the larger form CsoS2B (Figure 21B), yielding approximately 1 mg per liter of *E. coli* culture at a concentration of approximately 200 μg/ml. Because the desirable protein concentration is at least 1 mg/ml for crystallization trials, an attempt was made to concentrate the rCsoS2B protein. Unfortunately, like the His-tagged rCsoS2 preparation, this rCsoS2 protein sample also

formed an insoluble precipitate during the concentration process, and a final concentration higher than 400 μg/ml could not be obtained.

N-terminally truncated rCsoS2 protein generated in E. coli

According to the primary structure of *H. neapolitanus* CsoS2 protein, most of the V/I-T/S-G repeats exist between positions 270 and 869. This portion may have a more rigid conformation than the first 250 amino acids and may be crystallized more easily than the full length CsoS2 protein. Thus, an N-terminally truncated CsoS2 protein that does not contain the first 250 amino acids was expressed in *E. coli*. The truncated *csoS2* gene lacking the first 750 nucleotides but containing an added ATG start codon was cloned into the pET-22b expression vector. The C-terminus of the recombinant target protein was fused to a hexa-His tag. The N-terminally truncated CsoS2 protein was purified by chromatography on a Ni-NTA column using a 25 to 1000 mM imidazole gradient for elution. During the affinity purification, two polypeptides were eluted from the column (Figure 22). A polypeptide of observed molecular weight 38 kDa had a lower affinity to the Ni-NTA column and was eluted at a low concentration of imidazole (20 to 150 mM). A polypeptide of observed molecular weight 90 kDa was eluted at imidazole concentrations of 250 to 500 mM. The two abundant bands were excised from the gel and digested with trypsin for peptide mapping by MALDI-ToF. The larger polypeptide contained peptide sequences derived from CsoS2 and was covered to the arginine residue next to the last amino acid; therefore, it was identified as rΔ250CsoS2B protein. For the

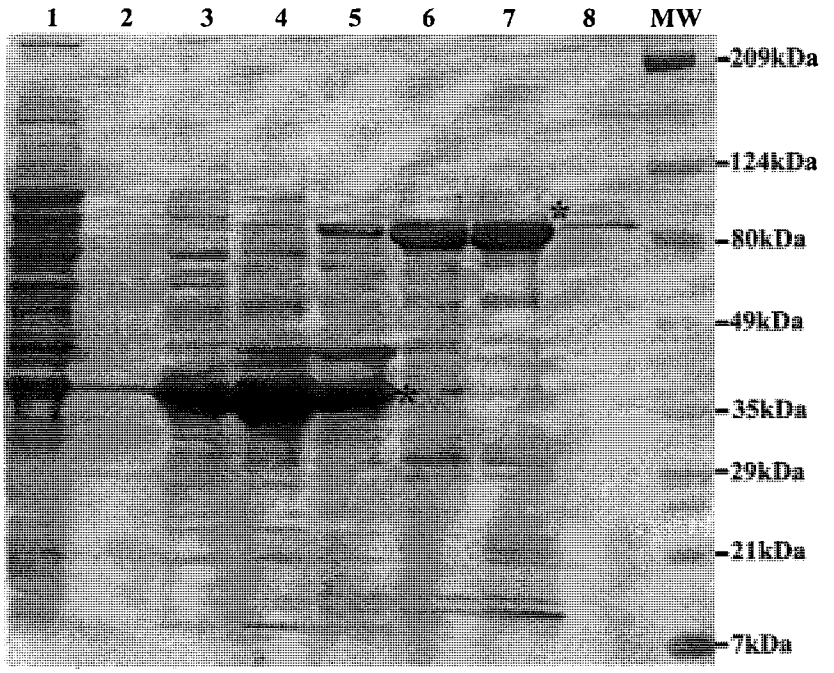


Figure 22. Expression of the N-terminally truncated rΔ250CsoS2B protein in *E. coli*.

A truncated *csoS2* gene that lacks the first 750 nucleotides was cloned into pET-22b vector and used to express a truncated version of CsoS2 with a C-terminal hexa-His tag. Lane 1: Ni-NTA column flow through fraction; Lane 2: Ni-NTA wash fraction eluted with 10 mM imidazole; Lanes 3 to 8: Elution fractions eluted with 20 mM, 50 mM, 150 mM, 250 mM, 500 mM and 1000 mM imidazole, respectively; MW: protein molecular weight standards. The protein bands labeled with an asterisk were excised and digested with trypsin for peptide mapping by MALDI-ToF. The polypeptide of observed molecular weight 90 kDa was identified as the rΔ250CsoS2B protein. The polypeptide of observed molecular weight 38 kDa appeared to be unrelated to CsoS2. Most of rΔ250CsoS2B protein appeared in the fractions eluted with 250 and 500 mM imidazole.

polypeptide of apparent molecular weight 38 kDa, the mass spectrometry data were inconclusive. Not enough peptide fragments were obtained to allow identification of this polypeptide. The protein may be an *E. coli* contaminant with some binding affinity for the Ni-NTA resin or may be a degraded polypeptide derived from the N-terminally truncated rCsoS2 protein. The concentration of rΔ250CsoS2B protein was approximately 1.5 mg/ml. Approximate 10 mg were obtained from one liter of *E. coli* culture. The relatively high concentration and yield of rΔ250CsoS2B suggested that the solubility of rΔ250CsoS2B is much better than that of full length CsoS2 protein. The soluble rCsoS2B and rΔ250CsoS2B protein have been sent to Dr. Cheryl Kerfeld's lab for the crystallization studies. To date, no protein crystals have been obtained.

Generation of the HnSPAS2 mutant

CsoS2 protein is present in two forms in *H. neapolitanus* carboxysomes [13] and so is the recombinant CsoS2 protein expressed in *E. coli* reported in the previous literature and this study [13]. Interestingly, CsoS2 protein displayed different expression patterns in several chemoautotrophic bacteria that harbor almost identical *cso* operons. CsoS2 protein yields two polypeptide species in the carboxysomes of the sulfur bacterium *Thiomonas intermedia*, but only one form is detected in purified carboxysomes from *Thiomicrospira crunogena* (Figure23, Balaraj Menon purified the carboxysomes from *T. crunogena* and performed SDS-PAGE and electron microscopy). Previously, Baker *et al.* had proposed that the two CsoS2 forms in *H. neapolitanus* carboxysomes are differentially glycosylated [13]. To further understand how the two forms of CsoS2

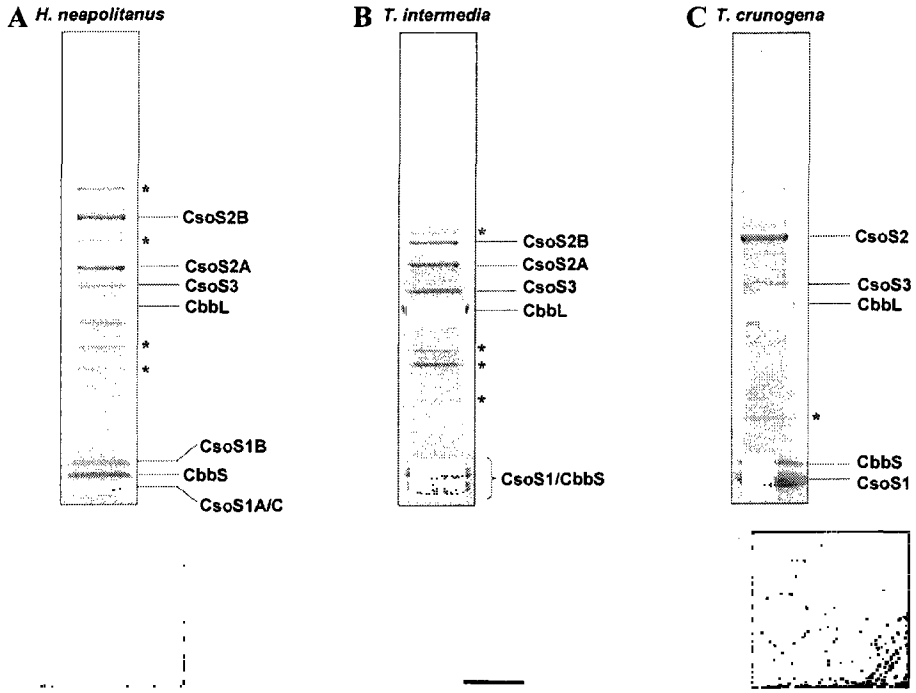


Figure 23. Comparison of composition and morphology of carboxysomes from *Halothiobacillus neapolitanus, Thiomonas intermedia* and *Thiomicrospira crunogena.*

Panels A, B and C: Protein composition of the purified carboxysomes of *H. neapolitanus*, *T. intermedia* and *T. crunogena*, respectively. Transmission electron micrographs of the corresponding purified carboxysomes, negatively stained with 1% (v/v) ammonium molybdate, are shown below the SDS-PAGE gel pictures.

CsoS2 protein is present as two polypeptides in purified carboxysomes of *H. neapolitanus* and *T. intermedia*, but appears to form only one polypeptide in *T. crunogena*. The shapes and sizes of the three carboxysomes are similar. The polypeptide bands labeled with an asterisk are likely carboxysome protein complexes that are resistant to the disruption by SDS or may be some contaminants that nonspecifically co-purify with the isolated carboxysomes. Bar: 100 nm. (Balaraj Menon purified the carboxysomes from *T. crunogena* and performed SDS-PAGE and electron microscopy.)

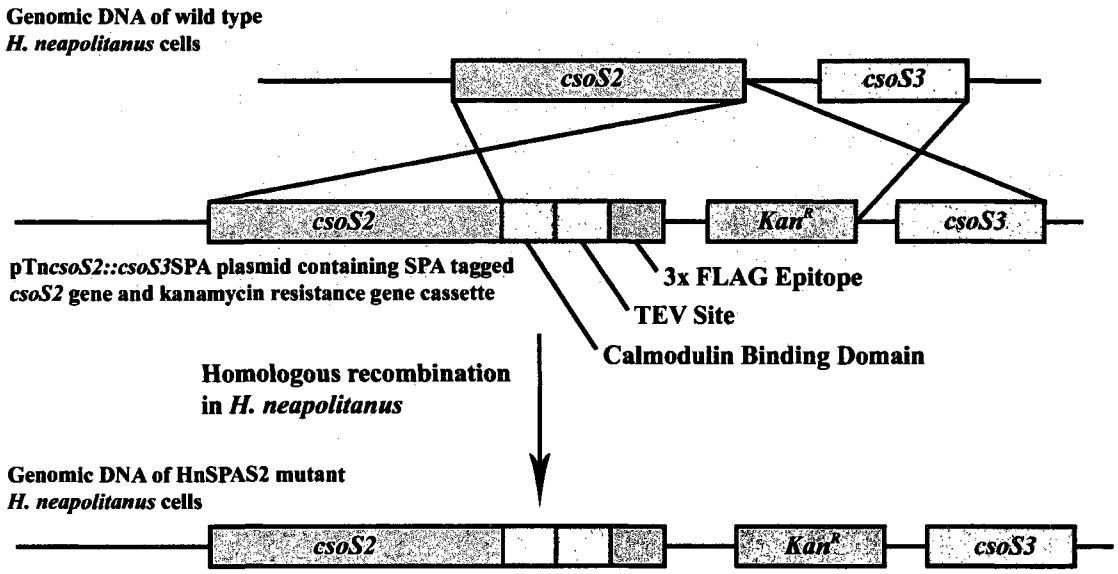


Figure 24. A sequential peptide affinity (SPA) tag sequence added at the 3' end of the *csoS2* gene by homologous recombination in *H. neapolitanus*.

The SPA tag contains a 3x FLAG epitope and a calmodulin binding domain separated by a TEV protease site [131]. A kanamycin resistance gene cassette follows the SPA tag for selection purpose. The SPA tag and the kanamycin resistance gene cassette first were integrated into the pTn*csoS2::csoS*3 plasmid [11] to obtain longer (approximately 2,000 bp) homologous flanking regions, which facilitate the subsequent gene replacement in *H. neapolitanus* by homologous recombination [46].

protein are generated from a single *csoS2* gene and to identify the location of CsoS2 protein in the carboxysome, a small sequential peptide affinity (SPA) tag was added at the C-terminus of CsoS2 protein by homologous recombination (Figure 24). The added affinity tag at the C-terminus of CsoS2 protein could also help trap intermediates of *in vitro* carboxysome assembly assays and reveal *bona fide* interactions among carboxysomal proteins *in vivo* that would enhance our understanding of carboxysome biogenesis.

The SPA tag contains a tandem 3x FLAG epitope, a tobacco etch virus (TEV) protease cleavage site and a calmodulin binding domain that allow the target protein and its interacting protein partners to bind to an anti-FLAG affinity column and to agarose beads that are conjugated with calmodulin, respectively. During the purification of protein complexes, the 3x FLAG epitope can be cleaved off by TEV protease, leaving protein complexes that only carry the calmodulin binding domain for a second round of affinity purification on agarose beads conjugated with calmodulin. These features of the tandem affinity tag should decrease nonspecific
binding by contaminants and eliminate them from the target protein complexes.

A pair of primers was designed to amplify the SPA tag and a kanamycin resistance gene cassette by PCR, using the pJL148 plasmid as the template [131]. To create an in-frame fusion between the PCR product and the 3' end of the *csoS2* coding sequence by homologous recombination, the forward primer contained 45 bp of nucleotide sequence corresponding to the region of the *csoS2* gene immediately upstream from the stop codon, followed by the first 17 bp of the SPA tag sequence. The reverse primer contained 45 bp of nucleotide sequence immediately downstream of the *csoS2*

stop codon, followed by the last 20 bp of the kanamycin resistance gene. The *E. coli* DY 330 strain was co-transformed with the resulting PCR product and with pTn*csoS2::csoS3* [11, 130]. This strain carries the lambda phage Red recombinase gene that facilitates precise homologous recombination between very short 45 bp homologous regions and was used to produce the SPA tagged version of *csoS2*, followed by a kanamycin resistance gene cassette in the pTn*csoS2::csoS*3SPA plasmid. In this plasmid, the SPA tag and kanamycin resistance gene cassette are flanked by the long (approximately 2,000 bp) homologous regions that facilitated the subsequent gene replacement in *H. neapolitanus* by homologous recombination (Figure 24). Wild type mid-exponential phase *H. neapolitanus* cells were electroporated with the pTn*csoS2::csoS*3SPA plasmid and transformants were selected in the presence of 50 μg/ml kanamycin and in 5% CO_2-enriched air. The genomic DNA of three mutant clones was screened by PCR amplification using the primers shown in Figure 25A. The pair of primers S2f4018 and SPAr3204 amplifies a 1.6 kbp fragment from mutant and no product from wild type DNA; the SPAf2212 and SPAr3204 pair yields a 1.9 kbp fragment with mutant and no product with wild type template; a 2.6 kbp mutant and a 800 bp wild type fragment are obtained with the S2f4018 and S3r4851primers; the primer pair SPAf2212 and S3r4851 yields a 1.0 kbp mutant amplicon and no product from wild type DNA (Figure 25B and C). The expected amplification products were obtained from all three HnSPAS2 mutant clones (Figure 25). The genomic DNA of two positive clones was sequenced, and one of these clones was used to establish a working culture. The resulting mutant was named HnSPAS2.

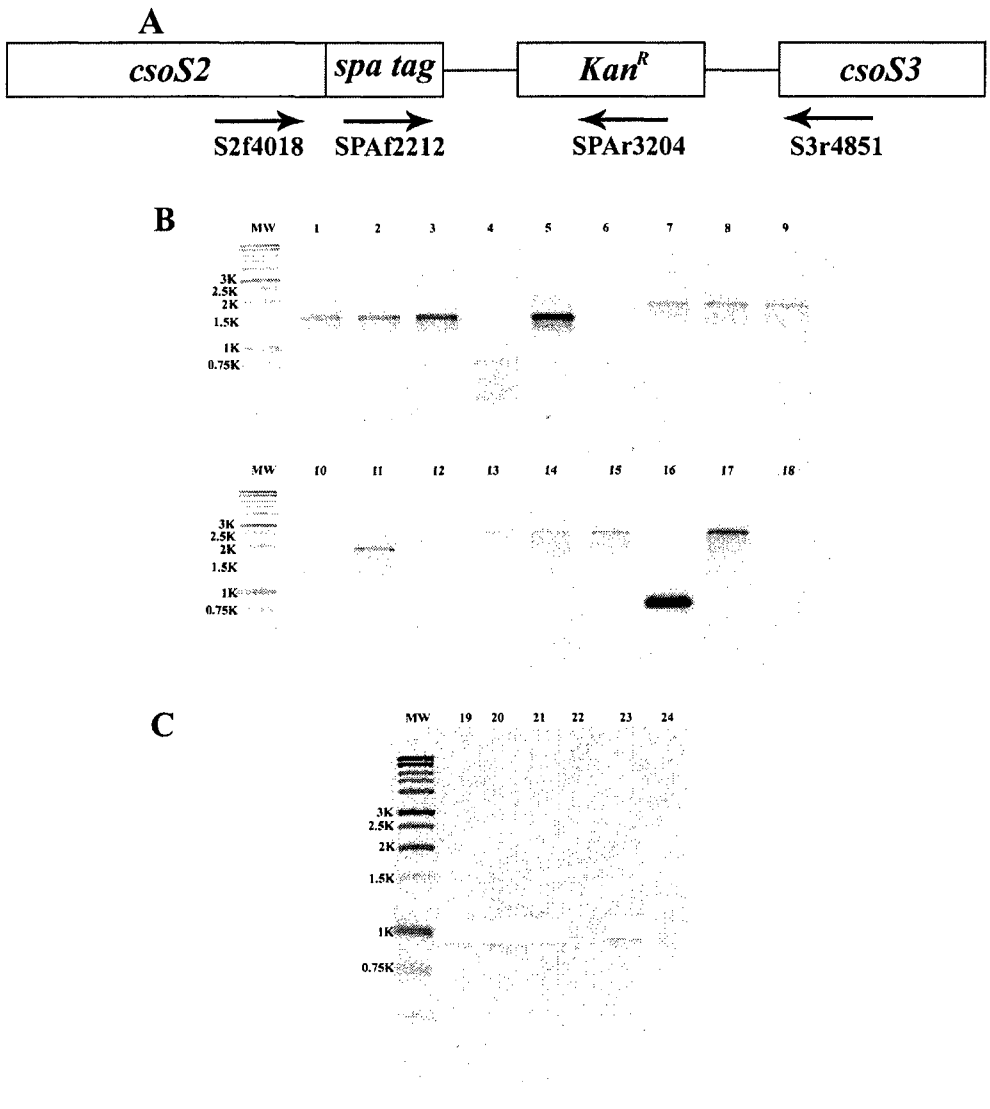


Figure 25. Screening of HnSPAS2 mutant clones by PCR amplification.

Two forward and two reverse primers (Table 1 in Experimental Procedures) were used to screen three HnSPAS2 transformants.
Panel A: The positions of the four primers in the genomic DNA of the HnSPAS2 mutant.
Panels B and C: Mutant screening by PCR amplification.

Lanes 1-6: S2f4018 and SPAr3204 primer pair; Lanes 7-12: SPAf2212 and SPAr3204 primer pair; Lanes 13-18: S2f4018 and S3r4851 primer pair; Lanes 19-24: SPAf2212 and S3r4851 primer pair. Lanes 1, 7, 13, 19: HnSPAS2 clone 1; Lanes 2, 8, 14, 20: HnSPAS2 clone 2; Lanes 3, 9, 15, 21: HnSPAS2 clone 3; Lanes 4, 10, 16, 22: wild type; Lanes 5, 11, 17, 23: pTn*csoS2::csoS3*SPA; Lanes 6, 12, 18, 24: negative control.
The expected amplification products were obtained from all three HnSPAS2 mutant clones.

Characterization of carboxysomes from the HnSPAS2 mutant

The HnSPAS2 mutant was able to grow in an ambient CO_2 environment. Unlike the high CO_2 requiring *csoS3::Km* mutant, it was not necessary to grow the HnSPAS2 mutant in 5% CO_2-enriched air and shift to ambient CO_2 environment prior to carboxysome preparation to increase the number of carboxysomes produced by the cells. The HnSPAS2 mutant was cultured like the wild type in a bioreactor aerated with air at a dilution rate of 0.08^{-1} at pH 6.4. Carboxysomes prepared by the standard method described in Experimental Procedures were obtained in amounts comparable to wild type carboxysomes, approximately 10 mg of carboxysome protein per 8 g of wet cell pellet. Transmission electron microscopy revealed that their shapes and sizes were indistinguishable from their wild type counterparts (Figure 26B). The composition of purified wild type and HnSPAS2 mutant carboxysomes was analyzed by SDS-PAGE. Interestingly, the CsoS2A protein in HnSPAS2 mutant and wild type carboxysomes had the same apparent molecular weight (approximately 80 kDa) (Figure 26A). However, the CsoS2B protein in the mutant carboxysomes with an observed molecular weight of approximately 140 kDa was slightly larger than wild type CsoS2B protein (130 kDa) (Figure 26A). To test whether the slower electrophoretic mobility of the HnSPAS2 mutant CsoS2B protein was due to the additional amino acids contributed by the SPA tag, cell extracts of wild type and two HnSPAS2 mutant clones were subjected to SDS-PAGE. Immunoblots were probed with anti-CsoS2 and anti-FLAG antibodies and showed that the CsoS2B protein in the HnSPAS2 mutants carries the SPA tag and migrates slightly more slowly than wild type CsoS2B protein. CsoS2A protein, on the other hand, does not

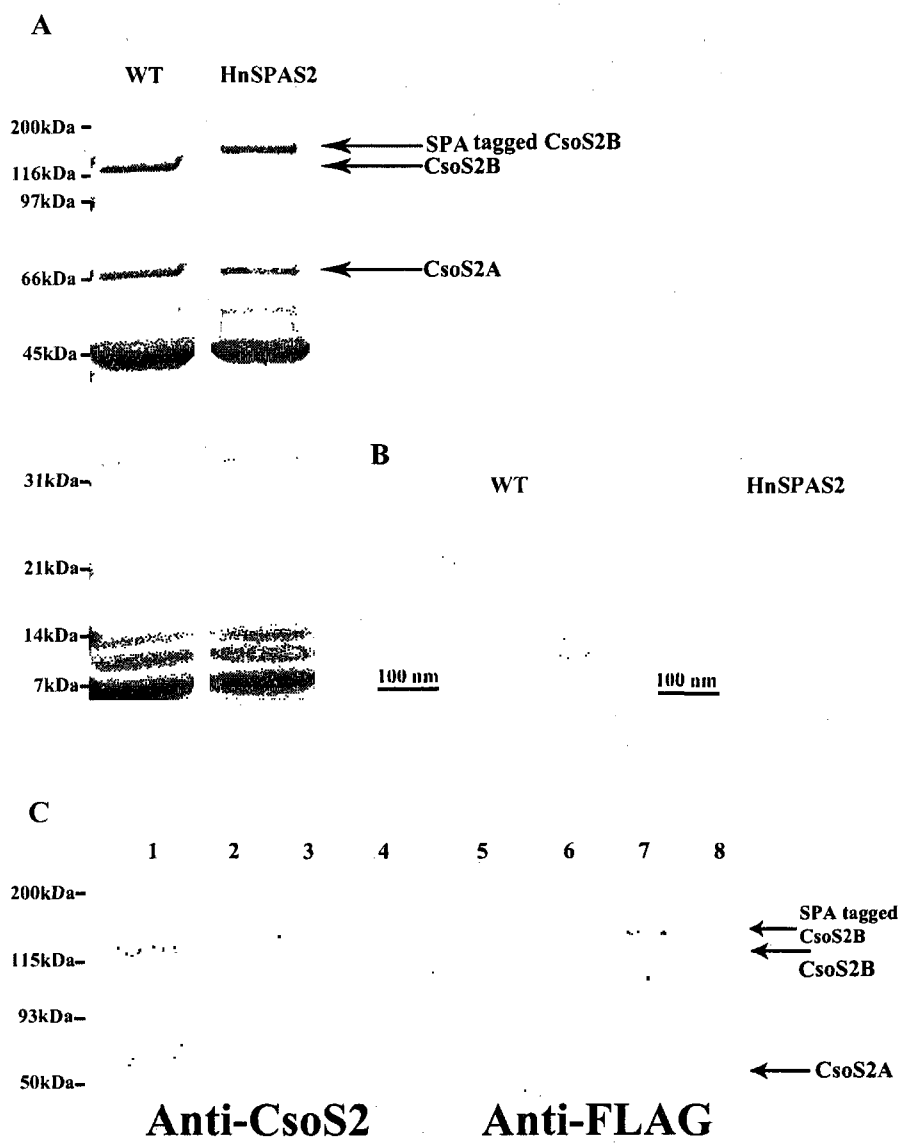


Figure 26. Morphology and composition of HnSPAS2 mutant carboxysomes.

A: SDS-PAGE analysis of purified wild type and SPAS2 mutant carboxysomes.
B: Morphology of purified wild type and SPAS2 mutant carboxysomes.
C. Immunoblot analysis of wild type and SPAS2 mutant cell extracts. Lanes 1 and 5: wild type; Lanes 2 and 6: HnSPAS2 mutant clone 1; Lanes 3 and 7: HnSPAS2 mutant clone 2; Lanes 4 and 8: purified wild type carboxysomes. Protein molecular weight standards are indicated on the left in Panels A and C.
Only the region corresponding to the higher molecular weight polypeptides is shown. No cross-reactivity of the anti-FLAG and anti-CsoS2 antibodies with small polypeptides was observed. CsoS2B protein is slightly larger in HnSPAS2 mutant than in wild type cells. Immunoblot analysis revealed that only CsoS2B protein contains the FLAG epitope tag.

carry the SPA tag (Figure 26C). These findings suggested that CsoS2B protein is a full length CsoS2 protein, and that CsoS2A protein is a C-terminally truncated version. To confirm that CsoS2B protein is a full length CsoS2 protein and attempt to reveal the last C-terminal amino acid of the C-terminally truncated CsoS2A protein, the wild type CsoS2A and CsoS2B polypeptide bands were excised from an SDS polyacrylamide gel and digested with trypsin for peptide mapping by MALDI-ToF. The results obtained with three independent preparations of CsoS2A and CsoS2B were combined, yielding the peptide fragments marked in red in Figure 27. CsoS2B protein was traced to the arginine residue next to the last amino acid of the CsoS2 protein sequence deduced from the nucleotide sequence of its gene. The CsoS2A protein could only be covered to the arginine residue at position 836 (Figure 27A and B). The mass spectrometry results confirmed that CsoS2B protein is the full length CsoS2 protein but could not reveal the last amino acid of CsoS2A. A C-terminal amino acid sequencing analysis may need to be adopted to reveal the last amino acid of CsoS2A protein.

The finding that CsoS2A protein is a C-terminally truncated CsoS2 protein helped to interpret the puzzling finding that only CsoS2B protein was eluted from the affinity column if the protein carried a C-terminal tag. Crude cell extracts were analyzed by SDS-PAGE, and immunoblots were probed with anti-CsoS2 antibody. The analyses showed that the majority of recombinant CsoS2 protein existed in the CsoS2A form in the induced cells (Figure 28). Therefore, most of the recombinant CsoS2 protein did not carry the chitin binding domain and could not bind to the chitin affinity column. Only a small amount of full length CsoS2B fusion protein carrying the C-terminal tag was recovered from the column (Figure 21).

A

```
  1 MPSQSGMNPA DLSGLSGKEL ARARRAALSK QGKAAVSNKT ASVNRSTKQA
 51 ASSINTNQVR SSVNEVPTDY QMADQLCSTI DHADFGTESN RVRDLCRQRR
101 EALSTIGKKA AKTTGKPSGR VRPQQSVVHN DAMIENAGDT NQSSSTSLNN
151 ELSEICSIAD DMPERFGSQA KTVRDICRAR RQALSERGTR AVPPKPQSQG
201 GPGRNGYQID GYLDTALHGR DAAKRHREML CQYGRGTAPS CKPTGRVKNS
251 VQSGNAAPKK VETGHTLSGG SVTGTQVDRK SHVTGNEPGT CRAVTGTEYV
301 GTEQFTSFCN TSPKPNATKV NVTTTARGRP VSGTEVSRTE KVTGNESGVC
351 RNVTGTEYMS NEAHFSLCGT AAKPSQADKV MFGATARTHQ VVSGSDEFRP
401 SSVTGNESGA KRTITGSQYA DEGLARLTIN GAPAKVARTH TFAGSDVTGT
451 EIGRSTRVTG DESGSCRSIS GTEYLSNEQF QSFCDTKPQR SPFKVGQDRT
501 NKGQSVTGNL VDRSELVTGN EPGSCSRVTG SQYGQSKICG GGVGKVRSMR
551 TLRGTSVSGQ QLDHAPKMSG DERGGCMPVT GNEYYGREHF EPFCTSTPEP
601 EAQSTEQSLT CEGQIISGTS VDASDLVTGN EIGEQQLISG DAYVGAQQTG
651 CLPTSPRFNQ TGNVQSMGFK NTNQPEQNFA PGEVMPTDFS IQTPARSAQN
701 RITGNDIAPS GRITGPGMLA TGLITGTPEF RHAARELVGS PQPMAMAMAN
751 RNKAAQAPVV QPEVVATQEK PELVCAPRSD QMDRVSGEGK ERCHITGDDW
801 SVNKHITGTA GQWASGRNPS MRGNARVVET SAFANRNVPK PEKPGSKITG
851 SSGNDTQGSL ITYSGGARG
```

B

```
  1 MPSQSGMNPA DLSGLSGKEL ARARRAALSK QGKAAVSNKT ASVNRSTKQA
 51 ASSINTNQVR SSVNEVPTDY QMADQLCSTI DHADFGTESN RVRDLCRQRR
101 EALSTIGKKA AKTTGKPSGR VRPQQSVVHN DAMIENAGDT NQSSSTSLNN
151 ELSEICSIAD DMPERFGSQA KTVRDICRAR RQALSERGTR AVPPKPQSQG
201 GPGRNGYQID GYLDTALHGR DAAKRHREML CQYGRGTAPS CKPTGRVKNS
251 VQSGNAAPKK VETGHTLSGG SVTGTQVDRK SHVTGNEPGT CRAVTGTEYV
301 GTEQFTSFCN TSPKPNATKV NVTTTARGRP VSGTEVSRTE KVTGNESGVC
351 RNVTGTEYMS NEAHFSLCGT AAKPSQADKV MFGATARTHQ VVSGSDEFRP
401 SSVTGNESGA KRTITGSQYA DEGLARLTIN GAPAKVARTH TFAGSDVTGT
451 EIGRSTRVTG DESGSCRSIS GTEYLSNEQF QSFCDTKPQR SPFKVGQDRT
501 NKGQSVTGNL VDRSELVTGN EPGSCSRVTG SQYGQSKICG GGVGKVRSMR
551 TLRGTSVSGQ QLDHAPKMSG DERGGCMPVT GNEYYGREHF EPFCTSTPEP
601 EAQSTEQSLT CEGQIISGTS VDASDLVTGN EIGEQQLISG DAYVGAQQTG
651 CLPTSPRFNQ TGNVQSMGFK NTNQPEQNFA PGEVMPTDFS IQTPARSAQN
701 RITGNDIAPS GRITGPGMLA TGLITGTPEF RHAARELVGS PQPMAMAMAN
751 RNKAAQAPVV QPEVVATQEK PELVCAPRSD QMDRVSGEGK ERCHITGDDW
801 SVNKHITGTA GQWASGRNPS MRGNARVVET SAFANRNVPK PEKPGSKITG
851 SSGNDTQGSL ITYSGGARG
```

Figure 27. Mapping of CsoS2A and CsoS2B derived peptides by MALDI-ToF mass spectrometry.

The polypeptides of purified carboxysomes were separated by the SDS-PAGE. The CsoS2A and CsoS2B bands were cut out of the gel and digested with trypsin. Peptides marked in red were identified in the digested protein mixtures by MALDI-ToF spectrometry. A: CsoS2A and B: CsoS2B. CsoS2B protein was traced to the arginine residue next to the last amino acid of the CsoS2 protein sequence deduced from the nucleotide sequence of its gene, which indicated that it is a full length CsoS2 protein.

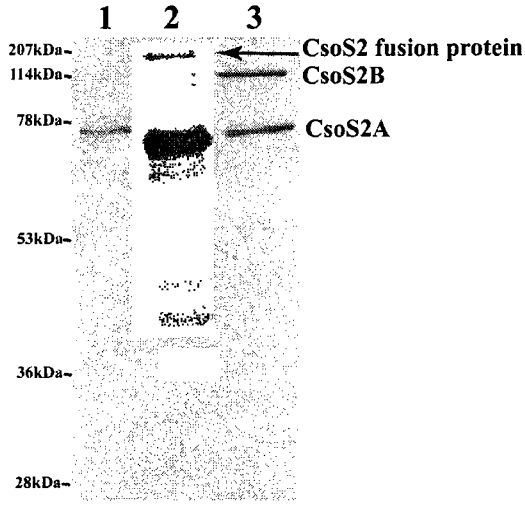


Figure 28. Immunoblot analysis of CsoS2 protein expressed in the IMPACT® system.

Uninduced and induced crude cell extracts were separated by SDS-PAGE and probed with polyclonal anti-CsoS2 antibody. In the induced *E. coli* cells, most of the recombinant CsoS2 protein was present in the CsoS2A form, and only a small amount of CsoS2B fusion protein was expressed. Because the CsoS2A protein does not carry the chitin binding tag, most of the recombinant CsoS2 protein does not bind to the affinity column, resulting in a low yield of CsoS2B form only (Figure 21). In the induced cells, several polypeptides recognized by anti-CsoS2 antibody may be CsoS2 protein degradation products or *E. coli* proteins that cross-react with the anti-CsoS2 antibody. Lane 1. Uninduced crude cell extract; Lane 2. Induced crude cell extract; Lane 3. Purified wild type carboxysomes.

Determination of the spatial location of CsoS2 protein in carboxysomes

In previous studies, CsoS2 protein was localized to the periphery of carboxysomes by immunoelectron microscopy [13]. Yeast two-hybrid analysis indicated that CsoS2 protein forms strong interactions with carboxysomal shell components CsoS1C and CsoS4B [127]. After the freeze-thaw treatment that breaks carboxysomes, CsoS2 protein is found to be strongly associated with the carboxysomal shell [114, 127]. The small SPA affinity tag added at the C-terminus of CsoS2 protein should be helpful in determining the spatial location of the C- terminus of full length CsoS2B protein in the carboxysome shell.

Purified wild type and HnSPAS2 mutant carboxysomes were mixed with agarose beads conjugated with anti-FLAG antibodies. Bound carboxysomes were eluted with 100 mM glycine buffer, pH 2.5. No detectable amount of wild type carboxysomes was recovered. HnSPAS2 mutant carboxysomes, on the other hand,
were retained on the column (Figure 29). Densitometric analysis revealed that the mass ratio of polypeptides in the eluted HnSPAS2 carboxysomes is 2: 1.5: 1: 11: 1.5: 7 for CsoS2B: CsoS2A: CsoS3: RuBisCO: CsoS1B: CsoS1A/C. These values are similar to those of purified intact HnSPAS2 carboxysomes (CsoS2B: CsoS2A: CsoS3: RuBisCO: CsoS1B: CsoS1A/C = 2.5: 2.0: 1: 13: 1: 6), suggesting that the trapped carboxysomes were likely intact.

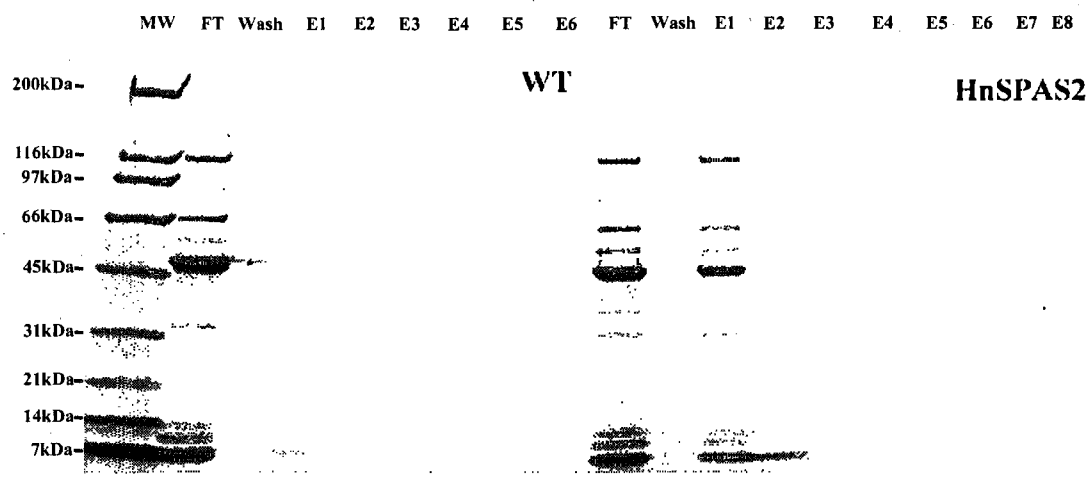


Figure 29. Trapping of SPA-tagged carboxysomes on an anti-FLAG affinity column.

Isolated wild type and HnSPAS2 mutant carboxysomes were mixed with anti-FLAG affinity beads. The trapped carboxysomes were eluted with 100 mM glycine buffer, pH 2.5.
MW: protein molecular weight standards; FT: flow-through fraction; Wash: last wash fraction; E: elution fraction. Purified HnSPAS2 carboxysomes bound to the affinity column, while wild type carboxysomes had no interaction with the resin, which suggested that the C-terminus of CsoS2 protein may be exposed to the outer surface of the carboxysome shell. There were extremely low amounts of polypeptide with an apparent molecular weight of approximately 7 kDa in the first elution fraction for wild type carboxysomes, which may represent the nonspecific binding of broken carboxysome shell component with the affinity column.

Assembly Studies of *H. neapolitanus* Carboxysomes

In previous studies, it was reported that the *H. neapolitanus* carboxysome shell can be disrupted into CsoS1A hexamers in the presence of 3.5 M urea, while the RuBisCO remains assembled as holoenzyme molecules [75, 127]. To determine whether during the removal of urea, the carboxysomal components interact with each other to re-assemble the carboxysomal shell, the SPA affinity tag at the C-terminus of CsoS2 protein was used to trap assembly intermediates.

In vitro assembly of the carboxysomal shell

Wild type and HnSPAS2 mutant carboxysomes were treated with 3.5 M urea. The turbidity of the sample significantly decreased, suggesting that the majority of carboxysomes was broken. Unbroken carboxysomes were pelleted by centrifugation at 21,000 x *g* for 30 minutes. The resulting supernatant was dialyzed against TEMB buffer without urea to permit protein complexes to re-form. After 12 hours of dialysis, the dialysate was mixed with anti-FLAG affinity beads to trap those protein complexes that contained accessible 3x FLAG epitopes. No protein assemblies from wild type carboxysomes were retained on the column. The trapped complexes derived from re-assembly of HnSPAS2 carboxysome proteins were composed of all carboxysomal proteins. Densitometric analysis showed that fewer RuBisCO polypeptides were present in these complexes (approximately 24% of the total protein in mass) comparable to intact HnSPAS2 mutant carboxysomes, in which RuBisCO represents approximately 55% of the total protein in mass (Figure 30). Transmission electron microscopy analysis revealed that the trapped protein complexes resembled the hexagonal shape of negatively stained

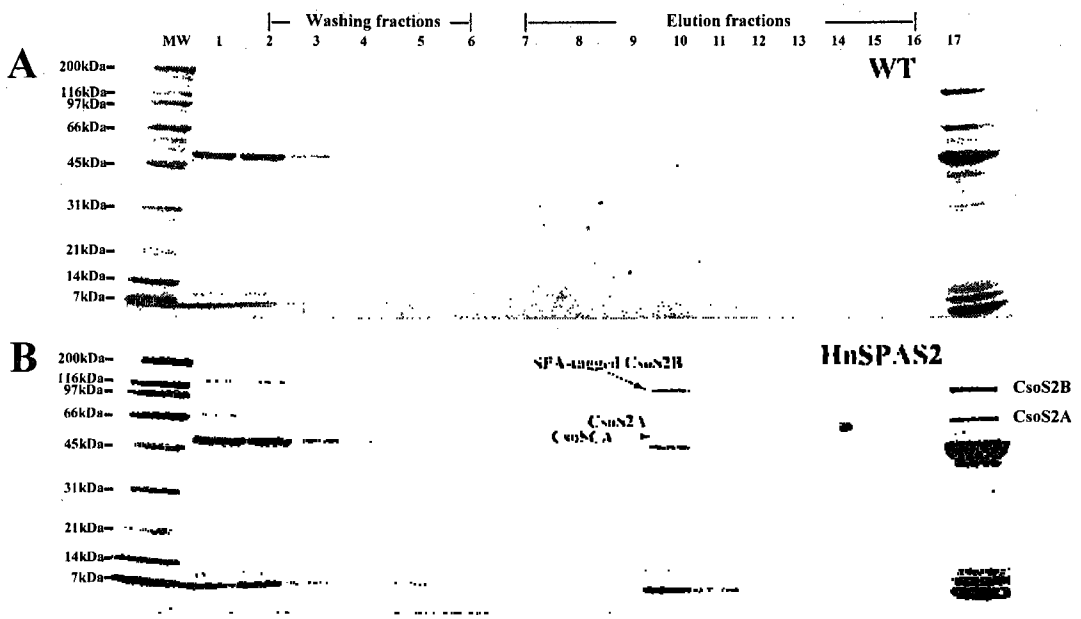


Figure 30. Affinity trapping of SPA-tagged protein complexes during *in vitro* re-assembly of HnSPAS2 mutant carboxysomal shell structures.

Isolated wild type and HnSPAS2 mutant carboxysomes were treated with 3.5 M urea to break the carboxysome shell. Dialysis was performed under non-denaturing conditions, followed by trapping of the re-formed protein complexes on an anti-FLAG affinity column. The trapped protein complexes derived from HnSPAS2 mutant carboxysomes were composed of all carboxysomal proteins. CsoSCA protein showed an extremely faint band in the trapped complexes due to its low abundance. Densitometric analysis showed that fewer RuBisCO polypeptides were present than in intact HnSPAS2 mutant carboxysomes.

Panel A: wild type carboxysomes; Panel B: HnSPAS2 mutant carboxysomes. MW: protein molecular weight standards; Lane 1: Flow-through of column; Lanes 2 to 6: washing fraction 1 to 5; Lanes 7 to 16: Elution fractions 1 to 10; Lane 17: purified wild type carboxysomes.

carboxysomes with diameters between approximately 100 to 120 nm (Figure 31). However, unlike purified carboxysomes, RuBisCO enzymes were not observed to fill the interior of these structures. These results suggested that the carboxysomal shell can self-assemble *in vitro* without any input of energy. Furthermore, the *in vitro* assembly of structures resembling carboxysomal shells does not depend on complete filling with RuBisCO holoenzyme molecules.

To date, *in vitro* carboxysome assembly from individual recombinant carboxysomal proteins has not been achieved, although all *H. neapolitanus* carboxysomal shell proteins have been expressed in *E. coli*. Previously, expression in *E. coli* of the entire *cso* operon under the control of its own promoter and in the presence of co-expressed chaperonin molecules resulted in carboxysome-like structures that were observed in the transmission electron microscope in thin sections of the bacteria [1]. However, the carboxysome-like structures have not been purified so far. The assembly of carboxysomes of *H. neapolitanus* in *E. coli* would be helpful in advancing our understanding of the biogenesis of carboxysomes *in vivo* and could provide guidance for the *in vitro* assembly of carboxysomes from individual recombinant proteins.

Expression of the cso operon in E. coli

A clone (pTn1) of the *cso* operon, including its promoter region, in a pET-derived vector, was a gift from Dr. Jessup Shively's lab. Given the complicated structure of Form I RuBisCO holoenzyme, which is composed of eight large and eight small subunits, high levels of chaperonin molecules in the cytoplasm of the host bacterium should help RuBisCO and the carboxysomal shell proteins fold and associate correctly to facilitate the

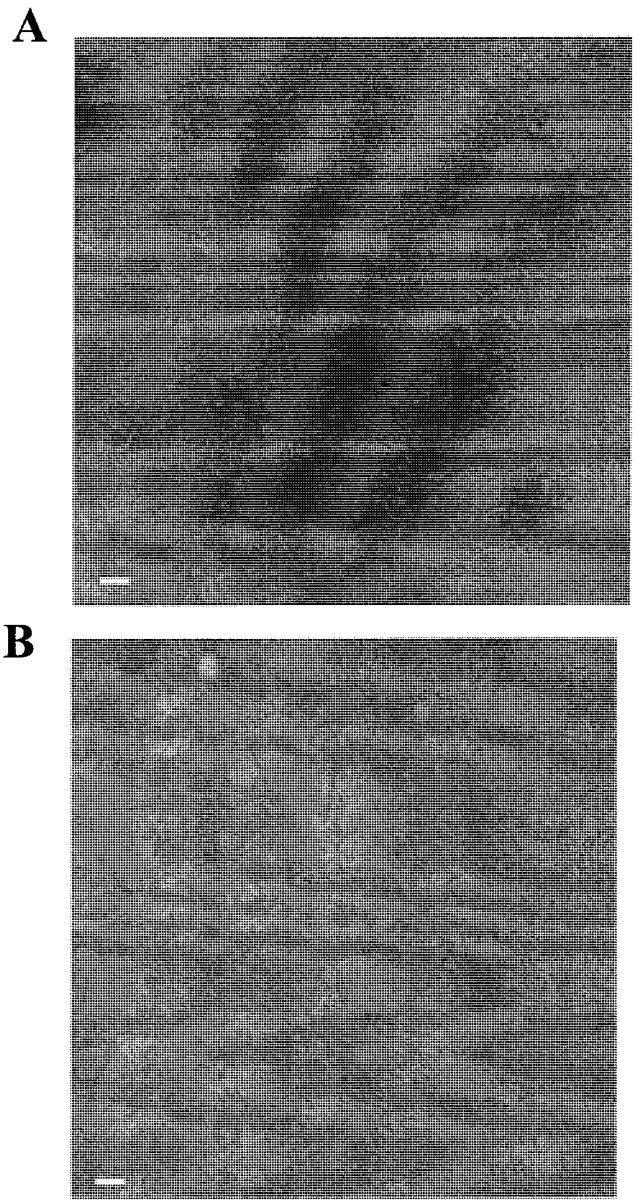


Figure 31. Transmission electron microscopy analysis of trapped intermediates formed during *in vitro* re-assembly of HnSPAS2 mutant carboxysomal shells.

Panel A: The trapped protein complexes of *in vitro* re-assembled HnSPAS2 mutant carboxysomal shells were eluted from the anti-FLAG affinity column with 200 mM glycine, pH 2.5. Before transmission electron microscopy, the fraction was neutralized with 200 mM NaOH and negatively stained with 1% (w/v) ammonium molybdate. Some trapped intermediates had a regular hexagonal structure and were 100 to 120 nm in diameter. Most of the protein complexes were aggregated, which may have been caused by denaturation due to the low pH.

Panel B: In the control experiments using wild type carboxysomes, no shell-like structures were observed. Bar: 100 nm.

assembly of carboxysomes [70]. An *E. coli* plasmid carrying *groESL* was therefore co-transformed into the *E. coli* BL21 host strain harboring pTn1. Cells were grown at 30°C to mid-exponential phase and induced with 1 mM IPTG for 24 hours at 16°C to promote a slower assembly of carboxysomes at the lower temperature. After 24 hours of induction, the culture was harvested, the cells were disrupted and subjected to a routine carboxysome purification procedure. Each sucrose gradient fraction was subjected to SDS-PAGE. Immunoblots were probed with anti-CsoS1A/C and anti-CbbL antibodies. The fraction from the middle of the sucrose gradient (fraction 5) showed the strong signals for CsoS1A/C and CbbL proteins (Figure 32). Most of CsoS1A/C proteins appeared to be in the bottom fraction, which indicated that CsoS1A/C proteins may not be efficiently utilized to participate in the self-assembly of the carboxysomes in *E. coli.* Given that wild type carboxysomes migrate to the middle of sucrose gradient, this result suggested that carboxysomes or carboxysome-like structures may have assembled in *E. coli.*

Characterization of carboxysome-like structures isolated from E. coli

To test whether the carboxysome-like structures assembled in *E. coli* are functional, RuBisCO CO_2 fixation assays were performed on a series of smaller fractions from the middle of a sucrose gradient. Fraction 13 had maximal CO_2 fixation activity, which indicated that the carboxysome-like structures formed in *E. coli* may be functional (Figure 33). When an aliquot of this fraction was stained with 1% (w/v) ammonium molybdate and observed under the transmission electron microscope, some regular hexagonal structures resembling carboxysomes were detected. In addition, some

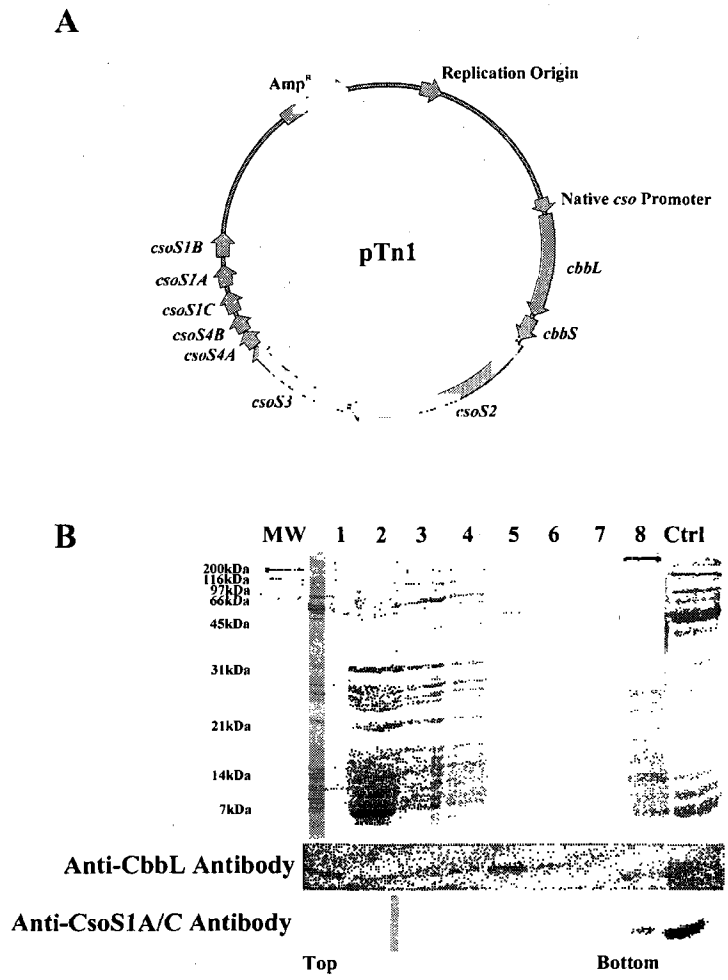


Figure 32. Expression of the entire *H. neapolitanus cso* operon in *E. coli*.

Panel A: The pTn1 clone of the *cso* operon in a pET-derived vector (a gift from Dr. Jessup Shively's lab). The expression of the *cso* operon is controlled by its own promoter. Panel B: Purification of carboxysome-like structures on a 10-60% sucrose gradient. A carboxysome-enriched pellet was loaded onto the gradient and centrifuged at 100,000 x *g* for 30 minutes. The gradient was fractionated into eight fractions from the top to the bottom. All fractions were loaded onto a 12% SDS-polyacrylamide gel. Blots were probed with anti-CbbL and anti-CsoS1A/C antibodies, respectively, to detect the distribution of RuBisCO enzyme and carboxysomal shell proteins in the gradient fractions. The fraction from the middle of the sucrose gradient (fraction 5) showed the strong signals for CsoS1A/C and CbbL proteins, suggesting carboxysome-like structures assembled in *E. coli* migrated similarly to wild type carboxysomes. Most of CsoS1A/C protein existed in the bottom fraction, indicating that the self-assembly of the carboxysomes is inefficient in *E. coli*.
MW: protein molecular weight standards; Lanes 1 to 8: Fractions 1 to 8 from the top to the bottom of a 10-60% sucrose gradient; Ctrl: purified wild type carboxysomes.

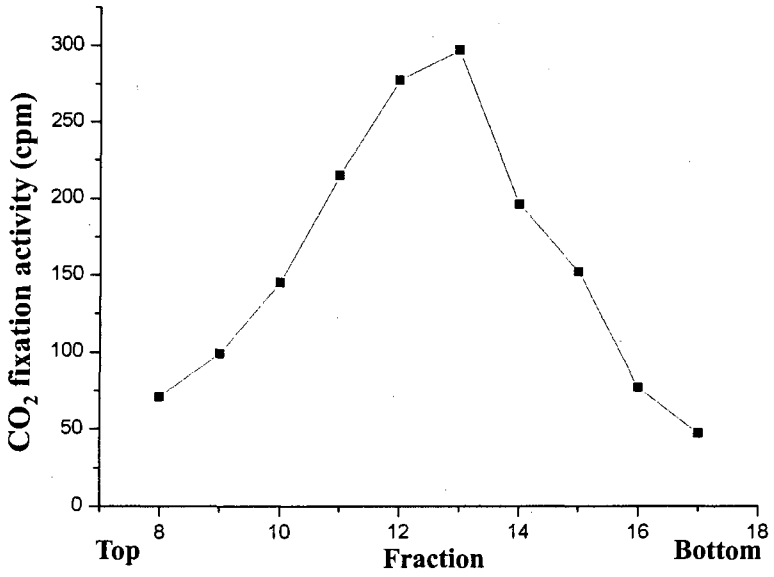


Figure 33. Carboxysome-like structures assembled in *E. coli* have CO_2 fixation activity.

Twenty-four fractions were collected from a 16 ml 10-60% sucrose gradient. RuBisCO CO_2 fixation assays were conducted with fractions 8 to 17. The fraction from the middle of the gradient (Fraction 13) had maximal CO_2 fixation activity.

larger, smaller and elongated carboxysome-like structures also were observed in this fraction. Interestingly, a lot of filaments co-purified with the regular hexagonal structures resembling carboxysomes (Figure 34). These results suggested that functional carboxysome-like structures can be assembled in the heterotrophic bacterium *E. coli*, but the self-assembly of the microcompartments is less efficient than it is in *H. neapolitanus*. Although the own promoter of the *cso* operon should be recognized by *E. coli*, it is not expected to yield high levels of transcription upon induction by IPTG [24]. Expression of the *cso* operon under the control of a strong IPTG-inducible promoter might help to produce larger amounts of recombinant carboxysomes or carboxysome-like structures in *E. coli*. Homogenous preparations of recombinant α-carboxysomes or α-carboxysome-like structures assembled in *E. coli* could facilitate self-assembly studies of β-carboxysomes in *E. coli* via an artificial gene transfer and enhance our understanding of protein-protein interactions and the three dimensional architecture of β-carboxysomes that are very difficult to purify to homogeneity in cyanobacteria due to the high content of internal membranes in the cells [28].

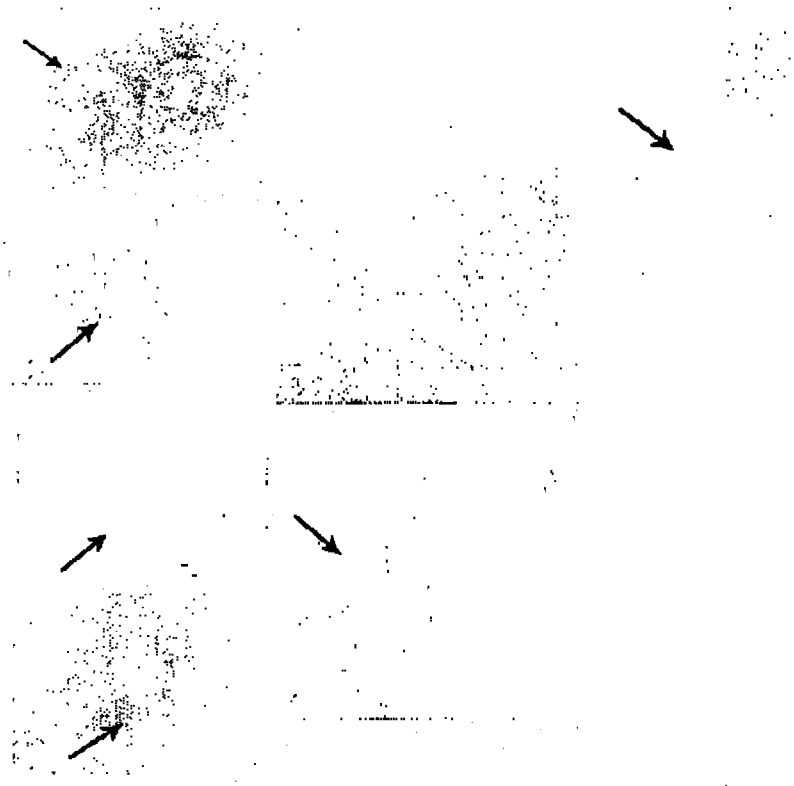


Figure 34. Transmission electron microscopy of carboxysome-like structures assembled in *E. coli.*

Fraction 13 with maximal CO_2 fixation activity from the middle of a sucrose gradient was negatively stained with 1% (w/v) ammonium molybdate before being observed under the transmission electron microscope. Carboxysome-like structures of the irregular shape and size (arrows) were observed (Panels C, D and F). Co-purifying filaments were observed in this fraction as well (Panel E). These results suggested that although carboxysome-like structures can be assembled in the heterotrophic bacterium *E. coli*, the self-assembly of the microcompartments is more error-prone than it is in *H. neapolitanus*. Bar: 100 nm.

CHAPTER V

DISCUSSION

Carboxysomes Are Essential in the CO_2-Concentrating Mechanism

The concentration of CO_2 is very low, approximately 400 ppm, in the environment surrounding autotrophic bacteria [47]. To efficiently utilize the available inorganic carbon, many photo- and chemoautotrophic bacteria adopt a CCM. In the bacterial cytoplasm, the inorganic carbon species are stored as bicarbonate ions. The lipid membrane has nearly 1,000 fold less permeability for the negatively charged ions than for the uncharged CO_2 molecules [23, 40, 49]. However, the substrate of RuBisCO enzyme is CO_2, not bicarbonate, so the bicarbonate needs to be dehydrated to CO_2 before it can be utilized by RuBisCO [50, 114]. Although this reaction proceeds uncatalyzed, the lack of a carboxysomal carbonic anhydrase in *H. neapolitanus* and *Synechocystis* PCC 6803 mutant results in a high CO_2-requiring (HCR) phenotype [41, 113], which indicates that the spontaneous conversion rate of bicarbonate to CO_2 is not fast enough to support the normal growth of these bacteria and that the carboxysomal carbonic anhydrase is very important in supporting growth at ambient CO_2 levels.

Transmission electron microscopy revealed that the carboxysomes assembled in *H. neapolitanus csoS3::Km* and in *Synechocystis* PCC 6803 *ccaA::Km* insertion mutants had the very similar size and shape to their wild type counterparts (Figure 7A and B, [41, 113]). The abundance of RuBisCO in the *csoS3::Km* mutant carboxysomes is almost the same as that in wild type (Figure 8). However, the CO_2 fixation activity of the *H. neapolitanus* mutant carboxysomes was compromised (Figure 12A). Using the V_{max} and K_c values of CO_2 fixation by wild type and mutant carboxysomes, the *in vivo* CO_2

fixation rates in air can be calculated according to the Michaelis-Menten equation. The concentration of the cytosolic C_i-pool is estimated to be approximately 10 mM in *Thiomicrospira crunogena* [40]. In the *Synechocystis* PCC 6803 mutant lacking carboxysomal carbonic anhydrase, the C_i-pool of the mutant is comparable in size to that of wild type cells [113]. In the *H. neapolitanus csoS3::Km* mutant, the inorganic carbon transport systems were not manipulated, so the mutant also should have a similar C_i-pool as wild type. The cytosolic pH of *T. crunogena* and *H. neapolitanus* is estimated at approximately 7.4 ([40] and G. C. Cannon, personal communication). According to the Henderson-Hasselbalch equation, at pH 7.4 the concentration of CO_2 dissolved in the cytoplasm is 560 μM (calculated from 10 mM bicarbonate at pH 7.4). The CO_2 fixation rate of wild type carboxysomes is approximately 1.3 $\mu mol \cdot min^{-1} \cdot mg^{-1}$ protein, which is nearly two fold higher than that of mutant carboxysomes (approximately 0.68 $\mu mol \cdot min^{-1} \cdot mg^{-1}$ protein) and indicates that the mutant cells fix carbon less efficiently than wild type cells in air (Figure 6), because their carboxysomes can not efficiently utilize the cytosolic C_i-pool. In 5% CO_2-enriched air, the ratio of soluble RuBisCO to carboxysome-encapsulated RuBisCO increases [25], therefore, the contribution of carboxysomal RuBisCO to inorganic carbon fixation in wild type and *csoS3::Km* mutant cells becomes less. The difference in efficiency of carbon fixation between mutant and wild type disappears (Figure 6).

In previous studies, the *cbbL* and *csoS1A* genes located in the *cso* operon were disrupted using the same cloning strategy, and the resulting *cbbL::Km* and *csoS1A::Km* mutants also showed high CO_2-requiring phenotype [12, 46]. "Empty" carboxysome structures were observed in the *cbbL::Km* mutant, and some irregular carboxysomes are

formed in the *csoS1A::Km* mutant. In some cyanobacterial high CO_2-requiring mutants, inorganic carbon species can be taken up normally and the accumulated bicarbonate ions are maintained in the cytosol at high concentration. However, the C_i-pool cannot be utilized efficiently in these mutants [84, 92]. These mutants contain morphologically altered carboxysomes or lack carboxysomes [84, 92]. The *csoS3::Km* mutant carboxysomes showed the similar size and shape, especially the same abundance of RuBisCO enzyme in total carboxysomal proteins as wild type, which facilitated the understanding of the effect of the missing carboxysomal carbonic anhydrase on CO_2 fixation of the packaged RuBisCO by CO_2 fixation kinetics and the role of carboxysomes in carbon metabolism in the bacteria quantitatively. Obviously, functional carboxysomes act as the terminal component of the CO_2-concentrating mechanism and play a very important role in efficient utilization of the accumulated cytosolic C_i-pool, therefore helping the autotrophic bacteria grow well in the low inorganic carbon levels in their environment.

The Carboxysomal Shell Is A Diffusion Barrier for Gas Molecules

Carboxysomal carbonic anhydrase is associated with the carboxysomal shell

The compromised CO_2 fixation activity of carboxysomes lacking carboxysomal carbonic anhydrase indicates that the encapsulated RuBisCO enzyme can not obtain sufficient CO_2 molecules to maintain its CO_2 fixation activity in the absence of carboxysomal carbonic anhydrase. Previously, expression of human carbonic anhydrase in the cytoplasm of the cyanobacterium *Synechococcus* PCC 7942 led to a high CO_2-requiring phenotype of the resulting mutant [88]. This result indicated that the carbonic

anhydrase can not exist in the cytoplasm of the cells. The presence of carbonic anhydrase can accelerate the conversion of cytosolic bicarbonate to CO_2 in the cytoplasm, and facilitate the diffusion of uncharged CO_2 molecules across the lipid plasma membrane, which will result in the loss of the cytosolic C_i-pool [88]. Subsequently, a quantitative mathematical model speculated that the carbonic anhydrase is located within the carboxysomes [50].

In 2000, Baker *et al.* revealed the CsoSCA protein is localized at the periphery of the *H. neapolitanus* carboxysomes by immunoelectron microscopy [14]. In addition, after the freeze-thaw treatment of carboxysomes, the CsoSCA protein was identified to be strongly associated with the carboxysome shell [57].Our experimental results revealed that the α-carboxysomal shell lacking carboxysomal carbonic anhydrase limits the diffusion of CO_2 and/or HCO_3^- across the carboxysome boundary and the carboxysomal shell-bound carbonic anhydrase increases the permeability of CO_2 for the shell (Figure 12A). In β-carboxysomes, a putative carboxysomal shell protein CcmK also forms a hexameric structure and could act as the major components of the shell [68]. But the location of the potential carboxysomal carbonic anhydrase is still unknown. If the β-carboxysome shell also shows a diffusion barrier for CO_2 molecules, the carboxysomal carbonic anhydrase can not be packaged into the interior of the carboxysomes as that quantitative mathematical model suggested [50]; otherwise it does not make the contribution to help the CO_2 molecules overcome the diffusion barrier.

The exogenously added recombinant CsoSCA protein does not complement the missing carbonic anhydrase function for the mutant carboxysomes *in vitro* (Figure 11), indicating that the carboxysomal carbonic anhydrase likely faces the interior of the

carboxysomes. This packing strategy of the carboxysomal carbonic anhydrase in the carboxysomal shell facilitates the diffusion CO_2 molecules across the shell by converting cytosolic bicarbonate to CO_2 to supply sufficient substrates to RuBisCO, but also prevents access to the cytosolic bicarbonate directly, thereby retaining the cytosolic C_i-pool. In addition, the carboxysomal shell prevents leakage of the trapped CO_2 molecules out of the carboxysomes and helps the encapsulated RuBisCO enzyme utilize the inorganic carbon accumulated in the bacteria more efficiently [8, 9, 95].

Transport of CO_2 and/or bicarbonate across the shell

The exact molecular mechanism by which CO_2 molecules cross the carboxysomal shell is still unknown. Cryoelectron tomography studies of the carboxysomes of *Synechococcus* WH 8102 indicated that the protein-protein contacts in the shell are likely quite flexible and may allow the passage of CO_2 and/or HCO_3^- across the shell [61]. Our experimental results revealed that exogenously added rCsoSCA protein does not change the kinetic parameters of CO_2 fixation by *csoS3::Km* mutant carboxysomes (Figure 11), which contradicts that proposed model.

Recent investigations have revealed that the carboxysomal carbonic anhydrase has a strong interaction with the carboxysome shell after a freeze-thaw treatment [114]. X-ray structural studies have revealed that CsoS1A, one of the three major carboxysomal shell components, forms hexamers [125]. Interestingly, there is a central, positively charged pore in each CsoS1A hexamer [125]. Co-crystallization experiments found that a sulfate ion can be trapped in the pore [125]. Given a similar size and a charge of the same sign of sulfate and bicarbonate ions, Tsai *et al.* speculated that the positively charged

pores can pull the cytoplasmic bicarbonate through them [125]. If CsoSCA is located under the pore, the enzyme could rapidly convert bicarbonate to CO_2 and release it to the sequestered RuBisCO as the substrate (Figure 35, [41]). According to the atomic level model of the carboxysome shell, there are 740 CsoS1A hexamers in one carboxysome of *H. neapolitanus* [124]. By contrast, there are approximately 40 CsoSCA dimers per carboxysomes [56]. Consequently, only approximately 5% of CsoS1A hexamers would be able to interact with CsoSCA.

Previous yeast two-hybrid results revealed that CsoSCA protein interacts only with CsoS1B, not with CsoS1A or CsoS1C [127]. This is surprising, since CsoS1A and CsoS1C differ only by two amino acids [127]. Therefore, CsoSCA protein may associate with the carboxysome shell via interaction with the unique C-terminal tail of CsoS1B [30]. If it does so, a CsoS1B C-terminal deletion that eliminates the extension should disrupt the interaction between CsoSCA proteins with the carboxysome shell and result in a damage in the location of CsoSCA protein in the carboxysomal shell. The resulting mutant carboxysomes should be compromised in their CO_2 fixation efficiency and show a higher K_c value of CO_2 fixation than the wild type.

Is CO_2 or HCO_3^- the preferred C_i species for the carboxysome?

Given that the carboxysomal shell is a diffusion barrier for CO_2 molecules and carboxysomal carbonic anhydrase convert bicarbonate to CO_2 as the substrate of the packaged RuBisCO, the carboxysomes may utilize bicarbonate more efficiently than CO_2. In *in vitro* CO_2 fixation assay, as the pH increases, the bicarbonate becomes more abundant. For example, when the pH increases from 8.0 to 9.0, the concentration of

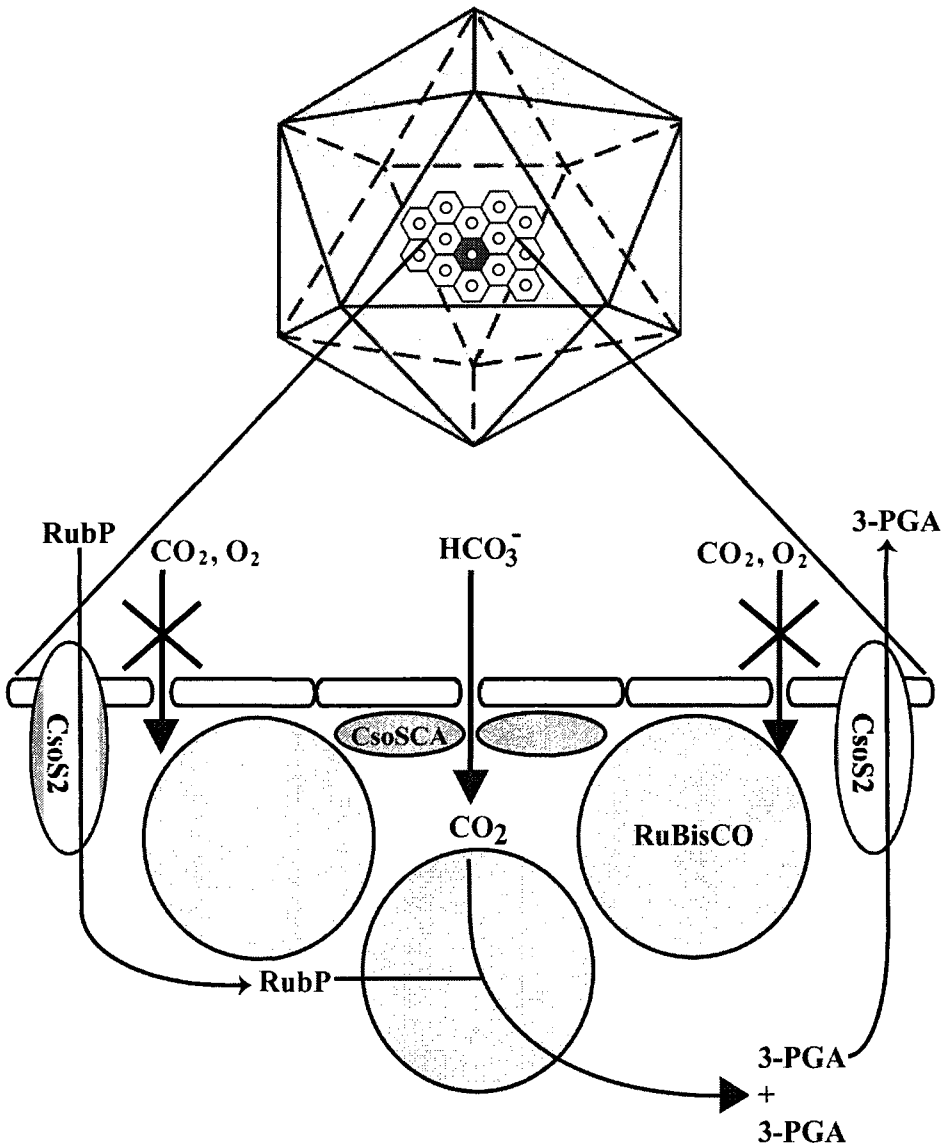


Figure 35. Working model for the location and function of CsoSCA and CsoS2 in the carboxysome shell. [41]

The carboxysomal shell is a diffusion barrier for CO_2 and O_2 molecules. The negatively charged bicarbonate ion could be pulled through the positively charged pore of the CsoS1 hexamers. The CsoSCA protein located underneath the pore catalyzes the conversion of bicarbonate to CO_2. The CsoS2 protein interacts with the carboxysomal shell and may play a role in the transportation of negatively charged larger molecules, such as RuBP and/or PGA across the carboxysome shell.

bicarbonate will increase ten fold. Therefore, CO_2 fixation activity of the carboxysomal RuBisCO should decrease less than that of free RuBisCO with increasing pH. In addition, the pH affects the activation rate of RuBisCO enzyme [81]. RuBisCO enzyme needs to be activated prior to performing its CO_2 fixation [53]. Activation process involves the reaction of a CO_2 molecule with the ε-amino group of the lysine in the active site of the large subunit to form a carbamate [53]. The magnesium ion binds to the carbamate to stabilize the carbamylated lysine [53]. The carbamylated lysine can catalyze the carboxylation of RuBP to produce two PGA molecules [53]. The formation of carbamate needs an uncharged amino group [76], therefore, as the pH increases, more of uncharged amino groups present [76]. Thus, activation of RuBisCO is favored at high pH [76, 81]. Previous studies revealed that increasing the pH decreases the apparent $K_{M(CO2)}$ [77, 119], that is consistent with this idea. In the *H. neapolitanus* cytoplasm, the pH is proximately 7.4. If the pH inside carboxysomes is higher than outside, the encapsulated RuBisCO is kept an easily activated state.

However, the plot of CO_2 fixation activity of carboxysomal and free RuBisCO over pH displayed a similar pH profile between carboxysomal and free RuBisCO (Figure 13B). Similarly, the carboxysomal carbonic anhydrase and recombinant carbonic anhydrase showed a very similar pH profile [57]. The similar pH profile between carboxysomal and free RuBisCO indicates that the pH inside carboxysomes is likely almost the same as that in the surrounding environment. The bicarbonate dehydration reaction catalyzed by carbonic anhydrase is an extremely rapid process [57]. If the pH between the two sides of the carboxysome shell is the same, it could result in the same concentration of CO_2 inside and outside the carboxysomes, therefore, no difference of

CO_2 fixation rates between carboxysomal and free RuBisCO is observed (Figure 13). The almost same K_c values of CO_2 fixation between carboxysomal RuBisCO and free RuBisCO did not also support that the encapsulated RuBisCO is kept in an easily activated state.

To determine carboxysomal RuBisCO utilizes bicarbonate more efficiently than CO_2, CO_2 fixation assay of the carboxysomal RuBisCO can be performed under $H[^{14}C]O_3^-$ or $H[^{14}C]O_3^-$ plus the identical concentration of HCl (for the CO_2 species) at low temperature (for example 5 °C) [36, 76]. Under that temperature, the equilibration of bicarbonate and CO_2 is in a slow equilibrium [36, 76]. Thus, if the carboxysomes utilize bicarbonate more efficiently than CO_2, the carboxysomes will synthesize more radioactive PGA using $^{14}HCO_3^-$ than $^{14}CO_2$. In the control experiments, free RuBisCO is treated with the same procedure and it should synthesize more radioactive PGA using $^{14}CO_2$ than $^{14}HCO_3^-$ because CO_2 is well known as the substrate of RuBisCO enzyme.

The carboxysomal shell also acts as a diffusion barrier for O_2

Given that the carboxysome shell is a diffusion barrier for CO_2 molecules, it could also limit the passage of O_2 into the interior of the carboxysomes. The CsoSCA protein facilitates the diffusion of CO_2 into the interior of the carboxysomes. Therefore, its discrimination between CO_2 and O_2 molecules may create a relatively higher $[CO_2]/[O_2]$ microenvironment inside the carboxysome than outside, therefore leading to an enhancement of the carboxylation reaction at the expense of the oxygenation reaction catalyzed by the encapsulated RuBisCO. The high resolution isotope-based simultaneous assay and classical inhibition assays were used to estimate the specificity factors and O_2

inhibition constants for carboxysomal and free RuBisCO enzyme, respectively. The specificity factor parameter represents the ratio of carboxylase to oxygenase activity of RuBisCO enzyme and reflects the substrate discrimination of RuBisCO between CO_2 and O_2 molecules. The O_2 inhibition constant reflects the sensitivity of the carboxylase activity to the competitive substrate O_2.

The specificity factor of wild type carboxysomes I measured in this study is very similar to that of free RuBisCO enzyme (Figure 16). The incubation time of specificity factor determination is 60 minutes because the extremely low amounts of 1-[^{3}H]RuBP is attempted to be completely converted to the products to increase the amount of radioactivity of the products and improve the resolution of PGA and PG peaks on the downstream anion exchange column [52, 65, 71]. The diffusion of gas molecules across the shell, however, is likely a very rapid event. During a 60-minute incubation, the concentration of O_2 would have reached equilibrium between the inside and the outside of the carboxysome, so that the effects of temporarily elevated $[CO_2]/[O_2]$ on RuBisCO catalysis would not have been detected over a long assay time.

In the following strategy, the incubation time was shortened to 3 minutes. In addition, the concentration of O_2 in the assay buffer was varied to create a concentration gradient of O_2 between the two sides of the carboxysome shell. Under the O_2 concentration in the range from 50 to 420 μM, carboxysomal RuBisCO produces 1.4 to 1.5 times more PGA per PG than free RuBisCO. However, the ratio of carboxylase to oxygenase activity between carboxysomal and free RuBisCO becomes close as the O_2 concentration increases to the range from 550 to 1050 μM (Figure 18 and Table 3). According to Fick's first law of diffusion, diffusion rate depends on the concentration

gradient over the distance [15]. As the concentration of O_2 increases, the diffusion rate becomes faster, the concentration of O_2 between the two sides of the carboxysome shell will be equilibrated quickly. Therefore, no significant difference of the ratio of carboxylase to oxygenase activity between carboxysomal and free RuBisCO was observed under the high concentration of O_2.

The V_{max} of *csoS3::Km* mutant carboxysomal RuBisCO is 2.2 μmol·min^{-1}·mg protein^{-1}(assumed that RuBisCO approximately accounts for 60% of total carboxysomal protein by weight), while the V_{max} of free RuBisCO purified from mutant carboxysomes is 3.0 μmol·min^{-1}·mg protein^{-1} (Table 2). According to Michaelis-Menten equation, at 50 μM CO_2, the ratio of CO_2 fixation between *H. neapolitanus csoS3::Km* carboxysomal and free RuBisCO is approximately 3.5, and it decreases to approximately 1.7 at 1050 μM CO_2. However, the ratio of oxygenase rates between carboxysomal and free RuBisCO is approximately 1.5 and 1.0 at 50 and 1050 μM O_2. This calculation shows that the carboxysomal shell does not show a strong diffusion barrier for O_2 molecules as for CO_2 molecules. CO_2 and O_2 molecules have a similar shape and electrostatic distribution, but the absolute electrostatic potential of CO_2 is ten fold higher than that of O_2 [67]. The diffusion of gas molecules across the shell may depend on the electrostatic interactions with the pores within the hexamers. Although both CO_2 and O_2 molecules do not carry the charge, the carbon in CO_2 molecule has a slight positive charge and each oxygen atom has a slight negative charge. Therefore, the negatively charged oxygen may form a hydrogen bond with the C-alpha hydrogen atom of the glycine residue surrounding the pores, but the positively charged basic amino acid residues could repulse the positively charged carbon atom. Thus, O_2 molecule could cross the central pore more

easily than CO_2 molecule, therefore, carboxysome shell displayed a stronger diffusion barrier for CO_2 than O_2.

The O_2 inhibition constant for carboxysomal RuBisCO was estimated to be almost three fold higher than for free RuBisCO using classic inhibition assays. The K_c values of carboxysomal RuBisCO increase less than those of free RuBisCO as the O_2 concentration increases (Figure19A and B). Similarly, in the blue-green alga *Coccochloris peniocystis*, the K_c values of a carboxysome-enriched fraction and of soluble RuBisCO were estimated in pure N_2, air and pure O_2. The K_c values of the carboxysome-enriched fraction increased much less than those of soluble RuBisCO, as the concentration of O_2 increased [35]. In addition, a carboxysome-enriched fraction from *Synechocystis* PCC 6803 showed similar results [79], which indicated that the carboxylase activity of carboxysomal RuBisCO is less sensitive to O_2 inhibition than that of free RuBisCO enzyme. It is worthy to emphasize that the carboxysome samples from the cyanobacteria are not homogeneous and are contaminated by internal membranes [28]. Some cytoplasmic RuBisCO molecules may nonspecifically bind to these membrane or some carboxysomes are surrounded with the contaminated membrane. If carboxysomes are surrounded with the membrane, the internal membrane will affect the access of CO_2 molecules to the encapsulated enzyme to change its CO_2 fixation kinetic parameters. Thus, the impurity of the samples could affect the outcome of RuBisCO CO_2 fixation assay.

For *H. neapolitanus* carboxysomal and free RuBisCO, the O_2 inhibition constant of carboxysomal RuBisCO is approximately two and a half times bigger than that of free RuBisCO (Figure 19). It indicates that the carboxysomal RuBisCO is less sensitive for O_2

inhibition than free RuBisCO. But the error bars of the K_c values for two RuBisCO forms estimated in different concentrations of O_2 overlapped, the difference of O_2 inhibition constants that are obtained by linear regression is not statistically significant. Therefore, a stopped flow technique that measures the enzymatic reaction over a short time frame may be adopted to measure the O_2 inhibition constants more accurately.

Is H. neapolitanus carboxysome shell necessary for a diffusion barrier for O_2?

The specificity factor of *H. neapolitanus* free RuBisCO enzyme is approximately 48.5. The cytosolic O_2 concentration is approximately equal to that in the growth medium when the cells are cultured in air (approximately 255 μM, according to Henry's law), and the cytosolic CO_2 concentration is 560 μM at pH 7.4. Thus, under physiological conditions, free RuBisCO synthesizes 107 PGA molecules for every PG molecule, according to the equation $v_c/v_o = SF \times [CO_2]/[O_2]$ (v_c: carboxylation rate; v_o: oxygenation rate). The specificity factor of wild type carboxysomes is approximately 49.3. Thus, the carboxysomal RuBisCO can synthesize 109 PGA molecules for every PG molecule, a value that is essentially identical to that calculated for the free enzyme. Such a large ratio of PGA to PG may indicate that it is not "necessary" for the carboxysomal shell to provide a strong diffusion barrier for O_2 molecules.

H. neapolitanus is a sulfur bacterium and does not produce oxygen like cyanobacteria. The concentration of O_2 in the cytoplasm of cyanobacteria should be much higher than in *H.* neapolitanus. However, when comparing the internal C_i-pool between *T. crunogena* and *Synechocystis* PCC 6803, the size of internal C_i-pool of *Synechocystis* PCC 6803 is approximately 4 to 5 fold higher than that in *T. crunogena* [40,

113]. Therefore, high concentrations of inorganic carbon in cyanobacteria may keep the $[CO_2]/[O_2]$ high to enhance the ratio of carboxylase to oxygenase activity. The specificity factors of RuBisCO from several cyanobacteria *Synechococcus*, *Aphanizomenon flos-aquae* and *Cocochloris peniocystis* are 42.5 ± 1.0, 48 ± 2 and 47 ± 2, respectively [64, 65], which are very similar to the specificity factor of *H. neapolitanus* RuBisCO measured in this study. However, carboxysomes from cyanobacteria have not been purified to homogeneity; therefore, the specificity factor of their carboxysomal RuBisCO can not be determined accurately and compared with that of free RuBisCO. In addition, little information is available regarding structure and composition of cyanobacterial carboxysomes. It is possible that the high concentration of inorganic carbon enhances the ratio of carboxylase to oxygenase activity of RuBisCO or the β-carboxysome shell has a unique mechanism of preventing O_2 from diffusing into the microcompartment interior.

The *pdu* and *eut* microcompartments participate in the catabolism of 1,2-propanediol and ethanolamine, respectively. Diol dehydratase and ethanolamine ammonia-lyase, the enzymes that are packaged inside the *pdu* and *eut* microcompartments, respectively, are sensitive to O_2. The shells of these microcompartments may also limit the diffusion of O_2 to protect the packaged enzymes and enhance their catalytic efficiencies [19, 20, 33].

CsoS2 Protein – A Potential Candidate to Transport RuBP and PGA

RuBP and PGA need to be transported into and out of carboxysomes

To maintain the carboxylation reaction proceeding continuously, a second substrate of RuBisCO, RuBP, and the carboxylation reaction product, PGA, need to cross

the carboxysome shell. To date, it is unknown how these negatively charged molecules are "transported" across the shell. The positively charged central pore is predicted to facilitate bicarbonate diffusion across the shell [125], because sulfate ions can be trapped in the pores of CsoS1A crystals [125]. RuBP and PGA also are negatively charged molecules, and they were not found within the pore of CsoS1A hexamer in the co-crystallization experiments [125]. Tsai *et al.* speculated that the pores within CsoS1A hexamers also facilitated the transportation of these larger metabolites across the shell [125]. However, RuBP molecule has two phosphate groups and a five-membered carbohydrate and PGA molecule has one phosphate group and a three-membered acid. These two molecules could be rod-like molecules. Therefore, they could not cross the shell as easily as bicarbonate ion. The unsuccessful co-crystallization of these two metabolites within the pores may also indicate that these two negatively charged molecules are relatively too large to accommodate within the pores. The CsoS2 protein, an abundant carboxysomal shell-associated protein [13, 127], has an unusually high pI point (Table 4), which indicates that CsoS2 protein is a potential candidate to transport RuBP and PGA into and out of the carboxysome [29].

Two forms of CsoS2 protein

In *H. neapolitanus* carboxysomes, CsoS2 protein is expressed as two distinct polypeptides of observed molecular weights 80 kDa and 130 kDa, respectively [13]. Two CsoS2 polypeptides have the similar epitope and the identical N-terminal 15 amino acids [13]. Two CsoS2 polypeptides was thought to have differential levels of glycosylation [13]. Expression of the *csoS2* gene in *E. coli* also resulted in two forms of CsoS2 protein

with the same observed molecular weights as those seen in *H. neapolitanus* carboxysomes (Figure 21 and 28). This finding makes differential glycosylation less likely because this modification of recombinant proteins overexpressed in *E. coli* is rarely seen.

Most of the recombinant N-terminally His-tagged CsoS2 protein was expressed in inclusion bodies [13]. Twenty cysteine residues are found in the deduced CsoS2 protein sequence; it might therefore be very difficult for the large nascent CsoS2 protein to pair so many disulfide bonds and reach its native tertiary structure in *E. coli*. Incorrect folding may cause the exposure of the hydrophobic patches at the surface of recombinant CsoS2 protein that lead to insoluble aggregates. So, the concentration of soluble CsoS2 protein is very low when using native method to purify rCsoS2 protein in *E. coli*. Also, when rCsoS2 protein was purified as the denatured form and was attempted to be re-folded to its native conformation *in vitro*, CsoS2 protein forms the white aggregates, which results in a low concentration of soluble rCsoS2.

To understand the difference between two CsoS2 forms, a small double affinity (SPA) tag sequence was added at the 3' end of the *csoS2* gene in *H. neapolitanus* by homologous recombination in this study. In addition, two CsoS2 proteins were analyzed with peptide mapping. These results revealed that CsoS2B protein is a full length CsoS2 protein, while CsoS2A protein lacks the C-terminal affinity tag and appears to be a C-terminally truncated CsoS2 protein (Figures 26 A, C and Figure 27).

It is unknown how the C-terminally truncated CsoS2A is formed. One possible explanation is that the translation of the *csoS2* transcript is terminated prematurely via ribosomal frameshifting event, which would result in the formation of a truncated protein.

For example, the τ and γ subunits of DNA polymerase III is encoded with a single gene in *E. coli* [73]. The γ subunit is about two-thirds of the τ subunit and shares all of its protein sequence with the τ subunit [73]. The ratio of these two subunits is approximately 1:1 [73]. Larsen *et al.* reported that the approximately 50% of initiating ribosomes translate the mRNA conventionally to synthesize the τ subunit and the other 50% shift into the -1 reading frame at a specific site in the mRNA to synthesize a truncated γ subunit [73]. In *H. neapolitanus*, the number of full length CsoS2B protein and the C-terminally truncated CsoS2A protein is approximately equal [56]. The C-terminal amino acid of CsoS2A was attempted be revealed by peptide mapping analysis. Three independent trypsin digestions of CsoS2A were analyzed by MALDI-ToF and resulted in the peptide coverage marked in red in Figure 27A. Given that the peptide VVETSAFANR at positions from 827 to 836 is far from the nearest upstream peptide MSGDER at positions from 568 to 573 and only appears in one digested sample, this peptide fragment may be derived from the contamination during the experimental manipulation. By translating the nucleotide sequence of the *csoS2* gene in the -1 and +1 reading frames, a potential stop codon (TGA) in the +1 reading frame starting at position 1826 (Figure 36) was found as the nearest one from the peptide MSGDER. The ribosome may shift the reading frame at this position and generate a C-terminally truncated CsoS2 protein. The peptide mapping analysis can not cover all amino acids in the digested polypeptide. There is only one arginine between the peptide MSGDER and the potential stop codon. Therefore, the peptide sequence between the peptide MSGDER and the potential stop codon could not be identified in three preparations of digested CsoS2A protein. More trials may be performed to increase the peptide coverage of the protein

1621 (540)

ggt ggc gtg gga aaa gtg cgc tca atg cgc acc ctt cgc ggc acc
G **G** **V** **G** **K** **V** **R** S M R **T** **L** **R** **G** **T**
tca gta tct ggc caa cag cta gat cat gcc cca aag atg tcc ggt
S **V** **S** **G** **Q** **Q** **L** **D** **H** **A** **P** **K** **M** **S** **G**
gac gag cgc ggc ggg tgc atg ccc gtc acc ggt aat gag tac tac
D **E** **R** G G C M P V T G N E Y Y
ggt cgt gaa cat ttc gaa ccg ttt tgt acg agc acc cca gag ccc
G R E H F E P F C T S T P E P
gaa gct caa tca act gaa caa tca t**TG** **A**cc tgt gaa gga caa att
E A Q S T E Q S L T C E G Q I

1845 (615)

Figure 36. The potential frame shifting stop codon in *H. neapolitanus csoS2* gene.

Partial nucleotide sequences of the *H. neapolitanus csoS2* gene (positions from 1621 to 1845) are shown in the figure. The numbers 1621 and 1845 are the beginning and ending numbers of the nucleotide fragment shown, and the numbers in parentheses are the corresponding numbers for the translated amino acids. Nucleotide sequences are translated into the corresponding amino acid in the regular reading frame. Peptides marked in red were identified by MALDI-ToF. A potential stop codon (TGA) in +1 reading frame starting at position 1826 is labeled in red. It is the nearest stop codon from the identified peptide MSGDER by MALDI-ToF. The ribosome may shift the reading frame at this position and generate a C-terminally truncated CsoS2 protein.

samples.

In *E. coli*, there is a potential slippery sequence AAAAAAG in the upstream of the stop codon for DNA replicase τ subunit [73]. The ribosome changes the reading frame at this specific site and generates a truncated protein. However, similar sequence can not be found upstream of the potential stop codon in *csoS2* gene. The *csoS2* gene may adopt a different strategy to switch the reading frame. In addition, 3' RNA pseudoknot also causes the different kinds of gene recoding events [16]. RNA pseudoknot is an RNA secondary structure that contains at least two stem-loop structures. The nucleotides in the loop can complement with a region of RNA downstream of this stem-loop [16]. The distance between the signal and frameshifting sequence is highly important. However, there is no available software for predicting the RNA pseudoknot pattern. It is unknown whether *csoS2* gene adopts this strategy to generate the C-terminally truncated CsoS2A protein.

Alternatively, a specific potential protease may exist in both *E. coli* and *H. neapolitanus* that cleaves the nascent full length CsoS2B protein into the shorter CsoS2A. Maybe due to the different abundance and activity of the protease in *E. coli* and *H. neapolitanus*, more CsoS2B protein is processed into CsoS2A in *E. coli* than in *H. neapolitanus*.

In *H. neapolitanus*, all or most of CsoS2 protein may be associated with the carboxysome shell. Although our experimental results indicated that the C-terminus of CsoS2B is likely to be exposed at the outer surface of the carboxysome shell (Figure 29), it is unknown how large a portion of the full length protein is exposed. However, the recombinant CsoS2 protein is completely exposed in the *E. coli* cytoplasm so that the

recombinant CsoS2 protein may be easily accessible to attack by a protease. Consequently, more CsoS2A than CsoS2B would be apparent protein in *E. coli.* Identifying the C-terminal amino acid of CsoS2A protein will help gaining an understanding of the mechanism by which two CsoS2 proteins are formed from one gene.

The observed molecular weight (approximately 130 kDa) of CsoS2B protein is significantly higher than the expected molecular weight (92 kDa) deduced from its amino acid sequence. The reason for this discrepancy is unknown. Yeast two-hybrid analysis revealed that CsoS2 protein has a strong interaction with itself [127]. Thus, one CsoS2B molecule could strongly interact with one CsoS2A and migrate abnormally on SDS-PAGE electrophoresis. Also, CsoS2 protein may be modified by a post-translational modification to show an abnormal observed molecular weight on SDS-PAGE gel. CsoS2 protein is less likely phosphorylated in *H. neapolitanus*, because protein phosphorylation will decrease its pI and do not support the proposed role for transporting negatively charged molecules into and out of carboxysomes. The CsoS2 protein may be palmitoylated in *H. neapolitanus* and *E. coli*. Palmitoylation generally refers to one post-translational modification, the protein could be covalently modified by a variety of lipids [83] and the lipid molecules could reversibly attach to the specific cysteine residues in protein substrates through thioester linkages. CsoS2 protein has twenty cysteine residues [83], therefore, it is very likely that this modification can be applied to this protein. In previous studies, pneumococcal surface adhesin A (PsaA) from *Streptococcus pneumoniae* was expressed in *E. coli* as a palmitoylated recombinant protein and showed a larger observed molecular weight on SDS-PAGE gel[38] The CsoS2 protein in *Acidithiobacillus ferrooxidans* was found to exist in the periplasm [34]. Therefore,

palmitoylation of CsoS2 protein may help deliver the nascent CsoS2 protein to the periplasm across the lipid membrane. The existence of CsoS2 protein in periplasm may also indicate that CsoS2 protein may act as a signaling molecule to sense the concentration of inorganic carbon in the environment to regulate the expression of the carboxysomal genes. In addition, the CsoS2B protein may resist the denaturation of SDS in the denaturing SDS-PAGE electrophoresis and displayed an abnormal migration pattern. Measurement of the exact molecular weight of CsoS2 protein by MALDI-ToF will help explaining the abnormal migration pattern of CsoS2B protein on SDS-PAGE.

By tagging the small affinity SPA peptide to the C-terminus of CsoS2, the C-terminal end of the full length CsoS2B protein is speculated to be exposed at the outer surface of the carboxysomes. The exposed C-terminal tail of CsoS2B protein may directly interact with the negatively charged molecules, such as RuBP in the cytoplasm. The strong interaction between CsoS2 protein and small subunit of RuBisCO enzyme identified by yeast two-hybrid analysis [127] indicated that the CsoS2 protein should also be exposed the inner side of the shell. Therefore, the CsoS2B protein may span the carboxysomal shell and act as a tunnel to "transport" the negatively charged RuBP into the carboxysomes. Similarly, the CsoS2 protein can take the PGA produced by encapsulated RuBisCO within the carboxysomes and deliver it out of the carboxysomes via the tunnel.

Carboxysomal Shell Assembly

The role of RuBisCO in carboxysome shell assembly

In previous studies, it was determined that the *H. neapolitanus* carboxysomal shell can be disrupted into CsoS1A hexamers in the presence of 3.5 M urea [127]; RuBisCO holoenzyme remains intact under these conditions [75]. During the removal of urea, the carboxysomal components could re-interact to from the carboxysomal shell. In this study, a small SPA double affinity tag (54 amino acids) was added to the C-terminus of CsoS2 protein (869 amino acids) and was used to trap the intermediates of *in vitro* carboxysomal shell assembly on the anti-FLAG affinity column. The relatively small tag obviously did not change the ratio of the compositions and morphology of the resulting HnSPAS2 mutant carboxysomes compared with the wild type carboxysomes (Figure 26A and B), which indicated that the mutant HnSPAS2 carboxysomes may have the similar package pattern as wild type carboxysomes. The trapped intermediates are composed of all carboxysomal protein, but less RuBisCO enzyme (24% of total proteins) compared with that in wild type carboxysomes (55% of total proteins) (Figure 30). The trapped intermediates resemble the carboxysomal shell with 100 to 120 nm in diameter, however, RuBisCO holoenzyme molecules are not be filled in the interior (Figure 31). These findings indicated the importance of some RuBisCO molecules for *in vitro* carboxysome assembly.

The *cso* operon in *H. neapolitanus* contains *cbbL* and *cbbS* gene encoding large and small subunit of RuBisCO enzyme, respectively [29]. Previously, Cai *et al* reported that all of genes in the entire *cso* operon are transcribed [24], therefore all carboxysome components are important for carboxysome assembly, structure and/or function [24].

However, in *Thiobacillus denitrificans*, a complete set of carboxysomal shell genes exist in the genome, but lack *cbbL and cbbS* genes [30]. Interestingly, to date, no carboxysomes have been detected in this bacterium [30]. It is unknown whether the absence of RuBisCO genes affects the formation of the carboxysomes in this bacterium.

In the trapped intermediates of *in vitro* carboxysome assembly, some RuBisCO molecules may strongly associate with the carboxysome shell proteins to facilitate the formation of the carboxysome shell *in vitro*. When the carboxysomes of *H. neapolitanus* and *Synechococcus* PCC 7942 are treated with several freeze-thaw cycles to release the encapsulated RuBisCO molecules, the encapsulated RuBisCO molecules can be divided into two different populations. One population interacts loosely or do not interact with the carboxysomal shell, while the other population shows a strong interaction with the carboxysomal shell [37, 114].

However, on the other hand, circular structures like partially assembled carboxysome shell intermediates were observed in *Anabaena variabilis* M3 [89] and "empty" carboxysomes were found in the *H. neapolitanus cbbL::Km* mutant [12]. Recently, the empty carboxysomes were observed and isolated from the *H. neapolitanus cbbLS::Km* mutant that is deficient in Form I RuBisCO [80]. The cryoelectron tomograms of carboxysomes of *Synechococcus* strain WH 8102 revealed that there do not seem to be strong interactions between RuBisCO and the carboxysome shell, which also supports a pathway of RuBisCO-independent formation of α-carboxysomes [61]. Although the "empty" carboxysomes show the similar shape and size as their wild type counterparts, it is unknown whether the three dimensional architecture of the empty carboxysomal shell is the same as that of the wild type. Cryoelectron tomography studies

of empty carboxysomal shells will help elucidate the role of RuBisCO in carboxysomal shell assembly. Similar observations were found in the *pdu* microcompartment of *Salmonella enterica* [20]. Deletion the genes for the sequestered enzymes that metabolize 1,2-propandiol do not affect the formation of the shell structures [20].

Carboxysome-like structures are assembled in E. coli

Previously, the *cso* operon was co-expressed in *E. coli* with chaperonin molecules and some carboxysome-like structures were observed in the micrographs of the bacteria [1]. In this study, recombinant carboxysome-like structures were recovered from a sucrose gradient (Figure 32). A lot of irregular carboxysome-like structures are observed in the middle fraction of 10-60% sucrose gradient (Figure 34). The size of these carboxysome-like structures is very irregular. The irregular size may be caused by the different transcription and translation patterns of the *cso* operon in *E. coli*. The entire *cso* operon was expressed the control of its own promoter. Although *E. coli* can recognize the own promoter of the *cso* operon, the transcriptional levels of individual carboxysomal gene in the operon may be different with those in *H. neapolitanus*. The different ratios of carboxysome components may change the package pattern of the carboxysome shell in *E. coli* and result in the formation of irregular carboxysome-like structures. In addition, Eric *et al.* identified that CsoS2 protein plays an important role in carboxysomal shell assembly with recombinant carboxysomal proteins *in vitro* [127]. When CsoS2 protein was expressed in *E. coli*, it was found to form inclusion bodies [13]. Therefore, the extremely low abundance of CsoS2 protein among the total carboxysome could also result in the formation of irregular carboxysomes.

In electron micrographs of negatively stained carboxysome-like structures purified from *E. coli* harboring the *cso* operon, the characteristic donut-shaped RuBisCO molecules that are clearly visualized in wild type carboxysomes are not visible. The interiors of the carboxysome-like structures could be filled with more irregularly shaped larger clusters that may represent the incorrectly folded RuBisCO molecules. Interestingly, a lot of filaments co-purified with carboxysome-like structures from *E. coli*. Similar filaments also were found in *H. neapolitanus* cells by cryoelectron tomography analysis (C. V. Iancu, personal communication). The composition of these filaments is still unknown. Given that recombinant CsoS1A protein can form disk- or rod-like structures *in vitro* [127], the filaments observed in *E. coli* and *H. neapolitanus* may represent misassembled carboxysomal shell proteins. Immunoelectron microscopic studies using anti-CsoS1A/C antibody may help reveal the composition of the assembled filaments.

CHAPTER VI

CONCLUSION AND FUTURE WORK

Successful disruption of *csoS3* gene in *H. neapolitanus* with a kanamycin resistance gene cassette caused a high CO_2-requirng phenotype in the resulting mutant. The missing of CsoSCA protein did not change the integrity of carboxysomes significantly. By comparing kinetic parameters of CO_2 fixation by wild type and mutant carboxysomes *in vitro*, the carboxysome shell lacking carboxysomal carbonic anhydrase showed a diffusion barrier for CO_2 molecules. The shell-associated carbonic anhydrase can increase the permeability of carboxysome shell for CO_2 molecules to enhance the catalytic efficiency of RuBisCO enzyme sequestered within the carboxysomes. Exogenously added recombinant CA did not functionally complement the missing carboxysomal component in the mutant carboxysomes, which suggested that the carboxysomal carbonic anhydrase may face the interior of the carboxysomes. Given that CsoSCA protein can interact with CsoS1B protein, not CsoS1A and CsoS1C protein via yeast two-hybrid analysis [127], CsoSCA protein may associate with the shell via the interaction with the unique 12 amino acid C-terminal tail of CsoS1B protein. A CsoS1B C-terminal 12 amino acid truncation mutant needs to be generated to identify the interaction between the C-terminal tail of CsoS1B with CsoSCA protein and its role in the function and formation of carboxysomes.

Carboxysome shell also acts as the barrier to protect the carboxysomal RuBisCO enzyme from its competitive substrate – O_2 to enhance its ratio of carboxylase to oxygenase activity. Isotope simultaneous assay revealed that wild type carboxysomes produced 1.4 to 1.5 times more PGA per PG than free RuBisCO. O_2 inhibition assay also

identified that the carboxysomal RuBisCO is less sensitive than free RuBisCO enzyme for O_2 inhibition. However, the carboxysomal RuBisCO only favor the carboxylation over the oxygenation reaction in a low O_2 concentration over a short time. Further characterization of the permeability of carboxysomal shell for O_2 can be performed by a stop-flowed technique that can measure the enzymatic reaction over a short time frame so that more accurate O_2 inhibition constants for carboxysomal and free RuBisCO could be obtained.

The precise mechanism of the transportation of phosphorylated RuBisCO substrates and products into and out of the carboxysomes is still unknown. The extraordinary high pI value of CsoS2 protein indicates that CsoS2 protein is a potential candidate to perform that role [30, 57]. Given that CsoS2 protein carries many V/I-T/S-G repeat motifs in its primary structure [30], recombinant CsoS2B and Δ250CsoS2B proteins were expressed in *E. coli* for their three dimensional structures determination. By adding a SPA affinity tag at the C-terminus of CsoS2 protein in *H. neapolitanus*, CsoS2B protein was identified as a full length CsoS2 protein, while CsoS2A protein is a C-terminally truncated CsoS2 protein. In addition, trapping the HnSPAS2 carboxysomes on an anti-FLAG affinity column revealed that the C-terminus of CsoS2B protein may be exposed to the outer surface of the carboxysomes. Further generation and characterization of a *csoS2* gene knockout mutant may help understand the role of CsoS2 protein in the biogenesis of carboxysomes and its function in the transportation of the negatively charged molecules. The crystal structure determination of CsoS2 protein also will enhance our understanding of CsoS2 protein's function. In this study, the last amino acid of CsoS2A protein was not identified by peptide mapping. A C-terminal sequencing

analysis could be adopted in the future to determine the C-terminal end of CsoS2A protein, which will help elucidate the mechanism of the formation of two CsoS2 polypeptides.

The *in vitro* re-assembly of carboxysome shell revealed that the carboxysome shell can be self-assembled without any input of energy and RuBisCO molecules are not filled into the carboxysome-like structure completely. To understand the assembly pathway of carboxysomes *in vivo*, HnSPAS2 mutants should be grown in CO_2-enriched environment first to suppress the expression and assembly of carboxysomes [18], and be switched to the ambient CO_2 environment to initiate the formation of carboxysomes *in vivo*. At different time points, mutant cells can be disrupted and the desired carboxysome intermediates at different stages can be caught on the affinity column. Their composition and morphology can be analyzed by SDS-PAGE electrophoresis and transmission electron microscopy, and this information will be crucial to resolve the biogenesis pathway of carboxysomes of *H. neapolitanus*.

Expression of the *cso* operon of *H. neapolitanus* in a heterotrophic bacterium *E. coli* yielded some carboxysome-like structures. Interestingly, lots of filaments co-purified with the carboxysome-like structures from 10-60% sucrose gradient. The similar filaments were also observed in *H. neapolitanus* cells (C. V. Cristina, personal communication). They may represent the mis-assembled carboxysomal shell proteins. Further, immunoelectron microscopy studies can be used to identify the composition of these filaments.

REFERENCES

1. Aldrich, H.C., Elvington, S., MacInnes, H.E., Szabady, R., Feder, K., McDowell, L. and Shively, J.M., (2001). Abstr. 59th Annu. Meet. Microsc. Soc. Am., abstr.

2. Amichay, D., Levitz, R. and Gurevitz, M., (1993). Construction of a *Synechocystis* PCC 6803 mutant suitable for the study of variant hexadecameric ribulose bisphosphate carboxylase/oxygenase enzymes. *Plant Mol Biol*, 23(3), 465-476.

3. Andrews, T.J. and Lorimer, G.H., in *The Biochemistry of Plants*. 1987, Academic Press. p. 131-218.

4. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K., *Current Protocols in Molecular Biology*. 1994: Wiley Interscience. Section 1.7.

5. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K., *Current Protocols in Molecular Biology*. 1994: Wiley Interscience. Section 2.4.

6. Badger, M.R., (1980). Kinetic Propeties of Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase from *Anabaena variabilis*. *Arch Biochem Biophys*, 201(1), 247-254.

7. Badger, M.R., Bassett, M. and Comins, H.N., (1985). A Model for HCO_3^- Accumulation and Photosynthesis in the Cyanobacterium *Synechococcus* sp: Theoretical Predictions and Experimental Observations. *Plant Physiol*, 77(2), 465-471.

8. Badger, M.R., Hanson, D. and Price, G.D., (2002). Evolution and diversity of CO_2 concentrating mechanisms in cyanobacteria. *Functional Plant Biology*, 29(3), 161-173.

9. Badger, M.R. and Price, G.D., (2003). CO_2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *J Exp Bot*, 54(383), 609-622.

10. Badger, M.R., Price, G.D., Long, B.M. and Woodger, F.J., (2006). The environmental plasticity and ecological genomics of the cyanobacterial CO_2 concentrating mechanism. *J Exp Bot*, 57(2), 249-65.

11. Baker, S.H., *Molecular Investigation of Carbon Dioxide Fixation in Thiobacilli.*, in *Doctoral dissertation. Clemson University, Clemson, S.C.* 1998.

12. Baker, S.H., Jin, S., Aldrich, H.C., Howard, G.T. and Shively, J.M., (1998). Insertion mutation of the form I *cbbL* gene encoding ribulose bisphosphate carboxylase/oxygenase (RuBisCO) in *Thiobacillus neapolitanus* results in expression of form II RuBisCO, loss of carboxysomes, and an increased CO_2 requirement for growth. *J Bacteriol*, 180(16), 4133-4139.

13. Baker, S.H., Lorbach, S.C., Rodriguez-Buey, M., Williams, D.S., Aldrich, H.C. and Shively, J.M., (1999). The correlation of the gene *csoS2* of the carboxysome operon with two polypeptides of the carboxysome in *Thiobacillus neapolitanus*. *Arch Microbiol*, 172(4), 233-239.

14. Baker, S.H., Williams, D.S., Aldrich, H.C., Gambrell, A.C. and Shively, J.M., (2000). Identification and localization of the carboxysome peptide CsoS3 and its corresponding gene in *Thiobacillus neapolitanus*. *Arch Microbiol*, 173(4), 278-283.

15. Ball, D.W., (2003). Physical Chemistry. 674.

16. Baranov, P.V., Gurvich, O.L., Hammer, A.W., Gesteland, R.F. and Atkins, J.F., (2003). RECODE 2003. *Nucleic Acids Res*, 31(1), 87-89.

17. Berry, S., Fischer, J.H., Kruip, J., Hauser, M. and Wildner, G.F., (2005). Monitoring cytosolic pH of carboxysome-deficient cells of *Synechocystis* sp. PCC 6803 using fluorescence analysis. *Plant Biol (Stuttg)*, 7(4), 342-347.

18. Beudeker, R.F., Cannon, G.C., Kuenen, J.G. and Shively, J.M., (1980). Relations between D-ribulose-1,5-bisphosphate carboxylase, carboxysomes, and CO_2 fixing capacity in the obligate chemolithotroph *Thiobacillus neapolitanus* grown under different limitations in the chemostat. *Arch Microbiol*, 124, 185-189.

19. Bobik, T.A., Havemann, G.D., Busch, R.J., Williams, D.S. and Aldrich, H.C., (1999). The propanediol utilization (*pdu*) operon of *Salmonella enterica* serovar Typhimurium LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme B_{12}-dependent 1, 2-propanediol degradation. *J Bacteriol*, 181(19), 5967-5975.

20. Bobik, T.A., (2006). Polyhedral organelles compartmenting bacterial metabolic processes. *Appl Microbiol Biotechnol*, 70, 517-525.

21. Brinsmade, S.R., Paldon, T. and Escalante-Semerena, J.C., (2005). Minimal functions and physiological conditions required for growth of *Salmonella enterica* on ethanolamine in the absence of the metabolosome. *J Bacteriol*, 187(23), 8039-8046.

22. Buan, N.R. and Escalante-Semerena, J.C., (2006). Purification and Initial Biochemical Characterization of ATP:Cob(I)alamin Adenosyltransferase (EutT) Enzyme of *Salmonella enterica*. *J Biol Chem*, 281(25), 16971-16977.

23. Burey, S.C., Poroyko, V., Ergen, Z.N., Fathi-Nejad, S., Schuller, C., Ohnishi, N., Fukuzawa, H., Bohnert, H.J. and Loffelhardt, W., (2007). Acclimation to low [CO_2] by an inorganic carbon-concentrating mechanism in *Cyanophora paradoxa*. *Plant Cell Environ*, 30(11), 1422-1435.

24. Cai, F., Heinhorst, S., Shively, J.M. and Cannon, G.C., (2008). Transcript analysis of the *Halothiobacillus neapolitanus cso* operon. *Arch Microbiol*, 189(2), 141-50.

25. Cannon, G.C., *Carboxysomes and CO_2 fixation in Thiobacillus neapolitanus.*, in *Doctoral dissertation. Clemson University, Clemson, S.C.* 1982.

26. Cannon, G.C. and Shively, J.M., (1983). Characterization of a homogenous preparation of carboxysomes from *Thiobacillus neapolitanus*. *Arch Microbiol*, 134(1), 52-59.

27. Cannon, G.C., English, R.S. and Shively, J.M., (1991). *In situ* assay of ribulose-1,5-bisphosphate carboxylase/oxygenase in *Thiobacillus neapolitanus*. *J Bacteriol*, 173(4), 1565-1568.

28. Cannon, G.C., Bradburne, C.E., Aldrich, H.C., Baker, S.H., Heinhorst, S. and Shively, J.M., (2001). Microcompartments in Prokaryotes: Carboxysomes and Related Polyhedra. *Appl Environ Microbiol*, 67(12), 5351-5361.

29. Cannon, G.C., Heinhorst, S., Bradburne, C.E. and Shively, J.M., (2002). Carboxysome genomics: a status report. *Functional Plant Biology*, 29(3), 175-182.

30. Cannon, G.C., Baker, S.H., Soyer, F., Johnson, D.R., Bradburne, C.E., Mehlman, J.L., Davies, P.S., Jiang, Q.L., Heinhorst, S. and Shively, J.M., (2003). Organization of carboxysome genes in the thiobacilli. *Curr Microbiol*, 46(2), 115-119.

31. Charlier, H.A., Runquist, J.A. and Miziorko, H.M., (1994). Evidence supporting catalytic roles for aspartate residues in phosphoribulokinase. *Biochemistry*, 33(31), 9343-9350.

32. Charlier, H.A.J., Runquist, J.A. and Miziorko, H.M., (1994). Evidence supporting catalytic roles for aspartate residues in phosphoribulokinase. *Biochemistry*, 33(31), 9343-9350.

33. Chen, P., Andersson, D. and Roth, J.R., (1994). The control region of the *pdu/cob* regulon on *Salmonella typhimurium J Bacteriol*, 176, 5674-5482.

34. Chi, A., Valenzuela, L., Beard, S., Mackey, A.J., Shabanowitz, J., Hunt, D.F. and Jerez, C.A., (2007). Periplasmic Proteins of the Extremophile *Acidithiobacillus ferrooxidans*. *Molecular & Cellular Proteomics*, 6, 2239-2251.

35. Coleman, J.R., Seemann, J.R. and Berry, J.A., (1982). RuBP carboxylase in carboxysomes of blue-green algae. *Carnegie Inst Wash Year Book*, 81, 83-87.

36. Cooper, T.G., Filmer, D., Wishnick, M. and Lane, M.D., (1969). The Active Species of "CO_2" Utilized by Ribulose Diphosphate Carboxylase. *J Biol Chem*, 244(4), 1081-1083.

37. Cot, S.S., So, A.K. and Espie, G.S., (2008). A multiprotein bicarbonate dehydration complex essential to carboxysome function in cyanobacteria. *J Bacteriol*, 190(3), 936-945.

38. Dea, B.K., Sampson, J.S., Ades, E.W., Huebner, R.C., Jue, D.L., Johnson, S.E., Espina, M., Stinsona, A.R., Briles, D.E. and Carlone, G.M., (2000). Purification and characterization of *Streptococcus pneumoniae* palmitoylated pneumococcal surface adhesin A expressed in *Escherichia coli*. *Vaccine*, 18, 1811-1821.

39. Dean, J.A., (1999). Lange's Handbook of Chemistry, 15th Ed. *McGraw-Hill, Inc., New York, NY*, 5.3–5.7.

40. Dobrinski, K.P., Longo, D.L. and Scott, K.M., (2005). The Carbon-Concentrating Mechanism of the Hydrothermal Vent Chemolithoautotroph *Thiomicrospira crunogena*. *J Bacteriol*, 187(16), 5761-5766.

41. Dou, Z., Heinhorst, S., Williams, E.B., Murin, C.D., Shively, J.M. and Cannon, G.C., (2008). 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(16), 10377-10384.

42. Drews, G. and Niklowitz, W., (1956). Beiträge zur Cytologie der Blaualgen. *Arch Microbiol*, 24(2), 147-162.

43. Ellis, R.J., (1979). The most abundant protein in the world. *Trends Biochem Sci*, 4, 241-244.

44. Emlyn-Jones, D., Woodger, F.J., Andrews, T.J., Price, G.D. and Whitney, S.M., (2006). A *Synechococcus* PCC 7942 ΔCcmM (Cyanophyceae) Mutant Pseudoreverts To Air Growth Without Regaining Carboxysomes. *J Phycol*, 42(4), 769-777.

45. English, R.S., Lorbach, S.C., Qin, X. and Shively, J.M., (1994). Isolation and characterization of a carboxysome shell gene from *Thiobacillus neapolitanus*. *Mol Microbiol*, 12(4), 647-654.

46. English, R.S., Jin, S. and Shively, J.M., (1995). Use of Electroporation To Generate a *Thiobacillus neapolitanus* Carboxysome Mutant. *Appl Environ Microbiol*, 61(9), 3256-3260.

47. Falkowski, P., (1997). The paradox of carbon dioxide efflux. *Curr Biol*, 7, 637-639.

48. Fan, C. and Bobik, T.A., (2008). The PduX enzyme of *Salmonella enterica* is an L-threonine kinase used for coenzyme B_{12} synthesis. *J Biol Chem*, 283(17), 11322-11329.

49. Fathinejad, S., Steiner, J.M., Reipert, S., Marchetti, M., Allmaier, G., Burey, S.C., Ohnishi, N., Fukuzawa, H., Loffelhardt, W. and Bohnert, H.J., (2008). A carboxysomal carbon-concentrating mechanism in the cyanelles of the 'coelacanth' of the algal world, *Cyanophora paradoxa*? *Physiol Plant*, 133(1), 27-32.

50. Fridlyand, L., Kaplan, A. and Reinhold, L., (1996). Quantitative evaluation of the role of a putative CO_2-scavenging entity in the cyanobacterial CO_2-concentrating mechanism. *Biosystems*, 37(3), 229-238.

51. Gonzales, A.D., Light, Y.K., Zhang, Z., Iqbal, T., Lane, T.W. and Martino, A., (2005). Proteomic analysis of the CO_2-concentrating mechanism in the open-ocean cyanobacterium *Synechococcus* WH 8102. *Can J Bot*, 83, 735-745.

52. Harpel, M.R., Lee, E.H. and Hartman, F.C., (1993). Anion-Exchange Analysis of Ribulose-Bisphosphate Carboxylase/Oxygenase Reactions: CO_2/O_2 Specificity Determination and Identification of Side Products. *Anal Biochem*, 209, 367-374.

53. Hartman, F.C. and Harpel, M.R., (1994). Structure, Function, Regulation, and Assembly of D-Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase. *Annu Rev Biochem*, 63, 197-234.

54. Havemann, G.D., Sampson, E.M. and Bobik, T.A., (2002). PduA is a shell protein of polyhedral organelles involved in the coenzyme B_{12}-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar Typhimurium LT2. *J Bacteriol*, 184, 1253-1261.

55. Havemann, G.D. and Bobik, T.A., (2003). Protein Content of Polyhedral Organelles Involved in Coenzyme B_{12}-Dependent Degradation of 1,2-Propanediol in *Salmonella enterica* Serovar Typhimurium LT2. *J Bacteriol*, 185(17), 5086-5095.

56. Heinhorst, S., Cannon, G.C. and Shively, J.M., *Carboxysomes and Carboxysome-like Inclusions*. 2006, Springer. p. 141-165.

57. Heinhorst, S., Williams, E.B., Cai, F., Murin, C.D., Shively, J.M. and Cannon, G.C., (2006). Characterization of the carboxysomal carbonic anhydrase CsoSCA from *Halothiobacillus neapolitanus. J Bacteriol*, 188(23), 8087-8094.

58. Heinhorst, S. and Cannon, G.C., (2008). A new, leaner and meaner bacterial organelle. *Nat Struct Mol Biol*, 15(9), 897-898.

59. Holthuijzen, Y.A., VanBreemen, J.F.L., Kuenen, J.G. and Konings, W.N., (1986). Protein composition of the carboxysomes of *Thiobacillus neapolitanus. Arch Microbiol*, 144, 398-404.

60. Holthuijzen, Y.A., Kuenen, J.G. and Konings, W.N., (1987). Activity of ribulose 1,5-bisphosphate carboxylase in intact and disrupted carboxysomes of *Thiobacillus neapolitanus. FEMS Microbiol Lett*, 42, 121-124.

61. Iancu, C.V., Ding, H.J., Morris, D.M., Dias, D.P., Gonzales, A.D., Martino, A. and Jensen, G.J., (2007). The structure of isolated *Synechococcus* strain WH 8102 carboxysomes as revealed by electron cryotomography. *J Mol Biol*, 372(3), 764-773.

62. Islam, T., Jensen, S., Reigstad, L.J., Larsen, O. and Birkeland, N.K., (2008). Methane oxidation at 55 °C and pH 2 by a thermoacidophilic bacterium belonging to the *Verrucomicrobia phylum. Proc Natl Acad Sci U S A*, 105(1), 300-304.

63. Johnson, C.L.V., Buszko, M.L. and Bobik, T.A., (2004). Purification and Initial Characterization of the *Salmonella enterica* PduO ATP:Cob(I)alamin Adenosyltransferase *J Bacteriol*, 186(23), 7881-7887.

64. Jordan, D.B. and Ogren, W.L., (1981). Species variation in the specificity of ribulose biphosphate carboxylase/oxygenase. *Nature*, 291, 513-515.

65. Kane, H.J., Viil, J., Entsch, B., Paul, K., Morell, M.K. and Andrews, T.J., (1994). An Improved Method for Measuring the CO_2/O_2 Specificity of Ribulosebisphosphate Carboxylase-Oxygenase. *Aust J Plant Physiol*, 21(4), 449-461.

66. Kaneko, Y., Danev, R., Nagayama, K. and Nakamoto, H., (2006). Intact carboxysomes in a cyanobacterial cell visualized by hilbert differential contrast transmission electron microscopy. *J Bacteriol*, 188(2), 805-808.

67. Kannappan, B. and Gready, J.E., (2008). Redefinition of Rubisco Carboxylase Reaction Reveals Origin of Water for Hydration and New Roles for Active-Site Residues. *J Am Chem Soc*, 130, 15063-15080.

68. Kerfeld, C.A., Sawaya, M.R., Tanaka, S., Nguyen, C.V., Phillips, M., Beeby, M. and Yeates, T.O., (2005). Protein structures forming the shell of primitive bacterial organelles. *Science*, 309(5736), 936-938.

69. Kofoid, E., Rappleye, C., Stojiljkovic, I. and Roth, J., (1999). The 17-gene ethanolamine (*eut*) operon of *Salmonella* typhimurium encodes five homologues of carboxysome shell proteins. *J Bacteriol*, 181(17), 5317-5329.

70. Kovács, E., van der Vies, S.M., Glatz, A., Török, Z., Varvasovszki, V., Horváth, I. and Vígh, L., (2001). The Chaperonins of *Synechocystis* PCC 6803 Differ in Heat Inducibility and Chaperone Activity. *Biochem Biophys Res Commun*, 289(4), 908-915.

71. Kreel, N.E. and Tabita, F.R., (2007). Substitutions at Methionine 295 of *Archaeoglobus fulgidus* Ribulose-1,5-bisphosphate Carboxylase/Oxygenase Affect Oxygen Binding and CO_2/O_2 Specificity. *J Biol Chem*, 282(2), 1341-1351.

72. Laing, W.A., Orgen, W.L. and Hageman, R.H., (1974). Regulation of Soybean Net Photosynthetic CO_2 Fixation by the Interaction of CO_2, O_2 and Ribulose 1,5-Diphosphate Carboxylase. *Plant Physiol*, 54, 678-685.

73. Larsen, B., Wills, N.M., Nelson, C., Atkins, J.F. and Gesteland, R.F., (2000). Nonlinearity in genetic decoding: Homologous DNA replicase genes use alternatives of transcriptional slippage or translational frameshifting. *Proc Natl Acad Sci U S A*, 97(4), 1683-1688.

74. Liu, Y., Leal, N.A., Sampson, E.M., Johnson, C.L.V., Havemann, G.D. and Bobik, T.A., (2007). PduL Is an Evolutionarily Distinct Phosphotransacylase Involved in B_{12}-Dependent 1,2-Propanediol Degradation by *Salmonella enterica* Serovar Typhimurium LT2. *J Bacteriol*, 189(5), 1589-1596.

75. Long, B.M., Badger, M.R., Whitney, S.M. and Price, G.D., (2007). Analysis of carboxysomes from *Synechococcus* PCC 7942 reveals multiple Rubisco complexes with carboxysomal proteins CcmM and CcaA. *J Biol Chem*, 282(40), 29323-29335.

76. Lorimer, G.H., Badger, M.R. and Andrews, T.J., (1976). The Activation of Ribulose 1,5-bisphosphate Carboxylase by Carbon Dioxide and Magnesium Ions. Equilibra, Kinetics, a Suggested Mechanism, and Physiological Implications. *Biochemistry*, 15(3), 529-536.

77. Lyttleton, J.W., (1973). Carbon dioxide incorporation by chloroplast extracts at high pH. *FEBS Lett*, 38(1), 4-6.

78. Maeda, S., Badger, M.R. and Price, G.D., (2002). Novel gene products associated with NdhD3/D4-containing NDH-1 complexes are invovled in photosynthetic

CO_2 hydration in the cyanobacterium *Synechococcus* sp. PCC7942. *Mol Microbiol*, 43, 425-435.

79. Marcus, Y., Berry, J.A. and Pierce, J., (1992). Photosynthesis and photorespiration in a mutant of the cyanobacterium *Synechocystis* PCC 6803 lacking carboxysomes. *Planta*, 187, 511-516.

80. Menon, B.B., Dou, Z., Heinhorst, S., Shively, J.M. and Cannon, G.C., (2008). *Halothiobacillus neapolitanus* Carboxysomes Sequester Heterologous and Chimeric RubisCO Species. *PLoS ONE*, 3(10), e3570.

81. Mott, K.A. and Berry, J.A., (1986). Effects of pH on Activity and Activation of Ribulose 1,5-Bisphosphate Carboxylase at Air Level CO_2. *Plant Physiol*, 82, 77-82.

82. Murin, C.D., *Mutation Study of the Carboxysome Carbonic Anhydrase of Halothiobacillus neapolitanus.*, in *Bachelor thesis. The University of Southern Mississippi, Hattiesburg, M.S.* 2008.

83. Nadolski, M.J. and Linder, M.E., (2007). Protein lipidation. *FEBS J*, 274(20), 5202-5210.

84. Ohkawa, H., Sonoda, M., Katoh, A. and Ogawa, T., (1998). The use of mutants in the analysis of the CO_2-concentrating mechanism in cyanobacteria. *Can J Bot*, 76, 1035-1042.

85. Omata, T., Price, G.D., Badger, M.R., Okamura, M., Gohta, S. and Ogawa, T., (1999). Idetification of an ATP-binding cassette transporter invovled in bicarbonate uptake in the cyanobacterium *Synechococcus* sp. strain PCC 7942. *Proc Natl Acad Sci U S A*, 96, 13571-13576.

86. Orus, M.I., Rodriguez, M.L., Martinez, F. and Marco, E., (1995). Biogenesis and Ultrastructure of Carboxysomes from Wild Type and Mutants of *Synechococcus* sp. Strain PCC 7942. *Plant Physiol*, 107(4), 1159-1166.

87. Penrod, J.T. and Roth, J.R., (2006). Conserving a volatile metabolite: a role for carboxysome-like organelles in *Salmonella enterica*. *J Bacteriol*, 188(8), 2865-2874.

88. Price, G.D. and Badger, M.R., (1989). Expression of Human Carbonic Anhydrase in the Cyanobacterium *Synechococcus* PCC 7942 Creates a High CO_2-Requiring Phenotype : Evidence for a Central Role for Carboxysomes in the CO_2 Concentrating Mechanism. *Plant Physiol*, 91(2), 505-513.

89. Price, G.D. and Badger, M.R., (1991). Evidence for the role of carboxysomes in the cyanobacterial CO_2-concentrating mechanism. *Can J Bot*, 69, 963-973.

90. Price, G.D., Coleman, J.R. and Badger, M.R., (1992). Association of Carbonic Anhydrase Activity with Carboxysomes Isolated from the Cyanobacterium *Synechococcus* PCC 7942. *Plant Physiol*, 100(2), 784-793.

91. Price, G.D., Howitt, S.M., Harrison, K. and Badger, M.R., (1993). Analysis of a genomic DNA region from the cyanobacterium *Synechococcus* sp. strain PCC 7942 involved in carboxysome assembly and function. *J Bacteriol*, 175(10), 2871-2879.

92. Price, G.D., Sultemeyer, D., Klughammer, B., Ludwig, M. and Badger, M.R., (1998). 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, 973-1002.

93. Price, G.D., Maeda, S.I., Omata, T. and Badger, M.R., (2002). Modes of inorganic carbon uptake in the cyanobacterium *Synechococcus* sp. PCC 7942. *Functional Plant Biology*, 29, 131-149.

94. Price, G.D., Woodger, F.J., Badger, M.R., Howitt, S.M. and Tucker, L., (2004). Identification of SulP-type bicarbonate transporter in marine cyanobacteria. *Proc Natl Acad Sci U S A*, 101, 18228-18233.

95. Price, G.D., Badger, M.R., Woodger, F.J. and Long, B.M., (2008). Advances in understanding the cyanobacterial CO_2-concentrating-mechanism (CCM): functional components, C_i transporters, diversity, genetic regulation and prospects for engineering into plants. *J Exp Bot*, 59(7), 1441-1461.

96. Purohit, K., McFadden, B.A. and Shaykh, M.M., (1976). D-Ribulose-1,5-bisphosphate carboxylase and polyhedral inclusion bodies in *Thiobacillus intermedius*. *J Bacteriol*, 127(1), 516-522.

97. Ronen-Tarazi, M., Lieman-Hurwitz, J., Gabay, C., Orus, M.I. and Kaplan, A., (1995). The genomic region of *rbcLS* in *Synechococcus* sp. PCC 7942 contains genes involved in the ability to grow under low CO_2 concentration and in chlorophyll biosynthesis. *Plant Physiol*, 108(4), 1461-1469.

98. Sambrook, J., Fritsch, E.F. and Maniatis, T., *Molecular Cloning - A Laboratory Manual (Second Edition)*. 1989: Cold Spring Harbor Laboratory Press. Section 1.40.

99. Sampson, E.M. and Bobik, T.A., (2008). Microcompartments for B_{12}-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. *J Bacteriol*, 190(8), 2966-2971.

100. Sawaya, M.R., Cannon, G.C., Heinhorst, S., Tanaka, S., Williams, E.B., Yeates, T.O. and Kerfeld, C.A., (2006). 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(11), 7546-7555.

101. Schmid, M.F., Paredes, A.M., Khant, H.A., Soyer, F., Aldrich, H.C., Chiu, W. and Shively, J.M., (2006). Structure of *Halothiobacillus neapolitanus* carboxysomes by cryo-electron tomography. *J Mol Biol*, 364(3), 526-535.

102. Shibata, M., Ohkawa, H., Kaneko, T., Fukuzawa, H., Tabata, S., Kaplan, A. and Ogawa, T., (2001). Distinct constitutive and low-CO_2-induced CO_2 uptake systems in cyanobacteria: genes involved and their phylogenetic relationship with homologous genes in other organisms. *Proc Natl Acad Sci U S A*, 98, 11789-11794.

103. Shibata, M., Katoh, H., Sonada, M., Ohkawa, H., Shimoyama, M., Fukuzawa, H., Kaplan, A. and Ogawa, T., (2002). Genes essential to sodium-dependent bicarbonate transport in cyanobacteria - function and phylogenetic analysis. *J Biol Chem*, 277, 18658-18664.

104. Shibata, M., Ohkawa, H., Katoh, H., Shimoyama, M. and Ogawa, T., (2002). Two CO_2-uptake systems: four systems for inorganic carbon acquisition in *Synechocystis* sp. strain PCC 6803. . *Functional Plant Biology*, 29, 123-129.

105. Shively, J.M., Ball, F., Brown, D.H. and Saunders, R.E., (1973). Functional Organelles in Prokaryotes: Polyhedral Inclusions (Carboxysomes) of *Thiobacillus neapolitanus. Science*, 182(4112), 584-586.

106. Shively, J.M., Ball, F.L. and Kline, B.W., (1973). Electron microscopy of the carboxysomes (polyhedral bodies) of *Thiobacillus neapolitanus. J Bacteriol*, 116(3), 1405-1411.

107. Shively, J.M., (1974). Inclusion Bodies of Prokaryotes. *Annu Rev Microbiol*, 28(1), 167-188.

108. Shively, J.M. and English, R.S., (1991). The carboxysomes, a prokaryotic organelle: a mini-review. *Can J Bot*, 69, 957-962.

109. Shively, J.M., Lorbach, S.C., Jin, S. and Baker, S.H., *Carboxysomes: the genes of Thiobacillus neapolitanus.*, in *M. E. Lidstrom and F. R. Tabita (ed.), Microbial growth on C1 compounds. Kluwer, Dordrecht, The Netherlands.* 1996. p. 56-63.

110. Shively, J.M., van Keulen, G. and Meijer, W.G., (1998). Something from almost nothing: carbon dioxide fixation in chemoautotrophs. *Annu Rev Microbiol*, 52, 191-230.

111. Shively, J.M., English, R.S., Baker, S.H. and Cannon, G.C., (2001). Carbon cycling: the prokaryotic contribution. *Curr Opin Microbiol*, 4(3), 301-306.

112. So, A.K. and Espie, G.S., (1998). Cloning, characterization and expression of carbonic anhydrase from the cyanobacterium *Synechocystis* PCC 6803. *Plant Mol Biol*, 37(2), 205-215.

113. So, A.K., John-McKay, M. and Espie, G.S., (2002). Characterization of a mutant lacking carboxysomal carbonic anhydrase from the cyanobacterium *Synechocystis* PCC 6803. *Planta*, 214(3), 456-467.

114. So, A.K., Espie, G.S., Williams, E.B., Shively, J.M., Heinhorst, S. and Cannon, G.C., (2004). A novel evolutionary lineage of carbonic anhydrase (epsilon class) is a component of the carboxysome shell. *J Bacteriol*, 186(3), 623-630.

115. Spreitzer, R.J. and Salvucci, M.E., (2002). Rubisco: structure, regulatory interactions, and possibilities of a better enzyme. *Annu Rev Plant Biol*, 53, 449-475.

116. Sriramulu, D.D., Liang, M., Hernandez-Romero, D., Raux-Deery, E., Lunsdorf, H., Parsons, J.B., Warren, M.J. and Prentice, M.B., (2008). *Lactobacillus reuteri* DSM 20016 Produces Cobalamin-Dependent Diol Dehydratase in Metabolosomes and Metabolizes 1,2-Propanediol by Disproportionation. *J Bacteriol*, 190(13), 4559-4567.

117. Starai, V.J., Garrity, J. and Escalante-Semerena, J.C., (2005). Acetate excretion during growth of *Salmonella enterica* on ethanolamine requires phosphotransacetylase (EutD) activity, and acetate recapture requires acetyl-CoA synthetase (Acs) and phosphotransacetylase (Pta) activities. *Microbiology*, 151, 3793-3801.

118. Stojiljkovic, I., Baumler, A.J. and Heffron, F., (1995). Ethanolamine utilization in *Salmonella* typhimurium: nucleotide sequence, protein expression, and mutational analysis of the *cchA cchB eutE eutJ eutG eutH* gene cluster. *J Bacteriol*, 177(5), 1357-1366.

119. Sugiyama, T., Nakayama, N. and Akazawa, T., (1968). Structure and function of chloroplast proteins. V. Homotropic effect of bicarbonate in RuDP carboxylase reaction and the mechanism of activation by magnesium ions. *Arch Biochem Biophys*, 126(3), 737-745.

120. Sutter, M., Boehringer, D., Gutmann, S., Gunther, S., Prangishvili, D., Loessner, M.J., Stetter, K.O., Weber-Ban, E. and Ban, N., (2008). Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat Struct Mol Biol*, 15(9), 939-947.

121. Tabita, F.R., (1988). Molecular and cellular regulation of autotrophic carbon dioxide fixation in microorganisms. *Microbiol Rev*, (52), 155-189.

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

123. Tabita, F.R., Hanson, T.E., Li, H., Satagopan, S., Singh, J. and Chan, S., (2007). Function, Structure, and Evolution of the RubisCO-Like Proteins and Their RubisCO Homologs. *Microbiol Mol Biol Rev*, 71(4), 576-599.

124. Tanaka, S., Kerfeld, C.A., Sawaya, M.R., Cai, F., Heinhorst, S., Cannon, G.C. and Yeates, T.O., (2008). Atomic-level models of the bacterial carboxysome shell. *Science*, 319(5866), 1083-1086.

125. Tsai, Y., Sawaya, M.R., Cannon, G.C., Cai, F., Williams, E.B., Heinhorst, S., Kerfeld, C.A. and Yeates, T.O., (2007). Structural analysis of CsoS1A and the protein shell of the *Halothiobacillus neapolitanus* carboxysome. *PLoS Biol*, 5(6), e144.

126. Wackett, L.P., Frias, J.A., Seffernick, J.L., Sukovich, D.J. and Cameron, S.M., (2007). Genomic and biochemical studies demonstrating the absence of an alkane-producing phenotype in *Vibrio furnissii* M1. *Appl Environ Microbiol*, 73, 7192-7198.

127. Williams, E.B., *Identification and Characterization of Protein Interactions in the Carboxysome of Halothiobacillus neapolitanus.*, in *Doctoral dissertation. The University of Southern Mississippi, Hattiesburg, M.S.* 2006.

128. Woodger, F.J., Badger, M.R. and Price, G.D., (2005). 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(4), 1959-1969.

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

130. Yu, D., Ellis, H.M., Lee, E., Jenkins, N.A., Copeland, N.G. and Court, D.L., (2000). An efficient recombination system for chromosome engineering in *Escherichia coli.*. *Proc Natl Acad Sci U S A*, 97(11), 5978-5983.

131. Zeghouf, M., Li, J., Butland, G., Borkowska, A., Canadien, V., Richards, D., Beattie, B., Emili, A. and Greenblatt, J.F., (2004). Sequential Peptide Affinity (SPA) system for the identification of mammalian and bacterial protein complexes. *J Proteome Res*, 3(3), 463-468.