Fall 12-2009

Transcript Analysis of the CSO Operon and Characterization of Two Sets of Conserved Bacterial Microcompartment Genes in the Model Organism *Halothiobacillus neapolitanus*

Fei Cai
University of Southern Mississippi

The University of Southern Mississippi

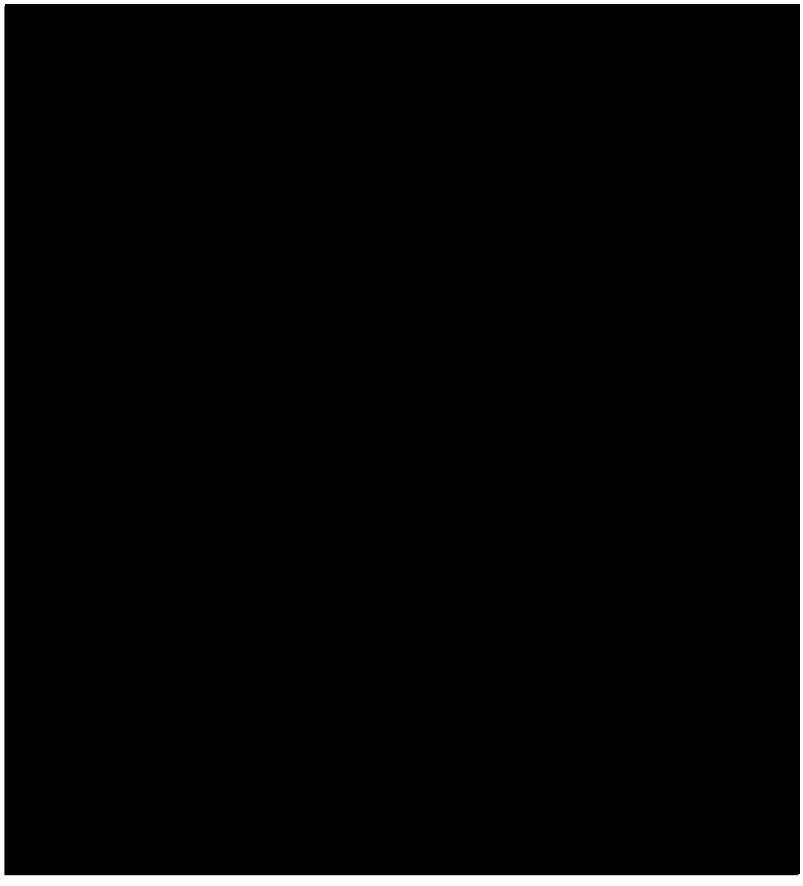
TRANSCRIPT ANALYSIS OF THE *CSO* OPERON AND CHARACTERIZATION OF TWO SETS OF CONSERVED BACTERIAL MICROCOMPARTMENT GENES IN THE MODEL ORGANISM *HALOTHIOBACILLUS NEAPOLITANUS*

by

Fei Cai

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

Approved:



December 2009

The University of Southern Mississippi

TRANSCRIPT ANALYSIS OF THE *CSO* OPERON AND CHARACTERIZATION OF TWO SETS OF CONSERVED BACTERIAL MICROCOMPARTMENT GENES IN THE MODEL ORGANISM *HALOTHIOBACILLUS NEAPOLITANUS*

by

Fei Cai

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

December 2009

ABSTRACT

TRANSCRIPT ANALYSIS OF THE *CSO* OPERON AND CHARACTERIZATION OF TWO SETS OF CONSERVED BACTERIAL MICROCOMPARTMENT GENES IN THE MODEL ORGANISM *HALOTHIOBACILLUS NEAPOLITANUS*

by Fei Cai

December 2009

This study was designed to achieve better understanding of (1) how carboxysome genes are regulated and expressed to yield with precise relative ratios and (2) the *in vivo* roles of two sets of conserved bacterial microcompartment genes, namely the three *csoS1* and two *csoS4* genes of *H. neapolitanus*, in the biogenesis and function of the carboxysome. For the first goal, a detailed transcriptional profile of carboxysomal genes in *H. neapolitanus* was established using absolute quantification real-time RT-PCR and transcript ends analysis. This transcriptional profile revealed that a single promoter, denoted *cso* promoter, was located upstream from the clustered carboxysomal genes. Transcripts of all nine carboxysomal genes were detectable but were present at different levels. *In vivo* activities of the *cso* promoter and selected internal non-coding regions within the carboxysome operon were further examined by using a promoter reporter vector and by generating a *cso* promoter deletion mutant. Both experiments established the *cso* promoter as the only promoter in the carboxysome operon. This result, together with the transcriptional profile of carboxysomal genes, strongly implied that the primary transcript of the *H. neapolitanus cso* operon undergoes differential processing and/or decay and results in different expression levels of individual carboxysomal proteins. To

address the functions of three CsoS1 paralogs, *in vivo* and *in vitro* experiments were performed. Fusing a tetracysteine motif to the C-terminus of CsoS1A confirmed that all three CsoS1 proteins are carboxysome components. A truncated *csoS1B* mutant and a site-directed mutant of *csoS1A* were generated to address the functions of the C-terminal extension of the CsoS1B polypeptide and of the central pore in the CsoS1A hexamer, respectively. An attempt was made to gain better insights into the spatial orientations of CsoS1 proteins and other carboxysome shell components via chemical modification on the carboxysome surface. Finally, single and double knockout mutants were generated for functional study of *csoS4A* and *csoS4B* genes. Phenotypic and kinetic characterization of these mutants revealed that lack of CsoS4 proteins resulted in a more permeable shell which cannot provide the catalytic advantage to ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO).

ACKNOWLEDGEMENTS

My most sincere thanks go to my advisors and mentors, Drs. Gordon C. Cannon and Sabine Heinhorst, for their guidance, encouragement, and support during the development of this study.

I am also grateful to my committee members, Drs. Faqing Huang, Kenneth J. Curry, and Mohamed O. Elasri, for their insights and suggestions.

I would like to thank current and past members of the Cannon/Heinhorst labs – Martha, Scott, Zhicheng, Raj, Kristina, Nicole, Eric, Steve, Andrea, Daniel, Sara, Lakeshia, Sarah, Maria, Jenifer, and Evan. Special thanks to Sarah Neidler who worked with me during the year of 2007 and generated the *cso* promoter deletion mutant.

I would also like to gratefully acknowledge the NSF-MRSEC program and the Department of Chemistry and Biochemistry for providing me financial support throughout my Ph.D. study.

Countless thanks go to my husband, Bing Yu, and my parents, for their love, support, and understanding.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Figure

LIST OF TABLES

Table

LIST OF ABBREVIATIONS

1,2-PD	1,2-propanediol
3-PGA	3-phosphoglycerate
β-ME	β-mercaptoethanol
AP	alkaline phosphatease
APS	ammonium persulfate
BB	bromophenol blue
BCA	bicinchoninic acid
BMC	bacteria micro-compartment
BSA	bovine serum albumin
CA	carbonic anhydrase
CBB Cycle	Calvin-Benson-Bassham Cycle
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CTAB	cetyltrimethylammonium bromide or hexadecyltrimethylammonium bomide
DEPC	diethyl pyrocarbonate
dI	deionized
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
GFP	green fluorescent protein
HCR	high CO2-requiring
HRP	horseradish peroxidase
IAA	isoamyl alcohol
IEF	isoelectric focusing
IPTG	isopropyl-β-D-thiogalactoside
MW	molecular weight
MWCO	molecular weight cut off
MOPS	3-(N-morpholino)propanesulfonic acid
NaAc	sodium acetate

PAGE	polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCR	polymerase chain reactions
PEG	polyethylene glycol
PIPES	piperazine-N,N'-bis-(2-ethanesulfonic acid)
PMSF	phenylmethylsulfonylfluoride
PTSF	p-toluenesulfonylfluoride
RLM-3'RACE	RNA Ligase-Mediated 3' Rapid Amplification of cDNA Ends
RuBisCO	$_{D}$-ribulose-1,5- bisphosphate carboxylase/oxygenase
RuBP	ribulose-1,5-bisphosphate
SCAR	smart cycler additive reagent
SDS	sodium dodecyl sulfate
SNB	sulfo-NHS-biotin
SOB	super optimal broth
SOC	super optimal with catabolite repression
Sulfo-NHS-Biotin	N-hydroxysulfosuccinimidebiotin
TC	tetracysteine
TEM	transmission electron microscopy
TEMB	Tris-EDTA-magnesium-bicarbonate
TEMED	N,N,N,N'-Tetramethylethylenediamine
TEV	Tobacco Etch Virus
Tris	tris-(hydroxymethyl)-aminomethane base
TSS	transcription start site
XC	xylene cyanole

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

1. Introduction

For decades, the absence of intracellular organelles with specific function has been considered as one of the major features of prokaryotic cells and has been used to distinguish prokaryotic cells from eukaryotic cells. The discovery of carboxysome from *Halothiobacillus neapolitanus* [26], a chemoautotrophic bacteria species, brought up the definition of "Prokaryotic Organelles". Unlike eukaryotic organelles, carboxysome is not enclosed by lipid bi-layer membrane, but composed entirely of protein: a 3-4 nm thick protein mono-layer shell filled with CO_2 fixation enzyme $_D$-ribulose-1,5- bisphosphate carboxylase/oxygenase (RuBisCO, EC 4.1.1.39) [13]. Carboxysome is not limited to chemoautotrophic bacteria. It can be found in all cyanobacteria [13]. Further investigation revealed, surprisingly, orthologous genes of some carboxysome shell proteins can also be found in many heterotrophic bacteria [32], or even certain thecate filose amoeba's chromatophore (*Paulinella chromatophora*) [48], and such protein-based micro-compartments might be widely used by bacteria to optimize and/or facilitate particular metabolic reactions.

Carboxysome is the best studied bacteria micro-compartment (BMC), and the investigation of carboxysome is important in several aspects. First, the CO_2 fixation ability of carboxysome-containing bacteria makes them contribute significantly to the global carbon cycle. Thus, the increasing concern about global warming, which is arguably considered mainly caused by the increased CO_2 emissions, makes researchers pay more attention to carboxysomes. Second, carboxysome is the simplest polyhedral

body among all known BMCs, and also the most understood one. Knowledge gained from investigating carboxysome structure, assembly, and function will facilitate the understanding of other BMCs as well. Finally, carboxysomes are nanometer-scaled particles with enzyme activity even after being isolated and this character is likely common to all BMCs. The ability of self-assembling thousands of protein components into one complex bio-reactor makes BMC a good candidate to be engineered as nanometer scale devices with desired biological activities.

2. Carboxysome Composition and Gene Arrangement

Based on the classification of encapsulated RuBisCO, carboxysome composition, and gene arrangement, carboxysomes can be divided into two types: the *cso*-type or α-carboxysomes and the *ccm*-type or β-carboxysomes.

2.1. The cso-type Carboxysomes (α-carboxysomes)

The carboxysomes present in chemoautotrophs and many marine cyanobacteria are α-carboxysoms, which encapsulate Form 1A RuBisCO [3]. Their genes are arranged in a single operon. In the case of model organism, *Halothiobacillus neapolitanus*, these genes are *cbbL*, *cbbS*, *csoS2*, *csoS3*, *csoS4A*, *csoS4B*, *csoS1C*, *csoS1A*, and *csoS1B*, in the direction of transcription [12]. Further investigation of α-carboxysome operon content suggests that there are six types of gene arrangement, as shown in Figure 1. Type (i) is the major type, including *H. neapolitanus*, *Hydrogenovibrio marinus* str. MH-110, *Thiomicrospira crunogena* str. XCL-2, and *Nitrococcus mobilis* Nb-231.

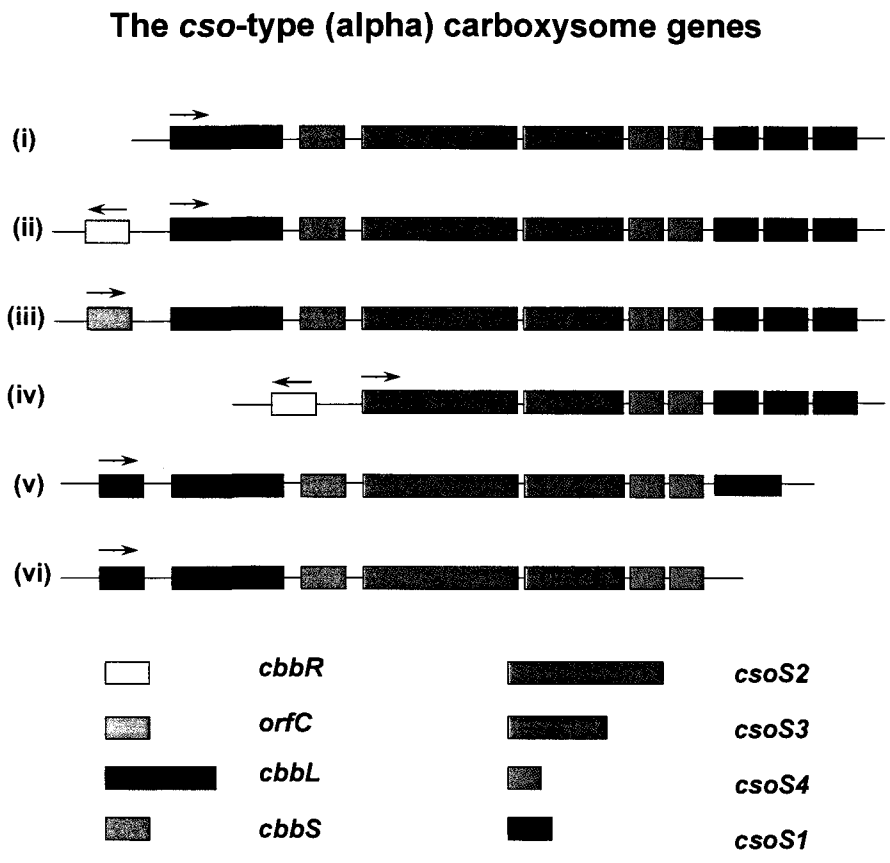


Figure 1. The *cso*-type (alpha) carboxysome genes.

The arrangement of *cso*-type (alpha) carboxysome genes in known species can be divided into six sub-groups. Arrows indicate direction of transcription. The boxes denote the individual genes, and the lengths of each box are proportional to the number of amino acids in the gene products. Different colors represent different homology.

Acidithiobacillus ferrooxidans (formerly *Thiobacillus ferrooxidans*), and *Thiomonas intermedia* (formerly *Thiobacillus intermedium*) have the same gene arrangement as type (i) expect the former one has *cbbR* gene right upstream from the carboxysome operon [49] and, in the case of the latter one, carboxysome operon begins with *orfC*, the homologues of which have not been found in any other carboxysome-containing species [49]. *Thiobacillus denitrificans* is another exception [type (iv)], in which the RuBisCO gene *cbbL* and *cbbS* are missing in the operon. It is not clear, at this point, whether this species forms carboxysome under certain metabolic condition [12]. Unlike the carboxysome-containing chemoautotrophs which have *csoS1* gene(s) at the end of operon (see type i-iv), α-cyanobacteria have the *csoS1* gene at the very beginning of the operon (see type v-vi). Two *Prochlorococcus* species, *Prochlorococcus marinus* str. MED4 (also called *P. marinus* str. CCMP1986) and *P. marinus* str. MIT9303, have gene organization as shown in type (v). Most *Synechococcus* species (including sp. WH5701, WH7803, WH7805, WH8102, RS9916, RS9917, RCC307, CC9311, CC9902, CC9605, and BL107) and two *Prochlorococcus* species, *P. marinus* str. MIT9313 and *P. marinus* str. MIT9211, share the gene organization as shown in type (vi).

2.2. The ccm-type Carboxysomes (β-carboxysomes)

The β-carboxysome contains Form 1B RuBisCO and can be found mostly in freshwater cyanobacteria [3]. Their genes are present in multiple clusters and distribute in whole genome [3,13]. As shown in Figure 2, the major carboxysome gene cluster is *ccmKLMN* in all six sub-groups. *Synechocystis* sp. PCC6803, *Cyanothece* sp. ATCC51142, *Cyanothece* sp. CCY0110, and *Crocosphaera watsonii* WH8501 fall into

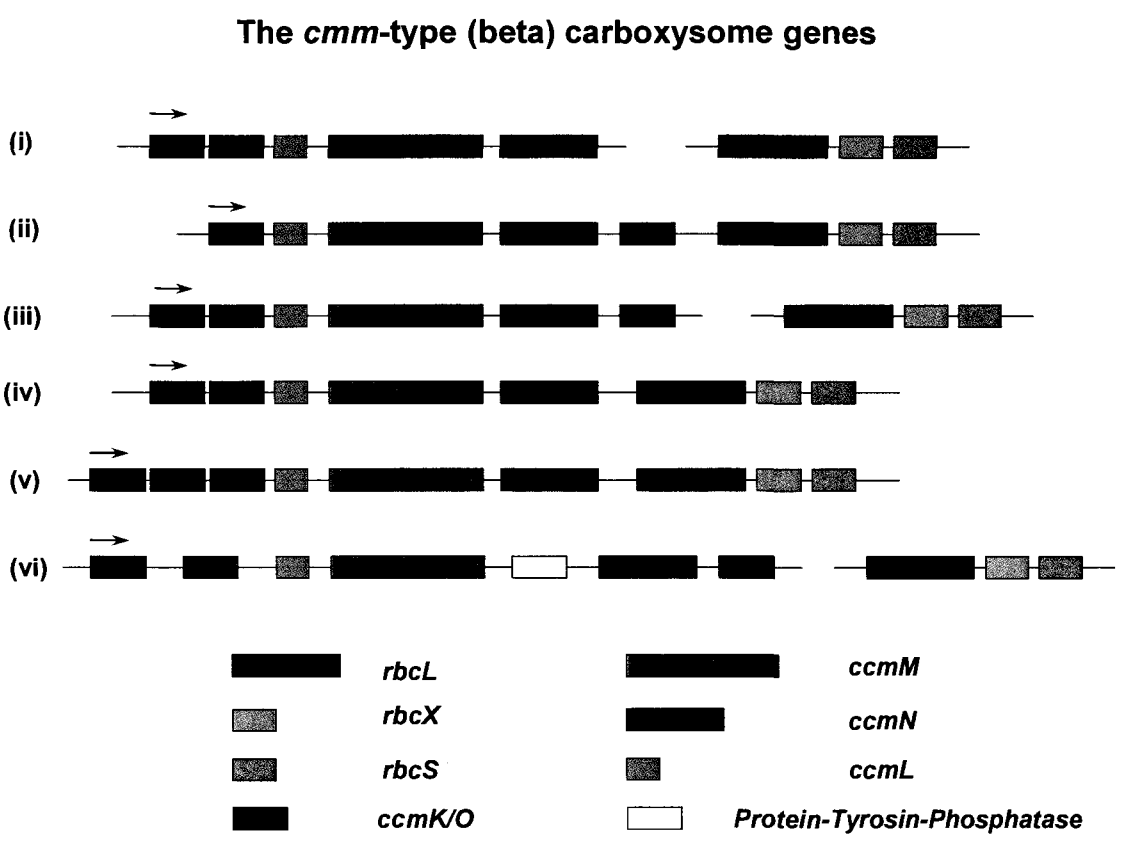


Figure 2. The *ccm*-type (beta) carboxysome genes.

The arrangement of *ccm*-type (beta) carboxysome genes in known species can be divided into six sub-groups. The boxes denote the individual genes, and the lengths of each box are proportional to the number of amino acids in the gene products. Different colors represent different homology. Discontinued line indicates different location in genome.

type (i), in which the RuBisCO gene cluster *rbcLXS* is separated from the carboxysome gene cluster. Type (iii), including *Synechococcus* sp. JA-2-3B'a(2-13), *Synechococcus* sp. JA-3-3Ab, *Gloeobacter violaceus* PCC7421, *Cyanothece* sp. PCC7425, *Nostoc punctiforme* PCC73102, and *Acaryochloris marina* MBIC11017, is similar to type (i), which has *rbcLXS* separated from carboxysome genes, but they contain one more *ccmK/O* gene at the end of *ccmKLMN* cluster. Type (ii), ex. *Synechococcus elongates* PCC7942, type (iv), including *Synechococcus* sp. PCC7002, *Microcystis aeruginosa* NIES-843, *Cyanothece* sp. PCC8801, *Cyanothece* sp. PCC8802, and type (v), including *Cyanothece* sp. PCC7822 and *Cyanothece* sp. PCC7424, have *rbcLXS* immediately following *ccmKLMN* cluster; and the only difference among these three groups is the location and/or the number of *ccmK/O* gene. *Trichodesmium erythraeum* IMS101 is a very special exception, the gene organization of which is similar to type (iii) but there is an open reading frame insertion, annotated as *Protein-Tyrosine-Phosphatase*, between *ccmM* and *ccmN*. Furthermore, the distance between the first two *ccmK* genes is surprisingly large compared with the other groups. Although the literature confirming the presence of carboxysome in *T. erythraeum* IMS101, as a marine cyanobacterium, is still limited, it is supposed to be able to form carboxysome. It will be very interesting to investigate if Protein-Tyrosine-Phosphatase is also a carboxysome component in this organism and how this gene contributes to carboxysome function. Besides the genes listed in Figure 2, in most species there is extra *ccmK/O* copies located at different place in genome, in all six sub-groups.

2.3. Two Conserved Carboxysome Genes

Comparing the carboxysome genes in α- with those in β-carboxysome reveals two conserved carboxysome genes: *csoS1* (α-carboxysome) and its counterpart *ccmK/O* (β-carboxysome), and *csoS4* and the orthologous *ccmL*. *csoS1* and *ccmK/O* contain BMC domain (Pfam ID PF00936), while *csoS4* and *ccmL* belong to "EutN_CcmL" protein family (Pfam ID PF03319). Furthermore, BMC proteins and EutN_CcmL proteins are not limited in carboxysome-containing bacteria, but can be found in all species having bacteria micro-compartment.

3. Carboxysome Shell Proteins

By far, the isolation of homogeneous carboxysome only succeeded in chemolithotrphs *H. neapolitanus* and *N. winogradskyi* (formerly *N. agilis*) [3, 52]. The purified carboxysome of *H. neapolitanus* has been used for a series of biochemical analysis mothods and the carboxysome genes have been assigned to protein compositions of the purified particle via N-terminal amino acid sequencing [3, 18, 26]. As shown in Figure 3, the carboxysomes from *H. neapolitanus* rang from 100 to 120 nm in diameter with a shell of 3-4 nm thickness [63]. The encapsulated RuBisCO can be released via freeze-thaw method. The separation of purified carboxysome and released RuBisCO on SDS-polyacrylaminde gel electrophoresis are shown in Figure 4. The large and small subunits of RubisCO together represent approximately 60-70 wt% of the carboxysome protein and the other proteins are considered as shell components: CsoS2A, CsoS2B, CsoSCA (formerly CsoS3), CsoS1A/C, and CsoS1B [3, 2]. Among the shell proteins, CsoS1A, CsoS1B and CsoS1C together compose over 50% of weight (Table 1) [32].

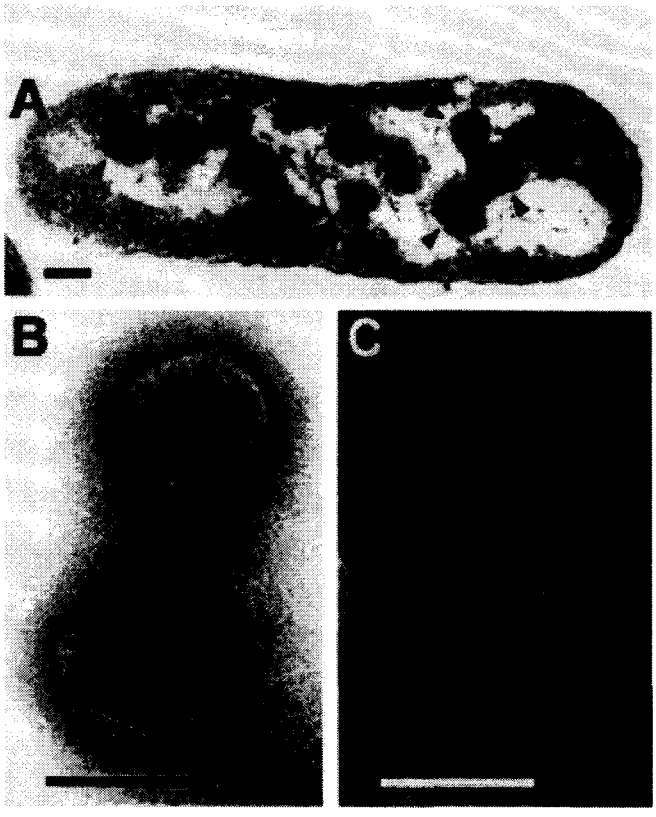


Figure 3. Carboxysomes from *H.neapolitanus*.

(A) Transmission electron micrograph of an *H.neapolitanus* cell containing numerous polyhedral carboxysomes (indicated by arrowheads). (B) Purified, negatively stained intact carboxysomes. (C) Negatively stained carboxysome shells after freeze-thaw treatment. Bars present 100 nm. Cited from Ref [63].

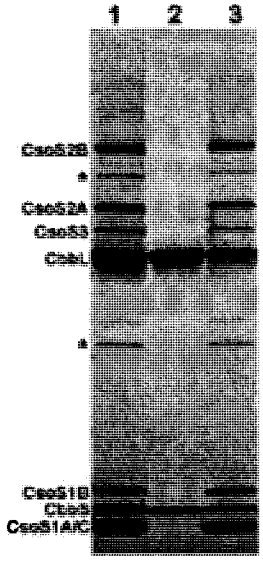


Figure 4. Carboxysome proteins separate by SDS-polyacrylamide gel electrophoresis. Lane 1, intact carboxysomes; lane 2, released RuBisCO separated from broken carboxysomes; lane 3, shell enriched fraction of broken carboxysomes. The identifications are shown on the left. The polypeptide bands marked with *asterisks* are carboxysome protein aggregates that were not disrupted during electrophoresis. Cited from Ref [63].

Table 1. *Halothiobacillus neapolitanus* carboxysome composition. (Cited from Ref [32].)

Protein [a]	M_r theor.[b] (kDa)	M_r app.[b] (kDa)	Wt% [c]	Copies per carboxysome	Theor. pI [d]	Comments
CbbL	52.3	55	70	2160	5.82	270 form 1A RuBisCO per carboxysome
CbbS	12.8	12.6		2160	6.06	
CsoS2A	92	130	6	143	9.06	329 CsoS2 protein per carboxysome
CsoS2B		85	6	186		
CsoSCA	57	64	2.3	81	6.08	40 CsoSCA dimmers per carboxysome ?
OrfA	8.9	--	--	--	5.72	
OrfB	8.8	--	--	--	5.15	
CsoS1C	9.9	6.4	13	2970	5.60	3510 CsoS1 copies per carboxysome
CsoS1A	9.9	6.4			5.58	
CsoS1B	11.3	15	3.7	540	4.86	

[a] Proteins are listed in the order in which their genes are located in the putative *cso* operon

[b] M_r theor. is the molecular mass calculated from the protein's deduced amino acid sequence; M_r app. is its apparent molecular mass based on its SDS-PAGE migration behavior

[c] Wt% is the percent of the carboxysome molecular mass (2.35 × 108 kDa) contributed by individual polypeptides. The values in parentheses were calculated by Cannon and Shively (1983) [3]

[d] The pI was calculated from the protein's deduced amino acid sequence

Considering that their molecular weights are relatively small to other shell proteins, CsoS1A, CsoS1B, and CsoS1C are major shell component based on stoichiometric calculation. CsoS2A and CsoS2B are both encoded by *csoS2* gene. Although the calculated molecular weight is 92 kDa, CsoS2A and Cso2B migrate on SDS-PAGE with observed molecular weight of 85 kDa and 130 kDa, respectively (Table 1). The deduced amino acid sequence of *csoS2* has unique repeating motifs, which might be glycosylation and phosphorylation sites. The calculated pI of *csoS2* is unusually high, 9.06, in comparison with the calculated pI of other carboxysome proteins, from 4.9 to 6.1. These unusual characters might be the cause of the two different product species, but the exact mechanism behind their different migration behavior is not clear. To date, the function of CsoS2 protein is also unclear, however, an intriguing suggestion is that CsoS2 protein might play an important role in transporting RuBisCO substrate ribulose-1,5-bisphosphate (RuBP), the product 3-phosphoglycerate (3-PGA), and/or phosphoglycolate across carboxysome shell [63]. But this assumption is awaiting for precise biochemical analysis. The last identified shell protein, CsoSCA (formerly CsoS3) was reported as a shell-associated carbonic anhydrase. Originally, it was thought to be a new class of CA [63], but a recent crystallographic study showed that it is distantly related to the β class of CA enzymes [69]. This carboxysomal CA is located at interior side of carboxysome shell and dehydrates cytosolic bicarbonate to CO_2, thus supporting efficient carbon fixation activity of RuBisCO. Then far, the remaining two deduced polypeptides, OrfA and OrfB, have not been experimentally identified either in intact carboxysome or in shell enriched fraction.

In the case of β-carboxysomes, the isolation of homogeneous carboxysomes has been

surprisingly difficult. The photosynthetic membranes present in cyanobacteria complicate the purification of carboxysome components from co-purified contaminate proteins [32]. Recently, B. M. Long et al. confirmed the presence of CcmM, CcaA besides RuBisCO large subunit, small subunit, and CcmK protein in the carboxysome-enrich fraction of *Synechococcus* PCC7942 [46]. The CcaA (IcfA) is about 31 kDa and belongs to β-type CA [74]. Although it was reported to be associated with carboxysome in *Synechococcus* PCC7942, it is too early to conclude, at this point, that CcaA is present in all β-carboxysomes. It is also uncertain whether CcmM protein functions as carboxysomal CA in β-carboxysome, as the CsoSCA in α-carboxysome; and there is no significant similarity between CcmM's and CsoSCA's sequences. However, like the *csoSCA* insertion mutant of *H.neapolitanus* and *ccaA* insertion mutant of *Synechococcus* PCC7942, *ccmM* insertion mutant of *Synechococcus* PCC7002 have a high CO_2-requiring phenotype (HRP), which indicates the crucial role of this protein in carbon fixation [316, 39, 318]. To date, there is no CsoS2 protein homologues found in β-carboxysomes.

4. The Carbon Concentrating Mechanism

RuBisCO enzyme catalyzes the first step, also the rate limiting step, in Calvin-Benson-Bassham (CBB) Cycle. Many autotrophic organisms use this cycle for CO_2 fixation and carbon acquisition [76]. However, RuBisCO is an inefficient catalyst, and O_2 is a competing substrate which results in an energy-wasting side reaction:

$$\mathrm{RuBP} \xrightarrow[+\ CO_2]{\mathrm{RuBisCO}} 2\ \mathrm{3\text{-}PGA}$$

$$\mathrm{RuBP} \xrightarrow[+\ O_2]{\mathrm{RuBisCO}} \mathrm{3\text{-}PGA + phosphoglycolate}$$

To survive under limited CO_2 concentrations, autotrophic bacteria (including many chemoautotrophic and cyanobacteria) which rely on RuBisCO for CO_2 fixation developed a strategy called Carbon Concentrating Mechanism (CCM). CCM can be briefly described in Figure 5: inorganic carbon from the environment is first transported across the cell membrane by CO_2 or HCO_3^- transporter; then the HCO_3^- crosses the shell of carboxysome, and is subsequently converted to CO_2 via carbonic anhyrase; the elevated CO_2 concentration inside of carboxysome ensures sufficient substrate for the RuBisCO enzyme. In another word, CCM is composed of two elements: Ci uptake system and carboxysome.

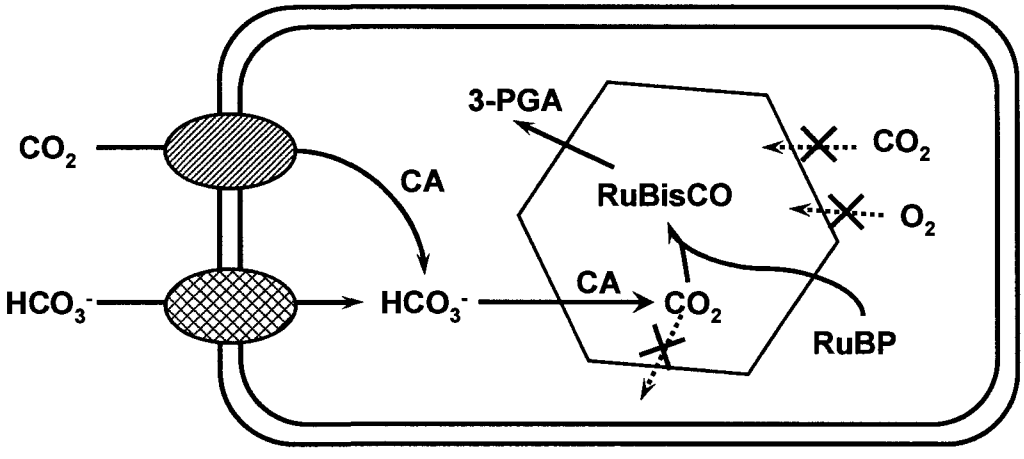


Figure 5. The Carbon Concentrating Mechanism

Oval shapes represent transporter; hexagon represents carboxysome. CA: carbonic anhydrase; RuBisCO: D-ribulose-1,5- bisphosphate carboxylase/oxygenase; RuBP: ribulose-1,5-bisphosphate; 3-PGA: 3-phosphoglycerate.

Previous investigation of Ci uptake system mainly focused on cyanobacteria, especially β-cyanobacteria [17, 61, 99, 102, 174, 98]. To date, a total of five different Ci uptake systems have been identified, including three HCO_3^- and two CO_2 uptake systems

[61]. Some β-cyanobacteria have all five systems, e.g. *Synechocystis* PCC6803, while α-cyanobacteria tend to have fewer Ci uptake systems and many of them lack CO_2 transporter, e.g. oceanic species of *Prochlorococcus* [61]. Among the five Ci uptake systems, BCT1 complex is an inducible high affinity HCO_3^- transporter encoded by *cmpABCD* operon and belongs to traffic ATPase family [56]. SbtA and BicA are both inducible Na^+-dependent HCO_3^- transporters; the former has high affinity and the latter has low affinity but high flux rate [102, 320]. The constitutive expressed complex NDH-I_4 and inducible complex NDH-I_3 are both CO_2 uptake systems. NDH-I_4 is located in the plasma membrane in *Synechococcus elongates* PCC7942 [47,63] while NDH-I_3 of *Synechocystis* PCC6803 was identified on thylakoid membrance [65,88].

On the contrary, little is known about Ci uptake system in carboxysome containing chemoautotrophic bacteria. BLAST searching *chpX* and *chpY* gene, which is a component of NDH-I_4 and NDH-I_3 complex, respectively, against *H. neapolitanus* and *T. intermedia* genome yielded no hit. But it is possible that chemoautotrophic bacteria have their own unique CO_2 uptake system. In the case of HCO_3^- transporter, no homologue of *sbtA* gene was found in these two genomes. Two *cmpA* homologues (*geneor0409* and *geneor1146*) and three *cmpB* homologues (*geneor1147*, *geneor0408,* and *geneor0549*) were present in *H. neapolitanus* genome, while only one *cmpB* homologue (*Tint2376*) and 56 *cmpC/D* homologues were found in *T. intermedia*. However, it's too early to conclude that BCT1 complex can be found in these two species without further investigation. Applying BicA protein sequence from *Synechococcus* sp. PCC 7002 as query sequence for BLAST search resulted single hit in *H. neapolitanus* (*geneor0263*), *T. intermedia* (*Tint1386*), and *T. crunogena* (*Tcr_1533*), with identity/similarity percentage of 30%/48%, 24%/45%, and

27%/45%, respectively. However, *Tcr_1533* was annotated as a sulphate transporter instead of a bicarbonate transporter. Figure 6 clearly showed that three BicA homologues found in chemoautotrophic bacteria are obviously quite distinct from the ones found in cyanobacteria. Since both sulphate transporter and bicarbonate transporter belong to Xan_ur_permease superfamily (Pfam ID: PF00860), assignment of substantial function of these homologues in chemoautotroph requires thorough biochemical analysis.

5. Structural Studies and Sarboxysome Function

Previous electron microscopic studies indicated that carboxysomes exhibited a hexagonal shape, in 2-D, with a diameter of approximately 100-120 nm and a shell thickness of 3-4 nm, in both thin section and isolated sample (Figure 3). However, the symmetry of carboxysome remains unsolved: the carboxysomes of *Nitrobacter agilis* are reported as icosahedral [58], while Holthuijzen et al. believed carboxysomes of *H. neapolitanus* are dodecahedral [32]. Although the icosahedral symmetry is more accepted, in 2006 Schmid and co-workers first argued strongly, with the help of cryo-election tomography, that isolated carboxysome of *H. neapolitanus* are icosahedral in 3-D [70]. They also reported this icosahedral body is slightly thicker at vertices [70]. In negatively stained transmission electron microscopic (TEM) pictures, holoenzyme molecules of RuBisCO can be clearly distinguished from each other. However, the information about how RuBisCO molecules pack in the interior space of carboxysomes remains insufficient. There are reports stating that the RuBisCO fills the entire volume with paracrystalline array [51], while some others suggested that RuBisCO molecules are only arranged in one or a few layers adjoining the inside surface of the shell and the core was hollow [32].

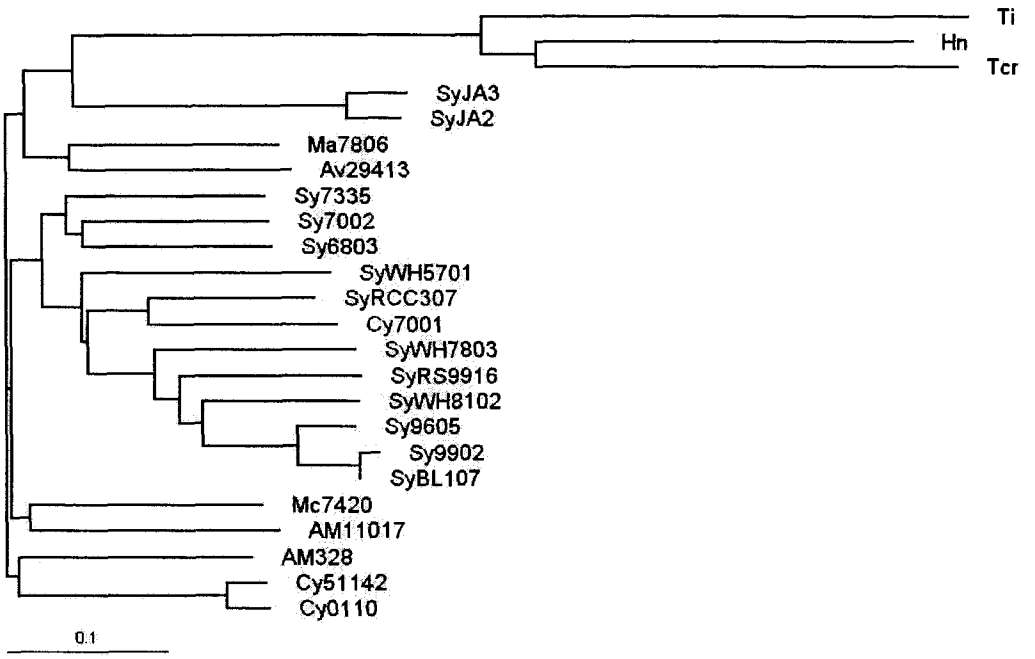


Figure 6. A phylogenetic tree of the BicA homologous

Proteins were aligned using CLUSTALX 1.83 and the dendrogram was generated in the TreeView program. Abbreviations: Ti, *Thiomonas intermedia* K12; Hn, *Halothiobacillus neapolitanus* C2; Tcr, *Thiomicrospira crunogena* XCL-2; SyJA3, *Synechococcus* sp. JA-3-3Ab; SyJA2, *Synechococcus* sp. JA-2-3B'a(2-13); Ma7806, *Microcystis aeruginosa* PCC 7806; Av29413, *Anabaena variabilis* ATCC 29413; Sy7335, *Synechococcus* sp. PCC 7335; Sy7002, *Synechococcus* sp. PCC 7002; Sy6803, *Synechocystis* sp. PCC 6803; SyWH5701, *Synechococcus* sp. WH 5701; SyRCC307, *Synechococcus* sp. RCC307; Cy7001, *Cyanobium* sp. PCC 7001; SyWH7803, *Synechococcus* sp. WH 7803; SyRS9916, *Synechococcus* sp. RS9916; SyWH8102, *Synechococcus* sp. WH 8102; Sy9605, *Synechococcus* sp. CC9605; Sy9902, *Synechococcus* sp. CC9902; SyBL107, *Synechococcus* sp. BL107; Mc7420, *Microcoleus chthonoplastes* PCC 7420; AM11017, *Acaryochloris marina* MBIC11017; AM328, *Arthrospira maxima* CS-328; Cy51142, *Cyanothece* sp. ATCC 51142; Cy0110, *Cyanothece* sp. CCY0110.

Cryo-election tomographic study concluded RubisCO is probably mostly located at the first layer under the shell but also able to populate the entire volume [36,70]. Iancu et al. suggested that packing forces alone could account for this layered organization. But partially filled carboxysomes are also observed [70], and it is not clear if they are just artifacts from preparation.

Crystal structure determination of carboxysome proteins had shed some light on understanding more details of the delicate structure. Monomers of BMC protein from both α-carboxysome (CsoS1A from *H. neapolitanus* [80], Protein Data Bank ID 2G13) and β-carboxysome (CcmK2 and CcmK4 from *Synechocystis* PCC6803 [42], Protein Data Bank ID 2A1B and 2A18, respectively) were revealed to arrange in circular manner and form hexamers with positively charged central pore (Figure 7A). This hexameric arrangement is conserved across all the BMC structures reported and can be closely superimposed to each other [78]. The two sides of the hexamer differ from each other in the distribution of charged amino acid residues and shape (Figure 7B). The concave side shows a strong positive electrostatic potential, while the other side is convex and slightly negatively charged. In all the crystal structures that have been determined, the hexamers pack side by side tightly and form molecular sheets with 2-3 nm thickness. This packing can be either in uniform orientation or in strips of alternative orientation [42,78,80]. The observation that a sulfate ion is found in the pore of CsoS1A hexamer triggered the idea that this central pore might serve as transport channel for small negatively charged molecules to cross the carboxysome shell. This suggestion is consistent with the hypothesis that carboxysome shell functions as a barrier for both CO_2 and O_2, while bicarbonate can be transported across the shell and subsequently converted to CO_2 which

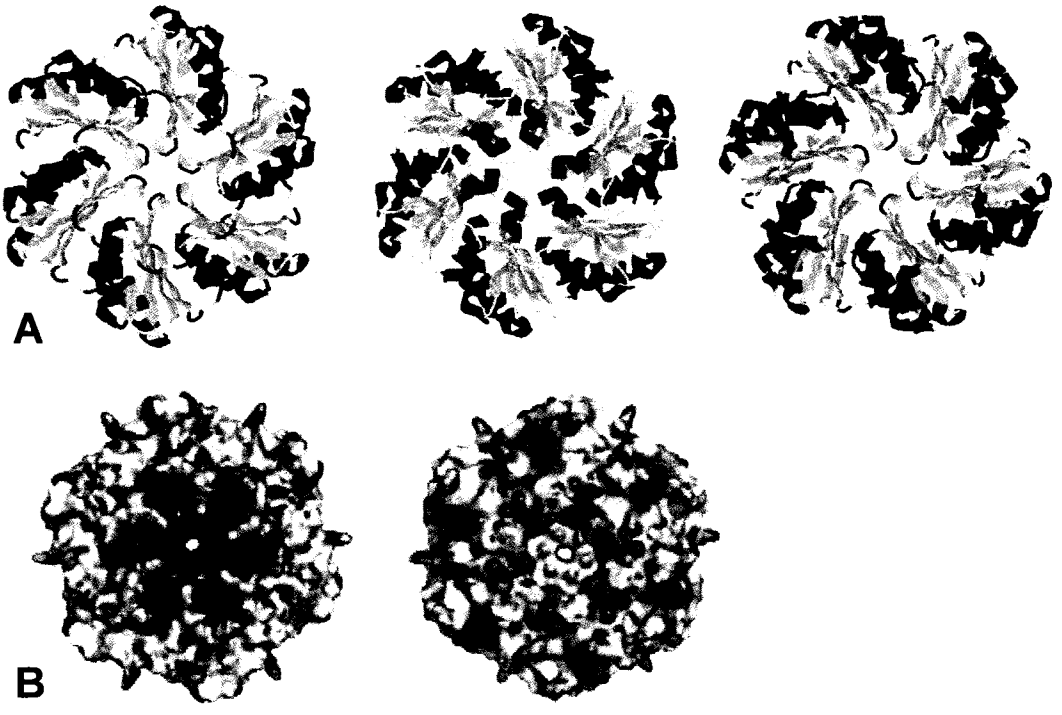


Figure 7. The crystal structure of BMC proteins.

(A) Hexameric structure of CsoS1A from *H. neapolitanus* (left), CcmK2 (middle) and CcmK4 (right) from *Synechocystis* PCC6803. Red helical ribbons represent alpha helices, yellow straight ribbons represent beta strands, and blue indicate turns. (B) Distribution of electrostatic potential on concave (left) and convex (right) sides of CsoS1A hexamer. Blue indicates positive electrostatic potential while red indicates negative electrostatic potential.

is used as substrate for RuBisCO enzyme.

The three-dimensional structures of CsoS4A (formerly OrfA) from *H. neapolitanus* and CcmL from *Synechocystis* PCC6803 have also been published recently [77]. Both CsoS4A and CcmL form pentamers, with one side broader than the other (Figure 8). The size and shape of this pentameric structure make it suitable to be inserted into hexagonally packed sheets of BMC proteins. With these crystal structures, a preliminary assembly model of carboxysome at atomic level has been established [77,86]. In this model, e.g. the carboxysome of *H. neapolitanus*, tightly packed BMC hexamers form the facet of icosahedron, CsoS4A pentamers serve as vertice, and RuBisCO is packed inside (Figure 9). At this point, how other shell proteins integrate into such a tightly packed array is unclear.

6. Regulation of Carboxysome Gene Expression

Like any other metabolic processes, carboxysome gene expression requires precise regulation in response to different environmental conditions. Generally speaking, since the main function of carboxysome is believed to be facilitating CO_2 fixation, it is not surprising that many investigators reported the observation of an increased number of carboxysomes under limited Ci condition [1, 331, 330, 42, 85]. However, the regulation of carboxysome gene expression, or the regulation of CCM system, is not limited to Ci variation.

6.1. Response to External Environmental Changes

For cyanobacteria, the external environmental changes include the change of Ci

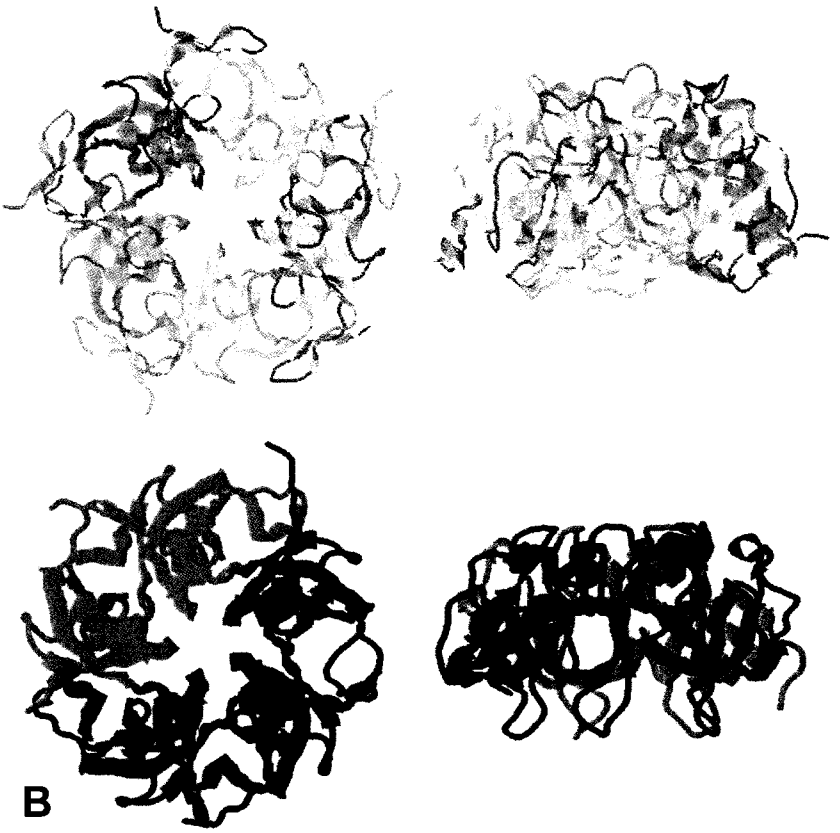


Figure 8. The crystal structure of CcmL and CsoS4A.

(A) Pentameric structure of CcmL from *Synechocystis* PCC6803, top view shown on left and side view shown on right. (B) Pentameric structure of CsoS4A from *H. neapolitanus*, top view shown on left and side view shown on right.

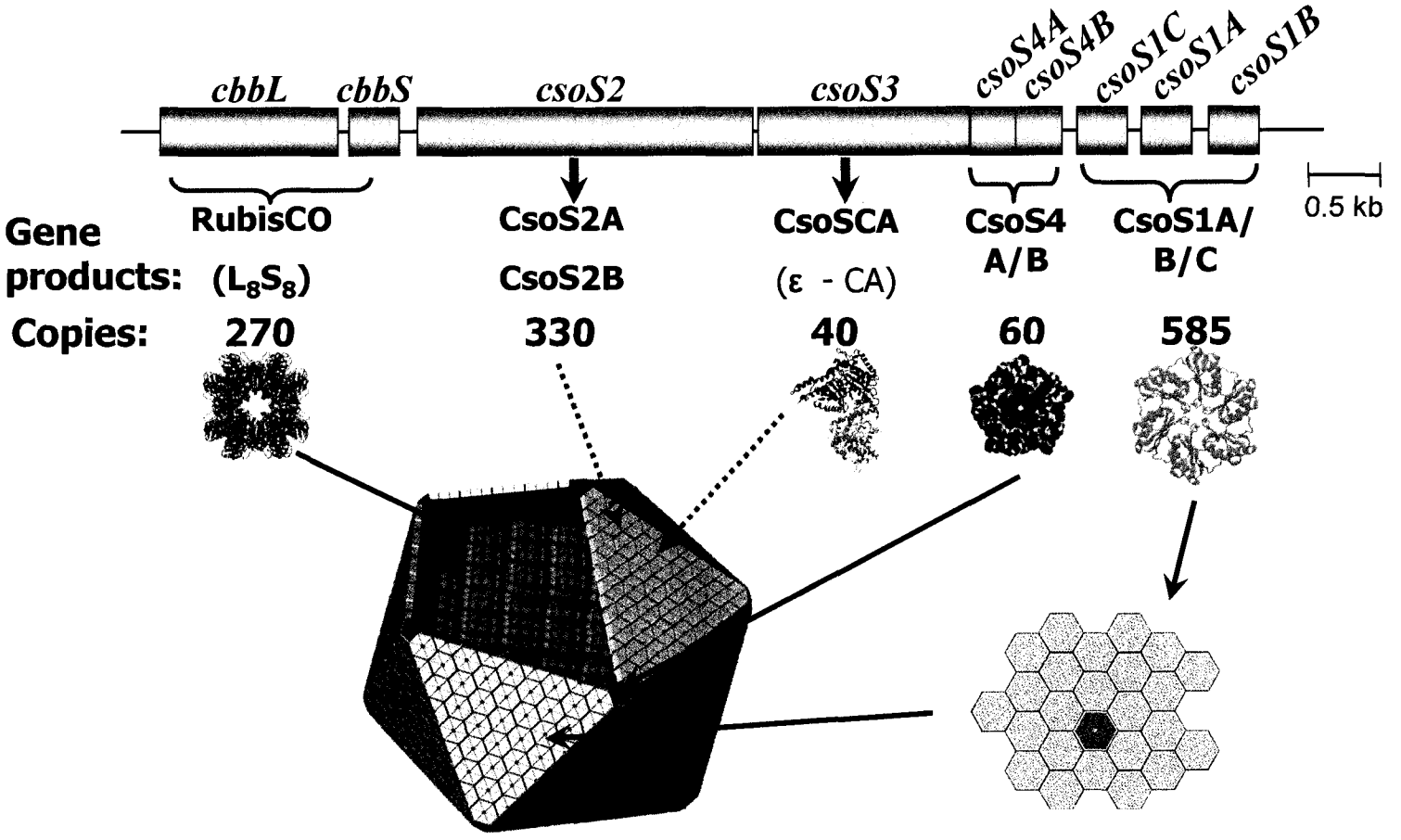


Figure 9. The preliminary atomic model for alpha-carboxysome.

H. neapolitanus was used for the model. Blue boxes and black straight lines represent the carboxysome operon, each gene product is labeled under each gene. Copy numbers indicate calculated oligomer or holoenzyme copies per carboxysome.

concentration and light exposure. To date, the efforts to elucidate these regulatory mechanisms were made mainly within three model organisms: *Hydrogenovibrio marinus* str. MH-110 which produces α-carboxysome and two β-cyanobacteria, *Synechocystis* sp. PCC6803 and *Synechococcus elongates* PCC7942.

H. marinus str. MH-110 is an obligatory lithoautotrophic hydrogen-oxidizing bacterium, which has two sets of form I RubisCO genes (*cbbLS-1* and *cbbLS-2)* and a form II RubisCO gene (*cbbM)*. Both *cbbLS-1* and *cbbM* were followed by *cbbQO* gene and preceded by *cbbR* gene which encodes LysR-type regulator, in the opposite and same directions, respectively. *cbbLS-2* was followed by α-carboxysomal shell genes. Arai's group reported that CbbM was the only RubisCO expressed when this strain was cultivated in 15% CO_2. In 2% CO_2, CbbLS-1 was expressed in addition to CbbM. In the 0.15% CO_2 culture, CbbLS-2 was expressed and the expression of CbbM decreased. In the atmospheric CO_2 concentration, all three sets of RubisCOs were detectable [87]. Also, they indicated that such regulation was at the transcriptional level. In their follow-up work, Arai's group reported that both *cbbR1* and *cbbRm* were required for the activation of *cbbLS-1* and *cbbM* genes, respectively, and either of them was required for the expression of *cbbLS-2* (carboxysomal RubisCO). The expression of *cbbLS-2* increased significantly in the *cbbR1 cbbRm* double mutant. However, direct regulation element for carboxysomal RubisCO and carboxysome shell genes were not identified [79].

Cyanobacterium *Synechocystis* sp. PCC6803 has β-carboxysome genes and all five Ci uptake systems that have been discussed previously. Wang et al. reported that mild Ci limitation (3% to ambient CO_2 concentration) induced a dramatic up-regulation of both CO_2 and HCO_3^- uptake systems [83], while McGinn et al. observed similar up-regulation

at mRNA level when cells were incubated with CO_2-free air under light [51]. However, surprisingly, McGinn et al. reported that carboxysome genes were insensitive to such CO_2 limitation treatments [51]. A similar phenomenon was observed by Omata et al. as well: in *cmpR* mutant, where the expression of *cmp* operon (an inducible high affinity HCO3- transportor) was abolished, the carboxysome operon expression was unaffected under CO_2 limitation [55]. Inverstigators have not come up with a consolidated conclusion on whether high light treatment has an effect on CCM system in *Synechocystis* sp. PCC6803. While McGinn et al. concluded that high light treatment alone could not trigger the expression of CCM-related transcripts [51]. Tu et al. reported that carboxysome genes were slightly down-regulated after high light treatment, while the expression of SbtA, a Na^+-dependent HCO_3^- uptake system, was up-regulated [81]. On the contrary, several independent studies based on DNA microarray profiles suggested that both RubisCO and carboxysome genes are up-regulated after high light treatment [328, 327, 332].

The investigation of carboxysome gene regulation in *Synechococcus elongates* PCC7942 was not as extended as in *Synechocystis* sp. PCC6803. Under CO_2 limitation condition, carboxysome genes and/or *cmp* operon were up-regulated predominantly at transcriptional level [331, 77, 98].

6.2. CbbR: A LysR-type Regulator

LysR-type transcriptional regulator CbbR, the gene of which usually locates upstream from *cbb* (Calvin-Benson-Bassham cycle) operon in the opposite direction, is able to activate the expression of *cbb* related genes in many bacterial species [29], so it is reasonable to assume that CbbR might play an important role in CCM. To date, however,

sulfur bacteria *A. ferrooxidans* and *T. denitrificans* are the only two species in which *cbbR* monolog is found directly upstream from carboxysome operon or *ccmKLMN* cluster. Further analysis is needed to determine whether this *cbbR* gene product is directly involved in carboxysome gene expression in these two species [12]. In many other carboxysome-containing bacteria, *cbbR* homologs have been found located upstream from non-carboxysomal RubisCO genes, for example, in *H. marinus* str. MH-110, as previously mentioned, two copies of *cbbR* homologs are present in genome, but neither of them is required for the activation of carboxysome gene [79]. Apparently, in this case, an alternative mechanism for sensing Ci availability and controlling carboxysome expression must exist.

6.3. Regulation within cso-operon and ccm-gene Cluster

Although the gene organization and subtle composition for α- and β-carboxysome are different, one common characteristic feature is shared by both types: the copy number of each carboxysome protein per carboxysome is not equal [3, 47, 52]. This led to the argument that carboxysome genes must be differentially regulated in both *cso* operon and *ccm* cluster. Several possible strategies might be involved: (i) additional promoters, which can initiate different transcription of downstream gene(s), might exist within the *cso* operon/*ccm* cluster; (ii) RNA secondary structure located at the 3' ends of some gene coding sequences might act as terminator and then cause mRNA premature; (iii) master mRNA might be cut into smaller species with different half lives; (iv) individual gene might have different translation efficiencies due to the different strength of ribosomal binding sites [13].

In the case of *H. neapolitanus*, although a transcription start site for the putative operon has been experimentally determined, the *cbbL* and *cbbS* gene were found co-transcribed on a mRNA approximately 2 kb nucleotides in size while the whole putative operon is over 8 kb nucleotides in size [4]. Additionally, although *cbbS* gene was not expressed in a *cbbL* insertion mutant and functional carboxysomes were not present, three specieses of *csoS1* transcripts with different sizes were detectable in this mutant by northern blotting [18, 60]. These observations implied the presence of an internal promoter within the operon, but experimental evidence is not available yet. The carboxysomal RubisCO genes (*rbcL* and *rbcS*) in four α-cyanobacteria, *Prochlorococcus* MED4, *Prochlorococcus* MIT9313, *Synechococcus* WH7803, and *Synechococcus* WH8102, have also been found co-transcribed, while the evidence about possible additional promoters downstream from the RubisCO genes in these bacteria is still not available [82,84]. On the other hand, transcript analysis related to differential expression of carboxysomal genes in β-cyanobacteria was very limited.

7. Other Bacterial Microcompartments

As previously mentioned, carboxysome is the best understood bacterial microcompartment but not the only BMC. These bacterial proteinaceous polyhedral organelles were also found in heterotrophic bacteria. With the aid of bioinformatic investigation, it is believed that they are involved in at least seven different metabolic processes and the number is still growing [9,18]. In spite of their distinctly different enzyme components and physiological functions, one unifying feature of them is that the genes encoding BMCs must contain two conserved genes: BMC domain-containing gene

and *eutN_ccmL* gene. This feature is also used for the computational prediction of new BMCs.

7.1. 1, 2-propanediol Utilization (pdu) BMC

Salmonella enterica serovar typhimurium LT2 is the first bacteria that is found to form 1, 2-propanediol (1, 2-PD) utilization polyhedral bodies during the growth on 1, 2-PD as a sole carbon and energy source [73]. Later on, the genes for 1, 2-PD degradation and the formation of polyhedral bodies were also found present in the genomes of a number of bacterial genera, including *Salmonella*, *Klebsiella*, *Shigella*, *Yersinia*, and at least one strain of *E. coli* (E24377A) [9]. In *S. enterica*, 1, 2-PD utilization (*pdu*) operon is consisted of 23 putative genes, among which the gene *pduABJKTU* is homologous to *csoS1/ccmK* and the gene *pduN* is homologous to *orfA/B* of *H. neapolitanus* [9]. Havemann and Bobik have reported that the purified Pdu micro-compartment from this bacterius is composed of 14 different major polypeptides, including carboxysome-related proteins PduABB'JKTU. The enzyme components can be further divided into four categories: (i) B_{12}-dependent diol dehydratase (PduCDE); (ii) diol dehydratase reactivase (PduGH); (iii) adenosyl-transferase (PduO); (iv) propionaldehyde dehydrogenase (PduP) [95, 73]. This enzymatic arrangement supports the idea that the first two steps of 1, 2-PD degradation occur within the Pdu micro-compartment, which has been confirmed experimentally [73, 335]. However, the remaining enzymes of the 1, 2-PD degradation pathway are not detected in the purified Pdu micro-compartment [73]. The proposed function of Pdu micro-compartment is to protect cytoplasmic components from aldehyde intermediates produced in the 1, 2-PD degradation pathway, which were known as

cytotoxins. [68].

7.2. B_{12}-dependent Ethanolamine Degradation (eut) BMC

Studies on *E. coli* and *Salmonella typhimurium* have also demonstrated that a similar micro-compartment is involved in ethanolaine utilization (*eut*) [44,57]. The DNA sequence of the *eut* operon in *S. typhimurium* included 17 genes, among which the gene *eutSMLK* is homologous to *csoS1/ccmK* and the gene *eutN* is homologous to *orfA/B* of *H.neapolitanus* [44]. Interestingly, *eutL*, from which the deduced protein sequence is 219 amino acids in length, contains two BMC domains. A function of protecting cells against aldehyde toxicity was also proposed for Eut micro-compartment earlier [94], however, based on the experimental observation that Eut micro-compartment mutants lose acetaldehyde to the environment and are unable to grow due to the loss of carbon, an alternative function was proposed, i.e. it prevents the loss of volatile metabolic intermediates [57].

7.3. BMC Domain

As mentioned previously, BMC proteins are crucial shell components in all known BMCs. The BMC pattern was defined by the ProSite (http://www.expasy.ch/prosite/): D-x(0,1)-M-x-K-[89](2)-x-[IV]-x-[LIVM]-[LIVMA]-[GCSY]-x(4)-[GDE]-[SGPDR]-[G A]. The average length of BMC domain is 76 amino acids and the length of BMC domain-containing proteins vary from 91 to 363 amino acids. Up to year 2008, there were a total of 829 BMC domain-containing proteins banked in Pfam (http://pfam.sanger.ac.uk/ ID PF00936). Among these reported proteins, 652 contain only

one BMC domain at N-/C- terminal or the protein functions as the BMC domain itself, 176 contain two BMC domains, and only 1 contains a BMC domain at N terminal and a EutN_CcmL domain at C-terminus (NCBI protein ID AAO68112, from *Salmonella enterica* subsp. *enterica* serovar Typhi Ty2).

7.4. EutN_CcmL Family

Another conserved gene related to BMC-crossing species has not gained a unified name yet: it is called *orfA/B* in α-carboxysome, *ccmL* in β-carboxysome, *pduN* in *pdu* operon, and *eutN* in *eut* operon. The average length of EutN_CcmL protein is 80 amino acids and, to date, there are a total of 243 members banked in Pfam (http://pfam.sanger.ac.uk/ ID PF03319). Unlike BMC domain genes, the function of this conserved gene has not been assigned yet. However, its product is considered to play an important role in the formation or assembly of BMCs, due to the fact that it is present in all the confirmed or putative BMC operons/clusters.

7.5. Distribution and Functional Diversity of BMC Domain and EutN_CcmL Family

The genes containing BMC domain or belonging to EutN_CcmL family have been found in the Actinobacteria, Bacteroidetes, Proteobacteria, Planctomycetes, Firmicutes, Cyanobacteria, Acidobacteria, Fusobacteria, and Verrucomicrobia, including a total of over 150 species. Such a wide distribution and the fact that only three BMCs have been experimentally confirmed to date implied that related but distinct BMCs with novel functions might exist and await for discovery. Although bioinformatics analysis will facilitate the prediction of new BMCs, experimental evidence is needed to support this

hypothesis. A hollow structure able to accommodate certain metabolic enzyme(s) and its substrate and a protein shell with selective permeability might be a common strategy in the bacterial kingdom.

CHAPTER II

EXPERIMENTAL PROCEDURES

1. Routine Material

Routine chemicals were purchased from VWR, Fisher Science, and Sigma. RNase-free Turbo DNase and RNase-free water were purchased from Ambion Inc. iScript cDNA Synthesis Kit, iQ SYBR Green Supermix, iTaq DNA polymerase, ReadyStrip IPG strip pH 4-7, Bio-Lyte 3/10 ampholyte, 40% Acrylamide/Bis 19:1, 40% Acrylamide/Bis 29:1 (3.3% C), iodoacetamide, IEF pH 4-7 IPG strip, Criterion PreCast 4-20% polyacrylamide gradient Tris-HCl protein gels, and nitrocellulose blotting paper were purchased from BioRad. DNase/RNase-free BioPur 1.5 ml and 0.6 ml tubes were purchased from Eppendorf. 10 ml polystyrene tubes were purchased from BD Falcon. MicroSpin G-25 columns were purchased from GE Healthcare. SuperScript II reverse transcriptase, Zero Blunt TOPO PCR Cloning Kit, and Alexa Fluor 546 were purchased from Invitrogen. Oligonucleotides were ordered from IDT DNA. 0.2 μm filters and 0.45μm filters were purchased from Millipore. Restriction Endonucleases, T4 polynucleotide kinase and T4 RNA ligase were purchased from New England BioLab. CL XPosure Film, Sulfo-NHS-Biotin, HRP conjugated Neutr-Avidin, TEMED, and one step NBT/BCIP reagent were purchased from Pierce. RNasin and MMLV H^- reverse transcriptase were purchased from Promega. RNAprotect Bacteria Reagent, RNeasy Bacteria Mini Kit, QIAprep Miniprep Kit, QIAquick Gel Extraction Kit, and QIAquick PCR purification kit were purchased from QIAGEN Inc. Non-acetylated BSA, trehalose, Tween-20, and Diethyl Pyrocarbonate were purchased from Sigma. GelCodeBlue stain, SuperSignal West Pico Chemiluminescent Substrate, BCA protein assay reagents, and

B-PER II Bacterial Protein Extraction Reagent were purchased from Thermo Scientific. Deoxynucleotide triphosphate (dNTPs) mixture was ordered from TAKARA Bio Inc. Sequenase Version 2.0 DNA Sequencing Kit was purchased from USB Corporation. Zymoclean Gel DNA Recovery Kit, Zyppy Plasmid Miniprep Kit, and ZR Fungal/Bacterial DNA Kit were purchased from ZymoResearch Corporation. Ammonium molybdate stain, 300 mesh formvar/carbon coated copper grids, and Kodak EM film 4489 were purchased from Electron Microscopy Services. Rabbit anti-CsoS1 polyclonal antiserum and rabbit anti-CsoS4 polyclonal antiserum were generated at Cocalico Biologicals. Chicken anti-RbcL antibodies (product No. AS01017) were puchased from Agrisera. Goat anti-chicken and anti-rabbit HRP-conjugated antibodies were ordered from Santa Cruz Biotechnology. Anti-6X-His Epitope Tag rabbit polyclonal antibodies were purchased from Affinity BioReagents.

2. Instrumentation and Equipment

Mass measurements were done using Balance AB54-S or PV303-S from Mettler Toledo. Microliter volumes were measured and pipetted using Rainin Pipetman P-1000, P-200, and P20 pepettes, or a FinnPipette P-5 from Thermo Fisher Scientific. Milliliter volumes were measured and pipetted with a Drummond Pipet-aid. pH measurements were made by using an Orion 8102BNUWP Ross Ultra Combination pH probe connected to an Orion 720A+ pH meter from Termo Electron. Optical density measurements were made in a Beckman Coulter DU800 UV-vis spectrophotometer. DNA or RNA concentration measurements were done using a NanoDrop ND-1000 UV–Vis Spectrophotometer from NanoDrop Technologies. Deionized water was obtained from a

Barnstead NANOpure Diamond water system. Cultures of *E. coli* were grown in a Fisher Scientific Isotemp or an Equaatherm 1572 incubator, a New Brunswick tC-7 culture rotator, or a New Brunswick Series 25 incubator shaker. Liquid cultures of wild type and mutant *H. neapolitanus* were grown in an Infors-HT Multitron shaker or maintained in an Infors-HT labfors two liter fermentor. Solid cultures of mutant *H. neapolitanus* were grown on plates incubated in an Equatherm 1572 or a Fisher Scientific Isotemp Plus D650 incubator. Long term storage of cultures, clones, plasmid constructions, primers, and antiserum was accomplished using a Thermo Scientific Revco Elite series -80°C freezer. Short term storage of plasmids, primers, and antiserum was done using Whirlpool -20°C freezer. Short term storage of purified proteins or cell extracts at 4°C was done by using a Sears or Whirlpool refrigerator. Cell breaking was done using a Branson model 450 sonifier. Large scale high speed centrifugations were carried out in Beckman Avanti J-30I and Avanti J-26 XP centrifuges using the JA-14, JA-20, JA-25.50, JS-24.38, and JSP-F250 rotors. Small scale centrifugations were done in an Eppendorf 5417C and 5417R microcentrifuge. Ultracentrifugation was performed in a Beckman L7-65 ultracentrifuge and a bench-top Beckman Optima Max ultracentrifuge. Nucleic acid or protein samples were heated prior to gel electrophoresis in a VWR Heating blocks Model 12621-088. Polymerase chain reactions for DNA amplification and first strand cDNA synthesis were carried out on a MyCycler thermal cycler from BioRad. Real-time RT-PCR was carried out on a MyiQ single-color real-time cycler from BioRad. DNA restriction digests (37°C), ligation reactions (16°C), and heat-shock step of transformation (42°C) were incubated in Fisher Versa-Bath model 137 or 133 waterbaths. DNA or RNA horizontal gel electrophoresis was done with Horizontal 58 Model 200 Electrophoresis

Gel Box from GibcoBRL or Horizontal Electrophoresis Gel Box Model 75.710 from Biotix Continental Lab Products. DNA vertical gel electrophoresis was carried out with Vertical Gel Electrophoresis System Model V16-2 from BRL Bethesda Research Laboratories, and the resulting gel was dried using BioRad Gel Dryer Model 583 and scanned with BioRad Phosphorimager Molecular Imager FX. Protein gel electrophoresis was performed using a BioRad Criterion or Mini-Protean III cell. Western blotting was done by using BioRad Mini Trans-Blot or Criterion electrophoretic transfer cell. Immunoblots were developed in a GenHunter PerfectWestern immunoblot development box. Visualization and image capturing of nucleic acid or protein gels and immunoblots were done by using BersaDoc Imaging System from BioRad. Concentrating wild type or mutant *H. neapolitanus* cells was carried out with a Millipore Pellicon filtration unit. Sucrose gradient was made in a Gradient former from Bethesda Research Laboratories. Electroporations were performed with using BioRad Gene Pulser. Vigorously mixing solutions or cell resuspensions were done with using a VWR Vortex Genie. Sterilization of media and solutions were done in a Steam sterilizer Amsco Lab 250 or a 3031-S autoclave. Protocols necessitating sterile techniques were carried out in a SterilGuard Class II Type A/B3 hood by Baker Company Inc. Zeiss EM109 electron microscopy was used for electro microscopy. Scanning of negative images was performed by using an Epson Perfection V700 Flatbed Photo scanner. Zeiss LSM 510 confocal laser scanning microscope was used for visualization and image capturing of green fluorescent protein (GFP) containing cells.

Primers for Real-time RT PCR were designed with Beacon Designer 4.0 software by PREMIER Biosoft International. Real-time RT PCR result was recorded and analyzed

with MyiQ software v1.0 by Bio-Rad. Densitometry or annotation of DNA and protein bands were performed with Quantity One v 4.6.3 by BioRad. Generation of DNA vector maps was done using pDRAW32 v 1.1.101 by Acaclone software. DNA and protein sequence alignments was done using ClustalX v 1.83. RNA secondary structure prediction was done by using RNAdraw program v 1.1.

3. Bacterial Strains

Halothiobacillus neapolitanus c2 (ATCC23641), and mutant *H. neapolitanus* strains used in this study: *csoS4A::Km* mutant, *csoS4B::Km* mutant, *csoS4AB::Km* mutant, *csoS1A*-TCtag mutant (*Kan*^{r}), and *csoS1B*-truancated mutant (*Kan*^{r}).

Escherichia coli strains:

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.*

BL21(DE3): *F– ompT hsdS$_B$(r$_B$-, m$_B$-) dcm ga, λ(DE3).*

BL21(DE3) pLysS: *F– ompT hsdS$_B$(r$_B$-, m$_B$-) dcm gal λ(DE3) pLysS (Cm^r).*

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

4. Media and Buffers

(Note: all media and buffers listed below are in alphabetical order)

Agarose gel tracking dye (10 X)

100 mM EDTA (pH 8.0)
50% (v/v) glycerol
1% (w/v) SDS
0.1% (w/v) bromophenol blue

Antibiotic stock solutions

Ampicillin: 25 mg/ml in H_2O, working concentration was 100 μg/ml
Kanamycin: 50 mg/ml in H_2O, working concentration was 50 μg/ml
Antibiotic stock solutions were filter-sterilized by passing through a 0.2 μm filter and stored at -20°C until use.

BEMB buffer, pH 8.0

10 mM Bicine
1 mM EDTA
10 mM $MgCl_2$
20 mM $NaHCO_3$
PH was adjusted to 8.0 with 1 M NaOH as needed.

Colony digestion buffer

0.2% (v/v) Trion X-100
0.1 mg/ml RNase A
Colony digestion buffer was stored at 4°C.

CTAB/NaCl solution

0.7 M NaCl
10% (w/v) CTAB
NaCl was first dissolved in 0.8 volume of H_2O. Then CTAB was added slowly while heating to 65°C and stirring.

Dialysis tubing preparation

Dialysis tubing was boiled in 2% (w/v) sodium bicarbonate and 10 – 50 mM EDTA (pH 8.0) for at least 10 min. The tubing was stored at 4°C in 50% ethanol with 1 mM EDTA and 0.02% (w/v) sodium azide.

Diethyl pyrocarbonate (DEPC) treatment of liquid solution used for RNA analysis

1 mg/L DEPC was added to liquid solutions and mixed rigorously. Treated solutions were set overnight, then autoclaved for 20 min at 121°C and 15 psi.

8% Denaturing polyacrylamide gel electrophoresis (PAGE) mix

7 M Urea
1X TBE buffer
20% (v/v) 40% Acrylamide/Bis 19:1
Store at 4°C
To make a gel, 0.01 gel volume of 10% (w/v) ammonium persulfate (APS) and 0.001 gel volume of N,N,N,N'-Tetramethylethylenediamine (TEMED) were added right before pouring the gel.

***Halothiobacillus neapolitanus* liquid growth 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% (w/v) phenol red solution in 50% ethanol was added as pH indicator when needed. pH was adjusted to 6.8 with 1 M KOH if needed. The medium was autoclaved for 20 min at 121°C and 15 psi. Sterilization time was increased to 90 min for batches of 8 L media.

***H. neapolitanus* low salt solid medium**

0.2 g/L KH_2PO_4
0.8 g/L K_2HPO_4
0.01 g/L NH_4Cl
0.24 g/L $MgSO_4$
10 g/L $Na_2S_2O_3 \cdot 5H_2O$
1 ml/L trace element solution
15 g/L agar
4 ml/L 1% (w/v) phenol red solution in 50% Ethanol
pH was adjusted to 6.8 with 1 M KOH if needed. The medium was autoclaved for 20 min at 121°C and 15 psi.

***H. neapolitanus* trace element solution (100 X)**

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 first dissolved in H_2O, and pH was adjusted to 6.0 with 20% (w/v) KOH. Then the rest chemicals were added sequentially as listed above, and pH was adjusted to 6.0 with 20% (w/v) KOH after each step. Final pH was adjusted to 6.8, and the solution was parafilmed and stored at 4°C prior to use.

Glycerol-tolerant PAGE buffer (20 X)

1.78 M Tris base
575 mM Taurine
8.6 mM EDTA (pH 8.0)

Luria-Bertani (LB) broth

10 g/L tryptone
5 g/L yeast extract
10 g/L NaCl
For solid LB medium, 1.5% (w/v) agar was added.
Medium was autoclaved for 20 min at 121°C and 15 psi.

Nucleic acid denaturing PAGE running dye (2 X)

7 M urea
1X TBE buffer
0.05% (w/v) bromophenol blue (BB)
0.05% (w/v) xylene cyanole (XC)

Overlay agarose solution

1% (w/v) low-melting point agarose
1X SDS-PAGE loading buffer
1X SDS-PAGE running buffer
This is solid buffer with blue color at room temperature. It should be microwaved till turning to clear solution prior use.

Phosphate-buffered saline (PBS), pH 7.4 (10 X)

100 mM Na_2HPO_4
20 mM KH_2PO_4
1.37 M NaCl
27 mM KCl
pH was adjusted to 7.4 with 1 M NaOH if needed.

PMSF/PTSF stock solution

100 mM phenylmethylsulfonylfluoride
100 mM p-toluenesulfonylfluoride
This solution was made with 100% ethanol instead of H_2O.

pProEX lysis buffer, pH 8.0

50 mM Tris-HCl (pH 8.0)
5 mM β-ME
1 mM PMSF/PTSF

pProEX wash buffer A / equilibration buffer

20 mM Tris-HCl (pH 8.0)
100 mM KCl
5 mM β-ME
10% (v/v) glycerol
20 mM imidazole

pProEX wash buffer B

20 mM Tris-HCl (pH 8.0)
100 mM KCl
5 mM β-ME
10% (v/v) glycerol

pProEX elution buffer

20 mM Tris-HCl (pH 8.0)
100 mM KCl
5 mM β-ME
10% (v/v) glycerol
100 - 250 mM imidazole

Rehydration buffer for IPG strip 3-10 or 4-7 pH range (1 X)

8 M Urea
2% (w/v) 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS)
30 mM dithiothreitol (DTT)
0.2% (v/v) Bio-Lyte 3/10 ampholyte
This buffer can be prepared without DTT and Bio-Lyte 3/10 ampholyte first and kept at room temperature. DTT and Bio-Lyte 3/10 ampholyte should only be added immediately prior use.

Rich broth

10 g/L Tryptone
5 g/L Yeast extract
5 g/L NaCl
2 ml/L 1 M NaOH
This medium was autoclaved for 20 min at 121°C and 15 psi.

SDS (sodium dodecyl sulfate)-PAGE equilibration buffer I and II

6 M Urea
375 mM Tris-HCl (pH 8.8)
2% (w/v) SDS
20% (v/v) glycerol
Above chemicals can be prepared in H_2O first, then buffer I was made by adding 2% (w/v) DTT, and buffer II was made by adding 2.5% (w/v) iodoacetamide. Both buffer I and II should be prepared fresh prior use.

SDS-PAGE loading buffer (4 X)

200 mM Tris-HCl (pH 8.3)
8% (w/v) sodium dodecyl sulfate (SDS)
0.4% (w/v) bromophenol blue
40% (w/v) glycerol
400 mM DTT or β-mercaptoethanol (β-ME)

SDS-PAGE running buffer (10 X)

250 mM Tris-HCl (pH 8.3)
2.50 M glycine
10% (w/v) SDS

Smart cycler additive reagent (SCAR) (5 X)

1 mM Tris-HCl, pH 8.0
1 mg/ml non-acetylated BSA
0.75 M trehalose
1% Tween-20
This reagent can be stored at 4°C for up to one week, and at -20°C for long-term storage. Do not freeze/thaw repeatedly more than 10 times. SCAR is an additive developed by Cepheid to enhance product yield and sensitivity of PCR.

Super optimal broth (SOB) medium

20 g/L Tryptone
5.0 g/L Yeast extract

0.5 g/L NaCl
This medium was autoclaved for 20 min at 121°C and 15 psi, then 0.01 volume of filter sterilized 1 M $MgCl_2$ and 0.01 volume of filter sterilized 1 M $MgSO_4$ were added to the medium.

Super optimal with catabolite repression (SOC) medium

To make 100 ml SOC medium, add 1 ml of filter sterilized 2 M glucose or 2 ml of 20% (w/v) glucose to SOB medium to a final volume of 100 ml. This medium should be prepared immediately before use.

Sodium azide stock solution

1% (w/v) in H_2O
Final working concentration was 0.02%.

TAE buffer (50 X)

2 M Tris-HCl (pH 8.3)
54.4 ml/L glacial acetic acid
50 mM EDTA

TBE buffer (10 X)

108 g/L Tris base
55 g/L Boric acid
20 mM EDTA

TE buffer

10 mM Tris-HCl
1 mM EDTA, pH 8.0

TEMB buffer

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

Terrific broth for low copy number plasmid

12 g/L Tryptone
24 g/L Yeast extract
4 ml/L glycerol
2.31 g/L KH_2PO_4

12.54 g/L K_2HPO_4
This medium was autoclaved for 20 min at 121°C and 15 psi.

TFB I

30 mM Potassium acetate
100 mM RbCl
10 mM $CaCl_2$
50 mM $MnCl_2$
15% (v/v) glycerol
pH was adjusted to 5.8 with acetic acid. The solution was then filter-sterilized.

TFB II

10 mM MOPS or PIPES
75 mM $CaCl_2$
10 mM RbCl
15% (v/v) glycerol
pH was adjusted to 6.5 with 1 M KOH. The solution was then filter-sterilized.

8% Sequencing PAGE with formamide

7 M Urea
1X Glycerol-tolerant PAGE buffer
20% (v/v) 40% Acrylamide/Bis 19:1
40% (v/v) formamide
This solution was stored where temperature is greater than 28°C.
To make a gel, 0.01 gel volume of 10% (w/v) APS and 0.001 gel volume of TEMED were added immediately before pouring the gel.

Western blot blocking buffer

5% non-fat milk in 1X PBS with 0.1% (v/v) Triton X-100
This buffer was stored at 4°C.

Western blot transfer buffer

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

5. Methods

H. neapolitanus growth

Wild type *H. neapolitanus* str. c2 (ATCC 23641) and the *csoS4AB::Km* mutant were grown at 30°C in a 2 L chemostat in *H. neapolitanus* medium with a dilution rate of 0.05 – 0.08 h^{-1}. Stationary phase cells were collected in an 8 L carboy connected to the bioreactor. Other *H. neapolitanus* mutants were grown as 750 ml batch cultures in 2 L baffled flasks at 30°C in a shaker incubator (200 rpm) with 5% CO_2 supplement.

Isolation of genomic DNA from *H. neapolitanus*

Mini-scale

For sequencing of specific genomic regions, a 50 ml culture was grown to saturation and harvested by centrifuging at 9,000 rpm in rotor JA25.50 for 15 min at 4°C. A Zymo Fungal/Bacterial DNA kit was used for purifying genomic DNA. The manufacturer's protocol was followed with one modification: an additional 1 min spin at 10,000 rpm, in a fresh 1.5 ml tube, was performed immediately before adding 50°C H_2O for elution.

Large-scale

A 750 ml culture was grown to saturation, and cells were harvested by centrifuging at 10,000 rpm in rotor JLA16.25 for 15 min at 4°C. The cell pellet was resuspended in 6 ml TE buffer, and the OD_{600} was recorded. TE buffer was used to dilute the cell suspension to a OD_{600} of around 1.0. Usually the final volume of cell suspension was approximately 20 ml. Freshly prepared 100 mg/ml lysozyme was added (400 µl), and the cell suspension was incubated at room temperature for 5 min. Then 800 µl of a 10% SDS solution, and 160 µl of 10 mg/ml Proteinase K were added and mixed well, and the cell suspension was

incubated for 1 h at 37°C. To this slurry, 2 ml of 5 M NaCl was added, then 2 ml of CTAB/NaCl (heated to 65°C) was added. The mixture was incubated at 65°C for 10 min. An equal volume of phenol:chloroform:isoamyl alcohol (IAA) (25:24:1) was added and mixed by shaking vigorously. The mixture was transferred into a 45 ml glass centrifuge tube and centrifuged at 10,000 rpm in rotor JA25.50 for 20 min at 20° C. The aqueous upper phase was carefully transferred to a new centrifuge tube and extracted again with chloroform:IAA (24:1) as described earlier. The aqueous phase was transferred to a 15 ml glass centrifuge tube, 0.1 volume of 3 M NaAc (pH 5.2) and 0.6 volume of cold (-20°C) isopropanol were added. The mixture was shaken gently until a white DNA precipitate formed. The resulting DNA precipitate was transferred carefully with transfer pipette to a new 15 ml centrifuge tube, and the chloroform:IAA (24:1) extraction and iso-propanol precipitation steps were repeated. The final DNA pellet obtained was dried under a stream of N_2, and allowed to resuspend in 3 ml TE buffer overnight. The next day, 30 µl of 10 mg/ml RNase A was added to the DNA solution, and the mixture was incubated at 37°C for 20 min. Then 0.1 volume of 3 M NaAc (pH 5.2) and 2 volumes of cold (-20°C) 100% ethanol were added and mixed well. The precipitated DNA was pelleted by centrifuging at 10,000 rpm for 15 min at 4°C in a bench-top centrifuge. The resulting DNA pellet was again dried under a stream of N_2, and resuspended in 2 ml TE buffer.

Mini-scale plasmid DNA preparation

A 3 ml or 5 ml for low-copy number plasmids *E. coli* overnight culture was pelleted by centrifugation in a microcentrifuge at 14,000 rpm for 1 min. The cell pellet was resuspended in 600 µl H_2O. And the cell suspension applied to a Zyppy Plasmid Miniprep

Kit according to the manual. Usually, each preparation yielded 3 to 5 μg of plasmid DNA.

Restriction digest of plasmid DNA

For diagnostic purposes, plasmid DNA was digested with the appropriate restriction endonucleases in a total reaction volume of 10 μl. Typically, the reaction consisted of 0.5-1.0 μg plasmid DNA, 1 μl of the appropriate 10 X digestion buffer, 1 μl of bovine serum albumin (10 mg/ml) if needed, 10 units of each restriction endonuclease. The reactions were incubated in a 37° C water bath for 2 h. If a DNA fragment needed to be recovered for ligation, the digest reaction was scaled up to 50 μl with 3-5 μg of plasmid DNA. The resulting DNA fragments were separated by agarose gel electrophoresis for visualization and/or recovery.

Colony digestion [20]

Single colonies were picked up with pipette tips and resuspended in 12 μl colony digestion buffer in a 0.6 ml microfuge tube. The samples were heated at 100° C for 15 s, and allowed to cool to room temperature. The appropriate 10X restriction enzyme buffer (1.3 μl) was added, followed by 0.5 μl of each restriction enzyme (5-10 units). The digestion was carried out at 37° C for 30 min before separation of the DNA fragments by agarose gel electrophoresis.

DNA ligation

DNA ligation reactions were set up with a vector:insert molar ratio of 1:3 or 1:5. The

total reaction volume of 10 µl contained 100 ng of vector DNA, the corresponding amount of insert DNA, 1 µl of 10X T4 DNA ligase buffer and 400-800 units of T4 DNA ligase. Reactions were incubated in a 16° C water bath for 20 h. A 5 µl aliquot of the reaction was used to transform chemically competent *E. coli* cells.

The Zero Blunt TOPO PCR Cloning Kit was used to ligate PCR products into the pCR-Blunt II TOPO vector. PCR products were either used directly, or purified with one of following methods: QIAquick PCR purification kit, gel recovery, or ethanol precipitation. Typically, 4 µl PCR product, 1 µl of salt solution and 1 µl TOPO vector were mixed and incubated for 30 min at room temperature. The reaction was used either immediately to transform chemically competent *E. coli* TOP10 cells or stored at -20° C overnight before transformation.

Polymerase chain reaction (PCR)

Typically, 20 µl polymerase chain reactions were set up in 0.2 ml thin-walled PCR tubes. Each reaction contained 1X Taq or Pfu buffer, 250 µM of each dNTP, 200 nM each of forward and reverse primer, 10 ng of plasmid DNA or 100 ng of genomic DNA as template, and 2 units of Taq DNA polymerase or Pfu Ultra II DNA polymerase. SCAR reagent was added if necessary. The thermal cycler was pre-heated to 95° C before inserting the PCR tubes. A typical PCR protocol consisted of a initial denaturation step at 95° C for 5 min, 35 repeat cycles of denaturation (95°C for 20 s), annealing (5° C below the melting temperature of the primers, 20 s), and extension (72° C for 30-120 s, depending on the size of the region to be amplified), a final extension step at 72° C for 5 min, and a hold step at 4°C. If necessary, a temperature gradient was applied for the

annealing step.

Agarose gel electrophoresis, nucleic acid visualization, and recovery

DNA or RNA samples were prepared in 10X agarose gel tracking dye and separated by electrophoresis on a agarose gel prepared in TAE or TBE buffer. The choice of percentage of the gel (0.8-2% (w/v)) depended on the size of the nucleic acid to be separated. Electrophoresis was performed at 5 V/cm until the bromophenol blue dye had migrated half the length of the gel. The gel was stained with 0.5 μg/ml ethidium bromide in H_2O for 12 min and then rinsed in H_2O for 30 min. The stained gel was place on a UV transilluminator for visualization. Images were taken with a VersaDoc imaging system and Quantity One v 4.6.3 software. Stained bands of interest were excised from the gel with clean razor blade. Gel slices were weighed and further processed using a Zymoclean Gel DNA Recovery kit.

Nucleic acid denaturing polyacrylamide gel electrophoresis

Two glass plates were assembled with 0.4 mm spacers, an 8% denaturing PAGE mix was prepared and injected between the plates with a 60 ml syringe. A comb was inserted and the assembled gel unit set horizontally to allow the polyacrylamide gel to solidify. The gel was pre-run in TBE buffer at 500 V for 20 min. Typically, 2 μl sample mixed with 2 μl 2X loading buffer were loaded into the wells, and 500 V was applied until the BB dye had migrated approximately half of the gel length. After completion of electrophoresis, one of the two glass plates was removed, and plastic wrap was used to cover the gel; then the other glass plates was removed and the gel was covered with a

piece of whatman filter paper of appropriate size. The gel was dried on a gel dryer for 8-10 min and then exposed to Kodak radioactive film for 0.5 – 2 min. The film was finally scanned with Molecular Imager FX.

Oligomer purification from denaturing polyacrylamide gel

Single stranded DNA oligomers were resuspended in the appropriate volume of H_2O to yield 100 μM concentration. Then 10 μl of oligomer were mixed with 10 μl of 2X loading buffer and applied to an 8% denaturing polyacrylamide gel. Gel electrophoresis conditions were as described. After electrophoresis, one glass plate was removed and replaced with plastic wrap. Short wavelength UV light was used to visualize the oligomer in the gel, and the position and shape of the oligomer shadow were marked with permanent marker on the plastic wrap. The gel slice was excited corresponding to the location of the oligomer with a clean razor blade, crushed in a 1.5 ml tube, and 50 μl of H_2O was added. The gel suspension was heated at 80° C for 5 min, and vortexed briefly every min before the gel pieces were pelleted in a microcentrifuge at 14,000 rpm for 5 min. The supernatant contained the purified oligomer.

DNA/RNA precipitation

Precipitation of DNA or RNA required adding of 3 volumes of 0.5 M sodium acetate (NaAc) (pH 5.2) and 7 volumes of ethanol. After mixing and incubation for at least 30 min at -20° C, the sample was centrifuged for 10 min at 14,000 rpm at 4°C. The supernatant was removed and nucleic acid pellet was dissolved in the solvent of choice.

Preparation of chemically competent *E. coli* cells

A single colony from the agar plate was used as an inoculation for 3 ml RB medium in a 10 ml culture tube and the culture was shaken overnight at 225 rpm at 37°C. The next day subculture the overnight culture with 1:100 dilution in 250 ml RB medium which contains 2 mM $MgSO_4$. Culture was continuously incubated until OD^{600} reading reach 0.4-0.6. Cells were harvested by centrifuge at 5 k rpm for 5 min at 4°C. Resulting cell pellet was gently resuspended in 100 ml ice-cold TFB I buffer. Keep all steps on ice from this point on. Cell resuspension was incubated on ice for 5 min, then pelleted down by centrifuging at 5 k rpm for 5 min at 4°C. Resulting pellet was then gently resuspended in 10 ml ice-cold TFB II buffer, and incubated on ice for 60 min. Assemble 1.5 tubes with lid open in rack and put in liquid nitrogen bath, distribute the cells and then close the tubes. After the tubes are well-frozen, transfer the tubes and store in -80°C freezer. These competent cells last at least a year.

Chemical transformation of *E. coli* cells

10 ng of plasmid or 5 µl of ligation reaction was mixed with 25 µl commercial purchased or 50 µl home-prepared chemically competent *E. coli* cells in pre-cold 10 ml round-bottom polystyrene tube, and incubated on ice for 60 min. Cells were heat-shocked in a 42° C water bath for exactly 30 s and then immediately kept on ice for 5 min. A 250 µl aliquot of room temperature SOC medium or RB medium was added to the cells and the suspension was incubated in shaker at 225 rpm for 1 h at 37° C. 50-100 µl of cells were plated on LB plate containing appropriate antibiotic according to the resistance marker of the plasmid. Plates were incubated at 37°C overnight.

Electroporation of *E.coli* DY330 cells

A saturated overnight DY330 culture was used as a 1:50 inoculation to generate 50 ml subculture, and the subculture was incubated in a 30° C shaker at 225 rpm until the OD_{600} reading reached 0.5. 25 ml of this culture were transferred to a sterilized 250 ml baffled flask and incubated in a 42° C water bath with constant shaking for 15 min, while the remaining 25 ml culture was allowed to continue growing at 30°C as uninduced control. Both flasks were then transferred to an ice-water slurry, shaken to cool, and incubated for an extra 15 min. Cells were harvested by centrifuging at 4,000 rpm for 10 min at 4° C, and then resuspended in 1 ml ice-cold sterilized H_2O. The cell suspension was transferred to a 1.5 ml tube and centrifuged at 14,000 rpm for 20 s at 4°C. The supernatant was discarded, and the cell pellet was washed with 1 ml ice-cold sterilized H_2O. The wash step was repeated two more times. The final cell pellet was resuspended in 200 µl ice-cold sterilized H_2O and kept on ice until use. For electroporation, 100 µl of induced or uninduced cells were mixed with 1 µg each of linear donor DNA and plasmid acceptor DNA and then transferred into a 0.1 cm gap pre-cooled electroporation cuvette. Cells were electroporated at 2.0 kV, 25 µF with the pulse controller set at 200 Ohms and then immediately allowed to recover in 1 ml LB medium on ice for 5 min before incubation in a 30° C shaker for 1.5 h. A 200 µl aliquot of cells was plated on LB agar containing the appropriate antibiotic. Plates were incubated at 30° C overnight.

Electroporation of *H. neapolitanus* cells

A 40 ml aliquot of exponentially growing *H. neapolitanus* cells was collected from the chemostat and harvested by centrifuging at 10,000 rpm in rotor JA25.50 for 10 min at

4° C. The cell pellet was resuspended in 40 ml of ice-cold sterilized H_2O and centrifuged at 10,000 rpm in rotor JA25.50 for 10 min at 4° C. The supernatant was discarded and the cell pellet was washed and centrifuged two more times, and the final resulting pellet was resuspended in 200 µl ice-cold sterilized H_2O and kept on ice until use. For electroporation, 60 µl of cells were mixed with 1 µg plasmid DNA (or H_2O as negative control), and transferred into a 0.1 cm gap pre-cooled electroporation cuvette. Cells were electroporated at 1.9 kV, 25 µF with the pulse controller set at 200 Ohms, and then immediately allowed to recover in 5 ml ice-cold *H. neapolitanus* medium on ice for 5 min before incubation in a 30°C shaker in air supplemented with 5% CO_2. Cells were allowed to recover for 24 h before adding kanamycin to 50 µg/ml. The cultures were incubated for 4 to 5 days in CO_2-supplemented air until a drop in pH indicated of growth. of this culture, 100 µl were spread on *H. neapolitanus* low salt solid medium containing 50 µg/ml kanamycin. Plates were incubated at 30° C until single colonies were visible. Individual colonies were selected and patched on *H. neapolitanus* low salt solid medium plates containing 50 µg/ml kanamycin or kanamycin and 100 µg/ml ampicillin. Only those colonies that were resistant to kanamycin, but sensitive to ampicillin, were selected for PCR amplification and genomic DNA sequencing.

Bicinchoninic Acid (BCA) Assay

Protein concentrations were determined using the bicinchoninic acid assay, using duplicates for the standards and triplicates for all other samples. BCA protein assay reagents A and B were mixed in a 50:1 ratio, and 950 µl of this mixture were distributed into 1 ml plastic cuvettes. A BSA stock (1 mg/ml) was used to prepare standards of 1 µg,

2 μg, 5 μg, 10 μg, 20 μg, and 40 μg. Usually, 2 to 50 μl of protein samples were analyzed. All sample volumes were adjusted to 1 ml with H_2O. Reactions were incubated at 37°C for 30 min, mixed briefly and read at 562 nm in a DU 800 UV/vis spectrophotometer.

Protein separation by SDS-PAGE

12% SDS-polyacrylamide gels were made and used in a Mini Protean III cell, and 10-20% or 4-20% precast gradient Tris-HCl gels were used in a Criterion cell. Electrophoresis running condition was 110 V for 1.5 h for the 12% home-made gels and 150 V for 1.6 hrs for the precast gradient gels. After electrophoresis, gels were rinsed in H_2O three times 10 min each, and then stained with GelCode Blue stain for 1 to 1.5 h. Gels were de-stained in H_2O overnight before visualization and image-capture.

Electroblotting

For immunoblotting purpose, proteins separated by SDS-PAGE were electro-transferred to 0.45 μm pore sized nitrocellulose membranes with Mini Trans Blot cell or Criterion transfer cell filled with pre-cold transfer buffer. Electroblotting condition were 250 mA for 1 h at 4° C. Blots were air-dried overnight at room temperature.

Immunoblotting

Dried nitrocellulose membranes with electro-transferred proteins were slowly shaken in 5% (w/v) non-fat milk in PBS, 0.1% (v/v) Triton X-100 for 45 min at room temperature in a PerfectWestern immunoblot development box. The blots were subsequently incubated for 1 h with a 1:2,000 to 1:10,000 dilution of primary polyclonal

antibody (IgG raised in rabbit) in 5 to 15 ml of PBS/Triton. Blots were washed with PBS/Triton for 15 min, 5% (w/v) non-fat milk in PBS/Triton for 15 min, and then again with PBS/Triton for another 15 min. Blots were then incubated with 5 to 15 ml of PBS/Triton containing the secondary antibody at a 1:10,000 dilution (goat anti-rabbit IgG alkaline phosphatase [AP]- or horseradish peroxidase [HRP]- conjugated), for 1 h at room tempature. After another series of wash cycles as described previously, the membranes were incubated with 2 ml of the NBT/BCIP reagent for 5 min before visualization and image capturing (AP-conjugated secondary antibody). When HRP-conjugated secondary antibody was used, membranes were incubated with 2-4 ml of freshly prepared SuperSignal West Pico Chemiluminescent Substrate for 5 min before visualizing immuno-reactive bands with the VersaDoc imaging system or exposure to CL XPosure Film.

RNA isolation

Wild-type *H. neapolitanus* was grown in a two-liter chemostat at a dilution rate of 0.05–0.08 h^{-1} as previously described [14]. Two volumes of RNAprotect Bacteria Reagent were added directly to one volume of log-phase cell culture to immediately stabilize the RNA prior to isolation using the RNeasy Bacteria Mini Kit. To remove genomic DNA contamination, the RNA was treated with 10 units of RNase-free Turbo DNase. The enzyme was inactivated by 10 min incubation at 75 °C in the presence of 5 mM EDTA and removed by phenol/chloroform extraction of the sample. The RNA was precipitated with ethanol. The RNA was either used immediately as a 0.1 to 0.2 μg/μl solution in RNase-free water, or stored overnight at -80°C as a dry pellet and resuspended

immediately before use.

cDNA synthesis

First-strand cDNA synthesis was conducted using the iScript cDNA Synthesis Kit according to the manufacturer's protocol. One microliter (140–230 ng) RNA was used in each 20 µl cDNA synthesis reaction. Mock cDNA synthesis reactions without reverse transcriptase were used as negative controls in the real-time PCR experiments. The ten-fold diluted reaction products were used directly as a template in real-time RT-PCR assays.

Real-time PCR

The Beacon Designer 4.0 software (PREMIER Biosoft International, Palo Alto, CA) was used to design primers that were 18–24 nucleotides in length, had melting temperatures between 55 and 60°C, and would produce 150–300 bp PCR products. A total of 17 primer pairs specific for individual carboxysome genes or for the junctions between adjoining genes were selected (Figure 10). Real-time PCR reactions using primers specific for each carboxysome gene or spanning adjacent genes (Appendix) were performed on a MyiQ single-color real-time cycler. Amplification reactions were carried out in a 20 µl volume and contained 1X iQ SYBR Green Supermix, 2 µl of ten fold diluted cDNA template, 200 nmol each of forward and reverse primer, and SCAR reagent. Amplification conditions were 3 min denaturation at 95 °C for the first step and 10 s at 95°C for the remaining 42 cycles, 30 s annealing and extension at 60 °C. To confirm the specificity of each primer pair, the PCR product was analyzed by gel electrophoresis and

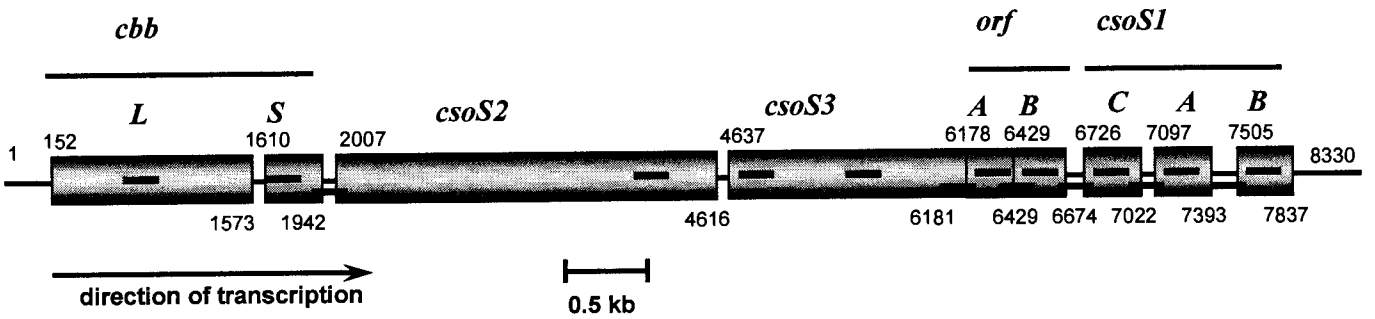


Figure 10: Amplicon coverage of the *H. neapolitanus cso* gene cluster for real time RT-PCR.

The numbers denote beginning (top) and end (bottom) of each carboxysome gene according to GeneBank record AF038430. Bars show the amplified regions obtained with the primer pairs listed in Appendix.

by establishing a melt curve after the final amplification cycle (0.2 °C s^{-1} ramp rate to 95 °C from a starting temperature of 55 °C). All real-time PCR reactions were carried out in triplicate for plasmid DNA standards and in quadruplicate for all cDNA samples.

Absolute cDNA quantification

Plasmid pTn1 [5] containing the entire *H. neapolitanus* carboxysomal gene cluster was isolated with a QIAprep Miniprep Kit and linearized with a restriction endonuclease of choice. The linearized plasmid digestion products were separated by agarose gel electrophoresis and the desired fragment was extracted from the excised gel slices with the QIAquick Gel Extraction Kit. The DNA concentration was determined with a NanoDrop ND-1000 UV–Vis Spectrophotometer, and the copy number of the linearized plasmid DNA calculated using 660 g/mol as the average Mr of a base pair. A serial dilution of the plasmid DNA was prepared in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) covering a copy number range from 10^6 to 10^2. Each dilution was divided into aliquots that were stored at -80 °C no longer than one week and thawed only once before being used in real-time PCR reactions. The Ct values of these standards were plotted against their copy numbers, and the slope of the linear regression was used to determine the amplification efficiency $E = 10^{-1/\text{slope}}$. The starting copy numbers of the genes of interest were calculated from the standard curve and normalized to the ng of input RNA. All Ct values were calculated using the MyiQ software (version 1.0, Bio-Rad) in default "maximum correlation coefficient" mode.

RNA secondary structure and promoter prediction

Stem-loop structures in transcripts derived from the *cso* gene cluster of *H. neapolitanus* were predicted using the RNAdraw program (beta2 version of V1.1; http://www.rnadraw.com). Free energy values were calculated for 37 °C, allowing G:U pairs. Putative promoters within the carboxysome gene cluster were analyzed with the BDGP: Neural Network Promoter Predictor program (http://www.fruitXy.org/seq_tools/promoter.html) using the default settings for prokaryotic promoters.

Primer extension

Three and four primers were used in primer extension reactions targeting the regions upstream from *cbbL* and *csoS1C*, respectively. Primers were 5'-end labeled with [γ-^{32}P]dATP using T4 polynucleotide kinase according to the manufacturers' protocol. Unincorporated nucleotides were removed with a MicroSpin G-25 column. The labeled primer (10^5 cpm) was annealed to 10 µg of *H. neapolitanus* RNA at 65 °C for 10 min. Primer extension reactions were carried out with SuperScript II reverse transcriptase according to the manufacturer's protocol. The same primers were used in sequencing reactions with pTn1 plasmid as template to identify the 5'-end of the transcript with Sequenase Version 2.0 DNA Sequencing Kit.

RNA Ligase-Mediated 3' Rapid Amplification of cDNA Ends (RLM-3'RACE)

To identify the 3' ends of selected carboxysome gene transcripts, RLM-3'RACE was used. Adaptor oligo (Appendix) with phosphorylated 5' and 3' ends (100 pmol) was ligated to 10 µg of freshly isolated RNA using 80 units of T4 RNA ligase in the presence

of 1unit/μl of RNasin in a total reaction volume of 20 μl. cDNA synthesis was primed with adaptor-REV (Appendix), which is complementary to adaptor oligo. First strand reverse transcription reactions were carried out according to the manufacturer's instructions using MMLV H^- reverse transcriptase. One percent of the cDNA was used for each PCR reaction with iTaq DNA polymerase, primed by adaptorREV and an appropriate gene-specific primer complementary to a sequence near the stop codon of the gene of interest (Appendix). PCR products were cleaned up with the QIAquick PCR purification kit or were eluted from agarose gels with the Zymoclean Gel DNA Recovery Kit before insertion into the pCR-Blunt II-TOPO vector with the Zero Blunt TOPO PCR Cloning Kit. The sequences of the inserts were determined at the DNA Sequencing Facility, University of Maine.

Test promoter activity of selected target regions from the *H. neapolitanus cso* operon

Three DNA test fragments from the *H. neapolitanus cso* gene cluster and the *E. coli lac* (Pro_{lac}, a positive control) were cloned into HindIII / EcoRI sites of the promoter probe vectors pPROBE-KT, pPROBE-NT, and pPROBE-OT [53]. The *cso* promoter region (Pro_{cso}) of *H. neapolitanus* was PCR-amplified with primers pTn1f2811 and pTn1r3012. A 190 bp fragment upstream from the *csoS1C* gene ($Frag_{1C}$) was amplified with primers 1Cf-202 and 1Cr-13, and a 185 bp fragment upstream from the *csoS4A* gene ($Frag_{4A}$) with primers orfAf-210 and orfAr-26. PCR amplification of the *lac* promoter from the *E. coli* vector pPROEX-HTb was performed with primers pTrcf11 and pTrcr252. Each resulting construction was transferred into *E. coli* TOP10 cells.

An *E. coli* overnight culture was used to inoculate subculture (1:50), and in the case

of Pro_{lac} containing constructions, 0.4 mM IPTG was added to induce the promoter activity. Subculture was incubated at 37° C for 1.5 h, and a 500 µl aliquot was centrifuged at 14,000 rpm in a bench-top microcentrifuge for 30 sec at room temperature. The medium (450 µl) was removed with a pipette and the cell pellet resuspend with remaining 50 µl medium. A 7 µl aliquot of cell suspension was dropped onto the glass slide and covered with a cover slip. Cells were observed with a Zeiss LSM 510 confocal laser scanning microscope.

Generation of a *H. neapolitanus cso* promoter deletion mutant

E. coli DY330 cells were transformed with plasmid pUC18-*cbbLS*, which contains the *cso* promoter region. An approximately 1,000 bp fragment was constructed that consists of a *Kan*^{r} cassette with its own promoter. The cassette was PCR-amplified from the pCR-Blunt II vector with forward primer ProKmF and reverse primer ProKmR. The resulting PCR fragment, which was flanked by short *H. neapolitanus* genomic DNA regions derived from both sides of the *cso* promoter, was introduced into competent *E. coli* DY330 cells that contained pUC18-*cbbLS*. Tthe -35 box and -10 box of the *cso* promoter of that plasmid was replaced with the *Kan*^{r} cassette by homologous recombination (Figure 11). The resulting plasmid, which contains the *Kan*^{r} cassette in the opposite direction from that of the *cso* operon, was electroporated into exponentially growing *H. neapolitanus* cells. Success of mutagenesis was confirmed by PCR amplification and genomic sequencing.

Generation of polyclonal antiserum against CsoS1 proteins

Recombinant CsoS1 Protein Purification

The *csoS1A, csoS1B, and csoS1C* genes of *H. neapolitanus* were cloned into the pProEX-HTb prokaryotic expression vector and introduced into DH5α competent *E. coli* (Eric B. Williams, dissertation). The resulting recombinant proteins have an N-terminal his_6 tag that can be cleaved with Tobacco Etch Virus (TEV) protease. For protein expression, four 3-ml-overnight cultures were used to inoculate four 1 L-baffled flasks, each containing 250 ml LB medium with 100 μg/ml ampicillin. Subcultures were incubated in a 37° C shaker at 225 rpm until the OD_{600} reading reached 0.5. Isopropyl-β-d-thiogalactoside (IPTG) was added to each culture to 0.4 mM concentration to induce recombinant protein expression. Cultures were incubated at room temperature for 4 h with shaking. The cells were harvested by centrifuging at 10,000 rpm in rotor JLA16.25 for 10 min at 4° C. Cell pellets were kept frozen (-20° C) if not been processed immediately. pProEX lysis buffer 30 ml was used to resuspend the cells, and the suspension was sonicated three times for 30 s each on ice, with 1 min intervals between each burst (30% sonicator power output setting). The slurry was centrifuged at 10,000 rpm in rotor JA25.50 for 10 min at 4° C. The resulting supernatant was mixed with 2 ml Ni-NTA agarose affinity resin that had been equilibrated with pProEX wash buffer A, and then incubated at 4° C overnight while rotating gently. The slurry was loaded into a 1.5 ×10 cm column and the resin allowed settle. The flow through was collected and re-applied to the column for two more times. Twenty bed volumes (40 ml) wash buffer A, then 20 ml wash buffer B, and finally another 20 ml wash buffer A were used to remove non-specifically bound proteins. The resin was resuspended in four bed volumes of

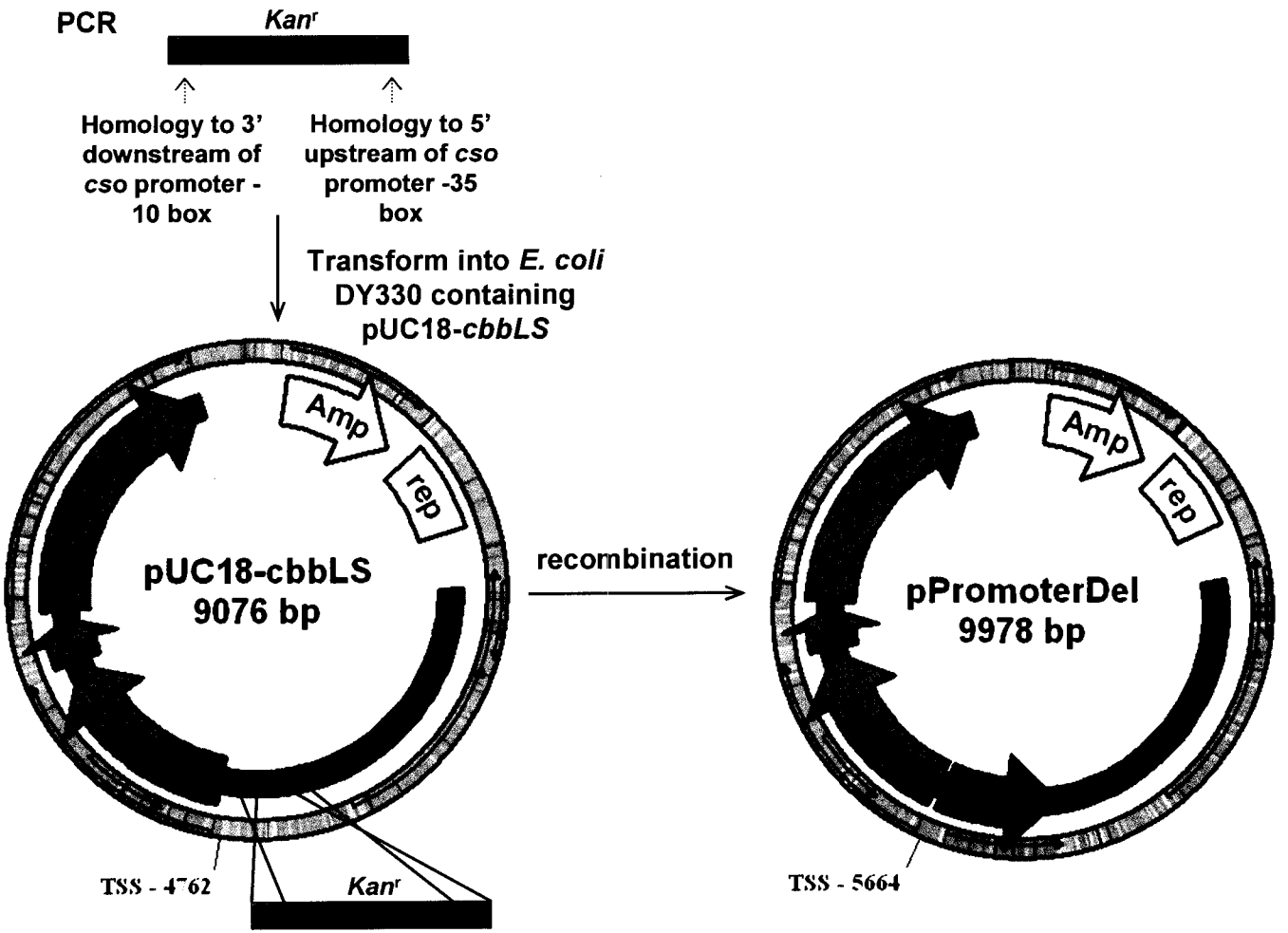


Figure 11. Strategy used for constructing the *H. neapolitanus cso* promoter deletion mutant.

TSS represents transcript start site determined in this study.

elution buffer containing 150 mM imidazole by pipetting up and down. The column was incubated at 4° C for 30 min while the resin was allowed to settle. Elution was collected and dialyzed twice against 10 mM Tris-HCl buffer (pH 8.0) at 4° C. The protein samples were stored at -20° C until needed.

AcTEV Cleavage

AcTEV protease was used to cleave the his_6 tag from purified recombinant CsoS1 proteins. rCsoS1A and rCsoS1C proteins were mixed in a 1:1 molar ratio before the reaction. rCsoS1B was cleaved separately. Typically, a 150 µl reaction was set up as following: 33 µg of recombinant protein, TEV buffer, 1 mM DTT, and 10 units of AcTEV. Sterilized H_2O was added to adjust the volume. This reaction was incubated in a 30° C oven for 6 h, then the reaction mixture was mixed with 0.5 volume of Ni-NTA resin that had been equilibrium with TEV buffer. The slurry was incubated at 4° C overnight while rotating gently. The flow-through fraction was collected, three 0.5 volumes of pProEX wash buffer B were used to wash the resin, and the flow-through fraction from these steps were pooled with the original flow-through. The protein in the flow-through fraction was concentrated in a dialysis tubing (MWCO 1,000) that was placed on a bed of dry polyethylene glycol (PEG) 8,000 for 2 h at 4° C. The samples were dialyzed twice against 10 mM Tris-HCl (pH 8.0) at 4° C.

Generation and Testing of Polyclonal Antibodies against CsoS1A/C and CsoS1B

Rabbit anti-CsoS1A/C and anti-CsoS1B polyclonal antibodies were raised at Cocalico Biologicals, Inc. from approximately 1 mg of de-tagged recombinant CsoS1A/C and CsoS1B. Antisera USM74 and USM75 were raised against CsoS1A/C without the his_6 tag, and USM76 and USM77 were raised against CsoS1B without the his_6 tag. First

and second test bleeds and the final bleeds of all four antisera were tested against purified wild type carboxysomes and recombinant CsoS1 proteins.

Generation of polyclonal anti-serum against CsoS4A and CsoS4B

Cloning and Expression the csoS4A and csoS4B Genes of H. neapolitanus in pProEX-HTb vector

The *csoS4A* and *csoS4B* expression clones in the pProEX-HTb vector were generated. The *csoS4A* and *csoS4B* genes were amplified from pTn1 [5] using Pfu UltraII DNA polymerase with primer pairs oAfBamHI / oArHindIII, and oBfBamHI / oBrHindIII, respectively (primer sequence see Appendix). Forward and reverse primers contained engineered restriction sites BamHI and HindIII, respectively, that allowed for in-frame ligation into the multi-cloning site (BamHI and HindIII) of the prokaryotic expression vector pProEX-HTb. Resulting recombinant plasmids were transferred into chemically competent *E. coli* DH5α cells and protein expression at small scale were tested. Recombinant protein should contain N-terminal his_6 tag.

Cloning and Expression of csoS4A and csoS4B in the pET22b Vector

The *csoS4A* and *csoS4B* expression clones in the pET22b vector were generated by following the same procedure described as above that outlined for pProEX-HTb, except that different primer pairs were used. A second set of constructs was made to generate fusion proteins with a N-terminal *pelB* signal sequence for periplasmic localization. All four constructs should generate recombinant protein with a C-terminal his_6 tag. Primer pairs oAfNdeI / oArXhoIns and oBfNdeI / oBrXhoIns were used for *csoS4A* and *csoS4B* constructs respectively including the signal sequence. Primer pairs oAMscI / oArXhoIns

and oBfNcoI / oBrXhoIns were used for constructs of *csoS4A* and *csoS4B* respectively without signal sequence. The resulting plasmids were transferred into chemically competent *E. coli* BL21(DE3) cells and protein expression induced as described. Clones that expressed CsoS4A and CsoS4B without *pelB* signal sequence were used for all subsequent work.

Recombinant CsoS4 Protein Purification

For protein expression, four 3-ml-overnight cultures were used to inoculate four 1 L-baffled flasks, each containing 250 ml LB medium with 100 μg/ml ampicillin. Subcultures were incubated in a 37° C shaker at 225 rpm until the OD_{600} reading reached 0.5. Isopropyl-β-d-thiogalactoside (IPTG) was added to each culture to 1.0 mM concentration to induce recombinant protein expression. Cultures were incubated at room temperature for 4 h with shaking. The cells were harvested by centrifuging at 10,000 rpm in rotor JLA16.25 for 10 min at 4° C. Cell pellets were kept frozen (-20° C) if not been processed immediately. Cell pellets were resuspended in 30 ml of B-PERII buffer, and the suspension was shaken vigorously at room temperature for 10 min. Then 600 μl of freshly prepared 10 mg/ml lysozyme was added, and the cell slurry was shaken for another 5 min at room temperature. The cell slurry was sonicated three times for 30 s each on ice, with 1 min intervals between each burst (30% sonicator power output setting). The lysate was shaken for another 10 min before centrifugation at 14,000 rpm in rotor JA25.50 for 20 min at 4° C. The resulting supernatant was mixed with 2 ml Ni-NTA agarose affinity resin. The remaining the purification step was as described in CsoS1 protein purification section.

Generation and Testing of Polyclonal Antibodies against CsoS4A and CsoS4B

Rabbit anti-CsoS4A and anti-CsoS4B polyclonal antibodies were generated as described previously in the section of the generation of antibodies against CsoS1 protein. Antisera USM61 and USM62 were raised against rCsoS4A (with his_6 tag), and USM59 and USM60 were raised against rCsoS4B (with his_6 tag). First and second test bleeds and the final bleeds of all four antisera were tested against purified wild type carboxysomes and recombinant CsoS4 proteins.

Two-dimensional (2 D) electrophoresis

A purified carboxysome sample was centrifuged at 14,000 rpm for 30 min at 4° C in a microcentrifuge. The supernatant was removed with pipette and the carboxysome pellet incubated at -20° C for 30 min. The frozen carboxysome pellet was thawed by adding one volume of TEMB buffer. During this freeze-thaw cycle carboxysomes broke and released RubisCO into the solution. Another centrifugation step at 14,000 rpm for 30 min at 4° C separated free RubisCO in the supernatant and the resulting pellet was referred as shell enriched pellet fraction. By determining the carboxysome protein concentration and that of the free RubisCO fraction, the mass of the shell enriched fraction could be calculated. An aliquot of this fraction corresponding to 88 μg was resuspended in 200 μl of freshly prepared rehydration buffer. After adding 1 μl of 0.5% (w/v) bromophenol blue to facilitate visualization, the sample was loaded in a clean lane of a Protean IEF focusing tray and distributed along the entire length of the channel except for approximately 1 cm at each end. Care was taken to avoid bubble formation. The cover sheet was peeled off a 11 cm pH 4-7 ReadyStrip IPG strip with clean forceps, and the strip was placed in the

tray gel side down and touching the sample solution. Again, introduction of bubbles was avoided in this step. The strip was overlayed with 1 ml mineral oil and the focusing tray covered with a lid and placed in the Protean IEF cell. Activation proceeded at 50 V, overnight at 20° C. After the rehydration step was complete, two paper wicks were wetted with H_2O and inserted with forceps between the IPG strip and the electrodes at both ends. Isoelectric focusing was done with a one-step program setting as follows: 50 V with constant current of 50 μA/strip and rapid ramping, maximum voltage 8,000 V, and 40,000 total V-h at 20°C. Following the focusing step, the mineral oil was removed by placing the stript gel side up onto a piece of dry filter paper and keeping it in a piece of folded parafilm, sealed by pressing the end of a forceps against the parafilm. The sealed strip was frozen at -20° C until ready for the next step. The frozen IPG strip was allowed to thaw for 10 min at room temperature. To a clean channel of an equilibration/rehydration tray, 5 ml of equilibrium buffer I with DTT, which reduces sulfhydryl groups, was added, and the thawed IPG strip was transferred into the channel gel side up. The tray was placed on an orbital shaker and gently shaken for 10 min before being transferred to another clean channel filled with 5 ml of equilibrium buffer II containing iodoacetamide which alkylates any reduced sulfhydryl groups. The strip was incubated for another 10 min. During this time, overlay agarose solution was melted in a microwave oven, and a precast 10-20% IPG+1 criterion gel was removed from the box. Excess water remaining inside the IPG loading well was removed with a thin piece of Whatman filter paper. The IPG strip was removed from the equilibration/rehydration tray, dipped briefly into a 10 ml graduated cylinder filled with SDS-PAGE running buffer, and laid onto the back plate of the criterion gel above the IPG well. The strip was pushed into the well on its plastic

backing with a forceps to ensure the gel matrix was in contact with the criterion gel surface. A transfer pipette was used to place overlay agarose solution into the IPG well and completely cover the strip, carefully avoiding bubbles introduced. The criterion gel was set vertically for at least 10 min to allow the agarose to solidify before proceeding. Running conditions for the criterion gel were the same as described previously for SDS-PAGE.

Carboxysome surface labeling with the amine-reactive fluorophore Alexa Fluor 546

The contents of one vial of Alexa Fluro 546 were dissolved in 1 ml TEMB buffer. Only 250 µl was used for the labeling reaction, and the remaining 750 µl were kept at -20 ° C until used again. The dissolved Alexa Fluro 546 dye was mixed with 500 µl freshly prepared carboxysomes in TEMB to allow the labeling reaction to occur. A 100 µl aliquot was removed after 10 s, 20 s, 30 s, 1 min, 2 min, 5 min and 15 min of incubation and added to 10 µl of 2 M glycine (pH 8.0) to quench the reaction. All fractions were centrifuged at 14,000 rpm for 60 min at 4° C in microcentrifuge. The pellets containing carboxysomes were washed gently with 50 µl TEMB buffer three times, then resuspended in 30 µl TEMB containing 200 mM glycine. The carboxysome proteins were separated by SDS-PAGE on 12% gel. Illumination with short wave length UV light was used to visualize Alexa Fluor 546-stained protein bands. A 100 µl aliquot of the pooled labeling reactions was loaded on a 700 µl 10-60% (w/v) mini sucrose gradient and centrifuged in a MLS50 rotor at 31,000 rpm for 30 min at 4° C. The gradient was fractionated in fifteen 50 µl fractions, and the pellet was resuspend in 50 µl TEMB buffer. Aliquots (20 µl) of all sixteen samples were separated on 10-20% precast gradient gel and

examined after staining and illumination with UV light.

Carboxysome labeling with Sulfo-NHS-Biotin

Sulfo-NHS-Biotin (SNB) was prepared in TEMB, Bicine-$MgCl_2$ solution (20 mM Bicine, 2 mM $MgCl_2$, pH 8.5) or PBS (pH 8.5) buffer as 2-5 mM. Carboxysome samples were resuspended in this solution and incubated at room temperature for 1 h. An equal volume of 2 M glycine was added to quench the reaction. Proteins were separated on SDS-PAGE as described. SNB-labeled protein bands were visualized with HRP-conjugated Neutr-Avidin (1:25,000 dilution).

Dynamic Light Scattering

Dynamic light scattering (DLS) experiments were conducted on a Malvern Instruments Zetasizer Nano-ZS (red badge) instrument operating with a 633 nm laser. Data was collected and processed with the Dispersion Technology software V5.10. Unlabeled or SNB-labeled carboxysomes were diluted 1,000 fold with TEMB buffer and their particle size distributions were measured.

Construction of pTn1 subclones pHnHE1.3 and pHnPE2.0

pTn1 subclones pTnCXY1 and pTnCXY1H from Dr. Shievly's lab [23] were used to prepare pHnPE2.0 and pHnHE1.3, respectively. pTnCXY1 was digested with PstI and EcoRI, and the 2.0 kb fragment that contains *csoS1C*, *A* and *B* was inserted into PstI / EcoRI sites of pUC18. The resulting plasmid was named pHnPE2.0 plasmid. pTnCXY1H was digested with HindIII and EcoRI, and the 1.3 kb fragment that contains the

C-terminal portion of *csoS1C* and all of *csoS1A* and *B* was inserted into HindIII / EcoRI sites of pUC18. The resulting plasmid, which was named pHnHE1.3 was electroporated into *E. coli* DY330 cells. Plasmid maps are shown in Figure 12.

Generation of the *H. neapolitanus csoS1A*-TC tag mutant

An approximately 1,000 bp fragment containing a *Kan*^{r} cassette was constructed by PCR amplification of the pCR-BluntII-TOPO vector with primers 1A-TC-tag-KmF and 1A-KmR. The resulting fragment, which was flanked by *H. neapolitanus* genomic DNA regions derived from both sides of the *csoS1A* stop codon, was transferred into *E. coli* DY330 cells containing pHnHE1.3 to engineer a TC-tag (tetracysteine motif: Cys-Cys-Pro-Gly-Cys-Cys) on the C-terminus of CsoS1A protein and insert a *Kan*^{r} cassette downstream from the *csoS1A* gene by homologous recombination (Figure 13). The resulting plasmid, pHnHE-1A-TCtag-Km, was then used for electroporation of exponentially growing *H. neapolitanus* cells. Success of mutagenesis was confirmed by PCR amplification and genomic sequencing.

Generation of *H. neapolitanus csoS1B* truncated mutant

The *Kan*^{r} cassette was amplified with primers 1BtKmF and 1BtKmR as described in the section of generating the *csoS1A*-TC tag mutant, generating a fragment flanked by *H. neapolitanus* genomic DNA regions derived from both sides of the coding sequence for the last 12 amino acids of *csoS1B*. This fragment was transferred into *E. coli* DY330 cells containing pHnHE1.3, and resulted in the formation of plasmid pHnHE-1Bt-Km, in which the last 12 amino acid codons of the *csoS1B* gene were deleted (Figure 14).

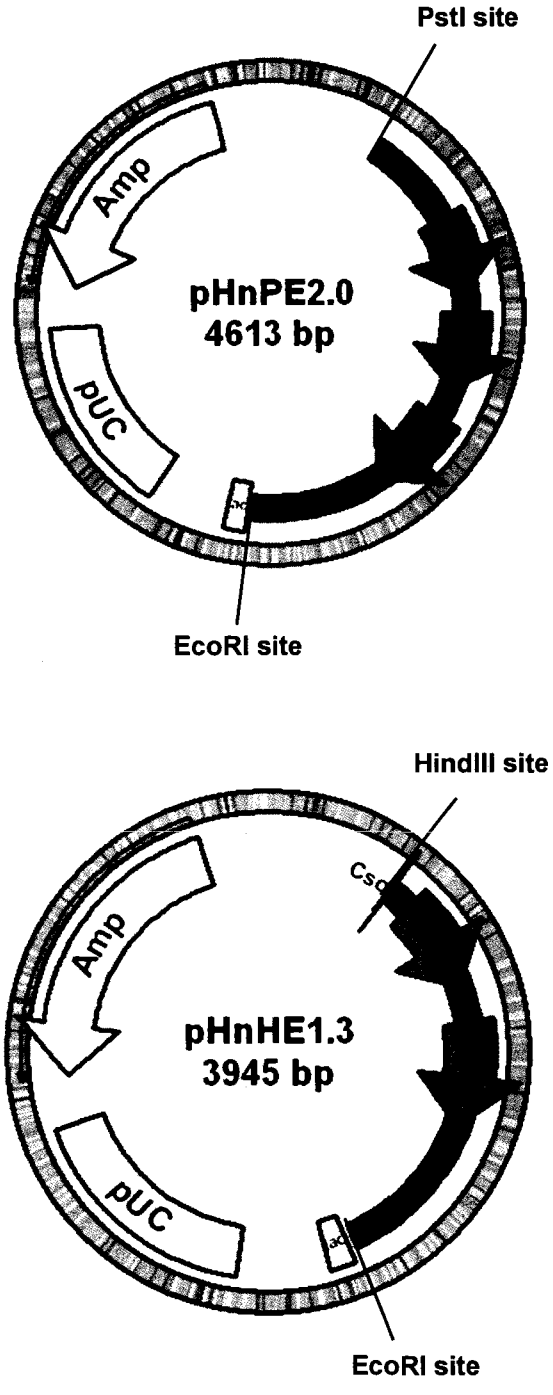


Figure 12. Maps for plasmid pHnPE2.0 and pHnHE1.3.

pHnPE2.0 and pHnHE1.3 generated in this study have pUC18 backbone and contain 2.0 kb PstI/EcoRI fragment and 1.3 kb HindIII/EcoRI fragment of *cso* operon of *H. neapolitanus*, respectively.

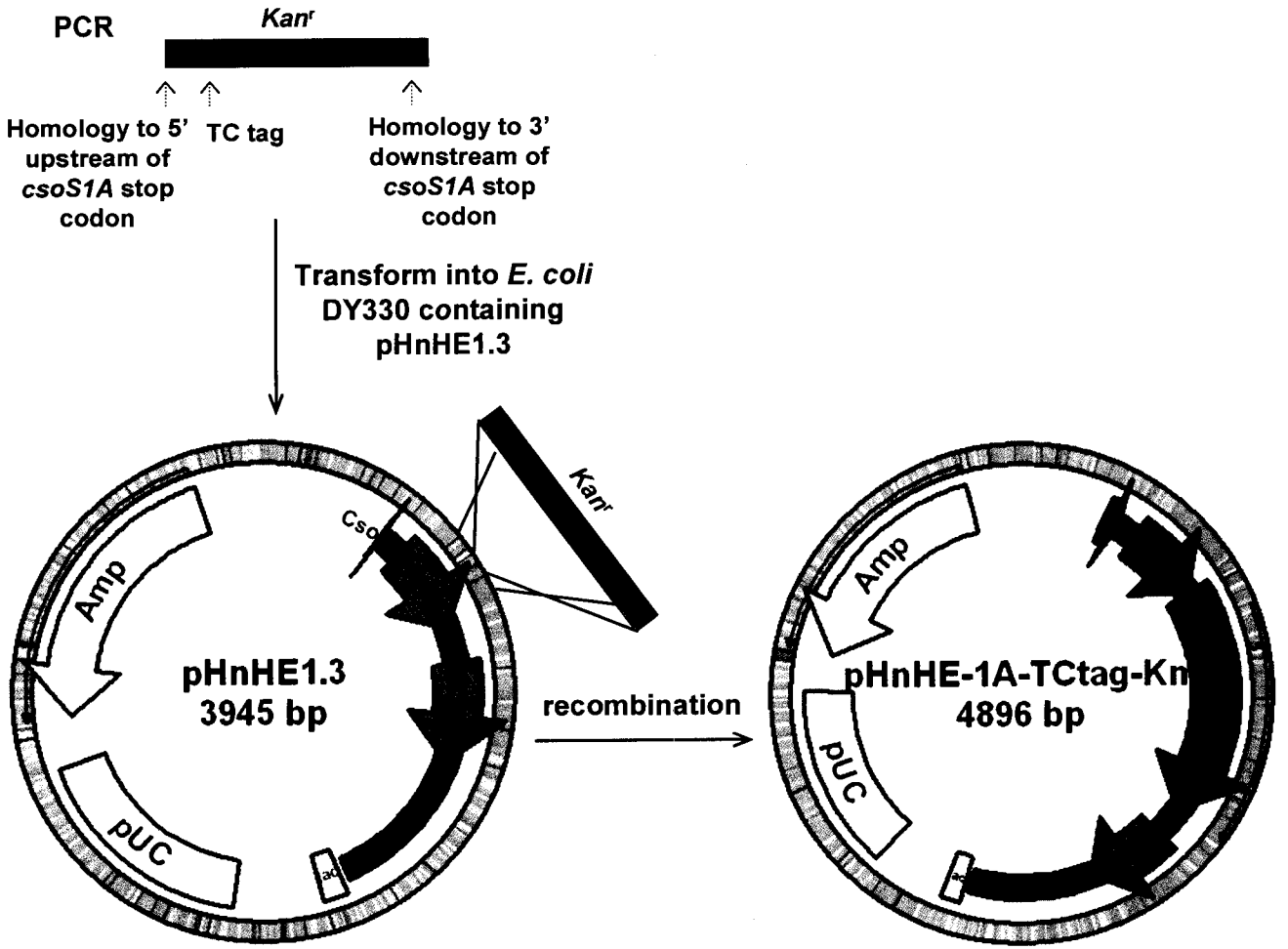


Figure 13. Strategy used for constructing the *H. neapolitanus csoS1A*-TCtag mutant.

A tetracysteine motif: Cys-Cys-Pro-Gly-Cys-Cys was enginerred at the C-terminusof *csoS1A* gene.

Plasmid pHnHE-1Bt-Km was then electroporated into exponentially growing *H. neapolitanus* cells to generate the mutant.

Generation of the *H. neapolitanus csoS1A::G43A* mutant

The pTOPO-*csoS1A::G43A* plasmid (generated by Sara Johnson), which contains a *csoS1A* site-specific mutation, was digested with PmlI and Bsu36I to excise a 236 bp fragment that includs the mutated site. This fragment was ligated into the PmlI / Bsu36I sites of pHnHE1.3 (Figure 15). The resulting plasmid, pHnHE-*csoS1A::G43A,* was transferred into *E. coli* DY330 cells. A *Kan*^{r} cassette was PCR amplified with primers 1A-KmF and 1A-KmR and introduced into this strain as previously described in the section of generation of *csoS1A*-TC tag mutant. The resulting plasmid pHnHE-*csoS1A::G43A*-Km was later used to transform exponentially growing *H. neapolitanus* cells to generate the mutant.

Generation of *H. neapolitanus csoS4A*, *csoS4B*, and *csoS4AB* deletion mutants

H. neapolitanus csoS4A, *csoS4B*, and *csoS4AB* deletion mutants were generated by replacing *csoS4A*, *csoS4B*, and both *csoS4* genes with a *Kan*^{r} cassette (Figure 16). The *Kan*^{r} gene was PCR amplified with primer pairs oAKmF / oAKmR, oBKmF2 / oBKmR, and oAKmF / oBKmR, in each case. The PCR products were used to transform *E. coli* DY330 cells containing the pTnE4.3 plasmid, in which the target region was replaced with the respective PCR product. The resulting plasmids, pTnE4.3-oAKm, pTnE4.3-oBKm2, and pTnE4.3-oABKm were introduced into exponentially growing *H. neapolitanus* cells by electroporation to generate the corresponding mutants.

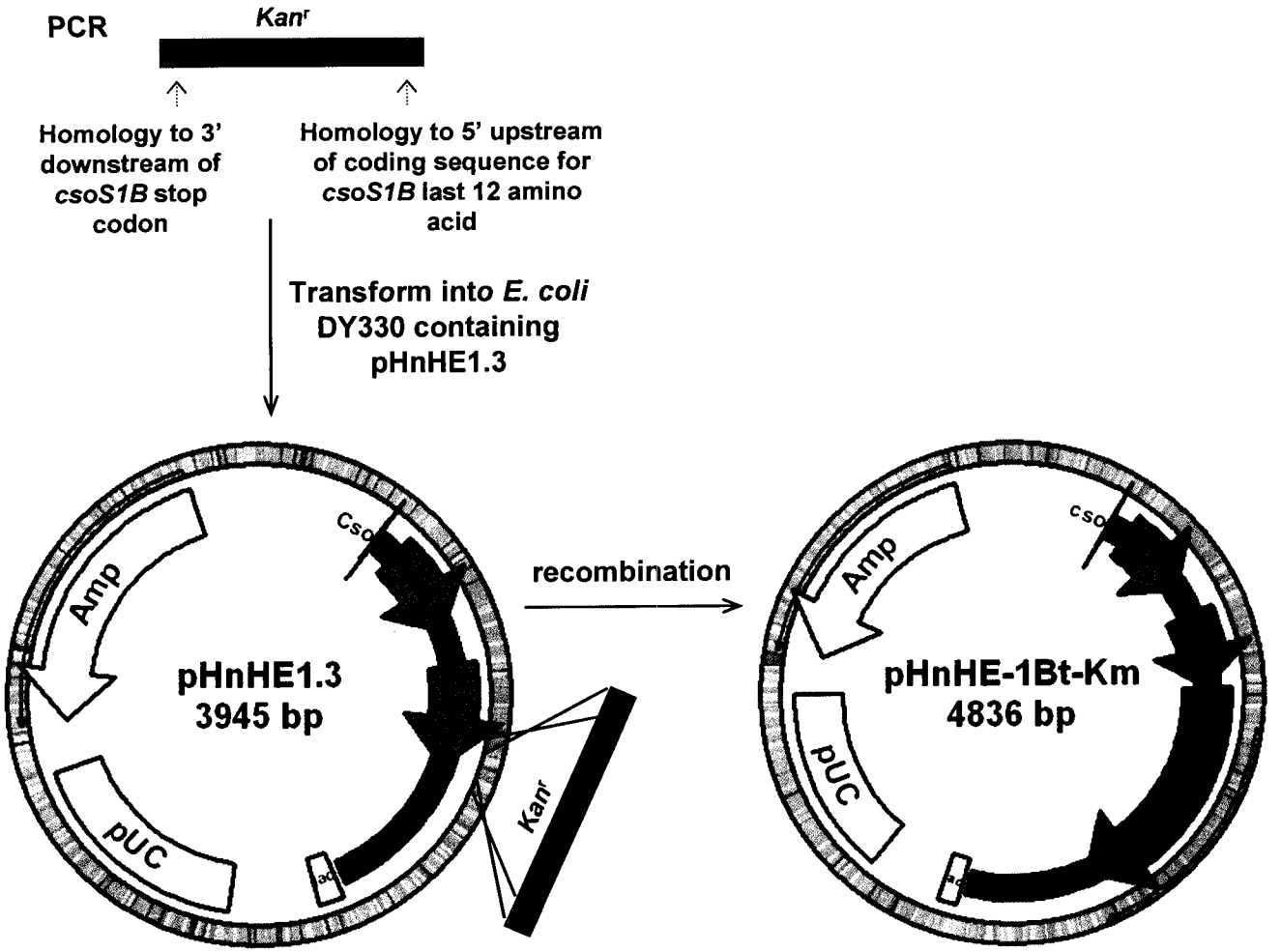


Figure 14. Strategy used for constructing the *H. neapolitanus csoS1B* truncated mutant.

The last 12 amino acid at C-terminal of the *csoS1B* gene was deleted.

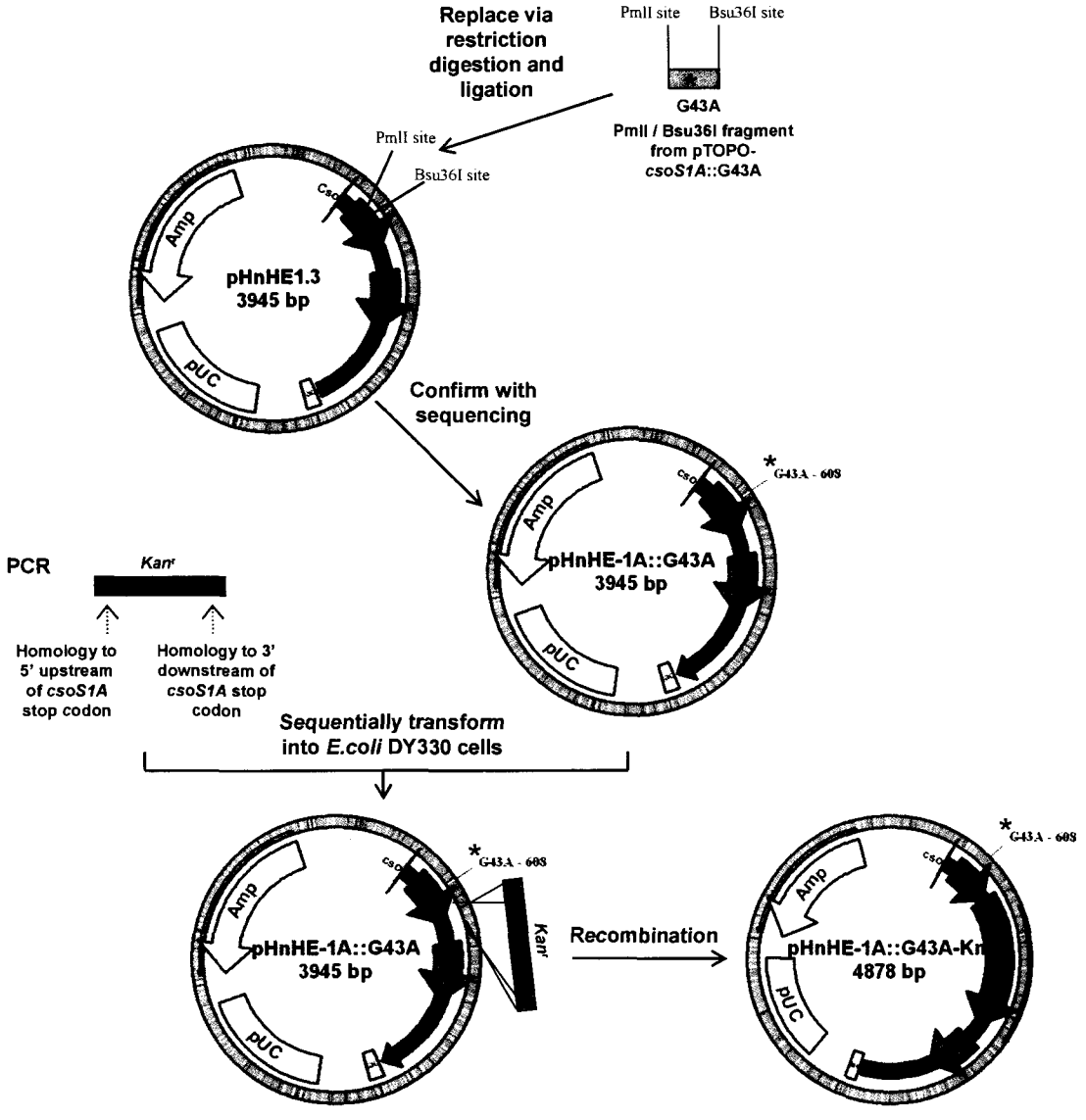


Figure 15. Strategy used for constructing the *H. neapolitanus csoS1A*::G43A mutant. One of the three glycines which locate at the pore of a CsoS1A hexamer was mutated to alanine.

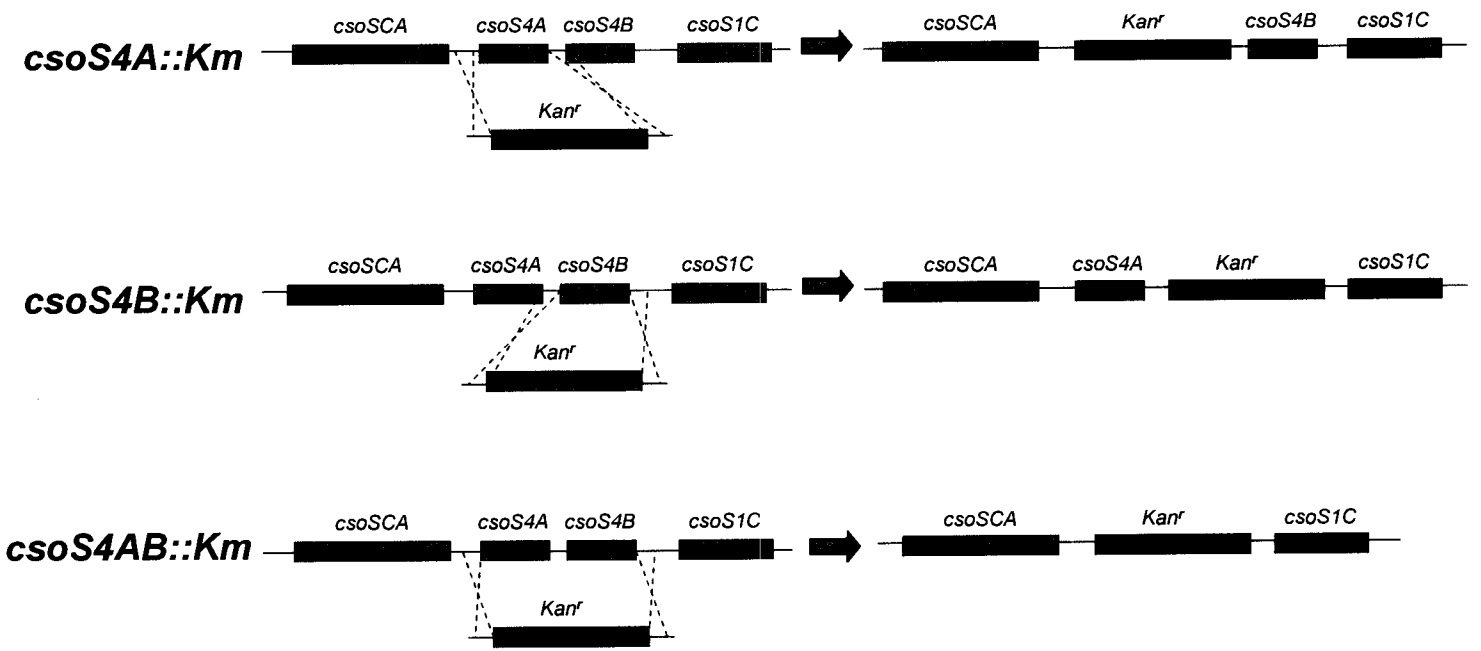


Figure 16. Generating the *H. neapolitanus csoS4* gene deletion mutants.

csoS4A, *csoS4B*, or *csoS4A* and *csoS4B* gene(s) was replaced by a *Kan*^{r} cassette via recombination.

Carboxysome enrichment from batched culture

Mutant *H. neapolitanus* cells were grown in 750 ml batch cultures in 2-L baffled flasks. The flasks were shaken at 200 rpm in a incubator at 30° C in air supplied with 5% CO_2. Cells were harvested by centrifugation at 12,000 rpm in rotor JLA16.25 for 5 min at 4°C. The cell pellet was washed in 10 ml BEMB and re-centrifuged at 10,000 rpm in rotor JA25.50 for 10 min at 4° C before being resuspend in 5 ml BEMB buffer with 2 mg/ml lysozyme and 5 mg/ml $MgSO_4$. An equal volume of B-PER II reagent was mixed with the cells and the lysate was sonicated with 3 mm tapered microtip disruptor horn for 10 s or until the suspension was no longer viscous. At that time, 15 µl of 1 mg/ml bovine pancreatic DNase I in BEMB was added. The mixture was shaked vigorously for 30 min at room temperature and cleared by centrifuging at 9,000 rpm in rotor JA25.50 for 10 min at 4°C. The carboxysomes were pelleted from the supernatant by centrifugation at 20,000 rpm in rotor JA25.50 for 35 min at 4°C. The resulting pellet was resuspended in 100-200 µl BEMB and subjected to a brief clearing spin at 3,300 rpm for 3 min at 4°C in a microcentrifuge. The supernatant from this low-speed spin was referred as P_{20} fraction. If necessary, this carboxysome-enriched fraction was loaded onto a 4-ml 10-60% (w/v) or 10-65% (w/w) continuous sucrose gradient prepared in BEMB buffer, in a 13 × 51 mm ultraclear centrifuge tube. The gradient was centrifuged at 35,000 rpm for 35 min at 4°C in a MLS50 swinging bucket rotor in a bench-top ultracentrifuge. The carboxysome band was collected or the entire gradient was fractionated into 500 µl fraction. Fractions of interested were transferred to a 11 × 34 mm PC centrifuge tube, diluted 1:2 with BEMB buffer, and centrifuged at 100,000 rpm for 2 h at 4°C in a MLA130 fixed angle rotor. The final carboxysome pellet was resuspended in 100 µl of BEMB and stored at 4°C for

further analysis.

Purification of mutant carboxysomes from fermentor cluture

Mutant *H. neapolitanus* cells were collected from a 8 L carboy connected to the chemostat and concentrated with a Millipore Pellicon filtration unit (0.2 μm pore size) driven by a peristaltic pump. The concentrated cell suspension was centrifuged at 10,000 rpm for 10 min at 4°C. Approximately 6 g (weight wet) cells can be yielded. The resulting cell pellet was resuspended in 15 ml of freshly prepared BEMB buffer containing 2 mg/ml lysozyme and 10 mM $MgSO_4$. A equal volume of B-PER II bacterial protein extraction reagent was added and the tube was inverted several times to mix the detergent with the cells. The cell suspension became very viscous at this point, which indicated the release of genomic DNA from the cells lysed. The cell suspension was sonicated for 10 s or until the slurry was no longer viscous, with a Branson model 450 sonifer set at a constant duty cycle with 30% power output. After sonication, 150 μl of 1 mg/ml bovine pancreatic DNase I in BEMB was added, and the cell lysate was shaken vigorously for 30 min at room temperature on a Reliable Scientific D55 shader set at maximum speed. The cell lysate was cleared by centrifugation at 10,000 rpm in rotor JLA16.25 for 10 min. The supernatant was transferred with a transfer pipette to a clean polypropylene centrifuge tube and centrifuged at 20,000 rpm in rotor JA25.50 for 30 min at 4° C. The pellet obtained from this step (P20 fraction) was resuspended in 3 ml BEMB buffer and centrifuged at 3,300 rpm for 3 min. The resulting supernatant was highly enriched with carboxysomes and was loaded on a 36-ml 10-60% (w/v) continuous sucrose gradient in BEMB buffer for further purification. The gradient was centrifuged in a JS 24.38

swinging bucket rotor at 24,000 rpm for 35 min at 4°C, and the white, milky band of carboxysomes approximately half way down the gradient was collected with a transfer pipette, transferred to a polycarbonate centrifuge bottle and diluted 1:2 with BEMB buffer. The carboxysomes were pelleted by ultra centrifugation at 35,000 rpm for 2 h at 4°C in a 70Ti rotor, resuspended in 500 µl BEMB buffer, and stored at 4°C for further analysis.

Dot Blots

When quick analysis of fractionized gradient samples was needed, dot blots were performed. A Dot Blotter was assembled with a piece of nitrocellulose membrane and two pieces of Whatman filter paper underneath, and then sealed with transparent cover film from which the area to be used was cut out with a clean razor blade. A vacuum was applied to the blotter and the samples loaded into the wells were 800 µl aliquot at a time. When a sample was very viscous, e.g. contained a high percentage of sucrose, it was diluted prior to adding to the well. The vacuum was kept on for an additional 10 min to ensure that the sample solution completely went through the membrane. The nitrocellulose membrane was allowed to air dry overnight before being developed by the same procedure as described for immunoblots.

Fixing and embedding of *H. neapolitanus* cells

H. neapolitanus wild type or mutant cells were collected from 50 ml cultures in early stationary phase by centrifuging at 10,000 rpm for 15 min at 4°C.

Step one was fixing the cell pellets. Cell pellets were fixed in 2.5% (w/v) glutaraldehyde, 2.0% (w/v) paraformaldehyde in 0.1 M sodium cacodylate buffer, pH 7.0

for 2 h. To prepare 2.0% (w/v) paraformaldehyde, 20 ml of 10% (w/v) paraformaldehyde in H_2O was heated to 60-65°C and a few drops of 1.0 M NaOH were added to clear the solution. The pellets were rinsed in 0.1 M sodium cacodylate buffer, pH 7.0 three times for 2 min, 3 min and 5 min, respectively. Finally pellets were postfixed with 1% (w/v) osmium tetroxide in 0.1 M cacodylate buffer, pH 7.0 for 45 min and rinsed in H_2O as described.

Step two was dehydrating the cell pellets. Pellets were dehydrated by consecutive incubations (5 min each) in 50%, 70%, 85%, 95% ethanol. Then pellets were subsequently incubated in 100% ethanol for three times for 5 min, 10 min, and 15 min, respectively; followed by a series of incubations in 100% propylene oxide or acetone for 5 min, 10 min, and 15 min. (Note: After incubation in 70% ethanol, samples can be stored at 4°C for up to one week before resuming the procedure.)

Step three was infiltrating and curing samples in ERL 4206 (Spurr's) epoxy. To prepare plastic resin, 26.0 g nonenyl succinic anhydride (NSA), 10.0 g vinyl cyclohexene dioxide (VCD, ERL 4206), and 6.0 g diglycidyl ether (DER 736, flexibilizer) were mixed by stirring. The catalyst, 0.20 g dimethylaminoethanol (DMAE) was added and stirring continued for several minutes to ensure complete mixing. Samples were first infiltrated with propylene oxide : plastic (1:1) for 4 h while on a rotator, then in propylene oxide : plastic (1:3) overnight under rotation. Samples were subsequently transferred to 100% plastic for 3 h on a rotator, then to fresh 100% plastic for another 5 h. Finally, fresh 100% plastic was added to the samples and cured at 70°C for 36 h.

Transmission electron microscopy (TEM)

Carboxysome samples were diluted 1:50 with BEMB or TEMB buffer, depending on the buffer in which they were suspended, and 4 µl of the diluted samples were loaded onto formvor/carbon coated copper grids and allowed to stand for 4 min. Excess solution was removed with the edge of a kimwipe, and the grid was allowed to air-dry. Samples were negatively stained with 4 µl of 1% (w/v) ammonium molybdate in BEMB or TEMB buffer for 45-60 s. Excess stain solution was removed with kimwipe and the grid was allowed to air-dry before being observed with a Zeiss EM-109 transmission electron microscope. Images were taken on Kodak EM film 4489 with a 2 s exposure time. Exposed films were developed in Kodak D-19 developer solution for 5 min, rinsed in running tap water for 2 min, fixed in Kodak fixer solution for 10 min, and rinsed in running tap water for 20 min. After drying in air for at least 3 h, films were scanned with an Epson Perfection V700 photo scanner.

CHAPTER III

RESULTS

Experiments designed in this study were targeted to promote the current understanding of how α-carboxysome genes are regulated and expressed to yield precise relative ratios of each protein product, which further enable the correct assembly, and to explore the *in vivo* functions of two sets of conserved bacterial microcompartment genes, *csoS1/ccmK* and *csoS4/ccmL*. For these purposes, the chemoautotrophic bacterium, *Halothiobacillus neapolitanus*, for which a genetic engineering method and a protocol for carboxysome isolation have been well established [14,22], was used as the model organism. Results of this study were grouped into three sections: (1) transcript analysis of the *cso* operon; (2) composition study of purified carboxysomes; and (3) investigation of carboxysome function with focuses on two sets of conserved genes.

1. Transcript Analysis of the *Halothiobacillus neapolitanus cso* Operon

Carboxysomes are polyhedral protein microcompartments that consist of thousands of polypeptide copies. While the exact protein composition of β-carboxysome is not very clear due to technical difficulties with their purification, the protein composition of α-carboxysomes in the model organism *Halothiobacillus neapolitanus* is well established. Stoichiometries of carboxysome polypeptides, which were based on densitometries of stained polypeptide bands on SDS-PAGE, revealed that all protein products of carboxysome genes are not present in equal molar ratios in purified carboxysomes. CbbL and CbbS, which together correspond to approximately 70% of carboxysome mass, are present in over 2000 copies per carboxysome. In other words, there are approximately

300 RubisCO holoenzyme molecules per carboxysome. The CsoS1 proteins 1A, 1B, and 1C are the most abundant shell components, with a combined number of over 3500 monomers [14]. Among them, around 500 copies are CsoS1B, but the relative ratio of CsoS1A to CsoS1C is unknown due to the fact that these two proteins only differ in two amino acids and they have almost identical calculated molecular weight and pI. Approximately 330 copies of CsoS2 (CsoS2A and CsoS2B) and 40 dimer copies of the carboxysomal carbonic anhydrase CsoSCA [14,33] are present in the shell as well. To date, the products of *csoS4A* and *csoS4B* (formerly *orfA* and *orfB*) have not been detected in *H. neapolitanus* cell extracts or purified carboxysomes. However, these proteins are expected to be important for carboxysome structure and/or function, because their homologs are present in both α- and β- carboxysome gene clusters and in gene clusters encoding all other known bacterial microcompartments. The carboxysomal genes of *H. neapolitanus* are tightly clustered and transcribed in the same direction, and they are known to be coordinately regulated. However, whether they form an operon has not been resolved. The experiments described in the following sections were undertaken to establish a detailed profile of carboxysome gene transcription and to examine the potential regulation strategy that might give rise to the differential levels of carboxysome protein constituents.

1.1 Steady-state Transcript Levels Indicate All Nine Carboxysomal Genes Are Transcriptionally active

Primer pairs for real-time RT-PCR were chosen using the Beacon Designer v4.0 software to probe individual carboxysomal genes as well as each intergenic region in

order to cover the entire *cso* cluster in *H. neapolitanus* (Figure 10). The specificity of each PCR primer pair was confirmed by melt curve analysis and also by electrophoresis on 2% agarose gels (data not shown). Absolute quantitative real-time reverse transcript (RT)-PCR was performed with linearized plasmid pTn1 [5] as standard. The standard curve for each primer pair spans five orders of magnitude while the linear regression value was equal to or greater than 0.98 (data not shown). Mock cDNA prepared without reverse transcriptase was used as an indicator of genomic DNA contamination. The absolute levels of transcripts were calculated from quadruplicate reactions performed with three independent RNA preparations (Figure 17), and all primer pairs tested consistently yielded Ct values that were 6-13 units higher than those obtained with mock cDNA. This indicated that the observed PCR amplifications were not from genomic DNA, and all carboxysomal genes were transcribed at detectable levels. The mRNA copy numbers of individual genes differed by one to two orders of magnitude, with mRNA copy numbers of *csoS2* and *csoS3* being lower than those of the upstream genes *cbbL* and *cbbS* and those of the downstream genes *csoS1C*, *csoS1A,* and *csoS1B*. Notably, transcripts were also measured for the *csoS4A* and *csoS4B* genes (formerly *orfA* and *orfB*, respectively), which were at levels comparable to those of *csoS2* and *csoS3*. This is the first experimental evidence showing that *csoS4A* and *csoS4B* are actively transcribed, since their deduced products have not been identified in both *H. neapolitanus* cell extract or purified carboxysomes. The relative transcript abundance of each carboxysome gene follows the trend of the calculated polypeptide composition of carboxysomes, except in the case of *csoS4A* and *csoS4B*. These results suggest that carboxysome gene expression in *H. neapolitanus* is regulated primarily at the transcriptional level.

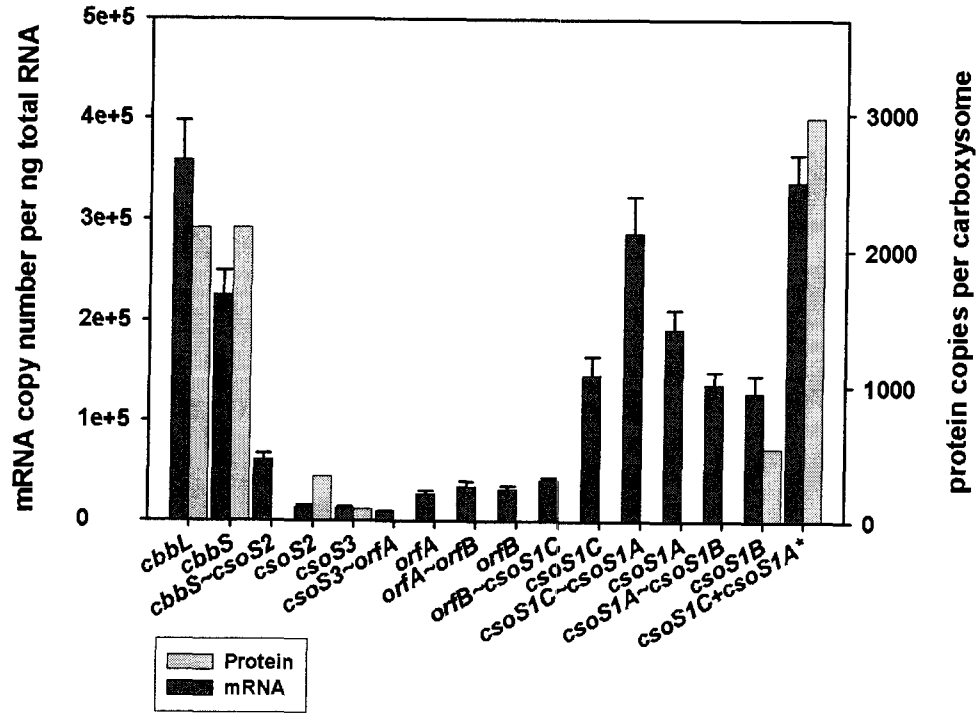


Figure 17. Transcript profile of the *Halothiobacillus neapolitanus cso* gene cluster.

Absolute steady state levels of individual gene transcripts and of the intergenic regions were determined by real time RT-PCR as detailed in Chapter II. The data represent the means obtained from quadruplicate determinations performed with three independent RNA preparations (± standard deviation). There is a strong correlation between transcript and carboxysome polypeptide abundance. The ***height*** of the ***bar*** indicated by * shows the sum of the *csoS1A* and *csoS1C* transcript levels for comparison with the combined CsoS1A and CsoS1C protein levels.

1.2 Transcript 5'-end Mapping Reveals a Single Transcription Start Site within the cso Gene Cluster

Previously, the promoter predictor program BDGP, which is based on *E. coli* promoters, had predicted 34 potential promoters for the carboxysome gene cluster in *H. neapolitanus* [12]. However, regulatory motifs vary from species to species, and *E. coli* based predictor programs including BDGP have been proven inaccurate for autotrophic bacteria [59,82]. To identify *H. neapolitanus* promoter(s) for the carboxysome gene cluster, the 5'-ends of transcripts were mapped via primer extension. According to the transcript levels of each carboxysomal gene measured with real-time RT-PCR, potential locations of promoters are the upstream regions of the *cbbL* gene and the *csoS1C* gene. To map the 5'-end of the transcript containing *cbbL*, primers designed to anneal to three positions (-16 to +13, +70 to +92, and +199 to +233, respectively, relative to the start codon of the *cbbL* gene, sequences see Appendix) were used. An adenine located 55 bases upstream from the ATG codon of *cbbL* was identified as the transcription start site (TSS) (Figure 18a). To identify potential internal promoter(s), which could explain the observed significant rise in steady state levels of the csoS1C/A/B transcripts, four primers which are complementary to different regions (-1 to +20, +93 to +120, +193 to +214, and +321 to +344, respectively, relative to the start codon of *csoS1C* gene, sequences see Appendix) were used in mapping the 5'-end(s) of mRNA that initiate upstream from *csoS1C*. However, no TSS was detected with these primers.

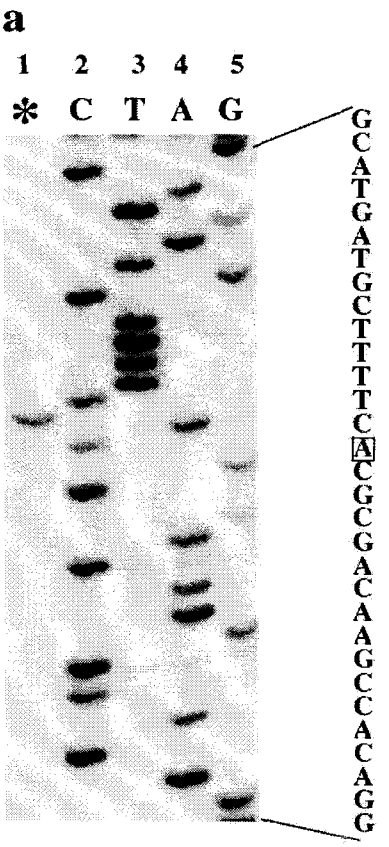


Figure 18. Transcript 5'-end mapping by primer extension.

a. A single extension product was detected with primer cbbLr164 upstream from the *cbbL* gene (lane 1) that corresponds to an A at position -55 relative to the *cbbL* translation start codon. Lanes 2-5 show sequencing reactions reformed with plasmid pTn1, which contains the *cso* operon of *H. neapolitanus*, as a template.

b. Comparison of -10 and -35 regions (*underlined*) upstream from the transcription start sites (*bent arrows*). *H.n. cso*, *Halothiobacillus neapolitanus cso* operon; m.c. *csoS1*, region upstream from the *csoS1* gene that precedes the RubisCO genes in the *cso* operons of four marine cyanobacteria; *E.c. lacI*, *E.coli lacI* promoter; *E.c.* σ^{70} promoters.

1.3 RNA-ligase Mediated 3'-RACE Reveals Multiple 3'-ends Downstream from cbbS and a Single 3'-end Downstream from csoS1B

The observation of a sharp drop and a significant rising in mRNA copies from *cbbS* to *csoS2*, and from *csoS4B* to *csoS1C*, respectively (Figure 17), suggested the occurrences of transcriptional termination in these two regions. Stem-loop structures, which might serve as Rho-independent bacterial terminators [54], are predicted to exist in the intergenic regions between *cbbS* and *csoS2* (Figure 19a) and between *csoS4B* and *csoS1C* (Figure 19b) by the RNAdraw program. Although no stable stem-loop structure is predicted to exist downstream of *csoS1B*, three subsequent stop codons in the same reading frame as the *csoS1B* gene (Figure 20b) are found immediately downstream from the *csoS1B* stop codon. Therefore, potential transcription termination events at three locations (between *cbbS* and *csoS2*, between *csoS4B* and *csoS1C*, and downstream from *csoS1B*) were probed by RNA ligase mediated 3' rapid amplification of cDNA ends(RLM-3'-RACE) [45]. A DNA adaptor was ligated to the 3'-end of mRNA by RNA ligase, and the first strand cDNA was synthesized with random primers. Then adaptorREV, which is designed to specifically anneal to the adaptor, was paired with one of three gene specific forward primers (cbbSf1872, orfBf6559, and 1Bf7641, sequences see Appendix) to amplify cDNA 3'-ends. Four abundant PCR products were obtained that represent 3'-ends located in the 5'-terminal portion of the *csoS2* coding region with primer cbbSf1872 (Figure 20a). This result, which suggested that bicistronic *cbbL/cbbS* transcripts frequently end in this area, could explain the significantly lower mRNA levels of the downstream genes (Figure 17). Despite the prediction of similar stem-loop structures downstream from *csoS4B*, no PCR products could be amplified reproducibly

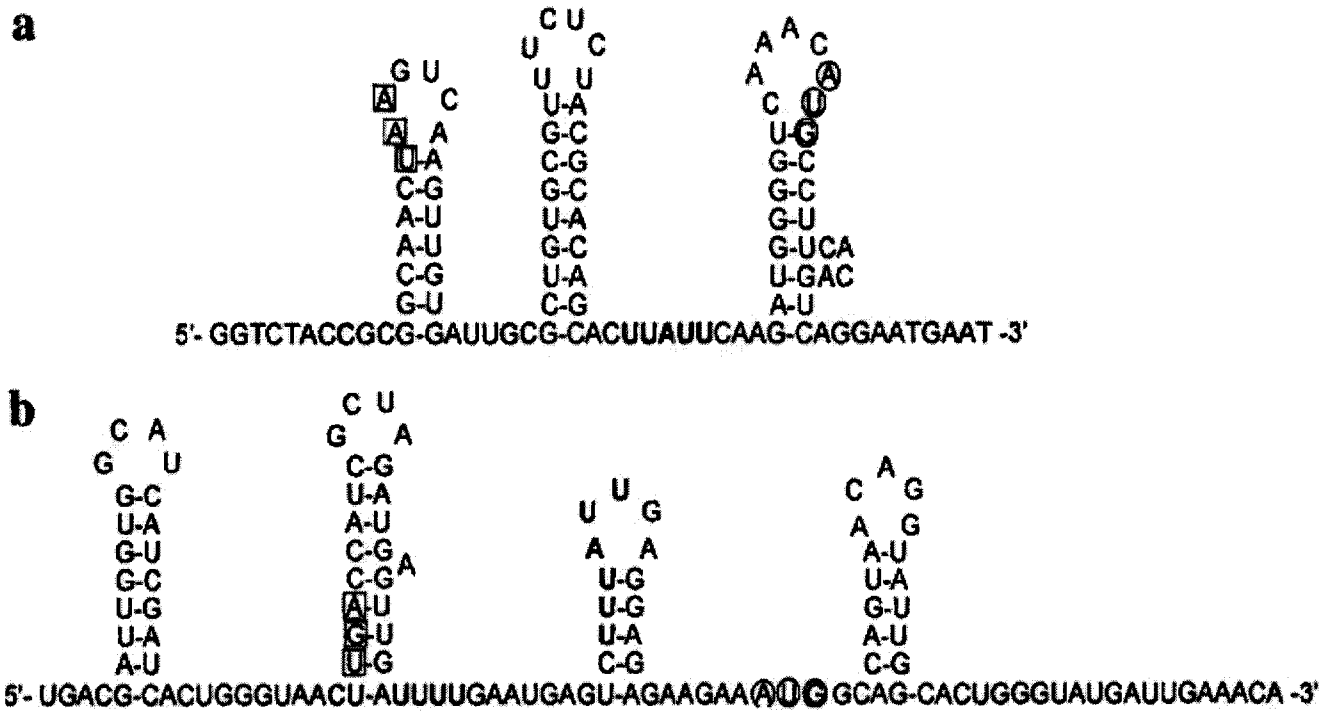


Figure 19. Predicted secondary structures of transcripts in the *Halothiobacillus neapolitanus cso* gene cluster and sequence downstream of *csoS1B*.

Predicted stem-loops between **a.** *cbbS* and *csoS2* (-98.24 kJ/mol at 37° C), and **b.** *csoS4B* and *csoS1C* (-93.87 kJ/mol at 37° C). The nucleotides of the stop codon for the respective upstream and of the start codon for the respective downstream gene are *boxed* and *circled* in, respectively. A uracil tract (*bold*) follows the G/C-rich stem-loop that may function as an intrinsic transcriptional terminator.

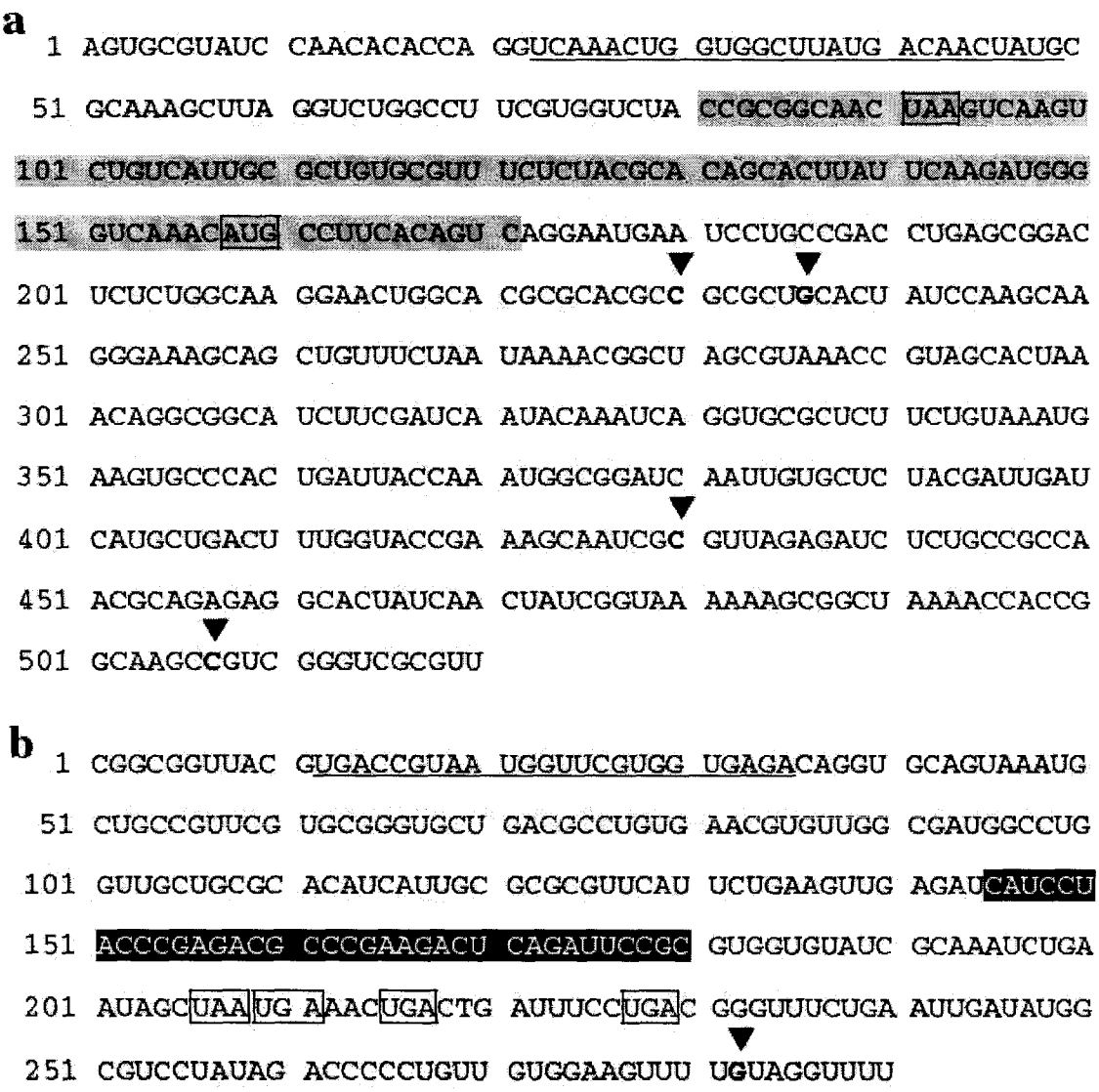


Figure 20. Transcript 3'-end mapping using RLM-3'-RACE.

Transcript 3'-ends are indicated in *bold letters* under the *down arrows*. Stop and start codons are *boxed* in, and primer annealing sites are *underlined*. a. Four prominent transcripts ending in the *csoS2* coding sequence were identified with primer cbbSf1872. The area of the predicted stem-loop structure is highlighted in grey. b. One stop site was identified down stream from the multiple stop codons of *csoS1B* using primer 1Bf7641. No potential Rho-independent terminator structures are predicted for this region. The C-rich region that may serve as a *rut* site is shown in white font in a *black box*.

that would be indicative of transcript species with 3' ends located in the intergenic region between *orfB* and *csoS1C*. On the other hand, although no stable stem-loop structure is predicted to exist downstream of *csoS1B*, a single transcript 3'-end was mapped with primer 1Bf7641 to 74 bp downstream from the first of four consecutive *csoS1B* stop codons (Figure 20b).

1.4 Constructions in Promoter Probe Vectors Prove that the Main cso Promoter Is Active in Escherichia coli but Fail to Identify Any Internal Promoter

Although the searching for promoter(s) within carboxysome gene cluster via primer extension resulted in only one main promoter upstream from *cbbL* gene and no internal promoter upstream from *csoS1C* gene, two selected regions upstream from *csoS1C* and *csoS4A* genes, respectively, were tested for potential promoter activity *in vivo*. Three DNA fragments from *H. neapolitanus*, Pro_{cso}, $Frag_{1C}$ and $Frag_{4A}$, and *lac* promoter sequence (Pro_{lac}) from *E. coli* as a positive control were cloned into HindIII / EcoRI sites of broad-host-range promoter probe vectors pPROBE-KT, pPROBE-NT, and pPROBE-OT [53], where the *gfp* gene was used as promoter activity reporter gene. Promoter activities of all the constructs in *E. coli* TOP10 cells were estimated by observing fluorescence of the green fluorescent protein (GFP) by confocal microscopy. The pPROBE-KT and pPROBE-NT constructs did not show any green fluorescence except in the case of the positive control Pro_{lac}. On the other hand, the Pro_{cso} in pPROBE-OT did show restricted promoter activity in *E. coli* (Figure 21b), as one would expect from comparison of the *cso* promoter from *H. neapolitanus* and the *trc* promoter from *E. coli* (Figure 18b). Neither $Frag_{1C}$ nor $Frag_{4A}$ showed promoter activity in *E. coli*

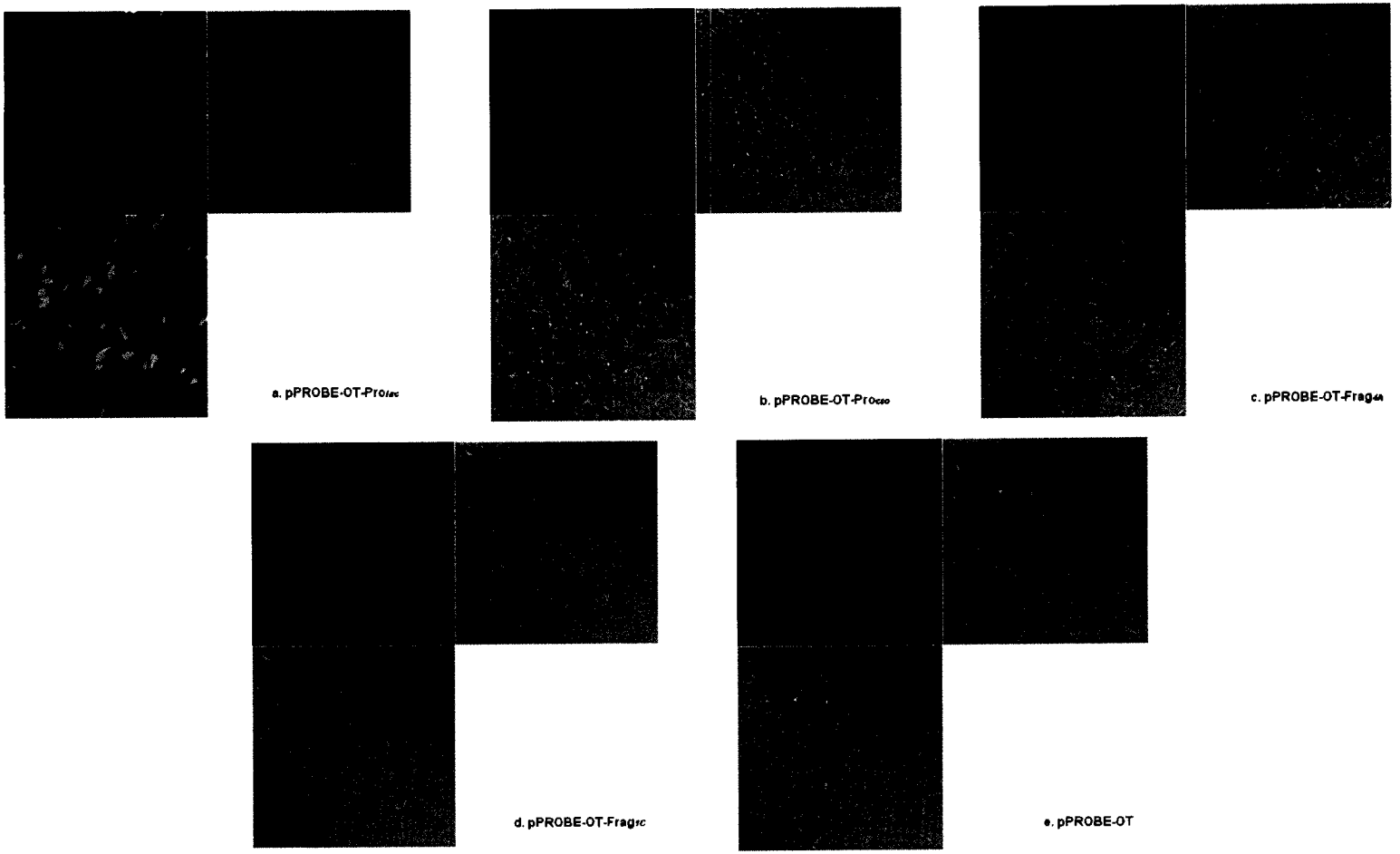


Figure 21. *E. coli* TopTen cells transformed with promoter constructs in pPROBE-OT vector.

a. *lac* promoter as positive control; b. *cso* promoter from *H. neapolitanus*; c. DNA fragment from upstream of *csoS4A* gene; d. DNA fragment from upstream of *csoS1C* gene; e. no insertion as negative control.

under the conditions tested (Figure 21c, d). Promoter probe vectors were also electroporated into *H. neapolitanus* cells; however recovery of transformed plasmid was not successful.

1.5 No Internal Promoter Activity Is Detected in the cso Promoter Deletion Mutant under the Condition Tested

The promoter probe vector failed to detect promoter activity of any selected internal region of the carboxysome operon in *E. coli*; however, the existence of internal promoter(s) cannot be ruled out since sigma factors in *E. coli* might not be able to recognize the *H. neapolitanus* promoter. To address this issue, an *H. neapolitanus cso* promoter deletion mutant was generated by replacing the *cso* promoter (-35 box till -10 box) with a Kan^r cassette transcribed in the opposite direction and its own promoter (mutant generated by Sarah Neidler). In this mutant, any carboxysomal gene expression should be the result of unknown cryptic promoter(s). The *cso* promoter deletion mutant does not grow in air but is able to grow in air supplemented with 5% CO_2 (data from Sarah Neidler, Figure 22). Immunoblot analysis of crude extract revealed no carboxysomal protein expression, including carboxysomal RubisCO (Figure 23 b, d, and e). The mutant is able to survive without the carboxysomal RubisCO because it possesses a second RubisCO-encoding gene (*cbbM*) that is not part of the *cso* operon and encodes a Form II RubisCO enzyme [4]. CbbM is expressed in the mutant at higher levels than in the wild type (Figure 23c), presumably to compensate for the absence of Form I RubisCO. Absence of carboxysomal protein expression in the *cso* promoter deletion mutant also indicated no internal promoter activity in the *cso* operon under the conditions tested.

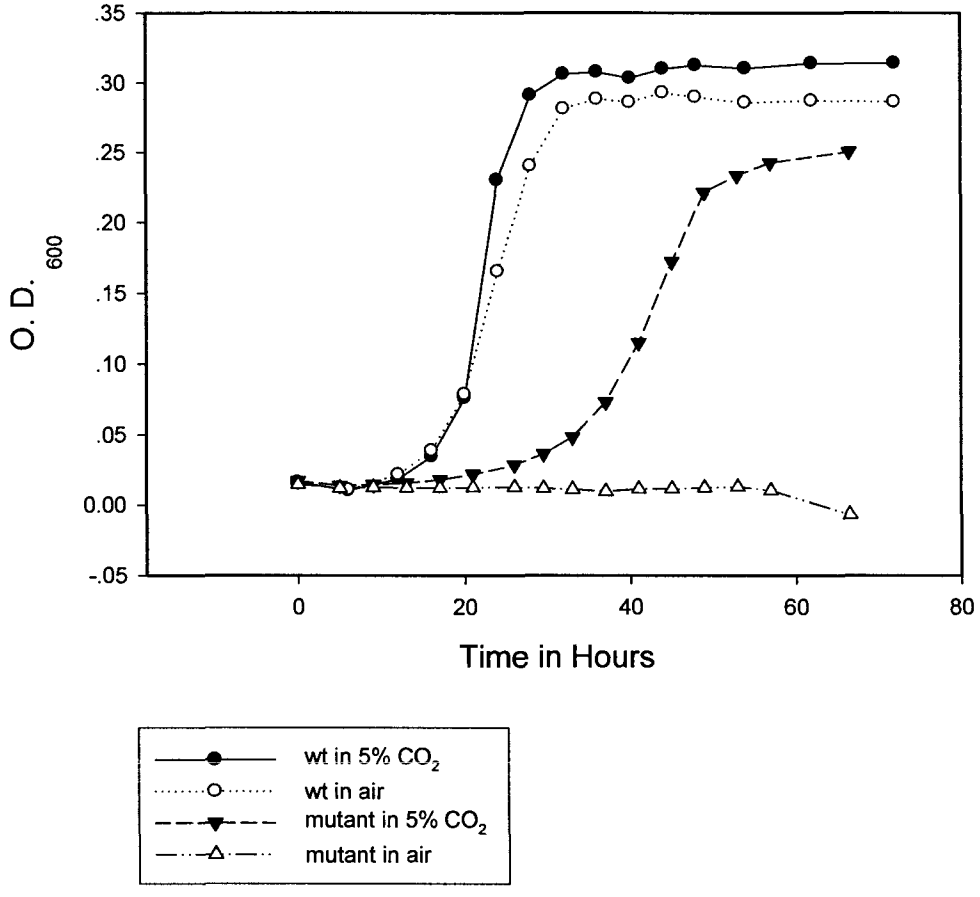


Figure 22. Growth curve of *H. neapolitanus* wild type and *cso* promoter deletion mutants.

Wild type (wt) and *cso* promoter deletion mutant (mutant) *H. neapolitanus* grow in air with or without 5% CO_2. The mutant cannot grow in air without an extra supplement of CO_2, and it also grows slightly slower than wild type with 5% CO_2.

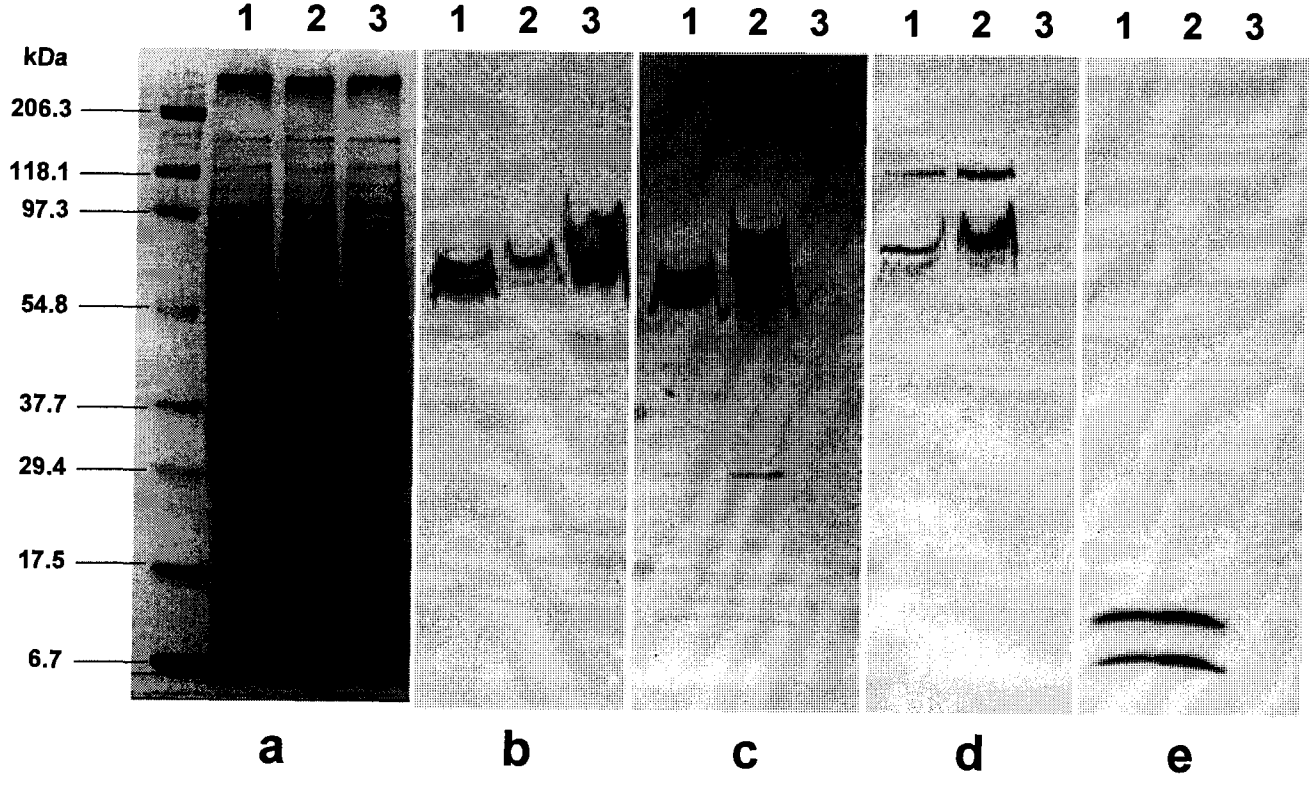


Figure 23. Immunoblots of crude extract with *H. neapolitanus* wild type and *cso* promoter deletion mutant.

a. Crude extracts separated on SDS-PAGE; b-e. immunoblot with antibody against CbbM (b), CbbL (c), CsoS2 (d), and CsoS1 (e). Lane 1, wild type cells grown with 5% CO_2; lane 2, wild type cells grown under air; lane 3, *cso* promoter deletion mutant cells grown with 5% CO_2.

2. Composition Study of the Carboxysome of *Halothiobacillus neapolitanus*

As shown in the previous section, steady-state transcript levels indicated all nine carboxysomal genes are transcriptionally active, including *csoS4A*, *csoS4B*, and all three BMC genes (*csoS1C*, *csoS1A*, and *CsoS1B*). The duplicate genes *csoS4A* and *csoS4B* are transcribed slightly higher than the upstream genes, *csoS2* and *csoSCA*, but considerably lower than the mRNA levels for the downstream BMC gene set. However, previous investigations have not detected their gene products in either whole cells or purified carboxysomes. This might be due to the difficulty in separating the 9 kDa polypeptide (CsoS4A or CsoS4B) from the 10 kDa polypeptide (CsoS1A or CsoS1C) on SDS-PAGE In addition, the *csoS4* gene products could be extremely less in abundance, thereby precluding their detection. On the other hand, although the transcript analysis suggested all three BMC gene paralogs are transcribed at approximately the same levels, it is still not clear if both CsoS1A and CsoS1C are shell components, due to difficulty in separating and distinguishing them from each other, since they differ by only two amino acids, the 3rd and the 97th amino acid residues (Figure 24). If both CsoS1A and CsoS1C are shell components, then at what ratio are they packed within the shell would be another question. To answer these questions, the following mutagenesis, recombination, and immunoblotting experiments were performed.

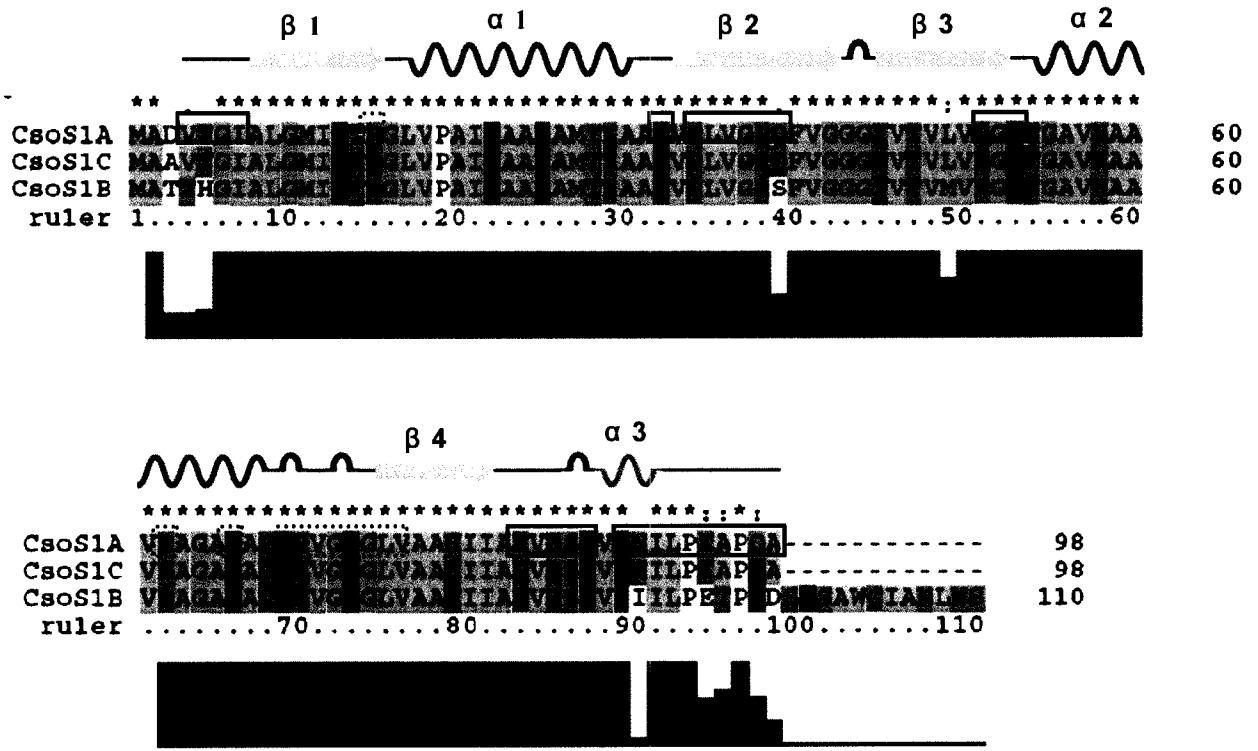


Figure 24. Amino acid sequence alignment of three BMC proteins from *H. neapolitanus*.

CsoS1A and its paralogs CsoS1B, and CsoS1C in carboxysome operon of *H. neapolitanus* belong to protein family BMC (Pfam ID: 00936). CsoS1A and CsoS1C are nearly identical, and only differ in two amino acids. And CsoS1B has a 12 amino acid C-terminal extension. Sequence alignment generated with Clustal X v1.83. Indicated above the alignment are the helical (waves) and β-sheet (arrows) regions deduced from the CsoS1A crystal structure [80]. According to this structure, accessible residues on the concave side are in solid box, while those on the convex side are in dashed box [80].

2.1 The csoS1A-TC Tag Mutant Confirms that All Three BMC Paralogs in H. neapolitanus Are Present in Carboxysomes

Previous studies clearly showed that the CsoS1 proteins are major shell components, and that they are responsible for polypeptide bands migrating at around 7 kDa and 15 kDa when separated by SDS-PAGE (Table 1). The 15 kDa polypeptide band was assigned as the CsoS1B protein; whether or not the 7 kDa band contains both CsoS1A and CsoS1C polypeptides was unknown since they are identical in primary structure. To answer this question, an *H. neapolitanus csoS1A*-TC tag mutant was generated by fusing a tetracysteine (TC) motif Cys-Cys-Pro-Gly-Cys-Cys (RM#341 and #342) on the C-terminal of the *csoS1A* gene. This mutant grows as well as wild type in air with or without supplemented with 5% CO_2 (Figure 25). Immunoblotting analysis of crude extract with anti-CsoS1 antiserum revealed that both CsoS1A-TC and CsoS1C are expressed (Figure 26a) and that they are present at approximately a ratio of 1:6, respectively, as estimated by densitometry (Table 2) in enriched mutant carboxysome separated by SDS-PAGE (Figure 26b).

2.2 CsoS4A and CsoS4B Are Minor Components of the Carboxysome

Both *csoS4A* and *csoS4B* genes were inserted first into the pProEX-HTb prokaryotic expression vector, which results in an N-terminal his_6 tagged fusion of the gene preducts. Various expression conditions were tested; however, none of them resulted in recombinant protein expression. Then, the two genes were inserted into the pET-22b(+) vector and heterogeneously expressed in the *E. coli* BL21(DE3) strain with a C-terminal his_6-tag (Leu-Glu-His_6). Typically, 20-50 mg of recombinant protein can be purified from

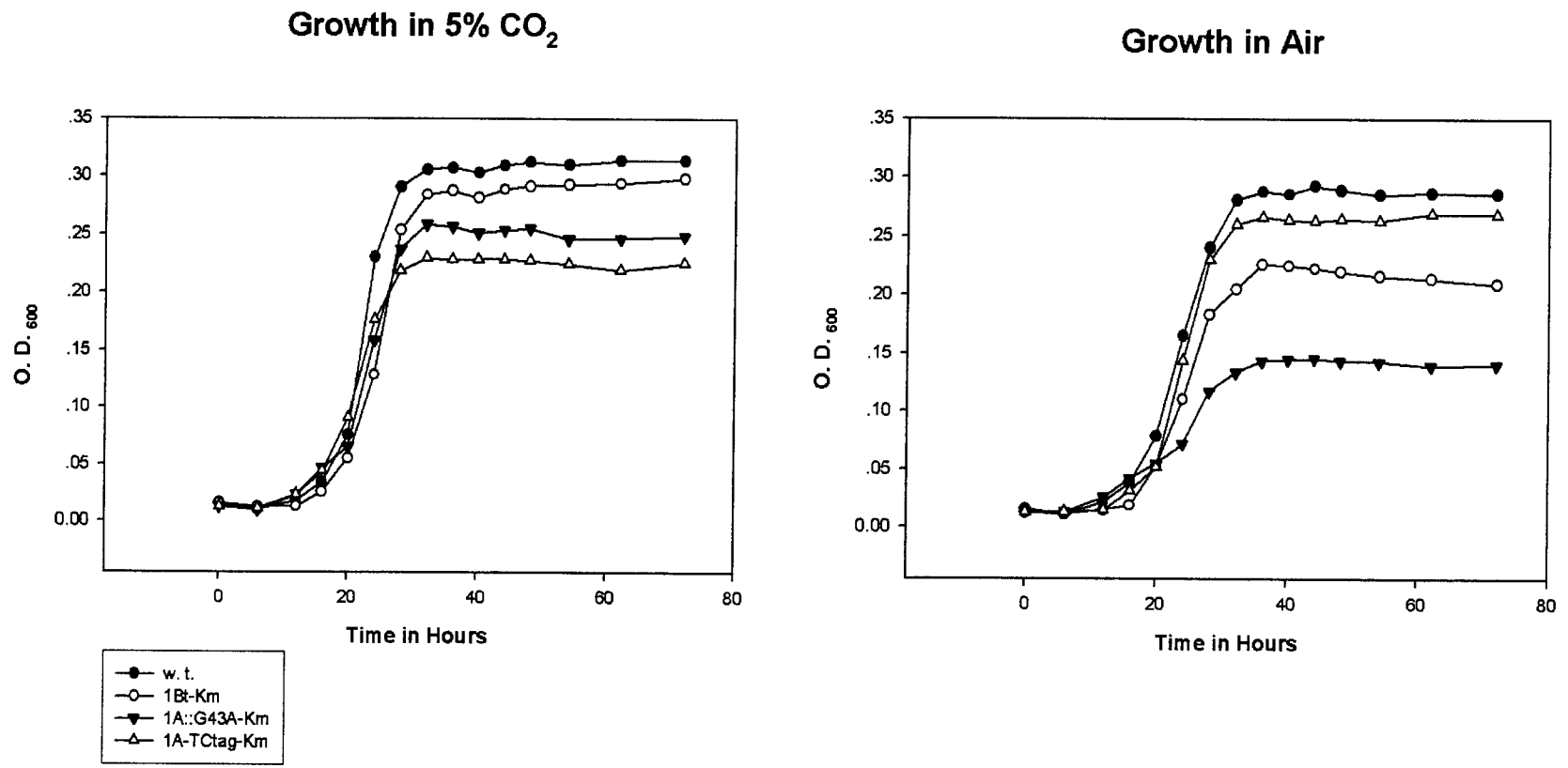


Figure 25. Growth curve of *H. neapolitanus csoS1* gene mutants.

Wild type (w. t.) and mutants *H. neapolitanus* grow in 5% CO_2 (left) and in air (right). 1Bt-Km: *csoS1B* truncated mutant; 1A::G43A-Km: *csoS1A::G43A* mutant; 1A-TC tag-Km: *csoS1A* tetracysteine fusion mutant. All three mutants grow as well as wild type *H. neapolitanus* with 5% CO_2, while in air, *csoS1B* truncated mutant and *csoS1A::G43A* mutant cannot reach same cell density as wild type.

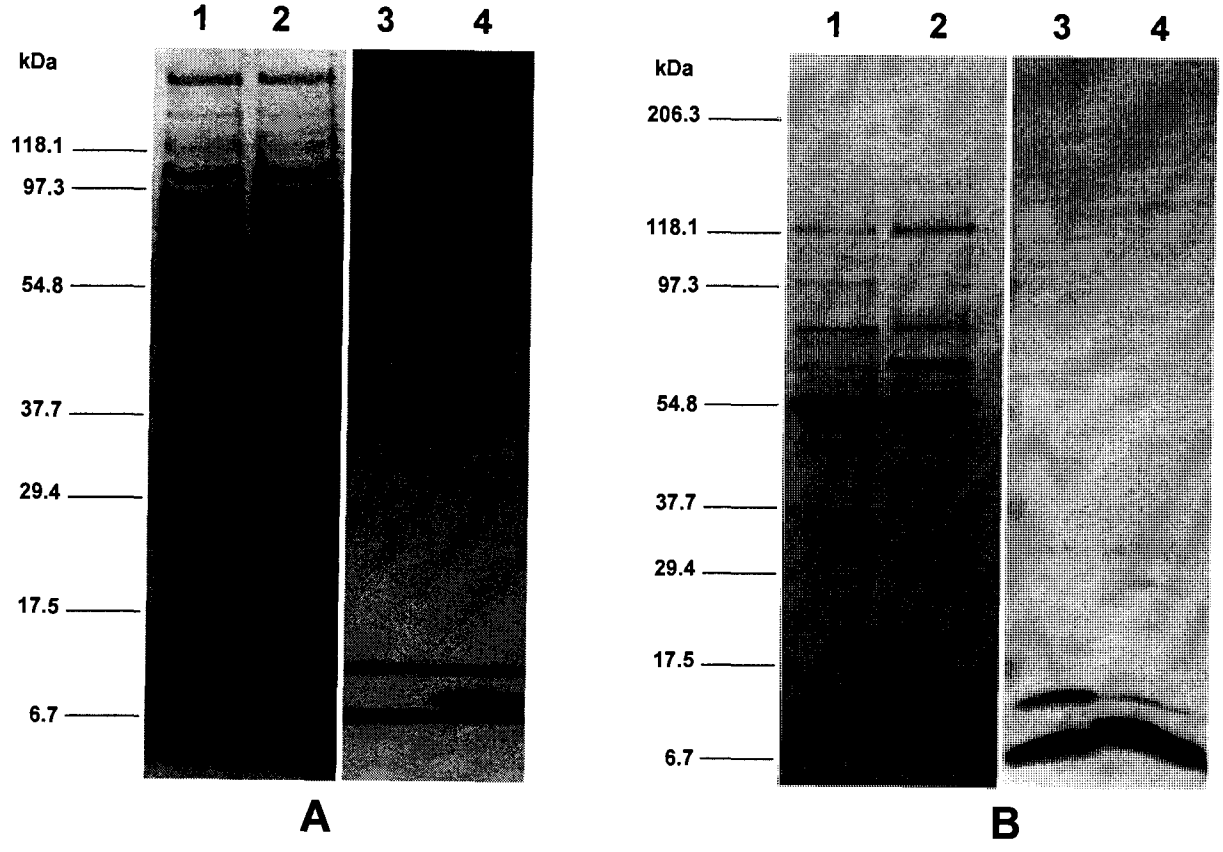


Figure 26. Immunoblots of crude extract and purified carboxysome samples with *H. neapolitanus csoS1A*-TC tag mutant. The polypeptide composition of crude cell extract (A) and purified carboxysomes (B), respectively, from wild type (lanes 1 and 3) and *csoS1A*-TC tag mutant (lanes 2 and 4). Lanes 1 and 2, stained polypeptides separated by SDS-PAGE; lanes 3 and 4, immuno-blot probed with anti-CsoS1A antiserum.

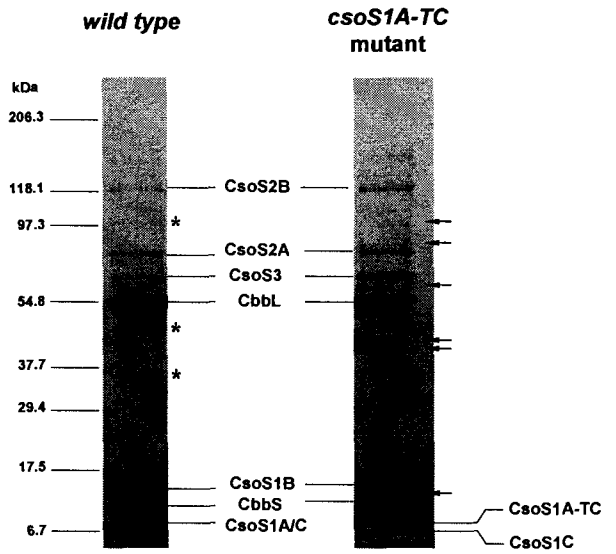


Table 2. Stoichiometric distributions of carboxysomal components in *H. neapolitanus wild type* and *csoS1A*-TC tag mutant.

Carboxysomal Components	Abundance in wild type	Abundance in *csoS1A*-TC mutant
CsoS2B	2.9%	9.8%
CsoS2A	3.9%	5.6%
CsoS3	2.1%	3.7%
CbbL	49.3%	39.3%
CsoS1B	6.4%	21.0%
CbbS	11.6%	9.0%
CsoS1A	23.9%	1.7% [a.]
CsoS1C		9.9%

Purified wild type carboxysomes (10 μg) and carboxysome enriched sample (10 μg) from *csoS1A*-TC mutant were separated on a denaturing 4-20% gradient SDS-PAGE and stained with GelCode Blue. The relative abundance of carboxysomal components was estimated by densitometry of stained polypeptide patterns using BioRad's Quantity One v 4.6.3 software. The CsoS4A and CsoS4B proteins were not listed as they represent very low abundance proteins and cannot be separated from CsoS1A/C protein band here.

[a.] abundance for fusion protein CsoS1A-TC.

1 liter of induced culture (Figure 27). The recombinant proteins were used for generating antisera in rabbits. USM#61 and USM#62 were antisera generated against recombinant CsoS4A, and USM#59 and USM#60 were antisera generated against recombinant CsoS4B. All four antisera recognized both rCsoS4 proteins; however, the CsoS4A and CsoS4B could not be distinguished from one another based on signal strength. Among the four antisera, USM#59 gave the best signal strength, and therefore it was used for detecting the CsoS4 proteins in subsequent studies. Since the deduced gene product of *csoS4A* or *csoS4B* should migrate at the same position as CsoS1A/C protein when separated by SDS-PAGE, it was important to test for cross-activity of this USM#59 with the CsoS1 proteins (Figure 28). USM#59 did recognize recombinant CsoS1A/C and CsoS1B; however, after cleavage of the his_6-tag with rTEV protease from these CsoS1 proteins, this binding activity of USM#59 also lost (Figure 28). This result suggested that the rCsoS4 antiserum does bind to native CsoS1 proteins. USM#59 was used for detecting the CsoS4 proteins in carboxysome shell-enriched fraction via immunoblotting coupled to two dimensional gel electrophoresis (Figure 29). Two protein spots were recognized on 2-D PAGE, and the observed pI values of these two were close to the calculated pI values of CsoS4A and CsoS4B peptides, 5.7 and 5.2, respectively. This correlation suggested that both CsoS4A and CsoS4B, although extremely low in abundance, are indeed components of the carboxysomes shell and that these two proteins are present at approximately 1:1 ratio.

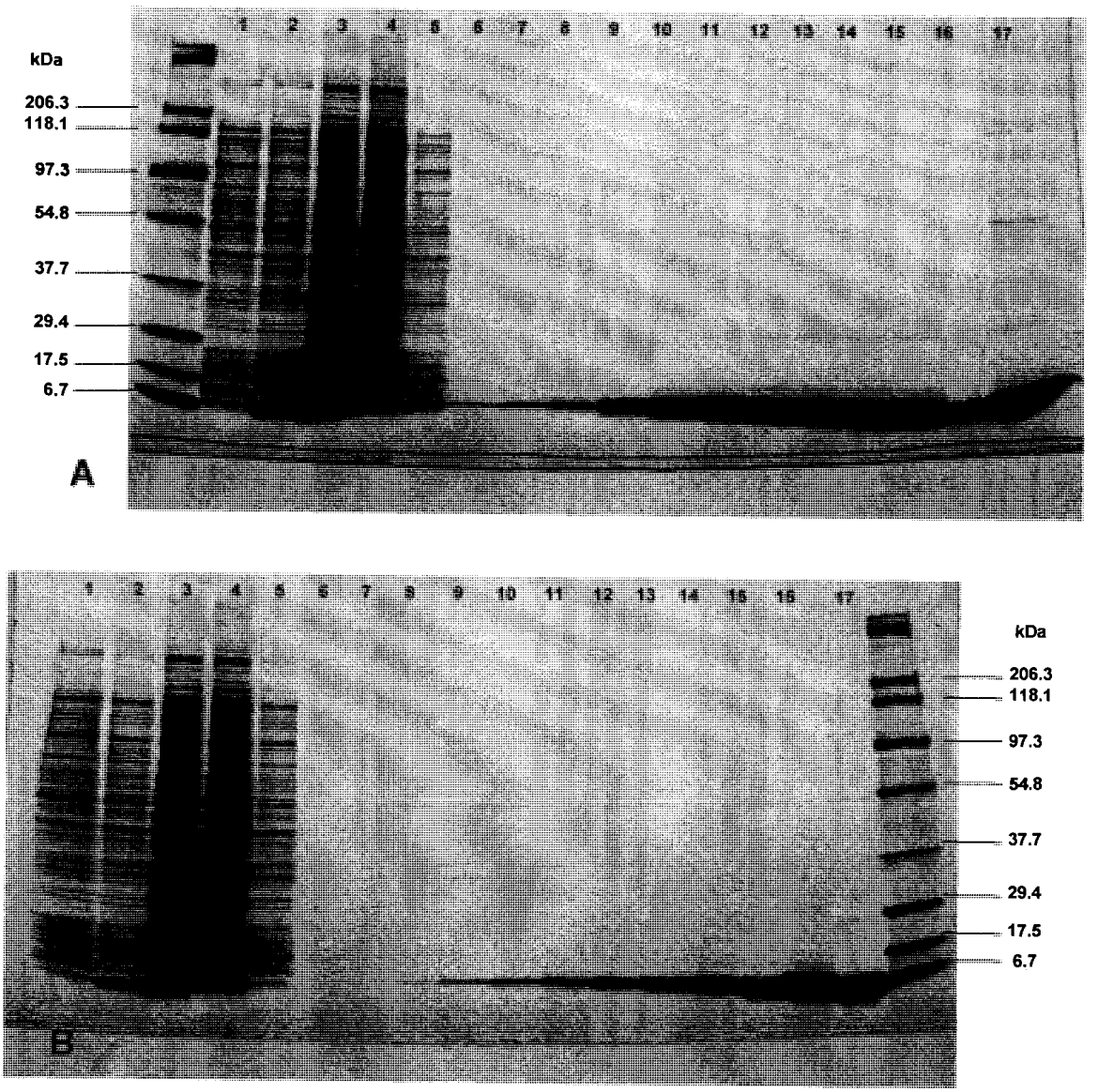


Figure 27. Purifications of recombinant CsoS4A and CsoS4B.

Purifications of recombinant CsoS4A (A) and CsoS4B (B) from *E. coli* cells. Lane 1, un-induced cell; lane 2, induced cell; lane 3, supernatant of cell lysate; lane 4, flow through from loading the Ni-NTA resin column; lane 5, wash I; lane 6, wash II; lane 7, wash III; lane 8-17, elutions with 20, 30, 40, 50, 60, 70, 80, 90, 100, and 250 mM imidazole, respectively.

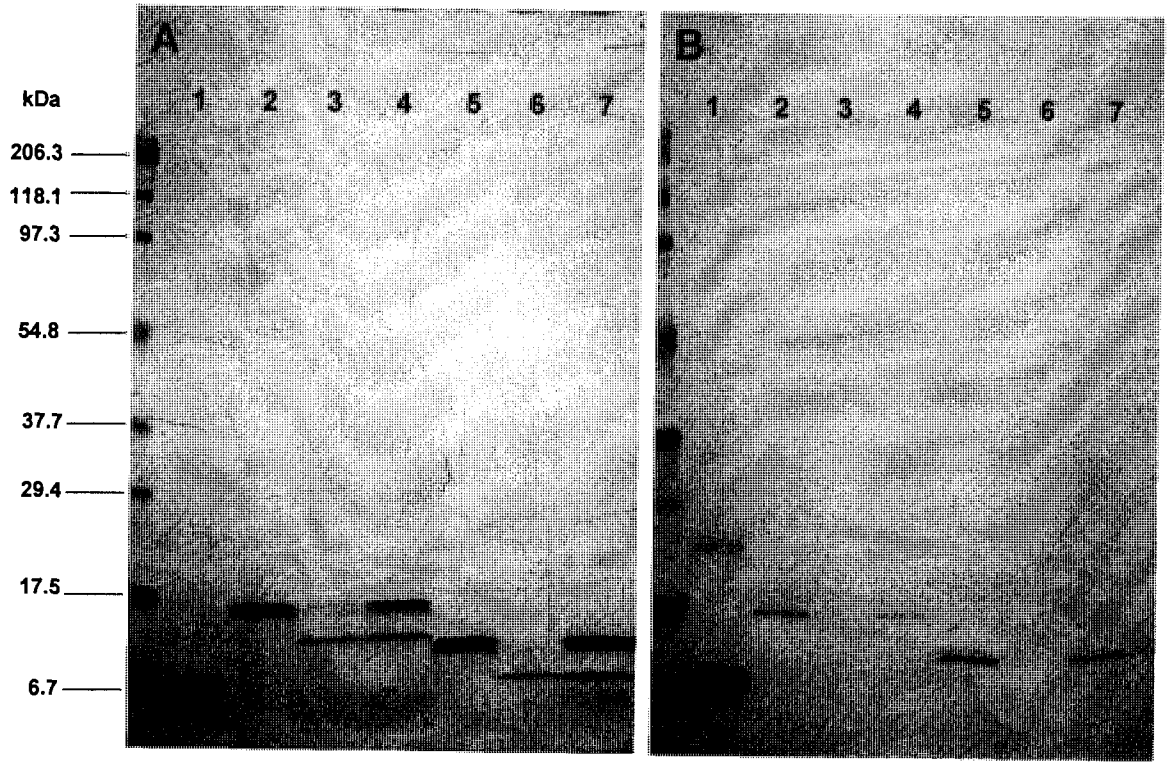


Figure 28. CsoS4 antiserum cross-reactivity test.

CsoS4 antiserum was test for its cross-reactivity against recombinant CsoS1 protein with or without his-tag. A, SDS-PAGE separation of tested samples; B, immunoblot developed with CsoS4 antiserum. Lane 1, 500 ng of rCsoS4A as positive control; lane 2, rCsoS1B with his-tag; lane 3, rCsoS1B without his-tag; lane 4, rCsoS1B digested with rTEV protease; lane 5, rCsoS1A/C with his-tag; lane 6, rCsoS1A/C without his-tag; lane 7, rCsoS1A/C digested with rTEV protease.

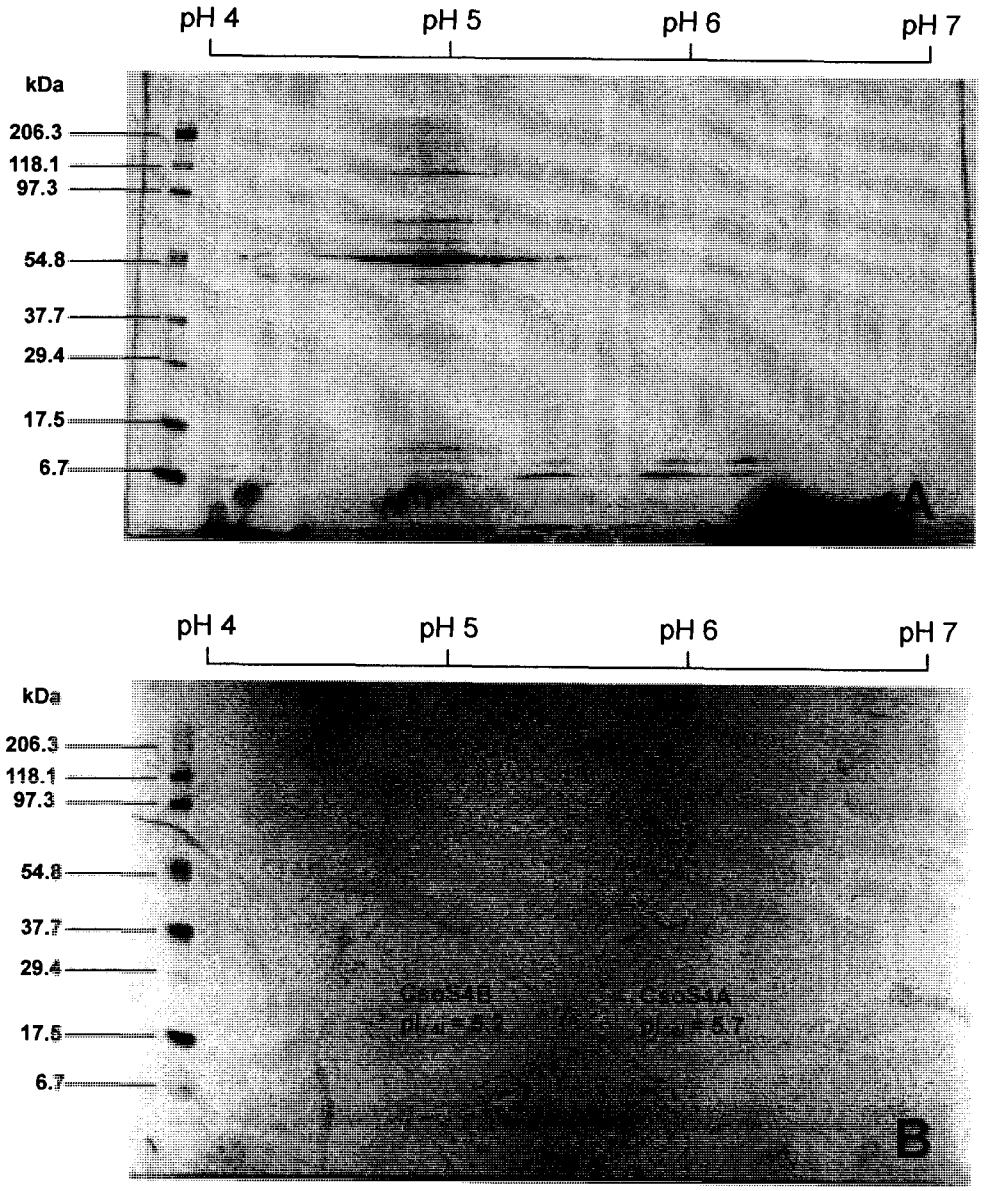


Figure 29. Detection of CsoS4A and CsoS4B proteins in carboxysome shell. Two-dimensional gel electrophoresis (A) combined with immunoblotting (B) indicated that the proteins CsoS4A and CsoS4B, although not abundant carboxysome components, are present in the shell.

3. Investigation of Carboxysome Function in *Halothiobacillus neapolitanus*

Comparative biochemical and genomic analysis shows that a BMC domain containing gene is always present in multiple copies within the BMC gene cluster or operon. For example, in the case of the carboxysome operon in *H. neapolitanus*, there are three BMC genes in the operon and the protein products of all three are shell components. The biological and/or structural implication of this observation of various paralogs present in the shell is not understood yet. Previous yeast two hybrid assays of carboxysomal proteins from *H. neapolitanus* provided some clue: RubisCO small subunit CbbS and carboxysomal carbonic anhydrase CsoSCA both interacted with CsoS1B, but not with CsoS1A/C *in vivo* (Eric B. Williams, dissertation). Since the only difference between CsoS1B and CsoS1A/C is the 12 amino acid C-terminal extension in the former. This observation suggests that the C-terminal of CsoS1B might have a distinct biological function within the carboxysome other than just being a shell component. On the other hand, crystal structures of BMC proteins have recently been determined for both α-carboxysome (CsoS1A of *H. neapolitanus*) [80] and β-carboxysome (CcmK1, CcmK2, and CcmK4 of *Synechocystis* sp. PCC 6803) [42,77]. All of these BMC proteins were shown to form hexamers arranged in a cyclic manner with a 4-7 Å pore at the central axis of symmetry (Figure 7). These hexamers were observed to pack tightly into two-dimensional molecular layers, which led to the hypothesis that they are the building blocks of the flat facets of the icosahedral carboxysome shell [42,80], and the central pore, which has a conserved positive charge, might serve as transport channels for substrates and products of CO_2 fixation reaction in and out of the carboxysome [42,80]. Although crystal structures of BMC proteins provided us a detailed insight of the carboxysome

shell, two questions still remain unanswered: (i) Do the BMC hexamers pack in a uniform orientation or in alternating orientation to form the shell? (ii) Given the fact that the two sides of BMC hexamers differ in shape and electrostatic potential, which side of the molecular layer faces outward if the hexamers pack in a uniform orientation. Later, the elucidation of crystal structures of recombinant CsoS4A and its homolog CcmL in β-carboxysome provided the missing piece to establish a structure based on a preliminary model of carboxysome shell assembly [77]. In this model, CsoS4A and CcmL constitute the 12 pentameric vertices of the carboxysome shell and tend to close the icosahedral body [77]. However, the proposed function of CsoS4/CcmL protein has not been confirmed experimentally. The following experiments were undertaken to collect information that might help answer the "multiplicity" question, the "sidedness" question, and molecular transport question of the carboxysome shell as well as to investigate the CsoS4 protein function *in vivo*.

3.1 The CsoS1B C-terminal Extension Is Not Essential for Packing CsoSCA within Carboxysome Shell

The reason for *H. neapolitanus* including all three BMC proteins in the carboxysome shell is unknown; however, yeast two-hybrid assays suggested that the 12 amino acid C-terminal extension of CsoS1B might be involved in structural or biochemical function. Modeling of the CsoS1B strucutre using the CsoS1A crystal structure as template revealed that the two structures are nearly identical until residue 91 (Ile91) in both sequences (Figure 30). Beginning from this point, the CsoS1A C-terminus is extended as

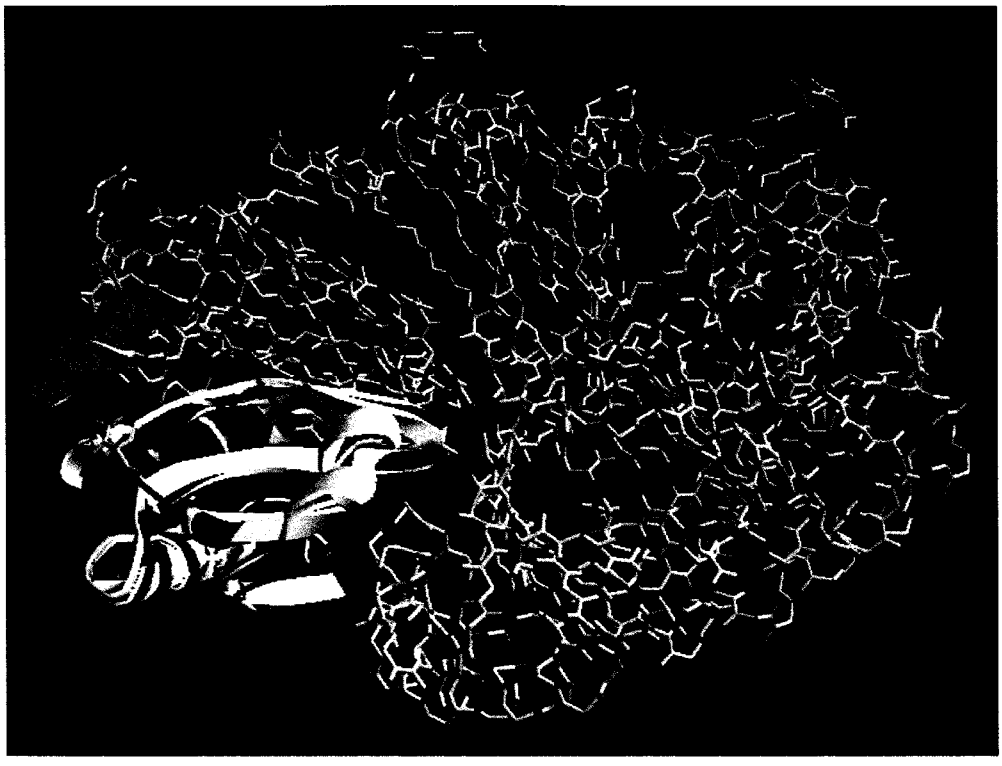


Figure 30. Modeling CsoS1B structure over reported CsoS1A structure.
Two structures are nearly identical until the Ile91 residues on both proteins. Residues before and after Ile91 of CsoS1A are shown in purple and cyan color, respectively, and residues before and after Ile91 of CsoS1B are showed in yellow and green color, respectively. CsoS1A hexamer is shown in gray backbone only, and the concave side is facing upward and outward in this figure.

a coil towards the central axis of the hexameric symmetry, while CsoS1B C-terminus is extended toward the edge of the hexamer. Moreover, residues 100-106 form an extra helix (Figure 30). Therefore, the predicted CsoS1B structure is more similar to the reported structure of CcmK4 from *Synechocystis* sp. PCC6803 [42]. To test this hypothesis, a CsoS1B truncated mutant was generated by deletion of the last 12 amino acids. This mutant can grow similarly to the wild type in air supplemented with 5% CO_2. When grown in ambient CO_2, the mutant does not reach the same cell density as wild

type cells (Figure 25). Carboxysomes from the CsoS1B truncated mutant did not show any morphological difference from wild type carboxysomes (Figure 31), which is in agreement with SDS-PAGE separation of mutant carboxysome and wild type carboxysome (Figure 32). Immunoblotting with anti-CsoSCA serum revealed that CsoSCA is still a component of carboxysomes even in the absence of the CsoS1B C-terminal extension (Figure 32). Comparing the polypeptide patterns and microscopic examination of intact mutant carboxysomes also suggested that the interaction between the C-terminal of CsoS1B and RubisCO small subunit CbbS is not essential for packing RuBisCO enzyme in carboxysomes either, even if such an interaction does exist.

3.2 Selective Labeling of Carboxysome Surface with Primary Amine Reactive Chemicals

The crystal structure of CsoS1A raised the sidedness question about the carboxysome shell [80]. One side of the hexamer has a major depression at the center appearing to be strongly concave while the other side has a minor depression appearing to be nearly convex (Figure 7B). Moreover, an uneven distribution of charged amino acid side chains resulted in an electrostatic difference between the two sides: the concave side is characterized by mostly positive electrostatic potential, whereas the convex side is characterized by slightly negative electrostatic potential (Figure 7B, 24). Considering most of carboxysomal proteins carry a net negative charge at the pH of the *H. neapolitanus* cytosol, which is about 7.8 [35], Tsai et al. argued that the positively

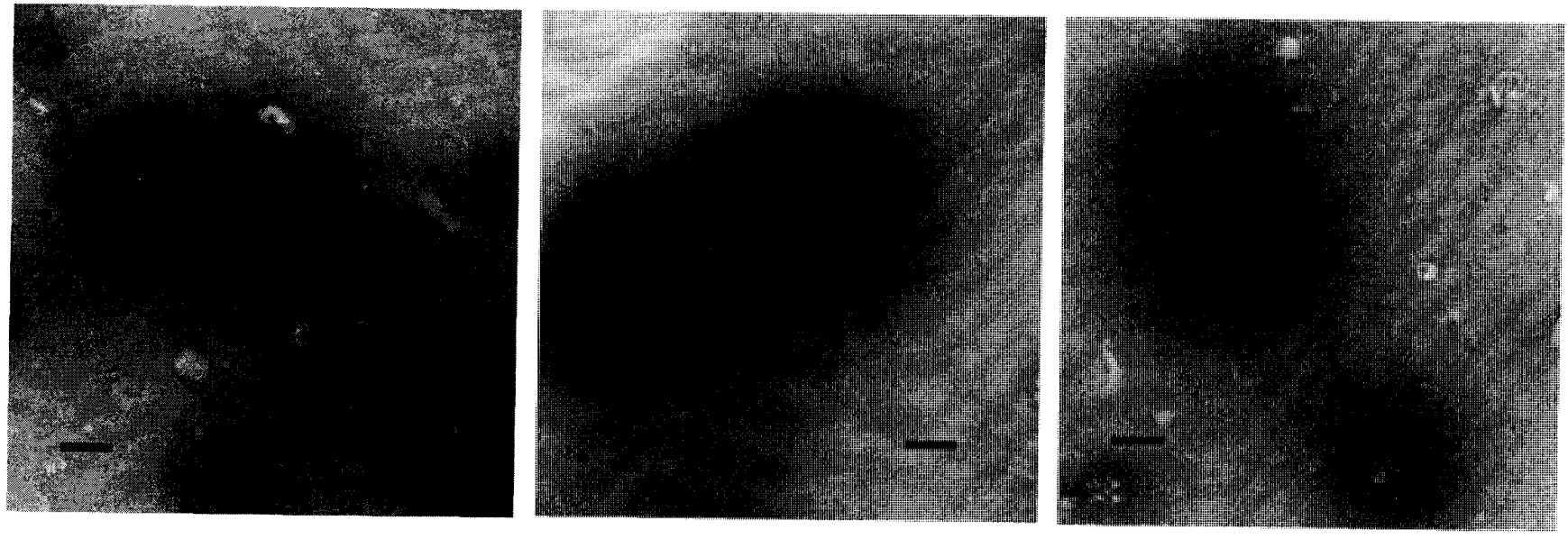


Figure 31. Observing carboxysomes from *csoS1B* truncated mutant under TEM.

Mutant carboxysome enrichment sample was negatively stained with 1% (w./v.) ammonium molybdate and observed under TEM. Bar indicates 100 nm.

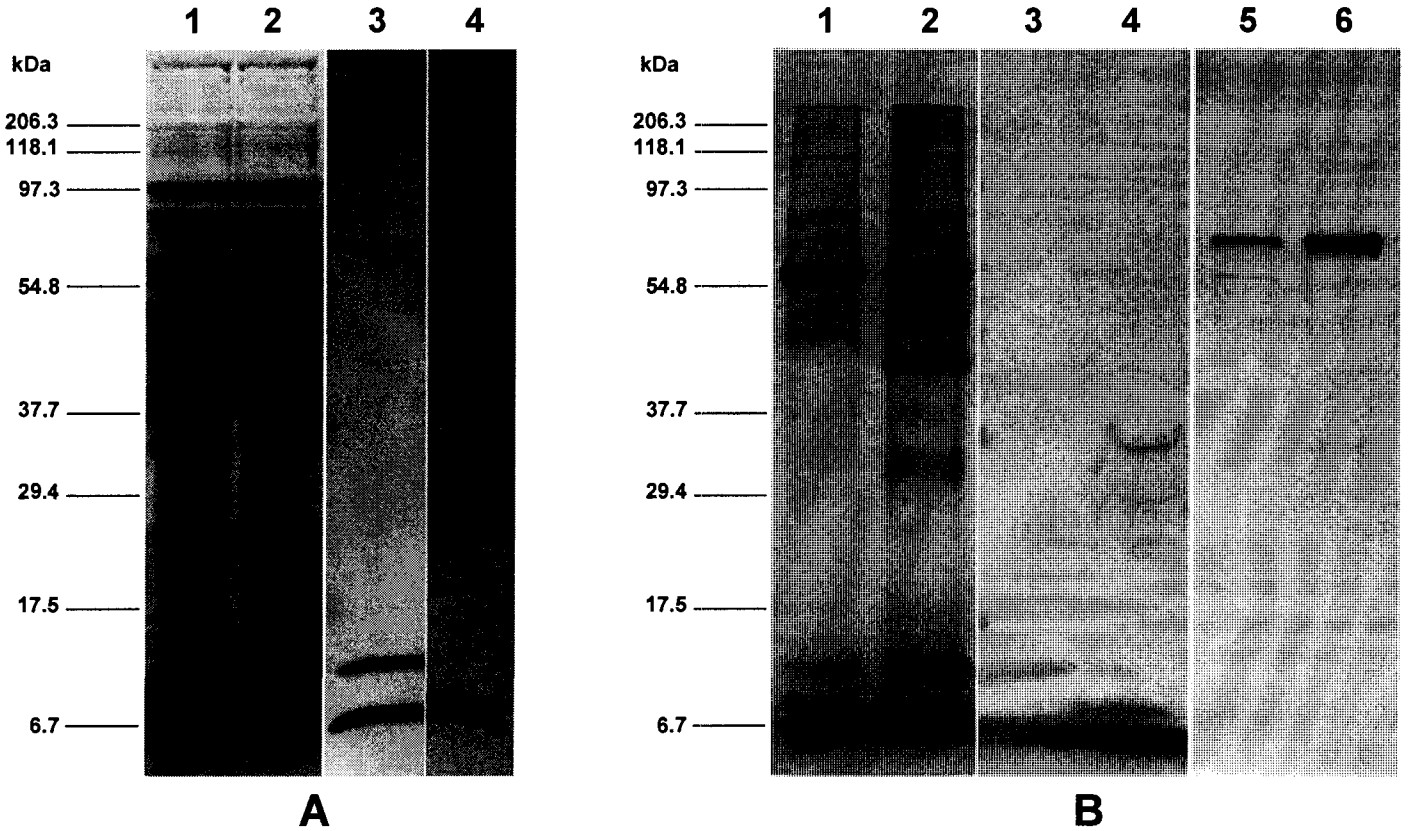


Figure 32. Immunoblots of crude extract and purified carboxysome samples with *H. neapolitanus csoS1B* truncated mutant. The polypeptide composition of crude cell extract (A) and purified carboxysomes (B), respectively, from wild type (lanes 1, 3, and 5) and *csoS1B* truncated mutant (lanes 2, 4 and 6). Lanes 1 and 2, stained polypeptides separated by SDS-PAGE; lanes 3 and 4, immuno-blot probed with anti-CsoS1 antiserum; lane 5 and 6, immunoblot probed with anti-CsoSCA antiserum.

charged concave side might face inward [80]. However, sequence comparisons of 20 BMC protein sequences revealed that only 55% of accessible residues on the concave side are conserved, while all the accessible ones on the convex side are conserved (Figure 24) [80]. This suggested that the concave side might face outward, which completely contradicts the previous argument. To test which case is true, selective labeling of the carboxysome surface was attempted. CsoS1A structure revealed that none of the primary amine containing group resides on the convex side: the N-terminus and Lys94 are accessible from concave side, whereas the Lys29 lies on the hexamer-hexamer interface and therefore is inaccessible (Figure 24, 33A). Because of this finding, amine reactive chemicals were used for surface labeling carboxysome (Figure 33B). Only in the case of the concave side facing outward, the CsoS1A/C protein can be labeled (Figure 33C), and whether or not RubisCO can be labeled as well can be used as an indicator of labeling reagent accessibility to interior space of the carboxysome. The fluorophore Alexa Fluor 546, a water soluble succinimidyl ester, was used to label freshly prepared intact carboxysome. Under UV light visualization, every single band of the carboxysomal proteins was labeled, and the intensity increased with reaction time (Figure 34). Labeled carboxysome was also separated on 10-60% (w/v) mini sucrose gradient, and SDS-PAGE analysis of each fraction from sucrose gradient also suggested that Alexa Fluor 546 stained free RubisCO as well as RubisCO inside of intact carboxysomes (Figure 35). These observations suggested Alexa Fluor 546 labeled both proteins on the surface of carboxysome and those inside of carboxysome. Therefore it is not suitable for surface labeling purposes. *N*-hydroxysulfosuccinimidebiotin (Sulfo-NHS-Biotin) was also used as labeling reagent. Sharp et al. have used Sulfo-NHS-Biotin to biotinylate virus surface

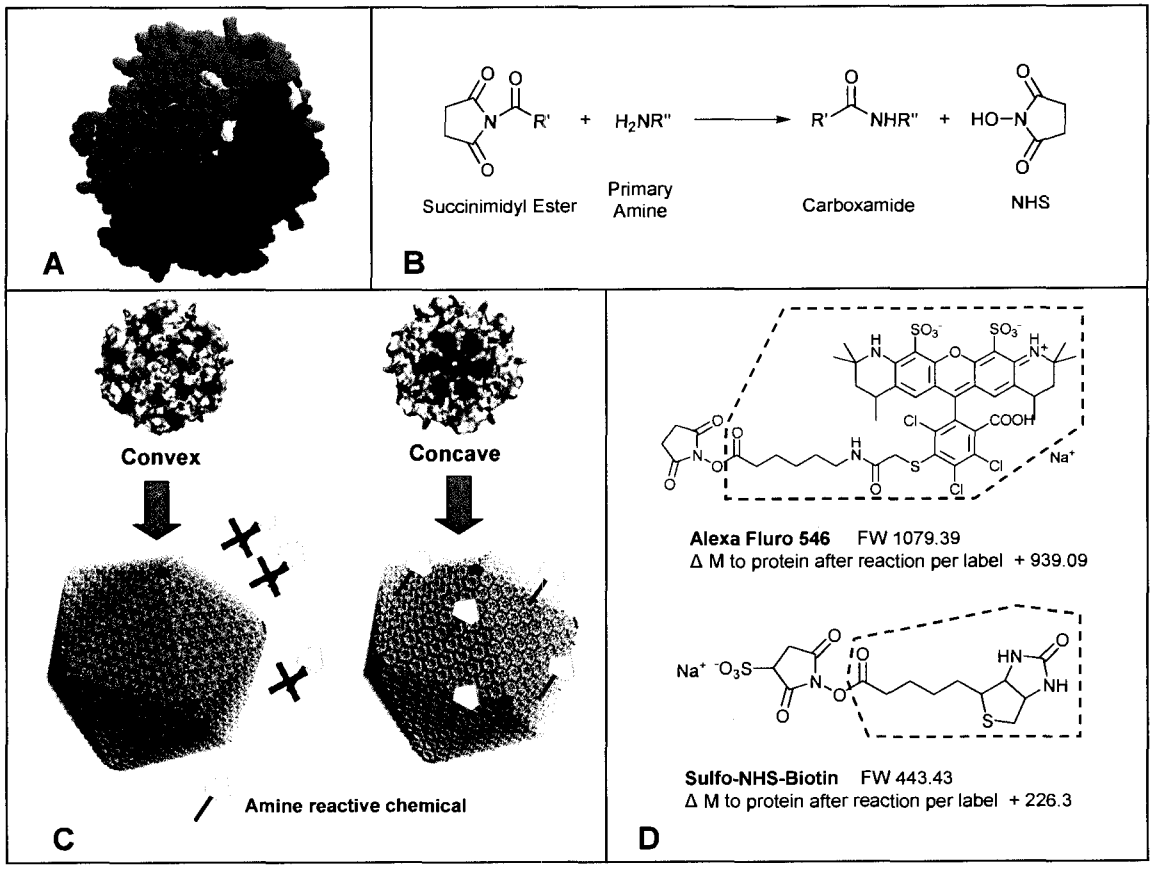


Figure 33. Labeling carboxysome surface with amine reactive chemicals.

A. Locations of primary amine containing groups and terminal groups on CsoS1A hexamer. Pink: N-terminal; white: C-temrinus; element color: Lysine residue. B. Equation for labeling reaction. C. Stratergy of selective labeling. D. Amine reactive chemicals used in this study. upper: Alexa Fluro 546 (Invitrogen); lower: Sulfo-NHS-Biotin (Thermo). Dashed boxes indicate the R' groups.

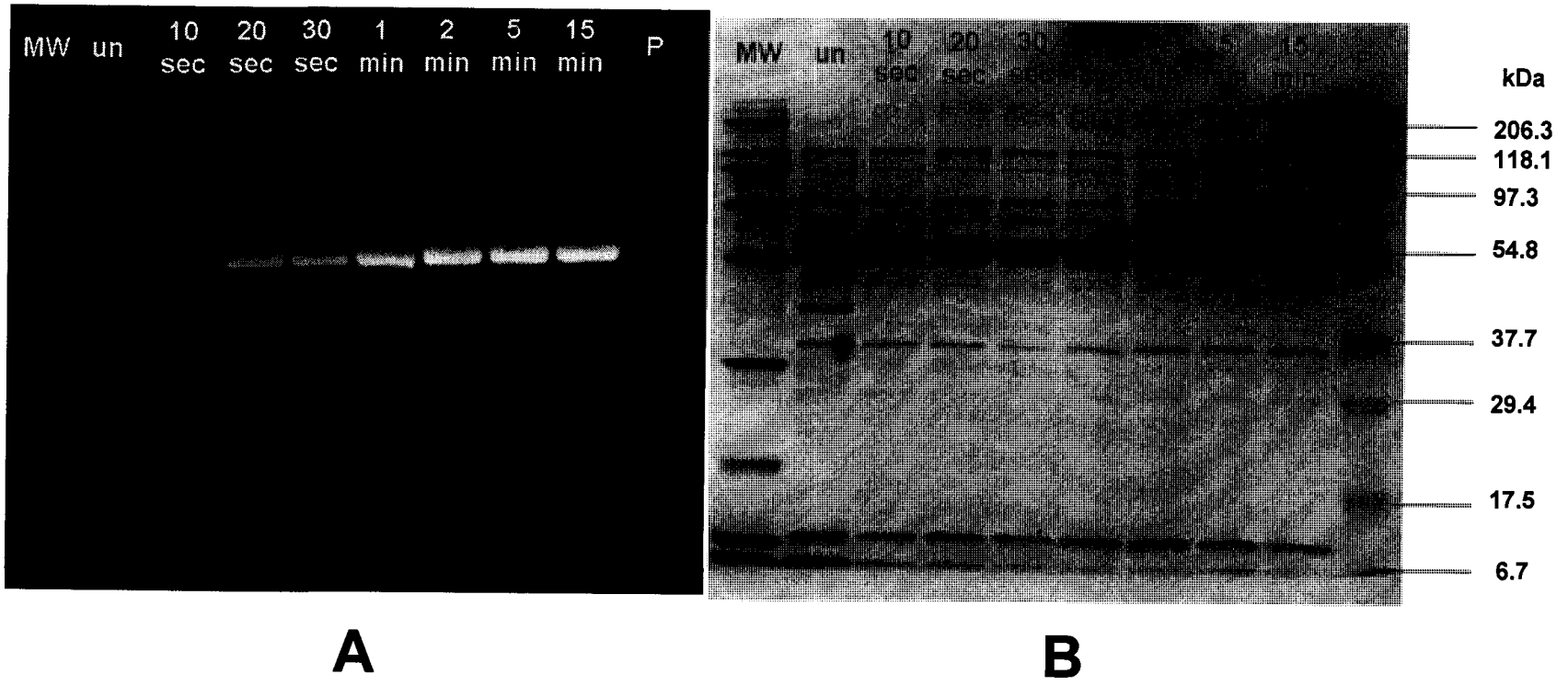


Figure 34. Labeling carboxysome with Alexa Fluor 546.

Reactions were terminated at varies time points. A. visualizing fluorescent bands under UV light; B. same SDS-PAGE stained with GelCode Blue. un: unlabeled sample.

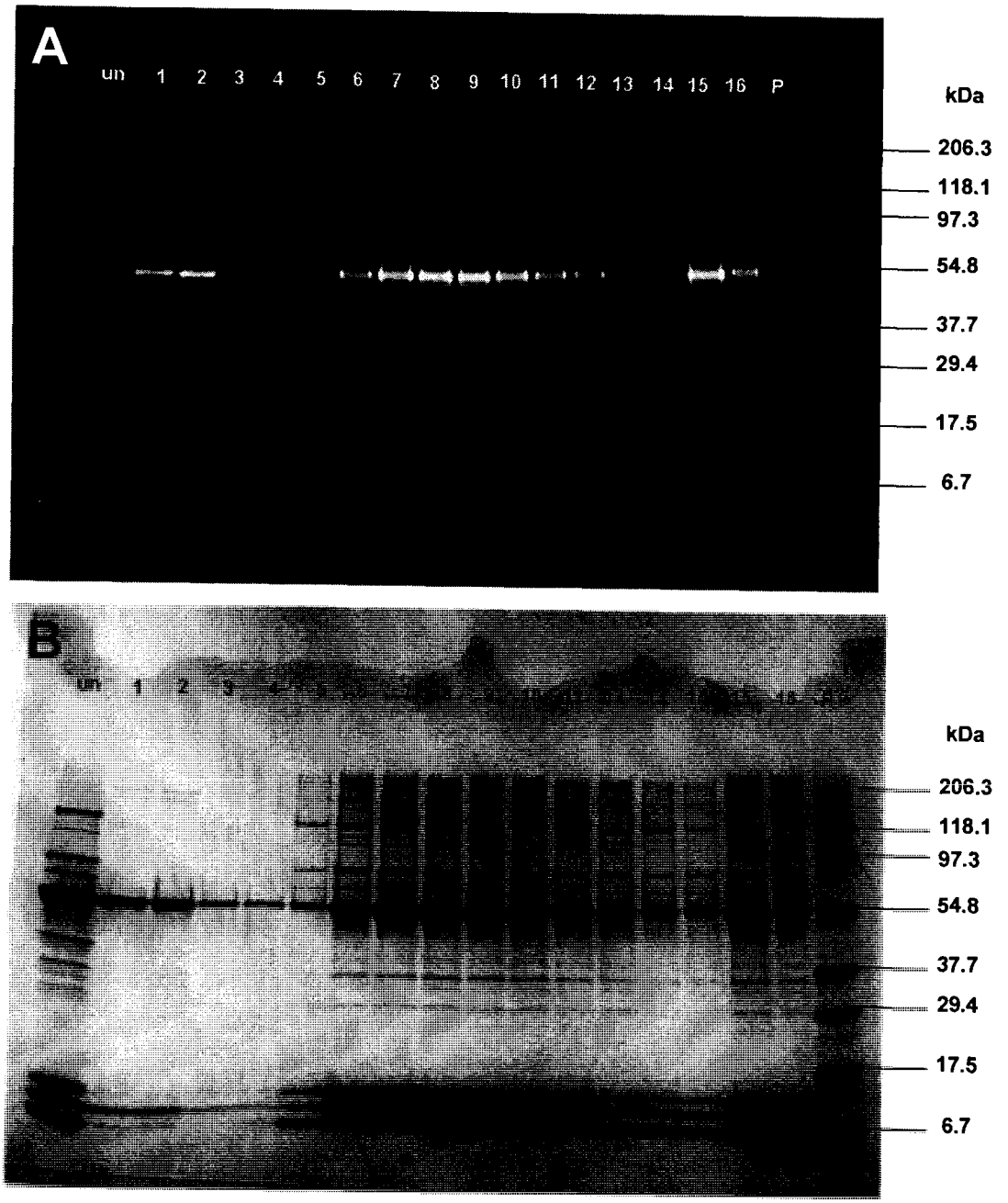


Figure 35. Separating Alexa Fluor 546 labeled carboxysome from free RubisCO on 10-60% (w/v) sucrose gradient.

A. visualizing fluorescent bands under UV light; B. same SDS-PAGE stained with GelCode Blue. un: unlabeled sample. Increasing number indicates from top to bottom of the gradient.

successfully without affecting structural integrity and infectious ability [72]. Expected results for biotinylating carboxysome are listed in Table 3. If the concave side of CsoS1 hexamer faces outward, the CsoS1 protein of intact and broken carboxysomes both should be labeled. If the concave side faces inward, only CsoS1 protein of broken but not intact carboxysome should be labeled. RubisCO molecules were treated as a reaction control, which should not be labeled in the intact carboxysome sample.

Table 3. Expected labeling results based on the sidedness of CsoS1 hexamer

Concave side outward / inward	CsoS1 protein	RubisCO large subunit
Intact carboxysome	+ / -	- / -
Broken carboxysome	+ / +	+ / +
Free RubisCO	N/A	+ / +

+: accessible to amine active reagent, should be labeled;
-: not accessible to amine active reagent, should not be labeled.

The condition for biotinylation had been optimized first via a series of experiments. First, carboxysome shell enriched fraction and free RubisCO sample, in which both CsoS1 protein and RubisCO should be accessible to labeling reagent, were used for biotinylation with varying sulfo-NHS-biotin concentrations (from 0.1 mM to 10 mM) in amine free BEMB buffer (Figure 36). However, only CsoS2A and CsoS2B proteins are consistently labeled under all conditions tested, and using 0.5 mM to 1.0 mM sulfo-NHS-biotin gave the strongest biotinylation signal. On the other hand, RubisCO large subunit CbbL did not seem to be biotinylated. Then, labeling CbbL was attempted with using varying sulfo-NHS-biotin concentration (0.1 mM to 100 mM) in both BEMB

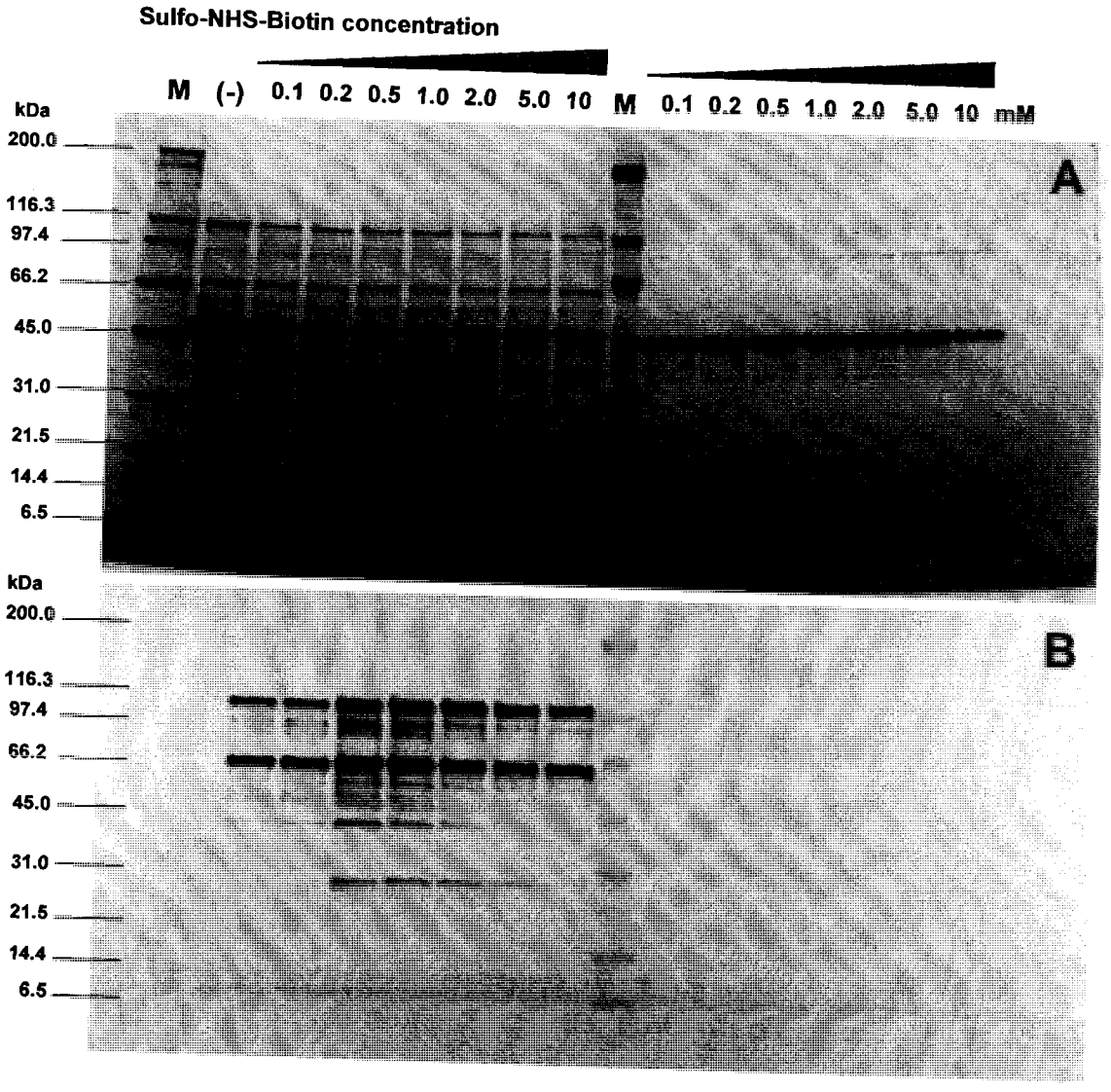


Figure 36. Labeling carboxysome proteins with varies Sulfo-NHS-Biotin concentration in BEMB buffer.

A. proteins separated on precast 10-20% gradient SDS-PAGE; B. immunoblot developed with HRP-conjugated Neutravidin. Left half, shell in-rich fraction; right half, free RubisCO. (-), negative control, intact carboxysome without labeling reaction; M, protein molecular weight marker.

and TEMB buffer (Figure 37). Surprisingly, biotinylation with 1 mM labeling reagent in TEMB worked, where 5 mM primary amine from buffer should act as a competitive substrate, but not in amine free BEMB buffer. Therefore 1 mM sulfo-NHS-biotin and TEMB buffer was used as experimental condition thereafter. Finally, biotinylated free RubisCO under such condition has been loaded with a series of 1/10 dilutions to test signal strength and sensitivity (Figure 38). Signal could be detected with as low as 10 ng labeled RubisCO, and signal from CbbL band looked ambiguous when 10 μg was loaded. After optimization, bioinylation was performed with intact carboxysomes, shell enriched fraction, and free RubisCO samples (Figure 39). CsoS2A and CsoS2B were both strongly biotinylated in all samples, which supported the idea that CsoS2 proteins are at least partially exposed on the carboxysome surface. CsoS1 proteins were labeled at approximately same level in both intact and broken carboxysomes, which suggested the concave side is always accessible or facing outward. However, RubisCO molecules of intact carboxysomes were also biotinylated as well as the ones from broken carboxysomes. This observation again complicated the interpretion of results.

3.3 Disrupting the Positively Charged Pore in CsoS1 Hexamer via Mutagenesis Partially Affected Cell Growth

One of the common features of the three BMC protein hexamers that have been reported so far is the small pores at the central symmetry axis [42,80]. These pores might differ in size, from approximately 4 Å to nearly 7 Å, but they all tend to carry a positive electrostatic potential due to conserved lysine or arginine residues [86]. Therefore, a hypothesis was proposed that these pores served as a transport channel, especially for the

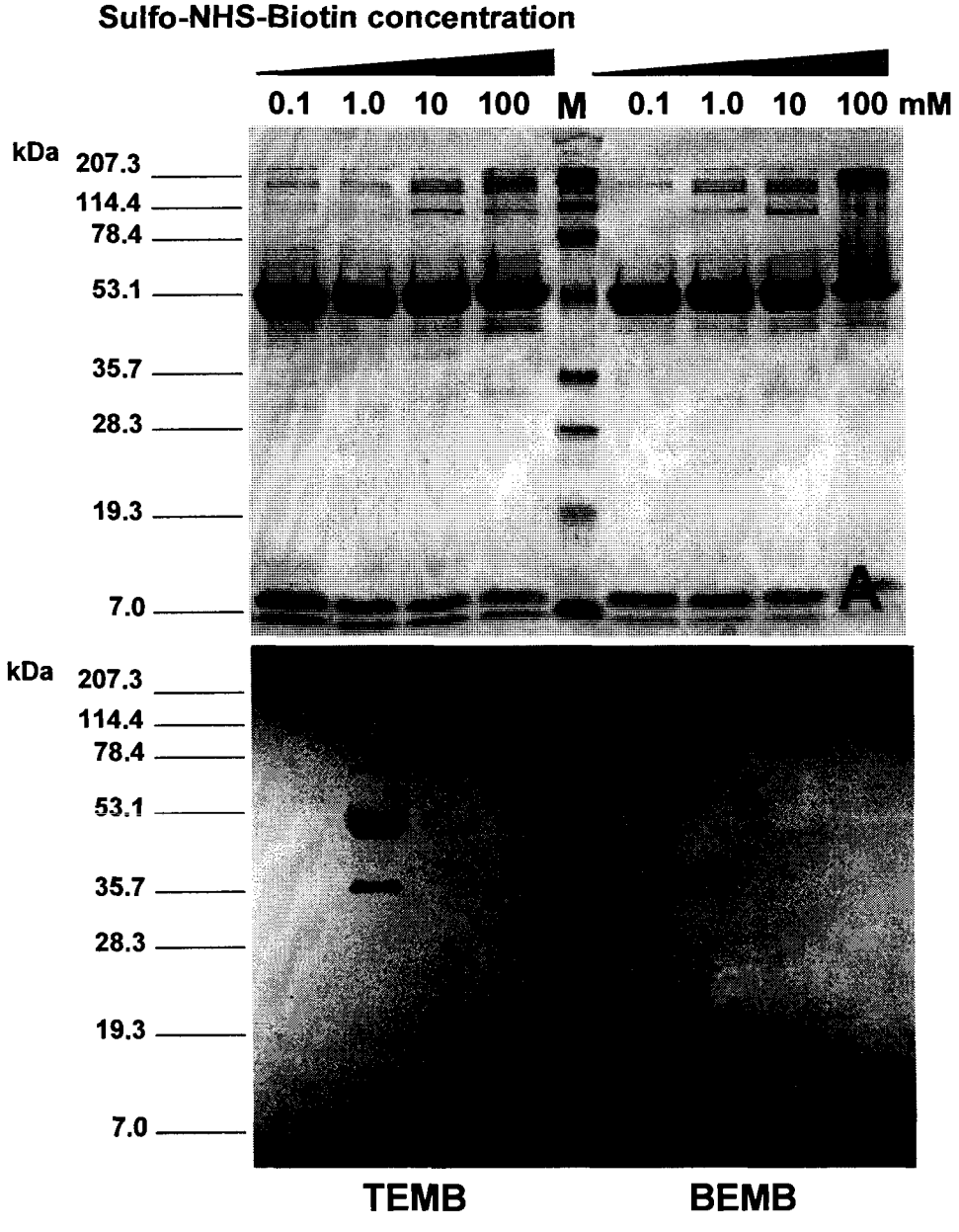


Figure 37. Biotinylation of free RubisCO with Sulfo-NHS-Biotin in TEMB and BEMB buffer.

A. proteins separated on 12% SDS-PAGE; B. immunoblot developed with HRP-conjugated Neutravidin. Left half, sample prepared and reaction performed in TEMB buffer; right half, in BEMB buffer. M, protein molecular weight marker.

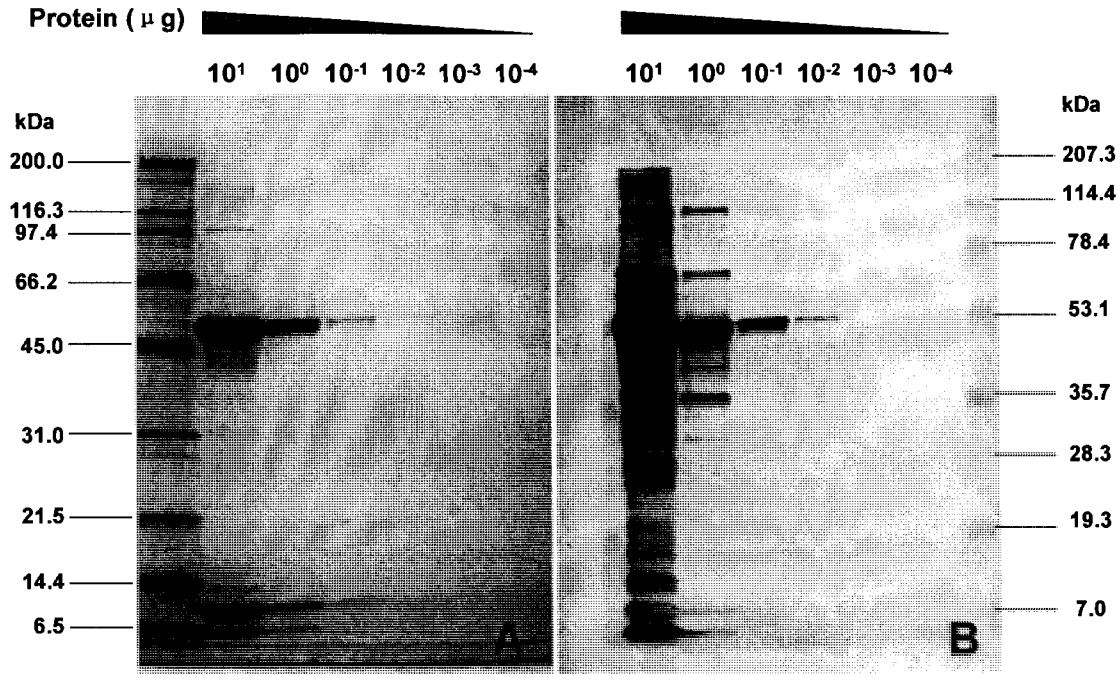


Figure 38. Free RubisCO labeled with 1 mM Sulfo-NHS-Biotin in TEMB buffer. A. proteins separated on precast 10-20% SDS-PAGE; B. immunoblot developed with HRP-conjugated Neutravidin. 1/10 dilution series were made and loaded.

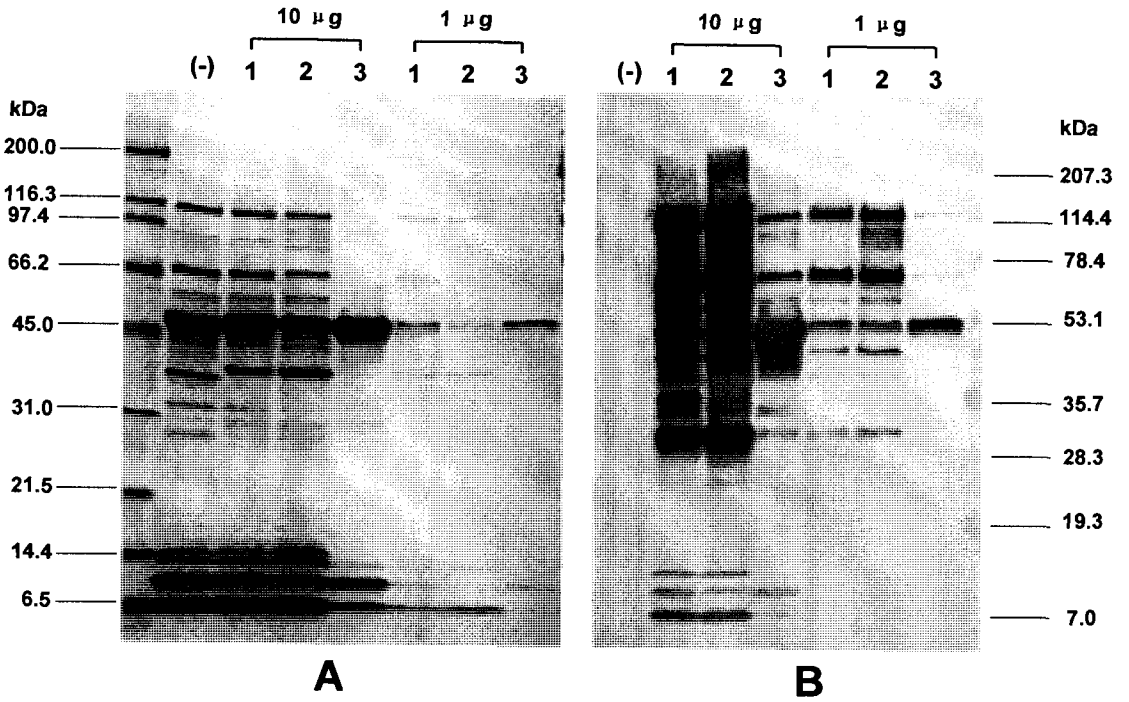


Figure 39. Labeling carboxysome protein with 1 mM Sulfo-NHS-Biotin in TEMB buffer.

A. proteins separated on precast 10-20% SDS-PAGE; B. immunoblot developed with HRP-conjugated Neutravidin. (-), negative control, intact carboxysome without labeling reaction. Lane 1, intact carboxysome; lane 2, shell in-rich carboxysome; lane 3, free RubisCO. Each reaction was loaded in both 10 µg and 1µg.

negatively charged metabolites that have to cross the carboxysome shell, like bicarbonate, ribulose bis-phosphate (RuBP), and 3-phosphoglycerate (3-PGA) [42]. The observation of sulfate ions in the pore of the CsoS1A hexamer contacting Gly43 residue via H-bonds supported the above hypothesis [80]. This Gly43 residue is in the middle of a Gly-Gly-Gly motif, which is conserved across the three BMC proteins in *H. neapolitanus* and is supposed to form the narrowest point of the pore [80]. In order to test the proposed hypothesis, a mutagenesis experiment was designed to target this residue. In the mutant, Gly43 was successfully mutated to Ala43 via PCR and DNA recombination, and a Kan^r cassette placed under its own promoter was inserted downstream from the mutated *csoS1A* gene to ensure transcription and expression of the *csoS1B* gene. Considering the typical length of carbon-carbon bond and carbon-hydrogen bond to be 1.54 Å and 1.09 Å [21], respectively, the extra methyl-group on Ala residue should be sufficient to block or close the pore without disrupting the overall hexamer structure. This mutant has been named *csoS1A::G43A*. It grew similar to wild type *H. neapolitanus* in air supplemented with 5% CO_2; however, it grew poorly and could not reach the same cell density as wild type in ambient CO_2 (Figure 25). TEM pictures revealed that carboxysomes from the *csoS1A::G43A* mutant did not show morphological difference from wild type carboxysome (Figure 40).

3.4 Characterization of the csoS4 Gene Deletion Mutants

Based on the pentamer crystal structure, the function of CsoS4 protein in *H. neapolitanus* has been proposed to form the vertices of the icosahedral body of the

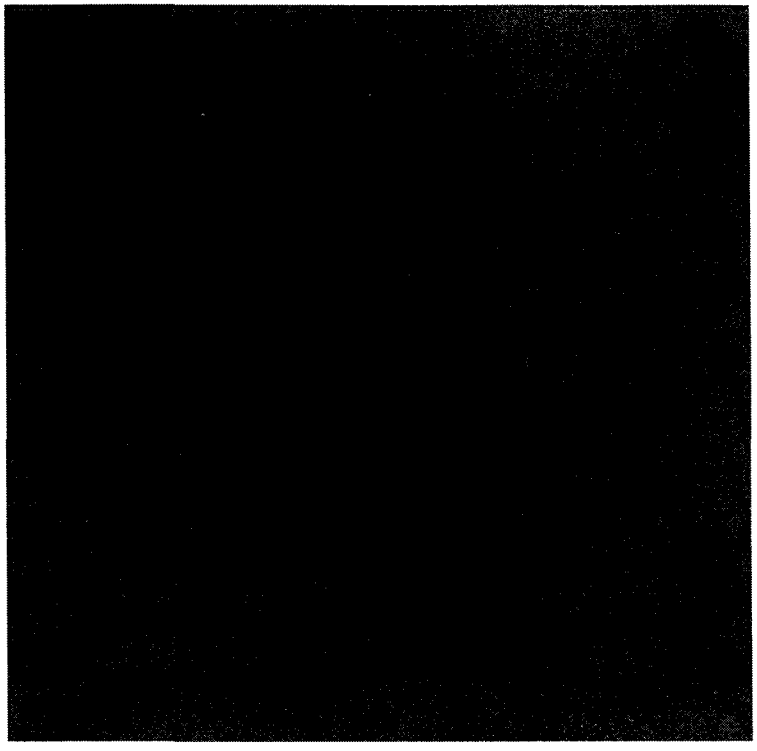


Figure 40. Observing carboxysomes from *csoS1A::G43A* mutant under TEM. Mutant carboxysome enrichment sample was negatively stained with 1% (w/v) ammonium molybdate and observed under TEM. Bar indicates 100 nm.

carboxysome [77]. The presence of both CsoS4 proteins in carboxysome shells was confirmed with two dimensional electrophoresis and immunoblotting. However, the function of CsoS4 protein *in vivo* has to be tested experimentally. Thus, deletion mutants, *csoS4A::Km*, *csoS4B::Km*, and *csoS4AB::Km*, were generated via PCR and DNA recombination, in which one or both of csoS4 genes were knocked out in *H. neapolitanus*. In these mutants, coding sequence of the csoS4 genes was replaced by a Kan^r cassette, which has its own promoter. Correct mutants were confirmed by PCR and genomic DNA sequencing. All three mutants displayed a *h*igh *C*O_2-*r*equiring (*hcr*) phenotype, which means they fail to grow in air. When they grew in air supplemented with 5% CO_2 their

grow rates were similar to the wild type culture; however, they could not reach same cell density as the wild type (Figure 41). This *hcr* growth behavior implied that delection of one or both *csoS4* genes resulted in either inability to assemble carboxysomes or assembly of dysfunctional carboxysomes. However, observing thin sections of three mutants under TEM revealed that all of them do produce carboxysomes (Figure 42). Many cells produced elongated carboxysomes (Figure 42 A-C, red arrows). In case of the *csoS4AB::Km* double mutant, a long mutant carboxysome extended into both mother and daughter cells and prevented cell division (Figure 42 D, red arrows). Since the function of CsoS4 protein has been proposed to introduce curvature to the carboxysome shell by inserting itself as pentamer into CsoS1 hexamer sheets and therefore to close the icosahedral body, one would expect to see an abnormally shaped mutant carboxysome. Surprisingly, however, carboxysomes with apparent icosahedral shape were also observed in all three deletion mutants, which included the double deletion mutant (Figure 42). Immunoblot of crude cell extract revealed that the CsoS4 protein was not present in either the double deletion mutant (*csoS4AB::Km*) or the singe deletion mutant (*csoS4A::Km* and *csoS4B::Km*), while CsoS1 proteins were expressed at levels comparable to the wild type (Figure 43). Analysis of enriched mutant carboxysome further suggested the single and double deletion mutants had similar phenotypes: absence of both CsoS4 proteins and coexist of icosahedral and elongated carboxysome (Figure 44). Therefore, the double deletion mutant, *csoS4AB::Km*, was used for further characterization. To obtain a quantitative idea about the abundance of enlongated carboxysomes over the icosahedral ones, both species were counted in *csoS4AB::Km* mutant in both thin sections and purified carboxysome samples, in random fields of vision under the transmission electron

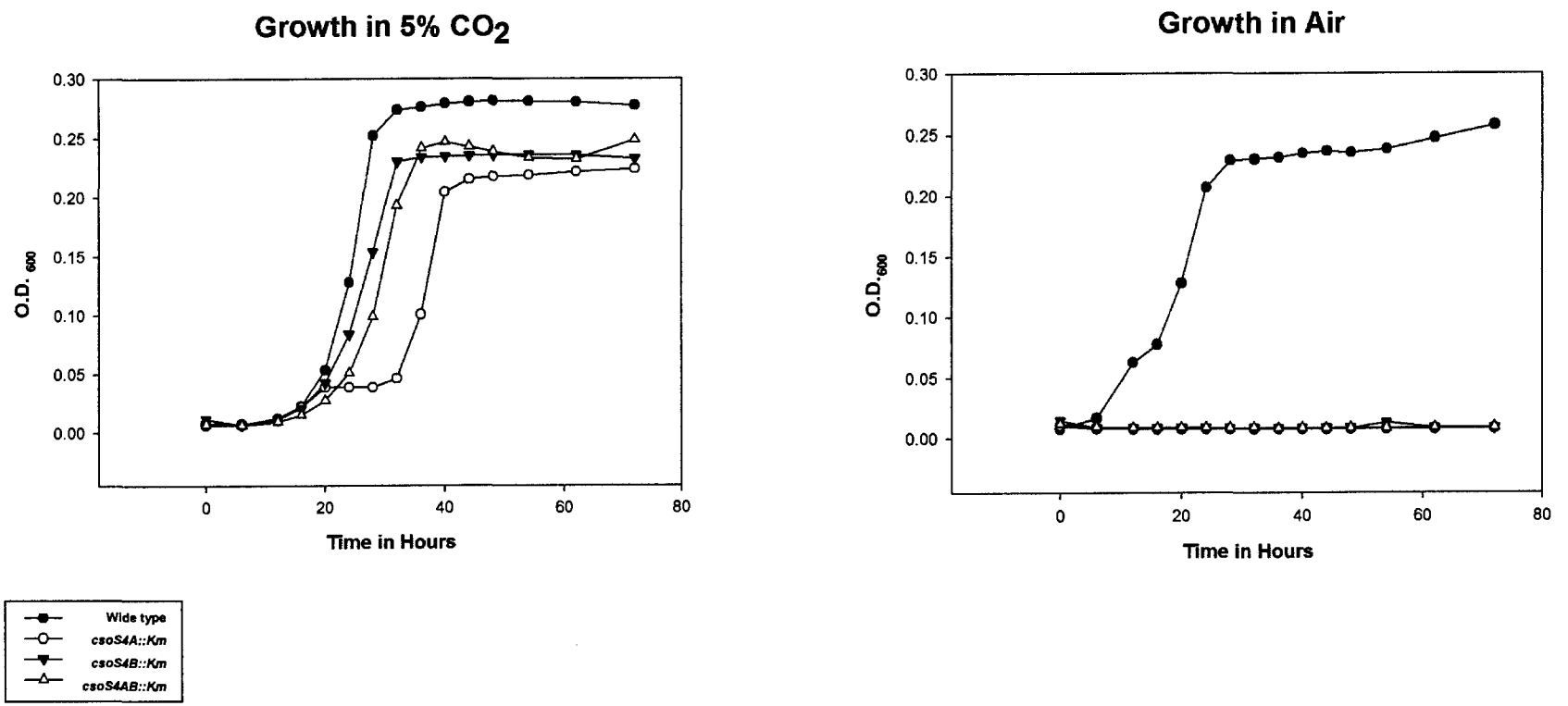


Figure 41. Growth curve of *H. neapolitanus csoS4* gene(s) deletion mutants.

Wild type and mutants *H. neapolitanus* grow in 5% CO_2 (left) and in air (right). *csoS4A::Km*, *csoS4A* deletion mutant; *csoS4B::Km*, *csoS4B* deletion mutant; and *csoS4AB::Km*, *csoS4AB* deletion mutant. All three mutants displayed a strict *hcr* phenotype.

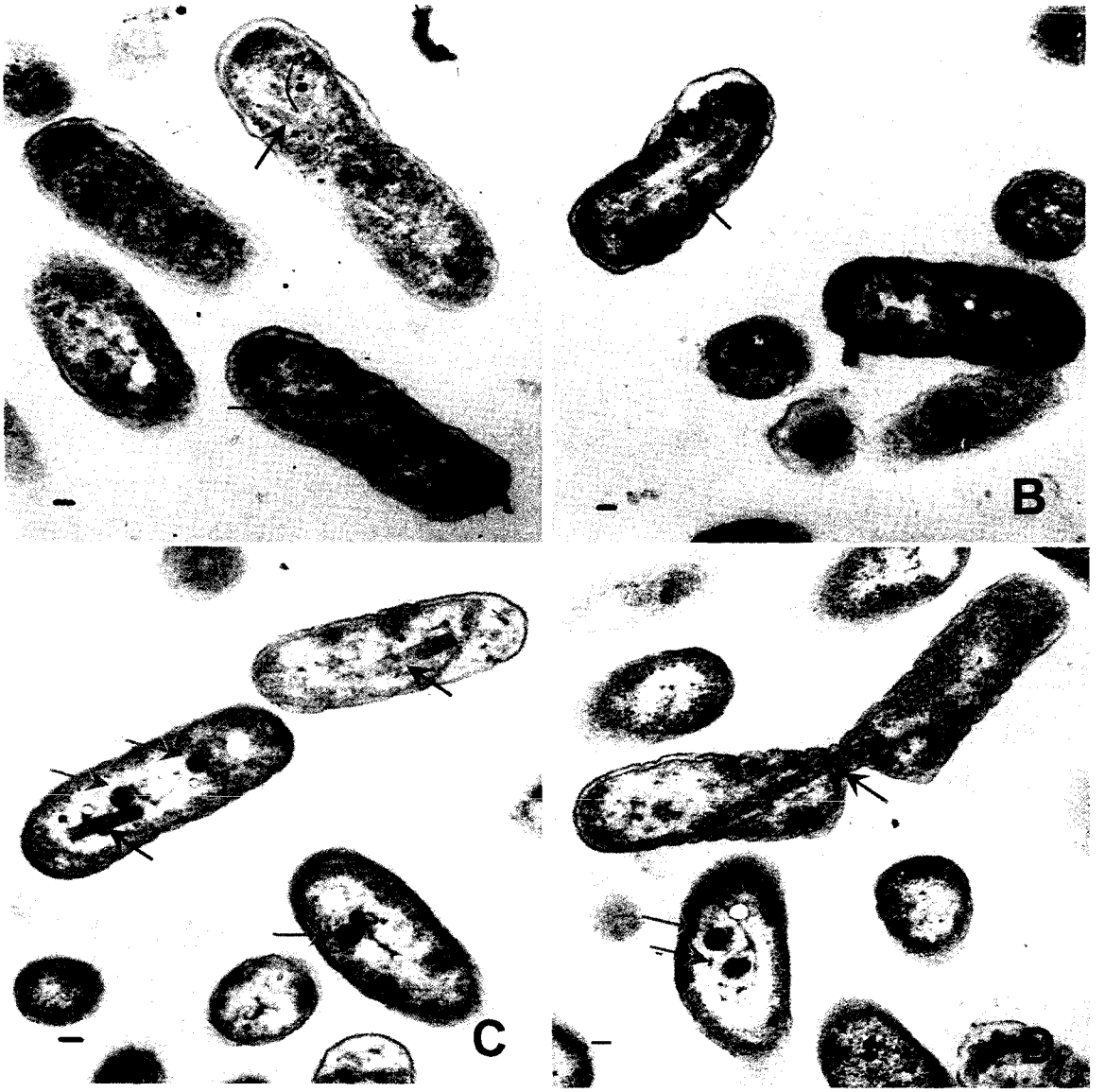


Figure 42. Transmission electron micrographs of *csoS4* gene(s) deletion mutants. Thin sections of mutant cells containing elongated carboxysomes (red arrow) as well as carboxysomes of apparently icosahedral shape (green arrows). A. *csoS4A::Km* mutant; B. *csoS4B::Km* mutant; C and D, *csoS4AB::Km* mutant. The scale bar indicates 100 nm.

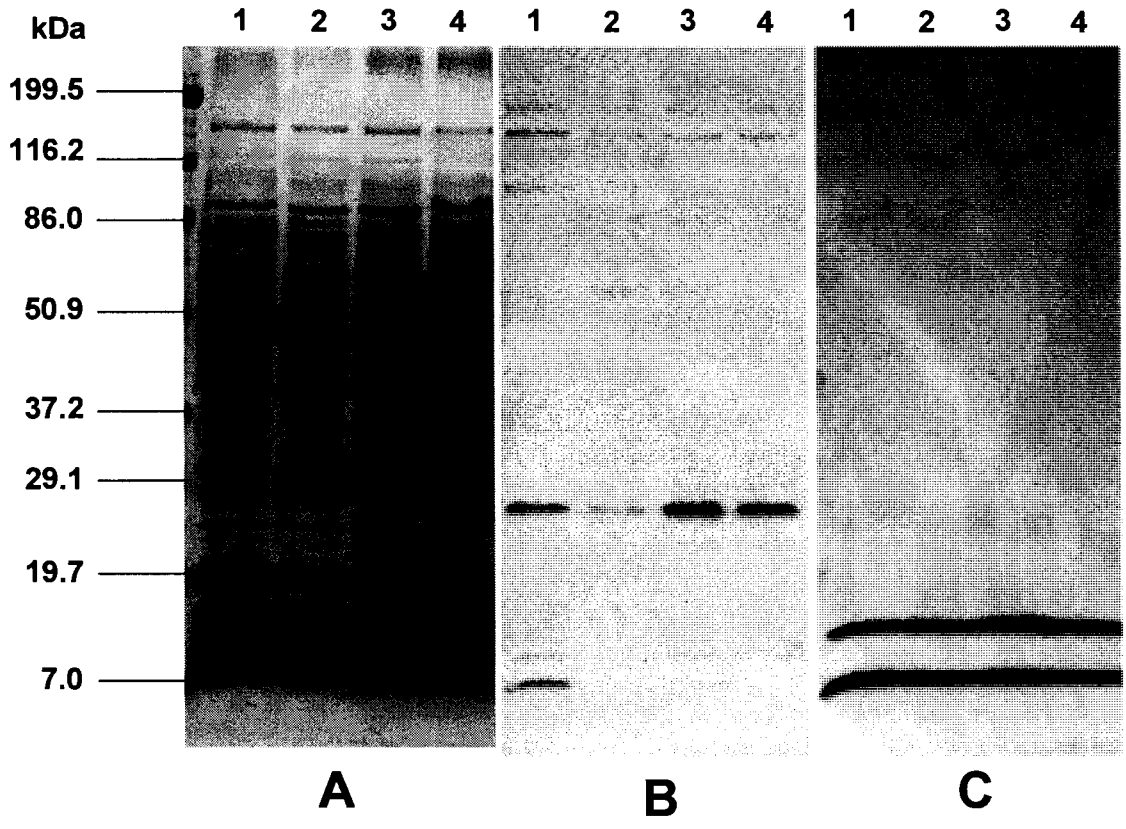


Figure 43. Immunoblots of crude extract with *H. neapolitanus* wild type and *csoS4* gene(s) deletion mutants.

Crude extracts separated on SDS-PAGE (A) and immunoblots of same samples with antibody against CsoS4 (B), and against CsoS1 (C). Lane 1, wild type cells; lane 2, *csoS4A::Km* mutant; lane 3, *csoS4B::Km* mutant; and lane 4, *csoS4AB::K*m mutant.

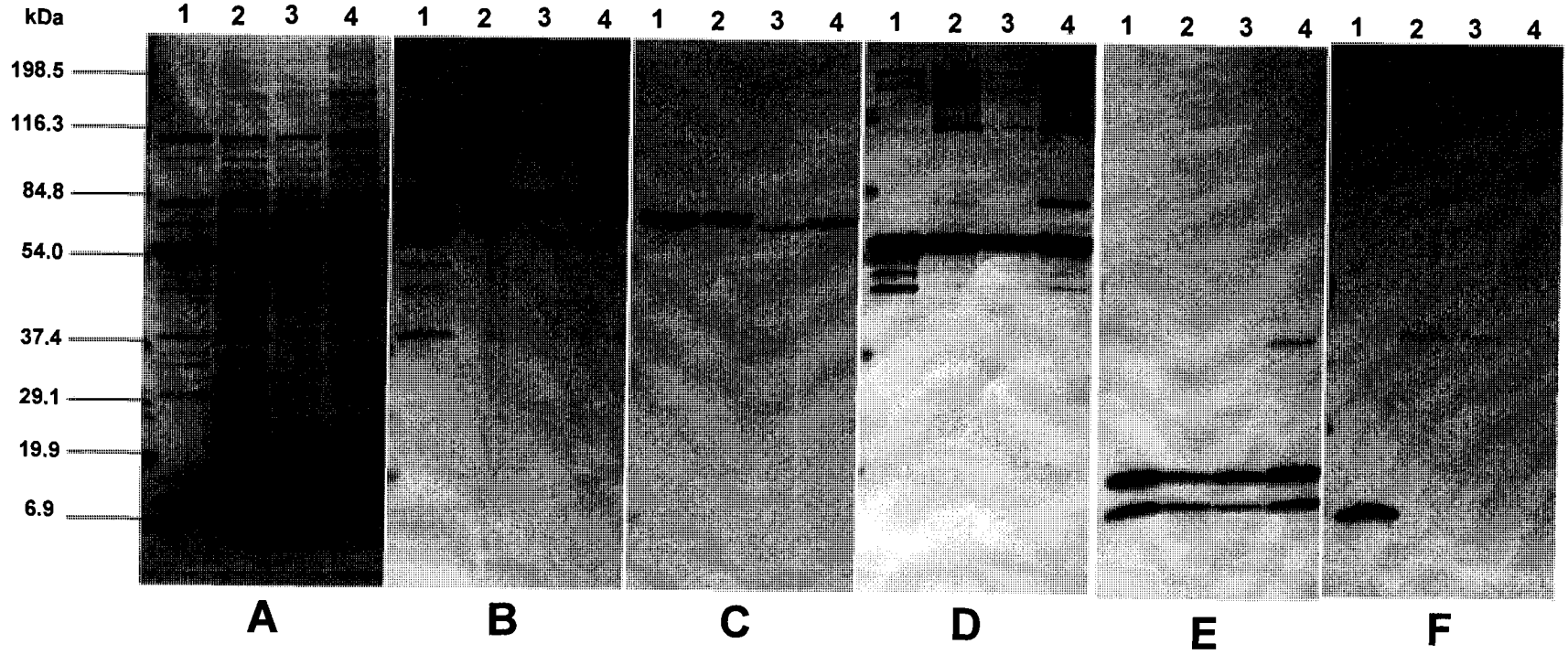


Figure 44. Comparison of *H. neapolitanus* wild type carboxysome and enrichment of *csoS4* gene(s) deletion mutant carboxysomes.

A. Mutant carboxysome enrichments and wild type carboxysome separated on SDS-PAGE. B-F, immunoblots of same samples with antibody against CsoS2 (B), CsoSCA (C), CbbL (D), CsoS1 (E), and CsoS4 (F). Lane 1, wild type cells; lane 2, *csoS4A::Km* mutant; lane 3, *csoS4B::Km* mutant; and lane 4, *csoS4AB::*Km mutant.

microscope, the resulting numbers were compared to the numbers obtained from the wild type (Table 4). In thin sections, 376 *csoS4AB::Km* mutant cells were counted and they contained 0.375 carboxysomes per cell on average, regardless of carboxysome shape. Meanwhile, 340 wild type cells were counted and they contained on average 0.576 carboxysomes per cell; no abnormally shaped carboxysomes were observed. The average number of carboxysomes per cell in the mutant was 30% less than that in the wild type.

Table 4. Counting of wild type and mutant carboxysome with help of Transmission electron microscopy.

	Thin Section				Purified carboxysome	
	cell	carboxysome[a]	%[b]	Average[c]	carboxysome[a]	%[b]
wild type	340	196 (0)	0%	0.576	325 (0)	0%
csoS4AB::Km	376	141 (18)	12.8%	0.375	312 (4)	1.3%

Note:

Both wild type and *csoS4AB::Km* mutant *H. neapolitanus* cultures were grown with 5% CO_2 supplyment.

a. The number in "carboxysome" column is the number of total counted carboxysomes regardless of shape. The number in bracket is the number for elongated carboxysomes.

b. "%" indicates percentage of elongated carboxysomes in total counted carboxysomes.

c. Average value indicates on average how many carboxysomes present in one cell under tested growth condition.

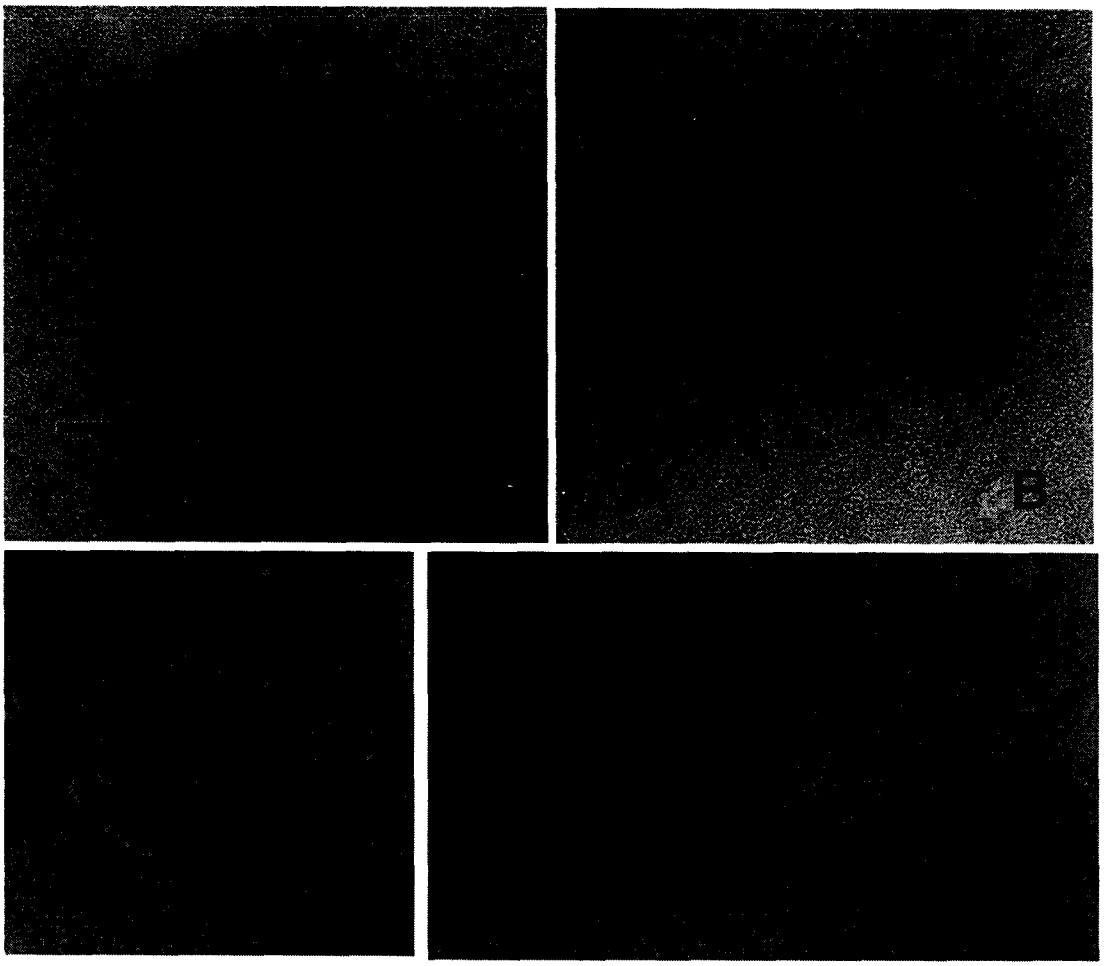


Figure 45. Observing carboxysomes from *csoS4AB::Km* mutant under TEM. Mutant carboxysome isolated from *H. neapolitanus csoS4AB::Km* culture was negatively stained with 1% (w/v) ammonium molybdate and observed under TEM. Bar indicates 100 nm. A and B were taken by Balaraj Menon.

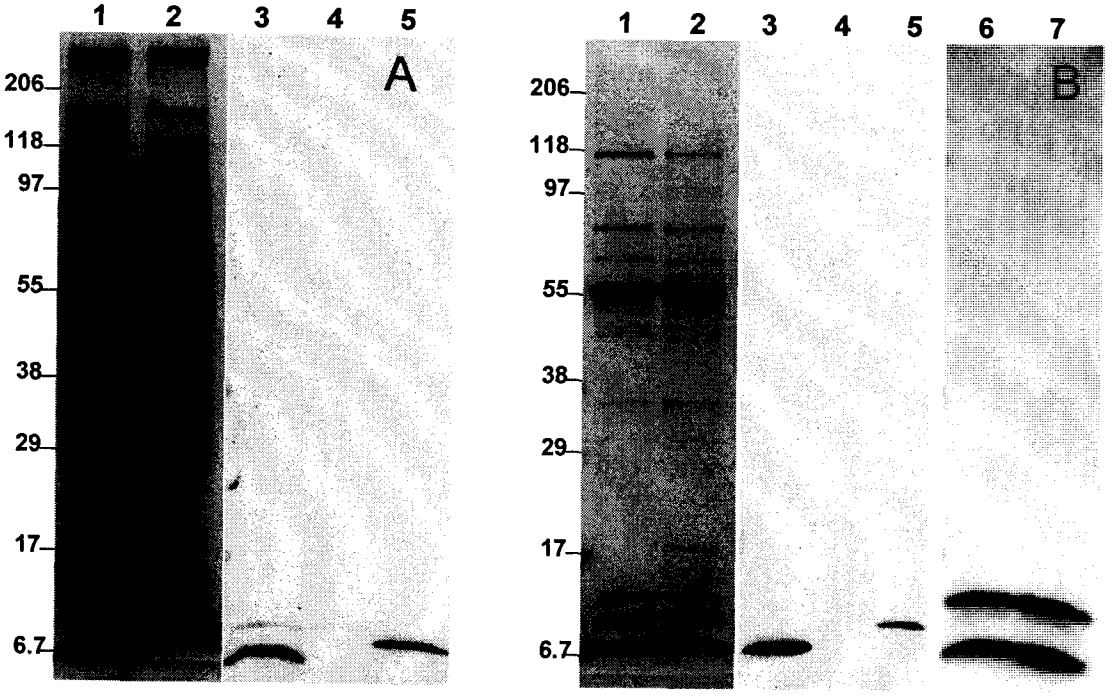


Figure 46. Composition of wild type and *csoS4AB::Km* mutant carboxysomes. A and B show the polypeptide composition of crude cell extract (*A*; 50 μg of protein) and purified carboxysomes (*B*; 10 μg of protein), respectively, from wild type (*lanes 1, 3, and 6*) and *csoS4AB::Km* mutant (*lanes 2, 4, and 7*). *Lanes 1* and *2*, stained polypeptides separated by SDS-PAGE; *lanes 3* and *4*, immuno-blots probed with anti-CsoS4A/B antiserum; *lane 5*, 10 ng of recombinant CsoS4A and CsoS4B protein mixed with 1:1 molar ratio as a positive control; *lanes 6* and *7*, immuno-blots probed with anti-CsoS1B antiserum. (*note: rCsoS4A/B has his-tag which makes them slightly bigger. Based on preliminary model, 10 μg of carboxysome should contain 23 ng of CsoS4A/B.)

Moreover, only 12.8% observed mutant carboxysomes were abnormally shaped. In other words, the majority of the *csoS4AB::Km* mutant carboxysomes were icosahadral or "normal" shaped. Mutant carboxysomes were isolated from a chemostat culture via a series of differential centrifugation and sucrose gradient separation. Transmission electron microscopy was used to observe negatively stained mutant carboxysomes, and it revealed that both elongated and "normal" carboxysomes co-purified (Figure 45). There were only 4 elongated carboxysomes in a total of 312 (Table 4). So the percentage of abnormally shaped carboxysomes in purified sample is a lot less than in cells. A closer investigation of mutant carboxysome composition was done via electrophoresis and immunoblot. Comparison of wild type and mutant carboxysome by SDS-PAGE suggested they were indistinguishable in composition (Figure 46A). However, immunoblotting with CsoS4 antiserum clearly showed that the CsoS4 protein was absent in mutant carboxysomes in both wild type and mutant. The major shell components, CsoS1, were present at about same level (Figure 46 B). These results suggested that the CsoS4 protein does not occupy the vertices of the icosahedral shaped mutant carboxysomes.

CHAPTER IV

DISCUSSION

It was generally assumed that the clustered carboxysomal genes found in many chemolithotrophs and all known α-cyanobacteria constitute an operon [8,12,17,71]. However, the regulatory mechanism behind the observation that the relative amounts of each gene product in the isolated carboxysomes are not equal was unclear. In the first part of this study, an extensive analysis of carboxysome gene expression was performed in *H. neapolitanus*, the findings of which included: (1) measurable transcript levels that were detected for all genes in the cluster (Figure 17); (2) a single transcription start site that was identified upstream from the first gene of this cluster (Figure 18a); and (3) a single transcript 3'-end was identified downstream from the last gene of this cluster (Figure 20b). These results led to the first conclusion of this study that the *cso* gene cluster in *H. neapolitanus* is an operon. The second conclusion was that the full-length *cso* transcript undergoes mRNA processing and the resulting smaller mRNA species have different stabilities. This conclusion was drawn based on the fact that the following results ruled out other possible regulatory schemes. The results are: (1) multiple transcript 3'-ends were identified downstream from the stem-loop structures which were predicted between the *cbbS* and *csoS2* genes (Figure 20a); (2) no transcript 3'-end was identified downstream from the similar stem-loop structures predicted to exist between the *csoS4B* and *csoS1C* genes; (3) no transcription start site was identified upstream from the *csoS1C* gene; (4) selected internal regions within the *cso* operon did not shown any promoter activity in *E. coli* (Figure 21b), while the *cso* promoter was shown to function in the same promoter reporter vector (Figure 21c-d); and (5) expression of carboxysomal genes was

not detected in *cso* promoter deletion mutant of *H. neapolitanus* (Figure 23).

The second part of this study focused on understanding *in vivo* functions of the *csoS1* genes and *csoS4* genes in *H. neapolitanus*, which are two sets of conserved genes for bacterial microcompartment. For the first time, direct experimental evidence was provided to indicate that protein products of all carboxysomal genes are present in functional carboxysome particles. This conclusion completed the inventory of carboxysome components based on these immunological results: (1) both CsoS1C and TC tag fused CsoS1A were detected in purified carboxysomes of the *csoS1A*-TC tag mutant (Figure 26); and (2) both CsoS4A and CsoS4B were present in carboxysome shell enriched samples as minor components (Figure 29). Further functional studies of the *csoS4* genes via mutagenesis yielded the following major findings: (1) the expression of CsoS4A and CsoS4B is tightly correlated; (2) *H. neapolitanus* mutant cells lacking CsoS4 polypeptides produce both elongated and normally shaped mutant carboxysomes, and the latter outnumbers the former; and (3) shells of carboxysomes from *csoS4A* and *csoS4B* double knockout mutant cannot function as a diffusion barrier for both inorganic carbon species, bicarbonate and carbon dioxide. Last, some clues have been provided by this study to obtain a better understanding of protein-based carboxysome shell in terms of its architecture and function. Those clues are: (1) the CsoS1B C-terminal tail was not required for CsoSCA protein functioning as a shell component; (2) the central pores of CsoS1A hexamers are important for carboxysome function; (3) both shell proteins, CsoS2A and CsoS2B, have access to the cytosol; and (4) the concave sides of CsoS1 hexamers most likely face outwards.

1. The *cso* Gene Cluster in *H. neapolitanus* Is an Operon

A bacterial operon usually presents the following three features [37,64,67,90]: a relatively short intergenic distance, a conserved gene order and orientation found across species, and a functional correlation among clustered genes. The gene cluster, which encodes for components of the α-carboxysome has all these features, therefore it is predicted to be an operon in general [8,12,17,71]. Table 4 and Table 5 list 20 putative carboxysome operons which include genes encoding Form IA RubisCO. And notably, the arrangement of the *csoS2* gene through *csoS4B* gene is highly conserved in both carboxysome containing chemoautotrophs (Table 4) and α-cyanobacteria (Table 5). Of these listed clusters, the average intergenic distances between *csoS2* and *csoS3*, *csoS3* and *csoS4A*, and *csoS4A* and *csoS4B* are 16 bp, 3 bp, and 2 bp, respectively. However, the intergenic distances beyond this region are significantly larger, and in some cases the value is over 50 bp. Although individual cases may vary, statistically two genes with intergenic distance larger than 50 bp have less probability of being in a same operon [67]. To test the operon assumption experimentally, *H. neapolitanus* was used as model organism to establish the transcriptional profile of carboxysome genes. With the aid of primer extension experiments, a single transcription start site (TSS) was identified as an adenine at -54 position according to the start codon of *cbbL* gene (Figure 18a), and no transcription 5'-ends in the region upstream from *csoS1C* gene was detected. These results favored a single common promoter, named *cso* promoter, for the carboxysome gene cluster. However, previously a cytosine at -53 position of *cbbL* was reported as TSS [12]. This could be an incorrectly identified TSS, because the preferred nucleotide for

Table 5. Comparison of intergenic regions in *cso* operons of selected chemoautotrophs.

Intergenic region	*H.n.* AF038430	*T. i.* AF046933	*T. c.* NC_007520	*T. d.* NC_007404	*A. F.* AF129925	*H. m.* AB122070
		12[a]				
cbbL-cbbS	36	67	45	-[d]	91	46
cbbS-csoS2	64	115	58	-[d]	70	61
csoS2-csoS3	20	16	27	25	55	19
csoS3-csoS4A	-4	10	14	9	-4	14
csoS4A-csoS4B	-1	-1	-20	-1	10	-20
csoS4B-csoS1C	51	105	38	-16	31	41
csoS1C-csoS1A	74	45	26	35	39	30
csoS1A-cosS1B	111	12	26	13	39	99
			58[b]	22[e]	-50[e]	54[e]
			-4[c]	-8[f]	-4[f]	-4[f]
				2[g]	36[g]	11[g]

The numbers indicate the number of nucleotides between adjacent genes, with negative numbers signifying overlapping reading frames. HP denotes a gene for a conserved protein of unknown function (Pfam database ID PB040172).
H.n., Halothiobacillus neapolitanus; T.i., Thiomonas intermedia; T.C., Thiomicrospira crunogena; T.D., Thiobacillus denitrificans; A.F., Acidithiobacillus ferrooxidans; H.m., Hydrogenovibrio marinus.
[a] *orfC-cbbL*; [b] *csoS1B*-rubrerythrin; [c] rubrerythrin-HP; [d] gene not present in operon; [e] *csoS1B*-bacteroferritin; [f] bacteroferritin-HP; [g] HP-partition protein A.

Table 6. Comparison of intergenic regions in *cso* operons of selected α-cyanobacteria.

Intergenic region	*P.m.* MIT 9301 NC_009091	*P.m.* MIT 9303 NC_008820	*P.m.* MIT 9312 NC_007577	*P.m.* MIT 9313 NC_005071	*P.m.* MIT 9515 NC_008817	*P. m.* NATL1A NC_008819	*P.m.* NATL2A NC_007335	*P.m.* AS 9601 NC_008816	*P.m.p.* CCMP 1986 NC_005072
csoS1-cbbL	69	69	69	69	66	71	71	68	65
cbbL-cbbS	89	108	94	107	93	100	100	90	89
csbbS-csoS2	91	111	79	111	92	99	99	83	90
csoS2-csoS3	7	7	7	7	64	7	7	7	7
csoS3-csoS4A	2	1	2	1	2	0	0	2	3
csoS4A-csoS4B	17	-1	19	-1	5	-1	-1	18	32
csoS4B-csoS1	-[a]		-[a]	56	-[a]	44	44	-[a]	-[a]
	75[b]		74[b]	143[c]	100[b]	75[c]	75[c]	74[b]	100[b]
				-17[d]					

The numbers indicate the number of nucleotides between adjacent genes, with negative numbers signifying overlapping reading frames. HP denotes a gene for a conserved protein of unknown function (Pfam database ID PB040172) that in *P. m. m.* CCMP 1375 is annotated as pterin 4α-carbinolamine dehydratase-like protein.

P.m., Prochlorococcus marinus; *P.m.p., Prochlorococcus marinus* subsp. *pastoris*; *P.m.m., Prochlorococcus marinus* subsp. *marinus*; *S.* sp., *Synechococcus* sp..

[a] gene not present in operon; [b] *orfB*-HP; [c] *csoS1*-HP; [d] HP-*cbbX*; [e] *csoS1*-Carbinolamine Dehydratase-like Protein; [f] *csoS1-ndhF1*; [g] *ndhF1-ndhD2*; [h] *ndhD2-chpX*; [i] *chpX*-HP; [j] HP-*ATPase*.

Table 6. (Cont'd)

Intergenic region	*P.m.m.* CCMP 1375 NC_005042	*S.* sp. CC 9311 NC_008319	*S.* sp. CC 9605 NC_007516	*S.* sp. CC 9902 NC_007513	*S.* sp. WH 8102 NC_005070
csoS1-cbbL	68	68	71	69	73
cbbL-cbbS	103	103	59	119	59
csbbS-csoS2	102	99	101	104	101
csoS2-csoS3	7	7	7	7	7
csoS3-csoS4A	1	2	2	2	2
csoS4A-csoS4B	-1	-1	-1	-1	-1
csoS4B-csoS1	45	154	53	52	52
	72[e]	79[f]	111[f]	112f	111[f]
		7[g]	10[g]	8[g]	10[g]
		-4[h]	-4[h]	-4[h]	-4[h]
			34[i]	14[i]	25[i]
			-17[j]	-17[j]	-17[j]

transcription initiation generally is a purine (A>>G) [31,34]. Another possibility that the *H. neapolitanus cso* operon may have two adjacent start sites, such as those being reported for some *Prochlorococcus* genes [27,82], cannot be completely ruled out, because the gene organizations of α-carboxysomes are so conserved that a horizontal gene transfer event of the *cso* operon was proposed earlier [12,13].

The -10 signature motif of the *H. neapolitanus cso* promoter is very similar to the hexa-nucleotide motifs found in several marine cyanobacteria (Figure 18b) [82]. It also matches that of the *E. coli lacI* promoter, but differs from the consensus σ^{70} Pribnow box in *E. coli* [31,34] (Figure 18b). On the other hand, the -35 region of the *cso* promoter is quite divergent from the -35 motif of the *E. coli* σ^{70} promoter (Figure 18b). The *E. coli* σ^{70} promoter is well-known as the "housekeeping" promoter which initiates transcription of most genes in growing *E. coli* cells [30,41], and promoter prediction algorithms take both -35 and -10 motifs into account. These explain why the promoter and operon prediction programs, which are mainly based on experimental data of *E. coli* promoters, are not very accurate for *H. neapolitanus*.

As expected for an operon, measurable transcript levels were detected for all carboxysome genes (Figure 17). The presence of *csoS4A* and *csoS4B* transcripts at levels comparable to those *csoS2* and *csoS3* was the first evidence showing that these two genes are transcriptionally active. Although the function of CsoS4 proteins was not clear at this point, they are expected to be important for carboxysome biogenesis and/or function. The transcripts of *csoS1A*, *csoS1B*, and *csoS1C* were present at about 1:1:1 ratio (Figure 17). This indicated that all three proteins, although almost identical in amino acid sequence, are likely necessary to form the carboxysome shell. Since CsoS1A and CsoS1C only

differ in two amino acid residues and cannot be distinguished immunologically [32], the finding that comparable amounts of CsoS1A and CsoS1C both are needed was significant.

Steady state levels of individual *cso* gene transcripts in *H. neapolitanus* differed from each other by at least an order of magnitude, but they matched protein levels of corresponding gene products measured in the purified carboxysome (Figure 17). This clearly indicated that the control of carboxysomal protein expression occurs at the transcriptional level. A sharp drop of mRNA levels after *cbbS* gene as well as a significant rise of transcript levels for the last *csoS1* gene set was observed. This implied the occurrence of mRNA processing after the primary polycistronic mRNA being synthesized from the *cso* promoter, and it led to the next section of discussion.

2. The Master *cso* Transcript Undergoes mRNA Processing and the Resulting Smaller mRNA Species Have Different Stabilities

In general, the observation that protein products of the first set of genes and the last set of genes in an operon are much more abundant than proteins encoded by those genes located in the middle of the same operon could be explained by several regulatory mechanisms: (1) Shine-Dalgarno sequences for the individual genes of the polycistronic mRNA have different ribosome binding abilities; (2) internal cryptic promoter(s) initiate the transcription of its downstream genes; (3) RNA secondary structures, e.g. stem-loop structures, act as terminators of transcription and result in mRNAs prematurely; and (4) the master transcript undergoes mRNA processing and the resulting smaller mRNA species have different stabilities. In case of the *cso* operon in *H. neapolitanus*,

interestingly, putative Shine-Dalgarno sequences are only found for those genes which encode for abundant carboxysome proteins, RubisCO (CbbL and CbbS), CsoS1A, CsoS1B, and CsoS1C (Table 7). This observation seemed to support the regulatory mechanism (1) that different ribosome binding abilities cause differential protein expressions. However, the observed correlation between transcript abundance and carboxysome protein ratios indicated that regulation of the *cso* gene expression occurs mainly at the transcriptional level. Therefore, this mechanism was ruled out.

To test if mechanism (2) is true, both *in vitro* and *in vivo* experiments were conducted for extensive searching of internal promoter(s). In primer extension experiments no transcription start site was identified upstream from the *csoS1C* gene. Also, promoter activity was not observed either from sequence upstream of *csoS4A* or from sequence upstream of *csoS1C* in *E. coli* (Figure 21b), while the *cso* promoter was shown to

Table 7. Putative ribosome binding sites in the *cso* operon of *H. neapolitanus*.

gene	upstream sequence	start codon
cbbL	---CATCCTTTCAGGAGGAACTC	*ATG*GCAGTT---
cbbS	---ACATACTAAGGTGAGTAACC	*ATG*GCTGAA---
csoS1C	---CTTTATTGAGGAGAGAAGAA	*ATG*GCAGCA---
csoS1A	---CAGCTTAGGAGTTTATTTAA	*ATG*GCTGAT---
csoS1B	---CTAGTTTTAGAGGATCTGTT	*ATG*GCAACG---

Putative Shine-Dalgarno sequences are underlined. And the start codon of each gene is boxed.

function in same promoter probe vector (Figure 21c-d). However, one might argue that the failure of identifying internal TSS(s) may be caused by mispriming within the highly homologous (≥ 70%) *csoS1* gene coding regions, and promoters of *H. neapolitanus* may not function appropriately in *E. coli* to produce the green fluorescence protein (GFP) at a detectable level. Well, an unambiguous conclusion is that the *cso* operon of *H. neapolitanus* does not contain any internal cryptic promoter can be drawn with a piece of additional evidence from the *cso* promoter deletion mutant, where no carboxysome protein was detected immunologically (Figure 23).

The possibility of mechanism (3) that RNA secondary structures act as transcriptional terminators was also considered. In bacterial operons, RNA secondary structures serving as Rho-independent transcriptional terminators are often found [1,15,66]. They modulate the transcript levels of downstream genes and when such terminating event occurs, only partial-length transcripts can be detected [1,15,66]. In the *cso* operon of *H. neapolitanus*, the predicted stem-loop structures between *cbbS* and *csoS2* (Figure 19) may function in the same way, although the four major 3'-ends are located further downstream within the *csoS2* coding region instead of within the short uracil tract that follows the G/C-rich stem-loop (Figure 20a). This highly implied that the function of the *cbbS* to *csoS2* stem-loop structure is protecting degradation of the *cbbL/S* transcript rather than causing transcriptional termination. On the other hand, despite the prediction of similar stem-loop structures between *csoS4B* and *csoS1C*, no transcript 3'-end was found near this region, which again suggested that the role of this stem-loop structure is not that of a transcriptional terminator.

To conclude, the possible regulatory strategy could be that the primary transcript of

the *H. neapolitanus cso* operon undergoes differential processing and decay which results in several shorter fragments with different half lives. Endo-ribonucleases may cleave the full-length transcript at intergenic region between *csoS4B* and *csoS1C* immediately upstream from the predicted stem-loop structures, and this cleavage results in two mRNA fragments. The newly exposed 3'-end then is attacked by exo-ribonucleases and the transcript degrades rapidly from the 3' to 5' direction in the region extended from *csoS4B* to *csoS2* genes. Further degradation of *cbbL/S* transcript could be prevented by the predicted *cbbS* to *csoS2* stem-loop structures which are located at the 3'-end of *cbbL/S*. The predicted *csoS4B* to *csoS1C* stem-loop structures which are located at the 5'-end, on the other hand, stabilize the remaining *csoS1C/A/B* transcript, although the mechanism behind this stabilization is not clear at this point. Although it has been widely accepted that bacteria do not have 5' to 3' exo-ribonuclease [6,7,19], a recent article reported 5' to 3' exoribonuclease activity in *Bacillus subtilis* [50]. So the stem-loop structures between *csoS4B* and *csoS1C* protect the transcript of *csoS1C/A/B* from 5' to 3' exoribonucleolytic activity is possible. Furthermore, according to the fact that putative Shine-Dalgarno sequences are found for genes *cbbL*, *cbbS*, *csoS1C*, *csoS1A*, and *csoS1B*, a series of ribosomes could bind to these transcripts simultaneously and provide protection for these regions from endonucleolytic cleavage as well.

3. All Carboxysomal Gene Products Are Present in Functional Carboxysome Particles of *H. neapolitanus*

Presence of multiple *csoS1/ccmK* genes in one genome is a common feature shared by carboxysome-containing bacteria. CsoS1/CcmK paralogs usually are highly similar to

each other within one genome [12,42,71]. However, whether or not all copies are required for carboxysome biogenesis and/or structure remained unclear. CcmK1 was identified as carboxysome component by proteomic techniques in *Synechococcus* sp. PCC 7942, while its homolog protein CcmO has not been found during carboxysome isolation procedure [46]. Analysis of purified carboxysomes from *Thiomicrospira crunogena* via electrophoresis and immunoblotting confirmed the presence of CsoS1 protein at levels comparable to those of *H. neapolitanus,* but failed to provide further information on which one(s) of those three CsoS1 proteins are present (Balaraj Menon, dissertation). In the case of the model organism *H. neapolitanus*, although CsoS1B has been unambiguously identified as major shell component for decades, whether both of CsoS1A and CsoS1C are present in shell and what is the ratio of their relative amount remained two open questions [14,32]. The detection of transcripts for both *csoS1A* and *csoS1C*, and the fact that their total transcript levels match the amount of polypeptide bands that migrated around 7 kDa on a 12% denaturing polyacrylamide gel of purified carboxysome, could only be explained that both CsoS1A and CsoS1C, although nearly identical, are present in wild type carboxysome shell. More direct evidence was provided with the *csoS1A*-TC tag mutant, where a tetracysteine (TC) motif Cys-Cys-Pro-Gly-Cys-Cys [2,25] was fused on the C-terminal of *csoS1A* gene. This TC tag provides efficient mass difference to distinguish CsoS1A-TC from CsoS1C, and immunoblotting confirmed the presence of both protein in mutant cell crude extract and mutant carboxysomes (Figure 26). However, since CsoS1A and CsoS1C are nearly identical, what kind of advantage it might provide to acquiring both proteins in shell remains unknown at this point. Or CsoS1A and CsoS1C can be functionally redundant to

each other, but the ratio of available CsoS1B protein to CsoS1A/C protein in the cytosol is essential for correct assembly of carboxysomes. If this is the case, containing two copies of *csoS1A/C* gene instead of one in the *cso* operon would diminish post-transcriptional regulation. The comparable transcriptional levels of *csoS1A* and *csoS1C* genes measured with real-time PCR (Figure 17) and their similar putative Shine-Dalgarno sequences (Table 6) together suggested that in wild type carboxysome CsoS1A and CsoS1C protein are present with 1:1 ratio. However, in *csoS1A*-TC mutant, CsoS1A-TC fusion protein and CsoS1C protein are present with approximately 1:6 ratios in enriched carboxysome particles based on densitometries of stained polypeptide bands (Figure 26b). Observed CsoS1A-TC to CsoS1C ratio differed from the expected CsoS1A to CsoS1C ratio could be explained with one or some the following possibilities. Firstly, the addition of a TC tag, though only six amino acid residues in length, largely reduces the packing ability of CsoS1A-TC fusion protein into the carboxysome shell compared to the ability of native CsoS1A protein. Secondly, CsoS1A-TC protein may not stain as well as CsoS1C protein, and this could result in under-estimating the amount of CsoS1A-TC protein. Thirdly, containment proteins which migrate to same positions as CsoS1C protein does may be present in the carboxysome enriched sample, and this could result in over-estimating the amount of CsoS1C protein. To test the first possibility, one would need to measure association and/or dissociation constant between tagged or native CsoS1A protein and other carboxysomal shell proteins. The observed relative signal strength on immunoblot for CsoS1A-TC to CsoS1C is 5:1 and it is opposite to the 1:6 ratio calculated with stained polypeptide pattern. This fact indicated that anti-CsoS1 antiserum recognizes these two proteins with different strength and therefore cannot be

used for estimating their present amounts. So, to verify the second possibility, one would have to establish a correlation between amount of protein and signal strength from immunoblot for both CsoS1C and CsoS1A-TC proteins. The good news is that the third possibility can be ruled out by further purifying mutant carboxysomes from the enriched sample.

Little was known about the other microcompartment-conserved genes, *csoS4A/B* (formerly *orfA/B*), until the steady state transcript profile of carboxysome operon was established [10]. The fact that *csoS4A* and *csoS4B* also are transcriptionally active suggested that the CsoS4 proteins may play an important role in carboxysome biogenesis and/or function. Expression of the CsoS4 protein heterogeneously was the first step to investigate their function. The prokaryotic expression vector pProEX has been successfully used to express recombinant CsoSCA [75] and CsoS1 proteins [80], but failed in producing recombinant CsoS4 proteins. However, soluble recombinant CsoS4A and CsoS4B proteins were generated with the pET22b(+) vector, where an affinity tag was added at C-terminal. Later, crystal structure of the CsoS4A protein revealed that the protein was a pentamer, and that the N-terminus of CsoS4A was relatively hydrophobic and buried inside while the C-terminus was more exposed to the surface [77]. In the pProEX vector, a peptide with 25 amino acid residues in length, including a his_6-tag, was designed to be relatively hydrophilic in order to be exposed on the surface of recombinant protein for affinity purification purpose. When this hydrophilic tag fused to the N-terminus of CsoS4 protein which is hydrophobic, it may prevent the recombinant CsoS4 protein from folding appropriately making it an easy target for endogenous protease. This might explain why recombinant CsoS4 protein could not be produced in

pProEX vector. Since CsoS4A and CsoS4B only share approximately 47% homology [13], it is unexpected to see rabbit antisera generated against recombinant CsoS4A or CsoS4B cross-react equally strongly with these two proteins thereby failing to distinguish them. With the aid of two dimensional electrophoresis and imunoblotting, both CsoS4A and CsoS4B were identified in the carboxysome as minor shell components with similar copy numbers. This is the first direct evidence to show the presence of *csoS4/ccmL/eutN* gene product in a bacterial microcompartment, and therefore it is a significant finding for understanding the role that this gene plays.

4. The Expression of CsoS4A and CsoS4B Is Tightly Correlated

In the *cso* operon of *H. neapolitanus*, the *csoS4A* and *csoS4B* genes feature overlapping coding sequences, which is quite common in α-carboxysome operons (Table 4, Table 5): 14 out of 20 examined cases have this overlapping pattern, among which over 90% (13 out of 14) overlaps by one nucleotide. The extent of this overlapping coding sequence not only implied a horizontal gene transfer event of carboxysome operon across species, but also suggested that *csoS4A* and *csoS4B* might undergo a translation coupled event, since *csoS4B* does not have its own Shine-Dalgarno sequence. Furthermore, overlapping tandem *csoS4* genes are a unique feature in α-carboxysome, while only one copy of *csoS4* homolog can be found in β-carboxysome gene clusters and in *pdu* or *eut* operons. Considering that ribosomal frameshifting might result in a single protein from two or more overlapping genes [24,28], *csoS4A* and *csoS4B* genes were co-constructed into *E. coli* expression vector, but no fusion protein was found in cell crude extract (data not shown). An intriguing finding with *csoS4A* and *csoS4B* gene knockout mutants was

that replacement of either gene with a *Kan*^{r} cassette resulted in absence of expression of the other gene. In the *csoS4A* knockout mutant, *csoS4B* gene was not expressed due to the lack of its own ribosome binding site. On the contrary, the *csoS1* genes, which reside downstream from *csoS4B* and have their own Shine-Dalgarno sequences were expressed at comparable levels as in wild type (Figure 43). However, the same explanation cannot explain the absence of the CsoS4A protein in the *csoS4B* knockout mutant where the *csoS4A* gene coding sequence was unchanged and located upstream from the *Kan*^{r} insertion. One possibility is that the local RNA secondary structure of *csoS4A*, which will not form in the presence of *csoS4B* transcript on same polycistronic mRNA, blocks the movement of ribosomes and causes the ribosome's release before full-length CsoS4A is synthesized. Whether this assumption is true or false, it is clear that the expression of CsoS4A and CsoS4B polypeptides is tightly correlated. Although the importance of this correlated expression is still not clear at this point, the reasons for *H. neapolitanus* to ensure the producing of CsoS4 proteins, which only have extremely low abundance in the functional carboxysomes, will be discussed in the next section.

5. CsoS4 Polypeptides Are Not Required for Compartmentalization of RubisCO

Recently a preliminary model of carboxysome shell at atomic-level was proposed mainly based on crystal structures of both α- and β-carboxysome shell proteins [77]. In this model, CsoS1 or CcmK proteins from α- or β-carboxysomes, respectively, form hexamers and then constitute facets of the icosahedral body. And the pentamers of CsoS4 or CcmL from α- or β-carboxysomes, respectively, was proposed to occupy the vertices [77]. This model agreed with the previous observation that *Synechococcus* sp. PCC7942

ccmL insertion mutants produce aberrantly shaped β-carboxysomes [60,62]. Therefore, it is not surprising to observe elongated α-carboxysomes in the *csoS4* gene(s) knockout mutants of *H. neapolitanus*. What is really unexpected is the observation of mutant carboxysomes with apparently normal size and shape as well (Figure 42). If these mutant carboxysomes have icosahedral geometry, a further question will be which protein, other than CsoS4, can fulfill the role of forming pentamers at the vertices. Quasi-equivalence could be one explanation, which means that the interchangeable formation of hexamer and pentamer by the same protein molecules [16,40]. Quasi-equivalence is quite common in the assembly of some virus capsid where more than 60 copies of single gene product arrange to form icosahedral symmetry [38,39]. The EutN protein from *E. coli*, which is a homolog of the CsoS4 protein, was identified as having a similar three-dimensional fold as CsoS4A, but it crystallized into a hexamer instead of a pentamer [77,85]. With this example, it is possible that one of those CsoS1 proteins can assume pentamer conformation under certain conditions which are not favored during crystal formation. An alternative possibility which cannot be completely ruled out without further experiments is that an unknown protein of *H. neapolitanus*, other than any well-known carboxysomal protein, could fulfill the role of pentamer. This protein cannot be easily identified with its extremely low abundance, which explains the observation that the protein compositions of wild type and mutant carboxysomes seemed to be the same except for the presence or absence of the CsoS4 protein, respectively (Figure 46). Regardless of which protein forms pentamers, the ratio of available pentamers to hexamers in the cytosol could be essential for the correct assembly of an intact icosahedral structure. When fewer required pentamers are available, hexamer sheets might extend further until the insertion of a

pentamer provides curvature and that may be how elongated structures form. Different from the above explanations, a third possibility would be the mutant carboxysome with "normal" shape is rather a round structure instead of being icosahedral. If this is the case, pentamer units are not required to close the structure, with the penalty of losing distinct vertices. At this point, limited information is available to determine which explanation is correct.

During the purification of mutant carboxysomes lacking CsoS4, significant amounts of elongated ones were lost (Table 4). This implied that the elongated carboxysomes are more fragile and less resistant or stable to mechanical force during purification steps. Kinetic characterization of purified mutant carboxysomes only can reveal the nature of those with "normal" appearance, but could not represent the elongated ones because less than 2% of the purified mutant carboxysomes are elongated (Table 4). It will be interesting to compare these two groups in protein composition and kinetics of CO_2 fixation. To be able to do this comparison, the technical difficulties of purifying elongated mutant carboxysomes must be overcome first.

6. Carboxysome Shells Lacking CsoS4 Proteins Cannot Function as a Diffusion Barrier

As mentioned previously, the *csoS4AB::Km* mutant still produces carboxysomes without the CsoS4 proteins, and the majority of mutant carboxysomes have normal appearance and composition. It was surprising to find that the *csoS4AB::Km* mutant has a strict *hcr* (high CO_2 requiring) phenotype, especially considering the fact that *H. neapolitanus* mutants where the carboxysomal carbonic anhydrase (CsoSCA) is deleted or the native carboxysomal RubisCO is replaced are able to grown in air [21,52]. The

observation of elongated mutant carboxysomes preventing cells from division only explains why mutant cells cannot reach the same cell density as wild type cells but cannot cause strict *hcr* phenotype if mutant carboxysomes are fully functional. Radiometric RubisCO activity assays revealed that the Kc for mutant carboxysomes is 124.0 ± 5.7 μM CO_2, which was lower than 163.3 ± 5.7 μM CO_2 measured for intact wild type carboxysomes [11]. In fact, this Kc value for mutant carboxysomes is very close to the value measured for broken wild type carboxysomes (126.7± 10.8 μM CO_2) [21]. This led to the assumption that regardless of the "normal" appearance or near identical protein composition to their wild type counterparts, mutant carboxysome without CsoS4 protein are more permeable to HCO_3^-. Stopped-flow changing pH indicator assays for carbonic anhydrase activity clearly showed both faster bicarbonate dehydration rate and faster CO_2 rehydration rate of mutant carboxysomes compared to that of wild type carboxysomes [11], and this result supported the assumption that the shell of mutant carboxysomes with normal appearance is leaky. Obviously, the absence of CsoS4 proteins caused the loss of important protein-protein interactions where CsoS4 may or may not evolved, and further resulted in a loosely packed shell. In addition, RubisCO molecules sequestered in mutant carboxysomes are more exposed to cytoplasmic HCO_3^-, CO_2, and O_2. A previous study on a carboxysomal carbonic anhydrase (CsoSCA) deletion mutant suggested that the carboxysome shell can function as a diffusion barrier of CO_2 into and out of the carboxysome and therefore provides catalytic advantage to RubisCO [21]. The leaky shell of a mutant carboxysome cannot function as a diffusion barrier and therefore cannot provide the same catalytic advantage. This would be the main reason for the strict *hcr* phenotype.

7. Carboxysome Shell Surface: More Questions than Answers

The current model for the carboxysome shell, which is mainly based on crystal structures [77], is incomplete in several respects. First, the question of which side of CsoS1 hexamer faces the cytosol remained unanswered. Second, this model needed to be refined to explain the strategy of integrating other shell components, CsoS2A/B and CsoSCA, into such tightly packed matrix. Third, the function of the CsoS1 hexamer central pores and the way through which metabolite flux gets into and out of carboxysome are not clear. And last but not the least, what is the potential function or structure advantage to have C-terminal extension on some major shell proteins (e.g. CsoS1B and CcmK4). These questions represent different aspects of the model and also are closely related to each other.

The answer of "sidedness" question is essential for understanding structure and function of carboxysome shell, although to date no solid conclusion can be drawn yet from various approaches that have been taken. However, the concave side facing cytosol seemed to be more favored in this study. In that case, Lys94 which is on the concave side of CsoS1A/C is always accessible to exterior solution, and therefore CsoS1A/C proteins were modified by amine-active chemical sulfo-NHS-biotin at approximately same level in both intact and broken carboxysome (Figure 36). And there is also made the observation in *csoS1B* truncated mutant, that the presence or absence of CsoS1B C-terminal extension does not affect integration of CsoSCA in the carboxysome shell (Figure 32), reasonable. Because if the C-terminus of CsoS1B, which is located on the concave side, is facing outward, it presumably is not accessible for the CsoSCA protein which was shown be in the interior side of the carboxysome shell [21]. A previous finding

in yeast twohybrid assay that CsoSCA could specifically interact with CsoS1B but not CsoS1A or CsoS1C could be a false positive result. The CsoS1B tail may interact with CsoSCA somehow with a conformation that is not physically and spatially accessible in real carboxysome structure.

The conserved positively charged pore in the CsoS1/CcmK hexamer structure combined with the visualization of sulfate ions inside the pore of CsoS1A suggested the role this pore plays in transporting metabolites related to CO_2 fixation [42,80]. Site-direct mutant *csoS1A::G43A* which is designed to close the pore in the CsoS1A hexamer did exhibit defective growth (Figure 25), while the appearance and size of mutant carboxysomes are similar to their wild type counterparts (Figure 40). This is the first experimental evidence that supported the transport channel hypothesis, and it suggested that the pores do play an important role in functional carboxysomes. Further characterization needs to be done of this mutant to further elucidate the function of the pore. A rigorous functional study should also include mutants where pores of two BMC proteins or all three BMC proteins are disrupted. Recently, a putative carboxysome protein CsoS1D of *Prochlorococcus* MED4 was reported that has a unique crystal structure, which contains two BMC structures [43]. The structure of N-terminal BMC can be superposed on any BMC protein structures resolved to date, but the amino acid sequence does not share any homology to known BMC domains. This tandem BMC structure trimerizes to form a pseudohexameric unit which can be incorporated into the CsoS1A layer, and the pore at its 3-fold axis of symmetry appears to have both gated or open conformation. The *csoS1D* homolog, *gene2345*, has also been found in the *H. neapolitanus* genome recently, located at about 11 kb downstream from the carboxysome

operon. (GeneBank ID EEG96231.1). The deduced peptide sequence of *gene2345* shares 60% identities and 77% similarities to CsoS1D from *Prochlorococcus* MED4. It would be interesting to investigate the function of this protein in *H. neapolitanus.* However preliminary evidence suggested it was not expressed in wild type cells grown in either ambient CO_2 or 5% CO_2[11].

CHAPTER V

CONCLUSIONS AND FUTURE WORK

The first conclusion drawn from this study was that the nine carboxysomal genes of *H. neapolitanus* were transcribed from a single promoter located upstream from the first gene (*cbbL*), and that the carboxysomal genes form an operon. The results of promoter activities examined *in vivo* with a promoter reporter vector and the characterization of a *cso* promoter deletion mutant led to the conclusion that the *cso* promoter is the only promoter present in the *cso* operon. Therefore, the fact that the abundance of carboxysomal gene transcripts match the levels of their protein products in purified carboxysomes is likely the result of primary transcript processing and differential decay. The present study also provided clear evidence for the first time that all three CsoS1 paralogs and both CsoS4 proteins are present in the carboxysome shell. This conclusion was based on the preliminary characterization of a *csoS1A*-TC tag mutant and the result of immunoblots coupled with two-dimensional electrophoresis of a carboxysome shell enriched fraction. An additional significant finding of this study was that the presence of CsoS4 proteins was not required for assembly of carboxysomes with apparently normal shape. However, lack of CsoS4 proteins did result in a more permeable shell which cannot provide the catalytic advantage of the wild type shell to RubisCO. Lastly, although more information is still needed to obtain a comprehensive understanding of how shell components are integrated, some light has been cast on details of carboxysome shell architecture: (i) the association of CsoSCA with the inner side of the shell was shown not to be mediated through its interaction with the C-terminal extension of CsoS1B protein; (ii) both CsoS2A and CsoS2B protein were at least partially, if not

completely, exposed on outer surface of the shell; and (iii) the experimental observations collected so far favor a shell model in which the concave side of CsoS1 hexamers face the cytosol.

Although the BMC proteins CsoS1A, CsoS1B, and CsoS1C, which are almost identical except the 12 amino acid C-terminal extension on CsoS1B, were identified as the major shell components, it remains unclear if they play different roles related to carboxysome function. One possibility is that CsoS1B, because of its C-terminal tail, cannot be replaced by CsoS1A or CsoS1C, while CsoS1A and CsoS1C are functionally redundant. Having both CsoS1A and CsoS1C in the shell may be a result of a gene regulation strategy to maintain the appropriate ratio of CsoS1A/C to CsoS1B. Further characterization of the *csoS1B*-truncated mutant could help understand the function of the CsoS1B tail. Deletion mutations where one, two, or all three *csoS1* genes are knocked out could test the assumption about functional redundancy of CsoS1 proteins. Recently, a *csoS1D* homolog, *gene2345*, was also found in *H. neapolitanus* genome. Based on homology modeling, the protein product of this gene contains same tandem BMC structures as being observed in CsoS1D protein from *Prochlorococcus* MED4 [43]. It will be interesting to investigate transcription and expression of this gene in wild type *H. neapolitanus* as well as in single, double, or triple *csoS1* gene deletion mutants.

The reasons why α-carboxysome operon contains two adjacent *csoS4* genes was mysterious in light of the fact that there is only one *csoS4* ortholog in β-carboxysome gene clusters and other BMC operons. Characterization of *csoS4A::Km* and *csoS4B::Km* mutants clearly indicated that *csoS4A* and *csoS4B* genes were transcriptionally and/or translationally coupled, although the mechanism behind this phenomenon remains

unknown. Another unsolved question is that if there is functional redundancy between two CsoS4 proteins. To gain better insight into the three-dimensional architecture of the elongated carboxysomes observed in the *csoS4AB::Km* mutant, cryo-electron tomography studies, which can deliver three-dimensional information of objects' near-native structures, will be necessary. This technique could also facilitate an evaluation of icosahedral geometry among mutant carboxysomes with similar size and appearance as wild type carboxysomes. If their icosahedral geometry is confirmed, a following question will be which protein fulfills the role of pentamers at the vertices.

Previous structural studies have established a preliminary model of the carboxysome shell [77,80]: hexameric and pentameric building blocks formed by CsoS1 and CsoS4 proteins, respectively, pack tightly to form the icosahedral shell. The positively charged pores of CsoS1 hexamerss might act as transport channels for small metabolites involved in the CO_2 fixation reaction. Partially closing the pores did affect growth of the *csoS1A*::G43A mutant, which provided the first experimental evidence that the pores are important for carboxysome function. However, further characterization in single, double, or even triple pore mutations will be needed to better elucidate the function of CsoS1 pores in RubisCO substrate and products transferring across the carboxysome shell. Furthermore, although preliminary data in this study favors the concave side of CsoS1A hexamers facing outward, the "sidedness" question still begs an answer. To clarify the ambiguous observations gained so far, a study involving labeling the carboxysome surface with fluorescent tags and identifying labeling position(s) with the aid of enzymatic digestions and mass spectrometry techniques is currently being conducted in our lab. The *csoS1A*-TC tag mutant can be used for an alternative approach to localize the

concave side of CsoS1 hexamers with electron microscopy, because the tetracysteine (TC) tag can be specifically stained with the biarsenical reagents ReAsH or FIAsH and further be counter-stained with osmium tetroxide after photoconversion [2,25]. Last but not least, a minor modification of the sulfo-NHS-biotin labeling experiment might also provide a more interpretable result: immobilizing sulfo-NHS-biotin on streptavidin beads before using it to label the carboxysome surface could prevent the reagent from gaining access to the interior of unbroken carboxysomes.

APPENDIX *

A1. Oligonucleotides used for real-time RT-PCR

Gene/range	Primer name	Sequence
cbbL	cbbLf1151	5'-TCTACTCTGGGCTGGATTGATTTG-3'
	cbbLr1420	5'-CGCTTCTACGCAGGCTTCC-3'
cbbS	cbbSf1630	5'-CAAGCAAAGCCTCAAATATG-3'
	cbbSr1864	5'-TGTTGGATACGCACTACG-3'
cbbS–csoS2 I	cbbSf1878	5'-TGGTGGCTTATGACAACTATG-3'
	csoS2r2109	5'-CTGCTTTCCCTTGCTTGG-3'
cbbS–csoS2 II	cbbSf1877	5'-CTGGTGGCTTATGACAACTATGC-3'
	S2r2154	5'-CCTGTTTAGTGCTACGGTTTACG-3'
csoS2	S2f4018	5'-ACACCAATCAGCCAGAAC-3'
	S2r4275	5'-CTTGAGCAGCCTTATTACG-3'
csoS3 I	S3f5368	5'-CACCTCTGACCCGACACACTCTG-3'
	S3r5545	5'-GCGAATGGCATCCGTATCCGTATC-3'

The numbers in the primer names refer to the positions of the primer 5'-ends with respect to the numbering system used in Figure 10 and in Genbank (accession no. AF038430); the letters "f" and "r" indicate forward and reverse primer, respectively.

* All oligonucleotides were synthesized by Integrated DNA Technologies (Coralville, IA)

A1. Oligonucleotides used for real-time RT-PCR (Cont'd)

Gene/range	Primer name	Sequence
csoS3 II	S3f4702	5'-TCGGCTTGACTTGATTGAG-3'
	S3r4851	5'-CGGTTCTTGACGCTATCC-3'
csoS3–orfA	S3f5978	5'-TTACCGATTCCGATTGCC-3'
	orfAr6268	5'-GTTTCTCCCATACCACTAATAG-3'
orfA (*csoS4A*)	orfAf6211	5'-TCAACAAACCGTATTGCTGATATG-3'
	orfAr6351	5'-TGCTGCCGATGACCCAAC-3'
orfA–orfB	orfAf6327	5'-TTTGTGCGTTGGGTCATC-3'
	orfBr6472	5'-CTGCGTGTTGCGATTAGG-3'
orfB (*csoS4B*)	orfBf6460	5'-TCGCAACACGCAGGATTC-3'
	orfBr6618	5'-AATCACCCACGCCAAACC-3'
orfB–csoS1C	orfBf6523	5'-GTAAGGTCAGTGTCGCTTGC-3'
	1Cr6735	5'-CTGCTGCCATTTCTTCTCTCC-3'

A1. Oligonucleotides used for real-time RT-PCR (Cont'd)

Gene/range	Primer name	Sequence
csoS1C	1Cf6725	5'-AATGGCAGCAGTAACAGGTATTG-3'
	1Cr6935	5'-TCGTTCGCAGGCATCAGC-3'
csoS1C–csoS1A	1Cf6869	5'-TTTGGTCCGTGGTGAAAC-3'
	1Ar7161	5'-TCAATCGCAGGAACTAAGC-3'
csoS1A	1Af7106	5'-GTAACTGGTATTGCTCTGGGTATG-3'
	1Ar7405	5'-GGAATATCTGACTTAGGCTTGTGG-3'
csoS1A–csoS1B	1Af7381	5'-GCCACAAGCCTAAGTCAGATATTC-3'
	1Br7666	5'-TGTCTCACCACGAACCATTACG-3'
csoS1B	1Bf7619	5'-TCATTTGTTGGCGGCGGTTAC-3'
	1Br7794	5'-TCGGGCGTCTCGGGTAGG-3'

A2. Oligonucleotides used for primer extension

Gene/range	Primer name	Sequence
cbbL upstream	cbbLr164	5'-TTTTAACTGCCATGAGTTCCTCCTGAAAG-3'
	cbbLr243	5'-GCAAGGATGTCGGAATCCAACGG-3'
	cbbLr374	5'-CTTTGTAGTAGTCCATGTCGGTCAG-3'
csoS1C upstream	1Cr6746	5'-ATGCAAAGAACCGGAACAAGCCT-3'
	1Cr6845	5'-AAATTGACGGCCAACCAACGTACTTCG-3'
	1Cr6939	5'-CGACTCGTTCGCAGGCATCAGC-3'
csoS1C–csoS1A	1CAr7069	5'-ATGCAAAGAACCGGAACAAGCCT-3'

The numbers in the primer names refer to the positions of the primer 5'-ends with respect to the numbering system used in Figure 10 and in Genbank (accession no. AF038430); the letters "f" and "r" indicate forward and reverse primer, respectively.

A3. Oligonucleotides used for RLM-3'RACE

Gene/range	Primer name	Sequence
-	adaptor oligo	5'-pGCCTAGGCTTAAGCTCGTAGCTTTCGACGC CGGCGp-3'
-	adaptorREV	5'-GAAAGCTACGAGCTTAAGCCTA-3'
cbbS downstream	cbbSf1872	5'-TCAAACTGGTGGCTTATGACAACTATG-3'
orfB downstream	orfBf6559	5'-CCGAGGGATGTTGGGTCTTTACG-3'
csoS1B downstream	1Bf7641	5'-TGACCGTAATGGTTCGTGGTGAGA-3'

The numbers in the primer names refer to the positions of the primer 5'-ends with respect to the numbering system used in Figure 10 and in Genbank (accession no. AF038430); the letters "f" and "r" indicate forward and reverse primer, respectively.

A4. Oligonucleotides used for plasmid constructions

Construction of promoter probe vectors	
Primer name	**Sequence**
pTrcf11	5'-AAT*AAGCTT***TTATCATCGACTGCACG**-3'
pTrcr252	5'-AAT*GAATTC***GTGAAATTGTTATCCGCT**C-3'
pTn1f2811	5'-AAT*AAGCTT***GTGATCGACCAAGCCAGC**-3'
pTn1r3012	5'-GGA*GAATTC***AAAGGATGGGGGATTCGC**-3'
1Cf-202	5'-AAT*AAGCTT***TAAGGTCAGTGTCGCTTGC**-3'
1Cr-13	5'-GGA*GAATTC***CAATAAAGACTCATTCAAAATC**-3'
orfAf-210	5'-AAT*AAGCTT***GAGTCGAGGTTTACCGATTC**-3'
orfAr-26	5'-GGA*GAATTC***CTTCCGATATCCTGCG**-3'

Cloning *csoS4A* and *csoS4B* into pProEX-HTb vector	
Primer name	**Sequence**
oAfBamHI	5'-CCTCCT*GGATCC***ATGAAAATCATGC**-3'
oArHindIII	5'-CCTCCT*AAGCTT***TTACTCACCATTCCAC**-3'
oBfBamHI	5'-CCTCCT*GGATCC***ATGGAAGTAATGCGCG**-3'
oBrHindIII	5'-CCTCCT*AAGCTT***GATGGTCAAGTTACCC**-3'

The numbers in the primer names refer to the position of the primer 5'-end with respect to the first codon of the specific gene; the letters "f" and "r" indicate forward and reverse primer, respectively. Underline remarks restrition site, and bold letters indicate homologues region to PCR template.

A4. Oligonucleotides used for plasmid constructions (Cont'd)

Cloning *csoS4A* and *csoS4B* into pET-22b(+) vector	
Primer name	**Sequence**
oAfNdeI	5'-GGTGGT*CATATG***AAAATCATGCAAGTTGAG**-3'
oAfMscI	5'-CGA*TGGCCA***TGAAAATCATGC**-3'
oArXhoIns	5'-GGTC*CTCGAG***CTCACCATTCC**-3'
oBfNdeI	5'-GGTGGT*CATATG***GAAGTAATGCGCGTTCG**-3'
oBfNcoI	5'-CGA*CCATGG***AAGTAATGCGCG**-3'
oBrXhoIns	5'-GGTC*CTCGAG***AGTTACCCAGTGATCG**-3'

A5. Oligonucleotides used for generating *Halothiobacillus neapolitanus* mutants

Mutant	Primer name	Sequence
cso promoter deletion mutant	ProKmF	5'-**CGTATCTAACGAGAACCAGATCATAGATTAGTTCCTATGTCGAAGAT AGAAAATATCGA**TCAGAAGAACTCGTCAAGAAGGCG-3'
	ProKmR	5'-**GGATGGGGGATTCGCGCAACGCCCCTGTGGCTTGTCGCGTGAAAA G**CCGGAATTGCCAGCTGGG-3'
csoS1B truncated mutant	1BtKmF	5'-**CGTTCATTCTGAAGTTGAGATCATCCTACCCGAGACGCCCGAAGA C**TAATGA[a]CCGGAATTGCCAGCTGGGGC-3'
	1BtKmR	5'-**CCACAACAGGGGGTCTATAGGACGCCATATCAATTCAGAAACCCGT CAGGAAATCAG**TCAGAAGAACTCGTCAAGAAGG-3'
	CsoR8101	5'-**GCAAATCCCTTGAACTGATCCTGTCT**-3'
	1Bf7641	5'-**TGACCGTAATGGTTCGTGGTGAGA**-3'

The letters "F" and "R" indicate forward and reverse primers, respectively. Underline remarks homologous region to *Kan*^{r} cassette, and bold letters indicate homologous region to *H. neapolitanus* genomic DNA.

[a.] This box contains two engineered stop codon.

[b.] This shading box contains engineered codons for tetracysteine (TC) motif: Cys-Cys-Pro-Gly-Cys-Cys.

A5. Oligonucleotides used for generating *Halothiobacillus neapolitanus* mutants (Cont'd)

Mutant	Primer name	Sequence
csoS1A-TCtag mutant	1A-TCtag-KmF	5'-**CGTGTCCACTCAGAAGTAGAAAACATTCTGCCTAAGGCGCCAC AAGCC**TGCTGTCCGGGTTGCTGTTAA[b]CCGCAATTGCCAGCTGGG-3'
csoS1A::G43A mutant	1A-KmF	5'-**CCACTCAGAAGTAGAAAACATTCTGCCTAAGGCGCCACAAGCCT AA**CCGCAATTGCCAGCTGGG-3'
	1A-KmR	5'-**GGTTTCGGCGACGGGTGTCGCCGAAGTGAGCCGTCTTAGGAAT ATCTGAC**TCAGAAGAACTCGTCAAGAAGGCG-3'
csoS4A, *csoS4B*, and *cso4AB* deletion mutant	oAKmF	5'-**GGCGCAGGATATCGGAAGCCCGATTGAAGAGGTTGCATCCGCA TGA**CCGGAATTG CCAGCTGGG-3'
	oAKmR	5'-**CCGGGAATCCTGCGTGTTGCGATTAGGTCGGAACGAACGCGCA TTACTTCCA**TCAGAAGAACTCGTCAAGAAGGCG-3'

A5. Oligonucleotides used for generating *Halothiobacillus neapolitanus* mutants (Cont'd)

Mutant	Primer name	Sequence
csoS4A, *csoS4B,* and *cso4AB* deletion mutant	oBKmF2	5'-**CTGATTTGACGATTATCGGGATTATTGATCAGTGGAATGGTGAG TAA**CCGGAATTGCCAGCTGGG-3'
	oBKmR	5'-**CTCTCCTCAATAAAGACTCATTCAAAATCAACTCATCTAGCGATG G**TCAGAAGAAC TCGTCAAGAAGGCG-3'

REFERENCES

1. Abe H, Abo T, and Aiba H. 1999 Regulation of Intrinsic Terminator by Translation in *Escherichia coli*: Transcription Termination at a Distance Downstream. *Genes Cells* **4**: 87-97.

2. Adams S R, Campbell R E, Gross L A, Martin B R, Walkup G K, Yao Y, Llopis J, and Tsien R Y. 2002 New Biarsenical Ligands and Tetracysteine Motifs for Protein Labeling in Vitro and in Vivo: Synthesis and Biological Applications. *J.Am.Chem.Soc.* **124**: 6063-6076.

3. 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.

4. 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**: 4133-4139.

5. 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**: 233-239.

6. Bechhofer D H. 1993 5' mRNA Stabilizers. *Control of Messenger RNA Stability.* 31-52.

7. Belasco J. 1993 mRNA degradation in Prokaryotic Cells: an Overview. *Control of*

Messenger RNA Stability. 3-11.

8. Beller H R, Chain P S G, Letain T E, Chakicherla A, Larimer F W, Richardson P M, Coleman M A, Wood A P, and Kelly D P. 2006 The Genome Sequence of the Obligately Chemolithoautotrophic Facultatively Anaerobic Bacterium *Thiobacilus denitrificans. J.Bacteriol.* **188 (4)**: 1473-1488.

9. Bobik T A. 2006 Polyhedral organelles compartmenting bacterial metabolic processes. *Appl.Environ.Microbiol* **70**: 517-525.

10. Cai F, Heinhorst S, Shively J M, and Cannon G C. 2008 Transcript analysis of the *Halothiobacillus neapolitanus cso* Operon. *Arch.Microbiol.* **189(2)**: 141-150.

11. Cai F, Menon B B, Shively J M, Curry K J, Cannon G C, and Heinhorst S. 2009 The Carboxysomal Shell Proteins CsoS4A and CsoS4B: A Case of Inconvenient Genetics.(running title). Manuscript in preparation.

12. 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. Current Microbiol.* **46**: 115-119.

13. 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.

14. Cannon G C and Shively J M. 1983 Characterization of a homogenous preparation of carboxysomes from *Thiobacillus neapolitanus. Arch.Microbiol.* **134**: 52-59.

15. Carafa Y, Brody E, and Thermes C. 1990 Prediction of Rho-Independent *Escherichia coli* Transcription Terminators: a Statistical Analysis of Their RNA

Stem-loop Structures. *J.Mol.Biol.* **216**: 835-858.

16. Caspar D L D and Klug A. 1962 Physical Principles in the Constrction of Regular Viruses. *Cold Spring Harbor Symp.Quant.Biol.* **27**: 1-24.

17. Chen X, Su Z, Dam P, Palenik B, Xu Y, and Jiang T. 2004 Operon prediction by comparative genomics: an application to the *Synechococcus* sp. WH8102 genome. *Nucleic Acids Res.* **32**: 2147-2157.

18. Cheng S, Liu Y, Crowley C S, Yeates T O, and Bobik T A. 2008 Bacterial Microcompartments: Their Properties and Paradoxes. *BioEssays* **30**: 1084-1095.

19. Deana A and Belasco J G. 2005 Lost in Translation: the Influence of Ribosomes on Bacterial mRNA Decay. *Genes Dev.* **19**: 2526-2533.

20. deLeseleuc L and Denis F. 2005 Restriction analysis of recombinant plasmids from colonies in less than 30 min. *Anal.Biochem.* **340**: 178-180.

21. 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.Bio.Chem* **283(16)**: 10377-10384.

22. 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.

23. 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**: 647-654.

24. Flower A M and McHenry C S. 1990 The γ Subunit of DNA Polymerase III

Holoenzyme of *Escherichia coli* Is Produced by Ribosomal Frameshifting. *PNAS* **87**: 3713-3717.

25. Gaietta G, Deerinck T J, Adams S R, Bouwer J, Tour O, Laird D W, Sosinsky G E, Tsien R Y, and Ellisman M H. 2002 Multicolor and Electron Microscopic Imaging of Connexin Trafficking. *Science* **296**: 503-507.

26. Garczarek L. 2001 Differential Expression of Antenna and Core Genes in *Prochlorococcus* PCC 9511 (Oxyphotobacteria) Grown under a Modulated Light-Dark Cycle. *Environ.Microbiol.* **3**: 168-175.

27. Garczarek L. 2001 Differential Expression of Antenna and Core Genes in *Prochlorococcus* PCC 9511 (Oxyphotobacteria) Grown under a Modulated Light-Dark Cycle. *Environ.Microbiol.* **3**: 168-175.

28. Gesteland R F and Atkins J F. 1996 Recoding: Dynamic Reprogramming of Translation. *Annu.Rev.Biochem.* **65**: 741-768.

29. Gibson J L and Tabita F R. 1996 The molecular regulation of the reductive pentose phosphate pathway in Proteobacteria and Cyanobacteria. *Arch.Microbiol.* **166**: 141-150.

30. Gruber T M and Gross C A. 2003 Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. *Annu.Rev.Microbiol.* **57**: 441-466.

31. Harley C B and Reynolds R P. 1987 Analysis of *E. coli* Promoter Sequences. *Nucleic Acids Res.* **15**: 2343-2361.

32. Heinhorst S, Cannon G C, and Shively J M. 2006 Carboxysomes and Carboxysome-Like Inclusions. *In: Shively JM, ed.Complex Intracellular Structures in Prokaryotes.Berlin/Heidelberg: Springer.* pp 141-164.

33. 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**: 8087-8094.

34. Hershberg R, Bejerano G, Santos-Zavaleta A, and Margalit H. 2001 PromEC: an Updated Database of *Escherichia coli* mRNA Promoters with Experimentally Identified Transcriptional Start Sites. *Nucleic Acids Res.* **29**: 277.

35. Holthuijzen Y A, VanDissel-Emiliani E F M, Kuenen J G, and Konings W N. 1987 Energetic aspects of CO_2 uptake in *Thiobacillus neapolitanus*. *Arch.Microbiol.* **147**: 285-290.

36. 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 WH8102 Carboxysomes as Revealed by Electron Cryotomography. *J.Mol.Biol.*

37. Jacob E, Sasikumar R, and Nair K N R. 2005 A Fuzzy Guided Genetic Algorithm for Operon Prediction. *Bioinformatics* **21**: 1403-1407.

38. Johnson J E. 1996 Functional Implications of Protein Protein Interactions in Icosahedral Viruses. *PNAS* **93**: 27-33.

39. Johnson J E and Fisher A J. 1994 Principles in Virus Structure. *Encyclopedia of Virology* : 1573-1586.

40. Johnson J E and Speir J A. 1997 Quasi-equivalent Viruses: A Paradigm for Protein Assemblies. *J.Mol.Biol.* **269**: 665-675.

41. Kapanidis A N, Margeat E, and Laurence T A. 2005 Retenion of Transcription Initiation Factor Sigma70 in Transcription Elongation: Single-Molecule Analysis. *Mol.Cell* **20(3)**: 347-356.

42. Kerfeld C A, Sawaya M R, Tanaka S, Nguyen C V, Phillips M, Beeby M, and Yeates T O. 2005 Protein Structures Forming the Shell of Primitive Bacterial Organelles. *Science* **309**: 936-938.

43. Klein M G, Zwart P, Bagby S C, Cai F, Chisholm S W, Heinhorst S, Cannon G C, and Kerfeld C A. 2009 Identification and Structural Analysis of a Novel Carboxysome Shell Protein with Implications for Metabolite Transport. *J.Mol.Biol.* **In Press**.

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

45. Liu X and Gorovsky M A. 1993 Mapping the 5' and 3' Ends of *Tetrahymena thermophila* mRNas Using RNA Ligase Mediated Amplification of cDNA Ends (RLM-RACE). *Nucleic Acids Res.* **21(21)**: 4954-4960.

46. Long B M, Badger M R, Whitney S M, and Price G D. 2007 Analysis of Carboxysomes from *Synechococcus* PCC7942 Reveals Multiple Rubisco Complexes with Carboxysomal Proteins CcmM & CcaA. *J.Bio.Chem.*

47. Maeda S, Badger M R, and Price G D. 2002 Novel gene products associated with NdhD3/D4-containing NDH-1 complexes are involved in photosynthetic CO_2 hydration in the cyanobacterium, *Synechococcus* sp. PCC7942. *Mol.Microbiol.* **43(2)**: 425-435.

48. Marin B, Nowack E C, Glockner G, and Melkonian M. 2007 The Ancestor of the *Paulinella* Chromatophore Obtained a Carboxysomal Operon by Horizontal Gene Transfer from a *Nitrococcus*-like γ-Proteobacterium. *BMC Evol.Biol.* **7**: 85.

49. Marlovits T C, Kubori T, Lara-Tejero M, Thomas D, Unger V M, and Galan J E. 2006 Assembly of the inner rod determines needle length in the type III secretion injectisome. *Nature* **441**: 637-640.

50. Mathy N, Benard L, Pellegrini O, Daou R, Wen T, and Condon C. 2007 5'-to-3' Exoribonuclease Activity in *Bacteria*: Role of RNase J1 in rRNA Maturation and 5' Stability of mRNA. *Cell* **129(4)**: 681-692.

51. McGinn P J, Price G D, Maleszka R, and Badger. 2003 Inorganic Carbon Limitation and Light Control the Expression of Transcripts related to the CO_2-concentrating mechanism in the Cyanobacterium *Synechocystis* sp. strain PCC 6803. *Plant Physiol.* **132**: 218-229.

52. 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.

53. Miller W G, Leveau J H J, and Lindow S E. 2000 Improved *gfp* and *inaZ* Broad-Host-Range Promoter-Probe Vectors. *MPMI* **13(11)**: 1243-1250.

54. Nudler E and Gottesman M E. 2002 Transcription termination and antitermination in *E. coli. Genes Cells* **7**: 755-768.

55. Omata T, Gohta S, Takahashi Y, Harano Y, and Maeda S-I. 2001 Involvement of a CbbR homolog in low CO_2-induced activation of the bicarbonate transporter operon in cyanobacteria. *J.Bacteriol.* **183**: 1891-1898.

56. Omata T, Price G D, Badger M R, Okamura M, Gohta S, and Ogawa T. 1999 Identification of an ATP-binding cassette transporter involved in bicarbonate uptake in the cyanobacterium *Synechococcus* sp. strain PCC 7942.

Proc.Natl.Acad.Sci.USA **96**: 13571-13576.

57. 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.

58. Peter K R. 1974 Characteriazation of a Phage-Like Particle from Cells of Nitrobacter. II. Structure and Size (Author's Transl). *Arch.Microbiol.* **97**: 129-140.

59. Prestridge D. 1991 SIGNAL SCAN: a computer program that scans DNA sequences for eukaryotic transcriptional elements. *CABIOS* **7**: 203-206.

60. Price G D and Badger M R. 1989 Isolation and characterization of high CO_2 -requiring-mutants of the cyanobacterium *Synechococcus* PCC 7942: two phenotypes that accumulate inorganic carbon but are apparently unable to generate CO_2 within the carboxysome. *Plant Physiol.* **91**: 514-525.

61. 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, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. *J.Exp.Bot.* **59(7)**: 1441-1461.

62. 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**: 2871-2879.

63. Price G D, Maeda S, Omata T, and Badger M R. 2002 Modes of Active Inorganic Carbon Uptake in the Cyanobacterium. *Funct.Plant Biol.* **29**: 131-149.

64. Price M N, Huang K H, Alm E J, and Arkin A P. 2005 A Novel Method for Accurate Operon Predictions in all Sequenced Prokaryotes. *Nucleic Acids Res.* **33**:

880-892.

65. Prommeenate P, Lennon A M, Markert C, Hippler M, and Nixon P J. 2004 Subunit Composition of NDH-1 Complexes of *Synechocystis* sp. PCC 6803: Identification of Two New *ndh* Gene Products with Nuclear-Encoded Homologues in the Chloroplast Ndh Complex. *J.Bio.Chem* **279**: 28165-28173.

66. Ramamoorthy R, McClain N A, Gautam A, and Scholl-Meeker D. 2005 Expression of the *bmpB* Gene of *Borrelia burgdorferi* is Modulated by Two Distinct Transcription Termination Events. *J.Bacteriol.* **187**: 2592-2600.

67. Salgado H, Moreno-Hagelsieb G, Smith T F, and Collado-Vides J. 2000 Operons in *Escherichia coli*: Genomic Analyses and Predictions. *PNAS* **97(12)**: 6652-6657.

68. 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.

69. Sawaya M R, Cannon G C, Heinhorst S, Tanaka S, Williams E B, Yeates T O, and Kerfeld C A. 2006 The Structure of β-Carbonic Anhydrase from the Carboxysomal Shell Reveals a Distinct Subclass with One Active Site for the Price of Two. *J.Biol.Chem.* **281 (11)**: 7546-7555.

70. 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**: 526-535.

71. Scott K M, Sievert S M, Abril F N, Ball L A, Barrett C J, Blake R A, Boller A J, Chain P S G, Clark J A, Davis C R, Detter C, Do K F, Dobrinski K P, Faza B I, Fitzpatrick K A, Freyermuth S K, Harmer T L, Hauser L J, Hugler M, Kerfeld C A,

Klotz M G, Kong W W, Land M, Lapidus A, Larimer F W, Longo D L, Lucas S, Malfatti S A, Massey S E, Martin D D, McCuddin Z, Meyer F, Moore J, Ocampo L H, Paul J H, Paulsen T, Reep D K, Ren Q, Ross R L, Sato P Y, Thomas P, Tinkham L E, and Zeruth G T. 2006 The Genome of Deep-Sea Vent Chemolithoautotroph *Thiomicrospira crunogena* XCL-2. *PLoS Biology* **4 (12)**: e383.

72. Sharp J S, Nelson S, Brown D, and Tomer K B. 2006 Structural Characterization of the E2 Glycoprotein from Sindbis by Lysine Biotinylation and LC-MS/MS. *Virol.* **348**: 216-223.

73. Shively J M, Bradburne C E, Aldrich H C, Bobik T A, Mehlman J L, Jin S, and Baker S H. 1998 Sequence homologs of the carboxysomal polypeptide CsoS1 of the thiobacilli are present in cyanobacteria and enteric bacteria that form carboxysomes-polyhedral bodies. *Can.J.Bot.* **76**: 906-916.

74. So A K C and Espie G S. 2005 Cyanobacterial carbonic anhydrases. *Can.J.Bot.* **83**: 721-734.

75. So A K C, Espie G S, Williams E B, Shively J M, Heinhorst S, and Cannon G C. 2004 A Novel Evolutionary Lineage of Carbonic Anhydrase (ε Class) Is a Component of the Carboxysome Shell. *J.Bacteriol.* **186(3)**: 623-630.

76. Tabita F R. 1999 Microbial rebulose 1,5-bisphosphate carboxylase/oxygenase: A different perspective. *Photosynth.Res.* **60**: 1-28.

77. 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**: 1083-1086.

78. Tanaka S, Sawaya M R, Phillips M, and Yeates T O. 2009 Insights from Multiple Structures of the Shell Proteins from the β-Carboxysome. *Protein Sci.* **18**: 108-120.

79. Toyoda K, Yoshizawa Y, Arai H, Ishii M, and Igarashi Y. 2006 The role of two CbbRs in the transcriptional regulation of three ribulose-1,5-bisphosphate carboxylase/oxygenase genes in *Hydrogenovibrio marinus* strain MH-110. *Microbiology* **151**: 3615-3625.

80. 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 Biology* **5 (6)**: e144.

81. Tu C, Shrager J, Burnap R L, Postier B L, and Grossman A R. 2004 Consequences of a Deletion in *dspA* on Transcript Accumulation in *Synechocystis* sp. Strain PCC6803. *J.Bacteriol.* **186(12)**: 3889-3902.

82. Vogel J, Axmann I M, Herzel H, and Hess W R. 2003 Experimental and computational analysis of transcriptional start sites in the cyanobacterium *Prochlorococcus* MED4. *Nucleic Acids Research* **31(11)**: 2890-2899.

83. Wang H L, Postier B L, and Burnap R L. 2004 Alterations in Global Patterns of Gene Expression in *Synechocystis* sp. PCC6803 in Response to Inorganic Carbon Limitation and the Inactivation of *ndhR*, a LysR Family Regulator. *J.Bio.Chem.* **279(7)**: 5739-5751.

84. Watson G M and Tabita F R. 1996 Regulation, Unique Gene Organization, and Unusual Primary Structure of Carbon Fixation Genes from a Marine Phycoerythrin-Containing Cyanobacterium. *Plant Mol.Biol.* **32**: 1103-1115.

85. Wunderlich Z, Acton T B, Liu J, Kornhaber G, Everett J, Carter P, Lan N, Echols N, Gerstein M, Rost B, and Montelione G T. 2004 The Protein Target List of the Northeast Structural Genomics Consortium. *Proteins.* **56(2)**: 181-187.

86. Yeates T O, Kerfeld C A, Heinhorst S, Cannon G C, and Shively J M. 2008 Protein-based Organelles in Bacteria: Carboxysomes and Related Microcompartments. *Nat.Rev.Microbiol.* **6(9)**: 681-691.

87. Yoshizawa Y, Toyoda K, Arai H, Ishii M, and Igarashi Y. 2004 CO_2-Responsive Expression and Gene Organization of Three Ribulose-1,5-bisphosphate Carboxylase/Oxygenase Enzymes and Carboxysomes in *Hydrogenovibrio marinus* Strain MH-110. *J.Bacteriol.* **186(17)**: 5685-5691.

88. Zhang P, Battchikova N, Jansen T, Appel J, Ogawa T, and Aro E M. 2004 Expression and Functional Roles of the Two Distinct NDH-1 Complexes and the Carbon Acquisition Complex NdhD3/NdhF3/CupA/Sll1735 in *Synechocystis* sp. PCC 6803. *Plant Cell* **16**: 3326-3340.

89. Zhang Z and Dietrich F S. 2005 Mapping of transcription start sites in *Saccharomyces cerevisiae* using 5' SAGE. *Nucleic Acids Res.* **33(9)**: 2838-2851.

90. Zheng Y, Szustakowski J D, Fortnow L, Roberts R J, and Kasif S. 2002 Computational Identification of Operons in Microbial Genomes. *Genome Res.* **12**: 1221-1230.