Synthetic engineering of biological carbon-fixing organelle shells

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by Qiuyao Jiang
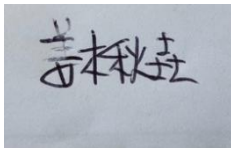
March 2022

Department of Biochemistry and Systems Biology, Institute of

Systems, Molecular and Integrative Biology

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Liverpool or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Liverpool or any other University or similar institution except as declared in the Preface and specified in the text. This thesis does not exceed the prescribed word limit of 100,000 words.

Qiuyao Jiang

Publications and Author's Contributions

The content of Chapter 4 is based on a manuscripted article entitled: Structural basis for the CsoS2-mediated α-carboxysome shell assembly.

The work described in Chapter 3 and Chapter 5 is unpublished.

I acknowledge the following collaborators for their contributions to the outcome described in the following chapters. Unless specified below, all works were completed by the author.

Name and organization	Contribution	Related Chapter
Prof Luning Liu University of Liverpool, UK	Guidance of research, revision of the manuscript	All Chapters
Prof Andrew I. Cooper University of Liverpool, UK	Guidance of hydrogen research	Chapter 3
Dr Reiner Sebastian Sprick University of Liverpool, UK	Assistance of hydrogen testing	Chapter 3
Prof Peijun Zhang University of Oxford, UK	Assistance of bioinformatic analysis	Chapter 4
Dr Tao Ni University of Oxford, UK	Assistance of bioinformatic analysis	Chapter 4
Dr Taiyu Chen	Assistance of Rubisco activity testing	Chapter 5
Gregory Dykes University of Liverpool, UK	Assistance of TEM imaging	Chapter 3, 4, and 5

Other publications produced during PhD studies but not included in my thesis:

Incorporation of Functional Rubisco Activases into Engineered Carboxysomes to Enhance Carbon Fixation. Taiyu Chen, Yi Fang, Qiuyao Jiang, Gregory F. Dykes, Yongjun Lin, G. Dean Price, Benedict M. Long, and Lu-Ning Liu. *ACS synthetic biology*. 2022; 11: 154-161.

I performed protein purification.

Reprograming prokaryotic organelles as a new nanoreactor for hydrogen production. Tianpei Li, Qiuyao Jiang, Jiafeng Huang, Catherine M. Aitchison, Fang Huang, Mengru Yang, Gregory F. Dykes, Hai-Lun He, Qiang Wang, Reiner Sebastian Sprick, Andrew I. Cooper, and Lu-Ning Liu. *Nature Communications*. 2020; 11, 5448.

I performed hydrogen-evolution activity test

Engineering and modulating functional cyanobacterial CO_2-fixing organelles. Fang, Yi, Fang Huang, Matthew Faulkner, Qiuyao Jiang, Gregory F. Dykes, Mengru Yang, and Lu-Ning Liu. *Frontiers in Plant Science*, 2018; 9: 739.

I performed the protein purification

Acknowledgement

Abstract

Bacterial microcompartments (BMCs) are "proteinaceous organelles" that play a crucial role in the improvement and regulation of metabolic processes in a variety of prokaryotic cells. BMCs are composed of a thin layer protein shell that encapsulates metabolic enzymes for the related metabolic pathway. It is known that BMCs are necessary in autotrophic CO_2 fixation and catabolic processes and promote bacterial fitness in specific environmental niches. Carboxysomes are a classical example of BMCs that can be found in all identified cyanobacteria as well as some chemoautotrophs, such as Halothiobacillus neapolitanus, that are involved in the fixation of CO_2 in the environment. The semi-permeable carboxysome shell plays a role in encapsulating ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) and accumulating CO2 and other substrate molecules to substantially enhance RuBisCO carboxylation. Despite the importance of carboxysome shells, we know very little about their structure, protein organisation, self-assembly principle and permeability.

In my PhD studies, the reprogramming of the α-carboxysomes shells (~ 100 nm in diameter) was performed by encapsulating the [NiFe]-hydrogenase from *E. coli*, via the C-terminus of CsoS2 as an encapsulation peptide, to produce a new nanobioreactor for hydrogen production. More than 10-fold increase in the yield of hydrogen was obtained by the hydrogenase-shell biocatalyst, highlighting the great potential of α-carboxysomes shells as nanoscale "factories" in biotechnology applications and bioenergy production. Moreover, the self-assembly of shell hexametric and pentameric proteins as well as the scaffolding protein CsoS2 resulted in the production of α-carboxysome mini-shell structures with varying icosahedral symmetry and diameters ranging from 23 to 40 nm. We also showed that the mini-shells could encapsulate foreign cargo proteins within the shell architecture, representing a new protein caging system with controllable features. Furthermore, the engineering of entire α-carboxysomes shells and mini-shell structures has sparked interest in the development of simplified α-carboxysomes. We showed that Rubisco, CsoS2, and main shell hexamers can form simplified α-carboxysomes, with a diameter of ~ 100 nm and carbon-fixation activities.

Our study provides new insights into the self-assembly, modularity, and permeability of carboxysome shells. The bioengineering strategies and synthetic carboxysome structures developed in this study will empower our synthetic biology approaches to repurpose carboxysomes for new functions.

Table of contents

Lists of Figures

Lists of Tables

Alphabetical List of Abbreviations and Acronyms

BMCs	Bacterial microcompartments
RuBisCO	1, 5-ribulose bisphosphate carboxylase/oxygenase
CA	carbonic anhydrase
GRMs	Glycyl radical enzyme-associated microcompartments
Cut BMCs	Choline utilizing bacterial microcompartments
TMA	Glycyl radical choline-trimethylamine lyase
Grp	Glycyl-radical propanediol
RMM	*Rhodococcus* and *Mycobacterium* bacterial microcompartment
AlcDH	Alcohol dehydrogenase
AldDH	Aldehyde dehydrogenase
PTAC	Phosphotrasnacylase
TEM	Transmission electron microscopy
MCPdb	Bacterial microcompartment database
BMC-H	Bacterial microcompartment hexametric protein
BMC-P	Bacterial microcompartment pentameric protein
BMC-T	Bacterial microcompartment trimeric protein
BMC-H$_c$	Bacterial microcompartment canonical BMC-H
BMC-H$_p$	Bacterial microcompartment permuted BMC-H
BMC-H$_{Fe}$	Bacterial microcompartment BMC-H with an [Fe-S] cluster
BMC-H$_{EX}$	Unknown function protein of BMC-H
EPs	Encapsulation peptides
Raf 1	Rubisco accumulation factor 1
CBs	Carboxysomes
α-CBs	α-Carboxysomes
β-CBs	β-Carboxysomes
RuBP	Ribulose-1,5-bisphosphate
CBB	Calvin-Benson-Bassham
3-PGA	3-phosphoglyceric acid
2-PG	2-phosphoglycolate
SSUL	Rubisco small subunit-like modules
γCAL	γ-type carbonic anhydrase-like domain

IPTG	Isopropyl β-D-thiogalactopyranoside
DO	Dissolved oxygen
DLS	Dynamic light scattering
CTF	Contrast transfer function
MV	Methyl viologen
SD	Standard deviation

Chapter 1 Introduction

1.1 Bacterial microcmpartments

1.1.1 Overview

Bacterial microcompartments (BMCs) are special "proteinaceous organelles", which were first discovered in cyanobacteria by electron microscopy in 1956 (Drews and Niklowitz, 1956a). They feature a thin protein shell that sequester metabolic enzymes, facilitating a series of metabolic processes (Kerfeld *et al.*, 2010; Rae *et al.*, 2013; Axen *et al.,* 2014; Chowdhury *et al.*, 2014; Kerfeld and Erbilgin, 2015; Zarzycki *et al.*, 2015; Kerfeld and Melnicki, 2016; Turmo *et al.*, 2017b; Liu, 2021a). The protein shells of BMCs are thought to function as a selectively "barrier", playing roles in regulating the passage of substrates and products into and out of BMCs (Jakobson *et al.*, 2017b). BMCs were found in 45 established bacterial phyla and can be divided into 68 classes or subclasses that are functionally variable and participate in CO_2 fixation and catabolic processes (Abdul-Rahman *et al.*, 2013; Zarzycki *et al.*, 2015; Sutter *et al.*, 2021).

One type of BMC is the carboxysomes (Figure 1-1a, 1b, and 1c), which can be found in all identified cyanobacteria and some chemoautotrophs, as the machinery for CO_2 fixation (Rae *et al.*, 2013; Kerfeld and Melnicki, 2016; Turmo *et al.*, 2017b; Liu, 2021a). Carboxysomes are composed of the enzymes 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) and carbonic anhydrase (CA) (Cannon *et al.*, 2010) and function in improving CO_2 fixation in the Calvin-Benson-Bassham cycle (Figure 1-2b) (Rae *et al.*, 2013; Long *et al.*, 2016b). Another family of BMCs is catabolic BMCs, also known as metabolosomes (Figure 1-1d), which are present in a wide range of bacteria and archaea (Chowdhury *et al.*, 2014; Kerfeld *et al.*, 2018; Liu, 2021a; Liu *et al.*, 2021b). The function of metabolosomes is metabolizing the substrates such as propanediol (1,2-propanediol utilization BMC, Pdu BMC), ethanolamine (ethanolamine utilization BMC, Eut BMC), fucose, and rhamnose to promote bacteria growth

(Figure 1-2c) (Bobik *et al.*, 1999; Kofoid *et al.*, 1999; Petit *et al.*, 2013; Erbilgin *et al.*, 2014).

Metabolosomes are often present in the human gut bacteria and are thought to be involved in some diseases (Chowdhury *et al.*, 2014, Herring *et al.*, 2018).
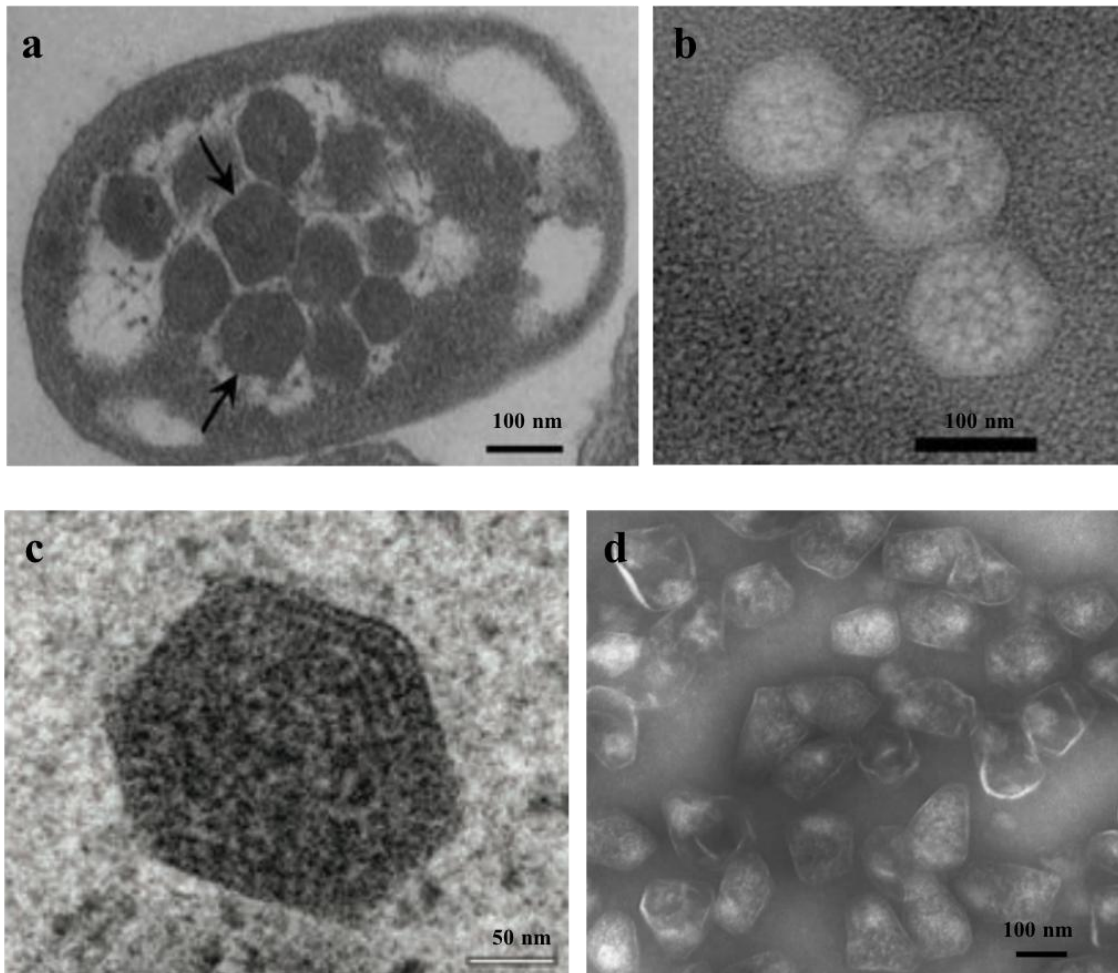
Figure 1-1. Different types of BMCs. (a) Transmission electron microscopic image of α-carboxysomes from thin-sectioned *Halothiobacillus neapolitanus* (*H. neapolitanus*). Scale bar is 100 nm. This image was taken from Menon *et al.* (2008). (b) The purified α-carboxysomes of *H. neapolitanus*. Scale bar is 100 nm. This image was taken from Menon *et al.* (2008). (c) The β-carboxysomes of *Synechocystis* PCC 6803 (Syn6803). Scale bar is 50 nm. This image was taken from Tanaka *et al.* (2008b). (d) The purified 1,2-propanediol utilization (Pdu) microcompartment of *Salmonella enterica* serovar Typhimurium LT2. Scale bar is 100 nm. This image was taken from Havemann and Bobik (2003)

Bioinformatic analysis revealed that the genes encoding proteins for the carboxysomes and metabolosomes are organized in consolidated BMC locus (Figure 1-2a) (Sutter *et al.*, 2021), including a series of genes responsible for shell proteins, metabolic pathway enzymes, and

auxiliary proteins. The genes encoded glycyl radical enzymes (GREs), as the metabolic enzymes, were found in the glycyl radical enzyme-associated microcompartments (GRMs) (Jorda *et al.*, 2013; Craciun and Balskus, 2012; Zarzycki *et al.*, 2015). GREs employ the radicals of glycine and cysteine, together with activating enzyme that are relate with GREs for the instalment of glycyl radical (Selmer *et al.*, 2005). Characterized GREs composed of choline trimethylamine (TMA)-lyase that function in producing TMA and acetaldehyde, and 1,2-propanediol dehydratase that function in producing propionaldehyde (Jorda *et al.*, 2013; Petit *et al.*, 2013; Axen *et al.*, 2014; Kalnins *et al.*, 2015; Martinez-del Campo *et al.*, 2015; Zarzycki *et al.*, 2017). The GRMs are further classified based on different substrates being catalyzed. One type of GRM is the choine utilizing bacterial microcompartments (Cut BMCs) (Figure 1-2d), which was found in *Desulfovibrio desulfuricans* firstly and function in catalyzing choline to TMA and acetaldehyde via choline TMA-lyase (Selmer *et al.*, 2005). The production of TMA in the gut as a result of the presence of Cut BMCs is thought to be the cause of a rise in trimethylamine N-oxide of serum, which has been show to correlate with human heart and liver diseases (Bae *et al.*, 2014; Trøseid *et al.*, 2015; Chen *et al.*, 2016).

Glycyl-radical propanediol (Grp) BMC is another type of GRM that has been discovered in *Rhodobacter capsulatus* (Jorda *et al.*, 2013; Zarzycki *et al.*, 2017; Schindel *et al.*, 2019; Lundin *et al.*, 2020) (Figure 1-2d). This type is responsible for converting 1,2-propanediol to propionaldehyde by 1,2-propanediol dehydratase. Recent studies also reported that a GRM (GRM5) similar to Grp BMCs found in *Clostridum phytofermentans*, which has the ability to convert L-fuculose-phosphate to lactaldehyde via fuculose phosphate aldolase, and subsequently lactaldehyde can be converted to 1,2-propanediol via lactaldehyde reductase. Apart from GRMs, *Rhodococcus* and *Mycobacterium* BMC (RMM) (Figure 1-2d) is a novel

type of BMC that use S-1-amino-2-propanol kinases to metabolize aminoacetone, which has

just been discovered (Mallette and Kimber, 2018; Sutter *et al.*, 2021).
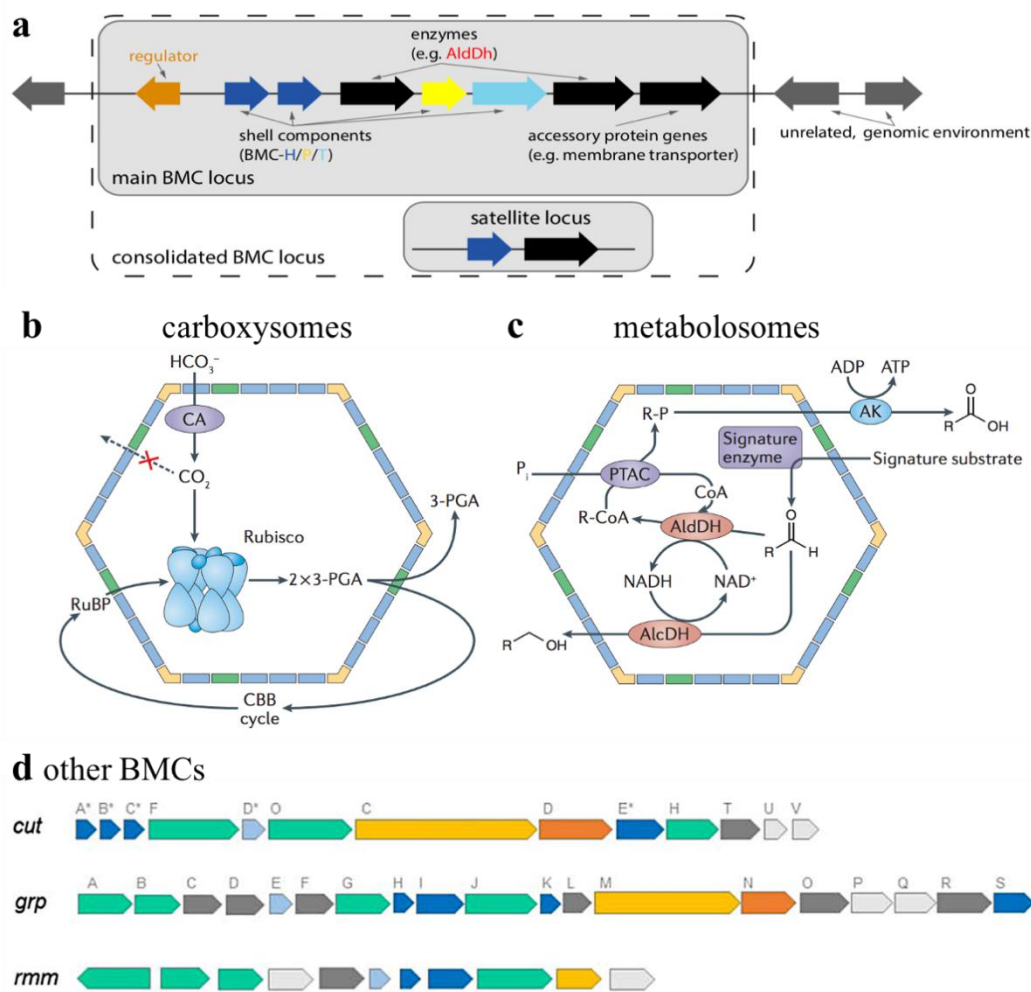
Figure 1-2. Schematic diagram of the composition, metabolic pathways of BMCs and the genes arrangement of new discovered BMCs. (a) The common BMC locus containing the genes that encoding the shell proteins (BMC-H in blue, BMC-P in yellow, BMC-T in cyan), metabolic enzymes (e.g. alcohol dehydrogenase (AlcDH) in black), and accessory proteins (e.g. membrane transporter in black). This diagram was taken from Sutter *et al.* (2021). (b) The metabolic pathway of carboxysomes was taken from Kerfeld *et al.* (2018). The carbonic anhydrase and Rubisco are encapsulated inside carboxysomes to fix CO_2 as a part of the Calvin-Benson-Bassham cycle. The shell of carboxysomes is responsible for CO_2 accumulation. (c) The metabolic pathway of metabolosomes was drawn from taken from Kerfeld *et al.* (2018). The internal enzymes contain signature enzyme, alcohol dehydrogenase (AlcDH), aldehyde dehydrogenase (AldDH), and phosphotransacylase (PTAC). The aldehyde is produced via a signature enzyme in the shell. The shell prevents the loss of aldehyde to the cytoplasm for the protection of the cell. Shell hexamer protein labeled in cyan, pentamer protein in yellow, and trimer protein in green were showed in model of (b) and (c). (d) The operon layouts of the new types of BMCs. Cut BMC (*cut*) operon encoded hexamer shell proteins CutA[*]/B[*]/C[*]/E[*] (dark blue), pentamer shell protein CutD[*] (cyan), metabolic enzyme and its activating enzyme CutC (yellow) and CutD (orange), accessory enzyme CutF/H/O (light green). Grp BMC

(*grp*) operon encoded hexamer shell protein GrpH/I/K/S (blue), pentamer shell protein GrpE (cyan), metabolic enzyme and its activating enzyme GrpM (yellow) and GrpN (orange).This diagram was taken from Stewart *et al.*(2021).

1.1.2 Bacterial microcompartment shell

Transmission electron microscopy (TEM) and X-ray crystallography have provided detailed structural information about the building blocks of highly conserved BMC shells (Kerfeld, 2005; Lassila *et al.*, 2014; Sutter *et al.*, 2017). In addition this, the bacterial microcompartment database (MCPdb) compiled in-depth information on 163 distinct microcompartment proteins and related microcompartments structures (Ochoa *et al.*, 2021); a few synthetic BMCs shells have been generated, for example, shell protien PduA of Pdu BMCs from *Salmonella typhimurium* could assemble a 13 nm in diameter nano-synthetic BMCs shell with 180 nm^3 of inner volume (Jorda *et al.*, 2016); a 6.5 MDa BMCs shell with 40 nm in diameter was generated, which contains one hexamer protein, one pentamer protein and three trimer proteins of *Haliangium orchraceum* (HO) BMCs (Sutter *et al.*, 2017); moreover, the repetitions of hexamer protein from HO BMCs connected via the short linker could assemble a smaller synthetic shell with a diameter of 25 nm (Sutter *et al.*, 2019); the synthetic β-carboxysome shells contain CcmK1/2 and CcmO of *Halothece* sp. PCC 7418 β-carboxysome was constructed with T=3 symmetry with 210 Å in diameter (Sutter *et al.*, 2019a); in 2021, the minimal α-carboxysome shell was produced based on the assembly of shell protien CsoS1A and CsoS4A of *H. neapolitanus* α-carboxysomes (Tan *et al.*, 2021a);

All bacterial microcompartment shells include three kinds of proteins: hexametric proteins (BMC-H), pentameric proteins (BMC-P), and trimeric proteins (BMC-T) (Figure 1-3). BMC-H proteins are the most abundant proteins in the shell, and they include a Pfam00936 domain

(~90 amino acids) that assemblies into a homohexamer with a convex and a concave face (Kerfeld, 2005). BMC-H can be divided into three subgroups (Figure 1-3). the first is the canonical BMC-H protein (BMC-H$_C$), which has a charges pore and is present in most BMCs (Kerfeld and Erbilgin, 2015). The second is the permuted BMC-H protein (BMC-H$_P$), which is composed of canonical BMC-H protein with the permuted domain (Crowley *et al.*, 2008). Structural analysis revealed that BMC-H$_P$ has an N-terminal extension that forms a beta barrel structure and can cover the pore, assumed to prevent metabolites transport via forming a beta barrel structure (Crowley *et al.*, 2008; Pitts *et al.*, 2012). The third subgroup is composed of a canonical BMC-H protein and an [Fe-S] cluster (BMC-H$_{Fe}$) that transports electrons and metal clusters into and out the shell (Cai *et al.*, 2013). Finally, another type of BMC-H protein is an unknown function protein termed BMC-H$_{EX}$, only known it has a C-terminal extension (Kerfeld and Erbilgin, 2015).
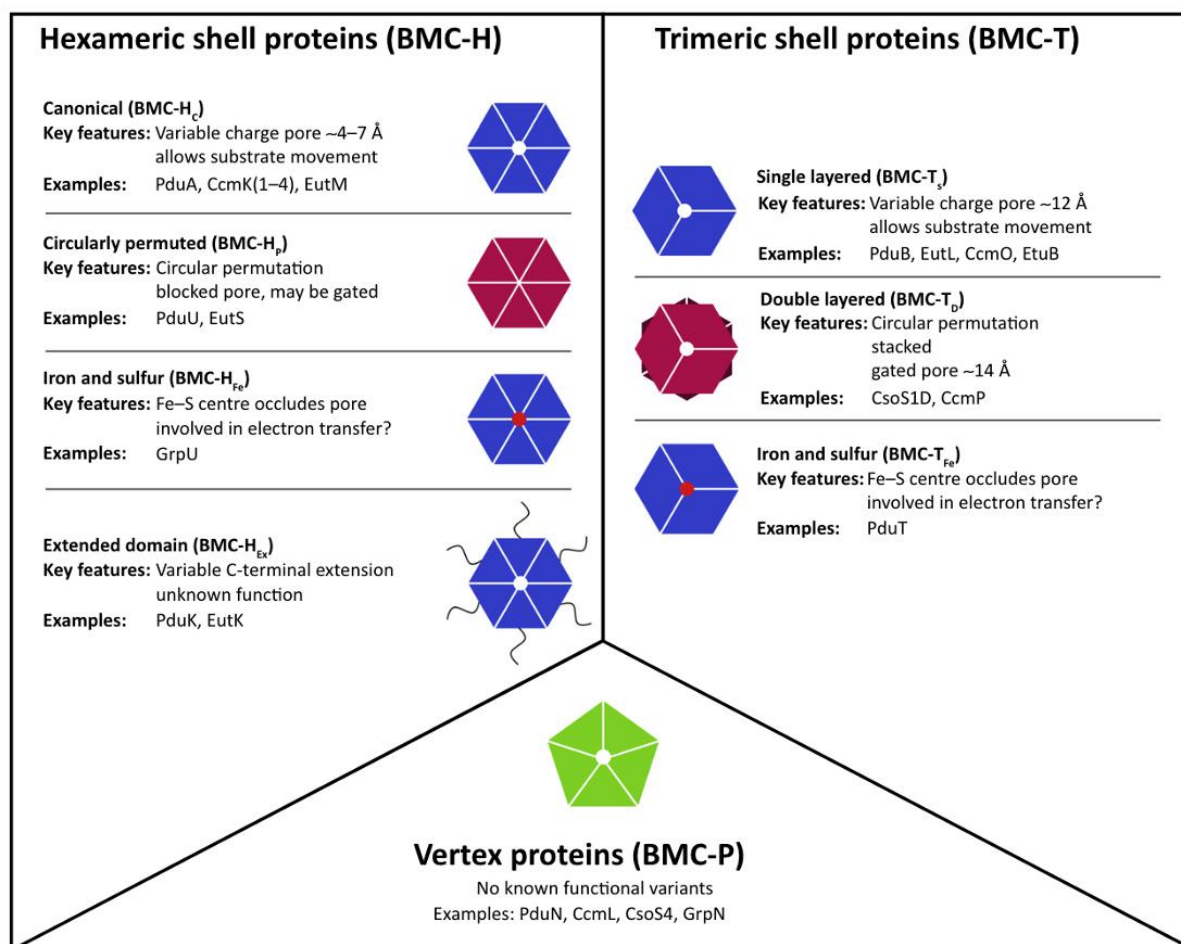
Figure 1-3. The bacterial microcompartment shell protein classes. This diagram was taken from Lee *et al.* (2019). BMCs shell proteins have three types including hexameric shell proteins (BMC-H) that contain a single Pfam00936 domain, trimeric shell proteins (BMC-T) that contain two fused Pfam00936 domains, and vertex proteins (BMC-P) that contain a single Pfam03319 domain. BMC-H and BMC-T serve as the facets of the BMCs shell (Yeates *et al.*, 2013; Sutter *et al.*, 2016); BMC-P was required as the vertices to cap the BMCs shell (Tanaka *et al.*, 2008b; Sutter *et al.*, 2013b; Wheatley *et al.*, 2013).

The BMC-T protein comprises two BMC-H domains that assemble into pseudohexamers (Klein *et al.*, 2009; Pang *et al.*, 2011; Pang *et al.*, 2012; Cai *et al.*, 2013) (Figure 1-3). There are three subtypes of BMC-T proteins (Figure 1-3). One is BMC-T_S, which has two BMC-H domains that form a single layer trimer. BMC-T_S has the central pores that are bigger than those found in BMC-H_C, presumably serving as gated portals for the passage of large metabolites (Klein *et al.*, 2009; Cai *et al.*, 2013; Larsson *et al.*, 2017; Mallette and Kimber, 2017). The second subclass is BMC-T_D, which contains two BMC-T_S that combine to produce a double layer trimer that also comprises the larger pores. The third subtype termed

BMC-T_{Fe}, containing the [Fe-S] center similar with BMC-H_{Fe}, is suggested to participate in electron transfer and [Fe-S] cluster accumulation within the BMCs (Parsons *et al.*, 2008; Pang *et al.*, 2011).

BMC-P proteins assemble into pentamers and contain a Pfam03319 domain (Figure 1-3) (Tanaka *et al.*, 2008a; Sutter *et al.*, 2013a; Mallette and Kimber, 2017). BMC-P proteins are present at the vertices of BMCs and function in capping and closing BMCs to separate metabolites (Cai *et al.*, 2009).

1.1.3 Bacterial microcompartments encapsulation and formation

It is now known that encapsulation peptides (EPs) are critical for targeting proteins into BMCs. Earlier studies found that multiple protiens within the β-carboxysomes and metabolosomes contain EPs at the N-terminus or C-terminus (Fan *et al.*, 2012; Kinney *et al.*, 2012; Aussignargues *et al.*, 2015). These EPs are generally short sequences about 15-20 amino acid and form an amphipathic helix to bond to the shell protein (Kinney *et al.*, 2012; Lawrence *et al.*, 2014; Jakobson *et al.*, 2015). EPs have been the subject of structural investigations, which show that they have a dual function in the nucleation of enzyme cores and protein encapsulation by interacting with the shell proteins (Erbilgin *et al.*, 2016; Juodeikis *et al.*, 2020). As shown in Figure 1-4a, EPs aggregates the enzymes firstly, and then interact with the shell proteins to form complete functional BMCs (Fan *et al.*, 2010; Fan and Bobik, 2011; Lawrence *et al.*, 2014; Aussignargues *et al.*, 2015, Erbilgin *et al.*, 2016; Wang *et al.*, 2019b). However, not all proteins are able to enter BMCs via using EPs. Figure 1-4b illustrates another pathway in which a disordered protein interacts with both enzymatic cores and shell proteins to form a complete BMC. This process is dependent on protein-protein

interactions between shell proteins and enzymatic cores (Cai *et al.*, 2015a, Chaijarasphong *et al.*, 2016, Liu *et al.*, 2018b).

1.1.4 The functional diversity of BMCs

Recently, bioinformatic studies confirmed that the main locus of BMCs contain shell genes and core enzymatic genes (Figure 1-2a) (Sutter *et al.*, 2021). Apart from these main genes, there are also genes that are involved in protein regulation such as RuBisCO accumulation factor 1 (Raf1) of β-carboxysomes from *Synechococcus elongatus* PCC 7942 (Syn7942), which enhances Rubisco assembly and β-carboxysomes biogenesis (Huang *et al.*, 2020). There are also factors that are involved in maintaining the metabolic activity and communication between BMCs and the cells, such as RuBisCO activases (CbbO and CbbQ) with the α-carboxysomes (Chen *et al*., 2022; Tsai *et al.*, 2022). Moreover, identification of the new class of BMC (GRMs) by loci composition analysis and experimentally characterizing has broadened the research of BMCs to human diseases (Bae *et al.*, 2014; Trøseid *et al.*, 2015; Chen *et al.*, 2016).
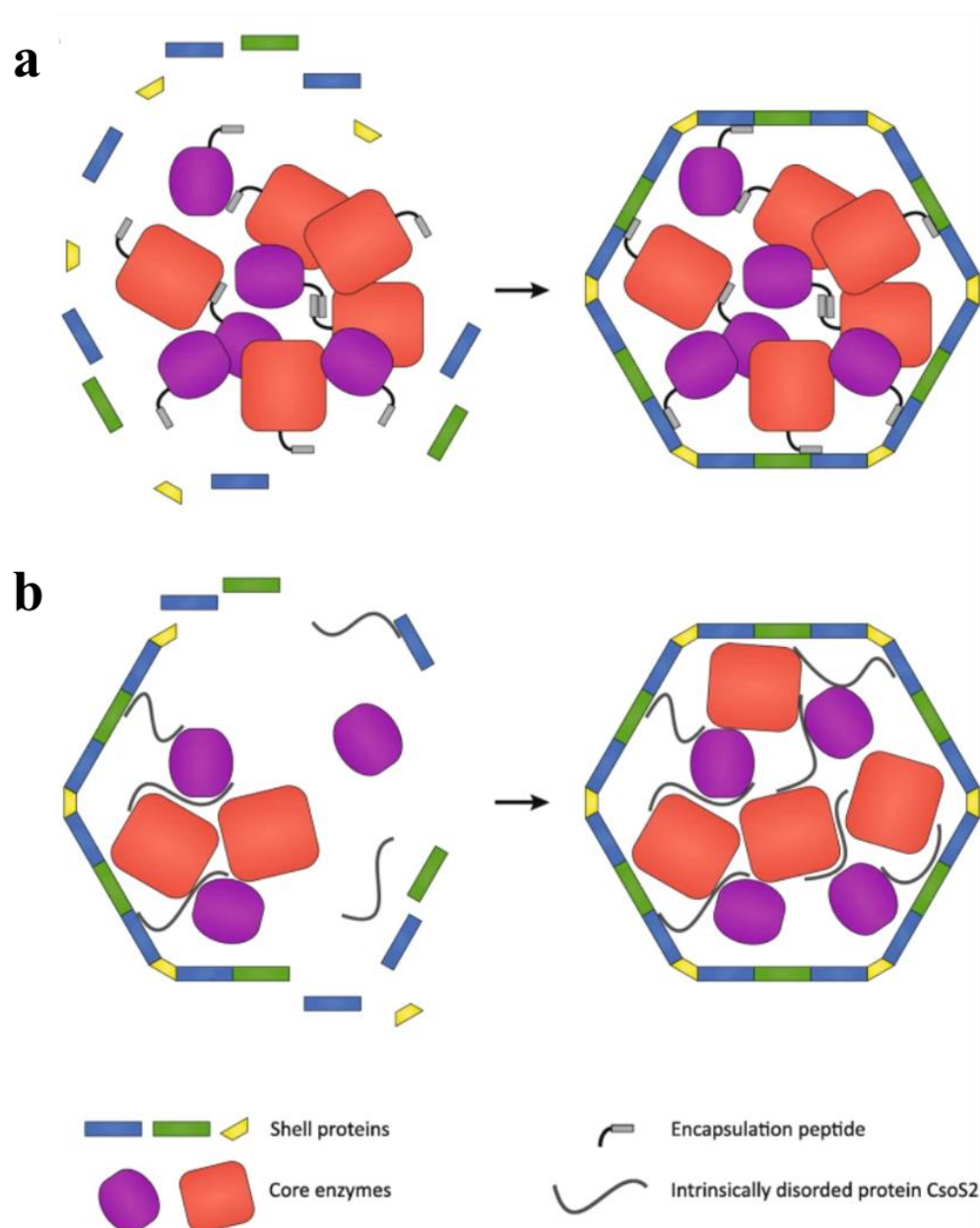
Figure 1-4. Schematics of bacterial microcompartments formations. This schematic diagram was taken from Kirst and Kerfeld (2019). (a) The assembly pathway of β-carboxysomes and metabolosomes. Using the encapsulating peptide in conjunction with other proteins, the enzymatic core was aggregated. Following aggregation, the encapsulation peptide linked to shell proteins, resulting in the functional BMCs formation. (b) The assembly pathway of α-carboxysomes. The protein CsoS2 makes enzymatic core aggregation and shell proteins recruitment to occur at the same time.

1.2 Carboxysomes (CBs)

1.2.1 Various types of CBs

The carboxysomes were initially considered as inclusion bodies by electron microscopy observations in early days (Drews and Niklowitz, 1956b). The following experiments confirmed that these inclusion bodies included the CO_2-fixation enzyme RuBisCO (Shively *et al.*, 1973). To date, there are at least two classes of carboxysomes based on the phyletic distributions (Castresana, 2000; Ronquist and Huelsenbeck, 2003; Wu and Eisen, 2008; Gouy *et al.*, 2010) and protein composition of the carboxysomes (Badger *et al.*, 2002; Cannon *et al.*, 2002). One class is the α-carboxysomes (α-CBs), which include form 1A RuBisCO and are encoded by a *cso* operon (Figure 1-5). Another class is the β-carboxysomes (β-CBs), which contain form 1B RuBisCO and are encoded by predominantly a *ccm* operon (Figure 1-5). Despite the variations in RuBisCO types, α- and β-carboxysomes showed similar RuBisCO content under lower CO_2 environment (Whitehead *et al.*, 2014). The α-carboxysomes were discovered in α-cyanobacteria and chemoautotrophs such as *Synechococcus* WH 8102 (Syn8102) and *H. neapolitanus* (Figure 1-5). Syn7942 and Syn6803 are two model species that contain β-carboxysomes (Figure 1-5).
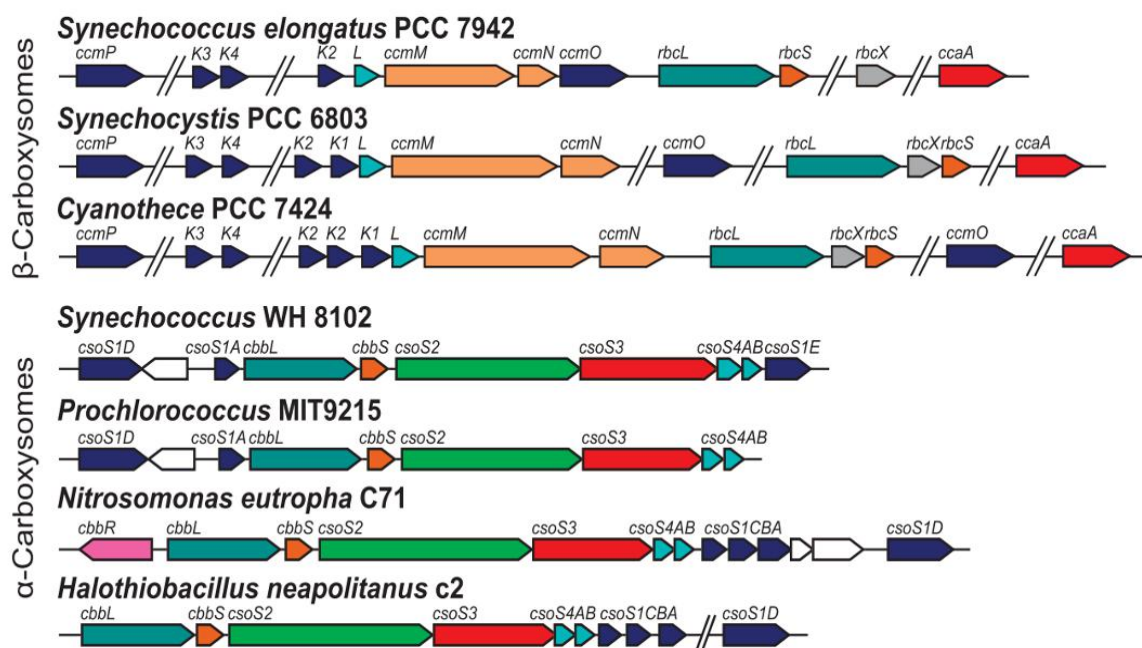
Figure 1-5. The organizations of the gene operons encoding α-carboxysome and β-carboxysome proteins in different species. This diagram was taken from Rae *et al.* (2013). The *ccm* operon of β-carboxysomes (top) and the *cso* operon of α-carboxysones (bottom) represented in variety of species. The *cso* operon includes the genes of large and small subunit of RuBisCO (*cbbL*, dark green; *cbbS*, orange), a gene of carbonic anhydrases (*csoS3*, red), the gene of intrinsically disordered protein (*csoS2*, green) and the genes of shell proteins (*csoS1*, dark blue; *csoS4*, cyan). It is a little less clear in the case of *ccm* operon, since the genes that encoding for their proteins are scattered across the genome. The genes of large and small subunit of RuBisCO (*rbcL*, dark green; *rbcS*, orange), the genes of shell proteins (*ccmK/P/O*, dark blue), the genes of carbonic anhydrases (*ccaA*, red), the core enzyme (*ccmM/N*, dark yellow)

1.2.2 The CB-encapsulated enzymes

Both types of carboxysomes share similar protein shell structures that encapsulate RuBisCO and carbonic anhydrases (CA) (Figure 1-2a) (Cannon *et al.*, 2010). As a central CO_2-fixing enzyme in the Calvin-Benson-Bassham cycle, Rubisco catalyses the carboxylation of ribulose-1,5-bisphosphate (RuBP) by CO_2 to produce two molecules of 3-phosphoglycerate (PGA) (Tabita, 1999; Atomi, 2002); RuBisCO also has oxygenase activity, which catalyzes the oxygenation of RuBP by O_2 to produce one molecules of 2-phosphoglycolate. Due to the toxicity of 2-phosphoglycolate, it should be converted back to 3-PGA by the photorespiration

process (Figure 1-6) (Spreitzer and Salvucci, 2002, Xu *et al.*, 2015). Form 1A and 1B

RuBisCO are composed of eight large and eight small subunits to form an L_8S_8 structure

(Andersson, 1996). The active sites are located in the large subunits (Andersson *et al.*, 1989;

Andersson and Backlund, 2008), whereas the small subunits play roles in the structural

stability and enhancing catalytic efficiency (Schneider *et al.*, 1990; Kanevski *et al.*, 1999;

Andersson and Backlund, 2008; Morita *et al.*, 2014). Despite its significant contribution to

global carbon fixation, RuBisCO is an inefficient enzyme due to its low turnover number and

poor capacity of differentiating CO_2 and O_2 (Cannon *et al.*, 2010).
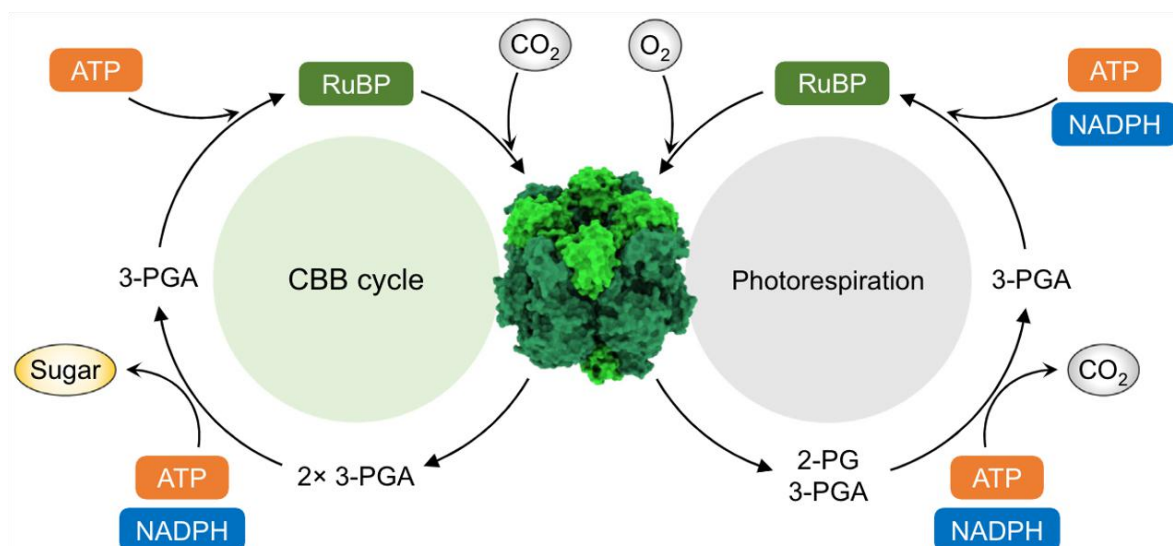
Figure 1-6. The role of Rubisco in Calvin-Benson-Bassham cycle and photorespiration pathway. This diagram was taken from Liu (2021a). The carboxylase activity of Rubisco is enhanced as a result of its association with CO_2 in the Calvin–Benson–Bassham (CBB) cycle. Here, 3-phosphoglyceric acid (3-PGA) is generated by Rubisco converting ribulose-1,5-bisphosphate (RuBP) and CO_2. For the photorespiration pathway, 3-PGA and 2-phosphoglycolate (2-PG) are produced by Rubisco binding to O_2. Comparing with CBB cycle, converting 2-PG to 3-PGA will consume more energy, which is a energy-waste process (Tabita, 1999).

Carbonic anhydrase (CA) is another essential enzyme that is encased in carboxysomes, and it

is an absolutely conserved structure (Cannon *et al*., 2010; Kimber, 2014). CA showed a low

abundance compared to total proteins of carboxysomes and is associated with the shell

proteins (Li *et al.*, 2020; Sun *et al.*, 2021). CsoSCA in α-carboxysomes belongs to the β-type

CA enzyme according to X-ray crystallographic analysis and is encoded by the *csoSCA* gene

(Baker *et al.*, 2000; So *et al.*, 2004; So and Espie, 2005; Sawaya *et al.*, 2006). CsoSCA is composed of three domains: an N-terminal domain, a C-terminal domain, and a catalytic domain (Heinhorst *et al.*, 2006). Sawaya *et al.* revealed that the N-terminal domain of CsoSCA is important for interactions with RuBisCO and shell proteins. Removing the N-terminal domain of CsoSCA led to the disrupted association of CsoSCA with other protein and reduced bacterial growth (Cannon *et al.*, 2010). However, removal of CsoSCA in *H. neapolitanus* did not affect notably the formation of carboxysomes (Dou *et al.*, 2008).

The β-carboxysomal CA, CcaA, is a β-type CA enzyme and is encoded by the *ccaA* gene (So and Espie, 1998; Long *et al.*, 2007). Comparison of sequences from multiple strains showed that CcaA has a unique C-terminal extension with unknown function (Kimber and Pai, 2000). Earlier studies found that CcaA of Syn7942 and Syn6803 is important for CO_2 fixation; the absence of *ccaA* gene led to the growth requirement of high CO_2 levels (Yu, 1992; So *et al.*, 2002b). CcmM is another type of CA that can be present in some (but not all) β-carboxysomes. CcmM of *Synechococcus* PCC7942 contains two forms CcmM58 (~58 kDa) and CcmM35 (~37 kDa). CcmM58 contains a γ-type carbonic anhydrase-like (γCAL) domain and three RuBisCO small subunit-like (SSUL) modules (Figure 1-7). The role of the γ-type carbonic anhydrase-like domain in the production of CcmM58 has been determined (Long *et al.*, 2007; Cot *et al.*, 2008). The N-terminus of CcmM58 links to the carboxysome shell (Cot *et al.*, 2008; Kinney *et al.*, 2012) and its C-terminal region bonds to RuBisCO also were confirmed (Long *et al.*, 2007; Cot *et al.*, 2008; Long *et al.*, 2010b). Smaller form CcmM35 includes only small subunit-like modules (Figure 1-7) and operates RuBisCO condensates via salt bridge and van der Waals interaction between arginine and phehylalanine of SSUL modules with the larger and small subunit of RuBisCO (Wang *et al.*, 2019b). In addition, experimental studies confirmed that protein-protein interaction between CcmM58 and CcaA

is the result of the γCAL domain of CcmM58 binding to the C-terminus of CcaA; CcmM58 connects with other CcmM58s via their γCAL domains, which is a head-to-head association to greatly increase the local concentration of the SSUL modules, resulting in a higher affinity of CcmM58 to Rubisco (Zang *et al.*, 2021). Like CsoSCA, the deletion of CcaA in Syn6803 has no significant effects on carboxysome formation (So *et al.*, 2002a). However, abolished carboxysomes appeared by deleting CcmM (Ludwig *et al.*, 2000).
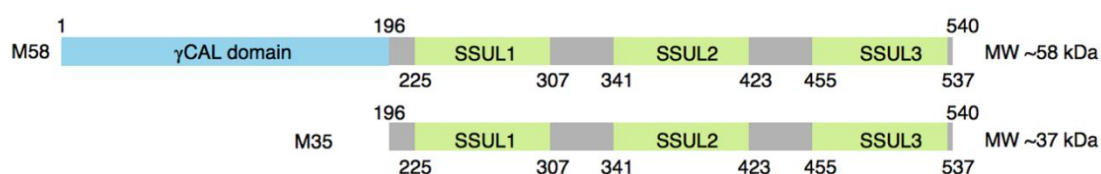
Figure 1-7. CcmM58 and CcmM35 domain structures. This diagram was organized from Zang *et al.* (2021). The γCAL located in the N-terminal of CcmM58 follows by three SSUL modules. Three SSUL modules are connected by the complex linkers. CcmM35 only contains SSUL modules, which is produced from the internal ribosome binding site (RBS) (Long *et al.*, 2010a)

1.2.3 The CB function and shell permeability

Carboxysomes are essential for CO_2 fixation, which is the final stages of the carbon-concentrating mechanism (CCM) in cyanobacteria (Rae *et al.*, 2017). The cyanobacterial CCM comprises two parts (Figure 1-8), the first part is bicarbonate is actively pumped inside the cell via bicarbonate transporters; Furthermore, CO_2 was transformed by NDH-1 into HCO_3^-, allowing HCO_3^- accumulation in the cytosol and limiting CO_2 leakage (Long *et al.*, 2016a; Wang *et al.*, 2019a; Sui *et al.*, 2020; Fang *et al.*, 2021); the second part is the accumulated bicarbonate in the cell cytoplasm then diffuses into the carboxysome across the shell, which is converted to CO_2 by CA around RuBisCO (Cannon *et al.*, 2001; Rae *et al.*, 2013; Kerfeld and Melnicki, 2016; Long *et al.*, 2016b). As a result of this CCM mechanism, a high concentration of CO_2 is supplied to the inefficient RuBisCO for improving RuBisCO carboxylation while reducing photorespiration (Price and Badger, 1989; Kaplan *et al.*, 1991;

Badger *et al.*, 1998; Cannon *et al.*, 2001; Badger and Price, 2003; Tcherkez *et al.*, 2006). The carboxysomes shell has selective permeability to metabolites and products, due to the charged central pores of shell proteins (Frey *et al.*, 2016; Faulkner *et al.*, 2020). Bicarbonate, 3-Phosphoglyceric acid (3-PGA) and Ribulose 1,5-bisphosphate (RuBP) can diffuse in and out of shell effectively (Heinhorst *et al.*, 2006; Dou *et al.*, 2008; Faulkner *et al.*, 2020). However, whether the carboxysomes shell prevents CO_2 leaking still remains unclear.
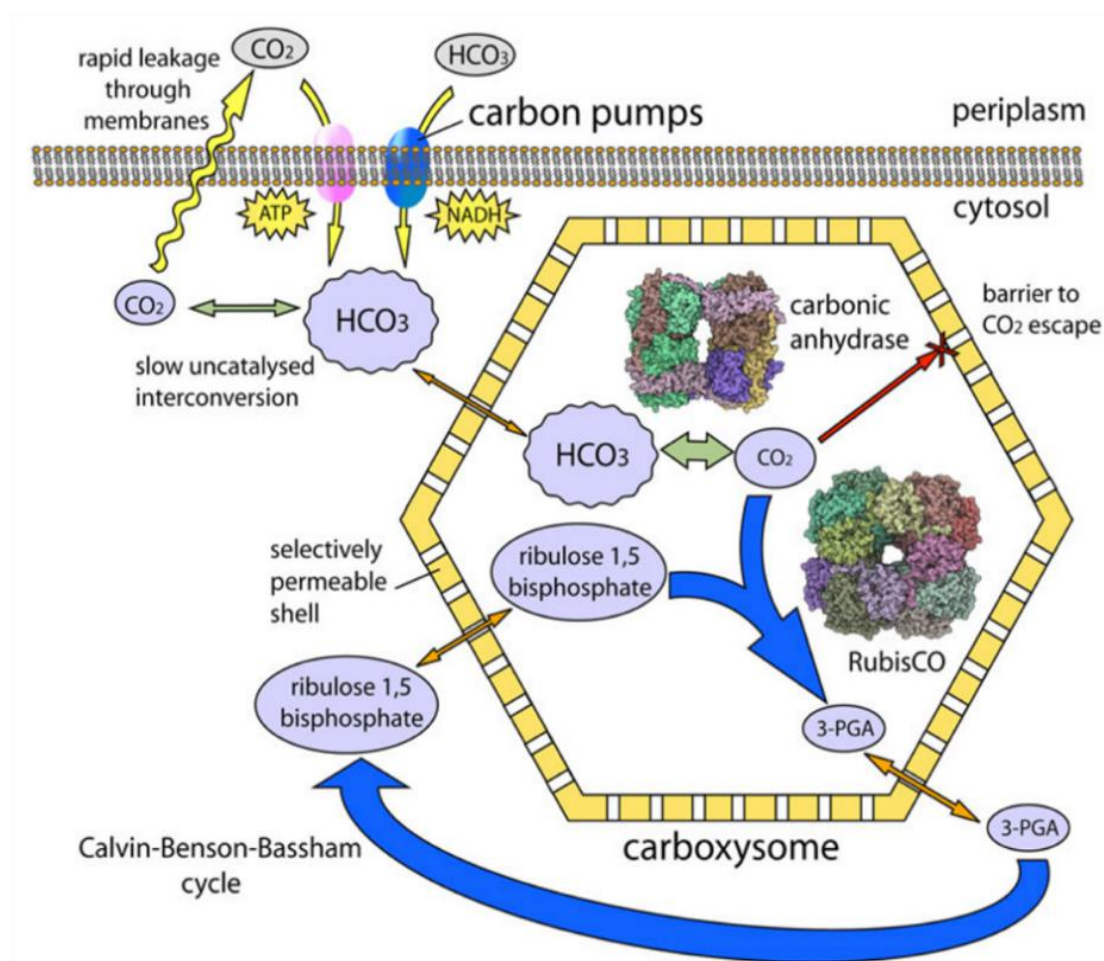
Figure 1-8. Model of CO_2-concentrating mechanism (CCM). This diagram was modified from Espie and Kimber (2011). Inorganic carbon enters and is accumulated in the cytosol. A part of HCO_3^- is converted to CO_2 and uptake CO_2 is transformed HCO_3^- via NDH-1. Accumlated HCO_3^- enters the carboxysome via the central pores of shell proteins. Within the carboxysome, HCO_3^- is converted CO_2 by encapsulated carbonic anhydrase. The carboxysomes shell may prevent CO_2 leakage to the cytosol and permit CO_2 accumulation.

1.2.4 CB structures

Our knowledge about the biochemical processes and biosynthesis of carboxysomes and other BMCs have been greatly improved by structural studies using TEM and X-ray crystallography. The isolated complete carboxysomes of *H. neapolitanus* and Syn8102 have been obtained as stable and regular icosahedrons. Moreover, the single-layer carboxysome shell is about 3-6 nm thick and RuBisCO molecules are arranged in layers within the shell (Schmid *et al.*, 2006; Iancu *et al.*, 2007). Another isolated β-carboxysome samples from Syn7942 also showed the feature of stable and regular icosahedrons for intact β-carboxysome structures (Faulkner *et al.*, 2017). The intact β-carboxysomes structure demonstrated the polyhedral and regular shape with diameter from 100 nm to 200 nm. TEM images showed the intact β-carboxysomes contains an outer shell that the average thickness is about 4.5 nm, ordered Rubisco arrays,and interval "gap" around 2 nm between the outer shell and RuBisCO (Figure 1-9b). Moreover, this isolated samples also found the partial β-carboxysome module that has the shell facets that the edges are straight and regular. Additionally, 1 or 2 layers of RuBisCO also were seen in the partial β-carboxysome structures (Figure 1-9a). A change in the size of intact β-carboxysomes from Syn7942 may be caused by the growth of strains in an exponential phase from one day to four days under changing environments, such as level of CO2 and light intensities (Figure 1-9c and 9d) (Sun *et al.*, 2019).
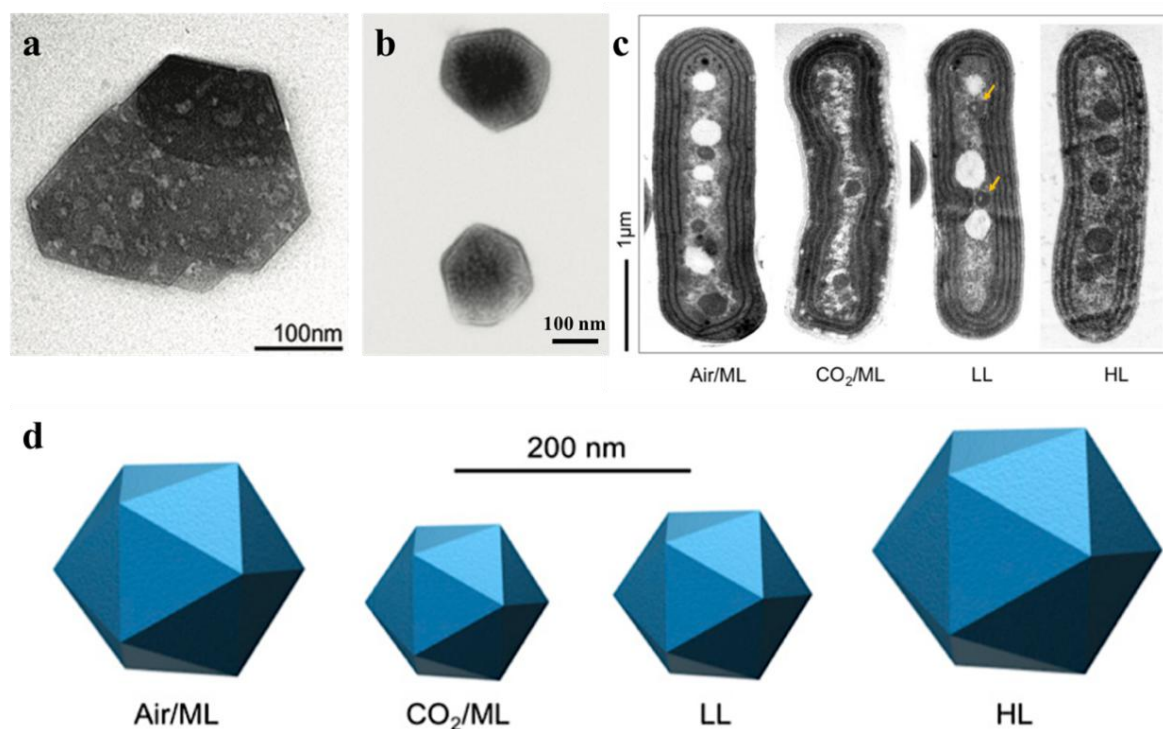
Figure 1-9. The different structures of β-carboxysomes from *Synechococcus elongatus* PCC 7942. (a) The partial β-carboxysomes modules. The shell facets with straight and regular edges as well as Rubisco enzymes aggregate within the shell also was seen. Scale bar is 100 nm. (b) Intact β-carboxysomes. The outer shell with regular and straight edges was observed and the ordered Rubisco also was seen within the shell. Scale bar is 100 nm. Figure 1-9a and 1-9b were modified from Faulkner *et al.* (2017). (c) The image of thin-section of β-carboxysomes from *Synechococcus elongatus* PCC 7942 under different treatments.The size of β-carboxysomes could be altered depending on the changing of environment condition. The size is large under both moderate light (ML) and high light (HL) in air treatments compared with moderate light in 3% CO_2 and low light (LL) in air treatments. Scale bar, 1 µm. (d) The models of size of β-carboxysomes. Scale bar, 200 nm. Figure 1-9c and 9d were taken from Sun *et al.* (2019)

Analyzing each protein structure allows for a better understanding of the overall structure and function of carboxysomes, as well as the self-assembly principles of carboxysomes at the atomic level. Following a series of observations and analysis using X-ray crystallography, the carboxysome shell proteins from *H. neapolitanus* and Syn6803 were successively observed to form hexamers and pentamers (Kerfeld *et al.*, 2005; Tsai *et al.*, 2007b; Tanaka *et al.*, 2008a). The hexamer proteins packed tightly to form the hexameric blocks, which subsequently formed the flat facets of the carboxysomes shell (Kerfeld *et al.*, 2005; Tsai *et al.*, 2007b)

(Figure 1-10a). Identification of the pentamer proteins in *H. neapolitanus* and Syn6803 was deemed to be important since they occupied the vertices of the icosahedral shell with low abundance, resulting in a closed shell structure (Parsons *et al.*, 2008) (Figure 1-10a). Apart from hexamers and pentamers, another shell protein appeared in the form of trimer that contains two hexamers (Klein *et al.*, 2009; Cai *et al.*, 2013). This trimer could form pseudohexamers and incorporate into the shell structure (Figure 1-10b). Additionally, the dimerization of the trimer could pack in the layer of the shell (Figure 1-10c).
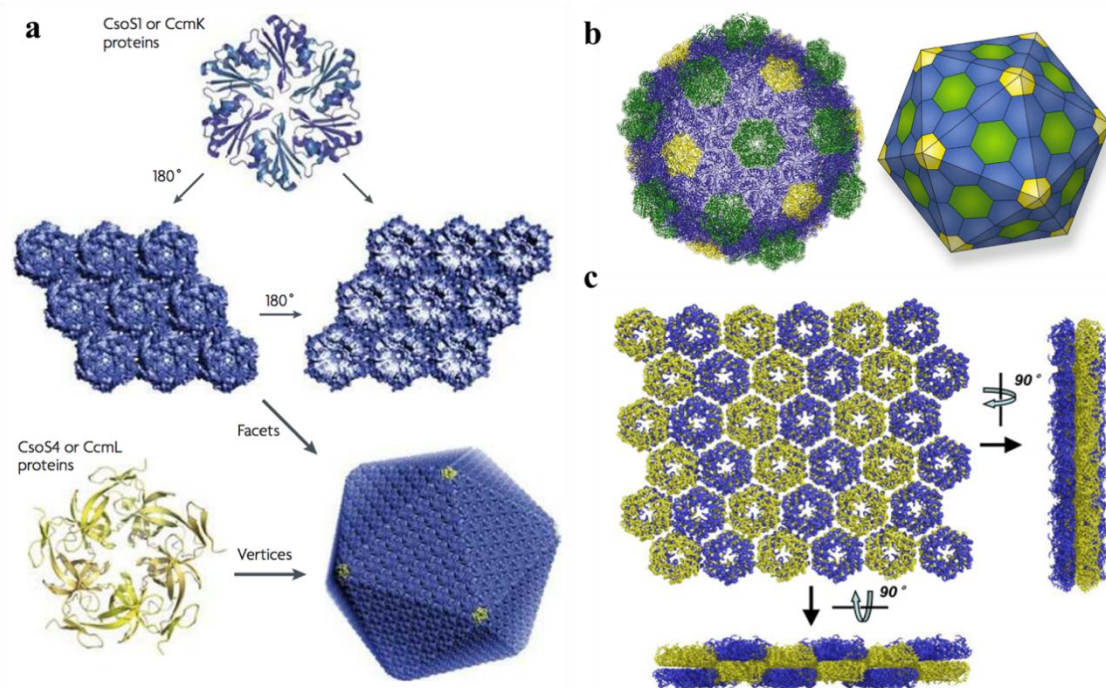
Figure 1-10. Carboxysome shell protein structure. (a) The hexamer proteins CsoS1 or CcmK togethers to create the shell facets. CsoS4 or CcmL are pentamer proteins that are suggested to be responsible to for the vertices of the carboxysomes shell. This diagram was modified from Yeates *et al.* (2008a). (b) The polyhedral shell of *Haliangium ochraceum* (HO). Hexamer protein showed in blue, pentamer protein showed in yellow, and trimer protein showed in green. This model was made from Sutter *et al.* (2017). (c) The dimerize of the trimer proteins were packed in the single layer of carboxysome shell. This model was made from Cai *et al.* (2013).

1.3 The α-carboxysome (α-CB)

1.3.1 The structure of α-CB

The α-carboxysome from *H. neapolitanus* is composed of the large and small subunits of RuBisCO (CbbL and CbbS), shell proteins (CsoS1 and CsoS4), and shell-associated proteins (CsoS2 and CsoSCA). These proteins components were encoded in a single *cso* operon (Figure 1-5 and 1-11). Rubisco filled the interior of the α-carboxysomes, which occupied around 66% of the total carboxysomes protein weight (Sun *et al.*, 2021). The thin shell of the α-carboxysome is created by the shell proteins CsoS1 and CsoS4, which have 9- to 11-kDa (Price *et al.*, 1993; English *et al.*, 1994; Tanaka *et al.*, 2008a). The CsoS1 proteins are highly conserved and have three protein paralogs termed CsoS1A, CsoS1C, and CsoS1B (Shively and English, 1991). There are only two amino acid differences between the sequences of CsoS1A and CsoS1C, whereas CsoS1B has an extra 12-aa carboxyl terminus and is 90% homologous to CsoS1A and CsoS1C (Cannon *et al.*, 2003, Shively *et al.*, 1996). Together, these three CsoS1 proteins represent to 17.1% of the total carboxysomes protein weights (Sun *et al.*, 2021). CsoS1D is a pseudohexamer that is not including the *cso* operon and is a low abundant protein in the shell (Sun *et al.*, 2021). CsoS4 pentamers contain CsoS4A and CsoS4B, which is the essential structural protein occupying the vertices of the shell (Tanaka *et al.*, 2008a). The intrinsically disordered protein, CsoS2, is essential for Rubisco packing and shell assembly (Cai *et al.*, 2015a, Oltrogge *et al.*, 2020) (discussed below).
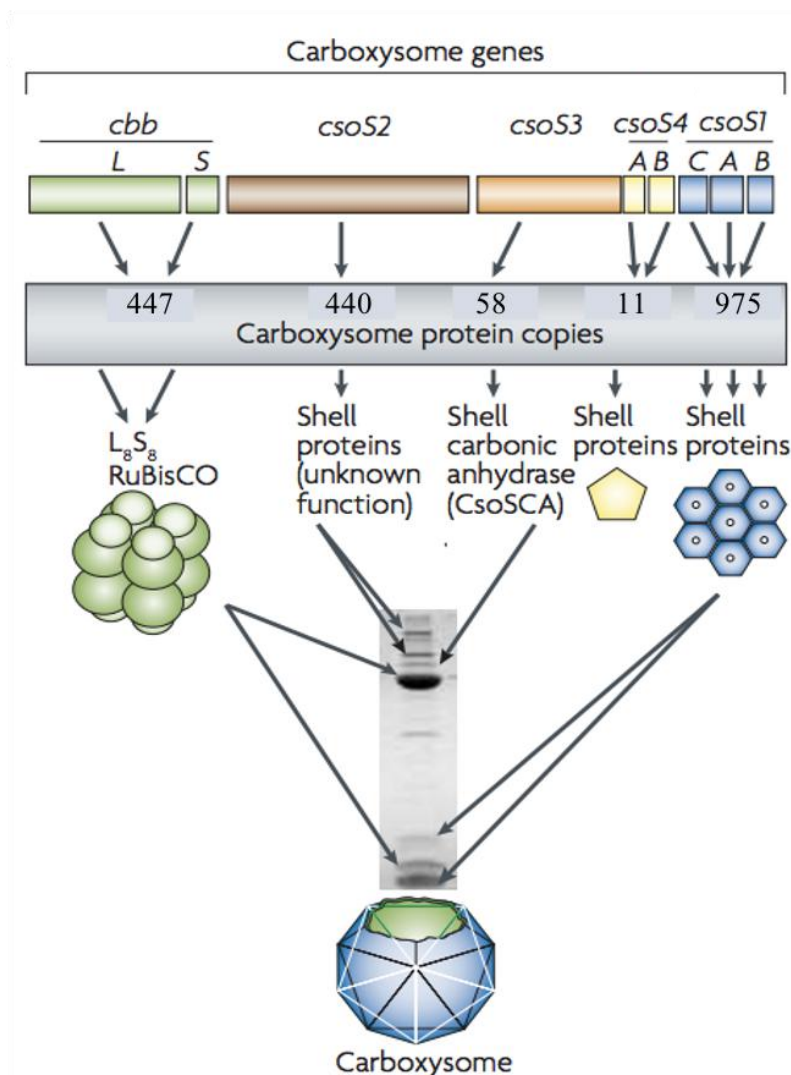
Figure 1-11. The genes arrangement of the α-Carboxysome from *H. neapolitanus*. This diagram was modified from Yeates *et al.* (2008b). The paralogous genes are labelled by using the same colour. The SDS-PAGE gel shows the identified carboxysomes proteins from purified carboxysomes. The carboxysome protein copies were modified based on the QconCAT strategy (Sun *et al.*, 2021). According to the QconCAT determination of protein abundances, the α-carboxysome from *H. neapolitanus* contains 447 copies of Rubisco, 863 copies of CsoS1A/C, 112 copies of CsoS1B, 248 copies of CsoS2A, 192 copies of CsoS2B, 58 copies of CsoSCA, 8.8 copies of CsoS4A, 2.2 copies of CsoS4B, and 2.9 copies of CsoS1D.

1.3.2 The α-CB shell

The α-carboxysome shell is composed of the shell proteins CsoS1 and CsoS4. The structures of CsoS1A and CsoS1C exhibited the characteristic of regular hexamers (Figure 1-12). The CsoS1A (PDB: 2EWH) protein structure in *H. neapolitanus* contains α/β fold (Figure 1-12a).

CsoS1A forms the cyclic hexameric protein with a central pore formed by six monomeric subunits. This pore is around 4 Å wide at its narrowest point in hexamers (Tsai *et al.*, 2009). Further investigation revealed that CsoS1A hexamers interact tightly with other hexamers to form a flat molecular layer (Tsai *et al.*, 2009) (Figure 1-10a). Each CsoS1A hexamer has a concave side that is mainly positively charged and a convex side that is slightly negatively charged (Figure 1-12b). CsoS1C (PDB: 3H8Y) is similar to CsoS1A in terms of sequence and structure. There is only one amino acid change at residue 97 (glutamate to glutamine), and this difference did not result in any structural changes (Tsai *et al.*, 2007b). However, a difference of electrostatics of CsoS1A/C could be observed on the concave faces (Figure 1-12b).
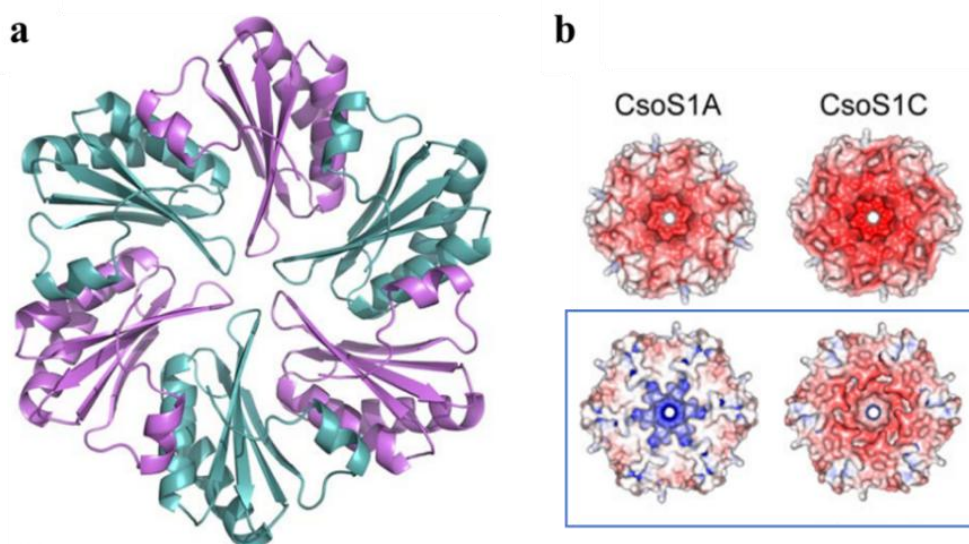
Figure 1-12. The structure of CsoS1A protein (PDB: 2EWH). (a) The secondary structure of CsoS1A of *H. neapolitanus*. (b) The charge distribution of convex (top) and concave (bottom) sides of Cso1A/C. Both models were made from Tsai *et al.* (2007b). Red shows negative charge and blue shows positive charge.

The trimer protein CsoS1D (PDB: 3FCH) of *Prochlorococcus marinus* MED4 was firstly identified in 2009 (Klein *et al.*, 2009). Sequence analysis revealed that CsoS1D possesses two BMC-H domains, which are located at residues 50-150 and 190-265, respectively. Two distinct conformations of CsoS1D have been reported, one with a closed central pore and one

with an open central pore with approximately 14 Å (Figure 1-13a). These two conformations were assumed to represent the opening of the gate to allow larger molecules enter and exit the shell (Klein *et al.*, 2009). CsoS1D trimers form pseudohexamers that resemble hexameric proteins and are thought to be responsible for the formation of shell facets given the opened CsoS1D trimer clashing with CsoS1 protein and the closed CsoS1D trimer easily fitted into the shell layer (Figure 1-13b). This is due to some side chains of opened CsoS1D trimer disrupt the interaction between opened CsoS1D and CsoS1, while the lysine side chain of closed CsoS1D could interact well with CsoS1 (Klein *et al.*, 2009).
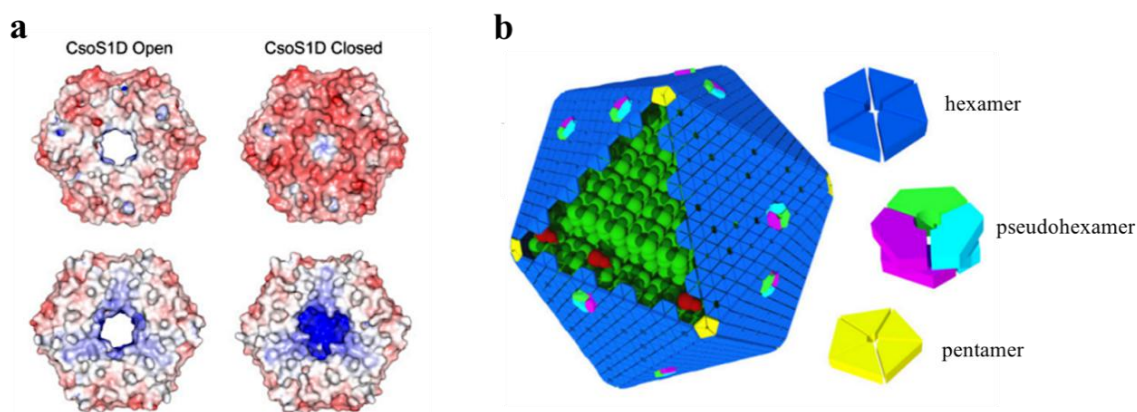
Figure 1-13. The structure of CsoS1D (PDB: 3FCH). (a) The electrostatic potentials of CsoS1D in two conformations, red shows negative charge and blue shows positive charge. This model was made from Klein *et a*l. (2009). (b) The model of the α-carboxysome, including core enzymes and outer shell. The facets of the outer shell has hexamers and pseudohexamers. This model was made from Lin (2018)

CsoS4A/B, the pentamer protein, has low copy numbers in α-carboxysomes (Cai *et al.*, 2009). The structure of CsoS4A (PDB: 2RCF) was determined at a resolution of 2.15 Å, containing β strands and α helices (Tanaka *et al.*, 2008a). CsoS4A is a disk-shaped symmetrical pentamer (Figure 1-14) and has a central pore with the diameter of 3.5 Å (Tanaka *et al.*, 2008a). CsoS4B (PDB: 6JY5) displays a similar structure to CsoS4A (Zhao *et al.*, 2019a) (Figure 1-14). CsoS4A and CsoS4B have structural similarities (Figure 1-14a), however, the pore of CsoS4B (2.9 Å) is smaller than that of CsoS4A. (Figure 1-14a). According to the

crystal structures of CsoS4A and CsoS4B, they are assumed to sit vertices in the α-carboxysome shell and are important for shaping the icosahedral shell made of CsoS1 proteins (Figure 1-13b). In *H. neapolitanus*, a double knockout mutant of CsoS4A/B resulted in elongated carboxysomes (Cai *et al.*, 2009) (Figure 14b), indicating that CsoS4A/B are responsible for the shell curvature (Figure 1-10 and 1-13b).
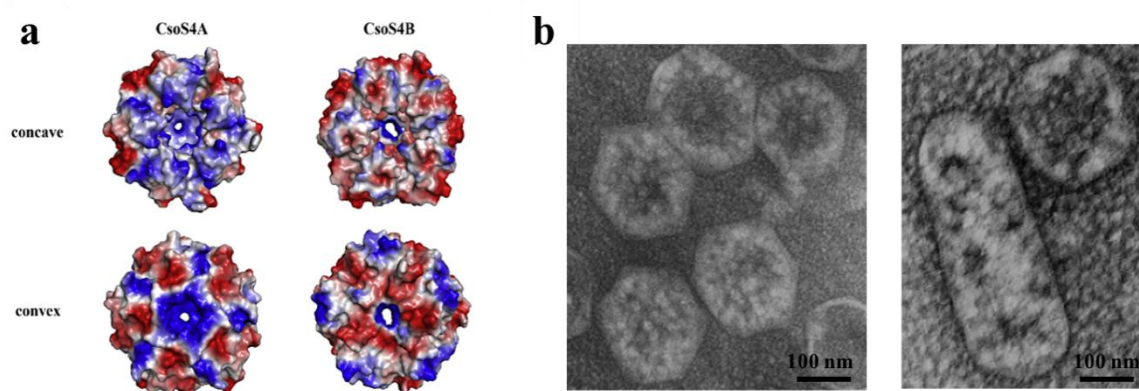
Figure 1-14. The structure of CsoS4. (a) The electrostatic potentials of CsoS4A and CsoS4B. CsoS4 proteins are pentamers proteins with a variety of edges, concave (top) and convex (bottom) conformation that exhibit different electrostatic potentials, red shows negative charge and blue shows positive charge. This model was made from Zhao *et al.* (2019a). (b) The TEM image of purified α-carboxysome from *H. neapolitanus*. Left shows the purified wild type of α-carboxysome from *H. neapolitanus* and right shows the CsoS4A/B double knockout of α-carboxysome from *H. neapolitanus*. Scale bar is 100 nm. These images were took from Cai *et al.* (2009).

1.3.3 The composition of α-CB interior

Apart from RuBisCO and CsoSCA, CsoS2 is also found within the α-carboxysome (Cai *et al.*, 2015a, Chaijarasphong *et al.*, 2016). In *H. neapolitanus*, CsoS2 appears two different forms: a longer form (CsoS2B) and a shorter form (CsoS2A) with a 1 to 1.3 ratio (Sun *et al.*, 2021). CsoS2B and CsoS2A are translated via programmed ribosomal frame shifting (Chaijarasphong *et al.*, 2016), both forms share the same N-terminus and CsoS2A is a C-terminal truncated protein with 570 amino acids (Baker *et al.*, 1999; Gonzales *et al.*, 2005; Cai *et al.*, 2015a; Chaijarasphong *et al.*, 2016). Analysis of the proteins secondary structures of different strains (e.g. *H. neapolitanus* and *Prochlorococcus* strains) showed that CsoS2 is

composed of the N-terminal region enriched with α-helices, the middle region that is formed

by β-strands, and the C-terminal region (Figure 1-15a) (Cai *et al.*, 2015a). In addition, the

primary structure of CsoS2 comprises multiple sequence repeats (Cai *et al.*, 2015a) (Figure 1-

15a and 15b). The N-region of *H. neapolitanus* CsoS2 includes four repetitive motifs, each

containing 16 amino acids. The M-region has 6 repetitive motifs with several short repeats

(three amino acids as a unit) (Cai *et al.*, 2015a) (Figure 1-15b).
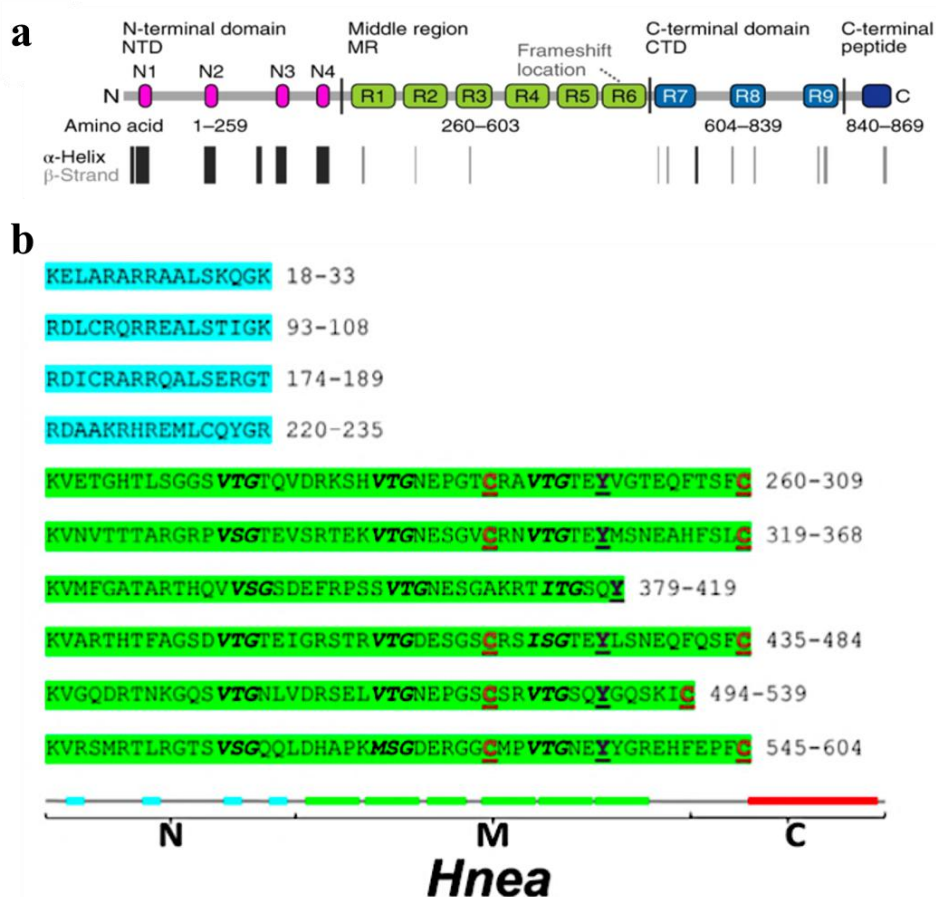
Figure 1-15. Analysis of the structure of *H. neapolitanus* CsoS2. (a) The CsoS2 secondary structure with repeat motifs. CsoS2 has three domains: the N-terminal domain, the middle region, and the C-terminal domain. The N-terminal domain contains the four repeating motifs with enriched α-helix, and the protein sequences was highlighted in (b) by light blue. The middle region contains the six repeat motifs, which also is presented in (b) by green. Every repeat includes three short repeats that is made up via three amino acids. There are eight amino acids as the "linker" to connect each short repeat. The last repeat in the middle region showed a frameshift location. A programmed ribosomal frameshift site appeared in this location, which is important for CsoS2A and CsoS2B expression(Chaijarasphong *et al.*, 2016). The C-terminal domain contains three repeats, which might drive the function of encapsulation peptide (Li *et al.*, 2020, Tan *et al.*, 2021a). This figure was adapted from Cai *et al.* (2015a) and Oltrogge *et al.* (2020)

CsoS2 has been found to interact with CsoS1 and RuBisCO (Cai *et al.*, 2015a). The mixing of different molar amounts of CsoS2 and CsoS1 *in vitro* resulted in interactions between CsoS2 and CsoS1 and precipitates with changed turbidity. Formation of the CsoS2-CsoS1 assemblies is dependent on the presence of shell proteins CsoS1D (Cai *et al.*, 2015a). Recent study found that the variable numbers of the N-terminal repeats of CsoS2 interacted with Rubisco effects the formation of α-carboxysomes (Oltrogge *et al.*, 2020). Specially, two N-terminal repeating motifs are required for α-carboxysome formation (Figure 1-16a). On the basis of these findings, a proposed model is presented for the function and interior arrangement of CsoS2 in α-carboxysomes (Figure 1-16b). The partial shells form firstly, followed by binding of the N-terminal region of CsoS2 to shell proteins; CsoS2-mediated association triggers RuBisCO nucleation and form liquid-like matrix to promote complete carboxysome assembly.
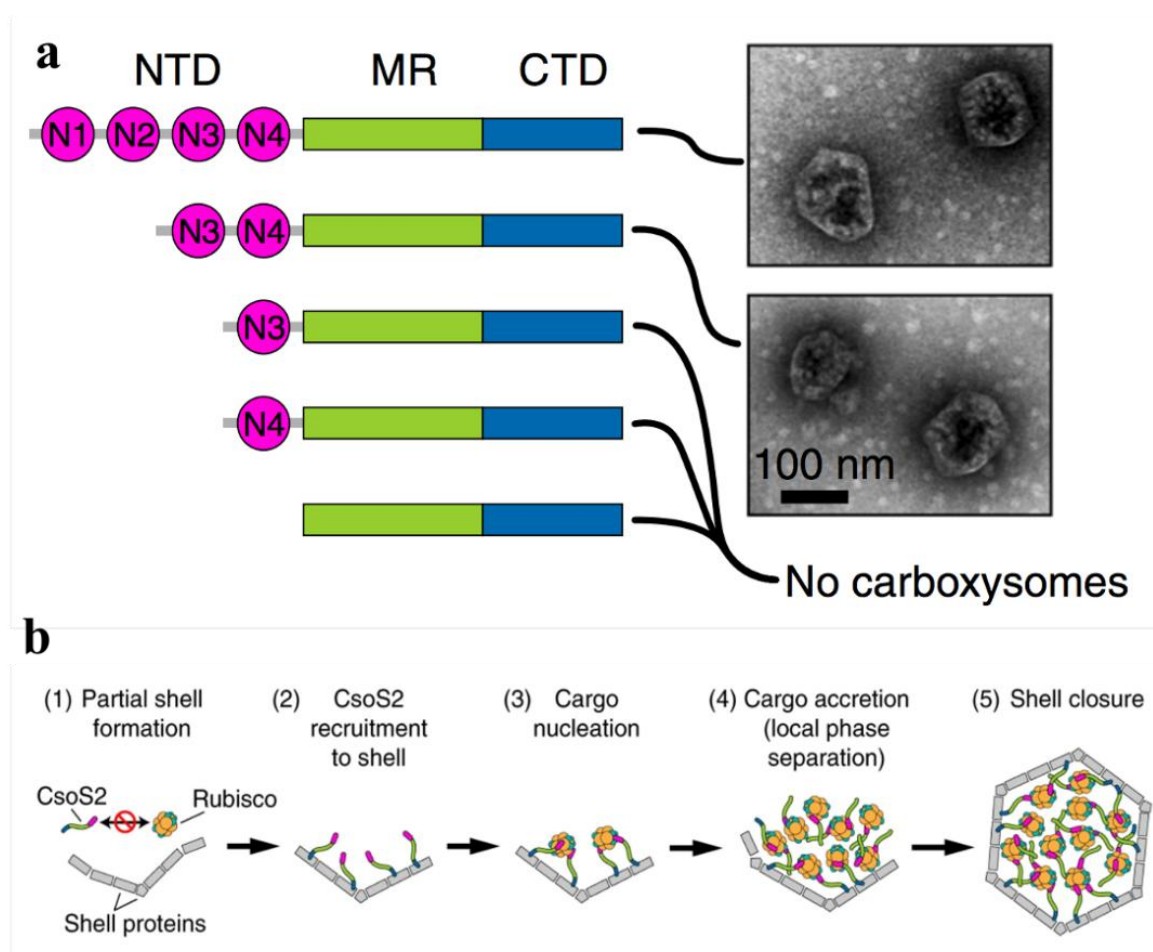
Figure 1-16. The role of CsoS2 in the formation and assembly of α-carboxysomes. (a) The variants of CsoS2 with variety numbers of the N-terminal repeat motifs were constructed. These variants were produced heterologously in *E. coli* through protein expression. TEM images showed the α-carboxysome formation related to the numbers of the N-terminal repeat motifs of CsoS2. Scale bar is 100 nm. (b) The assembly pathway of α-carboxysomes. This pathway responses to the assembly of the core enzyme and shell simultaneously (also showed in Figure 1-4b). In the pathway, shell proteins like CsoS1 tiles together to be the sheet, following by the C-terminus of CsoS2 connects to the shell proteins. Simultaneously, RuBisCO appears in the nucleation and interacts with the N-terminal of the CsoS2. Further nucleation of Rubisco and continued shell assembly. The shell closure via shell protein CsoS4. Both figures were modified from Oltrogge *et al.* (2020)

1.3.4 The relationships between α-CB shell with β-CB shell

The major shell proteins in the carboxysome are known as CcmK1/2/3/4 and CcmO for β-CB or CsoS1A/B/C/D for α-CB. Section 1.3.2 mentioned CsoS1A/C have the sidedness of concave and convex, which also observed for the CcmK2 (PDB: 2A1B) of Syn6803 (Tanaka

et al., 2009). Besides, the analysis of the primary structures of CcmK1, and CcmK4 found that they contain a C-terminus extension around 10 amino acid (Tanaka *et al.*, 2009), which is similar to shell protein CsoS1B. CcmL of β-CB is similar to CsoS4A, both of them form the pentamer with the sidedness of concave and convex (Tanaka *et al.*, 2008b). Both shells also contain trimeric shell proteins, CsoS1D for α-CB shell and CcmP for β-CB shell. The structural investigation of CsoS1D revealed that it has pseudo-sixfold symmetry with the gated pores (Klein *et al.*, 2009). CcmP is homologous protein of CsoS1D, which performs the role of forming a gated, separate nanocompartment inside the β-CB shell (Cai *et al.*, 2013; Larsson *et al.*, 2017).

1.4 Engineering of bacterial microcompartments (BMCs)

1.4.1 The engineering of recombinant bacterial microcompartments

BMCs have long been considered an ideal platform for constructing metabolic factories within non-native hosts by synthetic biology. In 2008, the first recombinant expression of BMC was achieved by transferring the entire *Pdu* operon from *Citrobacter freundii* to *E. coli* (Parsons *et al.*, 2008). This recombinant Pdu BMC has similar characters like size, shape and metabolic pathways as the wild type Pdu BMC (Mayer *et al.*, 2016). Likewise, the α-carboxysome operon from *H. neapolitanus* was transformed into *E. coli* and the gram-positive bacterium *Corynebacterium glutamicum* to generate carboxysome-like structures with detectable CO_2 fixation activity (Bonacci *et al.*, 2012; Baumgart *et al.*, 2017). Recently, the Liu lab at the University of Liverpool co-expressed Rubisco activases with α-carboxysomes in *E. coli* to further improve Rubisco activity (Chen *et al.*, 2022), and also introduced the β-carboxysome operon from Syn7942 into *E. coli* to generate the β-carboxysome-like particles with RuBisCO activity (Fang *et al.*, 2018). Moreover, several

attempts have been made to express carboxysome genes in plant chloroplasts (Lin *et al.*, 2014b; Long *et al.*, 2018), with the intent of enhancing CO_2 fixation and agricultural yields.

1.4.2 Synthetic engineering of empty BMC shells

Since BMCs are capable of encapsulating metabolic enzymes, they are the perfect tool for compartmentalizing synthetic multienzyme pathways and the development of new nano-bioreactors (Rae *et al.*, 2013; Axen *et al.*, 2014; Chowdhury *et al.*, 2014; Kerfeld and Erbilgin, 2015; Zarzycki *et al.*, 2015; Kerfeld and Melnicki, 2016; Turmo *et al.*, 2017b). Empty BMC shells with native-like large internal space can be constructed by expression of shell proteins only in *E. coli*, such as Pdu shell proteins from *Citrobacter freundii* (Parsons *et al.*, 2010), Eut shell proteins from *Salmonella enterica* (Choudhary *et al.*, 2012; Sargent *et al.*, 2013), unknown-function BMC shell proteins from *Haliangium ochraceum* (Lassila *et al.*, 2014), β-carboxysome shell proteins from *Halothece* sp. PCC7418 (Sutter *et al.*, 2017), α-carboxysome shell proteins of *H. neapolitanus* (Li *et al.*, 2020).

Moreover, chimeric shell structures have been proved to be constructed by combining CsoS1 proteins of α-carboxysomes and CcmK2/4 proteins of β-carboxysomes (Cai *et al.*, 2015b), or by incorporating CsoS1A of α-carboxysomes and CcmK4 of β-carboxysomes in *Synechococcus* PCC7942 (Fang *et al.*, 2018), as well as by using EutM of Eut microcompartment to replace PduA of Pdu microcompartment in *Salmonella enterica* (Slininger Lee *et al.*, 2017). These chimeric shell structures formed by shell proteins from different origins may suggest that the various permeability of BMCs could be achieved and modulated. On the other hand, the pores of BMC shell proteins act as the portal that allow molecule transport in and out of the BMCs (Kerfeld *et al.*, 2005). It has been revealed that the

modified size and electrostatic properties of the pores, by mutation of key pore residues, could adjust the shell permeability and BMC metabolic performance (Chowdhury *et al.*, 2015; Cai *et al.*, 2016; Slininger Lee *et al.*, 2017; Faulkner *et al.*, 2020); binding a FeS cluster to the pores could improve electron flow across the BMC shell (Aussignargues *et al.*, 2015). Modifying substrate specificity is necessary for repurposing BMC shells for biotechnological applications.

1.4.3 The assembly of BMCs *in vitro*

Rapid prototyping has been facilitated by newly established technologies of constructing BMC shells *in vitro*. Three assembled rapid prototypes contain a metabolosome shell, a β-carboxysome shell, and a BMC protein-based nanotube were produced by mixing various types individual shell protein from different types of BMCs *in vitro* (Figure 1-17a) (Hagen *et al.*, 2018b). Researchers employed the single hexamer protein of *Rhodococcus* and *Mycobacterium* BMCs, BMC-H$_{Rmm}$, coupled to a short ubiquitin-like modifier domain that blocks the production of supermolecules to produce a soluble and high yield of BMC-H$_{Rmm}$ protein. Following the removal of this short ubiquitin-like modifier, free BMC-H$_{Rmm}$ proteins self-assemble into the BMC protein-based nanotube structure (Figure 1-17b). In addition, the same process was applied to a BMC-H protein from *Haliangium ochraceum*, resulting in the production of a new native BMC-H protein that was incapable of forming supermolecules. The produced BMC-H protein mixed to BMC-T protein and BMC-P protein from *H. ochraceum* was able to construct a minimal *H. ochraceum* shell (Figure 1-17c). By combining CcmK1 with CcmK2 of *Halothece* PCC7418, it was possible to establish *in vitro* self-assembly of the β-carboxysome shell (Figure 1-17d). Protein nanotubes were produced by using the affinity tags such as the hexa-histidine tag fuse to Pdu shell proteins; PduA-His$_6$

and PduB-His$_6$ were produced *in vivo*, followed by purifying and assembling under low salt

conditions *in vitro* after the His$_6$-tag was removed (Uddin *et al.*, 2018).
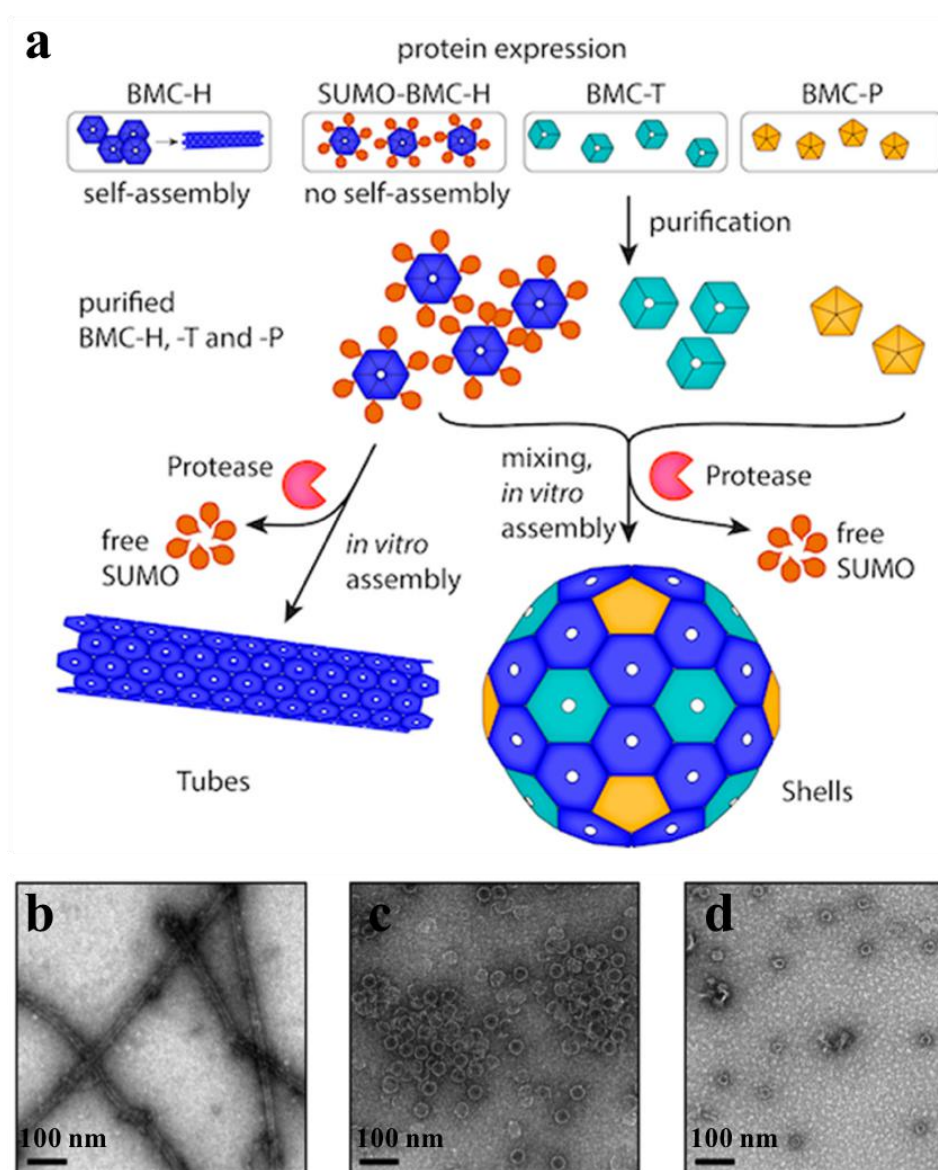
Figure 1-17. *In vitro* assembly strategy for BMC shells. This figure was modified from Hagen *et al.* (2018b). (a) Different individual BMC shell proteins were mixed *in vitro* to produce nanotube and minimal shells. Here, individual BMC shell protein firstly was fused with short ubiquitin-like modifier to prevent the supermolecules formation. Following by using protease to remove the short ubiquitin-like modifier (SUMO) from individual BMC shell proteins. Then the free BMC shell proteins self-assemble into nanotube and minimal shell structure. (b) The BMC protein-based nanotube was assembled by using the single hexamer protein of *Rhodococcus* and *Mycobacterium* BMCs. (c) The minimal *Haliangium ochraceum* shell was assembled based on the mixture of BMC-H, BMC-P, and BMC-T protein of *H. ochraceum*. (d) The minimal β-carboxysome shell was assembled by one CcmK1 and one CcmK2 protein of *Halothece* PCC7418. All scale bars are 100 nm.

1.4.4 The engineering of cargo proteins encapsulation within BMCs

The synthetic shells and their corresponding encapsulation peptides (EPs) were utilized in earlier studies to design novel BMCs. The first proof-of-concept made is to use the N-terminal 18 amino acids of PduP from *Citrobacter freundii* to load pyruvate decarboxylase and alcohol dehydrogenase from *Zymomonas mobilis* within a Pdu BMC shell for ethanol production (Lawrence *et al.*, 2014). We have recently reported that the CsoS2 C-terminus of α-carboxysomes from *H. neapolitanus* can be used as the EP to load foreign hydrogenases within the recombinant α-carboxysome shells in *E. coli* to create a new nanobioreactor for hydrogen production (Li *et al.*, 2020). In other examples, EPs have been utilized to encapsulate polyphosphate kinases and exopolyphosphatases in Pdu BMCs for polyphosphate accumulation (Liang *et al.*, 2017) and to fuse fluorescent proteins to recombinant Eut shells (Choudhary *et al.*, 2012; Quin *et al.*, 2016). Moreover, researchers employed PduP[1-18] as the template and operated rational and library-based methods to design the N-terminus of PduP as EP to enclose fluorescent protein within Pdu BMCs (Jakobson *et al.*, 2017a). Non-native protein encapsulation within Pdu BMCs through the genomic integration platform resulted in varied cargo encapsulation levels efficiencies, suggesting the potential strategies for modulating encapsulation efficiency (Nichols *et al.*, 2020).

Additionally, there are other cargo encapsulation methods to be taken into consideration. For example, co-expression of SpyTag-fused fluorescent proteins with SpyCatcher-fused BMC-T protein from *H. ochraceum in vivo* allowed the fluorescent proteins to be incorporated into the *H. ochraceum* BMCs (Young *et al.*, 2020). The *de novo* coiled-coil peptides CC-Di-B and CC-Di-A were used to fuse PduA and fluorescent proteins, respectively; protein co-expression resulted in the formation of the PduA shell-fluorescent protein assemblies (Lee *et al.*, 2018).

Each cargo encapsulation strategy has pros and cons. Native EPs do not require the modification of BMCs shell proteins, however, it is hard to modulate the EP specificity and encapsulation efficiency. Using designed EPs based on various protein–protein interactions may result in varying cargo-loading levels into BMCs. However, this will require a better understanding of the structures, abundance and organization of shell proteins in the BMCs.

1.4.5 The potential of BMCs in biomedicine

Development of nanoparticles for drug delivery and vaccination platforms is a branch of nanomedicine (Shi *et al.*, 2017; Steele *et al.*, 2017; Shen *et al.*, 2018). Recent developments have yielded a wide range of useful nanostructures, like virus-like nanoparticles, that may enclose therapeutic cargos (Aftab *et al.*, 2018). The BMC shells have the potential to be used as functional nanoparticles for applications in nanomedicine (Figure 1-18a and b). BMC shell proteins could be modified to integrate a series of appropriate peptides or drug molecules to promote the precise recognition of shells to pathogens and cancer cells (Sun *et al.*, 2018). The construction of BMCs *in vitro* also makes it possible to encapsulate cytotoxic medications used in cancer treatment. However, BMCs shell applicability in nanomedicine might be limited by its antigenicity, therefore modifying this antigenicity is expected to be a hot trend in the future (Veronese and Pasut, 2005). In the vaccine field, nanoparticles may be used as a vehicle to deliver antigens for diseases in modern vaccinations (López-Sagaseta *et al.*, 2016). It is possible to use synthetic BMC shells as scaffolds in biomedicine since it can assemble effectively and form robust nanoparticles. Furthermore, the permeability of BMC shells allows substrates to enter and exit the shell via the pore of the shell, avoids the complex process of BMC shells disassembling first and then releasing substrates (Sutter *et al.*, 2017; Hagen *et al.*, 2018a; Hagen *et al.*, 2018b).
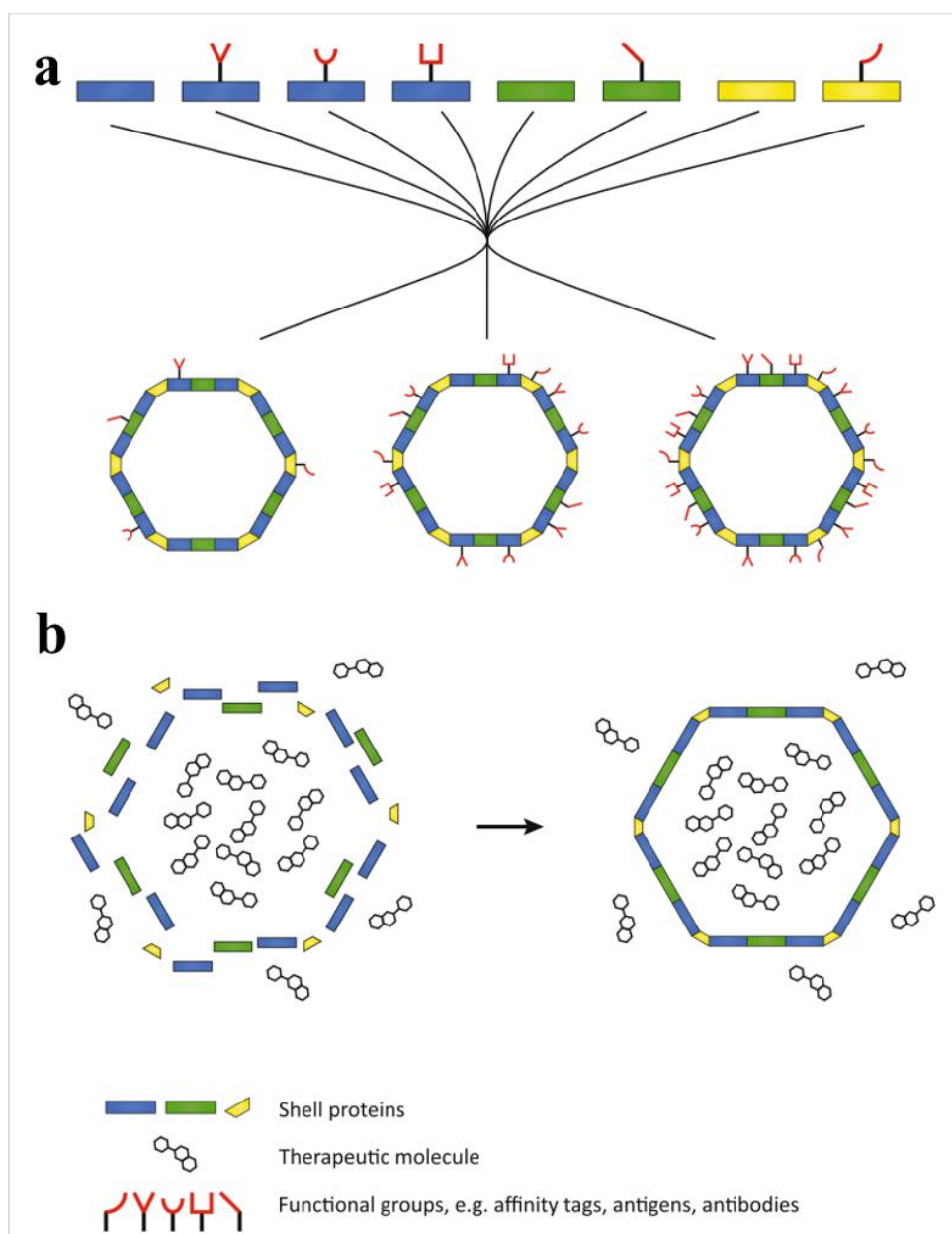
Figure 1-18. The BMCs shell assemblies for nanomedicine applications. (a) Nanomedicine can be built by fusing a set of developed peptides from existed nanoparticles to the engineered shell proteins. These developed peptides might be affinity tags, antigens, and antibodies. (b) BMC shell proteins mixed with therapeutic molecules assembled the nanomedicine for drug delivery. This figure was adapted from Kirst and Kerfeld (2019)

1.5 The overall aims of this thesis

Carboxysomes are the central CO_2-fixing organelles in cyanobacterial and proteobacteria.

The carboxysome shell is semi-permeable and serve as a physical barrier to allow passage of

substrates and products in and out of the carboxysome. The aim of my PhD research is to investigate in-depth the assembly mechanisms and engineering strategies of α-carboxysome shell in *E. coli* using synthetic biology, molecular biology, biochemistry, microscopy, and catalytic assays.

The first objective is to engineer the α-carboxysome shells in *E. coli* and incorporate hydrogenases into the recombinant shells to generate new nanobioreactors for biohydrogen production (Chapter 3). Both the large subunit (HyaB) and small subunit (HyaA) of [NiFe]-hydrogenase 1 from *E. coli* were fused with the C-terminus of CsoS2, and were co-expressed with empty α-carboxysome shells in *E. coli*, leading to the encapsulation of [NiFe]-hydrogenase into α-carboxysome shells. Hydrogenase kinetics and hydrogen production of the generated nanobioreactors were determined at different conditions. The study highlighted the possibility to encapsulate non-native cargoes into α-carboxysome shells to enhance hydrogen production.

The second objective is to produce synthetic α-carboxysome shells by expressing different combinations of shell proteins in *E. coli* for publishing (Chapter 4). Here, three α-carboxysome mini-shells were constructed, isolated, and analyzed by biochemical approaches and cryo-electron microscopy (cryo-EM). The results demonstrated that the formation of α-carboxysome mini-shells only require two types of shell proteins, hexamers and pentamers. Structural analysis also revealed the association of shell proteins with three domains of CsoS2 at the C-terminal region. Furthermore, the truncated C-terminal of CsoS2 and shell proteins also was analyzed to determine the roles of the identified CsoS2 binding domains in shell assembly.

The last objective is to construct simplified α-carboxysomes in *E. coli* (Chapter 5). We hypothesis that the simplified α-carboxysome can be synthetised in *E. coli* based on the mini-shell structures that we have discovered (Chapter 4). To test it, I produced a simplified α-carboxysome structure in *E. coli* by expressing CsoS2 and the major shell hexamer CsoS1A together with Rubisco CbbL and CbbS subunits. The formed assemblies showed carboxysome-like structures with Rubisco activity, highlighting the plasticity of the α-carboxysome protein composition and architecture.

Overall, the thesis provides new insight into the assembly, encapsulation, and modulation of the α-carboxysome shells and functional α-carboxysomes. It offers directly the synthetically engineered carboxysome structures and new tools that will aid engineering of carboxysome-based nanomaterials and diverse biotechnological applications.

Chapter 2 Material and Methods

2.1 Medium and Culture of *E. coli*

The *E. coli* Top 10 or BL21(DE3) cells were grown aerobically in lysogeny broth (LB) medium or LB agar plates (Sigma-Aldrich, UK). Antibiotics were added in medium or agar plates at the following final concentrations: 50 $\mu g \cdot mL^{-1}$ spectinomycin, 50 $\mu g \cdot mL^{-1}$ chloramphenicol, 100 $\mu g \cdot mL^{-1}$ ampicillin. All strains used in study were list in Table 2-1.

Table 2-1. Strains of E. coli with plasmids. Relevant antibiotic resistances are indicated by R: Ap, ampicillin; Ch, chloramphenicol; Sp, spectinomycin.

Strains	Description	Reference/origin	
E. coli BL21(DE)	*E. coli* BL21(DE3), WT	(Miroux and Walker, 1996)	
E. coli BL21(DE)-*hyaB::hyaA*	BL21(DE3) derivative with HyaB and HyaA; Sp^R	This study	
E. coli BL21(DE)-*hyaB-EP::hyaA-EP*	BL21(DE3) derivative with HyaB-EP and HyaA-EP; Sp^R	This study	Chapter 3
E. coli BL21(DE)-*shell-1*	BL21(DE3) derivative with Shell-1 operon; Ap^R	(Li *et al.*, 2020)	
E. coli BL21(DE)-*shell-2*	BL21(DE3) derivative with Shell-2 operon; Ap^R	(Li *et al.*, 2020)	
E. coli Top10	*E. coli* Top 10, WT	(Casadaban and Cohen, 1980)	
E. coli Top10-*csoS4A::csoS1A*	Top10 derivative with CsoS4A and CsoS1A; Ap^R	This study	
E. coli Top10-*csoS2::csoS4A::csoS1A*	Top10 derivative with CsoS2, CsoS4A and CsoS1A; Ap^R	This study	
E. coli Top10-*csoS2::csoS1A*	Top10 derivative with CsoS2 and CsoS1A; Ap^R	This study	Chapter 4
E. coli Top10-*csoS2-C1::csoS4A::csoS1A*	Top10 derivative with CsoS2-C1, CsoS4A and CsoS1A; Ap^R	This study	
E. coli Top10-*csoS2-C2::csoS4A::csoS1A*	Top10 derivative with CsoS2-C2, CsoS4A and CsoS1A; Ap^R	This study	
E. coli Top10-*csoS2-C3::csoS4A::csoS1A*	Top10 derivative with CsoS2-C3, CsoS4A and CsoS1A; Ap^R	This study	
E. coli Top10-*pHluorin2::csoS2-C*	Top10 derivative with CsoS2-C terminal fused with pHluorin2; Ch^R	This study	
E. coli Top10-*cbbLS::csoS2::csoS1A*	Top10 derivative with CbbL/S, CsoS2, and CsoS1A; Ap^R	This study	Chapter 5
E. coli Top10-*cso*	Top10 derivative with *cso* operon; Ch^R	(Chen *et al.*, 2022)	

2.2 *E. coli* genome extraction, plasmid extraction and competent cell preparation

The genomic DNA of *E. coli* was extracted from *E. coli* BL21(DE3) cell culture by using GenElute^{TM} Bacterial Genomic DNA kit (Sigma-Aldrich, UK). Plasmids were extracted by using GeneJET Plasmid Miniprep Kit (Thermo-Fisher, UK). Both processes were followed the standard manufacture protocols.

For electrocompetent cell preparation, 100 mL of *E. coli* cells at OD_{600}=0.4-0.6 were placed

on ice for 10 minutes, and then were harvested by centrifugation at 7,000 g for 10 minutes at

4 °C. The supernatant was discarded, and the pellet was resuspended in 20 mL of cold

autoclave MilliQ-H_2O, and incubated on ice for 10 minutes. This step was repeated twice.

After another centrifugation, the supernatant was discarded, and the pellets was resuspended

in 2 mL of cold 10% (v/v) glycerol solution. 100 µL cells were pipetted into 1.5 mL tube and

placed immediately into the liquid nitrogen. The frozen competent cells transferred to -80 °C

(Gonzales *et al.*, 2013).

2.3 Construction of plasmids

2.3.1 Construction of recombinant of [NiFe]-hydrogenase 1 (HyaAB and HyaAB-EP)

Genes encoding HyaA and HyaB were amplified via using CloneAmp HiFi polymerase

(Takara, Japan) from the BL21(DE3) genome. The *hyaB* gene was ligated to pCDFDuet-1

vector linearized by EcoRI and HindIII to produce the pCDF-*hyaB* vector. A 6x His tag was

added in front of the *hyaA* gene and ligated to the pCDF-*hyaB* vector linearized by NdeI and

XhoI to produce the pCDF-*hyaAB* vector. The C-terminus of full-length CsoS2 served as an

encapsulation peptide (EP) was amplified from the pHnCBS1D plasmid (Addgene, US). The

EP sequence was fused to the 3'-end of *hyaB* followed by ligating to the pCDFDuet-1 vector

linearized by EcoRI and HindIII to produce the pCDF-*hyaB*-EP vector. The *hyaA* gene with a

6x His tag was fused with the C-terminus of CsoS2 followed by ligating to pCDF-*hyaB*-EP

vector linearized by NdeI and XhoI to produce pCDF-*hyaAB*-EP. Each amplified gene

fragment contains its 50-bp upstream and 20-bp downstream sequences to keep native

ribosome-binding site sequences of the genes. The resulting PCR products and linearized

vectors were purified by using DNA gel extraction kit (New England BioLabs, UK)

following the standard manufacture protocol. All vectors were verified by PCR and DNA

sequencing (IDT, US) and transformed into *E. coli* TOP 10 cells. All the PCR primers are

listed in Table 2-2.

Table 2-2. ssDNA Oligonucleotides used in this study (overlap sequence for Gibson assembly are underlined, 6xHis Tags are highlighted).

Primers	Sequences (5' to 3')	Comment
pCDFDuet-HyaB-F	CCATCATCACCACAGCCAGGATCCGATGAGCACTCAGTACGAA	
pCDFDuet-HyaB-R	CTTAAGCATTATGCGGCCGCAAGCTTTAACGCACCTGCACGGA	
pCDFDuet-HyaA-F	TTAAGTATAAGAAGGAGATATACAT ATGCATCACCATCATCACCACAATAACGAGGAAACA	pCDF-hyaAB
pCDFDuet-HyaA-R	GCAGCGGTTTCTTTACCAGACTCGATCATGCCTGTTTATCCTC	
pCDFDuet-HyaB-EP-F	CCATCATCACCACAGCCAGGATCCGATGAGCACTCAGTACGAA	
HyaB-R	TTAACGCACCTGCACGGA	
EP-F-1	GCTGATCTCCGTGCAGGTGCGTTAA ATGACGAGCACCCCAGAG	
pCDFDuet-HyaB-EP-R	GCAGCGGTTTCTTTACCAGACTCGATCAACCGCGCGCGCCGCC	pCDF-hyaAB-EP
pCDFDuet-HyaA-EP-F	TTAAGTATAAGAAGGAGATATACAT ATGCATCACCATCATCACCACAATAACGAGGAAACA	
HyaA-R	TCATGCCTGTTTATCCTC	
EP-F-2	AGGCAATGAGGATAAACAGGCATGAATGACGAGCACCCCAGAG	
pCDFDuet-HyaA-EP-R	GCAGCGGTTTCTTTACCAGACTCGA TCAACCGCGCGCGCCGCC	
pBAD-CsoS4A-CsoS1A-F	TTTTGGGCTAACAGGAGGAATTAACTTTGAGCGTTCAGGCGCA	
CsoS4A-R	TTACTCACCATTCCACTG	
CsoS1A-F	TATTGATCAGTGGAATGGTGAGTAAGGATTGGGAAAGACGAAC	Mini-1 construct
pBAD-CsoS4A-CsoS1A-R	TTTTGTTCTACGTAAGCTTCGAATT TTAGGCTTGTGGCGCCTT	
pBAD-CsoS2-CsoS4A-CsoS1A-F	TTTTGGGCTAACAGGAGGAATTAACATGGGGTCAAACATGCCT	
CsoS2-R	TCAACCGCGCGCGCCGCC	
CsoS4A-CsoS1A-F	TTACTCCGGCGGCGCGCGCGGTTGATTTGAGCGTTCAGGCGCA	Mini-2 construct
pBAD-CsoS2-CsoS4A-CsoS1A-R	TTTTGTTCTACGTAAGCTTCGAATT TTAGGCTTGTGGCGCCTT	
pBAD-CsoS2-CsoS1A-F	TTTTGGGCTAACAGGAGGAATTAACATGGGGTCAAACATGCCT	
CsoS2-R	TCAACCGCGCGCGCCGCC	
CsoS1A-F	TTACTCCGGCGGCGCGCGCGGTTGAGGATTGGGAAAGACGAAC	Mini-3 construct
pBAD-CsoS2-CsoS1A-R	TTTTGTTCTACGTAAGCTTCGAATTTTAGGCTTGTGGCGCCTT	
pBAD-CsoS2-C1-CsoS4A-CsoS1A-F	TTTTGGGCTAACAGGAGGAATTAACATGCTTCCCACTAGTCCACGC	amplification of truncated of CsoS2-C terminus from CTTCCC start
pBAD-CsoS2-C2-CsoS4A-CsoS1A-F	TTTTGGGCTAACAGGAGGAATTAACATGCACGCTGCGCGCGAGTTG	amplification of truncated of CsoS2-C terminus from CACGCT start
pBAD-CsoS2-C3-CsoS4A-CsoS1A-F	TTTTGGGCTAACAGGAGGAATTAACATGAATGCGCGTGTGGTCGAA	amplification of truncated of CsoS2-C terminus from AATGCG start
pBAD33- F	GCACTTATTCAGGCGTAGCAAC	
pBAD33- R	CATATGTATATCTCCTTCTTAAAGT	
pBAD33-pHluorin2-F	ACTTTAAGAAGGAGATATACATATGGTGAGCAAGGGCGAG	
pHluorin2-R	TCACTTGTACAGCTCGTC	pBAD33-pHluorin2-EP construct
pHluorin2-CsoS2-C-F	CGGCATGGACGAGCTGTACAAGTGAATGACGAGCACCCCAGAG	
pHluorin2-CsoS2-C-R	CTGGTTGCTACGCCTGAATAAGTGCTCAACCGCGCGCGCCGCC	
pBAD-CbbLS-CsoS2-CsoS1A-FW	TTTTGGGCTAACAGGAGGAATTAACATGGCAGTTAAAAGTAT	
CbbLS-CsoS2-RW	TCAACCGCGCGCGCCGCC	Mini-CB construct
CsoS1A-FW	TTACTCCGGCGGCGCGGCGGTTGAGGATTGGGAAAGACGAAC	

pBAD-CbbLS-CsoS2-CsoS1A-RW	<u>TTTTGTTCTACGTAAGCTTCGAATT</u>TTTAGGCTTGTGGCGCCTT	
pCDFDuet-seq-F-1	GGATCTCGACGCTCTCCCT	
pCDFDuet-seq-R-1	CGATTATGCGGCCGTGTACAA	hyaAB and hyaAB-EP
pCDFDuet-seq-F-2	TTGTACACGGCCGCATAATC	sequences primers
pCDFDuet-seq-R-2	GCTAGTTATTGCTCAGCGG	
pBAD-seq-F	ATGCCATAGCATTTTTATCC	Mini-shells sequences
pBAD-seq-R	GATTTAATCTGTATCAGG	primers

2.3.2 Construction of recombinant Mini-shells

The genes encoding CsoS4A and CsoS1A from *H. neapolitanus*, together with their 50-bp upstream sequences that include the native promoter and ribosome-binding site sequences, were amplified from the pHnCBS1D plasmid (Addgene, US). This operon was inserted into a pBAD vector linearized by NcoI and EcoRI to produce mini-shell-1 vector. The gene encoding CsoS2 together with its 50-bp upstream sequence was amplified from pHnCBS1D. The nucleotide sequences encoding CsoS4A and CsoS1A were amplified from the mini-shell-1 vector. The gene encoding CsoS2 was fused with the *csoS4A-csoS1A* operon by the 50-bp upstream sequence of the *csoS4A* gene. This operon was inserted into the pBAD vector linearized by NcoI and EcoRI to produce mini-shell-2 vector. The genes encoding CsoS2 and CsoS1A, together with their native promoter and ribosome-binding site, were amplified from pHnCBS1D. The gene encoding CsoS2 was fused with *csoS1A* by the 50-bp upstream sequence of the *csoS1A* gene. This operon was inserted into the pBAD vector linearized by NcoI and EcoRI to produce mini-shell-3 vector. The truncated nucleotide sequences of the C-terminus of CsoS2 and nucleotide sequences of CsoS4A and CsoS1A were amplified from the mini-shell-2 vector. These operons were inserted into the pBAD vector linearized by NcoI and EcoRI to generate mini-shell-4/5/6 vectors, respectively. The truncated nucleotide sequences of the C-terminus of CsoS2 are listed in Table 2-3.

Table 2-3. The truncated nucleotide sequences of the C-terminus of CsoS2.

Protein	Sequences
CsoS2-C1	LPTSPRFNQTGNVQSMGFKNTNQPEQNFAPGEVMPTDFSIQTPARSAQNRITGNDIAP SGRITGPGMLATGLITGTPEFRHAARELVGSPQPMAMAMANRNKAAQAPVVQPEVV ATQEKPELVCAPRSDQMDRVSGEGKERCHITGDDWSVNKHITGTAGQWASGRNPS MRGNARVVETSAFANRNVPKPEKPGSKITGSSGNDTQGSLITYSGGARG
CsoS2-C2	HAARELVGSPQPMAMAMANRNKAAQAPVVQPEVVATQEKPELVCAPRSDQMDRV SGEGKERCHITGDDWSVNKHITGTAGQWASGRNPSMRGNARVVETSAFANRNVPK PEKPGSKITGSSGNDTQGSLITYSGGARG
CsoS2-C3	NARVVETSAFANRNVPKPEKPGSKITGSSGNDTQGSLITYSGGARG

The pHluorin2 gene was amplified from the pME-pHluorin 2 plasmid (Addgene, US) and the nucleotide sequence of the C-terminus of full-length CsoS2 was amplified from pHnCBS1D plasmid. Both genes were inserted into a pBAD33 vector linearized by PCR to generate the pBAD33-pHluorin2-EP vector. All the connections between genes and vectors were performed by Gibson assembly (Gibson *et al.*, 2010, Gibson *et al.*, 2009) (Gibson assembly kit, New England BioLabs, UK). All the PCR fragments were collected via DNA gel extraction (DNA gel extraction kit, New England BioLabs, UK) following the standard manufacture protocol. All of vectors were verified by PCR and DNA sequencing and transformed into *E. coli* TOP 10 cells. All the PCR-used primers were listed in Table 2-2.

2.3.3 Construction of recombinant of Simplified α-carboxysomes

The genes encoding CbbL and CbbS were amplified from the pHnCBS1D plasmid and the genes encoding CsoS2 and CsoS1A were amplified from the mini-shell-3 vector. These gene fragments were inserted into the pBAD vector linearized by NcoI and EcoRI to produce simplified α-carboxysome-expressing vector (Simpl α-CB). All the PCR fragments were collected via DNA gel extraction (DNA gel extraction kit, New England BioLabs, UK) following the standard manufacture protocol. All the vectors were verified by PCR and DNA sequencing and transformed into *E. coli* TOP 10 cells. All the PCR primers are listed in Table 2-2.

2.4 Expression and purification of mini-shells and Simplified α-carboxysomes

2.4.1 Expression of recombinant [NiFe]-hydrogenase (HyaAB and HyaAB-EP) and α-carboxysome shells

For expression of mature [NiFe]-hydrogenases, *E. coli* strains containing HyaAB and HyaAB-EP plasmids were grown in LB medium containing 50 μg·mL$^-$ spectinomycin, 0.03 mM (final) ferric ammonium citrate and 0.03 mM (final) nickel chloride monohydrate, and expression was induced by adding 0.05 mM isopropyl β-D-thiogalactopyranoside (IPTG) at $OD_{600} = 0.6$. For the co-expression of α-carboxysome shells and mature [NiFe]-hydrogenases, two types of empty shell expression plasmids from Liu Lab: Shell-1 including *csoS2*, *csoSCA*, *csoS4AB*, *csoS1ABC* and *csoS1D* genes (Li *et al.*, 2020), and Shell-2 including *csoS2*, *csoS4AB*, *csoS1ABC* and *csoS1D* genes (Li *et al.*, 2020). The HyaAB-EP plasmid was transformed into *E. coli* BL21(DE3) competent cells containing Shell-1 or Shell-2. Cells grown at 37 °C in LB medium supplemented with 50 μg·mL^{-1} spectinomycin, 100 μg·mL^{-1} ampicillin, 0.03 mM (final) ferric ammonium citrate and 0.03 mM (final) nickel chloride monohydrate. At $OD_{600}=0.6$, expression of HyaAB-EP was induced by adding IPTG at the final concentration of 0.05 mM for cell growth at 25 °C. 1 mM (final) L-arabinose was added after adding IPTG 4 hours for shell expression.

2.4.2 Purification of recombinant [NiFe]-hydrogenase and α-carboxysome shells (HyaAB-EP-Shell-1/2)

Cells were harvested by centrifugation at 5, 000 x *g* for 10 minutes at 4 °C. The pellets were resuspended with 20 ml of TMB buffer (10 mM Tris-HCl pH = 8, 10 mM $MgCl_2$, 20 mM $NaHCO_3$) to wash twice. The resuspended cells with 10 % (v/v) CelLytic B cell Lysis reagent (Sigma-Aldrich) were lysed by French press (STANSTED FPG12800 pressure cell

homogeniser) followed the standard recommend process. Cell debris was removed by an initial centrifugation at 10,000 x g for 10 minutes at 4 °C. After centrifugation at 50,000 x g for 30 minutes at 4 °C, the supernatant was discarded and the pellets were gently resuspended in 2 mL TMB buffer (10 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 20 mM $NaHCO_3$) using a soft brush. A 2-minutes spin at 4 °C was performed to remove the insoluble fraction. The soluble pellet fraction was applied on step gradients of 10-50 % sucrose in TMB buffer with 10 % step size for ultracentrifuge 35 min at 105,000 x g. Sucrose gradients were separately collected and stored at 4 °C. A whole purification process was operated under normal air condition. Every sucrose gradient fraction was collected, analyzed via SDS-PAGE, and stored at 4 °C. The supernatant containing free HyaAB-EP was applied on a HisTrap column (HisTrap HP chromatography column, Cytiva, UK) equilibrated in TMB buffer A (10 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 20 mM $NaHCO_3$, 5 mM imidazole). Hydrogenases were eluted by applying a 0-100 % gradient of TMB buffer B (10 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 20 mM $NaHCO_3$, 1 M imidazole) about 25 column volumes. Eluted fractions were collected and stored at 4 °C for further experimental analysis.

2.4.3 Heterogeneously generation of α-carboxysome mini-shells

E. coli TOP 10 strains including mini-shell constructs were cultivated at 37 °C in LB medium containing 100 $\mu g\,mL^{-1}$ ampicillin. When the cells density reached OD_{600} = 0.8, the temperature was reduced to 22 °C, and protein expression was induced via adding L-arabinose (1 mM, final) and cells were grown overnight.

The cultures were harvested by centrifugation at 5, 000 x *g* for 10 minutes at 4 °C. The pellets were resuspended with 30 ml of TEMB buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 10

mM $MgCl_2$, 20 mM $NaHCO_3$) to wash twice. The resuspended pellets supplemented with 10% (v/v) CelLytic B cell lysis reagent (Sigma-Aldrich) and 0.1% protein inhibitor cocktail (PIC, 10x) (Sigma-Aldrich). The resuspended pellets were lysed by sonication (30s ON, 20s OFF, 6 cycles). Cell debris was removed by centrifugation at 27,000 x *g* for 30 min at 4 °C. Following the supernatants were loaded on 30% sucrose cushion (30% sucrose in TEMB buffer) and were ultra-centrifuged at 250,000 x *g* for overnight at 4 °C. The supernatants were kept as control group and the pellets were gently resuspended in 5 ml TEMB buffer using a soft brush. A 2-minute centrifugation at 14,000 x g, 4 °C was applied to remove insoluble pellets. The soluble pellets fractions were loaded onto step sucrose gradients of 10-50 % sucrose in TEMB buffer with 10 % step size, followed by ultra-centrifugation at 70,000 x g for overnight at 4 °C. Sucrose gradient fractions were collected, analyzed via SDS-PAGE, and stored at 4 °C. The shell protein-enriched sucrose fractions (40% sucrose fractions) were applied on a HiTrap Q column (HiTrap Q FF anion exchange chromatography column, Cytiva, UK) equilibrated in TEMB buffer A (10 mM Tris-HCl pH = 8, 1 mM EDTA, 10 mM $MgCl_2$, 5 mM $NaHCO_3$, 5 mM NaCl). The mini-shells were eluted by applying a 0 - 40% gradient of TEMB buffer B (10 mM Tris-HCl pH = 8, 1 mM EDTA, 10 mM $MgCl_2$, 5 mM $NaHCO_3$, 1 M NaCl) about 25 column volumes. Eluted fractions were collected and stored at 4 °C for further experimental analysis.

2.4.4 Expression of pHluorin2 and generation of mini-shells with encapsulated pHluorin2

For the expression of pHluorin2 proteins, cells containing the pBAD33-pHluorin2-EP vector were grown in LB medium with 100 $\mu g\,mL^{-1}$ chloramphenicol and expression was induced by adding 1 mM (final) L-arabinose at OD_{600} = 0.6. For the generation of mini-shell-

pHluorin2, *E. coli* strains containing the pBAD33-pHluorin2-EP vector and mini-shell-1 or mini-shell-2 vectors were grown in LB medium containing 100 μg mL^{-1} ampicillin and 100 μg mL^{-1} chloramphenicol. Both proteins expression was induced by adding 1 mM (final) L-arabinose at OD_{600} = 0.6, and then the cell grown at 22 °C overnight. The purification process was same with the mini-shells generation mentioned above.

2.4.5 The heterogeneously generation of Simpl α-carboxysomes

Strains of *E. coli* TOP 10 including Simpl α-carboxysome expressing construct were cultivated at 37 °C in LB medium containing 100 μg mL^{-1} ampicillin. When the cells density reached OD_{600} = 0.8 temperature was reduced to 22 °C, and protein expression was induced via adding L-arabinose (1 mM, final) and cells were grown overnight.

Cells were harvested by centrifugation at 5, 000 x *g* for 10 minutes at 4 °C. The pellets were resuspended with 20 ml of TEMB buffer supplemented with 10 % (v/v) CelLytic B cell lysis reagent (Sigma-Aldrich) and 0.1% PIC (10x) (Sigma-Aldrich) and were washed twice, and were lysed by sonication (30s ON, 20s OFF, 6 cycles). Cell debris was removed by centrifugation at 10, 000 x *g* for 10 minutes at 4 °C. The supernatant was centrifuged at 50, 000 x *g* for 30 minutes at 4 °C. Following the supernatant was discarded and the pellet was resuspended in 3 mL TEMB buffer with a short spin at 2 minutes to remove insoluble pellet. The resuspended pellet was applied onto step gradients of 10-50 % sucrose in TEMB buffer with 10 % step size, followed by ultra-centrifugation at 105, 000 x g for 35 minutes at 4 °C. Sucrose fractions were collected and stored at 4 °C for next experiments.

2.5 Enzyme activity assays

2.5.1 Hydrogenase activity assay

2.5.1.1 *In vivo* H$_2$-evolution assay

Strains expressing free HyaAB, free HyaAB-EP, HyaAB-EP-Shell-1 and HyaAB-EP-Shell-2 were grown 37 °C aerobically in 200 mL flask with 50 mL LB medium supplemented with 0.03 mM (final) ferric ammonium citrate and 0.03 mM (final) nickel chloride monohydrate, 50 μg·mL^{-1} spectinomycin, and/or 100 μg mL^{-1} ampicillin. At OD$_{600}$ = 0.6, 30 mL culture was transferred to 50 mL falcon tube with sealed rubber closure (Sigma-Aldrich). The pure N$_2$ (100%, v/v) into the falcon tube 5 min via using needles (0.88 x 120 mm, BRAUN) for degassing, followed by addition of 1 mM IPTG. For aerobic treatment, 30 mL culture was transferred to falcon tube with sealed rubber closure without N$_2$ degassing. 1 mM L-arabinose was added 4 hours after the addition of IPTG for strains containing HyaAB-EP-Shell-1 and HyaAB-EP-Shell-2. Cells were grown at 25 °C for 16 hours with shaking following induction. The produced H$_2$ gas was measured using a gas-syringe and was detected using a Bruker 450-GC gas chromatography. For each experiment, at least three biological replicates were examined.

2.5.1.2 *In vivo* dissolved oxygen (DO) measurement

A polarographic DO probe (New BrunswickTM BioFlo/CellGen 115 Fermentor, Eppendorf) was used to measure the level of dissolved oxygen (DO) in strains expressing free HyaAB-EP, HyaAB-EP-Shell-1 and HyaAB-EP-Shell-2 for in vivo activity assays. Oxygen saturated LB medium (100 % DO) and sodium dithionite treated LB medium (0% DO) were used to calibrate the polarized polarographic DO probe, respectively. The DO levels of strains expressing free HyaAB-EP, HyaAB-EP-Shell-1 and HyaAB-EP-Shell-2 were determined

before adding IPTG and every four hours after adding IPTG. For each experiment, at least three biological replicates were examined.

2.5.1.3 *In vitro* H_2-evolution assay

Cells were grown aerobically at 37 °C in 2 L flask containing 800 mL LB medium with corresponding antibiotics, 0.03 mM (final) ferric ammonium citrate and 0.03 mM (final) nickel chloride monohydrate, 50 μg·mL^{-1} spectinomycin, and/or 100 μg mL^{-1} ampicillin. At OD_{600} = 0.6, HyaAB expression was induced by the addition of IPTG (1 mM). For strains contain HyaAB-EP-Shell-1 and HyaAB-EP-Shell-2, 1 mM IPTG was added for HyaAB-EP expression. After 4 hours, 1 mM (final) L-arabinose was added for shell expression. Cells were grown at 25 °C with shaking following induction.

The whole purification process was performed under aerobic conditions, followed by *in vitro* hydrogenase H_2-evlution activity assays. In a 10-mL sealed vial (Agilent), protein sample (0.5 mL, ~ 10 mg/mL) in TMB buffer was supplemented anaerobically with 2 mL nitrogen degassed methyl viologen (MV) (50 mM in TMB buffer, final) and 0.5 mL sodium dithionite (500 mM in TEM buffer, final). The reaction of hydrogenase activity was performed in an anaerobic glove bag (Sigma, UK). The samples were incubated at 37 °C for 20 minutes and hydrogen evolution was measured by using a gas-syringe and was detected by using a Bruker 450-GC gas chromatograph. For each experiment, at least three biological replicates were examined.

2.5.1.4 Hydrogenase kinetic assay

0.5 mL protein sample (~10 mg mL^{-1}) in TMB buffer and 2 mL nitrogen degassed MV (1.5-200 mM in TMB buffer, final) and 0.5 mL sodium dithionite (500 mM in TMB buffer, final) were anaerobically added in 10 mL sealed vial (Agilent) for a total reaction of 3 mL. The vials were incubated at 37 ℃ for 20 minutes and hydrogen evolution was measured via using a Bruker 450-GC gas chromatography. Hydrogenase activity was analyzed using a standard Michaelis-Menten model under various concentrations of MV. In addition, hydrogen evolution was measured every 20 minutes by gas chromatography as a function of the times. In this process, 0.5 mL protein samples (~10 mg mL^{-1}) mixed with 2 mL nitrogen degassed MV (50 mM in TMB buffer, final) and 0.5 mL sodium dithionite (500 mM in TMB buffer, final) in 10 mL sealed vials at 37 ℃. For each experiment, at least three biological replicates were examined.

2.5.1.5 Heat treatment and stability of hydrogenase activity

For heat treatment, the protein samples (~10 mg mL^{-1}) were heat treated at 65 ℃ for 20 minutes. For stability, the protein samples (~10 mg mL^{-1}) were kept at 4 ℃ for 7 days. Then the samples were incubated with 2 mL nitrogen degassed MV (50 mM in TMB buffer, final) and 0.5 mL sodium dithionite (500 mM in TMB buffer, final) in 10 mL sealed vials at 37 ℃ for 20 minutes and was then subjected to hydrogen evolution assays. For each experiment, at least three biological replicates were examined.

2.5.1.6 Oxygen exposure treatment

0.5 mL free HyaAB-EP and HyaAB-EP-Shell-1/2 samples (~10 mg mL^{-1}) were exposed to the air at 4 ℃ for 24 hours, respectively. Following, the samples were transferred into a 10

mL sealed vials and were degassed 5 minutes by pure nitrogen. The samples were incubated with 2 mL nitrogen degassed MV (50 mM in TMB buffer, final) and 0.5 mL sodium dithionite (500 mM in TMB buffer, final) at 37 °C for 20 minutes and was then subjected to hydrogen evolution assays. For each experiment, at least three biological replicates were examined.

2.5.1.7 Gas chromatography

Gas samples (1 mL) were taken with a gas-tight syringe and run on a Bruker 450-GC gas chromatograph equipped with a molecular sieve 13 × 60-80 mesh 1.5 m × 1/8 inch × 2 mm ss column at 50 °C with an argon flow of 40.0 mL min^{-1}. Hydrogen was detected with a thermal conductivity detector referencing against standard gas with a known concentration of hydrogen.

2.5.2 Rubisco activity assays

The protein concentration of purified Simpl α-carboxysomes was determined by Bradford assays. Rubisco activity assays were carried out as previously described (Chen *et al.*, 2021, Faulkner *et al.*, 2017, Sun *et al.*, 2016). In brief, 5 µL purified mini α-carboxysomes in the Rubisco assay buffer (100 mM EPPS, pH 8.0, 20 mM $MgCl_2$, 3.5 mM ATP) were aliquoted into scintillation vials containing $NaH^{14}CO_3$ at a final concentration of 25 mM and then incubated at 30 °C for 2 minutes. The carbon fixation process was operated by adding the D-ribulose 1,5-bisphosphate sodium salt hydrate (RuBP, Sigma-Aldrich, UK) to the purified samples at a range of concentrations (0.05 - 1mM). After 5 minutes of incubation, the reaction was stopped by adding 10% (v/v) formic acid to the mixture. Unfixed $NaH^{14}CO_3$

was removed by the samples were dried on heat blocks at 95 °C about 30 minutes, and the fixed pellet were resuspended with distilled water and 2 mL the scintillation cocktail (Ultima Gold XR; Perkin-Elmer, USA). Following, radioactivity measurements were operated by the scintillation counter (Tri-Carb; Perkin-Elmer, USA). The counts per minute was employed to calculate the quantity of fixed ^{14}C based on the standard curve and then converted to the total CO_2 fixation rates. For each experiment, at least three biological replicates were examined.

2.6 SDS-PAGE and immunoblot analysis

Protein samples were prepared and then mixed with 4x SDS-PAGE sample loading buffer (250 nM Tris pH 6.8, 8% SDS, 0.2% bromophenol blue, 40% glycerol, 20% mercaptoethanol). After 95 °C heating for 10 minutes, samples were centrifuged at 12,000 x g for 2 minutes and then loaded on 15% SDS-PAGE gels to analyze their composition. About 75 μg proteins was loaded in each SDS-PAGE gel well. 3 μL of unstained protein ladder (10-250 kD from NEB) was loaded as marker. SDS-PAGE was run at 200 V 45 minutes with transfer buffer (25 mM Tris, 192 mM glycine and 1% SDS). The gels were stained through Coomassie blue stain buffer (0.25% Coomassie Brilliant Blue R-250, 20% methanol, 10% acetic acid) and distained by destaining buffer (20% methanol, 10% acetic acid).

Immunoblot analysis was performed by loading 30 μg total proteins into each well. SDS-PAGE gels were transferred to a PVDF membrane (Bio-Rad). Following the transfer, membranes were immunoblotted analysed by using primary mouse monoclonal anti-Histag (Life Technologies, 69-74-9, UK), primary rabbit polyclonal anti-GFP (Agrisera, AS204443, dilution 1:2,000, US), anti-CsoS1A/B/C (Agrisera, AS142760, dilution 1:5,000, US), anti-

RbcL (Agrisera, AS03037A, dilution 1:3,000, US). For primary anti-body detection, horseradish peroxidase-conjugated goat anti-mouse IgG (Agrisera, AS111772 dilution 1:5,000, US) or anti-rabbit IgG (Agrisera, AS09602, dilution 1:10,000, US) were as the secondary antibody. After immunoblotting, membranes were washed with TBS buffer (10 mM Tris-HCl pH 7.4, 150 mM NaCl) and TBST buffer (10 mM Tris-HCl pH 7.4, 150 mM NaCl, 2% Tween-20) and signals were visualized using a Bio-Rad chemiluminescence kit and images were obtained via ImageQuant LAS 4000. The protein quantification was calculated by using ImageJ. For each experiment, at least three biological replicates were examined.

2.7 Transmission electron microscopy (TEM)

Isolated protein samples (1-2 mg mL^{-1} total protein) were stained on carbon grids (Carbon Films on 300-mesh Grids Copper, Agar Scientific) for 40s, followed by staining with 2% (v/v) uranyl acetate (Sigma-Aldrich). The carbon grids were then washed with distilled water and then dried using 0.2 μm filter paper. Images were recorded using FEI Tecnai G2 Spirit Bio TWIN transmission electron microscope equipped with a Gatan Rio 16 camera. ImageJ was used to analyse images. Statistical analysis was calculated by using Student's t-test.

2.8 Dynamic light scattering (DLS) analysis

Briefly, 1 mL (5-10 mg mL^{-1} total protein) of mini-shells or Simpl α-CB samples were analysed by Dynamic light scattering (Malvern DLS ZetaSizer) to measure the size distribution and average size of the particles. For each experiment, at least three biological replicates were examined.

2.9 Cryo-EM analysis

2.9.1 Cryo-EM data collection

Cryo-EM work in this PhD thesis was performed in collaboration with Professor Peijun Zhang at eBIC. The cryo-EM sample grids were prepared using Vitrobot. Three microliters of mini-shell samples were applied to glow-discharged Quantifoil grids (R2/1) and blotted with filter paper for 3 seconds before plunge freezing with liquid nitrogen-cooled ethane. The temperature is set to 20 °C and humidity at 100% during plunge freezing. The micrographs were taken using Krios microscope equipped with a Gatan K3 director electron camera and Bioquantum energy filter or Falcon 4 with Selectris X energy filter. The energy filter slit was set to 20 eV.

2.9.2 Data processing

Data processing was performed mainly using Relion3. The micrograph movies were gain normalized and motion corrected with MogtionCor2. Contrast transfer function (CTF) was estimated using CtfFind4. Two different sized mini-shells were observed on the raw micrographs and processed independently.

For the large sized particles, a subset of particles was picked manually in Relion to generate initial 2D class averages for auto-picking. Two rounds of 2D classification were performed, resulting a final dataset with 143,769 particles. An ab initial model was generated with I1 symmetry in Relion. 3D auto refine was carried out with the initial model reconstructed with I1 symmetry in Relion, which resulted in a density map with a mixed handedness. The resulted refined particles dataset was 3D classified into 10 classes, skipping alignment, which revealed two major classes, with opposite handedness. These two classes of particles were

refined separately with per-particle CTF refinement and polishing. To combine the two classes, particles with opposite handedness were inverted by changing the refined Euler angle in Relion star file (phi and tilt). The two half maps were reconstructed using Relion_reconstruct, with CTF and eward sphere correction. The final combined density maps were masked, and b-factor was sharpened with Relion_postprocess, which resulted in a final map at 2.1 Å resolution.

The small mini-shells were processed in a similar way to large Shell, except for particle picking step. A small number of small shell particles were manually picked in EMAN2.3 to train neural network, which was subsequently used to pick against the whole dataset. The coordinates of particles (EMAN2 box files) were imported into Relion for further processing in the same way as large shells. Similarly, a small portion of particles were found in opposite handedness after 3D refinement and classification, which was then corrected by updating the Euler angles in the Relion star file as above. The final map after per-particle CTF refinement and polishing is at 2.5 Å resolution.

2.9.3 Model building and refinement

Initial models from crystal structures of hexamer (PDB 2EWH) and pentamer (PDB 2RCF). For the large shell, CsoS2 was traced manually into the density map in Coot. At this resolution, the side chains of CsoS2 can be unambiguously placed (Extended Data Figx). One asymmetric unit of the icosahedral shell with additional surrounding subunits were further refined in Phenix.refine. Water molecules were placed into density manually. The final icosahedral models were reconstructed in Chimera with symmetry command with I1

symmetry. Model alignment and comparison were performed in Chimera and the figures are

rendered in Chimera.

Chapter 3 Developing a carboxysome-based nanobioreactor for hydrogen production

3.1 Introduction

Currently, energy and environment are two hot points related to the development of the economy and society around the world (Khan *et al.*, 2018; Li *et al.*, 2018). About 80% energy consume from fossil fuels that are not renewable and environmentally friendly (Zhu *et al.*, 2020). Renewable and clean energy is a potential possibility to take place of the fossil fuels. Hydrogen (H_2) was considered as an alternative based on that H_2 is generated only from water without toxic by-products and H_2 has higher energy than other energy fuels, for example, hydrogen has about three times as much energy as gasoline (Hallenbeck, 2009; Mudhoo *et al.*, 2011). Biohydrogen is produced through biological routes such as biophotolysis (Akkerman *et al.*, 2002; Hallenbeck, 2009; Torzillo *et al.*, 2009; Dasgupta *et al.*, 2010), fermentation (Guo *et al.*, 2014; Rosa *et al.*, 2014; Han *et al.*, 2015; Uyar *et al.*, 2015), microbial electrolysis (Heidrich *et al.*, 2013; Modestra *et al.*, 2015), which will decrease the consumption of fossil fuels and limit greenhouse gas emissions. However, biophotolysis and microbial electrolysis required high light intensity (Meherkotay and Das, 2008) and high voltage (Catal *et al.*, 2015), and fermentation showed a slow H_2-producing rate (Venkata Mohan *et al.*, 2007). Therefore, a growing amount of focus has been placed on developing novel hydrogen-producing strategies that produce hydrogen.

Synthetic nanoparticles (NPs) (Han *et al.*, 2011; Nasr *et al.*, 2015; Lin *et al.*, 2016; Mishra *et al.*, 2018) like FeO NPs and Ni NPs are known to improve the amount of biohydrogen produced by substantially boosting the activity of important enzymes involved in the biohydrogen generation process, such as [FeFe]- or [NiFe]-hydrogenase. They also engage in the usage of feedstock for the enhancement of hydrogen production. Recent research found that FeO NPs were employed in the process of producing biohydrogen of different types of bacteria (e.g. *Enterobacter aerugenes* and *Clostridium butyricum*) that are known to produce

[FeFe]-hydrogenase, which resulted in the significantly improved hydrogen (Yang and Shen, 2006; Han *et al.*, 2011; Zhang *et al.*, 2018). These studies suggested that the conductivity of FeO NPs enhances the electron transfer rate efficiently, allowing the metal centre of [FeFe]-hydrogenase to collect an appropriate number of electrons to promote hydrogen generation. Similarly, Ni nanoparticles co-operated with [NiFe]-hydrogenases also could obtain biohydrogen production (Mullai *et al.*, 2013). Although the yield of biohydrogen production improved significantly through interaction between NPs and hydrogenases, the NPs' concentration affects the production (Han *et al.*, 2011). Moreover, the higher concentration of NP's will damage the regulatory system of microorganisms and change bacterial physiology (Lin *et al.*, 2016).

Scaffolding platforms (Menzel *et al.*, 2014; Wilkerson *et al.*, 2018) also using hydrogenase to obtain enormous biohydrogen. Recently, a synthetic scaffolding platform, named the protein-based capsid of bacteriophage P22, free [NiFe]-hydrogenase from *E. coli* was encapsulated within the capsid of bacteriophage P22 by a truncating scaffold protein led to a higher concentration of [NiFe]-hydrogenases in a microenvironment to rise the yield of biohydrogen compared with free [NiFe]-hyrogenase (Jordan *et al.*, 2016). This synthetic scaffolding platform not only ensures packaging and condensing of enzymes, but it also allows for regulating enzymatic activities. However, it is important to keep in mind that the virus capsids do not possess permeability.

A virus-like proteinaceous organelle, named carboxysomes, could be a novel platform for using hydrogenase to promote the yield of biohydrogen. The reasons are i) carboxysomes contain a thin selective permeable protein shell, made of various shell proteins, to allow metabolic substrates and products diffuse inside and outside the carboxysomes (Kerfeld, 2005;

Tanaka *et al.*, 2008a; Faulkner *et al.*, 2017, Sun *et al.*, 2019); ii) carboxysomes shell encapsulates ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO) and carbonic anhydrase (CA) to improve carbon dioxide (CO_2) fixation, which demonstrates the good mechanism of enzyme encapsulation and condensation (Rae *et al.*, 2013; Kerfeld *et al.*, 2018; Huang *et al.*, 2020); iii) carboxysomes shell also could limit O_2 input and reduce CO_2 output (Dou *et al.*, 2008); iv) carboxysomes assembly pathway and purification process are more clear than others. Recently, we have generated recombinant α-carboxysome shells that encapsulate the oxygen-sensitive [FeFe]-hydrogenases as a new nanobioreactor for biohydrogen production (Li *et al.*, 2020).

Here, we developed this strategy by using another hydrogenase, [NiFe]-hydrogenase 1 from *E. coli*, to optimize the biohydrogen-produced system of α-carboxysome shells. Compared with [FeFe]-hydrogenases whose activity can be easily lost in the presence of O_2 (Melis and Happe, 2001; Ghirardi *et al.*, 2009; Melis, 2009; Tamagnini *et al.*, 2007), [NiFe]-hydrogenase 1 is an oxygen-tolerant enzyme (Menon *et al.*, 1991) and is involved in hydrogen recycling (Gary, 1985; Sawers and Boxer, 1986; Forzi and Sawers, 2007). It has previously been shown that recombinant [NiFe]-hydrogenases could be heterologously expressed in *E. coli* and that this can result in hydrogen generation under microaerobic condition (Jaoon YH Kim, 2010). Moreover, a truncating scaffolding protein fused to the C-terminal of the the large (HyaB) and small (HyaA) subunits of the recombinant [NiFe]-hydrogenase of *E. coli*, respectively, which was encapsulated encase the bacteriophage P22 capsids for boosting the hydrogen production (Jordan *et al.*, 2016) In this work, high H_2 production was obtained *in vitro* compared with free hydrogenase, confirming the protein-based capsids nanoparticles was also suitable to use [NiFe]-hydrogenase for the hydrogen production. In our investigation, [NiFe]-hydrogenase of *E. coli* was encapsulated inside the nanobioreactor, α-carboxysomes

shell, by employing the encapsulation peptide the C-termin of CsoS2 of α-carboxysomes from *Halothiobacillus neapolitanus*. The higher hydrogen production was obtained *in vitro* based on aerobic process of protein purification as well as *in vivo* based on *E. coil* aerobically expressed [NiFe]-hydrogenase within α-carboxysome shells.

3.2 Results

3.2.1 The generation of recombinant [NiFe]-hydrogenase-1

The [NiFe]-hydrogenase-1 in *E. coli* is encoded by the *hya* operon that consists of *hya*A/B/C/D/E/F genes (Menon NK, 1991). HyaA is the small core enzyme subunit including a [Fe-S] cluster and is involved in electron transfer, whereas HyaB is the large core enzyme subunit includes a biometallic center and is involved in enzyme activation. The *hyaA* and *hyaB* genes were firstly cloned into the linearized vector pCDFDuet by EcoRI and HindIII for *hyaB* as well as NdeI and XhoI for *hyaA* by under the control of the T7 promoters to generate the *hyaAB* vector with 6×His tags that fused to the N-terminal of *hyaA* and *hyaB*, respectively (Figure 3-1a and b). Previous studies confirmed that the C-terminus of full length CsoS2 could serve as an encapsulation peptide (EP) to recruit cargoes within the α-carboxysome shells (Li *et al.*, 2020). The C-terminus of CsoS2 was fused to the C-termini of HyaA and HyaB to generate the *hyaAB*-EP vector (Figure 3-1a and b). The IPTG-induced expression of both recombinant [NiFe]-hydrogenase plasmids in *E. coli* were achieved via adding IPTG at OD_{600} of up to 0.6 under 25 ℃ for 16 hours. Whole cells were collected and immunoblot analysis revealed that the HyaA/B and HyaA-EP/HyaB-EP were heterologously expressed in *E. coli* by employing an anti-Histag antibody (Figure 3-1c), consistent with previous findings (Menon, 1991; Jaoon YH Kim, 2010; Trchounian *et al.*, 2012a; Wulff *et al.*, 2016).
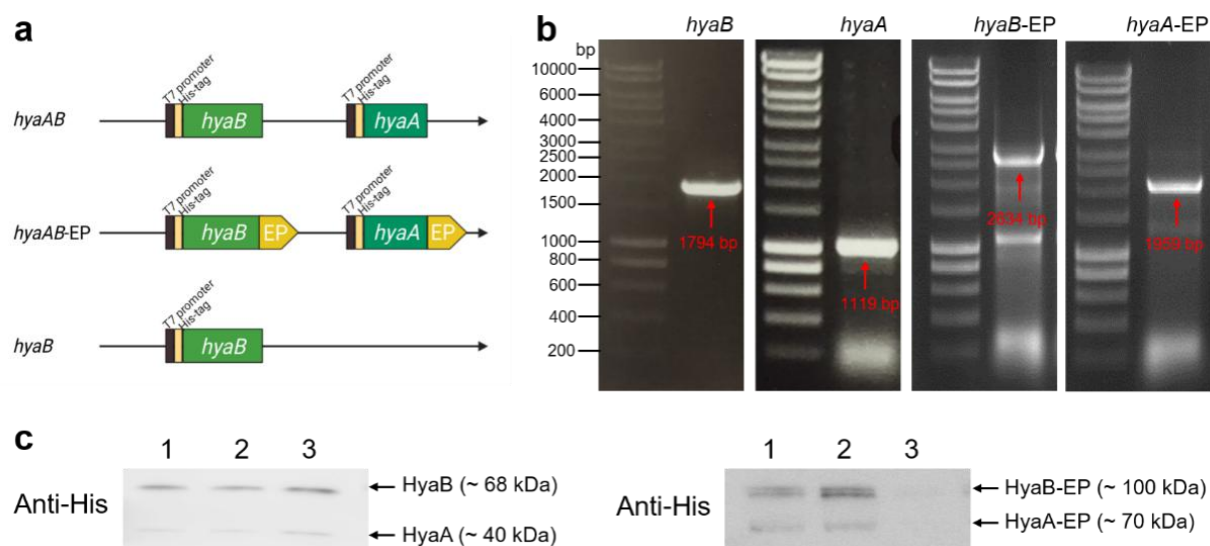
Figure 3-1. Construction of recombinant [NiFe]-hydrogenase-1 vectors. (a) The genetic organizations of the pCDFDuet-*hyaAB*/-*hyaAB*-EP/-*hyaB* constructs. EP is encapsulated peptide (the C-terminus of full length CsoS2, which is shown on the Figure 1-15a). (b) PCR products of *hyaA*, *hyaB*, *hyaA*-EP, *hyaB*-EP of the recombinant [NiFe]-hydrogenase-1 vectors. (c) The large (HyaB) and small (HyaA) subunit of [NiFe]-hydrogenase-1 were determined by using anti-Histag antibody (Life Technologies, USA). The numbers show three single colonies screened.

3.2.2 CsoS2-C mediates the encapsulation of [NiFe]-hydrogenase into the shells

We have recently proved that the oxygen-sensitive [FeFe]-hydrogenase could be encapsulated within the α-carboxysome shell mediated by CsoS2 C-terminus and showed improved H_2 evolution (Li *et al.*, 2020). The oxygen-tolerant [NiFe]-hydrogenases can be generated under aerobic conditions (Kim, 2010), and have potential application in fuel cells. To determine H_2 evolution activity of [NiFe]-hydrogenases in shell encapsulation, the *hya*AB-EP plasmid was transformed into strains containing either the *cso-1* or *cso-2* (Shell-1/2) constructs (Figure 3-2a) which were created in our previous work (Li *et al.*, 2020). Expression of the *hya*AB-EP plasmid was induced by the addition of IPTG for 4 hours before the expression of the shell induced by L-arabinose to ensure maturation and activation of hydrogenases prior to shell formation and encapsulation. Recombinant Shell-1 and Shell-2 with incorporated HyaAB-EP were purified from *E. coli*. After 50,000× g centrifugation, the

components of Shell-1-HyaAB-EP and Shell-2-HyaAB-EP including shell proteins and HyaAB-EP were detected in the pellet, whereas unencapsulated HyaAB-EP were only present in the supernatant (Figure 3-2b), implicating the formation of Shell-HyaAB-EP assemblies.
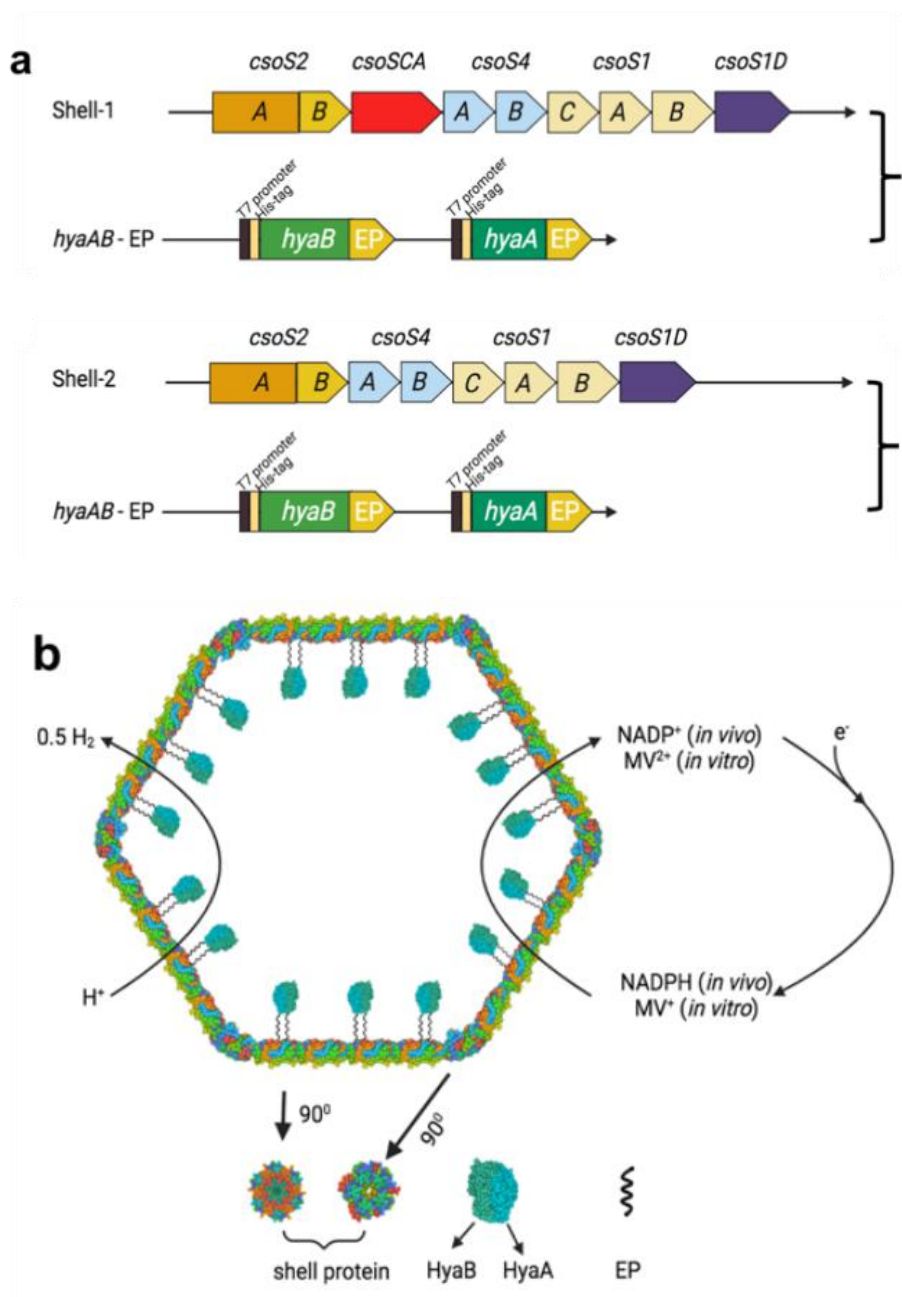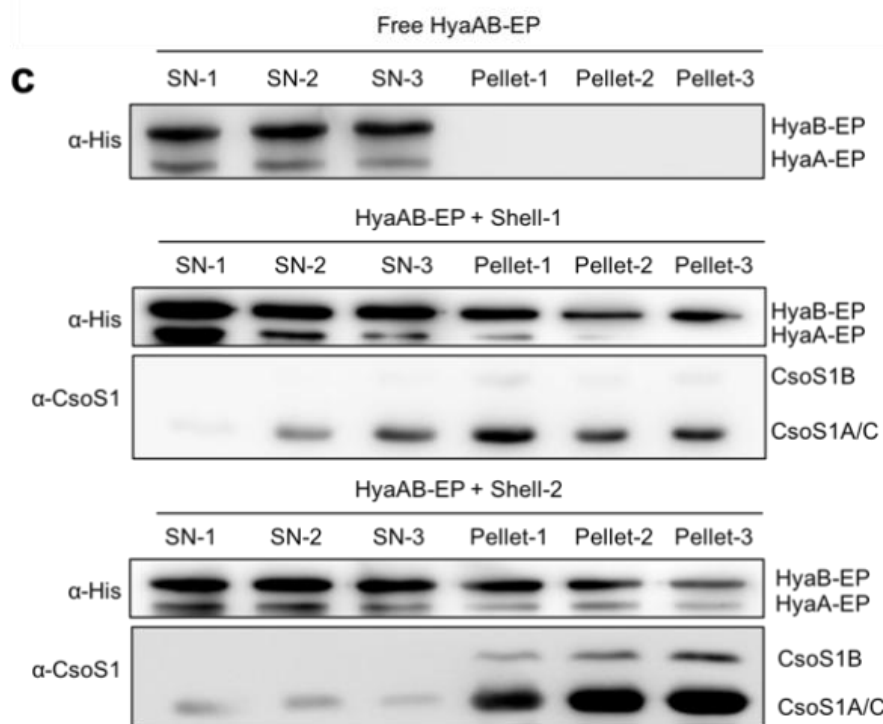
Figure 3-2. The determination of the encapsulation of recombinant [NiFe]-hydrogenase-1 within two types of α-carboxysome shells. (a) The schematic of the two types of shells encapsulating HyaAB-EP. The diagram was drawn by www.biorender.com. (b) The schematic of the encapsulated HyaAB-EP in Shells-HyaAB-EP nanoreactor systems. Hydrogen generation (Figure 3-6b, 3-10a, and 3-10b) by the nanoreactor systems were evaluated utilizing NADPH from cells as an electron source for *in vivo* tests and methyl viologen (MV+) as an electron donor, chemically reduced by sodium dithionite for *in vitro* tests. (c) Immunoblot analysis of the supernatants and pellets of cell extracts after 50,000 g centrifugation from strains producing HyaAB alone or co-expressing HyaAB with Shell-1 and Shell-2 (including three biological repeats) using anti-His antibody and anti-CsoS1 antibody, respectively.

We further isolated HyaAB-EP-Shell-1/2 from the pellet samples using sucrose gradient ultracentrifugation (Figure 3-3a). The empty Shell-1 as a control were also isolated by using the same process with HyaAB-EP-Shell-1/2. Immunoblot analysis of each sucrose fraction revealed that HyaB-EP and HyaA-EP encapsulated inside the shells were detected in the 10%−50% sucrose fractions (Figure 3-3b). In contrast, free HyaB-EP and HyaA-EP were undetectable in all sucrose fractions, suggesting that the EP mediated HyaB and HyaA encapsulation into both shells. EM of the 20% sucrose fractions further confirmed the incorporation of *hya*AB-EP in the shell interior (Figure 3-3c).
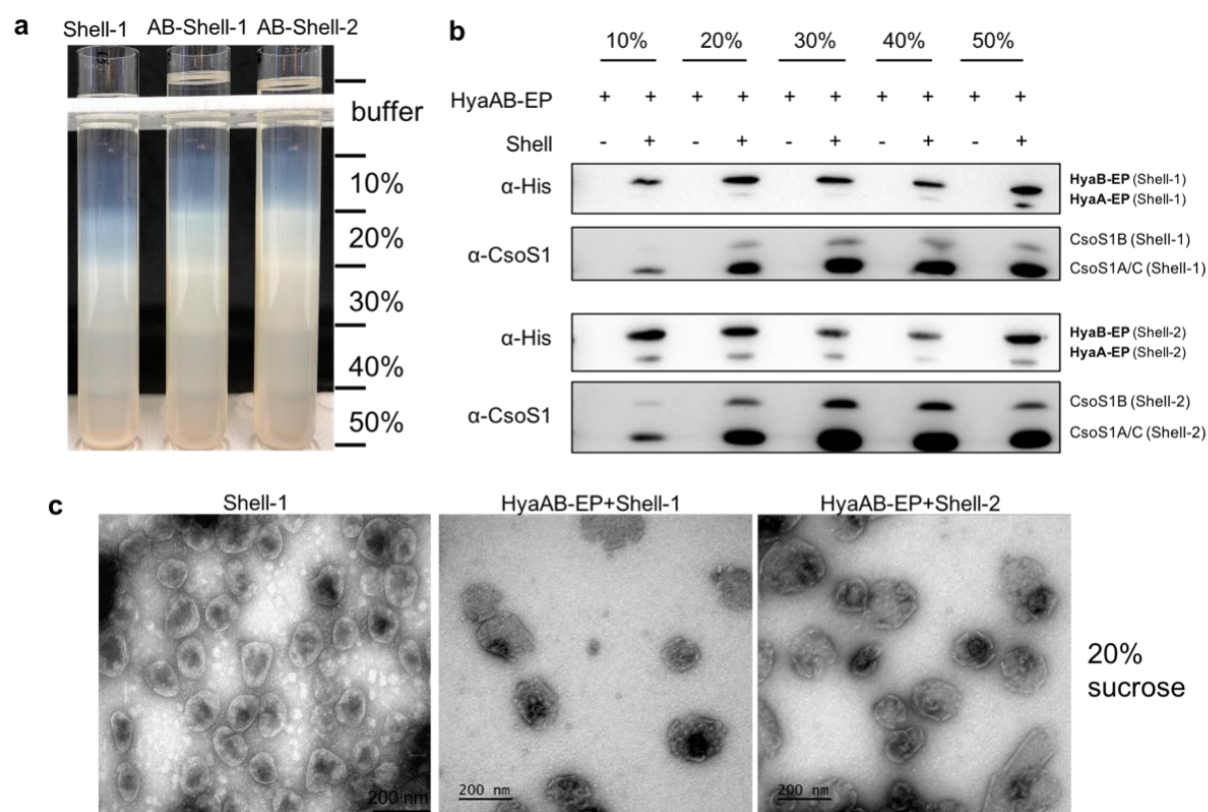
Figure 3-3. The purification of HyaAB-EP-Shell-1/2. (a) The sucrose gradients ultracentrifuge of HyaAB-EP-Shell-1/2. Sucrose gradients were set from 10% to 50% based on previous studies (Li *et al.*, 2020). (b) Immunoblot analysis of HyaAB-EP-Shell-1/2 confirmed the presence of HyaAB-EP and shell proteins in the samples purified from sucrose gradients ultracentrifugation. (c) The images of electron microscopy of empty Shell-1, HyaAB-EP-Shell-1 and HyaAB-EP-Shell-2.

3.2.3 *In vivo* hydrogen production of [NiFe]-hydrogenases encapsulated within carboxysome shells

The H_2 production activity of the generated *E. coli* cells expressing HyaAB-EP-Shell-1, HyaAB-EP-Shell-2 and free HyaAB-EP were assayed using endogenous NADPH in *E. coli* cells as the electron source (Figure 3-2a and 2b). The cells were grown and induced for 16 hours under aerobic or anaerobic conditions. The generated H_2 production was determined by gas chromatography.

It has been reported that the catalytic activity of the enzymes might be affected by the tags fused with them (Lee *et al.*, 2016). To investigate whether the encapsulation peptide has a negative effect on the catalytic activity of NiFe-hydrogenase, we transformed *hyaAB* and *hyaAB*-EP individually into the BL21(DE3) strain and induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) under either anaerobic or aerobic conditions for 16 hours (Figure 3-4a). Under anaerobic conditions, the H_2-productivity of cells expressing *hyaAB* is 1824.63 ± 161.89 nmol L^{-1} h^{-1} (mean $\pm$ standard deviation (s.d.), $n = 3$) at pH 7, similar to that of cells expressing *hya*AB-EP (1780.13 ± 197.87 nmol L^{-1} h^{-1}) at pH 7 (**$p = 0.21$, two-tailed unpaired *t*-test) (Figure 3-4b). Similarly, cells producing *hya*AB (285.15 ± 23.03 nmol L^{-1} h^{-1} at pH 7) have a minor advantage in H_2 production over cells producing *hya*AB-EP (269.16 ± 15.19 nmol L^{-1} h^{-1} at pH 7) under aerobic conditions (**$p = 0.49$, two-tailed unpaired *t*-test). These results suggest that EP fused at the C-terminal of HyaA and HyaB has negligible influence on the catalytic activity of [NiFe]-hydrogenases. Moreover, both cell types could produce H_2 under aerobic conditions (Figure 3-4b), indicating that an anoxic environment facilitates the catalytic activity of [NiFe]-hydrogenases though it has been classified as oxygen-tolerant hydrogenases (Peters *et al.*, 2015).
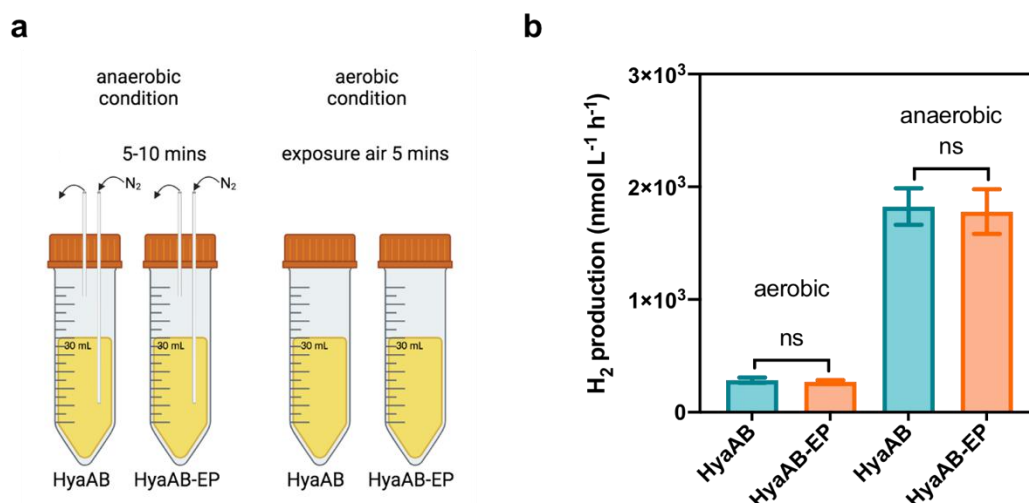
Figure 3-4. No significant difference in hydrogen productivity between HyaAB and HyaAB-EP. (a) The schematic diagram of hydrogen production under aerobic and anaerobic conditions *in vivo*. For the aerobic condition, 30 ml of culture was exposed to air about 5 mins then closing the seal cap. For the anaerobic condition, 30 ml of culture was closed by a seal cap then using pure N_2 (100%) degas 5-10 mins. The diagram was drawn by www.biorender.com. (b) *In vivo* hydrogenase activity assays. The H_2 production (normalized by inducing 30 ml liquid cells about 16 hours) of *E. coli* cells expressing HyaAB or HyaAB-EP grown under anaerobic (**$p = 0.21$, two-tailed unpaired *t*-test) or aerobic (**$p = 0.49$, two-tailed unpaired *t*-test) conditions were measured by gas chromatography.

Next, we measured the levels of dissolved oxygen (DO) in the three different types of cell cultures to confirm the level of aerobic and anaerobic conditions (Figure 3-5a and 5b). After 16 hours of incubation, the three types of cells culture consumed a significant amount of oxygen, with DO dropping from $38.4 \pm 0.4\%$ to $1.2 \pm 0.1\%$ for cells expressed free HyaAB-EP, from $37.5 \pm 0.4\%$ to $1.3 \pm 0.1\%$ for cells expressed HyaAB-EP-Shell-1, and from $37.2 \pm 0.1\%$ to $1.2 \pm 0.1\%$ for cells expressed HyaAB-EP-Shell-2. However, the final DO levels of all cell types culture under aerobic conditions were significantly higher than those maintained constant under anaerobic conditions. At the same time, immunoblot analysis verified the presence of HyaA-EP, HyaB-EP, and major shell proteins in the samples (Figure 3-5c).
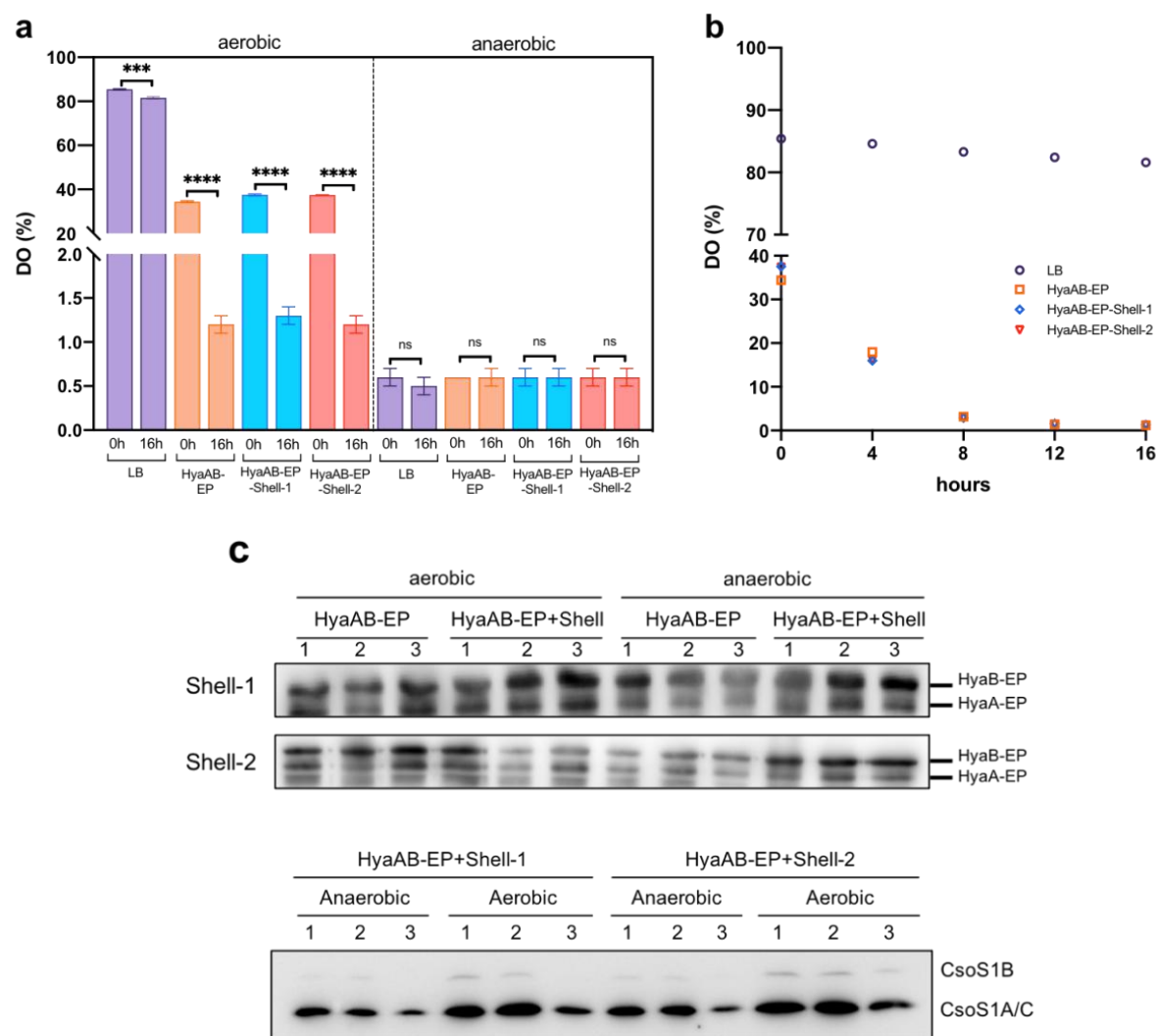
Figure 3-5. The oxygen consumption of cells expressing HyaAB-EP or HyaAB-EP-Shell-1/2 under aerobic and anaerobic condition. (a) The levels of dissolved oxygen (DO) in LB medium (control), LB medium containing the *E. coli* expressing HyaAB-EP or Shells-HyaAB-EP under aerobic and anaerobic condition during 16-hour induction. LB medium saturated DO was set as 100%. The data are presented as the average of nine (or three for control) DO measurements of distinct cell cultures. The data were compared by two-tailed unpaired *t*-test for aerobic and anaerobic conditions. For anaerobic condition, all comparisons showed no significantly difference between 0 hour and 16 hours $p < 0.05$. For aerobic condition, LB medium as control ***$p = 0.0008$, free HyaAB-EP, HyaAB-EP-Shell-1/2 all showed **** $p < 0.0001$. (b) DO changes in the LB medium (purple), cells producing HyaAB-EP (orange) or Shell-1-HyaAB-EP (blue) or Shell-2-HyaAB-EP (red) under aerobic conditions during 16-hour incubation. The data are presented as three DO measurements of three distinct cell cultures. (c) The recombinant [NiFe]-hdyroenase-1 and shell proteins were determined by western blot using anti-6xHis-tag and anti-CsoS1A/C/B antibody, respectively (Agrisera, USD).

The H_2 production of free HyaAB-EP and HyaAB-EP- Shell-1/2 were investigated *in vivo* (Figure 3-6a). Under anaerobic conditions, The H_2 production rates of cells expressing HyaAB-EP-Shell-1 and HyaAB-EP-Shell-2 at pH 7 are $36,977.52 \pm 986.71$ nmol L^{-1} h^{-1} and $41,777.52 \pm 1753.36$ nmol L^{-1} h^{-1}, respectively ($n = 3$), which are more than 20-fold greater than that of cells expressing free HyaAB-EP (1780.14 ± 197.88 nmol L^{-1} h^{-1} at pH 7). Under aerobic conditions, the H_2 production rate of cells producing HyaAB-EP-Shell-1 (7723.92 ± 352.55 nmol L^{-1} h^{-1}) and HyaAB-EP-Shell-2 (8937.37 ± 414.48 nmol L^{-1} h^{-1}) are about 30-fold higher than that of cells expressing free HyaAB-EP (269.17 ± 15.20 nmol L^{-1} h^{-1}) at pH 7. These results imply that [NiFe]-hydrogenase catalysis is more efficient, likely owing to the increased local enzyme and substrate concentration within the carboxysome shell lumen. It is note worthy that H_2 productivity of cells expressing HyaAB-EP-Shell-2 is greater than that of cells expressing HyaAB-EP-Shell-1, the consistent results also was found in the encapsulated [FeFe]-hdyrogenase inside α-carboxysome shells (Li *et al.*, 2020). The size of Shell-2 is larger than Shell-1, as was confirmed in previous study (Li *et al.*, 2020), possibly due to the larger size of Shell-2 than Shell-1 allowing more cargo enzymes and substrates condensed inside the shell and thereby further enhanced catalytic activity of hydrogenase (Figure 3-6b).
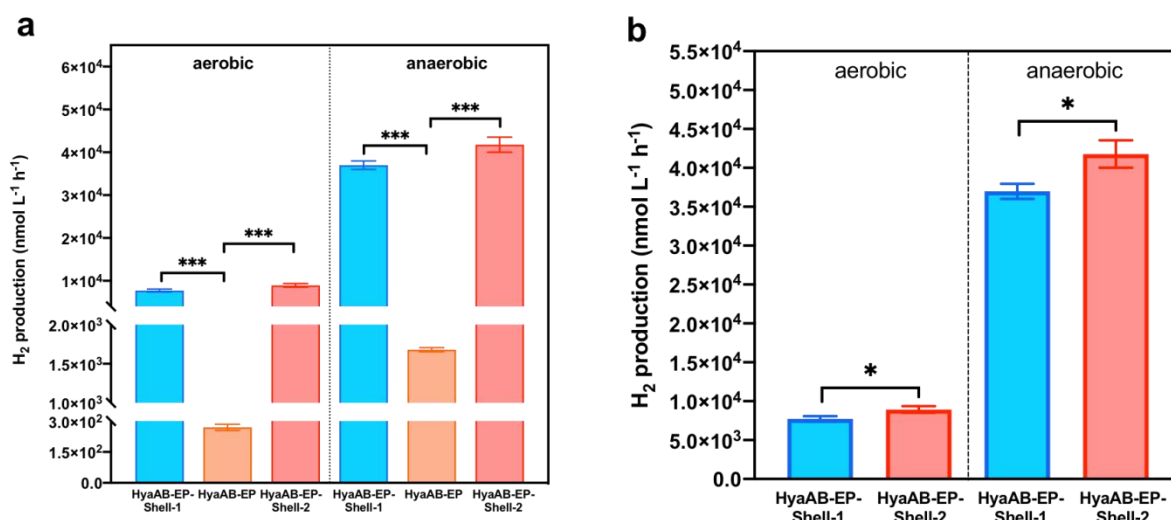
Figure 3-6. *In vivo* **hydrogen production of the carboxysome shell-based nanoreactor.** (a) The gas chromatography was used to measure H_2 production (normalized by inducing 30 ml cultures about 16 hours) of *E. coli* cells expressing HyaAB-EP, HyaAB-EP-Shell-1, and HyaAB-EP-Shell-2 grown under anaerobic (***p = 0.0003, ***p = 0.0006, two-tailed unpaired *t*-test) or aerobic (***p = 0.0007, ***p = 0.0008, two-tailed unpaired *t*-test) conditions. Error bars represent standard errors of the mean of three biological replicates. (b) The difference in H_2 production between *E. coli* cells expressing HyaAB-EP-Shell-1 and HyaAB-EP-Shell-2 under aerobic (*p = 0.0190, two-tailed unpaired *t*-test) or anaerobic (*p = 0.0234, two-tailed unpaired *t*-test)

3.2.4 The effect of other factors on *in vivo* H_2 production of [NiFe]-hydrogenase-containing nanobioreactors

The [NiFe]-hdyrogenase 1 of *E. coli* is a reversible hydrogenase that is expected to work largely in one direction depending on pH and the substrates used in the fermentation process (King, 1999; Trchounian *et al.*, 2011; Sanchez-Torres *et al.*, 2013; Jordan *et al.*, 2016; Trchounian *et al.*, 2017). Based on above results, HyaAB-EP-Shell-2 nanoreactor system could obtain more H2 production than HyaAB-EP-Shell-1, we first investigated the influence of pH on *in vivo* H_2 production for HyaAB-EP- Shell-2 under anaerobic conditions. Increasing H_2 generation was seen in Figure 3-7 as the pH decreased. The H_2 evolution rate of cells expressing HyaAB-EP-Shell-2 is 252,915.64 ± 2773.44 nmol H_2 L^{-1} h^{-1} (*n* = 3) at pH 5, which was almost 6-times greater than that at pH 7 (41773.74 ± 1781.84 nmol H_2 L^{-1} h^{-1}, *n*

= 3). The H_2 evolution rate measured at pH 8 was 10289.31 ± 317.98 nmol H_2 L^{-1} h^{-1}, 4 times lower than the H_2 evolution rate at pH 7.
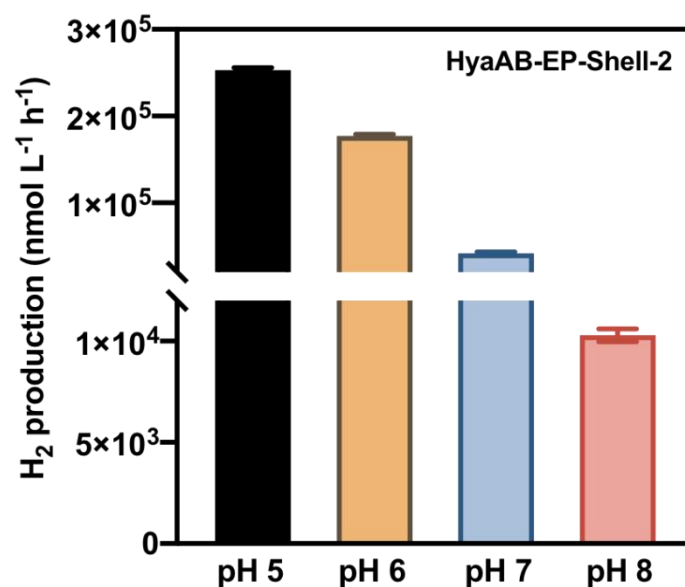
Figure 3-7. The effect pH on H_2 production for HyaAB-EP-Shell-2. *E. coli* expressing Shell-2-HyaAB-EP growing in different pH of LB medium under 37 °C until the OD_{600} up to 0.6~0.8. IPTG inducing recombinant HyaAB-EP expression 4 hours after the Shell-2 was induced via adding L-arabinose for total 16 hours under anaerobic conditions. The H_2 production was test by gas chromatography. Error bars represent standard errors of the mean of three biological replicates.

Previous studies have revealed that biohydrogen produced by hydrogenases could be enhanced by the aid of formate hydrogenlyase (FHL) complex in anaerobic conditions (Sawers, 2005; Yoshida *et al.*, 2005). The FHL complex components found in *E. coli* include formate dehydrogenase (FDH-H), hydrogenase 3 (Hyd-3), and electron transfer mediators (Böhm R, 1990, Sauter M, 1992, Zinoni F, 1986). Formate is oxidised to CO_2 and proton by FDH-H, and proton is converted to H_2 by Hyd-3 in *E. coli* (Sawers, 2005; Lacasse *et al.*, 2016). We investigated the effect of exogenous formate on hydrogen production in the *E. coli* expressing HyaAB-EP-Shell-2 (Figure 3-8) at pH 5 under anaerobic conditions. Adding 20 mmol exogenous foramte into the LB medium enhanced H_2 production by ~1.5 folds (370,145.41 $\pm$ 6969.38 nmol H_2 L^{-1} h^{-1}, *n* = 3) compared with no foramte addition

$(252,915.64 \pm 2772.44$ nmol H_2 L^{-1} h^{-1}, n = 3, ***p = 0.0003, two-tailed unpaired *t*-test).

Adding 100 mmol exogenous foramte resulted in 1.53 ± 0.01 folds H_2 production (387396.84

± 1503.73 nmol H_2 L^{-1} h^{-1}, n = 3) compared with no formate (****p < 0.0001, two-tailed

unpaired *t*-test). Our results showed a significantly increase in hydrogen production by adding

formate to 20 mmol final concentration similar to previous findings (Yoshida *et al.*, 2005;

Jaoon YH Kim, 2010; Trchounian *et al.*, 2012b).
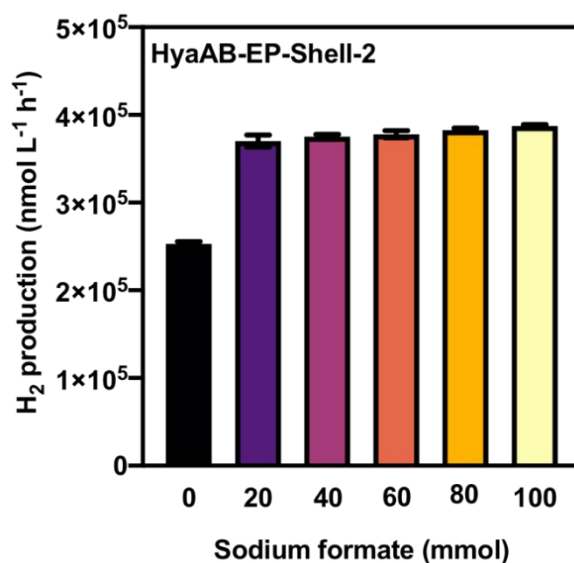
Figure 3-8. The effect exogenous foramte on H_2 production for HyaAB-EP-Shell-2.
Error bars represent standard errors of the mean of three biological replicates.

Formate is the product of glucose fermentation, which is widely applied in hydrogen

generation in *E. coli* under different conditions (Maeda *et al.*, 2007a; Juanita Mathews, 2010;

Trchounian *et al.*, 2012a; Trchounian *et al.*, 2012b). We further studied the impact of 0.5%

and 1% (w/v) of glucose on H_2 production for HyaAB-EP-Shell-2 under anaerobic conditions

at pH 5 (Figure 3-9). The H_2 evolution rate at pH 5 is $280,267.5 \pm 645.7$ nmol L^{-1} h^{-1} (n = 3)

at 0.5% glucose and $301,430.0 \pm 536.2$ nmol L^{-1} h^{-1} at 1% glucose (n = 3), higher than the H_2

production yield obtained without exogenous glucose (** p = 0.0023 for 0.5% glucose; *** p

= 0.0008 for 1% glucose).
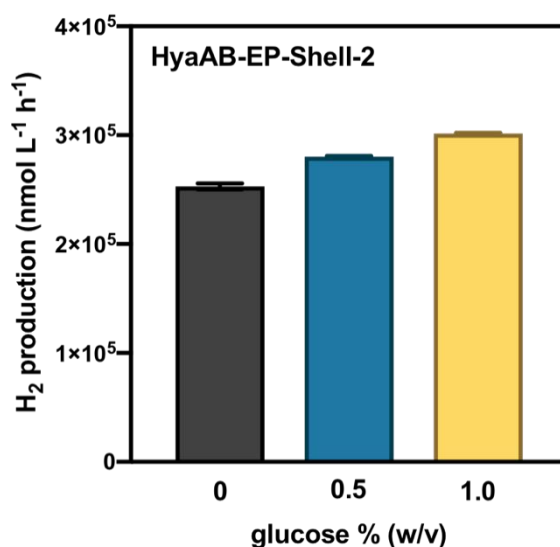
Figure 3-9. The effect of glucose on H$_2$ production for HyaAB-EP-Shell-2. Error bars represent standard errors of the mean of three biological replicates.

3.2.5 Shell encapsulation facilitates the catalytic performance of recombinant [NiFe]-hydrogenase (*in vitro*)

The H$_2$-evolution activities of isolated HyaAB-EP-Shell-1 and HyaAB-EP-Shell-2 were determined by gas chromatography using varying concentrations of methyl viologen (MV) as the electron donor (Figure 3-2a), which was chemically reduced by sodium dithionite. In the assays, the HyaB-EP content was normalised and quantified using purified HyaAB-EP as reference (Figure 3-10a, 3-10b). The maximum activity of free HyaAB-EP is 47.3 ± 1.6 nmol H$_2$ mg^{-1} min^{-1} at pH 8 (n = 3). It was shown, however, that the maximum hydrogen evolution rates of HyaAB-EP-Shell-1 and HyaAB-EP-Shell-2 were 522.7 ± 47.9 nmol H$_2$ mg^{-1} min^{-1} (n = 3) and 543.4 ± 67.7 nmol H$_2$ mg^{-1} min^{-1} (n = 3) at pH 8, respectively, around 11-fold greater than that of free HyaAB-EP (Figure 3-10c), respectively. The amount of hydrogen produced by both Shell-HyaAB-EP assemblies measured at 50 mM MV increased linearly as a function of time, indicating that the hydrogen evolution process is catalytic (Figure 3-10d). Under these conditions, both shell-HyaAB-EP assemblies produce remarkably more hydrogen than free HyaAB-EP, demonstrating that the carboxysome shell-based nanoreactors

could facilitate the H₂ production activity of oxygen-tolerant [NiFe]-hydrogenases (Figure 3-10a and 10b).
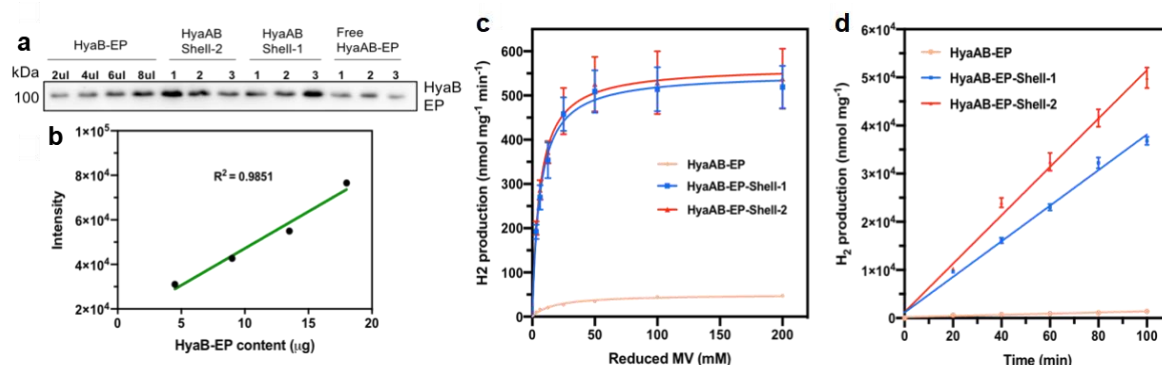
Figure 3-10. *In vitro* **hydrogen production of the carboxysome shell-based nanoreactors.** (a) Immunoblot analysis of purified HyaAB-EP using an anti-His antibody. Free HyaAB-EP was purified by immobilized-nickel affinity chromatography followed by quantification by BCA protein assay kit. 2μl, 4μl, 6μl, and 8μl of HyaAB was loaded onto each well of the gel, respectively. (b) Linear relationship between HyaAB content quantified by BCA protein assay kit and the corresponding HyaB-EP content quantification by immunoblot. The immunoblot analysis represent three biologically independent experiments. Densitometric quantitation of HyaB-EP levels were determined using ImageJ.(c) H₂ production activities (nmol H₂ mg⁻¹ min⁻¹) of free HyaAB-EP, Shell-1-HyaAB-EP and Shell-2-HyaAB-EP at pH 8 using different concentrations of reduced MV+ (sodium dithionite) as the electron donor, fitted with Michaelis-Menten kinetics. Error bars show standard deviation of the mean of three biological replicates. (d) Kinetic hydrogen production of free HyaAB-EP, Shell-1-HyaAB-EP and Shell-2-HyaAB-EP using 50 mM MV as electron donor at pH 8. Error bars show standard deviations of the mean of three biological replicates.

Then, we examined the O₂ tolerance of HyaAB-EP-Shell-1 and HyaAB-EP- Shell-2 by exposing the purified HyaAB-EP- Shell-1/2 and free HyaAB-EP to air for 24 hours, followed by a degassing process and an *in vitro* activity assay (Figure 3-11a, 11b and 11c). Compared to the activity observed before O₂ exposure (Figure 3-11a), there was a reduction of activity occurred for all the three sample after being exposed to O₂. Free HyaAB-EP has only 12.9 ± 1.8% (4.2 ± 0.3 nmol H₂ mg⁻¹ min⁻¹ (n = 3)) hydrogen production activity, both HyaAB-EP- Shell-1/2 assemblies remained 79.6 ± 5.5% (371.3 ± 15.1 nmol H₂ mg⁻¹ min⁻¹ (n = 3)) and 88.3 ± 2.6% (433.6 ± 8.3 nmol H₂ mg⁻¹ min⁻¹ (n = 3)) hydrogen production activities at pH 8, respectively (Figure 3-11b and 11c). These results not only confirmed that the oxygen-tolerant feature of [NiFe]-hydrogenase but also the shells provides an enhanced oxygen-

tolerant stability to the encased recombinant hydrogenase. Notably, the comparable phenomenon was not present in the enzymatic activity of [FeFe]-hydrogenase (Li *et al.*, 2020).

Next, the HyaAB-EP- Shell-1/2 and free HyaAB-EP were treated at 65 °C for 20 mins under anaerobic condition. The free HyaAB-EP preserved $33.0 \pm 5.0\%$ of activity (from 32.9 ± 2.7 nmol H_2 mg^{-1} min^{-1} to 10.8 ± 1.4 nmol H_2 mg^{-1} min^{-1} at pH 8, (n = 3)) (Figure 3-11a, 11e and 11f). HyaAB-EP- Shell-1/2, in contrast, preserved $52.6 \pm 3.2\%$ and $54.0 \pm 4.4\%$ of the activity were maintained (from 467.0 ± 16.8 nmol H_2 mg^{-1} min^{-1} to 245.1 ± 6.6 nmol H_2 mg^{-1} min^{-1} for HyaAB-EP-Shell-1 and from 491.4 ± 21.0 nmol H_2 mg^{-1} min^{-1} to 264.4 ± 12.1 nmol H_2 mg^{-1} min^{-1} for HyaAB-EP-Shell-2, at pH 8, (n = 3)), respectively. Our results indicate that high temperature could affect the activity and stability of [NiFe]-hydrogenase 1 of *E. coli*, which is similar to previous studies (?); the protein shells could serve as the thermal protector to maintain the encapsulated HyaAB-EP activity.
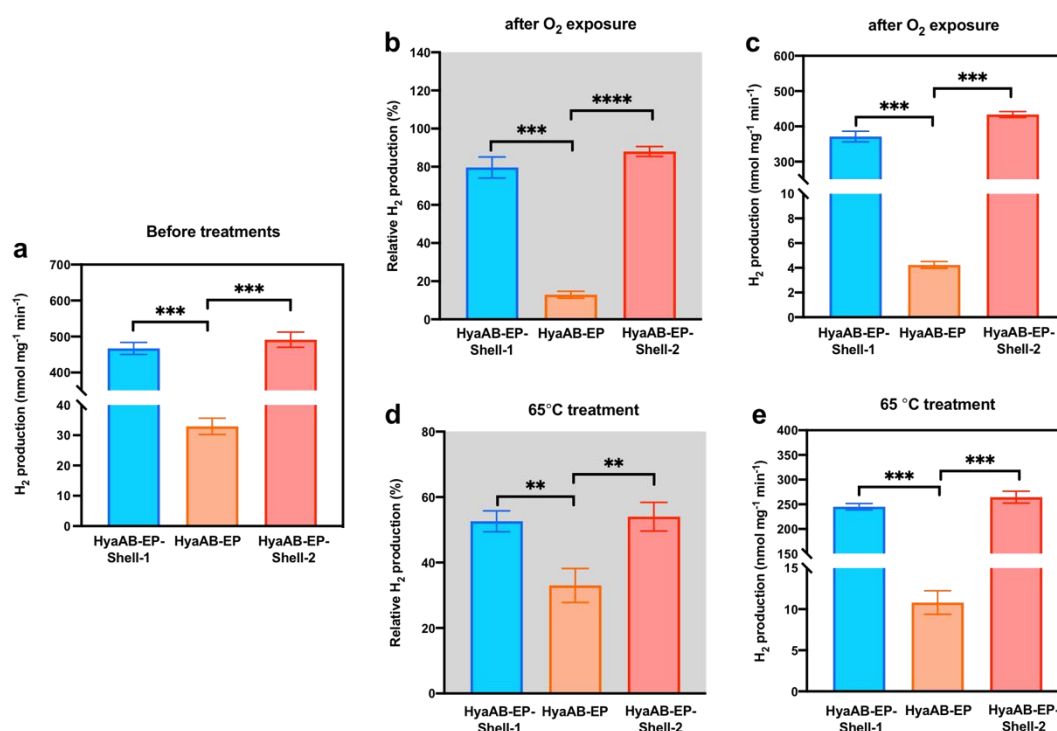
Figure 3-11. The difference in [NiFe]-hydrogenase activity before and after O_2 exposure. (a) The activity of [NiFe]-hydrogenase of free HyaAB-EP and HyaAB-EP-Shell-1/2 before O_2 exposure and 65 °C treatment. Each value represent the mean of three independent samples and standard deviation. *** $p = 0.0004$ (the left), *** $p = 0.0006$ (the right) (two-tailed unpaired t-test). (b) The relative activity of free HyaAB-EP and HyaAB-EP-Shell-1/2 after oxygen exposure for 24 hours at 4 °C, as a relative percentage of total activities measured under anaerobic conditions in Figure 3-5c. Each value represent the mean of three independent samples and standard deviation. *** $p = 0.004$ (the left), **** $p < 0.0001$ (the right) (two-tailed unpaired t-test). (c) The activity of [NiFe]-hydrogenase of free HyaAB-EP and HyaAB-EP-Shell-1/2 after O_2 exposure for 24 hours at 4 °C. Each value represent the mean of three independent samples and standard deviation. *** $p = 0.0006$ (the left), *** $p = 0.0001$ (the right) (two-tailed unpaired t-test). (d) The relative activity of free HyaAB-EP and HyaAB-EP-Shell-1/2 after 65 °C treatment for 20 mins, as a relative percentage of total activities measured under anaerobic conditions in Figure 3-5e. Each value represent the mean of three independent samples and standard deviation. *** $p = 0.001$ (the left), **** $p < 0.0001$ (the right) (two-tailed unpaired t-test). (e) The activity of [NiFe]-hydrogenase of free HyaAB-EP and HyaAB-EP-Shell-1/2 after 65 °C treatment for 20 mins. Each value represent the mean of three independent samples and standard deviation. *** $p = 0.0001$ (the left), *** $p = 0.0006$ (the right) (two-tailed unpaired t-test).

To test whether the shells would maintain the activity of the encased HyaAB-EP for an extended period of time, the isolated HyaAB-EP- Shell-1/2 and free HyaAB-EP were stored at 4 °C for 7 days under anaerobic condition (Figure 3-12a, 12b and 12c). HyaAB-EP-Shell-1/2 retained $16.7 \pm 1.0\%$ (77.6 ± 2.5 nmol H_2 mg^{-1} min^{-1} at pH 8, (n = 3)) and $16.9 \pm 0.3\%$ (83.2 ± 2.0 nmol H_2 mg^{-1} min^{-1} at pH 8, (n = 3)) activity. However, free HyaAB-EP only

maintained $1.7 \pm 0.4\%$ (0.6 ± 0.1 nmol H_2 mg^{-1} min^{-1} at pH 8, (n = 3)) activity after 7 days under anaerobic condition. These results demonstrated that the shells could maintain the enzymatic activity of encased HyaAB-EP.
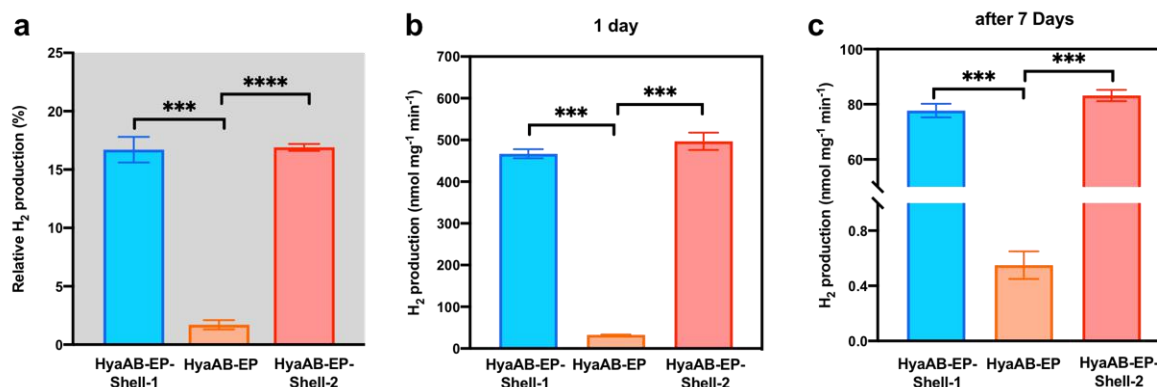
Figure 3-12. The difference in [NiFe]-hydrogenase activity for time treatment. (a) The relative activity of free HyaAB-EP and HyaAB-EP-Shell-1/2 after 7 days at 4 °C under anaerobic condition, as a relative percentage of total activities measured under anaerobic conditions in Figure 3-6c. Each value represent the mean of three independent samples and standard deviation. *** $p = 0.0004$ (the left), **** $p < 0.0001$ (the right) (two-tailed unpaired t-test). (b) The activity of [NiFe]-hydrogenase of free HyaAB-EP and HyaAB-EP-Shell-1/2 for 0 day at 4 °C under anaerobic condition. Each value represent the mean of three independent samples and standard deviation. *** $p = 0.0002$ (the left), *** $p = 0.0006$ (the right) (two-tailed unpaired t-test). (c) The activity of [NiFe]-hydrogenase of free HyaAB-EP and HyaAB-EP-Shell-1/2 after 7 days at 4 °C under anaerobic condition. Each value represent the mean of three independent samples and standard deviation. *** $p = 0.0003$ (left), *** $p = 0.0002$ (the right) (two-tailed unpaired t-test).

3.3 Discussion

In our study, we investigated the H_2 production activity of recombinant [NiFe]-hydrogenase 1 (*hyaAB*). We concluded that recombinant [NiFe]-hydrogenase 1 of *E. coli* could be functionally expressed aerobically and anaerobically. Utilization of EPs is an effective method for enzyme encapsulation, which showed in carboxysomes and other proteinaceous organelles (Bobik *et al.*, 1997; Sauvageot *et al.*, 2002; Fan *et al.*, 2010; Choudhary *et al.*, 2012; Kinney *et al.*, 2012). Our results also confirmed that the C-termin of CsoS2 of α-carboxysomes could function as an encapsulation peptide (EP) to encase [NiFe]-hydrogenases into the shells (Figure 3-2c and 3-3).

Our *in vivo* assays showed that there is a large amount of hydrogen production from Shells-[NiFe]-hydrogenases under aerobic and anaerobic conditions as well as hydrogen production is higher under anaerobic condition than aerobic condition (Figure 3-6b). [NiFe]-hydrogenase of *E. coli* is an oxygen-tolerant hydrogenase and it operates H_2 consumption under the presence of O_2 (Cracknell *et al.*, 2009; Tatyana, 2001; Lukey *et al.*, 2010; Evans *et al.*, 2013, Flanagan *et al.*, 2016). In dissolved oxygen (DO) test (Figure 3-5), there were less air left in the culture of Shells-[NiFe]-hydrogenase under aerobic condition after 16 hours (Figure 3-5a). Besides, [NiFe]-hydrogenase was induced prior to the shells 4 hours for hydrogenase maturation. During these four hours, the DO value from ~37% dropping in ~16% for Shells-[NiFe]-hydrogenase cultures (Figure 3-5b), confirming it was still a aerobic condition. Thus, it is a plausible explanation why hydrogen production under anaerobic condition more higher than under aerobic condition.

We also tested the effect of the different pH of LB medium on H_2 production under anaerobic condition. The results shown the H_2 evolution rate response to the shift of pH under anaerobic conditions and a low pH could stimulate increased hydrogen production. Our results agrees with the earlier study found that the H_2 mechanism via [NiFe]-hydrogenase 1 of *E. coli* dependent on the pH of the medium (King, 1999; Trchounian *et al.*, 2012a).

Previous studies indicated the formate hydrogenlyase (FHL) complex is induced in *E. coli* by exogenous formate (Rossmann, 1991; Mnatsakanyan *et al.*, 2002; Mnatsakanyan N, 2004), which promotes H_2 production of hydrogenases in *E. coli* (Sawers, 2005; Yoshida *et al.*, 2005). Our experiments confirmed that the FHL complex plays a role in assisting recombinant [NiFe]-hdyrogenase-1 to produce H_2 by stimulation of sodium formate. Notably, the price of exogenous formate (100g/£34.7 in Sigmaaldrich.com) is too high to benefit

engineering applications. However, the fermentation of glucose into formate in *E. coli* provides a significant output of hydrogen (Maeda *et al.*, 2007a; Juanita Mathews, 2010; Trchounian *et al.*, 2012a; Trchounian *et al.*, 2012b). Thus, we tested the effect of adding 0.5% and 1.0% (w/v) glucose to the medium on hydrogen production (Figure 3-9). There was an increase in H_2 production by adding 0.5% and 1.0% (w/v) glucose compared with the LB medium without glucose. The substantial H_2 production was verified once again by a similar discovery (Bisaillon *et al.*, 2006).

Our *in vitro* assays showed that Shell-[NiFe]-hydrogenases have a maximum H_2 evolution rate that is more than 10-fold greater than that of the free [NiFe]-hydrogenases (Figure 3-10), indicating the α-carboxysomes shell plays a role in condensation of enzymes to boost activity. In addition, the encapsulation of the α-carboxysome shell structure helped to retain the [NiFe]-hydrogenases activity under O_2 exposure, time, and temperature treatment (Figure 3-11 and 3-12). Moreover, the selective permeability of carboxysome shells plays important roles in preventing O_2 diffuse into the shell and CO_2 leakage (Dou *et al.*, 2008; Mahinthichaichan *et al.*, 2018; Gao *et al.*, 2022), providing a catalytically favorable microenvironment for [NiFe]-hydrogenases to enhance hydrogen production.

Immunoblot analysis found that the purified HyaA-EP protein content was lower than the purified HyaB-EP (Figure 3-2c and 3-3b) for pellet fractions, which might be a contributing factor to the poor encapsulation effectiveness of the small subunit HyaA. However, HyaA is a significant factor in hydrogen production because it contains [Fe-S] clusters that transfer electrons to the catalytic active site of the large subunit HyaB (Adams and Hall, 1979; Volbeda *et al.*, 2013; Murphy, 2014). Furthermore, a recent study discovered that only encapsulated HyaB encased in bacteriophage P22 capsid was unable to produce high-yield

hydrogen (Jordan *et al.*, 2016). Thus, it will be essential to increase expression of HyaA in the future employing various promoters or expression circumstances (e.g. induction time, induction temperature, the concentration of inducer). Previous investigation that the activity of encapsulated [FeFe]-hydrogenase and free [FeFe]-hydrogenase was lost under O_2 exposure (?), whereas [NiFe]-hydrogenase was found to be unaffected, making [NiFe]-hydrogenase a better candidate for biohydrogen generation (Cracknell *et al.*, 2009; Lukey *et al.*, 2010; Evans *et al.*, 2013).

Chapter 4 Structural basis for the CsoS2-mediated α-carboxysome shell assembly

4.1 Introduction

Organelles confine specific biochemical pathways within the cell to enhance metabolic efficiency, alleviate metabolic crosstalk, and facilitate spatiotemporal regulation of sequestered pathways (Kerfeld *et al.*, 2010). Apart from eukaryotes, in the past decade, advances in bioinformatics, imaging, and cell physiology have demonstrated that bacteria have also evolved subcellular organelles, including bacterial microcompartments (BMCs), to compartmentalize metabolism (Greening and Lithgow, 2020).

Carboxysomes are anabolic BMCs for autotrophic CO_2 fixation in all cyanobacteria and many chemoautotrophs (Hennacy and Jonikas, 2020; Liu *et al.*, 2021a; Liu, 2021b). The carboxysome is composed of a polyhedral shell that encapsulates the key CO_2-fixation enzyme, ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), and carbonic anhydrase (CA) (Rae *et al.*, 2013; Faulkner *et al.*, 2017; Sun *et al.*, 2019). The carboxysome shell acts as a selectively permeable barrier, allowing the influx of HCO_3^- and ribulose 1,5-bisphosphate (RuBP) while precluding O_2 influx and CO_2 leakage (Faulkner *et al.*, 2020). The co-encapsulated CA then dehydrates HCO_3^- to CO_2, the substrate for Rubisco carboxyslation (Price *et al.*, 2008; Long *et al.*, 2021). The intriguing structural features of carboxysomes are fundamental for maximizing CO_2 assimilation and reducing the unproductive RuBisCO oxygenation, thereby allowing carboxysomes to make substantial contributions to global carbon fixation and primary production (Rae *et al.*, 2013).

The carboxysome from the chemoautotroph *Halothiobacillus neapolitanus*, which was classified as the α-type (Rae *et al.*, 2013; Kerfeld and Melnicki, 2016), has been used as a model system in fundamental studies and synthetic engineering. The *H. neapolitanus* α-carboxysome proteins are encoded by several genes that are mostly clustered in a single *cso*

operon. The shell comprises numerous protein paralogs, including three types of hexameric proteins (BMC-H, CsoS1A/B/C), one trimeric pseudohexamer CsoS1D (BMC-T), and two forms of pentameric proteins (BMC-P, CsoS4A/B). CsoS1A/B/C and CsoS1D assemble to form the shell facets, whereas CsoS4A/B proteins occupy the vertices of the polyhedral shell. It is presumed that α-carboxysome biogenesis adopts the 'Shell first' or 'Concomitant shell–core assembly pathways, which is supported by recent studies showing that native-like α-carboxysome shells can form without cargos (Li *et al.*, 2020).

In the α-carboxysome, a high-abundance component, CsoS2, functions as a linker protein to mediate the association between the shell and cargo enzymes. Previous results suggested that the N-terminal domain of CsoS2 binds with Rubisco and induces Rubisco condensation (Oltrogge *et al.*, 2020), whereas its C-terminus associates with the shell (Cai *et al.*, 2015a) and CsoS2 is necessary for the assembly of intact α-carboxysome shells (Li *et al.*, 2020). However, how CsoS2 interacts with the shell and governs shell assembly remains mysterious.

Production of synthetic BMC shells provide a means for evaluating the assembly mechanisms and pairwise interactions that drive shell formation (Sutter *et al.*, 2017; Greber *et al.*, 2019; Sutter *et al.*, 2019b; Kalnins *et al.*, 2020; Tan *et al.*, 2021c). Here, we synthetically engineered miniaturized α-carboxysome shells and conducted high-resolution cryo-electron microscopy, in collaboration with Prof Peijun Zhang at Oxford University, to characterize the structural variations of the resulting shell assemblies. This study provides a mechanistic insight into the binding of CsoS2 with multiple shell proteins and CsoS2-mediates α-carboxysome shell assembly.

4.2 Results

4.2.1 Generation of synthetic α-carboxysome mini-shells

It has been demonstrated that native-like α-carboxysome shells can be formed by expressing all the shell components encoded by the *cso* operon in *E. coli* (Li *et al.*, 2020) (Figure 4-1a). To investigate the molecular principles driving carboxysome shell formation and the role of CsoS2 in α-carboxysome shell assembly, we designed two types mini-shell constructs, mini-shell-1 and mini-shell-2 (Figure 4-1b and 4-1c), comprising two genes encoding CsoS1A and CsoS4A, the major BMC-H (pfam00936) and BMC-P (pfam03319) proteins respectively (Sun *et al.*, 2021), as well as the *csoS2* gene encoding the linker protein CsoS2 (pfam12288) in the mini-shell-2 construct. CsoS1A is a cyclic hexamer with six monomeric subunits forming a central pore (Tsai *et al.*, 2007b); CsoS4A displays a pentagonal shape, capping the vertices of a polyhedral shell (Tanaka *et al.*, 2008a) (Figure 4-1d).
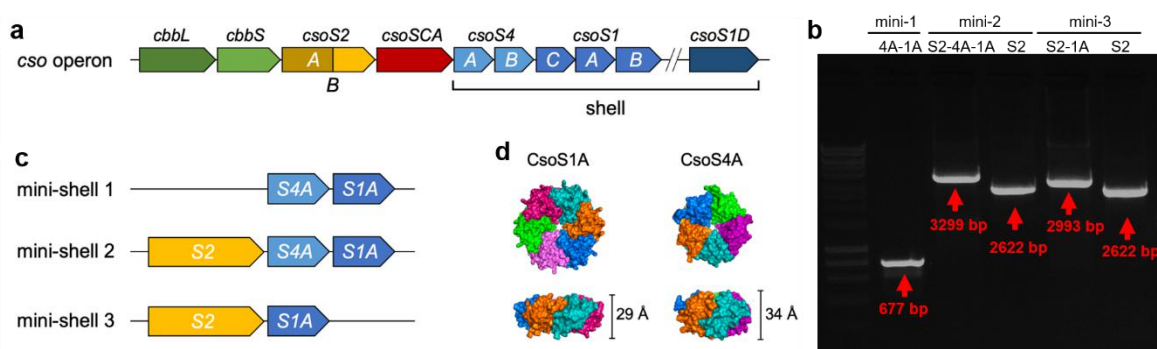
Figure 4-1. Constructs of α-carboxysome mini-shells. (a) Genetic organization of the native-like α-carboxysomes operon from *H. neapolitanus*. (b) PCR-based confirmation of mini-shell 1, mini-shell 2, and mini-shell 3. The size (bp) of the PCR fragments were indicated (red). (c) The synthetic mini-shells operons. S2 is *csoS2*, S4A is *csoS4A*, S1A is *csoS1A*. (d) Model of CsoS1A hexamer and CsoS4A pentamer (PDB ID: CsoS1A: 2EWH/2G13; CsoS4A: 2RCF).

Expression of the mini-shell-1 and mini-shell-2 constructs in *E. coli* both led to generation of higher-ordered shell architectures, as shown by protein electrophoresis and electron microscopy (EM) (Figure 4-2a and 4-2c). The results indicated that the hexamer CsoS1A and the pentamer CsoS4A represent the minimal components required for α-carboxysome shell assembly, supported by recent finding (Tan *et al.*, 2021b). Dynamic Light Scattering (DLS)

indicates that the majority of shell assemblies in mini-shell-1 have a diameter of ~23 nm, whereas the shell assemblies in mini-shell-2 have two major populations, with the diameters of ~35 nm and ~22 nm (Figure 4-2b). Variable shell sizes (T = 9, T = 4, and T = 3) in these assemblies are also apparent in cryo-EM micrographs (Figure 4-3).
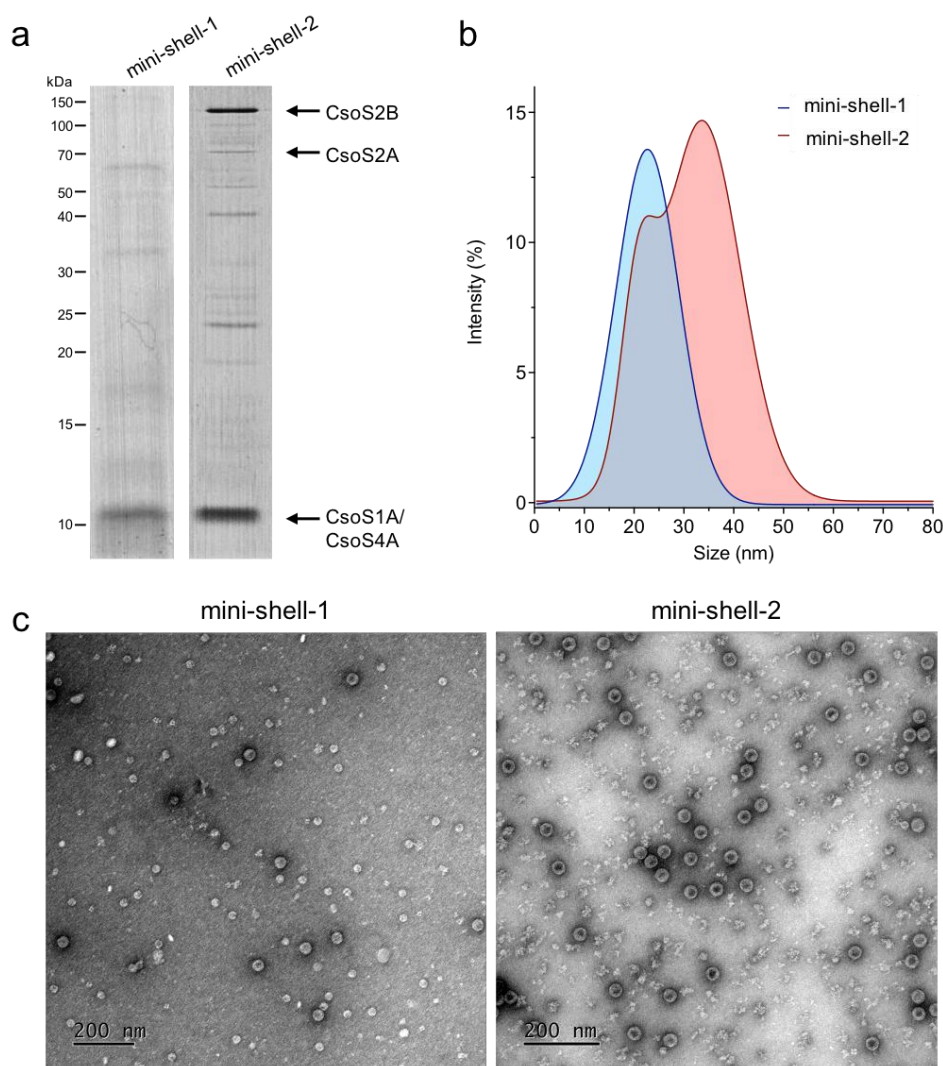
Figure 4-2. Characterisation of the mini-shells generated from mini-shell-1 (CsoS4A-CsoS1A) and mini-shell-2 (CsoS2-CsoS4A-CsoS1A). (a) SDS-PAGE revealed the major protein components of purified mini-shells from mini-shell-1 and mini-shell-2. (b) Dynamic light scattering (DLS) of the mini-shells from mini-shell-1 and mini-shell-2. (c) Electron microscopy (EM) images of purified mini-shells from mini-shell-1 and mini-shell-2. Scale bar is 200 nm.

Using single particle cryo-EM, we determined structures of three different shell assemblies with icosahedral symmetry of $T = 9$ from mini-shell-2 sample at 1.86 Å resolution, and $T = 4$

and $T = 3$ from mini-shell-1 sample at 3.54 Å and 2.79 Å resolution, respectively (Figure 4-3a). The $T=3$ shell has a diameter of 21.8 nm, containing 20 CsoS1A hexamers and 12 CsoS4A pentamers, whereas the $T=4$ shell, about 10% in the mini-shell-1 sample, has a diameter of 24.6 nm, composed of 30 CsoS1A hexamers and 12 CsoS4A pentamers (Figure 4-3b and 4-3d). Incorporation of CsoS2 within the shell assemblies (mini-shell-2 sample) resulted in predominantly the formation of larger $T=9$ shells with a diameter of 36.9 nm, in addition to the minor population of $T=4$ shells (~10%) (Figure 4-3c and 4-3d). The $T=4$ shell (mini-shell-2 sample) consists of 80 CsoS1A hexamers and 12 CsoS4A pentamers, as well as 60 CsoS2 proteins that are located at the inner surface of the shell. In contrast, the CsoS2 density is absent in the $T=3$ shells within the mini-shell-2 sample. This $T=4$ shell structure appears in mini-shell-2 identical to the $T=4$ shell derived from the mini-shell-1 construct, albeit at lower resolution (Figure 4-3b, 3c, 3d).
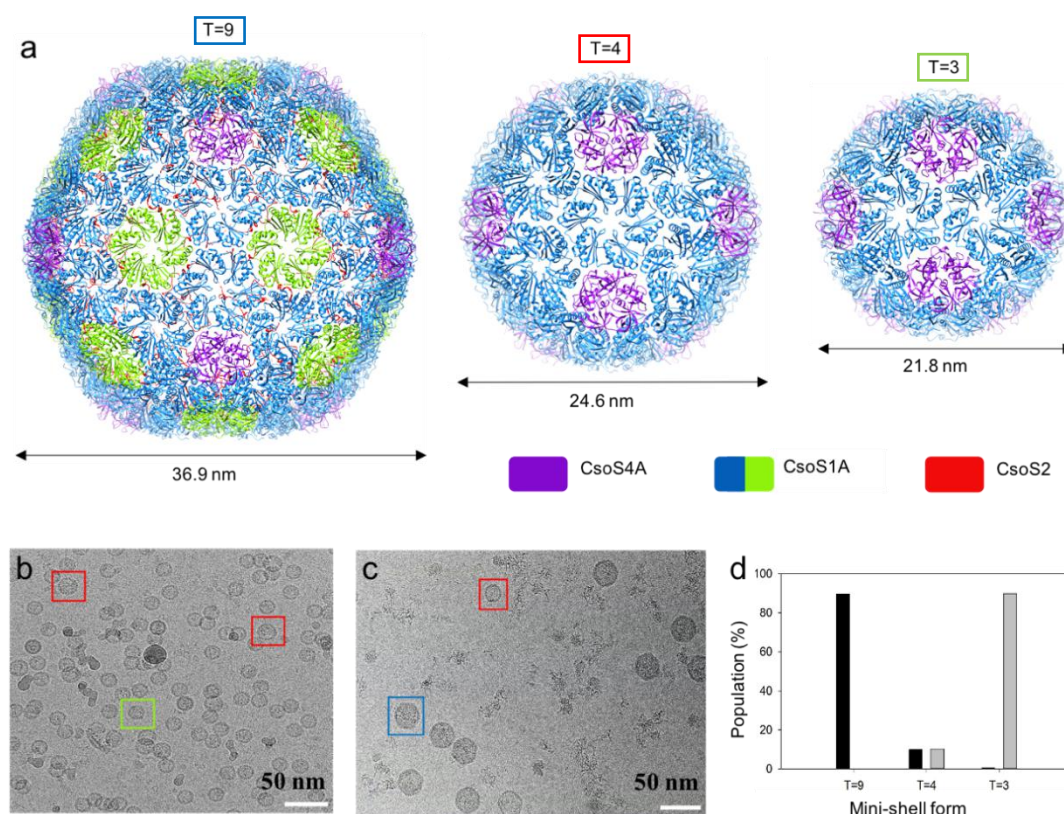
Figure 4-3. Cryo-EM analysis of α-carboxysome mini-shells. (a) Cryo-EM structures of three different mini-shell forms, with icosahedral symmetry of *T*=9, *T*=4 and *T*=3, respectively. The diameter of mini-shells are indicated. Mini-shell components are coloured in purple (CsoS4A pentamer), blue/green (quasi-equivalent CsoS1A hexamer) and red (CsoS2) which is only present in the *T*=9 mini-shell. (b-c) Representative cryo-electron micrographs of mini-shells produced from mini-shell-1 (b) and mini-shell-2 (c), respectively. Boxed particles have different sizes: blue, large mini-shells (*T*=9); red, medium mini-shell (*T*=4); and green, small mini-shell (*T*=3). Scale bars: 50 nm. (d). Distribution of mini-shell forms assembled with mini-shell-1 construct (gray, total 177,237 mini-shells) and mini-shell-2 construct (black, total 137,690 mini-shells)

All the shell proteins in these α-carboxysome mini-shells are in the same orientation (concave-out), similar to those in the previous studies (Kalnins *et al.*, 2020, Sutter *et al.*, 2017, Tan *et al.*, 2021c), indicating the universal "concave-out" orientation of shell proteins within diverse BMC architectures. Consequently, the outer surfaces of the mini-shells are formed by the concave sides of CsoS1A hexamers and CsoS4A pentamers, which are largely positively charged, whereas the inner surfaces of *T*=3 and *T*=4 shells are predominantly negatively charged (Figure 4-4). The presence of positively charged CsoS2 at the inner surface of the *T*=9 shell makes the inner surface less charged (Figure 4-4).
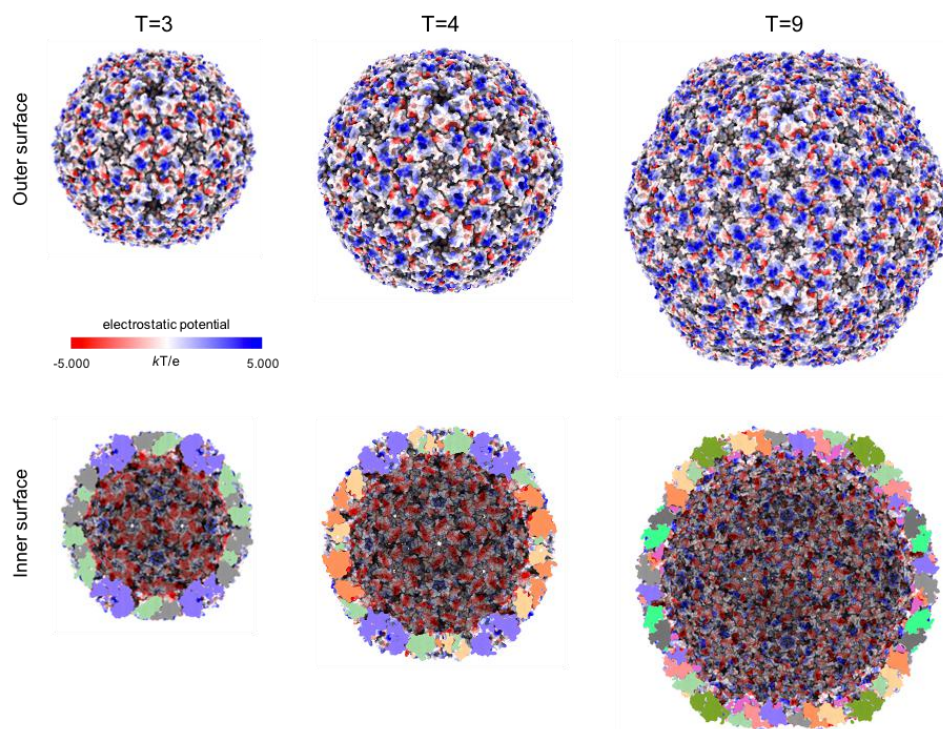
Figure 4-4. Electrostatic potential maps of the outer and inner surfaces (top and bottom) of the *T*=3, *T*=4, and *T*=9 mini-shells. Both CsoS1A and CsoS4A have concave and convex sides, which differ in the conformation and surface electrostatics. Negative potential is colored red and positive potential is colored blue over a ±5 kT/e range. Maps were calculated using the APBS plugin in Pymol.

4.2.2 Structural plasticity of shell proteins and protein-protein interactions

Our results showed that the three different sized shells are built of essentially the same CsoS1A hexamers and CsoS4A pentamers, following the *T*=3, *T*=4 and *T*=9 icosahedral symmetry (Figure 4-5a, 5b, 5c), with the RMSD range of 0.180-0.231 Å and 0.240-0.251 Å, respectively. Superimposing the cryo-EM structures of CsoS1A hexamers and CsoS4A pentamers in the shell assemblies with the X-ray crystal structures of CsoS1A in a P6 lattice (PDB: 2EWH) and CsoS4A (PDB: 2RCF) reveals slight deviations at their intermolecular interfaces, with the overall RMSD range of 0.290-0.337 and 0.370-0.437, respectively (Table 4-1). A comparison of the basic higher-order assembly unit (1 CsoS4A pentamer + 2 CsoS1A hexamer) of the three mini-shells suggests the interfaces are highly consistent (Figure 4-5d).
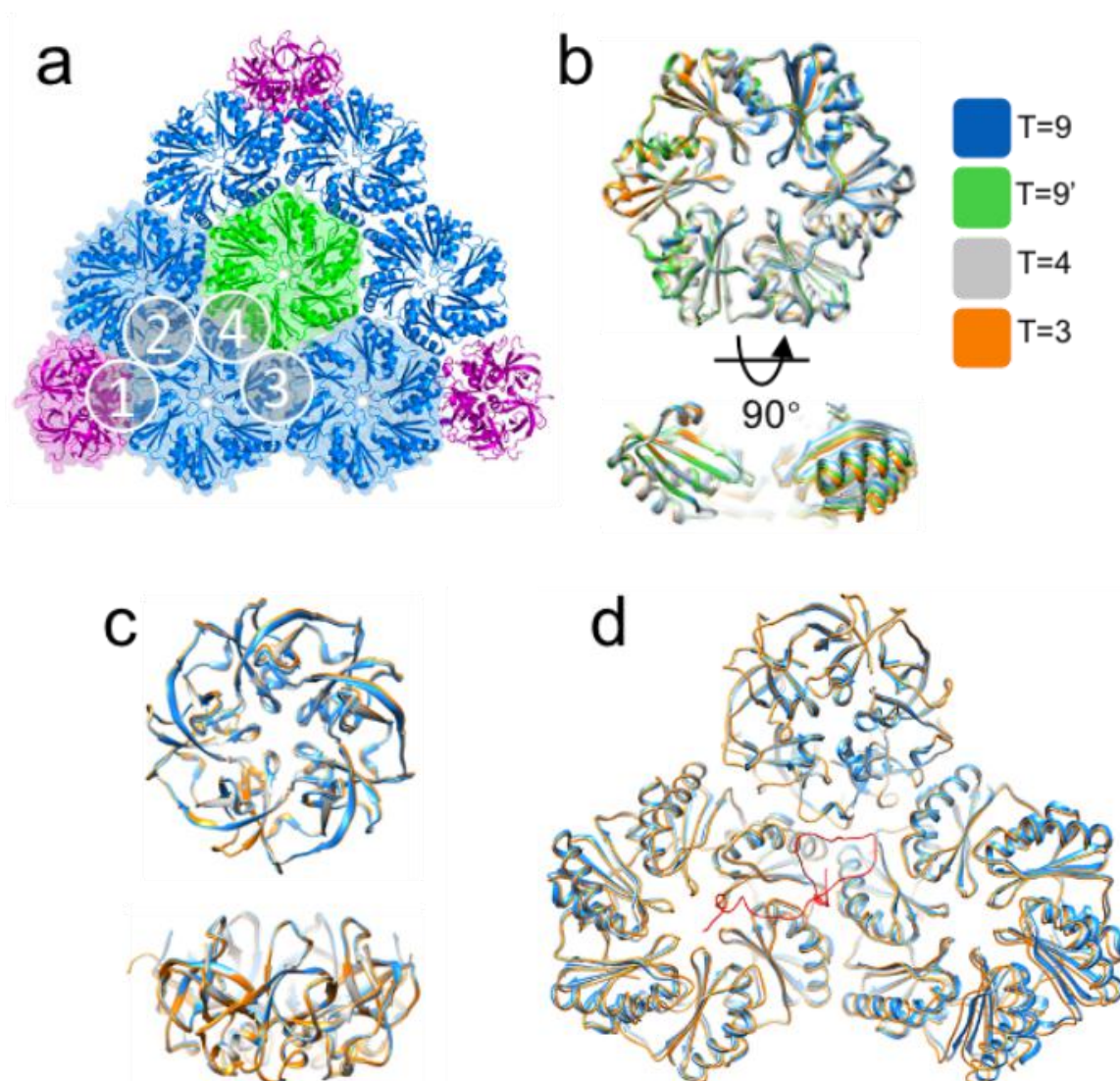
Figure 4-5. Structurally conserved shell proteins with plastic assembly interfaces. (a) The overall organization *T*=9 mini-shell, with assembly interfaces 1 to 4 between capsomeres labelled. Interfaces 3 and 4 are unique to *T*=9 shell. (b-c) Overlay of hexamer (CsoS1A) (b) and pentamer (CsoS4A) (c) structures from *T*=9 (blue/green), *T*=4 (gray), and *T*=3 (orange) mini-shells, in top and side views. (d) Overlay of the basic higher-order assembly unit comprising one pentamer (CsoS4A) and two hexamers (CsoS1A), from three mini-shells. CsoS2 (red) is only present in the *T*=9 mini-shell. *T*=9' means CsoS1A model of *H. neapolitanus* α-carboxysomes.

However, the three shells are of different diameters and different curvatures, with the *T*=9 shell least curved. We therefore analyzed the relative orientations between CsoS1A hexamers and between the Cso1A hexamer and CsoS4A pentamer. The angles between CsoS4A and CsoS1A at the hexamer-pentamer interfaces vary slightly from 30° to 35° (Figure 4-6a). At this interface, Lys29, Glu32, Arg62, Asp66, and Arg83 of CsoS1A interact with Thr9, Gly20,

His21, Gly43, Asp48, Ile77, and Asp78 of CsoS4A via hydrogen bonds and salt bridges (Figure 4-6d). In contrast, the angles between two neighbouring hexamers vary substantially in these different mini-shells (Figure 4-6b and 6c): the hexamers connecting pentamers between two vertices of the shell bind side-by-side in a planar fashion, whereas the hexamers surrounding the same pentamer are tilted by 30° to 43°, similar to the angles measured in other BMC mini-shells (Sutter *et al.*, 2017). The major interactions driving the inter-hexamer interaction are hydrogen bonds and salt bridges formed by residues Asn58, Arg62, Arg83 (Figure 4-6d). Despite a large deviation of the tilt angle, from 0° to 43°, the interactions at the interface between adjacent hexamers appear consistent (Figure 4-6c). Our results demonstrate that the same structural components, CsoS1A hexamers and CsoS4A pentamers, can construct three different mini-shell structures with variable sizes and symmetry, and suggest that the plasticity of hexamer interfaces contributes to the local curvature and thus, the polymorphism of carboxysomes.

Table 4-1. Root mean square deviation (RMSD) of superimposed CsoS1A and CsoS4A structures obtained from X-ray crystallography and cryo-EM (in this study)

		PDB: 2EWH	T3-H	T4-H
CsoS1A	T3-H	0.335	-	-
	T4-H	0.290	0.232	-
	T9-H	0.337	0.219	0.180
		PDB: 2RCF	T3-P	T4-P
CsoS4A	T3-P	0.435	-	-
	T4-P	0.437	0.240	-
	T9-P	0.370	0.241	0.251
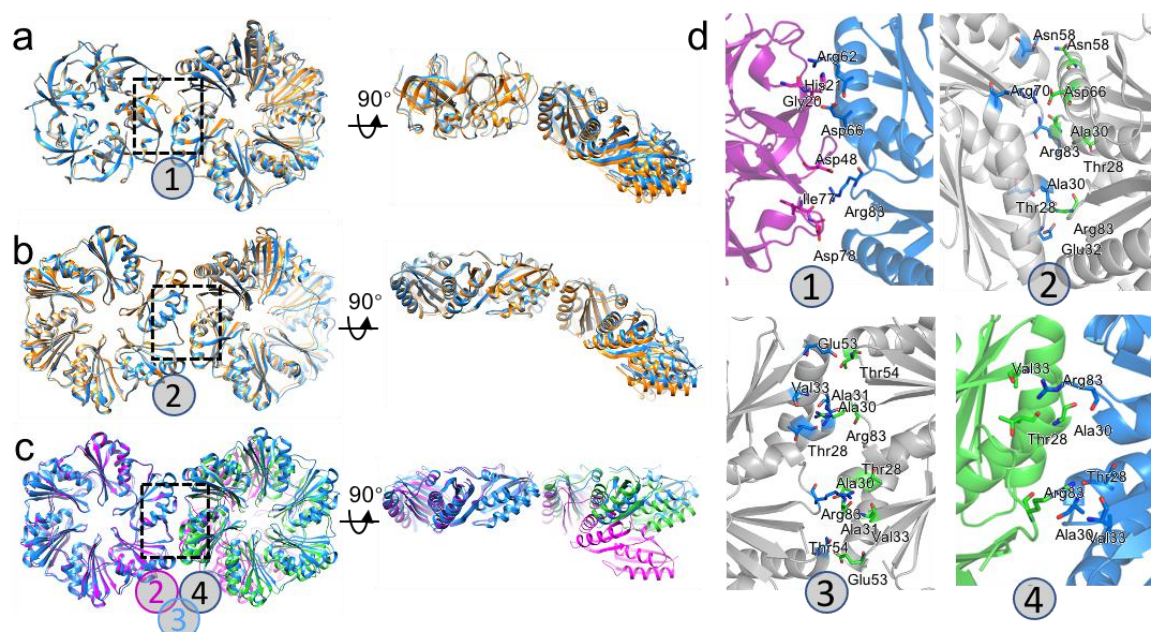
Figure 4-6. Variation of inter-hexamer and hexamer-pentamer interfaces in different sized mini-shells. Overlay of interface 1 (a) and interface 2 (b) from *T*=9 (blue), *T*=4 (gray) and *T*=3 (orange) mini-shells, viewed from top (left) and side (right). (c) Overlay of interfaces 2, 3 and 4 from *T*=9 only, aligned to the shared hexamer. There is a 43 ° difference in curvature. (d) Details of interacting residues in the dimer interfaces 1 to 4 in the *T*=9 mini-shell (dashed boxes in a-c).

4.2.3 Multivalent interactions of CsoS2 with shell proteins

It has been proposed that CsoS2 functions as the linker protein between the shell and cargo enzymes and is a vital component in α-carboxysome biogenesis (Cai *et al.*, 2015a). Genetic deletion of the *csoS2* gene resulted in the lack of carboxysomes in the *H. neapolitanus* cells and high CO_2-requiring phenotypes (Cai *et al.*, 2015a). CsoS2 is a large polypeptide (~900 residues) and is composed of three regions: a N-terminal region, a middle region, and a C-terminal region (Cai *et al.*, 2015a; Chaijarasphong *et al.*, 2016; Oltrogge *et al.*, 2020) (Figure 4-8a and 8b). Repetitive arrangements have been identified in the CsoS2 N- and M-regions, which have varying numbers among species; for example, the *H. neapolitanus* CsoS2 contains 4 N-repeats and 6 M-repeats (Cai *et al.*, 2015a) (Figure 4-7a and 7b). Previous studies have suggested that the CsoS2 N-terminal domain binds with Rubisco, playing roles

in mediating Rubisco condensation and α-carboxysome assembly (Oltrogge *et al.*, 2020), whereas the C-terminus of CsoS2 binds with the shell (Cai *et al.*, 2015a) and serves as an encapsulation peptide for cargo recruitment (Li *et al.*, 2020). However, how CsoS2 interacts with shell proteins to drive shell assembly has remained enigmatic.

The cryo-EM structure of the *T*=9 mini-shell at 1.86 Å resolution enables the first structural characterization of CsoS2 and its interactions with the shell at atomic level (Figure 4-7c). Despite the C-terminus of CsoS2 has been labelled as intrinsically disordered (Figure 4-7a and 7b), it is clearly resolved in the cryo-EM map (Figure 4-8d), allowing unambiguous assignment of amino acid residues of three regions of the CsoS2 C-terminus: F1, Arg712−Arg731; F2, Leu773−Gly823; F3, Glu829−Gly869 (Figure 4-7d, 7e, 7f, 7g, 7h). The region between F1 and F2 (including 42 amino acid residues) was not resolved likely due to its structural flexibility; the relative position of F1 and F2 from the same polypeptide is therefore uncertain. Although the density between F2 and F3 (6 amino acid residues) is weak, it still allows us to assign the connectivity between the two regions. Overall, it is explicit that the CsoS2 C-terminus is completely encapsulated inside the shell and that the C-terminal tail is not exposed to the cytoplasm as previously proposed (Cai *et al.*, 2015a).

The CsoS2 C-terminal regions form multiple hydrogen bonds and salt bridges with both CsoS1A and CsoS4A, implicating a strong anchoring of CsoS2 to the shell inner surface (Figure 4-7b and 7c). Detailed interface analysis reveals that the F1, F2, and F3 fragments form three distinct interfaces with the inner surface of the *T*=9 shell (Figure 4-7e, 7f, 7g). F1 interacts with the interfaces between three CsoS1A hexamers (H1, H5, H6), via hydrogen bonds and salt bridges driven by Gly, Leu, Thr, and Arg residues (Figure 4-7i). F2 form intensive interactions with the inter-hexamer interface between three hexamers (H1, H2, H4).

F3 connects the interfaces between one pentamer (P) and two hexamers (H1, H3), with the extreme C-terminus buried inside the cavity formed by the interface. In all the interfaces, CsoS2-C associates tightly with three shell proteins simultaneously, which essentially "stiches" all building blocks together at the inner surface of the T=9 shell, forming a "molecular thread" specific to the assembled carboxysomal shell.
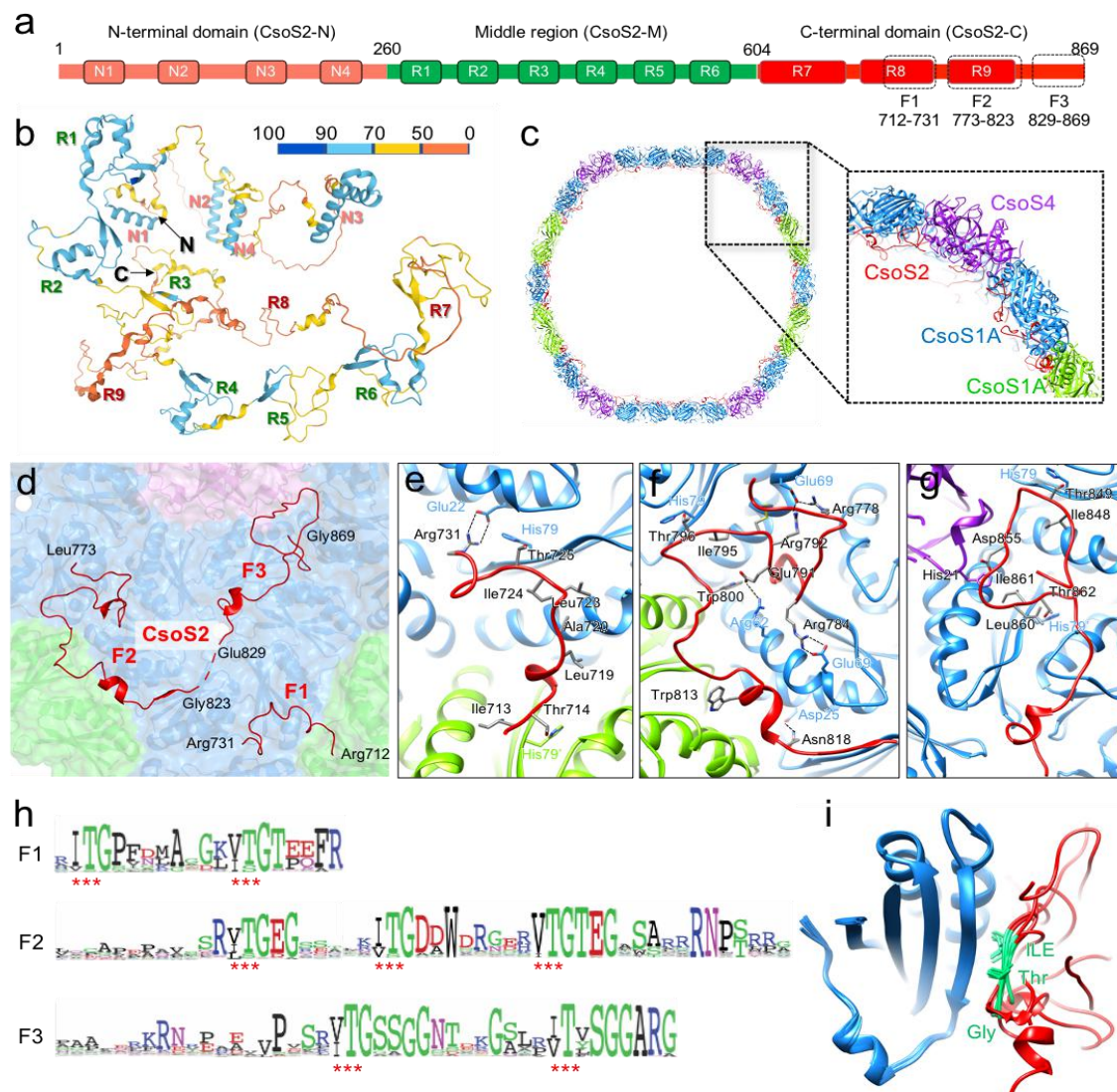
Figure 4-7. CsoS2 stabilizes the shell through multivalent interactions with shell proteins and highly conserved interfaces via novel I(V)TG repeats. (a) The domain arrangement of CsoS2. (b) AlphaFold structure prediction of CsoS2. The predicted model is coloured according to model confidence scores (pLDDT) as indicated. The N- and C-termini and domains are labelled. (c) Localization of CsoS2 at the inner surface of *T*=9 shell assembly, shown as a central slice. Shell proteins are coloured the same as in Figure 4-5a. Inset shows a close-up view. (d) CsoS2 interactions with shell components, viewed from inside. Three structured fragments in the C-terminal domain, F1, F2 and F3, are labelled. (e-g) Interaction interfaces between CsoS2 F1(e), F2 (f) and F3 (g) fragments with mini-shell components, CsoS1A (blue/green) and CsoS4A (purple). (h) Sequence of CsoS2 C-terminal F1, F2 and F3 fragments from 100 CsoS2 sequences, plotted with Weblog. Asterisks indicate the conserved repeating I(V)TG motif present in each fragment. (i) Alignment of CsoS2-CsoS1A interacting motifs marked by asterisks in (h).

4.2.4 CsoS2-C promotes shell assembly

Our cryo-EM and DLS results demonstrated that CsoS2 promotes generation of larger shell assemblies and that all three CsoS-C fragments, F1-F3, form contacts with the shell (Figure 4-2, 4-3, and 4-7). We further tested the effects of the anchoring of CsoS2-C fragments. Truncated CsoS2 C-termin with various numbers of the binding regions were generated (Figure 4-8a), and the efficiency of shell formation was evaluated by comparing the abundance of shell proteins in the free form and the assemblies (Figure 4-8b). With the decrease in the numbers of CsoS2 C-terminus binding regions, the formation efficiency of higher-ordered shell assemblies declines gradually, implying their roles in promoting shell assembly (Figure 4-8c). Taken together, our data demonstrate that the CsoS2-C not only serves as an encapsulation linker peptide for cargo recruitment as previously reported (Li *et al.*, 2020), but also plays a key role in α-carboxysome shell assembly and stability.

It appears that the interface created by two neighbouring hexamers with the same pentamer is very similar among these three mini-shell assemblies (Figure 4-5 and 4-6). This interface in the *T*=9 shell is responsible for the interactions with CsoS2-F3 (Figure 4-7g). In principle, the same interface in the *T*=3 and *T*=4 shells, as well as in the intact α-carboxysome shell, should enable the binding of F3. The interaction between F3 and the pentamer-hexamer interface implies the specific stoichiometric ratio of CsoS2 and BMC-P, and the number of CsoS2 in native α-carboxysome could be maximally 60 copies. However, it has been estimated that there are approximately 192 copies of CsoS2B within the α-carboxysome from *H. neapolitanus* (Sun *et al.*, 2021) and 163 copies of CsoS2 within the α-carboxysome from *Prochlorococcus marinus* MED4 (Roberts *et al.*, 2012), suggesting that other binding sites of CsoS2 to the α-carboxysome shell may exist. The other two inter-hexamer interfaces responsible for F1 and F2 interactions in the *T*=9 shell are absent in the *T*=3 and *T*=4 shells

(Figure 4-7e and 7f), therefore deterring efficient encapsulation of CsoS2 within these mini-shells. On the other hand, the smaller volume of the T=3 and T=4 shells would not allow efficient packaging of CsoS2 with a full occupancy. Given the molecular weight of intact CsoS2 (~92 kDa, equivalent to 110.769 nm^3/particle with a spherical diameter of ~6.0 nm), the T=9 shell with the inner radius of 14 nm could accommodate up to 103 copies of CsoS2, whereas the T=4 shell (inner radius of 8.8 nm) and the T=3 shell (inner radius of 7.2 nm) can maximally accommodate 26 and 14 copies, respectively.
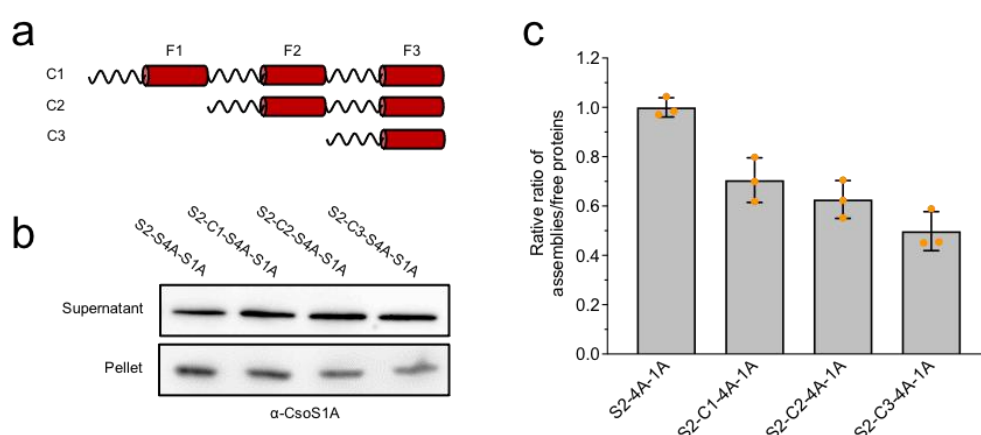
Figure 4-8. Mini-shell structures with minimal CsoS2 fragments. (a) Truncated CsoS2 C-terminus with various numbers of the binding regions were generated. (b) Immunoblot analysis of the formation of the mini-shell structures with truncated C-terminal of CsoS2 via using anti-CsoS1A antibody (Agrisera, AS142760, US). (c) The efficiency of the formation of mini-shell with truncated C-terminal of CsoS2. The efficiency of the formation decreases along with the reduction in the numbers of the C-terminal regions.

4.2.5 Foreign cargo proteins can be incorporated into the mini-shells

We have confirmed that CsoS2 C-terminus could as the encapsulated peptide to incorporate cargo into the α-carboxysome shells (Li *et al.*, 2020). We fused CsoS2 C-terminus to a pH-sensitive green fluorescent protein, pHluorin2 (Mahon, 2011), to determine the encapsulation ability of the mini-shells. The results showed that the pH-sensitive green fluorescent protein could be incorporated into the mini-shell-1 and -2 by CsoS2 C-terminus as encapsulated peptide. The pHluorin2 content in each sucrose gradient fraction was determined via

immunoblot analysis and fluorescence detection (Figure 4-9a). The structures of isolated mini-shells encapsulating pHluorin2 were examined by EM (Figure 4-9b). Moreover, DLS analysis showed that the average diameters of isolated mini-shell-1-pHluorin2 and mini-shell-2-pHluorin2 were 27.1 ± 0.07 nm ($n = 3$) and 40.2 ± 1.04 nm ($n = 3$), respectively (Figure 4-9c). DLS of mini-shell-1 and mini-shell-2 in Figure 4-2b and Cryo-EM results in Figure 4-3 showed the size of mini-shell-2 is large than mini-shell-1. Besides, there was a significant difference in size between mini-shell-1 ($T = 3$) with mini-shell-1-pHluorin2 ($p^{****} < 0.0001$, two-tailed unpaired *t-test*) and between mini-shell-2 ($T = 9$) with mini-shell-2-pHluorin2 ($p^{**} = 0.0048$, two-tailed unpaired *t-test*). These results indicated that the cargo protein could be encapsulated into mini-shells by the C-terminus of CsoS2 as encapsulated peptide. Further research is required to determine how cargo proteins encapsulate inside the shell and how the size of the shell change.
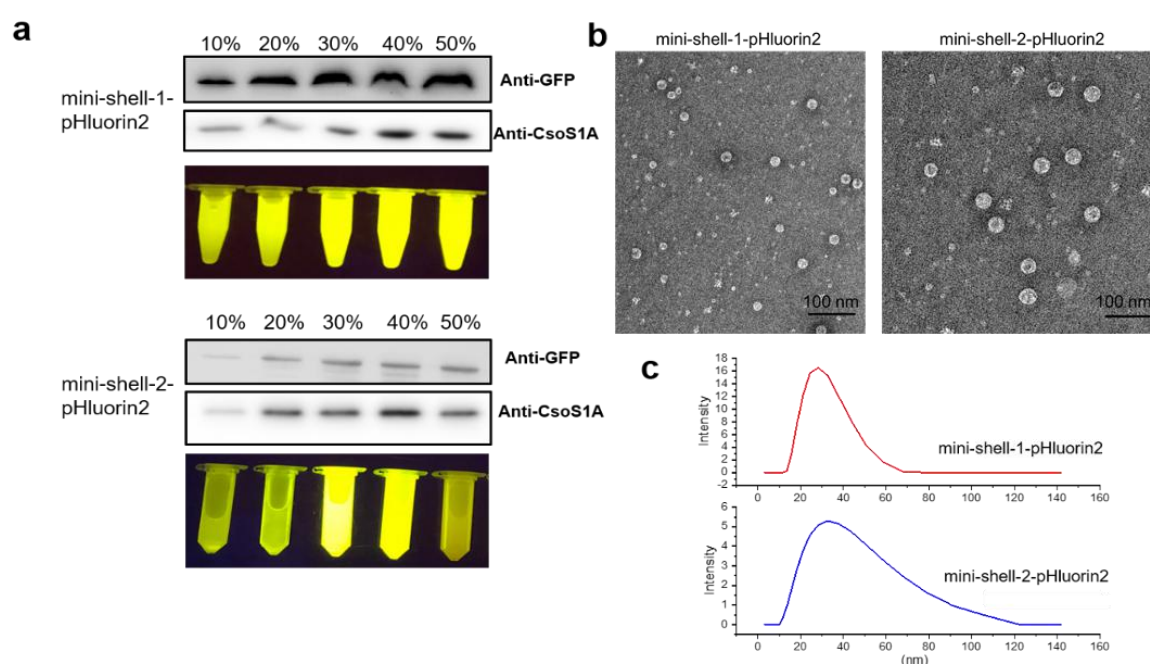
Figure 4-9. Cargo encapsulation within mini-shells. (a) Immunoblot analysis and fluorescence image of purified mini-shell-1/2-pHluorin2 samples. Each sucrose fractions from 10% to 50% was performed using anti-GFP antibody (Agrisera, AS204443, US) and anti-CsoS1A antibody (Agrisera, AS142760, US), respectively. pHluorin2 fluorescence signals of each sucrose fraction was shown by 480 nm LED illumination (Bio-Rad, US). (b) EM of pHluorin2 within mini-shells, using 40% sucrose fractions. Scale bar is 500 nm. (c) DLS of pHluorin2 within mini-shells, using 40% sucrose fractions.

4.3 Conclusion

Carboxysomes are a paradigm of self-assembling protein organelles designed in nature, offering compartmentalisation of enzymes and pathways to enhance carbon fixation. Given their significance in global carbon cycle, carboxysomes are gaining increasing attention from fundamental studies and synthetic engineering, with the intent of generating metabolic factories for sustainably turbocharging carbon fixation and primary production. In this study, we established the pipelines to produce recombinant α-carboxysome mini-shells and study the structural principles underlying shell assembly and encapsulation. Our results provide insight into the physical association between the shell and the linker protein CsoS2, which is vital for encapsulation of cargo enzyme, and the architectures and modulation of mini-shell assemblies mediated by CsoS2. Advanced knowledge may offer new strategies for design and engineering of carboxysome shell-based nanobioreactors and new cages in diverse biotechnological applications, such as enhancement of biocatalysis, food and energy production, molecule delivery, and therapeutics.

Chapter 5 Engineering of simplified α-

carboxysomes in *E. coli*

5.1 Introduction

Cells are able to delegate and govern metabolic processes by self-assemble protein-organelles and molecular-complex, which is important for environmental adaption (Chen and Silver, 2012; Long *et al.*, 2018; Li *et al.*, 2020). The protein-organelles, BMCs, are an excellent example to illustrate how a polyhedral protein shell is formed to enhance and modify cellular metabolism in the organism in question (Chen and Silver, 2012; Liu, 2016). BMCs are found among bacterial phyla (Axen *et al.*, 2014; Sutter *et al.*, 2021) and they sequester enzymes involved in CO_2 fixation and diseases (Yeates *et al.*, 2008a; Yeates *et al.*, 2010; Bobik *et al.*, 2015). Carboxysomes, a family of BMCs, are widespread in cyanobacteria and some chemoautotrophs (Axen *et al.*, 2014; Kerfeld *et al.*, 2018; Liu, 2021a; MacCready and Vecchiarelli, 2021; Sutter *et al.*, 2021). The carboxysome shell encases RuBisCO and concentrates CO_2 near to Rubisco to enhance the catalytic efficiency (Figure 5-1a) (Holthuijzen, 1987; Rae *et al.*, 2013; Fang *et al.*, 2018; Liu, 2021a). Carboxysomes can be divided into two categories based on the classes of RuBisCO contained. α-carboxysomes include Form 1A RuBisCO in α-cyanobacteria and some chemoautotrophs, whereas β-carboxysomes encapsulate Form 1B RuBisCO and are present in β-cyanobacteria (Kerfeld and Melnicki, 2016; Turmo *et al.*, 2017a).

Recombinant expression of α-carboxysomes has been performed in *E. coli* and *Corynebacterium glutamicum*, which demonstrated that α-carboxysomes could correctly assembled with detectable Rubisco activity (Cai *et al.*, 2008; Bonacci *et al.*, 2012; Baumgart *et al.*, 2017; Chen *et al.*, 2021). The β-carboxysome structures of *Synechococcus elongatus* PCC7942 have also been engineered in *E. coli* (Fang *et al.*, 2018). In addition, the β-carboxysomal Rubisco has been expressed in the chloroplasts of *Nicotiana benthamiana* in the presence of CcmM35 (Lin *et al.*, 2014a). A simplified α-carboxysome contain Rubisco

(RbcL and RbcS), CsoS1A and CsoS2 from *Cyanobium marinum* PCC7001 were also generated in the tobacco chloroplasts (Long *et al.*, 2018). Self-assembly allows carboxysomes to be excellent candidates in different hosts to improve photosynthesis (Lin *et al.*, 2014c; Long *et al.*, 2018) and to generate new bio-nanoreactors for the development of metabolic pathways (Gonzalez‑Esquer *et al.*, 2016).

Empty BMC shells have also been generated in *E. coli*, with the variety of functions and morphologies (Tanaka *et al.*, 2008a; Sutter *et al.*, 2017; Sutter *et al.*, 2019b; Kalnins *et al.*, 2020; Li *et al.*, 2020; Tan *et al.*, 2021a). A variety of expression systems have been used to produce the *cso* operon of *H. neapolitanus* encoded α-carboxysomes, demonstrating that the size of the α-carboxysome was variable (Chen *et al.*, 2021). Among the shell variants, the self-assembly of minimal-shell only required the presence of two shell proteins, BMC-H (CsoS1A) and BMC-P (CsoS4A) (Tan *et al.*, 2021a). BMC-H and BMC-P proteins have critical functions in forming the polyhedral shell and control the transit of substrate and product molecules in and out of the shell (Kerfeld *et al.*, 2018). For examples, CsoS1 of *H. neapolitanus* is the hexametric protein, which tile together to form the facets of the α-carboxysomes with the pores that transfer metabolic substrates in and out of the shell (Tsai *et al.*, 2007a; Mahinthichaichan *et al.*, 2018). CsoS4 of *H. neapolitanus* is a pentameric protein, which occupies the vertex of shell to close the carboxysomes (Cai *et al.*, 2009). Besides, the icosahedral shape of the α-carboxysomes is unaffected by the deletion of CsoS4 (Cai *et al.*, 2009). We have expanded the application of α-carboxysome shells by incorporating [FeFe]-hydrogenases (Li *et al.*, 2020) and [NiFe]-hydrogenases (in Chapter 3), using the encapsulation peptide (the C-terminus of CsoS2), to generate new nanobioreactors to boost hydrogen production. In Chapter 4, we have demonstrated the three identified regions of the C-terminus of CsoS2 binding to the shell protein CsoS1A and CsoS4A, resulting in

producing a large mini-shell (T = 9, 36.9 nm in diameter). Except for this, it has been established that the N-terminal of CsoS2 interacts with RuBisCO, and the ability to form carboxysomes was limited by the numbers of the N-peptide repeat of CsoS2 (Oltrogge *et al.*, 2020). These findings indicate that CsoS2, an intrinsically disordered protein (Cai *et al.*, 2015a), is required for α-carboxysome self-assembly. In-depth knowledge of carboxysome composition and assembly mechanism is essential for repurposing carboxysomes for new functions in synthetic biology.

Here, Simplified α-carboxysomes (Simpl-CBs) synthesis from *H. neapolitanus* is described, which might be used as a model system to provide insights into the assembly principles of α-carboxysomes and develop new routes for biotechnological applications. We showed that the Simpl-CBs can be produced utilizing RuBisCO (CbbL/S), CsoS2 and CsoS1A, as well as has RuBisCO activity. The study highlights the structural plasticity and flexibility of native α-carboxysomes.

5.2 Results

5.2.1 Expression of Simpl-CBs in *E. coli*

The *cso* operon of the carboxysome from *H. neapolitanus* was expressed in *E. coli* and resulted in the production of a native-like α-carboxysome that exhibited RuBisCO activity (Cai *et al.*, 2008; Bonacci *et al.*, 2012; Chen *et al.*, 2021) (Figure 5-1b). In Chapter Four, we showed that a mini-shell structure (T =3) formation only requires two shell genes (*csoS1A* and *csoS4A*). These findings point to a synthetic simple α-carboxysome generation as our primary focus. The N-terminal region of CsoS2 contains four repetitive motifs, which is

binding to Rubisco and the production of carboxysomes is restricted by the amount of the repeat motifs (Oltrogge *et al.*, 2020).

To generate a simplified α-carboxysome, the *cbbLS*, *csoS2*, and *csoS1A* genes were inserted into the plasmid pBAD that was driven by the araBAD promoter (Figure 5-1b and 1c) and expressed in *E. coli* induced by L-arabinose at 25 °C for 16 hours. SDS-PAGE of cell extracts of Simpl-CBs showed the expression of the shell proteins (CsoS1A/B/C), Rubisco large and small subunits (CbbL and CbbS), and the shell-associated protein CsoS2B and CsoS2A (Figure 5-1d), consistent with previous studies (Liu *et al.*, 2018a; Chen *et al.*, 2021; Sun *et al.*, 2021). Here, the plasmid pBAD33-S1D (α-CBs), from the Liu Lab at University of Liverpool as a control, was also expressed in *E. coli* by L-arabinose induction at the same condition. SDS-PAGE of cell extracts of synthetic α-CBs showed the genes encoding the protein components CsoS2, Rubisco large and small subunit proteins as well as CsoS1A/B/C were expressed (Figure 5-1d), consistent with previous finding (Chen *et al.*, 2021).
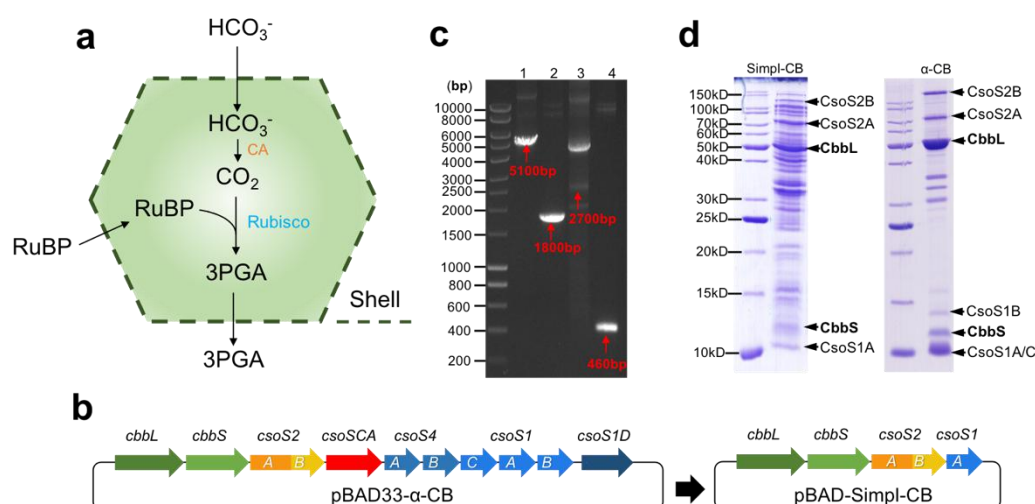
Figure 5-1. Construct of Simpl-CBs. (a) The schematic model of α-carboxysome and metabolic pathway. CA is a shell-associated β-carbonic anhydrase; Rubisco is Form 1A ribulose-1,5-bisphosphate carboxylase/oxygenase. RuBP is ribulose-1,5-bisphosphate; 3PGA is the 3-phosphoglycerate. (b) The genetic organizations of the synthetic cso operon and Simpl-CB operon. The synthetic *cso* operon was a gift from the Liu Lab at the University of Liverpool (Chen *et al*., 2021). (c) The PCR products of Simpl-CBs, 1 shows the PCR product of Simpl-CB operon genes using sequences primers in Table 2-2 (Chapter 2); 2 in Chapter 2 (Materials and Methods) is the PCR product of *cbbLS* genes via using cbbL gene forward primer and sequence reverse primer; 3 is the PCR product of *csoS2* gene via using csoS2 gene forward primer and sequence reverse primer; 4 is the PCR product of csoS1A via using *csoS1A* gene forward primer and sequence reverse primer. (d) SDS-PAGE results of whole cell of Simpl-CB and α-CB.

5.2.2 The purified synthetic Simpl α-carboxysomes

The synthetic Simpl-CBs can be isolated by following the α-carboxysome purification protocol as described in Chapter 2 (Materials and Methods). As a control, recombinant intact α-carboxysomes were expressed using a plasmid pBAD33-α-CB including the *cbbL/S*, *csoS2*, *csoSCA*, *csoS4A/B*, and *csoS1C/A/B/D* genes under the control of an araBAD promoter (Figure 5-1b) induced by L-arabinose, and were isolated using sucrose gradient centrifugation (see Chapter 2, Materials and Methods). SDS-PAGE was used to analyze every sucrose fraction of isolated Simpl-CBs and intact α-CBs (Figure 5-2). Using 75 ng of total protein content from each sucrose fraction, the CbbL/S, CsoS1A, and CsoS2B bands were shown to be stronger in 30 % and 40 % sucrose fractions than in the other sucrose fractions for isolated

Simpl-CBs. The CsoS2A band is weaker than that of CsoS2B, perhaps owing to low expression levels. For the isolated intact α-CBs, the CbbL/S, CsoS1A/B/C, CsoS2B and CsoS2A protein bands appeared in sucrose fractions ranging from 20% to 50%, with the strongest bands appearing in 30% and 40% of the sucrose fractions. Following, immunoblot analysis of CbbL by using an anti-RbcL antibody (Agrisera, US) indicated that the 40% sucrose fraction had the highest protein concentration for both Simpl-CBs and intact α-CBs compared to other fractions of CbbL. (Figure 5-2). Based on the findings of immunoblot analysis, both 40% samples were tested for the next step.
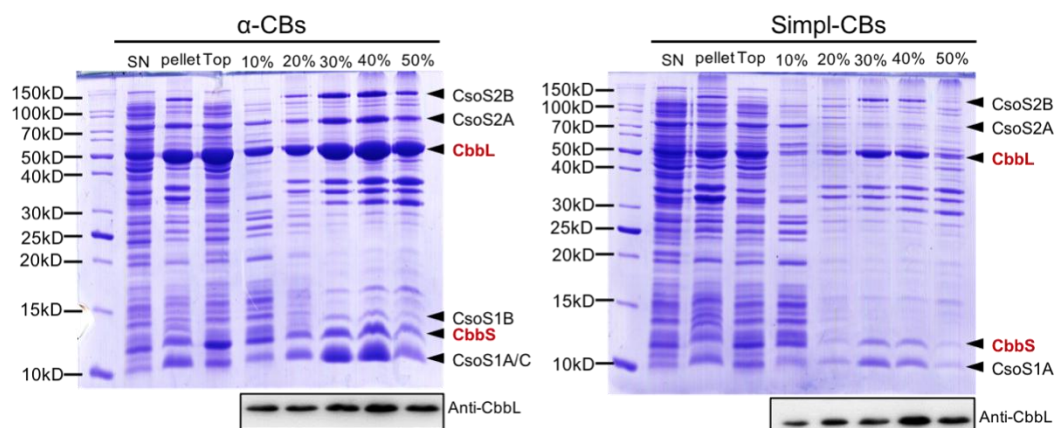
Figure 5-2. The expression of two kinds of recombinant carboxysomes. SDS-PAGE shows the main protein components of isolated two kinds of recombinant α-carboxysomes. Immunoblot analysis of isolated two kinds of recombinant α-carboxysomes by using anti-RbcL antibody (Agrisera, AS03037, US).

Then, we compared the structure and size between 40% fraction of purified Simpl-CBs and intact α-CBs. Both CBs exhibited icosahedral structures with clear edges, as evidenced by electron microscopy (TEM) images (Figure 5-3a). The diameters of the synthetic Simpl-CBs and α-CBs for 40% sucrose fraction are 124.2 ± 8.0 nm ($n = 62$) and 132.6 ± 9.0 nm ($n = 59$), respectively (Figure 5-3b), with $p^{**} = 0.0016$ using two-tailed unpaired *t-test*. These results indicated that the Simpl-CBs could be assembled in *E. coli* and that the reduction of genes might be effect on the size of Simpl-CBs.
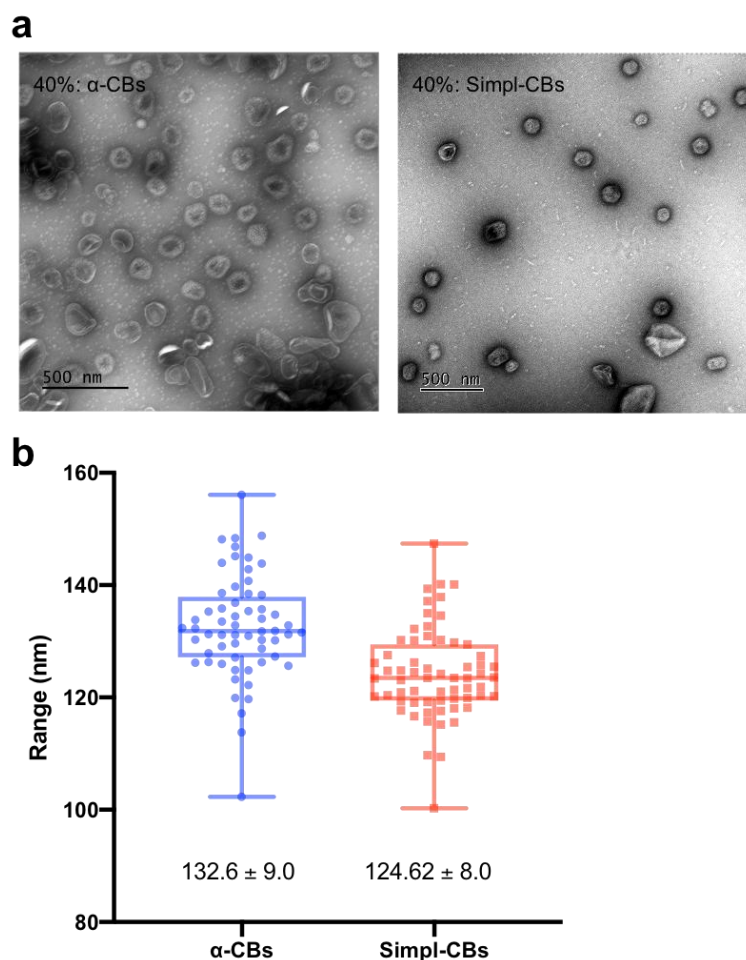
Figure 5-3. Calculation of the size of Simpl-CBs and α-CBs. (a) The TEM images of both recombinant α-CBs from 40% sucrose fraction. Scale bar show 500 nm. (b) The diameters of isolated Simpl-CBs and intact α-CBs were measured by ImageJ based on the TEM images. The data are presented as mean ± standard deviation (SD) , $p^{***} = 0.0016$ (two-tailed unpaired *t-test*).

Next, each isolated sucrose fraction of Simpl-CBs was characterized using electron microscopy. The icosahedral structures with straight edges were found in 20% to 50% sucrose fractions (Figure 5-4), and their shape is similar to that of native and recombinant α-carboxysomes discovered in previous studies (Shively, 1973; Zhao *et al.*, 2019a; Chen *et al.*, 2021). The vertices was seen in the icosahedral structure of Simpl-CBs, even without the pentameric CsoS4 protein. Previous studies also found *H. neapolitanus* cells without *csoS4A/B* could still preserve the regular shape of α-carboxysomes (Cai *et al.*, 2009); similar

structures have also been seen in the tobacoo engineered α-CBs without CsoS4 (Long *et al.*, 2018).
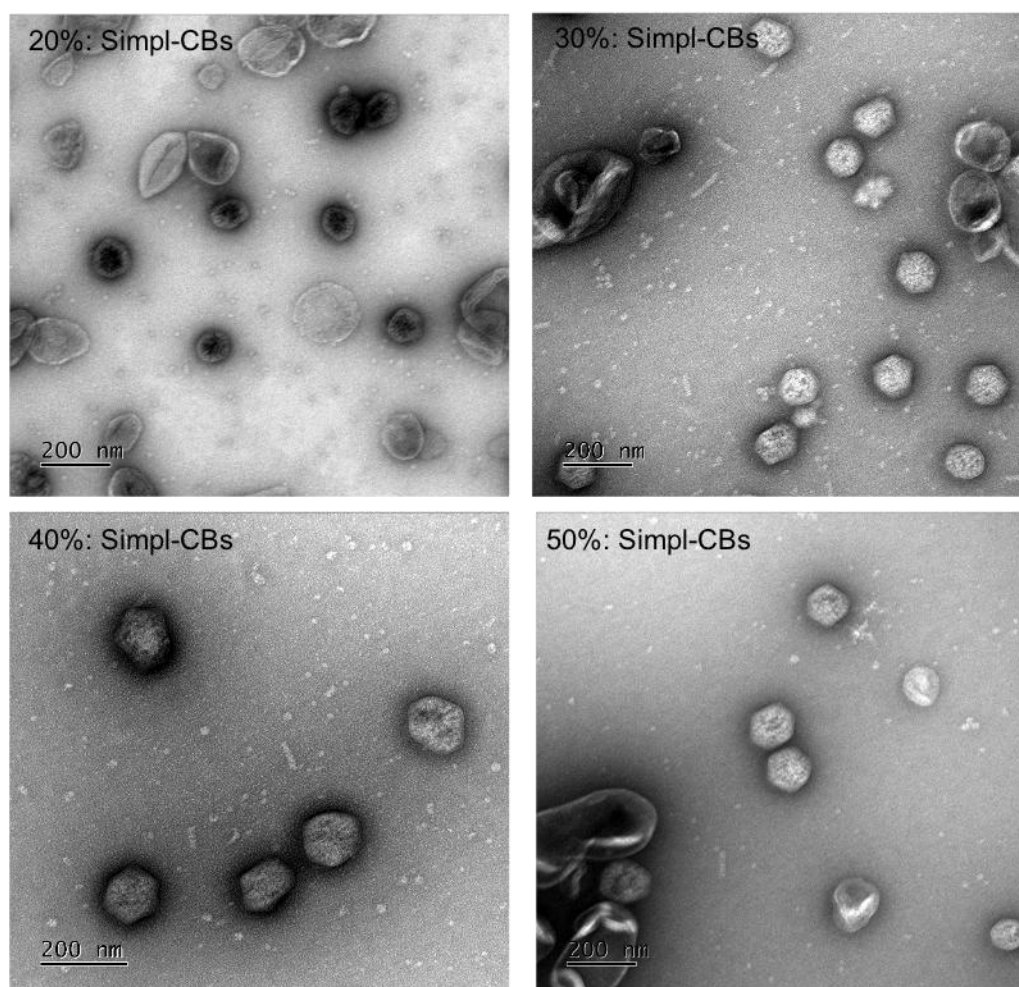
Figure 5-4. TEM images of Simpl-CBs in variety of sucrose fractions. Scale bar shows 200 nm.

We also calculated the size of Simpl-CBs particles in each sucrose fraction using ImageJ (Figure 5-5). The average diameter of the recombinant Simpl-CBs in the 20% sucrose fraction is 115.1 ± 6.4 nm (mean $\pm$ SD, $n = 34$), which is smaller than those of other fractions ($p^{***} = 0.0005$, two-way ANOVA). The average diameters of Simpl-CBs in 30% sucrose fraction is 123.5 ± 9.7 nm ($n = 55$), in 40% sucrose fraction is 124.2 ± 8.0 nm ($n = 62$), and in 50% sucrose fraction is 121.7 ± 8.4 nm ($n = 41$), one way ANOVA analyzed there was no significant difference in 30% to 50% sucrose fractions ($p = 0.424$).
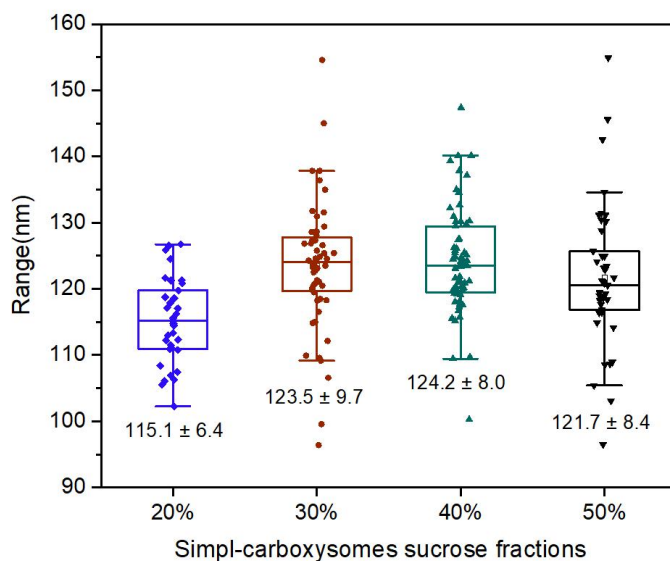
Figure 5-5. The diameters of isolated Simpl-carboxysomes measured by Image J based on the TEM images. The data are presented as mean ± standard deviation (SD).

5.2.3 Activity of synthetic Simpl α-carboxysomes (Simpl-CBs)

To further evaluate the Rubisco activity of synthetic Simpl-CBs, we carried out ^{14}C radiometric Rubisco assays to determine the carbon fixation activities of purified synthetic Simpl-CBs in all the sucrose fractions, and synthetic intact α-CBs (from the same fractions) were used as a control. The CO_2 fixation rate was monitored when the substrate $^{14}CO_2$ ($NaH^{14}CO_3$ dehydrated into $^{14}CO_2$) was provided to a variety of the concentration of D-ribulose 1,5-bisphosphate sodium salt hydrate (RuBP), and the results were fitted with a Michaelis-Menten kinetic model (Figure 5-6). We first determined the CO_2 fixation rate for each sucrose fraction under 0.6 mM of RuBP. It was discovered that the Rubisco activity of 40% sucrose fraction for α-CBs have maximum activity of 2.5 ± 0.06 μmol min^{-1} mg^{-1} and 40 % sucrose fraction for Simpl-CBs 1.2 ± 0.12 μmol min^{-1} mg^{-1} (Figure 5-6a). Then, the 40% sucrose fractions of Simpl-CBs and α-CBs were used to evaluate the CO_2 fixation rate under a range of RuBP concentrations. The synthetic Simpl-CBs has the detectable Rubisco activity, with a Vmax = 1.44 ± 0.12 μmol min^{-1} mg^{-1} and Km (RuBP) of 141.7 ± 11.7 μM (n = 3). In

contrast, the α-CBs have a $Vmax = 2.77 \pm 0.24$ µmol min^{-1} mg^{-1} and $Km_{(RuBP)}$ of 389.3 ± 45.3 µM (n = 3) (Figure 5-6b). The Rubisco activity of Simpl-CBs is lower than α-CBs, likely because Simpl-CBs presents an unclosed structure in the absence of the pentamer vertex protein, results in CO_2 leaking. These results were consistent with previous observations that deleting the pentamers resulted in the impaired function of α-carboxysomes (Cai *et al.*, 2009).
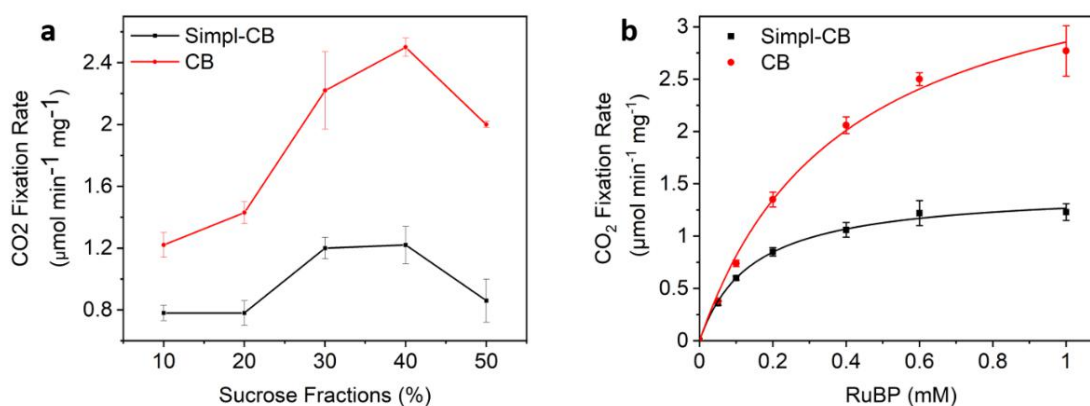
Figure 5-6. CO_2 fixation activity of purified synthetic Simpl-CBs. (a) CO_2 fixation rate of each sucrose fraction for both synthetic α-carboxysomes. (b) The Michaelis-Menten model was used to analysis the function of both synthetic α-carboxysomes under a variety of RuBP concentration. Intact α-carboxysome have a *Vmax* of 2.77 ± 0.24 µmol min^{-1} mg^{-1} and $K_{m(RuBP)}$ of 389.3 ± 45.3 µM (*n* = 3), synthetic simpl-α-carboxysomes have a V_{max} of 1.44 ± 0.12 µmol min^{-1} mg^{-1} and $K_{m(RuBP)}$ of 141.7 ± 11.7 µM (*n* = 3). The data are presented as mean ± standard deviation (SD) for at least three independent repeats.

5.3 Discussion

Prior studies have shown that recombinant α-carboxysomes can be generated by expressing the *cso* operon of *H. neapolitanu*s in *E. coli* (Cai *et al.*, 2008; Bonacci *et al.*, 2012; Chen *et al.*, 2021). A similar icosahedral shape to that of the native α-carboxysomes is maintained in the minimal α-carboxysomes presented here, without the interior enzyme carbonic anhydrase (CsoSCA), pentameric shell proteins (CsoS4), and gated shell protein (CsoS1D). Despite the absence of specific components, we confirmed that a simplified α-carboxysome can be self-assembled by Rubisco, CsoS1A, and CsoS2, with the ability of fixing CO_2.

Our data indicated that the Rubisco activity of Simpl-CBs is lower than α-CBs, which might be caused by unclosed structure. The optimization of the operon expressing Simpl-CBs could as a promising area for improving the function of synthetic Simpl-CBs. Previous studies have suggested that the CsoS4A/B pentamer proteins function as vertex to close the α-carboxysomes (Tanaka *et al.*, 2008a; Zhao *et al.*, 2019b). The loss of CsoS4A/B proteins may result in the impeded shell permeability and a reduction in the catalytic advantage of carboxysomes (Zhao *et al.*, 2019b). As shown in Chapter 4, only CsoS1A and CsoS4A proteins are required for mini-shell formation. Thus, a feasible option in the future would be to include *csoS4A* into the current operon. This new vector will allow us to detect whether introducing extra components could impact the size and shape of carboxysomes and to compare changes in Rubisco activity. CsoSCA is a major factor in Rubisco activity (Dou *et al.*, 2008). It has been recently proposed that the N-terminal peptide of CsoSCA interacts with Rubisco and aids in the encapsulation of Rubisco in carboxysomes (Blikstad *et al.*, 2021). The CO_2 fixation rate of Simpl-CBs may be affected in the absence of CsoSCA. Another optimization of the Simpl-CBs is therefore to add CsoSCA into the Simpl-CBs, which may boost the Rubisco activity of Simpl-CBs.

In summary, the generation of Simpl-CBs provide a framework for understanding the structural and functional requirements of each carboxysome component and modifying the engineering strategies to produce functional recombinant carboxysomes in heterologous hosts to boost carbon fixation.

Chapter 6 Conclusions and Perspectives

6.1 Conclusions

Carboxysomes are self-assembling proteinaceous organelles that play crucial roles in CO_2 fixation. Engineering carboxysomes in heterologous hosts for new functions has attracted increasing attention in recent years. In my PhD study, I performed synthetic engineering of empty α-carboxysome shells for hydrogen production in *E. coli* (Chapter 3), α-carboxysome mini-shells in *E. coli* to study self-assembly principles of shell proteins (Chapter 4), and simplified α-carboxysomes with carboxylation activities in *E. coli* (Chapter 5).

In Chapter 3, I constructed the pCDFDuet-*hyaAB*-EP vector (*hyaAB*-EP) to produce recombinant oxygen-tolerant [NiFe]-hydrogenase 1, with the C-terminus of CsoS2 (encapsulated peptide, EP) fused to the C-terminus of the large (HyaB) and small (HyaA) subunits of [NiFe]-hydrogenase 1. I co-expressed *hyaAB*-EP with two different form of α-carboxysome shells (Shell-1/-2) in *E. coli* BL21(DE3). The recombinant [NiFe]-hydrogenase 1 was encased inside the shells as confirmed by SDS-PAGE, immunoblot, and electron microscopy (EM). Furthermore, the activities of recombinant [NiFe]-hydrogenase within the carboxysome shells were determined by using methyl viologen (MV) *in vitro* and endogenous NADPH *in vivo*. The results showed that [NiFe]-hydrogenase-shells could enhance the hydrogenase activity by 10 folds *in vitro* and 20 folds *in vivo*, compared with free [NiFe]-hydrogenase activities. I also characterised the effects of temperature, O_2 exposure, and time on hydrogenase activity *in vitro*. These results indicated that α-carboxysome shells represent a promising system to be reprogrammed as nanoreactors in biotechnological applications, such as improvement of catalytic pathways, protection of enzymes, and development of drug delivery vehicles.

To carry out systematical structural analysis of α-carboxysome self-assembly, in Chapter 4, I generated three distinct types of α-carboxysome mini-shells (mini-shell-1/2/3) in *E. coli* and used biochemistry, negative staining EM and cryo-EM to study the mini-shell architectures. We found that the mini-shell formation requires only two types of shell proteins, CsoS1A hexamers and CsoS4A pentamers. The self-assembly of CsoS1A and CsoS4A resulted in formation of two types of mini-shells, T=3 and T=4 shells. In addition, addition of CsoS2 led to the formation of T=3 mini-shells. In addition, cro-EM analysis revealed three regions of CsoS2 C-terminus form specific binding with CsoS1A and CsoS4A. Characterization of the assembly of mini-shells with truncated C-terminus of CsoS2 revealed that the binding regions affect the efficiency of mini-shell formation. These results shed light on protein-protein interactions between shell proteins ad between CsoS2 and shells proteins, which are essential for shell formation and cargo encapsulation, as well as the structural plasticity of carboxysome shells.

In Chapter 5, I generated simplified α-carboxysomes (Simpl-CBs) in *E. coli*, which include Rubisco, CsoS2, and CsoS1A. The Simpl α-CBs have an icosahedral shape and possess a diameter of ~ 100 nm, roughly comparable to native and synthetic α-carboxysomes but drastically larger than mini-shells. The carbon-fixation activity of Simpl-CBs was confirmed by Rubisco assays *in vitro*, although their Rubisco activity was lower than synthetic α-carboxysomes. The production and characterization of Simpl-CBs CBs opened the door for in-depth understanding the assembly principles of α-carboxysomes, and highlighted the possibility to modulate the protein composition and function of engineered carboxysomes in non-native hosts, such as crop plants, to enhance photosynthesis and growth.

6.2 Perspectives

We have experimentally shown that the C-terminus of CsoS2 might function as encapsulated peptide for the incorporation of non-native enzymes into α-carboxysomes shells (Li *et al.*, 2020). Thus, it is possible to use CsoS2 C-terminus as an encapsulated peptide to direct various cargos into carboxysome shells to generate and modulate different nanobioreactors for new functions. Other encapsulation strategies may be also considered for optimizing cargo encapsulation efficiency. Moreover, advanced techniques will need to be developed to evaluate and adjust precisely the efficiency and quantity of non-native cargo encapsulation within α-carboxysomes.

Efficient production and function of simplified α-carboxysomes hold great promising for installing functional carboxysome structures in crop plants to maximise crop performance and yields, although the current Rubisco activity of simplified α-carboxysomes was relatively low. Recently, we have also shown incorporation of functional Rubisco activases, CbbO and CbbQ could enhance CO_2 fixation of α-carboxysomes (Chen *et al.*, 2021). It is thus possible to improve the activities of the simplified α-carboxysomes by adding essential shell components, such as CsoS4A, to completely seal the shells, adding carbonic anhydrase to accumulate CO_2 levels for further exploration.

References

ABDUL-RAHMAN, F., PETIT, E. & BLANCHARD, J. L. 2013. The distribution of polyhedral bacterial microcompartments suggests frequent horizontal transfer and operon reassembly. *J Phylogenetics Evol Biol,* 1**,** 118.

ADAMS, M. W. & HALL, D. O. 1979. Purification of the membrane-bound hydrogenase of Escherichia coli. *Biochem J,* 183**,** 11-22.

AFTAB, S., SHAH, A., NADHMAN, A., KURBANOGLU, S., OZKAN, S. A., DIONYSIOU, D. D., SHUKLA, S. S. & AMINABHAVI, T. M. 2018. Nanomedicine: An effective tool in cancer therapy. *International journal of pharmaceutics,* 540**,** 132-149.

AKKERMAN, I., JANSSEN, M., ROCHA, J. & WIJFFELS, R. H. 2002. Photobiological hydrogen production: photochemical efficiency and bioreactor design. *International Journal of Hydrogen Energy,* 27**,** 1195-1208.

ANDERSSON, I. 1996. Large structures at high resolution: the 1.6 Å crystal structure of spinach ribulose-1, 5-bisphosphate carboxylase/oxygenase complexed with 2-carboxyarabinitol bisphosphate. *Journal of molecular biology,* 259**,** 160-174.

ANDERSSON, I. & BACKLUND, A. 2008. Structure and function of Rubisco. *Plant Physiology and Biochemistry,* 46**,** 275-291.

ANDERSSON, I., KNIGHT, S., SCHNEIDER, G., LINDQVIST, Y., LUNDQVIST, T., BRÄNDÉN, C.-I. & LORIMER, G. H. 1989. Crystal structure of the active site of ribulose-bisphosphate carboxylase. *Nature,* 337**,** 229-234.

AUSSIGNARGUES, C., PAASCH, B. C., GONZALEZ-ESQUER, R., ERBILGIN, O. & KERFELD, C. A. 2015. Bacterial microcompartment assembly: The key role of encapsulation peptides. *Communicative & Integrative Biology,* 8**,** e1039755.

AXEN, S. D., ERBILGIN, O. & KERFELD, C. A. 2014. A Taxonomy of Bacterial Microcompartment Loci Constructed by a Novel Scoring Method. *PLOS Computational Biology,* 10**,** e1003898.

BADGER, M. R., ANDREWS, T. J., WHITNEY, S., LUDWIG, M., YELLOWLEES, D. C., LEGGAT, W. & PRICE, G. D. 1998. The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO2-concentrating mechanisms in algae. *Canadian Journal of Botany,* 76**,** 1052-1071.

BADGER, M. R., HANSON, D. & PRICE, G. D. 2002. Evolution and diversity of CO2 concentrating mechanisms in cyanobacteria. *Functional Plant Biology,* 29**,** 161-173.

BADGER, M. R. & PRICE, G. D. 2003. CO2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *Journal of experimental botany,* 54**,** 609-622.

BAE, S., ULRICH, C. M., NEUHOUSER, M. L., MALYSHEVA, O., BAILEY, L. B., XIAO, L., BROWN, E. C., CUSHING-HAUGEN, K. L., ZHENG, Y. & CHENG, T.-Y. D. 2014. Plasma choline metabolites and colorectal cancer risk in the Women's Health Initiative Observational Study. *Cancer research,* 74**,** 7442-7452.

BAKER, S. H., LORBACH, S. C., RODRIGUEZ-BUEY, M., WILLIAMS, D. S., ALDRICH, H. C. & SHIVELY, J. M. 1999. The correlation of the gene csoS2 of the carboxysome operon with two polypeptides of the carboxysome in Thiobacillus neapolitanus. *Archives of Microbiology,* 172**,** 233-239.

BAKER, S. H., WILLIAMS, D. S., ALDRICH, H. C., GAMBRELL, A. C. & SHIVELY, J. M. 2000. Identification and localization of the carboxysome peptide Csos3 and its corresponding gene in Thiobacillus neapolitanus. *Archives of microbiology,* 173**,** 278-283.

BAUMGART, M., HUBER, I., ABDOLLAHZADEH, I., GENSCH, T. & FRUNZKE, J. 2017. Heterologous expression of the Halothiobacillus neapolitanus carboxysomal gene cluster in Corynebacterium glutamicum. *Journal of Biotechnology,* 258**,** 126-135.

BISAILLON, A., TURCOT, J. & HALLENBECK, P. 2006. The effect of nutrient limitation on hydrogen production by batch cultures of Escherichia coli. *International Journal of Hydrogen Energy,* 31**,** 1504-1508.

BLIKSTAD, C., DUGAN, E., LAUGHLIN, T., LIU, M., SHOEMAKER, S., REMIS, J. & SAVAGE, D. 2021. Discovery of a carbonic anhydrase-Rubisco supercomplex within the alpha-carboxysome. bioRxiv.

BOBIK, T. A., HAVEMANN, G. D., BUSCH, R. J., WILLIAMS, D. S. & ALDRICH, H. C. 1999. The Propanediol Utilization Operon of Salmonella enterica Serovar Typhimurium LT2 Includes Genes Necessary for Formation of Polyhedral Organelles Involved in Coenzyme B_{12}-Dependent 1,2-Propanediol Degradation. *Journal of Bacteriology,* 181**,** 5967-5975.

BOBIK, T. A., LEHMAN, B. P. & YEATES, T. O. 2015. Bacterial microcompartments: widespread prokaryotic organelles for isolation and optimization of metabolic pathways. *Molecular Microbiology,* 98**,** 193-207.

BOBIK, T. A., XU, Y., JETER, R. M., OTTO, K. E. & ROTH, J. R. 1997. Propanediol utilization genes (pdu) of Salmonella typhimurium: three genes for the propanediol dehydratase. *J Bacteriol,* 179**,** 6633-9.

BÖHM R, S. M., BÖCK A. 1990. Nucleotide sequence and expression of an operon in Escherichia coli coding for formate hydrogenlyase components. *Molecular Microbiology,* 4**,** 13.

BONACCI, W., TENG, P. K., AFONSO, B., NIEDERHOLTMEYER, H., GROB, P., SILVER, P. A. & SAVAGE, D. F. 2012. Modularity of a carbon-fixing protein organelle. *Proc Natl Acad Sci U S A,* 109**,** 478-83.

CAI, F., BERNSTEIN, S. L., WILSON, S. C. & KERFELD, C. A. 2016. Production and Characterization of Synthetic Carboxysome Shells with Incorporated Luminal Proteins. *Plant Physiol,* 170**,** 1868-77.

CAI, F., DOU, Z., BERNSTEIN, S. L., LEVERENZ, R., WILLIAMS, E. B., HEINHORST, S., SHIVELY, J., CANNON, G. C. & KERFELD, C. A. 2015a. Advances in Understanding Carboxysome Assembly in Prochlorococcus and Synechococcus Implicate CsoS2 as a Critical Component. *Life (Basel),* 5**,** 1141-71.

CAI, F., HEINHORST, S., SHIVELY, J. M. & CANNON, G. C. 2008. Transcript analysis of the Halothiobacillus neapolitanus cso operon. *Archives of Microbiology,* 189**,** 141-150.

CAI, F., MENON, B. B., CANNON, G. C., CURRY, K. J., SHIVELY, J. M. & HEINHORST, S. 2009. The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a CO2 leakage barrier. *PLoS One,* 4**,** e7521.

CAI, F., SUTTER, M., BERNSTEIN, S. L., KINNEY, J. N. & KERFELD, C. A. 2015b. Engineering Bacterial Microcompartment Shells: Chimeric Shell Proteins and Chimeric Carboxysome Shells. *ACS Synthetic Biology,* 4**,** 444-453.

CAI, F., SUTTER, M., CAMERON, J. C., STANLEY, D. N., KINNEY, J. N. & KERFELD, C. A. 2013. The structure of CcmP, a tandem bacterial microcompartment domain protein from the β-carboxysome, forms a subcompartment within a microcompartment. *Journal of Biological Chemistry,* 288**,** 16055-16063.

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. & SHIVELY, J. M. 2003. Organization of Carboxysome Genes in the Thiobacilli. *Current Microbiology,* 46**,** 0115-0119.

CANNON, G. C., BRADBURNE, C. E., ALDRICH, H. C., BAKER, S. H., HEINHORST, S. & SHIVELY, J. M. 2001. Microcompartments in prokaryotes: carboxysomes and related polyhedra. *Applied and environmental microbiology,* 67**,** 5351-5361.

CANNON, G. C., HEINHORST, S., BRADBURNE, C. E. & SHIVELY, J. M. 2002. Carboxysome genomics: a status report. *Functional Plant Biology,* 29**,** 175-182.

CANNON, G. C., HEINHORST, S. & KERFELD, C. A. 2010. Carboxysomal carbonic anhydrases: structure and role in microbial CO2 fixation. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics,* 1804**,** 382-392.

CASADABAN, M. J. & COHEN, S. N. 1980. Analysis of gene control signals by DNA fusion and cloning in Escherichia coli. *Journal of molecular biology,* 138**,** 179-207.

CASTRESANA, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution,* 17**,** 540-552.

CATAL, T., LESNIK, K. L. & LIU, H. 2015. Suppression of methanogenesis for hydrogen production in single-chamber microbial electrolysis cells using various antibiotics. *Bioresour Technol,* 187**,** 77-83.

CHAIJARASPHONG, T., NICHOLS, R. J., KORTRIGHT, K. E., NIXON, C. F., TENG, P. K., OLTROGGE, L. M. & SAVAGE, D. F. 2016. Programmed Ribosomal Frameshifting Mediates Expression of the alpha-Carboxysome. *J Mol Biol,* 428**,** 153-164.

CHEN, A. H. & SILVER, P. A. 2012. Designing biological compartmentalization. *Trends in cell biology,* 22**,** 662-670.

CHEN, T., FANG, Y., JIANG, Q., DYKES, G. F., LIN, Y., PRICE, G. D., LONG, B. M. & LIU, L.N. 2021. Incorporation of Functional Rubisco Activases into Engineered Carboxysomes to Enhance Carbon Fixation. *ACS Synthetic Biology*.

CHEN T, FANG. Y., JIANG Q, DYKES GF, LIN Y, PRICE GD, LONG BM, LIU LN. 2022. Incorporation of Functional Rubisco Activases into Engineered Carboxysomes to Enhance Carbon Fixation. *ACS Synth Biol,* 11**,** 8.

CHEN, Y.M., LIU, Y., ZHOU, R.F., CHEN, X.L., WANG, C., TAN, X.Y., WANG, L.J., ZHENG, R.D., ZHANG, H.W. & LING, W.H. 2016. Associations of gut-flora-dependent metabolite trimethylamine-N-oxide, betaine and choline with non-alcoholic fatty liver disease in adults. *Scientific reports,* 6**,** 1-9.

CHERYL A. KERFELD, M. R. S., SHIHO TANAKA, CHAU V. NGUYEN, MARTIN PHILLIPS, MORGAN BEEBY, TODD O. YEATES 2005. Protein Structures Forming the Shell of Primitive Bacterial Organelles. *Science,* 309**,** 936-938.

CHOUDHARY, S., QUIN, M. B., SANDERS, M. A., JOHNSON, E. T. & SCHMIDT-DANNERT, C. 2012. Engineered protein nano-compartments for targeted enzyme localization. *PLoS One,* 7**,** e33342.

CHOWDHURY, C., CHUN, S., PANG, A., SAWAYA, M. R., SINHA, S., YEATES, T. O. & BOBIK, T. A. 2015. Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proc Natl Acad Sci U S A,* 112**,** 2990-5.

CHOWDHURY, C., SINHA, S., CHUN, S., YEATES, T. O. & BOBIK, T. A. 2014. Diverse bacterial microcompartment organelles. *Microbiol Mol Biol Rev,* 78**,** 438-68.

COLOMER-LLUCH, M., JOFRE, J. & MUNIESA, M. 2014. Quinolone resistance genes (qnrA and qnrS) in bacteriophage particles from wastewater samples and the effect of inducing agents on packaged antibiotic resistance genes. *J Antimicrob Chemother,* 69**,** 1265-74.

COT, S. S.W., SO, A. K.C. & ESPIE, G. S. 2008. A multiprotein bicarbonate dehydration complex essential to carboxysome function in cyanobacteria. *Journal of bacteriology,* 190**,** 936-945.

CRACIUN, S and BALSKUS, E.P. 2012. Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proceedings of the National Academy of Sciences*,109(52), 21307-21312.

CRACKNELL, J. A., WAIT, A. F., LENZ, O., FRIEDRICH, B. & ARMSTRONG, F. A. 2009. A kinetic and thermodynamic understanding of O_2 tolerance in [NiFe]-hydrogenases. *Proc Natl Acad Sci U S A,* 106**,** 20681-6.

CROWLEY, C. S., SAWAYA, M. R., BOBIK, T. A. & YEATES, T. O. 2008. Structure of the PduU shell protein from the Pdu microcompartment of Salmonella. *Structure,* 16**,** 1324-1332.

DASGUPTA, C. N., JOSE GILBERT, J., LINDBLAD, P., HEIDORN, T., BORGVANG, S. A., SKJANES, K. & DAS, D. 2010. Recent trends on the development of photobiological processes and photobioreactors for the improvement of hydrogen production. *International Journal of Hydrogen Energy,* 35**,** 10218-10238.

DOU, Z., HEINHORST, S., WILLIAMS, E. B., MURIN, C. D., SHIVELY, J. M. & CANNON, G. C. 2008. CO2 fixation kinetics of Halothiobacillus neapolitanus mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO2. *J Biol Chem,* 283**,** 10377-84.

DREWS, G. & NIKLOWITZ, W. 1956a. Beiträge zur Cytologie der Blaualgen. *Archiv für Mikrobiologie,* 24**,** 147-162.

DREWS, G. & NIKLOWITZ, W. 1956b. [Cytology of Cyanophycea. II. Centroplasm and granular inclusions of Phormidium uncinatum]. *Arch Mikrobiol,* 24**,** 147-62.

ENGLISH, R. S., LORBACH, S. C., QIN, X. & SHIVELY, J. M. 1994. Isolation and characterization of a carboxysome shell gene from Thiobacillus neapolitanus. *Molecular Microbiology,* 12**,** 647-654.

ERBILGIN, O., MCDONALD, K. L. & KERFELD, C. A. 2014. Characterization of a planctomycetal organelle: a novel bacterial microcompartment for the aerobic degradation of plant saccharides. *Applied and environmental microbiology,* 80**,** 2193-2205.

ERBILGIN, O., SUTTER, M. & KERFELD, C. A. 2016. The Structural Basis of Coenzyme A Recycling in a Bacterial Organelle. *PLOS Biology,* 14**,** e1002399.

ESPIE, G. S. & KIMBER, M. S. 2011. Carboxysomes: cyanobacterial RubisCO comes in small packages. *Photosynthesis Research,* 109**,** 7-20.

EVANS, R. M., PARKIN, A., ROESSLER, M. M., MURPHY, B. J., ADAMSON, H., LUKEY, M. J., SARGENT, F., VOLBEDA, A., FONTECILLA-CAMPS, J. C. & ARMSTRONG, F. A. 2013. Principles of sustained enzymatic hydrogen oxidation in the presence of oxygen--the crucial influence of high potential Fe-S clusters in the electron relay of [NiFe]-hydrogenases. *J Am Chem Soc,* 135**,** 2694-707.

FAN, C. & BOBIK, T. A. 2011. The N-Terminal Region of the Medium Subunit (PduD) Packages Adenosylcobalamin-Dependent Diol Dehydratase (PduCDE) into the Pdu Microcompartment. *Journal of Bacteriology,* 193**,** 5623-5628.

FAN, C., CHENG, S., LIU, Y., ESCOBAR, C. M., CROWLEY, C. S., JEFFERSON, R. E., YEATES, T. O. & BOBIK, T. A. 2010. Short N-terminal sequences package proteins into bacterial microcompartments. *Proc Natl Acad Sci U S A,* 107**,** 7509-14.

FAN, C., CHENG, S., SINHA, S. & BOBIK, T. A. 2012. Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. *Proc Natl Acad Sci U S A,* 109**,** 14995-5000.

FANG, S., HUANG, X., ZHANG, X., ZHANG, M., HAO, Y., GUO, H., LIU, L.-N., YU, F. & ZHANG, P. 2021. Molecular mechanism underlying transport and allosteric inhibition of bicarbonate transporter SbtA. *Proceedings of the National Academy of Sciences,* 118.

FANG, Y., HUANG, F., FAULKNER, M., JIANG, Q., DYKES, G. F., YANG, M. & LIU, L. N. 2018. Engineering and Modulating Functional Cyanobacterial CO2-Fixing Organelles. *Front Plant Sci,* 9**,** 739.

FAULKNER, M., RODRIGUEZ-RAMOS, J., DYKES, G. F., OWEN, S. V., CASELLA, S., SIMPSON, D. M., BEYNON, R. J. & LIU, L. N. 2017. Direct characterization of the native structure and mechanics of cyanobacterial carboxysomes. *Nanoscale,* 9**,** 10662-10673.

FAULKNER, M., SZABÓ, I., WEETMAN, S. L., SICARD, F., HUBER, R. G., BOND, P. J., ROSTA, E. & LIU, L.-N. 2020. Molecular simulations unravel the molecular principles that mediate selective permeability of carboxysome shell protein. *Scientific reports,* 10**,** 1-14.

FERLEZ, B., SUTTER, M. & KERFELD, C. A. 2019. Glycyl Radical Enzyme-Associated Microcompartments: Redox-Replete Bacterial Organelles. *mBio,* 10.

FLANAGAN, L. A., WRIGHT, J. J., ROESSLER, M. M., MOIR, J. W. & PARKIN, A. 2016. Re-engineering a NiFe hydrogenase to increase the H_2 production bias while maintaining native levels of O2 tolerance. *Chem Commun (Camb),* 52**,** 9133-6.

FORZI, L. & SAWERS, R. G. 2007. Maturation of [NiFe]-hydrogenases in Escherichia coli. *Biometals,* 20**,** 565-78.

TABITA, F.R. 1999. Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: a different perspective. *Photosynthesis Research,* 60**,** 28.

FREY 2002. Hydrogenases: Hydrogen-Activating Enzymes. *Journal of chemical biology,* 1**,** 2-3.

FREY, R., MANTRI, S., ROCCA, M. & HILVERT, D. 2016. Bottom-up construction of a primordial carboxysome mimic. *Journal of the American Chemical Society,* 138**,** 10072-10075.

GAO, R., TAN, H., LI, S., MA, S., TANG, Y., ZHANG, K., ZHANG, Z., FAN, Q., YANG, J., ZHANG, X. E. & LI, F. 2022. A prototype protein nanocage minimized from carboxysomes with gated oxygen permeability. *Proc Natl Acad Sci U S A,* 119.

GHIRARDI, M. L., DUBINI, A., YU, J. & MANESS, P. C. 2009. Photobiological hydrogen-producing systems. *Chem Soc Rev,* 38**,** 52-61.

GIBSON, D. G., GLASS, J. I., LARTIGUE, C., NOSKOV, V. N., CHUANG, R.-Y., ALGIRE, M. A., BENDERS, G. A., MONTAGUE, M. G., MA, L., MOODIE, M. M., MERRYMAN, C., VASHEE, S., KRISHNAKUMAR, R., ASSAD-GARCIA, N., ANDREWS-PFANNKOCH, C., DENISOVA, E. A., YOUNG, L., QI, Z.-Q., SEGALL-SHAPIRO, T. H., CALVEY, C. H., PARMAR, P. P., HUTCHISON, C. A., SMITH, H. O. & VENTER, J. C. 2010. Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science,* 329**,** 52-56.

GIBSON, D. G., YOUNG, L., CHUANG, R.-Y., VENTER, J. C., HUTCHISON, C. A. & SMITH, H. O. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods,* 6**,** 343-345.

GONZALES, A. D., LIGHT, Y. K., ZHANG, Z., IQBAL, T., LANE, T. W. & MARTINO, A. 2005. Proteomic analysis of the CO_2-concentrating mechanism in the open-ocean cyanobacterium Synechococcus WH8102. *Canadian Journal of Botany,* 83**,** 735-745.

GONZALES, M. F., BROOKS, T., PUKATZKI, S. U. & PROVENZANO, D. 2013. Rapid protocol for preparation of electrocompetent Escherichia coli and Vibrio cholerae. *Journal of visualized experiments: JoVE.*

GONZALEZ-ESQUER, C. R., NEWNHAM, S. E. & KERFELD, C. A. 2016. Bacterial microcompartments as metabolic modules for plant synthetic biology. *The Plant Journal,* 87**,** 66-75.

GOUY, M., GUINDON, S. & GASCUEL, O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution,* 27**,** 221-224.

GREBER, B. J., SUTTER, M. & KERFELD, C. A. 2019. The Plasticity of Molecular Interactions Governs Bacterial Microcompartment Shell Assembly. *Structure,* 27**,** 749-763 e4.

GREENING, C. & LITHGOW, T. 2020. Formation and function of bacterial organelles. *Nature Reviews Microbiology,* 18**,** 677-689.

GUO, X., TRABLY, E., LATRILLE, E., E., CARRÈRE, H. & STEYER, J.-P. 2014. Predictive and explicative models of fermentative hydrogen production from solid organic waste: Role of butyrate and lactate pathways. *International Journal of Hydrogen Energy,* 39**,** 7476-7485.

ATOMI, H. 2002. Microbial enzymes involved in carbon dioxide fixation. *Journal of Bioscience and Bioengineering* 94**,** 9.

HAGEN, A., SUTTER, M., SLOAN, N. & KERFELD, C. A. 2018a. Programmed loading and rapid purification of engineered bacterial microcompartment shells. *Nature communications,* 9**,** 1-10.

HAGEN, A. R., PLEGARIA, J. S., SLOAN, N., FERLEZ, B., AUSSIGNARGUES, C., BURTON, R. & KERFELD, C. A. 2018b. In Vitro Assembly of Diverse Bacterial Microcompartment Shell Architectures. *Nano Letters,* 18**,** 7030-7037.

HALLENBECK, P. C. 2009. Fermentative hydrogen production: Principles, progress, and prognosis. *International Journal of Hydrogen Energy,* 34**,** 7379-7389.

HAN, H., CUI, M., WEI, L., YANG, H. & SHEN, J. 2011. Enhancement effect of hematite nanoparticles on fermentative hydrogen production. *Bioresour Technol,* 102**,** 7903-9.

HAN, W., YE, M., ZHU, A. J., ZHAO, H. T. & LI, Y. F. 2015. Batch dark fermentation from enzymatic hydrolyzed food waste for hydrogen production. *Bioresour Technol,* 191**,** 24-9.

HAVEMANN, G. D. & BOBIK, T. A. 2003. Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1,2-propanediol in Salmonella enterica serovar Typhimurium LT2. *J Bacteriol,* 185**,** 5086-95.

HEIDRICH, E. S., DOLFING, J., SCOTT, K., EDWARDS, S. R., JONES, C. & CURTIS, T. P. 2013. Production of hydrogen from domestic wastewater in a pilot-scale microbial electrolysis cell. *Appl Microbiol Biotechnol,* 97**,** 6979-89.

HEINHORST, S., WILLIAMS, E. B., CAI, F., MURIN, C. D., SHIVELY, J. M. & CANNON, G. C. 2006. Characterization of the carboxysomal carbonic anhydrase CsoSCA from Halothiobacillus neapolitanus. *Journal of bacteriology,* 188**,** 8087-8094.

HENNACY, J. H. & JONIKAS, M. C. 2020. Prospects for engineering biophysical CO2 concentrating mechanisms into land plants to enhance yields. *Annual review of plant biology,* 71**,** 461-485.

HERRING, T. I., HARRIS, T. N., CHOWDHURY, C., MOHANTY, S. K. & BOBIK, T. A. 2018. A bacterial microcompartment is used for choline fermentation by Escherichia coli 536. *Journal of bacteriology,* 200**,** e00764-17.

HOLTHUIJZEN, Y. A., KUENEN, J.G. AND KONINGS, W.N. 1987. Activity of ribulose-1, 5-bisphosphate carboxylase in intact and disrupted carboxysomes of Thiobacillus neapolitanus. . *FEMS Microbiology Letters,* 42**,** 4.

HUANG, F., KONG, W. W., SUN, Y., CHEN, T., DYKES, G. F., JIANG, Y. L. & LIU, L. N. 2020. Rubisco accumulation factor 1 (Raf1) plays essential roles in mediating Rubisco assembly and carboxysome biogenesis. *Proc Natl Acad Sci U S A,* 117**,** 17418-17428.

IANCU, C. V., DING, H. J., MORRIS, D. M., DIAS, D. P., GONZALES, A. D., MARTINO, A. & JENSEN, G. J. 2007. The structure of isolated Synechococcus strain WH8102 carboxysomes as revealed by electron cryotomography. *J Mol Biol,* 372**,** 764-73.

JAKOBSON, C. M., KIM, E. Y., SLININGER, M. F., CHIEN, A. & TULLMAN-ERCEK, D. 2015. Localization of Proteins to the 1,2-Propanediol Utilization Microcompartment by Non-native Signal Sequences Is Mediated by a Common Hydrophobic Motif. *Journal of Biological Chemistry,* 290**,** 24519-24533.

JAKOBSON, C. M., SLININGER LEE, M. F. & TULLMAN-ERCEK, D. 2017a. De novo design of signal sequences to localize cargo to the 1,2-propanediol utilization microcompartment. *Protein Science,* 26**,** 1086-1092.

JAKOBSON, C. M., TULLMAN-ERCEK, D., SLININGER, M. F. & MANGAN, N. M. 2017b. A systems-level model reveals that 1,2-Propanediol utilization microcompartments enhance pathway flux through intermediate sequestration. *PLOS Computational Biology,* 13**,** e1005525.

JAOON YH KIM, B. H. J., HYUNG JOON CHA 2010. Production of biohydrogen by recombinant expression of [NiFe]-hydrogenase 1 in Escherichia coli. *Microbial Cell Factories,* 9**,** 23-32.

JORDA, J., LEIBLY, D. J., THOMPSON, M. C. & YEATES, T. O. 2016. Structure of a novel 13 nm dodecahedral nanocage assembled from a redesigned bacterial microcompartment shell protein. *Chem Commun (Camb),* 52**,** 5041-4.

JORDA, J., LOPEZ, D., WHEATLEY, N. M. & YEATES, T. O. 2013. Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci,* 22**,** 179-95.

JORDAN, P. C., PATTERSON, D. P., SABODA, K. N., EDWARDS, E. J., MIETTINEN, H. M., BASU, G., THIELGES, M. C. & DOUGLAS, T. 2016. Self-assembling biomolecular catalysts for hydrogen production. *Nat Chem,* 8**,** 179-85.

JUANITA MATHEWS, Q. L., GUANGYI WANG 2010. Characterization of hydrogen production by engineered Escherichia coli strains using rich defined media. *Biotechnology and Bioprocess Engineering,* 15**,** 10.

JUODEIKIS, R., LEE, M. J., MAYER, M., MANTELL, J., BROWN, I. R., VERKADE, P., WOOLFSON, D. N., PRENTICE, M. B., FRANK, S. & WARREN, M. J. 2020. Effect of metabolosome encapsulation peptides on enzyme activity, coaggregation, incorporation, and bacterial microcompartment formation. *MicrobiologyOpen,* 9**,** e1010.

KALNINS, G., CESLE, E.-E., JANSONS, J., LIEPINS, J., FILIMONENKO, A. & TARS, K. 2020. Encapsulation mechanisms and structural studies of GRM2 bacterial microcompartment particles. *Nature communications,* 11**,** 1-13.

KANEVSKI, I., MALIGA, P., RHOADES, D. F. & GUTTERIDGE, S. 1999. Plastome engineering of ribulose-1, 5-bisphosphate carboxylase/oxygenase in tobacco to form a sunflower large subunit and tobacco small subunit hybrid. *Plant Physiology,* 119**,** 133-142.

KAPLAN, A., SCHWARZ, R., LIEMAN-HURWITZ, J. & REINHOLD, L. 1991. Physiological and molecular aspects of the inorganic carbon-concentrating mechanism in cyanobacteria. *Plant Physiology,* 97**,** 851-855.

KERFELD, C. A., AUSSIGNARGUES, C., ZARZYCKI, J., CAI, F. & SUTTER, M. 2018. Bacterial microcompartments. *Nat Rev Microbiol,* 16**,** 277-290.

KERFELD, C. A. & ERBILGIN, O. 2015. Bacterial microcompartments and the modular construction of microbial metabolism. *Trends Microbiol,* 23**,** 22-34.

KERFELD, C. A., HEINHORST, S. & CANNON, G. C. 2010. Bacterial Microcompartments. *Annual Review of Microbiology,* 64**,** 391-408.

KERFELD, C. A. & MELNICKI, M. R. 2016. Assembly, function and evolution of cyanobacterial carboxysomes. *Current Opinion in Plant Biology,* 31**,** 66-75.

KERFELD, C. A., SAWAYA, M. R., TANAKA, S., NGUYEN, C. V., PHILLIPS, M., BEEBY, M. & YEATES, T. O. 2005. Protein Structures Forming the Shell of Primitive Bacterial Organelles. *Science,* 309**,** 936-938.

KHAN, M. A., ZHAO, H., ZOU, W., CHEN, Z., CAO, W., FANG, J., XU, J., ZHANG, L. & ZHANG, J. 2018. Recent Progresses in Electrocatalysts for Water Electrolysis. *Electrochemical Energy Reviews,* 1**,** 483-530.

KIM, J. Y., JO, B.H., CHA, H.J. 2010. Production of biohydrogen by recombinant expression of [NiFe]-hydrogenase 1 in *Escherichia coli. Microbial cell factories,* 9.

KIMBER, M. S. 2014. Carboxysomal carbonic anhydrases. *Carbonic Anhydrase: Mechanism, Regulation, Links to Disease, and Industrial Applications***,** 89-103.

KIMBER, M. S. & PAI, E. F. 2000. The active site architecture of Pisum sativum β-carbonic anhydrase is a mirror image of that of α-carbonic anhydrases. *The EMBO journal,* 19**,** 1407-1418.

KING, P. W. A. P., A.E. 1999. Response of hya Expression to External pH in Escherichia coli. *Journal of bacteriology,* 181**,** 7.

KINNEY, J. N., SALMEEN, A., CAI, F. & KERFELD, C. A. 2012. Elucidating essential role of conserved carboxysomal protein CcmN reveals common feature of bacterial microcompartment assembly. *J Biol Chem,* 287**,** 17729-17736.

KIRST, H., FERLEZ, B. H., LINDNER, S. N., COTTON, C. A. R., BAR-EVEN, A. & KERFELD, C. A. 2022. Toward a glycyl radical enzyme containing synthetic bacterial microcompartment to produce pyruvate from formate and acetate. *Proc Natl Acad Sci U S A,* 119.

KIRST, H. & KERFELD, C. A. 2019. Bacterial microcompartments: catalysis-enhancing metabolic modules for next generation metabolic and biomedical engineering. *BMC Biol,* 17**,** 79.

KLEIN, M. G., ZWART, P., BAGBY, S. C., CAI, F., CHISHOLM, S. W., HEINHORST, S., CANNON, G. C. & KERFELD, C. A. 2009. Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. *J Mol Biol,* 392**,** 319-33.

KOFOID, E., RAPPLEYE, C., STOJILJKOVIC, I. & ROTH, J. 1999. The 17-gene ethanolamine (eut) operon of Salmonella typhimurium encodes five homologues of carboxysome shell proteins. *Journal of bacteriology,* 181**,** 5317-5329.

LACASSE, M. J., DOUGLAS, C. D. & ZAMBLE, D. B. 2016. Mechanism of Selective Nickel Transfer from HypB to HypA, Escherichia coli [NiFe]-Hydrogenase Accessory Proteins. *Biochemistry,* 55**,** 6821-6831.

LARSSON, A. M., HASSE, D., VALEGÅRD, K. & ANDERSSON, I. 2017. Crystal structures of β-carboxysome shell protein CcmP: ligand binding correlates with the closed or open central pore. *Journal of experimental botany,* 68**,** 3857-3867.

LASSILA, J. K., BERNSTEIN, S. L., KINNEY, J. N., AXEN, S. D. & KERFELD, C. A. 2014. Assembly of robust bacterial microcompartment shells using building blocks from an organelle of unknown function. *Journal of Molecular Biology,* 426**,** 2217-2228.

LAWRENCE, A. D., FRANK, S., NEWNHAM, S., LEE, M. J., BROWN, I. R., XUE, W.-F., ROWE, M. L., MULVIHILL, D. P., PRENTICE, M. B., HOWARD, M. J. & WARREN, M. J. 2014. Solution Structure of a Bacterial Microcompartment Targeting Peptide and Its Application in the Construction of an Ethanol Bioreactor. *ACS Synthetic Biology,* 3**,** 454-465.

LEE, M. J., BROWN, I. R., JUODEIKIS, R., FRANK, S. & WARREN, M. J. 2016. Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1,2-propanediol production in Escherichia coli. *Metab Eng,* 36**,** 48-56.

LEE, M. J., MANTELL, J., BROWN, I. R., FLETCHER, J. M., VERKADE, P., PICKERSGILL, R. W., WOOLFSON, D. N., FRANK, S. & WARREN, M. J. 2018. De novo targeting to the cytoplasmic and luminal side of bacterial microcompartments. *Nature Communications,* 9**,** 3413.

LEE, M. J., PALMER, D. J. & WARREN, M. J. 2019. Biotechnological Advances in Bacterial Microcompartment Technology. *Trends in Biotechnology,* 37**,** 325-336.

LI, A., SUN, Y., YAO, T. & HAN, H. 2018. Earth-Abundant Transition-Metal-Based Electrocatalysts for Water Electrolysis to Produce Renewable Hydrogen. *Chemistry,* 24**,** 18334-18355.

LI, T., JIANG, Q., HUANG, J., AITCHISON, C. M., HUANG, F., YANG, M., DYKES, G. F., HE, H. L., WANG, Q., SPRICK, R. S., COOPER, A. I. & LIU, L. N. 2020. Reprogramming bacterial protein organelles as a nanoreactor for hydrogen production. *Nat Commun,* 11**,** 5448.

LIANG, M., FRANK, S., LÜNSDORF, H., WARREN, M. J. & PRENTICE, M. B. 2017. Bacterial microcompartment-directed polyphosphate kinase promotes stable polyphosphate accumulation in E. coli. *Biotechnology Journal,* 12**,** 1600415.

LIN, M. T., OCCHIALINI, A., ANDRALOJC, P. J., DEVONSHIRE, J., HINES, K. M., PARRY, M. A. & HANSON, M. R. 2014a. beta-Carboxysomal proteins assemble into highly organized structures in Nicotiana chloroplasts. *Plant J,* 79**,** 1-12.

LIN, M. T., OCCHIALINI, A., ANDRALOJC, P. J., DEVONSHIRE, J., HINES, K. M., PARRY, M. A. J. & HANSON, M. R. 2014b. β-Carboxysomal proteins assemble into highly organized structures in Nicotiana chloroplasts. *The Plant Journal,* 79**,** 1-12.

LIN, M. T., OCCHIALINI, A., ANDRALOJC, P. J., PARRY, M. A. J. & HANSON, M. R. 2014c. A faster Rubisco with potential to increase photosynthesis in crops. *Nature,* 513**,** 547-550.

LIN QIU JIAN, Z. P., DING A QIANG. 2018. Composition, structure, function and detection of carboxysome and their significance to nitrogen-removal bacteria. *Microbiology China,* 45**,** 11.

LIN, R., CHENG, J., DING, L., SONG, W., LIU, M., ZHOU, J. & CEN, K. 2016. Enhanced dark hydrogen fermentation by addition of ferric oxide nanoparticles using Enterobacter aerogenes. *Bioresour Technol,* 207**,** 213-9.

LIU, L.-N. 2016. Distribution and dynamics of electron transport complexes in cyanobacterial thylakoid membranes. *Biochimica et Biophysica Acta (BBA) - Bioenergetics,* 1857**,** 256-265.

LIU, L.-N., YANG, M., SUN, Y. & YANG, J. 2021a. Protein stoichiometry, structural plasticity and regulation of bacterial microcompartments. *Current Opinion in Microbiology,* 63**,** 133-141.

LIU, L. N. 2021a. Advances in the bacterial organelles for CO2 fixation. *Trends Microbiol.*

LIU, L. N. 2021b. Bacterial metabolosomes: new insights into their structure and bioengineering. *Microbial Biotechnology,* 14**,** 88-93.

LIU, L. N., YANG, M., SUN, Y. & YANG, J. 2021b. Protein stoichiometry, structural plasticity and regulation of bacterial microcompartments. *Curr Opin Microbiol,* 63**,** 133-141.

LIU, Y., HE, X., LIM, W., MUELLER, J., LAWRIE, J., KRAMER, L., GUO, J. & NIU, W. 2018a. Deciphering molecular details in the assembly of alpha-type carboxysome. *Scientific Reports,* 8.

LIU, Y., HE, X., LIM, W., MUELLER, J., LAWRIE, J., KRAMER, L., GUO, J. & NIU, W. 2018b. Deciphering molecular details in the assembly of alpha-type carboxysome. *Scientific Reports,* 8**,** 15062.

LONG, B. M., BADGER, M. R., WHITNEY, S. M. & PRICE, G. D. 2007. Analysis of carboxysomes from Synechococcus PCC7942 reveals multiple Rubisco complexes with carboxysomal proteins CcmM and CcaA. *Journal of Biological Chemistry,* 282**,** 29323-29335.

LONG, B. M., FÖRSTER, B., PULSFORD, S. B., PRICE, G. D. & BADGER, M. R. 2021. Rubisco proton production can drive the elevation of CO2 within condensates and carboxysomes. *Proceedings of the National Academy of Sciences,* 118.

LONG, B. M., HEE, W. Y., SHARWOOD, R. E., RAE, B. D., KAINES, S., LIM, Y.-L., NGUYEN, N. D., MASSEY, B., BALA, S., VON CAEMMERER, S., BADGER, M. R. & PRICE, G. D. 2018. Carboxysome encapsulation of the CO2-fixing enzyme Rubisco in tobacco chloroplasts. *Nature Communications,* 9**,** 3570.

LONG, B. M., RAE, B. D., ROLLAND, V., FORSTER, B. & PRICE, G. D. 2016a. Cyanobacterial CO2-concentrating mechanism components: function and prospects for plant metabolic engineering. *Curr Opin Plant Biol,* 31**,** 1-8.

LONG, B. M., RAE, B. D., ROLLAND, V., FÖRSTER, B. & PRICE, G. D. 2016b. Cyanobacterial CO2-concentrating mechanism components: function and prospects for plant metabolic engineering. *Current opinion in plant biology,* 31**,** 1-8.

LONG, B. M., TUCKER, L., BADGER, M. R. & PRICE, G. D. 2010a. Functional cyanobacterial beta-carboxysomes have an absolute requirement for both long and short forms of the CcmM protein. *Plant Physiol,* 153**,** 285-93.

LONG, B. M., TUCKER, L., BADGER, M. R. & PRICE, G. D. 2010b. Functional cyanobacterial β-carboxysomes have an absolute requirement for both long and short forms of the CcmM protein. *Plant Physiology,* 153**,** 285-293.

LÓPEZ-SAGASETA, J., MALITO, E., RAPPUOLI, R. & BOTTOMLEY, M. J. 2016. Self-assembling protein nanoparticles in the design of vaccines. *Computational and structural biotechnology journal,* 14**,** 58-68.

LUDWIG, M., SÜLTEMEYER, D. & PRICE, G. D. 2000. Isolation of ccmKLMN genes from the marine cyanobacterium, Synechococcus sp. PCC7002 (Cyanophyceae), and evidence that CcmM is essential for carboxysome assembly. *Journal of Phycology,* 36**,** 1109-1119.

LUKEY, M. J., PARKIN, A., ROESSLER, M. M., MURPHY, B. J., HARMER, J., PALMER, T., SARGENT, F. & ARMSTRONG, F. A. 2010. How Escherichia coli is equipped to oxidize hydrogen under different redox conditions. *J Biol Chem,* 285**,** 3928-3938.

LUNDIN, A. P., STEWART, K. L., STEWART, A. M., HERRING, T. I., CHOWDHURY, C. & BOBIK, T. A. 2020. Genetic Characterization of a Glycyl Radical Microcompartment Used for 1,2-Propanediol Fermentation by Uropathogenic Escherichia coli CFT073. *J Bacteriol,* 202.

MACCREADY, J. S. & VECCHIARELLI, A. G. 2021. Positioning the Model Bacterial Organelle, the Carboxysome. *mBio,* 12.

MAEDA, T., SANCHEZ-TORRES, V. & WOOD, T. K. 2007a. Enhanced hydrogen production from glucose by metabolically engineered Escherichia coli. *Appl Microbiol Biotechnol,* 77**,** 879-90.

MAEDA, T., SANCHEZ-TORRES, V. & WOOD, T. K. 2007b. Escherichia coli hydrogenase 3 is a reversible enzyme possessing hydrogen uptake and synthesis activities. *Appl Microbiol Biotechnol,* 76**,** 1035-42.

MAHINTHICHAICHAN, P., MORRIS, D. M., WANG, Y., JENSEN, G. J. & TAJKHORSHID, E. 2018. Selective Permeability of Carboxysome Shell Pores to Anionic Molecules. *J Phys Chem B,* 122**,** 9110-9118.

MAHON, M. J. 2011. pHluorin2: an enhanced, ratiometric, pH-sensitive green florescent protein. *Adv Biosci Biotechnol,* 2**,** 132-137.

MALLETTE, E. & KIMBER, M. S. 2017. A complete structural inventory of the mycobacterial microcompartment shell proteins constrains models of global architecture and transport. *Journal of Biological Chemistry,* 292**,** 1197-1210.

MALLETTE, E. & KIMBER, M. S. 2018. Structure and Kinetics of the S-(+)-1-Amino-2-propanol Dehydrogenase from the RMM Microcompartment of Mycobacterium smegmatis. *Biochemistry,* 57**,** 3780-3789.

MARTINEZ-DEL CAMPO, A., Bodea, S., Hamer, H.A., Marks, J.A., Haiser, H.J., Turnbaugh, P.J. and Balskus, E.P., 2015. Characterization and detection of a widely distributed gene cluster that predicts anaerobic choline utilization by human gut bacteria. *MBio*, 6(2), e00042-15.

MAYER, M. J., JUODEIKIS, R., BROWN, I. R., FRANK, S., PALMER, D. J., DEERY, E., BEAL, D. M., XUE, W.-F. & WARREN, M. J. 2016. Effect of bio-engineering on size, shape, composition and rigidity of bacterial microcompartments. *Scientific Reports,* 6**,** 36899.

MEHERKOTAY, S. & DAS, D. 2008. Biohydrogen as a renewable energy resource—Prospects and potentials. *International Journal of Hydrogen Energy,* 33**,** 258-263.

MELIS, A. 2009. Solar energy conversion efficiencies in photosynthesis: Minimizing the chlorophyll antennae to maximize efficiency. *Plant Science,* 177**,** 272-280.

MELIS, A. & HAPPE, T. 2001. Hydrogen Production. Green Algae as a Source of Energy. *Plant Physiology,* 127**,** 740-748.

MENON, B. B., DOU, Z., HEINHORST, S., SHIVELY, J. M. & CANNON, G. C. 2008. Halothiobacillus neapolitanus carboxysomes sequester heterologous and chimeric RubisCO species. *PLoS One,* 3**,** e3570.

MENON NK, R. J., WENDT JC, SHANMUGAM KT, PRZYBYLA AE. 1991. Mutational analysis and characterization of the Escherichia coli hya operon, which encodes [NiFe] hydrogenase 1. *Journals of Bacteriology,* 173**,** 11.

MENON, N. K., ROBBINS, J., WENDT, J. C., SHANMUGAM, K. T. & PRZYBYLA, A. E. 1991. Mutational analysis and characterization of the Escherichia coli hya operon, which encodes [NiFe] hydrogenase 1. *J Bacteriol,* 173**,** 4851-61.

MENON, N. K., ROBBINS, J., WENDT, J.C., SHANMUGAM, K.T. AND PRZYBYLA, A.E. 1991. Mutational analysis and characterization of the Escherichia coli hya operon, which encodes [NiFe] hydrogenase 1. *Journal of bacteriology,* 173**,** 11.

MENZEL, K., APFEL, U. P., WOLTER, N., RUGER, R., ALPERMANN, T., STEINIGER, F., GABEL, D., FORSTER, S., WEIGAND, W. & FAHR, A. 2014. [FeFe]-hydrogenase models assembled into vesicular structures. *J Liposome Res,* 24**,** 59-68.

MIROUX, B. & WALKER, J. E. 1996. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of molecular biology,* 260**,** 289-298.

MISHRA, P., THAKUR, S., MAHAPATRA, D. M., WAHID, Z. A., LIU, H. & SINGH, L. 2018. Impacts of nano-metal oxides on hydrogen production in anaerobic digestion of palm oil mill effluent – A novel approach. *International Journal of Hydrogen Energy,* 43**,** 2666-2676.

MNATSAKANYAN N, B. K., TRCHOUNIAN A. 2004. Hydrogenase 3 but not hydrogenase 4 is major in hydrogen gas production by Escherichia coli formate

hydrogenlyase at acidic pH and in the presence of external formate. *Cell Biochemistry and Biophysics,* 41**,** 10.

MNATSAKANYAN, N., VASSILIAN, A., NAVASARDYAN, L., BAGRAMYAN, K. & TRCHOUNIAN, A. 2002. Regulation of Escherichia coli formate hydrogenlyase activity by formate at alkaline pH. *Curr Microbiol,* 45**,** 281-6.

MODESTRA, J. A., BABU, M. L. & MOHAN, S. V. 2015. Electro-fermentation of real-field acidogenic spent wash effluents for additional biohydrogen production with simultaneous treatment in a microbial electrolysis cell. *Separation and Purification Technology,* 150**,** 308-315.

MORITA, K., HATANAKA, T., MISOO, S. & FUKAYAMA, H. 2014. Unusual small subunit that is not expressed in photosynthetic cells alters the catalytic properties of Rubisco in rice. *Plant Physiology,* 164**,** 69-79.

MUDHOO, A., FORSTER-CARNEIRO, T. & SANCHEZ, A. 2011. Biohydrogen production and bioprocess enhancement: a review. *Crit Rev Biotechnol,* 31**,** 250-63.

MULLAI, P., YOGESWARI, M. K. & SRIDEVI, K. 2013. Optimisation and enhancement of biohydrogen production using nickel nanoparticles - a novel approach. *Bioresour Technol,* 141**,** 212-9.

MURPHY, B. J., SARGENT, F. AND ARMSTRONG, F.A. 2014. Transforming an oxygen-tolerant [NiFe] uptake hydrogenase into a proficient, reversible hydrogen producer. . *Energy & Environmental Science,* 7**,** 8.

NASR, M., TAWFIK, A., OOKAWARA, S., SUZUKI, M., KUMARI, S. & BUX, F. 2015. Continuous biohydrogen production from starch wastewater via sequential dark-photo fermentation with emphasize on maghemite nanoparticles. *Journal of Industrial and Engineering Chemistry,* 21**,** 500-506.

NICHOLS, T. M., KENNEDY, N. W. & TULLMAN-ERCEK, D. 2020. A genomic integration platform for heterologous cargo encapsulation in 1,2-propanediol utilization bacterial microcompartments. *Biochemical Engineering Journal,* 156**,** 107496.

OCHOA, J. M., BAIR, K., HOLTON, T., BOBIK, T. A. & YEATES, T. O. 2021. MCPdb: The bacterial microcompartment database. *PLoS One,* 16**,** e0248269.

OLTROGGE, L. M., CHAIJARASPHONG, T., CHEN, A. W., BOLIN, E. R., MARQUSEE, S. & SAVAGE, D. F. 2020. Multivalent interactions between CsoS2 and Rubisco mediate alpha-carboxysome formation. *Nat Struct Mol Biol,* 27**,** 281-287.

PANG, A., LIANG, M., PRENTICE, M. B. & PICKERSGILL, R. W. 2012. Substrate channels revealed in the trimeric Lactobacillus reuteri bacterial microcompartment shell protein PduB. *Acta Crystallographica Section D: Biological Crystallography,* 68**,** 1642-1652.

PANG, A., WARREN, M. J. & PICKERSGILL, R. W. 2011. Structure of PduT, a trimeric bacterial microcompartment protein with a 4Fe–4S cluster-binding site. *Acta Crystallographica Section D: Biological Crystallography,* 67**,** 91-96.

PARSONS, J. B., DINESH, S. D., DEERY, E., LEECH, H. K., BRINDLEY, A. A., HELDT, D., FRANK, S., SMALES, C. M., LÜNSDORF, H. & RAMBACH, A. 2008. Biochemical and structural insights into bacterial organelle form and biogenesis. *Journal of biological chemistry,* 283**,** 14366-14375.

PARSONS, J. B., FRANK, S., BHELLA, D., LIANG, M., PRENTICE, M. B., MULVIHILL, D. P. & WARREN, M. J. 2010. Synthesis of Empty Bacterial Microcompartments, Directed Organelle Protein Incorporation, and Evidence of Filament-Associated Organelle Movement. *Molecular Cell,* 38**,** 305-315.

PAULETTE M. VIGNAIS, B. B., JACQUES MEYER 2001. Classification and phylogeny of hydrogenases1. *microbiology reviews,* 25**,** 455-501.

PETERS, J. W., SCHUT, G. J., BOYD, E. S., MULDER, D. W., SHEPARD, E. M., BRODERICK, J. B., KING, P. W. & ADAMS, M. W. 2015. [FeFe]- and [NiFe]-hydrogenase diversity, mechanism, and maturation. *Biochim Biophys Acta,* 1853**,** 1350-69.

PETIT, E., LATOUF, W. G., COPPI, M. V., WARNICK, T. A., CURRIE, D., ROMASHKO, I., DESHPANDE, S., HAAS, K., ALVELO-MAUROSA, J. G. & WARDMAN, C. 2013. Involvement of a bacterial microcompartment in the metabolism of fucose and rhamnose by Clostridium phytofermentans. *PLoS One,* 8**,** e54337.

PITTS, A. C., TUCK, L. R., FAULDS-PAIN, A., LEWIS, R. J. & MARLES-WRIGHT, J. 2012. Structural insight into the Clostridium difficile ethanolamine utilisation microcompartment. *PloS one,* 7**,** e48360.

PRICE, G. & BADGER, M. 1989. Isolation and characterization of high CO_2-requiring-mutants of the cyanobacterium Synechococcus PCC7942: two phenotypes that accumulate inorganic carbon but are apparently unable to generate CO_2 within the carboxysome. *Plant physiology,* 91**,** 514-525.

PRICE, G. D., BADGER, M. R., WOODGER, F. J. & 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. *Journal of experimental botany,* 59**,** 1441-1461.

PRICE, G. D., HOWITT, S. M., HARRISON, K. & BADGER, M. R. 1993. Analysis of a genomic DNA region from the cyanobacterium Synechococcus sp. strain PCC7942 involved in carboxysome assembly and function. *Journal of Bacteriology,* 175**,** 2871-2879.

QUIN, M. B., PERDUE, S. A., HSU, S.-Y. & SCHMIDT-DANNERT, C. 2016. Encapsulation of multiple cargo proteins within recombinant Eut nanocompartments. *Applied Microbiology and Biotechnology,* 100**,** 9187-9200.

R. GARY SAWERS, S. P. B. A. D. H. B. 1985. Differential expression of hydrogenase isoenzymes in Escherichia coli K-12: evidence for a third isoenzyme. *Journal of bacteriology,* 164**,** 1324-1331.

RAE, B. D., LONG, B. M., BADGER, M. R. & PRICE, G. D. 2013. Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO_2 fixation in cyanobacteria and some proteobacteria. *Microbiol Mol Biol Rev,* 77**,** 357-79.

RAE, B. D., LONG, B. M., FORSTER, B., NGUYEN, N. D., VELANIS, C. N., ATKINSON, N., HEE, W. Y., MUKHERJEE, B., PRICE, G. D. & MCCORMICK, A. J. 2017. Progress and challenges of engineering a biophysical CO_2-concentrating mechanism into higher plants. *J Exp Bot,* 68**,** 3717-3737.

RAVCHEEV, D. A., MOUSSU, L., SMAJIC, S. & THIELE, I. 2019. Comparative Genomic Analysis Reveals Novel Microcompartment-Associated Metabolic Pathways in the Human Gut Microbiome. *Front Genet,* 10**,** 636.

RG., S. 2005. Formate and its role in hydrogen production in Escherichia coli. *Biochemical Society Transactions* 33**,** 5.

ROBERTS, E. W., CAI, F., KERFELD, C. A., CANNON, G. C. & HEINHORST, S. 2012. Isolation and Characterization of the Prochlorococcus Carboxysome Reveal the Presence of the Novel Shell Protein CsoS1D. *Journal of Bacteriology,* 194**,** 787-795.

RONQUIST, F. & HUELSENBECK, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics,* 19**,** 1572-1574.

ROSA, P. R., SANTOS, S. C., SAKAMOTO, I. K., VARESCHE, M. B. & SILVA, E. L. 2014. Hydrogen production from cheese whey with ethanol-type fermentation: effect

of hydraulic retention time on the microbial community composition. *Bioresour Technol,* 161**,** 10-19.

ROSSMANN R, S. G., BÖCK A. 1991. Mechanism of regulation of the formate-hydrogenlyase pathway by oxygen, nitrate, and pH: definition of the formate regulon. . *Molecular microbiology,* 5**,** 8.

RURUP, W. F., SNIJDER, J., KOAY, M. S. T., HECK, A. J. R. & CORNELISSEN, J. J. L. M. 2014. Self-Sorting of Foreign Proteins in a Bacterial Nanocompartment. *Journal of the American Chemical Society,* 136**,** 3828-3832.

SANCHEZ-TORRES, V., MOHD YUSOFF, M. Z., NAKANO, C., MAEDA, T., OGAWA, H. I. & WOOD, T. K. 2013. Influence of Escherichia coli hydrogenases on hydrogen fermentation from glycerol. *International Journal of Hydrogen Energy,* 38**,** 3905-3912.

SARGENT, F., DAVIDSON, F. A., KELLY, C. L., BINNY, R., CHRISTODOULIDES, N., GIBSON, D., JOHANSSON, E., KOZYRSKA, K., LADO, L. L., MACCALLUM, J., MONTAGUE, R., ORTMANN, B., OWEN, R., COULTHURST, S. J., DUPUY, L., PRESCOTT, A. R. & PALMER, T. 2013. A synthetic system for expression of components of a bacterial microcompartment. *Microbiology (Reading, England),* 159**,** 2427-2436.

SAUTER M, B. R., BÖCK A. 1992. Mutational analysis of the operon (hyc) determining hydrogenase 3 formation in Escherichia coli. *Molecular Microbiology,* 6**,** 10.

SAUVAGEOT, N., PICHEREAU, V., LOUARME, L., HARTKE, A., AUFFRAY, Y. & LAPLACE, J. M. 2002. Purification, characterization and subunits identification of the diol dehydratase of Lactobacillus collinoides. *Eur J Biochem,* 269**,** 5731-7.

SAWAYA, M. R., CANNON, G. C., HEINHORST, S., TANAKA, S., WILLIAMS, E. B., YEATES, T. O. & 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. *Journal of Biological Chemistry,* 281**,** 7546-7555.

SAWERS, R. G. & BOXER, D. H. 1986. Purification and properties of membrane-bound hydrogenase isoenzyme 1 from anaerobically grown Escherichia coli K12. *Eur J Biochem,* 156**,** 265-75.

SCHMID, M. F., PAREDES, A. M., KHANT, H. A., SOYER, F., ALDRICH, H. C., CHIU, W. & SHIVELY, J. M. 2006. Structure of Halothiobacillus neapolitanus Carboxysomes by Cryo-electron Tomography. *Journal of Molecular Biology,* 364**,** 526-535.

SCHNEIDER, G., KNIGHT, S., ANDERSSON, I., BRÄNDÉN, C., LINDQVIST, Y. & LUNDQVIST, T. 1990. Comparison of the crystal structures of L2 and L8S8 Rubisco suggests a functional role for the small subunit. *The EMBO journal,* 9**,** 2045-2050.

SCHINDEL, H.S., Karty, J.A., McKinlay, J.B. and Bauer, C.E., 2019. Characterization of a glycyl radical enzyme bacterial microcompartment pathway in Rhodobacter capsulatus. *Journal of bacteriology*, 201(5), e00343-18.

SEED, K. D., FARUQUE, S. M., MEKALANOS, J. J., CALDERWOOD, S. B., QADRI, F. & CAMILLI, A. 2012. Phase variable O antigen biosynthetic genes control expression of the major protective antigen and bacteriophage receptor in Vibrio cholerae O1. *PLoS Pathog,* 8**,** e1002917.

SELMER, T., PIERIK, A.J. and HEIDER, J. 2005. New glycyl radical enzymes catalysing key metabolic steps in anaerobic bacteria. *Biological Chemistry*, 386(10): 981-988.

SHEN, Y., HAO, T., OU, S., HU, C. & CHEN, L. 2018. Applications and perspectives of nanomaterials in novel vaccine development. *MedChemComm,* 9**,** 226-238.

SHI, J., KANTOFF, P. W., WOOSTER, R. & FAROKHZAD, O. C. 2017. Cancer nanomedicine: progress, challenges and opportunities. *Nature reviews cancer,* 17**,** 20-37.

SHIVELY, J., BALL, F., BROWN, D. & SAUNDERS, R. 1973. Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of Thiobacillus neapolitanus. *Science***,** 584-586.

SHIVELY, J. M., BALL, F.L. AND KLINE, B.W. 1973. Electron microscopy of the carboxysomes (polyhedral bodies) of Thiobacillus neapolitanus. . *Journal of Bacteriology,* 116**,** 7.

SHIVELY, J. M. & ENGLISH, R. S. 1991. The carboxysome, a prokaryotic organelle: a mini-review. *Canadian Journal of Botany,* 69**,** 957-962.

SHIVELY, J. M., LORBACH, S. C., JIN, S. & BAKER, S. H. 1996. Carboxysomes: The Genes of Thiobacillus Neapolitanus. *In:* LIDSTROM, M. E. & TABITA, F. R. (eds.) *Microbial Growth on C1 Compounds: Proceedings of the 8th International Symposium on Microbial Growth on C1 Compounds, held in San Diego, U.S.A., 27 August – 1 September 1995.* Dordrecht: Springer Netherlands.

SLININGER LEE, M. F., JAKOBSON, C. M. & TULLMAN-ERCEK, D. 2017. Evidence for Improved Encapsulated Pathway Behavior in a Bacterial Microcompartment through Shell Protein Engineering. *ACS Synth Biol,* 6**,** 1880-1891.

SO, A. K. & ESPIE, G. S. 1998. Cloning, characterization and expression of carbonic anhydrase from the cyanobacterium Synechocystis PCC6803. *Plant molecular biology,* 37**,** 205-215.

SO, A. K. & ESPIE, G. S. 2005. Cyanobacterial carbonic anhydrases. *Canadian Journal of Botany,* 83**,** 721-734.

SO, A. K., JOHN-MCKAY, M. & ESPIE, G. S. 2002a. Characterization of a mutant lacking carboxysomal carbonic anhydrase from the cyanobacterium Synechocystis PCC6803. *Planta,* 214**,** 456-467.

SO, A. K., JOHN-MCKAY, M. & ESPIE, G. S. 2002b. Characterization of a mutant lacking carboxysomal carbonic anhydrase from the cyanobacterium Synechocystis PCC6803. *Planta,* 214**,** 456-67.

SO, A. K.-C., ESPIE, G. S., WILLIAMS, E. B., SHIVELY, J. M., HEINHORST, S. & CANNON, G. C. 2004. A novel evolutionary lineage of carbonic anhydrase (ε class) is a component of the carboxysome shell. *Journal of bacteriology,* 186**,** 623-630.

SOMMER, M., SUTTER, M., GUPTA, S., KIRST, H., TURMO, A., LECHNO-YOSSEF, S., BURTON, R. L., SAECHAO, C., SLOAN, N. B., CHENG, X., CHAN, L.-J. G., PETZOLD, C. J., FUENTES-CABRERA, M., RALSTON, C. Y. & KERFELD, C. A. 2019. Heterohexamers Formed by CcmK3 and CcmK4 Increase the Complexity of Beta Carboxysome Shells. *Plant Physiology,* 179**,** 156-167.

SPREITZER, R. J. & SALVUCCI, M. E. 2002. Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annu Rev Plant Biol,* 53**,** 449-75.

STEELE, J. F. C., PEYRET, H., SAUNDERS, K., CASTELLS-GRAELLS, R., MARSIAN, J., MESHCHERIAKOVA, Y. & LOMONOSSOFF, G. P. 2017. Synthetic plant virology for nanobiotechnology and nanomedicine. *WIREs Nanomedicine and Nanobiotechnology,* 9**,** e1447.

STEWART, A. M., STEWART, K. L., YEATES, T. O. & BOBIK, T. A. 2021. Advances in the World of Bacterial Microcompartments. *Trends in Biochemical Sciences.*

SUI, N., HUANG, F. & LIU, L. N. 2020. Photosynthesis in Phytoplankton: Insights from the Newly Discovered Biological Inorganic Carbon Pumps. *Mol Plant,* 13**,** 949-951.

SUN, H., DONG, Y., FEIJEN, J. & ZHONG, Z. 2018. Peptide-decorated polymeric nanomedicines for precision cancer therapy. *Journal of controlled release,* 290**,** 11-27.

SUN, Y., CASELLA, S., FANG, Y., HUANG, F., FAULKNER, M., BARRETT, S. & LIU, L.-N. 2016. Light modulates the biosynthesis and organization of cyanobacterial carbon fixation machinery through photosynthetic electron flow. *Plant Physiology,* 171**,** 530-541.

SUN, Y., HARMAN, V. M., JOHNSON, J. R., CHEN, T., DYKES, G. F., LIN, Y., BEYNON, R. J. & LIU, L.-N. 2021. Decoding the absolute stoichiometric composition and structural plasticity of α-carboxysomes. *bioRxiv.*

SUN, Y., WOLLMAN, A. J. M., HUANG, F., LEAKE, M. C. & LIU, L. N. 2019. Single-Organelle Quantification Reveals Stoichiometric and Structural Variability of Carboxysomes Dependent on the Environment. *Plant Cell,* 31**,** 1648-1664.

SUTTER, M., FAULKNER, M., AUSSIGNARGUES, C., PAASCH, B. C., BARRETT, S., KERFELD, C. A. & LIU, L. N. 2016. Visualization of Bacterial Microcompartment Facet Assembly Using High-Speed Atomic Force Microscopy. *Nano Lett,* 16**,** 1590-5.

SUTTER, M., GREBER, B., AUSSIGNARGUES, C. & KERFELD, C. A. 2017. Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science,* 356**,** 1293-1297.

SUTTER, M., LAUGHLIN, T. G., SLOAN, N. B., SERWAS, D., DAVIES, K. M. & KERFELD, C. A. 2019a. Structure of a Synthetic beta-Carboxysome Shell. *Plant Physiol,* 181**,** 1050-1058.

SUTTER, M., LAUGHLIN, T. G., SLOAN, N. B., SERWAS, D., DAVIES, K. M. & KERFELD, C. A. 2019b. Structure of a Synthetic β-Carboxysome Shell1 [OPEN]. *Plant Physiology,* 181**,** 1050-1058.

SUTTER, M., MELNICKI, M. R., SCHULZ, F., WOYKE, T. & KERFELD, C. A. 2021. A catalog of the diversity and ubiquity of bacterial microcompartments. *Nat Commun,* 12**,** 3809.

SUTTER, M., WILSON, S. C., DEUTSCH, S. & KERFELD, C. A. 2013a. Two new high-resolution crystal structures of carboxysome pentamer proteins reveal high structural conservation of CcmL orthologs among distantly related cyanobacterial species. *Photosynthesis research,* 118**,** 9-16.

SUTTER, M., WILSON, S. C., DEUTSCH, S. & KERFELD, C. A. 2013b. Two new high-resolution crystal structures of carboxysome pentamer proteins reveal high structural conservation of CcmL orthologs among distantly related cyanobacterial species. *Photosynth Res,* 118**,** 9-16.

TAMAGNINI, P., LEITAO, E., OLIVEIRA, P., FERREIRA, D., PINTO, F., HARRIS, D. J., HEIDORN, T. & LINDBLAD, P. 2007. Cyanobacterial hydrogenases: diversity, regulation and applications. *FEMS Microbiol Rev,* 31**,** 692-720.

TAN, Y. Q., ALI, S., XUE, B., TEO, W. Z., LING, L. H., GO, M. K., LV, H., ROBINSON, R. C., NARITA, A. & YEW, W. S. 2021a. Structure of a Minimal alpha-Carboxysome-Derived Shell and Its Utility in Enzyme Stabilization. *Biomacromolecules*.

TAN, Y. Q., ALI, S., XUE, B., TEO, W. Z., LING, L. H., GO, M. K., LV, H., ROBINSON, R. C., NARITA, A. & YEW, W. S. 2021b. Structure of a Minimal alpha-Carboxysome-Derived Shell and Its Utility in Enzyme Stabilization. *Biomacromolecules, https://doi.org/10.1021/acs.biomac.1c00533*.

TAN, Y. Q., ALI, S., XUE, B., TEO, W. Z., LING, L. H., GO, M. K., LV, H., ROBINSON, R. C., NARITA, A. & YEW, W. S. 2021c. Structure of a Minimal α-Carboxysome-Derived Shell and Its Utility in Enzyme Stabilization. *Biomacromolecules,* 22**,** 4095-4109.

TANAKA, S., KERFELD, C. A., SAWAYA, M. R., CAI, F., HEINHORST, S., CANNON, G. C. & YEATES, T. O. 2008a. Atomic-Level Models of the Bacterial Carboxysome Shell. *Science,* 319**,** 1083-1086.

TANAKA, S., KERFELD, C. A., SAWAYA, M. R., CAI, F., HEINHORST, S., CANNON, G. C. & YEATES, T. O. 2008b. Atomic-level models of the bacterial carboxysome shell. *Science,* 319**,** 1083-6.

TANAKA, S., SAWAYA, M. R., PHILLIPS, M. & YEATES, T. O. 2009. Insights from multiple structures of the shell proteins from the beta-carboxysome. *Protein Sci,* 18**,** 108-20.

TATYANA V LURINAVICHENE, A. A. T. 2001. H2 consumption by Escherichia coli coupled via hydrogenase 1 or hydrogenase 2 to different terminal electron acceptors. *FEMS Microbiology Letters,* 202**,** 5.

TCHERKEZ, G. G., FARQUHAR, G. D. & ANDREWS, T. J. 2006. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proceedings of the National Academy of Sciences,* 103**,** 7246-7251.

TORZILLO, G., SCOMA, A., FARALONI, C., ENA, A. & JOHANNINGMEIER, U. 2009. Increased hydrogen photoproduction by means of a sulfur-deprived Chlamydomonas reinhardtii D1 protein mutant. *International Journal of Hydrogen Energy,* 34**,** 4529-4536.

TRCHOUNIAN, K., PINSKE, C., SAWERS, R. G. & TRCHOUNIAN, A. 2012a. Characterization of Escherichia coli [NiFe]-hydrogenase distribution during fermentative growth at different pHs. *Cell Biochem Biophys,* 62**,** 433-40.

TRCHOUNIAN, K., POLADYAN, A., VASSILIAN, A. & TRCHOUNIAN, A. 2012b. Multiple and reversible hydrogenases for hydrogen production by Escherichia coli: dependence on fermentation substrate, pH and the F(0)F(1)-ATPase. *Crit Rev Biochem Mol Biol,* 47**,** 236-49.

TRCHOUNIAN, K., SANCHEZ-TORRES, V., WOOD, T. K. & TRCHOUNIAN, A. 2011. Escherichia coli hydrogenase activity and H2 production under glycerol fermentation at a low pH. *international journal of hydrogen energy,* 36**,** 4323-4331.

TRCHOUNIAN, K., SAWERS, R. G. & TRCHOUNIAN, A. 2017. Improving biohydrogen productivity by microbial dark- and photo-fermentations: Novel data and future approaches. *Renewable and Sustainable Energy Reviews,* 80**,** 1201-1216.

TRØSEID, M., UELAND, T., HOV, J., SVARDAL, A., GREGERSEN, I., DAHL, C., AAKHUS, S., GUDE, E., BJØRNDAL, B. & HALVORSEN, B. 2015. Microbiota-dependent metabolite trimethylamine-N-oxide is associated with disease severity and survival of patients with chronic heart failure. *Journal of internal medicine,* 277**,** 717-726.

TSAI, Y., SAWAYA, M. R., CANNON, G. C., CAI, F., WILLIAMS, E. B., HEINHORST, S., KERFELD, C. A. & YEATES, T. O. 2007a. Structural analysis of CsoS1A and the protein shell of the Halothiobacillus neapolitanus carboxysome. *PLoS Biol,* 5**,** e144.

TSAI, Y., SAWAYA, M. R., CANNON, G. C., CAI, F., WILLIAMS, E. B., HEINHORST, S., KERFELD, C. A. & YEATES, T. O. 2007b. Structural Analysis of CsoS1A and the Protein Shell of the Halothiobacillus neapolitanus Carboxysome. *PLOS Biology,* 5**,** e144.

TSAI, Y., SAWAYA, M. R. & YEATES, T. O. 2009. Analysis of lattice-translocation disorder in the layered hexagonal structure of carboxysome shell protein CsoS1C. *Acta Crystallographica Section D,* 65**,** 980-988.

TSAI, Y. C., LIEW, L., GUO, Z., LIU, D. & MUELLER-CAJAR, O. 2022. The CbbQO-type rubisco activases encoded in carboxysome gene clusters can activate carboxysomal form IA rubiscos. *J Biol Chem,* 298**,** 101476.

TURMO, A., GONZALEZ-ESQUER, C. R. & KERFELD, C. A. 2017a. Carboxysomes: metabolic modules for CO2 fixation. *FEMS Microbiol Lett,* 364.

TURMO, A., GONZALEZ-ESQUER, C. R. & KERFELD, C. A. 2017b. Carboxysomes: metabolic modules for CO2 fixation. *FEMS Microbiology Letters,* 364.

UDDIN, I., FRANK, S., WARREN, M. J. & PICKERSGILL, R. W. 2018. A Generic Self-Assembly Process in Microcompartments and Synthetic Protein Nanotubes. *Small,* 14**,** 1704020.

UYAR, B., GURGAN, M., OZGUR, E., GUNDUZ, U., YUCEL, M. & EROGLU, I. 2015. Hydrogen production by hup(-) mutant and wild-type strains of Rhodobacter capsulatus from dark fermentation effluent of sugar beet thick juice in batch and continuous photobioreactors. *Bioprocess Biosyst Eng,* 38**,** 1935-42.

VENKATA MOHAN, S., VIJAYA BHASKAR, Y. & SARMA, P. N. 2007. Biohydrogen production from chemical wastewater treatment in biofilm configured reactor operated in periodic discontinuous batch mode by selectively enriched anaerobic mixed consortia. *Water Res,* 41**,** 2652-64.

VERONESE, F. M. & PASUT, G. 2005. PEGylation, successful approach to drug delivery. *Drug discovery today,* 10**,** 1451-1458.

VOLBEDA, A., DARNAULT, C., PARKIN, A., SARGENT, F., ARMSTRONG, F. A. & FONTECILLA-CAMPS, J. C. 2013. Crystal structure of the O(2)-tolerant membrane-bound hydrogenase 1 from Escherichia coli in complex with its cognate cytochrome b. *Structure,* 21**,** 184-190.

WANG, C., SUN, B., ZHANG, X., HUANG, X., ZHANG, M., GUO, H., CHEN, X., HUANG, F., CHEN, T., MI, H., YU, F., LIU, L. N. & ZHANG, P. 2019a. Structural mechanism of the active bicarbonate transporter from cyanobacteria. *Nat Plants,* 5**,** 1184-1193.

WANG, H., YAN, X., AIGNER, H., BRACHER, A., NGUYEN, N. D., HEE, W. Y., LONG, B. M., PRICE, G. D., HARTL, F. U. & HAYER-HARTL, M. 2019b. Rubisco condensate formation by CcmM in β-carboxysome biogenesis. *Nature,* 566**,** 131-135.

WHEATLEY, N. M., GIDANIYAN, S. D., LIU, Y., CASCIO, D. & YEATES, T. O. 2013. Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins. *Protein Sci,* 22**,** 660-5.

WHITEHEAD, L., LONG, B. M., PRICE, G. D. & BADGER, M. R. 2014. Comparing the in vivo function of α-carboxysomes and β-carboxysomes in two model cyanobacteria. *Plant Physiology,* 165**,** 398-411.

WILKERSON, J. W., YANG, S. O., FUNK, P. J., STANLEY, S. K. & BUNDY, B. C. 2018. Nanoreactors: Strategies to encapsulate enzyme biocatalysts in virus-like particles. *N Biotechnol,* 44**,** 59-63.

WU, M. & EISEN, J. A. 2008. A simple, fast, and accurate method of phylogenomic inference. *Genome biology,* 9**,** 1-11.

WULFF, P., THOMAS, C., SARGENT, F. & ARMSTRONG, F. A. 2016. How the oxygen tolerance of a [NiFe]-hydrogenase depends on quaternary structure. *J Biol Inorg Chem,* 21**,** 121-34.

XU, Z., JIANG, Y. & ZHOU, G. 2015. Response and adaptation of photosynthesis, respiration, and antioxidant systems to elevated CO2 with environmental stress in plants. *Frontiers in Plant Science,* 6.

YANG, H. & SHEN, J. 2006. Effect of ferrous iron concentration on anaerobic bio-hydrogen production from soluble starch. *International Journal of Hydrogen Energy,* 31**,** 2137-2146.

YEATES, T. O., CROWLEY, C. S. & TANAKA, S. 2010. Bacterial Microcompartment Organelles: Protein Shell Structure and Evolution. *Annual Review of Biophysics,* 39**,** 185-205.

YEATES, T. O., JORDA, J. & BOBIK, T. A. 2013. The shells of BMC-type microcompartment organelles in bacteria. *J Mol Microbiol Biotechnol,* 23**,** 290-9.

YEATES, T. O., KERFELD, C. A., HEINHORST, S., CANNON, G. C. & SHIVELY, J. M. 2008a. Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nature Reviews Microbiology,* 6**,** 681-691.

YEATES, T. O., KERFELD, C. A., HEINHORST, S., CANNON, G. C. & SHIVELY, J. M. 2008b. Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nat Rev Microbiol,* 6**,** 681-91.

YOSHIDA, A., NISHIMURA, T., KAWAGUCHI, H., INUI, M. & YUKAWA, H. 2005. Enhanced hydrogen production from formic acid by formate hydrogen lyase-overexpressing Escherichia coli strains. *Appl Environ Microbiol,* 71**,** 6762-8.

YOUNG, E. J., SAKKOS, J. K., HUANG, J., WRIGHT, J. K., KACHEL, B., FUENTES-CABRERA, M., KERFELD, C. A. & DUCAT, D. C. 2020. Visualizing in Vivo Dynamics of Designer Nanoscaffolds. *Nano Letters,* 20**,** 208-217.

YU, J. W., PRICE, G.D., SONG, L. AND BADGER, M.R. 1992. Isolation of a putative carboxysomal carbonic anhydrase gene from the cyanobacterium Synechococcus PCC7942. *Plant Physiology,* 100**,** 7.

ZANG, K., WANG, H., HARTL, F. U. & HAYER-HARTL, M. 2021. Scaffolding protein CcmM directs multiprotein phase separation in beta-carboxysome biogenesis. *Nat Struct Mol Biol,* 28**,** 909-922.

ZARZYCKI, J., ERBILGIN, O. & KERFELD, C. A. 2015. Bioinformatic characterization of glycyl radical enzyme-associated bacterial microcompartments. *Applied and environmental microbiology,* 81**,** 8315-8329.

ZARZYCKI, J., SUTTER, M., CORTINA, N. S., ERB, T. J. & KERFELD, C. A. 2017. In Vitro Characterization and Concerted Function of Three Core Enzymes of a Glycyl Radical Enzyme - Associated Bacterial Microcompartment. *Sci Rep,* 7**,** 42757.

ZHANG, J., FAN, C., ZHANG, H., WANG, Z., ZHANG, J. & SONG, M. 2018. Ferric oxide/carbon nanoparticles enhanced bio-hydrogen production from glucose. *International Journal of Hydrogen Energy,* 43**,** 8729-8738.

ZHAO, Y.-Y., JIANG, Y.-L., CHEN, Y., ZHOU, C.-Z. & LI, Q. 2019a. Crystal structure of pentameric shell protein CsoS4B of Halothiobacillus neapolitanus α-carboxysome. *Biochemical and Biophysical Research Communications,* 515**,** 510-515.

ZHAO, Y. Y., JIANG, Y. L., CHEN, Y., ZHOU, C. Z. & LI, Q. 2019b. Crystal structure of pentameric shell protein CsoS4B of Halothiobacillus neapolitanus alpha-carboxysome. *Biochem Biophys Res Commun,* 515**,** 510-515.

ZHU, J., HU, L., ZHAO, P., LEE, L. Y. S. & WONG, K.-Y. 2020. Recent Advances in Electrocatalytic Hydrogen Evolution Using Nanoparticles. *Chemical Reviews,* 120**,** 851-918.

ZINONI F, B. A., STADTMAN TC, BÖCK A. 1986. Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from Escherichia coli. *PNAS,* 83**,** 5.