UNIVERSIDADE FEDERAL DO RIO DE JANEIRO CENTRO DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA VEGETAL E BIOPROCESSOS

# EXPLORANDO A COEXISTÊNCIA DE DOIS CICLOS DE FIXAÇÃO DE CARBONO NA BACTÉRIA TERMOFÍLICA *CARBONACTINOSPORA THERMOAUTOTROPHICA* STC

SULAMITA SANTOS CORREA

RIO DE JANEIRO
2024

SULAMITA SANTOS CORREA

# EXPLORANDO A COEXISTÊNCIA DE DOIS CICLOS DE FIXAÇÃO DE CARBONO NA BACTÉRIA TERMOFÍLICA *CARBONACTINOSPORA THERMOAUTOTROPHICA* STC

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Biotecnologia Vegetal e Bioprocessos, Centro de Ciências da Saúde da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Biotecnologia Vegetal e Bioprocessos.

Orientador: Prof. Dr. Alexandre Soares Rosado
Co-orientadora: Dra. Júnia Schultz

RIO DE JANEIRO
FEVEREIRO, 2024

FICHA CATALOGRÁFICA


CIP - Catalogação na Publicação

S949e

Santos Correa, Sulamita
EXPLORANDO A COEXISTÊNCIA DE DOIS CICLOS DE FIXAÇÃO DE CARBONO NA BACTÉRIA TERMOFÍLICA CARBONACTINOSPORA THERMOAUTOTROPHICA STC / Sulamita Santos Correa. -- Rio de Janeiro, 2024.
214 f.

Orientador: Alexandre Soares Rosado.
Coorientador: Júnia Schultz.
Tese (doutorado) - Universidade Federal do Rio de Janeiro, Decania do Centro de Ciências da Saude, Programa de Pós-Graduação em Biotecnologia Vegetal, 2024.

1. Bactéria termofílica. 2. Vias de fixação de carbono. 3. Carboxissomo. 4. Ômicas. I. Soares Rosado, Alexandre, orient. II. Schultz, Júnia, coorient. III. Título.

Elaborado pelo Sistema de Geração Automática da UFRJ com os dados fornecidos pelo(a) autor(a), sob a responsabilidade de Miguel Romeu Amorim Neto - CRB-7/6283.

FOLHA DE APROVAÇÃO

**Universidade Federal do Rio de Janeiro**

**Coordenação de Pós-Graduação em Biotecnologia Vegetal e Bioprocessos**

**ATA DA DEFESA DE TESE DE DOUTORADO DE SULAMITA SANTOS CORREA COMO PARTE DOS REQUISITOS NECESSÁRIOS À OBTENÇÃO DO GRAU DE DOUTORA EM CIÊNCIAS (BIOTECNOLOGIA VEGETAL E BIOPROCESSOS).**

Aos vinte e oito dias do mês de fevereiro de dois mil e vinte e quatro, às 09 horas, reuniu-se no Auditório do nstituto de Microbiologia Paulo Góes – Bloco I do Centro de Ciências da Saúde da Universidade Federal do Rio de Janeiro, a Banca Examinadora abaixo discriminada para avaliação da Tese de Doutorado da aluna **Sulamita Santos Correa** intitulada "Explorando a coexistência de dois ciclos de fixação de carbono na bactéria termofílica *Carbonactinosopora thermoautotrophica* stc", desenvolvida sob a orientação do **Prof. Alexandre Soares Rosado** e coorientação da **Drª. Júnia Schultz.** A apresentação feita pela candidata foi acompanhada da arguição pelos componentes da Banca. Em seguida, esta se reuniu para sua avaliação e a defesa foi considerada A (inserir letra apropriada).

A) Aprovado;
B) Aprovado com pequenas modificações* a serem combinadas com o Presidente da Banca dentro de um mês;
C) Não aprovado ainda, é necessária a apresentação das modificações/correções* em uma nova versão do documento para o Presidente da Banca e uma carta do Presidente com prazo de máximo 3 meses para ser aprovada; se não reprovada
D) Não aprovado ainda, é necessária uma nova apresentação*, oral e escrita, para a mesma ou nova banca examinadora dentro de um prazo combinado com a comissão do Programa.
E) Reprovado, razões da reprovação* escritas no espaço destinado.

*As modificações/correções/razões para reprovação precisam ser discriminadas e o Presidente da Banca e o(a) aluno(a) precisam estar cientes.

E, para constar, foi lavrada a presente ata que vai devidamente assinada pelo coordenador, pelos membros da Banca Examinadora, pelo Presidente da Banca Examinadora e pelo orientadores doa aluna, que ficará com uma via da ata e entregará outra para a secretaria do PBV.

Rio de Janeiro, 28 de fevereiro de 2024

Universidade Federal do Rio de Janeiro
Centro de Ciências da Saúde

Coordenação de Pós-Graduação em Biotecnologia Vegetal
e Bioprocessos

**ATA DA DEFESA DE TESE DE DOUTORADO DE SULAMITA SANTOS CORREA COMO PARTE DOS REQUISITOS NECESSÁRIOS À OBTENÇÃO DO GRAU DE DOUTORA EM CIÊNCIAS(BIOTECNOLOGIA VEGETAL E BIOPROCESSOS).**

Documento assinado digitalmente
gov.br
BIANCA ORTIZ DA SILVA
Data: 28/03/2024 09:03:46-0300
Verifique em https://validar.iti.gov.br

Drª. Bianca Ortiz da Silva – Coordenadora

Drª. Alane Beatriz Vermelho – PBV/UFRJ

Dr. Jean Luiz Simões de Araújo - PBV/UFRJ

Dr. Marcio Alves Ferreira - PBV/UFRJ

Dr. Ricardo Moreira Chaloub - UFRJ

Drª. Patrícia Moura - LNEG (Portugal)

Dr. Alexandre Soares Rosado (Orientador)

Drª Júnia Schultz (Co-orientadora)

Sulamita Santos Correa.
Sulamita Santos Correa (Doutoranda)

**LISTA DE MODIFICAÇÕES, CORREÇÕES OU RAZÕES PARA REPROVAÇÕES:**

Sem modificações. Sugestões aceitas.

**Ciente,**

**Presidente da Banca**

**Ciente,**

Sulamita Santos Correa.
**Aluno**

Centro de Ciências da Saúde – Bloco K
Sala K2-032 – 3º andar – Cidade Universitária
CEP: 21941-590 – Rio de Janeiro – RJ – Brasil
Tel: +552139386676 - E-mail: pbv@ccsdecania.ufrj.br

“Education is not preparation for life; education is life itself.” John Dewey

Aos meus pais Filomena e Aguinaldo. Às minhas irmãs Elen, Kely, Evelyn e Keverly e os meus Gatos.
Dedico essa Tese de Doutorado.

# AGRADECIMENTOS

Percorrer o caminho da pesquisa científica demanda não apenas dedicação individual, mas também a colaboração inestimável de pessoas, instituições, amigos e familiares. Desde a concepção do projeto até a conclusão deste trabalho, agradeço sinceramente a todos que, de alguma forma, contribuíram para o sucesso desta Tese. Seus apoios, conselhos e encorajamentos foram a força motriz que impulsionou cada etapa, tornando este percurso não apenas acadêmico, mas também enriquecedor e significativo. Expresso meus profundos agradecimentos, reconhecendo que esta jornada não seria completa sem a valiosa participação de cada um de vocês. Em especial, quero expressar meus sinceros agradecimentos:

Apesar de acreditar assiduamente na ciência e seus resultados, algo interno relacionado ao bem-estar traz um conforto sobre a existência de um ser supremo, por isso, não poderia deixar de agradecer ao supremo Deus e a Meishu Sama, o qual busco os ensinamentos para levar uma vida na busca do paraíso interior e na Terra. Expresso minha profunda gratidão aos meus pais, Filomena da Silva Santos e Aguinaldo Santiago Correa, à minha segunda mãe, Elen da Silva Santos Guimarães, minhas queridas irmãs Kely, Evelyn e Keverly, e aos meus cunhados, em especial ao Adriano Guimarães Pimentel. Seu apoio incansável, carinho e presença constante foram fundamentais para o alcance dos meus objetivos acadêmicos. A jornada foi mais significativa e vitoriosa graças ao amor e suporte inabaláveis que recebi de cada um de vocês. Muito obrigado por fazerem parte desta trajetória.

À Universidade Federal do Rio de Janeiro, que foi meu lar científico por seis anos. Obrigada pela oportunidade e formação;

Agradeço aos meus orientadores, Alexandre Soares Rosado e Júnia Schultz, pela inestimável aprendizagem proporcionada ao longo desta jornada acadêmica. Agradeço, também, pela confiança depositada em mim para a realização deste trabalho. Suas orientações e apoio foram essenciais para o meu crescimento profissional e para o sucesso deste projeto. Muito obrigado por serem mentores dedicados e inspiradores. Agradeço à Professora Raquel Peixoto por ter me concedido uma bolsa de estudo complementar. À Dr. Daniele Blasquez, agradeço sua generosidade e confiança, as quais foram fundamentais para o meu desenvolvimento acadêmico e profissional.

À querida Pós-graduação em Biotecnologia Vegetal e Bioprocessos da UFRJ merece meu mais profundo agradecimento. Em especial, expresso minha gratidão ao Professor Andrew Macrae, ao longo desses seis anos de jornada (mestrado e doutorado), espero ter conseguido honrar essa confiança e trazer orgulho à equipe PBV. Agradeço também à secretária Tayline por toda a ajuda incansável. Sempre pude contar com a paciência e suporte de vocês, o que tornou esta jornada ainda mais significativa. Muito obrigada por fazerem parte do meu percurso acadêmico.

Ao órgão de fomento Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), o qual forneceu minha bolsa de estudos durante os quatro anos de Doutorado e pela bolsa de Doutorado Sanduíche no Exterior por 6 meses; À Universidade King Abdullah de Ciência e Tecnologia (KAUST) a qual financiou minha pesquisa e me concedeu uma bolsa Doutorado Sanduíche por 6 meses e complementação por outros 6 meses.

Ao Laboratório Biociência Core da Universidade KAUST, em especial ao Anthony Weatherhead e Papita Mandal pelo treinamento no curso de proteômica e por me ajudarem com as amostras. Ao Laboratório de imagens também da Universidade KAUST, ao Brandon Huntington e ao professor Andreas Naschberger pelo treinamento e ajuda nas análises de microscopia. E agradeço ao Dr. Luis Arge por toda ajuda e conhecimento de bioinformática para me ajudar a processar os dados de genômica e transcriptômica. Agradeço por compartilharem seu conhecimento e experiência, tornando possível a exploração deste campo fascinante.

Aos amigos do Laboratório de Ecologia Microbiana Molecular (LEMM), Helena, Bárbara, Beatriz, Érika, Karen, Phelipe, Eikon, Laenne, Flúvio, Adam e os demais por compartilharem comigo toda experiência de laboratório. Aos amigos que fiz durante o intercâmbio na Universidade KAUST, Sharifah, Raquel, Ping, Leyla, Hamad, Nicholas e Norah, por terem me recebido de braços abertos e me apresentado à cultura Árabe.

Aos meus amigos da vida, Natasha, Milena, Dneson, Paloma, Thuane, Dora, Gilmara e demais por me apoiarem e me fortalecerem em todas as horas.

Por fim, agradeço a mim mesma: “Por último, mas não menos importante, quero me agradecer por acreditar em mim, por nunca desistir, por tentar fazer mais certo do que errado, por ser eu em todos os momentos”. “Quero agradecer a mim mesma, porque eu trabalhei duro”.

**RESUMO**

A fixação de carbono é um processo essencial para a vida na Terra, pois realiza a conversão de carbono inorgânico em moléculas orgânicas. Sete ciclos de fixação autotrófica de carbono foram descritos até o momento; entretanto, é raro encontrar um microrganismo que exiba mais de uma via para a fixação de carbono. O presente trabalho teve como objetivo avaliar a expressão de genes e proteínas relacionados com as vias de fixação de carbono em *Carbonactinospora thermoautotrofica* StC. A estirpe StC é uma *Actinomycetota* aeróbica quimoautotrófica termofilia (temperatura ótima entre 55° e 65° C), isolada de uma área com constante queima de material vegetal. Para entender melhor os mecanismos de fixação de carbono e potenciais aplicações biotecnológicas da espécie e da estirpe StC, foi conduzido uma análise de com multi-ômicas (genoma, transcriptoma e proteoma), em duas condições de cultivo, autotrófica e heterotrófica, usando a tecnologia Oxford Nanopore, Illumina HiSeq e LC- MS/MS, respectivamente. Além disso, foi purificado um microcompartimento de concentração de $CO_2$, chamado carboxissomo, o qual encapsula a enzima RuBisCo. Como resultado, observou- se uma superexpressão de genes e proteínas relacionados tanto ao ciclo de Calvin-Benson-Bassham (CBB) quanto a uma variante do Ciclo Redutor de Citrato (rTCA) em células cultivadas autotroficamente. Genes que codificam enzimas chave do ciclo CBB, incluindo RuBisCo forma I, fosforibuloquinase (PRK) e Gliceraldeído-3-fosfato desidrogenase (GAPDH), foram expressos, juntamente com genes que codificam proteínas fosfoenolpiruvato carboxilase e 2-oxoglutarato/2-oxoácido ferredoxina oxidoredutase, componentes cruciais para o ciclo rTAC, e estas proteínas também foram expressas. Estas descobertas sugerem que *C. thermoautotrofica* StC emprega ambas as vias (ciclos CBB e rTCA) que podem estar trabalhando simultaneamente para a fixação de carbono, sendo está a primeira vez que a expressão de genes e proteínas relacionadas com duas vias de fixação de carbono é demonstrada em um microrganismo. Adicionalmente, foi observada a possível presença de um carboxissomo ainda não descrito no filo *Actinomycetota*. Os dados dessas análises visam fornecer uma compreensão da fisiologia e do potencial metabólico da StC, e fornecem base e conhecimento para estudos futuros que buscam esclarecer os motivos para um microrganismo possuir vias duplas de fixação de carbono, pois isso requer mais energia ou potencialmente indica uma vantagem energética em certas fases do seu ciclo de vida.

**Palavras-chave:** bactéria termofílica, vias de fixação de carbono, carboxissomo, ômicas.

**ABSTRACT**

Carbon fixation, an essential process for life on Earth, involves the conversion of inorganic carbon to organic molecules. Seven cycles of autotrophic carbon fixation have been described so far; however, it is rare to find a microorganism that exhibits more than one carbon fixation pathway. Here, we evaluate the genes and protein expression related with carbon fixation pathways in *Carbonactinospora thermoautotrophica* strain StC. StC is a thermophilic chemoautotrophic aerobic *Actinomycetota*, isolated from a burning organic matter and thrives at temperatures between 55° and 65° C. To better understand the carbon fixation mechanisms and potential biotechnological applications, we conducted genomic, transcriptomic and proteomic analysis on autotrophically and heterotrophically grown bacterial cells using Oxford Nanopore, Illumina HiSeq technology and LC-MS/MS. In addition, we purified a microcompartment possible carboxysome which encapsulates the RuBisCo enzyme. Our results revealed the overexpression of genes and proteins related with Calvin- Benson-Bassham (CBB) cycle and a variant of Reductive Citrate Cycle (rTCA) in autotrophically cultivated cells. Genes encoding key enzymes of CBB cycle, including RuBisCo form I, phosphoribulokinase (PRK) and Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), were expressed, along with genes which encoding phosphoenolpyruvate carboxylase and 2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase, crucial components for the rTCA cycle, as these proteins were expressed as well. These findings suggest that *C. thermoautotrophica* StC employs both pathways (CBB and rTCA cycles) and may be working simultaneously for carbon fixation within single cells. Our results hint at a novel discovery, namely that genes and protein expression related to two carbon fixation pathways are active in the same microorganism and the possible presence of a carboxysome not described in the *Actinomycetota* phylum. The data from these analyzes aim to provide an understanding of the physiology and metabolic potential of strain StC, it sheds light on possible future studies seeking to clarify why a microorganism can have dual carbon fixation pathways, as this requires more energy or potentially indicates an energetic advantage in certain stages of its life cycle.

**Keywords:** thermophilic bacteria, carbon fixation pathways, carboxysome, omics.

LISTA DE FIGURAS

# LISTA DE ABREVIAÇÕES, SÍMBOLOS E FÓRMULAS

°C - Graus celsius

µg - Micrograma

µL - Microlitro

PG - Fosfoglicolato

HP - Biciclo 3-hydroxyproprionato

3-PGA - 3-fosfoglicerato

ATP - Trifosfato de adenosina

BMC - Microcompartimento

Br - Bromo

C - Carbono

Ca - Cálcio

CaCl2. 2H2O - Cloreto de cálcio diidratado

CBB - Ciclo de Calvin-Benson-Bassham

CCL - Citril-CoA liase

CCS - Citril-CoA sintetase

cDNA - DNA complementar

CDSs - Sequências codificadoras de proteínas

CETCH - Ciclo crotonil-(CoA)/etilmalonil-CoA/hidroxibutiril-CoA

CID - Colisão de dissociação

circRNA - RNAs circulares

cm - Centímetro

Co - Cobalto

CO - Monóxido de carbono

$CO_2$ - Dióxido de carbono

CoA - Acetil coenzima A

$CoCl_2$ - Cloreto de cobalto

Crio-ME - Microscopia eletrônica criogênica

Crio-ET - Tomografia crioeletrônica

CS - Citrato sintase

Cu - Cobre

$CuCl_2$ - Cloreto de cobre

DC/4-HB - Ciclo dicarboxilato/4-hidroxibutirato

DDa - Aquisição Dependente de Dados

DEPC - Pirocarbonato de Dietila

DIC - Microscopia de contraste de interferência diferencial

DNA - Ácido desoxirribonucleico

DTT - Ditiotreitol

E - Coeficiente de extinção

ECD - Dissociação por captura de elétrons

EDTA - Ácido etilenodiamino tetra-acético

EID - Dissociação induzida por elétrons

ETD - Dissociação de elétron de transferência

Eut - Etanolamina

FDR - Taxa de Descoberta Falsa

g - Grama

GAPDH - Gliceraldeído-3-fosfato desidrogenase

GC - Guanina, citosina

H - Hidrogênio

h - Hora

$H_3BO_3$ - Ácido bórico

HCD - Dissociação de colisão de energia mais alta

HCl - Ácido clorídrico

HR-ICP-MS - Espectrometria de massa por plasma acoplado indutivamente

K - Potássio

$K_2HPO_4$ - Fosfato dipotássico

kDa - Quilodalton

KH2PO4 - Fosfato monopotássico

kV - quilovoltz

L - Litro

LC-MS - Cromatografia líquida-espectrometria

lncRNAs - RNAs não codificantes longos

M - Molar

m/z - Relação massa-carga

Mg - Magnésio

mg - Miligrama

$MgSO_4$. $7H_2O$ - Magnésio hepta-hidratado

min - Minuto

miRNAs - microRNAs

mL - Mililitro

mM - Milimolar

$MnCl_2$. $4H_2O$ - Cloreto de manganês tetra-hidratado

Mo - Molibdênio

Moco - cofator de molibdênio

MS - Espectrometria de massas

ms - Milissegundos

MW - Peso molecular

$N_2$ - Nitrogênio

$Na_2MoO_4$ - Molibdato de sódio

$Na_2MoO_4$. $2H_2O$ - Molibdato de sódio di-hidratado

$Na_2SeO_3$ - Selenito de sódio

$NaCl_2$ - Cloreto de sódio

NADH - Dinucleótido de nicotinamida e adenina reduzido

NADP - Fosfato de dinucleótido de nicotinamida e adenina

$NH_4Cl_2$ - Cloreto de amónio

Ni - Níquel

$NiCl_2$ - Cloreto de níquel

NiFe - Níquel-ferro

nl/minute - Litros normais por minutos

nm - Nanômetro

nrGly - Ciclo redutor natural de glicina

$O_2$ - Oxigênio

ONT - Oxford Nanopore Tecnologia

P - Fósforo

Pdu - Propanodiol

ppm - Parte por milhão

PRK - Fosforribuloquinase

PSM - correspondências do espectro peptídico

R2A - Meio de cultura ágar Reasoner 2

RHP - Via redutora da hexulose-fosfato

RNA - Ácido ribonucleico

roTCA - ciclo oxidativo reverso do TCA

rpm - Rotações por minuto

rRNA - Ácido ribonucleico ribossômico

rTCA - Ciclo redutor de citrato

RuBisCo - ribulose-1,5-bisfosfato carboxilase oxigenasse

RuBP - Ribulose-1,5-bifosfato

S - Enxofre

s - Segundos

SACA - Ciclo sintético de acetil-CoA

SBP - Sedoheptulose-bifosfatase

SBPase - Sedoheptulose-1,7-bifosfatase

SDS - Dodecil sulfato de sódio

SPM - Microscopia crioeletrônica de partícula única

TEM - microscopia eletrônica de transmissão

TEV - microscopia eletrônica de varredura

TPM - Transcrição por milhão

Tris - Hidroximetil-aminometano

tRNA - Ácido ribonucleico transportador

V - Volt

Vol - Volume

WL - Wood-Ljungdahl

Zn - Zinco

$ZnSO_4 . 7H_2O$ - Sulfato de zinco hepta-hidratado

α - Alfa

β - Beta

ɣ - Gama

SUMÁRIO

# 1. INTRODUÇÃO

## 1.1 *Carbonactinospora thermoautotrophica*: visão geral

*Carbonactinospora thermoautotrophica* é uma *Actinomycetota* a qual inicialmente acreditava-se ser classificada como pertencente ao gênero *Streptomyces*, e na época foi nomeada de *Streptomyces thermoautotrophicus* (Gadkari et al., 1990; Volpiano et al., 2021). *Actinomycetota* é um filo de bactérias filamentosas pertencente à ordem *Actnomicetales* (Hazarika; Thakur, 2020), sendo um dos filos mais dominantes dentro do domínio das bactérias (Ventura et al., 2007). Seus representantes são conhecidos por possuir uma ampla diversidade bioquímica e nichos ecológicos que incluem ecossistemas terrestres e aquáticos, com crescimento em diferentes faixas de temperatura (Shivlata; Satyanarayana, 2015).

As *Actinomycetotas* possuem alto conteúdo GC (guanina; citosina), produzem micélios e um grande número dessas bactérias se reproduzem por esporulação (Hazarika; Thakur, 2020). Também são notáveis pela sua variedade morfológica, que inclui cocos ou bastonetes celulares (ex.: *Micrococcus*, *Mycobacterium* e *Arthrobacter*) e bactérias morfologicamente complexas (ex.: *Amycolatopsis*, *Frankia* e *Streptomyces*) (Yadav et al., 2018). Essa diversidade fisiológica e ecológica reflete no potencial metabólito e na versatilidade dos membros pertencentes ao filo *Actinomycetota* (Van Bergeijk et al., 2020).

A primeira estirpe de *Carbonactinospora thermoautotrophica* foi isolada de uma "pilha" de carvão vegetal constantemente em chamas, e foi estudada por membros do laboratório de Ortwin Meyer, na Universidade Bayreuth, na Alemanha (Gadkari et al., 1990). UBT1 foi a primeira estirpe de *C. thermoautotrophica*, descrita como uma *Actinomycetota* termofilia, quimiolitoautótrofo, ou seja, obtém energia a partir de compostos inorgânicos como hidrogênio e monóxido de carbono (H e CO respectivamente) e fixadora de carbono (C). Entretanto, a fixação de carbono não foi explorada e não foi descrita a via utilizada pela estirpe para fixar carbono. Posteriormente, duas outras estirpes - H1 e P1-2 - foram isoladas (também provenientes de ambiente de queima de carvão) por Mackellar et al. (2016). Nesse estudo foi mostrado que as estirpes UBT1 e H1 compartilhavam uma média de 95% de identidade entre si, e >99% de nucleotídeos com os genes da bactéria *Hydrogenibacillus schlegelii* (Mackellar et al., 2016). Dessa forma foi sugerido que a estirpe UBT1 não pertencia ao gênero *Streptomyces* e precisava de uma nova reclassificação.

A taxonomia da espécie de *Streptomyces thermoautotrophicus* foi revista pelo nosso grupo de pesquisa, e foi proposto uma reclassificação para *Carbonactinospora*

*thermoautotrophica*, um novo gênero e nova família (*Carbonoactinsporaceae*) de *Actinomycetota* (Volpiano et al., 2021). Recentemente, uma nova estirpe de *C. thermoautotrophica* também foi descrita, isolada a partir de um consórcio microbiano termofílico, quimiolitoautotrófico e aeróbio (Pinheiro et al., 2023) e foi nomeada como estirpe StC. Além da estirpe StC, outras três bactérias faziam parte desse consórcio microbiano, são elas: *Sphaerobacter thermophilus*, *Chelatococcus* sp., e *Geobacillus* sp. No entanto, StC se apresenta dominante no consórcio, sendo carboxidotrófico com capacidade de fixar dióxido de carbono (Pinheiro et al., 2023). De acordo com os autores, acredita-se que no consórcio, a *C. thermoautotrophica* estirpe StC fornece carbono e energia para a sobrevivência das outras bactérias.

*Carbonactinospora thermoautotrophica* StC é uma *Actinomycetota* filamentosa, que apresenta produção de esporos, e termofílica, com a temperatura ótima de crescimento entre 50 °C e 65 °C. StC é uma bactéria quimiolitoautotrófica, aeróbica, capaz de crescer em meio mineral mínimo sem a presença de fonte orgânica de carbono e nitrogênio, apenas com a atmosfera controlada, com adição dos gases $N_2$, CO, $CO_2$ e $O_2$ e em meio de cultura oligotrófico (Fig. 1a-b).

Trabalhos anteriores com microscopia eletrônica de transmissão (TEM) desenvolvido no Laboratório de Ecologia Microbiana Molecular da Universidade Federal do Rio de Janeiro mostraram que a parede celular da estirpe StC é Gram positiva, o citoplasma é separado por septo e há a presença de microcompartimento bacteriano (BMC) (Moreira, 2018) (Fig. 1c-d). Também foi observado através de microscopia eletrônica de varredura (MEV), a presença de substâncias extracelulares produzidas nos filamentos e projetadas para fora da colônia (Moreira, 2018) (Fig. 1e). Possivelmente, essas substâncias sejam lipopolissacarídeos, uma molécula constituída por lipídeos e polissacarídeos que confere flutuabilidade à colônia em meio líquido, devido sua hidrofobicidade (Wai et al., 2022). A presença de lipopolissacarídeos e os filamentos com presença de esporos em suas bordas são semelhantes às descritas para a estirpe UBT1 de *C. thermoautotrophica* (Gadkari et al., 1992; Mackellar et al., 2016).

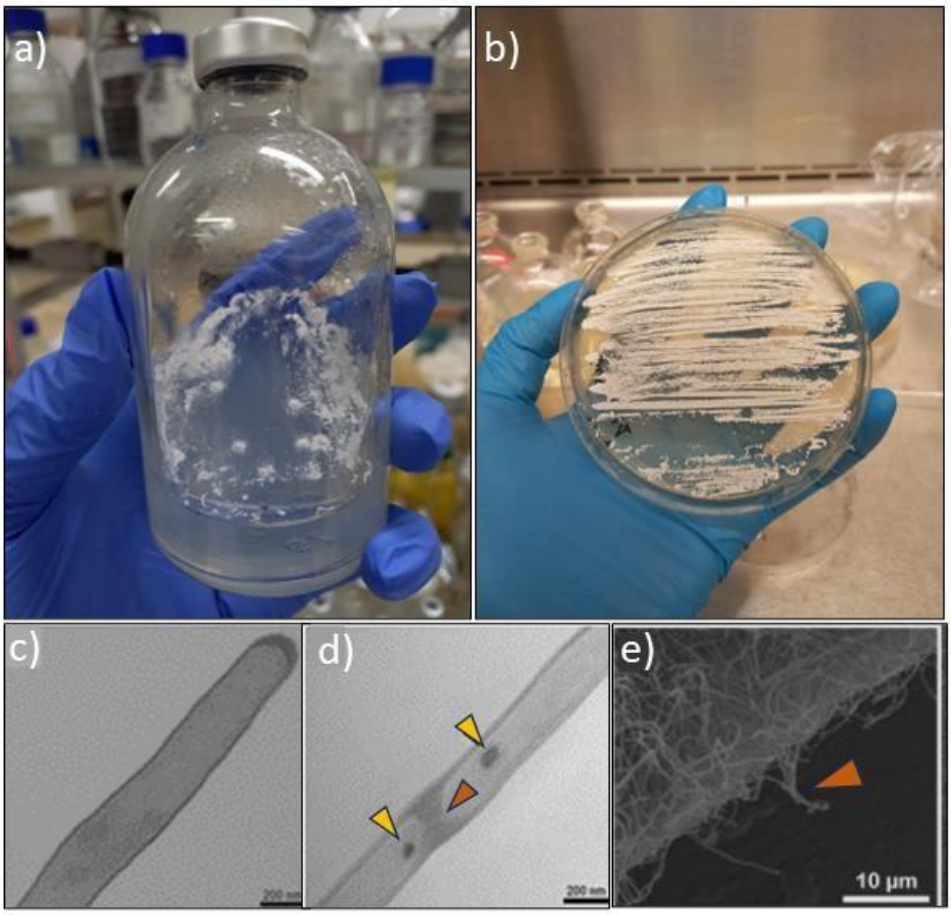



Figura 1: Visão geral morfológica da *Actinomycetota Carbonactinospora thermoautotrophica* StC. (a) Estirpe StC cultivada em meio de cultura N-FIX com adição dos gases CO, $CO_2$, $N_2$ e $O_2$ (sem fonte orgânica de carbono e nitrogênio). (b) Estirpe StC cultivada em meio de cultura oligotrófico R2A. (c) Microscopia eletrônica de transmissão 200 nm: morfologia do filamento da estirpe StC com parede celular gram-positiva. (d) Microscopia eletrônica de transmissão 200 nm: formação de septos comum entre o filo *Actinomycetota* (indicado pela seta laranja) e presença de BMC (indicado pela seta amarela). (e) Microscopia eletrônica de varredura 10 µm: formação de estruturas semelhantes a lipopolissacarídeo, como no filo *Actinomycetota* (indicado pela seta laranja). Imagens de microscopia realizada por Moreira, 2018.

A fixação de carbono por microrganismos quimioautotróficos tem um grande impacto na transição de compostos inorgânicos para orgânicos (Jiao et al., 2021). Organismos autotróficos assimilam carbono inorgânico em biomassa usando processos que são impulsionados pela luz ou energia química derivada da quebra de moléculas orgânicas. Quimioautotróficos são organismos que obtêm energia de compostos orgânicos ou inorgânicos (CO, $H_2$) para produzir moléculas orgânicas a partir do dióxido de carbono (Thevasundaram et al., 2022). Os microrganismos quimiolitoautotróficos usam compostos inorgânicos (CO, $CO_2$, $H_2$, sulfato, fosfito e Fe) como fontes de energia, enquanto os quimioorganotróficos dependem da energia produzida a partir de transformações químicas usando compostos orgânicos (açúcares, ácidos orgânicos e aminoácidos) como fonte de carbono. Entre os quimioautotróficos, o grupo de bactérias que utilizam o CO como doadores de elétrons para

realizar a fixação de $CO_2$ são definidos como carboxidotróficos/carboxidobactéria (Omae et al., 2019). As bactérias aeróbicas oxidantes de CO são capazes de crescer em meio de cultura suplementado com 5% a 20% (vol/vol) de oxigênio e 20% a 95% (vol/vol) de CO (Stevenson et al., 2004).

As carboxidobactérias possuem uma fisiologia diversa, como por exemplo *Pseudomonas carboxidoflava*, que contém CO oxidoredutase mesmo durante crescimento em piruvato, glicose, $H_2$ + $CO_2$ e outros substratos (Bowien; Schlegel, 1981), assim como a *C. thermoautotrophica* StC. Esses microrganismos com a habilidade de assimilar $CO_2$ pela via Calvin-Benson e a habilidade de usar uma ampla gama de outros substratos como energia, são conhecidos como quimioautotróficos facultativos (Pu; Han, 2022). Estas possuem uma enorme variedade metabólica e capacidade mixotrófica, ou seja, suportam o crescimento autotróficos e heterotróficos (Luedin et al., 2019).

Como previamente mencionado, a StC foi isolada a partir de um consórcio microbiano obtido a partir de um solo coletado embaixo de uma pilha de galhos e grama queimada, localizado no município de Seropédica, Rio de Janeiro, Brasil (Pinheiro et al., 2023). A StC pode ser considerada uma bactéria de vida livre do solo, e as bactérias autotróficas de vida livre requerem compostos inorgânicos para servirem como doador de elétrons e compostos orgânicos como suprimento de substratos. A comunidade microbiana do solo é dependente das bactérias quimioautotróficas, pois essas bactérias são responsáveis pela produção primária de compostos orgânicos, sugerindo que o metabolismo quimioautotrófico pode ser a base da cadeia alimentar nessas comunidades de solo (Wang et al., 2022b).

### 1.2. *Actinomycetotas* do solo e a relação com a fixação de carbono

O solo é um ambiente que apresenta grande diversidade de microrganismos vivendo de forma livre ou associativa. São as propriedades físicas, químicas e biológicas do solo que tornam a comunidade microbiana diversa nesse ecossistema. Dentre esses microrganismos, as bactérias são as que mais se destacam, seja pela sua abundância ou pela sua alta competitividade e simbiose (Barriuso et al., 2008). Embora as *Actinomycetotas* possam ser encontradas em diversos nichos ecológicos, o solo representa o habitat mais predominante para esse filo, com o gênero *Streptomyces* existindo em maior número (Panneerselvam et al., 2021). As *Actinomycetotas*, junto com outros organismos presentes no solo, desempenham um importante papel na decomposição da matéria orgânica e na reciclagem de nutrientes nos ciclos

biogeoquímicos (Adenan et al., 2020). A mineralização da matéria orgânica contribui para o fornecimento de energia para os microrganismos, tornando disponível o carbono, nitrogênio, e o fósforo, entre outros nutrientes (Zhang et al., 2016). As *Actinomycetotas* também produzem uma gama de enzimas e compostos bioativos que ajudam no equilíbrio da microbiota do solo, além de possuírem um grande potencial para serem usadas na biorremediação do solo, no uso industrial, farmacêutico e no sequestro de carbono (Priyaragini et al., 2013; Shekhar et al., 2014; Araujo-Melo et al., 2019).

A fixação de carbono está presente em muitos filos bacterianos, o que inclui além das *Actinomycetotas*, as *Cyanobacterias*, *Aquificota*, *Chloroflexota*, *Pseudomonadota*, *Bacillota*, entre outros (Saini et al., 2011). O carbono junto com o nitrogênio e fósforo são os principais nutrientes limitantes para a sobrevivência da comunidade microbiana do solo (Johnson et al., 2017; Romero-Rodrigues et al., 2018). Dentre o filo das *Actinomycetota*, os membros do gênero *Streptomyces* são os mais abundantes no solo, e participam da ciclagem do carbono preso em resíduos orgânicos insolúveis como plantas e fungos, sendo essa ação viabilizada pela produção de exoenzimas hidrolíticas (Barka et al., 2016). Ademais, as bactérias fixadoras de carbono realizam uma importante função ao sequestrar o carbono atmosférico, o que ajuda a reduzir o fenômeno do efeito estufa e consequentemente o aquecimento global. Estudos com metagenomas com amostras de solo do deserto Namíbia, mostram os altos níveis da transcrição de genes relacionados com a fixação de carbono em bactérias quimioautotróficas em contraste com a expressão de genes fotossintéticos, indicando que a quimioautotrofia é uma importante alternativa à fotossíntese para a ciclagem do carbono em solos (León-Sobrino et al., 2019). A fixação de carbono quimioautotrófica e/ou mecanismos de assimilação de $CO_2$ são responsáveis pela entrada de carbono inorgânico na comunidade microbiano do solo (King; Weber, 2007; Pratscher et al., 2011).

Até o momento sete vias de fixação de carbono foram descritas distribuídas entre diversos grupos de organismos, sendo eles: I) ciclo de Calvin-Benson-Bassham (CBB), II) ciclo redutor de citrato (rTCA), III) 3- hidroxipropionato biciclo (3-HP), IV) ciclo Wood-Ljungdahl (WL), V) ciclo dicarboxilato/4-hidroxibutirato, VI) ciclo 4-hidroxibutirato, e descrito mais recentemente VII) via redutora da glicina (Correa et al., 2023). Entre as vias de fixação de carbono o ciclo CBB é a mais abundante entre os micoorganismos (McKinlay; Harwood, 2010). Mais informações sobre as vias de fixação de carbono estão descritas no capítulo I desta tese.

No filo *Actinomycetotas* são encontradas algumas vias de fixação de carbono, conforme descrito a seguir. Genes envolvidos no ciclo rTCA também são transcritos amplamente em

vários táxons, e significativamente ativos em *Actinomycetotas*. Análise com metagenoma indicou a transcrição de genes pertencentes ao ciclo rTCA em muitas famílias de *Actinomycetotas*: *Streptomycetaceae*, *Thermomonosporaceae*, *Rubrobacteraceae*, *Frankiaceae*, entre outras (León-Sobrino et al., 2019). Enzimas chaves do 3-HP biciclo (metilfumaril -CoA isomerase, malonil-CoA redutase, e propionil-CoA sintase) também já foi encontrado em representante do filo *Actinomycetota* (Oren; Garrity, 2021) especulando a habilidade dessas bactérias de realizar autotrofia com esse ciclo (Ruiz-Fernández et al., 2020). Recentemente uma nova ordem/espécie de *Actinomycetota* foi proposta, *Anaerosoma tenue* (mesófila) e *Parvivirga hydrogeniphila* (termofilia), ambas descritas por supostamente codificarem genes para a via redutora da glicina para fixação autotrófica de $CO_2$ (Khomyakova et al., 2022).

### 1.3. Aplicação da tecnologia das ômicas para o estudo de microrganismos

O estudo de biologia de sistemas possibilitou avanços na área de biotecnologia e microbiologia. O projeto inicial do genoma humano impulsionou o conhecimento na área da biologia molecular (Venter et al., 2001). O rápido desenvolvimento de tecnologias na área, como o sequenciamento de alto rendimento e da espectrometria de massas, levaram ao conceito de ômicas, o que proporcionou o avanço metodológico do entendimento do sistema celular e metabólico (Dai; Shen, 2022). Atualmente, são conhecidas diferentes esferas dentro do campo das ômicas, que inclui: genômica, transcriptômica, proteômica, metabolômica, epigenômica, lipidômica, glicômica e metalômica, as quais permitiram uma visão holística sobre as comunidades microbianas que habitam ambientes diversos (De Sousa; Doolan 2016; Ferreira, et al., 2019; Gupta et al., 2020; Rodríguez et al., 2020). No presente estudo, vamos abordar com mais detalhes a genômica, transcriptômica e proteômica, as quais foram utilizadas para elucidar a taxonomia e o potencial metabólico de *C. thermoautotrophica* StC.

A primeira década do século XXI foi marcada pelo desenvolvimento tecnológico no campo da genética e da genômica; houve a transição do esforço internacional para o mapeamento do genoma humano o que demorou mais de uma década para ser concluído e custou milhares de milhões de dólares, para o sequenciamento individual a baixo custo e curto espaço de tempo (McGuire et al., 2020). Atualmente é possível obter sequências completas do genoma de milhares de espécies, bem como de muitos indivíduos dentro das espécies, e estas informações transformaram a compreensão da quantidade, distribuição e significado funcional da variação genética nas populações naturais (Allendorf et al., 2010).

O genoma contém a informação genômica completa de um organismo ou célula, essas informações são armazenas nos ácidos nucleicos de cadeia simples (RNA) ou dupla (DNA) em uma sequência linear ou circular (Giani et al., 2020). Para determinar com precisão as informações contidas nas sequencias do DNA (sequenciamento do genoma) foram desenvolvidas tecnologias progressivamente eficientes, caracterizadas por maior precisão, rendimento e velocidade de sequenciamento (Richards et al., 2010). O sequenciamento genômico representa a sequência completa de bases nucleotídicas para todos os cromossomos da espécie de interesse. Funciona como um "mapa físico" do conteúdo genético de um organismo (em oposição ao "mapa genético ou de ligação" que estabelece a ordem e as distâncias de recombinação entre genes marcadores) (Ekblom; Wolf, 2014).

A tecnologia que revolucionou o campo de sequenciamento de genomas e vem sendo amplamente utilizada, é conhecida como *Next Generation Sequencing* (NGS - Sequenciamento de Próxima Geração), essa plataforma foi desenvolvida para abordar genomas maiores, em um processo denominado *Whole Genome Sequencing* (WGS - Sequenciamento do Genoma Completo) (Giani et al., 2020). As tecnologias conhecidas como segunda geração como *Ion Torrent* e *Illumina*, são caracterizadas pela geração de *reads* curtas (leitura), enquanto as tecnologias de terceira geração, como PacBio e NanoPore, caracterizam-se pela síntese de *reads* longas (Wick et al., 2023). O sequenciamento do genoma revolucionou a biologia molecular, a medicina, a genômica e campos afins, levando ao desenvolvimento de novas plataformas e aprimoramento do sequenciamento de DNA. Estas tecnologias, juntamente com diversas ferramentas computacionais para análise e interpretação de dados, têm ajudado na compreensão dos genomas de vários organismos (Ekblom; Wolf, 2014).

O sequenciamento do genoma segue duas estratégias: a genômica estrutural e a genômica funcional. No genoma estrutural, o DNA genômico é fragmentado e sequenciado, gerando informações de regiões gênicas e intergênicas (Byrne et al., 2019). O genoma funcional, ou também chamado de transcriptoma, se refere ao sequenciamento dos genes transcritos em larga escala (Martin; Wang, 2011).

O transcriptoma é dinâmico e está sujeito a alterações impostas pelas condições fisiológicas, temporais e ambientais, as quais influenciam no estágio de desenvolvimento celular através de estímulos internos e/ou externos (Dai; Shen, 2022). Com a técnica de transcriptômica é possível detectar e quantificar moléculas de RNA transcritas em um genoma em um determinado momento e/ou condição (Wang et al., 2009). Através da identificação do conjunto completo de transcritos, que inclui grandes e pequenos RNAs, genes não anotados,

isoformas de "*splicing*", e fusão de genes, tornam possível o estudo abrangente do tradicional transcriptoma (Martin; Wang, 2011). O transcriptoma também se refere a outros tipos de transcrições, como microRNAs (miRNAs), assim como RNAs não codificantes longos (lncRNAs) e RNAs circulares (circRNA) (Wang et al., 2009). A análise transcriptômica permite o estudo em ampla escala das funções e interações de genes e proteínas, buscando entender e definir a função de um gene em um organismo (Paterson et al., 2009).

Em relação a proteômica, se refere a compreensão da expressão total de proteínas em um organismo (Dutt; Lee, 2000). A proteômica permite a análise simultânea de proteínas expressas nas células sobre diferentes condições fisiológicas ou pré-determinadas, interrogando o fluxo de informações através da sinalização de proteínas (Petricoin et al., 2002). A proteômica consiste no estudo de proteínas utilizando técnicas de separação e identificação, como eletroforese, cromatografia, espectrometria de massas, e bioinformática (Hein et al., 2013).

Os estudos qualitativos, quantitativos e a elucidação estrutural das proteínas são fundamentais para a compreensão da decodificação de mecanismos moleculares, vias metabólicas, e modificações pós-tradução dentro da célula (Chandran; Sharma, 2020). Através da espectrometria de massas (MS) é realizada a caracterização de proteínas por meio da determinação da sequência de aminoácidos de uma proteína (Kawashima et al., 2019). A eficácia da análise do proteoma e qualidade dos resultados também dependem da quantidade e qualidade das proteínas extraídas, assim como da técnica adotada para o fracionamento das proteínas, e da identificação dos peptídeos/proteínas por meio da pesquisa em banco de dados e interpretação de dados (Nickerson; Doucette, 2020).

### *3.3.1. Análise integrada de genômica, transcriptômica e proteômica no estudo das vias de fixação de carbono*

À medida que a disponibilidade das ômicas têm avançado e o custo têm reduzido, associado ao crescente conceito de que nenhuma das técnicas isoladas é capaz de lidar com a complexidade dos sistemas biológicos em sua totalidade (Bathke et al., 2019), uma abordagem combinada de genômica, transcriptômica e proteômica podem fornecer informações mais aprofundadas sobre as mudanças moleculares em microrganismos. A análise genômica, transcriptômica e a proteômica possuem características específicas e acumulam um grande volume de dados, o que permite o melhor entendimento dos processos celulares e metabólicos dos microrganismos (Liu et al., 2016). As análises moleculares integradas podem proporcionar uma melhor compreensão sobre a estrutura, o funcionamento e a dinâmica dos sistemas

metabólicos (Haider et al., 2013). Desta forma, uma abordagem combinada da quantificação dos transcritos e proteínas em uma escala genômica fornece um panorama mais complexo de múltiplos processos além da concentração de transcrições que contribuem para estabelecer o nível de expressão de uma proteína (McManus et al., 2015) (Fig. 2).

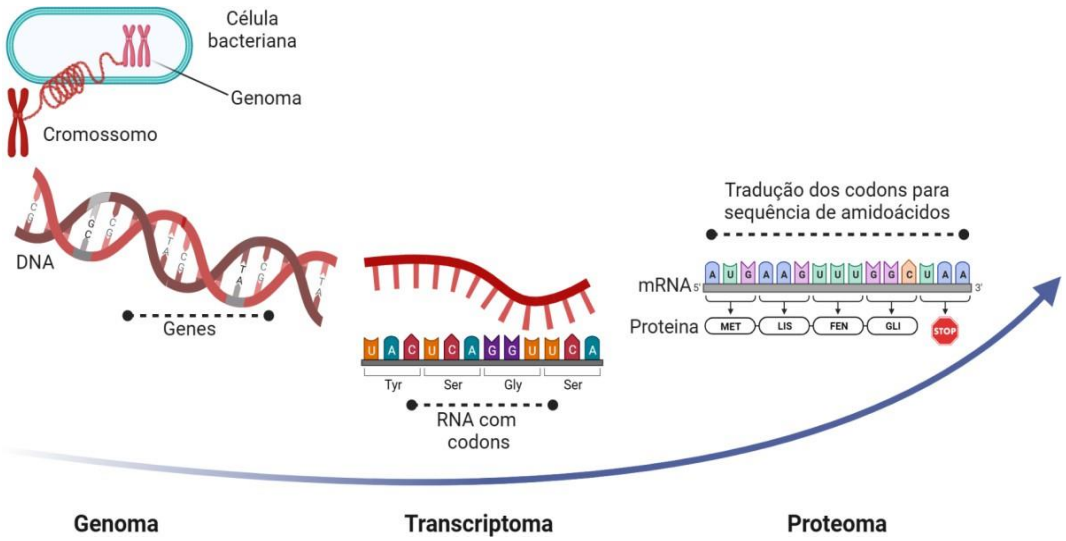



Figura 2: Ilustração do processo de integração entre: genoma, transcriptoma e proteoma. Figura criada com a plataforma BioRender.com.

O nível de expressão de um gene não reflete diretamente na quantidade de proteína sintetizada, sua localização e atividade biológica, sendo necessárias informações adicionais e complementares, as quais são obtidas pela análise do proteoma (Yan et al., 2022). Ainda há os níveis de transcrições gênicas que podem ocorrer de forma pós-transcricional ou pós-traducional, os níveis de expressão proteica dependem não apenas dos níveis de transcrição, mas também da eficiência translacional e da degradação regulada (Liu et al., 2016). Os organismos vivos constantemente remodelam mRNA e proteínas como resposta a fatores ambientais, ou no curso de desenvolvimento e diferenciação, como as bactérias que podem ajustar a expressão gênica nos níveis transcricionais e traducionais (Berghoff et al., 2013).

Além da regulação transcricional e traducional, os eventos de pós-transcrição e pós-tradução têm uma importância e não podem ser negligenciados, pois ajudam a explicar a discrepância entre o mRNA e os níveis de proteínas observados regularmente em sistemas biológicos (Maier et al., 2009). De fato, foi observado que a regulação pós-transcrição desempenha um papel importante na determinação de proteínas acumuladas em *Rhodobacter sphaeroides*, visto que o número de proteínas detectadas pela análise de proteômica havia apresentado apenas a metade dos transcritos e que a variação no acúmulo de proteínas durante a transição para a fase estacionária foi mais dinâmica (Bathke et al., 2019). A transcrição

generalizada do genoma cria um reservatório de ncRNAs, no entanto há um conhecimento limitado sobre as funções e mecanismos reguladores desses RNAs; portanto tem sido aceito que as interações entre a ligação do RNA e proteínas desempenham papéis vitais na manutenção da homeostase celular (Micheel et al., 2021).

O estudo do genoma, transcriptoma e proteoma tem possibilitado a identificação de genes envolvidos na fixação de carbono, permitindo a previsão de potenciais rotas de fixação de carbono em micróbios ambientais, bem como a descoberta e desenho de nova vias para a fixação de carbono (McAllister et al., 2020; Correa et al., 2023). Essas descobertas não eram possíveis apenas com as abordagens moleculares anteriores, principalmente em bactérias não cultiváveis (Hegde et al., 2003). A análise integrada de transcriptoma e proteoma de *Leptospirillum ferriphilum* DSM 14647T revelou que essa bactéria possui todos os genes e proteínas expressos do ciclo de fixação de carbono rTCA, o que antes não havia sido possível confirmar apenas com a anotação genômica (Hugler; Sievert, 2011; Christel et al., 2017). A descoberta da sétima via de fixação de carbono, a via redutiva da glicina, só foi possível através da aplicação de multi-ômicas (genômica, transcriptômica, proteômica e metabolômica) nos experimentos com *Desulfovibrio desulfuricans* (Sánchez-Andrea et al., 2020).

Até o momento existem especulações sobre microrganismos cultivados e isolados que usam mais de uma via de fixação de carbono para gerar energia, e esse entendimento vem se tornando possível com análise comparativa entre as ômicas. Estudos com metagenômica, metatranscriptoma e metaproteômica no verme tubular *Paraescarpia echinospica* (holobionte) indicou a existência das vias CBB e rTCA em seu simbionte (Yang et al., 2020). Em bactéria isolada, genes para esses mesmos ciclos (CBB e rTCA) foram detectados na gammaproteobacteria *Thioflavicoccus mobilis* (Rubin-Blum et al., 2019). Neste estudo análises de transcriptoma e proteoma seriam fundamentais para complementar os dados genômicos, com medição direta da abundância de genes e proteínas.

Os dados obtidos através da proximidade entre transcriptômica e proteômica possibilitam a realização do estabelecimento de associações entre o potencial genético e o fenótipo final. Esse conhecimento pode estabelecer a base para a descoberta de novos genes, proteínas e enzimas em uma escala muito maior em comparação com os esforços anteriores. De uma maneira mais ampla, o conhecimento obtido através da biologia de sistemas poderá otimizar processos biotecnológicos de origem microbiana, através de um melhor controle dos processos em cultura pura ou pelo uso de enzimas mais eficientes em aplicações industriais e de bioengenharia.

## 1.4. Carboxissomo: uma organela especializada entre os procariontes

Algumas bactérias possuem a habilidade de desenvolver microcompartimentos bacteriano (BMCs) que funcionam como organelas a base de proteínas, as quais segregam enzimas para otimizar reações metabólicas (Chen et al., 2023). Até o momento foram caracterizados alguns BMCs, entre eles o carboxissomo e metabolossomos. O carboxissomo foi o primeiro BMC descoberto através de micrografias eletrônicas, onde foi possível observar estruturas poliédricas em cianobactérias (Drews et al., 1956). Dentre os BMCs, o carboxissomo é o mais extensivamente caracterizado, e são encontrados em todas as cianobactérias e algumas bactérias quimioautotróficas (Shively, 1974; Kerfeld; Melnicki, 2016).

Os carboxissomos são BMCs proteicos poligonais de automontagem e desempenham um papel importante na fixação de carbono e utilização de fontes de carbono (MacCready; Vecchiarelli, 2021). O carboxissomo encapsula a enzima RuBisCo e a enzima anidrase carbônica para aumentar a taxa de fixação de carbono na célula e diminuir a taxa de fixação de oxigênio pela RuBisCo (Kinney et al., 2012) (Fig. 3). Ao encapsular a RuBisCo, o carboxissomo protege essa enzima contra a entrada de oxigênio no seu sítio ativo. O oxigênio compete com o carbono pelo mesmo sítio ativo da RuBisCo, o que leva a perda de carbono fixado (Ducat et al., 2012). Algumas bactérias desenvolveram esse mecanismo para aumentar a taxa de carboxilação da RuBisCo, o qual é uma importante fonte de energia para a célula (Hennacy et al., 2020).

Existem dois tipos de carboxissomo caracterizados até o momento: β-carboxissomo e α-carboxissomo (Badger et al., 2002). Ambos os compartimentos apresentam uma organização geral semelhante e operam de maneira semelhante (Ni et al., 2022). No β-carboxissomo a forma da RuBisCo é IB e γ-anidrase carbônica, enquanto no α-carboxissomo é encontrado a RuBisCo IA e β-anidrase carbônica (Pena et al. 2010; Zang et al., 2021). O carboxissomo é formado por um complexo de proteínas semipermeáveis que formam a "*shell*" e é composto por proteínas internas (Sutter et al., 2021). A concha (*shell*) do carboxissomo é composta pelas proteínas BMC-H (hexâmero), BMC-T (trímero) e BMC-P (pentâmero) (Sutter; Kerfeld, 2022) (Fig. 3). Juntas, essas proteínas formam os homo-oligômeros cíclicos típicos com poros que permitem o transporte de metabólitos através da concha (MacCready et al., 2021).

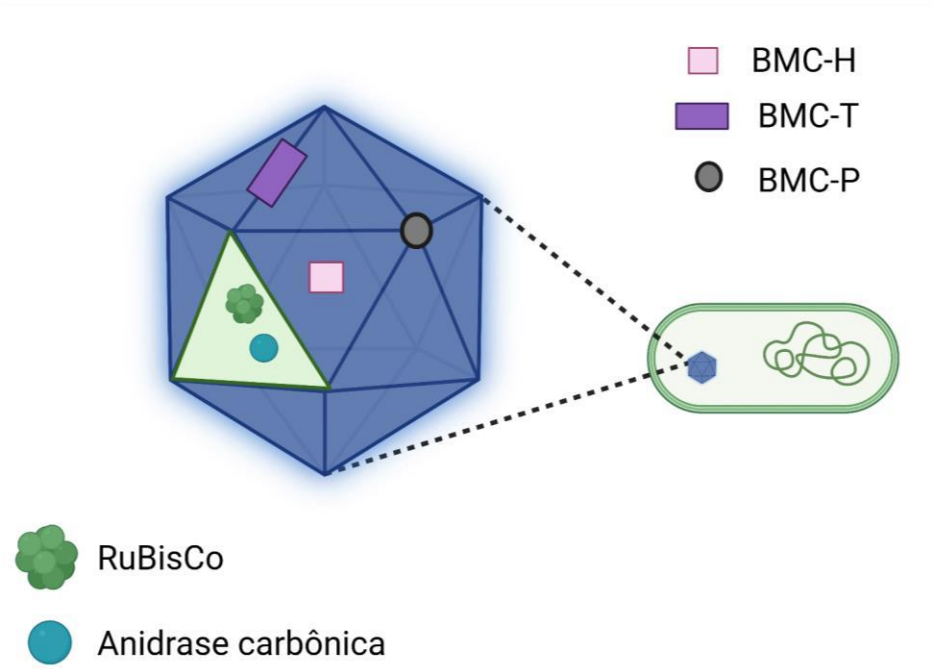



Figura 3: Imagem ilustrativa de um carboxissomo formado pelo complexo de proteínas comuns entre o β-carboxissomo e α-carboxissomo: BMC-H (quadrado rosa), BMC-T (retângulo roxo), BMC-P (círculo marrom) e as proteínas internas, RuBisCo e anidrase carbônica. Figura criada com a plataforma BioRender.com.

As proteínas do hexâmero BMC-H são formadas pelo domínio Pfam00936, no β-carboxissomo compreende as proteínas CcmK1/2/3/4/5/6; e no α-carboxissomo compreende as proteínas CsoS1A/B/C (Rae et al., 2013; Sommer et al., 2019). O trímero BMC-T consiste na fusão de dois domínios Pfam00936, no β-carboxissomo compreende as proteínas CcmP e CcmO; e no α-carboxissomo compreende a proteína csoS1D (Klein et al., 2009; Cai et al., 2013; Sun et al., 2021). Em relação às proteínas do pentâmero BMC-P, esse é formado por proteínas contidas no domínio Pfam03319, no β-carboxissomo compreende a proteína CcmL; e no α-carboxissomo constituem as proteínas CsoS4A/B (Tanaka et al., 2008; Zhao et al., 2019).

Internamente ao β-carboxissomo são encontradas as proteínas: CcmM, CcmN, CcaA, RbcL, and RbcX, enquanto no α-carboxissomo são encontradas as proteínas: CsoS2, CsoSCA, CbbL, CbbS, CbbQ e CbbO (Long et al., 2007; Kerfeld; Melnicki, 2016). As proteínas CcaA e CsoSCA são as anidrase carbônica dos carboxissomos β e α respectivamente (Long et al., 2007; Ludwig et al., 2000; Blikstad et al., 2021). No β-carboxissomo as proteínas RbcL e RbcS são a subunidade larga e pequena da RuBisCo respectivamente, enquanto RbcX são chaperones (Blikstad et al., 2019; Huang et al., 2019). No α-carboxissomo, a proteína CbbL forma a subunidade larga da RuBisCo, a CbbS forma a subunidade pequena da RuBiSCo e as proteínas CbbQ e CbbO são chaperones da RuBisCo (Roberts et al., 2012; Turmo et al., 2017). O α-carboxissomo é encontrado em alguns fototróficos anoxigênicos e α-cianobactérias, incluindo

α-cianobactérias do ecossistema marinho e em algumas bactérias quimioautotróficas, como em *Halothiobacillus neapolitanus* e *Escherichia coli* (Iancu et al., 2010; Turmo et al., 2017). Enquanto o β-carboxissomo somente foi descrito em β-cianobactérias que vivem em diferentes condições ambientais como em *Synechococcus elongatus* (PCC 7942) e *Synechocystis* sp. (PCC 6803) (Liu et al., 2022). Até o momento nenhuma das duas formas de carboxissomo foram encontrados em bactérias do filo *Actinomycetotas*.

No que se refere ao metabolossomo, esse pode estar envolvido no catabolismo do 1,2-propanodiol (Pdu metabolossomo) dentro do microcompartimento bacteriano proteico automontado (Mills et al., 2022) ou na catálise da etalonamina (Eut metabolossomo) para produção de amônia (fonte de nitrogênio), acetil coenzima A (acetil-CoA), e etanol (Pokhrel et al., 2021; Liu, 2021). Ao contrário das organelas presentes nos eucariontes, a membrana externa delimitadora de todos os BMCs é formada apenas por proteínas (BMC-H, BMC-T e BMC-P) (Melnicki et al., 2021). A diferença entre os BMCs, consiste no envolvimento em reações anabólicas (metabolossomo) e catabólicas (carboxissomo), e na enzima assinatura a qual é encapsulada para que a reação ocorra dentro do microcompartimento (Melnicki et al., 2021).

Novos BMCs identificáveis foram encontrados através de análise genômica do locu BMC, no entanto precisam ser caracterizados para encontrar as enzimas de assinatura (Sutter et al., 2021).

### *1.4.1. Caracterização de BMCs através da microscopia*

A microscopia vem desempenhando um grande papel na área da microbiologia desde o desenvolvimento inicial de microscópios simples, quando Van Leeuwenhoek observou pela primeira vez micróbios unicelulares por uma lente biconvexa com capacidade de aumentar a imagem cerca de 1000 vezes (Ford, 1989). Desde então, tem se buscado aprimorar as técnicas de contraste da microscopia para observar eventos celulares com cada vez mais resolução espacial e temporal (Lichtman; Conchello, 2005).

A microscopia de contraste de interferência diferencial (DIC), microscopia polarizada e a microscopia de fluorescência, proporcionaram o aumento do contraste inerente dos espécimes vivos para torná-los mais visíveis (Sanderson et al., 2014; Datta et al., 2020). No entanto, estas técnicas de microscopia estão sujeitas ao baixo contraste o que desencadeia uma baixa resolução das imagens, embora esteja-se buscando o aprimoramento da microscopia de fluorescência (Hell, 2003; Datta et al., 2020).

Alguns métodos de microscopia têm sido usados para estudos estruturais de carboxissomos. A microscopia eletrônica de transmissão (TEM) permitiu a visualização das interações entre proteína-proteína, seguida pelo encapsulamento para montagem do carboxissomo (Iancu et al., 2010; Chen et al., 2013). A microscopia de fluorescência viabilizou avaliar a funcionalidade do carboxissomo, incluindo a observação da dinâmica da fixação de $CO_2$, tempo de vida, e renovação das proteínas (Wang et al., 2019). Mas se tratando de BMCs, estas microscopias oferecem uma resolução limitada para observação do microcompartimento dentro da célula e proteínas isoladas (Evans et al., 2022). A alta resolução da microscopia é necessária para compreender a estrutura precisa dos conjuntos de proteínas para investigar a biossíntese e a função das proteínas (Nguyen; Ueno, 2018). Desde que a crio-microscopia eletrônica de transmissão (crio-ME) foi desenvolvido, tem sido utilizado para avaliar a estrutura, função e composição do carboxissomo (Schmid et al., 2006; Blikstad et al., 2021; Wang et al., 2022a). Crio-ME é descrito como uma técnica de imagem para realizar microscopia eletrônica em temperaturas criogênicas, adequada para a visualização de moléculas individuais, a qual possibilita a compreensão de suas interações dentro da célula (Gan; Jensen, 2012).

Dois subtipos de crio-ME são amplamente usados: tomografia crioeletrônica (crio-ET) e microscopia crioeletrônica de partícula única (SPM) (Chari; Stark, 2023; Hong et al., 2023). Com Crio-ET é possível visualizar células e seu interior em estado quase nativo, devido à ausência de fixadores químicos (Ting et al., 2007). A técnica de SPM tem sido empregada com sucesso para determinar estruturas de biomoléculas de alta resolução, e em alguns casos alcançando a resolução atômica (Nakane et al., 2020). A crio-ME SPM tem sido usada para gerar imagens em 3D de proteínas isoladas, ou seja, as proteínas devem ser extraídas e purificadas da célula para gerar a imagem, enquanto a crio-ET permite a visualização de estruturas macromoleculares em resolução em nanoescala, como complexos de proteínas, vírus e organelas (Nygaard et al., 2020; Turk; Baumeister, 2020).

Com o avanço das técnicas de crio-ME foi possível obter uma alta resolução das estruturas de BMCs, da organização das proteínas da concha e da dinâmica dessas proteínas para facilitar a entrada e saída do metabólito (Greber et al., 2019; Tan et al., 2021). No entanto, o uso de crio-ME também possui uma limitação em visualizar partículas no interior celular devido a fraca penetração de elétrons através de amostras biológicas, para resolver essa limitação, tem sido utilizado crio-ME CEMOVIS (microscopia crioeletrônica de seções vítreas) (Chlanda; Sachse, 2014; Liedtke et al., 2022). O método CEMOVIS, utiliza crio-ultramicrotomia e tem sido empregada para produzir partículas com espessuras de 40-100 μm

de material biológico vitrificado (congelado sem formar cristais de gelo) o qual é aplicado para crio-ME (Yusuf et al., 2020). Na técnica CEMOVIS a amostra vitrificada é posteriormente transferida para um crio-ultra-micrótomo e seccionada com uma faca de diamante, essa seção é fixada em uma grade de microscopia eletrônica (ME) e a grade é transferida para o microscópio crio-eletrônico para gerar as imagens (Al-Amoudi et al., 2004; Liedtke et al., 2022). Outra técnica de microscopia que vem sendo usada para observação de carboxissomo e as proteínas que o compõe é a microscopia eletrônica de coloração negativa (Chaijarasphong et al., 2016; Oltrogge et al., 2020). A microscopia eletrônica de coloração negativa permite a observação relativamente simples e rápida de macromoléculas e fornece um meio rápido de avaliar amostras para crio-ME (Scarff et al., 2018).

Os primeiros carboxissomos analisados por Crio-ET foram das bactérias *Halothiobacillus neapolitanus* e *Synechococcus* WH810219 (Schmid et al., 2006). Mais tarde, foi demonstrado por crio-ME que a formação de um invólucro de carboxissomo ocorre simultaneamente com a RuBisCo interna e provavelmente é guiada por interações específicas entre proteínas do invólucro e a RuBisCo (Iancu et al., 2010). A primeira estrutura intacta do α-carboxissomo de *Cyanobium* sp., uma cianobactéria marinha, foi proposta por crio-ME (SPM) (Evans et al., 2022). Imagens da partícula simples da RuBisCo no carboxissomo e no citosol da cianobactéria *Synechococcus* sp., também foi visualizada aplicando cryo-ET com contraste de fase de Zernike e anotação automatizada (Dai et al., 2018). Sabe-se que a dificuldade de visualizar uma enzima individual dentro da célula se deve em parte ao baixo contraste da imagem. No entanto, a técnica de crio-ME e crio-ET permite um contraste melhorado em comparação com a imagem convencional gerada por outros métodos de microscopia, o que possibilitou o estudo complexo de carboxissomos (Evans et al., 2022). Mais informações estão descritas no capítulo II desta tese.

O carboxissomo é um paradigma versátil em procariotos, servindo como um microcompartimento automontável composto por proteínas. Cumpre funções cruciais na fixação de carbono para todas as cianobactérias e algumas bactérias quimioautotróficas. Portanto, uma exploração abrangente nos avanços recentes da microscopia, juntamente com técnicas complementares, pode melhor contribuir para a compreensão da estrutura deste importante microcompartimento ou na descoberta de novos carboxissomos ou de outros BMCs.

## 2. JUSTIFICATIVA

Para a viabilização dos microrganismos no uso da biotecnologia e/ou suas interações

ecológicas é preciso entender o funcionamento da célula microbiana de maneira isolada e integrada com o ecossistema. Sabe-se que os principais desafios para predizer o comportamento, manipular e modificar microrganismos estão na compreensão de como funcionam esses organismos.

Nos últimos anos, grandes avanços foram feitos no entendimento do processo de fixação de $CO_2$ por bactérias nas áreas de fisiologia, bioquímica, biologia molecular e ômicas. A regulação e a interação entre as vias de fixação de carbono, em especial a via Calvin-Benson, bem como o envolvimento da organela especializada carboxissomo tem sido alvo de uma compreensão mais clara em vários nichos ecológicos ocupados por estes organismos.

O impacto do $CO_2$ na atmosfera está estritamente relacionado à energia do sistema e as interações gás-biota. As bactérias fixadoras de $CO_2$ além de representarem as forças motrizes para o estabelecimento de uma comunidade microbiana ativa, atuam junto com as plantas no "sequestro" do $CO_2$, ajudando a reduzir sua concentração na atmosfera, o que tem sido uma preocupação global por ser considerado nocivo em altas proporções. No cenário atual do planeta, tem se buscado soluções para reduzir o acelerado aumento de $CO_2$ na atmosfera; o qual é causado em sua maior parte por ações antropogênicas como desmatamento, processos industriais e queima de combustíveis fósseis, o qual ao liberar $CO_2$ para a atmosfera aumenta fenômeno de efeito estufa e consequentemente ao aquecimento global.

Atrelado a essa catástrofe ambiental, o rápido crescimento populacional e as mudanças climáticas provocam uma pressão adicional na busca de soluções alternativas sustentáveis para lidar com a emissão desregulada na atmosfera. Desta forma, novas abordagens exigirão a aplicação de soluções biológicas incluindo a manipulação e exploração dos microrganismos. Nesse sentido, o estudo com bactérias termofilias tem mostrado o alto potencial desses microrganismos em ser fonte de novos metabolismos e moléculas, exatamente por desenvolverem estratégias adaptativas alternativas para sobreviverem em ambientes com condições consideradas desfavoráveis. A capacidade adaptativa dos *Actinomycetotas* termofílicos conferem uma vantagem sobre os outros microrganismos; além de sobreviverem em condições extremas, possuem fisiologia e flexibilidade metabólica que ajudam a desencadear a produção de compostos bioativos e enzimas industrialmente valiosas para aplicação.

Diante desse contexto, *Carbonactinospora thermoautotrophica* estirpe StC está no grupo das bactérias termofílicas que possuem versatilidade metabólica, possui duas vias diferentes para fixar carbono atmosférico, o que ainda não foi demostrado em outra bactéria

isolada. Além de possuir um mecanismo de concentração de $CO_2$ o qual confere maior taxa de carboxilação. Dessa forma o presente estudo fornece uma visão holística do metabolismo do carbono e uma melhor compreensão do comportamento celular em *Carbonactinospora thermoautotrophica* estirpe StC.

## 3. OBJETIVOS

### 3.1. Objetivo Geral

Identificar as vias de fixação de carbono que a *Actinomycetota* termofílica *Carbonactinospora thermoautotrophica* StC utiliza para fixar carbono atmosférico e desenvolver um protocolo para isolar e caracterizar o microcompartimento que concentra carbono nos filamentos da StC.

### 3.2. Objetivos específicos

- Realizar o sequenciamento do genoma de *Carbonactinospora thermoautotrophica* estirpe StC através da tecnologia Illumina HiSeq e Oxford Nanopore Technologies, e análise dos dados gerados por ferramentas de bioinformática;
- Conduzir um experimento biológico com a estirpe StC cultivada em condições autotrófica e heterotrófica para extração de RNA e proteína;
- Realizar análise integrada do genoma, transcriptoma e proteoma da estirpe StC cultivada em condições autotróficas e heterotróficas através da tecnologia Oxford Nanopore, Illumina HiSeq, LC- MS/MS;
- Identificar os genes e proteínas expressos relacionados às vias naturais de fixação de carbono e microcompartimento que concentra carbono na célula;
- Desenvolver um protocolo para isolar e caracterizar o microcompartimento bacteriano que concentra carbono na estirpe StC; realizar um gradiente de glicose, western-blot, LC- MS/MS e microscopia.

## 4. MATERIAL E MÉTODOS

### 4.1. Área de estudo, amostragem e isolamento da *C. thermoautotriphica* estirpe StC

A estirpe StC foi previamente isolada pelo nosso grupo de pesquisa a partir de um consórcio microbiano (CNF) (Pinheiro et al., 2023) composto por quatro bactérias (*Carbonactinospora thermoautotrofica* StC, *Sphaerobacter thermophilus*, *Chelatococcus* sp. e *Geobacillus* sp.) encontrado em uma amostra de clareira composta por resíduos orgânicos, localizado no município de Seropédica, Rio de Janeiro, Brasil (coordenadas: -22.775892, -43.691673) (Fig. 4). Para mais detalhes sobre cultivo do CNF, checar Pinheiro et al. (2023).

Figura 4: Localização do ponto de coleta do consórcio microbiano. Área situada no município de Seropédica, estado do Rio de Janeiro, Brasil (coordenadas: -22.775892, -43.691673). Figura adaptada: Google Maps e Google Earth.

Para o isolamento da estirpe StC a partir do consórcio microbiano foi utilizado o meio de cultura sólido Reasoner 2 (R2A) (Kasvi) suplementado com antibiótico amoxicilina (1 μg/ml), e incubado à 53 °C por um dia. Esse cultivo foi então transferido para o meio mineral N-FIX (Gadkari et al., 1992) solidificado com 1,5% de ágar nobre de alta pureza (Sigma-Aldrich), e suplementado com 0,03% de clinoptilolita (uma zeólita eliminadora de amônia) (Liao et al., 2015; Dahal et al., 2017), a fim de obter a estirpe pura, pois esse meio tem condições fortemente restritivas e sem vestígios de amônia contaminante. A estirpe foi armazenada em nossa coleção microbiana no Laboratório de Ecologia Microbiana Molecular, da Universidade Federal do Rio de Janeiro, sob supervisão do Professor Alexandre Soares Rosado.

### 4.2. Extração de DNA da *C. thermoautotrophica* StC

Para a extração de DNA genômico, a estirpe StC foi previamente cultivada por 4 dias a

60 °C em frascos do “tipo” penicilina (100 mL) contendo 40 mL de meio de cultura N-FIX, suplementado com elementos traços (Apêndice A), solidificado com 30 g/L de Ágar Nobre (Sigma-Aldrich), sem adição de fontes orgânica de carbono e nitrogênio (Gadkari et al., 1990; Meyer, 1982; Meyer; Schegel, 1983). Os frascos foram fechados com tampa de borracha adequada e com lacre de alumínio. Com auxílio de uma seringa e agulha, 20 mL de ar foram aspirados dos frascos já contido de inóculo bacteriano, e foi adicionado os gases: 9 mL de CO, 10 mL N/O (proporção: 80% $N_2$ e 20% $O_2$), e 1 mL de $CO_2$ (Gadkari et al., 1990).

A extração total de DNA da estirpe StC foi realizada com o kit comercial QIAGEN® Genomic DNA com algumas adaptações. Anterior às etapas do kit comercial, o extrato bacteriano contendo 480 µL de EDTA 0.5 M e 120 µL de lisozima (100 mg/mL) foi incubado em “banho maria” a 37 °C durante 30 minutos, seguido por lise celular com perolas de vidro no equipamento FastPrep (MP Biomedicals) programado para *Actinomycetotas* (três ciclos de 40 segundos com intervalo de 2 minutos em cada ciclo).

O experimento foi realizado com três repetições biológicas e três repetições técnicas no Microbial Ecogenomics & Biotechnology Laboratory, na universidade King Abdullah University of Science and Technology (KAUST), localizado na Arábia Saudita. O DNA total da estirpe StC foi quantificado por espectrofotometria, através do NanoDrop One/One$^c$ Microvolume UV-Vis Spectrophotometers (Thermo Fisher Scientific), a partir da leitura da absorbância de 1 µL do DNA total, em um comprimento de onda de 260 nm (Apêndice B). O grau de pureza das amostras foi avaliado pela relação entre as leituras da absorbância a 260 e 280 nm, de forma que leituras entre 1.8 indicam alto grau de pureza (Desjardins; Conklin, 2010). Posteriormente, o DNA total foi purificado com o reagente à base de esferas AMPure XP (Beckman Coulter Life Sciences), quantificado com Qubit® dsDNA HS Assay Kit (Thermo Fisher Scientific) e examinado por eletroforese em gel de agarose a 1,5% (Apêndice C; D).

### *4.2.1. Sequenciamento, montagem e anotação do genoma*

O genoma da estirpe StC foi sequenciado por Oxford Nanopore Technologies (ONT) no Bioscience Core Laboratory (KAUST) e Illumina HiSeq. A montagem do genoma da *C. thermoautotrofica* foi realizado usando uma abordagem híbrida combinando *reads* curtas e longas, com alta profundidade e alta cobertura, respectivamente. *Reads* genômicas curtas (pareadas - 300 pb) foram produzidas usando o sequenciador Illumina HiSeq, e leituras longas usando Oxford Nanopore. As *reads* proveniente do sequenciamento por *Illumina* foram inicialmente submetidas a uma etapa de limpeza, para remoção de bases e adaptadores de baixa

qualidade, utilizando Trimmomatic Ver. 0,39 (Imm; Chang, 2023) com os seguintes parâmetros: tratado como paired-end (PE), ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:keepBothReads para remover adaptadores e manter ambas as *reads* emparelhadas, um valor LEADING e TRAILING de 20 para remover bases com qualidade acima de 20, e todas as leituras com comprimento inferior a 36 pares de bases também foram removidas. As *reads* proveniente do sequenciamento por Oxford Nanopore foram limpas inicialmente usando o basecaller Guppy Ver. 3.0.3+7e7b7d0 (Wick; Holt, 2019) para verificar e remover possíveis adaptadores, e bases de baixa qualidade foram removidas usando Porechop Ver. 0.2.4 (Wick et al., 2017), ambos com parâmetros padrão. A montagem do genoma foi realizada com *reads* limpas de ambos os tipos de leitura usando MaSuRCA Ver. 4.0.3 (Zimin et al., 2013). Para a escolha das melhores referências, os contigs gerados para a estirpe StC foram submetidos à ferramenta básica de busca de alinhamento local para nucleotídeos (BLASTn) (Altschul et al., 1990). Os *contigs* montados foram orientados para gerar um *scaffold* utilizando Medusa v. 1.3 (Bosi et al., 2015), e as estirpes *C. thermoautotrofica* UBT1 e H1 foram utilizadas como referência. As lacunas resultantes das montagens foram preenchidas manualmente usando o software CLC Genomics Workbench, para visualização, v. 7.0 (Qiagen, EUA), onde as leituras foram mapeadas contra um genoma de referência para gerar uma sequência de consenso, que foi então usada para fechar as lacunas.

O genoma da StC foi anotado automaticamente usando PROKKA versão 3 (Seemann, 2014). Os dados da sequência do genoma StC foram carregados no Type (Strain) Genome Server (TYGS) (//tygs.dsmz.de/user_requests/new) para análise taxonômica completa baseada no genoma (Meier-Kolthoff; Göker, 2019). Informações sobre nomenclatura taxonômica associada foram fornecidas pelo banco de dados LPSN (Lista de nomes procarióticos com posição na nomenclatura, disponível em https://lpsn.dsmz.de) (Meier-Kolthoff et al., 2022). A árvore filogenética foi gerada usando a ferramenta Phylogenomic Tree Tool no Pathosystems Resource Integration Center (PATRIC) (//www.patricbrc.org), versão 3.5.17 (Wattam et al., 2014). Todas as árvores foram visualizadas usando iTOL v4.261 (Letunic e Bork, 2016). A similaridade entre os genomas das estirpes StC, UBT1 e H1 foi realizada usando o Pathosystems Resource Integration Center (PATRIC). A anotação funcional das sequências do genoma foi realizada com KEGG usando BLASTKOALA (Kanehisa; Morishima, 2016). A reconstrução das vias metabólicas encontradas no genoma da StC foi realizada usando a ferramenta KEGG pathway e GhostKOALA (Kanehisa; Morishima, 2016; Aramaki et al., 2020).

### 4.2.2. Análise de transcriptômica: preparação das amostras e sequenciamento

O experimento de transcriptômica e proteômica foi realizado a fim de avaliar a expressão de genes e proteínas relacionados com as vias de fixação de carbono sob duas diferentes formas de cultivo. Para tanto, a estirpe StC foi cultivada autotroficamente e heterotroficamente com quatro réplicas biológicas e quatro réplicas técnicas: I) N-FIX - meio de cultura N-FIX sólido (descrito no tópico 4.2.), 40 mL por frasco, sem fontes orgânicas de carbono e nitrogênio, com adição dos gases $CO/CO_2/N_2/O_2$ e II) R2A - meio de cultura líquido R2A solidificado com 30g/L de Ágar Nobre (Sigma-Aldrich), 40 mL por placa (Fig. 5). Para cada réplica biológica havia cinco sub-réplicas totalizando 20 frascos para o tratamento N-FIX e 20 placas para o tratamento R2A, com a finalidade de enriquecimento de biomassa celular. O experimento foi conduzido em incubadora a 60 °C durante 4 dias, até atingir uma quantidade considerável de biomassa (aproximadamente 23 mg) para posterior extração de RNA e proteína.

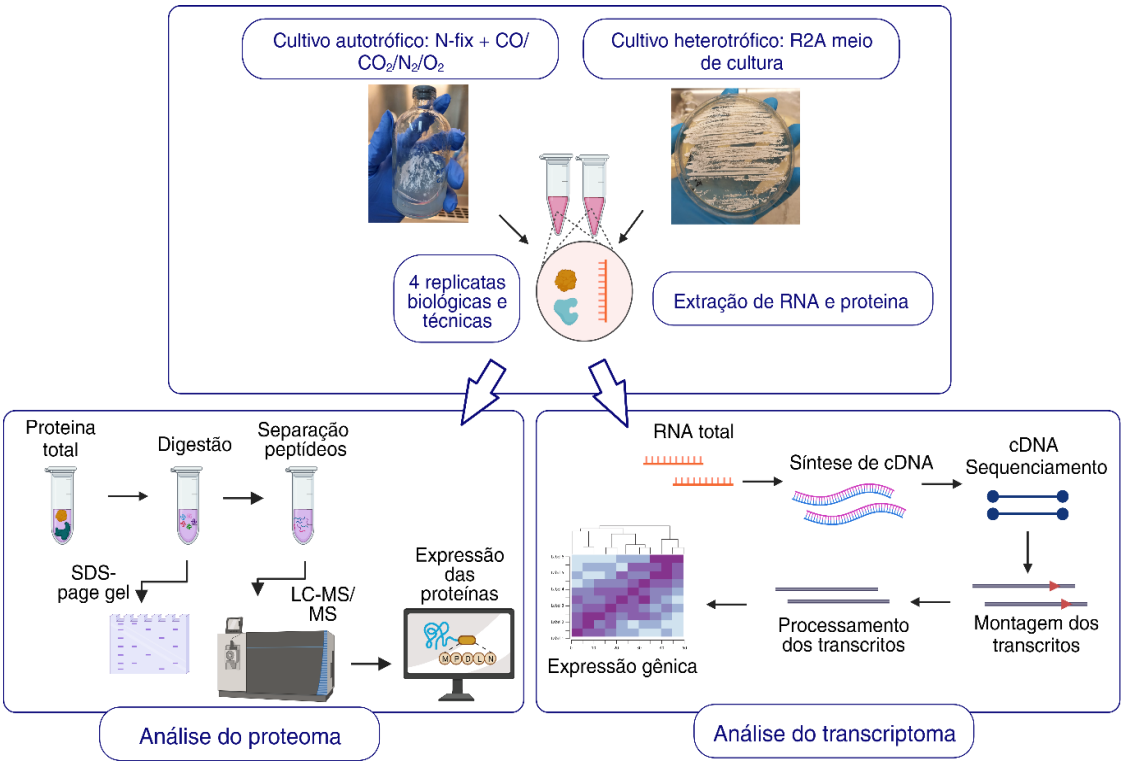



Figura 5: Visão geral do experimento para análise do transcriptoma e proteoma da estirpe StC. A estirpe StC foi cultivada em duas diferentes condições de cultivo: autotrófico e heterotrófico. Foi realizada a extração de RNA e proteína total da mesma amostra, para posterior sequenciamento do RNA e proteínas. Figura criada com a plataforma BioRender.com.

Para extração de RNA e proteína total foi utilizado o kit NucleoSpin TriPrep DNA/RNA/proteína Purification (Macherey-Nagel), o qual realiza a extração de RNA e proteína da mesma amostra. Antes de iniciar as etapas do kit foi realizada algumas adaptações

para melhorar a lise celular. A biomassa bacteriana contendo 480 µL de EDTA 0.5 M e 120 µL de lisozima (10 mg/mL) foi incubado em "banho maria" a 37 °C durante 30 minutos, seguido por lise celular com perolas de vidro no equipamento FastPrep (MP Biomedicals) programado para *Actinomycetotas* (três ciclos de 40 segundos com intervalo de 2 minutos em cada ciclo).

O RNA total foi tratado com o kit de tratamento de DNAse (Ambion's Dnase Treatment Kit, Thermo Fisher Scientific), a fim de remover possíveis resquícios de DNA. O RNA purificado foi quantificado por espectrofotometria, através do NanoDrop 2000 (Thermo Fisher Scientific) e Qubit® dsDNA HS Assay Kit (Thermo Fisher Scientific) e a qualidade do RNA foi examinado por eletroforese em gel de agarose a 1,5% livre de RNAse (Apêndice E; F; G). Para a preparação do gel de agarose de RNA, todo o suporte o que inclui a cuba e os pentes foram tratados durante 1 hora com SDS (dodecil sulfato de sódio) 0,1% preparado com DEPC (dietil pirocarbonato). O gel de agarose foi preparado com tampão TAE (Tris-Acetato-EDTA) 1x com DEPC, e o mesmo tampão foi utilizado para a corrida do gel. As amostras de RNA total purificada foram sequenciadas por Illumina HiSeq.

Para análise do transcriptoma, aproximadamente 100 µg/µL do RNA total purificado da estirpe StC foi usado para preparação da biblioteca e sequenciamento pela plataforma Illumina HiSeq. As quatro réplicas foram processadas independentemente. Para o controle de qualidade e mapeamento das *reads*, os arquivos fastQ library foram primeiro submetidos ao software FastQC v. 0.1.2 Brandine et al., 2019), seguido por trimagem usando o programa Trimmomatic v. 0.39 (Imm; Chang, Y., 2023). Posteriormente, as bibliotecas foram submetidas ao mapeamento contra o genoma da *C. thermoautotrophica* StC (obtido conforme descrito no tópico 4.2.1) através do programa Bowtie v. 2.3.4.3 (Li et al., 2023). Os arquivos de texto (SAM) foram convertidos para o formato binário (BAM), seguido pela exclusão da qualidade das *reads* mapeadas (MapQ) com valor dois, as quais representam *reads* que foram mapeadas em mais de três locais no alvo. Por último, foi realizada a etapa de classificação dos arquivos mapeados limpos usando Samtools v.1.9 (Li et al., 2019). O programa HTSeq v.2.0.3 (Siddique et al., 2021) foi utilizado para realizar a etapa de contagem das *reads* mapeadas sobre o genoma, e os arquivos de contagem foram normalizados para TPM (Transcritos por milhão). Os genes expressos relacionados com as vias de fixação de carbono foram preditos com BlastKOALA (KEGG) e os gráficos Heatmaps foram gerados com o programa Orange Data Mining versão 3.36.2 (Bar-Joseph et al., 2001).

### 4.4. Análise de proteômica: preparação das amostras e análises em LC-MS

Para a análise de proteômica, a proteína total extraída das mesmas amostras usadas para a extração de RNA, foram quantificadas pelo kit comercial Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific). Em quatro réplicas, um total de 12,5 µg/ml de solução de proteína foi submetido a digestão usando o banho seco com agitação a 37 °C e 900 rpm overnight. As amostras de proteínas foram então preparadas usando o protocolo Filter Aided Sample Preparation do laboratório Bioscience Core Laboratory (KAUST), baseado no método descrito por Wiśniewski et al. (2009). As proteínas foram reduzidas com DTT 10 mM e alquiladas com iodoacetamida 40 mM. A digestão foi realizada com a enzima tripsina, onde as amostras de proteínas foram diluídas com 0,02 µg/µL tripsina e incubadas overnight a 37 °C e à 900 rpm.

Os peptídeos foram purificados antes de injetar no espectrofotômetro de massas seguindo o protocolo do Bioscience Core Laboratory (KAUST). Os peptídeos foram dessalinizados usando uma ponteira do tipo ZipTip® com resina C18 (Merck Millipore). A mistura de peptídeos foi medida em um espectrômetro de massa Orbitrap Fusion Lumos (Thermo Fisher Scientific) acoplado ao UltiMate™ 3000 UHPLC (Thermo Fisher Scientific). Os peptídeos foram separados usando uma coluna Acclaim PepMapTM C18 (75 um ID X 50 cm, tamanhos de partícula de 2 µm, tamanhos de poro de 100 Å) com uma taxa de fluxo de 300 nl/minuto. Um gradiente de 100 minutos foi estabelecido usando fase móvel A (0,1% FA) e fase móvel B (0,1% FA em 99,9% ACN): 4-35% B por 60 minutos, 15 minutos aumentando para 90% B, 90% B por 5 minutos e 2% de B por 10 minutos de condicionamento de coluna.

Posteriormente, as amostras foram introduzidas no espectrômetro de massas Orbitrap Fusion Lumos através de um Nanospray Flex (Thermo Fisher Scientific) com potencial de eletrospray de 2,5 kV. A temperatura do tubo de transferência de íons foi ajustada em 160 °C. O Q-Exactive HF foi configurado para realizar aquisição de dados no modo Aquisição Dependente de Dados (DDa). Uma varredura MS completa (faixa de 200-2000 m/z) foi adquirida no Orbitrap com uma resolução de 60.000 (a 200 m/z) em modo de perfil, um tempo máximo de acumulação de íons de 50 minutos. A triagem do estado de carga para íon precursor foi ativada. Os vinte íons mais intensos acima do limite 1x106 e carregando múltiplas cargas foram selecionados para fragmentação usando dissociação de colisão de energia mais alta (HCD). A resolução foi definida como 15.000. A exclusão dinâmica para fragmentação HCD foi de 30 segundos. Outra configuração para íons fragmentados incluiu um tempo máximo de acumulação de íons de 60 minutos, um valor alvo de 2,50×103, uma energia de colisão normalizada em 30% e largura de isolamento de 1,6.

Os arquivos de MS provenientes do Orbitrap Fusion Lumos foram analisados usando o

programa Thermo Proteome Discoverer (versão 2.5.400) e pesquisados contra a anotação de proteínas da *C. thermoautotrophica* StC. Foi realizada uma busca com peptídeos totais e foram consideradas duas clivagens perdidas. A carbamidometilação (C) foi uma modificação fixa, enquanto a oxidação e acetilação da metionina (proteína N-terminal) foram selecionadas como modificações variáveis. A tolerância de massa para íons precursores foi de 10 ppm e 0,05 DDa para os íons fragmentados. A Taxa de Descoberta Falsa (FDR) foi fixada para proteínas e peptídeos com uma pontuação de corte menor que 1% nos níveis de proteína, peptídeo e PSM (correspondências do espectro peptídico). As proteínas foram agrupadas em proteínas mestres utilizando o princípio da máxima parcimônia. A identificação das proteínas foi realizada através do software Proteome Discoverer v. 2.5.400 e as proteínas expressas pertencentes as vias de fixação de carbono foram preditas com BlastKOALA (KEGG) (Kanehisa; Morishima, 2016).

### 4.5. Análise genômica, extração, purificação e visualização de BMC

Para extração e purificação do BMC presente na estirpe StC, foi desenvolvido um protocolo exclusivo por se tratar de uma bactéria termofilia com desafios no manuseio, além de pertencer a um filo bacteriano onde até o momento não foi descrito a existência de carboxissomo. Optamos por cultivar a estirpe StC em meio de cultura heterotrófico para extrair o microcompartimento pois essa extração requer grande quantidade de biomassa para conseguir extrair uma quantidade razoável de BMC, conforme previsto em estudos anteriores (Chaijarasphong et al., 2015; Sun et al., 2022). Em meio de cultura autotrófico as bactérias em geral produzem baixa biomassa, devido as condições sem fonte orgânica de carbono e nitrogênio. Pelo fato de o BMC não ser produto do metabolismo secundário, assumimos que a presença na célula bacteriana não depende das condições do meio.

A metodologia foi desenvolvida com base em protocolos anteriores descritos na literatura (Sommer et al., 2019; Flamholz et al., 2020; Oltrogge et al., 2020). As etapas de extração do microcompartimento e Western blot foram realizadas no Microbial Ecogenomics & Biotechnology Laboratory (KAUST). A análise por massa espectrometria foi realizada no Bioscience Core Laboratory (KAUST). As análises de microscopia foram realizadas no laboratório de Imaging and Characterization Core Lab (KAUST). O desenho experimental utilizado pode ser observado na Fig. 6, e está descrito nos tópicos a seguir.

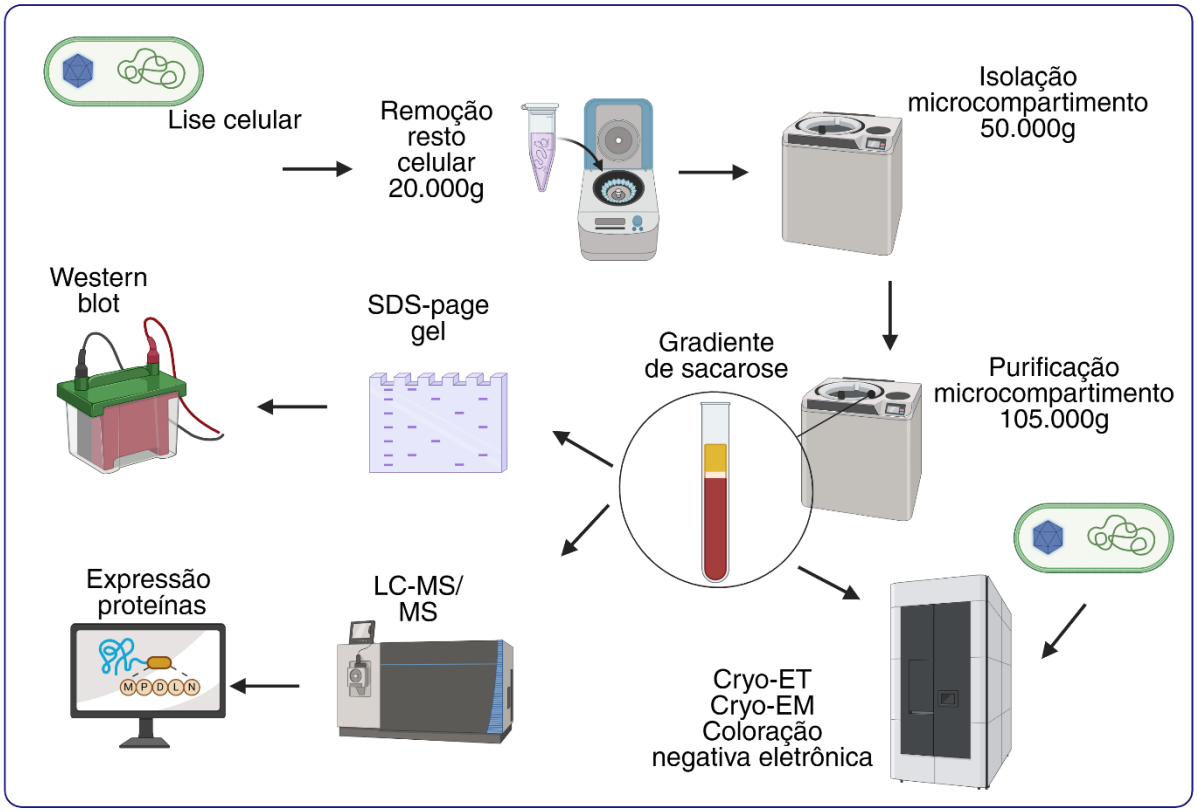



Figura 6: Visão geral do experimento para isolamento e caracterização do microcompartimento presente na estirpe StC. A estirpe StC foi cultivada em meio de cultura R2A: I) o isolamento do microcompartimento foi realizado com ultracentrifugação; II) purificação do microcompartimento com gradiente de sacarose; III) detecção do sinal da proteína Ccmk/CsoS1 por western blot; IV) identificação das proteínas presentes na amostra purificada por LC- MS/MS; V) Visualização do microcompartimento na amostra e dentro da célula bacteriana por microscopia. Figura criada com a plataforma BioRender.com.

### *4.5.1. Busca por genes que codificam para microcompartimentos bacterianos*

Inicialmente foi realizada uma busca manual de genes que codificam BMCs no genoma da estirpe StC. Quando possíveis genes e suas respectivas proteínas foram identificadas, sua sequência proteica prevista foi então usada para formular uma pesquisa BlastP (//www.ncbi.nlm.nih.gov), banco de dados não redundante do NCBI. A função da proteína, classificação da família e domínio previsto foram fornecidos pela InterPro (McDowall; Hunter, 2011). Para a organização dos agrupamentos de genes do carboxissomo e para prever o operon e as proteínas homólogas utilizamos o banco de dados MicrobesOnline (MicrobesOnline - A website for browsing and comparing microbial genomes) (Alm et al., 2005). A sequência das proteínas da StC foi comparada com perfis HMM (Hidden Markov Model) usando BMC Caller (BMC Caller (msu.edu)) (Sutter; Kerfeld, 2022). A sequência da proteína BMC-P foi alinhada

com outros 768 genes obtidos do PFAM PF03319 usando MUSCLE e a árvore filogenética foi construída com Phyml usando o modelo JTT (Sengupta; Azad, 2021). Também foi realizada uma busca manual das proteínas que compõem os BMCs nos dados de proteômica, para verificar se foram expressas.

### *4.5.2. Protocolo para extração de BMC da estirpe StC*

Para a extração do microcompartimento da estirpe StC, a bactéria foi previamente cultivada em meio de cultura líquido R2A solidificado com 30 g/L de Ágar Nobre (Sigma-Aldrich) durante 3 dias a 60 °C, com um total de 40 placas de petri utilizadas. Após 3 dias de cultivo, as colônias foram transferidas com auxílio de uma alça microbiológica de inoculação (10 µL) estéril, para microtubos de 2 mL estéreis, e a massa celular foi pesada, totalizando 325 mg de biomassa bacteriana. A massa celular foi transferida para microtubos de 2 mL estéreis contendo 480 µL de solução TEMB [5 mM Tris-HCl (pH 8.0), 1 mM EDTA, 10 mM $MgCl_2$, 20 mM $NaHCO_3$], 120 µL de lisozima (100 mg/mL), 2 mL de B-PER™ (Reagente de extração de proteínas bacterianas, Thermo Fisher Scientific) e 50 µL de inibidor de protease. Esse extrato celular foi incubado em banho maria a 37 °C por 1:30 horas, seguido por 1:30 horas em um sonicador com a programação pulsante e temperatura ajustada até 37 °C. Após essas duas etapas de lise celular inicial, o extrato celular foi homogeneizado no equipamento FastPrep (MP Biomedicals) programado para *Actinomycetota*.

Para a etapa após a lise celular, a amostra foi centrifugada a 4.000 x g por 10 minutos para sedimentar restos celulares, seguido de centrifugação a 20.000 x g por 20 minutos para clarificar a amostra. A etapa posterior seguiu com uma alta centrifugação a fim de formar o pellet de carboxissomo: a amostra clarificada foi centrifugada a 50.000 x g por 1 hora, quando se formou um pellet branco. O pellet foi ressuspenso em 1 mL de solução TEMB e quantificado por espectrofotometria NanoDrop One/One$^c$ Microvolume UV-Vis Spectrophotometers (Thermo Fisher Scientific) na função proteína>outras proteínas E (coeficiente de extinção) e MW (peso molecular) (Apêndice H). O coeficiente de extinção e peso molecular das proteínas da concha do carboxissomo foi obtido na ferramenta online na ProtParam Expasy (Expasy - ProtParam tool). Também foi realizada a quantificação das proteínas pelo kit comercial Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific) (Apêndice I). A amostra foi aplicada em SDS-PAGE gel para confirmar a presença e integridade das proteínas.

O pellet ressuspenso na solução TEMP foi concentrado usando o filtro Amicon® 10

kDa (Merck Millipore), e então foi aplicado em 15 mL de gradiente de sacarose preparado com a solução TEMP nas concentrações: 10%, 20%, 30%, 40%, 50% de sacarose. O gradiente de sacarose foi centrifugado a 105.000 x g por 30 minutos, e houve a formação de um pellet com aspecto leitoso entre as concentrações de 10% e 20%. Frações de 0,5 mL do gradiente foram coletadas e quantificadas novamente por espectrometria e pelo kit BCA Protein Assay. Essas frações de 0,5 mL também foram analisadas via SDS-PAGE para identificar a presença de proteínas que compõem o carboxissomo. As frações foram armazenadas a -30 °C e a fração que melhor correspondia ao BMC foi analisada por Western blot, espectrometria de massas e microscopia.

*4.5.3. Análise de BMC por Western blot*

A fração de proteínas correspondente ao BMC foi mixada com SDS loading buffer 2x (Bio-Rad) (30 µL de amostra e 10 µL de SDS loading) e aquecido a 95 °C em termociclador (Benchmark H5000-HC MultiTherm™) por 5 minutos. A amostra aquecida foi aplicada em 4-20% Tris-glicina Criterion™ TGX Stain free gel (Bio-Rad) e submetida a eletroforese a 120 V por 1 hora com tampão de corrida 1x Tris/glicina/SDS Running Buffer (Bio-Rad).

Para o Western blot, o SDS-PAGE gel da etapa anterior contido de proteínas foi mergulhado em tampão para western blot (b.w.b) (10% Tris/glicina e 20% de metanol) para retirar o excesso de SDS. Na cuba para *blotting*, o "sanduíche" foi montado da seguinte forma: esponja e filtro (previamente embebidos em b.w.b), gel de SDS-PAGE contido de proteínas, membrana de nitrocelulose (previamente embebida em metanol), esponja e filtro. A montagem do "sanduíche" foi realizada dentro de um suporte contendo tampão b.w.b. Proteínas separadas por SDS-PAGE foram transferidas para a membrana de nitrocelulose (Bio-Rad) por eletroforese 120 V por 80 minutos com tampão de corrida 1x Tris/glicina (Bio-Rad), e um suporte com gelo foi adicionado dentro da cuba para manter a temperatura baixa. Para a etapa de bloqueio para evitar a ligação não específica do anticorpo à membrana do *blotting*, a membrana de nitrocelulose contendo as proteínas foi lavada em TBST 1x posteriormente foi incubada a 4 °C overnight em 5% de BSA blocking buffer (BSA + TBST). Após essa etapa, a membrana foi lavada novamente com TBST 1X, por 3 vezes por 5 minutos cada lavagem, para retirar o excesso de BSA.

Com o propósito de detecção de proteína presente na concha do carboxissomo, foi confeccionado um anticorpo específico para a sequência de proteína CcmK/CsoS1 encontrada no genoma da estirpe StC, sendo está:

MPDLNGKALGLVETLGLVAATEAADAMVKAANVRLVTKQQVGGGL
ITVIVAGDVGAVKAAVDAGQTAGSAVGKVVSAHVIPRPHDDIPGILERPPVR.

Após a etapa descrita anteriormente, a membrana foi sondada com o anticorpo primário anti-CcmK/CsoS1 (GenScript) (soro derivado de coelho) seguindo as recomendações do fabricante, e agitado overnight a 4 °C (em sala refrigerada). Posteriormente, a membrana foi lavada 3 vezes por 5 minutos com TBST 1X, seguido por adição do anticorpo secundário igG anti-habbit (Agrisera), e agitado por uma 1 hora a 4 °C (em sala refrigerada). Por fim, a membrana foi lavada conforme descrito anteriormente. A membrana foi incubada por 1 minuto em ECL prime Western blotting detection (Cytiva Amersham) e o sinal da proteína foi detectado com o transiluminador (ThermoFisher Scientific™).

*4.5.4. Análise por espectrometria de massa (LC/MS) do BMC*

A amostra contendo o BMC foi analisada por LC-MS em Orbitrap Fusion Lumos (Thermo Fisher Scientific) acoplado a um UltiMate™ 3000 UHPLC (Thermo Fisher Scientific), a fim de detectar as proteínas da concha e do interior do microcompartimento. A análise foi procedida seguindo o protocolo do Bioscience Core Laboratory (KAUST) conforme descrito no item 4.4. com adaptação. Anterior a etapa da digestão das proteínas, a amostra contendo o microcompartimento foi lisada com tampão de lise SDT (10% w/v SDS, 100 mM Tris/HCl pH 7.6, 100 mM DTT), a fim de romper a concha do BMC. Para a lise, em 40 µL de amostra (0,22 mg/mL de proteínas) foi adicionado 40 µL de SDT, aquecido a 95 °C por 10 minutos em termociclador (Thermo Fisher Scientific), e o lisado foi centrifugado a 16.000 g por 5 minutos. Após essa etapa, procedeu-se o protocolo para processamento de digestão da proteína, protocolo de dessalinização e a mistura de peptídeos foi medida em um espectrômetro de massa Orbitrap Fusion Lumos (Thermo Fisher Scientific) acoplado a um UltiMate™ 3000 UHPLC (Thermo Fisher Scientific). Os arquivos de MS provenientes do Orbitrap Fusion Lumos foram analisados usando o Thermo Proteome Discoverer (versão 2.5.400) e pesquisados contra a anotação de proteínas da *C. thermoautotrophica* StC.

*4.5.5. Aquisição de imagens por microscopia: Microscopia óptica, microscopia eletrônica de coloração negativa, crio-ET, e crio-ME*

Microscopia óptica: para visualização dos filamentos e esporos da estirpe StC, foi

utilizado a técnica de microcultivo em placa de petri, adaptado do protocolo descrito por Su et al. (2012). Em uma placa de petri contendo 40 mL de meio de cultura R2A líquido e solidificado com ágar nobre de alta pureza (Sigma-Aldrich), foi adicionado uma lamínula de vidro previamente esterilizada com metanol, e na parte superior da lamínula foi inoculado a estirpe StC. Após dois dias de cultivo a 60 °C, a lamínula foi retirada cuidadosamente da placa de petri e adicionada com o cultivo celular voltado para cima em uma lâmina de vidro (previamente esterilizada com metanol), e as bordas da lamínula foram seladas com fita adesiva. O microcultivo foi visualizado em microscópio invertido com iluminação de LED (Leica DM IL LED) acoplado a uma câmera com sensor CCD para microscópios (Leica DFC3000 G), pelo qual foi obtido as imagens da morfologia da estirpe StC.

Microscopia eletrônica de coloração negativa: para o procedimento, 7 µL de amostra contendo BMC foram colocadas em grades de cobre descarregadas (15 mA, 30 segundos) (Cu, 300 mesh, filme de carbono, EMS) por 1 minuto, depois enxugadas com papel filtro (WhatmanTM). Em seguida, aproximadamente 5 µl de acetato de uranila a 2% foi adicionado à grade e imediatamente *blotted* (marcado). Em seguida, ~5 µl de acetato de uranila a 2% foi adicionado novamente e deixado para corar por 1 minuto. As grades foram então enxugadas para remover o excesso de líquido e secas ao ar. Amostras coradas negativamente foram examinadas em um microscópio eletrônico Tecnai T12 (Thermo Fisher Scientific) operado a uma voltagem de aceleração de 120 kV. As imagens foram obtidas em diferentes ampliações com desfocagem variando de -0,8 µm a -3,0 µm com uma câmera com sensor CCD US 4000 (Gatan).

Crio-ET: as amostras contendo BMC foram preparadas na grade criogênica e congeladas utilizando o Vitrobot Mark IV (Thermo Fisher Scientific) a 4 °C com 100% de umidade. As amostras foram preparadas aplicando 2 a 3 µL de amostra de proteína em grades de ouro com descarga luminosa (30 mA por 60 segundos) (QuantiFoil Au, malha 200, grades R2/2, EMS) e *blotting* por 3 segundos sem tempo de espera. As amostras foram preparadas como para microscopia crioeletrônica, porém, durante a geração de imagens, uma série de imagens foram coletadas no mesmo local da amostra, mas com diferentes ângulos de inclinação da platina, e essas imagens inclinadas foram utilizadas para gerar um único tomograma. As séries de inclinação foram coletadas usando Tomografia 5 (Thermo Fisher Scientific) com ângulos de inclinação de -60° a +60° com passos de 2 graus usando um esquema dose-simétrico com uma ampliação de 33kx. A dose total de elétrons foi em torno de 100e/A/segundos e os valores de desfocagem ficaram entre -8 a -4 µm. As séries de inclinação foram coletadas em

um Titan Krios G4 equipado com filtro de energia SelectrisX e detector direto de elétrons Falcon 4i. Os tomogramas foram gerados utilizando o software IMOD.

Crio-ME CEMOVIS: a estirpe StC foi cultivada em meio de cultura R2A solidificado com ágar nobre de alta pureza (Sigma-Aldrich) em placa de petri, a 60 °C durante 3 dias. Após esse tempo, as colônias foram transferidas para dentro de um operador CEMOVIS (instrumento de metal com furo na parte superior) e foi adicionado uma solução de Dextran. Em seguida, as bactérias foram congeladas em alta pressão e foram devidamente vitrificadas (congeladas sem formação de cristais de gelo). Então, foram feitas seções finas das bactérias vitrificadas usando uma faca de diamante (para criar as seções de 20-50 nm da bactéria) e as seções foram colocadas na grade EM. Os dados criogênicos foram coletados em um microscópio Titan Krios G4 (ThermoFisher Scientific) operando a 300 kV. Os conjuntos de dados foram adquiridos com uma ampliação nominal de 50.000x usando uma câmera com detector direto de elétrons Falcon IV (Thermo Fisher Scientific).

## 5. RESULTADOS

O tópico de resultados desta tese de doutorado está separado em capítulos, no qual o capítulo I e II foi escrito pela necessidade de um material de revisão para dar suporte na condução do presente trabalho. No capítulo III contém dados inéditos resultantes do delineamento experimental da presente tese.

- O capítulo I é um artigo de revisão publicado no ano de 2023 na revista Journal of Advanced Research (IF: 12.8), intitulado: “Natural carbon fixation and advances in synthetic engineering for redesigning and creating new fixation pathways”. Esse artigo consiste na descrição de todas as vias naturais e sintéticas de fixação de carbono descobertas até o presente momento e na aplicação das ômicas aliadas a bioquímica para estudar as vias já conhecidas e descobrir novas vias.
- O capítulo II é um artigo de revisão submetida à revista Biotechnology Advances (IF: 17.681), intitulado: “Carboxysomes: The next frontier in biotechnology and sustainable solutions”. Nesse artigo é descrito detalhadamente a estrutura bioquímica e genética do carboxissomo, bem como os métodos para isolamento a partir da célula bacteriana e visualização por miscroscopia.
- O capítulo III é um artigo a ser submetido, intitulado: “Co-existence of CBB cycle and rTCA carbon fixation pathway in thermophilic *Actinomycetota Carbonactinospora*

*thermoautotrophica* StC and correlation with the CO2-concentrating microcompartment.” Nesse artigo está descrito os resultados inéditos gerados pela análise de genômica, transcriptômica, proteômica e isolamento/caracterização do microcompartimento.

## Capítulo I

Journal of Advanced Research 47 (2023) 75-92

Review

# Natural carbon fixation and advances in synthetic engineering for redesigning and creating new fixation pathways

Sulamita Santos Correa[a], Junia Schultz[b,c], Kyle J. Lauersen[d], Alexandre Soares Rosado[b,c,e]

[a] *Laboratory of Molecular Microbial Ecology, Institute of Microbiology, Federal University of Rio de Janeiro, Rio de Janeiro 21941-902, Brazil*

[b] *Red Sea Research Center (RSRC), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia*

[c] *Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia*

[d] *Bioengineering Program, Biological and Environmental Sciences and Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST),*

*Thuwal 23955-6900, Saudi Arabia*

[e] *Bioscience Program, Biological and Environmental Sciences and Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia*

h i g h l i g h t s

g r a p h i c a l a b s t r a c t

Carbon circulation on Earth. Anthropogenic and natural carbon emissions and carbon sequestration routes.

- Carbon is often negatively associated with the anthropogenic emissions of greenhouse gases and, consequently, with environmental issues such as global warming. However, carbon is an element of paramount importance for life and ecological processes.
- The knowledge of carbon fixation cycles, in addition to the ecological and evolutionary studies on the emergence of the first beings that inhabited Earth and metabolized carbon, provides a detailed understanding of the enzymes and important intermediaries involved in these cycles.
- The central carbon metabolizing enzymes can be used for a variety of purposes; artificial $CO_2$ fixation pathways can be designed and implemented in suitable host organisms; new solutions are possible for manipulating carbon fixation pathways and increasing their final yields.
- The multidisciplinary integration of omics, synthetic biology, molecular biology, and biochemistry can enable the scientific community to describe new carbon fixation pathways and predict important routes and enzymes involved in these processes. This integration may further provide guidance for overcoming the limitations of natural carbon fixation pathways and allow the development of biotechnological applications.
- Investigating thermophilic microorganisms is crucial for understanding many evolutionary events. These microorganisms are investigated extensively for developing biotechnological applications because of their

metabolic diversity.
● In addition to the six well-known natural carbon fixation pathways, the development of omics and systems biology has spearheaded the discovery of new natural and synthetic pathways for carbon metabolism.

Peer review under responsibility of Cairo University.

* Corresponding author.

*E-mail address:* alexandre.rosado@kaust.edu.sa (A. Soares Rosado).

https://doi.org/10.1016/j.jare.2022.07.011



article info

*Article history:*
Received 24 April 2022
Revised 30 June 2022
Accepted 25 July 2022
Available online 30 July 2022

*Keywords:*
Carbon fixation
Autotrophic
Thermophiles
Omics
Biochemistry
Synthetic pathway

abstract

*Background:* Autotrophic carbon fixation is the primary route through which organic carbon enters the biosphere, and it is a key step in the biogeochemical carbon cycle. The Calvin-Benson-Bassham pathway, which is predominantly found in plants, algae, and some bacteria (mainly cyanobacteria), was previously considered to be the sole carbon-fixation pathway. However, the discovery of a new carbon-fixation pathway in sulfurous green bacteria almost two decades ago encouraged further research on previously overlooked ancient carbon-fixation pathways in taxonomically and phylogenetically distinct microorganisms. *Aim of Review:* In this review, we summarize the six known natural carbon-fixation pathways and outline the newly proposed additions to this list. We also discuss the recent achievements in synthetic carbon fixation and the importance of the metabolism of thermophilic microorganisms in this field.
*Key Scientific Concepts of Review:* Currently, at least six carbon-fixation routes have been confirmed in Bacteria and Archaea. Other possible candidate routes have also been suggested on the basis of emerging ‘‘omics” data analyses, expanding our knowledge and stimulating discussions on the importance of these pathways in the way organisms acquire carbon. Notably, the currently known natural fixation routes can not balance the excessive anthropogenic carbon emissions in a highly unbalanced global carbon cycle. Therefore, significant efforts have also been made to improve the existing carbon-fixation pathways and/or design new efficient *in vitro* and *in vivo* synthetic pathways.

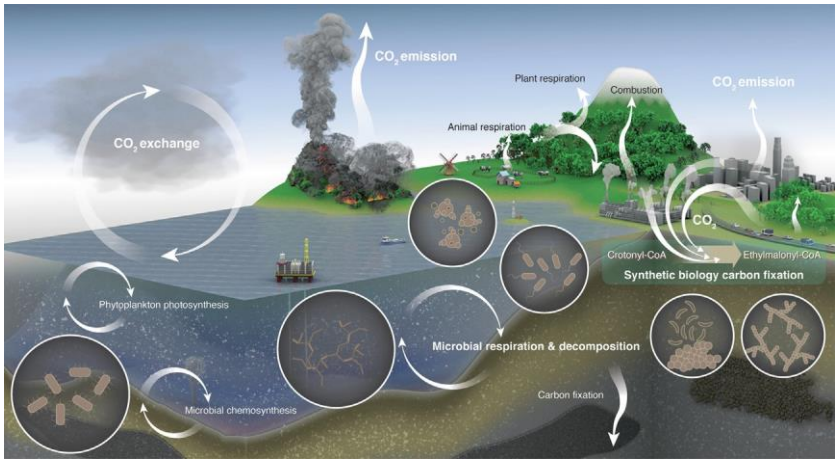


graphical abstract

Carbon circulation on Earth. Anthropogenic and natural carbon emissions and carbon sequestration routes.

Contents








## Introduction

Carbon is the fourth-most abundant chemical element on Earth, after hydrogen, oxygen, and helium [1,2], and it is considered as a base element for building organic compounds that are essential for life, such as proteins, carbohydrates, nucleic acids, and lipids. This element drives whole communities of living organisms and underpins the biogeochemical cycles on Earth [3].

The entry of inorganic carbon into the biosphere is related to carbon fixation. Primary production of organic compounds is directly related to autotrophic carbon fixation, which converts inorganic carbon into biomass [4] (Fig. 1). At present, six natural autotrophic carbon-fixation pathways have been described and accepted, while other candidate routes have been recently pro- posed. With advances in synthetic biology, bioinformatics, and biochemical analyses, various synthetic carbon-fixation routes have also been developed (see Table 1), including the crotonyl-( CoA)/ethylmalonyl-CoA/hydroxybutyryl-CoA (CETCH; here, CoA represents coenzyme A) cycle, reductive glycine, synthetic malyl-CoA-glycerate pathway, synthetic acetyl-CoA (SACA) cycle, and formalase pathway [5–6]. A timeline summary showing the key carbon fixation milestones discussed in our review is listed in Fig. 1.

The Calvin-Benson-Bassham (CBB) cycle is the most well-kwons mechanism of carbon assimilation and is found in all types of plants and algae and some prokaryotes [7]. However, new pathways of carbon fixation have also been elucidated in recent ecological, biochemical, and genomic studies [8]. In addition to the CBB cycle, autotrophic microorganisms can incorporate the carbon in biomass through five other natural carbon-fixation pathways: i) the reductive tricarboxylic acid (rTCA) cycle; ii) the reductive acetyl-CoA (also called the Wood–Ljungdahl [WL]) pathway; iii) the 3-hydroxypropionate [3-HP] bicycle; iv) the 3-hydroxypropionate/4-hydroxybutyrate (3-HP/4-HB) cycle; and v) the dicarboxylate/4-hydroxybutyrate (DC/4-HB) cycle [9–10]. Three previously unknown natural carbon-fixation routes have also been described recently: i) the reductive hexulose-phosphate pathway (RHP), ii) the natural reductive glycine (nrGly) cycle, and iii) the reverse oxidative TCA cycle (roTCA) [11–12].

Plants and photoautotrophic microorganisms are capable of fixing inorganic $CO_2$ via the CBB cycle, which uses energy in the form of NAD(P)H and adenosine triphosphate (ATP) derived from pho- tons. The global incident solar energy is 178,000 TW y-1, which is severalfold higher than that required by human society [13]. Oxy- genic phototrophs, such as cyanobacteria, plants, and algae, have the ability to use this incident solar energy as a free source for photosynthesis-induced energy generation. Due to the release of oxygen in the process, these organisms are considered as the primary producers of the current biosphere [14]. Conversely, anoxygenic photosynthetic bacteria are also able to fix $CO_2$ and grow autotrophically, typically using light as energy (ATP) and without producing oxygen in the process [15].

However, some autotrophic microorganisms are also capable of fixing inorganic carbon using other energy sources [16,17]. Autotrophic organisms assimilate inorganic carbon into

biomass using processes that are driven by light or chemical energy derived from organic molecule breakdown. Chemolithoautotrophic microbes use inorganic compounds (CO, $CO_2$, $H_2$, sulfate, phosphite, and Fe) as their energy sources, whereas chemoorganotrophic microorganisms rely on the energy produced from chemical transformations using organic compounds (sugars, organic acids, and amino acids) as carbon sources [17,18].

Autotrophic organisms are crucial components of the global carbon cycle since they assimilate the organic carbon that is released by other organisms in the biosphere [8–17]. The fossil fuel reserves that are essential in the current linear global economy represent carbon stored from the primary autotrophic production of biomass and sequestered in geological deposits. However, the unsustainable rates of global anthropogenic carbon release have now surpassed the ability of natural carbon-fixation pathways to recapture and fix this carbon in the biosphere [19]. Thus, an under- standing of the metabolic conversion of inorganic carbon into organic carbon is critical for developing potential alternative solutions for capturing and using these waste carbon sources.

The collective techniques that generate ‘‘omics” data on a massive scale are expected to facilitate the discovery of novel carbon- fixation pathways and efficient carbon-fixing enzymes [4–20]. Future trends are focused on manipulating the known autotrophic carbon-fixation pathways to increase their efficiency and to under- stand their components, as well as the development of new and more efficient natural and synthetic fixation cycles [21,22]. Moreover, thermophilic microorganisms are also valuable sources of metabolic diversity and are currently being investigated for alternative carbon-fixation pathways, because a chemolithoautotrophic thermophile is probably the most common precursor to life and provides the best model for investigating primordial metabolism [7]. This review summarizes and discusses the six known auto- trophic pathways of carbon fixation, the newly identified and pro- posed natural pathways, as well as recent advances in designing new pathways for biological and synthetic carbon fixation (Fig. 2). We discuss the continued progress of ‘‘omics” tools and their role in unveiling new routes of carbon fixation and the efforts devoted to identifying the enzymes that participate in carboxylation reactions. Furthermore, efforts to develop engineered enzymes with activities that surpass those found in natural carbon-fixation cycles are also discussed. Subsequently, we focus on carbon fixation performed by thermophilic Bacteria and Archaea to under- stand the carbon-fixation pathways and obtain insights for future attempts at synthetic engineering. Finally, we highlight the knowledge gaps in the field as well as the challenges and new directions for future studies on carbon fixation.

## Autotrophic carbon-fixation pathways: The six known cycles

### Calvin–Benson–Bassham cycle

The photosynthetic and pentose phosphate reducing pathway (CBB pathway) discovered by Melvin Calvin, Andrew Benson, and James Bassham in the 1940 s and 1950 s [23] was the first reported carbon-fixation pathway. The CBB cycle is present in algae, plants, cyanobacteria, phyla Proteobacteria and *Bacillota*, and photoautotrophic and chemoautotrophic bacteria [24–25]. In microorganisms, this cycle is the most important carbon dioxide fixation pathway [26]. The key enzymes in this pathway include ribulose- 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) and phosphoribulokinase (PRK) [27] (Fig. 2A). Another important enzyme in the CBB cycle is glyceraldehyde-3-phosphate dehydrogenase (GAPDH). GAPDH and PRK participate in the regeneration of RuBisCO and form an inactive complex with the regulatory protein CP12 when in a dark environment [28]. CP12 is formed by 80 amino acids and, after binding to GAPDH and PRK, becomes an important light/dark regulator [29]. In dark, the CP12 protein com- plex is deactivated, while under light conditions, this complex is dissociated, allowing the reactivation of GAPDH and PRK enzymes [30]. Most oxygenic photosynthetic organisms that use the CBB cycle to fix $CO_2$ contain CP12, whereas nonoxygenic phototrophic bacteria do not have this protein [28–30]. In the CBB cycle, three molecules of $CO_2$ are fixed by RuBisCO to form one molecule of glyceraldehyde 3-phosphate, in a pathway composed of a total of 29 reactions from 13 enzyme reactions [31]. RuBisCO drives the production of carbohydrates by catalyzing the removal of $CO_2$ from the atmosphere and its conversion into carbohydrates for plants and other organisms [32].

Table 1 Comparative description of the natural and synthetic carbon fixation pathways: microorganism examples, energy sources, necessary inputs, products of the cycle, reducing agents, main enzymes, and sensitivity of the enzymes to oxygen.

| | *Status* | *Organisms* | *Energy Source* | *Input* | *Output* | *Reductants* | *Key Enzyme* | *$O_2$ Sensitive* | *References* |
|---|---|---|---|---|---|---|---|---|---|
| Calvin-Benson | *Natural* | Plants, Algae, Cyanobacteria, Aerobic Proteobacteria, Purple bacteria | Light | 3 $CO_2$, 9 ATP,6 NAD(P)H | Glyceraldehyde-3-phosphate | NAD(P)H | RuBisCO | No | [23] |
| rTCA | *Natural* | Green sulfur bacteria, Proteobacteria, *Aquificae, Nitrospirae* | Light and Sulfur | 2 $CO_2$, 2 ATP,4 NAD(P)H | Pyruvate | NAD(P)H and ferredoxin | 2-Oxoglutarate synthase, Isocitrate dehydrogenase | Yes | [54] |
| Wood–Ljungdahl | *Natural* | Acetogenic, Methanogenic Archaea, Planctomycetes, Sulfate, *Archaeoglobales* | Hydrogen | 2 $CO_2$, 1 ATP, 4 NAD(P)H | Acetyl-CoA | Ferredoxin | NAD-independent formate dehydrogenase, Acetyl-CoA synthase-CO dehydrogenase | Yes | [9] |
| 3-HP | *Natural* | *Chloroflexaceae* | Light | 3 $HCO^-$, 5 ATP, 5 NAD(P)H | Pyruvate | NAD(P)H | Acetyl-CoA carboxylase, Propionyl-CoA Carboxylase | No | [72,187] |
| HP/HB | *Natural* | Aerobic *Sulfolobales* | Hydrogen and Sulfur | 2 $HCO^-$, 4 ATP, 4 NAD(P)H | Acetyl-CoA | NAD(P)H | Acetyl-CoA-Propionyl-CoA carboxylase | No | [78] |
| DC/HB | *Natural* | Anaerobic *Thermoproteale, Desulfurococcales* | Hydrogen and Sulfur | 1 $CO_2$, 1 $HCO^-$, 3 ATP, 4 NAD(P)H | Acetyl-CoA | NAD(P)H and Ferredoxin | Pyruvate synthase, PEP carboxylase | Yes | [79] |
| RHP | Candidate *Natural* | *Methanospirillum hungatei* | Hydrogen | $CO_2$, 3 ATP, 2 NAD(P)H | Gluconeogenesis and glycolysis | NAD(P)H | RuBisCO | No | [11] |
| Natural Reductive Glycine | Candidate *Natural* | *Candidatus phosphitivorax anaerolimi*, *Desulfovibrio desulfuricans* | Phosphite | $CO_2$, ATP, NAD(P)H | Formate/ Pyruvate | NAD(P)H and ferredoxin | $CO_2$-reducing formate dehydrogenase (fdhAB) | – | [92] |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Reverse oTCA | Candidate | *Desulfurella acetivorans* | Hydrogen | $CO_2$, ATP, NAD(P)H | Acetyl-CoA | Ferredoxin | Citrate synthase | – | [12] |
| CETCH | *Natural Synthetic* | Theoretical | – | 2 $CO_2$, 2 ATP, 3 NAD(P)H | Glyoxylate | NAD(P)H | CoA- dependent carboxylase | No | [104] |
| Reductive Glycine | *Synthetic* | Demonstrated in *E. coli* as host | – | $CO_2$, NADH | Pyruvate | Ferredoxin | Glycine cleavage system | No | [93] |
| Synthetic malyl-CoA-glycerate | *Synthetic* | Demonstrated in *E. coli* and *Synechococcus elongatus* PCC7942 host | – | $CO_2$, 3 ATP, 3 NADH | Acetyl-CoA | NAD(P)H | PEP-carboxylase, RuBisCO | No | [119] |
| SACA Pathway | *Synthetic* | Demonstrated in *E. coli* as host | – | $CO_2$ | Acetyl-CoA | – | NAD-independent formate dehydrogenase | No | [6] |
| Formolase pathway | *Synthetic* | Theoretical | – | $CO_2$, NADH, ATP | Dihydroxyacetone-phosphate | NADH | NAD-independent formate dehydrogenase | No | [5] |

3

3

3

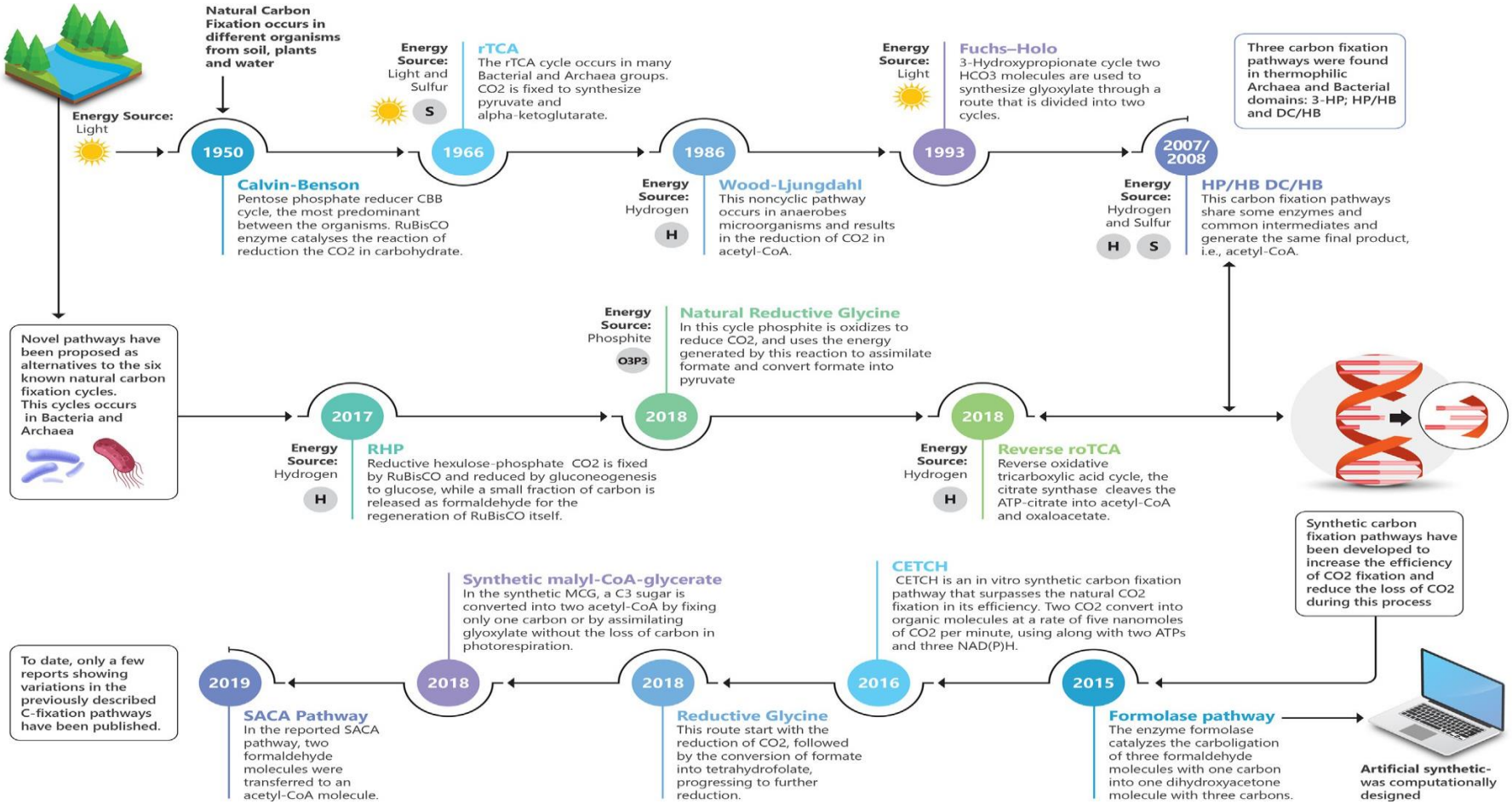



Fig. 1. A timeline summary showing key milestones and findings related to carbon fixation pathways discussed in our work.

Recently, a novel variant of the CBB cycle was described using computation of elementary flux modes or extreme paths to find all pathways leading to the formation of glyceraldehyde 3- phosphate from $CO_2$ [31]. Named as the S7P-removing transaldolase variant, it involves transaldolase, and it bypasses the intermediate enzyme fructose 1,6-bisphosphate [31]. The author suggested that transaldolase can be an alternative for enhancing photosynthetic carbon metabolism. Previously, this theoretical pathway proposed as the S7P-forming transaldolase variant because it requires an S7P-forming step [33], different from the variant described by Ohta [31]. The S7P-forming transaldolase variant works twice per three $CO_2$ molecules fixed, in which transaldolase transfers three carbons from fructose 6-phosphate to erythrose 4-phosphate to make S7P and glyceraldehyde 3- phosphate [33–34].

RuBisCO catalyzes both the carboxylation and oxygenation of its substrate [35]. $CO_2$ and O2 compete for the same active site on RuBisCO when reacting with the same substrate. The carboxylation results in the formation of two molecules of a three-carbon- containing organic acid, 3-phosphoglycerate (3-PGA), whereas RuBisCO oxygenation produces one molecule of 3-PGA and another of 2-phosphoglycolate (2-PG) using a process called photorespiration [36]. The 3-PGA molecule resulting from RuBisCO carboxylation generates carbohydrates, whereas 2-PG is not metabolized in the CBB cycle and is instead used in respiration, wherein it consumes O2 and releases already fixed $CO_2$ [37].

To increase the concentration of $CO_2$ near RuBisCO and decrease the photorespiration rate, cyanobacteria and some chemoautotrophs contain an intracellular microcompartment-carboxysome-that functioning as an organelle [38]. The carboxysomes encapsulate RuBisCO and the carbonic anhydrases and operate as semipermeable barriers that allow the passage of HCO3 and RuBisCO and exclude O2 that competes for the substrate. In the carboxysome, the carbonic anhydrases catalyze the conversion of HCO3 into $CO_2$, which is fixed by RuBisCO to form 3-PGA [39]. The efficient $CO_2$ fixation mechanism in the carboxysomes recently inspired their implementation in plants to increase crop productivity [40]. In some eukaryotic algae, carbon fixation is mediated by several mechanisms to enrich the $CO_2$ concentration locally as the pyrenoid, which appears as a dynamic compartment of mem- branes and crystalline starch under a microscope [41]. A recent study provided the first glimpse of the cycle of a carboxysome inside living cells through a fluorescence imaging platform that allows the simultaneous measurement of the numbers, positions, and activities of carboxysomes in cyanobacteria [42].

Importance of RuBisCO

Four forms of RuBisCO, which has a structure formed by eight large subunits and eight small subunits, are known [43]. The I and II forms participate in the autotrophic assimilation of $CO_2$. RuBisCO form I is more widely distributed and is found in eukaryotes and some prokaryotes. Additionally, form I RuBisCO is roughly 2.4 million years old and emerged from the great oxygenation event that occurred owing to the transforming action of cyanobacteria which began to introduce oxygen into the atmosphere driven by the abundant photosynthetic energy provided by the sunlight. Meanwhile, RuBisCO form II occurs in prokaryotes (mostly in Proteobacteria species) and microeukaryotes such as dinoflagellates (i.e., *Symbiodinium*) [44]; these are the only groups capable of processing RuBisCO form II, which is encoded by the nuclear DNA [45,46].

Banda et al. [47] discovered an uncharacterized clade sister of RuBisCO form I, i.e., RuBisCO I'. Their study involved the metage- nomic analysis of samples from environmental communities with a large amount of nonculturable bacteria. They reported that RuBisCO form I' may have evolved under anaerobic conditions before the evolution of cyanobacteria. Unlike RuBisCO form I, RuBisCO form I' consists of only eight large subunits without the small subunits. Twenty-four rbcL genes with gene products that share a high sequence homology (52 %–61 %) with the known RuBisCO form I were reported. However, a more detailed analysis of the genomes assembled from the metagenome indicated the absence of the rbcS genes [47]. These genes are always found in Bacteria and Archaea that use the CBB cycle to fix carbon [48]. On the basis of these results, the group suggested that RuBisCO I' probably represents a distinct form of RuBisCO that presumably diverged from form I before the origin of life.

According to previously reported studies, form III of RuBisCO does not participate in autotrophic $CO_2$ fixation; instead, it uses ribonucleotides via the pentose-bisphosphate pathway [49]. However, using ''omics" and biochemical approaches, Frolov et al. [25] identified an operative form III RuBisCO in a transaldolase variant of the CBB cycle in *Thermithiobacillus tepidarius* and *Thiobacillus thioparus*. Form III was found exclusively in Archaea [50], and since then, it has been hypothesized to exist in bacterial candidates [51]. RuBisCO form IV is known to not perform carboxylase and oxygenase activities and is distributed in seven phylogenetically distinct subgroups, namely, IV-Photo (phototrophic bacteria), IV- Nonphoto (nonphototrophic bacteria), IV-AMC (microbial consortia found in acid minutees lakes), IV-GOS (ocean sampling sequences), IV-DeepYkr (deep branch close to the YkrW clade), IV-Aful (Archaeoglobus fulgidus), IV-YkrW (Bacillus) [52], and a recently discovered form that acts as an oxygenase that converts 3-keto-D-ribitol-1,5- bisphosphate into two hydroxy acid phosphorylates: 3-D-phosphoglycerate and phosphoglycolate [53].

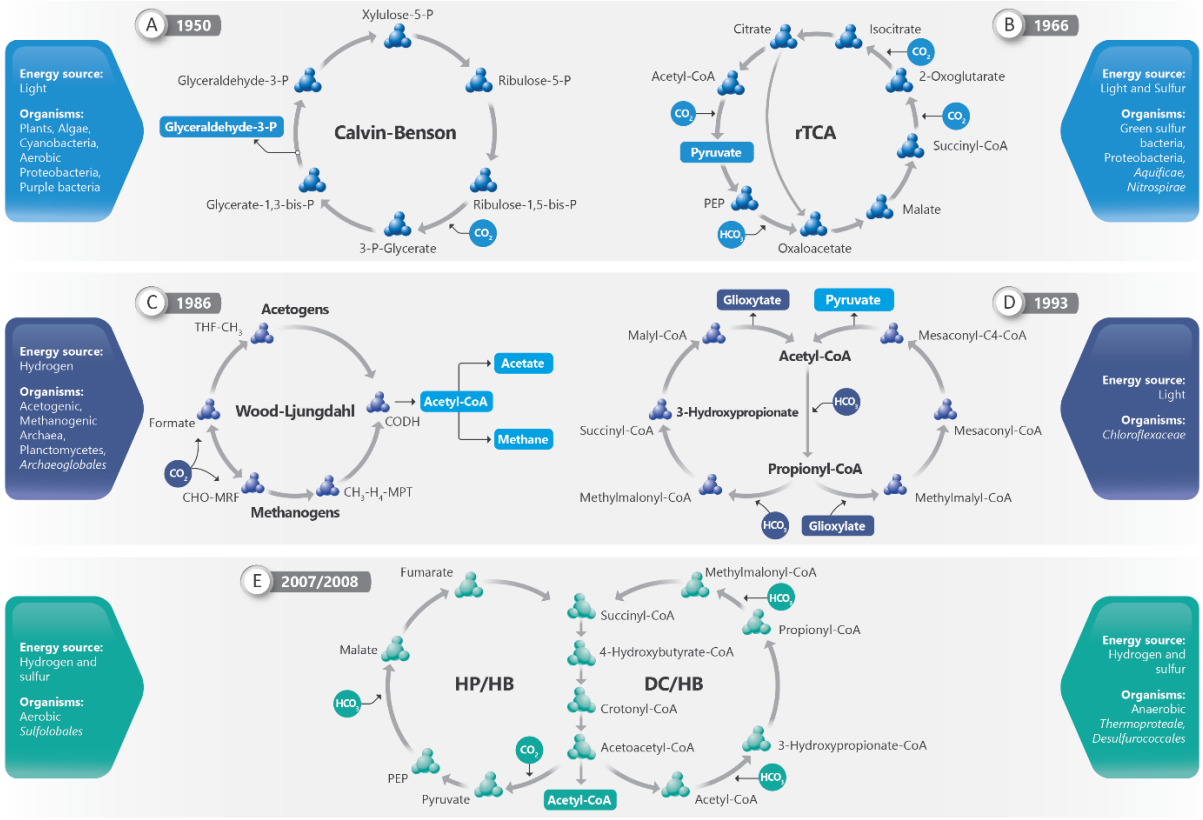



Fig. 2. Natural carbon fixation. (A) CBB cycle; enzymes: ribulose-1,5-bisphosphate carboxylase/oxygenase, 3-phosphoglycerate kinase, glyceraldehyde-3-phosphate dehydrogenase, ribulose-phosphate epimerase. (B) rTCA cycle; enzymes: ATP-citrate lyase, malate dehydrogenase, succinyl-CoA synthetase, ferredoxin (Fd)-dependent-2- oxoglutarate synthase, isocitrate dehydrogenase, PEP carboxylase. (C) Wood–Ljungdahl cycle, upstairs acetogens Archaea and downstairs methanogens Archaea; enzymes: MPT-methylene tetrahydromethopterin, MFR-methanofuran, THF, tetrahydrofolate. (D) 3HP cycle; enzymes: acetyl-CoA carboxylase, propionyl-CoA carboxylase, methylmalonyl-CoA epimerase, succinyl-CoA:(S)-malate-CoA transferase, trifunctional (S)-malyl-CoA, -methylmalyl-CoA, mesaconyl-CoA transferase, mesaconyl-C4-CoA hydratase. (E) HP/HP and DC/HP cycle; enzymes: pyruvate synthase, PEP-carboxylase, malate dehydrogenase, fumarate hydratase/reductase, acetyl-CoA/propionyl-CoA carboxylase, 3-hydroxypropionate-CoA ligase/dehydratase, methylmalonyl-CoA mutase, succinyl-CoA reductase, 4-hydroxybutyrate-CoA ligase, crotonyl-CoA hydratase, acetoacetyl-CoA-ketothiolase.

*Reductive tricarboxylic acid cycle (rTCA)*

The second carbon-fixation pathway was described in 1966 by Evans and collaborators [54]. This route was revealed in *Chlorobium limicola*, a green sulfur bacterium. Experiments with fermentative bacteria revealed a reaction that challenged the concepts of carbon fixation. Ferredoxin was found to serve directly as an electron donor for carbon fixation for the synthesis of pyruvate and alpha-ketoglutarate from $CO_2$ and reduced ferredoxin and acetyl-CoA or succinyl-CoA, respectively, through an irreversible pathway [55].

This route was named rTCA or the Arnon-Buchanan cycle and is believed to be the most plausible candidate for the first auto- trophic metabolism [56]. In the rTCA cycle, two cofactors

mediate the carbon fixation: thiamine pyrophosphate, which converts acetyl-CoA to pyruvate, and succinyl-CoA to produce ketoglutarate [57]. This cycle behaves in a direction opposite to that of the oxidative citric acid cycle (oTCA; one of the stages of cellular respiration in heterotrophs, which is performed in the presence of $O_2$ and releases fixed carbon) and forms acetyl-CoA from two $CO_2$ molecules.

Some key enzymes of the Krebs cycle are considered to react irreversibly and are replaced in the rTAC cycle to reverse the cycle; for example, succinate dehydrogenase is substituted by fumarate reductase; NAD 2-oxoglutarate dehydrogenase is replaced by ferredoxin 2-oxoglutarate synthase; and citrate synthase is replaced by ATP-citrate lyase [54,58,59]. The main product of carbon fixation in the rTCA cycle is acetyl-CoA, which is later converted into other central intermediates of the carbon metabolism, including pyruvate/phosphoenolpyruvate (PEP), oxaloacetate, and 2-oxoglutarate [56]. Acetyl-CoA is carboxylated into pyruvate by ferredoxin-dependent pyruvate synthase, and pyruvate can also be converted into PEP; oxaloacetate is synthesized in the reactions of pyruvate or PEP carboxylase [54] (Fig. 2B). The rTCA cycle occurs in many Bacteria and Archaea groups and includes enzymes that are sensitive to oxygen. For this reason, this cycle is observed in anaerobes or facultative anaerobes that can tolerate $O_2$ concentrations lower than that found in the air [60]. Although discovered in photosynthetic green sulfur bacteria, the rTCA cycle has been shown to be present in plenty of chemoautotrophic microbes [55]. More recently, it was assumed that the complete TCA cycle can occur in *Pandoravirus massiliensis*, which showed genes with low similarity to all enzymes of the cellular TCA cycle and, most importantly, a functional isocitrate dehydrogenase, a key enzyme of this cycle [61]. The gene ORF132 that encodes isocitrate dehydrogenase was cloned and expressed in *Escherichia coli* and was shown to encode an isocitrate dehydrogenase. The authors suggested the possible existence of an autonomous TCA pathway or another unknown metabolic pathway that may be involved in redirecting amoeba carbon metabolism where this virus was replicated [61].

*Reductive acetyl-CoA pathway*

In 1986, Ljungdahl and collaborators [9] described the third carbon-fixation route, i.e., the reductive acetyl-CoA or WL pathway. This noncyclic pathway occurs in anaerobic microorganisms and results in the reduction of $CO_2$ in acetyl-CoA. According to the reported studies, these microbes may have been the first autotrophs to use inorganic compounds such as CO and H2 as energy sources and $CO_2$ as an electron acceptor one billion years before the

appearance of $O_2$ [62].

In this cycle, two $CO_2$ molecules are converted into acetyl-CoA through two different branches that operate in parallel: the eastern or methyl branch and the western or carbonyl branch [63]. In the eastern branch, one $CO_2$ molecule undergoes six-electron reduction to form a methyl group, whereas the western branch involves the reduction of the other $CO_2$ molecule to CO, which is condensed with the methyl group and CoA to form cetyl-CoA [64]. Acetyl-CoA is then incorporated into cellular carbon or converted to acetylphosphate, from which the phosphoryl group is transferred to adenosine diphosphate (ADP) to generate ATP and acetate, which are the main growth products of acetogenic bacteria [9,65,66].

The eastern branch begins with the reduction of $CO_2$ in formate catalyzed by the formate dehydrogenase enzyme, which undergoes ATP-dependent condensation with tetrahydrofolate (H4F) to form 10-formyl-H4folate, which is converted to 5,10-methylenyl- H4folate. The next step is the reduction of 5,10-methylene- H4folate to 5-CH3-H4folate [63]. At the end of the branch, the methyltetrahydrofolate (corrinoid/iron–sulfur protein) methyl- transferase enzyme catalyzes the relocation of the methyl group on N5 of (6S)–CH3-H4folate to the cobalt center of the corrinoid iron-sulfur protein (CfeSP). In contrast, the western branch involves one enzyme, i.e., the CO-dehydrogenase/acetyl-CoA synthase, which catalyzes the conversion of $CO_2$ into CO [67].

When organisms are cultivated with CO, CO dehydrogenase generates $CO_2$, which is converted into formate in the eastern branch, and CO is incorporated directly into the carbonyl group of acetyl-CoA. In the next step, the acetyl-CoA synthase catalyzes the condensation of CO, CoA, and the methyl group of a methylated CfeSP to generate acetyl-CoA; at this point, the western branch meets the eastern branch.

In the Archaea *Methanosarcina acetivorans*, the conversion of methanogen to acetogen has been shown to completely dispense with the conserved methanogenesis energy [68]. *Methanosarcina acetivorans* uses the reducing acetyl-CoA to conserve energy in the form of acetate [69]. The genes encoding Mtr, i.e., disruption from the methyl branch toward methane, were eliminated in *M. acetivorans*, and all carbon from the methanogenesis substrate was diverted to flow through acetyl-CoA [68]. Thus, *M. acetivorans* could be converted into an acetogenic organism, which suggests that methanogenesis may have evolved from the acetyl-CoA path- way [68].

The reductive acetyl-CoA route is found in prokaryotes that live at thermodynamic limits such as the acetogenic and methanogenic Archaea [23], which use this route not only for the reduction of $CO_2$ in acetyl-CoA but also for the conservation of energy by generating an

electrochemical gradient [70]. In the energy-conservation process occurring during autotrophic growth or in the fermentation process, the acetogens and methanogens generate acetate and methane, respectively [71] (Fig. 2C).

*3-Hydroxypropionate bicycle*

The 3-hydroxypropionate bicycle (or the Fuchs-Holo route) was found in *Chloroflexus aurantiacus*, which is a green sulfur thermophilic bacterium, by Helge Holo and Georg Fuchs in 1989 [58,72]. In this pathway, two HCO3 molecules are used to synthesize glyoxylate through a route that is divided into two cycles. In the first cycle, glyoxylate is formed from the carboxylation of acetyl-CoA to produce malonyl-CoA, which is reduced to proprionyl-CoA via hydroxypropionate, after which the carboxylation of proprionyl-CoA produces succinyl-CoA. Subsequently, succinyl-CoA is converted to (S)-malyl-CoA, regenerating acetyl- CoA (starting molecule) and releasing glyoxylate as the first pro- duct of the carbon fixation. The second cycle begins with the assimilation of glyoxylate and its conversion to methylmalyl-CoA, which in turn is converted to form citramailyl-CoA by mesaconyl-CoA. Finally, the product citramailyl-CoA is converted into pyruvate and acetyl-CoA, thereby regenerating the precursors of the cycle [73] (Fig. 2D).

This pathway involves 19 steps and 13 enzymes; none of the steps are sensitive to oxygen and all the enzymes work under aerobic conditions [74]. This fourth carbon-fixation pathway appears to be exclusively restricted to a single clade of photosynthetic *Chloroflexus* [75]. This green sulfur bacterium uses the 3-hydroxypropionate bicycle in combination with the glyoxylate cycle to channel organic substrates such as glycolate, acetate, propionate, 3-hydroxypropionate, lactate, butyrate, or succinate for its central carbon metabolism under both autotrophic and hetero- trophic growth conditions [76].

*Hydroxypropionate/4-hydroxybutyrate (HP/HB) cycle and dicarboxylate/4-hydroxybutyrate (DC/HB) cycle*

The fifth and sixth carbon-fixation pathways share some enzymes and common intermediates and generate the same final product, i.e., acetyl-CoA [77] (Fig. 2E). The fifth carbon-fixation pathway the 3-hydroxypropionate/4-hydroxybutyrate cycle (HP/HB) was described in *Metallosphaera sedula*, which is a thermophilic Archaea belonging to the *Sulfolobales* order [78]. The sixth pathway (dicarboxylate/4-hydroxybutyrate [DC/HB] was

found in *Ignicoccus hospitalis*, which is an Archaea belonging to the order Desulfurococcales [79]. Existing variants of both cycles, the orders Sulfolobales and Desulfurococcales, did not evolve from a common ancestor, and the representatives of these two orders have phylogenetically unrelated enzymes [80].

In these two cycles, two $CO_2$ molecules are added to acetyl-CoA to produce succinyl-CoA and rearrange into acetoacetyl-CoA and cleave into two acetyl-CoA molecules [81]. In the HP/HB cycle, an acetyl-CoA, and the propionyl-CoA carboxylase, fixes two HCO3 molecules, and in the DC/HB cycle, the pyruvate synthase and PEP carboxylase catalyze this reaction. These pathways vary mainly in relation to oxygen tolerance and the use of reduction cofactors: NAD(P)H is used in the HP/HB cycle, and ferredoxin/NAD(P)H is employed in the DC/HB cycle [23,82]. The enzymes in the HP/HB cycle are oxygen-tolerant, whereas the 4-hydroxybutyryl-CoA dehydratase enzyme in the DC/HB cycle is inactivated in the presence of oxygen. Although the two cycles produce acetyl-CoA, their main difference lies in their link to the carbon metabolism. In the DC/HB pathway, pyruvate is synthesized from acetyl-CoA by using a pyruvate synthase, and in the HP/HB pathway, another half turn of the cycle is needed to produce succinyl-CoA, which is oxidized via succinate to produce pyruvate [83].

This cycle is also found in *Nitrosopumilus maritimus*, belonging to the order Sulfolobales. This species heterologously produces the protein Nmar_1028 that catalyzes the conversion of succinyl- CoA into two acetyl-CoA molecules. Nmar_1028 is homologous to the dehydrogenase domain of crotonyl-CoA hydratase/(S)-3-hydro xybutyryl-CoA dehydrogenase and seems to be the only (S)-3-hydroxybutyryl-CoA dehydrogenase in N. maritimus and is thus essential for the functioning of the 3-hydroxypropionate/4-hydroxybutyrate cycle [84]. In *Metallosphaera sedula*, Msed_2001 catalyzes the same reaction as Nmar_1028, and these two enzymes are homologous and have evolved independently from their respective bacterial homologs [15]. The existence of the HP/HB cycle in two distantly related archaeal groups with different physiologies is not evident in the course of evolution of autotrophic $CO_2$ fixation, because this process would need to be adapted to the corresponding ecological niche [15].

New candidate pathways for natural carbon fixation

Recent studies have revealed that carbon fixation is not limited to these six cycles (Tab. 1). Genome analyses have shown that several autotrophs do not present genes from any of the already described autotrophic $CO_2$-fixation pathways [23]. Therefore, investigations performed in this direction are of utmost importance for establishing novel ways of fixing inorganic carbon

and exploring alternative routes for metabolic energy supply [85].

*RHP pathway*

To date, three novel pathways have been proposed as alternatives to the six known natural carbon-fixation routes. Kono et al. [11] fully described the RHP pathway in the methanogenic Archaea, *Methanospirillum hungatei*. The carbon metabolism has been shown to involve the enzymes RuBisCO and phosphoribulokinase (PRK), which are the same enzymes found in the CBB cycle. RuBisCO of *M. hungatei* seems to form a new clade with RuBisCO of another methanogenic Archaea that also has PRK; this clade is different from RuBisCO form III. The coexistence of the RuBisCO and PRK in this Archaea led the authors to propose that *M. hungatei* fixed carbon. However, the enzymes transketolase, ribulose-5-phosphate and sedoheptulose-1,7-bisphosphatase, which participate in the CBB cycle, are absent from the genome. Archaea are known to lack genes for a transketolase, which is essential in the CBB cycle [86].

The proposed RHP pathway is noncyclical and shares the majority of the reactions of the CBB cycle, except for those in the regeneration stage. In the RHP cycle, $CO_2$ is fixed by RuBisCO and reduced by gluconeogenesis to glucose, while a small fraction of carbon is released as formaldehyde for the regeneration of RuBisCO itself [11]. Because the RHP cycle differs from the CBB cycle in some steps, the reported results provide insights into the evolutionary and functional associations between carbon metabolism and the pathways involving the RuBisCO enzyme present in photo- synthetic organisms and methanogenic Archaea. The authors speculate that the CBB cycle and RHP originated from a primitive carbon metabolic pathway that used RuBisCO, but with the replacement of a few steps and without carbon release. Although the energy requirements of the RHP pathway are significantly lower than those of the CBB cycle in plants and cyanobacteria, it does not proliferate as in the case of the CBB.

*Natural reductive glycine (nrGly) pathway*

Next-generation sequencing can provide a holistic approach to understand metabolisms through multiple ‘‘omics” datasets [87,88]. The data generated by genomics, proteomics, transcriptomics, and metabolomics have significantly increased the under- standing of cell physiology and provided answers to important questions about the metabolisms of

microorganisms [88,89]. Studies on the (meta)genome and (meta)transcriptome have also made it possible to identify the key genes involved in carbon fixation, allowing the prediction of potential carbon-fixation routes in (un- culturable) environmental microbes as well as the discovery and design of new pathways for carbon fixation [90,91].

Figueroa et al. [92] proposed another candidate carbon-fixation route, nrGly, using 16S rRNA region sequencing. The group pro- posed that inorganic carbon is assimilated into the glycine-reducing pathway found in the uncultivated bacterium *Candidatus Phosphitivorax anaerolimi* strain Phox-21, belonging to the class Deltaproteobacteria. This pathway was previously described as a synthetic route for carbon fixation [93]. In this proposed route, Phox-21 oxidizes phosphite to reduce $CO_2$, and uses the generated energy to assimilate formate through the glycine-reducing pathway. Due to the lack of alternative electron acceptors, Phox-21 has been suggested to couple phosphite oxidation to $CO_2$ reduction; in addition to the presence of $CO_2$ reductase and formate dehydrogenase (FdhAB) in its genome, this assumption could explain the mechanism by which this metabolism occurs [92]. The bacteria reduce $CO_2$ to formate by using FdhAB and convert formate into pyruvate via the glycine-reducing pathway. All the necessary precursors (acetyl-CoA, oxaloacetate, 2-oxoglutarate, and succinyl-CoA) can subsequently be generated from the pyruvate route in the partial rTCA cycle. The results of Phox-21 genomic analysis suggest that it may be possible for microorganisms to exploit the energy derived from phosphite oxidation, enabling autotrophic growth through the glycine-reducing pathway with $CO_2$ as the only electron acceptor [92].

Because of its very low redox potential, phosphite might be the only biological electron donor that can drive the fixation of $CO_2$ in chemotrophs through the reductive glycine pathway (an ATP-consuming network) in the absence of an additional energy source or electron acceptor. Furthermore, according to other reported studies, phosphite oxidation may be coupled with $CO_2$ reduction through the WL pathway [94]. Although genomic analysis has indicated a new carbon-fixation pathway, experimental evidence is required to verify the validity of this hypothesis.

Sánchez-Andrea et al. [95] described the reductive glycine path- way in *Desulfovibrio desulfuricans*. In addition to performing genomic, transcriptomic, proteomic, and metabolomic analyses, they also performed in vivo evaluation of autotrophic growth of bacteria. Unlike *C. Phosphitivorax anaerolimi*, the results of the study on *D. desulfuricans* demonstrated the enrichment of the culture medium with an organic carbon compound (cysteine and ruminal fluid) via autotrophic growth. The studies performed with labeled $CO_2$ and functional ‘‘omics” experiments proved the fixation of carbon through the reductive glycine pathway. The results

of this study demonstrate that the reductive glycine pathway is the only carbon-fixation pathway that allows autotrophic growth of *D. desulfuricans*. In the reductive glycine pathway, firstly, $CO_2$ is first reduced to formate, which is further reduced and condensed with a second $CO_2$ molecule to generate glycine. The resulting product is then reduced by the glycine reductase to acetyl-P and then to acetyl-CoA, which is further condensed with another $CO_2$ molecule to form pyruvate [95]. Thus far, it seems that this is the first reported study to prove the operation of the entire reductive glycine pathway in a microorganism.

Recently, Hao et al. [96] identified two *C. Phosphitivorax* strains (F81 and R76) as butyrate oxidizers. Transcriptome analyses were performed, and the entire ptx-ptd gene cluster for phosphite oxidation was reconstructed in strain F81; interestingly, this cluster was not found in strain R76. Genes involved in the glycine-reduction pathway were found in both the strains, whereas the essential genes for other autotrophic $CO_2$ pathways were absent. However, according to the authors, although genes for the glycine- reduction pathway were observed, they were not expressed. This observation indicates that these microorganisms, unlike the Phox-21 strain that was externally enriched with $CO_2$ and phosphate, live as heterotrophic syntrophs besides autotrophs. Other studies demonstrated that the glycine-cleavage system can sup- port glycine and serine biosynthesis from formate in an engineered.

E. coli strain at increased $CO_2$ concentration [97–98]. Nonetheless, it is necessary to verify the growth of the bacterium on formate (and also $CO_2$), which remains an unresolved challenge [99].

*Reverse oxidative tricarboxylic acid cycle (roTCA)*

Although metagenomic analyses have gained considerable acceptance in the studies of organisms and microbial communities, researchers have reported the need for joint efforts involving both ‘‘omics” and biochemistry [12]. In studying *Desulfurella acetivorans*, which is a sulfur reducing and thermophilic Deltaproteobacteria (optimal growth at 52 °C to 57 °C), the authors elucidated a new version of the rTCA cycle, which was named the reverse oTCA cycle (roTCA) (this report also mentions the oTCA cycle that provides energy for aerobic organisms and has citrate synthase as the key enzyme). The genome analysis of *D. acetivorans* revealed the absence of key genes and enzymes involved in the known path- ways of carbon fixation normally found in organisms exhibiting autotrophic growth. On the basis of this information, the authors investigated *D. acetivorans* in detail using classical biochemical techniques. *D. acetivorans* can grow heterotrophically using acetate as an electron donor and carbon source

[100] or autotrophically with H2 [101]. As previously described, the key enzyme in the rTCA cycle is the ATP-dependent citrate lyase enzyme; however, *D. acetivorans* do not have the genes that encode that enzyme.

In the roTCA pathway, citrate synthase is the key enzyme that cleaves ATP-citrate into acetyl-CoA and oxaloacetate. Using ultra- performance liquid chromatography, the cleavage of citrate into acetyl-CoA and oxaloacetate was observed in bacterial cell extracts, with the latter being further reduced to malate by malate dehydrogenase. The only enzyme in *D. acetivorans* that could possibly catalyze this ATP-independent cleavage is citrate synthase. In the cell extracts of *D. acetivorans*, fumarate reductase activity, which is another enzyme characteristic of the rTCA cycle, was also detected, and the regulation of carbon flow was proposed to occur as follows: the presence of H2 changes the flow toward the roTCA cycle or drives the exogenous acetate from the TCA cycle in the oxidative direction [12].

Zhang et al. [102] also identified the roTCA pathway in *Geobacter sulfurreducens*. Until then, it was believed that this group of bacteria was unable to fix carbon. However, the identification of the roTCA cycle in this bacterium suggested that it may grow chemolithoautotrophically. The authors used genomic analysis to investigate whether *G. sulfurreducens* could fix inorganic carbon, and serially transferred the strain to a chemolithoautotrophic culture medium containing formate as an electron donor and carbon source, as well as iron as an electron acceptor. Furthermore, enzymatic assays also corroborated the carbon fixation, showing that citrate synthase can achieve citrate cleavage, which is necessary for the function of the roTCA cycle. These findings showed that *G. sulfurreducens* can fix $CO_2$ through the roTCA cycle after adaptation; this previously unexplored metabolic pathway can be used for biotechnology and may elucidate previously unclear ecological functions for *Geobacter* [102].

## Development of synthetic pathways for carbon fixation

The global energy crisis and greenhouse effect have impeded the sustainable development of society [4]. Thus, reducing the emission of $CO_2$, which causes the greenhouse effect, is an urgent task. However, increasingly efficient carbon-fixation processes are also required because the rate of natural carbon fixation is inadequate to balance the $CO_2$ released from industries. Carbon fixation performed by photosynthetic organisms allows the recycling of greenhouse gas emissions into high value-added products. However, the dependence on light to drive carbon fixation can be limiting for industrial chemical synthesis [103]. Although natural

carbon-fixing processes are not feasible for use in industrial production, they offer a variety of carbon-fixing enzymes and corresponding pathways, opening options for artificial engineering [4]. Thus, in recent years, synthetic carbon-fixation pathways have been developed to increase the efficiency of $CO_2$ fixation and reduce the loss of $CO_2$ during this process.

*Crotonyl-(CoA)/ethylmalonyl-CoA/hydroxybutyryl-CoA (CETCH) cycle*

The CBB cycle is the most prevalent $CO_2$ assimilation mechanism in the biosphere and is found in all plants, algae, and some prokaryotes. However, its efficiency is limited for natural carbon fixation, rendering the productivity of this path extremely low despite its widespread ecological success owing to the availability of free energy through sunlight.

Schwander et al. [104] described an in vitro synthetic carbon- fixation pathway called the CETCH cycle, which surpasses natural $CO_2$ fixation in its efficiency. This pathway involves 17 enzymes that convert $CO_2$ into organic molecules at a rate of five nanomoles of $CO_2$ per minute, using two $CO_2$ molecules along with two ATPs and three NAD(P)H cofactors, which are the same energy carriers as for the oxygenic photoautotrophs.

The CETCH cycle was realized through a combinatorial approach involving enzyme engineering and metabolic review that was described as a radical and reductionist approach. This $CO_2$-fixation cycle was assembled from its main components in an ascending manner. First, a comparison was made between the kinetic and biochemical properties of all the known classes of carboxylases, and based on these analyses, the enzymes CoA-dependent carboxylases and enoyl-CoA carboxylases/reductases (ECRs) were chosen. ECRs are found in Alphaproteobacteria and Streptomyces, whose carboxylation activities surpass that of the propionyl-CoA carboxylase involved in the 3-HP bicycle, HP-HB cycle, and RuBisCO carboxylation [105,106]. In comparison with other carboxylases, ECRs work with a broad spectrum of substrates, are insensitive to oxygen, do not accept oxygen as a substrate, and catalyze $CO_2$ fixation with a high catalytic efficiency [107]. The CETCH pathway converts two $CO_2$ molecules from proprionyl-CoA to form a glyoxylate molecule through 13 main reactions catalyzed by 17 enzymes belonging to nine different organisms from four domains of life, including plants, humans, and microbes (Bacteria and Archaea), thereby indicating the potential for natural diversity.

The CETCH cycle shares four reactions and five intermediates with the HP-HB cycle [108]. At the time of writing this manuscript, the CETCH cycle was not implemented in a living host (neither a heterotrophic host, nor a photoautotrophic host). Indeed, balancing the

expression of 17 separate enzymes in a heterologous system is challenging even for the most genetically tractable organisms. Thus, the development of artificial carbon-fixation pathways may be limited by the gap between theoretical predictions and their experimental realization in synthetic biology. Moreover, attempts to synthesize new metabolic pathways in living organisms are challenging because of the limited understanding of the interactions between enzymes in heterologous systems [104]. Although the CETCH cycle presents attractive features for implementation in vivo, its synthetic engineering in microorganisms remains undemonstrated [109].

*Synthetic reductive glycine pathway (rGlyP)*

Carbon fixation can be implemented through assimilation routes based on natural or synthetic formats [110]. In recent syn- thetic biology studies, a formate-assimilation pathway, called the rGlyP pathway, that supports carbon fixation if coupled with $CO_2$ reduction has been demonstrated [93,110,111]. The rGlyP cycle is structurally similar to the most efficient carbon-fixation cycle known, i.e., the WL route. Both routes start with the reduction of $CO_2$, followed by the conversion of formate into tetrahydrofolate, and progressing to further reduction. This cycle uses an enzymatic complex to assimilate carbon and generate $CO_2$ compounds and convert them to pyruvate by condensation with another unit of carbon [110]. Formate is the product of $CO_2$ reduction by an electron pair and is considered to be the simplest organic compound that can supply cells with carbon and reducing power [110]. An efficient carbon-fixation pathway is one that combines $CO_2$ reduction with carboxylation, as shown by the rGlyP cycle [93]. Thus, the synthetic rGlyP pathway represents an efficient route because of the combination of ATP-free $CO_2$ reduction, carboxylation, and versatility.

The rGlyP route seems be a promising pathway for the assimilation of formate and other sustainable C1-feedstocks and could be applicable in future biotechnology-based attempts [112]. Modular engineering was used to implement the rGlyP for sup- porting synthetic formatotrophic growth as well as for growth on methanol in *Escherichia coli* [99]. The rGlyP was divided into four metabolic modules: 1) the C1 module (C1M), which consisted of the enzymes formate-THF ligase, methenyl-THF cyclohydrolase, and methylene-THF dehydrogenase from *Methylobacterium extorquens* that together converted formate into methylene-THF; 2) the C2 module (C2M), which consisted of endogenous enzymes of the GCS (GcvT, GcvH and GcvP) that converted methylene-THF with $CO_2$ and ammonia to form glycine; 3) the C3 module (C3M), which utilized serine hydroxy methyltransferase and serine

deaminase that condenses glycine and methylene-THF to generate serine and pyruvate; and 4) the energy module, which used formate dehydrogenase from Pseudomonas sp. [99]. This approach was used to redesign the central carbon metabolism of the model in E. coli that supported the growth of the bacteria with one carbon using rGlyP, and the growth in methanol and $CO_2$ was achieved by the additional expression of a methanol dehydrogenase. Biologically adapted microbial growth using formate and methanol has been investigated by the synthetic biology community in recent years [113]. Thus, this synthetic route appears to be finally becoming possible.

*Formaldehyde fixation 1 - the formolase pathway*

Even though formaldehyde is considered an organic carbon, sections 4.3 and 4.4 were included to discuss the complementary synthesis mechanisms of carbon fixation. To shorten the process of using carbon and accelerate growth, an artificial synthetic route the formolase pathway was computationally designed [17], and the computationally designed enzyme, called formolase, performs a reaction that catalyzes carboligation by directly fixing the units of one carbon in units of three carbons that feed the central metabolism. By integrating formolase with various naturally occurring enzymes, a new pathway for carbon fixation has been created that assimilates units of a carbon via formate. The enzyme formolase catalyzes the carboligation of three formaldehyde molecules with one carbon into one dihydroxyacetone molecule with three carbons [5].

*Formaldehyde fixation 2 - synthetic acetyl-CoA (SACA)*

The efficiency of carbon fixation can also be improved during the synthesis of acetyl-CoA, which is one of the central precursors involved in the biosynthesis of various products [114]. Recently, a synthetic route for acetyl-CoA (SACA) was designed using a glycolaldehyde synthase and an acetyl phosphate synthase and was introduced in *E. coli* [6]. In the reported SACA pathway, two formaldehyde molecules were transferred to an acetyl-CoA molecule in just three steps by using two ATP and two NADH molecules. First, the formaldehyde was condensed into glycolaldehyde by a glycolaldehyde synthase. Subsequently, the glycolaldehyde was converted to acetyl phosphate by an acetyl phosphate synthase. The glycolaldehyde synthase and acetyl phosphate synthase were selected and designed to increase their catalytic efficiency with their new substrates. At the end of the cycle, the acetyl phosphate group was

replaced by CoA in a process catalyzed by a phosphate acetyltransferase, and both in vitro and in vivo production of acetyl-CoA was achieved. The pathway exhibited high carbon- fixation capabilities [115,116]. The SACA pathway works with five reaction steps and enables acetyl-CoA production from formaldehyde; however, employing this route in *E. coli* can be challenging due to the low enzyme activity of glycolaldehyde synthase engineered, but glycolaldehyde synthase seems to show improved affinity to substrate formaldehyde [117,118]. Another question is whether the SACA pathway requires high intracellular concentration of formaldehyde, which can be toxic for the cells [118].

To overcome the deficiency of the CBB cycle for the efficient synthesis of acetyl-CoA, Yu and collaborators [119] designed the synthetic malyl-CoA-glycerate (MCG) pathway. The CBB cycle has not evolved for optimal production of acetyl-CoA; when 3- phosphoglycerate, which is the product of the CBB cycle, is con- verted to acetyl-CoA, one fixed carbon is lost as $CO_2$ [119]. In the synthetic MCG, a C3 sugar is converted into two acetyl-CoA by fixing only one carbon or by assimilating glyoxylate without the loss of carbon in photorespiration. In this synthetic carbon-fixation pathway, two oxaloacetate molecules are reduced to malate, which is then converted to malyl-CoA and later splits into two acetyl- CoAs and two glyoxylates. These two acetyl-CoAs are the final products of the pathway, and the two glyoxylates are used to regenerate PEP. The MCG pathway converts a PEP and HCO3 into an acetyl-CoA using three ATPs and three NADHs. The functionality of the MCG pathway was first demonstrated in vitro and in E. coli. Later, the pathway was implemented in the photosynthetic strain *Synechococcus elongatus* PCC7942, which showed an increase in the intracellular pool of acetyl-CoA as well as an improvement in the assimilation of HCO3 [119]. Recently, a system of $CO_2$ fixation oxygen insensitive, self-replenishing with opto-sensing was demonstrated through the junction of the MCG pathway and syn- thetic reductive glyoxylate and pyruvate synthesis in vitro. This system produced acetyl-CoA, pyruvate, and malate from $CO_2$, and its self-replenishing feature allowed every intermediate of system to be produced from $CO_2$ [120].

## Engineering RuBisCO for enhanced $CO_2$ fixation

The low efficiency of carbon fixation in autotrophs, limited by the enzyme RuBisCO, is a biotechnological short coming that has led to numerous studies focused on genetic engineering with different strategies to optimize the carboxylase function of this enzyme. RuBisCO is considered the most abundant enzyme on Earth, constituting 50 % of the total soluble protein in the chloroplasts of a C3 plant or in bacteria that use this cycle [121,122]. In

addition, RuBisCO also catalyzes a competitive reaction with oxy- gen and initiates the photorespiration process, which leads to a loss of fixed carbon and involves significant consumption of ATP to fix $CO_2$. However, in comparison with other enzymes in the CBB cycle, RuBisCO has a low turnover rate, implying that large amounts of this enzyme are needed to sustain the efficiency of $CO_2$ assimilation [123].

Improving RuBisCO by using bioengineering is a complex challenge due to two key aspects: 1) identification of the structural changes that promote performance and 2) identification of the ways to efficiently transplant these changes into RuBisCO within a target organism. Additionally, this task requires a satisfactory understanding of the regulatory pathways of the chloroplast gene, as well as the complex nature of catalysis and biogenesis promoted by this enzyme [32,124].

Despite the challenges in production, substantial progress has been made in bioengineering RuBisCO. In a previous study, the construction of a mutant RuBisCO was adopted as a strategy to improve the rate of carbon fixation in plants. Substitutions were made in the small subunit of the RuBisCO enzyme, thereby increasing the carboxylation activity by 85 % and enhancing the catalytic efficiency toward $CO_2$ by 45 % [125]. According to that study, the RuBisCO mutant could still be transplanted into higher plants if they shared the same hexadecameric L8S8 structure.

Another strategy already used to improve RuBisCO performance is the replacement of native RuBisCO by an exogenous homolog. Lin et al. [126] worked on replacing the RuBisCO of tobacco plants with the functional RuBisCO of the cyanobacterium *Synechococcus elongatus* PCC7942 (Se7942). The native gene encoding the large RuBisCO subunit was knocked out, and the genes for the large and small subunits of Se7942 corresponding to a chaperone, RbcX, or an internal carboxysomal protein were inserted. The trans- formed plants were photosynthetically competent and supported autotrophic growth, and the respective forms of RuBisCO exhibited greater rates of $CO_2$ fixation per unit of enzyme than the control plants [126].

The discovery of the new RuBisCO assembly factors (C3 and C4 plants) has provided next steps for improving this enzyme for higher plants [127,128]. For example, improving $CO_2$ assimilation now includes equipping C3 plants with a $CO_2$ concentration mechanism and generating alternative metabolic pathways to bypass the oxygenation [129,130].

Gene editing tools, such as chemical and physical mutagenic agents, that operate on nuclear DNA can be used to modify the family of multiple rbcS genes or insert new rbcS copies and are commonly applicable on various cultures [131]. Because the trans- formation of the

chloroplast is viable in only a few species, modifying the rbcL gene in the chloroplast genome is excessively complicated [132]. Thus, the implementation of RuBisCO in practical applications has necessitated the establishment of chloroplast transformation in several varieties of plants [133].

Mixotrophic-mediated inorganic carbon uptake enhancement

Many photosynthetic eukaryotic microorganisms are capable of consuming organic carbon sources in addition to $CO_2$. Higher plants arose from the first photosynthetic flagellated protists, which acquired a cyanobacterial endosymbiont that led to the chloroplast in photosynthetic eukaryotes. As these organisms are found in every environment, they possess the ability to consume organic carbon in the absence of light and $CO_2$, exhibiting an advantageous metabolic capacity. Within the green algae, the genus of Chlorella is ubiquitous around the world, and C. vulgaris has been generally regarded to show a safe status as an alternative protein-rich biomass, nutraceutical, and source of bio-compounds.

This organism can be fermented like yeasts on simple sugars as a carbon source [134,135]. Chlamydomonas is another genus of green algae, known for only being able to metabolize organic acetic acid via the glyoxylate cycle [136], whereas in the red algae, the extremophile *Galdieria sulphuraria* is known for its ability to metabolize over 50 different carbon sources [137]. The consumption of glucose in algae has been observed to correlate with a reduction in photosynthetic activity, in which the heterotrophic growth is favored over the autotrophic growth [138]. However, glycerol addition to mixotrophic cultures of the diatom *Phaeodactylum tricornutum* or *G. sulpheraria* increases their overall growth rates, resulting in higher growth rates in the presence of light and $CO_2$. *Chlamydomonas reinhardtii* demonstrates marked improvements in growth behavior in photo-mixotrophic conditions of high-$CO_2$ cultures fed with acetic acid [139,140]. Each of these organic carbon sources is metabolized through a different pathway and clearly stimulates different cellular responses. For example, although acetic acid mixotrophy reduces the chlorophyll content in exponential-phase *C. reinhardtii* cultures, the overall photosynthetic rates and biomass formation are improved [140]. We postulate that large amounts of organic carbon precursors can increase the amounts of CBB intermediates and, concomitantly, improve their ability to sequester $CO_2$ [140].

Algal cultivation is generally aimed at maximizing biomass pro- duction for a range of different applications, and now, with the advent of synthetic biological techniques, the engineering possibilities in these hosts have increased [141,142]. Examples of engineered algae

as sources of polyamines, isoprenoids, and recombinant proteins have been demonstrated using mixotrophic cultivation strategies [143–145]. Algae offer a unique platform for waste carbon conversion, and the study of mixotrophic carbon uptake enhancement can increase the yield of $CO_2$ and provide desired products via waste-stream conversions.

## Carbon fixation in thermophilic microorganisms

Thermophiles are microorganisms that can survive above the typical thermal limits of life, in inhospitable environments [146]. Although the temperature range for these microbes varies between 45 °C and 80 °C, the hyperthermophiles can survive even at temperatures above 80 °C [147,148]. They are found in volcanic areas, hydrothermal vents, solfatara fields, soils heated by steam, geothermal plants, and deserts [20]. Archaea, Bacteria, and Eukarya are the three domains of life in which extremophiles are found [149]. These microorganisms have genetic and physiological complexes that assist them in tolerating the environmental changes and adapting and repairing the damage caused by extreme environmental disturbances [150,151].

Thermophiles have attracted interest for biotechnological applications requiring whole cells, pure cultures, or consortia, and for applications wherein their macromolecules, metabolites and extreme enzymes are used [152,153]. Proteins and cell membranes with thermostable capabilities are not denatured at high temperatures; some can also resist proteolysis, making them highly desirable for industrial applications [20,154]. The discovery of the polymerase enzyme in the thermophiles *Thermus aquaticus* and *Pyrococcus furiosus* significantly influenced and revolutionized the use of the polymerase chain reaction owing to its stability and rea- sonable cost [62,155,156].

Life is believed to have originated in a hydrothermal environment, and chemoevolution began in volcanoes through a transition metal-catalyzed autocatalytic carbon-fixation cycle [77]. Three of the six carbon-fixation pathways were found in thermophilic microorganisms, in the Archaea and Bacteria domains. For example, approximately 50 % of the bacterial cellular carbon originates from pyruvate/PEP (a gluconeogenic substrate that forms cell wall components and nucleotides), followed by acetyl-CoA (approximately 30 %), oxaloacetate (approximately 13 %), and alpha- ketoglutarate (approximately 7 %) [157]. The 3-HP cycle was found in the bacterium *Chloroflexus aurantiacus*, known as a thermophilic filamentous anoxygenic photoheterotrophic bacterium, that was initially isolated in hot springs and found to develop at temperatures between 45 °C and 75 °C [158,159].

The carbon-fixation pathways HP/HB and DC/HB have been found in thermophilic Archaea. The HP/HB cycle was found in the Archaea *Metallosphaera sedula*, which grows autotrophically at 73 °C [160], and the DC/HB cycle was reported for the hyperthermophilic Archaea *Ignicoccus hospitalis* [79,161]. *Ignicoccus* was isolated from the submarine hydrothermal ventilation systems, which are anaerobic, hyperthermophilic, always chemolithoautotrophic, and exhibit an optimal growth temperature of 90 °C [162].

In contrast, the CBB cycle occurs in several thermophiles, but never in hyperthermophiles, showing a temperature limit between approximately 70 °C and 75 °C [8]. This temperature limit of the cycle can be explained by the heat-induced instabilities of some intermediates in the cycle. Such instabilities are mainly observed in glyceraldehyde-3-phosphate, which produces toxic methylglyoxal at high temperatures [163]. The thermophilic bacteria *Thermosulfidibacter takaii* ABI70S6T showed a greatly efficient and reversible citrate synthase that needs reduced ferredoxin in the pathway [12,164]. Until this discovery, it was believed that the CBB cycle can be reversed only to cleave citrate and fix $CO_2$ autotrophically, however, this can be achieved only with alternative enzymes (e.g., citrate lyase). According to the authors, this question cannot be answered by metagenomics approach, but classical biochemistry can fill the gaps between genome sequences and organism phenotypes.

Although most species that use the rTCA cycle are mesophilic, the representatives of Aquificae are thermophiles that grow best at ≥ 70 °C, whereas *Aquifex aeolicus* grows best at temperatures up to 95 °C [23]. *Hydrogenobacter thermophilus*, a member of the phylum Aquificae, invests additional ATP in the conversion of 2- oxoglutarate to form isocitrate by the combination of an irreversible biotin-dependent 2-oxoglutarate carboxylase and a noncarboxylating isocitrate dehydrogenase [165]. This process possibly becomes effective and irreversible at high temperatures [23]. A recent fixation cycle describing the roTCA cycle was discovered in a thermophilic Deltaproteobacterium, *Desulfurella acetivorans* (a sulfur-reducing thermophile showing optimum growth at 52 °C-57 °C), through metagenomic analysis. In addition to the carbon-fixation pathways, the key enzymes involved in these routes have also been described in these thermophiles. Aoshima et al. [166] reported a new key enzyme, called the citryl-CoA synthetase, which was isolated from *Hydrogenobacter thermophilus*. This enzyme, which is involved in the rTCA cycle, could reportedly catalyze the first citrate cleavage step.

Carboxydotrophic bacteria are also found in geothermal environments, suggesting that CO-oxidizing Bacteria and Archaea per- form an important role in thermophilic environments [167]. For example, acetogens and methanogens can survive in thermophilic systems and use

CO as a carbon source [168]. In the WL pathway, CO is a key intermediate in the fixation of $CO_2$ into acetyl-CoA in acetogens, methanogens, and sulfate-reducing bacteria [169,170]. Recently, a novel Archaea phylum from terrestrial hot spring and deep-sea hydrothermal vent sediments that presented unique and versatile carbon cycling pathways was described [171]. Metagenome-assembled genomes were reconstructed from 15 Archaea genomes, and a new phylum classified as Brockarcheota was proposed. The Brockarchaeota are capable of mediating non- methanogenic anaerobic methylotrophs via the tetrahydrofolate methyl branch of the WL pathway and the reductive glycine path- way. The members of this unique phylum exhibit a type of anaerobic methylotrophic metabolism that has not been described previously in Archaea. The methyltransferase process is composed of two key steps: first, specific methyltransferases break the C-R bonds in several substrates (e.g., methanol and trimethylamine) and transfer the methyl moieties to the subunit MtaC; then, the methyltransferase (MtaA for methanol and MtbA for methylamines) transfers the methyl group from the corrinoid protein to coen- zyme M in the methanogens or to tetrahydrofolate in the acetogens [171,172].

Brockarchaeota lack the key enzymes from the methyl and carbonyl branches of the WL pathway, which are involved in the transfer and reduction of the C1 fraction for methane production. Furthermore, Brockarchaeota do not encode the pyrroloquinoline quinone-linked methanol dehydrogenase pathway for aerobic methylotrophic reactions. This lack of methylotrophic pathways suggests that the new phylum metabolizes methanol and trimethylamine through the convergent action of the tetrahydrofolate (H4F) methyl branches of the WL and glycine-reduction path- ways [171].

The phylum Brockarchaeota also presents an unknown pathway for butanol metabolism, since key enzymes involved in the fer- mentation of pyruvate to butanol are absent in their genomes [171]. Nevertheless, most of the 15 genomes encode a putative aldehyde dehydrogenase that can convert butyraldehyde to butyric acid. In addition, a putative enoyl-CoA hydratase/isomerase protein has been found to be involved in the further conversion of butyric acid to acetyl-CoA, which suggests an alternative pathway for the oxidation of butanol. Members of this phylum can also perform the following activities: (i) degrade complex carbon com- pounds such as xylan, (ii) assimilate formaldehyde and formate. (iii) use arsenate as an electron acceptor, (iv) possibly reduce sulfur during fermentative growth and produce H2S, and (v) possibly use group 3b [NiFe]-hydrogenases for the oxidation of H2 with NADP or NAD(P) as the electron acceptors [171–173].

In the early 1990 s, Gadkari et al. [174] described a new species of Streptomyces, which

was a thermophilic-carboxidotrophic-chemolithoautotrophic microbe isolated from soil covering a burning charcoal pile and named it *Streptomyces thermoautrotophicus* UBT1. In contrast to the other carboxidotropics, this strain is unique in its exclusive use of a lithotrophic substrate with a new CO dehydrogenase. Several decades later, MacKellar et al. [175] described a new strain of *Streptomyces thermoautrotophicus* (strain H1), which was isolated from another burning charcoal pile; furthermore, multiple CO dehydrogenase gene clusters were identified in the genomes of the strains UBT1 and H1 [176].

Recently, Volpiano et al. [176] proposed a new reclassification of the species and genus *Streptomyces thermoautrotophicus*, which was modified to *Carbonactinospora thermoautotrophica* based on its phylogenetic placement and distinctive phenotypes among Actinomycetes. Multiple genes related to carbon metabolism were found in this species, including ribulose bisphosphate carboxylase (rbcL), phosphoenolpyruvate carboxykinase, glucose/mannose-6-phosphate isomerase and PFK 6-phosphofructokinase. Thus, more investigations are necessary to identify the carbon path used by these thermophilic strains [176].

These findings indicate that thermophilic microorganisms are important for identifying new ways of fixing carbon. Thus, it is crucial to identify microorganisms that can participate in the global carbon cycle and provide alternatives for applications in various sectors of biotechnology.

## Challenges and perspectives

The main challenge underlying the manipulation and modification of carbon-fixation routes is the insufficient understanding of how the pathways perform together within the studied organism. The use of fossil fuel resources is one of the main causes of global warming, being strictly related to the increase of atmospheric $CO_2$ concentration [110]. Natural photosynthesis can convert approximately 100 billion tons of $CO_2$ into biomass annually; however, RuBisCO, which is the most abundant enzyme on Earth, itself has a limited carboxylation rate, indicating an extremely low productivity of this pathway [126]. Synthetic biologists have made numerous efforts to overcome these challenges and improve the specificity of RuBisCO. Natural carbon-fixation processes are often slow, increasing the difficulty in designing genetically modified autotrophic plants and organisms and improving native $CO_2$ fixation [4,110].

Since the discovery of other natural pathways of carbon fixation besides the well-known CBB cycle, studies have attempted to identify the existence of other pathways in organisms that were not previously thought to fix carbon. This has led to a better understanding of the

functioning of the pathways and the catalytic power of key enzymes related to carbon fixation. In addition, these findings have raised the question of increasing the efficiency of already known pathways as well as redesigning new pathways of carbon fixation to naturally reduce the amount of anthropogenically emitted $CO_2$.

The global objective is to achieve liquid zero carbon emissions by the year 2050 [177]. This emitted carbon can be fixed, stored, and converted into sustainable feedstock. The present review aimed to elaborate the natural carbon pathways, as well as high- light the synthetic pathways that have been developed over the last decades. Given the need for shedding light on these two sub- jects, we collected and centralized the available data on the recent research in this field. In recent years, different pathways have emerged to improve carbon-fixation efficiency, demonstrating the importance of this discussion. Furthermore, the most important step will be to apply these efficient pathways in hosts and make them functional.

Based on the recent progress made in this field, certain problems related to biological carbon fixation have been identified [4]. Due to the low activities of the key carbon-fixing enzymes, there are limitations on natural carbon-fixation pathways. How- ever, the rapid advancements in microbiome studies and their related ‘‘omics” analyses have generated large amounts of data, from which novel and more efficient carbon-fixing enzymes can be identified. Natural carbon-fixation pathways have several short comings, with many reaction steps and low efficiencies. However, the development of synthetic biology permits the exploration of new natural carbon-fixation pathways and the design of novel synthetic carbon-fixation techniques. Another barrier is the low solar energy-absorption efficiency of carbon fixation, which significantly reduces the economic benefits. These limitations can be addressed by converting light to other sources such as electric energy and hydrogen for direct carbon fixation. Chemoautotrophs provide an alternative platform because they can be driven chemically by renewable H2 instead of light [103]. Electrical energy can also be applied to produce energy carriers like formate, hydrogen, carbon monoxide, methanol, and methane [178]. Li et al. [179] described *Ralstonia eutropha* H16, an autotrophic genetically engineered microorganism, that produced high levels of alcohol in an electric bioreactor using $CO_2$ as the sole carbon source and electricity as the sole energy input. Another approach could be to improve light-utilization efficiency. Recently, Nürnberg et al. [180] discovered specific photosystems I and II in *Chroococcidiopsis thermalis* that can use energy from far-infrared radiation.

Due to the low efficiencies and limited scope for performing genetic modifications of the natural carbon-fixation systems, various researchers are exploring these systems using different ways; for example, introducing partial or total carbon-fixation pathways in

heterotrophic chassis cells. Generally, heterotrophic microorganisms exhibit superior growth and production yields than autotrophic microorganisms [17]. The significant improvements in synthetic biology permit the engineering of novel functions and metabolic networks in heterotrophic microorganisms for innumerous biotechnological applications [178–182]. Novel genes mediating the $CO_2$ fixation pathway in autotrophic microorganisms may apply negative influences on the regulatory networks in cells, such as carbon/nitrogen metabolism; however, they have been widely used as chassis to express natural $CO_2$-fixation pathways in the model heterotrophic microbes Escherichia coli and Saccharomyces cerevisiae [183,184].

In the near future, novel enzymes and entire pathways of carbon fixation are expected to be identified with rapid development of ‘‘omics” tools and microbiome projects [4]. These advancements will facilitate the development of feasible methods for changing and/or increasing the production of the desired product. Moreover, the use and availability of ‘‘omics” tools have progressed rapidly, highlighting the limitations of single techniques in dealing with the complexities of biological systems [185].

One of the future possibilities is the integration of data obtained using the ‘‘omics” tools with studies on metabolism and kinetics, classical biochemistry, and molecular biological techniques. In addition, the development of mathematical models to analyze biological processes on a micro-scale will allow us to predict their effects on a macro scale. In the near future, these suggested methods may facilitate the exploration, modeling, manipulation, and engineering of metabolic pathways in natural and synthetic microbial systems. According to Bar-Even et al. [115] and Fuchs [7], the number of known autotrophic pathways is expected to increase with the aid of genome searching and the application of synthetic biological tools. Notably, ‘‘omics” analyses have revolutionized the research on carbon fixation and unveiled the genomic black box. Nevertheless, for better results, these analyses must be linked to bench tests and classical biochemical analyses to obtain detailed and robust results because some resources may not be identified solely with bioinformatics tools.

Although carbon-fixation pathways have not yet been used at the industry level owing to the low catalytic efficiencies of the enzymes and the need for a significant energy input, studies of these routes are anticipated to provide a plethora of possibilities for biotechnological advances. The major challenge, i.e., the joining of carbon-fixation pathways with synthetic pathways due to the differences between the metabolism of autotrophs and heterotrophs that are used as chassis, can be resolved. Carbon-fixation pathways based on synthetic biology have already been described, and the energy supply and more efficient enzymes can be rede- signed in these pathways [186].

## Conclusion (Remarks)

Carbon is often negatively associated with the anthropogenic emissions of greenhouse gasses; however, it is also an element of paramount importance for life and ecological processes. The knowledge of carbon-fixation cycles, in addition to ecological and evolutionary studies on the emergence of the first beings that inhabited Earth and metabolized carbon, can provide a detailed understanding of the enzymes and important intermediaries involved in these cycles.

The central carbon-metabolizing enzymes can be used for a variety of purposes; for example, artificial $CO_2$ fixation pathways can be designed and implemented in suitable host organisms. Additionally, new solutions are possible for manipulating carbon- fixation pathways and increasing their final yields. The multidisciplinary integration of ‘‘omics,” synthetic biology, molecular biology, and biochemistry can enable the scientific community to describe new carbon-fixation pathways and predict important routes and enzymes involved in these processes. This integration may provide further guidance for overcoming the limitations of natural carbon-fixation pathways and allow the development of biotechnological applications.

Nevertheless, investigation of thermophilic microorganisms is crucial for understanding many evolutionary events. Overall, these microorganisms have been extensively investigated for developing biotechnological applications because of their metabolic capabilities; however, thermophiles remain underexplored with regard to the identification of novel natural routes of carbon fixation and the development of efficient synthetic pathways.

## Funding

This study was funded by KAUST (BAS/1/1096-01-01). SSC was awarded scholarships from the Foundation for Research Support of the State of Rio de Janeiro (FAPERJ) and Coordination for the Improvement of Higher Education Personnel (CAPES).

## Compliance with Ethics Requirements

This article does not contain any studies with human or animal subjects.

## CRediT authorship contribution statement

Sulamita Santos Correa: Conceptualization, Methodology, Writing-original draft, Writing-review & editing. Junia Schultz: Data curation, Writing-original draft, Writing-review & editing. Kyle J. Lauersen: Validation, Writing-review & editing. Alexandre Soares Rosado: Supervision, Writing-review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We thank the Graduate Program of Plant Biotechnology and Bioprocesses (PBV) at the Federal University of Rio de Janeiro for supporting this work and Prof. Patricia Moura, Ricardo Chaloub and Alex Prast for their valuable suggestions to improve the manuscript. The graphical abstract was created by Heno Hwang, scientific illustrator at King Abdullah University of Science and Technology (KAUST).

## References

[1] Aversa R, Petrescu RV, Apicella A, Petrescu FI. The basic elements of life's. American J E and A S 2016;9(4):1189–97.
[2] Loeve, S., & Vincent, B. B. The multiple signatures of carbon. In *Res. Obj. in their Tech. Setting* (pp. 185-200). (2017).
[3] Anthony MA et al. Distinct assembly processes and microbial communities constrain soil organic carbon formation. One earth 2020;2(4):349–60.
[4] Gong F, Zhu H, Zhang Y, Li Y. Biological carbon fixation: from natural to synthetic. J of $CO_2$ Uti 2018;28:221–7.
[5] Siegel JB, Smith AL, Poust S, Wargacki AJ, Bar-Even A, Louw C, et al. Computational protein design enables a novel one-carbon assimilation pathway. Proc of the Nat Aca of Scie 2015;112(12):3704–9.
[6] Lu X, Liu Y, Yang Y, Wang S, Wang Q, Wang X, et al. Constructing a synthetic pathway for acetyl-coenzyme A from one-carbon through enzyme design. Natu Comm 2019;10(1).
[7] Fuchs G. Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? Annual Rev of Microb 2011;65:631–58.
[8] Hugler M, Sievert SM. Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. Ann Rev of Mar Scie 2011;3:261–89.
[9] Ljungdhal LG. The autotrophic pathway of acetate synthesis in acetogenic bacteria. Ann Rev of Microb 1986;40(1):415–50.
[10] Bar-Even A, Noor E, Milo R. A survey of carbon fixation pathways through a quantitative lens. J of Expe Bot 2012;63(6):2325–42.
[11] Kono T, Mehrotra S, Endo C, Kizu N, Matusda M, Kimura H, et al. A RuBisCO- mediated carbon metabolic pathway in methanogenic Archaea. Natu comm 2017;8(1):1–12.
[12] Mall A et al. Reversibility of citrate synthase allows autotrophic growth of a thermophilic bacterium. Scien 2018;359(6375):563–7.
[13] Kruse O, Rupprecht J, Mussgnug JH, Dismukes GC, Hankamer B. Photosynthesis: a blueprint for solar energy capture and biohydrogen production technologies. Photo & Photobi Scien 2005;4(12):957–70.
[14] Sciuto K, Moro I. Cyanobacteria: the bright and dark sides of a charminuteg group. Biodi and Conse 2015;24(4):711–38.
[15] Liu L, Brown PC, Könneke M, Huber H, König S, Berg IA. Convergent evolution of a promiscuous 3-hydroxypropionyl-CoA dehydratase/crotonyl-CoA hydratase in crenarchaeota and thaumarchaeota. Msphere 2021;6(1): e01079–10120.
[16] De Souza YPA, Rosado AS. Opening the Black Box of Thermophilic Autotrophic Bacterial Diversity. In: Microbial Diversity in the Genomic Era. Acad. Press; 2019. p. 333–43.
[17] Liang B, Zhao Y, Yang J. Recent Advances in Developing Artificial Autotrophic Microorganism for Reinforcing $CO_2$ Fixation. Front in Microb 2020;11:2848.
[18] Engel, A. S. Chemolithoautotrophy. *In Encyclopedia of Caves, Acad. Press,* 267-276. (2019).
[19] Irfan M, Bai Y, Zhou L, Kazmi M, Yuan S, Mbadinga SM, et al. Direct microbial transformation of carbon dioxide to value-added chemicals: a comprehensive analysis and application potentials. Bior Tech 2019;288:121401.

[20] Sysoev M, Grötzinger SW, Renn D, Eppinger J, Rueping M, Karan R. Bioprospecting of novel extremozymes from prokaryotes - The advent of culture-independent methods. Front in Microb 2021;12.
[21] Atsumi S, Higashide W, Liao JC. Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. Natu Biotec 2009;27(12):1177–80.
[22] Durall C, Lindblad P. Mechanisms of carbon fixation and engineering for increased carbon fixation in cyanobacteria. Algal Res 2015;11:263–70.
[23] Berg IA. Ecological aspects of the distribution of different autotrophic $CO_2$ fixation pathways. Appl and Envir Micro 2011;77(6):1925–36.
[24] Bjornsson L, Hugenholtz P, Tyson GW, Blackall LL. Filamentous Chloroflexi (green non-sulfur bacteria) are abundant in wastewater treatment processes with biological nutrient removal. Micro 2002;148(8):2309–18.
[25] Frolov EN, Kublanov IV, Toshchakov SV, Lunev EA, Pimenov NV, Bonch- Osmolovskaya EA, et al. Form III RuBisCO-mediated transaldolase variant of the Calvin cycle in a chemolithoautotrophic bacterium. Proc of the Nat Aca of Scien 2019;116(37):18638–46.
[26] Wang X, Li W, Xiao Y, Cheng A, Shen T, Zhu M, et al. Abundance and diversity of carbon-fixing bacterial communities in karst wetland soil ecosystems. Catena 2021;204:105418.
[27] Kusian B, Bowien B. Organization and regulation of cbb $CO_2$ assimilation genes in autotrophic bacteria. FEMS Micro Rev 1997;21(2):135–55.
[28] Gurrieri L, Fermani S, Zaffagnini M, Sparla F, Trost P. Calvin-Benson cycle regulation is getting complex. Trends Plant Sci 2021;26(9):898–912.
[29] Launay H, Shao H, Bornet O, Cantrelle FX, Lebrun R, Receveur-Brechot V, et al. Flexibility of Oxidized and Reduced States of the Chloroplast Regulatory Protein CP12 in Isolation and in Cell Extracts. Biomolecules 2021;11(5):701.
[30] Shao H, Huang W, Avilan L, Receveur-Bréchot V, Puppo C, Puppo R, et al. A new type of flexible CP12 protein in the marine diatom Thalassiosira pseudonana. Cell Communication and Signaling 2021;19(1):1–13.
[31] Ohta J. A novel variant of the Calvin-Benson cycle bypassing fructose bisphosphate. Sci Rep 2022;12(1):1–8.
[32] Sharwood RE. Engineering chloroplasts to improve RuBisCO catalysis: prospects for translating improvements into food and fiber crops. New Phyto 2017;213(2):494–510.
[33] Sharkey TD, Weise SE. The glucose 6-phosphate shunt around the Calvin- Benson cycle. J Exp Bot 2016;67(14):4067–77.
[34] Sharkey TD. Pentose phosphate pathway reactions in photosynthesizing cells. Cells 2021;10(6):1547.
[35] Cleland WW, Andrews TJ, Gutteridge S, Hartman FC, Lorimer GH. Mechanism of RuBisCO: the carbamate as general base. Chem Rev 1998;98(2): 549–62.
[36] Ducat DC, Silver PA. Improving carbon fixation pathways. Cur Opi in Chem Biol 2012;16(3–4):337–44.
[37] Caldwell PE, MacLean MR, Norris PR. Ribulose bisphosphate carboxylase activity and a Calvin cycle gene cluster in Sulfobacillus species. Micro 2007;153(7):2231–40.
[38] Yeates TO, Kerfeld CA, Heinhorst S, Cannon GC, Shively JM. Protein-based organelles in bacteria: carboxysomes and related microcompartments. Natu Rev Micro 2008;6(9):681–91.
[39] Savage DF, Afonso B, Chen AH, Silver PA. Spatially ordered dynamics of the bacterial carbon fixation machinery. Scien 2010;327(5970):1258–61.
[40] Long BM, Hee WY, Sharwood RE, Rae BD, Kaines S, Lim Y-L, et al. Carboxysome encapsulation of the CO 2-fixing enzyme RuBisCO in tobacco chloroplasts. Natu Comm 2018;9(1).
[41] Mackinder LCM, Chen C, Leib RD, Patena W, Blum SR, Rodman M, et al. A spatial interactome reveals the protein organization of the algal CO2- concentrating mechanism. Cell 2017;171(1):133–147.e14.
[42] Hill NC, Tay JW, Altus S, Bortz DM, Cameron JC. Life cycle of a cyanobacterial carboxysome. Scien Adv 2020;6(19):eaba1269.
[43] Tabita FR, Satagopan S, Hanson TE, Kreel NE, Scott SS. Distinct form I, II, III, and IV RuBisCO proteins from the three kingdoms of life provide clues about RuBisCO evolution and structure/function relationships. J of Expe Bot 2008;59(7):1515–24.
[44] Wilkes EB, Carter SJ, Pearson A. CO2-dependent carbon isotope fractionation in the dinoflagellate *Alexandrium tamarense*. Geoch et Cosm Acta 2017;212:48–61.
[45] Morse DP, Salois P, Markovic JWH. A nuclear-encoded form II RuBisCO in dinoflagellates. Science 1995;268:1622–4.
[46] Rowan RSM, Whitney A, Fowler DY. Rubisco in marine symbiotic dinoflagellates: form II

enzymes in eukaryotic oxygenic phototrophs encoded by a nuclear multigene family. Plant *Cell* 1996;8:539–53.
[47] Banda DM, Pereira JH, Liu AK, Orr DJ, Hammel M, He C, et al. Novel bacterial clade reveals origin of form I RuBisCO. Natu Plan 2020;6(9):1158–66.
[48] Whitney SM, Andrews TJ. The gene for the ribulose-1, 5-bisphosphate carboxylase/oxygenase (RuBisCO) small subunit relocated to the plastid genome of tobacco directs the synthesis of small subunits that assemble into RuBisCO. Plant Cell 2001;13(1):193–205.
[49] Yoshida S, Atomi H, Imanaka T. Engineering of a type III RuBisCO from a hyperthermophilic archaeon in order to enhance catalytic performance in mesophilic host cells. Appl and Enviro Micro 2007;73(19):6254–61.
[50] Aono R, Sato T, Imanaka T, Atomi H. A pentose bisphosphate pathway for nucleoside degradation in Archaea. Natu Chem Bio 2015;11(5):355–60.Wrighton KC et al. RuBisCO of a nucleoside pathway known from Archaea is found in diverse uncultivated phyla in bacteria. The ISME J 2016;10 (11):2702–14.
[51] Tabita FR, Hanson TE, Li H, Satagopan S, Singh J, Chan S. Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs. Microbiol Mol Biol Rev 2007;71(4):576–99.
[52] Kim SM, Lim HS, Lee SB. Discovery of a RuBisCO -like Protein that Functions as an Oxygenase in the Novel D-Hamamelose Pathway. Biotec and Biopr Eng 2018;23(5):490–9.
[53] Evans MC, Buchanan BB, Arnon DI. A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. *Proce. of the Nat.* Aca of Scien of the Unit Sta of Amer 1996;55(4):928.
[54] Buchanan BB, Sirevåg R, Fuchs G, Ivanovsky RN, Igarashi Y, Ishii M, et al. The Arnon-Buchanan cycle: a retrospective, 1966–2016. Photo Res 2017;134 (2):117–31.
[55] Kitadai N, Kameya M, Fujishima K. Origin of the reductive tricarboxylic acid (rTCA) cycle-type $CO_2$ fixation: a perspective. Life 2017;7(4):39.
[56] Kluger R, Tittmann K. Thiaminute diphosphate catalysis: enzymic and nonenzymic covalent intermediates. Chem Rev 2008;108(6):1797–833.
[57] Fuchs G. Alternative pathways of autotrophic carbon dioxide fixation in autotrophic bacteria. *In Bio. of Autotr. Bac.*, ed. HG Schlegel, 365-82. (1989).
[58] Ivanovsky RN, Fal YI, Berg IA, Ugolkova NV, Krasilnikova EN, Keppen OI, et al. Evidence for the presence of the reductive pentose phosphate cycle in a filamentous anoxygenic photosynthetic bacterium, Oscillochloris trichoides strain DG-6. Micro 1999;145(7):1743–8.
[59] Thauer RK. A fifth pathway of carbon fixation. Scien 2007;318(5857):1732–3.
[60] Aherfi S, Brahim Belhaouari D, Pinault L, Baudoin JP, Decloquement P, Abrahao J, et al. Incomplete tricarboxylic acid cycle and proton gradient in Pandoravirus massiliensis: is it still a virus? The ISME journal 2022;16 (3):695–704.
[61] Brock TD. The value of basic research: discovery of Thermus aquaticus and other extreme thermophiles. Genet 1997;146(4):1207.
[62] Adam PS, Borrel G, Gribaldo S. An Archaeal origin of the Wood-Ljungdahl H 4 MPT branch and the emergence of bacterial methylotrophy. Natu Micro 2019;4(12):2155–63.
[63] Youssef NH, Farag IF, Rudy S, Mulliner A, Walker K, Caldwell F, et al. The Wood-Ljungdahl pathway as a key component of metabolic versatility in candidate phylum Bipolaricaulota (Acetothermia, OP1). Enviro Micro Rep 2019;11(4):538–47.
[64] Drake HL, Daniel SL, Küsel K, Matthies C, Kuhner C, Braus-Stromeyer S. Acetogenic bacteria: what are the *in situ* consequences of their diverse metabolic versatilities? Biofac 1997;6(1):13–24.
[65] Shafiee S, Topal E. When will fossil fuel reserves be diminuteished? Ener Poli 2009;37(1):181–9.
[66] Roberts JR, Lu WP, Ragsdale SW. Acetyl-coenzyme A synthesis from methyltetrahydrofolate, CO, and coenzyme A by enzymes purified from Clostridium thermoaceticum: attainment of in vivo rates and identification of rate-limiting steps. J of Bac 1992;174(14):4667–76.
[67] Schöne, C., Poehlein, A., Jehmlich, N., Adlung, N., Daniel, R., von Bergen, M., ... & Rother, M. Deconstructing Methanosarcina acetivorans into an acetogenic archaeon. *Proceedings of the National Academy of Sciences*, 119(2), e2113853119. (2022).
[68] Orsi WD, Vuilleminute A, Rodriguez P, Coskun ÖK, Gomez-Saez GV, Lavik G, et al. Metabolic activity analyses demonstrate that Lokiarchaeon exhibits homoacetogenesis in sulfidic marine sediments. Nat Microbiol 2020;5 (2):248–55.
[69] Biegel E, Müller V. Bacterial Na+-translocating ferredoxin: NAD+ oxidoreductase. *Proce. of the Nat.* Aca of Scien 2010;107(42):18138–42.
[70] Borrel G, Adam PS, Gribaldo S. Methanogenesis and the Wood-Ljungdahl pathway: an ancient, versatile, and fragile association. Geno Bio and Evo 2016;8(6):1706–11.
[71] Holo H. *Chloroflexus aurantiacus* secretes 3-hydroxypropionate, a possible intermediate in the

assimilation of $CO_2$ and acetate. Arch of Micro 1989;151 (3):252–6.
[72] Zarzycki J, Brecht V, Müller M, Fuchs G. Identifying the missing steps of the autotrophic 3-hydroxypropionate $CO_2$ fixation cycle in *Chloroflexus aurantiacus*. Proce of the Nat Aca of Scie 2009;106(50):21317–22.
[73] Shih PM, Ward LM, Fischer WW. Evolution of the 3-hydroxypropionate bicycle and recent transfer of anoxygenic photosynthesis into the *Chloroflexi*. Proce of the Nat Aca of Scien 2017;114(40):10749–54.
[74] Ward LM, Shih PM. The evolution and productivity of carbon fixation pathways in response to changes in oxygen concentration over geological time. Free Rad Bio and Med 2019;140:188–99.
[75] Zarzycki J, Fuchs G. Coassimilation of organic substrates via the autotrophic 3-hydroxypropionate bi-cycle in Chloroflexus aurantiacus. Appl Environ Microbiol 2011;77(17):6181–8.
[76] Berg IA, Ramos-Vera WH, Petri A, Huber H, Fuchs G. Study of the distribution of autotrophic $CO_2$ fixation cycles in Crenarchaeota. Micro 2010;156 (1):256–69.
[77] Berg IA, Kockelkorn D, Buckel W, Fuchs G. A 3-hydroxypropionate/4- hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. Scien 2007;318(5857):1782–6.
[78] Huber H, Gallenberger M, Jahn U, Eylert E, Berg IA, Kockelkorn D, et al. A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum Ignicoccus hospitalis. Proc of the Nat Acad of Scien 2008;105(22):7851–6.
[79] Flechsler, J., Heimerl, T., Huber, H., Rachel, R., & Berg, I. A. Functional compartmentalization and metabolic separation in a prokaryotic cell. *Proceedings of the National Academy of Sciences*, 118(25), e2022114118. (2021).
[80] Hawkins AS, Han Y, Bennett RK, Adams MW, Kelly RM. Role of 4- hydroxybutyrate-CoA synthetase in the CO2 fixation cycle in thermoacidophilic Archaea. J of Bio Chem 2013;288(6):4012–22.
[81] Auernik KS, Kelly RM. Physiological versatility of the extremely thermoacidophilic archaeon Metallosphaera sedula supported by transcriptomic analysis of heterotrophic, autotrophic, and mixotrophic growth. Appl and Enviro Micro 2010;76(3):931–5.
[82] Estelmann S, Hu ̈gler M, Eisenreich W, Werner K, Berg IA, Ramos-Vera WH, et al. Labeling and enzyme studies of the central carbon metabolism in Metallosphaera sedula. J of Bac 2011;193(5):1191–200.
[83] Berg IA, Liu L, Schubert D, Koenneke M. (S)-3-Hydroxybutyryl-CoA dehydrogenase from the autotrophic 3-hydroxypropionate/4-hydroxybutyrate cycle in Nitrosopumilus maritimus. Front Microbiol 2021;12:1846.
[84] Rosgaard L, de Porcellinis AJ, Jacobsen JH, Frigaard NU, Sakuragi Y. Bioengineering of carbon fixation, biofuels, and biochemicals in cyanobacteria and plants. J of Biot 2012;162(1):134–47.
[85] Soderberg T. Biosynthesis of ribose-5-phosphate and erythrose-4-phosphate in archaea: a phylogenetic analysis of archaeal genomes. Arch 2005;1(5): 347–52.
[86] White III RA, Rivas-Ubach A, Borkum MI, Köberl M, Bilbao A, Colby SM, et al. The state of rhizospheric science in the era of multi-omics: A practical guide to omics technologies. Rhizo 2017;3:212–21.
[87] Palazzotto E, Weber T. Omics and multi-omics approaches to study the biosynthesis of secondary metabolites in microorganisms. Cur Opi in Micro 2018;45:109–16.
[88] Murugadas V, Prasad MM. Bioinformatics in Microbial systematics. Tec: ICAR- Cen. Inst. of Fish; 2018.
[89] Christel S, Herold M, Bellenberg S, El Hajjami M, Buetti-Dinh A, Pivkin IV, et al. Multi-omics reveals the lifestyle of the acidophilic, mineral-oxidizing model species *Leptospirillum ferriphilum* T. Appl and Enviro Micro 2017;84(3): e02091–10117.
[90] McAllister SM, Polson SW, Butterfield DA, Glazer BT, Sylvan JB, Chan CS. Validating the Cyc2 neutrophilic iron oxidation pathway using meta-omics of Zetaproteobacteria iron mats at marine hydrothermal vents. Msys 2020;5(1): e00553–e619.
[91] Figueroa IA, Barnum TP, Somasekhar PY, Carlström CI, Engelbrektson AL, Coates JD. Metagenomics-guided analysis of microbial chemolithoautotrophic phosphite oxidation yields evidence of a seventh natural $CO_2$ fixation pathway. Proc of the Nat Acad of Scien 2018;115 (1):92–101.
[92] Cotton CA, Edlich-Muth C, Bar-Even A. Reinforcing carbon fixation: CO2 reduction replacing and supporting carboxylation. Curr Opi in Biote 2018;49:49–56.
[93] Schink B, Thiemann V, Laue H, Friedrich MW. *Desulfotignum phosphitoxidans* sp. nov., a new marine sulfate reducer that oxidizes phosphite to phosphate. Arch of Micro 2002;177(5):381–91.
[94] Sánchez-Andrea I, Guedes IA, Hornung B, Boeren S, Lawson CE, Sousa DZ, et al. The reductive

glycine pathway allows autotrophic growth of *Desulfovibrio desulfuricans*. Natu Comm 2020;11(1):1–12.
[95] Hao L, Michaelsen TY, Singleton CM, Dottorini G, Kirkegaard RH, Albertsen M, et al. Novel syntrophic bacteria in full-scale anaerobic digesters revealed by genome-centric metatranscriptomics. *The*. ISME J 2020;14(4):906–18.
[96] Yishai O, Bouzon M, Döring V, Bar-Even A. In vivo assimilation of one-carbon via a synthetic reductive glycine pathway in Escherichia coli. ACS Synth Biol 2018;7(9):2023–8.
[97] Tashiro Y, Hirano S, Matson MM, Atsumi S, Kondo A. Electrical-biological hybrid system for CO2 reduction. Metab Eng 2018;47:211–8.
[98] Kim S, Lindner SN, Aslan S, Yishai O, Wenk S, Schann K, et al. Growth of E. coli on formate and methanol via the reductive glycine pathway. Nat Chem Biol 2020;16(5):538–45.
[99] Bonch-Osmolovskaya EA, Sokolova TG, Kostrikina NA, Zavarzin GA. Desulfurella acetivorans gen. nov. and sp. nov. a new thermophilic sulfur- reducing eubacterium. Arch of Micro 1990;153(2):151–5.
[100] Pradella S, Hippe H, Stackebrandt E. Macrorestriction analysis of *Desulfurella acetivorans* and *Desulfurella multipotens*. FEMS Micro Lett 1998;159 (1):137–44.
[101] Zhang T, Shi XC, Ding R, Xu K, Tremblay PL. The hidden chemolithoautotrophic metabolism of *Geobacter sulfurreducens* uncovered by adaptation to formate. The ISME J 2020;14(8):2078–89.
[102] Thevasundaram, K., Gallagher, J. J., Cherng, F., & Chang, M. C. Engineering nonphotosynthetic carbon fixation for production of bioplastics by methanogenic archaea. *Proceedings of the National Academy of Sciences*, 119 (23), e2118638119. (2022).
[103] Schwander T, von Borzyskowski LS, Burgener S, Cortina NS, Erb TJ. A synthetic pathway for the fixation of carbon dioxide in vitro. Scie 2016;354 (6314):900–4.
[104] Erb TJ, Berg IA, Brecht V, Müller M, Fuchs G, Alber BE. Synthesis of C5- dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/ reductase: the ethylmalonyl-CoA pathway. Proce of the Nat Acad of Scien 2007;104(25):10631–6.
[105] Peter DM, Schada von Borzyskowski L, Kiefer P, Christen P, Vorholt JA, Erb TJ. Screening and engineering the synthetic potential of carboxylating reductases from central metabolism and polyketide biosynthesis. Ange Chem Inter Ed 2015;54(45):13457–61.
[106] Erb TJ, Brecht V, Fuchs G, Müller M, Alber BE. Carboxylation mechanism and stereochemistry of crotonyl-CoA carboxylase/reductase, a carboxylating enoyl-thioester reductase. Proce of the Nat Acad of Scien 2009;106 (22):8871–6.
[107] Gong F, Li Y. Fixing carbon, unnaturally. Scien 2016;354(6314):830–1.
[108] Naduthodi MIS, Claassens NJ, D'Adamo S, van der Oost J, Barbosa MJ. Synthetic biology approaches to enhance microalgal productivity. Trends Biotechnol 2021;39(10):1019–36.
[109] Bar-Even A. Formate assimilation: the metabolic architecture of natural and synthetic pathways. Bioch 2016;55(28):3851–63.
[110] Bar-Even A, Noor E, Flamholz A, Milo R. Design and analysis of metabolic pathways supporting formatotrophic growth for electricity-dependent cultivation of microbes. Bioch et Bioph Acta (BBA)-Bioene 2013;1827(8– 9):1039–47.
[111] Claassens NJ. Reductive glycine pathway: a versatile route for one-carbon biotech. Trends Biotechnol 2021;39(4):327–9.
[112] Wang X, Wang Y, Liu J, Li Q, Zhang Z, Zheng P, et al. Biological conversion of methanol by evolved Escherichia coli carrying a linear methanol assimilation pathway. Bioresources and Bioprocessing 2017;4(1):1–6.
[113] Lan EI, Liao JC. Microbial synthesis of n-butanol, isobutanol, and other higher alcohols from diverse resources. Bior Tec 2013;135:339–49.
[114] Bar-Even A, Noor E, Lewis NE, Milo R. Design and analysis of synthetic carbon fixation pathways. Proce of the Nat Acad of Scien 2010;107(19):8889–94.
[115] Li D, Huang L, Liu T, Liu J, Zhen L, Wu J, et al. Electrochemical reduction of carbon dioxide to formate via nano-prism assembled CuO microspheres. Chemo 2019;237:124527.
[116] Bang J, Ahn JH, Lee JA, Hwang CH, Kim GB, Lee J, et al. Synthetic formatotrophs for one-carbon biorefinery. Adv Sci 2021;8(12):2100199.
[117] Zhao T, Li Y, Zhang Y. Biological carbon fixation: a thermodynamic perspective. Green Chem 2021;23(20):7852–64.
[118] Yu H, Li X, Duchoud F, Chuang DS, Liao JC. Augmenting the Calvin–Benson- Bassham cycle by a synthetic malyl-CoA-glycerate carbon fixation pathway. Natu Comm 2018;9(1):1–10.
[119] Luo S, Lin PP, Nieh LY, Liao GB, Tang PW, Chen C, et al. A cell-free self- replenishing CO2-fixing system. Nat Catal 2022;5(2):154–62.

[120] Sarles LS, Tabita FR. Derepression of the synthesis of D-ribulose 1, 5- bisphosphate carboxylase/oxygenase from Rhodospirillum rubrum. J of Bac 1983;153(1):458–64.
[121] Wildman SG. Along the trail from Fraction I protein to Rubisco (ribulose bisphosphate carboxylase-o xygenase). Photo Res 2002;73(1):243–50.
[122] Parry MA, Keys AJ, Madgwick PJ, Carmo-Silva AE, Andralojc PJ. RuBisCO regulation: a role for inhibitors. J of Expe Bot 2008;59(7):1569–80.
[123] Bock R. Genetic engineering of the chloroplast: novel tools and new applications. Cur Opi in Biotec 2014;26:7–13.
[124] Cai Z, Liu G, Zhang J, Li Y. Development of an activity-directed selection system enabled significant improvement of the carboxylation efficiency of RuBisCO. Prote & cell 2014;5(7):552–62.
[125] Lin MT, Occhialini A, Andralojc PJ, Parry MA, Hanson MR. A faster RuBisCO with potential to increase photosynthesis in crops. Natu 2014;513 (7519):547–50.
[126] Hauser T, Popilka L, Hartl FU, Hayer-Hartl M. Role of auxiliary proteins in RuBisCO biogenesis and function. Natu Plan 2015;1(6):1–11.
[127] Whitney SM, Birch R, Kelso C, Beck JL, Kapralov MV. Improving recombinant RuBisCO biogenesis, plant photosynthesis and growth by coexpressing its ancillary RAF1 chaperone. Proce of the Nat Acad of Scien 2015;112 (11):3564–9.
[128] Hanson MR, Lin MT, Carmo-Silva AE, Parry MA. Towards engineering carboxysomes into C3 plants. The Plant J 2016;87(1):38–50.
[129] Long BM, Rae BD, Rolland V, Förster B, Price GD. Cyanobacterial $CO_2$- concentrating mechanism components: function and prospects for plant metabolic engineering. Cur Opi in Plant Bio 2016;31:1–8.
[130] Songstad DD, Petolino JF, Voytas DF, Reichert NA. Genome editing of plants. Crit Rev in Plant Scien 2017;36(1):1–23.
[131] Bock R. Engineering plastid genomes: methods, tools, and applications in basic research and biotechnology. Ann Rev of Plant Bio 2015;66:211–41.
[132] Conlan B, Whitney S. Preparing RuBisCO for a tune up. Natu Plant 2018;4 (1):12–3.
[133] Hu J, Nagarajan D, Zhang Q, Chang JS, Lee DJ. Heterotrophic cultivation of microalgae for pigment production: A review. Biotec Adva 2018;36(1):54–67.
[134] Canelli G, Neutsch L, Carpine R, Tevere S, Giuffrida F, Rohfritsch Z, et al. Chlorella vulgaris in a heterotrophic bioprocess: study of the lipid bioaccessibility and oxidative stability. Algal Res 2020;45:101754.
[135] Lauersen KJ, Willamme R, Coosemans N, Joris M, Kruse O, Remacle C. Peroxisomal microbodies are at the crossroads of acetate assimilation in the green microalga Chlamydomonas reinhardtii. Algal Res 2016;16:266–74.
[136] Barbier G, Oesterhelt C, Larson MD, Halgren RG, Wilkerson C, Garavito RM, et al. Comparative genomics of two closely related unicellular thermo- acidophilic red algae, *Galdieria sulphuraria* and *Cyanidioschyzon merolae*, reveals the molecular basis of the metabolic flexibility of *Galdieria sulphuraria* and significant differences in carbohydrate metabolism of both algae. Plant Physi 2005;137(2):460–74.
[137] Sloth JK, Wiebe MG, Eriksen NT. Accumulation of phycocyanin in heterotrophic and mixotrophic cultures of the acidophilic red alga *Galdieria sulphuraria*. Enzy and Micro Tech 2006;38(1–2):168–75.
[138] Fields FJ, Ostrand JT, Mayfield SP. Fed-batch mixotrophic cultivation of *Chlamydomonas reinhardtii* for high-density cultures. Algal Res 2018;33:109–17.
[139] Puzanskiy R, Shavarda A, Romanyuk D, Shishova M. The role of trophic conditions in the regulation of physiology and metabolism of *Chlamydomonas reinhardtii* during batch culturing. J of App Phyco 2021;33(5):2897–908.
[140] Hallmann A. Algae biotechnology green cell-factories on the rise. Cur Biotec 2015;4(4):389–415.
[141] Südfeld C et al. High-throughput insertional mutagenesis reveals novel targets for enhancing lipid accumulation in *Nannochloropsis oceanica*. Meta Eng 2021;66:239–58.
[142] Lauersen KJ. Eukaryotic microalgae as hosts for light-driven heterologous isoprenoid production. Plant 2019;249(1):155–80.
[143] Specht E, Miyake-Stoner S, Mayfield S. Micro-algae come of age as a platform for recombinant protein production. Biotec Lett 2010;32(10):1373–83.
[144] Freudenberg RA, Baier T, Einhaus A, Wobbe L, Kruse O. High cell density cultivation enables efficient and sustainable recombinant polyaminutee production in the microalga *Chlamydomonas reinhardtii*. Bior Tech 2021;323:124542.
[145] Donato P, Finore I, Poli A, Nicolaus B, Lama L. The production of second generation bioethanol:

The biotechnology potential of thermophilic bacteria. J of Clea Prod 2019;233:1410–7.
[146] Holden, J. F. Hot Environments. In: Schaechter M (ed) Schmidt TM. *Top. in Eco. and Enviro. Micro. Elsevier Inc*, 313-332. (2012).
[147] Straub CT, Zeldes BM, Schut GJ, Adams MW, Kelly RM. Extremely thermophilic energy metabolisms: biotechnological prospects. Cur Opi in Biotec 2017;45:104–12.
[148] Robb F, Antranikian G, Grogan D, Driessen A, editors. Thermophiles: biology and technology at high temperatures. CRC Press; 2007.
[149] Ranawat P, Rawat S. Stress response physiology of thermophiles. Arch of Micro 2017;199(3):391–414.
[150] Schultz J, Rosado AS. Extreme environments: a source of biosurfactants for biotechnological applications. Extremo 2020;24(2):189–206.
[151] Urbieta MS, Donati ER, Chan KG, Shahar S, Sin LL, Goh KM. Thermophiles in the genomic era: biodiversity, science, and applications. Biotech Adv 2015;33 (6):633–47.
[152] Jorquera MA, Graether SP, Maruyama F. Bioprospecting and biotechnology of extremophiles. Front in Bioen and Biotec 2019;7:204.
[153] Sarmiento F, Peralta R, Blamey JM. Cold and hot extremozymes: industrial relevance and current trends. Front in Bioen and Biotec 2015;3:148.
[154] Bruins ME, Janssen AE, Boom RM. Thermozymes and their applications. Appl Bioch and biotechno 2001;90(2):155–86.
[155] Irwin JA, Baird AW. Extremophiles and their application to veterinary medicine. Irish Vet J 2004;57(6):1–7.
[156] Alber BE. Biotechnological potential of the ethylmalonyl-CoA pathway. Appl Micro and Biotec 2011;89(1):17–25.
[157] Pierson BK, Castenholz RW. A phototrophic gliding filamentous bacterium of hot springs, *Chloroflexus aurantiacus*, gen. and sp. nov. Arch of Micro 1974;100(1):5–24.
[158] Kawai S, Nishihara A, Matsuura K, Haruta S. Hydrogen-dependent autotrophic growth in phototrophic and chemolithotrophic cultures of thermophilic bacteria, *Chloroflexus aggregans* and *Chloroflexus aurantiacus*, isolated from Nakabusa hot springs. FEMS Micro Lett 2019;366(10):fnz122.
[159] Keller MW et al. Exploiting microbial hyperthermophilicity to produce an industrial chemical, using hydrogen and carbon dioxide. Proce of the Nat Acad of Scien 2013;110(15):5840–5.
[160] Jahn U, Huber H, Eisenreich W, Hügler M, Fuchs G. Insights into the autotrophic CO2 fixation pathway of the archaeon *Ignicoccus hospitalis*: comprehensive analysis of the central carbon metabolism. J of Bac 2007;189 (11):4108–19.
[161] Heimerl T et al. A complex endomembrane system in the archaeon *Ignicoccus hospitalis* tapped by *Nanoarchaeum equitans*. Front in Micro 2017;8:1072.
[162] Imanaka H, Fukui T, Atomi H, Imanaka T. Gene cloning and characterization of fructose-1, 6-bisphosphate aldolase from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. J of Biosc and Bioeng 2002;94(3):237–43.
[163] Nunoura T, Chikaraishi Y, Izaki R, Suwa T, Sato T, Harada T, et al. A primordial and reversible TCA cycle in a facultatively chemolithoautotrophic thermophile. Scien 2018;359(6375):559–63.
[164] Aoshima M, Igarashi Y. Non decarboxylating and decarboxylating isocitrate dehydrogenases: oxalosuccinate reductase as an ancestral form of isocitrate dehydrogenase. J of Bac 2008;190(6):2050–5.
[165] Aoshima M, Ishii M, Igarashi Y. A novel enzyme, citryl-CoA synthetase, catalysing the first step of the citrate cleavage reaction in Hydrogenobacter thermophilus TK-6. Mole Micro 2004;52(3):751–61.
[166] Robb, F. T., & Techtmann, S. M. Life on the fringe: microbial adaptation to growth on carbon monoxide. *F1000 Res.*, *7*. (2018).
[167] Liu Y, Whitman WB. Metabolic, phylogenetic, and ecological diversity of the methanogenic Archaea. *Annals of the New York Acad*. of Scien 2008;1125 (1):171–89.
[168] Seravalli J, Kumar M, Lu WP, Ragsdale SW. Mechanism of carbon monoxide oxidation by the carbon monoxide dehydrogenase/acetyl-CoA synthase from *Clostridium thermoaceticum*: kinetic characterization of the intermediates. Bioche 1997;36:11241–51.
[169] Techtmann SM, Colman AS, Robb FT. 'That which does not kill us only makes us stronger': the role of carbon monoxide in thermophilic microbial consortia. Enviro Micro 2009;11(5):1027–37.
[170] De Anda V, Chen L-X, Dombrowski N, Hua Z-S, Jiang H-C, Banfield JF, et al. *Brockarchaeota*, a novel Archaeal phylum with unique and versatile carbon cycling pathways. Natu Comm 2021;12(1).

[171] Sousa DZ, Visser M, van Gelder AH, Boeren S, Pieterse MM, Pinkse MWH, et al. The deep-subsurface sulfate *reducer Desulfotomaculum kuznetsovii* employs two methanol-degrading pathways. Natu Comm 2018;9(1).
[172] Van Haaster DJ, Silva PJ, Hagedoorn PL, Jongejan JA, Hagen WR. Reinvestigation of the steady-state kinetics and physiological function of the soluble NiFe-hydrogenase I of *Pyrococcus furiosus*. J of Bac 2008;190 (5):1584–7.
[173] Gadkari D, Schricker K, Acker G, Kroppenstedt RM, Meyer O. *Streptomyces thermoautotrophicus* sp. nov., a thermophilic CO-and H2-oxidizing obligate chemolithoautotroph. Appl and Enviro Micro 1990;56(12):3727–34.
[174] MacKellar D et al. *Streptomyces thermoautotrophicus* does not fix nitrogen. Sci Rep 2016;6:20086.
[175] Volpiano, C. G., Sant'Anna, F. H., da Mota, F. F., Sangal, V., Sutcliffe, I., Munusamy, M., ... & Rosado, A. S. Proposal of Carbonactinosporaceae fam. nov. within the class Actinomycetia. Reclassification of *Streptomyces thermoautotrophicus* as *Carbonactinospora thermoautotrophica* gen. nov., comb. nov. *Syst. and Appl. Micro.*, *44*(4), 126223. (2021).
[176] Guterres, A. United Nations Secretary-General. Carbon Neutrality by 2050: The World's Most Urgent Mission (United Nations, 2020);Carbon neutrality by 2050: the world's most urgent mission | United Nations Secretary- General.
[177] Chen SC, Sun GX, Yan Y, Konstantinidis KT, Zhang SY, Deng Y, et al. The Great Oxidation Event expanded the genetic repertoire of arsenic metabolism and cycling. Proce of the Nat Acad of Scien 2020;117(19):10414–21.
[178] Li H, Opgenorth PH, Wernick DG, Rogers S, Wu TY, Higashide W, et al. Integrated electromicrobial conversion of CO2 to higher alcohols. Scien 2012;335(6076):1596.
[179] Nürnberg DJ, Morton J, Santabarbara S, Telfer A, Joliot P, Antonaru LA, et al. Photochemistry beyond the red limit in chlorophyll f-containing photosystems. Scien 2018;360(6394):1210–3.
[180] Nielsen J, Keasling JD. Engineering cellular metabolism. Cell 2016;164 (6):1185–97.
[181] Clarke L, Kitney R. Developing synthetic biology for industrial biotechnology applications. Bioch Society Trans 2020;48(1):113–22.
[182] from heterologous RuBisCO expression to the Calvin-Benson-Bassham cycle. Antonovsky, N., Gleizer, S., Milo, R. Engineering carbon fixation in E. coli. Cur Opi in Biotec 2017;47:83–91.
[183] Song S, Timm S, Lindner SN, Reimann V, Hess WR, Hagemann M, et al. Expression of formate-tetrahydrofolate ligase did not improve growth but interferes with nitrogen and carbon metabolism of Synechocystis sp. PCC 6803. *Front.* Micro 2020;11:1650.
[184] Bathke J, Konzer A, Remes B, McIntosh M, Klug G. Comparative analyses of the variation of the transcriptome and proteome of *Rhodobacter sphaeroides* throughout growth. BMC Geno 2019;20(1):1–13.
[185] Fuchs G, Berg IA. Unfamiliar metabolic links in the central carbon metabolism. J of Biotec 2014;192:314–22.
[186] Strauss, G., & fuchs, G. Enzymes of a novel autotrophic CO2 fixation pathway in the phototrophic bacterium Chloroflexus aurantiacus, the 3- hydroxypropionate cycle. *European Journal of Biochemistry*, *215*(3), 633-643. (1993).

**Capítulo II**

## Carboxysomes: The next frontier in biotechnology and sustainable solutions

**Sulamita Santos Correa[1,2#], Júnia Schultz[1#], Brandon Huntington[1], Andreas Naschberger[1], Alexandre Soares Rosado[1*]**

[1]Bioscience Program, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Makkah, 23955, Saudi Arabia
[2]Laboratory of Molecular Microbial Ecology, Institute of Microbiology, Federal University of Rio de Janeiro, Rio de Janeiro, 21941-902, Brazil

# The authors contributed equally.

* Correspondence:
Alexandre Soares Rosado
alexandre.rosado@kaust.edu.sa

## ABSTRACT

Certain bacteria have the ability to develop microcompartments that function as protein-based organelles. These bacterial microcompartments (BMCs) sequester enzymes in order to optimize metabolic reactions. To-date, several BMCs have been characterized, including the carboxysome and metabolosome. New identifiable BMCs have been found via genomic analysis of the BMC locus; however, these need to be characterized in order to find their signature enzymes. Among the various BMCs, the carboxysome is the most extensively characterized and is found in all Cyanobacteria and some chemoautotrophic bacteria. Carboxysomes are self-assembling polyhedral proteinaceous BMCs that play an important role in carbon fixation and utilization of carbon sources. The carboxysome encapsulates the enzymes RuBisCo and carbonic anhydrase (CA) to increase the rate of carbon fixation in the cell and decrease the rate of oxygenation by RuBisCo. Carboxysomes are attractive targets for carbon assimilation bioengineering because of their ability to concentrate carbon in crops and industrially relevant microorganisms. Hence, the characterization of these BMCs is the first step for considering the applications of the carboxysomes. Therefore, the present review

provides a comprehensive exploration of carboxysome morphology, physiology, and biochemistry, along with recent advances in microscopy and complementary techniques for isolating and characterizing this important versatile class of prokaryotic organelles.

**Keywords:** Bacterial microcompartments; Carbon fixation; $CO_2$-concentrating mechanisms; Bioengineering; Bio-based solutions; Biotechnological applications.

## 1 INTRODUCTION

Autotrophic carbon fixation is the primary route through which organic carbon enters the biosphere, and it is a key process for sustaining life on Earth. To date, the following seven pathways of autotrophic carbon fixation have been described: (i) the Calvin-Benson-Bassham (CBB) cycle, (ii) the reductive tricarboxylic acid (rTCA) cycle, (iii) the 3-hydroxypropionate bi-cycle, (iv) the Wood-Ljungdahl (WL) pathway, (v) the dicarboxylate/4-hydroxybutyrate cycle, (vi) the 4-hydroxybutyrate cycle, and, more recently, (vii) the reductive glycine pathway (Garritano *et al.*, 2022; Correa *et al.*, 2023). Carbon drives whole communities of living organisms and underpins one of the most important biogeochemical cycles on Earth (Anthony *et al.*, 2020). Carbon-fixation pathways are essential for sequestering atmospheric carbon dioxide ($CO_2$) and converting it into organic carbon (Hügler & Sievert, 2011). The CBB cycle is quantitatively the most important mechanism of autotrophic $CO_2$ fixation. The enzyme that catalyzes the carboxylation of the substrate ribulose-1,5-bisphosphate, known as ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo), is the major driver of the CBB cycle and is found in all types of plants, algae, Cyanobacteria, and several other prokaryotes (Fuchs *et al.*, 2011; Hügler & Sievert, 2011).

Significant changes in environmental conditions have been observed over the past 3.5 billion years. These have been characterized by a decrease in $CO_2$ and an increase in oxygen ($O_2$) in the atmosphere, mainly due to the production of $O_2$ by photosynthetic organisms such as Cyanobacteria (Badger *et al.*, 2002). These changes induced the evolution of RuBisCo's kinetic properties, along with the co-evolution of $CO_2$-concentrating mechanisms (CCMs) (Badger & Bek, 2008). However, the specificity of RuBisCo to $CO_2$ over $O_2$ is low, which results in a significant amount of oxygenation of ribulose-1,5-bisphosphate due to the high $O_2$ concentrations in the atmosphere (Ducat *et al.*, 2012). This process is known as photorespiration and is also sometimes referred to as the $C_2$ cycle. Photorespiration leads to the production of the molecule 2-phosphoglycolate (2-PG), which inhibits several metabolic pathways in the cell

(Caldwell *et al.*, 2007) and needs to be recycled in an energetically wasteful process at the cost of ATP and NAD(P)H (Busch *et al.*, 2018; Eisenhut *et al.*, 2019). In addition, photorespiration results in a decrease in carbohydrate synthesis, along with the release of $CO_2$ (carbon loss) and nitrogen in the form of ammonia (Long *et al.*, 2021).

It is important to emphasize that photorespiration should not be considered a failure of RuBisCo. Instead, this pathway is linked to several other essential metabolic processes, such as nitrogen assimilation and the dissipation of excess energy (Busch *et al.*, 2020). However, to mitigate carbon loss and maximize carbon fixation, certain microorganisms, such as Cyanobacteria, some chemoautotrophs, and purple photoautotrophs, have developed internal bacterial microcompartments (BMCs) termed carboxysomes (Hennacy & Jonikas, 2020). As self-assembling polyhedral proteinaceous microcompartments, these are versatile examples of prokaryotic organelles and are crucial for carbon fixation in all Cyanobacteria and certain chemoautotrophs (Axen *et al.*, 2014). Globally, carboxysomes may contribute to more than 25% of carbon fixation through atmospheric $CO_2$ assimilation (Rae *et al.*, 2013). These BMCs encapsulate the central $CO_2$-fixing enzyme RuBisCo by using a selectively permeable polyhedral protein shell, thereby increasing the local $CO_2$ concentration and creating a favorable environment for RuBisCo carboxylation (Kinney *et al.*, 2012). The following two distinct lineages of carboxysomes have been described: (i) the α-carboxysomes and (ii) the β-carboxysomes. These differ in their protein compositions and the forms of RuBisCo and carbonic anhydrase (CA) that they enclose (Ni *et al.*, 2022).

The carboxysome's shell assembly and capacity to enhance $CO_2$ fixation make it an attractive engineering target. Efforts have been made to engineer the α- and β-carboxysome structures in various heterologous organisms, with strategies aimed at elucidating the principles of BMC assembly in order to generate approaches for repurposing BMCs in synthetic biology (Liu *et al.*, 2022). A central aim is to improve the efficiency of RuBisCo carboxylation in order to increase carbon fixation, thereby removing the excess amounts of accumulated $CO_2$ from the atmosphere and enabling its re-utilization by plants and/or bacteria to enhance their productivity (Dai *et al.*, 2018; Xiao *et al.*, 2021). Hence, the present review aims to provide an overview of the general organization of carboxysomes and their importance as a carbon concentration mechanism for microorganisms. Several suitable strategies for the *in vitro* and *in situ* analyses of the molecular structure of carboxysomes are discussed, and a comprehensive analysis of the characteristics of α- and β-carboxysomes is provided.

## 2 CARBOXYSOMES: THE $CO_2$-FIXING BMCs

Carboxysomes serve as essential BMCs dedicated to $CO_2$ fixation via RuBisCo within the CBB pathway and are major players in carbon assimilation on a global scale. While the CBB cycle is the most abundant $CO_2$ fixation pathway, the key enzyme RuBisCo exhibits a reduced carboxylation rate. This is attributed to $CO_2$ and $O_2$ competing for the same active site on the enzyme (Wang *et al.*, 2021), thus resulting in the formation of two 3-phosphoglycerate (3-PGA) molecules during carboxylation and one 3-PGA molecule and one 2-phosphoglycolate (2-PG) during photorespiration (Ducat *et al.*, 2012; Caldwell *et al.*, 2007). While the 3-PGA molecule is catalyzed to generate carbohydrates, the 2-PG molecule is not metabolized in the CBB cycle and is instead used in respiration, whereby it consumes $O_2$ and releases already-fixed $CO_2$ (Caldwell *et al.*, 2007).

In addition to the reduced efficiency due to photorespiration, RuBisCo's catalytic turnover rate is very poor (Hanson, 2016). RuBisCo is estimated to be one of the most inefficient enzymes described, with only three to ten molecules of $CO_2$ fixed per second per molecule of enzyme (Ellis, 2010; Ludwig *et al.*, 1998). The catalytic turnover of RuBisCo for $CO_2$ conversion is several orders of magnitude slower than the diffusion limit of enzymes, which means large amounts of the protein need to be produced to adequately support reasonable $CO_2$ fixation rates (Hanson, 2016). In fact, the enzyme was estimated to be the most abundant protein on Earth (Ellis, 1979).

RuBisCo enzymes are classified into four types based on the primary sequence of their large catalytic subunit (L), which weighs approximately 50 kDa. The common structural motif of all four types is a functional dimer of *rbcL* subunits. The small subunit (S) with a molecular weight of 15 kDa only occurs in the most abundant type, referred to as Form I, which consists of large subunits (four dimers) surrounded by eight small subunits ($L_8S_8$). Meanwhile, Form II normally consists of a functional dimer ($L_2$) but can also form hexamers $(L_2)_6$. Both Forms I and II participate in autotrophic $CO_2$ assimilation in the CBB cycle. Form III of RuBisCo is found in thermophilic archaea and is arranged mainly as functional dimers ($L_2$), although pentamers of dimers $(L_2)_5$ have also been reported (Tabita *et al.*, 2008). Notably, Form III demonstrates the highest carboxylation activity, which is likely due to the favorable carboxylation kinetics provided by elevated environmental temperatures (Ezaki *et al.*, 1999). Although Form III was believed to be mainly involved in nucleotide and nucleoside metabolism (Sato *et al.*, 2007; Zeng *et al.*, 2022), it has been shown to have a carboxylase function in chemolithoautotrophic bacteria (Frolov *et al.*, 2019). Form IV of RuBisCo exclusively forms dimers ($L_2$) and is known to lack both carboxylase and oxygenase activities and is therefore not

involved in the CBB cycle (Tabita *et al.*, 2008).

To increase the concentration of $CO_2$ near RuBisCO and decrease the photorespiration rate, some microorganisms developed an intracellular microcompartment called a carboxysome, which functions as an organelle (Yeates *et al.*, 2008). Carboxysomes are composed of thousands of structural proteins and functional enzymes. These BMCs are functional analogs of eukaryotic organelles that sequester specific metabolic pathways in bacterial cells, such as pyrenoids, which similarly concentrate $CO_2$ near the $CO_2$-fixing enzyme RuBisCo in many algae (Zarzycki *et al.*, 2015; Fei *et al.*, 2022). The carboxysome is an organized framework formed by several semipermeable shell proteins (Sutter *et al.*, 2021). Encapsulation of the enzymatic core within this shell protects the cell from reactive and toxic metabolic intermediates such as acetaldehyde in the ethanolamine pathway and propionaldehyde in the 1,2-propanediol pathway (Penrod & Roth, 2006). It also protects the cell from other encapsulation enzymes that might produce toxic metabolic intermediates, and prevents unwanted side reactions (Sutter & Kerfeld, 2022). Although the membrane shell is formed by highly conserved proteins, the BMCs are functionally diverse, as they are involved in both anabolic and catabolic processes. Carboxysomes are the only known examples of anabolic BMCs, and metabolosomes are catabolics that encapsulate core enzymes for the specific purpose of recycling cofactors (Axen *et al.*, 2014; Turmo *et al.*, 2017). The outer protein shell allows for diffusion of $HCO_3^-$ (bicarbonate) ions into the compartment, where they are converted into $CO_2$ by CA and subsequently used by RuBisCo (Long *et al.*, 2018). Thus, carboxysomes regulate the local concentration of $CO_2$ to enhance the efficiency and selectivity of RuBisCo. Inorganic carbon from outside the cell is transferred first into the cytosol and then into the carboxysomes. This mechanism elevates the inorganic carbon concentration above the $CO_2$ concentration required for saturating the active site of RuBisCO (Badger & Price, 2003; Price *et al.*, 2008) and increases the ratio of $CO_2$ to $O_2$ so that carboxylation overcomes oxygenation. Hence, carboxysomes serve a dual purpose: firstly, they improve carbon fixation by elevating local substrate concentrations and secondly, they reduce the $O_2$ concentrations near RuBisCo, which shifts the reaction from photorespiration to $CO_2$ fixation.

In microorganisms equipped with CCMs, the active transport of $HCO_3^-$ elevates its intracellular concentration (Scott *et al.*, 2019). The cell accumulates more $HCO_3^-$ because it is charged and less soluble on the bilipid outer membrane, in contrast to $CO_2$, which is highly permeable and uncharged, allowing it to pass easily through the membrane and escape from the cell (Volokita *et al.*, 1984; Price & Badger, 1989). The cytoplasmic $CO_2$ is actively converted to $HCO_3^-$ via membrane-associated CAs (Shibata *et al.*, 2001; Maeda *et al.*, 2002). Cytoplasmic

$HCO_3^-$ passively diffuses into the carboxysome, where it is dehydrated by CA and then fixed within the compartment by RuBisCo, with the shell helping to confine the $CO_2$ near the RuBisCO (Long *et al.*, 2007). The $CO_2$ inside the carboxysome is then carboxylated via RuBisCo, thus leading to the formation of 3-PGA for biosynthesis (Price, 2011). Notably, a recent study using cryo-electron tomography (cryo-ET) has suggested that the high concentration of RuBisCo inside of the α-carboxysomes results from its polymerization therein (Metskas *et al.*, 2022). The authors showed that RuBisCo can arrange into a low-periodicity lattice inside the BMC, thereby enhancing its encapsulation and boosting the efficiency of carbon fixation.

## 2.1 History of carboxysome research

An overview of the history of carboxysome discovery is presented in Figure 1. Thus, the first observation of polygonal bodies, now recognized as carboxysomes, in prokaryotes was reported for the cyanobacterium *Phormidium uncinatum* (Drews, 1956). Almost 20 years later, these polygonal body structures were identified as a shell made from a proteinaceous envelope containing the enzyme RuBisCo. To reflect this discovery, the term carboxysome was introduced in the literature; however, there was no differentiation between α- and β-carboxysomes at that time (Shively *et al.*, 1973). After this discovery, carboxysomes were found in several other Cyanobacteria and chemoautotrophic bacteria (Shively, 1974). At this time, the carboxysome was known to have a CCM, but the mechanisms of inorganic carbon transport and RuBisCo activity were not yet explored. In 1982, the hypothesis that carboxysomes function only as simple protein storage bodies was rejected (Coleman, 1982). In that study, carboxysomes were isolated from Cyanobacteria (blue-green algae) and the presence of RuBisCo therein was observed and seemed to impact the efficiency of $CO_2$ fixation. The same study also presented preliminary results on the kinetic characteristics of RuBisCo.

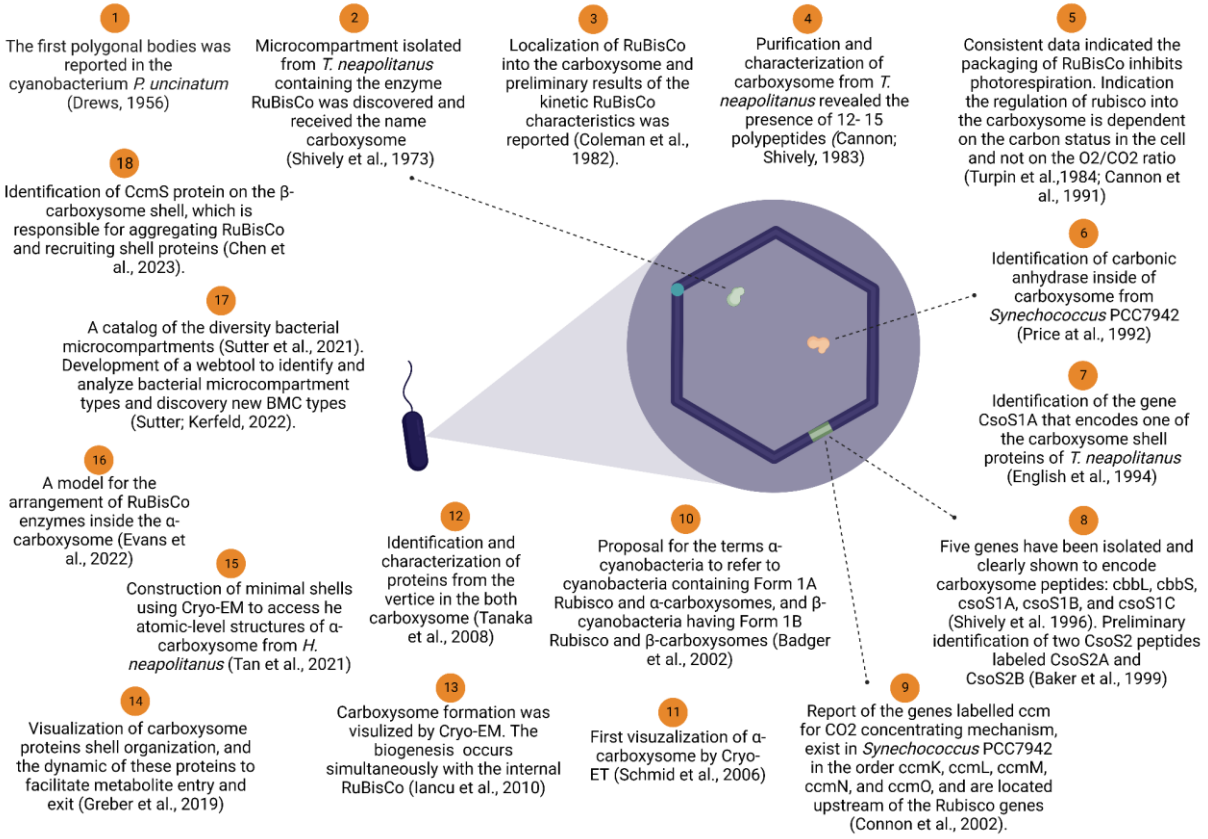



**Figure 1.** An overview of the history of carboxysome discovery, from the unknown microcompartment to the description of α- and β-carboxysomes, shape characterization, protein identification, and visualization of these structures via microscopic imaging. Created using Biorender.com.

Thenceforward, a representative α-carboxysome (although the terms α- and β-carboxysome had not yet been established) was more extensively characterized from cells of *Thiobacillus neapolitanus,* a specialized, obligately chemolithoautotrophic sulfur oxidizer (Cannon & Shively, 1983; Holthuijzen *et al.*, 1986). Then, in 1984, it was shown that the carboxysome kinetics in *Synechococcus leopoliensis* (Cyanobacteria) were related to inorganic carbon levels (Turpin *et al.*, 1984). Subsequently, the presence of a carboxysome in *Thiobacillus neapolitanus* was indicated via an *in-situ* assay, and the data were consistent with the hypothesis that $CO_2$ fixation was brought about by RuBisCo in the carboxysome, rather than by free RuBisCo in the cytoplasm (Cannon *et al.*, 1991).

Thereafter, the two carboxysome types (i.e., the α- and β-carboxysomes) were identified by Price *et al.* (1992) and Badger *et al.* (2002) in order to distinguish between Forms IA and IB of RuBisCo. The latter study also showed that this phylogenetic divide is further characterised the protein/gene components of the carboxysome. Thus, while the α- and β-carboxysomes have similar permeable shell layers, each type encapsulates a phyletically distinct form of the RuBisCo enzyme, and the proteins that form the structures such as the protein shell or the inner

RuBisCo matrix differ between the carboxysome types (Rae *et al.*, 2013). Moreover, each type of carboxysome has evolutionarily distinct forms of the same proteins (Rae *et al.*, 2013). The classification as α or β depends on the convergent evolution and the form of RuBisCo that they encapsulate; α-carboxysomes contain RuBisCo Form IA, and β-carboxysomes contain RuBisCo Form IB (Rae *et al.*, 2013). The carboxysomes also differ in the type of CA that they encapsulate (the β-class CsoSCA versus the γ-class CcmM (Kerfeld & Melnicki, 2016). Conversely, while each type of carboxysome differs from the other in terms of protein composition, they share the following two major structural and functional features: (i) the icosahedral BMC shell is composed of thousands of hexameric capsomers and 60 pentameric capsomers, and (ii) the lumen has several copies of the RuBisCo and CA enzymes (Cannon *et al.*, 2001).

The α-carboxysomes are present in marine Cyanobacteria, as well as many chemoautotrophs, a few purple sulfur bacteria, and some chemoautotrophic Proteobacteria (MacCready & Vecchiarelli, 2021). The sulfur oxidizer *Halothiobacillus neapolitanus* has been used as the model organism for the genetic and biochemical study of α-carboxysomes because it has fewer technical difficulties related to cell lysis and contamination with photosynthetic products (Roberts *et al.*, 2012). The α-carboxysome is thought to have arisen in Proteobacteria, after which the horizontal gene transfer of RuBisCo Form IA from Proteobacteria to α-Cyanobacteria resulted in the loss of the ancestral Cyanobacterial RuBisCo Form IB (Marin *et al.*, 2007). It has emerged that the α-Cyanobacteria are a monophyletic clade that diverged from β-Cyanobacteria approximately 1.0 Gya, when they possibly gained their cso operon (Figure 2) by horizontal gene transfer from a γ-proteobacteria genus such as *Nitrococcus* (Criscuolo & Gribaldo 2011; Rae *et al.*, 2013). Therefore, the α-carboxysome is thought to have evolved more recently than the β-carboxysome.

In brief, the two types of carboxysomes have distinct phyletic distributions, protein components and gene operons (Badger & Bek, 2008), and may have arisen via convergent evolution, perhaps after the divergence of the α- and β-Cyanobacteria 1.0 Gya. This is supported by the lack of β-carboxysome genes in the α-Cyanobacteria and the lack of α-carboxysome genes in the β-Cyanobacteria (Rae *et al.*, 2013). Experimentally, the β-carboxysome was distinguished from its α counterpart based on the biochemistry and observed interactions between carboxysomal BMC-H proteins (Samborska & Kimber, 2012), as discussed in detail in Section 3. This type of carboxysome is found in freshwater/estuarine Cyanobacteria (β-Cyanobacteria) such as strains of *Synechococcus elongatus* (PCC 7942) and *Synechocystis* sp. (PCC 6803), which live in diverse habitats subject to environmental dynamics such as nutrient

availability, desiccation, symbioses (Melnicki *et al*., 2021). The genes that encode the essential proteins for the β-carboxysome are found in the *ccm* operon, and there are significant variations between the loci. In addition, the RuBisCO *rbcL* and *rbcS* genes are only encoded in approximately a quarter of β-carboxysome loci, occasionally with the RuBisCO chaperone RbcX (Whitehead *et al.*, 2014; Axen *et al.*, 2014) (Figure 2).

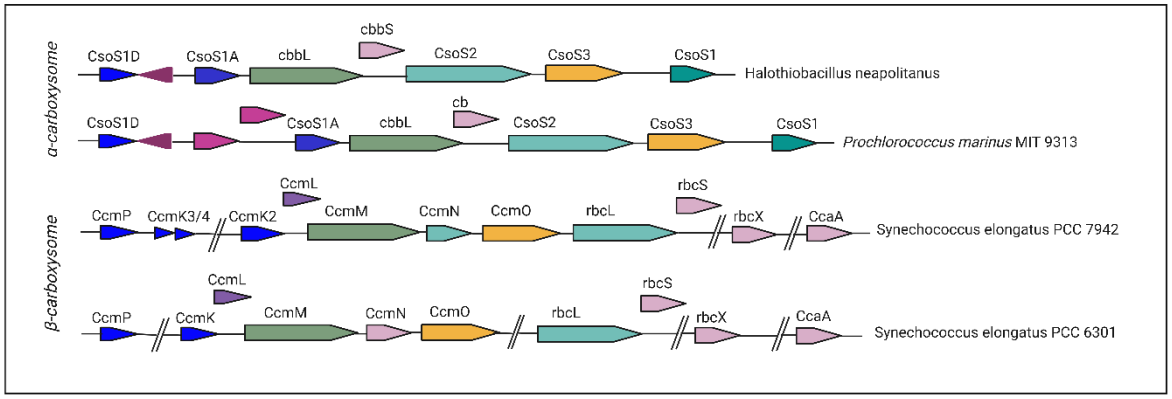



**Figure 2.** The gene organization schemes of (top) the cso operon from the α-carboxysome and (bottom) the ccm operon from the β-carboxysome. Genes with structurally and/or functionally similar products are the same color. The data were adapted from the MicrobesOnline database and edited using Biorender.com.

## 3 CARBOXYSOME STRUCTURE

Despite the differences between the α- and β-carboxysomes, both compartments show a similar overall organization and operate in a similar manner. The outer shell of each carboxysome is composed of hexameric BMC-H, pentameric BMC-T, and trimeric BMC-P proteins, which form the typical cyclic homo-oligomers with pores that enable metabolite transport across the shell (MacCready *et al.*, 2021) (Figure 3a). In detail, BMC-H consists of hexamer-forming proteins containing a single Pfam00936 domain, while BMC-T is a trimer/pseudo-hexamer-forming protein consisting of a fusion of two Pfam00936 domains, and BMC-P forms pentamers of Pfam03319 domains (Kerfeld *et al.*, 2005; Larsson *et al.*, 2017). The hexamer has a concave and convex side, and the facets of the BMC shells are comprised of a monolayer of proteins. The concave sides face the cytosol, and the pores pass through the cyclic axes of symmetry of the oligomers to act as the conduits for substrates and products (Sutter *et al.*, 2017; Sommer *et al.*, 2019) (Figure 3b).

**(a)**

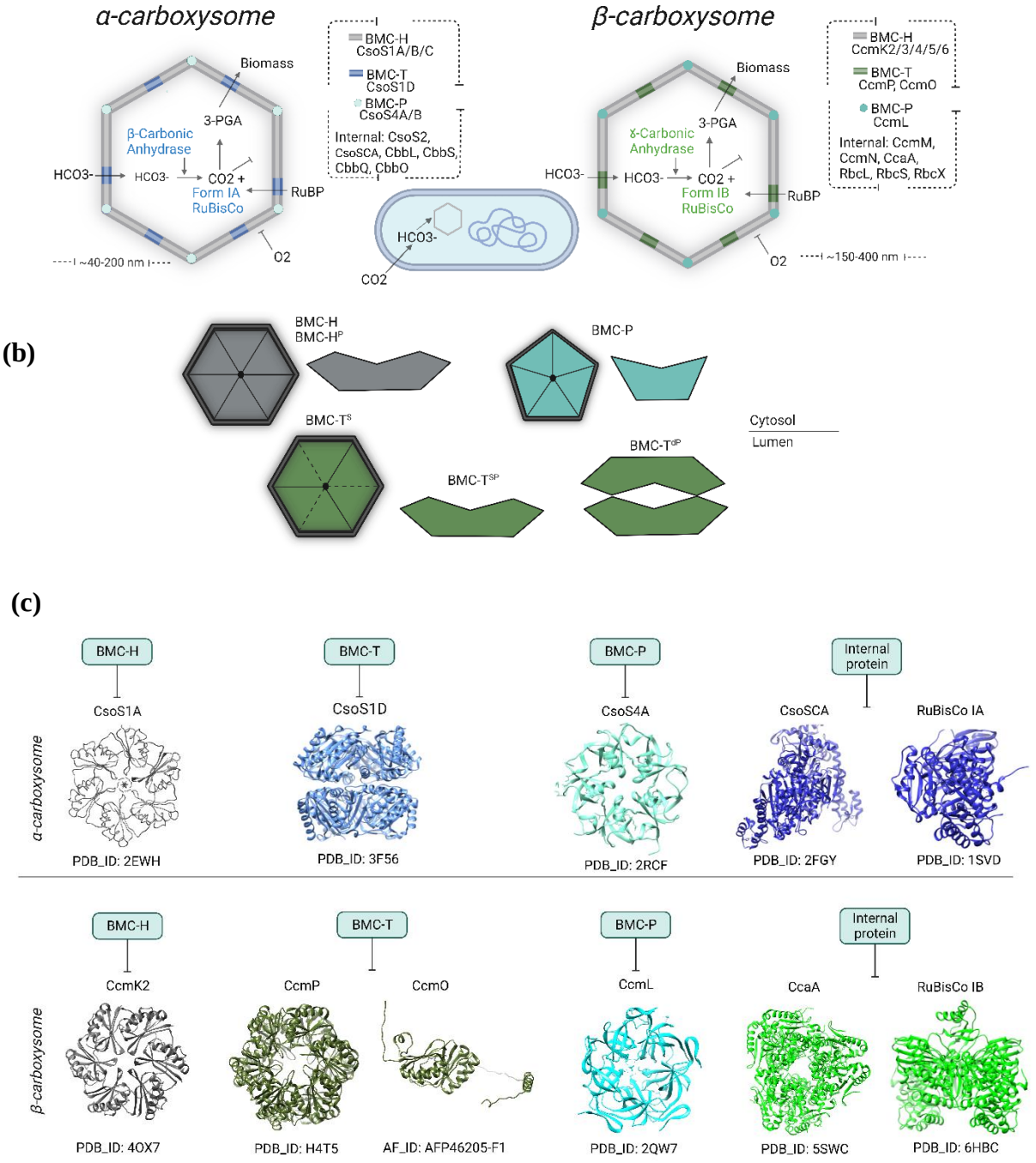



**Figure 3.** A schematic representation of the α- and β-carboxysome organization. (a) The general structure of each carboxysome and details of the reaction pathways. The shells are made of three types of building blocks, namely: BMC-H (gray), BMC-T (blue on α and green on β), and BMC-P (light cyan on α; cyan on β). (b) The structures of the distinct types of BMC-H, BMC-P, and BMC-T proteins. (c) The protein structures of the shell and internal proteins of each carboxysome. The data were adapted from the UNIPROT database using Chimera version 1.17.3. The figure a was created using Biorender.com.

## 3.1 BMC-H proteins

The BMC-H proteins constitute the building blocks of the polyhedral shell, providing flexibility in the permeability of the compartment (Axen *et al.*, 2014; Melnicki *et al.*, 2021). The BMC-H proteins form the hexameric complexes that constitute the bulk of the BMC shell facets (Figure 3a and b, left) (Sutter & Kerfeld, 2022). In α-carboxysomes, the BMC-H proteins are CsoS1A/B/C (upper row, Figure 3c). Studies have shown that the inactivation of these proteins in mutants can result in a reduction in the number of carboxysomes, thus requiring increased levels of $CO_2$ for growth in wild-type cells (English *et al.*, 1995). The CsoS1 paralogs form a very simple shell and have strong conservation upon the establishment of hexamers (Rae *et al.*, 2013). In the β-carboxysome, the BMC-H building blocks are Ccmk1, CcmK2, CcmK3, CcmK4, CcmK5, and CcmK6 (Figure 3a and b, right, and the bottom row of Figure 3c). The CcmK proteins possess a pfam00936 domain and form almost perfect symmetrical hexamers with a distinctive concave and convex side (Kerfeld *et al.*, 2005). The multiplicity of CcmK paralogs is believed to be an adaptation to changes in the carboxysome permeability in response to fluctuating environments (Sommer *et al.*, 2019). It has been demonstrated that CcmK3 and CcmK4 are not essential for carboxysome biogenesis in *Synechococcus elongatus* (PCC 7942); nevertheless, their presence is crucial for maximum carboxysome activity (Rae *et al.*, 2013). The pfam00936 domain members are diverse, and the most basic structure is formed by six units of BMC-H protein organized in a regular hexagon to form the bulk of the BMC shell facets (Sutter & Kerfeld, 2022). This hexamer has a diameter of 65 Å, with the convex side facing the lumen and the concave side facing the cytosol, along with a circular permutation termed BMC-$H^p$ in which two secondary structural elements are displaced from the C- to the N-terminus (Sutter & Kerfeld, 2022) (right-hand side, Figure 3b).

Recently, a new carboxysome shell protein termed CcmS was identified in the *Synechocystis* sp. strain (PCC 6803). This protein interacts with the CcmK1 and CcmM proteins from the β-carboxysome and is encoded by the gene slr1911 (Chen *et al.*, 2023). Specifically, CcmS appears to interact with the C-terminal extension of CcmK1 and is co-localized with the intermediate complexes of CcmK1, which can affect the accumulation and assembly of CcmK1 complexes. The same authors found that CcmS can also interact with CcmM, which is a key factor in β-carboxysome assembly, and that a mutation in CcmS can lead to an aberrant carboxysome that decreases the photosynthetic capacity of the organism (Chen *et al.*, 2023).

## 3.2 BMC-T proteins

The following three types of BMC-T proteins have been described: (i) BMC-$T^{s}$, which is a single trimer consisting of three copies of a fusion-protein with two pfam00936 domains; (ii) BMC-$T^{sp}$, which is a permuted form of BMC-$T^{s}$ in which each pfam00936 domain contains a circular permutation as in BMC-$H^{p}$, and (iii) BMC-$T^{dp}$, a permuted BMC-T variant in which two trimers dimerize across their concave faces to form an interior chamber (Sutter *et al.*, 2021) (Figure 3b). The opening of the BMC-$T^{dp}$ pores can be correlated with the binding of metabolites such as 3-PGA or adenosine diphosphate in a pocket in the inside cavity (Larsson *et al.*, 2017; Mallette & Kimber, 2017). These trimers are described as having a large pore through which the metabolites pass into the carboxysome. In the β-carboxysome, the BMC-T proteins are composed of CcmP and CcmO (Figure 3a and b, right, and the bottom row of Figure 3c). The CcmP protein forms a dimer of tightly stacked trimers that can weakly bind 3-PGA (Cai *et al.*, 2013). Biochemical assays have shown that CcmO is crucial for carboxysome formation, with knockout of the CcmO gene resulting in a severely high $CO_2$-requiring phenotype (Sun *et al.*, 2021). Deletion of CcmO also results in a high $CO_2$-requiring phenotype, and a homologous relationship between the CcmK and CcmO genes can be established (Marco *et al.*, 1994; Badger & Price, 2003).

The BMC-T in the α-carboxysome is constructed of CsoS1D proteins that form stacked trimers with distinct open and closed conformations (Klein *et al.*, 2009) (Figure 3a and b, left, and the top row of Figure 3c). The CsoS1D protein is encoded upstream of the cso operon in most α-Cyanobacteria and forms trimers with a relatively large central pore (14 Å in diameter) (Cai *et al.*, 2013). This protein provides selective permeability in the shell, thereby allowing the passage of $HCO_3^-$ and RuBisCo into the carboxysome and the release of the 3-PGA product (Klein *et al.*, 2009; Cai *et al.*, 2013). In the absence of the CsoS1D, the carboxysome is slightly rounded. Therefore, this protein likely contributes to shell rigidity (Bonacci *et al*., 2012; Long *et al.*, 2018). Despite the other proteins belonging to the *cso* operon, CsoS1D is encoded by a gene outside the canonical *cso* gene cluster and has two Pfam00936 domains (Klein *et al.*, 2009). CsoS1D is similar to CcmP, although the latter is not an essential shell protein (Rae *et al.*, 2013). Bioinformatic analyses have shown that CcmP is orthologous to CsoS1D and present in all β-carboxysome genomes from Cyanobacteria (Cai *et al.*, 2012). The protein CsoS1D forms similar trimers to CsoS1A, while the CsoS1C protein forms a flattened structure, and the CcmO protein seems to form a similar structure to that of the CcmK hexamer (Kinney *et al.*, 2011).

### 3.3 BMC-P proteins

It is believed that BMC-P is confined to the vertices of the assembled shells, and it appears to be essential for the function of the fully closed shell as a diffusive barrier (Cai *et al.*, 2009) (Figure 3a). These pentamer-forming proteins belong to the pfam03319 domain fold-containing family and have the shape of a truncated pyramid (Tanaka *et al.*, 2008) (Figure 3b). The BMC-P proteins are characterized into four or five distinct phylogenetic groups with conserved amino acid positions at the hexamer-pentamer interface and exhibit strong sequence similarity (Tanaka *et al.*, 2008; Sutter *et al.*, 2021) (Figure 3c). When CcmL proteins are absent, the carboxysome takes on a rod-like shape due to the cell's inability to close the vertices (Price *et al.*, 1993). The abundance of CcmL proteins can vary according to environmental conditions; the concentration of CcmL per carboxysome increases with higher light illumination and greater $CO_2$ availability (Sun *et al.*, 2019). Unlike the CcmL proteins from β-carboxysomes, which form a distinct class of BMC-P proteins, the BMC-Ps in the α-carboxysomes consist of CsoS4A and CsoS4B, which are similar to each other and distinct from other BMC-Ps (Zhao *et al.*, 2019) (Figure 3c). This observation provides evidence for divergence between the two existing types of carboxysomes. The BMC-P phylogeny suggests that the α-Cyanobacteria were probably the recipients of a transfer event involving CsoS4A/B and CsoS2 genes (Kerfeld & Melnicki, 2016). These proteins close the shell, thus preventing the escape of $CO_2$ from the α-carboxysomes. A study involving structural comparisons and multiple-sequence alignments has shown that CsoS4A and CsoS4B differ from each other in their component interactions in the α-carboxysome and probably work differently in the assembly of the α-carboxysome, even though they share a similar overall structure (Zhao *et al.*, 2019). The structural comparison indicated that CsoS4A has a similar structure to that of CcmL, but lacks two β strands at the C-terminus and has a smaller central pore diameter of ~3.5 Å. Meanwhile, CsoS4B has a central pore with a diameter 2.15 Å, with each subunit being composed of a β-barrel core domain and an inserted α helix, and the $Loop_{\alpha1\text{-}\beta5}$ is quite flexible (Zhao *et al.*, 2019).

### 3.4 The internal proteins of the β-carboxysome

An enhanced understanding the function of the internal proteins is limited by the difficulty of disrupting the carboxysome shell and accessing these proteins without damaging them (Heinhorst *et al.*, 2006). Therefore, the function of some internal proteins remains elusive as they are difficult to characterize effectively. As shown in Figure 3, the internal proteins in

the β-carboxysome comprise CcmM, CcmN, CcaA, and RuBisCo. In turn, the CcmM proteins contain an N-terminal γ-CA domain, which is involved in the aggregation of RuBisCo (Long *et al.*, 2007). There are two forms of CcmM, including a full-length form with a CA-like domain plus three RuBisCo small subunit-like domains, and another form which lacks the CA-like domain (Ludwig *et al.*, 2000). These proteins form a bond close to the equatorial region of RuBisCo and, contrary to expectation, do not replace RbcS subunits (Wang *et al.*, 2019). CcmM induces the separation of the phase into a liquid-like matrix and interacts with the additional carboxysomal carbonic anhydrase A (CcaA) (Long *et al.*, 2007; Ludwig *et al.*, 2000). Trimeric CcmM, consisting of γCAL oligomerization domains, interacts with the C-terminal tails of the CcaA subunits and additionally mediates a head-to-head association of CcmM trimers (Zang *et al.*, 2021). The CcaA tetramer is consistent with β-CAs functioning as dimers or higher-order oligomers. The γ-CAL domains of trimeric CcmM recruit CcaA by binding the C-terminal peptide sequence of the latter, and these interactions cooperate to facilitate the multiprotein coassembly of CcmM, CcaA, and RuBisCo for encapsulation into the β-carboxysome (Pena *et al.*, 2010; Zang *et al.*, 2021).

The CcmM proteins also work together with the CcmN proteins to build a bridge connecting the shell proteins. The latter contains approximately seventeen amino acids and folds into an amphipathic α-helix that is fully conserved in β-carboxysomal gene clusters (Kinney *et al.*, 2012). It forms a complex with CcmM, and acts as an adaptor that binds the enzymatic core and the shell proteins, thereby facilitating the recruitment of shell proteins and the subsequent assembly of the β-carboxysome (Sun *et al.*, 2021). Specifically, the N-terminal of the CcmN protein interacts with the CcmM protein and, by extension, also with RuBisCo, while the C-terminal, a 15–20 amino-acid peptide, interacts with the carboxysome shell (Aussignargues *et al.*, 2015). The absence of either the full-length CcmN or the C-terminal in *Synechococcus elongatus* (PCC 7942) can result in large protein aggregates, so that increased $CO_2$ supplementation becomes necessary for growth (Cameron *et al.*, 2013). However, the quantitative analysis of β-carboxysome proteins in cell lysates has revealed that CcmM is more abundant than CcmN (Jakobson *et al.*, 2015). The CcmN C-terminal can also be found among the encapsulated proteins of other BMCs (Faulkner *et al.*, 2017). These results indicate that CcmN has an important role in carboxysome assembly; however, further studies are necessary in order to understand these interactions completely.

Together with CcmM, the RuBisCo Form IB encoded by the RbcL gene constitutes the carboxysome lumen. In the RuBisCo-CcmM complex, the RuBisCo small subunit-like (SSUL) domains bind between two dimers of the large RuBisCo subunit (RbcL) via Van der Waals

interactions to form critical salt bridges (Blikstad *et al.*, 2019). The RuBisCo Form I enzyme is a hexadecamer complex (~550 kD) containing eight RbcL (~50 kD) units and eight RbcS (~15 kD) units (Bracher *et al.*, 2017). The RbcL subunits form a tetramer of antiparallel RbcL dimers, while four RbcS subunits cover the top and bottom. For stabilization of the RbcL dimer, some chaperones such as a homodimer of RbcX are necessary (Huang et al., 2019) (see Figure 4e below). The biogenesis of β-carboxysomes has been described as inside-out, which suggests that the RuBisCo-CcmM complex triggers the formation of a core followed by its encapsulation by shell proteins to form the complete carboxysome (Cameron *et al.*, 2013; Chen *et al.*, 2013). A study of molecular genetics, including biochemical assays and live-cell microscopic imaging, showed that the deletion of RbcX in *Synechococcus elongatus* (PCC 7942) results in a decrease in the abundance of RuBisCo along with the carboxysome number and size (Huang *et al.*, 2019). These experiments provide insights into the role of RbcX in carboxysome assembly.

### 3.5 The internal proteins of the α-carboxysome

In contrast to the β-carboxysome, the assembly process of the α-carboxysome is enigmatic. However, it is agreed that the components are mainly encoded by genes in the *cso* operon (Ni *et al.*, 2022). The biogenesis of the α-carboxysome is believed to involve simultaneous shell assembly and cargo enzyme aggregation, with the shell being constructed from hexameric CsoS1 and pentameric CsoS4 proteins (Kerfeld & Melnicki, 2016). The protein CsoS2 appears to function as a linker bridge between the shell and the cargo RuBisCo *in vitro*; the N-terminus of CsoS2 binds RuBisCo, while the C-terminus of CsoS2 is presumed to interact with the shell proteins (Oltrogge *et al.*, 2020; Li *et al.*, 2020). Due to the order of the *cso* locus, it is accepted that the α-carboxysomes of the Cyanobacteria share some ancestry with chemoautotrophic microbes such as Proteobacteria (Kerfeld & Melnicki, 2016).

The principal proteins found inside the α-carboxysome are CsoS2, CsoSCA, and RuBisCo (Figure 3). As with CcmM in the β-carboxysome, CsoS2 functions in α-carboxysome assembly and structure (Figure 2). Despite having neither sequence nor domain homology, both CcmM and CsoS2 are essentially conserved and are abundant in each carboxysome type (Rai *et al.*, 2013; Kerfeld & Melnicki, 2016). CsoS2 is a large polypeptide with around 900 residues and contains no structured domains due to its intrinsically disordered nature (Cai *et al.*, 2015). The CsoS2 form binds with the shell, and RuBisCo is responsible for its own encapsulation, thus facilitating carboxysome nucleation, which is an indispensable step for carboxysome assembly (Borden & Savage, 2021). Two different isoforms of CsoS2 exist, being encoded by

the ribosomal frameshifting of cis elements on the CSoS2 mRNA. These frameshifting elements lead to the formation of the full-length CsoS2B protein and the shorter CsoS2A protein due to a C-terminal truncation during translation to the alternative frame (Chaijarasphong *et al.*, 2016). Only CsoS2B can lead to the formation of carboxysome-like structural assemblies all by itself; CsoS2A is completely incapable of this feat (Chaijarasphong *et al.*, 2016). An X-ray structural analysis of CsoS2 from the α-carboxysomes of *Halothiobacillus neapolitanus* has demonstrated that the N-terminus of CsoS2 binds with RuBisCo (Oltrogge *et al.,* 2020). The authors suggested that CsoS2 operates as a hub for the condensation of RuBisCo, thereby allowing efficient α-carboxysome formation (Chaijarasphong *et al.*, 2016; Ni *et al.,* 2023). The high resolution cryo-electron microscopy (cryo-EM) structures of the shells show that the intrinsically disordered CsoS2-C makes multivalent contacts with the shell proteins, thereby functioning as a molecular thread to stitch the assembly interfaces and mediate the α-carboxysome shell assembly (Ni *et al.*, 2023).

Previously, CsoSCA was called CsoS3 and was thought to be associated with the shell. However, it was demonstrated in *Halothiobacillus neapolitanus* that CsoSCA is a novel ε-class CA that converts cytoplasmic $HCO_3^-$ into $CO_2$ and produces a high $CO_2$ concentration near RuBisCo, thus enhancing the catalytic efficiency of this enzyme (So *et al.*, 2004; Heinhorst *et al.*, 2006). CsoSCA was found to be incorporated into the α-carboxysome through its interaction with RuBisCo via an intrinsically disordered N-terminal domain (Blikstad *et al.*, 2023). In addition, cryo-EM has revealed that RuBisCo forms a hydrogen-bonding network with CsoSCA, and that the binding site of CsoSCA overlaps that of CsoS2. However, these proteins have different motifs and binding modes, thereby demonstrating the plasticity of the RuBisCo binding site. Thus, the CA-RuBisCo supercomplex performs protein-protein interactions that lead to the self-assembly of the α-carboxysome (Blikstad *et al.*, 2023).

The genes CbbL (large subunit) and CbbS (small subunit) encode Form IA of the RuBisCo enzyme in the α-carboxysome (Roberts *et al.*, 2012). The CbbQ and CbbO genes encode putative RuBisCo chaperones and RuBisCo activase (CbbX), and the α-carboxysome superloci normally include these genes (Turmo *et al.*, 2017). Further, CbbQ has been found to have protein subtypes and can occur with both Forms IA and II of RuBisCo (Sutter *et al.*, 2015). The CbbQ protein is a hexamer of the typical AAA+ domain (ATPases associated with ATP-driven dissociation, unfolding, and remodeling of macromolecules) and interacts with CbbO to form the bacterial microcompartment associated with ATPase activity (Snider & Houry, 2008; Sutter *et al.*, 2015).

# 4 CARBOXYSOMES AS ESSENTIAL BACTERIAL ORGANELLES

Several microorganisms harbor carboxysomes, including Cyanobacteria, autotrophic Proteobacteria, Planctomycetes, green sulfur bacteria, Aquificales, and Archaea. Specifically, α-carboxysomes are found in certain anoxygenic phototrophs and a wide range of α-Cyanobacteria, including the marine α-Cyanobacteria that dominate oceanic ecosystems, while β-carboxysomes exist in a variety of β-Cyanobacteria that live in various habitats with fluctuating environmental conditions, including freshwater (Liu *et al.*, 2022). In particular, the Cyanobacteria, along with some chemoautotrophic microorganisms, have developed the strategy of using the Calvin-Benson pathway in the presence of oxygen, thus reducing carbon loss. Moreover, the very recent increase in available metagenomic data has made it possible to predict the existence of more BMCs, including in other phyla than previously described (Chen *et al.*, 2021). Indeed, with comprehensive bioinformatic analysis, it has been shown that more than 7000 BMC loci cluster into 68 BMC types or subtypes, including 29 new functional BMC types or subtypes in 45 phyla across the bacterial tree of life (Sutter *et al.*, 2021). The same study revealed a large BMC function in certain phyla such as Actinobacteria, Proteobacteria, and Firmicutes, thus reflecting the importance of metabolic flexibility across disparate niches. Recently, a microbial consortium dominated by the thermophilic and chemolithoautotrophic Actinobacteria *Carbonactinospora thermoautotrophica* StC has been associated with the possible presence of carboxysomes therein (Pinheiro *et al.*, 2023). In that study, the use of genomic data and visualization of polyhedral structures in the cellular cytoplasm suggested the presence of carboxysomes in other bacterial groups not previously described. The ability of these microorganisms to fix $CO_2$ at high rates when external $CO_2$ concentrations are very low, and to reduce or eliminate photorespiration under ambient conditions, can be explained by the active transport of inorganic carbon (Coleman and Colman, 1981). The success of this strategy depends on a set of adaptations termed the CCM, where the RuBisCO enzyme is dissolved and localized in the carboxysome.

Together, all the photosynthetic Cyanobacteria are responsible for ~ 25% of the organic carbon fixation on Earth, and their survival today depends on the existence of the carboxysome BMC (Kerfeld *et al.*, 2018). Nevertheless, these photosynthetic microorganisms emerged 2.4 billion years ago without either carboxysomes or a CCM. The evolution of the CCM seems to have occurred after the origin of the chloroplast (Rae *et al.*, 2013). When Cyanobacteria started to utilize RuBisCo, the atmosphere contained a high level of $CO_2$ and the level of oxygen was irrelevant. This condition presumably meant that RuBisCo operated efficiently in its primary

role as a carboxylase (Badger *et al.*, 2002). However, as the Cyanobacteria consumed the $CO_2$ via photorespiration, the oxygen levels in the atmosphere increased, first gradually and then dramatically, eventually reaching similar levels to today. Consequently, the cyanobacteria sought alternative ways to evolve by enhancing the kinetic properties of RuBisCo or using different types of CCMs (Shih *et al.*, 2016). For chemolithoautotrophic bacteria, which grow in habitats with low concentrations of $CO_2$, the CCMs can facilitate rapid autotrophic growth by utilizing bicarbonate transporters and $CO_2$ traps to generate high intracellular concentrations of carbon (Dobrinski *et al.*, 2005). The CCM can also facilitate growth in an environment where the concentration of $CO_2$ varies spatially and temporally (Dobrinski et al., 2005; Wang et al., 2020). All Cyanobacteria, and some autotrophic bacteria, are now dependent on carboxysomes, and their existence and productivity would not be possible without the CCM (Rae *et al.*, 2013).

Presently, two of the main questions remaining are: (i) how did the carboxysome evolve into two subgroups, and (ii) how it exists in various microorganisms with ecologically distinct niches. The evolution of carboxysomes may have started with other smaller shell proteins or more distantly related BMCs (Melnicki *et al.*, 2021). Because the diffusion of $CO_2$ is slower in water than in the air, the changes in atmospheric conditions were likely more challenging for aquatic phototrophs, which were also subject to other factors that affect $CO_2$ availability, such as pH and temperature (Rae *et al.*, 2013). Meanwhile, the RuBisCO Form I sequence are arranged in a cluster with four distinct evolutionary groups, one of which is that of the oceanic Cyanobacteria that with RuBisCO genes with phylogenetic affinity towards proteobacterial strains (Tabita *et al.*, 1999). It is believed that the α-Cyanobacteria evolved from a monophyletic clade of freshwater β-Cyanobacteria, possibly gaining their *cso* operon by horizontal gene transfer from a γ-proteobacterium to form the α-carboxysomes (Badger & Bek, 2008). This may explain why there are two types of carboxysomes in distinct niches. Thus, it could be argued that the two groups of Cyanobacteria developed CCMs independently of each other and not from a common ancestor having a carboxysome (Badger *et al*., 2002). More studies are required to elucidate the origin and evolution of carboxysomes.

## 5 STRUCTURAL CHARACTERIZATIONS OF CARBOXYSOMES

This section examines the various approaches that are currently used for structural studies on carboxysomes and their building blocks. Despite their first visualization as polygonal bodies on thin sections from bacteria cells via electron microscopy in 1956, most structural studies have been conducted using X-ray crystallography on subcomponents of the

carboxysomes (Drews, 1956; Rae *et al.*, 2013). Indeed, from the 1980s until 2013, X-ray crystallography was the method of choice for investigating carboxysome proteins at high resolution. At that time, studies using cryo-EM were primarily employed to resolve biomolecules at very low resolution, and therefore lacked structural details (Schmid *et al*., 2006). The first crystal structure that was determined from a carboxysome-derived protein was RuBisCo from the α-proteobacterium *Rhodospirillum rubrum* (Schneider *et al.*, 1986). The structure of Form II RuBisCo revealed an elongated dimeric organization and provided the first 3D visualization of the active site. The first high-resolution crystal structures of the hexameric BMC-H shell proteins CcmK2 and CcmK4 were published much more recently by Kerfeld *et al*. (2005). This study demonstrated how multiple hexamers arrange in a higher-order shell, and suggested that the narrow, positively charged pore at the center of each hexamer allows the passage of charged metabolites such as bicarbonate and PG-3 while excluding uncharged molecules such as $CO_2$ and $O_2$. One year later, the first structure of a carboxysomal β-CA was determined (Sawaya *et al*., 2006). This study revealed that, in contrast to other β-CAs, only one active site is present per dimer, rather than two. Subsequently, the first crystal structures of the BMC-P proteins CcmL and CsoS4A were published at 2.4 Å and 2.15 Å resolution, respectively (Tanaka *et al.*, 2008). Both complexes displayed the characteristic pentameric arrangement, prompting the authors to suggest that they act as vertices and work together with the hexameric BMC-H proteins in the construction of icosahedral carboxysome shells. The first structure of a BMC-T protein, CsoS1D, was then published by Klein et al. (2009), who revealed that the protein consists of two fused BMC domains which trimerize to form a pseudo hexamer. Further, the resulting trimers subsequently dimerize to create a channel-like barrel structure which implies the presence of a gated transport mechanism across the carboxysome shell.

Due to the difficulty in crystallizing the fully assembled carboxysomes, structural studies have mainly been restricted to the individual subunits that are typically recombinantly expressed in bacterial hosts such as *Escherichia coli*. However, following the introduction of direct electron detectors (Kühlbrandt, 2014) and advanced computer algorithms (Scheres, 2012), it became possible to determine structures at near-atomic resolution by using cryo-EM without the need to crystallize the respective proteins or protein complexes. This methodological advancement allowed researchers to study BMCs in the context of their fully assembled structures and led to several remarkable high-resolution studies including α- and β-carboxysome-derived protein shells (Sutter *et al.*, 2019). The prerequisites for such studies are the biochemical isolation and purification of carboxysomes from their native sources. Hence, the currently known methods for purifying the fully assembled carboxysomes from native

sources are outlined in the following section, after which Section 5.2 discusses the recent studies in which cryo-EM/ET have been used to elucidate the overall structure and organization of carboxysomes at intermediate to high resolutions.

## 5.1 Methods for the extraction and purification of native bacterial carboxysomes for structural studies

The challenge of characterizing the structure of fully assembled carboxysomes begins with their extraction and purification from the microbial cell. The crucial initial step in this process is the efficient lysis of the cells to release the carboxysomes. The choice of method depends on the bacterial species and the composition of their cell walls and should be ideally optimized for each cell type. Various methods for lysing cells have been applied in order to isolate carboxysomes. For example, Metskas *et al.* (2022) used chemical lysis to purify the α-carboxysomes from *Halothiobacillus neapolitanus* cells, while Ni *et al.* (2022) used a combination of lysozyme and mechanical lysis with glass beads in a bead-beating device. For the same cell type, Flamholzt *et al*. (2020) performed chemical lysis via a combination of a non-denaturing detergent and lysozyme, along with vigorous stirring. The addition of a detergent serves to solubilize the lipid bilayers from the cell wall or the thylakoid membranes from photosynthetic organisms, and helps to remove lipid contaminants (Chaijarasphong *et al.*, 2016). In addition, successful lysis has been performed by using sonication, without the need for enzymes (Cannon & Shively, 1983; So *et al.*, 2004). The quality of the carboxysomes obtained using sonication alone has been shown to be equal to that obtained via the simultaneous use of a detergent in the purification buffer (Cannon & Shively, 1983). This suggests that detergents could potentially be omitted in all protocols without compromising the quality. For other bacteria, such as Cyanobacteria, similar protocols have been successfully employed to lyse the cells. For isolating the α-carboxysomes from *Cyanobium* sp. PCC 7001, a combination of lysozyme with mechanical lysis has been similarly effective (Ni *et al.*, 2022; Evans *et al.*, 2023). Meanwhile, the commonly used non-denaturing zwitterionic detergent formulation known as Cell lytic B was successfully applied in combination with sonication to lyse *Synechococcus elongatus* (PCC 7942) cells and isolate the β-carboxysomes (Faulkner *et al.*, 2017).

For all protocols, protease inhibitors can be used throughout the lysis process to prevent protein degradation, and any released DNA can be enzymatically degraded by the addition of DNAse to decrease the viscosity of the lysate (Roberts *et al.*, 2012; Faulkner *et al.*, 2017;

Flamholzt *et al.*, 2020; Ni *et al*., 2022). In addition, all available protocols use the same two subsequent steps after cell lysis, regardless of whether the α- or β-carboxysome is involved (Figure 4a). First, a low-speed centrifugation step (10,000–12,000 g, ~20min) is used to clear the lysate, after which carboxysome pellets are obtained via high-speed centrifugation at 40,000–50,000 g for 30–60 min.

For further improvement of the carboxysome purity at this stage, various techniques have been employed. One technique involves the $Mg^{2+}$-dependent aggregation of carboxysomes with Percoll beads, followed by washing steps using Triton with ethylenediaminetetraacetic acid (EDTA) and subsequent $Mg^{2+}$ precipitation (Price *et al.*, 1992; Long *et al.*, 2005). However, it was later found that the use of Triton-Percoll is too harsh, resulting in a large reduction in certain shell proteins and, hence, damage to the ultrastructure of the BMC (Rae *et al.*, 2013). Therefore, a gentler technique is more commonly employed, utilizing a sucrose gradient (10% to 50%) followed by collection and analysis of the various obtained fractions by sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and/or western blotting (Chaijarasphong *et al.*, 2016; Faulkner *et al.*, 2017 Sun *et al.*, 2022). Specific antibodies can bind with the shell proteins, and the signal reveals the presence of carboxysome in the sample. However, only antibodies for the shell proteins of the α-carboxysome are presently commercially available (Long *et al.*, 2018). To detect the shell proteins of the β-carboxysome, it is necessary to customize the antibodies (Chen *et al.*, 2023). To overcome this problem, bacterial strains with shell proteins fused to fluorescent tags such as the enhanced green fluorescent protein (eGFP) have been generated in order to monitor the presence of the β-carboxysome throughout the purification process (Faulkner, 2017).

In conclusion, protocols for isolating α- and β-carboxysomes from native sources (e.g., bacteria) are well-established and have been used to generate samples for analyzing their structures via cryo-electron-microscopy. The recent advancements in cryo-electron microscopy are expected improve the process, making it more accessible for researchers to isolate and analyze BMC structures from various organisms, encompassing various sizes and forms. This will ultimately enhance our understanding of the molecular mechanisms and functions of these compartments.

### 5.2 Structural investigations on carboxysomes using single particle cryo-EM and cryo-ET.

Currently, two main methods are used for high-resolution structural studies on carboxysomes and other BMCs, namely: (i) single-particle cryo-EM (SPM) (Chari & Stark,

2023) and (ii) cryo-ET, which is often combined with subtomogram averaging (STA) (Hong *et al.*, 2023). With the development of new-generation electron detectors and advanced algorithms in SPM, the technique has been successfully employed to determine high-resolution structures of biomolecules, achieving atomic resolution in some cases (Nakane *et al.*, 2020). However, the use of SPM for high-resolution studies of carboxysomes is limited to shell-only compartments lacking internal proteins. This is because the recorded signal in the micrograph projection is dominated by the heterogeneously organized internal proteins, making the computational alignment of particles onto the much thinner carboxysome shell signal extremely challenging, if not impossible. To address this, there have been significant advancements utilizing synthetic BMCs with minimal shell components that lack internal enzymes to study the BMC shell architecture via SPM (Sutter *et al.*, 2017; Greber *et al*., 2019; Tan *et al.*, 2021; Ni *et al*., 2023)

The first breakthrough in determining a full BMC protein shell via SPM was reported in 2017, in which the authors used a minimal shell component strategy (Sutter *et al*., 2017). In this study, a low-resolution SPM reconstruction (8.7 Å) of a recombinantly produced 6.5 Mda BMC shell of *Haliangium ochraceum* was used to determine the crystal structure of the very same BMC by molecular replacement at 3.5 Å resolution. Notably, this remains the only published crystal structure of a BMC shell available in the Protein Data Bank (PDB) to date. The same structure was subsequently determined by SPM at an improved encapsulation of 3.0 Å (Greber *et al.*, 2019). Although the structure represents a metabolosome microcompartment, it captures the overall structure of a BMC at high-resolution, including interactions at the interfaces between BMC-H, BMC-P, and BMC-T (Figure 4b). Unlike crystallography, SPM allows for the classification and separation of different sub-classes of the BMC shell within the same dataset. The BMC shell exhibits both local and global heterogeneity with different shell compositions. Interestingly, the BMC-T dimeric pseudo-hexamer can adopt the following three distinct gating conformations: (i) semi-open, (ii) intermediate, and (iii) closed. Studies have correlated the semi-open conformation (outer ring closed, inner ring open) with fuzzy density, likely originating from a contaminant trapped in the lumen of the shell, and the gating arginines at the outer ring were in a closed position. To open the channel fully, an additional co-factor or metabolite needs to bind in the interior of the channel. This observation aligns with previously proposed airlock-like gating mechanisms (Klein *et al.*, 2009; Cai *et al*., 2013; Larsson *et al*., 2017).

Additionally, the structure of another BMC, the GRM2 metabolosome from *Klebsiella pneumoniae*, was determined by SPM of recombinantly produced shell proteins with

encapsulated core enzymes (Kalnins *et al.*, 2020). This study also found a large amount of heterogeneity in the BMC shell. While most of the particles adopted a pseudo-icosahedral T = 4 structure with characteristic BMC-H and BMH-P organization, BMCs containing missing BMC-P components and elongated shells with additional BMC-H components were also observed (Kalnins *et al.*, 2020). As expected, the enzymatic core could not be resolved informatively, likely due to the aforementioned alignment challenges. These results suggest that the experimental design of carboxysome studies can be informed by investigating other types of BMCs, determining how they can be isolated, and determining their structures.

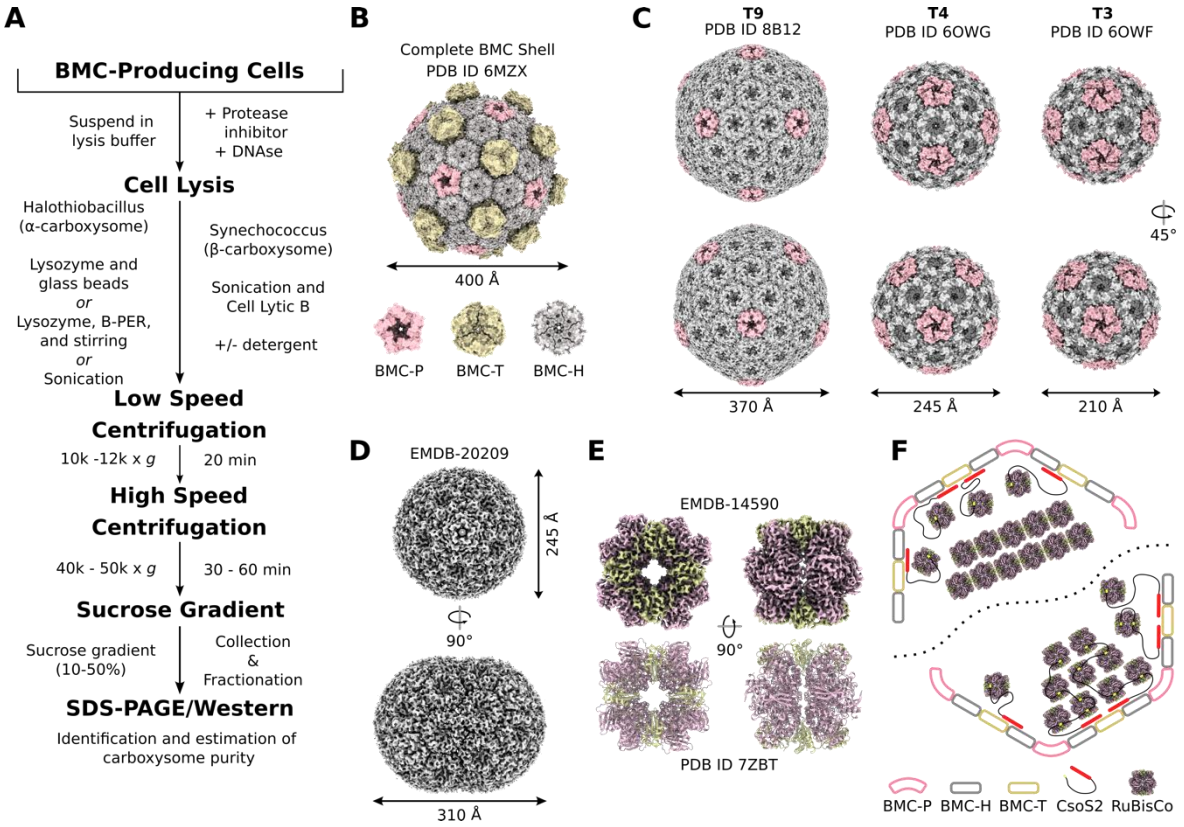



**Figure 4** Isolation protocols and structural analyses of carboxysomes: (A) a flowchart for native carboxysome purifications; (B) the high-resolution (3.0 Å) SPM structure of a fully intact BMC shell, where BMC-P, BMC-T, and BMC-H are colored pink, yellow, and grey, respectively; (C) the SPM structures of T9 (1.9 Å), T4 (2.6 Å), and T3 (3.0 Å) α- and β-carboxysome shells, using the same color code as in part (B); (D) the SPM structure of a prolate beta carboxysome shell (4.1 Å); (E) a 3.3 Å map of RuBisCo reconstructed from cryo-ET images with STA (top) and its associated atomic model (bottom), with RbcL and RbcS colored pink and yellow, respectively; (F) a schematic diagram of the two proposed internal RuBisCo arrangement models for α-carboxysomes, using the same color code as in part (B), along with the CsoS2 scaffold depicted as red anchors with black flexible lengths and yellow RuBisCo attachment sites.

The first high-resolution SPM structures of synthetic β-carboxysome shells were published by Sutter *et al.* in 2019. This study unveiled the following three major structural classes: (i) a population with a pseudo-icosahedral T = 4 structure (signifying that three proteins form the asymmetric unit, but not all subunits, are identical) with a diameter of 245 Å (Figure

4c), (ii) a small population (3.4%) with a pseudo-icosahedral T = 3 shell of 210 Å diameter, and (iii) a small population (4.5%) of elongated shells with dimensions of 245 x 310 Å (Figure 4d). Although BMC-T proteins were co-expressed during shell formation in *E. coli*, only BMC-H and BMC-P proteins were present in the final reconstruction, indicating that internal proteins are also necessary for the complete assembly of an intact β-carboxysome outer shell. Nevertheless, these structures offer a rational explanation for the permeability of metabolites. The external part of the BMC-H shell is highly negatively charged, while the inside of the pore is positively charged, thus providing favorable permeability for negatively charged metabolites that need to cross the shell. Subsequently, two high-resolution SPM structures (T = 3 and T = 4) of a synthetic α-carboxysome shell at 3.24 Å and 2.9 Å, respectively, were published (Tan *et al.*, 2021). These two structures exhibit good agreement with the T = 3 and T = 4 structures of synthetic β-carboxysome shells (Sutter *et al.*, 2019) (Figure 4c). Similarly, it was found that, despite overexpression of the BMC-T subunit, none of these proteins were present in the final reconstruction. Thus, despite significant efforts to elucidate the high-resolution structures of α- and β-carboxysome shells via SPM, a complete structure including the BMC-T subunits remains unavailable to date.

Recently, an SPM study was conducted using minimal shell constructs with only one type of BMC-H (CsoS4A) and BMC-P (CsoS1A) with or without the internal protein CsoS2A/B in order to decipher the α-carboxysome shell assembly and encapsulation principles (Ni *et al.*, 2023). Three structures were determined at 1.86, 3.54, and 2.79 Å resolution for T = 9, T = 4, and T = 3 shell assemblies, respectively. When only the BMC-H and BMC-P components are present, the BMC shell adopts a T = 4 or T = 3 configuration, with T = 3 predominating. These structures closely resemble the T = 4 and T = 3 configurations of synthetic α- and β-carboxysomes (Sutter *et al.*, 2019; Tan *et al.*, 2021) (Figure 4c). However, when CsoS2A or B are present, the BMC shell either adopts an identical, lowly populated, T = 4 configuration without the incorporation of CsoS2A/B, or a highly populated T = 9 configuration with CsoS2B intimately integrated at the tri-capsomere interface (Ni *et al.*, 2023) (Figure 4c). Despite only resolving three fragments of the CsoS2 C-terminus, their results demonstrate that, in addition to its role as a linker between the enzymatic core and the BMC shell, CsoS2 also acts as a molecular thread, facilitating the assembly of the α-carboxysome shell. These results also highlight the successful advancements being made toward a comprehensive understanding of fundamental carboxysome principles, enabled by a reductionist, bottom-up approach.

The only SPM study of the structure of a fully intact carboxysome from the

Cyanobacteria *Cyanobium* was undertaken by Evans *et al.* (2023). They reconstructed RuBisCo at a resolution of 2.9 Å from purified but broken carboxysomes, as well as the overall shell structure and internal RuBisCo-layers from the intact carboxysomes at a resolution of 18 Å. The high-resolution RuBisCo structure agreed well with previously published crystalline structures (L8S8, PDB: 1SVD). Density was observed in some subunits in the active site, suggesting that the enzymes were active and bound to the substrate. The arrangement of RuBisCo inside the carboxysome was found to be in concentric layers. However, 18 Å resolution was not sufficient to elucidate whether other proteins were bound to RuBisCo within these layers.

While SPM is the preferred method for studying the structures of empty carboxysome shells at high resolution, cryo-ET is preferred for complete natively purified carboxysomes or *in situ* studies (inside the cells). Although the latter technique often results in lower resolution reconstructions due to lower data contrast (Schmid *et al.*, 2006; Ting *et al.*, 2007; Lancu *et al.*, 2010; Dai *et al.*, 2018), recent examples have demonstrated that RuBisCo proteins from tomograms of natively purified carboxysomes can be aligned and averaged via STA, thus leading to a significantly improved resolution (Metskas *et al.*, 2022; Ni *et al.*, 2022). For example, Metskas *et al.* (2022) utilized cryo-ET to reconstruct the complete α-carboxysome from *H. neapolitanus*, using STA of the purified carboxysomes to achieve a 4.5 Å reconstruction of the RuBisCo and elucidate its arrangement within the BMC. RuBisCo polymerizes inside the carboxysome through its small subunit to form higher-order fibrils with six-fold pseudosymmetry. This arrangement seems to facilitate accessibility to the active site of RuBisCo and provide space for interactions with CsoS2 and CA (Blikstad *et al.*, 2023), although no density was found for those proteins in the cryo-EM map. In addition, a second study analyzed purified intact carboxysomes from two distinct species, namley *Cyanobium* sp. PCC 7001 and *H. neapolitanus* (Ni *et al.*, 2022), with STA giving RubisCo reconstructions at resolutions of 3.8 Å and 3.3 Å, respectively (Figure 4e). This study confirmed the fibrillar arrangement of RuBisCo in *H. neapolitanus*. However, in contrast to Metskas *et al*. (2022), the density for the CsoS2 scaffolding protein was found in proximity to the carboxysome shell, thus suggesting that CsoS2 holds the RuBisCo network in place while the fibrils form without the need for CsoS2. The internal arrangement of RuBisCo in *Cyanobium* was revealed to be organized in concentric layers, in agreement with the SPM study (Ni *et al.*, 2022; Evans *et al.*, 2023). The high-resolution reconstruction also revealed extra density near RuBisCo, likely corresponding to CsoS2, which appears to scaffold RuBisCo independently whether the proteins are near the shell or inside the lumen.

Thus, α-carboxysomes can be structurally divided into two groups, one in which CsoS2 acts as a scaffold for every RuBisCo inside the lumen, and one in which only the RuBisCos near the protein shell are scaffolded by CsoS2, while the luminal RuBisCos are self-scaffolded into fibrils (Figure 4f). However, the functional differences between these two groups need to be determined in future studies. Moreover, none of the abovementioned high-resolution cryo-ET studies used *in situ* prepared α-carboxysomes. Therefore, further cryo-ET studies are nneded in order to determine whether the findings from purified carboxysomes correspond to the structures inside the cell. In addition, no cryo-ET or SPM studies have been conducted on β-carboxysomes, likely due to their soft nanomechanical properties (Faulkner *et al.*, 2017) and resulting difficulties in the purification and grid-freezing process. This challenge might be addressed by utilize an *in-situ* approach or exploring different species in which β-carboxysomes exhibit greater stability.

# 6 MOVING FORWARD: BIOTECHNOLOGICAL APPLICATIONS OF CARBOXYSOMES AND FUTURE TRENDS

As our understanding of carboxysome structure and functionality continues to evolve, researchers are exploring innovative ways to harness their capabilities and pave the way for future advancements. Notably, several interesting biotechnological applications, such as carbon sequestration, agriculture, biofuel and bioenergy, biomedicine, biocatalysis, and materials science, have been discussed as promising uses of carboxysomes (Lin *et al.*, 2014a; Liu *et al*., 2022; Gao *et al.*, 2022). The unique properties of carboxysomes and their modularity and ability to self-assemble have gained increasing attention from the community of synthetic biologists and metabolic engineering scientists, with the aim of engineering and optimizing their functions. Interestingly, due to their relative simplicity, carboxysomes offer the potential for creating *in vitro* structures that can perform carbon concentration and fixation, thereby allowing for better understanding and further engineering of these functions. Such applications have been focused on improving crop productivity, and other agricultural demands, as well as enhancing carbon sequestration through the bioengineering of the carboxysome structure (Lin *et al.*, 2014a; Lin *et al.*, 2014b; Occhialini *et al.*, 2016; Cai *et al.*, 2015; Baumgart *et al*., 2017).

The major focus of these endeavors has been on reconstituting the carboxysome in an alternative host organism. Achieving a carboxysomal CCM in a heterologous bacterial host entails incorporating additional genes beyond those found in the major carboxysomal operon. For example, the successful reconstitution of the CCM from *H. neapolitanus* in *E. coli* involved

expressing a secondary operon alongside the major operon, which included the DAB inorganic carbon transporter, the CbbO and CbbQ RuBisCo activase complex, and the potential RuBisCo chaperone, acRAF (Wheatley *et al.*, 2014; Flamholz *et al.*, 2020). Interestingly, the DAB complex and the bicarbonate transporter SbtA are the only active transporters demonstrated in a heterologous system, making them promising candidates for CCM engineering (Du *et al.*, 2014; Desmarais *et al.*, 2019). The ability to reconstitute a carboxysomal CCM in *E. coli* signifies advancement in understanding the relationship between carboxysome structure and CCM function.

Other studies have already shown that carboxysome genes can be heterologously produced in *E. coli.* A genetic system for the heterologous expression of functional carboxysomes from *H. Neapolitanus* in *E. coli* has been well established, demonstrating the ability to produce similar icosahedral complexes to those from the native host (Bonacci *et al.*, 2012). Moreover, attempts to create the encapsulation of heterologous pathways in microbial model organisms have not been restricted to *E coli*. For example, Baumgart *et al.* (2017) was the first to introduce carboxysome genes in gram-positive bacteria, namely *Corynebacterium glutamicum*, thereby suggesting the formation of BMC-like structures in cells expressing either the complete carboxysome operon or only the shell proteins. However, the reconstitution of a heterologous organelle is challenging, as it requires the incorporation of a synthetically simplified organelle in order to facilitate the engineering process, along with extensibility to enable the improvement of a variety of functions and applications. The reconstruction is particularly difficult for carboxysomes because it requires a set of molecular constraints, including the description and number of genes necessary for functional expression in a host, which must be clarified by genetic and protein engineering (Bonacci *et al.*, 2012). Consequently, even though photosynthetic CCM-containing microorganisms are responsible for half of the daily $CO_2$ fixation that occurs on Earth (Hennacy & Jonikas, 2020), the attempts to reconstitute a heterologous carboxysome have not been restricted to microorganisms. Instead, an alternative approach has been sought to increase crop yields by inserting the carboxysome into plants via genetic engineering (Occhialini, *et al.*, 2016). By exploiting the efficient CCM of carboxysomes, scientists aim to engineer synthetic systems that enhance photosynthetic efficiency in crops.

The first introduction of the carboxysome set in a eukaryotic host was constructed via a multigene transformation for the tobacco chloroplast (Long *et al.*, 2018). Genes for the large and small RuBisCo subunits from *Cyanobium*, with or without the α-carboxysome proteins CsoS1A and CsoS2, were used to produce simplified carboxysomes. The resulting genetic

expression constructs were introduced into the tobacco plastome in order to increase crop yields by improving the photosynthesis in the chloroplast. Although this was not a new strategy, previous research was focused on engineering the plant RuBisCo, but this proved challenging due to the oxygenation activity of RuBisCo (Parry *et al.*, 2003; Whitney *et al.*, 2011; Ort *et al.*, 2015). The RuBisCo-mediated $CO_2$ fixation in C3 chloroplasts is catalytically slow compared with carboxysome RuBisCo (Tcherkez *et al.*, 2006; Whitehead *et al.*, 2014; Flamholz *et al.*, 2019). From an agricultural perspective, the large investment in RuBisCo to overcome its poor kinetics and loss of water from the open the stomata lead to a loss of biomass yield (Long *et al.*, 2015; Ort *et al.*, 2015). Hence, plant transformation with carboxysome seems to be more attractive and less challenging. Indeed, the feasibility of introducing carboxysomes into chloroplasts for the potential compartmentalization of RuBisCo or other proteins has already been demonstrated by the successful introduction of β-carboxysome genes from *Synechococcus elongatus* (PCC 7942) into *Nicotiana benthamiana* (Lin *et al.*, 2014). More recently, Chen *et al*. (2023) introduced engineered α-carboxysomes from *H. neapolitanus* into the chloroplasts of *Nicotiana tabacum* (tobacco) to support and enhance crop autotrophic photosynthesis and yields. The transgenic lines generated could facilitate further development of carboxysome engineering and provide an ideal host for installing fully functional CCM into chloroplasts to improve plant photosynthesis and productivity.

Researchers have also attempted to improve the carboxysome itself, with recent findings highlighting the possibility of optimizing the carboxysome by focusing on RuBisCo and protein shell biogenesis. For example, Nguyen *et al.* (2023) demonstrated a hybrid carboxysome with shell proteins from the *Cyanobium* (PCC700) α-carboxysome (specifically, CsoS1A and CsoS2), for which the assembly process was simpler than for the β-carboxysome and enabled the encapsulation of more active variants of RuBisCo, particularly Form IB from the *T. elongatus* β-carboxysome. The study was focused on the potential for improving the fixation flux and proposed an appropriate carboxysome functionality. The type of RuBisCo encapsulated within the carboxysome appears to impact its functionality, as evidenced by (i) the impaired growth of cells expressing an orthologous Form IA RuBisCo in an α-carboxysome and (ii) the disappearance of the carboxysome from the cell upon replacement with Form II (Pierce *et al*., 1989; Menon *et al.*, 2008). Additionally, the expression of RuBisCo activase, which corresponds the recombinant RuBisCo, is crucial for proper functioning within the carboxysome.

An expansion of the above knowledge to industrial hosts such as *Corynebacterium glutamicum*, or eukaryotic hosts such as yeast, could lead to the development of bioindustrial

strains that are better equipped to utilize carboxylation as part of metabolic engineering strategies. Furthermore, the introduction of CCM into autotrophic strains such as *Cupriavidus necator* could enhance their $CO_2$-dependent growth behavior. Researchers are also exploring the possibility of constructing minimal carboxysome systems by eliminating unnecessary proteins or creating fusions, which could prove advantageous in recombinant CCM engineering due to reduced DNA payloads.

Advancements in understanding carboxysome structure and assembly have now made it possible to re-engineer them for alternative metabolisms, e.g., for sustainable energy supplies, an area that was previously limited to other types of BCMs. For example, the expression of [FeFe]-hydrogenase and ferredoxin within the α-carboxysome shell of *E. coli* has been shown to enhance hydrogen ($H_2$) production while protecting the hydrogenase from $O_2$ inactivation (Li *et al.*, 2020). Hydrogen is considered a green energy source compared to traditional fossil fuels (Hallenbeck, 2009; Mudhoo *et al.*, 2011). Hence, clean combustion for a low-carbon future depends on the development of novel catalysts for processes such as hydrogen production from water (Das; Peu, 2022). The following three classes of hydrogenases exist, based on their active site metals: (i) [Fe]-hydrogenases, (ii) [FeFe]-hydrogenases, and (iii) [NiFe]-hydrogenases. Due to their desirable electrocatalytic activities, these hydrogenases provide great promise for constructing new hydrogen-evolution catalysts (Tard & Pickett, 2009). However, these enzymes have limitations, such as high sensitivity to oxygen (Peters *et al.*, 2015). To solve this problem, many studies have invested in encapsulating these enzymes in the carboxysome to improve the $H_2$ production and $O_2$ tolerance of the hybrid biocatalyst (Li *et al.*, 2020). For example, a novel hybrid catalyst was developed by introducing the enzyme [NiFe]-hydrogenase-1 from *E. coli* into an α-carboxysome from *Halothiobacillus neapolitanus* (Jiang *et al.*, 2023). The [NiFe]-hydrogenase is not completely sensitive to oxygen and can catalyze $H_2$ oxidation in the presence of $O_2$; under anoxic conditions, however, it has a high activity (Menon *et al.*, 1991; Zhang *et al.*, 2020). These studies have opened up exciting possibilities for biotechnological applications of carboxysomes and engineered structures derived from them, while also contributing to an understanding of basic carboxysome biology. Additionally, they suggest that carboxysomes create an environment that excludes $O_2$, a hypothesis that has sparked significant discussion.

Another area of interest is the development of targeted drug delivery systems by using carboxysome structures as configurable nanometer-scale molecular devices carrying therapeutic molecules (Bonacci *et al.*, 2012). These microcompartments display interesting characteristics applicable to nanotechnology and can be engineered to encapsulate and

selectively deliver medicinal molecules to specific cells or tissues, as desired. The BMCs offer the same supramolecular advantages as viral capsids, but assemble and can be used for the design, synthesis, and assembly of nanowires (Flynn, 2003; Nam *et al.*, 2006; Gao *et al.*, 2022). With precise control over the encapsulation process and surface modification, carboxysomes offer a versatile platform for drug delivery, reducing side effects and improving treatment efficacy (Bonacci *et al.*, 2012). For instance, although certain $O_2$-sensitive bioprobes and clinical drugs are particularly effective, they face severe stability problems or side effects due to reactions with oxygen, whereas the enzymes are extremely sensitive and can be irreversibly inactivated by the presence of oxygen (Esselborn *et al*., 2019). Inspired by the selectivity, permeability, and structures of the carboxysomes, Gao *et al.* (2022) used self-assembling proteins to construct artificial oxygen-impermeable protein nanocages (OIPNCs) and used these to develop nanodevices for broad application. The authors obtained a nanostructure that was able to control the oxygen permeability, which would be useful for experimental studies on the permeability of carboxysomes and BMC-derived shells, as well as offering a paradigm for solving the structures of proteins complexed with inorganic nanoparticles at atomic resolution, thereby improving our understanding of protein-nanoparticle interactions.

The approach of bioengineering the carboxysomes of microorganisms has the potential to revolutionize not only agriculture and carbon fixation but also bioenergy production by increasing and improving the production of renewable fuels. Carboxysomes could be used as nanoreactors for biocatalytic reactions, such as the production of biofuels or other sustainable and high-value chemicals. While the development of these microbial factories incurs significant costs, the long-term benefits include high product values, low material costs, reduced pollution compared to chemical catalysts, use of waste and their valorization, and a cleaner greener process (Huffine *et al.*, 2023). This has prompted increased attention and research in the field of microbial production for sustainable manufacturing strategies. In this way, the enzymes used in industrial processes that commonly require extreme conditions (e.g., pH, temperature, solvents, salinity) could be encapsulated in the carboxysome to protect them from degradation or denaturation, and the concentrated substrate inside the microcompartment could increase the reaction rate and efficiency.

Looking ahead, several future trends are expected to shape the biotechnological applications of carboxysomes. Advances in synthetic biology and genetic engineering techniques will enable the precise manipulation of carboxysome components, allowing for the creation of customized microcompartments with tailored functionalities. Also, advancements in high-resolution imaging techniques will facilitate a deeper understanding of carboxysome

assembly, dynamics, and interactions within cells, thereby enabling researchers to optimize their performance and applications. Furthermore, the integration of carboxysomes with other cellular compartments or synthetic structures holds great promise. By combining the unique attributes of carboxysomes with other organelles or engineered systems, researchers can create hybrid architectures that synergistically enhance their functionality, opening a black box of unimaginable applications.

## FINAL REMARKS

Carboxysomes are the key metabolic modules for carbon fixation in Cyanobacteria and other microorganism groups and show great promise for synthetic engineering to improve the catalytic efficiencies of enzymes in non-native hosts for industrial biocatalysis. Thus, by encapsulating enzymes within the microcompartments, researchers can enhance their functionality and stability, protecting them from degradation, and creating enzyme cascades for efficient biotransformation. This approach opens doors to the development of enhanced technologies, as well as greener and more sustainable industrial processes, enabling the production of valuable compounds and reducing the reliance on harsh chemical methods. Overall, the biotechnological applications of carboxysomes and bioengineered microcompartments represent a rapidly expanding field with tremendous potential for a range of industries, from carbon fixation and biofuel production to drug delivery and industrial biocatalysis, where the carboxysomes offer versatile platforms for innovation. In this respect, an understanding of the evolution, physiology, and morphology of the carboxysome is the first step towards establishing it as a product for application. Hence, the present review began by examining the various studies over multiple decades that have built up this essential knowledge base. Moving forward, the recent progress in synthetic biology and imaging techniques were described, and continued improvements, along with interdisciplinary collaborations, will undoubtedly lead to exciting breakthroughs in carboxysome research and drive their integration into various industries, thereby fostering a sustainable and biotechnologically advanced future.

## FUNDING

This work was financially supported by KAUST Baseline Grant BAS/1/1096-01-0 (to Prof. A. S. Rosado). SSC was funded by the National Council for the Improvement of Higher Education (CAPES).

**Capítulo III**

# Co-existence of CBB cycle and rTCA carbon fixation pathway in thermophilic *Actinomycetota Carbonactinospora thermoautotrophica* StC and correlation with the CO2- concentrating microcompartment

Sulamita Santos Correa[1,4#]; Luís Arge[2]; Fabio Mota[3]; Brandon Huntington[4]; Andreas Naschberger[4]; Júnia Schultz[4]; Alexandre Soares Rosado[4*]

[1]Institute of Microbiology, Postgraduate in Plant Biotechnology and Bioprocesses, Federal University of Rio de Janeiro, Brazil.
[2]Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, Minnesota, 55108-6026, USA.
[3]Oswaldo Cruz Foundation, Rio de Janeiro, Brazil
[4]Bioscience Program, Biological and Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology, the Kingdom of Saudi Arabia.

*Corresponding author Alexandre Soares Rosado
Biological and Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology, the Kingdom of Saudi Arabia.
E-mail: alexandre.rosado@kaust.edu.sa

## ABSTRACT

The exploration of inorganic carbon fixation holds paramount significance in unraveling the mysteries of early cellular life, evolution and addressing contemporary global concerns, notably climate change. Up to the present, researchers have identified and characterized seven distinct carbon autotrophic fixation pathways. Nevertheless, the simultaneous existence of multiple cycles within a single cell is a rarity in the scientific landscape. In this context, we present *Carbonactinospora thermoautotrophica* StC, a thermophilic bacterium characterized by its distinctive and adaptable carbon metabolism. Isolated from a consortium thriving in a blazing organic pile, StC thrives optimally at temperatures ranging from 55 °C to 65 °C. Genomic analyses indicate that the strain StC potentially performs two carbon fixation pathways the Calvin-Benson-Bassham (CBB) cycle and the Reductive Citrate Cycle (rTCA). Additionally,

it has a microcompartment associated with $CO_2$ concentration, underscoring its unique features in carbon utilization. In order to gain a comprehensive understanding of carbon fixation in the StC strain and explore its potential applications in biotechnological sectors, an analysis was conducted on the expression of genes and proteins in bacterial cells grown under autotrophic and heterotrophic conditions. To our surprise, the data revealed two carbon fixation pathways - CBB and rTCA cycles. The functionality of these two cycles holds the potential to enhance $CO_2$ fixation. Furthermore, this investigation sheds light on the identification of a carboxysome, a bacterial microcompartment linked to carbon fixation, a feature not previously documented in members of the *Actinomycetota* phylum until now. The revelation of multiple carbon fixation pathways within a single cell prompts the intriguing question of why microorganisms employ multiple pathways to fix carbon and the potential advantages it confers. Delving deeper into this inquiry is crucial for gaining insights into the biological mechanisms of carbon fixation in thermophiles, with implications for harnessing enzymatic processes in synthetic biology for biotechnological applications.

**Keywords:** carbon fixation, autotrophic thermophile, carboxysome, omics.

## INTRODUCTION

Carbon is an essential element for living organisms, and carbon fixation represents the most important biogeochemical cycle in the ecosystem. An increase in carbon dioxide in the atmosphere due to anthropogenic emissions leads to significant changes in global climate (Yoro; Daramola, 2020). As per the data presented in the annual Global Carbon Budget during the United Nations Conference of the Parties (COP28), it is projected that total fossil carbon dioxide emissions will surge to an unprecedented level of 36.8 billion tons in the year 2023 (International Energy Agency, ANO). The increased emission of gasses from industrial processes in an overpopulated world, currently accelerates the effects of climate change and environmental damage, including the potential extinction of more sensitive organisms (Knowlton et al., 2021; Hoegh-Guldberg et al., 2019).

When addressing the task of sequestering global carbon emissions, autotrophic carbon-fixing bacteria emerge as crucial contributors, significantly influencing carbon sequestration and contributing to the overall global net carbon fixation (Wang et al., 2021). Microorganisms play a pivotal role in driving crucial metabolic processes within the carbon cycle. They actively regulate various metabolic pathways associated with the carbon cycle, highlighting their ability

to respond to global climate change (Carvalhais et al., 2014). This microbial involvement is integral to maintaining the functionality and stability of ecosystems. The microbes have significant impact on the chemical cycling of carbon between the Earth's biosphere and biogeography and can also influence climate change through carbon cycle feedback (Zhou et al., 2012; Sokol et al., 2022). At least seven natural carbon fixation routes have been described and are distributed in all tree life: i) Calvin-Benson-Bassham cycle (CBB), ii) Reductive citrate cycle (rTCA), iii) 3-Hydroxypropionate bicycle (3-HP), iv) Wood- Ljungdahl (WL) pathway, v) Dicarboxylate/4-hydroxybutyrate, vi) 4-Hydroxybutyrate cycle and vii) Glycine reducing pathway (Correa et al., 2023).

Despite the diversity of multiple metabolic pathways identified for autotrophic carbon fixation, it has been previously observed that most autotrophic organisms usually harbor genes for only one of these pathways, and the presence of multiple carbon fixation genes is uncommon. For instance, the bacterium *Thioflavicoccus mobilis*, recognized as the first cultivable and free-living gammaproteobacterial sulfur oxidizer, harbors genes for two distinct carbon fixation pathways-the reductive citrate cycle (rTCA) and the Calvin-Benson-Bassham (CBB) cycle (Rubin-Blum et al., 2019). This characteristic has also been noted in certain uncultured bacterial symbionts found in the deep-sea worm *Raftia*, as well as in the free-living giant bacteria *Beggiatoa* and *Thiomargarita* (Thiel et al., 2017). Besides CBB and rTCA cycles, a study with *Thermovibrio ammonificans* (Giovannelli et al., 2017) provides evidence of ancestral and acquired metabolic traits in the genome of this bacterium. The reductive citrate cycle's origins trace back to a reducing anaerobic environment, and over time, evolutionary adaptations, coupled with the emergence of acetyl-CoA pathways, facilitate a gradual liberation from the constrained chemical conditions of the original habitat of life (Wächtershäuser, 1990). The conquest of a non-reductive or even aerobic environments was only possible after the emergence of an entirely new autocatalytic carbon cycle fixation, the CBB cycle, or a variant (Kandler, 1981; Wächtershäuser, 1990).

More the twenty years ago, an intriguing thermophilic aerobic chemolithoautotrophic bacterium *Streptomyces thermoautotrophicus* strain UBT1, isolated from the covering soil of a burning charcoal pile was described to have an interesting CO dehydrogenase activity and able to reduce low potential electron acceptors such as methyl and benzyl viologens, but the carbon fixation was not investigated (Gadkari et al., 1990). The species *S. thermautotrophicus* has been reclassified by our group, now identified as a member of a new family within *Actinomycetes*, named *Carbonactinosporaceae*, and it has been assigned a new genus and species designation, officially recognized as *Carbonactinospora thermoautotrophica* (Volpiano et al., 2021).

Our research group recently isolated a new strain of *C. thermoautotrophica*, designated as StC, from a consortium of bacteria discovered in actively burning organic matter (Pinheiro et al., 2023). *C. thermoautotrophica* StC thrives at temperatures between 55° and 65° C, is aerobic and chemolithoautotrophic bacterium possessing unique traits associated with autotrophic carbon assimilation and a Bacterial microcompartment (BMC) associated with $CO_2$ concentration mechanism. To comprehensively explore the carbon fixation mechanisms employed by StC, we conducted a multi-omics analysis encompassing genomic, transcriptomic, and proteomic approaches. Simultaneously, we isolated and characterized the bacterial microcompartments (BMC) present in the filamentous structure of StC. This investigation aimed to provide a thorough understanding of the strain's carbon fixation pathways and shed light on the functional attributes of the identified BMCs.

## METHODS

### Bacterial strain and culture conditions

The strain StC was isolated from a thermophilic microbial consortium (CNF) (Pinheiro et al., 2023) composed of four bacteria (*Carbonactinospora thermoautotrophica* StC, *Sphaerobacter thermophilus*, *Chelatococcus* sp., and *Geobacillus* sp.), in Brazil, Rio de Janeiro, coordinates: -22.775892, -43.691673. The StC strain was isolated from the microbial consortium using Reasoner 2 (R2A) solid culture medium (Kasvi), supplemented with the antibiotic amoxicillin (1 μg/ml). The culture was then incubated at 53 °C for one day. This culture was then transferred to N-FIX mineral medium (Gadkari et al., 1992) solidified with 1.5% high purity noble agar (Sigma-Aldrich) and supplemented with 0.03% clinoptilolite (a zeolite ammonia scavenger) (Liao et al., 2015; Dahal et al., 2017). The isolated StC strain has been preserved and is currently stored in our microbial collection at the Molecular Microbial Ecology Laboratory, located at the Federal University of Rio de Janeiro.

### DNA extraction, genome sequencing, assembly, and annotation

For the extraction of the genomic DNA, the StC strain was cultivated in N-FIX culture medium supplemented with $CO_2$/CO/$N_2$/$O_2$ gasses as described by Gadkari et al. (1990; Meyer, 1982; Meyer; Schegel, 1983), for 4 days at 60 °C. Total DNA was extracted with the QIAGEN® Genomic DNA kit with some modifications, as follows: the bacterial biomass was firstly lysed

with 480 µL of 0.5M EDTA and 120 µL of lysozyme (100 mg/mL), in bath incubated at 37 °C for 30 minutes, followed by FastPrep homogenization (MP Biomedicals). The quantification of total DNA was carried out using a NanoDrop spectrophotometer from Thermo Fisher Scientific. Following this, the total DNA underwent purification using the AMPure XP bead-based reagent from Beckman Coulter Life Sciences, and its concentration was further quantified using the Qubit® dsDNA HS Assay Kit, also from Thermo Fisher Scientific. The integrity of the extracted DNA was evaluated by non-denaturing agarose gel electrophoresis (1.5%).

Following the aforementioned steps, the genome of the StC strain underwent sequencing utilizing Oxford Nanopore Technologies (ONT) at the Bioscience Core Laboratory of the King Abdullah University of Science and Technology, Saudi Arabia, as well as being sequenced using Illumina HiSeq technology. The assembly of the *C. thermoautotrophica* genome employed a hybrid approach that combined short reads and long reads, ensuring both high depth and high coverage in the sequencing process. Genomic short reads (paired-end - 300bp) were generated using the Illumina HiSeq sequencer, while long reads were obtained through Oxford Nanopore sequencing. The Illumina reads underwent an initial cleaning step to eliminate low-quality bases and adapters, a process carried out using Trimmomatic Ver. 0.39. (Imm; Chang, 2023) with the following parameters: treated as paired-end (PE), ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:keepBothReads to remove adapters and keep both reads as paired, a LEADING and TRAILING value of 20 to remove bases with quality above 20, and all reads with a length less than 36 base pairs were also removed. Oxford Nanopore reads were cleaned initially using the basecaller Guppy Ver. 3.0.3+7e7b7d0 (Wick; Holt, 2019) to check and remove possible adapters, and low-quality bases were removed using Porechop Ver. 0.2.4 (Wick et al., 2017), both with default parameters. The genome assembly was performed with cleaned reads of both read types using MaSuRCA Ver. 4.0.3 (Zimin et al., 2013). For the choice of the best references, the contigs generated for the strain were submitted to the Basic Local Alignment Search Tool for nucleotides (BLASTn) (Altschul et al., 1990). The assembled contigs were oriented to generate a scaffold using Medusa v. 1.3 (Bosi et al., 2015), and the strains *C. thermoautotrophica* UBT1 and H1 were used as reference. Gaps resulting from the assemblies were manually filled using the CLC Genomics Workbench software, for visualization, v. 7.0 (Qiagen, USA), where the reads were mapped against a reference genome to generate a consensus sequence, which was then used to close the gaps.

The annotation of the *C. thermoautotrophica* StC genome was carried out through automated processes using PROKKA v. 3.0 (Seemann, 2014). Subsequently, the genomic sequence data for StC were submitted to the Type (Strain) Genome Server (TYGS)

(https://tygs.dsmz.de/user_requests/new) to undergo a whole-genome-based taxonomic analysis, as proposed by Meier-Kolthoff and Göker in 2019. Information on nomenclature, synonymy and associated taxonomic literature was provided by LPSN database (List of Prokaryotic names withstanding in Nomenclature, available at https://lpsn.dsmz.de) (Meier-Kolthoff et al., 2022). The phylogenomic tree was generated using the Phylogenomic Tree Tool in Pathosystems Resource Integration Center (PATRIC) (http://www.patricbrc.org), version 3.5.17 (Wattam et al., 2014). All the trees were visualized using iTOL v4.261 (Letunic and Bork, 2016). The similarity of StC, UBT1 and H1 genome was generated using the Pathosystems Resource Integration Center (PATRIC). The functional annotation of genome sequences was performed with KEGG using BLASTKOALA (Kanehisa; Morishima, 2016). The metabolic pathway was reconstructed to gap filling using the web tool KEGG pathway and GhostKOALA (Kanehisa; Morishima, 2016; Aramaki et al., 2020).

**Transcriptomics and proteomics analyses**

The cells of StC were grown for four days at 60 °C under two different conditions: i) autotrophically in a N-FIX solid culture medium with a gas atmosphere ($CO/CO_2/N_2/O_2$) and ii) heterotrophically in R2A culture medium (glucose and peptone as carbon and nitrogen source respectively) solidified with Noble Agar (Sigma-Aldrich). The experiment was conducted in quadruplicate, and RNA and protein were extracted from the same sample using NucleoSpin TriPrep kit (Macherey-Nagel) with some modifications, as follows: 480 µL of 0.5M EDTA and 120 µL of lysozyme (100 mg/mL) were used to lyse the bacterium cells by incubating at 37 °C for 30 minutes, followed by FastPrep grind and homogenization (MP Biomedicals). Total RNA was then purified with the Ambion's DNAse Treatment Kit (Thermo Fisher Scientific) and quantified by spectrophotometry, using NanoDrop 2000 (Thermo Fisher Scientific) and Qubit® dsDNA HS Assay Kit (Thermo Fisher Scientific). The RNA integrity was evaluated by non-denaturing agarose gel electrophoresis (1.5%) RNAse free. Subsequently, approximately 100 ug/ul of the RNA total was used for the library preparation and sequencing on Illumina HiSeq.

For quality control and mapping of reads, fastQ library files were firstly submitted to the software FastQC v.0.11.9 (Brandine et al., 2019) to assess the quality of bases and sequencing artifacts, followed by the library cleaning step performed using Trimmomatic v.0.39 (Imm; Chang, 2023). The library cleaning step was performed using Trimmomatic Ver. 0.39, where possible adapters and low-quality bases were removed from the raw data files. The

following parameters were used: treated as paired-end (PE), ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:keepBothReads to remove adapters and keep both reads as paired, a LEADING and TRAILING value of 20 to remove bases with quality above 20, and all reads with a length less than 36 base pairs were also removed. After the cleaning step, reads from each library were mapped against the *C. thermoautotrophica* StC genome using Bowtie2 Ver. 2.3.4.3 (Li et al., 2023). Mapping text files (SAM) were converted to binary format (BAM) using Samtools Version 0.1.20 (Li et al., 2009), and then used to count mapped reads by genes. HTSeq Ver. 2.0.3 (Siddique et al., 2021) was used to perform the step of counting of mapped reads over the genome, and the expression values were normalized to TPM (Transcript Per Million).

Genes involved in $CO_2$ fixation were used as query sequences to search the genome sequence of StC, using TBlastN and BlastP, respectively, with default parameters. Candidate genes and their translated proteins were further characterized employing the following bioinformatic tools: ClustalW (Thompson et al., 2003) for primary structure similarity relations, InterPro (McDowall; Hunter, 2011) and Pfam (Mistry et al., 2021) for domain and homologous superfamily predictions.

Total proteins were quantified with Pierce™ BCA Protein Assay Kit (ThermoFisher), and a total of 12,5 µg/ml protein solution was subjected to digestion using a thermomixer 35 °C, 900 rpm overnight. The protein samples were prepared following Wiśniewski et al. (2009), with modifications. Firstly, the proteins were reduced with 10 mM DTT and alkylated with 40 mM iodoacetamide. For trypsin digestion, the protein sample was diluted with 0,02 µg/µL trypsin overnight at 35° C, 900 rpm. Subsequently, peptides were desalinated using ZipTip® with resin C18 (Merck Millipore). The obtained peptides were analyzed in an Orbitrap Fusion Lumos (Thermo Fisher Scientific) coupled to an UltiMate™ 3000 UHPLC (Thermo Fisher Scientific). The peptides samples were separated using an Acclaim PepMapTM C18 column (75 um I.D. X 50 cm, 2 µm particle sizes, 100 Å pore sizes) with a flow rate of 300 nl/minute. A 100-minute gradient was established using mobile phase A (0.1% FA) and mobile phase B (0.1% FA in 99.9% ACN): 4-35% B for 60 minute, 15-minute ramping to 90% B, 90% B for 5 minute, and 2% B for 10-minute column conditioning. The sample was introduced into the Orbitrap Fusion Lumos Mass Spectrometer through a Nanospray Flex (Thermo Fisher Scientifc) with an electrospray potential of 2.5 kV. The ion transfer tube temperature was set at 160 °C. The Q-Exactive HF was set to perform data acquisition in Data Dependent Acquisition (DDA) mode. A full MS scan (200-2000 m/z range) was acquired in the Orbitrap at a resolution of 60,000 (at 200 m/z) in a profile mode, a maximum ion accumulation time of 50 minute. Charge state screening for precursor ions was activated. The twenty most intense ions above a

1x106 threshold and carrying multiple charges were selected for fragmentation using higher energy collision dissociation (HCD). The resolution was set as 15,000. Dynamic exclusion for HCD fragmentation was 30 seconds. Other settings for fragment ions included a maximum ion accumulation time of 60 minute, a target value of 2.50×e3, a normalized collision energy at 30%, and isolation width of 1,6.

The raw data from the mass spectrometry were imported to Thermo Proteome Discoverer v. 2.5.400 (Thermo Fisher Scientific) and searched against to the StC protein annotation database to perform protein identification analysis. A search was executed with total peptides and two missed cleavages were considered. Carbamidomethylation (C) was a fixed modification, whereas methionine oxidation and acetylation (protein N-terminal) were selected as variable modifications. The mass tolerance for precursor ions was 10 ppm and 0.05 Da for the fragment ions. The False Discovery Rate (FDR) was fixed for protein and peptide with a cutoff score minor to 1% at the protein, peptide and PSM (peptide-spectrum matches) levels. Proteins were grouped in master proteins using the maximum parsimony principle. The proteins identification was generated with Proteome Discoverer v. 2.5.400 software and the protein expression of the carbon fixation pathways were predicted with BlastKOALA (KEGG).

**Microcompartment (BMC) isolation, purification, and visualization**

Firstly, the proteins involved with shell carboxysome, and its internal proteins were used as query sequences to search the genome sequence of StC. When prospective candidate proteins were identified, its predicted protein sequence was then used to formulate a BlastP (http://www.ncbi.nlm.nih.gov) search of the nonredundant database at NCBI. The protein function, classification of family and predicted domain was provided by InterPro (McDowall; Hunter, 2011). To organize carboxysome gene clusters and to predict the operon and homologous proteins we used the MicrobesOnline database (MicrobesOnline - A website for browsing and comparing microbial genomes) (Alm et al., 2005). The protein sequence of StC was compared with HMM profiles (Hidden Markov Model) using BMC Caller (BMC Caller (msu.edu)) (Sutter; Kerfeld, 2022). The sequence of the BMC-P protein was aligned with 768 genes obtained from PFAM PF03319 using MUSCLE and a phylogenetic tree was built with Phyml using JTT model for identify BMC-P proteins more related with that found in StC strain (Sengupta; Azad, 2021). To extract and purify the BMC proteins from StC strain cells, a protocol was developed. We opted to cultivate the StC strain in a heterotrophic culture medium for this experiment. This choice was driven by the necessity of generating substantial biomass

for microcompartment extraction. As indicated in prior studies (Chaijarasphong et al., 2015; Sommer et al., 2019; Sun et al., 2022), a lot of biomasses is required to obtain a reasonable yield of BMC. In autotrophic conditions, the StC strain exhibits low biomass production, attributed to the absence of an organic source of carbon and nitrogen. Given that the production of bacterial microcompartments (BMC) is not linked to secondary metabolism, we assume that the presence of BMC within the bacterial cell is independent of environmental conditions. The methodology was developed based on previous described protocols for *Synechococcus elongatus* PCC7942 and *Escherichia coli* (Sommer et al., 2019; Flamholz et al., 2020; Oltrogge et al., 2020).

BMC were isolated from bacterial cells grown at 60 °C for 48 hours in R2A culture medium solidified with Noble Agar (Sigma-Aldrich). The cells (~1,4 g) were resuspended in 480 mL of TEMB buffer (5 mM Tris-HCl [pH 8.0], 1 mM EDTA, 10 mM MgCl2, 20 mM NaHCO3), 2 mL B-PER™ II Bacterial Protein Extraction Reagent (Sigma-Aldrich), 120 µL of lysozyme (100 mg/mL) (Sigma-Aldrich), and 50 µL of protease inhibitor (Sigma-Aldrich). Cell lysis was performed in the bath incubator at 37 °C for 1 hour and 30 minutes, followed by 1 hour and 30 minutes at the sonicator bath at 37 °C, and homogenized by FastPrep-24™ (MP Biomedicals) with three cycles of 40 seconds recommended for *Actinomycetota*. The crude cell extract was centrifuged at 4,000 x g for 10 minutes to remove cell debris, and subsequently the supernatant was centrifuged for 20 minutes at 20,000 x g for clarify the sample. The clarified supernatant was centrifuged at 50,000 x g for 1 hour until form a pellet. The pellet was resuspended in 1 mL of TEMB solution and quantified by NanoDrop One/One$^{c}$ Microvolume UV-Vis Spectrophotometers (Thermo Fisher Scientific) in the function protein>other proteins E (extinction coefficient) and MW (molecular weight). The extinction coefficient and molecular weight of the carboxysome shell proteins were obtained from the ProtParam Expasy web tool.

Proteins were also quantified using the Pierce™ BCA Protein Assay Kit (ThermoFisher Scientific™). A fraction of the sample was applied to SDS-PAGE gel to confirm the presence and integrity of the proteins. The pellet was resuspended in 1 ml of TEMB buffer and concentrated with the Amicon® Pro Purification System with 10 kDa (Merck Millipore). After the concentration, the sample was transferred to 15 ml sucrose gradient prepared with TEMB solution in the concentration: 10%, 20%, 30%, 40%, and 50% of sucrose. The gradient was centrifuged at 105,000 x g for 30 minutes, and a white pellet was formed within the concentration range of 10% to 20%. Fractions of 0,5 ml were collected from each gradient and analyzed for the presence of pure carboxysomes via SDS-PAGE gel. The samples were quantified with NanoDrop™ One (ThermoFisher Scientific™) and Pierce™ BCA Protein

Assay Kit (ThermoFisher Scientific™). The fractions containing the carboxysomes were then resuspended in 1 ml TEMB and stored at -30 °C for posteriorly western blot and mass spectrometry analysis.

*Western blotting*

The protein fraction corresponding to BMC was mixed with SDS-PAGE loading buffer 2X (Bio-Rad) (30 µL of sample and 10 µL de SDS loading) and heated at 95 °C in a thermocycler (Benchmark H5000-HC MultiTherm™) for 5 minutes. The samples were electrophoresed on a 4-20% TGX Criterion stain free gel (Bio-Rad) at 120 V for 1 hour with Tris/Glycine/SDS buffer (Bio-Rad). The proteins from SDS-page gel were transferred to a nitrocellulose membrane by electrophoresed at 120 V for 80 minutes with a Tris/Glycine buffer (Bio Rad). Immunoblot analyses were performed using primary specific antibody customized for CcmK/CsoS1 (from StC genome) protein carboxysome shell (rabbit-derived serum) (GenScript) and anti-rabbit igG secondary antibody (Agrisera). The membrane was incubated for 1 minute at ECL prime Western blotting detection (Cytiva Amersham) and the signals were visualized using a transilluminator (ThermoFisher Scientific™). [CcmK/CsoS1 protein sequence:
>MPDLNGKALGLVETLGLVAATEAADAMVKAANVRLVTKQQVGGGLITVI
VAGDVGAVKAAVDAGQTAGSAVGKVVSAHVIPRPHDDIPGILERPPVR].

*LC/MS analysis of BMC*

The fractions containing the carboxysomes were analyzed on an Orbitrap Fusion Lumos Mass Spectrometer (ThermoFisher Scientific™) coupled with an UltiMate™ 3000 UHPLC (ThermoFisher Scientific™), to detect the shell and internal proteins from the microcompartment. Before the protein digestion step, the sample containing the microcompartment was lysed with SDT-lyses buffer (10% w/v SDS, 100 mM Tris/HCl pH 7.6, 100 mM DTT), to disrupt the BMC shell. For lysis, 40 µL of sample (0.22 mg/mL of proteins) was added to 40 µL SDT-lyses buffer, heated at 95 °C for 10 minutes in a thermocycler (Thermo Fisher Scientific), and the lysate was centrifuged at 16,000 x g for 5 minutes. After the cell lysis: I) protein digestion; II) desalination and peptides separation was done and III) peptides were measured on an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific) coupled to an UltiMate™ 3000 UHPLC (Thermo Fisher Scientific); IV) raw MS files were

analyzed using Thermo Proteome Discoverer (version 2.5.400) and searched against the StC protein annotation; V) Search for shell and internal carboxysome proteins.

*Microscopy analysis*

*Optical microscopy of StC cells*: For optical microscopy, a microculture technique in a petri dish was used, according to Su et al. (2012), with adaptations. In a petri dish containing 40 mL of R2A culture medium solidified with high purity Noble Agar (Sigma- Aldrich), a glass coverslip previously sterilized with methanol was added, and the StC strain was inoculated on the top of the coverslip. After two days of cultivation at 60 °C, the coverslip was carefully removed and visualized using an inverted microscope with LED illumination (Leica DM IL LED) coupled to a camera with a CCD sensor for microscopes (Leica DFC3000 G), through which images of the filaments and spores of the StC strain were obtained.

*Negative Stain Electron Microscopy of carboxysome*: for negative stain preparation, 7 µL of purified BMC sample was placed onto glow discharged (15 mA, 30 s) copper grids (Cu, 300 mesh, carbon film, EMS) for 1 minute, then blotted with Whatman blotting paper. Next, ~5 µl of 2% uranyl acetate was added to the grid, and immediately blotted. Then, ~5 µl of 2% uranyl acetate was added again and left to stain for 1 minute. Grids were then blotted to remove excess liquid and air-dried. Negatively stained samples were screened on a Tecnai T12 electron microscope (Thermo Fisher Scientific) operated at an acceleration voltage of 120 kV. Images were taken at different magnifications with defocus ranging from -0.8 µm to -3.0 µm with a US4000 CCD camera (Gatan).

*Cryo-electron tomography (cryo-ET) of carboxysome*: for cryogenic grid preparation, purified BMC samples were plunge-frozen using the Vitrobot Mark IV (Thermo Fisher Scientific) at 4 °C with 100% humidity. Samples were prepared by applying 2-3 µL of protein sample to glow-discharged (30 mA for 60 seconds) gold grids (QuantiFoil Au, 200 mesh, R2/2 grids, EMS) and blotting for 3 seconds, no wait time. The samples were prepared as for cryo-electron microscopy, however, during imaging, a series of images were collected on the same specimen location, but with different stage tilt angle, and these tilted images were used to generate a single tomogram. Tilt-series were collected using Tomography 5 (Thermo Fisher Scientific) with tilt angles from -60° to +60° with 2 degrees steps using a dose-symmetric scheme at a magnification of 33kx. The total electron dose was around 100e/A/sec and the defocus values were between -8 to -4 µm. The tilt-series were collected on a Titan Krios G4 equipped with SelectrisX energy filter and Falcon 4i direct electron detector. The tomograms

were generated using the IMOD software.

*Cryo-EM CEMOVIS*: StC cells was taken from the R2A culture medium plate and scraped them into the CEMOVIS carrier, topped the carrier up with Dextran, and then high-pressure froze the bacteria. This step worked, and the bacteria were properly vitrified. Then thin sections of the vitrified bacteria were made using the diamond knife and put the sections onto the EM grid. Then the EM grid was introduced into the Krios to see the vitrified samples under cryo-TEM. Cryogenic data were collected on a Titan Krios G4 microscope (Thermo Fisher Scientific) operating at 300 kV. The data sets were acquired at a nominal magnification of 50,000x using a Falcon IV direct electron detector camera (Thermo Fisher Scientific).

## RESULTS

### Filamentous morphology of the STC strain

To visualize the StC strain, the bacterium was cultivated on slides submerged in R2A culture medium. Images of the developing bacterium were captured using phase-contrast light microscopy. Figures 1a-b depict the morphology of StC, showcasing its filamentous structure and notable spore production, characteristic of *Actinomycetota*.

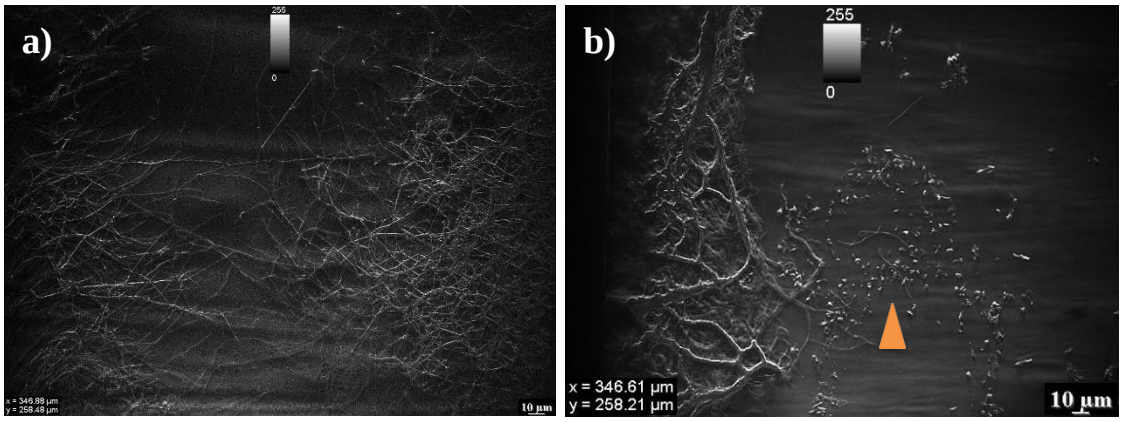


Figure 1: Optical microscopy of StC strain micro-cultive. a) filamentous aggregation structure; b) presence of spore production (orange arrow).

### StC genome features and relatedness indices

The identification of closely related type strains was conducted using the Type (Strain) Genome Server (TYGS) web tool (https://tygs.dsmz.de/user_requests/new) based on the genomic data of *C. thermoautotrophica* StC. The analysis revealed that the StC genome formed

a cluster with the *C. thermoautotrophica* strain UBT1 (*Streptomyces thermoautotrophicus* UBT1), while the *C. thermoautotrophica* strain H1 did not fall within identified species and subspecies clusters. The resulting clusters, comprising 18 species clusters, are illustrated in Figure 2. The StC strain was assigned to one of these species clusters and located within one of the 18 subspecies clusters.

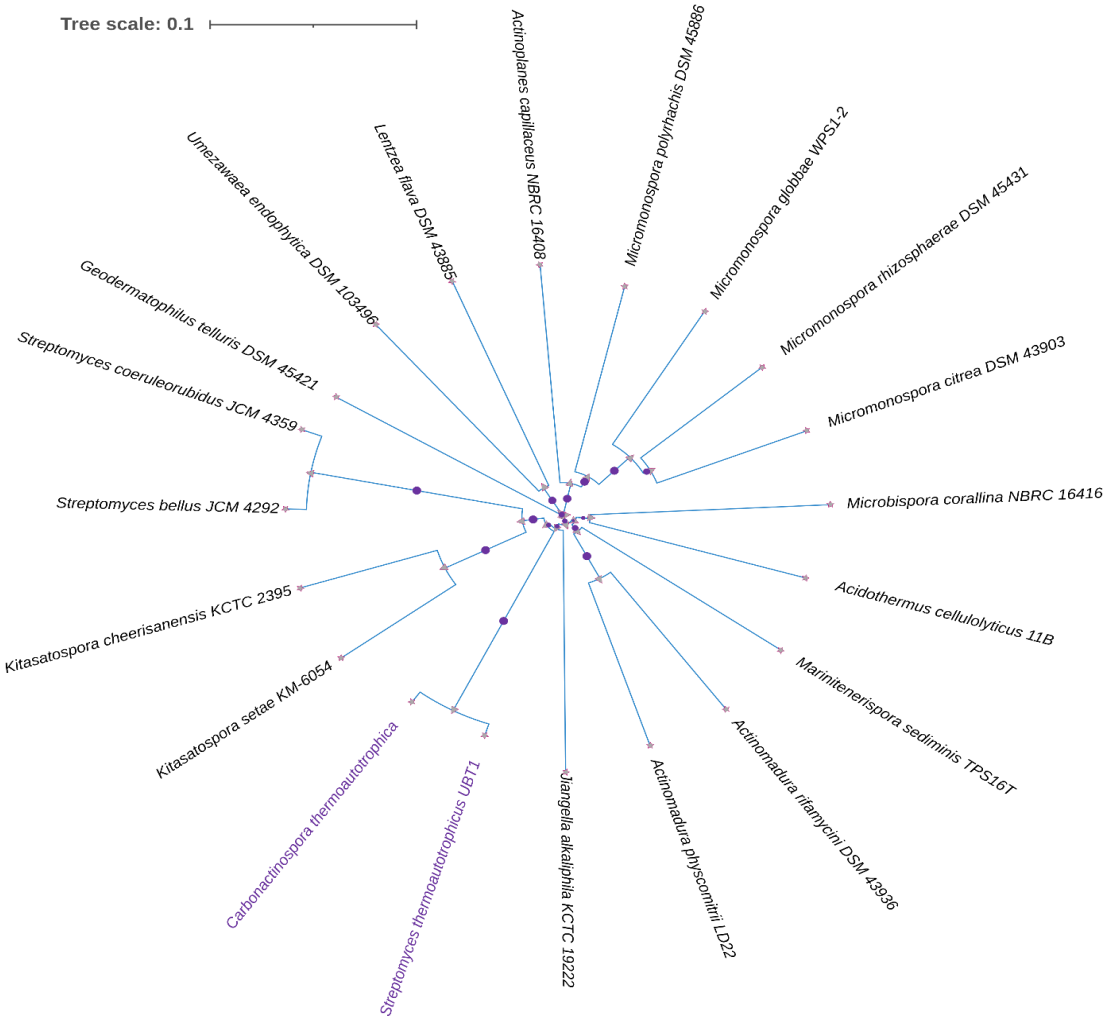



Figure 2: *C. thermoautotrophica* strain StC genome tree. StC strain formed cluster with UBT1 strain. The image was generated by Genome Server (TYGS) (https://tygs.dsmz.de/user_requests/new) and was visualized using iTOL v4.261 (Letunic and Bork, 2016).

The StC strain genome comprises 4,728,204 base pairs with a GC content of 70.95%, encompassing 4,895 coding sequences. Phylogenomic analysis indicated that the StC strain exhibited greater similarity to strain UBT1 (digital DNA-DNA hybridization, dDDH 83.2%) compared to strain H1, which did not fall within identified species and subspecies clusters (Supplementary Table 1). Genome annotation and gene prediction highlighted a remarkably close resemblance between the StC genome and those of *C. thermoautotrophica* UBT1 and H1

(distance ~0.016). Table 1 presents an overview of the key characteristics of the StC strain in comparison to UBT1 and H1 strains.

Table 1: Statistical comparison of information and characteristics of genomes *C. thermoautotrophica* StC, UBT1 and H1 strains. For that purpose, similar genome was generated using the Pathosystems Resource Integration Center (PATRIC) (http://www.patricbrc.org).

| Characteristics | StC | UBT1 | H1 |
|---|---|---|---|
| GenBanck accessions | | JYIJ00000000 | LAXD00000000 |
| Contigs | 1 | 19 | 7 |
| Genome size | 4,728.204 | 5,134.156 | 5,014.930 |
| GC content (%) | 70.95 | 70.97 | 64.85 |
| CDS | 4,895 | 5,187 | 4,708 |
| tRNA | 50 | 50 | 44 |
| rRNA | 8 | 6 | 4 |
| Distance | | 0.0139988 | 0.0164427 |

The phylogenetic analysis involved the utilization of the whole genome of the *C. thermoautotrophica* StC strain, alongside seventeen other bacteria belonging to *Streptomyces*, *Planomospora*, and *Carbonactinospora*, including the *C. thermoautotrophica* strains UBT1 and H1. All genomes were sourced from the NCBI database, and alignments were performed at the most similar level. The Maximum Likelihood method was employed with the Automated Progressive Refinement option selected. In the comparisons, the species *C. thermoautotrophica* displayed the most divergent genome concerning the closest available complete sequences of the same family in the database, suggesting the presence of a new strain. The phylogenetic tree in Figure 3 illustrates that *C. thermoautotrophica* StC forms a clade within its respective genus. Notably, the genome of StC is grouped with the *Carbonactinospora* genus, indicating its affiliation with this genus.

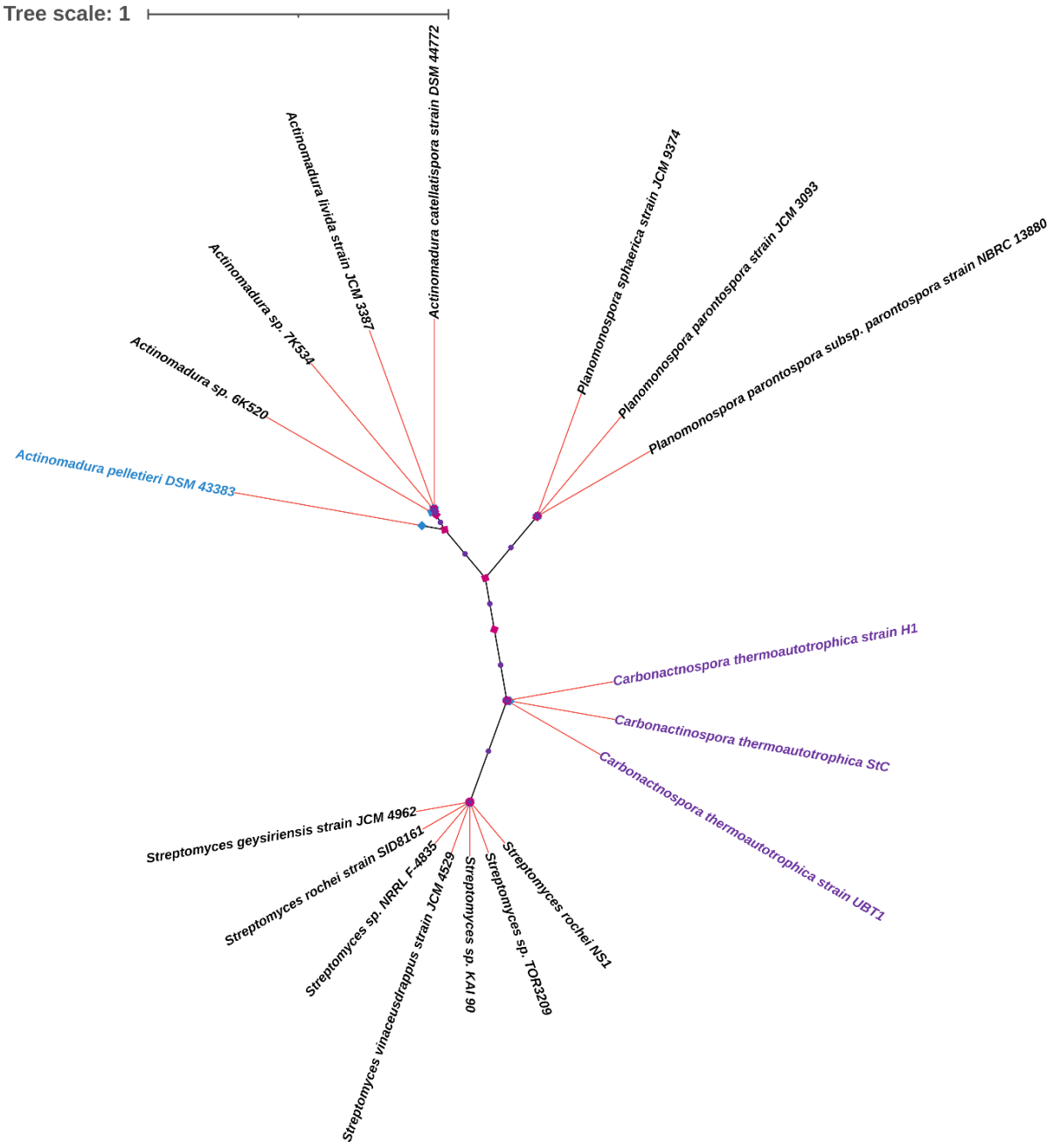



Figure 3: Phylogenic analysis based in the Maximum Likelihood method. The *Carbonactinospora* genus is highlighted in purple, along with the genome of StC strain. The genome of *Actnomadura pelletieri* DMS 43383 (highlighted in blue), members of *Thermomonosporaceae* family was used as an outgroup. The Bootstraps are represented by a purple circle; The Node are by a blue rhombus. The internal symbols are represented by a pink rhombus. The resulting Newick tree file was visualized using iTOL v4.261 (Letunic and Bork, 2016).

To provide additional insights into the genome of StC, a functional annotation of genome sequences was conducted using KEGG with BLASTKOALA. Gene clustering analysis (Supplementary Table 2) identified a total of 1,884 KEGG-associated genes. The top categories included carbohydrate metabolism (225 genes), protein families associated with genetic information processing (217 genes), genetic information processing (182 genes), and amino acid metabolism (151 genes), energy metabolism (142 genes), proteins families: signaling and

cellular processes (136 genes), metabolism of cofactors and vitamins (126 genes), environmental information processes (115 gene), and other categories (590 genes) (Fig. 4).

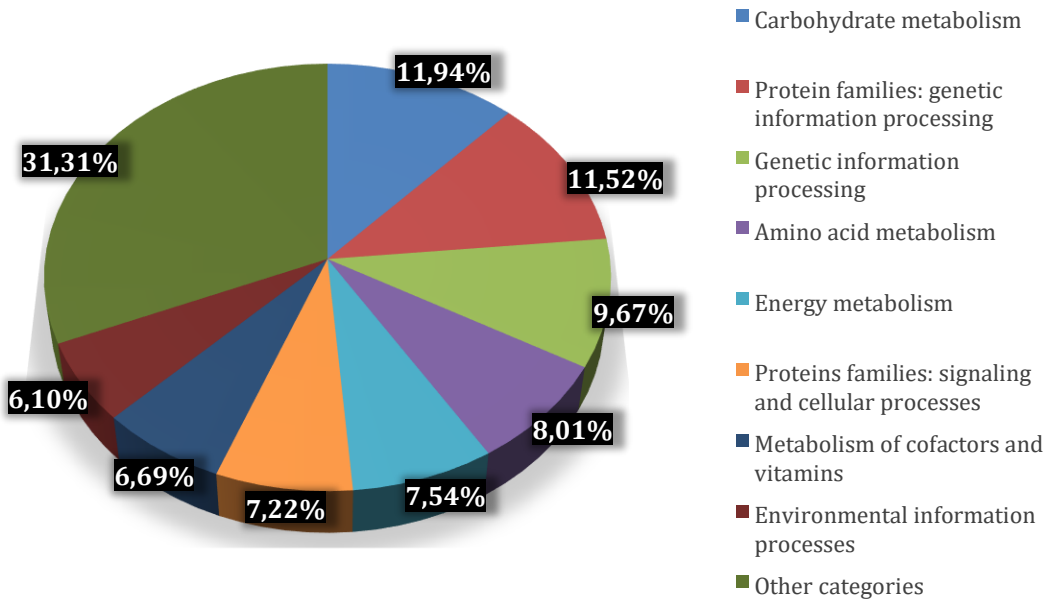



Figure 4: Functional Annotation of *C. thermoautotrophica* StC genes. This figure was prepared by analysis of the whole genome and relative abundance and distribution of KEGG categories plotted using Microsoft Excel.

**Predicted genes involved with CO2 fixation and BMC**

Through genome analysis, genes associated with three natural carbon fixation pathways were identified in the StC strain: the Calvin-Benson-Bassham (CBB) cycle, the reductive tricarboxylic acid (rTCA) cycle, and the 3-hydroxypropionate (3-HP) bicycle. The 11 genes involved in the canonical prokaryotic CBB cycle are present in the StC genome. Similarly, the StC genome harbored a total of 12 genes related to a variant of the rTCA cycle, and 6 genes are predicted for the 3-HP bicycle.

Furthermore, the metabolic pathway reconstruction, conducted using KEGG (KEGG Pathway and GhostKOLA), predicted the presence of enzymes associated with the 3-HP bicycle (citryl-CoA synthetase and citryl-CoA lyase) in the StC genome (Supplementary Fig. 1). The reconstructed metabolic pathway indicated the presence of all 3-HP bicycle genes (Supplementary Fig. 1). Table 2 provides an overview of the genes associated with the three carbon fixation cycles found in the StC genome.

Table 2: Predicted genes involved in CBB cycle, rTCA cycle and 3-Hydroxyproprionate bicycle in StC genoma.

| Cycle | Gene function | KEGG ID |
|---|---|---|
| CBB cycle | RuBisCo form I, large and small chain | [EC:4.1.1.39] |
| | Phosphoribulokinase (PRK) | [EC:2.7.1.19] |
| | Phosphoglycerate kinase | [EC:2.7.2.3] |
| | Glyceraldehyde 3-phosphate dehydrogenase | [EC:1.2.1.12] |
| | Triosephosphate isomerase | [EC:5.3.1.1] |
| | Fructose-bisphosphate aldolase | [EC:4.1.2.13] |
| | Fructose-1,6-Bisphosphatase | [EC:3.1.3.11] |
| | Transketolase | [EC:2.2.1.1] |
| | Ribulose-phosphate 3-epimerase | [EC:5.1.3.1] |
| | Ribose 5-phosphate isomerase | [EC:5.3.1.6] |
| | Sedoheptulose | [EC: 3.1.3.37] |
| | Pyruvate synthase | [EC:1.2.7.1] |
| rTCA cycle | Pyruvate, orthophosphate dikinase | [EC:2.7.9.1] |
| | Pyruvate, water dikinase | [EC:2.7.9.2] |
| | Pyruvate carboxylase | [EC:6.4.1.1] |
| | Phosphoenolpyruvate carboxylase | [EC:4.1.1.31] |
| | Malate dehydrogenase | [EC:1.1.1.37] |
| | Fumarate hydratase, class I | [EC:4.2.1.2] |
| | Succinate dehydrogenase/fumarate reductase | [EC:1.3.5.1] |
| | 2-Oxoglutarate/2-oxoacid ferredoxin oxidoreductase | [EC:1.2.7.3] |
| | Isocitrate dehydrogenase | [EC:1.1.1.42] |
| | Aconitate hydratase | [EC:4.2.1.3] |
| | Succinyl-CoA synthetase alpha subunit | [EC:6.2.1.5] |
| 3-Hydroxyproprionate Bicycle | Acetyl-CoA carboxylase | [EC:6.4.1.2] |
| | Methylmalonyl-CoA/ethylmalonyl-CoA epimerase | [EC:5.1.99.1] |
| | Methylmalonyl-CoA mutase | [EC:5.4.99.2] |
| | Succinyl-CoA synthetase | [EC:6.2.1.5] |
| | Succinate dehydrogenase | [EC:1.3.5.1] |
| | Fumarate hydratase | [EC:4.2.1.2] |

In addition to the genes associated with carbon fixation pathways, the analysis uncovered evidence of gene clusters in the StC strain with the potential to form bacterial microcompartments (BMCs). The StC BMC gene cluster consists of two BMC-P genes located in separate clusters. The BMC-P genes, encoding for CDS 1469144.17.peg.758 (exclusive to StC) and 1469144.17.peg.4339 (shared with UBT1 and H1 strains), were aligned with 768 protein sequences obtained from PFAM PF03319 (EutN_CcmL) using MUSCLE. A maximum likelihood tree was constructed with Phyml using the JTT model (Fig. 5).

While the BMC-P gene (1469144.17.peg.4339) exhibited taxonomic proximity to strain H1 (A0A132NDA6) and *Rhodococcus jostii* RHA1 (Q0RVS4_RHO), the BMC-P gene (1469144.17.peg.758) was more closely related to *Methylibium petroleiphilum* PM1 (A2SNY5_MET) (Fig. 5). RAST annotation identified 1469144.17.peg.758 (exclusive to StC)

as an ethanolamine utilization (EutN) polyhedral-body-like protein EutN. In contrast, 1469144.17.peg.4339, shared with UBT1 and H1 strains, was annotated as a carbon dioxide concentrating mechanism protein CcmL.

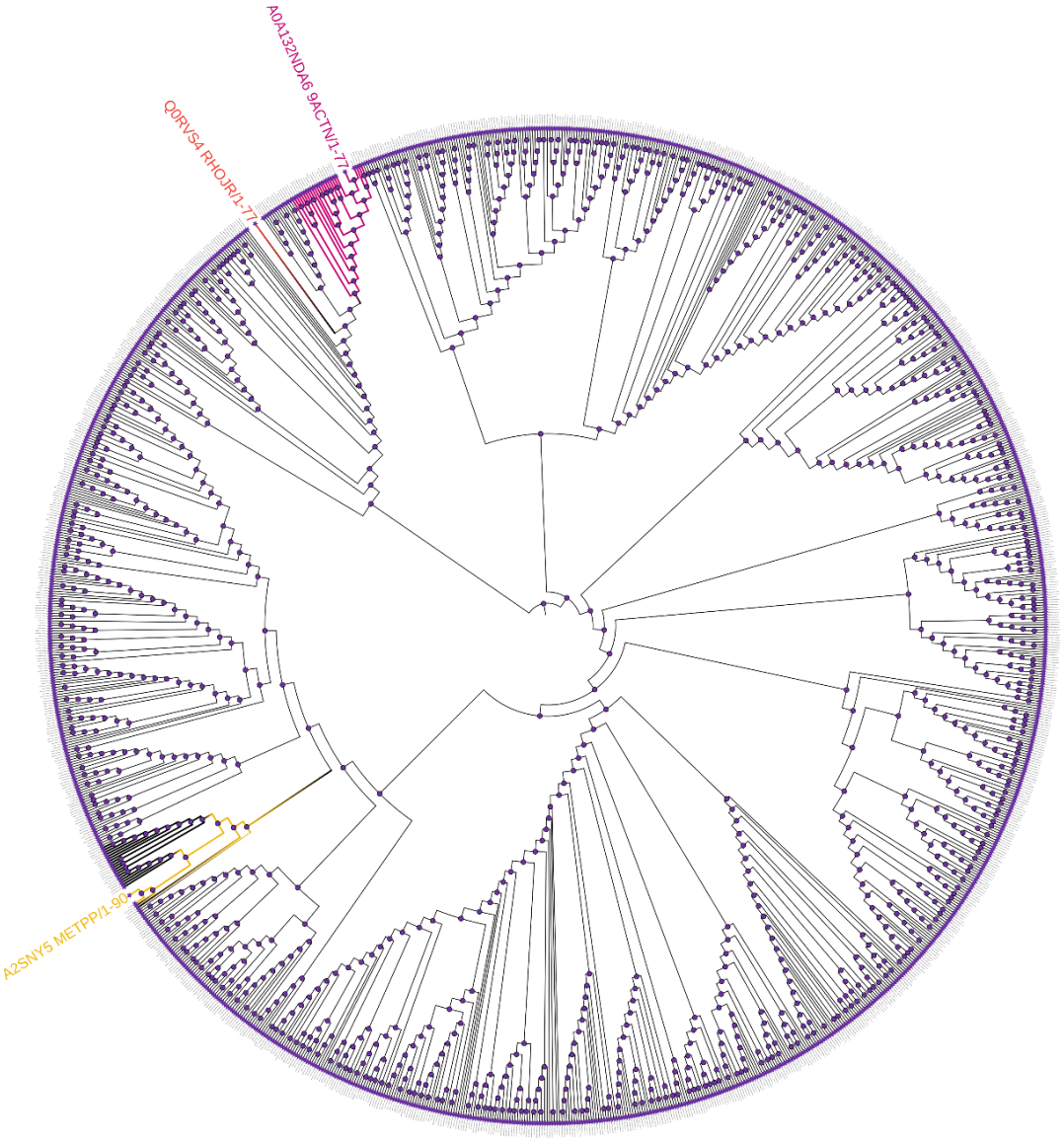



Figure 5: The maximum-likelihood phylogenetic tree generated for BMC-P shell proteins. The CDS (1469144.17.peg.758) exclusive to StC has a BMC-P gene sequence closest to *Methylibium petroleiphilum* PM1 (yellow) the DCS (1469144.17.peg.4339) shared with UBT1 and H1 strains demonstrated the BMC-P gene is closer to H1 strain (pink) and *Rhodococcus jostii* RHA1 (orange). The construction of the ML tree involved the analysis of 768 genes sequences belonging to the BMC-P protein family (Pfam03319).

The StC genome was found to contain two genes encoding for BMC-H shell proteins and three for BMC-$T^{dp}$. The potential presence of a carboxysome in StC is supported by the existence of 9 genes, which encode shell proteins, along with 3 genes encoding enzymes related to RuBisCO (cbbL, cbbS, and CbbX), and 2 genes encoding carbonic anhydrase (β and ɣ)

(Table 3). The BMC-H proteins are characterized by a single Pfam00936 domain, while BMC-T proteins consist of a fusion of two Pfam00936 domains, and BMC-P is formed by Pfam03319 domains. Within the StC genome, four genes were identified containing the Pfam00936 domain, and two genes contained the PF03319 domain, both homologous to the prototypical BMC shell hexamer. Through BlastP analysis aligning the gene sequences containing the Pfam00936 domain, the following similarities were observed: one gene (IKOGDBKN_00107) displayed 91.01% identity with *Chloroflexota carboxysome*; another gene (IKOGDBKN_00291) exhibited 87.50% identity with a Kineosporiaceae BMC domain-containing protein; a third gene (IKOGDBKN_00099) demonstrated 78.02% identity with a BMC domain-containing protein from *Bradyrhizobium*; and the fourth gene (IKOGDBKN_00288) had 82.09% identity with a BMC domain-containing protein from *Hypericibacter terrae*. Regarding the two genes containing the PF03319 domain, one gene (IKOGDBKN_00106) showed 60% identity with a *Nocardioides sp*. EutN/CcmL family microcompartment protein, while the other gene (IKOGDBKN_00290) displayed 78.48% identity with the carbon dioxide concentrating mechanism protein CcmL from the putative uncultured *Nocardioidaceae* bacterium.

Table 3: Potential genes encoding for carboxysome in StC strain. Homologues of all these genes were found in the genomes of StC. This inference is based on matching to profile hidden Markov models (HMMs) in the PFAM and TIGRFAM databases and InterPro.

| Gene ID | Gene function | Domain | Homologous superfamily | PFAM |
|---|---|---|---|---|
| IKOGDBKN_00099 | Major carboxysome shell protein 1A | BMC / CcmK /CsoS1 | CcmK | Bacterial microcompartment domain: CcmK-like superfamily-CcmK/CsoS1 |
| IKOGDBKN_00106 | Ethanolamine utilization protein EutN | EutN_CcmL Ethanolamine | EutN_CcmL | Ethanolamine utilization protein EutN/carboxysome shell vertex protein CcmL-EutN/Ccml superfamily |
| IKOGDBKN_00290 | Hypothetical protein | EutN_CcmL Ethanolamine | EutN_CcmL | Ethanolamine utilization protein EutN/carboxysome shell vertex protein-CcmL-EutN/Ccml superfamily |
| IKOGDBKN_00107 | Hypothetical protein | BMC / CcmK /CsoS1 | CcmK | Bacterial microcompartments protein, conserved site |

| | | | | |
|---|---|---|---|---|
| | | | | CcmK-like superfamily CcmK/CsoS1 |
| IKOGDBKN_00110 | Hypothetical protein | BMC_CP | Ccmk_like | Bacterial microcompartment domain-CcmK-like superfamily-(BMC) circularly permuted domain |
| IKOGDBKN_00288 | Hypothetical protein | BMC | CcmK | Bacterial microcompartment domain-CcmK-like superfamily-Bacterial microcompartment (BMC) circularly permuted domain |
| IKOGDBKN_00289 | Hypothetical protein | None predicted | None predicted | EutN/CcmL family microcompartment protein |
| IKOGDBKN_00291 | Propanediol utilization protein PduA | BMC / CcmK | CcmK | Bacterial microcompartment domain-CcmK-like superfamily-CcmK/CsoS1 |
| IKOGDBKN_00292 | Hypothetical protein | BMC | CcmK | CcmK-like superfamily-Bacterial microcompartment (BMC) circularly permuted domain |
| IKOGDBKN_03845 | Carbonic anhydrase | Carbonic anhydrase | Beta-Carbonic anhydrase | Beta-Carbonic anhydrase |
| IKOGDBKN_01139 | Gama Carbonic anhydrase | Gama Carbonic anhydrase | Gama Carbonic anhydrase | Gama Carbonic anhydrase |
| IKOGDBKN_03889 | Ribulose bisphosphate carboxylase large | RuBisCo | RuBisCo large subunit | Ribulose bisphosphate carboxylase, large subunit |
| IKOGDBKN_03890 | Ribulose bisphosphate carboxylase small | RuBisCo | RuBisCo small | Ribulose bisphosphate carboxylase small |
| IKOGDBKN_03891 | CbbX | CbxX/CfxQ_ATPASE_AAA_core | CbxX/CfxQ P-Loop_NATPae | CbxX/CfxQ |

The protein sequences of the StC genes containing the Pfam00936 and PF03319 domains were utilized to determine the functionality of the related genes. In the operon viewer, it became evident that the gene IKOGDBKN_00107 (CcmK/CsoS1) is strongly associated with CcmL genes, featuring an intergenic distance of forty. The CcmK/CsoS1 from StC exhibited

three clusters, each associated with an operon. One operon contained genes for malate/lactate dehydrogenase-like protein, ethanolamine utilization protein EutE, and deoxycytidylate deaminase-like protein, with an intergenic distance of 176. Another operon included genes related to a putative PTS system, DNA (apurinic or apyrimidinic site) lyase, and a putative metallo-beta-lactamase family protein. A third operon featured genes cbbL, cbbS, cbbX, cbbY, cbbE, cbbZ, cbbF, cbbP, cbbT, cbbG, cbbK, and cbbA. Additionally, the CcmK/CsoS1 from StC demonstrated 87.06% homology with shell protein H from *Desulfitobacterium hafniense* Y51 and 84.95% identity with putative carboxysome structural-like protein from *Methylibium petroleiphilum* PM1.

The gene encoding bacterial microcompartment (IKOGDBKN_00291) forms a cluster with an operon containing genes for a hypothetical protein and aminoglycoside phosphotransferase. This gene (IKOGDBKN_00291) demonstrates 95% homology with microcompartment protein H from *Mycobacterium flavescens* PYR-GCK.

The gene IKOGDBKN_00099, encoding the major carboxysome shell protein, forms an operon with nine genes. These include three genes related to the carbon dioxide concentrating mechanism CcmK, one gene related to propanediol utilization protein, one hypothetical protein, one uncharacterized conserved protein, one gene related to NAD-dependent aldehyde dehydrogenases, one alcohol dehydrogenase, and one predicted NADH ubiquinone oxidoreductase. It is interesting to note the homology (100%) of IKOGDBKN_00099 with shell proteins from *Yersinia intermedia* ATCC 29909 and *Yersinia frederiksenii* ATCC 33641 (70,93%). Additionally, the absence of matches for IKOGDBKN_00106 (containing the PF03319 domain - Ethanolamine utilization protein EutN/carboxysome shell vertex protein CcmL) in MicrobesOnline Genomes suggests a potentially unique or less characterized feature in the context of available genomic data.

The gene IKOGDBKN_00290, encoding Ethanolamine utilization protein EutN/carboxysome shell vertex protein CcmL, forms a cluster with four genes in the same operon. These include one hypothetical protein, one aminotransferase, one gene related to aldehyde dehydrogenase, and one gene for a microcompartment protein.

Using the BMC Caller, the StC genome was identified as matching a type of BMC termed metabolosome with incomplete core (MIC), with less than 80% HMM match (Supplementary Fig. 2). This type of metabolosome is exclusive to members of the *Actinomycetota* phylum. Unlike the conventional metabolosome, the MIC microcompartment does not encode a complete biochemical pathway for a functional BMC, and the enzyme signature is unknown. The MIC metabolosome lacks a phosphotransacetylase (PTAC) and a

signature enzyme that generates the aldehyde; however, these may be encoded elsewhere in the genome. This analysis indicated the presence of two BMC-P, three BMC-H, and two BMC-$T^{dp}$ shell proteins.

**Transcriptomic and Proteomic Data**

In the current study, we employed a comprehensive approach, integrating genomics, transcriptome and proteome analyses, to elucidate the overexpression patterns of genes and proteins associated with carbon fixation pathways. The study focused on monitoring transcriptomic and proteomic responses under two divergent culture conditions for *C. thermoautotrophica strain* StC: autotrophic growth in an atmosphere of $CO/CO_2/N_2/O_2$ (referred to as N-FIX) and heterotrophic growth in R2A culture medium already supplemented with organic sources of carbon and nitrogen (referred to as R2A).

In the transcriptome analysis, four RNA-Seq libraries were generated for each growth condition. Following the filtration of raw reads, an average of 23,515,190.25 clean reads was obtained for the growth condition with gases (N-FIX), while 28,668,253.25 reads were obtained for the growth condition without gases (R2A). The clean reads were then mapped to the genome sequence of the StC strain, achieving mapped rates of at least 21,849,638.25 for the N-FIX condition and 108,253,231.00 for the R2A condition.

In the proteomic analysis, a total of 240,412 spectra generated from four biological replicates of the two growth conditions were examined using Proteome Discoverer 2.5. Among them, 18,039 spectra were matched with proteins of StC, encompassing 6,496 peptides. Subsequently, 2,052 proteins were identified in the N-FIX group, and 1,569 proteins were identified in the R2A group. In both groups, the functional categories with the most proteins included carbohydrate metabolism, genetic information processing, amino acid metabolism, metabolism of cofactors and vitamins, and energy metabolism. Both culturing conditions exhibited functional enrichment with genes and proteins related to the CBB cycle, rTCA cycle, and other carbon fixation pathways such as the 3-HP bicycle, Dicarboxylate-hydroxybutyrate, Wood-Ljungdahl pathway, and BMC, which were upregulated.

**Comparative omics suggest two CO2 fixation pathways.**

To discern the $CO_2$ fixation pathway's identity beyond genome annotation, we adopted a comprehensive approach integrating genome, transcriptome, and proteome analyses. Initially,

the StC genome was scrutinized to identify potential genes associated with carbon fixation. Subsequently, a comparative analysis was conducted on the transcriptome and proteome of the bacterium under autotrophic and heterotrophic growth conditions. This approach aimed to provide a more comprehensive understanding of the $CO_2$ fixation pathway in StC.

Consistently, we observed a significant upregulation of several genes and proteins belong to carbon fixation under both autotrophic and heterotrophic conditions. These pathways include the rTCA cycle, CBB cycle, 3-Hydroxypropionate bicycle, Hydroxypropionate-hydroxybutyrate cycle, Dicarboxylate-hydroxybutyrate, and Wood Ljungdahl pathway. Nevertheless, StC possesses a comprehensive set of genes and proteins specifically related to the Calvin-Benson-Bassham (CBB) cycle and a variant of the reductive citrate cycle (rTCA), ensuring the presence of all necessary components for an effective carbon fixation process by these two cycles. Specifically, genes encoding enzymes such as RuBisCo form IA (large and small chains), phosphoribulokinase (PRK), phosphoglycerate kinase, glyceraldehyde 3-phosphate dehydrogenase (GAPDH), triosephosphate isomerase, fructose-bisphosphate aldolase, fructose-1,6-bisphosphatase, transketolase, ribulose-phosphate 3-epimerase, and ribose 5-phosphate isomerase were highly overexpressed in both autotrophic and heterotrophic conditions (Fig. 6a). These findings strongly suggest that StC possesses the metabolic potential for carbon fixation through the CBB cycle.

The transcriptome analyses also revealed the expression of genes encoding a variant of the rTCA cycle in both autotrophic and heterotrophic cultivation conditions (Fig. 6b). In StC cultivated under autotrophic conditions, all genes and proteins associated with this variant of the rTCA cycle were expressed. However, for StC cells cultivated heterotrophically, the proteins phosphoenolpyruvate carboxylase and 2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase was downregulated (Supplementary Fig. 3), despite the corresponding genes being upregulated in both heterotrophic and autotrophic conditions. It is noteworthy that although the gene encoding pyruvate synthase found in the StC genome was upregulated in both cultivation conditions, the abundance of this protein was not observed in either autotrophic or heterotrophic conditions.

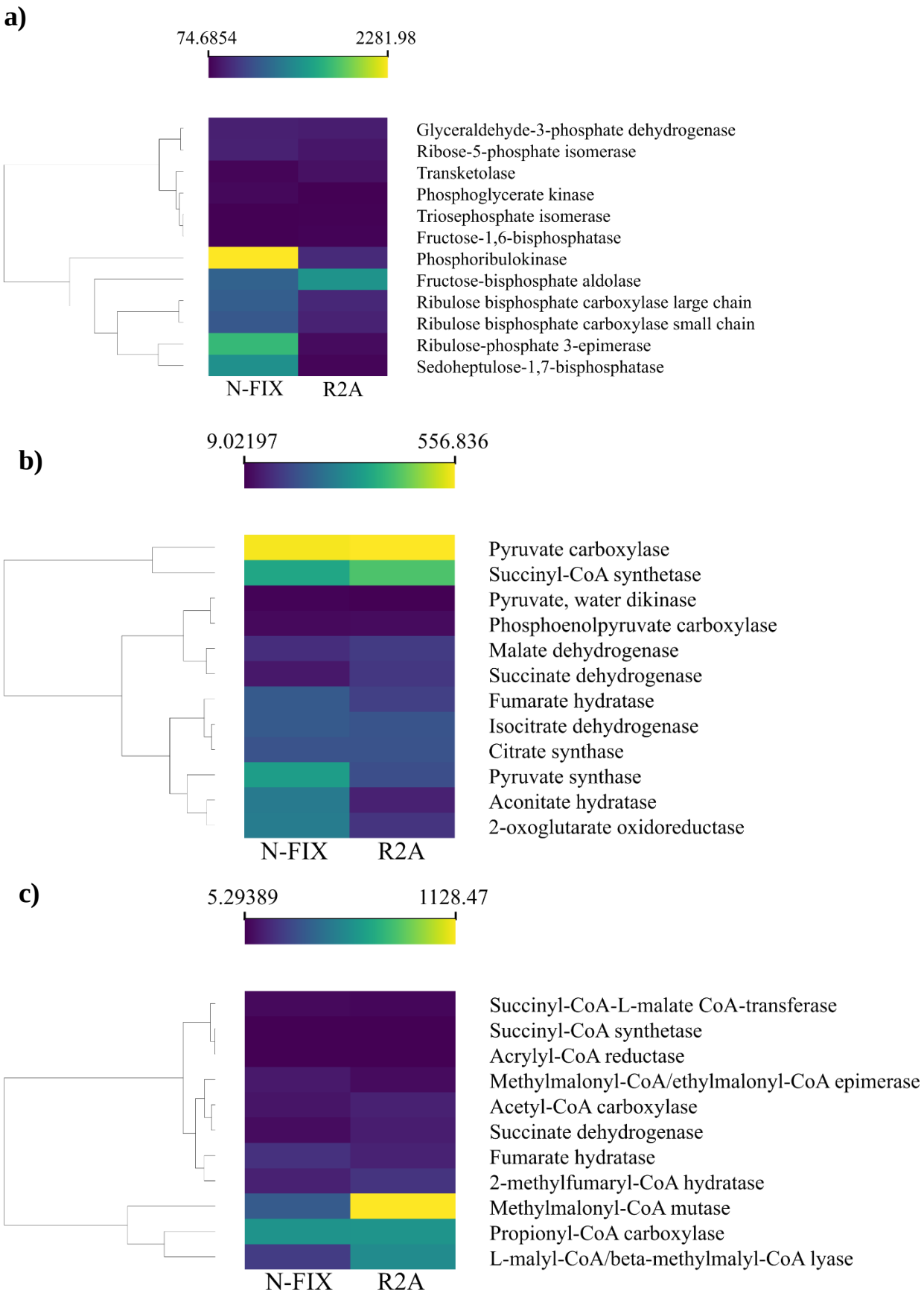
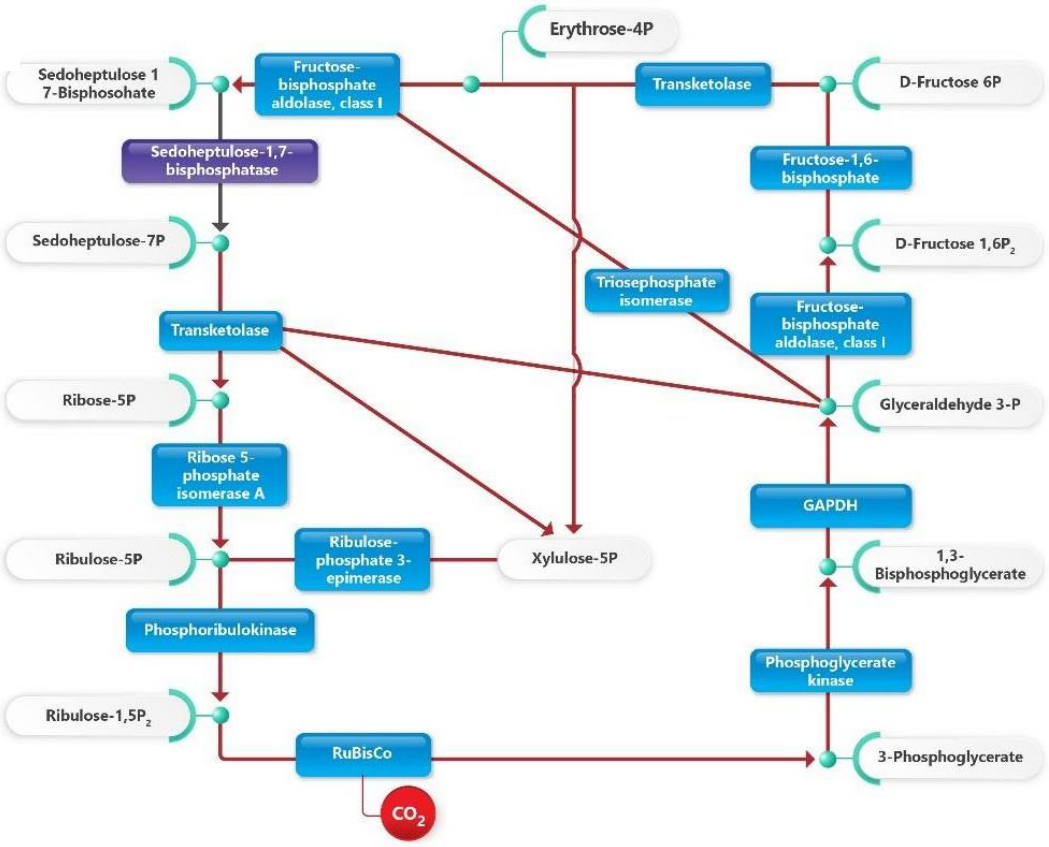



Fig 6: Related carbon fixation genes expressions in StC cell. Heatmaps of the genes belonging to CBB (a), rTCA cycle (b) and 3-HP; carbon fixation pathways in the StC strain cultivated autotrophically (N-FIX) and heterotrophically (R2A).

In addition to the upregulation of genes associated with the CBB cycle, the corresponding proteins encoded by these genes were also markedly upregulated in StC cells grown under both autotrophic and heterotrophic conditions. Utilizing BlastKOALA to predict carbon fixation pathways, we observed the expression of all proteins belonging to the CBB cycle, except for sedoheptulose-1,7-bisphosphatase, in the two contrasting cultivation conditions (Fig. 7a).

**a)**



**b)**

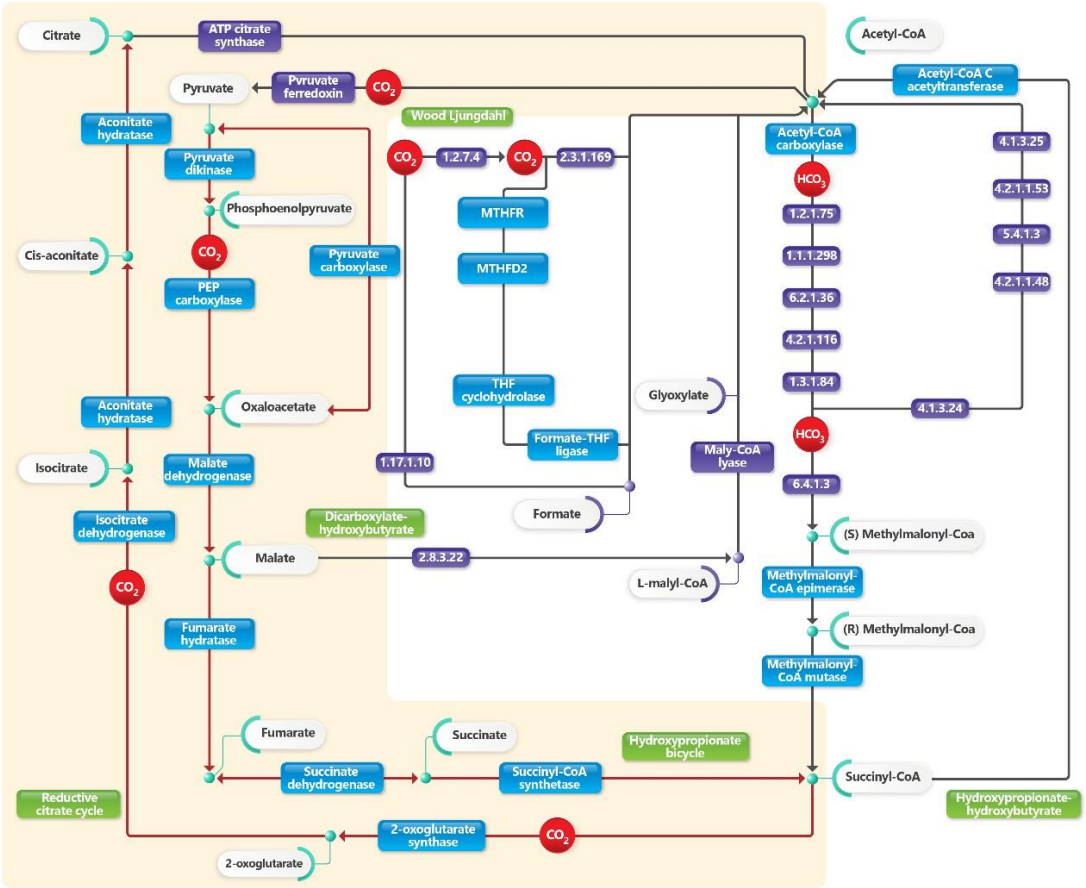



Fig 7: Proteins predicted for carbon fixation in the StC strain using BlastKOALA. a) CBB cycle; b) rTCA cycle. StC cultivated in autotrophic conditions. Abundant proteins in blue and no abundant proteins in purple; the red line represents the active reaction and black line the not active reaction. The highlight in yellow represents the rTCA reaction. Figure adapted from BlastKOALA.

Based on our integrated omics analysis, it can be inferred that the StC strain has the potential to fix carbon through the CBB cycle and may also utilize the rTCA pathway for carbon fixation. The strain possesses a functional rTCA cycle, where pyruvate can serve as a substrate to form phosphoenolpyruvate through the action of pyruvate dikinase enzyme. This initial reaction fixes the first $CO_2$, and subsequently, phosphoenolpyruvate carboxylase (PEP-carboxylase) converts phosphoenolpyruvate to oxaloacetate. Alternatively, pyruvate can be directly converted to oxaloacetate by pyruvate carboxylase. The cycle continues with the conversion of oxaloacetate to malate, and the enzyme malate dehydrogenase catalyzes this reaction. Subsequently, fumarate hydratase condenses malate to fumarate, which is then converted to succinate and succinyl-CoA. The enzymes succinate dehydrogenase and succinyl-CoA synthetase catalyze these reactions, respectively. In the next step, the second $CO_2$ is introduced into the cycle, and succinyl-CoA is converted to 2-oxoglutarate by the enzyme 2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase. Finally, the cycle can either stop with 2-oxoglutarate as a product or proceed to the production of citrate, which can initiate the cycle as

acetyl-CoA or oxaloacetate. In the scenario where the cycle proceeds to produce citrate, 2-oxoglutarate is converted into isocitrate, and a third $CO_2$ molecule is fixed. This is followed by the formation of cis-aconitate and then citrate. The enzymes isocitrate dehydrogenase and aconitate hydratase catalyze these reactions. However, it is worth noting that the enzymes citryl-CoA synthetase and citryl-CoA lyase, responsible for converting citrate to acetyl-CoA, are missing in the StC genome. Nevertheless, the metabolic pathway reconstruction predicts the presence of these enzymes. The expressed proteins for the rTCA cycle and the associated reactions in StC are depicted in Fig. 7b.

The metabolic reconstruction indicated the presence of all genes required for a third cycle, the 3-Hydroxypropionate cycle. However, only a few genes for the corresponding proteins were expressed in both autotrophic and heterotrophic growth conditions. The expressed genes include acetyl-CoA carboxylase, methylmalonyl-CoA/ethylmalonyl-CoA epimerase, methylmalonyl-CoA mutase, succinyl-CoA synthetase, succinate dehydrogenase flavoprotein, fumarate hydratase, acrylyl-CoA reductase, propionyl-CoA carboxylase, succinyl-CoA-L-malate CoA-transferase, L-malyl-CoA/beta-methylmalyl-CoA lyase, and 2-methylfumaryl-CoA hydratase (Fig. 6c). However, the proteome analysis revealed that, among these upregulated genes, only the proteins related to acetyl-CoA carboxylase, methylmalonyl-CoA/ethylmalonyl-CoA epimerase, methylmalonyl-CoA mutase, succinyl-CoA synthetase, succinate dehydrogenase flavoprotein, and fumarate hydratase were abundantly present in StC cells (Fig. 7b).

**Purification and single-particle analysis of carboxysomes from the strain StC**

StC is a thermophilic chemoautotrophic bacterium with mixotrophic capabilities, exhibiting slow growth and producing low biomass when subjected to autotrophic cultivation. As carboxysome formation is not a product of secondary metabolism, it could potentially be induced in heterotrophic conditions. To isolate the carboxysome, we cultivated StC heterotrophically in the R2A culture medium.

Native carboxysomes were isolated from StC using sucrose gradient ultracentrifugation, and the structures were enriched in the 10%-50% sucrose fraction. In the fraction between 10% and 20%, a white pellet is visible (Supplementary Fig. 4). SDS-PAGE of 10%-50% fractions demonstrated the presence of carboxysome components cbbL, cbbS, and CcmK/CsoS1 (Fig. 8a), and immunoblot analysis confirmed the presence of CcmK/CsoS1 protein (Fig. 8b). Mass spectrometry analysis of the isolated carboxysome indicated the presence of only the cbbL

protein. However, the proteomic data of the StC cells cultivated autotrophically indicated the abundance of cbbL, cbbS, CbbX, and CcmK/CsoS1 proteins.

With Cryo-EM CEMOVIS, structures that could potentially be BMC and broken membrane fragments with disassembled BMC outside of the filamentous were observed (Fig. 8c). Negative-stain EM showed that the isolated BMC have a canonical polyhedral BMC shape, with an average diameter of 60 nm (Fig. 8d). The samples containing the potential carboxysomes were then subjected to single-particle cryo-ET analysis to study their 3D architecture, and the initial screening showed a heterogeneous sample, containing structures strongly resembling carboxysomes (Fig. 8e).

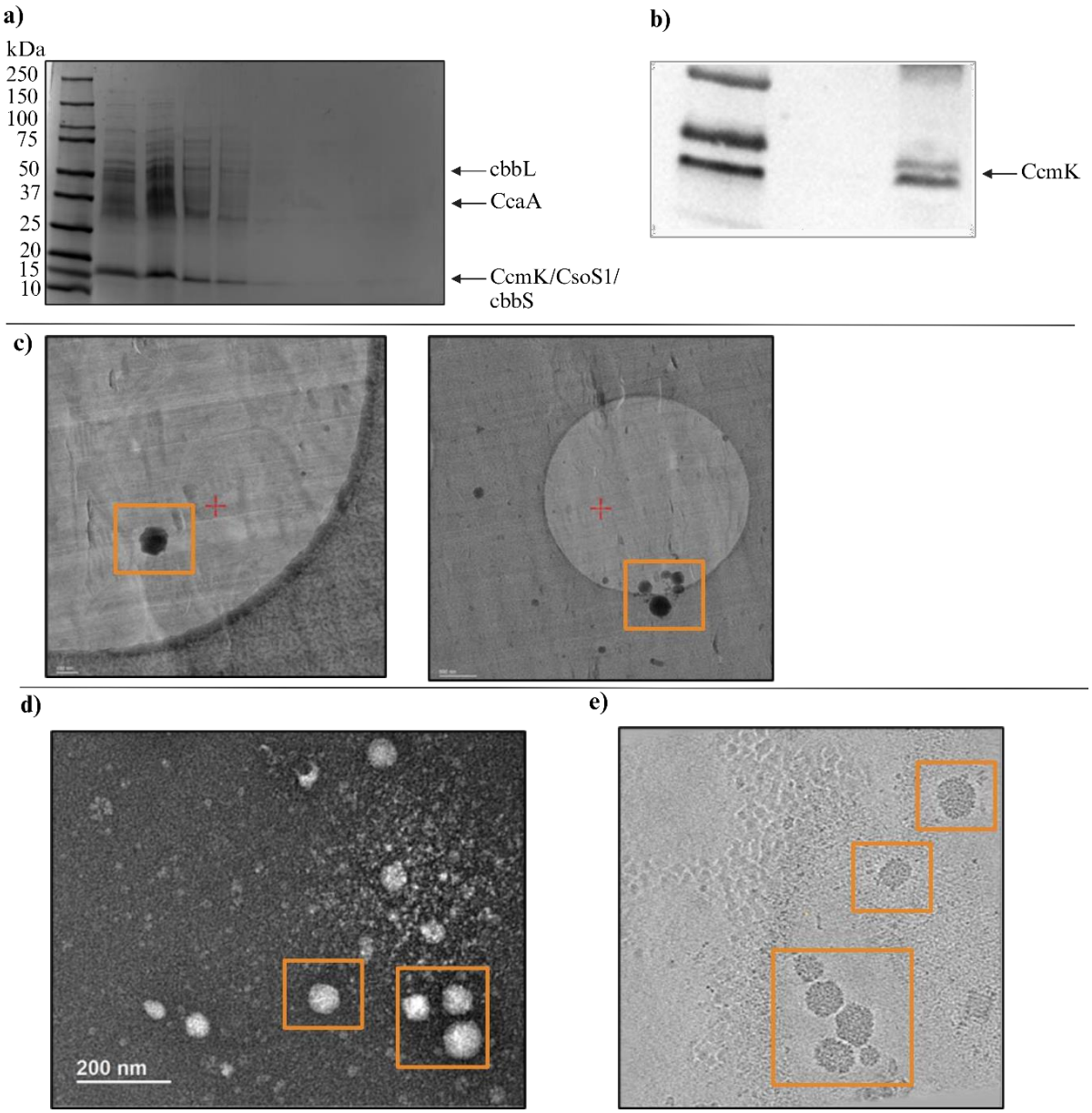



Figure 8: Purification and microscopy analysis of the StC microcompartment. (a) SDS-PAGE of the purified carboxysomes, bands for proteins rbcL, CcaA (carbonic anhydrase) and CcmK proteins. (b) Western blotting of

the purified a-carboxysome complex, using antibodies raised against peptides from CcmK, confirming the presence of protein. (c) Cryo-EM CEMOVIS of frozen-hydrated StC cell samples with visible structure of carboxysome or metabolosome (orange box). Scale bar: 100 nm (left); 500 nm (right). (d) Negative-stain EM of carboxysome samples from StC (orange box) Scale bar: 200 nm. f) Cryo-ET of frozen-hydrated carboxysome samples (orange box) Scale bar: 100 nm.

## DISCUSSION

Was the origin of life heterotrophic or autotrophic? The heterotrophic origin life is the prevailing theory. However, if the autotrophic arrived later, the carbon fixation in chemoautotrophic was the first biogeochemical cycle to emerge (Popper, 1990). There is great interest and advances in studies on chemoautotrophic bacteria to understand the physiology, biochemistry and molecular biology of the $CO_2$ fixation process, regulation, and interaction between the pathways.

Currently, there are seven known carbon fixation pathways, where some bacteria may possess genes for more than one carbon cycle (Thiel et al., 2017). In this study, the high-quality genome annotation, represented by a single contig, suggests that the cultured bacterium *Carbonactinospora thermoautotrophica* strain StC has genes for two functional carbon fixation pathways: the Calvin-Benson-Bassham (CBB) cycle and the reductive tricarboxylic acid (rTCA) cycle. To confirm this, transcriptome and proteome analyses were conducted, revealing the cooccurrence of the complete set of genes and related proteins for both the CBB cycle and rTCA cycle. In addition, we isolated a BMC related to a carboxysome microcompartment in from StC.

### *Genomic insights*

The Calvin-Benson-Bassham (CBB) cycle is the predominant carbon fixation pathway found in plants, algae, cyanobacteria, as well as photo- and chemoautotrophic bacteria. The enzyme responsible for catalyzing the $CO_2$ fixation reaction in this pathway is ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo), which is the most abundant protein in ecosystems and plays a crucial role in the global carbon cycle and crop production (Andersson, 2008). The cbb genes, encoding RuBisCo, are organized in cbb operons, which may vary in size and composition, though they exhibit conserved features (Bowien; Kusian, 2002). In the genome of *Carbonactinospora thermoautotrophica* strain StC, the cbb operon consists of cbbL, cbbS, and cbbX genes. The cbbX gene is positioned adjacent to the cbbS gene. The cbbX gene encodes a red type form I (L8S8) RuBisCO, which, despite having unknown functions, is

crucial for the autotrophic growth of the organism (Watson; Tabita 1997).

In the carbon dioxide fixation process through the Calvin cycle, RuBisCo plays a crucial role by catalyzing the carboxylation of ribulose-1,5-bisphosphate (RuBP), resulting in the formation of two molecules of 3-phosphoglycerate. To regenerate RuBisCo, several enzymes in the Calvin-Benson-Bassham (CBB) cycle are involved. Three glycolytic enzymes: phosphoglycerate kinase, glyceraldehyde-3-phosphate dehydrogenase, and triosephosphate isomerase, work together to convert two molecules of 3-phosphoglycerate into glyceraldehyde-3-phosphate. Subsequent rearrangement reactions lead to the production of ribulose-5-phosphate. Transketolase is responsible for converting glyceraldehyde-3-phosphate and dihydroxyacetone phosphate into xylulose-5-phosphate and erythrose-4-phosphate. This process is followed by the formation of xylulose-5-phosphate and ribulose-5-phosphate from dihydroxyacetone phosphate and erythrose-4-phosphate. Sedoheptulose-1,7-bisphosphatase (SBPase) then catalyzes the dephosphorylation of sedoheptulose-bisphosphate (SBP). The last step involves the conversion of ribulose-5-phosphate back to RuBP, with phosphoribulokinase (PRK) phosphorylating ribulose-5-phosphate (Berg, 2011).

In the genome of StC, the absence of the sedoheptulose-1,7-bisphosphatase (SBPase) protein expression suggests an alternative to the pentose phosphate cycle. In this alternative scenario, the SBPase enzyme is replaced by the transketolase enzyme. Transketolase facilitates the transfer of two carbons from fructose-6-phosphate to glyceraldehyde 3-phosphate, producing xylulose-1,5-bisphosphate, and erythrose-4-phosphate. In the subsequent step, transketolase acts again to form ribose-5-phosphate and xylulose-5-phosphate. This cycle, referred to as the transaldolase cycle, operates without the participation of SBPase.

The transaldolase cycle represents a variant of the canonical Calvin-Benson-Bassham (CBB) cycle, where SBPase is replaced by transketolase. The main distinction is that the variant renders the second aldolase-phosphate transaldolase (APT) unit reversible. This proposed CBB cycle variant with the absence of SBPase is known as the transaldolase cycle, and its existence was suggested based on observations that the Calvin-Benson cycle did not become fully labeled in experiments with $14CO_2$ or $13CO_2$ in photosynthesizing leaves. Additionally, the finding that a high cyclic photosynthetic electron flow phenotype resulted from a lack of the enzyme fructose bisphosphatase supported this idea. The variation of the CBB cycle with the absence of SBPase has been demonstrated in *Thermithiobacillus tepidarius* and *Thiobacillus thioparus* (Livingston et al., 2010; Boden et al., 2016; Hutt et al., 2017).

*Genomics, transcriptomics, and proteomics insights into dual carbon fixation in a single cell*

More recently, a variant of the Calvin-Benson-Bassham (CBB) cycle was demonstrated in the chemolithoautotrophic bacterium *Thermodesulfobium acidiphilum* using multiomics and biochemical approaches (Frolov et al., 2019). In the case of the chemoautotrophic StC strain, we have provided evidence of genes and proteins expression, as well as cultivation of the strain in a culture medium without organic carbon sources, using only CO, $CO_2$, $O_2$ and $N_2$ gases. These findings further support the existence of a variant of the CBB cycle in StC, emphasizing its adaptability to chemoautotrophic conditions. It is interesting to note that StC also can grow without an organic source of nitrogen, only with $N_2$ gas. This ability also was observed in UBT1 and H1 strain (Gadkari et al., 1990), but all the three strains of *Carbonactinospora thermoautotrophica* (StC, UBT1 and H1) do not have the conventional enzyme to fix nitrogen, other enzyme not described until the moment could catalyze this reaction.

Genes and proteins specific to the reductive tricarboxylic acid (rTCA) pathway were also observed in StC. The rTCA cycle represents a reversal of the tricarboxylic acid (TCA) cycle, also known as the Krebs cycle. While the Krebs cycle occurs in heterotrophic organisms to oxidize acetyl-CoA to $CO_2$ for generating reducing power and ATP synthesis, the rTCA cycle operates in the reverse direction, producing acetyl-CoA from two molecules of $CO_2$ (Hugler; Sievert, 2011). In the rTCA cycle, there are two carboxylation reactions: I) The first carbon fixation occurs with the reductive carboxylation of succinyl-CoA to 2-oxoglutarate, catalyzed by 2-oxoglutarate synthase. II) The second carboxylation occurs with reductive carboxylation of 2-oxoglutarate to isocitrate. This second carboxylation can be accomplished either by isocitrate dehydrogenase or by the enzymes 2-oxoglutarate carboxylase and oxalosuccinate reductase, with oxalosuccinate as a free intermediate (Ju et al., 2023). The rTCA cycle with the production of isocitrate-by-isocitrate dehydrogenase has been demonstrated in *Chlorobium limicola*, and the pathway involving 2-oxoglutarate carboxylase and oxalosuccinate reductase, with oxalosuccinate as an intermediate, has been described in *Hydrogenobacter thermophilus* (Kanao et al., 2002; Aoshima, 2007).

In the StC strain, the rTCA pathway appears to involve the first and second carboxylation reactions. The first carbon fixation is catalyzed by 2-oxoglutarate synthase, converting succinyl-CoA to 2-oxoglutarate. The second carboxylation is facilitated by isocitrate dehydrogenase, which reduces 2-oxoglutarate to isocitrate. Some bacteria, such as *Mycobacterium tuberculosis, Geobacter sulfurreducens*, and Chlorobium *limicola*, utilize an additional step to fix a third carbon in the rTCA cycle (Holmes et al., 2017; Hu et al., 2020). This step involves the ATP-dependent cleavage of citrate to acetyl-CoA and oxaloacetate. The reaction can be performed by ATP citrate synthase or by the combined action of citryl-CoA

synthetase and citryl-CoA lyase (Kitadai et al., 2017). To complete the process of obtaining pyruvate from acetyl-CoA in the rTCA cycle, another carboxylation reaction is required, and this reaction is catalyzed by pyruvate synthase (Evans et al., 1966).

Despite the fact that the genes for these enzymes (ATP citrate synthase, citryl-CoA synthetase, and citryl-CoA lyase) were not annotated in the StC genome, the metabolic pathway reconstruction predicted the presence of citryl-CoA synthetase and citryl-CoA lyase enzymes. The citryl-CoA synthetase (CCS) enzyme is composed of two subunits that share sequence similarities with succinyl-CoA synthetase, and it is believed that CCS may have evolved from succinyl-CoA synthetase after a gene duplication event (Aoshima, 2007). The sequence of citrate synthase (CS) shows similarities to the sequence of citryl-CoA lyase (CCL), making it quite reasonable to believe that it evolved from CCL (Kitadai et al., 2017). The StC strain has the genes for succinyl-CoA synthetase and citrate synthase in its genome, which suggests that the database for metabolic reconstruction aligned the sequences from these enzymes. However, StC strain showed genes and protein expression to proceed with the rTCA cycle from isocitrate to citrate. The reaction is catalyzed by aconitate hydratase to form cis-aconitate from isocitrate, followed by the formation of citrate. Furthermore, the cycle could start with pyruvate forming phosphoenolpyruvate by pyruvate dikinase, initiating the first carboxylation of the cycle by PEP carboxylase. In this case, oxaloacetate is formed, and the cycle proceeds to the next carboxylation by 2-oxoglutarate synthase and isocitrate dehydrogenase. StC is an aerobic bacterium, and it has already demonstrated in other bacteria that the oxygen-sensitive rTCA cycle enzyme (ferredoxin oxidoreductase) has an isoform of 2-oxoglutarate: ferredoxin oxidoreductase with different sensitivities to molecular oxygen (Yamamoto et al., 2006).

In addition to the variant of the CBB cycle and rTCA, the metabolic reconstruction of the StC genome showed genes for a third carbon fixation pathway, the 3-hydroxypropionate bicycle (3-HP). However, not all the genes and proteins required for the functionality of this cycle were observed. The 3-HP bicycle starts from acetyl-CoA, with acetyl-CoA and propionyl-Co-A carboxylases acting as carboxylating enzymes (Zarzyck et al., 2009). In the StC strain, the genes for both enzymes (acetyl-CoA and propionyl-Co-A carboxylases) were expressed in autotrophic and heterotrophic conditions; however, the enzyme propionyl-Co-A carboxylases was not abundant. From two molecules of bicarbonate and acetyl-CoA, (S)-malyl-CoA is formed (Xu et al., 2023). In StC strain (S)-malyl-CoA could not be formed because only some intermediary genes for acrylyl-CoA reductase, 2-methylfumaryl-CoA hydratase, malyl-CoA/beta methylmalyl-CoA lyase and propionyl-CoA carboxylase were expressed, but these proteins were not abundant, thus, the formation of (S)-malyl-CoA could not occur. The reaction

sequence of bicycle from the (S)-malyl-CoA to malate, all genes and proteins were upregulated. Therefore, there is no evidence that the 3-HP bicycle could function as a third carbon fixation pathway in the StC strain, at least under the tested conditions. Among the seven-carbon fixation pathway he 3-HP pathway has been the most suitable pathway for aerobic $CO_2$ fixation and more efficient than those natural carbon fixation pathways using $CO_2$ (Bar-Even et al., 2010; Liu et al., 2020). For microorganisms that can use more than one carbon fixation pathway, could be interesting arrange the 3-HP bicycle with another carbon cycle, since in 3-HP bicycle $CO_2$-fixing steps are rapid; and not to have reactions that are affected by oxygen (Mattozzi et al., 2013). Compared with RuBisCo ((kcat≈1–10/s), carboxylases of 3-HP bicycle are more rapid (kcat≈30–80/s) and has high rate of $CO_2$ fixation (Sage, 2002; Hugler et al., 2003).

We discovered, using comparative omics (genome, transcriptome, and proteome) analysis, the co-occurrence of the Calvin-Benson-Bassham (CBB) cycle and the reductive tricarboxylic acid (rTCA) cycle in *Carbonactinospora thermoautotrophica* strain StC. The presence of a complete set of genes for both carbon fixation pathways is not commonly observed in microorganisms. This study not only describes the functionality of these genes and proteins but also represents the first time that protein expression has been accessed to corroborate gene expression in a cultivated bacterium. Other studies have demonstrated the presence of genes related to the Calvin-Benson-Bassham (CBB) cycle and the reductive tricarboxylic acid (rTCA) cycle in bacteria. Notably, endosymbiotic bacteria residing in deep-sea tubeworms *Riftia pachyptila* and *Tevnia jerichonana* were found to harbor genes associated with both carbon fixation pathways. The analysis in these cases involved proteogenomic comparison, providing insights into the carbon metabolism of these symbiotic bacteria (Gardebrecht et al., 2012). Additionally, similar occurrences of the CBB and rTCA cycles were reported in endosymbiotic bacteria associated with Lamellibrachia tubeworms. These findings were based on carbon isotopic data and the analysis of genes involved in carbon metabolism (provide reference if available). The interaction between the Calvin-Benson-Bassham (CBB) cycle and the reductive tricarboxylic acid (rTCA) cycle in tubeworm symbioses has been identified as a key factor contributing to high carbon fixation efficiency. This synergy supports the rapid growth of tubeworms, outpacing other invertebrates (Klatt, Polerecky, 2015; Bright et al., 2013). However, challenges arise in investigating this co-occurrence of two carbon fixation pathways, especially in cases where the symbiotic bacteria cannot be cultivated (Rubin-Blum et al., 2019). This limitation hinders a comprehensive understanding of the molecular and metabolic mechanisms underlying the concurrent operation of CBB and rTCA cycles in these symbiotic systems. Studies on free-living bacteria, including *Candidatus Thiomargarita*

*nelso*nii and *Thioflavicoccus mobilis*, have demonstrated the co-occurrence of the Calvin-Benson-Bassham (CBB) cycle and the reductive tricarboxylic acid (rTCA) cycle. Genomic data, as well as single-cell genomic and transcriptomic analyses, have provided evidence for the existence of multiple carbon fixation pathways in bacteria (Flood et al., 2016; Winkel et al., 2016; Rubin-Blum et al., 2019). The presence of genes encoding enzymes for the rTCA cycle in the aerobic strain StC suggests its ability to survive in environments with low oxygen levels, as the rTCA cycle is known to function under anaerobic and microaerobic conditions (Steffens et al., 2022).

The simultaneous use of two different carbon fixation pathways, such as the Calvin-Benson-Bassham (CBB) cycle and the reductive tricarboxylic acid (rTCA) cycle, raises intriguing questions about the evolution and metabolic strategies of bacteria. The reasons why a single cell might utilize multiple pathways, even if it requires more energy, remain mysterious. Evolutionary considerations suggest that the rTCA cycle could be one of the earliest autotrophic metabolisms in the history of life, with the possibility that the CBB cycle emerged later. As new enzymes appeared, cells might have retained both cycles to enhance carbon fixation efficiency (Wächtershäuser, 1990; Kitadai et al., 2017).

The efficiency of the rTCA cycle in terms of ATP consumption per $CO_2$ molecule fixed, running near the thermodynamic edge, may provide an advantage. The coexistence of CBB and rTCA cycles could represent an evolutionary strategy to optimize carbon fixation. Genetic engineering studies in *E. coli* have demonstrated that combining CBB and rTCA cycles can lead to improved carbon fixation capacity (Lo et al., 2019). The mystery behind why cells may choose to utilize multiple carbon fixation pathways and how these pathways function together or at separate times requires further investigation.

As previously mentioned in regard of the study with tubeworm symbionts, due to the fragmented nature could not determinate whether the genes for both pathways are present in a single genome or if the two pathways are distributed in a strain (Klatt; Polerecky, 2015; Bright et a., 2013). Another study with a single cell of *Thioflavicoccus mobilis* provides evidence with sequencing coverage for the genes and demonstrated the expression of these genes (Rubin-Blum et al., 2019). Here we provide evidence with genes and proteins expression of two carbon fixation pathways can occur in *a Carbonactinospora thermoautotrophica* strain StC single microorganism.

*Evidence for Bacterial microcompartment*

In addition to the evidence for a dual carbon fixation pathway in the StC strain,

noteworthy is the presence of genes encoding a microcompartment more closely related to carboxysomes. Various microorganisms employ different strategies to obtain carbon and energy, including operating the CBB cycle within the carboxysome bacterial microcompartment (BMC). Despite the CBB cycle being the predominant pathway among organisms, the efficiency of the RuBisCo enzyme in fixing carbon is limited due to competition with oxygen for the same active site, resulting in carbon loss when RuBisCo fixes oxygen. In contrast, other carbon fixation pathways like the rTCA cycle, the 3-HP bicycle, or the 3-HP/4-HB cycle can facilitate carbon assimilation, requiring less energy to synthesize a three-carbon unit from $CO_2$ (Hugler; Sievert, 2011). BMCs are large macromolecular structures resembling organelles but lacking a lipid bilayer, consisting entirely of proteins (Evans et al., 2023). These BMCs can sequester metabolic enzymes from the cytosol using virus-like protein shells and play roles in various enzymatic pathways, including autotrophic $CO_2$ fixation, as seen in carboxysomes (Borden; Savage, 2021).

There are two known types of carboxysomes: alpha (α) and beta (β). They are distinguished by their respective operons, with α-carboxysomes encoded by the cso operon and β-carboxysomes encoded by the ccm operon (Kerfeld; Melnick, 2016). These two groups of carboxysomes also differ in the type of RuBisCo they encapsulate (form IA for α-carboxysomes and form IB for β-carboxysomes) and the type of carbonic anhydrase they utilize (β-carbonic anhydrase for α-carboxysomes and ɣ-carbonic anhydrase for β-carboxysomes) (Chen et al., 2023). In the StC strain, we consistently found genes related to carboxysomes, Pdu, and Eut microcompartments. The availability of genomic sequence data has facilitated the discovery of bacterial microcompartments (BMCs), and an effective approach is to search for shell protein genes clustered with genes for putative enzymes (Sutter et al., 2021). The CcmK/CsoS1 proteins in StC form clusters that include an operon gene for RuBisCo (cbbL, cbbS, cbbX, cbbY, cbbE, cbbZ, cbbF, cbbP, cbbT, cbbG, cbbK, cbbA).

In addition, BMC loci can be identified in genomes by searching for the presence of shell proteins belonging to pfam00936 and pfam03319 (Evans et al., 2023). StC has genes for both domains, as described in the results. Although the RuBisCo in StC is not associated with the shell proteins, occasionally BMC-associated genes are found in a distal, satellite location (Sommer et al., 2019). The analysis, which includes the purification of the BMC, western-blot, and LC/MS-MS, indicates that the BMC present in the StC strain is closer to the carboxysome microcompartment. Until now, carboxysomes were not described to exist in the *Actinomycetota* phylum, but the occurrence of BMCs with an unknown function was demonstrated in this phylum (Sutter et al., 2021). However, studies have already demonstrated that the phylum

*Actinomycetota* has loci for α-carboxysomes (Axen et al., 2014). In fact, StC has more genes related to α-carboxysomes, which include the form IA of RuBisCo and two carbonic anhydrases. However, in the transcriptome analysis, only the β-carbonic anhydrase was expressed.

In the StC genome, the gene CcmK/CsoS1 for the putative carboxysome structural-like protein is followed by the CcmL protein shell, and on the opposite side, it is followed by a cluster with genes for RuBisCo (cbbL and cbbS). CbbL and CbbS are sequestered in carboxysomes, which help the enzyme fix $CO_2$ at low concentrations. StC cells were harvested for Cryo-EM CEMOVIS, and polyhedral particles showing the typical shape and approximate size of carboxysomes were observed. In Cryo-EM CEMOVIS, it is difficult to determine the structure; it might be expected to be inside the bacteria rather than free-floating in the BMC. We postulate that the broken membrane observed in the study may result from sample handling or freezing, as demonstrated in other studies (Evans et al., 2022). Previous estimations of carboxysome proteins using immunoblot and mass spectrometry also demonstrated the relative abundance of carboxysome proteins from either whole cell lysates or isolated forms (Long et al., 2005, 2011; Rae et al., 2012; Faulkner et al., 2017). In our analysis, isolated forms were used for mass spectrometry, and the low abundance of the protein shell in the sample may have interfered with the results. Furthermore, it seems that the best way to perform MS analysis is with PAGE bands of purified shell proteins matched to a theoretical molecular weight library generated from the protein sequences to identify peptides specific to each isoform (Chaijarasphong et al., 2015).

Isolated BMCs were also observed with negative stain microscopy and Cryo-ET. These results suggest that the StC strain may have carboxysomes or other BMCs similar to carboxysomes. Future work will focus on clarifying the function of BMCs in StC, whether it is a type of carboxysome, the presence of another yet undescribed BMC, or if this atypical BMC has modified functions, as described by Kerfeld and Scott (2022).

## CONCLUSION

Our study represents a significant advancement in understanding carbon fixation pathways, particularly the co-occurrence of the Calvin-Benson-Bassham (CBB) cycle and the reverse tricarboxylic acid (rTCA) cycle in a cultivated thermophilic chemoautotrophic bacterium, *Carbonactinospora thermoautotrophica* strain StC. This new understanding is based on integrated analyses of genomics, transcriptomics, and proteomics, providing for a first time

the dual co-occurrence of genes and proteins expression for CBB and rTCA cycles in a cultivated microorganism. Additionally, our work proposes the first evidence of a bacterial microcompartment (BMC) related to carboxysomes in the *Actinomycetota* phylum. The identification of BMCs in StC raises intriguing questions about their function and carboxysome classification within the bacterial doimain.

Furthermore, this study demonstrates that StC utilizes both the CBB cycle and the rTCA pathway for carbon fixation, shedding light on the versatility of microbial carbon assimilation strategies. This insight into the simultaneous operation of two distinct carbon fixation pathways in a single isolated bacterium expands our understanding of microbial metabolic diversity. While these findings contribute to the growing body of knowledge on carbon fixation, there is still much to explore. Future investigations, including physiological and biochemical studies, are essential to elucidate the interplay between these pathways in StC filaments. Additionally, further research is needed to characterize the BMC identified in StC, determining its type (α or β carboxysome) and unraveling its specific functions.

These results reinforce previous studies that many microorganisms have developed the ability to use multiple carbon fixation and open highlights for future exploration in other organisms and guide bioengineering projects towards optimal production using integrated carbon fixation cycles to further improve the ability carbon fixation.

FUNDING

This work was financially supported by KAUST Baseline Grant BAS/1/1096-01-0 (to Prof. A. S. Rosado). SSC was funded by the National Council for the Improvement of Higher Education (CAPES).

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Sulamita Santos Correa: Conceptualization, Methodology, Writing-original draft, Writing-review & editing. Junia Schultz: Data curation, Writing-review & editing. Brandon Huntington: Methodology with negative stain EM, Cryo-EM, and Cryo-ET microscopy. Andreas Naschberger: Methodology with Cryo-EM and Cryo-ET microscopy. Luis Arge: Methodology with genomic and transcriptomic assembly. Fabio Mota: Methodology with carboxysome phylogeny. Alexandre Soares Rosado: Supervision, Writing-review & editing.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing monetary interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGMENTS

We thank the King Abdullah University of Science and Technology (KAUST) and Graduate Program of Plant Biotechnology and Bioprocesses (PBV) at the Federal University of Rio de Janeiro for supporting this work. We thank Ashraf Alamoudi for the support with Cryo-EM and Cryo-Et microscopy analysis. We thank Anthony Weatherhead and Papita Mandal for the support with proteomic analysis. We thank Eikon Alves and Edir Ferreira for the initial support with bacteria cultivation.

# SUPPLEMENTARY DATA

Supplementary Table 1: Pairwise comparisons of StC genome vs. type strain genomes

| Query strain | Subject strain | dDDH (d0, in %) | C.I. (d0, in %) | dDDH (d4, in %) | C.I. (d4, in %) | dDDH (d6, in %) | C.I. (d6, in %) | G+C content difference (in %) |
|---|---|---|---|---|---|---|---|---|
| 'Carbonactinospora thermoautotrophica.fa' | Streptomyces thermoautotrophicus UBT1 | 83.2 | [79.4 - 86.5] | 87.4 | [84.8 - 89.6] | 86.9 | [83.8 - 89.4] | 0.03 |
| 'Carbonactinospora thermoautotrophica.fa' | Acidothermus cellulolyticus 11B | 12.7 | [10.0 - 16.0] | 21.7 | [19.4 - 24.1] | 13.1 | [10.8 - 15.9] | 4.04 |
| 'Carbonactinospora thermoautotrophica.fa' | Kitasatospora setae KM-6054 | 13.8 | [11.0 - 17.1] | 20.7 | [18.5 - 23.2] | 14.1 | [11.7 - 16.9] | 3.25 |
| 'Carbonactinospora thermoautotrophica.fa' | Kitasatospora cheerisanensis KCTC 2395 | 13.8 | [11.0 - 17.2] | 20.6 | [18.3 - 23.0] | 14.1 | [11.7 - 16.9] | 2.6 |
| 'Carbonactinospora thermoautotrophica.fa' | Marinitenerispora sediminis TPS16T | 13.5 | [10.7 - 16.8] | 20.4 | [18.2 - 22.8] | 13.8 | [11.4 - 16.6] | 2.93 |
| 'Carbonactinospora thermoautotrophica.fa' | Actinomadura physcomitrii LD22 | 13.5 | [10.8 - 16.9] | 20.1 | [17.9 - 22.5] | 13.8 | [11.4 - 16.6] | 1.55 |
| 'Carbonactinospora thermoautotrophica.fa' | Streptomyces coeruleorubidus JCM 4359 | 13.5 | [10.7 - 16.8] | 20.0 | [17.8 - 22.4] | 13.8 | [11.4 - 16.6] | 0.19 |
| 'Carbonactinospora thermoautotrophica.fa' | Actinomadura rifamycini DSM 43936 | 13.5 | [10.8 - 16.9] | 19.9 | [17.7 - 22.3] | 13.8 | [11.4 - 16.6] | 3.15 |
| 'Carbonactinospora thermoautotrophica.fa' | Streptomyces bellus JCM 4292 | 13.5 | [10.8 - 16.9] | 19.8 | [17.6 - 22.2] | 13.8 | [11.4 - 16.6] | 0.19 |
| 'Carbonactinospora thermoautotrophica.fa' | Microbispora corallina NBRC 16416 | 13.6 | [10.9 - 17.0] | 19.8 | [17.6 - 22.2] | 13.9 | [11.5 - 16.7] | 1.43 |
| 'Carbonactinospora thermoautotrophica.fa' | Micromonospora citrea DSM 43903 | 13.6 | [10.9 - 17.0] | 19.7 | [17.5 - 22.1] | 13.9 | [11.5 - 16.7] | 2.87 |
| 'Carbonactinospora thermoautotrophica.fa' | Micromonospora polyrhachis DSM 45886 | 13.1 | [10.4 - 16.5] | 19.7 | [17.5 - 22.1] | 13.5 | [11.1 - 16.3] | 1.72 |
| 'Carbonactinospora thermoautotrophica.fa' | Jiangella alkaliphila KCTC 19222 | 13.5 | [10.7 - 16.8] | 19.6 | [17.4 - 22.0] | 13.8 | [11.4 - 16.6] | 1.26 |
| 'Carbonactinospora thermoautotrophica.fa' | Micromonospora globbae WPS1-2 | 13.5 | [10.7 - 16.8] | 19.5 | [17.3 - 21.9] | 13.8 | [11.4 - 16.6] | 2.24 |
| 'Carbonactinospora thermoautotrophica.fa' | Micromonospora rhizosphaerae DSM 45431 | 13.6 | [10.8 - 16.9] | 19.4 | [17.2 - 21.8] | 13.9 | [11.5 - 16.7] | 0.66 |
| 'Carbonactinospora thermoautotrophica.fa' | Geodermatophilus telluris DSM 45421 | 13.6 | [10.9 - 17.0] | 19.4 | [17.2 - 21.8] | 13.9 | [11.5 - 16.7] | 4.77 |
| 'Carbonactinospora thermoautotrophica.fa' | Actinoplanes capillaceus NBRC 16408 | 13.2 | [10.4 - 16.5] | 18.8 | [16.7 - 21.2] | 13.5 | [11.1 - 16.3] | 0.08 |

Supplementary Table 2. Functional characterization of genome sequences of C. thermoautotrophica StC connected with Kyoto Encyclopedia of Genes and Genomes (KEGG) using BLASTKOALA.

| KEGG category | Genes |
|---|---|
| Carbohydrate metabolism | 225 |
| Protein families: genetic information processing | 217 |
| Genetic information processing | 182 |
| Amino acid metabolism | 151 |
| Energy metabolism | 142 |
| Proteins families: signaling and cellular processes | 136 |
| Metabolism of cofactors and vitamins | 126 |
| Environmental information processes | 115 |
| Unclassified: metabolismo | 106 |
| Nucleotide metabolismo | 70 |
| Protein families: metabolismo | 52 |
| Lipid metabolismo | 50 |
| Unclassified: genetic information processing | 44 |
| Xenobiotics biodegradation and metabolismo | 37 |
| Cellular processes | 33 |
| Unclassified: signaling and cellular processes | 25 |
| Glycan biosynthesis and metabolismo | 23 |
| Metabolism of other amino acids | 23 |
| Metabolism of terpenoids and polyketides | 12 |
| Biosynthesis of other secondary metabolites | 12 |
| Organismal systems | 6 |
| Human deseases | 3 |
| Unclassified | 95 |

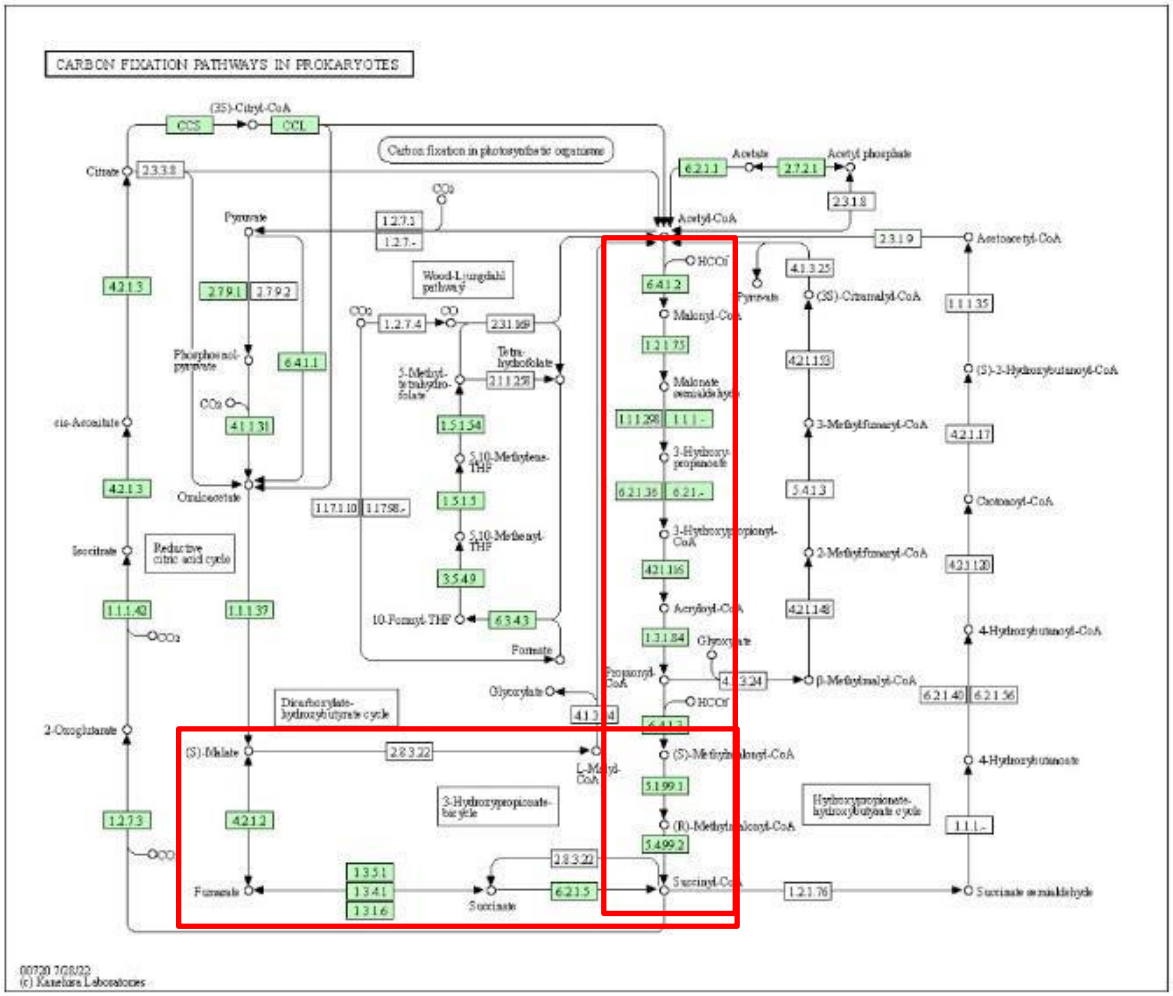



Figure 1: Metabolic pathway reconstruction of StC gemona show the genes for 3-Hydroxyproprionate bicycle.

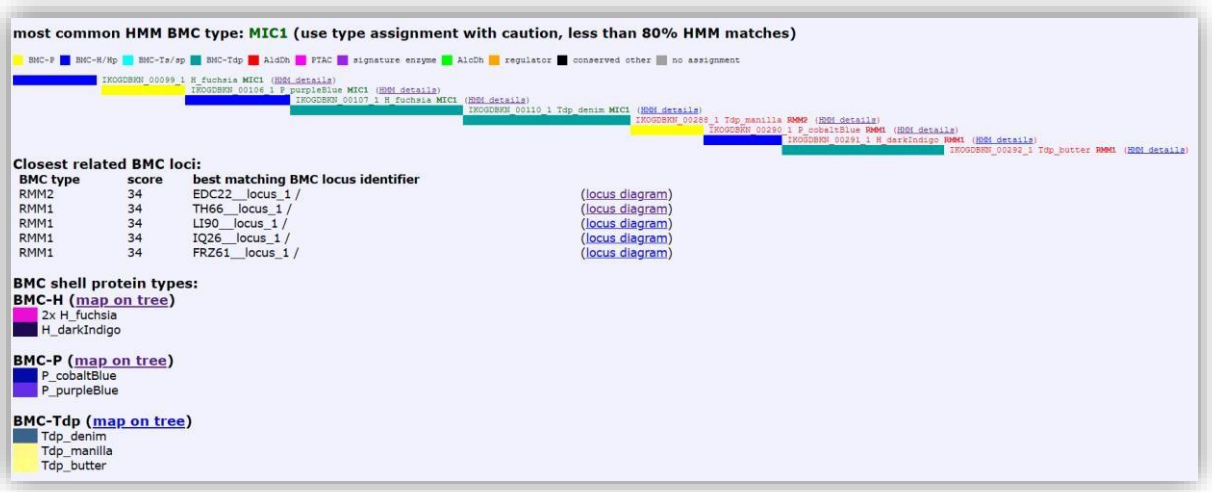



Figure 2: Predicted MIC microcompartment generated using the BMC caller tool (BMC Caller (msu.edu)). (Sutter; Kerfeld, 2022).

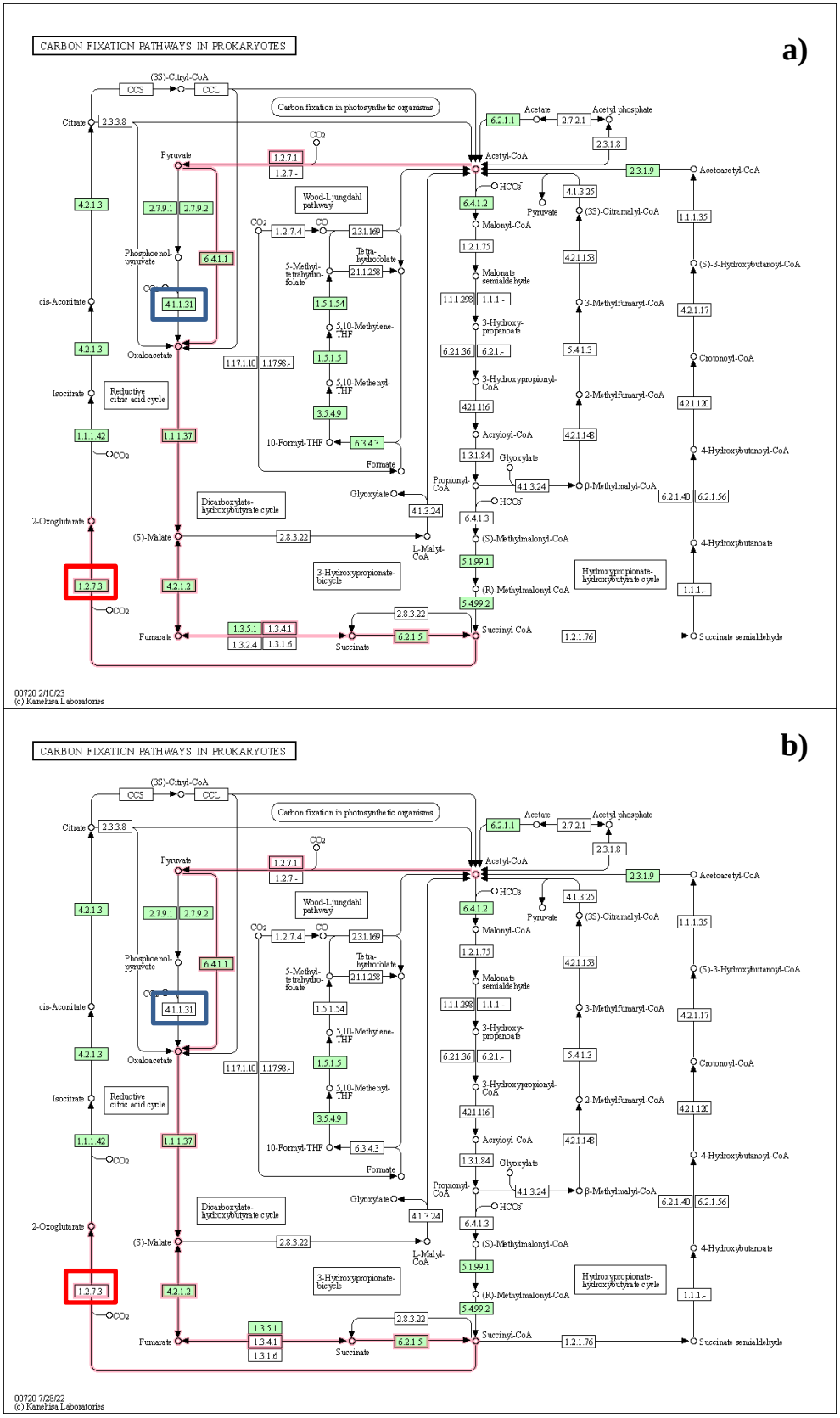



Figure 3: Expression protein predicted by KEGG KOALA. a) the upregulated proteins phosphoenolpyruvate carboxylase and 2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase in StC strain cultivated autotrophically; b) downregulated proteins phosphoenolpyruvate carboxylase and 2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase in StC strain cultivated heterotrophically. Phosphoenolpyruvate carboxylase (squad blue); 2-oxoglutarate/2-oxoacid ferredoxin (squad red).

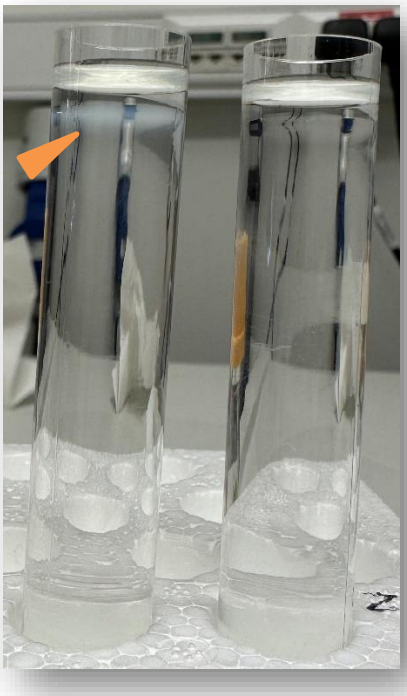


Fig. 4: Gradient ultracentrifugation enriched with 10%-50% sucrose fraction. In the fraction between 10% and 20%, a white pellet is visible (orange arrow) representing the microcompartment; the right tube represents the control.

## 6. DISCUSSÃO GERAL

A fixação de carbono rege o ecossistema produzindo alimentos para os organismos ao transformar o composto inorgânico em orgânico. No entanto, o desequilíbrio do ciclo global do carbono é a principal causa das alterações climáticas e, além de afetar o ciclo do carbono dos ecossistemas, também afeta o ambiente de vida dos seres humanos (Martínez et al., 2023). O dióxido de carbono ($CO_2$) tem sido foco de muitos dos desafios sociais atuais. A liberação massiva de $CO_2$ antropogênico vem ameaçando o equilíbrio climático do Planeta, pois como consequência, pode levar ao aumento calamitoso das temperaturas globais (Kim et al., 2020). Ainda assim, o carbono é essencial para a vida, sendo necessário redobrar a atenção para com esse elemento. O carbono é o quarto elemento químico mais abundante na Terra, depois do hidrogênio, oxigênio e hélio (Aversa et al., 2016; Loeve; Vicent, 2017), e é considerado um elemento base para a construção de compostos orgânicos essenciais à vida, como proteínas, carboidratos, ácidos nucleicos e lipídios. O carbono impulsiona comunidades inteiras de organismos vivos e sustenta os ciclos biogeoquímicos na Terra (Anthony et al., 2020).

A entrada de carbono inorgânico na biosfera está relacionada à fixação biológica de carbono. A produção primária de compostos orgânicos está diretamente relacionada à fixação autotrófica de carbono, a qual converte carbono inorgânico em biomassa (Gong et al., 2018). Dentre as setes vias de fixação de carbono conhecidas, o ciclo de Calvin-Benson-Bassham, que é predominantemente encontrado em plantas, algas e em algumas bactérias (principalmente cianobactérias), foi a primeira via de fixação de $CO_2$ descrita, e desde seu descobrimento, é a via mais estudada (Fuchs, 2011). A descoberta de uma nova via de fixação de carbono em bactérias verdes sulfurosas há quase duas décadas depois do ciclo CBB encorajou novas pesquisas sobre antigas vias de fixação de carbono anteriormente negligenciadas em microrganismos taxonomicamente e filogeneticamente distintos, bem como impulsionou o entendimento da interação entre as vias.

Os microrganismos autotróficos podem incorporar o carbono na biomassa através das vias naturais de fixação de carbono: I) ciclo CBB; II) o ciclo redutivo de citrato (rTCA); III) Wood-Ljungdahl (WL); IV) biciclo 3-hidroxipropionato (3-HP); V) ciclo 3-hidroxipropionato/4-hidroxibutirato (3-HP/4-HB); VI) ciclo dicarboxilato/4-hidroxibutirato (DC/4-HB); VII) ciclo redutor natural de glicina (nrGly) (Ljungdhal, 1986; Bar-Even et al., 2012). Duas outras vias naturais de fixação de carbono anteriormente desconhecidas também foram descritas: a via redutora da hexulose-fosfato (RHP) e o ciclo oxidativo reverso do TCA (roTCA) (Kono et al., 2017; Mall et al., 2018).

Além das vias naturais de fixação de carbono, nos últimos anos, foram desenvolvidas vias sintéticas de fixação de carbono para aumentar a eficiência da fixação de $CO_2$ e reduzir a perda de $CO_2$ durante este processo. Embora os processos naturais de fixação de carbono não sejam viáveis para uso na produção industrial, eles oferecem uma variedade de enzimas fixadoras de carbono e caminhos correspondentes, abrindo opções para a engenharia artificial (Gong et al., 2018). Por exemplo, através da biologia sintética, bioinformática e análises bioquímicas, foram desenvolvidas as seguintes vias: ciclo Acetil-(CoA)/etilmalonil-CoA/hidroxibutiril-CoA (CETCH), glicina redutora sintética, via sintética de malil-CoA-glicerato, ciclo sintético de Acetil-CoA (SACA), e via formalase (Siegel et al., 2015; Lu et al., 2019).

No presente trabalho, foi descrita a capacidade da bactéria termofilia e quimiolitoautotrófica *Carbonactinospora thermoautotrophica* StC de fixar carbono através de duas vias diferentes de fixação de carbono - CBB e rTCA- assim como indícios de possível existência de carboxissomo presente nos filamentos da StC. StC possui ótima temperatura de crescimento entre 55 °C e 65 °C, e foi isolada de um consórcio microbiano termofílico obtido em um local com constante queima de material vegetal.

O ciclo rTCA é o inverso do ciclo TCA (também conhecido como ciclo de Krebs). O ciclo de Krebs ocorre em organismos heterotróficos para oxidar acetil-CoA em $CO_2$ para gerar poder redutor para a síntese de ATP, enquanto o ciclo rTCA se comporta no processo reverso para produzir acetil-CoA a partir de duas moléculas de $CO_2$ (Hugler; Sievert, 2011). O ciclo rTCA possui duas reações de carboxilação: I) a primeira fixação de carbono ocorre com a carboxilação redutiva de succinil-CoA a 2-oxoglutarato, esta reação é catalisada pela 2-oxoglutarato sintase; II) a segunda carboxilação ocorre com carboxilação redutiva de 2-oxoglutarato a isocitrato (Hugler; Sievert, 2011). Esta segunda carboxilação pode ser realizada tanto pela isocitrato desidrogenase, pelas enzimas 2-oxoglutarato carboxilase e oxalosuccinato redutase, tendo o oxalosuccinato como intermediário livre (Ju et al., 2023). O ciclo do rTCA com produção de isocitrato pela isocitrato desidrogenase, já foi demonstrado em *Chlorobium limicola* e pela oxoglutarato carboxilase e oxalosuccinato redutase, sendo que o oxalosuccinato foi descrito em *Hydrogenobacter thermophilus* (Kanao et al., 2002; Aoshima, 2007). O ciclo rTCA ocorre em muitos grupos de Bactérias e Archaeas anaeróbicas ou pouco sensível ao oxigênio.

O ciclo CBB é a via de fixação de dióxido de carbono mais presente entre os organismos. As principais enzimas nesta via incluem RuBisCo, PRK, GAPDH e triosefosfato isomerase, essas três últimas enzimas participam na regeneração da RuBisCo (Kusian; Bowien, 1997;

Gurrieri et al., 2021). No ciclo CBB, três moléculas de $CO_2$ são fixadas pela RuBisCo para formar uma molécula de gliceraldeído-3-fosfato (Ohta, 2022). O ciclo CBB está presente em algas, plantas, cianobactérias, filos *Pseudomonadota* e *Bacillota* e bactérias fotoautotróficas e quimioautotróficas. Algumas bactérias, em especial cianobactérias e bactérias quimioautotróficas desenvolveram uma estratégia para usar o ciclo CBB e aumentar sua taxa de fixação de carbono e consequentemente não perder o carbono fixado. Essa estratégia partiu da evolução de estruturas que funcionam como organela (mas não possuem bicamada lipídica e são compostos inteiramente de proteínas) que podem funcionar como catalítica na célula, a qual recebeu o nome de microcompartimento (BMC) (Evans et al., 2022). Entre os BMCs, o carboxissomo foi o primeiro descrito e vem sendo o mais estudado. O carboxissomo encapsula a enzima RuBisCo e a enzima anidrase carbônica para aumentar a taxa de fixação de carbono na célula e diminuir a taxa de oxigenação pela RuBisCo (Borden; Savage, 2021).

A estirpe StC vem sendo cultivada em meio de cultura mínimo com metais traços e sem fontes orgânicas de carbono e nitrogênio, apenas com adição dos gases CO, $CO_2$, $N_2$ e $O_2$. Para a investigação de qual via de fixação de carbono é utilizada pela estirpe StC para fixar carbono atmosférico, foi conduzido um experimento in vivo integrando as ômicas: genômica, transcriptômica e proteômica. Para o sequenciamento do genoma, a estirpe StC foi cultivada autotroficamente, foi obtido um *contig* e análise filogenia indicou que a estirpe está no mesmo clado das outras duas estirpes UBT1 e H1 já descrita. Para o experimento de transcriptômica e proteômica a estirpe StC foi cultivada em meio de cultura autotrófico e heterotrófico à 60 °C para que fosse realizado uma comparação sob essas duas condições de cultivo. De forma pioneira, foi encontrado genes e proteínas expressos não apenas para uma via de fixação de carbono, mas para duas vias, indicando a coexistência funcional das vias na estirpe.

No genoma da estirpe StC foi encontrado a expressão dos 11 genes pertencentes ao canonical ciclo CBB, no entanto, a enzima SBPase não foi expressa, o que leva ao entendimento que a StC utiliza uma variante já descrita do ciclo CBB (Sharkey; Weise, 2016). As reações do canonical ciclo CBB envolvem a fixação do $CO_2$, catalisando a carboxilação da RuBP para formar duas moléculas de 3-fosfoglicerato (Sharkey, 2019). As enzimas responsáveis pela regeneração da RuBisCo incluem PPK, GAPDH e triosefosfato isomerase. Duas moléculas de 3-fosfoglicerato são convertidas em gliceraldeído-3-fosfato, seguido por uma série de reações de rearranjo que resultam na produção de ribulose-5-fosfato (Finn; Tabita, 2004). Neste processo, a transketolase converte gliceraldeído-3-fosfato e diidroxiacetona fosfato em xilulose-5-fosfato e eritrose-4-fosfato, seguido pela formação de xilulose-5-fosfato e ribulose-5-fosfato a partir de diidroxiacetonafosfato e eritrose-4-fosfato (Berg, 2011). A enzima SBPase

catalisa a desfosforilação da enzima SBP (Fuchs, 2011). As reações seguintes convertem a ribulose-5-fosfato convertida em RuBP, esta etapa final é catalisada pela PRK, que fosforila a ribulose-5-fosfato (Berg, 2011).

Como a enzima SBPase não foi expressa na StC, isto pode levar a uma alternativa do ciclo das pentoses fosfato. Neste cenário a enzima SBPase é substituída pela enzima transketolase, a qual transfere dois carbonos da frutose-6-fosfato para o gliceraldeído-3-fosfato para produzir xilulose-1,5-bifosfato e eritrose-4P, e na próxima etapa a transketolase atua novamente para formar ribose-5-fosfato e xilulose-5-fosfato (Sharkey; Weise, 2016). Neste ciclo, a SBPase não participa. A variante do ciclo CBB foi proposta como ciclo da transaldolase (Sharkey; Weise, 2016). A diferença entre o ciclo canônico CBB que usa uma SBPase e a variante que usa transketolase, é que a variante torna a segunda unidade ATP reversível (Ohta, 2022). A proposta de que o CBB poderia funcionar sem algumas enzimas intermediárias ocorreu com a observação de que o ciclo de CBB não se torna totalmente marcado em um experimento com $14CO_2$ ou $13CO_2$ com folhas fotossintetizantes e a descoberta de que a causa de um fenótipo de alto fluxo de elétrons fotossintético cíclico resulta da falta da enzima frutose-bifosfatase (Livingston et al., 2010). A variação do ciclo CBB com ausência da enzima SBPase foi demonstrada no genoma de *Thermithiobacillus tepidarius* e *Thiobacillus thioparusare* (Boden et al., 2016; Hutt et al., 2017). Mais recentemente, a variante do ciclo CBB foi demonstrada usando abordagens multi-ômicas e bioquímicas na bactéria quimiolitoautotrófica *Thermodesulfobium acidiphilum* (Frolov et al., 2019). Aqui, demonstramos a expressão de genes e proteínas na estirpe quimiolitoautotrófica StC cultivada em meio de cultura sem fonte orgânica de carbono, apenas gases CO e $CO_2$, o que sugere que a variante do CBB é funcional nessa estirpe.

A expressão de genes e proteínas relacionados a via do rTCA também foi observada na estirpe StC. No ciclo rTCA da estirpe StC, a primeira fixação de carbono é catalisada pela 2-oxoglutarato sintase para formar 2-oxoglutarato a partir de succinil-CoA. A segunda carboxilação é catalisada pela isocitrato desidrogenase para reduzir o 2-oxoglutarato em isocitrato. Outra etapa para fixar um terceiro carbono no ciclo rTCA pode ser usada por algumas bactérias, como *Mycobacterium tuberculosis*, *Geobacter sulfurreducens* e *Chlorobium limicola* (Holmes et al., 2017; Hu et a., 2020). Esta reação é mais complexa que as duas primeiras carboxilações; envolve a clivagem do citrato dependente de ATP em Acetil-CoA e oxaloacetato, esta reação pode ser realizada pela ATP citrato sintase ou pela ação combinada da citril-CoA sintetase (CCS) e citril-CoA liase (CCL) (Kitadai et al., 2017). Para que esta reação seja obtida a partir do piruvato a partir do Acetil-CoA, é necessária outra reação de

carboxilação, catalisada pela piruvato sintase (Evans et al., 1966). Apesar do gene para as enzimas - ATP citrato sintase; citril-CoA sintetase e citril-CoA liase - não terem sido anotados no genoma da estirpe StC, a reconstrução da via metabólica previu as enzimas citril-CoA sintetase e citril-CoA liase. A enzima CCS é composta por duas subunidades que possuem semelhanças de sequência com a succinil-CoA sintetase, acredita-se que o CCS pode ter evoluído da succinil-CoA sintetase após um evento de duplicação genética (Aoshima, 2007). A sequência da enzima citrato sintase (CS) apresenta semelhanças com a sequência da CCL, tornando bastante razoável acreditar que evoluiu a partir da CCL (Kitadai et al., 2017). A estirpe StC possui no genoma o gene das enzimas succinil-CoA sintetase e citrato sintase, o que leva a acreditar que o banco de dados para reconstrução metabólica alinhou a sequência dessas enzimas. No entanto, a StC exibiu expressão de genes e proteínas para prosseguir com o ciclo do rTCA de isocitrato para citrato, reação catalisada pela enzima aconitato hidratase para formar cis-aconitato a partir do isocitrato, seguido pela formação de citrato. O ciclo começa com o piruvato formando fosfoenolpiruvato pela piruvato diquinase, após a primeira carboxilação do ciclo pela PEP carboxilase. Neste caso, oxaloacetato é formado, e o ciclo prossegue para a próxima carboxilação pela 2-oxoglutarato sintase e isocitrato desidrogenase, como já foi demostrado em *Chlorobium limícola* (Aoshima, 2007). O ciclo do rTCA pode ser encontrado em muitos grupos de bactérias, no entanto, parece estar restrito a bactérias anaeróbicas ou microaeróbicas (Nunoura et al., 2018). A StC é uma bactéria aeróbica, e já foi demostrado em outras bactérias que a enzima do ciclo rTCA sensível ao oxigênio (ferredoxina oxidorredutase) possui uma isoforma de 2-oxoglutarato insensível ao oxigênio (Yamamoto et al., 2006).

Além da variante do ciclo CBB e do rTCA, a reconstrução metabólica do genoma do StC mostrou genes para uma terceira via de fixação de carbono, o biciclo 3-HP. Entretanto, não foi observada a expressão de todos os genes e proteínas para a funcionalidade deste ciclo nas condições testadas. O biciclo 3-HP começa a partir de acetil-CoA, com acetil-CoA e propionil-Co-A carboxilases atuando como enzimas carboxilantes (Zarzyck et al., 2009). Na StC, o gene para as enzimas acetil-CoA e propionil-Co-A carboxilases foi expresso em condições autotróficas e heterotróficas, no entanto, a enzima propionil-Co-A carboxilase não foi abundante na análise de proteômica. A partir de duas moléculas de bicarbonato e acetil-CoA, forma-se o (S)-malil- CoA (Xu et al., 2023). Na cepa StC (S)-malil-CoA não pode ser formada, pois apenas alguns genes intermediários para as enzimas acrilil-CoA redutase, 2-metilfumaril-CoA hidratase, malil-CoA/beta metilmalil-CoA liase e propionil-CoA carboxilase foram expressos, mas a expressão dessas proteínas não foi observada nos dados de proteômica, sendo assim, a formação de (S)-malil-CoA não poderia ocorrer, pelo menos nas condições testadas.

Para sequência de reações a partir do (S)-malil-CoA ao malato, todos os genes e proteínas foram regulados positivamente. Dentre as sete vias naturais de fixação carbono reconhecidas, a via 3-HP tem sido a via mais adequada para a fixação aeróbica de $CO_2$ e mais eficiente do que as outras vias naturais de fixação de carbono usando $CO_2$ (Bar-Even et al., 2010; Liu et al., 2020). Para microrganismos que podem utilizar mais de uma via de fixação de carbono, poderia ser interessante acoplar o biciclo 3-HP com outro ciclo de carbono, uma vez que no biciclo 3-HP as etapas de fixação de $CO_2$ são relativamente rápidas; e não são reações afetadas pelo oxigênio (Mattozzi et al., 2013). Comparadas com RuBisCo ((kcat≈1–10/s), as carboxilases da biciclo 3-HP são relativamente mais rápidas (kcat≈30–80/s) e possuem alta taxa de fixação de $CO_2$ (Sage, 2002; Hugler et al., 2003).

Através da análise comparativa das ômicas utilizadas (genoma, transcriptoma e proteoma), foi identificada a co-ocorrência dos ciclos CBB e rTCA na estirpe StC. O conjunto de genes para mais de um ciclo fixação de carbono já foi encontrado em alguns microrganismos, aqui descrevemos a funcionalidade desses genes e proteínas, e pela primeira vez a expressão proteica foi acessada para corroborar a expressão gênica em uma bactéria cultivada. Outros estudos demonstraram genes relacionados ao ciclo CBB e rTCA em bactérias endossimbióticas que vivem nos vermes tubulares de águas profundas *Riftia pachyptila* e *Tevnia jerichonana*, essas bactérias não foram isoladas dos vermes tubulares e a análise foi feita com comparação proteogenômica (Gardebrecht et al., 2012). A ocorrência dessas duas vias de fixação de carbono (CBB e rTCA) também foi demonstrada em outra bactéria endossimbiótica presente no verme tubular *Lamellibrachia*, neste estudo foram analisados dados isotópicos do carbono, bem como genes envolvidos no metabolismo do carbono. Provavelmente a interação entre os dois ciclos de fixação de carbono (ciclo CBB e rTCA) em simbioses de vermes tubulares contribui para a alta eficiência de fixação de carbono e consequentemente para o crescimento desses vermes tubulares, que crescem mais rapidamente que outros invertebrados (Klatt; Polerecky, 2015; Bright et al., 2013). No entanto, nestes estudos as bactérias não puderam ser cultivadas, o que dificultou a investigação aprofundada desta co-ocorrência de duas fixações de carbono em uma única bactéria ou em bactérias separadas (Rubin-Blum et al., 2019). Alguns estudos com bactérias de vida livre demonstraram genes para o ciclo CBB e rTCA usando dados do genoma de *Candidatus Thiomargarita nelsonii* (Flood et al., 2016; Winkel et al., 2016). Mais recentemente, a co-ocorrência do ciclo CBB e rTCA foi demonstrada na bactéria *Thioflavicoccus mobilis*, através análises genômicas (Rubin-Blum et al., 2019). Todos estes estudos comprovam que é possível a existência de mais uma via de fixação de carbono em bactérias, necessitando apenas de uma investigação aprofundada. Em relação à estirpe StC que

é aeróbia, a detecção de genes que codificam enzimas para o ciclo do rTCA leva ao entendimento de que esta estirpe pode sobreviver em ambiente com pouco oxigênio, uma vez que o rTCA é descrito para funcionar em condições anaeróbicas e microaeróbicas (Steffens et al., 2022).

Porém, não podemos determinar se esses ciclos funcionam simultaneamente ou em momento ou tempo diferencialmente. Além disso, há um mistério: "Por que uma única célula microbiana opta por usar duas vias diferentes de fixação de carbono se necessita de mais energia?" A resposta para esta questão ainda não é conhecida, mas se for pensado na evolução, o ciclo rTCA é o candidato mais plausível para ser o primeiro metabolismo autotrófico a surgir no ecossistema (Kitadai et al., 2017). O ciclo CBB possivelmente surgiu posteriormente ao ciclo rTCA e quando novas enzimas apareceram, a célula manteve ambos os ciclos para melhorar a fixação de carbono (Wächtershäuser, 1990). Quanto à energia, o ciclo rTCA é conhecido por ser o caminho mais eficiente em termos de consumo de ATP por moléculas de $CO_2$ fixadas e correr próximo ao limite termodinâmico (Bar-Even, et al., 2012). Uma combinação eficiente do ciclo CBB e rTCA para fixar carbono já foi realizada usando engenharia genética em *E. coli in silico* para aumentar a capacidade de fixação de carbono (Lo et al., 2019).

Além da evidência da dupla via de fixação de carbono, também foram encontrados no genoma da estirpe StC, genes que codificam um BMC relacionado ao carboxissomo. Alguns microrganismos possuem estratégias para obtenção de carbono e energia, como por exemplo, para a utilização do ciclo CBB dentro do microcompartimento carboxissomo. Pois apesar do ciclo CBB ser o ciclo mais predominante entre os organismos, a enzima carboxilase RuBisCo não é totalmente eficiente para fixar carbono, existe uma competição entre o oxigênio e o carbono pelo mesmo sítio ativo da enzima (Lo et al., 2019). Quando a RuBisCo fixa o oxigênio, leva à perda de carbono. Comparado com ciclo CBB, outros ciclos de fixação de carbono, como o ciclo rTCA, 3-HP biciclo ou o ciclo 3-HP/4-HB, podem facilitar a assimilação de carbono, exigindo menos energia para sintetizar uma unidade de três carbonos a partir de $CO_2$ (Hugler; Sievert, 2011). Uma dessas estratégias é o desenvolvimento de BMCs, os quais segregam enzimas metabólicas do citosol usando invólucros de proteínas semelhantes a estruturas de vírus e desempenham papéis em diversas vias enzimáticas, incluindo fixação autotrófica de $CO_2$ (Borden; Savege, 2021). Existem dois tipos conhecidos de carboxissomos, o alfa e o beta. Eles se distinguem pelo operon: α-carboxissomo, codificado pelo operon cso, e β-carboxissomo, codificado pelo operon ccm (Kerfeld; Melnick, 2016). Esses dois grupos de carboxissomos também se distinguem tipo de RuBisCo que é encapsulada (forma IA para α-carboxissomo e

forma IB para β-carboxissomo); e o tipo de anidrase carbônica (β-anidrase carbônica para α-carboxissomo e ɣ-anidrase carbônica para β-carboxissomo (Chen et al., 2023). No genoma da estirpe StC encontramos consistentemente genes relacionados aos BMCs carboxissomo, Pdu e Eut. A disponibilidade de dados de sequência genômica permitiu a descoberta de BMCs entre as bactérias, no nível genômico, uma boa maneira é pesquisar genes de proteínas da concha grupados com genes para enzimas putativas (Sutter et al., 2021). As proteínas da concha do carboxissomo CcmK/CsoS1 em StC formam agrupamentos com genes para RuBisCo (cbbL, cbbS, cbbX, cbbY, cbbE, cbbZ, cbbF, cbbP, cbbT, cbbG, cbbK, cbbA). Além disso, loci BMC podem ser encontrados em genomas procurando a presença das proteínas da concha que pertencem ao domínio pfam00936 e pfam03319 (Evans et al., 2023). StC possui genes para ambos os domínios conforme descrito nos resultados. A RuBisCo em StC não está próxima as proteínas da concha, mas ocasionalmente há genes associados ao BMC encontrados em uma localização distal de satélite (Sommer et al., 2019). A análise que inclui a purificação do BMC, western-blot e LC/MS-MS, também indicou que o BMC presente na StC está mais próximo do BMC carboxissomo. Até o momento, não foi descrita a existência de carboxissomo no filo *Actinomycetota*, mas a ocorrência de BMC com função desconhecida foi demonstrada neste filo e estudos já demonstraram que o filo Actinomycetota possui loci para α-carboxissomo (Axen et al., 2014; Sutter et al., 2021). De fato, a estirpe StC possui mais genes relacionados com α-carboxissomo, que incluem a forma IA da RuBisCo e duas anidrases carbônicas (β e ɣ), mas na análise do transcriptoma, apenas a β-anidrase carbônica foi expressa.

No genoma da StC, um gene CcmK/CsoS1 para uma suposta proteína estrutural do carboxissomo é seguido pelo invólucro da proteína CcmL e no lado oposto seguido por um cluster com genes para RuBisCo. CbbL e cbbS provavelmente são sequestrados nos carboxissomos e ajudam a enzima a fixar $CO_2$ em baixas concentrações de $CO_2$ (Yoshizawa et al., 2004). As observações dos filamentos da StC por Crio-ME CEMOVIS mostraram uma partícula poliédrica mostrando a forma típica de BMC, no entanto foi difícil determinar a estrutura, pois se esperava encontrar o BMC dentro da bactéria, e não flutuante fora da célula. O BMC isolado da estirpe também foi observado com microscopia de coloração negativa e Crio-ET com tamanho aproximado dos carboxissomos, 60 nm. Estes resultados sugerem que a estirpe StC possui carboxissomos ou outro BMC semelhante ao carboxissomo. Trabalhos futuros são necessários para o esclarecimento da função do BMC em StC, se é um tipo de carboxissomo, ou pode ser um outro BMC ainda não descrito.

## 7. CONCLUSÃO GERAL

No presente estudo apresentamos pela primeira vez dados de genes e proteínas expressas que corroboram a coexistência de duas vias funcionais de fixação de carbono - CBB e rTCA - em uma única bactéria - *Carbonactinospora thermoautotrophica* StC. Propomos evidências de um BMC relacionado ao carboxissomo pela primeira vez descrito em uma bactéria do filo *Actinomycetota*. Adicionalmente, descrevemos estudos recentes sobre a fixação natural de carbono e o desenvolvimento de vias sintética de fixação de carbono, bem como a importância do uso de novas técnicas de microscopia para caracterização de BMC presentes em células bacterianas, e concluímos com a importância do estudo da fixação biológica de $CO_2$ para que o sequestro do carbono ocorra naturalmente em grande escala.

Esses resultados reforçam estudos anteriores de que alguns microrganismos desenvolveram a capacidade de usar a fixação múltipla de carbono e abrir destaques para exploração futura em outros organismos, além de abrir caminhos para projetos de bioengenharia para a produção ideal usando ciclos integrados de fixação de carbono para melhorar ainda mais a capacidade de fixação de carbono em um organismo.

## 8. PERSPECTIVAS E TRABALHOS FUTUROS

A partir da realização do presente trabalho, observou-se que pode ser interessante novos estudos bioquímicos e fisiológicos para a compreender se as vias de fixação de $CO_2$ CBB e rTCA funcionam simultaneamente ou em espaço de tempo ou momento diferente, dependendo da necessidade fisiológica na estirpe StC. Em relação ao BMC, futuros estudos bioquímicos e microscopia são necessários para melhor caracterização desse BMC e para compreender a função deste BMC, se é um α ou β carboxissomo, ou ainda um novo tipo de BMC ainda não descrito. Dessa maneira, a partir dos resultados obtidos neste trabalho, propõe-se:

- Testar a atividade catalítica das enzimas do ciclo CBB e rTCA.
- Realizar testes com $CO_2$ marcado para percorrer o caminho desse elemento dentro da célula bacteriana.
- Medir a concentração de $CO_2$ marcado nos aminoácidos.
- Determinar o custo de energia dos ciclos CBB e rTCA.
- Determinar se a RuBisCo está no citosol e/ou dentro do BMC.
- Melhorar as imagens de microscopia do BMC.

## 9. REFERÊNCIAS


Adenan, N. H., Lim, Y. Y., & Ting, A. S. Y. (2020). Discovering decolorization potential of triphenylmethane dyes by Actinomycetota from soil. Water, Air, & Soil Pollution, 231(12), 560.

Al-Amoudi, A. S. (2004). Cryo-electron microscopy of vitreous sections of native biological cells and tissues: methodology and applications (Doctoral dissertation, Université de Lausanne, Faculté de biologie et médecine).

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. Nature reviews genetics, 11(10), 697-709.

Alm, E. J., Huang, K. H., Price, M. N., Koche, R. P., Keller, K., Dubchak, I. L., & Arkin, A. P. (2005). The MicrobesOnline Web site for comparative genomics. Genome research, 15(7), 1015-1022.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of molecular biology, 215(3), 403-410.

Andersson, I., & Backlund, A. (2008). Structure and function of Rubisco. Plant Physiology and Biochemistry, 46(3), 275-291.

Anthony, M. A., Crowther, T. W., Maynard, D. S., van den Hoogen, J., & Averill, C. (2020). Distinct assembly processes and microbial communities constrain soil organic carbon formation. One Earth, 2(4), 349-360.

Aoshima, M. (2007). Novel enzyme reactions related to the tricarboxylic acid cycle: phylogenetic/functional implications and biotechnological applications. Applied microbiology and biotechnology, 75, 249-255.

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., & Ogata, H. (2020). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics, 36(7), 2251-2252.

Araujo-Melo, R. D. O., de Oliveira, T. H. B., de Oliveira, C. V. J., de Araújo, J. M., de Sena, K. X., & Coelho, L. C. B. B. (2019). Actinomycetota: a renewable source of bioactive molecules with medical, industrial, and pharmacological importance. Advances and Trends in Biotechnology and Genetics, 1, 63-79.

Aussignargues, C., Paasch, B. C., Gonzalez-Esquer, R., Erbilgin, O., & Kerfeld, C. A. (2015). Bacterial microcompartment assembly: the key role of encapsulation peptides. Commun Integr Biol 8: e1039755.

Aversa, R., Petrescu, R. V., Apicella, A., & Petrescu, F. I. (2016). The basic elements of life's. American Journal of Engineering and Applied Sciences, 9(4), 1189-1197.

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(10), e1003898.

Badger, M. R., & Bek, E. J. (2008). Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO2 acquisition by the CBB cycle. Journal of experimental botany, 59(7), 1525-1541.

Badger, M. R., & Price, G. D. (2003). CO2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. Journal of experimental botany, 54(383), 609-622.

Badger, M. R., Hanson, D., & Price, G. D. (2002). Evolution and diversity of CO2 concentrating mechanisms in cyanobacteria. Functional Plant Biology, 29(3), 161-173.

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.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., ... & Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to

single-cell sequencing. Journal of computational biology, 19(5), 455-477.

Bar-Even, A., Flamholz, A., Noor, E., & Milo, R. (2012). Thermodynamic constraints shape the structure of carbon fixation pathways. Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1817(9), 1646-1659.

Bar-Even, A., Noor, E., Lewis, N. E., & Milo, R. (2010). Design and analysis of synthetic carbon fixation pathways. Proceedings of the National Academy of Sciences, 107(19), 8889-8894.

Bar-Joseph, Z., Gifford, D.K., Jaakkola, T.S. (2001) Fast optimal leaf ordering for hierarchical clustering, Bioinformatics, 17, 22-29.

Barka, E. A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H. P., ... & van Wezel, G. P. (2016). Taxonomy, physiology, and natural products of Actinomycetota. Microbiology and molecular biology reviews, 80(1), 1-43.

Barriuso, J., Ramos Solano, B., Lucas, J. A., Lobo, A. P., García-Villaraco, A., & Gutiérrez Mañero, F. J. (2008). Ecology, genetic diversity and screening strategies of plant growth promoting rhizobacteria (PGPR). Plant-bacteria interactions: Strategies and techniques to promote plant growth, 1-17.

Bathke, J., Konzer, A., Remes, B., McIntosh, M., & Klug, G. (2019). Comparative analyses of the variation of the transcriptome and proteome of Rhodobacter sphaeroides throughout growth. BMC genomics, 20, 1-13.

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.

Berg, I. A. (2011). Ecological aspects of the distribution of different autotrophic CO2 fixation pathways. Applied and environmental microbiology, 77(6), 1925-1936.

Berg, I. A., Kockelkorn, D., Ramos-Vera, W. H., Say, R. F., Zarzycki, J., Hügler, M., ... &

Fuchs, G. (2010). Autotrophic carbon fixation in archaea. Nature Reviews Microbiology, 8(6), 447-460.

Berghoff, B. A., Konzer, A., Mank, N. N., Looso, M., Rische, T., Förstner, K. U., ... & Klug, G. (2013). Integrative "omics"-approach discovers dynamic and regulatory features of bacterial stress responses. PLoS genetics, 9(6), e1003576.

Blikstad, C., Dugan, E. J., Laughlin, T. G., Liu, M. D., Shoemaker, S. R., Remis, J. P., & Savage, D. F. (2021). Discovery of a carbonic anhydrase-Rubisco supercomplex within the alpha-carboxysome. bioRxiv.

Blikstad, C., Dugan, E. J., Laughlin, T. G., Turnšek, J. B., Liu, M. D., Shoemaker, S. R., ... & Savage, D. F. (2023). Identification of a carbonic anhydrase–Rubisco complex within the alpha-carboxysome. Proceedings of the National Academy of Sciences, 120(43), e2308600120.

Blikstad, C., Flamholz, A. I., Oltrogge, L. M., & Savage, D. F. (2019). Learning to Build a β-Carboxysome. Biochemistry, 58(16), 2091-2092.

Boden, R., Hutt, L. P., Huntemann, M., Clum, A., Pillay, M., Palaniappan, K., ... & Kyrpides, N. (2016). Permanent draft genome of Thermithiobacillus tepidarius DSM 3134T, a moderately thermophilic, obligately chemolithoautotrophic member of the Acidithiobacillia. Standards in genomic sciences, 11(1), 1-8.

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. Proceedings of the National Academy of Sciences, 109(2), 478-483.

Borden, J. S., & Savage, D. F. (2021). New discoveries expand possibilities for carboxysome engineering. Current Opinion in Microbiology, 61, 58-66.

Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M. F., Lió, P., ... & Fondi, M. (2015). MeDuSa: a multi-draft-based scaffolder. Bioinformatics, 31(15), 2443-2451.

Bowien, B., & Schlegel, H. G. (1981). Physiology and biochemistry of aerobic hydrogen-

oxidizing bacteria. Annual Reviews in Microbiology, 35(1), 405-452.

Bracher, A., Whitney, S. M., Hartl, F. U., & Hayer-Hartl, M. (2017). Biogenesis and metabolic maintenance of Rubisco. Annual review of plant biology, 68, 29-60.

Bright, M., Klose, J., & Nussbaumer, A. D. (2013). Giant tubeworms. Current Biology, 23(6), R224-R225.

Busch, F. A. (2020). Photorespiration in the context of Rubisco biochemistry, CO2 diffusion and metabolism. The Plant Journal, 101(4), 919-939.

Busch, F. A., Sage, R. F., & Farquhar, G. D. (2018). Plants increase CO2 uptake by assimilating nitrogen via the photorespiratory pathway. Nature plants, 4(1), 46-54.

Byrne, A., Cole, C., Volden, R., & Vollmers, C. (2019). Realizing the potential of full-length transcriptome sequencing. Philosophical Transactions of the Royal Society B, 374(1786), 20190097.

Caemmerer, S. V., Farquhar, G. D., Christeiler, J. T., Laing, W. A., Field, C., Berry, J. A., ... & Stumpf, D. K. (1977). Rubp carboxylase in carboxysomes of blue green algae. Annual report of the director department of plant biology, 78, 83.

Cai, F., Kerfeld, C. A., & Sandh, G. (2012). Bioinformatic identification and structural characterization of a new carboxysome shell protein. Functional genomics and evolution of photosynthetic systems, 345-356.

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(10), e7521.

Cai, F., Sutter, M., Bernstein, S. L., Kinney, J. N., & Kerfeld, C. A. (2015). Engineering bacterial microcompartment shells: chimeric shell proteins and chimeric carboxysome shells. ACS synthetic biology, 4(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(22), 16055-16063.

Caldwell, P. E., MacLean, M. R., & Norris, P. R. (2007). Ribulose bisphosphate carboxylase activity and a Calvin cycle gene cluster in Sulfobacillus species. Micro., 153(7), 2231-2240.

Cameron, J. C., Wilson, S. C., Bernstein, S. L. & Kerfeld, C. A. (2013). Biogenesis of a bacterial organelle: the carboxysome assembly pathway. Cell, 155, 1131-1140.

Cannon, G. C., & Shively, J. M. (1983). Characterization of a homogenous preparation of carboxysomes from Thiobacillus neapolitanus. Archives of microbiology, 134(1), 52-59.

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(12), 5351-5361.

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

Carvalhais, N., Forkel, M., Khomik, M., Bellarby, J., Jung, M., Migliavacca, M., ... & Reichstein, M. (2014). Global covariation of carbon turnover times with climate in terrestrial ecosystems. Nature, 514(7521), 213-217.

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 α-carboxysome. Journal of Molecular Biology, 428(1), 153-164.

Chandran, H., Meena, M., & Sharma, K. (2020). Microbial biodiversity and bioremediation assessment through omics approaches. Frontiers in Environmental Chemistry, 1, 570326.

Chari, A., & Stark, H. (2023). Prospects and Limitations of High-Resolution Single-Particle

Cryo-Electron Microscopy. Annual Review of Biophysics, 52, 391-411.

Chen, A. H., Robinson-Mosher, A., Savage, D. F., Silver, P. A., & Polka, J. K. (2013). The bacterial carbon-fixing organelle is formed by shell envelopment of preassembled cargo. PloS one, 8(9), e76127.

Chen, Q., Yan, G., Gao, M., & Zhang, X. (2015). Ultrasensitive proteome profiling for 100 living cells by direct cell injection, online digestion and nano-LC-MS/MS analysis. Analytical chemistry, 87(13), 6674-6680.

Chen, X., Zheng, F., Wang, P., & Mi, H. (2023). Novel protein CcmS is required for stabilization of the assembly of β-carboxysome in Synechocystis sp. strain PCC 6803. New Phytologist, 239(4), 1266-1280.

Chlanda, P., & Sachse, M. (2014). Cryo-electron microscopy of vitreous sections. Electron Microscopy: Methods and Protocols, 193-214.

Christel, S., Herold, M., Bellenberg, S., El Hajjami, M., Buetti-Dinh, A., Pivkin, I. V., ... & Dopson, M. (2018). Multi-omics reveals the lifestyle of the acidophilic, mineral-oxidizing model species Leptospirillum ferriphilum T. Applied and environmental microbiology, 84(3), e02091-17.

Coleman, J. R. (1982). RuBP carboxylase in carboxysomes of blue-green algae. Carnegie Inst. Washington Year Book, 81, 83-87.

Coleman, J. R., & Colman, B. (1981). Inorganic carbon accumulation and photosynthesis in a blue-green alga as a function of external pH. Plant physiology, 67(5), 917-921.
Correa, S. S., Schultz, J., Lauersen, K. J., & Rosado, A. S. (2023). Natural carbon fixation and advances in synthetic engineering for redesigning and creating new fixation pathways. Journal of Advanced Research, 47, 75-92.

Criscuolo, A., & Gribaldo, S. (2011). Large-scale phylogenomic analyses indicate a deep origin of primary plastids within cyanobacteria. Molecular biology and evolution, 28(11), 3019-3032.
d Mattozzi, M., Ziesack, M., Voges, M. J., Silver, P. A., & Way, J. C. (2013). Expression of

the sub-pathways of the Chloroflexus aurantiacus 3-hydroxypropionate carbon fixation bicycle in E. coli: Toward horizontal transfer of autotrophic growth. Metabolic engineering, 16, 130-139.

Dahal, B., NandaKafle, G., Perkins, L., & Brözel, V. S. (2017). Diversity of free-Living nitrogen fixing Streptomyces in soils of the badlands of South Dakota. Microbiological Research, 195, 31-39.

Dai, W., Chen, M., Myers, C., Ludtke, S. J., Pettitt, B. M., King, J. A., ... & Chiu, W. (2018). Visualizing individual RuBisCO and its assembly into carboxysomes in marine cyanobacteria by cryo-electron tomography. Journal of molecular biology, 430(21), 4156-4167.

Dai, X., & Shen, L. (2022). Advances and trends in omics technology development. Frontiers in Medicine, 9, 911861.

Das, A., & Peu, S. D. (2022). A comprehensive review on recent advancements in thermochemical processes for clean hydrogen production to decarbonize the energy sector. Sustainability, 14(18), 11206.

Datta, R., Heaster, T. M., Sharick, J. T., Gillette, A. A., & Skala, M. C. (2020). Fluorescence lifetime imaging microscopy: fundamentals and advances in instrumentation, analysis, and applications. Journal of biomedical optics, 25(7), 071203-071203.

De Sena Brandine, G., & Smith, A. D. (2019). Falco: high-speed FastQC emulation for quality control of sequencing data. F1000Research, 8.

De Sousa, K. P., & Doolan, D. L. (2016). Immunomics: a 21st century approach to vaccine development for complex pathogens. Parasitology, 143(2), 236-244.

Desjardins, P., & Conklin, D. (2010). NanoDrop microvolume quantitation of nucleic acids. JoVE (Journal of Visualized Experiments), (45), e2565.

Desmarais, J. J., Flamholz, A. I., Blikstad, C., Dugan, E. J., Laughlin, T. G., Oltrogge, L. M., ... & Savage, D. F. (2019). DABs are inorganic carbon pumps found throughout prokaryotic phyla. Nature Microbiology, 4(12), 2204-2215.

Dobrinski, K. P., Longo, D. L., & Scott, K. M. (2005). The carbon-concentrating mechanism of the hydrothermal vent chemolithoautotroph Thiomicrospira crunogena. Journal of bacteriology, 187(16), 5761-5766.

Drews, G. & Niklowitz, W. (1956). Beiträge zur Cytologie der Blaualgen. II. Zentroplasma und granulare Einschlüsse von Phormidium uncinatum. Arch. Mikrobiol. 24, 147-162.

Du, J., Förster, B., Rourke, L., Howitt, S. M., & Price, G. D. (2014). Characterisation of cyanobacterial bicarbonate transporters in E. coli shows that SbtA homologs are functional in this heterologous expression system. PloS one, 9(12), e115905.

Ducat, D. C., & Silver, P. A. (2012). Improving carbon fixation pathways. Cur. Opi. in Chem. Biol., 16(3-4), 337-344.

Dutt, M. J., & Lee, K. H. (2000). Proteomic analysis. Current opinion in biotechnology, 11(2), 176-179.

Eisenhut, M., Roell, M. S., & Weber, A. P. (2019). Mechanistic understanding of photorespiration paves the way to a new green revolution. New Phytologist, 223(4), 1762-1769.

Ellis, R. J. (1979). The most abundant protein in the world. Trends in biochemical sciences, 4(11), 241-244.

Ellis, R. J. (2010). Tackling unintelligent design. Nature, 463(7278), 164-165.

English, R. S., Jin, S., & Shively, J. M. (1995). Use of electroporation to generate a Thiobacillus neapolitanus carboxysome mutant. Applied and environmental microbiology, 61(9), 3256-3260.

Esselborn, J., Kertess, L., Apfel, U. P., Hofmann, E., & Happe, T. (2019). Loss of specific active-site iron atoms in oxygen-exposed [FeFe]-hydrogenase determined by detailed X-ray structure analyses. Journal of the American Chemical Society, 141(44), 17721-17728.

Evans, M. C., Buchanan, B. B., & Arnon, D. I. (1966). A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. Proceedings of the National Academy of Sciences, 55(4), 928-934.

Evans, S. L., Al-Hazeem, M. M., Mann, D., Smetacek, N., Beavil, A. J., Sun, Y., ... & Bergeron, J. R. (2023). Single-particle cryo-EM analysis of the shell architecture and internal organization of an intact α-carboxysome. Structure, 31(6), 677-688.

Evans, S. L., Al-Hazeem, M. M., Mann, D., Smetacek, N., Beavil, A. J., Sun, Y., ... & Bergeron, J. R. (2022). Single-particle cryo-EM analysis of the shell architecture and internal organization of an intact α-carboxysome. bioRxiv, 2022-02.

Ezaki, S., Maeda, N., Kishimoto, T., Atomi, H., & Imanaka, T. (1999). Presence of a structurally novel type ribulose-bisphosphate carboxylase/oxygenase in the hyperthermophilic archaeon, Pyrococcus kodakaraensis KOD1. Journal of Biological Chemistry, 274(8), 5078-5082.

Ekblom, R., & Wolf, J. B. (2014). A field guide to whole-genome sequencing, assembly and annotation. Evolutionary applications, 7(9), 1026-1042.

Faulkner, M., Rodriguez-Ramos, J., Dykes, G. F., Owen, S. V., Casella, S., Simpson, D. M., ... & Liu, L. N. (2017). Direct characterization of the native structure and mechanics of cyanobacterial carboxysomes. Nanoscale, 9(30), 10662-10673.

Fei, C., Wilson, A. T., Mangan, N. M., Wingreen, N. S., & Jonikas, M. C. (2022). Modelling the pyrenoid-based CO2-concentrating mechanism provides insights into its operating principles and a roadmap for its engineering into crops. Nature Plants, 8(5), 583-595.

Fernández-Martínez, M., Peñuelas, J., Chevallier, F., Ciais, P., Obersteiner, M., Rödenbeck, C., ... & Janssens, I. A. (2023). Diagnosing destabilization risk in global land carbon sinks. Nature, 615(7954), 848-853.

Ferreira, N. G., Morgado, R. G., Cunha, L., Novo, M., Soares, A. M., Morgan, A. J., ... & Kille, P. (2019). Unravelling the molecular mechanisms of nickel in woodlice. Environmental

research, 176, 108507.

Finn, M. W., & Tabita, F. R. (2004). Modified pathway to synthesize ribulose 1, 5-bisphosphate in methanogenic archaea. Journal of bacteriology, 186(19), 6360-6366.

Firmino, A. A. P., Gorka, M., Graf, A., Skirycz, A., Martinez-Seidel, F., Zander, K., ... & Beine-Golovchuk, O. (2020). Separation and paired proteome profiling of plant chloroplast and cytoplasmic ribosomes. Plants, 9(7), 892.

Flamholz, A. I., Dugan, E., Blikstad, C., Gleizer, S., Ben-Nissan, R., Amram, S., ... & Savage, D. F. (2020). Functional reconstitution of a bacterial CO2 concentrating mechanism in Escherichia coli. Elife, 9, e59882.

Flamholz, A. I., Prywes, N., Moran, U., Davidi, D., Bar-On, Y. M., Oltrogge, L. M., ... & Milo, R. (2019). Revisiting trade-offs between Rubisco kinetic parameters. Biochemistry, 58(31), 3365-3376.

Flood, B. E., Fliss, P., Jones, D. S., Dick, G. J., Jain, S., Kaster, A. K., ... & Bailey, J. (2016). Single-cell (meta-) genomics of a dimorphic Candidatus Thiomargarita nelsonii reveals genomic plasticity. Frontiers in microbiology, 7, 603.

Flynn, C. E., Lee, S. W., Peelle, B. R., & Belcher, A. M. (2003). Viruses as vehicles for growth, organization and assembly of materials. Acta Materialia, 51(19), 5867-5880.

Ford, B. J. (1989). Antony van Leeuwenhoek-Microscopist and visionary scientist. Journal of Biological Education, 23(4), 293-299.

Frolov, E. N., Kublanov, I. V., Toshchakov, S. V., Lunev, E. A., Pimenov, N. V., Bonch-Osmolovskaya, E. A., ... & Chernyh, N. A. (2019). Form III RubisCO-mediated transaldolase variant of the Calvin cycle in a chemolithoautotrophic bacterium. Proceedings of the National Academy of Sciences, 116(37), 18638-18646.

Fuchs, G. (2011). Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? Annual review of microbiology, 65, 631-658.

Gadkari, D. I. L. I. P., Mörsdorf, G., & Meyer, O. R. T. W. I. N. (1992). Chemolithoautotrophic assimilation of dinitrogen by Streptomyces thermoautotrophicus UBT1: identification of an unusual N2-fixing system. Journal of bacteriology, 174(21), 6840-6843.

Gadkari, D., Schricker, K., Acker, G., Kroppenstedt, R. M., & Meyer, O. (1990). Streptomyces thermoautotrophicus sp. nov., a thermophilic CO-and H2-oxidizing obligate chemolithoautotroph. Applied and environmental microbiology, 56(12), 3727-3734.

Gan, L., & Jensen, G. J. (2012). Electron tomography of cells. Quarterly reviews of biophysics, 45(1), 27-56.

Gao, R., Tan, H., Li, S., Ma, S., Tang, Y., Zhang, K., ... & Li, F. (2022). A prototype protein nanocage minuteimized from carboxysomes with gated oxygen permeability. Proceedings of the National Academy of Sciences, 119(5), e2104964119.

Gardebrecht, A., Markert, S., Sievert, S. M., Felbeck, H., Thürmer, A., Albrecht, D., ... & Schweder, T. (2012). Physiological homogeneity among the endosymbionts of Riftia pachyptila and Tevnia jerichonana revealed by proteogenomics. The ISME Journal, 6(4), 766-776.

Garritano, A. N., Song, W., & Thomas, T. (2022). Carbon fixation pathways across the bacterial and archaeal tree of life. PNAS nexus, 1(5), pgac226.

Giani, A. M., Gallo, G. R., Gianfranceschi, L., & Formenti, G. (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. Computational and Structural Biotechnology Journal, 18, 9-19.

Giessen, T. W., Orlando, B. J., Verdegaal, A. A., Chambers, M. G., Gardener, J., Bell, D. C., ... & Silver, P. A. (2019). Large protein organelles form a new iron sequestration system with high storage capacity. Elife, 8, e46070.

Giovannelli, D., Sievert, S. M., Hügler, M., Markert, S., Becher, D., Schweder, T., & Vetriani, C. (2017). Insight into the evolution of microbial metabolism from the deep-branching bacterium, Thermovibrio ammonificans. Elife, 6, e18990.

Gong, F., Zhu, H., Zhang, Y., & Li, Y. (2018). Biological carbon fixation: from natural to synthetic. Journal of CO2 Utilization, 28, 221-227.

Greber, B. J., Sutter, M., & Kerfeld, C. A. (2019). The plasticity of molecular interactions governs bacterial microcompartment shell assembly. Structure, 27(5), 749-763.

Gupta, K., Biswas, R., & Sarkar, A. (2020). Advancement of omics: prospects for bioremediation of contaminuteated soils. Microbial bioremediation & biodegradation, 113-142.

Gurrieri, L., Fermani, S., Zaffagnini, M., Sparla, F., & Trost, P. (2021). Calvin–Benson cycle regulation is getting complex. Trends in Plant Science, 26(9), 898-912.

Haider, S., & Pal, R. (2013). Integrated analysis of transcriptomic and proteomic data. Current genomics, 14(2), 91-110.

Hallenbeck, P. C. (2009). Fermentative hydrogen production: principles, progress, and prognosis. International Journal of Hydrogen Energy, 34(17), 7379-7389.

Hanson, D. T. (2016). Breaking the rules of Rubisco catalysis. Journal of experimental botany, 67(11), 3180-3182.

Hazarika, S. N., & Thakur, D. (2020). Actinomycetota. In Beneficial microbes in agro- ecology (pp. 443-476). Academic Press.

Hegde, P. S., White, I. R., & Debouck, C. (2003). Interplay of transcriptomics and proteomics. Current opinion in biotechnology, 14(6), 647-651.

Hein, M. Y., Sharma, K., Cox, J., & Mann, M. (2013). Proteomic analysis of cellular systems. In Handbook of systems biology: concepts and insights (pp. 3-25). Academic Press.

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(23), 8087-8094.

Hell, S. W. (2003). Toward fluorescence nanoscopy. Nature biotechnology, 21(11), 1347-1355.

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.

Hill, N. C., Tay, J. W., Altus, S., Bortz, D. M., & Cameron, J. C. (2020). Life cycle of a cyanobacterial carboxysome. Science Advances, 6(19), eaba1269.

Hoegh-Guldberg, O., Jacob, D., Taylor, M., Guillén Bolaños, T., Bindi, M., Brown, S., ... & Zhou, G. (2019). The human imperative of stabilizing global climate change at 1.5
C. Science, 365(6459), eaaw6974.

Holmes, D. E., Shrestha, P. M., Walker, D. J., Dang, Y., Nevin, K. P., Woodard, T. L., & Lovley, D. R. (2017). Metatranscriptomic evidence for direct interspecies electron transfer between Geobacter and Methanothrix species in methanogenic rice paddy soils. Applied and environmental microbiology, 83(9), e00223-17.

Holthuijzen, Y. A., Van Breemen, J. F., Kuenen, J. G., & Konings, W. N. (1986). Protein composition of the carboxysomes of Thiobacillus neapolitanus. Archives of microbiology, 144(4), 398-404.

Hong, Y., Song, Y., Zhang, Z., & Li, S. (2023). Cryo-Electron Tomography: The Resolution Revolution and a Surge of In Situ Virological Discoveries. Annual Review of Biophysics, 52, 339-360.

Hu, J., Jin, K., He, Z. G., & Zhang, H. (2020). Citrate lyase CitE in Mycobacterium tuberculosis contributes to mycobacterial survival under hypoxic conditions. PLoS One, 15(4), e0230786.

Huang, F., Vasieva, O., Sun, Y., Faulkner, M., Dykes, G. F., Zhao, Z., & Liu, L. N. (2019). Roles of RbcX in carboxysome biosynthesis in the cyanobacterium Synechococcus elongatus PCC7942. Plant Physiology, 179(1), 184-194.

Huffine, C. A., Zhao, R., Tang, Y. J., & Cameron, J. C. (2023). Role of carboxysomes in cyanobacterial CO2 assimilation: CO2 concentrating mechanisms and metabolon implications. Environmental Microbiology, 25(2), 219-228.

Hugler, M., & Sievert, S. M. (2011). Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. Annual review of marine science, 3, 261-289.

Hugler, M., Huber, H., Stetter, K. O., & Fuchs, G. (2003). Autotrophic CO 2 fixation pathways in archaea (Crenarchaeota). Archives of microbiology, 179, 160-173.

Hutt, L. P., Huntemann, M., Clum, A., Pillay, M., Palaniappan, K., Varghese, N., ... & Boden, R. (2017). Permanent draft genome of Thiobacillus thioparus DSM 505T, an obligately chemolithoautotrophic member of the Betaproteobacteria. Standards in genomic sciences, 12(1), 1-8.

Iancu, C. V., Morris, D. M., Dou, Z., Heinhorst, S., Cannon, G. C., & Jensen, G. J. (2010). Organization, structure, and assembly of α-carboxysomes determined by electron cryotomography of intact cells. Journal of molecular biology, 396(1), 105-117.

Imai, K., Asano, A., & Sato, R. (1967). Oxidative phosphorylation in Micrococcus denitrificans. I. Preparation and properties of phosphorylating membrane fragments. Biochimica et Biophysica Acta (BBA)-Bioenergetics, 143(3), 462-476.

Imm, S., & Chang, Y. (2023). Complete Genome Sequence of the Bacillus cereus Temperate Bacteriophage BSG01. Microbiology Resource Announcements, 12(6), e00217-23.

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(40), 24519-24533.

Jiang, Q., Li, T., Yang, J., Aitchison, C. M., Huang, J., Chen, Y., ... & Liu, L. N. (2023). Synthetic engineering of a new biocatalyst encapsulating [NiFe]-hydrogenases for enhanced hydrogen production. Journal of Materials Chemistry B, 11(12), 2684-2692.

Jiao, J. Y., Fu, L., Hua, Z. S., Liu, L., Salam, N., Liu, P. F., ... & Li, W. J. (2021). Insight into the function and evolution of the Wood-Ljungdahl pathway in Actinomycetota. The ISME journal, 15(10), 3005-3018.

Johnson, R. M., Ramond, J. B., Gunnigle, E., Seely, M., & Cowan, D. A. (2017). Namib Desert edaphic bacterial, fungal and archaeal communities assemble through determinuteistic processes but are influenced by different abiotic parameters. Extremophiles, 21, 381-392.

Ju, Y., Zhang, H., Jiang, Y., Wang, W., Kan, G., Yu, K., ... & Jiang, J. (2023). Aqueous microdroplets promote C–C bond formation and sequences in the reverse tricarboxylic acid cycle. Nature Ecology & Evolution, 7(11), 1892-1902.

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), 388.

Kanao, T., Kawamura, M., Fukui, T., Atomi, H., & Imanaka, T. (2002). Characterization of isocitrate dehydrogenase from the green sulfur bacterium Chlorobium limicola: A carbon dioxide-fixing enzyme in the reductive tricarboxylic acid cycle. European Journal of Biochemistry, 269(7), 1926-1931.

Kandler, 0. (1981) in Photosynthesis, ed. Akoyunoglou, G. Balaban lnt. Sci. Services, Philadelphia, Vol. 1, pp. 3-14.

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. Journal of molecular biology, 428(4), 726-731.

Kawashima, Y., & Ohara, O. (2018). Development of a NanoLC–MS/MS system using a nonporous reverse phase column for ultrasensitive proteome analysis. Analytical chemistry, 90(21), 12334-12338.

Kawashima, Y., Watanabe, E., Umeyama, T., Nakajima, D., Hattori, M., Honda, K., & Ohara, O. (2019). Optimization of data-independent acquisition mass spectrometry for deep and highly

sensitive proteomic analysis. International journal of molecular sciences, 20(23), 5932.

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., & Scott, K. M. (2022). Atypical Carboxysome Loci: JEEPs or Junk? Insights in Microbial Physiology and Metabolism: 2021.

Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F., & Sutter, M. (2018). Bacterial microcompartments. Nature Reviews Microbiology, 16(5), 277-290.

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(5736), 936-938.

Khomyakova, M. A., Zavarzina, D. G., Merkel, A. Y., Klyukina, A. A., Pikhtereva, V. A., Gavrilov, S. N., & Slobodkin, A. I. (2022). The first cultivated representatives of the Actinomycetotal lineage OPB41 isolated from subsurface environments constitute a novel order Anaerosomatales. Frontiers in Microbiology, 13, 1047580.

Kim, S., Lindner, S. N., Aslan, S., Yishai, O., Wenk, S., Schann, K., & Bar-Even, A. (2020). Growth of E. coli on formate and methanol via the reductive glycine pathway. Nature chemical biology, 16(5), 538-545.

King, G. M., & Weber, C. F. (2007). Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. Nature Reviews Microbiology, 5(2), 107-118.

Kinney, J. N., Axen, S. D., & Kerfeld, C. A. (2011). Comparative analysis of carboxysome shell proteins. Photosynthesis Research, 109(1), 21-32.

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.

Kitadai, N., Kameya, M., & Fujishima, K. (2017). Origin of the reductive tricarboxylic acid (rTCA) cycle-type CO2 fixation: A perspective. Life, 7(4), 39.

Klatt, J. M., & Polerecky, L. (2015). Assessment of the stoichiometry and efficiency of CO2 fixation coupled to reduced sulfur oxidation. Frontiers in Microbiology, 6, 484.

Klein, M. G., Zwart, P., Bagby, S. C., Cai, F., Chisholm, S. W., Heinhorst, S., ... & Kerfeld, C. A. (2009). Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. Journal of molecular biology, 392(2), 319-333.

Knowlton, N., Corcoran, E., Felis, T., de Goeij, J., Grottoli, A., Harding, S., ... & Ferse, S. (2021). Rebuilding coral reefs: a decadal grand challenge.

Kono, T., Mehrotra, S., Endo, C., Kizu, N., Matusda, M., Kimura, H., ... & Ashida, H. (2017). A RuBisCO-mediated carbon metabolic pathway in methanogenic archaea. Nature communications, 8(1), 14007.

Kühlbrandt, W. (2014). The resolution revolution. Science, 343(6178), 1443-1444.

Kusian, B., & Bowien, B. (1997). Organization and regulation of cbb CO2 assimilation genes in autotrophic bacteria. FEMS microbiology reviews, 21(2), 135-155.

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(14), 3857-3867.

León-Sobrino, C., Ramond, J. B., Maggs-Kölling, G., & Cowan, D. A. (2019). Nutrient acquisition, rather than stress response over diel cycles, drives microbial transcription in a hyper-arid Namib Desert soil. Frontiers in microbiology, 10, 1054.

Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic acids research, 44(W1), W242-W245.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & 1000 Genome Project

Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. bioinformatics, 25(16), 2078-2079.

Li, J., Liu, T., McIlroy, S. J., Tyson, G. W., & Guo, J. (2023). Phylogenetic and metabolic diversity of microbial communities performing anaerobic ammonium and methane oxidations under different nitrogen loadings. ISME communications, 3(1), 39.

Li, T., Jiang, Q., Huang, J., Aitchison, C. M., Huang, F., Yang, M., ... & Liu, L. N. (2020). Reprogramminuteg bacterial protein organelles as a nanoreactor for hydrogen production. Nature communications, 11(1), 5448.

Liao, Z. L., Chen, H., Zhu, B. R., & Li, H. Z. (2015). Combination of powdered activated carbon and powdered zeolite for enhancing ammonium removal in micro-polluted raw water. Chemosphere, 134, 127-132.

Lichtman, J. W., & Conchello, J. A. (2005). Fluorescence microscopy. Nature methods, 2(12), 910-919.

Liedtke, J., Depelteau, J. S., & Briegel, A. (2022). How advances in cryo-electron tomography have contributed to our current view of bacterial cell biology. Journal of Structural Biology: X, 6, 100065.

Lin, M. T., Occhialini, A., Andralojc, P. J., Devonshire, J., Hines, K. M., Parry, M. A., & Hanson, M. R. (2014a). β-Carboxysomal proteins assemble into highly organized structures in Nicotiana chloroplasts. The Plant Journal, 79(1), 1-12.

Lin, M. T., Occhialini, A., Andralojc, P. J., Parry, M. A., & Hanson, M. R. (2014b). A faster Rubisco with potential to increase photosynthesis in crops. Nature, 513(7519), 547-550.

Liu, L. N. (2021). Bacterial metabolosomes: new insights into their structure and bioengineering. Microbial Biotechnology, 14(1), 88-93.

Liu, L. N. (2022). Advances in the bacterial organelles for CO2 fixation. Trends in microbiology, 30(6), 567-580.

Liu, Y., Beyer, A., & Aebersold, R. (2016). On the dependency of cellular protein levels on mRNA abundance. Cell, 165(3), 535-550.

Liu, Z., Wang, K., Chen, Y., Tan, T., & Nielsen, J. (2020). Third generation biorefineries as the means to produce fuels and chemicals from CO2. Nature catalysis, 3(3), 274-288.

Livingston, A. K., Cruz, J. A., Kohzuma, K., Dhingra, A., & Kramer, D. M. (2010). An Arabidopsis mutant with high cyclic electron flow around photosystem I (hcef) involving the NADPH dehydrogenase complex. The Plant Cell, 22(1), 221-233.

Ljungdhal, L. G. (1986). The autotrophic pathway of acetate synthesis in acetogenic bacteria. Annual Reviews in Microbiology, 40(1), 415-450.

Lo, S. C., Huang, C. C., Ho, T. Y., & Yang, Y. T. (2019). Detailed profiling of carbon fixation of in silico synthetic autotrophy with reductive tricarboxylic acid cycle and Calvin-Benson-Bassham cycle in Esherichia coli using hydrogen as an energy source. Synthetic and Systems Biotechnology, 4(3), 165-172.

Loeve, S., & Vincent, B. B. (2017). 12 The multiple signatures of carbon. Research objects in their technological setting, 185.

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(40), 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(18), e2014406118.

Long, B. M., Hee, W. Y., Sharwood, R. E., Rae, B. D., Kaines, S., Lim, Y. L., ... & Price, G. D. (2018). Carboxysome encapsulation of the CO2-fixing enzyme Rubisco in tobacco chloroplasts. Nature communications, 9(1), 1-14.

Long, B. M., Price, G. D., & Badger, M. R. (2005). Proteomic assessment of an established

technique for carboxysome enrichment from Synechococcus PCC7942. Canadian Journal of Botany, 83(7), 746-757.

Long, S. P., Marshall-Colon, A., & Zhu, X. G. (2015). Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. Cell, 161(1), 56-66.

Lu, X., Liu, Y., Yang, Y., Wang, S., Wang, Q., Wang, X., ... & Jiang, H. (2019). Constructing a synthetic pathway for acetyl-coenzyme A from one-carbon through enzyme design. Nature communications, 10(1), 1378.

Lu, Y., Yeung, N., Sieracki, N., & Marshall, N. M. (2009). Design of functional metalloproteins. Nature, 460(7257), 855-862.

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(6), 1109-1119.

Ludwig, M., Yellowlees, D., Leggat, W., & Price, G. D. (1998). The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO2-concentrating mechanisms in algae. Can J Bot 76: 1052-1071. Can. J. Bot, 76, 1052-1071.

Luedin, S. M., Storelli, N., Danza, F., Roman, S., Wittwer, M., Pothier, J. F., & Tonolla, M. (2019). Mixotrophic growth under micro-oxic conditions in the purple sulfur bacterium “Thiodictyon syntrophicum”. frontiers in microbiology, 10, 384.

MacCready, J. S., & Vecchiarelli, A. G. (2021). Positioning the model bacterial organelle, the carboxysome. MBio, 12(3), 10-1128.

MacCready, J. S., Basalla, J. L., & Vecchiarelli, A. G. (2020). Origin and evolution of carboxysome positioning systems in cyanobacteria. Molecular biology and evolution, 37(5), 1434-1451.

MacCready, J. S., Tran, L., Basalla, J. L., Hakim, P., & Vecchiarelli, A. G. (2021). The McdAB system positions α-carboxysomes in proteobacteria. Molecular Microbiology, 116(1), 277-297.

MacKellar, D., Lieber, L., Norman, J. S., Bolger, A., Tobin, C., Murray, J. W., ... & Prell, J. (2016). Streptomyces thermoautotrophicus does not fix nitrogen. Scientific reports, 6(1), 20086.

Maeda, S. I., Badger, M. R., & Price, G. D. (2002). Novel gene products associated with NdhD3/D4-containing NDH-1 complexes are involved in photosynthetic CO2 hydration in the cyanobacterium, Synechococcus sp. PCC7942. Molecular microbiology, 43(2), 425-435.

Maier, T., Güell, M., & Serrano, L. (2009). Correlation of mRNA and protein in complex biological samples. FEBS letters, 583(24), 3966-3973.

Mall, A., Sobotta, J., Huber, C., Tschirner, C., Kowarschik, S., Bačnik, K., ... & Berg, I. A. (2018). Reversibility of citrate synthase allows autotrophic growth of a thermophilic bacterium. Science, 359(6375), 563-567.

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(4), 1197-1210.

Mangan, N. M., & Brenner, M. P. (2014). Systems analysis of the CO2 concentrating mechanism in cyanobacteria. Elife, 3, e02043.

Mangan, N. M., Flamholz, A., Hood, R. D., Milo, R., & Savage, D. F. (2016). pH determinutees the energetic efficiency of the cyanobacterial CO2 concentrating mechanism. Proceedings of the National Academy of Sciences, 113(36), E5354-E5362.

Marco, E., Martinez, I., Ronen-Tarazi, M., Orus, M. I., & Kaplan, A. (1994). Inactivation of ccmO in Synechococcus sp. strain PCC 7942 results in a mutant requiring high levels of CO (2). Applied and Environmental Microbiology, 60, 1018–1020.

Marin, B., Nowack, E., Glöckner, G., & Melkonian, M. (2007). The ancestor of the Paulinella chromatophore obtained a carboxysomal operon by horizontal gene transfer from a Nitrococcus-like γ-proteobacterium. BMC Evolutionary Biology, 7(1), 1-14.

Martin, J. A., & Wang, Z. (2011). Next-generation transcriptome assembly. Nature Reviews Genetics, 12(10), 671-682.

McAllister, S. M., Polson, S. W., Butterfield, D. A., Glazer, B. T., Sylvan, J. B., & Chan, C. S. (2020). Validating the Cyc2 neutrophilic iron oxidation pathway using meta-omics of Zetaproteobacteria iron mats at marine hydrothermal vents. Msystems, 5(1), e00553-19.

McDowall, J., & Hunter, S. (2011). InterPro protein classification. Bioinformatics for Comparative Proteomics, 37-47.

McGuire, A. L., Gabriel, S., Tishkoff, S. A., Wonkam, A., Chakravarti, A., Furlong, E. E., ... & Kim, J. S. (2020). The road ahead in genetics and genomics. Nature Reviews Genetics, 21(10), 581-596.

McKinlay, J. B., & Harwood, C. S. (2010). Carbon dioxide fixation as a central redox cofactor recycling mechanism in bacteria. Proceedings of the National Academy of Sciences, 107(26), 11669-11675.

McManus, J., Cheng, Z., & Vogel, C. (2015). Next-generation analysis of gene expression regulation–comparing the roles of synthesis and degradation. Molecular BioSystems, 11(10), 2680-2689.

Meier-Kolthoff, J. P., & Göker, M. (2019). TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nature communications, 10(1), 2182.

Meier-Kolthoff, J. P., Carbasse, J. S., Peinado-Olarte, R. L., & Göker, M. (2022). TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic acids research, 50(D1), D801-D807.

Melnicki, M. R., Sutter, M., & Kerfeld, C. A. (2021). Evolutionary relationships among shell proteins of carboxysomes and metabolosomes. Current Opinion in Microbiology, 63, 1-9.

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(10), e3570.

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. Journal of bacteriology, 173(15), 4851-4861.

Metskas, L. A., Ortega, D., Oltrogge, L. M., Blikstad, C., Lovejoy, D. R., Laughlin, T. G., ... & Jensen, G. J. (2022). Rubisco forms a lattice inside alpha-carboxysomes. Nature Communications, 13(1), 4863.

Meyer, O. (1982). Chemical and spectral properties of carbon monoxide: methylene blue oxidoreductase. The molybdenum-containing iron-sulfur flavoprotein from Pseudomonas carboxydovorans. Journal of Biological Chemistry, 257(3), 1333-1341.

Meyer, O., & Schlegel, H. G. (1983). Biology of aerobic carbon monoxide-oxidizing bacteria. Annual review of microbiology, 37(1), 277-310.

Micheel, J., Safrastyan, A., & Wollny, D. (2021). Advances in non-coding RNA sequencing. Non-coding RNA, 7(4), 70.

Mills, C. E., Waltmann, C., Archer, A. G., Kennedy, N. W., Abrahamson, C. H., Jackson, A. D., ... & Tullman-Ercek, D. (2022). Vertex protein PduN tunes encapsulated pathway performance by dictating bacterial metabolosome morphology. Nature communications, 13(1), 3746.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L., ... & Bateman, A. (2021). Pfam: The protein families database in 2021. Nucleic acids research, 49(D1), D412-D419.

Moreira, R. G. G. (2018). Isolamento e identificação microscópica e molecular de Actinomycetota componente de consórcio microbiano termofílico quimiolitoautotrófico obtido do solo. Repositório Institucional Pantheon.

Mudhoo, A., Forster-Carneiro, T., & Sánchez, A. (2011). Biohydrogen production and

bioprocess enhancement: a review. Critical reviews in biotechnology, 31(3), 250-263.

Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P. M., ... & Scheres, S. H. (2020). Single-particle cryo-EM at atomic resolution. Nature, 587(7832), 152-156.

Nam, K. T., Kim, D. W., Yoo, P. J., Chiang, C. Y., Meethong, N., Hammond, P. T., ... & Belcher, A. M. (2006). Virus-enabled synthesis and assembly of nanowires for lithium-ion battery electrodes. science, 312(5775), 885-888.

Nguyen, N. D., Pulsford, S. B., Hee, W. Y., Rae, B. D., Rourke, L. M., Price, G. D., & Long, B. M. (2023). Towards engineering a hybrid carboxysome. Photosynthesis Research, 156(2), 265-277.

Nguyen, T. K., & Ueno, T. (2018). Engineering of protein assemblies within cells. Current Opinion in Structural Biology, 51, 1-8.

Ni, T., Jiang, Q., Ng, P. C., Shen, J., Dou, H., Zhu, Y., ... & Zhang, P. (2023). Intrinsically disordered CsoS2 acts as a general molecular thread for α-carboxysome shell assembly. Nature Communications, 14(1), 5512.

Ni, T., Sun, Y., Burn, W., Al-Hazeem, M. M., Zhu, Y., Yu, X., ... & Zhang, P. (2022). Structure and assembly of cargo Rubisco in two native α-carboxysomes. Nature Communications, 13(1), 4299.

Nickerson, J. L., & Doucette, A. A. (2020). Rapid and quantitative protein precipitation for proteome analysis by mass spectrometry. Journal of proteome research, 19(5), 2035-2042.

Nunoura, T., Chikaraishi, Y., Izaki, R., Suwa, T., Sato, T., Harada, T., ... & Takai, K. (2018). A primordial and reversible TCA cycle in a facultatively chemolithoautotrophic thermophile. Science, 359(6375), 559-563.

Nygaard, R., Kim, J., & Mancia, F. (2020). Cryo-electron microscopy analysis of small membrane proteins. Current opinion in structural biology, 64, 26-33.

Occhialini, A., Lin, M. T., Andralojc, P. J., Hanson, M. R., & Parry, M. A. (2016). Transgenic tobacco plants with improved cyanobacterial Rubisco expression but no extra assembly factors grow at near wild-type rates if provided with elevated CO 2. The Plant Journal, 85(1), 148- 160.

Ohi, M., Li, Y., Cheng, Y., & Walz, T. (2004). Negative staining and image classification-powerful tools in modern electron microscopy. Biological procedures online, 6, 23-34.

Ohta, J. (2022). A novel variant of the Calvin–Benson cycle bypassing fructose bisphosphate. Scientific reports, 12(1), 3984.

Oltrogge, L. M., Chaijarasphong, T., Chen, A. W., Bolin, E. R., Marqusee, S., & Savage, D. F. (2020). Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation. Nature structural & molecular biology, 27(3), 281-287.

Omae, K., Fukuyama, Y., Yasuda, H., Mise, K., Yoshida, T., & Sako, Y. (2019). Diversity and distribution of thermophilic hydrogenogenic carboxydotrophs revealed by microbial community analysis in sediments from multiple hydrothermal environments in Japan. Archives of microbiology, 201, 969-982.

Oren, A., & Garrity, G. M. (2021). Valid publication of the names of forty-two phyla of prokaryotes. International journal of systematic and evolutionary microbiology, 71(10), 005056.

Ort, D. R., Merchant, S. S., Alric, J., Barkan, A., Blankenship, R. E., Bock, R., ... & Zhu, X. G. (2015). Redesigning photosynthesis to sustainably meet global food and bioenergy demand. Proceedings of the national academy of sciences, 112(28), 8529-8536.

Panneerselvam, P., Selvakumar, G., Ganeshamurthy, A. N., Mitra, D., & Senapati, A. (2021). Enhancing pomegranate (Punica granatum L.) plant health through the intervention of a Streptomyces consortium. Biocontrol Science and Technology, 31(4), 430-442.

Parry, M. A. J., Andralojc, P. J., Mitchell, R. A., Madgwick, P. J., & Keys, A. J. (2003). Manipulation of Rubisco: the amount, activity, function and regulation. Journal of experimental botany, 54(386), 1321-1333.

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., ... & Rokhsar, D. S. (2009). The Sorghum bicolor genome and the diversification of grasses. Nature, 457(7229), 551-556.

Pena, K. L, Castel, S. E, de Araujo C., Espie, G. S., Kimbera, M. S. (2010). Structural basis of the oxidative activation of the carboxysomal γ -carbonic anhydrase, CcmM. P Natl Acad Sci, 107, 2455-60.

Penrod, J. T.; Roth, J. R. (2006). Conserving a Volatile Metabolite: A Role for Carboxysome-Like Organelles in Salmonella enterica. Journal of Bacteriology, 188(8), 2865-2874.

Peters, J. W., Schut, G. J., Boyd, E. S., Mulder, D. W., Shepard, E. M., Broderick, J. B., ... & Adams, M. W. (2015). [FeFe]-and [NiFe]-hydrogenase diversity, mechanism, and maturation. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 1853(6), 1350-1369.

Petricoin, E. F., Zoon, K. C., Kohn, E. C., Barrett, J. C., & Liotta, L. A. (2002). Clinical proteomics: translating benchside promise into bedside reality. Nature reviews Drug discovery, 1(9), 683-695.

Pierce, J., Carlson, T. J., & Williams, J. G. (1989). A cyanobacterial mutant requiring the expression of ribulose bisphosphate carboxylase from a photosynthetic anaerobe. Proceedings of the National Academy of Sciences, 86(15), 5753-5757.

Pinheiro, Y., Faria da Mota, F., Peixoto, R. S., van Elsas, J. D., Lins, U., Mazza Rodrigues, J. L., & Rosado, A. S. (2023). A thermophilic chemolithoautotrophic bacterial consortium suggests a mutual relationship between bacteria in extreme oligotrophic environments. Communications biology, 6(1), 230.

Pokhrel, A., Kang, S. Y., & Schmidt-Dannert, C. (2021). Ethanolamine bacterial microcompartments: from structure, function studies to bioengineering applications. Current Opinion in Microbiology, 62, 28-37.

Popper, K. R. (1990). Pyrite and the origin of life. Nature, 344(6265), 387-387.

Pratscher, J., Dumont, M. G., & Conrad, R. (2011). Ammonia oxidation coupled to CO2 fixation by archaea and bacteria in an agricultural soil. Proceedings of the National Academy of Sciences, 108(10), 4170-4175.

Price GD, Howitt SM, Harrison K, Badger MR. (1993). Analysis of a genomic DNA region from the cyanobacterium Synechococcus sp. Strain PCC 7942 involved in carboxysome assembly and function. J. Bacteriol. 175:2871-2879.

Price, G. D. (2011). Inorganic carbon transporters of the cyanobacterial CO 2 concentrating mechanism. Photosynthesis research, 109, 47-57.

Price, G. D., & Badger, M. (1989). Expression of human carbonic anhydrase in the cyanobacterium Synechococcus PCC7942 creates a high CO2-requiring phenotype: evidence for a central role for carboxysomes in the CO2 concentrating mechanism. Plant physiology, 91(2), 505-513.

Price, G. D., Badger, M. R., Woodger, F. J., & Long, B. M. (2008). Advances in understanding the cyanobacterial CO2-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. Journal of experimental botany, 59(7), 1441-1461.

Price, G. D., Coleman, J. R., & Badger, M. R. (1992). Association of carbonic anhydrase activity with carboxysomes isolated from the cyanobacterium Synechococcus PCC7942. Plant Physiology, 100(2), 784-793.

Priyaragini, S., Sathishkumar, S. R., & Bhaskararao, K. V. (2013). Biosynthesis of silver nanoparticles using Actinomycetota and evaluating its antimicrobial and cytotoxicity activity. Int J Pharm Pharm Sci, 5(2), 709-712.

Pu, X., & Han, Y. (2022). Promotion of carbon dioxide biofixation through metabolic and enzyme engineering. Catalysts, 12(4), 399.

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 CO2

fixation in cyanobacteria and some proteobacteria. Microbiology and Molecular Biology Reviews, 77(3), 357-379.
Richards, C. L., Bossdorf, O., & Pigliucci, M. (2010). What role does heritable epigenetic variation play in phenotypic evolution? BioScience, 60(3), 232-237.

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(4), 787-795.

Rodríguez, A., Castrejón-Godínez, M. L., Salazar-Bustamante, E., Gama-Martínez, Y., Sánchez-Salinas, E., Mussali-Galante, P., ... & Ortiz-Hernández, M. L. (2020). Omics approaches to pesticide biodegradation. Current Microbiology, 77, 545-563.

Romero-Rodríguez, A., Maldonado-Carmona, N., Ruiz-Villafán, B., Koirala, N., Rocha, D., & Sánchez, S. (2018). Interplay between carbon, nitrogen and phosphate utilization in the control of secondary metabolite production in Streptomyces. Antonie Van Leeuwenhoek, 111, 761-781.

Rubin-Blum, M., Dubilier, N., & Kleiner, M. (2019). Genetic evidence for two carbon fixation pathways (the Calvin-Benson-Bassham cycle and the reverse tricarboxylic acid cycle) in symbiotic and free-living bacteria. Msphere, 4(1), e00394-18.

Ruiz-Fernández, P., Ramírez-Flandes, S., Rodríguez-León, E., & Ulloa, O. (2020). Autotrophic carbon fixation pathways along the redox gradient in oxygen-depleted oceanic waters. Environmental microbiology reports, 12(3), 334-341.

Sage, R. F. (2002). Variation in the k cat of Rubisco in C3 and C4 plants and some implications for photosynthetic performance at high and low temperature. Journal of Experimental Botany, 53(369), 609-620.

Saini, R., Kapoor, R., Kumar, R., Siddiqi, T. O., & Kumar, A. (2011). CO2 utilizing microbes-a comprehensive review. Biotechnology advances, 29(6), 949-960.

Samborska, B., & Kimber, M. S. (2012). A dodecameric CcmK2 structure suggests β-

carboxysomal shell facets have a double-layered organization. Structure, 20(8), 1353-1362.

Sánchez-Andrea, I., Guedes, I. A., Hornung, B., Boeren, S., Lawson, C. E., Sousa, D. Z., ... & Stams, A. J. (2020). The reductive glycine pathway allows autotrophic growth of Desulfovibrio desulfuricans. Nature Communications, 11(1), 5090.

Sanderson, M. J., Smith, I., Parker, I., & Bootman, M. D. (2014). Fluorescence microscopy. Cold Spring Harbor Protocols, 2014(10), pdb-top071795.

Sato, T., Atomi, H., & Imanaka, T. (2007). Archaeal type III RuBisCOs function in a pathway for AMP metabolism. Science, 315(5814), 1003-1006.

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(11), 7546-7555.

Scarff, C. A., Fuller, M. J., Thompson, R. F., & Iadanza, M. G. (2018). Variations on negative stain electron microscopy methods: Tools for tackling challenging systems. JoVE (Journal of Visualized Experiments), (132), e57199.

Scheres, S. H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determinuteation. Journal of structural biology, 180(3), 519-530.

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(3), 526-535.

Schneider, G., Lindqvist, Y., Brändén, C. I., & Lorimer, G. (1986). Three-dimensional structure of ribulose-1, 5-bisphosphate carboxylase/oxygenase from Rhodospirillum rubrum at 2.9 Å resolution. The EMBO journal, 5(13), 3409-3415.

Scott, K. M., Leonard, J. M., Boden, R., Chaput, D., Dennison, C., Haller, E., ... & Whittaker, R. (2019). Diversity in CO2-concentrating mechanisms among chemolithoautotrophs from the

genera Hydrogenovibrio, Thiomicrorhabdus, and Thiomicrospira, ubiquitous in sulfidic habitats worldwide. Applied and Environmental Microbiology, 85(3), e02096-18.
Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics, 30(14), 2068-2069.

Sengupta, S., & Azad, R. K. (2021). A Protocol for Phylogenetic Reconstruction. In Plant Metabolic Engineering: Methods and Protocols (pp. 35-46). New York, NY: Springer US.

Sharkey, T. D. (2019). Discovery of the canonical Calvin–Benson cycle. Photosynthesis Research, 140(2), 235-252.

Sharkey, T. D., & Weise, S. E. (2016). The glucose 6-phosphate shunt around the Calvin–Benson cycle. Journal of experimental botany, 67(14), 4067-4077.

Shekhar, S. K., Godheja, J., Modi, D. R., & Peter, J. K. (2014). Growth potential assessment of Actinomycetes isolated from petroleum contaminuteated soil. J Bioremed Biodeg, 5(259), 2.

Shibata, M., Ohkawa, H., Kaneko, T., Fukuzawa, H., Tabata, S., Kaplan, A., & Ogawa, T. (2001). Distinct constitutive and low-CO2-induced CO2 uptake systems in cyanobacteria: genes involved and their phylogenetic relationship with homologous genes in other organisms. Proceedings of the National Academy of Sciences, 98(20), 11789-11794.

Shih, P. M., Occhialini, A., Cameron, J. C., Andralojc, P. J., Parry, M. A., & Kerfeld, C. A. (2016). Biochemical characterization of predicted Precambrian RuBisCO. Nature Communications, 7(1), 1-11.

Shively, J. M. (1974). Inclusion bodies of prokaryotes. Annual review of microbiology, 28(1), 167-188.

Shively, J. M., Ball, F., Brown, D. H., & Saunders, R. E. (1973). Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of Thiobacillus neapolitanus. Science, 182(4112), 584-586.

Shivlata, L., & Satyanarayana, T. (2015). Thermophilic and alkaliphilic Actinomycetota:

biology and potential applications. Frontiers in microbiology, 6, 1014.

Siddique, K., Ager-Wick, E., Fontaine, R., Weltzien, F. A., & Henkel, C. V. (2021). Characterization of hormone-producing cell types in the teleost pituitary gland using single-cell RNA-seq. Scientific Data, 8(1), 279.

Siegel, J. B., Smith, A. L., Poust, S., Wargacki, A. J., Bar-Even, A., Louw, C., ... & Baker, D. (2015). Computational protein design enables a novel one-carbon assimilation pathway. Proceedings of the National Academy of Sciences, 112(12), 3704-3709.

Snider, J., & Houry, W. A. (2008). AAA+ proteins: diversity in function, similarity in structure. Biochemical Society Transactions, 36(1), 72-77.

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(3), 623-630.

Sommer, M., Sutter, M., Gupta, S., Kirst, H., Turmo, A., Lechno-Yossef, S., ... & Kerfeld, C. A. (2019). Heterohexamers formed by CcmK3 and CcmK4 increase the complexity of beta carboxysome shells. Plant physiology, 179(1), 156-167.

Sokol, N. W., Slessarev, E., Marschmann, G. L., Nicolas, A., Blazewicz, S. J., Brodie, E. L., ... & Pett-Ridge, J. (2022). Life and death in the soil microbiome: how ecological processes influence biogeochemistry. Nature Reviews Microbiology, 20(7), 415-430.

Steffens, L., Pettinato, E., Steiner, T. M., Eisenreich, W., & Berg, I. A. (2022). Tracking the Reversed Oxidative Tricarboxylic Acid Cycle in Bacteria. Bio-protocol, 12(6), e4364-e4364.

Steven, A. C., & Aebi, U. (2003). The next ice age: cryo-electron tomography of intact cells. Trends in cell biology, 13(3), 107-110.

Stevenson, B. S., Eichorst, S. A., Wertz, J. T., Schmidt, T. M., & Breznak, J. A. (2004). New strategies for cultivation and detection of previously uncultured microbes. Applied and environmental microbiology, 70(8), 4748-4755.

Su, P. T., Liao, C. T., Roan, J. R., Wang, S. H., Chiou, A., & Syu, W. J. (2012). Bacterial colony from two-dimensional division to three-dimensional development. PloS one, 7(11), e48098.

Sun, H., Cui, N., Han, S. J., Chen, Z. P., Xia, L. Y., Chen, Y., ... & Zhou, C. Z. (2021). Complex structure reveals CcmM and CcmN form a heterotrimeric adaptor in β-carboxysome. Protein Science, 30(8), 1566-1576.

Sun, Y., Harman, V. M., Johnson, J. R., Brownridge, P. J., Chen, T., Dykes, G. F., ... & Liu, L. N. (2022). Decoding the absolute stoichiometric composition and structural plasticity of α-carboxysomes. Mbio, 13(2), e03629-21.

Sun, Y., Wollman, A. J., Huang, F., Leake, M. C., & Liu, L. N. (2019). Single-organelle quantification reveals stoichiometric and structural variability of carboxysomes dependent on the environment. The Plant Cell, 31(7), 1648-1664.

Sutter, M., & Kerfeld, C. A. (2022). BMC Caller: a webtool to identify and analyze bacterial microcompartment types in sequence data. Biology direct, 17(1), 1-6.

Sutter, M., Greber, B., Aussignargues, C., & Kerfeld, C. A. (2017). Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. Science, 356(6344), 1293-1297.

Sutter, M., Laughlin, T. G., Sloan, N. B., Serwas, D., Davies, K. M., & Kerfeld, C. A. (2019). Structure of a synthetic β-carboxysome shell. Plant physiology, 181(3), 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. Nature communications, 12(1), 3809.

Sutter, M., Roberts, E. W., Gonzalez, R. C., Bates, C., Dawoud, S., Landry, K., ... & Kerfeld, C. A. (2015). Structural characterization of a newly identified component of α-carboxysomes: the AAA+ domain protein CsoCbbQ. Scientific Reports, 5(1), 1-14.

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

Tabita, F. R., Hanson, T. E., Li, H., Satagopan, S., Singh, J., & Chan, S. (2007). Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs. Microbiology and Molecular Biology Reviews, 71(4), 576-599.

Tabita, F. R., Satagopan, S., Hanson, T. E., Kreel, N. E., & Scott, S. S. (2008). Distinct form I, II, III, and IV RuBisCO proteins from the three kingdoms of life provide clues about RuBisCO evolution and structure/function relationships. J. of Expe. Bot., 59(7), 1515-1524.

Tan, Y. Q., Ali, S., Xue, B., Teo, W. Z., Ling, L. H., Go, M. K., ... & Yew, W. S. (2021). Structure of a minuteimal α-carboxysome-derived shell and its utility in enzyme stabilization. Biomacromolecules, 22(10), 4095-4109.

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

Tard, C., & Pickett, C. J. (2009). Structural and functional analogues of the active sites of the [Fe]-[NiFe]-, and [FeFe]-hydrogenases. Chemical reviews, 109(6), 2245-2274.

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(19), 7246-7251.

Thevasundaram, K., Gallagher, J. J., Cherng, F., & Chang, M. C. (2022). Engineering nonphotosynthetic carbon fixation for production of bioplastics by methanogenic archaea. Proceedings of the National Academy of Sciences, 119(23), e2118638119.

Thiel, V., Hügler, M., Ward, D. M., & Bryant, D. A. (2017). The dark side of the mushroom spring microbial mat: life in the shadow of chlorophototrophs. II. Metabolic functions of abundant community members predicted from metagenomic analyses. Frontiers in microbiology, 8, 943.

Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2003). Multiple sequence alignment using ClustalW and ClustalX. Current protocols in bioinformatics, (1), 2-3.

Ting, C. S., Hsieh, C., Sundararaman, S., Mannella, C., & Marko, M. (2007). Cryo-electron tomography reveals the comparative three-dimensional architecture of Prochlorococcus, a globally important marine cyanobacterium. Journal of bacteriology, 189(12), 4485-4493.

Turk, M., & Baumeister, W. (2020). The promise and the challenges of cryo-electron tomography. FEBS letters, 594(20), 3243-3261.

Turmo, A., Gonzalez-Esquer, C. R., & Kerfeld, C. A. (2017). Carboxysomes: metabolic modules for CO2 fixation. FEMS microbiology letters, 364(18).

Turpin, D. H., Miller, A. G., & Canvin, D. T. (1984). Carboxysome content of Synechococcus leopoliensis (cyanophyta) in response to inorganic carbon 1. Journal of phycology, 20(2), 249-253.

Van Bergeijk, D. A., Terlouw, B. R., Medema, M. H., & van Wezel, G. P. (2020). Ecology and genomics of Actinomycetota: new concepts for natural product discovery. Nature Reviews Microbiology, 18(10), 546-558.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., ... & Kalush, F. (2001). The sequence of the human genome. science, 291(5507), 1304-1351.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G. F., Chater, K. F., & Van Sinderen, D. (2007). Genomics of Actinomycetota: tracing the evolutionary history of an ancient phylum. Microbiology and molecular biology reviews, 71(3), 495-548.

Volokita, M., Zenvirth, D., Kaplan, A., & Reinhold, L. (1984). Nature of the inorganic carbon species actively taken up by the cyanobacterium Anabaena variabilis. Plant Physiology, 76(3), 599-602.

Volpiano, C. G., Sant'Anna, F. H., da Mota, F. F., Sangal, V., Sutcliffe, I., Munusamy, M., ... & Rosado, A. S. (2021). Proposal of Carbonactinosporaceae fam. nov. within the class Actinomycetia. Reclassification of Streptomyces thermoautotrophicus as Carbonactinospora thermoautotrophica gen. nov., comb. nov. Systematic and applied microbiology, 44(4), 126223.

Wächtershäuser, G. (1990). Evolution of the first metabolic cycles. Proceedings of the National Academy of Sciences, 87(1), 200-204.

Wang, H., & Hayer-Hartl, M. (2022a). Phase Separation of Rubisco by the Folded SSUL Domains of CcmM in Beta-Carboxysome Biogenesis. In Phase-Separated Biomolecular Condensates: Methods and Protocols (pp. 269-296). New York, NY: Springer US.

Wang, H., Yan, X., Aigner, H., Bracher, A., Nguyen, N. D., Hee, W. Y., ... & Hayer-Hartl, M. (2019). Rubisco condensate formation by CcmM in β-carboxysome biogenesis. Nature, 566(7742), 131-135.

Wang, L., Liu, J., Yuan, J., & Wang, N. (2022b). The Study of Soil Bacterial Diversity and the Influence of Soil Physicochemical Factors in Meltwater Region of Ny-Ålesund, Arctic. Microorganisms, 10(10), 1913.

Wang, P., Ma, J., Wang, X., & Tan, Q. (2020). Rising atmospheric CO2 levels result in an earlier cyanobacterial bloom-maintenance phase with higher algal biomass. Water Research, 185, 116267.

Wang, X., Li, W., Xiao, Y., Cheng, A., Shen, T., Zhu, M., & Yu, L. (2021). Abundance and diversity of carbon-fixing bacterial communities in karst wetland soil ecosystems. Catena, 204, 105418.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews genetics, 10(1), 57-63.

Watson, G. M., & Tabita, F. R. (1996). Regulation, unique gene organization, and unusual primary structure of carbon fixation genes from a marine phycoerythrin-containing cyanobacterium. Plant molecular biology, 32, 1103-1115.

Wattam, A. R., Abraham, D., Dalay, O., Disz, T. L., Driscoll, T., Gabbard, J. L., ... & Sobral, B. W. (2014). PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic acids research, 42(D1), D581-D591.

Wheatley, N. M., Sundberg, C. D., Gidaniyan, S. D., Cascio, D., & Yeates, T. O. (2014). Structure and identification of a pterin dehydratase-like protein as a ribulose-bisphosphate carboxylase/oxygenase (RuBisCO) assembly factor in the α-carboxysome. Journal of Biological Chemistry, 289(11), 7973-7981.

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(1), 398-411.

Whitney, S. M., Houtz, R. L., & Alonso, H. (2011). Advancing our understanding and capacity to engineer nature's CO2-sequestering enzyme, Rubisco. Plant physiology, 155(1), 27-35.

Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M., ... & Kuster, B. (2014). Mass-spectrometry-based draft of the human proteome. Nature, 509(7502), 582-587.

Wick, R. R., Judd, L. M., & Holt, K. E. (2023). Assembling the perfect bacterial genome using Oxford Nanopore and Illumina sequencing. PLOS Computational Biology, 19(3), e1010905.

Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS computational biology, 13(6), e1005595.

Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome biology, 20, 1-10.

Winkel, M., Salman-Carvalho, V., Woyke, T., Richter, M., Schulz-Vogt, H. N., Flood, B. E., ... & Mußmann, M. (2016). Single cell sequencing of Thiomargarita reveals genomic flexibility for adaptation to dynamic redox conditions. Frontiers in microbiology, 7, 964.

Wiśniewski, J. R., Zougman, A., Nagaraj, N., & Mann, M. (2009). Universal sample preparation method for proteome analysis. Nature methods, 6(5), 359-362.

Xiao, L., Liu, G., Gong, F., Zhu, H., Zhang, Y., Cai, Z., & Li, Y. (2021). A minuteimized

synthetic carbon fixation cycle. ACS Catalysis, 12(1), 799-808.

Xu, S., Qiao, W., Wang, Z., Fu, X., Liu, Z., & Shi, S. (2023). Exploiting a heterologous construction of the 3-hydroxypropionic acid carbon fixation pathway with mesaconate as an indicator in Saccharomyces cerevisiae. Bioresources and Bioprocessing, 10(1), 1-12.

Yadav, A. N., Verma, P., Kumar, S., Kumar, V., Kumar, M., Sugitha, T. C. K., ... & Dhaliwal, H. S. (2018). Actinomycetota from rhizosphere: molecular diversity, distributions, and potential biotechnological applications. In New and future developments in microbial biotechnology and bioengineering (pp. 13-41). Elsevier.

Yamamoto, M., Arai, H., Ishii, M., & Igarashi, Y. (2006). Role of two 2-oxoglutarate: ferredoxin oxidoreductases in Hydrogenobacter thermophilus under aerobic and anaerobic conditions. FEMS microbiology letters, 263(2), 189-193.

Yan, S., Bhawal, R., Yin, Z., Thannhauser, T. W., & Zhang, S. (2022). Recent advances in proteomics and metabolomics in plants. Molecular Horticulture, 2(1), 17.

Yang, Y., Sun, J., Sun, Y., Kwan, Y. H., Wong, W. C., Zhang, Y., ... & Qian, P. Y. (2020). Genomic, transcriptomic, and proteomic insights into the symbiosis of deep-sea tubeworm holobionts. The ISME journal, 14(1), 135-150.

Yasuda, G., Shionoiri, H., Hamada, K., Umemura, S., Hiroto, S., Takasaki, I., & Kaneko, Y. (1985). Bilateral pheochromocytoma associated with papillary adenocarcinoma of the thyroid gland, report of an unusual case. Endocrinologia japonica, 32(3), 399-404.

Yeates, T. O., Kerfeld, C. A., Heinhorst, S., Cannon, G. C., & Shively, J. M. (2008). Protein-based organelles in bacteria: carboxysomes and related microcompartments. Nature Reviews Microbiology, 6(9), 681-691.

Yoshizawa, Y., Toyoda, K., Arai, H., Ishii, M., & Igarashi, Y. (2004). CO2-responsive expression and gene organization of three ribulose-1, 5-bisphosphate carboxylase/oxygenase enzymes and carboxysomes in Hydrogenovibrio marinus strain MH-110. Journal of bacteriology, 186(17), 5685-5691.

Yusuf, M., Farooq, S., Robinson, I., & Lalani, E. N. (2020). Cryo-nanoscale chromosome imaging-future prospects. Biophysical Reviews, 12, 1257-1263.

Zang, K., Wang, H., Hartl, F. U., & Hayer-Hartl, M. (2021). Scaffolding protein CcmM directs multiprotein phase separation in β-carboxysome biogenesis. Nature Structural & Molecular Biology, 28(11), 909-922.

Zarzycki, J., Brecht, V., Müller, M., & Fuchs, G. (2009). Identifying the missing steps of the autotrophic 3-hydroxypropionate CO2 fixation cycle in Chloroflexus aurantiacus. Proceedings of the National Academy of Sciences, 106(50), 21317-21322.

Zarzycki, J., Erbilgin, O., & Kerfeld, C. A. (2015). Bioinformatic characterization of glycyl radical enzyme-associated bacterial microcompartments. Applied and environmental microbiology, 81(24), 8315-8329.

Zeng, R., Lv, C., Zang, J., Zhang, T., & Zhao, G. (2022). Designing stacked assembly of type III rubisco for CO2 fixation with higher efficiency. Journal of Agricultural and Food Chemistry, 70(23), 7049-7057.

Zhang, C., Liu, G., Xue, S., & Wang, G. (2016). Soil bacterial community dynamics reflect changes in plant community and soil properties during the secondary succession of abandoned farmland in the Loess Plateau. Soil Biology and Biochemistry, 97, 40-49.

Zhang, L., Morello, G., Carr, S. B., & Armstrong, F. A. (2020). Aerobic photocatalytic H2 production by a [NiFe] hydrogenase engineered to place a silver nanocluster in the electron relay. Journal of the American Chemical Society, 142(29), 12699-12707.

Zhao, Y. Y., Jiang, Y. L., Chen, Y., Zhou, C. Z., & Li, Q. (2019). Crystal structure of pentameric shell protein CsoS4B of Halothiobacillus neapolitanus α-carboxysome. Biochemical and biophysical research communications, 515(3), 510-515.

Zhou, J., Xue, K., Xie, J., Deng, Y. E., Wu, L., Cheng, X., ... & Luo, Y. (2012). Microbial mediation of carbon-cycle feedback to climate warming. Nature Climate Change, 2(2), 106-110.

Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler. Bioinformatics, 29(21), 2669-2677.

## 10. APÊNDICES

Apêndice A: Meio mineral N-FIX e solução de elementos traços. Descrição e quantidade dos componentes. Fonte: Adaptado de Gadkari et al. (1990).

| Meio de cultura N-FIX | Composição do meio |
|---|---|
| Solução A<br>Preparar 798 mL | $MgSO_4$. $7H_2O$ 0,3 g/L<br>$NaCl_2$ 0,2 g/L<br>$CaCl_2$. $2H_2O$ 0,1 g/L<br>$Na_2MoO_4$. $2H_2O$ 12,6 mg/L<br>Ágar nobre 30/L |
| Solução B<br>Preparar 200 mL | $K_2HPO_4$ 0,9 g/L<br>$KH_2PO_4$ 0,1 g/L |
| As duas soluções foram preparadas separadamente e autoclavadas, e misturadas em fluxo laminar com adição de 2 mL de solução de elementos traços, pH 7.8. Composição da solução de elementos traços está descrita abaixo. | |
| Solução traços | $ZnSO_4$. $7H_2O$ 0,1 g<br>$MnCl_2$. $4H_2O$ 0,03 g<br>$H_3BO_3$ 0,3 g<br>$CoCl_2$ 0,2 g<br>$CuCl_2$ 0,01 g<br>$NiCl_2$ 0,02 g<br>$Na_2MoO_4$ 0,9 g<br>$Na_2SeO_3$ 0,02 g |

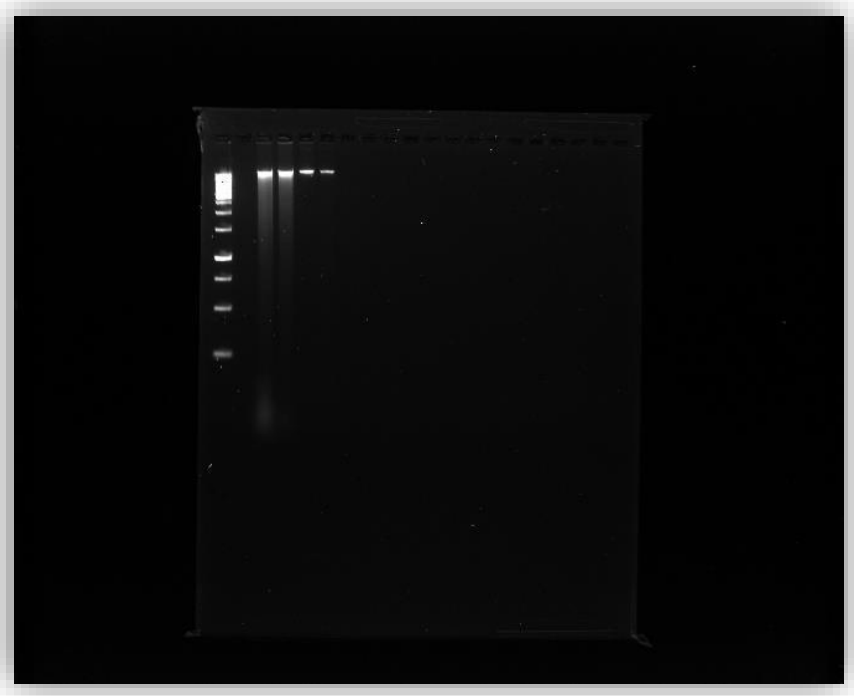
Apêndice B: Quantificação do DNA total da estirpe StC por Nanodrop antes da purificação

| Amostra | ng/ul | 260/280 | 230/260 |
|---|---|---|---|
| A | 430.8 | 2.11 | 2.23 |
| B | 164.7 | 2.08 | 2.39 |
| C | 43.3 | 2.04 | 2.23 |

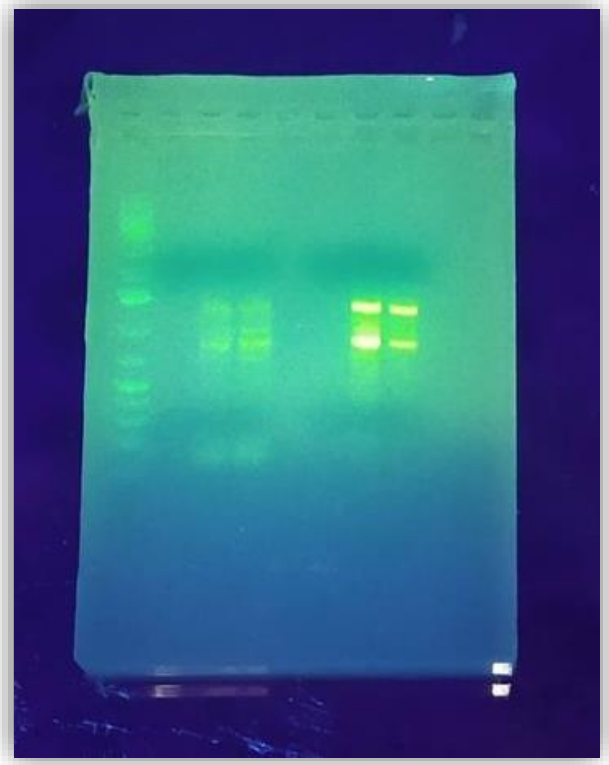
Apêndice C: Quantificação do DNA total da estirpe StC por Qubit® dsDNA HS Assay Kit (Thermo Fisher Scientific) depois da purificação com AMPure XP (Beckman Coulter Life Sciences).

| Amostra | ng/ul |
|---|---|
| A | 64.4 |
| B | 16.6 |
| C | 3.16 |

Apêndice D: Eletroforese em gel de agarose a 1,5% do DNA total da estirpe StC. Foram utilizados para o sequenciamento o DNA presente nas três últimas bandas do gel.



Apêndice E: Eletroforese em gel de agarose a 1,5% livre de RNAse do RNA total da estirpe StC após a purificação com Ambion's DNAse Treatment Kit (Thermo Fisher Scientific).



Apêndice F: Quantificação do RNA total da estirpe StC por Nanodrop após a purificação com Ambion's DNAse Treatment Kit (Thermo Fisher Scientific), Quatro réplicas para cada tratamento.

| Tratamento | ng/ul | 230/260 | 260/280 |
|---|---|---|---|
| N-FIX1 | 51.4 | 0.18 | 1.98 |
| N-FIX2 | 22.9 | 0.08 | 1.84 |
| N-FIX3 | 56.7 | 0.91 | 1.83 |
| N-FIX4 | 33.0 | 0.08 | 1.88 |
| R2A1 | 154.9 | 1.27 | 2.05 |
| R2A2 | 179.8 | 0.95 | 2.06 |
| R2A3 | 132.2 | 1.22 | 2.07 |
| R2A4 | 230.8 | 0.68 | 2.10 |

Apêndice G: Quantificação do RNA total da estirpe StC por Qubit® dsDNA HS Assay Kit (Thermo Fisher Scientific) após a purificação com Ambion's Dnase Treatment Kit (Thermo Fisher Scientific), Quatro réplicas para cada tratamento.

| Tratamento | µg/ul |
|---|---|
| N-FIX1 | 50,7 |
| N-FIX2 | 9,20 |
| N-FIX3 | 12.5 |
| N-FIX4 | 11.5 |
| R2A1 | 79.4 |
| R2A2 | 120 |
| R2A3 | 140 |
| R2A4 | 160 |

Apêndice H: Quantificação em Nanodrop da amostra de purificação de carboxissomo extraído da estirpe StC

| | ng/ul | 230/260 | 260/280 |
|---|---|---|---|
| Carboxissomo | 1031.6 | 0.79 | 0.77 |

Apêndice I: Quantificação na amostra contento carboxissomo com o kit Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific), Os 16 primeiros poços são os padrões, os 2 poços circulados em vermelho são as amostras em duplicata contento BMC. Foi realizado uma curva de calibração para prever o valor da concentração de proteína a partir da resposta analítica medida, e 0,69 mg/mL foi a concentração de proteínas na amostra.